การประมวลผลข้อมูลขนาดใหญ่โดยใช้ภาษา R และแพ็กเกจ Dplyr

BIG DATA COMPUTING WITH R PROJECT AND DPLYR

ธัญชนก ตรีอภิบาลวงศา

พรวลัย ภูชมศรี

ศิริวรรณ พัฒสังวาล

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การประมวลผลข้อมูลขนาดใหญ่โดยใช้ภาษา R และแพ็กเกจ Dplyr

BIG DATA COMPUTING WITH R PROJECT AND DPLYR

ผู้จัดทำ

1. นางสาวธัญชนก ตรีอภิบาลวงศา รหัสนักศึกษา 60010451
2. นางสาวพรวลัย ภูชมศรี รหัสนักศึกษา 60010665
3. นางสาวศิริวรรณ พัฒสังวาล รหัสนักศึกษา 60010981

อาจารย์ที่ปรึกษา

(ผศ.ดร.สุรินทร์ กิตติธรกุล)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

การประมวลผลข้อมูลขนาดใหญ่โดยใช้ภาษา R และแพ็กเกจ Dplyr

ธัญชนก	ตรีอภิบาลวงศา	รหัสนักศึกษา 60010451
พรวลัย	ภูชมศรี	รหัสนักศึกษา 60010665
ศิริวรรณ	พัฒสังวาล	รหัสนักศึกษา 60010981
ผศ.ดร.สุรินทร์	กิตติธรกุล	อาจารย์ที่ปรึกษา

ปีการศึกษา 2563

บทคัดย่อ

การใช้งานและพัฒนาโปรแกรมสำหรับการประมวลผลข้อมูลขนาดใหญ่ด้วยภาษา R สามารถช่วยวิเคราะห์ข้อมูลขนาดใหญ่ในแง่มุมต่าง ๆ โดยแบ่งการวิเคราะห์เบื้องต้นเป็น 2 วิธี คือ การวิเคราะห์ข้อมูลโดยใช้แพ็กเกจ Dplyr และโดยใช้แพ็กเกจ Rodbc เพื่อทำการเปรียบเทียบประสิทธิภาพการทำงานของแพ็กเกจ Dplyr และแพ็กเกจ Rodbc ด้วยการทำ Profiling ในการวัดเวลาและหน่วยความจำที่ใช้ในการทำงานของฟังก์ชันในการทำ Data cleaning และอื่น ๆ จากนั้นโครงงานจึงเลือกใช้แพ็กเกจ Dplyr เนื่องจากมีประสิทธิภาพที่ดีกว่า เมื่อวิเคราะห์ข้อมูลด้วย Dplyr แล้วเสร็จจึงแสดงผลโดยใช้รูปแบบ Visualization ที่เหมาะสมและเข้าใจง่าย

ปริญญานิพนธ์ฉบับนี้เป็นการวิเคราะห์ข้อมูลเกี่ยวกับโรคติดเชื้อไวรัสโคโรนา 2019 ประกอบด้วยข้อมูล จำนวนผู้ติดเชื้อ จำนวนผู้เสียชีวิต จำนวนผู้รักษาหาย โดยแสดงข้อมูลตามแต่ละประเทศทั่วโลก เพื่อหาเลือก 20 อันดับประเทศที่ติดเชื้อมากที่สุดในโลก และวิเคราะห์ว่าผลิตภัณฑ์มวลรวมในประเทศ อุณหภูมิและอันดับระบบการจัดการด้านสาธารณสุขมีผลต่อจำนวนการติดเชื้อ 20 อันดับประเทศที่ติดเชื้อมากที่สุดอย่างไร โดยแสดงผลการวิเคราะห์ในรูปแบบของกราฟชนิดต่าง ๆ เช่น กราฟเส้น Heatmap Correlation และแผนที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Big Data Computing with R Project and Dplyr

Thanchanok	Triapibanwongsa	60010451
Phonwalai	Phuchomsri	60010665
Siriwan	Phatsangwan	60010981
Asst.Prof.Dr.Surin	Kittitornkun	Advisor

Academic Year 2020

ABSTRACT

The use and development of R scripts can be applied to analyze various aspects of Big data processing. Preliminary data analysis can be divided into two approaches, namely: using the Dplyr vs. using Rodbc. Performance comparison can be carried out by profiling in terms of run time and working memory space of Data cleaning and other steps. Eventually, Dplyr package is selected due to its performance. Upon completion with Dplyr, results can be visualized in various formats, making it easy to understand.

This project is mainly based on 2019 Coronavirus data from John Hopkins University. The number of global infected, dead, recovered cases and several others are included. Top 20 most infected countries are chosen as our case studies. Gross domestic product, average temperature and public health system rankings are incorporated for deeper analyses. The results can be visualized in line graphs, heatmap, correlation and maps.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

สารบัญ.....III

สารบัญตาราง.....V

สารบัญภาพ.....VI

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 วัตถุประสงค์.....2

1.3 ประโยชน์ที่คาดว่าจะได้รับ.....2

1.4 ข้อจำกัดของโครงงาน.....2

1.5 ขอบเขตของโครงงาน.....2

1.6 แผนการดำเนินงานของโครงงาน.....3

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....7

2.1 ข้อมูลที่ใช้ในการวิเคราะห์.....7

2.2 ภาษา R และเทคโนโลยีที่เกี่ยวข้อง.....16

2.3 เอกสารและงานวิจัยที่เกี่ยวข้อง.....40

บทที่ 3 การออกแบบและการพัฒนา.....42

3.1 ภาพรวมของระบบ.....42

3.2 การตั้งสมมติฐาน.....42

3.3 การประมวลผลข้อมูล (DataProcessibg).....43

3.4 การวิเคราะห์ข้อมูล.....45

3.5 การสร้างแบบจำลองทางคณิตศาสตร์.....54

3.6 แพ็กเกจที่ใช้ใน R 60

3.7 การเชื่อมต่อฐานข้อมูล.....63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 4 การออกแบบและการพัฒนา.....64
4.1 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ทั่วโลก.....64
4.2 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา.....71
4.3 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย.....80
4.4 ผลลัพธ์ที่ได้จาก SEIRD model of covid-19 in United States.....86
4.5 ผลลัพธ์ที่ได้จาก SEIR model of covid-19 in Thailand.....90
4.6 การวัดประสิทธิภาพโดยใช้ Profiling.....94

บทที่ 5 บทสรุปและข้อเสนอแนะ.....100
5.1 บทสรุป.....101
5.2 ปัญหาและอุปสรรค.....104
5.3 แนวทางการพัฒนาต่อ.....105

บรรณานุกรม.....106
ภาคผนวก ก.....109
ภาคผนวก ข.....126

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตาราง **หน้า**

1.1 แผนการดำเนินงานระยะเวลา 2 เทอม....3
2.1 การพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์....36
3.1 ข้อมูลติดเชื้อทั่วโลก....48
3.2 แพ็กเกจที่ใช้ร่วมกัน....61
3.3 แพ็กเกจเพิ่มเติมแบบไม่ใช้แพ็กเกจ Dplyr....61
3.4 แพ็กเกจเพิ่มเติมแบบใช้แพ็กเกจ Dplyr....62
4.1 ผลการทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr....95
4.2 รายละเอียด Environment ที่ใช้ในการเปรียบเทียบเพื่อวัดประสิทธิภาพ....96
4.3 ผลการทดสอบด้านเวลาที่ใช้ในการประมวลผลตั้งแต่การเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด....98

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูป **หน้า**

2.1 เปรียบเทียบ GDP ของ IMF ในปี 2019....8
2.2 เปรียบเทียบ GDP ของ IMF ในแต่ละปี....8
2.3 คะแนนอันดับระบบสาธารณสุขทั่วโลกปี 2019....11
2.4 20 อันดับประเทศที่เข้าชมเว็ปไซต์ Pornhun 2019....12
2.5 รูปทางเดินหายใจเฉียบพลันโรคซาร์ส....13
2.6 รายงาน Covid-19 ประจำวันประเทศไทย....14
2.7 รายงาน Covid-19 ในสหรัฐอเมริกา....14
2.8 เครื่องหมายการค้าของ R programming....16
2.9 เครื่องหมายการค้าของ RStudio....17
2.10 การทำงานของ R Notebook....18
2.11 ODBC....23
2.12 ตัวอย่างข้อมูลแบบ Wide format....25
2.13 ตัวอย่างข้อมูลแบบ Long format....25
2.14 ข้อมูลที่แสดงในแถบ Flame Graph....31
2.15 ข้อมูลที่แสดงในแถบ Data....31
2.16 กราฟ Logistic Curve....34
2.17 กราฟ N ฟังก์ชันซิกมอยด์และ tan h....35
2.18 ปัจจัยที่กำหนดว่าโรคระบาด (R0) ถูกควบคุมด้วยค่าคงที่ β และ γ....38
3.1 ภาพรวมของระบบ....42
3.2 การประมวลผลข้อมูลของโลก....43
3.3 การประมวลผลข้อมูลของสหรัฐอเมริกา....44
3.4 การประมวลผลข้อมูลของประเทศไทย....44
3.5 ฟังก์ชันสำหรับทำความสะอาดข้อมูล....46
3.6 ข้อมูลหลังการทำ Data Cleaning....46
3.7 ข้อมูลผู้ติดเชื้อทั่วโลก....47
3.8 ข้อมูลหลังจากการเพิ่มคอมลัมน์ current.confirmed....48
3.9 หลังจากการวิเคราะห์ข้อมูลส่วนต่าง ๆ....51
3.10 SEIR Diagram....55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป (ต่อ)

รูป **หน้า**

3.11 SEIRD Diagram……...58
4.1 จำนวนผู้ติดเชื้อปัจจุบัน(สีแดง) จำนวนผู้รักษาหาย(สีเขียว) และจำนวนผู้เสียชีวิต(สีเทา)ทั่วโลกรูปแบบ stack และจำนวนผู้ติดทั้งหมด(สีม่วง) จำนวนผู้ติดเชื้อปัจจุบัน(สีแดง) จำนวนผู้รักษาหาย(สีเขียว)และจำนวนผู้เสียชีวิต(สีดำ)ทั่วโลกในรูปแบบ log scale.....65
4.2 (A) อัตราการติดเชื้อ (B)อัตราการรักษาหาย และ(C) อัตราการเสียชีวิตของประเทศที่มีผู้ติดเชื้อไวรัสโคโรนาสูงสุด 20 ประเทศแรกและประเทศอื่น ๆนอกเหนือจากประเทศที่ติดอันดับ…...…...66
4.3 จำนวนผู้ติดเชื้อปัจจุบัน(สีแดง) จำนวนผู้รักษาหาย(สีเขียว) และจำนวนผู้เสียชีวิต(สีเทา) ของ 20 ประเทศที่พบผู้ติดเชื้อไวรัสโคโรนา2019 สูงสุดทั่วโลก……..67
4.4 ค่า Correlation ของข้อมูลที่ใช้ในการวิเคราะห์……68
4.5 ความถี่ข้อมูลที่เกี่ยวข้องของประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับ……...69
4.6 ความถี่ข้อมูลที่เกี่ยวข้องของประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับและ Hierarchical clustering…...70
4.7 กราฟแสดงจำนวนผู้ติดเชื้อ (สีแดง) และจำนวนผู้เสียชีวิต (สีเทา) ในสหรัฐอเมริกา……71
4.8 กราฟเส้นจำนวนผู้ติดเชื้อรายวันในสหรัฐอเมริกา……...73
4.9 กราฟ Correlation แสดงความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับโรคติดเชื้อไวรัสโคโรนา 2019 ใน 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา……..74
4.10 กราฟแท่งแสดงจำนวนผู้ติดเชื้อสูงสุด 20 รัฐในสหรัฐอเมริกา และแสดงร้อยละของจำนวนผู้ติดเชื้อในแต่ละรัฐเมื่อเทียบกับจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา……...75
4.11 แผนภูมิแท่งแสดงอัตราการติดเชื้อและอัตราการเสียชีวิตแบ่งตามรัฐของ 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา……...76
4.12 กราฟ Heatmap แสดงความถี่ของข้อมูลที่เกี่ยวข้องกับโรคติดเชื้อไวรัสโคโรนา 2019 ของ 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา……78
4.13 ความถี่ข้อมูลที่เกี่ยวข้องของประเทศที่สหรัฐอเมริกาและ Hierarchical clustering……..79
4.14 จำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย……80
4.15 จำนวนกลุ่มความเสี่ยงจากการติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย……..81
4.16 อายุ เพศชาย(สีฟ้า)และเพศหญิง(สีส้ม) ของผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย……83
4.17 สัญชาติผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย……84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป (ต่อ)

รูป **หน้า**

4.18 จังหวัดที่รักษา(สีส้ม)และจังหวัดที่พบ(สีฟ้า) ของ10 อันดับจังหวัดที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทย.....85

4.19 สัดส่วนประชากรแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ยังไม่มีมาตรการป้องกัน covid-19 กรณีนี้กำหนดให้ $\beta = 0.18, \gamma = 0.07, \alpha = 0.19$ และ $\sigma = 0.02$.....87

4.20 สัดส่วนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่มาตรการป้องกัน กรณีนี้กำหนดให้ $\beta = 0.18, \gamma = 0.07, \alpha = 0.18, \sigma = 0.02$ และ $\rho = 0.8$.....88

4.21 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน กรณีนี้กำหนดให้มีมาตรการป้องกัน $\rho = 0.8$ และไม่มีมาตรการป้องกัน $\rho = 1$.....89

4.22 สัดส่วนประชากรแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ยังไม่มีมาตรการป้องกัน covid-19 กรณีนี้กำหนดให้ $\beta = 0.46, \gamma = 0.07$ และ $\alpha = 0.19$.....91

4.23 สัดส่วนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่มาตรการป้องกัน กรณีนี้กำหนดให้ $\beta = 0.46, \gamma = 0.07, \alpha = 0.19$ และ $\rho = 0.8$.....92

4.24 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน กรณีนี้กำหนดให้มีมาตรการป้องกัน $\rho = 0.8$ และไม่มีมาตรการป้องกัน $\rho = 1$.....93

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในยุคปัจจุบันที่โลกและธุรกิจกำลังถูกขับเคลื่อนด้วยข้อมูลจำนวนมหาศาลจนทำให้คำว่า “Big Data” เป็นที่รู้จักมากขึ้น หลายคนไม่ได้สนใจว่าจะนำข้อมูลเหล่านั้นไปใช้ประโยชน์ได้อย่างไร และวิธีการจัดการข้อมูลจำนวนมากนี้ทำอย่างไร ทำให้ข้อมูลที่มีไร้ซึ่งประโยชน์ และกลายเป็นขยะข้อมูลสูญเปล่า จึงทำให้มีนักพัฒนาต่าง ๆ ได้พัฒนาภาษาคอมพิวเตอร์ขึ้นมา เพื่อนำมาประมวลผลและจัดการกับสิ่งเรียกว่า “Big Data” ภาษา R เป็นหนึ่งในภาษาที่ได้รับความนิยมในการนำมาจัดการกับข้อมูลขนาดใหญ่เนื่องจากเป็นภาษาที่เข้าใจง่าย มีความสามารถในการวิเคราะห์ Data Frame อีกทั้งยังรองรับระบบปฏิบัติการหลายระบบ ปฏิบัติการได้ทั้ง Windows, Mac, OS หรือ Linux และเหตุผลสำคัญที่สุดคือสามารถดาวน์โหลด R และ RStudio มาใช้งานได้ฟรีเนื่องจากเป็น Open Source

เนื่องจากภาษา R มีความสามารถในการวิเคราะห์ข้อมูล แต่ยังมีข้อจำกัดหลายประการ ดังนั้นจึงจำเป็นจะต้องใช้แพ็กเกจอื่นร่วมด้วยซึ่ง Dplyr เป็นแพ็กเกจหนึ่งที่สามารถติดตั้งและใช้งานได้ง่าย มีแหล่งข้อมูลจำนวนมากให้ศึกษา และสามารถเชื่อมต่อกับฐานข้อมูลได้สะดวก ด้วยเหตุผลดังกล่าวนี้ คณะผู้จัดทำจึงเลือกแพ็กเกจนี้มาช่วยในการวิเคราะห์จัดการ Data Frame โดยเรียกใช้ความสามารถของแพ็กเกจ Dplyr และแพ็กเกจอื่น ๆ ร่วมกัน และเมื่อเปรียบเทียบกับ R ที่ไม่มีการเรียกใช้แพ็กเกจแล้ว สามารถวัดประสิทธิภาพการทำงานของแพ็กเกจได้ โดยทางคณะผู้จัดทำได้เลือกข้อมูลเกี่ยวกับการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 หรือ “COVID-19” มีข้อมูลที่เกี่ยวข้องกับไวรัสชนิดนี้เป็นจำนวนมาก ไม่ว่าจะเป็นจำนวนผู้ป่วยที่เพิ่มขึ้น จำนวนผู้เสียชีวิต จำนวนผู้ที่รักษาหายและข้อมูลที่น่าสนใจที่อาจเป็นปัจจัยในการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 นับได้ว่าเป็นข้อมูลขนาดใหญ่และมีแนวโน้มที่จะเปลี่ยนแปลงตลอดเวลาในอนาคต นอกจากนี้คณะผู้จัดทำได้ทำการ Visualization ข้อมูลที่วิเคราะห์ แสดงบน Dashboard โดยอาศัยแพ็กเกจ Shiny มาใช้งานร่วมด้วย

ทางผู้จัดทำจึงเล็งเห็นว่า หากสามารถดึงประสิทธิภาพแพ็กเกจออกมาใช้งานได้อย่างครบถ้วนจะส่งผลให้นำไปใช้งานง่ายยิ่งขึ้น ลดขั้นตอนการเรียกใช้งานจากหลาย ๆ แพ็กเกจ ซึ่งจะเป็นประโยชน์แก่ผู้ที่ต้องการใช้งานภาษา R เพื่อวิเคราะห์ข้อมูลที่มีความหลากหลายมากขึ้น และแสดงข้อมูลต่าง ๆ ให้กับคนทั่วไปสามารถเข้าถึงข้อมูลจำนวนมากนี้ออกมาในรูปแบบที่ทุกคนสามารถเข้าใจได้ง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.2 วัตถุประสงค์

1) เพื่อศึกษาวิธีการวิเคราะห์ข้อมูลโดยการใช้ภาษา R
2) เพื่อวิเคราะห์ข้อมูลเชิงลึกประเทศที่น่าสนใจและสร้างแบบจำลองทางคณิตศาสตร์คาดการณ์เกี่ยวกับการแพร่ระบาดโรคติดเชื้อไวรัสโคโรนา 2019 ในสหรัฐอเมริกาและสหราชอาณาจักรไทย
3) เพื่อศึกษาแนวทางการนำข้อมูลจากการวิเคราะห์มาทำการแสดงเป็น Visualization ในรูปแบบต่าง ๆ ตามความเหมาะสมของข้อมูล
4) เพื่อนำข้อมูลจากการวิเคราะห์มาแสดงผลในรูปแบบของ Interactive Dashboard และแสดงข้อมูลแบบ Realtime เกี่ยวกับการแพร่ระบาดของไวรัสโคโรนา 2019
5) เพื่อศึกษาความสามารถในการใช้แพ็กเกจ Dplyr โดยวัดประสิทธิภาพจากการทำ Profiling ภายใต้สภาพแวดล้อมที่ต่างกัน

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้รับความรู้ความเข้าใจเกี่ยวกับการใช้ภาษา R การใช้แพ็กเกจ Dplyr และการใช้แพ็กเกจอื่น ๆ ที่ช่วยเสริมประสิทธิภาพการทำงานของแพ็กเกจ Dplyr
2) สามารถนำข้อมูลขนาดใหญ่มาวิเคราะห์และแสดงผลในรูปแบบที่เข้าถึงได้ง่าย ได้แก่ การทำ Data Visualization และการแสดงผลบน Dashboard
3) ได้นำความรู้จากการเรียนภายในห้องเรียนมาประยุกต์ใช้งานเพื่อวิเคราะห์ข้อมูลจริง
4) สามารถเปรียบเทียบประสิทธิภาพการทำงานของแพ็กเกจ Dplyr โดยการทำ Profiling ได้และศึกษาการใช้งานเพิ่มเติม เพื่อเป็นแนวทางในการใช้แพ็กเกจ Dplyr ได้มีประสิทธิภาพมากยิ่งขึ้น

1.4 ข้อจำกัดของโครงงาน

1) ไม่สนับสนุนการเชื่อมต่อกับ Hadoop การประมวลผลใน Cluster
2) เนื่องจากใช้หลายแพ็กเกจร่วมกัน จึงทำให้ต้องมีการศึกษาเพิ่มเติมเกี่ยวกับแนวทางการประยุกต์ใช้ร่วมกันของแพ็กเกจต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.5 ขอบเขตของโครงงาน

1) สามารถแสดงให้เห็นถึงความแตกต่างระหว่างการเรียกใช้แพ็กเกจ Dplyr และการใช้ภาษา R ในการวิเคราะห์ข้อมูลโดยตรง
2) สามารถนำแพ็กเกจ Dplyr ที่เลือกมาใช้ในการวิเคราะห์ข้อมูลร่วมกับแพ็กเกจอื่น เพื่อแสดงประสิทธิภาพในการทำงานร่วมกันออกมา
3) สามารถแสดงข้อมูลจากการวิเคราะห์ออกมาในรูปแบบที่สามารถเข้าใจได้ง่าย

1.6 แผนการดำเนินงานของโครงงาน

โดยมีระยะเวลาดำเนินงานทั้งหมด 2 เทอม

ตาราง 1.1 แผนการดำเนินงานระยะเวลา 2 เทอม

แผนการดำเนินงาน	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ม.ค.	ก.พ.	มี.ค.
1. ค้นหาข้อมูลที่ต้องการนำมาใช้งาน	←	→					
1.1 หาข้อมูลที่น่าสนใจ	←	→					
1.2 หาแหล่งที่มาของข้อมูล	←	→					
2. ศึกษาเครื่องมือที่ต้องการใช้งาน	←	→					
2.1 ทดลองใช้งานแพ็กเกจ Dplyr	←	→					
2.2 ทดลองใช้งาน Profiling ใน R Studio	←	→					
2.3 ศึกษาการทำVisualization และเครื่องมือที่ใช้ในการทำ	←	→					
2.4 ศึกษาทฤษฎีที่เกี่ยวข้อง	←	→					
3.ศึกษาและทำความเข้าใจข้อมูลที่ต้องการนำมาใช้งาน	←	→	→				
3.1 เลือกข้อมูลที่ต้องการนำมาใช้งาน	←	→					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 1.1 แผนการดำเนินงานระยะเวลา 2 เทอม (ต่อ)

แผนการดำเนินงาน	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ม.ค.	ก.พ.	มี.ค.
3.2 กำหนดช่วงเวลาของข้อมูลที่ต้องการนำมาใช้งาน	←	→					
3.3 จัดเตรียมข้อมูลให้พร้อมใช้งาน		←	→				
3.4 ออกแบบแบบจำลองแสดงความสัมพันธ์ของข้อมูลอธิบายโครงสร้างของฐานข้อมูล		←	→				
3.5 แสดงความสัมพันธ์ของข้อมูล		←	→				
4. นำข้อมูลที่ได้จัดเก็บลงใน Local Server		←	→				
4.1 เลือก Local server ที่จะนำมาใช้งาน		←	→				
4.2 นำข้อมูลมาจัดเก็บใน Local server		←	→				
5. สรุปผลจากการศึกษาข้อมูลที่ต้องการใช้งานและหาวิธีแก้ไขปัญหา		←	→				
5.1 แลกเปลี่ยนความรู้ที่ได้จากการทดลอง		←	→				
5.2 นำข้อมูลใน Local server มาใช้งาน		←	→				
5.3 เลือกแพ็กเกจเสริมที่ต้องการนำมาใช้งานร่วมกับแพ็กเกจหลัก		←	→				
6. พัฒนากลุ่มแพ็กเกจให้สามารถใช้งานได้หลากหลายยิ่งขึ้น		←	→				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 1.1 แผนการดำเนินงานระยะเวลา 2 เทอม (ต่อ)

แผนการดำเนินงาน	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ม.ค.	ก.พ.	มี.ค.
7. นำแพ็กเกจที่เลือกมาทดลองใช้งานกับข้อมูลขนาดเล็ก		←	—	→			
7.1 ชุดข้อมูลผู้ติดเชื้อไวรัสโคโรนา 2019		←	→				
7.2 ชุดข้อมูล GDP			←	→			
7.3 ชุดข้อมูลอุณหภูมิ			←	→			
7.4 ชุดข้อมูลอันดับสาธารณสุข			←	→			
7.5 นำผลการวิเคราะห์ข้อมูลขนาดเล็กมาใช้ในการทำ Visualization			←	→			
8. เปรียบเทียบประสิทธิภาพของแพ็กเกจ				↔			
8.1 เปรียบเทียบประสิทธิภาพความแม่นยำ (ใช้การคำนวณทางคณิตศาสตร์)				↔			
8.2 เปรียบเทียบประสิทธิภาพการทำงาน (ใช้ประกอบการทำ Profiling)				↔			
9. สรุปผลจากการเปรียบเทียบประสิทธิภาพแพ็กเกจ				↔			
10. สรุปผลจากการวิเคราะห์ข้อมูลขนาดเล็ก				↔			
11. สรุปปัญหาและข้อเสนอแนะเพื่อนำไปใช้พัฒนาต่อ				↔			
12.เพิ่มขนาดข้อมูลและวิเคราะห์ข้อมูล					↔		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 1.1 แผนการดำเนินงานระยะเวลา 2 เทอม (ต่อ)

แผนการดำเนินงาน	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ม.ค.	ก.พ.	มี.ค.
12.1เพิ่มระยะเวลาของข้อมูลการแพร่ระบาด covid-19					↔		
12.2 ชุดข้อมูลโรคซาร์ส					↔		
12.3 ชุดข้อมูลอันดับการเข้าชมเว็บไซต์ Pornhub.com					↔		
12.4 ชุดข้อมูลการแพร่ระบาดเชื้อไวรัสโคโรนาในประเทศไทย					↔		
13. ทำ Visualization เพื่อแสดงผลข้อมูลขนาดใหญ่						↔	
14. ออกแบบ Model สำหรับวางมาตรการป้องกัน covid-19						↔	↔
15. ทำ Dashboard เพื่อแสดงผลข้อมูล						↔	
15.1 ออกแบบ Dashboard						↔	
15.2 ทำ Dashboard ที่สามารถ Interactive และแสดงข้อมูลแบบเรียลไทม์						↔	
16. สรุปผลจากการวิเคราะห์ข้อมูลขนาดใหญ่						↔	
17. สรุปปัญหาและข้อเสนอแนะเพื่อนำไปใช้ในการพัฒนาต่อไป						↔	↔
18. จัดทำรูปเล่มรายงาน			↔	↔	↔	↔	↔
18.1 บทที่ 1			↔				
18.2 บทที่ 2				↔			
18.3 บทที่ 3					↔		
18.4 บทที่ 4						↔	
18.5 บทที่ 5							↔

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 ข้อมูลที่ใช้ในการวิเคราะห์

2.1.1 โรคติดเชื้อไวรัสโคโรนา 2019

โรคติดเชื้อไวรัสโคโรนา 2019 (Novel coronavirus 2019, 2019-nCoV) หรือที่รู้จักกันในชื่อโควิด-19 ที่เริ่มมีการระบาดในเมืองอู่ฮั่น (Wuhan) มณฑลหูเป่ย (Hubei) ประเทศจีน จึงทำให้ไวรัสตัวดังกล่าวถูกเรียกอีกชื่อว่า ไวรัสอู่ฮั่น ก่อนจะแพร่กระจายไปทั่วโลก ซึ่งต่อมาได้เปลี่ยนชื่อเป็นโคโรนาไวรัสสายพันธุ์กลุ่มอาการทางเดินหายใจเฉียบพลันรุนแรง 2 (Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)) ในช่วงแรกคาดว่าเป็นการติดเชื้อจากสัตว์สู่คน ไวรัสชนิดนี้มีสารพันธุกรรมชนิด RNA (Ribonucleic Acid) มีเปลือกหุ้มด้านนอกประกอบด้วยโปรตีนคลุมด้วยคาร์โบไฮเดรต เมื่อส่องดูด้วยกล้องจุลทรรศน์อิเล็กตรอนจึงเห็นเป็นลักษณะคล้ายมงกุฎ ซึ่งในภาษาลาตินคำว่า Corona แปลว่ามงกุฎ จึงเป็นที่มาของชื่อไวรัสในกลุ่มนี้ โอกาสเกิดการกลายพันธุ์มีได้สูง สามารถก่อให้เกิดการติดเชื้อข้ามสปีชีส์ได้ เป็นไวรัสโคโรนาสายพันธุ์ที่ก่อโรคในสัตว์ทั้งระบบทางเดินหายใจและทางเดินอาหารเกิดการแพร่จากสัตว์สู่คนจนก่อให้เกิดโรคในคน (Zoonotic Infection) โดยรายแรกที่ถูกระบุว่าติดเชื้อมาจากคนงานและลูกค้าในตลาดขายส่งอาหารทะเลฮั่วนาน (Huanan Seafood Wholesale Market) ในเดือนธันวาคม พ.ศ. 2561 และต่อมาเชื้อมีการกลายพันธุ์และกระจายมากขึ้น จึงสามารถติดต่อจากคนสู่คน

ปัจจุบันแหล่งที่มาของเชื้อไวรัส SARS-CoV-2 หลักฐานที่มีอยู่ทั้งหมดแสดงให้เห็นว่า SARS-CoV-2 มีต้นกำเนิดจากสัตว์ตามธรรมชาติและไม่ใช่ไวรัสที่สร้างขึ้น และมีองค์ประกอบทางนิเวศวิทยาคล้ายคลึงกับในค้างคาว เป็นกลุ่มไวรัสที่มีความเกี่ยวข้องกับ SARS-CoV และไวรัสชนิดอื่น ๆ ที่อยู่ในค้างคาว รวมถึง MERS-CoV ด้วย แต่อาจมีความเกี่ยวข้องไม่มากนัก

โดยอาการของผู้ติดเชื้อไวรัสชนิดนี้ อาจจะมีอาการเป็นไข้ ไอ มีน้ำมูก หายใจถี่เหนื่อยง่าย ในรายที่หนักมากการติดเชื้อสามารถก่อให้เกิดโรคปอดบวม ปอดอักเสบ ผู้เสียชีวิตส่วนใหญ่เป็นผู้สูงอายุและมีโรคประจำตัว หรือเกิดภาวะแทรกซ้อนต่าง ๆ ระยะเวลาติดเชื้อของไวรัสจะแตกต่างกันไปแต่ละบุคคล อาการไม่รุนแรงสำหรับผู้ที่สุขภาพแข็งแรง หลายอาการคล้ายคลึงกับไข้หวัดใหญ่ สำหรับผู้ที่มีปัญหาทางสุขภาพ การฟื้นจากอาการป่วยอาจใช้เวลานานกว่าคนที่สุขภาพแข็งแรง การป้องกันการติดเชื้อสามารถทำได้โดยการล้างมือด้วยสบู่เป็นเวลา 20 วินาที หรือใช้แอลกอฮอล์ล้างมือ เช็ดทำความสะอาดจุดรวมสัมผัส ของใช้ส่วนตัวที่ใช้มือหยิบจับบ่อย ๆ ห้าม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ใช้มือสัมผัสหน้าหากยังไม่มีการทำความสะอาด และผู้ที่เดินทางมาจากพื้นที่เสี่ยงควรแยกตัวออกจากผู้อื่นอย่างน้อยเป็นเวลา 14 วัน

แหล่งที่มาข้อมูลการระบาด Novel Coronavirus (COVID-19) ตั้งแต่วันที่ 22 มกราคม 2020 ถึงวันที่ 22 มกราคม 2021 ข้อมูลนี้ถูกรวบรวมโดยศูนย์วิทยาศาสตร์และวิศวกรรม มหาวิทยาลัย Johns Hopkins (JHU CCSE) จากแหล่งต่าง ๆ รวมถึงองค์กรการอนามัยโลก World Health Organization (WHO) ข้อมูลนี้ประกอบด้วย Province/State, Country/Region, Last Update, Confirmed, Suspected, Recovered, Deaths

2.1.2 ผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Product: GDP)

ผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Product: GDP) หมายถึง มูลค่าทางตลาดของสินค้าและบริการขั้นตอนสุดท้ายที่ได้ผลิตขึ้นในประเทศในช่วงเวลาหนึ่ง ๆ โดยไม่คำนึงว่าผลผลิตนั้นจะผลิตขึ้นมาด้วยทรัพยากรของประเทศอื่น ซึ่งคิดค้นโดย Simon Kuznets นักเศรษฐศาสตร์ชาวรัสเซีย ผลิตภัณฑ์มวลรวมในประเทศสามารถใช้เป็นตัวที่บ่งชี้ถึงมาตรฐานการครองชีพของประชากรและสะท้อนถึงภาพรวมของเศรษฐกิจในประเทศนั้น ๆ ได้

รูป 2.1 เปรียบเทียบ GDP ของ IMF ในปี 2019

รูป 2.2 เปรียบเทียบ GDP ของ IMF ในแต่ละปี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

การวัดผลิตภัณฑ์มวลรวมภายในประเทศ

การวัดผลิตภัณฑ์มวลรวมภายในประเทศ สามารถวัดได้ ดังต่อไปนี้

1) การวัดรายจ่ายภายในประเทศ (Expenditure Approach) คือรายจ่ายให้สินค้าและบริการขั้นสุดท้ายในประเทศคำนวณได้ตามสมการนี้

GDP = รายจ่ายเพื่อบริโภค (Consumption) + รายจ่ายเพื่อการลงทุน (Investment) + รายจ่ายของรัฐบาล (Government spending) + รายจ่ายสุทธิของต่างประเทศที่ซื้อสินค้าผลิตในประเทศ (Exports – Imports)

2) การวัดรายได้ภายในประเทศ (Resource Cost - Income Approach) คือรายได้จากการขายสินค้าและบริการขั้นสุดท้ายในประเทศ คำนวณได้ตามสมการนี้

GDP = ค่าจ้างและเงินเดือนลูกจ้าง + รายได้เจ้าของธุรกิจส่วนตัว + รายได้ผู้ถือหุ้น + ดอกเบี้ย + ค่าเช่า + ภาษีธุรกิจทางอ้อม + ค่าเสื่อมราคา + รายได้สุทธิของคนต่างชาติในประเทศ

สามารถอธิบายตัวแปรในสมการที่คำนวณ GDP ดังต่อไปนี้

$$GDP = C + I + G + (X - M) \tag{2.1}$$

โดยที่

C คือ Consumption หรือ การบริโภค คือ การบริโภคภาคเอกชน (Private consumption) ซึ่งรวมค่าใช้จ่ายส่วนบุคคลแทบทั้งหมด เช่น ปัจจัยสี่ ค่าอาหาร ยารักษาโรค แต่ไม่รวมการซื้อที่อยู่อาศัยหลังใหม่

I คือ Investment หรือ การลงทุน คือ การลงทุนธุรกิจในสินค้าทุน เช่น การก่อสร้างเหมืองแร่ใหม่ การซื้อซอฟต์แวร์ การซื้ออุปกรณ์เครื่องจักรสำหรับโรงงาน เป็นต้น การใช้จ่ายโดยครัวเรือนเพื่อซื้อบ้านหลังใหม่รวมไว้ในการลงทุน แต่การซื้อผลิตภัณฑ์ทางการเงินเช่น การซื้อหุ้นไม่จัดว่าเป็นการลงทุนแต่เป็นการออม จึงไม่รวมในผลิตภัณฑ์มวลรวมในประเทศ เพราะเป็นเพียงการสับเปลี่ยนเอกสารทางกฎหมายเท่านั้น ซึ่งเงินนั้นไม่ได้แปลงให้กลายเป็นสินค้าหรือบริการ จึงไม่เป็นส่วนหนึ่งของเศรษฐกิจที่แท้จริง และจัดให้เป็นรายจ่ายประเภทเงินโอน (Transfer payment)

G คือ Government Spending หรือ รายจ่ายรัฐบาล คือ ค่าใช้จ่ายทั้งหมดของรัฐบาลที่ใช้ซื้อสินค้าในประเทศ ซึ่งรวมถึงเงินเดือนของข้าราชการ การซื้ออาวุธทางทหาร และค่าใช้จ่ายลงทุนของรัฐ แต่ไม่รวมรายจ่ายประเภทโอนเงินอย่างเช่น สวัสดิการสังคมหรือผลประโยชน์จากการว่างงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

X คือ Export คือ การส่งสินค้าออกภายในประเทศ

M คือ Import คือ การนำเข้าสินค้าภายในประเทศ

แหล่งที่มาข้อมูลผลิตภัณฑ์มวลรวมในประเทศหรือ GDP มาจาก (IMF International Monetary Fund) คือกองทุนการเงินระหว่างประเทศ ซึ่งเป็นที่รู้จักกันในฐานะกองทุนการเงิน ก่อตั้งเมื่อวันที่ 22 กรกฎาคม 1944 โดยแนวคิดการก่อตั้ง IMF มาจากที่ประชุมว่าด้วยเรื่องการเงินและการคลังแห่ง สหประชาชาติ หรือ United Nations Monetary and Financial Conference ณ เมืองเบรตตัน วูดส์ รัฐ นิวแฮมเชียร์ ประเทศสหรัฐอเมริกา หรือที่รู้จักกันในนาม การประชุมเบรตตัน วูดส์ (Bretton Woods Conference) ข้อมูลประกอบด้วยรายชื่อประเทศ ข้อมูล GDP ปี 2019

2.1.3 อุณหภูมิ (Temperature)

อุณหภูมิ (Temperature) หมายถึง การวัดค่าเฉลี่ยของพลังงานจลน์ซึ่งเกิดขึ้นจากอะตอมแต่ละตัว หรือแต่ละโมเลกุลของสสาร เมื่อเราใส่พลังงานความร้อนให้กับสสาร อะตอมของสสารจะเคลื่อนที่เร็วขึ้น ทำให้อุณหภูมิสูงขึ้น แต่เมื่อเราลดพลังงานความร้อน อะตอมของสสารจะเคลื่อนที่ช้าลง ทำให้อุณหภูมิลดลง เนื่องจากอุณหภูมิอากาศมีเปลี่ยนแปลงในแต่ละช่วงเวลา เช่น ปี ฤดูกาล เดือน วัน หรือแม้กระทั่งรายชั่วโมง นักอุตุนิยมวิทยาจึงศึกษาค่าเฉลี่ยของอุณหภูมิอากาศดังนี้

1) อุณหภูมิเฉลี่ยในแต่ละวัน (Daily Mean Temperature) ใช้ค่าอุณหภูมิสูงสุดและอุณหภูมิต่ำสุดรวมกันหารสอง
2) อุณหภูมิเฉลี่ยของเดือน (Monthly Mean Temperature) ใช้ค่าเฉลี่ยอุณหภูมิของแต่ละวันรวมกันหารด้วยจำนวนวัน
3) อุณหภูมิเฉลี่ยของปี (Yearly Mean Temperature) ใช้ค่าเฉลี่ยอุณหภูมิของแต่ละเดือนรวมกันหารด้วยจำนวนเดือน

แหล่งที่มาข้อมูลรายชื่อเมืองตามอุณหภูมิเฉลี่ย (รายเดือนและรายปี) จากWikipedia อุณหภูมิที่ระบุเป็นค่าเฉลี่ยของความสูงและต่ำในแต่ละวัน ดังนั้นอุณหภูมิตอนกลางวันจริงในเดือนหนึ่ง ๆ จะสูงกว่าอุณหภูมิที่ระบุไว้ในที่นี้ประมาณ 2 ถึง 10 ° C (4 ถึง 18 ° F) ขึ้นอยู่กับความแตกต่างระหว่างความสูงและต่ำในแต่ละวัน แหล่งข้อมูลอุณหภูมิประกอบด้วยรายชื่อประเทศ เมือง วันที่และค่าเฉลี่ย อุณหภูมิในแต่ละวัน รวบรวมข้อมูลโดยมหาวิทยาลัย Dayton

2.1.4 ระบบสาธารณสุขทั่วโลกปี 2019

แหล่งที่มาข้อมูลจากนิตยสาร CEOWORLD นิตยสารด้านธุรกิจของสหรัฐอเมริกา ได้จัดอันดับประเทศที่มีระบบสุขภาพดีที่สุดในโลก ประจำปี 2019 โดยมีการสำรวจทั้งหมด 89 ประเทศทั่วโลก พิจารณาจากหลักการต่าง ๆ ดังต่อไปนี้

1) โครงสร้างพื้นฐานด้านสาธารณสุข
2) ประสิทธิภาพด้านสาธารณสุข แพทย์ พยาบาลและบุคลากรอื่น ๆ
3) ค่าใช้จ่ายในระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4) การเข้าถึงยาคุณภาพ
5) ความพร้อมของรัฐบาลในการจัดการระบบ
6) ปัจจัยอื่น ๆ เช่น สิ่งแวดล้อม การเข้าถึงแหล่งน้ำสะอาด ระบบสุขอนามัย และควบคุมปัจจัยเสี่ยงด้านสุขภาพ เช่น การจัดการกับบุหรี่ ยาสูบ และการจัดการโรคอ้วน

Rank	Country	Health Care Index (Overall)	Infrastructure	Professionals	Cost	Medicine Availability
1	Taiwan	78.72	87.16	14.23	83.59	82.3
2	South Korea	77.7	79.05	13.06	78.39	78.99
3	Japan	74.11	90.75	30.01	82.59	92.06
4	Austria	71.32	86.18	20.25	78.99	88.23
5	Denmark	70.73	78.77	21.6	74.88	74.18
6	Thailand	67.99	92.58	17.37	96.22	67.51
7	Spain	65.38	77.86	13.24	71.82	55.1
8	France	64.66	86.28	34.25	75.81	83.82

รูป 2.3 คะแนนอันดับระบบสาธารณสุขทั่วโลกปี 2019 แหล่งที่มาข้อมูลจากนิตยสาร CEOWORLD

2.1.5 จำนวนประชากร

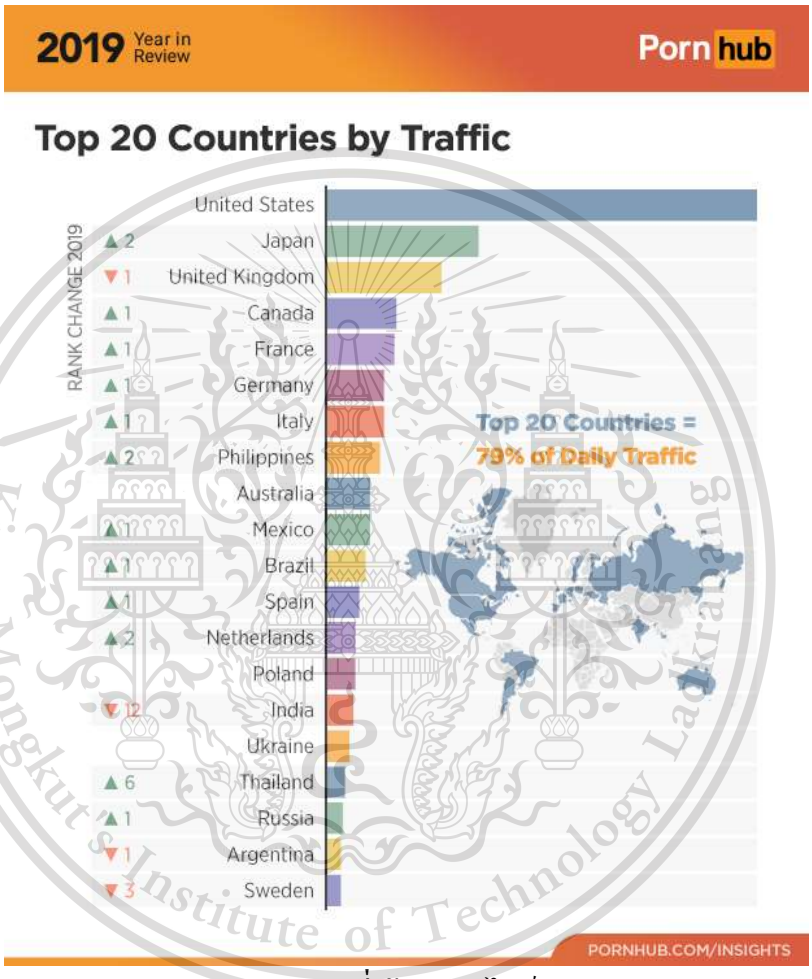
แหล่งที่มาข้อมูลจำนวนประชากร 2019 ครั้งที่ 26 การประมาณการและการคาดการณ์จำนวนประชากรอย่างเป็นทางการขององค์การสหประชาชาติ United Nations (UN) เป็นศูนย์การสำหรับการประสานงานของการกระทำของชาติต่าง ๆ ซึ่งเป็นตัวแทนในระดับสากลน่าเชื่อถือมากที่สุดจัดทำโดยกองประชากรของกรมเศรษฐกิจและสังคมของสำนักเลขาธิการสหประชาชาติ

ผลลัพธ์หลักจะถูกนำเสนอในชุดไฟล์ Excel ที่แสดงตัวบ่งชี้ทางประชากรที่สำคัญสำหรับแต่ละกลุ่มการพัฒนาของสหประชาชาติกลุ่มรายได้ของธนาคารโลกภูมิภาคทางภูมิศาสตร์เป้าหมายการพัฒนาที่ยั่งยืน (SDGs) ภูมิภาคและประเทศหรือพื้นที่สำหรับช่วงเวลาหรือวันที่ที่เลือกภายใน 1950 - 2100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.6 ข้อมูลการเข้าชมเว็บไซต์ Pornhub 2019

แหล่งที่มาข้อมูลการเข้าชมเว็บไซต์ Pornhub 2019 วิเคราะห์ข้อมูลโดยเว็บไซต์ Pornhub ที่มีการจัดทำขึ้นทุกปี มีข้อมูลที่น่าสนใจหลากหลาย ได้แก่ จัดอันดับประเทศที่เข้าชมเว็บไซต์ Pornhub มากที่สุด 20 อันดับ อันดับช่องทางที่เข้าชมเว็บไซต์ Pornhub อันดับดารานักแสดงที่ได้รับการรับชม อันดับหมวดหมู่ที่ถูกค้นหา เป็นต้น



รูป 2.4 20 อันดับประเทศที่เข้าชมเว็บไซต์ Pornhub 2019

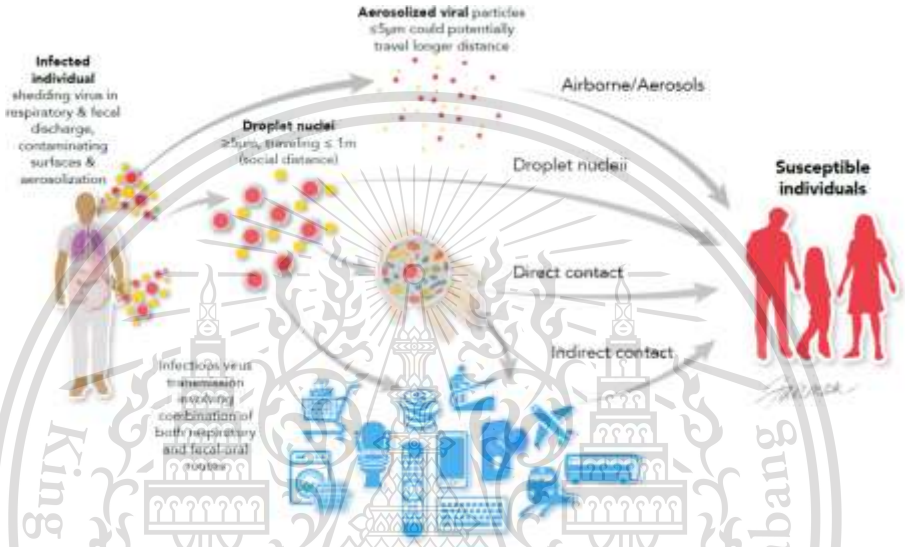
แหล่งข้อมูลที่เลือกมาใช้ในการวิเคราะห์เป็นเพียง 20 อันดับประเทศที่เข้าชมเว็บไซต์ Pornhub ปี 2019 จากทุกประเทศที่เข้าชมทั่วโลก

2.1.7 ข้อมูลการโรคระบาดซาร์ส

โรคทางเดินหายใจเฉียบพลันหรือโรคซาร์ส (Severe acute respiratory syndrome : SARS) ผู้ป่วยมีอาการติดเชื้อในทางเดินหายใจอย่างรุนแรง และอาการในระบบทางเดินอาหาร เชื้อที่เป็นสาเหตุคือ เชื้อไวรัสโคโรนา (Coronavirus) เกิดการระบาดครั้งแรกในมณฑลกวางตุ้งของประเทศจีน ในปลายปี พ.ศ. 2545 โดยพบผู้ป่วยปอดบวมไม่ตอบสนองต่อยาปฏิชีวนะ ต่อมาเกิดการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ระบาดของโรคปอดบวมในประเทศเวียดนาม ฮ่องกง สิงคโปร์ แคนาดา สามารถเชื่อมโยงเหตุการณ์ต่างๆได้ว่ามาจากแพทย์ท่านหนึ่งที่ดูรักษาผู้ป่วยในมณฑลกวางตุ้ง เดินทางมายังฮ่องกงขณะมีอาการไข้ และเข้าพักโรงแรมก่อนจะถูกนำส่งโรงพยาบาลและเสียชีวิตในเวลาต่อมา และพบว่าคนในโรงแรมหลายคนได้ติดเชื้อและนำกลับไปยังประเทศของตน จนกระทั่ง ณ วันที่ 6 กรกฎาคม พ.ศ. 2546 ได้มีการแพร่ระบาดไปยัง 29 ประเทศ สรุปโดยรวมมีรายงานผู้ป่วย 8,098 ราย และเสียชีวิต 774 ราย อัตราผู้เสียชีวิตจากการติดเชื้อร้อยละ 9.6



รูป 2.5 โรคทางเดินหายใจเฉียบพลันรุนแรงหรือโรคซาร์ส

แหล่งที่มาข้อมูลโรคทางเดินหายใจเฉียบพลันหรือโรคซาร์สมาจากองค์การอนามัยโลก (World Health Organization : WHO) โดยชุดข้อมูลประกอบไปด้วย วันที่พบผู้ติดเชื้อ ประเทศที่พบผู้ติดเชื้อ จำนวนผู้ติดเชื้อ จำนวนผู้เสียชีวิตและจำนวนผู้รักษาหาย

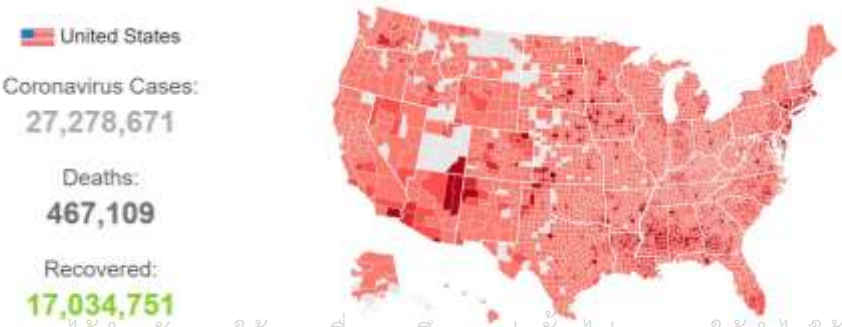
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.8 ข้อมูลการแพร่ระบาดของไวรัสโคโรนา 2019 ในประเทศไทย

รูป 2.6 รายงาน Covid-19 ประจำวันประเทศไทย

ประเทศไทยมีการแพร่ระบาดของไวรัสโคโรนา 2019 ซึ่งมีผู้ติดเชื้อภายในประเทศเป็นจำนวนมาก กรมควบคุมโรคแห่งประเทศไทย (ศบค.) ได้มีการจัดเก็บข้อมูลตั้งแต่เริ่มมีผู้ติดเชื้อภายในประเทศโดยเริ่มเก็บข้อมูลตั้งแต่วันที่ 20 มีนาคม พ.ศ.2563 จนถึงปัจจุบัน โดยปรับปรุงข้อมูลความถี่แบบรายวัน โดยชุดข้อมูลประกอบด้วย จำนวนผู้ติดเชื้อ วันที่ติดเชื้อ สัญชาติผู้ติดเชื้อ เพศ อายุ จังหวัดที่พบผู้ติดเชื้อ จังหวัดที่รักษาผู้ติดเชื้อ อำเภอที่พบผู้ติดเชื้อ และสาเหตุความเสี่ยงของผู้ติดเชื้อ

2.1.9 ข้อมูลการแพร่ระบาดเชื้อไวรัสโคโรนา 2019 ในสหรัฐอเมริกา



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

รูป 2.7 รายงาน Covid-19 ในสหรัฐอเมริกา

สหรัฐอเมริกามีการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ในระดับที่สูงมากเป็นอันดับแรกของโลก จากรายงานพบว่าเริ่มตรวจพบผู้ติดเชื้อเป็นครั้งแรกในวันที่ 21 มกราคม ค.ศ. 2020 หลังจากนั้นในสหรัฐอเมริกามีอัตราการติดเชื้อรวมถึงจำนวนผู้เสียชีวิตเพิ่มสูงขึ้นอย่างรวดเร็ว และกลายเป็นประเทศที่มีผู้ติดเชื้อสูงที่สุดในโลกในเวลาต่อมา ชุดข้อมูลโรคติดเชื้อไวรัสโคโรนา 2019 ในสหรัฐอเมริกาประกอบด้วยวันที่ รายชื่อรัฐ จำนวนผู้ติดเชื้อ และจำนวนผู้เสียชีวิต และชุดข้อมูลอื่น ๆ ในสหรัฐอเมริกา

2.1.9.1 ข้อมูลอัตราของประชากรแบ่งตามเพศ

ข้อมูลของประชากรในสหรัฐอเมริกาแบ่งตามเพศในแต่ละรัฐคิดเป็นร้อยละ โดยข้อมูลประกอบด้วยรายชื่อรัฐ จำนวนประชากรเพศชายคิดเป็นร้อยละ และจำนวนประชากรหญิงคิดเป็นร้อยละ

2.1.9.2 ข้อมูลจำนวนประชากร

ข้อมูลจำนวนประชากรที่อาศัยอยู่ในสหรัฐอเมริกาในแต่ละรัฐ โดยข้อมูลประกอบด้วยรายชื่อรัฐ และจำนวนประชากรในรัฐนั้น ๆ

2.1.9.3 ข้อมูลการล็อกดาวน์

ข้อมูลการล็อกดาวน์หรือการปิดเมืองในสหรัฐอเมริกา ซึ่งเป็นการเพิ่มการป้องกันการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 และจำกัดให้สามารถควบคุมการแพร่ระบาดได้ แต่อาจส่งผลกระทบต่อเศรษฐกิจองค์รวมและมีการจำกัดสิทธิส่วนบุคคล โดยการล็อกดาวน์อาจมีระดับการบังคับใช้ที่เข้มงวดแตกต่างกันในแต่ละพื้นที่ อาจอยู่ในรูปแบบของการจำกัดเวลาในการเดินทาง การปิดสถานศึกษาหรือพื้นที่สาธารณะ เป็นต้น

ชุดข้อมูลในส่วนนี้ประกอบด้วยรายชื่อรัฐ เดือนที่ใช้การล็อกดาวน์ วันที่ล็อกดาวน์ วันที่แบบระบุวันเดือนปีที่ล็อกดาวน์ ประเภทของการล็อกดาวน์ จำนวนวันที่ล็อกดาวน์

2.1.9.4 ผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Product: GDP) ในสหรัฐอเมริกา

ข้อมูลผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ของสหรัฐอเมริกา ใช้เป็นตัวที่บ่งชี้ถึงมาตรฐานการครองชีพของประชากรและสะท้อนถึงภาพรวมของเศรษฐกิจในสหรัฐอเมริกา โดยข้อมูลชุดนี้ประกอบด้วยรายชื่อรัฐ, อันดับ GDP, ค่า GDP, เปอร์เซ็นเมื่อเทียบแต่ละรัฐในสหรัฐอเมริกา, เปอร์เซ็นต์การเติบโตของ GDP และจำนวนประชากรในสหรัฐอเมริกาประจำปี 2020

2.1.9.5 ข้อมูลประชากรไร้บ้าน (Homeless)

ข้อมูลประชากรที่ไร้บ้านหรือไร้ที่อยู่อาศัยในสหรัฐอเมริกา ชุดข้อมูลประกอบด้วยจำนวน CoC (Number of CoCs) โดยเป็นข้อมูลจำนวนหน่วยงานที่ดูแลและเยียวยาบุคคลไร้บ้านในระยะยาว (Continuum of Care) และจำนวนประชากรไร้บ้านในแต่ละรัฐ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.9.6 ข้อมูลอันดับสาธารณสุขของสหรัฐอเมริกา

ข้อมูลอันดับสาธารณสุขของสหรัฐอเมริกาแบ่งตามรัฐ จัดอันดับโดย Agency for Healthcare Research and Quality (AHRQ) โดยอันดับคิดจากองค์ประกอบทางสาธารณสุข ได้แก่ ประเภทการดูแล (Type of care), ระบบการดูแล (Setting of care), พื้นที่ที่มีบริการสถานพยาบาล (Clinical areas)

2.2 ภาษา R และเทคโนโลยีที่เกี่ยวข้อง

รูป 2.8 เครื่องหมายการค้าของ R programming

R programming หรือ ภาษา R เป็นซอฟต์แวร์ Open Source ที่สามารถเข้าถึงและสามารถใช้งานง่าย ไม่เสียค่าใช้จ่าย ไม่มีการละเมิดลิขสิทธิ์ อีกทั้งยังมีความสามารถสูง มีความยืดหยุ่น สามารถแสดงผลด้วยกราฟิก มีแพ็กเกจให้เลือกใช้จำนวนมาก และยังมี Built-in ฟังก์ชัน มากมาย สาเหตุนี้จึงทำให้ R ได้รับความนิยมเพื่อใช้ในการวิเคราะห์ข้อมูล ใช้คำนวณทางสถิติ

R พัฒนามาจากภาษา S ซึ่งได้รับการพัฒนาโดย Robert Gentleman และ Ross Ihaka โดย R อยู่ภายใต้ข้อกำหนดของ GUN R สามารถใช้งานได้กับหลายระบบปฏิบัติการ ไม่ว่าจะเป็น UNIX platforms หรือระบบที่คล้ายคลึงกัน รวมทั้ง Windows และ MacOS เนื่องจากภาษา S และภาษา R มีคามคล้ายคลึงกันมาก และจากภาษา S มักเป็นสื่อของตัวเลือกสำหรับการวิเคราะห์ข้อมูลทางสถิติ แต่เนื่องจาก R เป็น Open Source จึงมาความเอื้ออำนวยให้สามารถใช้งานได้มากกว่า ดังนั้น R จึงเป็นที่รู้จักและได้รับความนิยมมากกว่านั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.2.1 R Studio

รูป 2.9 เครื่องหมายการค้าของ RStudio

RStudio คือ Integrated Development Environment (IDE) สำหรับ R ซึ่งเป็นภาษาสำหรับการคำนวณทางสถิติและกราฟิก มีรูปแบบ คือ RStudio Desktop (ทำงานบน Desktop ใน Windows, Mac และ Linux) และ RStudio Server (ทำงานบนเบราว์เซอร์ที่เชื่อมต่อกับเซิร์ฟเวอร์ เช่น Debian / Ubuntu, Red Hat / CentOS และ SUSE Linux) โดยการทำงานร่วมกับ R นั้นในส่วนของการทำงานและการคำนวณจะดำเนินการอยู่ที่ R แต่การสื่อสาร การแสดงผล จะปรากฏผ่าน RStudio เป็นซอฟต์แวร์ Open source ถูกพัฒนาขึ้นในปี 2009 โดย J.J. Allaire โดยปัจจุบันเป็น AGPL เวอร์ชัน 3

สามารถแบ่งหน้าที่ ออกเป็น 4 ส่วน คือ

1) Source
2) Console เป็นส่วนที่ใช้เขียนคำสั่ง
3) Environment และ History เป็นส่วนที่แสดงผลของค่าที่กำหนด และเก็บไว้โดยโปรแกรม
4) File, Plot, Package, Help และ View โดยแท็บ Plot เป็นส่วนที่แสดงผลเชิงกราฟิกต่าง ๆ แท็บ File เป็นส่วนแสดง File directory ที่ทำงานอยู่ แท็บ Package แสดงแพ็กเกจที่มีอยู่ และยังสามารถดาวน์โหลดแพ็กเกจเพื่อใช้งาน หรือติดตั้งแพ็กเกจและอัพเดตแพ็กเกจได้ด้วย

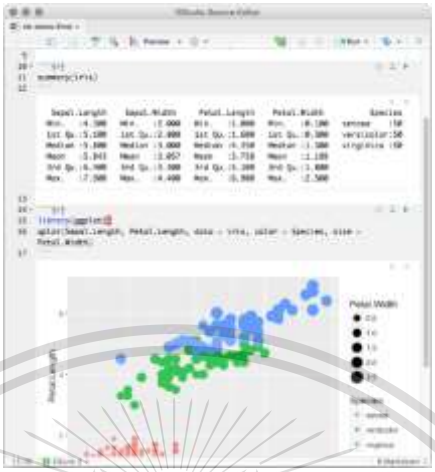
2.2.1.1 R Notebook

R Notebook เป็น Markdown ที่แบ่งการดำเนินการแต่ละส่วนออกจากกันอย่างอิสระ แสดงผลลัพธ์ให้เห็นได้ทันทีหลังจากใส่อินพุต โดยไฟล์ R Notebook สามารถแบ่งออกได้เป็น 3 ส่วนคือ

1) YAML metadata คือ ส่วนอธิบายข้อมูลเพื่อสร้างไฟล์ที่ต้องการ
2) R Markdown สามารถกำหนดรูปแบบผลลัพธ์ที่ต้องการได้
3) Text คือ ส่วนพิมพ์ข้อความหรือคำอธิบายรายละเอียดของ R Notebook นั้น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4) Chunks คือ ส่วนพิมพ์คำสั่งภาษา R สามารถเลือกประมวลผล Chunks ที่ต้องการได้เนื่องจากแต่ละ Chunks นั้นเป็นอิสระต่อกัน



รูป 2.10 การทำงานของ R Notebook

R Markdown สามารถใช้เป็น Notebook และ R Notebook ทั้งหมดสามารถแสดงผลเป็น Markdown อื่น ๆ ได้จึงถือว่าเป็นตัวเลือกที่ดีและมีความรวดเร็ว ถ้าหากต้องการเผยแพร่สามารถทำได้โดยง่าย นอกจากนี้ยังสามารถแสดงผลออกมาในรูปแบบของ html ได้อีกด้วย

สาเหตุที่ผู้จัดทำเลือกใช้ RStudio เนื่องจากเป็นซอฟต์แวร์ Open Source ที่สามมารถใช้งานได้ฟรี รองรับการใช้งานในหลาย Platforms ไม่ว่าจะเป็น Windows, Mac หรือ Linux กราฟที่ได้มีการการันตีคุณภาพ มีความพร้อมในการรองรับข้อมูลและความพร้อมทางด้านอัลกอริทึม ผู้ใช้งานสามารถสร้างแพ็กเกจได้ และ RStudio ก็ยังถูกออกแบบมาเพื่อ R อีกด้วย

2.2.2 Dplyr

Dplyr เป็นแพ็กเกจสำหรับจัดการข้อมูลในภาษา R โดยพัฒนามาจาก Plyr มีความแตกต่างคือมุ่งเน้นที่ Data Frames และ Tibbles เป็นหลัก Dplry อยู่ในกลุ่ม Tidyverse ของแพ็กเกจในภาษา R ที่ใช้สำหรับ Data Science หากข้อมูลมีขนาดใหญ่เกินกว่าที่จะแสดงออกมาทั้งหมด Dplyr จะช่วยให้สามารถเข้าถึงข้อมูลที่ไม่สามารถแสดงออกมาได้ ช่วยจัดการข้อมูลก่อนนำไปวิเคราะห์ต่อไป Dplyr นั้นเขียนด้วย C ++ จึงทำให้มีการประมวลผลเร็ว และสามารถเขียนได้ด้วยภาษา R จึงทำให้เขียนได้ง่ายมากยิ่งขึ้น นอกจากนี้ยังสนับสนุนการทำงานร่วมกับฐานข้อมูล (ตารางเทียบเท่ากับ Data frame) การคำนวณถูก Offloaded ไปยังฐานข้อมูล โดยการแปล R เป็น SQL แต่จะรวบรวมผลลัพธ์และแสดงผลในส่วนของ R

ประกอบไปด้วย 5 ฟังก์ชันหลัก คือ

1) mutate() ใช้เพิ่มตัวแปรใหม่ ซึ่งส่วนใหญ่เป็นการสร้างคอลัมน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้นำไปดัดแปลง เผยแพร่ หรือนำออกเผยแพร่ต่อในครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2) arrange() ใช้เรียงลำดับของแถว ซึ่งโดยปกติแล้วจะเป็นเรียงข้อมูลจากต่ำไปสูง

3) select() ใช้เลือกตัวแปรตามชื่อ ส่วนใหญ่จะเป็นการเรียกเพื่อเลือกคอลัมน์ที่ต้องการ

4) filter() ใช้เลือกตามเงื่อนไขที่สร้างขึ้น

5) summarise() ใช้ลดค่าจำนวนหลายๆค่าให้เหลือเพียงข้อสรุปเดียว

ฟังก์ชันที่ใช้ในการทำงาน

2.2.2.1 ฟังก์ชัน rename

ใช้สำหรับการเปลี่ยนชื่อคอลัมน์

โปรแกรม 2.1 การใช้งานฟังก์ชัน rename()

```
rename(.data, replace, warn_missing = TRUE,
warn_duplicated = TRUE)
```

Arguments

1) .data: ชุดข้อมูลที่ต้องการเลือก Data

2) replace: ชื่อใหม่ที่ต้องการตั้ง (char)

3) warn_missing: พิมพ์ข้อความหากไม่มีชื่อเก่าอยู่ใน Data

4) warn_duplicated: พิมพ์ข้อความหากมีคอลัมน์นั้นใน Data มากกว่า 1 ครั้ง

ตัวอย่าง : rename(confirmed = count)

2.2.2.2 ฟังก์ชัน gather()

ใช้เพื่อทำให้ตารางข้อมูลที่มีความกว้างกลายเป็นข้อมูลที่มีความยาวแทน คือ ทำการเปลี่ยนคอลัมน์จำนวนมากให้มาอยู่ในรูปแถวและใช้ Key เพื่อรวบรวม

โปรแกรม 2.2 การใช้งานฟังก์ชัน gather()

```
gather(data, key = "key", value = "value", ..., na.rm =
FALSE,convert = FALSE, factor_key = FALSE)
```

Arguments

1) key, value: ชื่อคอลัมน์ที่ต้องการใช้เป็น Key สามารถเลือกค่าได้เหมือนใช้ select ()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2) na.rm: ถ้ากำหนดค่าให้เป็น TRUE จะลบแถวนั้นออกจากเอาต์พุต ถ้าค่าเป็น NA
3) convert: ถ้ากำหนดค่าให้เป็น TRUE จะเรียกใช้ type.convert() ที่ Key Column อัตโนมัติ ช่วยตอนคอลัมน์เป็น int, num หรือ logic
4) factor_key: ถ้ากำหนดค่าให้เป็น FALSE ค่าของ Key จะเป็น Char ถ้ากำหนดค่าเป็น TRUE ค่าของ Key จะเป็น Factor เอามาใช้เรียงลำดับ

2.2.2.3 ฟังก์ชัน setdiff()

ใช้เพื่อลบข้อมูลคอลัมน์ที่ไม่ต้องการออก

โปรแกรม 2.3 การใช้งานฟังก์ชัน setdiff()

```
setdiff(x, …) / setdiff(x, y, …)
```

Arguments

1) (x,y) หมายถึงใน x แต่ไม่ใช่ใน y

2.2.2.4 ฟังก์ชัน select()

ฟังก์ชัน select ใช้ในการเลือกข้อมูลคอลัมน์ที่ต้องการ มีการทำงานคล้ายคลึงกับฟังก์ชัน rename

โปรแกรม 2.4 การใช้งานฟังก์ชัน select()

```
select(.data, ...)
```

ฟังก์ชันย่อยที่สามารถใช้งานได้

1) starts_with(), ends_with(), contains()
2) matches()
3) num_range()
4) one_of()
5) everything()

ถ้าต้องการลบคอลัมน์สามารถใช้เครื่องหมาย “–” นำหน้าตัวแปรตัวนั้นใช้ “c()” เพื่อเลือกตัวแปรที่ต้องการ การใช้ select () จะเก็บตัวแปรเฉพาะตัวที่เลือกเท่านั้น แต่ rename()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เก็บตัวแปรทุกตัวและเปลี่ยนเฉพาะตัวที่ถูกเปลี่ยนชื่อเท่านั้น และใช้เครื่องหมาย ":" เพื่อเลือกตัวแปรเป็นช่วง

2.2.2.5 ฟังก์ชัน mutate()

ฟังก์ชัน mutate ใช้เพื่อแทนที่คอลัมน์เดิมด้วยค่าใหม่และสร้างคอลัมน์ใหม่ขึ้นมา

โปรแกรม 2.5 การใช้งานฟังก์ชัน mutate()

```
mutate(
      .data,
      ...,
      .keep = c("all", "used", "unused", "none"),
      .before = NULL,
      .after = NULL
)
```

Arguments

1) keep: เลือกเก็บข้อมูลคอลัมน์ที่ต้องการ โดยกำหนด ดังนี้

"all" เก็บทั้งหมด

"used" เก็บตัวแปรเอาไว้สร้างตัวแปรใหม่ เอาไว้ช่วยตรวจสอบข้อมูลเปรียบเทียบอินพุตและเอาต์พุต แบบ Side-by-side

"unused" เก็บเฉพาะตัวแปรที่มีอยู่ ไม่ใช่เพื่อสร้างตัวแปรใหม่

"none" เก็บเฉพาะ Key เหมือนการใช้ transmute()

2) before / after: การย้ายคอลัมน์ไปซ้ายหรือขวา ไม่ควรใช้พร้อมกัน

2.2.2.6 ฟังก์ชัน group_by()

ใช้ในการจัดกลุ่มข้อมูล

โปรแกรม 2.6 การใช้งานฟังก์ชัน group_by()

```
group_by(.data, ..., .add = FALSE, .drop =
group_by_drop_default(.data))
```

Arguments

1) .data: ชุดข้อมูลที่ต้องการเลือก data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2) .add: ถ้ากำหนดค่าให้เป็น FALSE จะแทนที่กลุ่มที่มีอยู่ ถ้ากำหนดให้เป็น TRUE จะเพิ่มจากกลุ่มที่มีอยู่

3) .drop: default = TRUE

2.2.2.7 ฟังก์ชัน summarise()

ฟังก์ชัน summarise โดยทั่วไปจะใช้ summary() กับข้อมูลที่จัดกลุ่มโดย group_by() เอาต์พุตจะมีหนึ่งแถวเท่านั้นจากหนึ่งกลุ่ม

โปรแกรม 2.7 การใช้งานฟังก์ชัน summarise()

```
summarise(.data, ..., .groups = NULL)
```

Arguments

1) .data : ชุดข้อมูลที่ต้องการเลือก data

2) .groups : สามารถกำหนดเป็นค่า NULL/ drop_last / drop / keep / rowwise ได้

ฟังก์ชันที่เกี่ยวข้องที่สามารถใช้งานร่วมกันได้

1) Center ได้แก่ mean(), median()

2) Spread ได้แก่ sd(), IQR(), mad()

3) Range ได้แก่ min(), max(), quantile()

4) Position ได้แก่ first(), last(), nth(),

5) Count ได้แก่ n(), n_distinct()

6) Logical ได้แก่ any(), all()

2.2.2.8 ฟังก์ชัน filter()

ใช้เลือกเฉพาะข้อมูลที่ต้องการออกมาแสดง

โปรแกรม 2.8 การใช้งานฟังก์ชัน filter()

```
arrange(.data, ..., .by_group = FALSE)
```

Arguments

1) .data ชุดข้อมูลที่ต้องการเลือก data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.2.2.9 ฟังก์ชัน arrange()

ใช้เพื่อเรียงข้อมูลตามเงื่อนไขที่ต้องการ

โปรแกรม 2.9 การใช้งานฟังก์ชัน arrange()

```
arrange(.data, ..., .by_group = FALSE)
```

Arguments

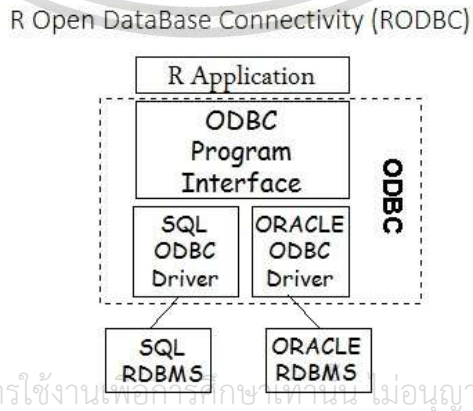
1) .data: ชุดข้อมูลที่ต้องการเลือก data เพื่อนำมาเรียง
2) .by_group: ถ้ากำหนดค่าให้เป็น TRUE จะเรียงลำดับโดยการจัดกลุ่มตัวแปร

2.2.3 Library ที่ใช้งานนอกเหนือจาก Dplyr

2.2.3.1 ODBC

ODBC มีจุดประสงค์เพื่อสร้าง API ที่ใช้ร่วมกันเพื่อใช้ในการเชื่อมต่อกับฐานข้อมูลในรูปแบบ SQL-Based Database Management Systems (DBMSs) เช่น MySQL2, MySQL, PostgreSQL เป็นต้น เป็นแพลตฟอร์มที่สามารถทำงานได้กับแพลตฟอร์มอื่น ๆ ได้หลากหลาย และมีโครงสร้างเป็นแบบ Client-server ช่วยให้สามารถเชื่อมต่อกับ Remote server ได้โดยไม่จำเป็นต้องระบุตัวตนของ End user โดยการเชื่อมต่อไปยัง DBMSs จะต้องใช้ ODBC driver ซึ่งอาจมีอยู่ในตัวจัดการ DBMS ตัวจัดการ ODBC server หรืออาจเป็น Driver แยกจากผู้พัฒนา ในส่วนการเชื่อมต่อไปยังฐานข้อมูลที่ระบุแบบจำเพาะจะถูกเรียกว่า Data Source Name หรือ DSN

2.2.3.2 RODBC



รูป 2.11 ODBC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

RODBC เป็นแพ็กเกจที่ใช้ในการเชื่อมต่อและเข้าถึงฐานข้อมูลที่อยู่ในระบบฐานข้อมูลเชิงสัมพันธ์ (Relational database) เช่น SQL Server, Oracle โดยเชื่อมต่อระหว่าง R และระบบฐานข้อมูลผ่านทาง ODBC Interface ในแพ็กเกจ RODBC ประกอบด้วยฟังก์ชันคำสั่ง 2 กลุ่มใหญ่ คือ คำสั่ง odbc* ที่ใช้สำหรับจัดการการเชื่อมต่อกับฐานข้อมูลในระดับล่าง (Low-level access) เพื่อเชื่อมต่อกับ ODBC และคำสั่ง sql* ที่ใช้สำหรับดำเนินการเกี่ยวกับฐานข้อมูลในระดับสูงเพื่ออ่าน บันทึก คัดลอก หรือแก้ไขข้อมูลระหว่าง Data Frame และตาราง SQL

โดยฟังก์ชันหลักที่ใช้ มีดังนี้

1) odbcConnect() ใช้สำหรับเชื่อมต่อกับฐานข้อมูล ODBC
2) sqlQuery() ใช้ส่งกลับข้อมูลที่ทำการ Query ไปยัง ODBC

2.2.3.3 Lubridate

เป็นแพ็กเกจที่สามารถแยกส่วนข้อมูลวันที่และเวลา และอัพเดตข้อมูลแต่ละส่วนเกี่ยวกับวันที่และเวลาได้ ได้แก่ ปี เดือน วัน ชั่วโมง นาที และวินาที ใช้สำหรับจัดการวันที่และเวลา หรือช่วงเวลาให้อยู่ในรูปแบบที่ต้องการ เพื่อให้นำข้อมูลในส่วนวันที่ เวลา หรือช่วงเวลาอยู่ในรูปแบบที่ต้องการ และสามารถนำไปใช้ในการวิเคราะห์ข้อมูลต่อไปได้ mdy() ใช้สำหรับแปลงรูปแบบของวันที่ใหม่ โดยให้เรียงลำดับเป็น เดือน วัน และปี ตามลำดับ

2.2.3.4 Magrittr

เป็นแพ็กเกจที่มีจุดมุ่งหมายในการสร้างเครื่องมือที่ช่วยลดระยะเวลาการพัฒนาโค้ด และช่วยเพิ่มความสะดวกในการอ่านโค้ด เพื่อให้ง่ายต่อการดูแลและแก้ไข โดย Magrittr มีตัวปฏิบัติการที่เรียกว่า “Pipe operator” ซึ่งทำหน้าที่เสมือนท่อลำเลียงที่ช่วยในการส่งค่าไปยังโค้ดหรือฟังก์ชันเรียกที่อยู่ถัดไป โดยพื้นฐานแล้ว ข้อมูลที่อยู่ทางด้านซ้ายมือ (Left-Handed Side: LHS) จะถูก pipe หรือถูกส่งต่อไปเป็น Argument แรกของฟังก์ชันที่อยู่ทางด้านขวามือ (Right-Hand Side: RHS) โดยแพกเกจ Magrittr สามารถช่วยปรับปรุงโค้ดได้ดังนี้

1) ช่วยจัดการโครงสร้างการปฏิบัติการข้อมูลให้อยู่ในรูปแบบจากซ้ายไปขวา
2) หลีกเลี่ยงการเรียกใช้ฟังก์ชันที่ซ้ำซ้อน (Nested Function Calls)
3) ลดการใช้ตัวแปรและฟังก์ชัน
4) ง่ายต่อการเพิ่มขั้นตอนหรือเพิ่มโค้ดในการจัดการข้อมูล

ตัวอย่างการใช้งานเบื้องต้น

x %>% f มีค่าเท่ากับ f(x) หรือการใส่ข้อมูล x เข้าไปในฟังก์ชัน

x %>% f(y) มีค่าเท่ากับ f(x, y) หรือการใส่ข้อมูล x และ y เข้าไปในฟังก์ชัน

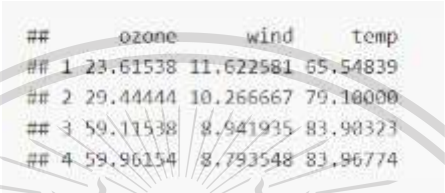
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

x %>% f %>% g %>% h มีค่าเท่ากับ h(g(f(x))) หรือการใส่ข้อมูลและเข้าไปในฟังก์ชัน และนำข้อมูลที่ได้จากฟังก์ชัน f(x) ไปใส่ในฟังก์ชัน g() และนำค่าที่ได้จากฟังก์ชัน g() ไปใส่ในฟังก์ชัน h()

2.2.3.5 Reshape2

Reshape2 เป็นแพ็กเกจที่ถูกเขียนขึ้นโดย Hadley Wickham เพื่อใช้ในการแปลงรูปแบบข้อมูลระหว่างรูปแบบกว้าง (Wide format) และรูปแบบยาว (Long format)

Wide format: เป็นข้อมูลที่มีคอลัมน์ของแต่ละตัวแปร



```
##      ozone     wind     temp
## 1 23.61538 11.622581 65.54839
## 2 29.44444 10.266667 79.10000
## 3 59.11538  8.941935 83.90323
## 4 59.96154  8.793548 83.96774
```

รูป 2.12 ตัวอย่างข้อมูลแบบ Wide format

Long format: เป็นข้อมูลที่ไม่มีคอลัมน์ของตัวแปร ใช้ทุกส่วนเป็นตัวแปรในการวัดค่า โดยคอลัมน์ที่ปรากฏจะเป็นคอลัมน์ที่ใช้บอกประเภทของตัวแปร

```
##    variable    value
## 1     ozone 23.61538
## 2     ozone 29.44444
## 3     ozone 59.11538
## 4     ozone 59.96154
## 5      wind 11.62258
## 6      wind 10.26667
## 7      wind  8.94193
## 8      wind  8.79355
## 9      temp 65.54839
## 10     temp 79.10000
## 11     temp 83.90323
## 12     temp 83.96774
```

รูป 2.13 ตัวอย่างข้อมูลแบบ Long format

ในการวิเคราะห์ข้อมูลส่วนมากต้องมีการแปลงรูปแบบข้อมูลจาก Wide format เป็นแบบ Long format เพื่อนำไปใช้ต่อไป เช่น ในแพ็กเกจ ggplot2 ที่ต้องใช้ข้อมูลแบบ Long format เพื่อนำไปใช้ในการพล็อตกราฟในรูปแบบต่าง ๆ เป็นต้น หรืออาจใช้แปลงข้อมูลจาก Long format เป็นแบบ Wide format เพื่อความสะดวกในการอ่านในส่วนของฟังก์ชันที่ใช้ โดยใช้ฟังก์ชัน melt() แปลงข้อมูลจาก Wide format เป็น Long format

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โปรแกรม 2.10 การใช้ (melt) แปลงข้อมูลจาก Wide format เป็น Long format

```
melt(data, ..., na.rm = FALSE, value.name = "value")
```

Arguments

1) data: ชุดข้อมูลที่ต้องการแปลง
2) ... : Argument อื่น ๆ ที่ใช้
3) na.rm: ลบข้อมูลที่มีค่าเป็น NA ในชุดข้อมูล
4) value.name: ชื่อของตัวแปรที่ใช้สำหรับเก็บค่า

2.2.3.6 ggplot2

ggplot2 เป็นระบบที่ช่วยในการสร้างกราฟิกจากชุดข้อมูล โดยมีพื้นฐานจากไวยากรณ์กราฟิก (The Grammar of Graphics) สามารถใช้ในการสร้างกราฟิกออกมาได้หลากหลายรูปแบบ ปรับแต่งองค์ประกอบของกราฟ เช่น ขนาด สี ได้ โดยเราสามารถป้อนข้อมูล และเขียนคำสั่งให้แพ็กเกจ ggplot2 ว่าต้องการจับคู่ตัวแปรในการพล็อตกราฟอย่างไร บอกรูปแบบทางด้านกราฟิกพื้นฐานที่ต้องการใช้ และใส่รายละเอียดของกราฟเพิ่มเติมได้ ฟังก์ชันที่ใช้ในการทำงาน ใช้ในการสร้าง ggplot object โดยการป้อนข้อมูล Data frame ที่ต้องการใช้สำหรับสร้างกราฟิก และบอกเซ็ตของตัวแปรที่ต้องการนำมาพล็อต

โปรแกรม 2.11 การใช้ ggplot

```
ggplot(data = NULL, mapping = aes(), ...)
```

Arguments

1) data: ชุดข้อมูลที่ต้องการนำมาพล็อต โดยชุดข้อมูลที่ใช้ต้องอยู่ในรูปแบบ Data Frame หากชุดข้อมูลไม่อยู่ในรูปแบบของ Data Frame จะถูกนำไปแปลงให้เป็น Data Frame โดยใช้ฟังก์ชัน fortify() ที่อยู่ในแพ็กเกจ ggplot2 หรือหากไม่ได้ระบุชุดข้อมูลในฟังก์ชันนี้ จะต้องมีชุดข้อมูลจากเลเยอร์อื่นเพิ่มเข้ามาในการพล็อต
2) mapping: รายการของคู่พิกัดที่ใช้ในการพล็อต โดย aes() เป็นการกำหนดแกนในการพล็อต เช่น aes(display, hwy) หมายถึงกำหนดให้แกน x เป็นค่าของตัวแปร display และแกน y เป็นค่าของตัวแปร hwy หากไม่ได้ระบุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

พิกัดที่ใช้ในฟังก์ชันนี้ จะต้องมีข้อมูลพิกัดจากเลเยอร์อื่นเพิ่มเข้ามาในการพล็อต

3) … : Argument อื่น ๆ ที่ใช้

ส่วนมากการใช้งานฟังก์ชันนี้ มักจะตามด้วยเครื่องหมาย '+' เพื่อเพิ่มคอมโพเนนต์ในการพล็อต

2.2.3.7 Plyr

Plyr เป็นแพ็กเกจที่เป็นชุดเครื่องมือในการแก้ไขปัญหาทั่วไป สามารถแบ่งส่วนข้อมูลเพื่อลดขนาดของปัญหาที่ใหญ่ให้มีขนาดเล็กลง จัดการกับข้อมูลแต่ละส่วนได้ และสามารถนำข้อมูลที่จัดการเรียบร้อยแล้วรวบรวมกลับเป็นกลุ่มข้อมูลได้สามารถประยุกต์ใช้กับข้อมูลได้หลากหลายรูปแบบ เช่น การทำการสรุปข้อมูล หรือการแยกชิ้นส่วนอาเรย์หลายมิติให้อยู่ในรูปแบบเชิงสถิติที่เข้าใจง่าย การพัฒนาของแพ็กเกจ Plyr นั้นได้รับการสนับสนุนโดย Becton Dickinson ฟังก์ชันที่ใช้ในการทำงานคือฟังก์ชัน arrange() ใช้สำหรับเรียงข้อมูลใน Data frame โดยใช้คอลัมน์

โปรแกรม 2.12 การใช้งาน arrange()

```
arrange(df, ...)
```

2.2.3.8 ddply()

ใช้แบ่ง Data frame ออกเป็นส่วน เพื่อนำไปใช้ในฟังก์ชันเพื่อจัดการข้อมูลเฉพาะส่วน และทำการรวมผลลัพธ์ที่ได้เข้าไปใน Data frame อีกครั้ง

โปรแกรม 2.13 การใช้งาน ddply()

```
ddply( .data, .variables, .fun = NULL, ..., .progress =
"none", .inform = FALSE, .drop = TRUE, .parallel =
FALSE, .paropts = NULL )
```

Arguments

1) .data: Data Frame ที่ต้องการดำเนินการ
2) .variable: ตัวแปรที่ต้องการแบ่ง
3) .fun: ฟังก์ชันที่ต้องการใช้กับข้อมูลที่แบ่ง
4) ... : Argument อื่นที่เพิ่มเข้ามาใน .fun

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

5) .progress: ชื่อของแถบแสดงความคืบหน้า (Progress Bar) ที่ต้องการให้แสดง

6) .inform: ต้องการให้แสดงข้อความเมื่อเกิดข้อผิดพลาดขึ้นหรือไม่ (Error Message) โดยค่าพื้นฐานจะถูกปิดไว้ เนื่องจากจะไปลดความเร็วในการรัน แต่มีประโยชน์อย่างมากในการใช้ Debug

2.2.3.9 sqldf

sqldf เป็นแพ็กเกจสำหรับรันคำสั่ง SQL บน Data frame โดยสามารถใช้ได้กับฐานข้อมูลแบบ SQLite, H2, PostgreSQL หรือ MySQL และใช้คำสั่งใน SQL เพื่อจัดการกับ data frame ได้โดยที่ข้อมูลไม่ต้องอยู่ในรูปแบบตาราง ใช้เวลาเร็วกว่าการคำนวณด้วยภาษา R พื้นฐาน ฟังก์ชันที่ใช้ คือ ฟังก์ชัน sqldf() ใช้ในการเลือกข้อมูลใน Data frame โดยใช้คำสั่ง SQL โดยทั่วไปใช้

โปรแกรม 2.14 การใช้งาน mutate()

```
sqldf(x)
```

Arguments

1) x: คำสั่งใน SQL ที่ต้องการใช้กับ Data Frame

2.2.3.10 shiny

shiny สามารถสร้าง Dashboard หรือการแสดงผลลัพธ์เป็นเว็บแอปพลิเคชันที่สามารถโต้ตอบได้ในภาษา R โดยโครงสร้างประกอบไปด้วย 3 ส่วนหลัก คือ

1. User interface (UI) คือ ส่วนที่คอยรับข้อมูลและแสดงผลให้ผู้ใช้งานได้ โดยการเขียนเป็น Object และสามารถเพิ่มข้อความ รูปภาพ หรือการเขียนภาษาอื่นๆแทรกเพิ่มเพื่อความสวยงามของหน้าแสดงผล
2. Server function คือ ส่วนที่ใช้มนการคำนวณและส่งผลลัพธ์กลับไปเพื่อนำไปแสดงผล
3. ShinyApp function คือ ส่วนที่รวบรวมทั้ง UI และ Server function ไว้ด้วยกัน ถ้าหากไม่มีการเรียกใช้จะไม่สามารถทำงานได้

2.2.3.11 lsoda

ฟังก์ชันสำหรับแก้สมการเชิงอนุพันธ์สามัญ (ODE) การสลับโดยอัตโนมัติระหว่างวิธีการ stiff และ non-stiff การแก้ปัญหาค่าเริ่มต้นสำหรับระบบที่ stiff หรือ non-stiff ของสมการเชิงอนุพันธ์สามัญลำดับที่หนึ่ง (ODEs)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โปรแกรม 2.13 การใช้งาน lsoda()

```
lsoda(y, times, func , parms)
```

Arguments

1) y ค่าเริ่มต้นสำหรับระบบ ODE ถ้า y มีแอตทริบิวต์ name ชื่อจะถูกใช้เพื่อติดกำกับเมทริกซ์เอาต์พุต
2) times เวลาที่ต้องการประมาณการอย่างชัดเจนสำหรับ y ค่าแรกในครั้งต้องเป็นเวลาเริ่มต้น
3) func ฟังก์ชัน R ที่คำนวณค่าของอนุพันธ์ในระบบ ODE (นิยามแบบจำลอง) ณ เวลา t หรืออักขระแสดงชื่อของฟังก์ชันที่คอมไพล์แล้วในไลบรารีที่โหลดแบบไดนามิก
4) parms เวกเตอร์หรือรายการพารามิเตอร์ที่ใช้ใน func หรือ jacfunc

2.2.4 Profiling

Profiling เป็นทฤษฎีในการวิเคราะห์โปรแกรมเพื่อตรวจสอบว่าแต่ละส่วนของโปรแกรมที่ใช้เวลาในการดำเนินการเท่าไหร่ อย่างไร ส่วนใดใช้เวลามากที่สุด แบ่งเป็นสัดส่วนต่อโปรแกรมโดยรวมได้อย่างไร ซึ่งสามารถใช้เป็นแนวทางในการพัฒนาเพื่อเพิ่มความเร็วและประสิทธิภาพของโปรแกรม โดยใน R จะมีกลไกการ Profiling เป็นแบบสุ่ม (Sampling) ซึ่งเป็นการบันทึกข้อมูลเกี่ยวกับฟังก์ชันเรียก Stack ในแต่ละช่วงเวลา หากมีข้อมูลในช่วงเวลานั้น ๆ จะมีโค้ดที่ทำงานในช่วงเวลานั้นปรากฏในจุดที่ทำการสุ่ม (Sampling Point) ถูกเก็บไว้ หรือกระทั่งข้อมูลเกี่ยวกับช่วงเวลาที่ใช้ในตัวจัดการขยะ (GC) ก็อาจถูกเก็บด้วยเช่นกัน ข้อมูล Profiling ที่ถูกเก็บจะถูกเขียนลงบนไฟล์ โดยเริ่มต้นจะถูกเก็บในไฟล์ชื่อ Rprof.out และใช้ฟังก์ชัน summaryRprof ที่อยู่ใน base R ในการสร้าง Interface อย่างง่ายในการตรวจสอบข้อมูลดังกล่าว ในส่วนของแพ็กเกจ จะใช้แพ็กเกจ profvis ในการสร้างส่วนแสดงผล (Visualization) รวมถึงข้อมูลในส่วนอื่น ๆ สำหรับเครื่องมือที่ใช้ในการ Profiling จะเรียกว่า Profiler

2.2.4.1 การใช้ Profiler

Profiler โดยพื้นฐานจะแสดงข้อมูล 2 ส่วน คือ

1) Memory (MB): หน่วยความจำที่ใช้ในการเรียกฟังก์ชันนั้น ๆ มีหน่วยเป็นเมกะไบต์
2) Time (ms): เวลาที่ใช้ในการเรียกฟังก์ชันนั้น ๆ มีหน่วยเป็นมิลลิวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สำหรับการใช้งาน Profiler จะมีอยู่ทั้งหมด 2 แท็บด้วยกัน ได้แก่

1) Flame Graph

2) Data

2.2.4.2 การใช้งานแถบ Flame graph

Flame graph คือส่วนที่แสดงการโค้ดในแต่ละส่วนออกเป็นบล็อก แต่ละบล็อกใน Flame graph แสดงถึงการเรียกไปยังฟังก์ชัน หรือการเรียกซ้ำของฟังก์ชันเดิม ความกว้างของบล็อกจะเป็นสัดส่วนเดียวกับเวลาที่ใช้ในการเรียกฟังก์ชันนั้น ๆ กล่าวคือ หากบล็อกนั้นใช้เวลาในการเรียกฟังก์ชันค่อนข้างมาก ความกว้างของบล็อกนั้นจะกว้าง ในทางกลับกัน หากโค้ดในบล็อกนั้นใช้เวลาในการเรียกฟังก์ชันน้อย ความกว้างของบล็อกก็จะน้อยตามไปด้วย และเมื่อฟังก์ชันมีการเรียกฟังก์ชันอื่นร่วมด้วย จะมีบล็อกอีกบล็อกเพิ่มที่ด้านบนของ Flame graph นั้น ๆ

รูป 2.14 ข้อมูลที่แสดงในแถบ Flame graph

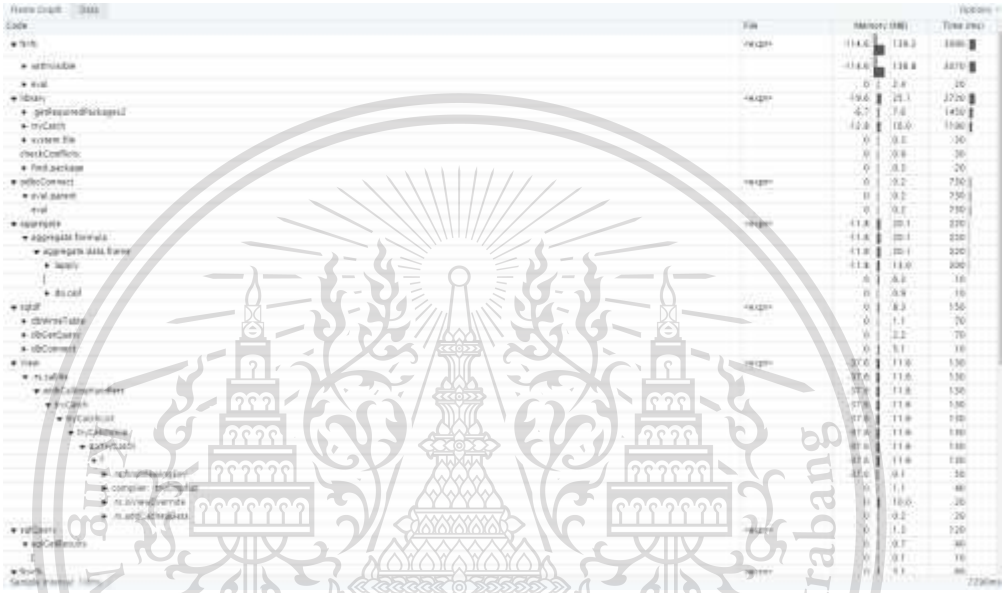
เนื่องจากในแถบ Flame graph มีการทำงานเชิงโต้ตอบ (Interactive) จึงสามารถใช้งานได้หลากหลาย เช่น สามารถทราบข้อมูลโดยละเอียดของแต่ละบล็อกได้โดยการนำเมาส์ไปวางบนบล็อกที่ต้องการดูรายละเอียด หรือหากนำเมาส์วางบนบล็อกสีเหลืองจะแสดงโค้ดที่มีความเกี่ยวข้องกับการเรียกฟังก์ชันนั้น เป็นต้น

2.2.4.3 การใช้งานแถบ Data

แถบ Data จะแสดงมุมมองในรูปแบบของตารางจากบนลงล่าง โดยสามารถคลิกที่คอลัมน์ ‘Code’ เพื่อขยายส่วนเรียก Stack ซึ่งสามารถใช้ในการวิเคราะห์ได้ และมีคอลัมน์ที่แสดงถึงการจัดสรรทัพยากร (Resource allocation) ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1) Memory: แสดงหน่วยความจำที่ถูกใช้งาน (Allocated) หรือถูกยกเลิกการจัดสรร (Deallocated) โดยหน่วยความจำที่ถูกยกเลิกการจัดสรรจะแสดงเป็นจำนวนลบ และตัวเลขที่แสดงเป็นหน่วยความจำทั้งหมดที่ใช้ในการเรียกฟังก์ชันในบรรทัดนั้น ๆ มีหน่วยเป็นเมกะไบต์
2) Time: เวลาที่ใช้ในหน่วยมิลลิวินาที ซึ่งเป็นเวลารวมที่ใช้เรียกฟังก์ชัน ในโค้ดบรรทัดนั้น ๆ



รูป 2.15 ข้อมูลที่แสดงในแถบ Data

2.2.4.4 การเลือกข้อมูลสำหรับการ Profiling

สามารถอธิบายการเลือกข้อมูลในการ Profiling ได้ ดังนี้

1) Profiler จะใช้ Rprof ในการเลือกข้อมูล (เป็นส่วนหนึ่งของ R distribution)
2) ในแต่ละ Time interval (default : 10 ms)
3) Profiler จะหยุดการทำงานของ R Interpreter ก่อน
4) ดูที่ฟังก์ชันเรียก Stack ปัจจุบัน
5) บันทึกเวลาและหน่วยความจำที่ Stack นั้นใช้
6) ทำงานเป็นแบบสุ่มตัวอย่าง (Sampling) ทำให้ผลที่ได้ไม่มีค่าตายตัว แต่ละครั้งที่ทำการ Profiling ค่าอาจจะต่างกันเล็กน้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

7) ผลลัพธ์ในส่วนของหน่วยความจำนั้นได้มาจากพฤติกรรมแบบไม่ตายตัว (Non-deterministic behavior) ของตัวจัดการขยะ ซึ่งทำให้การยกเลิกการจัดในหน่วยความจำค่อนข้างเป็นแบบสุ่ม

2.2.5 SQL

SQL หรือ Structured Query Language เป็นภาษามาตรฐานในการเข้าถึง ฐานข้อมูล ภาษา SQL ออกแบบมาเพื่อทำการจัดการข้อมูลที่อยู่ใน Relational Database Management System (RDBMS) หรือไว้สำหรับค้นหาข้อมูล เปลี่ยนแปลง เพิ่มและลบข้อมูลที่ถูกเก็บอยู่ในฐานข้อมูลในรูปแบบตารางที่มีลักษณะเป็น Column และ Row เรียกข้อมูลเหล่านี้ว่าถูกเก็บอยู่ใน Table ด้วยความสามารถของ SQL ยังสามารถสร้างตารางขึ้นมาใหม่ ลบตาราง และเปลี่ยนแปลงค่า ของ Table ได้ สามารถใช้งานภาษา SQL ได้จากโปรแกรมต่าง ๆ ที่ต้องทำการกับระบบฐานข้อมูล เช่น ใช้ SQL ในการทำการดึงข้อมูล จากฐานข้อมูลและเป็นมาตรฐานกลางที่ใช้ในระบบฐานข้อมูลต่าง ๆซึ่งเป็นมาตรฐานของ ANSI (American National Standard Institute)

2.2.5.1 SQL syntax

ภาษา SQL จะแบ่งออกมาเป็นส่วนประกอบต่าง ๆ เรียกส่วนเหล่านี้ตามรูปแบบเช่น Clauses คือองค์ประกอบหนึ่งของ Statement และ Query และ Expressions คือการสร้างผลลัพธ์ออกมาในรูป Table ที่ประกอบด้วย Column และ Row จากข้อมูล ส่วน Predicates คือรูปแบบเงื่อนไขที่มีผลลัพธ์เป็น True / False / Unknown หรือที่เรียกกันว่า Boolean Queries เป็นส่วนสำคัญใน SQL Statements เพราะมีผลต่อโครงสร้างข้อมูล การจัดการข้อมูล Transactions, Program Flow, Session หรือแม้กระทั่งการวิเคราะห์ปัญหา โดยจำเป็นต้องจบด้วย Semicolon (;) ซึ่งเป็นสิ่งจำเป็นที่ต้องมีทุกครั้งสำหรับ SQL Insignificant Whitespace หรือช่องว่างสำหรับใน SQL statement และ Query จะไม่สนใจ ทำให้ SQL สามารถเขียนในรูปแบบที่หลากหลาย

2.2.5.2 ประเภทของ SQL Query

โปรแกรม 2.15 Select Query ใช้ดึงข้อมูลในรูป Table คือเป็น Row และ Column

```
SELECT column1, column2, ...
FROM table_name;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โปรแกรม 2.16 Update Query ใช้แก้ไขข้อมูลที่มีอยู่แล้วใน Table

```
UPDATE table_name
SET column1 = value1, column2 = value2, ...
WHERE condition;
```

โปรแกรม 2.17 Insert Query ใช้เพิ่มข้อมูลใน Table

```
INSERT INTO table_name (column1, column2, column3,...)
VALUES (value1, value2, value3, ...);
```

โปรแกรม 2.18 Delete Query ใช้ลบข้อมูลใน Table

```
DELETE FROM table_name
WHERE condition;
```

2.2.5.3 ข้อมูล SQL command เบื้องต้น

SQL – Data Types เป็นการกำหนดชนิดค่าตัวแปรในตารางข้อมูล เช่น ข้อมูลตัวเลข ตัวอักษร วันเวลา หรือแบบไม่มีโครงสร้าง ซึ่งสิ่งเหล่านี้จำเป็นตั้งแต่เริ่มสร้างตารางฐานข้อมูล ข้อมูลที่ใส่ลง Table มีความถูกต้องตามที่วางไว้ ยังช่วยให้ฐานข้อมูลทำงานได้ง่ายขึ้นในการจัดเก็บ โดย Data types บนฐานข้อมูลมีหลายชนิด ขึ้นอยู่กับชนิดของฐานข้อมูล SQL – Operator เป็นการกำหนดกระบวนการในการเลือกข้อมูลโดยระบบฐานข้อมูล จะมีการจอง Key Word บางคำหรือ ตัวอักษรไว้สำหรับให้ SQL Statement ใช้ระบุเงื่อนไขเพื่อเป็นการสั่งระบบให้ทำตามเงื่อนไขที่วางไว้ เช่น การเปรียบเทียบ การคำนวณทางคณิตศาสตร์ การทำกระบวนการเหล่านี้มักจะใช้ระบุเงื่อนไขภายใต้ SQL Statement และสามารถกำหนดได้หลายเงื่อนไขภายใน Statement นั้น SQL Expression เป็นการสร้างเงื่อนไขจากกระบวนการเพื่อเลือกข้อมูล การรวมเอาค่าตัวแปรหรือข้อมูลตั้งแต่ 1 ตัวขึ้นไป รวมถึง SQL Operator ส่วน SQL Expression เป็นเหมือนประโยคในภาษาเขียนซึ่งใช้ในการดึงข้อมูล (Query) ทำให้สามารถเลือกข้อมูลที่ต้องการจากฐานข้อมูลได้

2.2.5.4 ประเภทของคำสั่ง SQL

ภาษานิยามข้อมูล (Data Definition Language : DDL) เป็นคำสั่งที่ใช้ในการสร้างฐานข้อมูล กำหนดโครงสร้างข้อมูลว่ามีแอตทริบิวต์ใดชนิดของข้อมูล รวมทั้งการเปลี่ยนแปลงตาราง และการสร้างดัชนี

คำสั่ง : CREATE, DROP, ALTER

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาษาจัดการข้อมูล (Data Manipulation Language :DML) เป็นคำสั่งที่ใช้ในการเรียกใช้ เพิ่ม ลบ และเปลี่ยนแปลงข้อมูลในตาราง

คำสั่ง : SELECT, INSERT, UPDATE, DELETE

ภาษาควบคุมข้อมูล (Data Control Language : DCL) เป็นคำสั่งที่ใช้ในการกำหนดสิทธิการอนุญาติ หรือ ยกเลิก การเข้าถึงฐานข้อมูล เพื่อป้องกันความปลอดภัยของฐานข้อมูล

คำสั่ง : GRANT, REVOKE

2.2.6 ฟังก์ชันซิกมอยด์ (Sigmoid Function)

ฟังก์ชันซิกมอยด์เป็นฟังก์ชันทางคณิตศาสตร์ที่เป็น Curve รูปตัว S หรือเส้นโค้งซิกมอยด์ และเนื่องจากเอาท์พุตของฟังก์ชันซิกมอยด์มีค่าระหว่าง 0 – 1 จึงเหมาะที่จะถูกใช้ในงานที่ต้องการเอาท์พุต เป็นความน่าจะเป็น (Probability) หรือใช้เป็นเอาท์พุตว่า 1 = Yes, 0 = No ตัวอย่างที่นิยมทั่วไปของฟังก์ชันซิกมอยด์ คือฟังก์ชันโลจิสติก ดังกราฟที่แสดงต่อไปนี้

รูป 2.16 กราฟ Logistic Curve

มีการใช้ฟังก์ชันซิกมอยด์ที่หลากหลายฟังก์ชันเป็นที่นิยมได้แก่ฟังก์ชันโลจิสติกส์ (Logistic function) และไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic tangent) โดยมีสมการดังต่อไปนี้

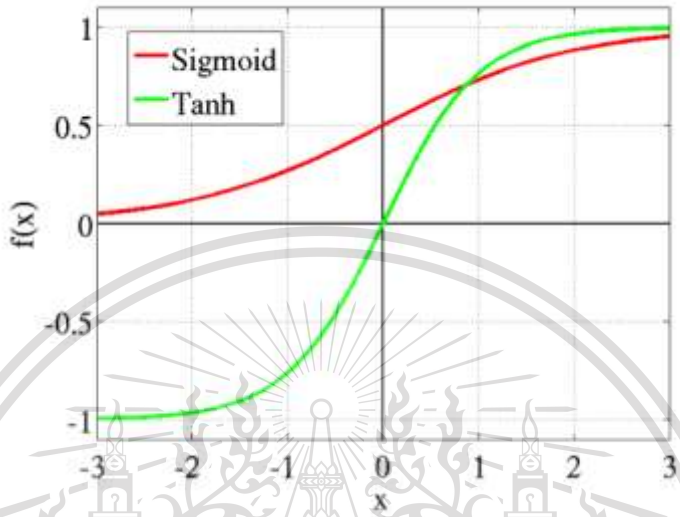
1) Logistic function เป็นฟังก์ชันที่เปลี่ยนค่าทั้งหมดบนเส้นจำนวนให้กลายเป็นช่วง (0, 1)

$$\text{Sigmoid(x)} = \frac{1}{1+e^{-x}} \tag{2.2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2) Hyperbolic tangent ทำงานคล้าย Sigmoid แต่เปลี่ยนช่วงของเอาท์พุตเป็น (-1, 1)

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (2.3)$$



รูป 2.17 กราฟ N ฟังก์ชันซิกมอยด์และ tanh

ข้อดีของฟังก์ชันซิกมอยด์

1) เอาท์พุตอยู่ในช่วง 0-1 ใช้งานได้หลากหลาย ทั้ง Probability และ Boolean
2) เข้าใจง่าย
3) หา Derivative ไม่ยาก

ข้อเสียของฟังก์ชันซิกมอยด์

1) ถ้าอินพุตน้อยกว่า -5 หรือมากกว่า 5 ความชัน Slope จะเข้าใกล้ 0 จน Gradient หายไปหมด ทำให้โมเดล Train ไม่ไปไหน เรียกว่า Vanishing Gradient จะอธิบายต่อไป
2) เอาท์พุตไม่สมดุล มี Mean ไม่เท่ากับ 0 เพราะมีแต่ค่าเป็นบวก ทำให้ Optimize ยาก
3) ใช้แล้วโมเดลมักจะ Converge ช้า

2.2.7 สหสัมพันธ์ (Correlation)

สหสัมพันธ์ (Correlation) เป็นการหาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป เพื่อพิจารณาระหว่างตัวแปรว่ามีมากน้อยแค่ไหน ค่าที่วัดความสัมพันธ์ เรียกว่าค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ในการวัดความสัมพันธ์แต่ละประเภทนั้น จำเป็นจะต้องมีการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ทดสอบนัยสำคัญก่อนเสมอ สรุปได้ว่าตัวแปรคู่ใดมีความสัมพันธ์มากน้อยเพียงใด แต่ไม่สามารถระบุได้ว่าตัวแปรใดเป็นตัวแปรต้นหรือตัวแปรตาม

ค่าสัมประสิทธิ์สหสัมพันธ์ ใช้สัญลักษณ์ r แทนสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง และแทนสัมประสิทธิ์สหสัมพันธ์ของประชากร ค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้วัดขนาดของความสัมพันธ์กันระหว่างตัวแปร มี 2 แบบ คือระหว่าง -1 ถึง 1 และ ระหว่าง 0 ถึง 1

การบอกขนาดความสัมพันธ์จะใช้ตัวเลขของค่าสัมประสิทธิ์สหสมพันธ์หากค่าสัมประสิทธิ์สหสัมพันธ์นั้นเข้าใกล้ -1 หรือ 1 แสดงว่ามีความสัมพันธ์กันในระดับสูงแต่หากเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันระดับน้อย หรือแทบไม่มีเลย สำหรับการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ มีเกณฑ์การพิจารณาดังนี้ (อ้างอิงตาม Hinkle D.E. 1998)

ตาราง 2.1 การพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์

ค่า r	ระดับความสัมพันธ์
0.90 - 1.00	ความสัมพันธ์ระดับสูงมาก
0.70 - 0.90	ความสัมพันธ์ระดับสูง
0.50 - 0.70	ความสัมพันธ์ระดับปานกลาง
0.30 - 0.50	ความสัมพันธ์ระดับต่ำ
0.00 - 0.30	ความสัมพันธ์ระดับต่ำมาก

เครื่องหมาย+,-หน้าตัวเลขสัมประสิทธิ์สหสัมพันธ์แสดงถึงทิศทางของความสัมพันธ์

เครื่องหมาย + แสดงว่า มีความสัมพันธ์กันไปในทิศทางเดียวกัน

เครื่องหมาย - แสดงมีความสัมพันธ์กันไปในทิศทางตรงกันข้าม

ค่าสัมประสิทธิ์สหสัมพันธ์จะใช้ได้อย่างเหมาะสมกับข้อมูลที่มีความสัมพันธ์เชิงเส้นเท่านั้น ดังนั้น ถ้าหากค่า r = 0 หมายความว่าข้อมูลไม่มีความสัมพันธ์กันนั้นไม่ถูกต้อง เนื่องจากอาจเป็นไปได้ว่าข้อมูลมีความสัมพันธ์กันในลักษณะไม่ใช่เชิงเส้นแต่เป็นลักษณะอื่น

2.2.8 การแสดงผล

1) Heatmap คือ การแสดงภาพของข้อมูลแนวโน้มหรือความสัมพันธ์อย่างง่าย โดยจะแสดงออกมาในรูปแบบของ “คู่สี” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่หรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ปริมาณของข้อมูลที่เกิดขึ้นโดยพื้นฐานของแผนภูมินี้ไม่จำเป็นต้องระบุตัวเลขลงไปในแต่ละช่องของตารางทำ Heatmap มาเพื่อแสดงให้เห็นชัดเจนว่าข้อมูลแต่ละส่วนมีค่ามากน้อยเพียงใด สามารถนำมาวิเคราะห์ประเมินความสัมพันธ์ระหว่างข้อมูลว่าไปในทิศทางเดียวกัน ตรงกันข้าม หรือไม่เกี่ยวข้องกัน

2) Bar chart หรือกราฟแท่ง ความสูงของกราฟจะเปลี่ยนแปลงตามจำนวนของข้อมูล แต่ความกว้างของกราฟแต่ละแท่งเท่ากัน และจำเป็นต้องมีการแบ่งของกราฟแต่ละแท่งอย่างชัดเจน ใช้สีหรือแรเงาเพื่อให้กราฟแท่งแสดงถึงข้อมูลแต่ละชุด กราฟชนิดนี้ใช้เพื่อเปรียบเทียบข้อมูลที่มีความหลากหลาย ความสูงของกราฟแต่ละแท่งแสดงถึงความต่างของข้อมูลประเภทนั้น ๆ

3) Stack bar chart คือการนำเสนอข้อมูลในรูปแบบการนำข้อมูลมาต่อกัน หรือนำมาเปรียบเทียบ เนื่องจากสามารถทำให้ผู้อ่านสามารถเห็นถึงความแตกต่างของจำนวนข้อมูลได้ง่ายและเป็นการแสดงถึงผลรวมของข้อมูล เหมาะสำหรับการนำเสนอที่แสดงให้เห็นยอดรวมทั้งหมดของข้อมูลอีกด้วย

4) Histogram คือกราฟแท่งแสดงความสัมพันธ์ระหว่างข้อมูลเพื่อช่วยกระจายข้อมูลสามารถเห็นความถี่ได้อย่างชัดเจน แต่ละแท่งของกราฟมีความกว้างเท่ากัน ส่วนความสูงของแท่งจะขึ้นอยู่กับความถี่ของข้อมูลทำให้สามารถตีความข้อมูลได้ง่ายจึงเหมาะแก่การเปรียบเทียบข้อมูล

5) Line chart คือกราฟเส้นที่สามารถแสดงแนวโน้มของข้อมูลได้อย่างชัดเจน เหมาะสำหรับข้อมูลที่อยู่ในรูปแบบช่วงเวลา ใช้เพื่อดูแนวโน้มของข้อมูลและทำนายข้อมูลต่อไปในอนาคต สามารถนำเอาไปใช้กับกราฟแท่งหรือกราฟประเภทอื่นได้ตามความเหมาะสม

6) Scatter plot คือกราฟที่ใช้เพื่อแสดงให้เห็นความสัมพันธ์ของข้อมูลหรือการกระจายของข้อมูลเป็นกลุ่มก้อน และสามารถบอกได้ว่าค่าของทั้ง 2 แกนมีความสัมพันธ์กันอย่างไร ซึ่งทำให้สามารถจำแนกข้อมูลได้อย่างชัดเจน สำรวจข้อมูลได้ง่ายและสามารถตรวจสอบการกระจายของข้อมูลได้

2.2.9 ค่าระดับการติดเชื้อ (Reproduction Number)

ค่าระดับการติดเชื้อ (Ro) คือค่าเฉลี่ยที่ผู้ป่วยหนึ่งคนจะสามารถแพร่เชื้อให้ผู้คนกลุ่มเสี่ยงเป็นจำนวนกี่คนในช่วงเวลาที่ยังติดเชื้ออยู่ หรืออีกนัยหนึ่งคือค่าที่บ่งบอกว่าไวรัสมีความสามารถการแพร่ระบาดได้มากน้อยเพียงใด โดยการกำหนดค่าระดับการติดเชื้อ (Ro) ถือเป็นเป้าหมายหลักของการศึกษาการแพร่ระบาดของไวรัส

โดยถ้าค่าระดับการติดเชื้อ (Ro) < 1 จะหมายถึง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

- โดยเฉลี่ยผู้ป่วยหนึ่งคน สามารถแพร่ไวรัสให้แก่ผู้คนในกลุ่มเสี่ยงได้ไม่เกินหนึ่งคน

- การที่ค่าระดับการติดเชื้อ (Ro) น้อยกว่า 1 หมายถึงว่าไวรัสชนิดนี้สามารถหยุดการแพร่ระบาดไว้ได้

หากค่าระดับการติดเชื้อ (Ro) = 1 จะหมายถึง

- โดยเฉลี่ยผู้ป่วยหนึ่งคน สามารถแพร่ไวรัสให้แก่ผู้คนในกลุ่มเสี่ยงได้ประมาณหนึ่งคน

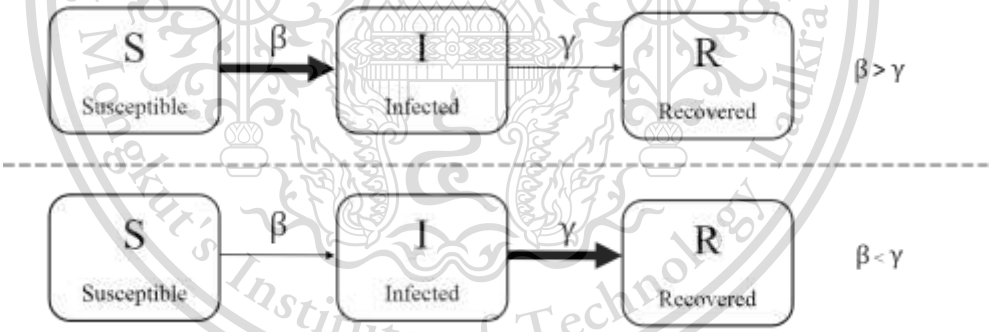
- การที่ค่าระดับการติดเชื้อ (Ro) เท่ากับ 1 นั้นหมายถึงว่าไวรัสชนิดนี้อยู่เข้าสภาวะสมดุล

และหากค่าระดับการติดเชื้อ (Ro) > 1 จะหมายถึง

- โดยเฉลี่ยผู้ป่วยหนึ่งคน สามารถแพร่ไวรัสให้แก่ผู้คนในกลุ่มเสี่ยงได้มากกว่าหนึ่งคน

- การที่ค่าระดับการติดเชื้อ (Ro) มากกว่า 1 นั้นหมายถึงว่าไวรัสชนิดนี้ยังมีการแพร่ระบาดที่เพิ่มขึ้นอยู่

ค่าระดับการติดเชื้อ (Ro) ของโรคระบาด จะถูกควบคุมด้วยค่าคงที่ β และ γ โดยสามารถพิสูจน์ได้ว่า ถ้า $\beta > \gamma$ แล้วจะทำให้เกิดโรคระบาด โดยที่ β คือค่าคงที่บอกถึงอัตราการการแพร่เชื้อ (Transmission rate) เป็นตัวกำหนดว่าประชากรจะติดเชื้อมากเพียงใด โดยการนำระยะเวลาที่ทำให้เกิดการติดเชื้อมาหาค่าเฉลี่ย ส่วน γ คืออัตราการหายจากอาการติดเชื้อ (Recovery Rate)



รูป 2.18 ปัจจัยที่กำหนดว่าโรคจะระบาด (Ro) ถูกควบคุมด้วยค่าคงที่ β และ γ

ในช่วงเริ่มต้นของการระบาด หาก $\beta > \gamma$ จำนวนผู้ป่วยจะเพิ่มเร็วกว่าคนที่หายป่วย ทำให้เกิดโรคระบาด แต่ถ้า $\beta < \gamma$ จำนวนคนที่หายป่วยจะเพิ่มเร็วกว่าจำนวนคนป่วย ทำให้โรคไม่ระบาดเพิ่ม เรียกอัตราส่วนระหว่าง β และ γ ว่า Ro หรือ Basic reproduction number

$$R_0 = \frac{\beta}{\gamma} \tag{2.4}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ในปัจจุบันนั้นค่าระดับการติดเชื้อ (Ro) ของโรคโควิด-19 นั้นอยู่ที่ประมาณ 2 – 3 โดยการกำหนดค่าระดับการติดเชื้อ (Ro) ที่ถูกต้องได้นั้นถือเป็นปัจจัยที่สำคัญที่สุดในการทำนายโรคระบาด การที่ไวรัสค่าระดับการติดเชื้อ (Ro) สูงนั้นเป็นปัจจัยที่ต้องคำนึงถึงเป็นอย่างมาก ค่าระดับการติดเชื้อ (Ro) นั้นเป็นค่าเฉลี่ยทำให้บางครั้งอาจมีค่าสูงเกินกว่าปกติได้เนื่องจากหลายปัจจัย

ดังนั้นค่าระดับการติดเชื้อ (Ro) สามารถแปรเปลี่ยนได้อยู่ตลอดเวลา และการติดตามผู้ป่วยทุกรายนั้นเป็นไปได้ยาก ทำให้การประมาณค่าระดับการติดเชื้อ (Ro) นั้นซับซ้อน และเป็นไปได้ยากที่จะคาดคะเนได้อย่างถูกต้อง เพราะค่าระดับการติดเชื้อ (Ro) จะเปลี่ยนไป เมื่อมีข้อมูลชุดใหม่เข้ามาส่งผล

2.2.10 Root Mean Square Error (RMSE)

Metrics สำหรับปัญหา Regression เราสร้างโมเดล Regression เพื่อทำนายตัวแปร y แบบ Continuous (ตัวเลข) สามารถเขียนสมการคำนวณค่า Error ได้ดังนี้

$$Error = prediction - actual \tag{2.5}$$

RMSE ย่อมาจาก “Root Mean Square Error” เป็นการใส่ Square root บนค่า MSE เพื่อให้ได้ค่า Loss ที่มีหน่วยเดียวกับตัวแปร y เหตุผลที่ต้องทำ Square root เนื่องจากมีการยกกำลังสองค่า Error ก่อนหาค่าเฉลี่ย ทำให้หน่วยเปลี่ยนไปจากเดิม

$$RMSE = \sqrt{\frac{1}{n} \times \sum (Error)^2} \tag{2.6}$$

RMSE มีค่ายิ่งต่ำยิ่งดี ถ้าเท่ากับ 0 แปลว่าโมเดลทำนายค่า y ได้ถูกต้อง 100% ในทางปฏิบัติโอกาสที่จะเทรนโมเดลได้ Loss = 0 เป็นไปได้ยากมาก เพราะอาจนำไปสู่ปัญหา Overfitting

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.3 เอกสารและงานวิจัยที่เกี่ยวข้อง

[1] ดร. ไกรชาติ ตันตระการอาภา และ พ.ท.ดร.นพ.ภพกฤต ภพธรอังกูร (2020) ศึกษาการแพร่กระจายของไวรัสโคโรนา 2019 ในประเทศไทย ศึกษากรณีผู้ติดเชื้อกับข้อมูลสภาพอากาศที่เกี่ยวข้อง และศึกษาการใช้แบบจำลองทางคณิตศาสตร์ SIR Model เพื่อทำนายผู้ติดเชื้อในอนาคต

[2] ผศ.ดร.อานนท์ ศักดิ์วรวิชญ์ (2018) การ Normalize ตัวแปรต่างจากการ Standardize ตัวแปร เนื่องจากการ Normalize เป็นการทำให้ตัวแปรอยู่บนบรรทัดฐานเดียวกันเพื่อให้ตัวแปรเดียวกันเปรียบเทียบกันได้ ส่วนการ Standardize ตัวแปรนั้นเป็นการทำให้ตัวแปรเป็นค่ามาตรฐาน เช่น ค่ามาตรฐาน z ที่ทำโดยหักออกด้วยค่าเฉลี่ยของตัวแปรแต่ละตัวหารด้วยส่วนเบี่ยงเบนมาตรฐานในแต่ละตัวแปร การ Standardize ต้องทำหลังจาก Normalize และการ Normalize ทำให้ตัวแปรเปรียบเทียบกันในตัวแปรเดียวกันข้ามแถวได้ ในขณะที่ Standardize ทำให้ตัวแปรมีสเกลเดียวกัน คือ มีค่าเฉลี่ยและความแปรปรวนเท่ากันทุกตัวแปร เพื่อให้สามารถเปรียบเทียบกันข้ามตัวแปรได้

[3] Andrew Atkeson (2020) งานวิจัยเกี่ยวกับ SEIR Model พฤติกรรมการแพร่ระบาดของไวรัสโคโรนา 2019 ในสหรัฐอเมริกาและสหราชอาณาจักรในช่วงเวลา 12 เดือน เพื่อหาปัจจัยที่ส่งผลต่ออัตราการสียชีวิตรายวันและอัตราการแพร่เชื้อ ตั้งแต่วันที่ 15 กุมภาพันธ์จนถึงปัจจุบันและได้ทำการคาดการณ์จำนวนผู้ติดเชื้อในอีกสองปีข้างหน้า

[4] Antoine Soetewey (2020) ศึกษา Top 35 R resources ที่ใช้ในการวิเคราะห์ข้อมูลเกี่ยวกับโรคติดเชื้อไวรัสโควิด 2019 เป็นแหล่งข้อมูลที่น่าเชื่อถือเหมาะสำหรับนำมาวิเคราะห์ ได้แก่ R Shiny apps และ dashboards ถูกพัฒนาโดย John Coene เป็นแอพติดตามการแพร่กระจายของโรคติดเชื้อไวรัสโคโรนา 2019 จากแหล่งข้อมูลสามแหล่ง (John Hopkins, Weixin และ DXY Data) นำเสนอข้อมูลจำนวนผู้เสียชีวิต ผู้อยู่ในกลุ่มเสี่ยงและผู้ป่วยที่ติดเชื้อตามเวลาและภูมิภาคต่าง ๆ, COVID-19 Global Cases, Visualization of Covid-19 Cases, COVID-19 Data Visualization Platform เป็นต้น

[5] Dan Connelly (2020) การสร้างแบบจำลองทางคณิตศาสตร์ SIR SEIR และ SEIRD การแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศเกาหลีใต้ การคำนวณสมการทางคณิตศาสตร์ การวิเคราะห์ค่าต่าง ๆ ที่คำนวณได้ และวิธีหาค่า Ro หรือ Basic reproductive number

[6] Jen Ciarochi (2019) ศึกษาการสร้างแบบจำลองทางคณิตศาสตร์เกี่ยวกับการแพร่กระจายโรคระบาดต่างๆรวมทั้งการแพร่ระบาดไวรัสโคโรนา 2019 โดยอ้างอิงข้อมูลที่ใช้จากองค์กรอนามัยโลก (WHO) วิธีหาค่า Ro หรือ Basic reproductive number และการสร้างแบบจำลองต่าง ๆ ได้แก่ SIR และ SEIR Model

[7] Mykyta Solonko (2020) ศึกษาเกี่ยวกับอายุและเพศของผู้ติดเชื้อไวรัสโควิด-19 โดยใช้ Rstudio วิเคราะห์อายุอัตราการเสียชีวิตจากผู้ติดเชื้อ ได้ผลสรุปว่าผู้ที่มีอายุมากจะมีโอกาสสูงกว่าผู้ที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

มีอายุน้อยกว่า โดยค่าเฉลี่ยของผู้ที่เสียชีวิตมีอายุประมาณ 68.6 ปี และผู้ติดเชื้อที่ยังมีชีวิตอยู่มีอายุเฉลี่ยประมาณ 48 ปี แสดงให้เห็นว่าคนที่อายุมากมีโอกาสเสี่ยงเสียชีวิตมากกว่า สำหรับอัตราการเสียชีวิตจากผู้ติดเชื้อเพศชายมีอัตราการเสียชีวิต 8.5% เทียบกับอัตราการเสียชีวิตในเพศหญิง 3.7% สรุปได้ว่าเพศชายมีโอกาสเสียชีวิตจากการติดเชื้อมากกว่าเพศหญิง

[8] Nasir Mustafa (2020) งานวิจัยทางสถิติเกี่ยวกับการแพร่ระบาดไวรัสโคโรนาสายพันธุ์ใหม่ SARS-CoV-2 ทั่วโลก วิเคราะห์หาปัจจัยที่ส่งผลต่ออัตราการติดเชื้อและอัตราการเสียชีวิตจากการติดเชื้อ ได้มีการเปรียบกับโรคอื่น ๆ ที่มีลักษณะคล้ายกัน แบ่งกลุ่มการติดเชื้อตามทวีป กลุ่มอายุ เพศ และอาการจากการติดเชื้อ

[9] Santhiya (2020) วิเคราะห์ผลกระทบของโรคติดต่อเชื้อไวรัสโคโรนา 2019 ทั่วโลก วิเคราะห์อัตราการตายจากผู้ติดเชื้อในช่วงเดือนมกราคมถึงมีนาคม การทำ Data Visualization โดยใช้การวิเคราะห์แบบ Plotly และเปรียบเทียบกับโรคซาร์ส โดยใช้ข้อมูลเชิงลึกซึ่งนำเสนอข้อมูลเหล่านี้ออกมาให้สามารถเข้าใจได้ง่ายโดยใช้ Plotly ผลสรุปพบว่าโรคติดเชื้อไวรัสโคโรนา 2019 มีอัตราการติดเชื้อเพิ่มขึ้นเรื่อย ๆ จากเดือนมกราคมและสถานที่ที่พบแพร่เชื้อในวงกว้างขึ้นเรื่อย ๆ จากเดิมที่มีสถานการณ์แย่แค่ในประเทศจีนเมื่อนำไปวิเคราะห์เปรียบเทียบกับโรคซาร์สแล้วพบว่ามีลักษณะอาการบางอย่างคล้ายคลึงกันแต่โรคติดเชื้อไวรัสโคโรนา 2019 สามารถแพร่เชื้อได้เร็วและมีอัตราการตายสูงขึ้นอย่างชัดเจนกว่าโรคซาร์ส

[10] University of Dayton (2020) ชุดข้อมูลค่าอุณหภูมิเฉลี่ยระดับรายวัน ภาวะโลกร้อน คือ การเพิ่มขึ้นอย่างต่อเนื่องของอุณหภูมิเฉลี่ยของระบบภูมิอากาศของโลกและแสดงให้เห็นโดยการวัดอุณหภูมิโดยตรงและโดยการวัดผลกระทบต่าง ๆ ของภาวะโลกร้อน ดังนั้นชุดข้อมูลเกี่ยวกับอุณหภูมิของเมืองใหญ่ของโลกจะช่วยวิเคราะห์ได้เช่นเดียวกัน นอกจากนี้ข้อมูลสภาพอากาศยังมีประโยชน์สำหรับงานด้านวิทยาศาสตร์ข้อมูลมากมาย เช่น การพยากรณ์การขายโลจิสติกส์ เป็นต้น

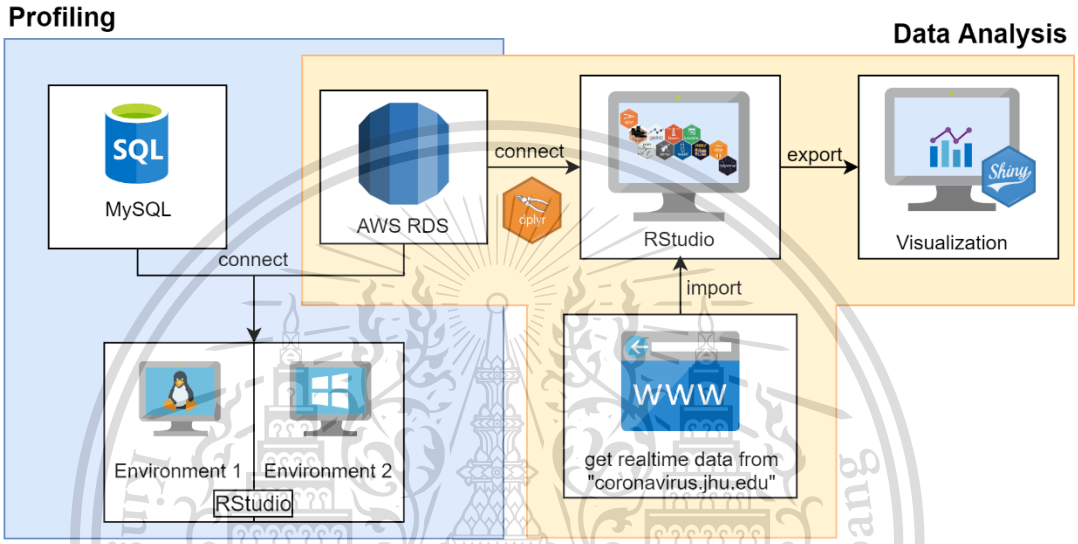
[11] Yanchang Zhao (2020) ศึกษาการทำ Data analysis เรื่องไวรัสโควิด-19 ด้วยภาษา R ซึ่งวิเคราะห์จากผู้ติดเชื้อทั่วโลกเพื่อแสดงข้อมูลการประมวลผล โดยใช้แพ็กเกจมาช่วยในการวิเคราะห์และแสดงข้อมูลได้แก่ tidyverse และ ggplot2 ซึ่งเป็นงานวิจัยที่ยังต้องปรับปรุงเพราะมีข้อมูลใหม่เพิ่มขึ้นมาเรื่อย ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3

การออกแบบและการพัฒนา

3.1 ภาพรวมของระบบ



รูป 3.1 ภาพรวมของระบบ

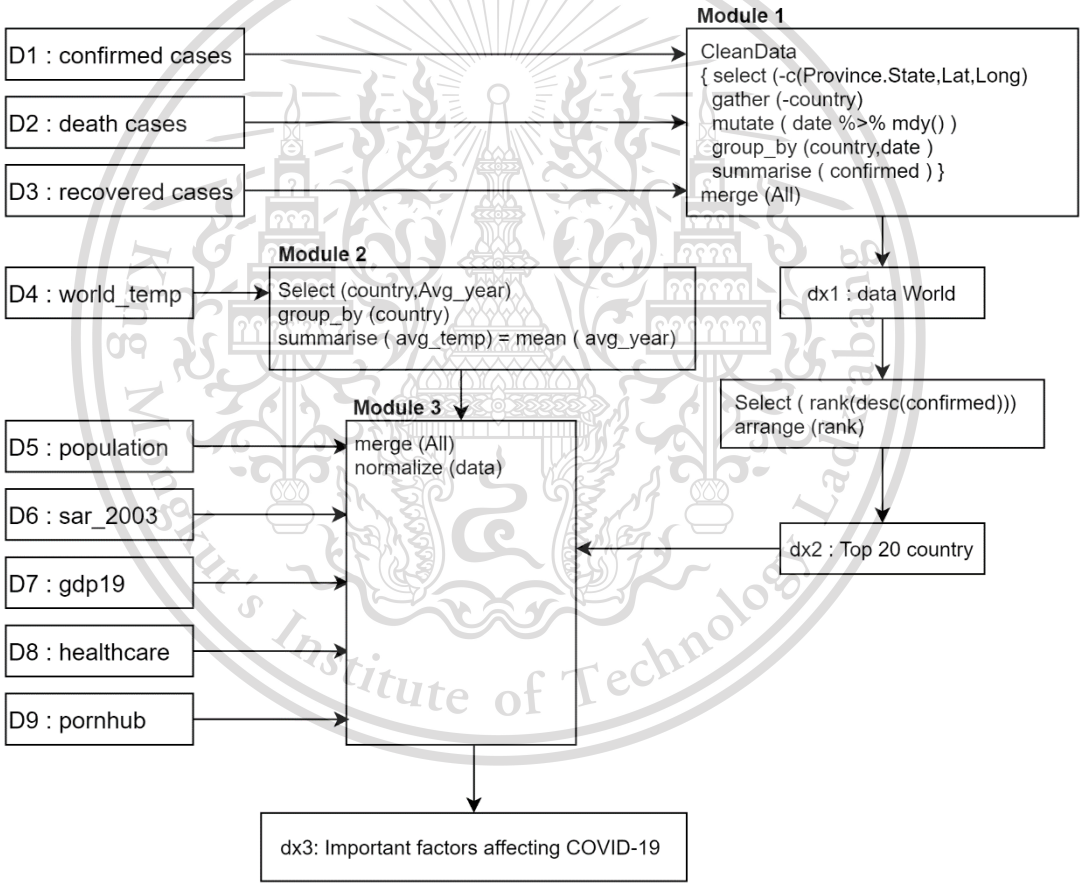
3.2 การตั้งสมติฐาน

1) การวิเคราะห์ 20 ประเทศที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุด และปัจจัยที่มีผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ได้แก่ อุณหภูมิ ผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) อันดับของระบบการจัดการด้านสาธารณสุข จำนวนผู้เข้าชม Pornhub จำนวนประชากรแต่ละประเทศและการแพร่ระบาดของโรคซาร์ส
2) การวิเคราะห์ 20 รัฐที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในสหรัฐอเมริกาและปัจจัยที่มีผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา ได้แก่ จำนวนประชากรไร้บ้าน ผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) จำนวนวันที่ Lockdown จำนวนประชากรในแต่ละรัฐ และจำนวนประชากรแบ่งตามเพศ
3) การวิเคราะห์ 10 จังหวัดที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทยและปัจจัยที่มีผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย ได้แก่ ความเสี่ยงจากการติดเชื้อ เพศ ช่วงอายุ และสัญชาติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

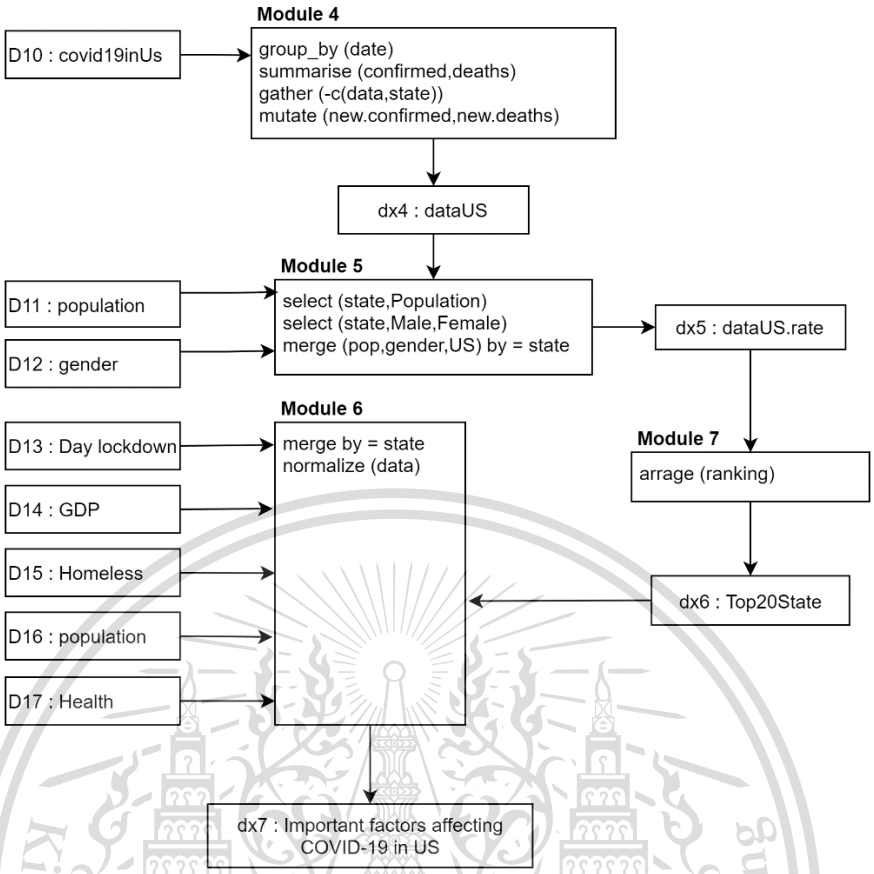
4) ผลการคาดการณ์การแพร่ระบาดโคโรนา 2019 ในประเทศไทยแบบมีมาตรการป้องกันดีกว่าแบบไม่มีมาตรการป้องกัน โดยใช้แบบจำลองทางคณิตศาสตร์ SEIR Model
5) ผลการคาดการณ์การแพร่ระบาดโคโรนา 2019 ในสหรัฐอเมริกาแบบมีมาตรการป้องกันดีกว่าแบบไม่มีมาตรการป้องกัน โดยใช้แบบจำลองทางคณิตศาสตร์ SEIRD Model
6) การวัดประสิทธิภาพ Profiling แบบใช้แพ็กเกจ Dplyr ดีกว่าแบบไม่ใช้แพ็กเกจ Dplyr ทั้งด้านหน่วยความจำและความเร็ว แล้วเมื่อมีปัจจัยอื่นมาเกี่ยวข้องจะส่งผลต่อความเร็วและเวลาในการประมวลผล

3.3 การประมวลผลข้อมูล (Data Processing)

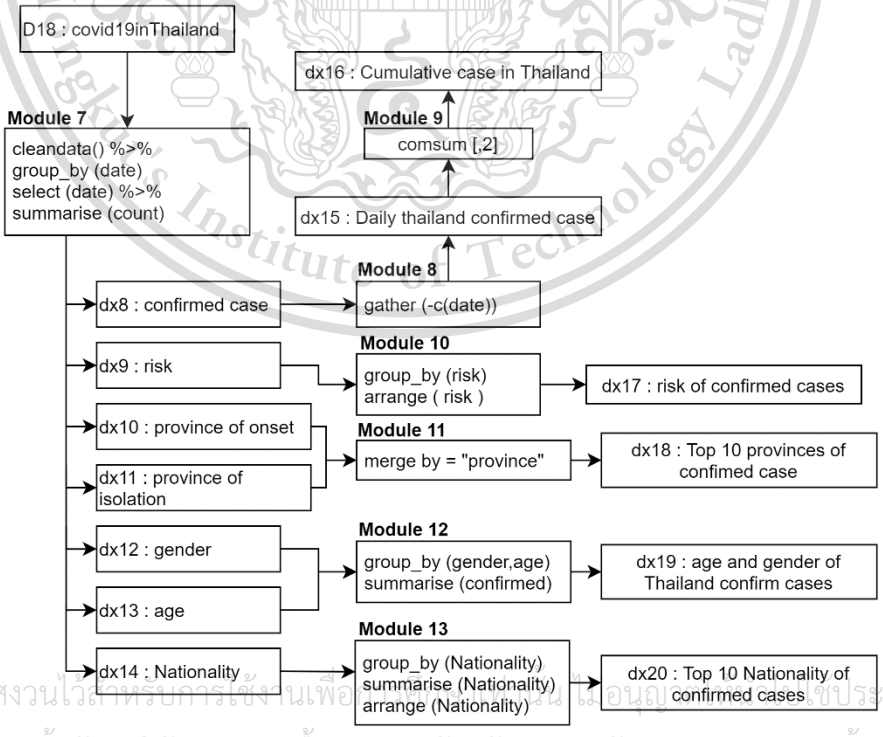


รูป 3.2 การประมวลผลข้อมูลของโลก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 3.3 การประมวลผลข้อมูลของสหรัฐอเมริกา



รูป 3.4 การประมลผลข้อมูลของประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.4 การวิเคราะห์ข้อมูล

3.4.1 ข้อมูลโรคติดเชื้อไวรัสโคโรนา 2019

ข้อมูลการระบาด Novel Coronavirus (COVID-19) ตั้งแต่วันที่ 22 มกราคม 2020 ข้อมูลนี้ถูกรวบรวมโดยศูนย์วิทยาศาสตร์และวิศวกรรมมหาวิทยาลัย Johns Hopkins (JHU CCSE) จากแหล่งต่าง ๆ รวมถึงองค์กรการอนามัยโลก World Health Organization (WHO) ข้อมูลนี้ประกอบด้วย Province/State (ชื่อเมืองหรือรัฐ), Country/Region (ชื่อประเทศ), Last Update (วันที่อัพเดตล่าสุด), Confirmed (จำนวนผู้ติดเชื้อ), Suspected (จำนวนผู้เสี่ยงติดเชื้อ), Recovered (จำนวนผู้รักษาหาย) และ Deaths (จำนวนผู้เสียชีวิต)

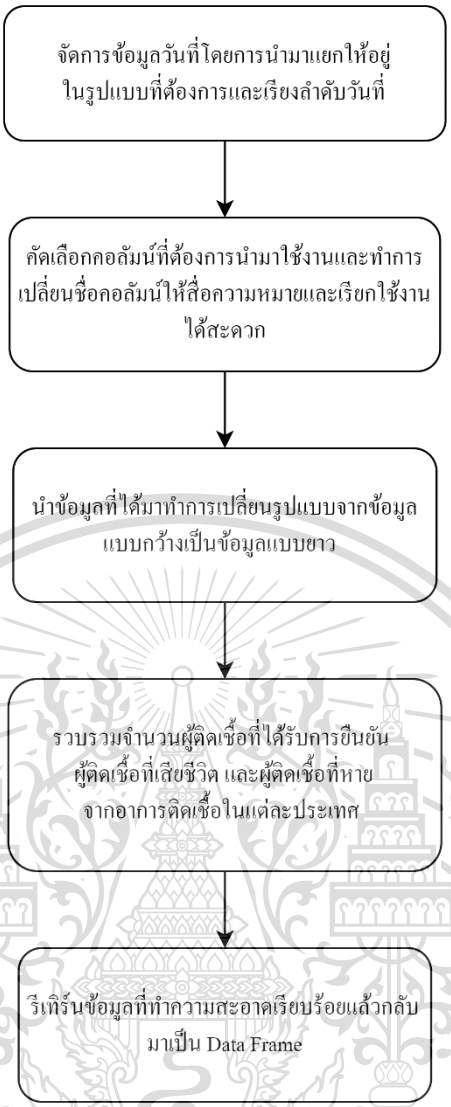
3.4.1.1 การทำความสะอาดข้อมูล (Data Cleaning)

การทำความสะอาดข้อมูล จะนำข้อมูลหลักที่ใช้ในการวิเคราะห์จำนวน 3 ชุด รวบรวมให้เป็นข้อมูลชุดเดียวกัน และจัดการข้อมูลบางส่วนให้อยู่ในรูปแบบที่พร้อมจะนำไปวิเคราะห์ในขั้นตอนต่อไป โดยข้อมูลที่ใช้ ได้แก่

1) ข้อมูลผู้ติดเชื้อไวรัสโคโรนาที่ได้รับการยืนยัน (Confirmed)
2) ข้อมูลผู้ที่เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา (Deaths)
3) ข้อมูลผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา (Recovered)

โดยมีการสร้างฟังก์ชันสำหรับทำความสะอาดข้อมูล ซึ่งมีขั้นตอนภายในฟังก์ชัน ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 3.5 ฟังก์ชันสำหรับทำความสะอาดข้อมูล

country <chr>	date <date>	confirmed <int>	deaths <int>	recovered <int>
Afghanistan	2020-04-01	237	4	5
Afghanistan	2020-04-02	273	6	10
Afghanistan	2020-04-03	281	6	10
Afghanistan	2020-04-04	299	7	10
Afghanistan	2020-04-05	349	7	15
Afghanistan	2020-04-06	367	11	18
Afghanistan	2020-04-07	423	14	18
Afghanistan	2020-04-08	444	14	29
Afghanistan	2020-04-09	484	15	32
Afghanistan	2020-04-10	521	15	32

1-10 of 28,764 rows Previous 1 2 3 4 5 6 ... 100 Next

รูป 3.6 ข้อมูลหลังการทำ Data Cleaning

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.4.1.2 ข้อมูลประเทศ 20 อันดับแรกที่มีผู้ติดเชื้อไวรัสโคโรนาสูงสุด

การวิเคราะห์ประเทศ 20 อันดับแรกที่มีผู้ติดเชื้อไวรัสโคโรนาสูงสุด เป็นการนำข้อมูลจากข้อ 3.5.1.1 มาทำการเรียงลำดับประเทศ 20 อันดับแรกที่มียอดผู้ติดเชื้อสูงสุด โดยมีขั้นตอนดังนี้

1) รวมจำนวนผู้ติดเชื้อ จำนวนผู้เสียชีวิต และจำนวนผู้ที่หายจากเชื้อไวรัสโคโรนาในแต่ละประเทศตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021
2) ทำการเรียงอันดับจากยอดรวมผู้ติดเชื้อ โดยประเทศที่มียอดรวมผู้ติดเชื้อสูงที่สุดจะเป็นอันดับแรก เรียงลำดับลงมาจนครบ 20 อันดับ

3.4.1.3 ข้อมูลผู้ติดเชื้อทั่วโลก

เพิ่มข้อมูลผู้ติดเชื้อทั่วโลกลงใน Data frame ที่ผ่านการทำความสะอาดแล้ว มาทำการรวบรวมจำนวนผู้ติดเชื้อ ผู้ที่เสียชีวิต และผู้ที่หายจากเชื้อไวรัสโคโรนาของทุกประเทศทั่วโลกตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021 โดยสร้างเป็นแถวใหม่ที่มีรายละเอียดในแต่ละคอลัมน์ ดังนี้

Confirmed : ยอดรวมจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ทั่วโลก

Deaths : ยอดรวมจำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 ทั่วโลก

Recovered : ยอดรวมจำนวนผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา 2019 ทั่วโลก

country <chr>	date <date>	confirmed <int>	deaths <int>	recovered <int>
World	2020-04-01	952188	49863	192918
World	2020-04-02	1033586	56101	209967
World	2020-04-03	1116829	62205	225415
World	2020-04-04	1197166	68351	245832
World	2020-04-05	1269049	73540	259672
World	2020-04-06	1342333	79481	276252
World	2020-04-07	1420111	87712	299643
World	2020-04-08	1504095	94464	328359
World	2020-04-09	1590519	102237	353707
World	2020-04-10	1678051	109629	375509

1-10 of 153 rows Previous 1 2 3 4 5 6 ... 16 Next

รูป 3.7 ข้อมูลผู้ติดเชื้อทั่วโลก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 3.1 ข้อมูลผู้ติดเชื้อทั่วโลก

country	date	confirmed	deaths	recovered
World	2020-01-22	ยอดผู้ติดเชื้อทั่วโลกในวันที่ 22 มกราคม 2020	ยอดผู้เสียชีวิตทั่วโลกในวันที่ 22 มกราคม 2020	ยอดผู้ที่หายทั่วโลกในวันที่ 22 มกราคม 2020
World	2020-01-23	ยอดผู้ติดเชื้อทั่วโลกในวันที่ 23 มกราคม 2020	ยอดผู้เสียชีวิตทั่วโลกในวันที่ 23 มกราคม 2020	ยอดผู้ที่หายทั่วโลกในวันที่ 23 มกราคม 2020
...	...	...	...	...

รวมถึงเพิ่มคอลัมน์ current.confirmed หรือจำนวนผู้ติดเชื้อปัจจุบันซึ่งเป็นการหาจำนวนผู้ติดเชื้อสุทธิ โดยนำจำนวนผู้ติดเชื้อลบกับจำนวนผู้เสียชีวิตและลบกับจำนวนผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา 2019 สามารถเขียนเป็นสมการได้ ดังนี้

$$\text{current.confirmed} = \text{confirmed} - \text{deaths} - \text{recovered} \quad (3.1)$$

โดย

current.confirmed คือ จำนวนผู้ติดเชื้อปัจจุบัน

confirmed คือ จำนวนผู้ติดเชื้อไวรัสโคโรนา 2019

deaths คือ จำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019

recovered คือ จำนวนผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา 2019

country <chr>	date <date>	confirmed <int>	deaths <int>	recovered <int>	current.confirmed <int>
Afghanistan	2020-04-01	237	4	5	228
Afghanistan	2020-04-02	273	6	10	257
Afghanistan	2020-04-03	281	6	10	265
Afghanistan	2020-04-04	299	7	10	282
Afghanistan	2020-04-05	349	7	15	327
Afghanistan	2020-04-06	367	11	18	338
Afghanistan	2020-04-07	423	14	18	391
Afghanistan	2020-04-08	444	14	29	401
Afghanistan	2020-04-09	484	15	32	437
Afghanistan	2020-04-10	521	15	32	474

1-10 of 28,917 rows 1 2 3 4 5 6 100 Next

รูป 3.8 ข้อมูลหลังจากเพิ่มคอลัมน์ current.confirmed

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.4.1.4 การวิเคราะห์ข้อมูลโรคติดเชื้อไวรัสโคโรนา 2019 เบื้องต้น

นำข้อมูลที่ทำการรวมแล้ว ได้แก่ ข้อมูลผู้ติดเชื้อไวรัสโคโรนาที่ได้รับการยืนยัน ข้อมูลผู้ที่เสียชีวิตจากเชื้อไวรัสโคโรนา ข้อมูลผู้ที่หายจากโรคเชื้อไวรัสโคโรนามาทำการวิเคราะห์ในแง่มุมต่าง ๆ ดังนี้

1) จำนวนผู้ที่ติดเชื้อไวรัสโคโรนาที่เพิ่มขึ้นในแต่ละวัน
เป็นการนำจำนวนผู้ติดเชื้อในวันนั้น ๆ ลบกับจำนวนผู้จำนวนติดเชื้อในวันก่อนหน้า สามารถเขียนเป็นสมการได้ ดังนี้

$$\text{new.confirmed} = \text{today.confirmed} - \text{previous.confirmed} \quad (3.2)$$

โดย

new.confirmed	คือ จำนวนผู้ที่เสียชีวิตที่เพิ่มขึ้นในแต่ละวัน
today.confirmed	คือ จำนวนผู้เสียชีวิตในวันนั้น ๆ
day.before.confirmed	คือ จำนวนผู้ติดเชื้อในวันก่อนหน้า

2) จำนวนผู้ที่เสียชีวิตจากเชื้อไวรัสโคโรนาที่เพิ่มขึ้นในแต่ละวัน
เป็นการนำจำนวนผู้ที่เสียชีวิตในวันนั้น ๆ ลบกับจำนวนผู้ที่เสียชีวิตในวันก่อนหน้า สามารถเขียนเป็นสมการได้ ดังนี้

$$\text{new.deaths} = \text{today.deaths} - \text{previous.deaths} \quad (3.3)$$

โดย

new.deaths	คือ จำนวนผู้ที่เสียชีวิตที่เพิ่มขึ้นในแต่ละวัน
today.deaths	คือ จำนวนผู้เสียชีวิตในวันนั้น ๆ
previous.deaths	คือ จำนวนผู้เสียชีวิตในวันก่อนหน้า

3) จำนวนผู้ที่หายจากเชื้อไวรัสโคโรนาที่เพิ่มขึ้นในแต่ละวัน
เป็นการนำจำนวนผู้ที่หายในวันนั้น ๆ ลบกับจำนวนผู้ที่หายในวันก่อนหน้า สามารถเขียนเป็นสมการได้ ดังนี้

$$\text{new.recovered} = \text{today.recovered} - \text{previous.recovered} \quad (3.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โดย

new.recovered คือ จำนวนผู้ที่รักษาหายที่เพิ่มขึ้นในแต่ละวัน

today.recovered คือ จำนวนผู้ที่รักษาหายในวันนั้น ๆ

previous.recovered คือ จำนวนผู้ที่รักษาหายในวันก่อนหน้า

4) อัตราการเสียชีวิตเทียบกับจำนวนผู้เสียชีวิตและผู้ที่รักษาหาย

เป็นการคำนวณอัตราการเสียชีวิต โดยใช้ผลรวมของผู้เสียชีวิตและผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนาเป็นส่วนเทียบ สามารถคำนวณได้จาก

$$\text{rate.upper} = 100 * \text{deaths}/(\text{deaths} + \text{recovered}) \quad (3.5)$$

โดย

rate.upper คือ อัตราการเสียชีวิตเทียบกับจำนวนผู้เสียชีวิตและผู้ที่หายจากเชื้อไวรัสโคโรนา

deaths คือ จำนวนผู้เสียชีวิต

recovered คือ จำนวนผู้ที่รักษาหาย

5) อัตราการเสียชีวิตเทียบกับจำนวนผู้ติดเชื้อ

เป็นการคำนวณอัตราการเสียชีวิต โดยใช้จำนวนผู้ติดเชื้อเป็นส่วนเทียบ สามารถคำนวณได้จาก

$$\text{rate.lower} = 100 * \text{deaths}/\text{confirmed} \quad (3.6)$$

6) อัตราการเสียชีวิตในแต่ละวัน

เป็นการคำนวณอัตราการเสียชีวิตในแต่ละวัน โดยเทียบกับผลรวมของจำนวนผู้เสียชีวิตรายวันและจำนวนผู้ติดเชื้อรายวัน สามารถ

$$\text{rate.daily} = 100 * \text{new.deaths}/(\text{new.deaths} + \text{new.recovered}) \quad (3.7)$$

โดย

rate.daily คือ อัตราการเสียชีวิตในแต่ละวัน

new.deaths คือ จำนวนผู้เสียชีวิตในแต่ละวัน

new.recovered คือ จำนวนผู้ที่รักษาหายในแต่ละวัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

	country	date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.upper	rate.lower	rate.daily
1	Afghanistan	2020-04-01	237	4	5	228	NA	NA	NA	44.4	1.7	NA
2	Afghanistan	2020-04-02	273	6	10	257	36	2	5	37.5	2.2	28.6
3	Afghanistan	2020-04-03	281	6	10	265	8	0	0	37.5	2.1	NaN
4	Afghanistan	2020-04-04	299	7	10	282	18	1	0	41.2	2.3	100.0
5	Afghanistan	2020-04-05	349	7	15	327	50	0	5	31.8	2.0	0.0
6	Afghanistan	2020-04-06	367	11	18	338	18	4	3	37.9	3.0	57.1
7	Afghanistan	2020-04-07	423	14	18	391	56	3	0	43.8	3.3	100.0
8	Afghanistan	2020-04-08	444	14	29	401	21	0	11	32.6	3.2	0.0
9	Afghanistan	2020-04-09	484	15	32	437	40	1	3	31.9	3.1	25.0
10	Afghanistan	2020-04-10	521	15	32	474	37	0	0	31.9	2.9	NaN

รูป 3.9 หลังจากการวิเคราะห์ข้อมูลส่วนต่าง ๆ

3.4.1.5 การทำ Data Reduction

นำข้อมูลมา normalization โดยใช้การ normalization แบบ Sigmoid ปรับค่าข้อมูลให้อยู่ระหว่างช่วงที่ -1 ถึง 1 ทั้งหมดเพื่อให้สามารถเปรียบเทียบกันได้

3.4.1.6 การทำ Visualization

ตัดชื่อประเทศออก เปลี่ยนตารางที่ได้จากการ normalization ให้อยู่ในรูปของ matrix เพื่อแสดงความสัมพันธ์แบบ correlation ว่าข้อมูลแต่ละแกนมีความเกี่ยวข้องกันเพียงใด

3.4.2 ข้อมูลจำนวนประชากร

ข้อมูลจำนวนประชากร 2020 ครั้งที่ 26 การประมาณการและการคาดการณ์จำนวนประชากรอย่างเป็นทางการขององค์การสหประชาชาติ United Nations (UN) เป็นศูนย์การสำหรับการประสานงานของการกระทำของชาติต่าง ๆ ซึ่งเป็นตัวแทนในระดับสากลน่าเชื่อถือมากที่สุด จัดทำโดยกองประชากรของกรมเศรษฐกิจและสังคมของสำนักเลขาธิการสหประชาชาติ

ผลลัพธ์หลักจะถูกนำเสนอในชุดไฟล์ Excel ที่แสดงตัวบ่งชี้ทางประชากรที่สำคัญสำหรับแต่ละกลุ่มการพัฒนาของสหประชาชาติกลุ่มรายได้ของธนาคารโลกภูมิภาคทางภูมิศาสตร์ เป้าหมายการพัฒนาที่ยั่งยืน (SDGs) ภูมิภาคและประเทศหรือพื้นที่สำหรับช่วงเวลาหรือวันที่ที่เลือกภายใน 1950- 2100

3.4.2.1 การทำ Data Cleaning

1) เลือกคอลัมน์ข้อมูลที่ต้องการนำมาใช้งาน คือ คอลัมน์ ‘Population (2020)’ ซึ่งเป็นคอลัมน์แสดงจำนวนประชากรในแต่ละประเทศ
2) นำข้อมูลไปรวมกับข้อมูลโรคติดเชื้อไวรัส 2019 เพื่อนำไปคำนวณค่าต่าง ๆ ต่อไป

3.4.2.2 การวิเคราะห์ข้อมูลจำนวนประชากร

คำนวณอัตราการติดเชื้อ โดยนำข้อมูลจำนวนผู้ติดเชื้อเทียบกับจำนวนประชากรแต่ละประเทศ โดยคำนวณจากสูตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$confirmed.rate = 100 * confirmed/population \quad (3.8)$$

โดย

confirmed.rate	คือ อัตราการติดเชื้อ
confirmed	คือ จำนวนผู้ติดเชื้อ
population	คือ จำนวนประชากรทั้งประเทศ

3.4.3 ข้อมูลผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ปี 2019

แหล่งข้อมูลผลิตภัณฑ์มวลรวมในประเทศหรือ GDP มาจาก (IMF International Monetary Fund) คือกองทุนการเงินระหว่างประเทศ ซึ่งเป็นที่รู้จักกันในฐานะกองทุนการเงิน ก่อตั้งเมื่อวันที่ 22 กรกฎาคม 1944 โดยแนวคิดการก่อตั้ง IMF มาจากที่ประชุมว่าด้วยเรื่องการเงินและการคลังแห่ง สหประชาชาติ หรือ United Nations Monetary and Financial Conference ณ เมืองเบรตตันวูดส์ รัฐ นิวแฮมเชียร์ ประเทศสหรัฐอเมริกา หรือที่รู้จักกันในนาม การประชุมเบรตตัน วูดส์ (Bretton Woods Conference) ข้อมูลประกอบด้วยรายชื่อประเทศ ข้อมูล GDP ปี 2019

3.4.3.1 การทำ Data Cleaning

1) เลือกคอลัมน์ข้อมูลที่ต้องการนำมาวิเคราะห์ ได้แก่ GDP
2) ดึงข้อมูลที่ได้ในข้อ 1) มาเปลี่ยนชื่อให้คอลัมน์ประเทศชื่อเป็น country และคอลัมน์ GDP (millions of US dollars) เป็นชื่อ GDP

3.4.3.2 การทำ Data Integration

ทำการ Merge ตารางที่ได้จาก 3.4.3.2 เข้ากับตาราง 20 อันดับจำนวนผู้ติดเชื้อมากที่สุดในโลก โดยใช้ Key เป็น Country

3.4.3.3 การทำ Data Reduction

นำข้อมูลมา Normalization โดยใช้การ Normalization แบบ Sigmoid ปรับค่าข้อมูลให้อยู่ระหว่างช่วงที่ -1 ถึง 1 ทั้งหมดเพื่อให้สามารถเปรียบเทียบกันได้

3.4.3.4 การทำ Visualization

1) เพิ่มคอลัมน์ชื่อประเทศ
2) กลับตารางในข้อ 3.4.3.3 จากรูปแบบกว้างให้อยู่ในรูปแบบยาว เพื่อใช้ในการแสดงกราฟ Heatmap ในแกน x และแกน y

3.4.4 ข้อมูลอุณหภูมิ

รายชื่อเมืองตามอุณหภูมิเฉลี่ย (รายเดือนและรายปี) จาก Wikipedia อุณหภูมิที่ระบุเป็นค่าเฉลี่ยของความสูงและต่ำในแต่ละวัน ดังนั้นอุณหภูมิตอนกลางวันจริงในเดือนหนึ่ง ๆ จะสูงกว่าอุณหภูมิที่ระบุไว้ที่นี่ 2 ถึง 10 ° C (4 ถึง 18 ° F) ขึ้นอยู่กับความแตกต่างระหว่างความสูงและต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ในแต่ละวัน แหล่งข้อมูลอุณหภูมิ ประกอบด้วย รายชื่อประเทศ เมือง วันที่ และค่าเฉลี่ยอุณหภูมิในแต่ละวัน รวบรวมข้อมูลโดยมหาวิทยาลัย Dayton

เนื่องจากข้อมูลอุณหูมิประจำปี 2020 เป็นข้อมูลที่รวบรวมอุณหภูมิเฉลี่ยของแต่ละเดือนใน ปี 2020 เราจึงต้องทำการคัดเลือกข้อมูลที่ต้องการใช้งานให้อยู่ในช่วงระหว่างเดือนเมษายนถึงเดือนสิงหาคม เพื่อนำไปใช้วิเคราะห์ร่วมกับข้อมูลโรคติดเชื้อไวรัสโคโรนา 2019 ที่อยู่ในช่วงเดียวกัน รวมถึงในบางประเทศมีข้อมูลอุณหภูมิเฉลี่ยจากหลายรัฐ ทำให้ต้องมีการหาค่าเฉลี่ยของแต่ละรัฐ เพื่อให้สามารถนำไปใช้เป็นข้อมูลอุณหภูมิเฉลี่ยของประเทศต่อไป โดยมีการทำความสะอาดข้อมูลก่อนนำไปใช้งาน ดังนี้

3.4.4.1 การทำ Data Cleaning

1) ตัดคอลัมน์เดือนจากเดือนมกราคมถึงเดือนธันวาคมให้เหลือเพียงค่าอุณหภูมิเฉลี่ยรายปี
2) หาค่าเฉลี่ยอุณหภูมิของแต่ละประเทศ โดยนำข้อมูลที่ได้จากข้อ 2) มาทำการหาค่าเฉลี่ยรวมของทุกรัฐในประเทศนั้น ๆ

3.4.4.2 การทำ Data Integration

ทำการ Merge ตารางที่ได้จากข้อ 3.5.4.1 เข้ากับตาราง 20 อันดับจำนวนผู้ติดเชื้อมากที่สุดในโลก โดยใช้ Key เป็น Country

3.4.4.3 การทำ Data Reduction

นำข้อมูลมา Normalization โดยใช้การ Normalization แบบ Sigmoid ปรับค่าข้อมูลให้อยู่ระหว่างช่วงที่ -1 ถึง 1 ทั้งหมดเพื่อให้สามารถเปรียบเทียบกันได้

3.4.4.4 การทำ Visualization

1) ตัดชื่อประเทศออก เปลี่ยนตารางที่ได้จากการ Normalization ให้อยู่ในรูปของ Matrix เพื่อแสดงความสัมพันธ์แบบ Correlation ว่าข้อมูลแต่ละแกนมีความเกี่ยวข้องกันเพียงใด
2) เพิ่มคอลัมน์ชื่อประเทศ
3) กลับตารางในข้อ 3.4.4.3 จากรูปแบบกว้างให้อยู่ในรูปแบบยาว เพื่อใช้ในการแสดงกราฟ Heatmap ในแกน x และแกน y

3.4.5 ข้อมูลระบบสาธารณสุขทั่วโลกปี 2019

ข้อมูลจากนิตยสาร CEOWORLD นิตยสารด้านธุรกิจของสหรัฐอเมริกา ได้จัดอันดับประเทศที่มีระบบสุขภาพดีที่สุดในโลก ประจำปี 2019 โดยมีการสำรวจทั้งหมด 89 ประเทศทั่วโลก พิจารณาจากหลักการต่าง ๆ ได้แก่

1) โครงสร้างพื้นฐานด้านสาธารณสุข
2) ประสิทธิภาพด้านสาธารณสุข แพทย์ พยาบาลและบุคลากรอื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3) ค่าใช้จ่ายในระบบ

4) การเข้าถึงยาคุณภาพ

5) ความพร้อมของรัฐบาลในการจัดการระบบ

6) ปัจจัยอื่น ๆ เช่น สิ่งแวดล้อม การเข้าถึงแหล่งน้ำสะอาด ระบบสุขอนามัย และควบคุมปัจจัยเสี่ยงด้านสุขภาพ เช่น การจัดการกับบุหรี่ ยาสูบ และการจัดการโรคอ้วน

3.4.5.1 การทำ Data Cleaning

1) เลือกคอลัมน์ข้อมูลที่ต้องการนำมาวิเคราะห์ ได้แก่ healthCareIndex

2) ดึงข้อมูลที่ได้ในข้อ 1) มาเปลี่ยนชื่อคอลัมน์ประเทศจาก Country.Region ให้เป็น Country

3.4.5.2 การทำ Data Integration

ทำการ Merge ตารางที่ได้จากข้อ 3.4.5.1 เข้ากับตาราง 20 อันดับจำนวนผู้ติดเชื้อมากที่สุดในโลก โดยใช้ Key เป็น Country

3.4.5.3 การทำ Data Reduction

นำข้อมูลมา Normalization โดยใช้การ Normalization แบบ Sigmoid ปรับค่าข้อมูลให้อยู่ระหว่างช่วงที่ -1 ถึง 1 ทั้งหมดเพื่อให้สามารถเปรียบเทียบกันได้

3.4.5.4 การทำ Visualization

1) เพิ่มคอลัมน์ชื่อประเทศ

2) กลับตารางในข้อ 3.4.5.3 จากรูปแบบกว้างให้อยู่ในรูปแบบยาว เพื่อใช้ในการแสดงกราฟ Heatmap ในแกน x และแกน y

3.5 การสร้างแบบจำลองทางคณิตศาสตร์

3.5.1 SEIR model [1,3,6]

การจำลองสถานการณ์เกี่ยวกับไวรัสผ่านโมเดลทางคณิตศาสตร์ที่จะนำมาใช้นั้นจะเป็นการจำลองการแพร่กระจายในระดับประชากรผ่านสมมติฐาน โดยจะใช้ SEIR model (Susceptible, Exposed, Infectious, Recovered) โดยโมเดลจะแบ่งประชากรออกเป็น 4 ประเภท ได้แก่

1) กลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible)

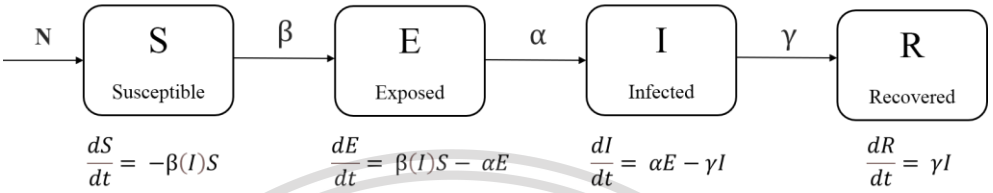
2) กลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) คือผู้ที่ติดเชื้อแล้ว แต่ยังไม่อยู่ในระยะที่สามารถแพร่เชื้อได้

3) กลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4) กลุ่มที่หายจากการติดเชื้อแล้ว และไม่มีโอกาสเป็นซ้ำ รวมถึงไม่สามารถแพร่เชื้อไปยังผู้อื่นได้ (Recovered)

โดย SEIR model นั้น ประชากรแต่ละประเภทนั้นจะมีการเปลี่ยนสถานะอยู่ตลอดเวลา ขึ้นอยู่กับการกำหนดสมมติฐาน ปัจจัย และอัตราในการเพิ่มหรือลดของประชากรแต่ละประเภท โดยสามารถแสดงออกมาเป็นแผนภูมิเบื้องต้นได้ดังภาพด้านล่าง



รูป 3.10 SEIR Diagram

หลังจากมีการกำหนดสมมติฐาน รวมถึงปัจจัยเพิ่มลดของประชากรแต่ละประเภทผ่านสมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equations) และจำนวนประชากรแต่ละประเภทในระบบที่ช่วงเวลาเริ่มต้นแล้ว โมเดลจะคำนวณออกมาเป็นผลทำนายว่าเมื่อเวลาผ่านไปช่วงหนึ่ง ประชากรแต่ละประเภทจะมีอัตราเพิ่มหรือลดไปเท่าใด เพื่อให้สามารถประเมินสถานการณ์ และออกมาตรการควบคุมได้ดียิ่งขึ้น

3.5.1.2 การกำหนดสมมติฐาน และปัจจัยเพิ่มลดของประชากรแต่ละประเภทผ่านสมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equations)

$$N = S + E + I + R \tag{3.9}$$

N คือ ประชากรในประเทศทั้งหมด

S คือ จำนวนประชากรเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible)

E คือ จำนวนประชากรที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed)

I คือ จำนวนประชากรที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

R คือ จำนวนประชากรที่หายจากการติดเชื้อแล้ว และไม่มีโอกาสเป็นซ้ำ รวมถึงไม่สามารถแพร่เชื้อไปยังผู้อื่นได้ (Recovered)

$$\beta(I) = \frac{R_0 I}{D_I N} \tag{3.10}$$

$$\alpha = \frac{1}{D_E} \tag{3.11}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$\gamma = \frac{1}{D_I} \tag{3.12}$$

$\beta(I)$ คือ อัตราที่ประชากรกลุ่มเสี่ยง (Susceptible) กลายเป็นผู้ติดเชื้อระยะฟักตัว (Exposed) หรือสามารถเรียกได้ว่า ความรุนแรงในการแพร่ (Force of infection) โดย I จะไม่ใช่ค่าคงที่ เนื่องจากขึ้นอยู่กับจำนวนผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) ด้วย

D_E คือ ช่วงเวลาเฉลี่ยของการฟักตัวของเชื้อ

D_I คือ ช่วงเวลาเฉลี่ยที่ผู้ป่วยจะสามารถแพร่เชื้อได้

α คือ อัตราการฟักตัวของเชื้ อ หรือคืออัตราที่ประชากรกลุ่มผู้ติดเชื้อระยะฟักตัว (Exposed) กลายเป็นผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

γ คือ อัตราการหายจากอาการติดเชื้อ (Recovery Rate) หรืออัตราที่กลุ่มผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) กลายเป็นผู้ที่หายจากการติดเชื้อแล้ว (Recovered)

กลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) โดยปัจจัยหลักของการที่กลุ่มเสี่ยงนั้นจะติดเชื้อ เกิดจากการที่กลุ่มเสี่ยงอยู่ใกล้ผู้ที่ติดเชื้อ ดังนั้นเมื่อมีผู้ติดเชื้อในระบบมากขึ้น รวมถึงโอกาสที่จะกลุ่มเสี่ยงจะติดเชื้อเพิ่มสูงขึ้น จะทำให้มีประชากรเปลี่ยนประเภทจากกลุ่มเสี่ยง (Susceptible) เป็นผู้ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) มากขึ้น

$$\frac{dS}{dt} = -\beta(I)S \tag{3.13}$$

กลุ่มติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สามารถเพิ่มขึ้นในระบบจากการที่ประชากรจากกลุ่มเสี่ยงติดเชื้อเพิ่มขึ้น และจะลดลงก็ต่อเมื่อผ่านระยะฟักตัว (Incubation period) กลายเป็นประชากรกลุ่มติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

$$\frac{dE}{dt} = \beta(I)S - \alpha E \tag{3.14}$$

กลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) สามารถเพิ่มขึ้นได้ในระบบจากการที่ประชากรจากกลุ่มติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) นั้นผ่านระยะการฟักตัวไปแล้ว และจะสามารถลดลงได้เมื่อผ่านระยะเวลาที่ติดเชื้อ (Infectious period) ไปแล้ว กลายเป็นประชากรกลุ่มที่รักษาจนหายแล้ว (Recovered)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$\frac{dI}{dt} = \alpha E - \gamma I \tag{3.15}$$

กลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) สามารถเพิ่มขึ้นได้ในระบบ โดยการที่มีผู้ติดเชื้อที่รักษาจนหายแล้วในระบบมากขึ้น และจะไม่ลดลงเนื่องจากผู้ที่รักษาจนหายแล้วจะมีภูมิต้านทานโรค ทำให้ไม่สามารถเป็นซ้ำ และไม่สามารถแพร่เชื้อได้อีก

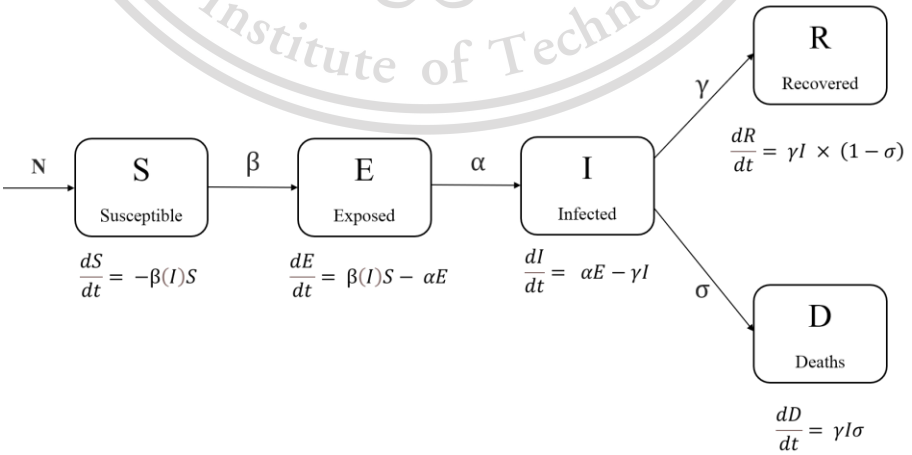
$$\frac{dR}{dt} = \gamma I \tag{3.16}$$

3.5.2 SEIRD model [5]

แบบจำลอง SEIRD นี้ใช้สมการเชิงเส้นที่ไม่ขึ้นกับเวลาลำดับแรกเพื่อจำลองว่าผู้ป่วยดำเนินไปอย่างไร จะมีความคล้ายกับ SEIR model แต่จะเพิ่มประเภทที่เสียชีวิตจากโรคระบาด เนื่องจากสหรัฐอเมริกามีจำนวนผู้เสียชีวิตสูงมาก จึงนำมาใช้ในการคำนวณด้วย โดยโมเดลจะแบ่งประชากรออกเป็น 5 ประเภท ได้แก่

1) กลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible)
2) กลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) คือผู้ที่ติดเชื้อแล้ว แต่ยังไม่อยู่ในระยะที่สามารถแพร่เชื้อได้
3) กลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)
4) กลุ่มที่หายจากการติดเชื้อแล้ว และไม่มีโอกาสเป็นซ้ำ รวมถึงไม่สามารถแพร่เชื้อไปยังผู้อื่นได้ (Recovered)
5) กลุ่มที่เสียชีวิตจากโรคระบาด (Deaths)

โดย SEIRD model นั้น ประชากรแต่ละประเภทนั้นจะมีการเปลี่ยนสถานะอยู่ตลอดเวลา ขึ้นอยู่กับการกำหนดสมมติฐาน ปัจจัย และอัตราในการเพิ่ม หรือลด ของประชากรแต่ละประเภท โดยสามารถแสดงออกมาเป็นแผนภูมิเบื้องต้นได้ดังภาพด้านล่าง



รูป 3.11 SEIRD Diagram

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

หลังจากมีการกำหนดสมมติฐาน รวมถึงปัจจัยเพิ่มลดของประชากรแต่ละประเภทผ่านสมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equations) และจำนวนประชากรแต่ละประเภทในระบบที่ช่วงเวลาเริ่มต้นแล้ว ทางโมเดลจะคำนวณออกมาเป็นผลทำนายว่าเมื่อเวลาผ่านไปนั้น ประชากรแต่ละประเภทจะมีอัตราเพิ่ม หรือลดไปเท่าใด เพื่อที่จะได้ประเมินสถานการณ์ และออกมาตรการควบคุมได้ดียิ่งขึ้น

3.5.2.2 การกำหนดสมมติฐาน และปัจจัยเพิ่มลดของประชากรแต่ละประเภทผ่านสมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equations)

$$N = S + E + I + R + D \tag{3.17}$$

N คือ ประชากรในประเทศทั้งหมด

S คือ จำนวนประชากรเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible)

E คือ จำนวนประชากรที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed)

I คือ จำนวนประชากรที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

R คือ จำนวนประชากรที่หายจากการติดเชื้อแล้ว และไม่มีโอกาสเป็นซ้ำ รวมถึงไม่สามารถแพร่เชื้อไปยังผู้อื่นได้ (Recovered)

D คือ จำนวนประชากรที่เสียชีวิตจากการติดเชื้อ (Deaths)

$$\beta(I) = \frac{R_0 I}{D_I N} \tag{3.18}$$

$$\alpha = \frac{1}{D_E} \tag{3.19}$$

$$\gamma = \frac{1}{D_I} \tag{3.20}$$

$$\sigma = \frac{1}{D_D} \tag{3.21}$$

$\beta(I)$ คือ อัตราที่ประชากรกลุ่มเสี่ยง (Susceptible) กลายเป็นผู้ติดเชื้อระยะฟักตัว (Exposed) หรือสามารถเรียกได้ว่า ความรุนแรงในการแพร่ (Force of infection) โดย I จะไม่ใช่ค่าคงที่เนื่องจากขึ้นอยู่กับจำนวนผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) ด้วย

D_E คือ ช่วงเวลาเฉลี่ยของการฟักตัวของเชื้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

D_I คือ ช่วงเวลาเฉลี่ยที่ผู้ป่วยจะสามารถแพร่เชื้อได้

D_D คือ ช่วงเวลาเฉลี่ยที่ผู้ป่วยเสียชีวิตหลังจากได้ติดเชื้อ

α คือ อัตราการฟักตัวของเชื้อ หรืออัตราที่ประชากรกลุ่มผู้ติดเชื้อระยะฟักตัว (Exposed) กลายเป็นผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

γ คือ อัตราการหายจากอาการติดเชื้อ (Recovery rate) หรืออัตราที่กลุ่มผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) กลายเป็นผู้ที่หายจากการติดเชื้อแล้ว (Recovered)

σ คือ อัตราการเสียชีวิตจากอาการติดเชื้อ (crude fatality Rate : CFR) หรืออัตราที่กลุ่มผู้ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) กลายเป็นผู้ที่เสียชีวิตจากการติดเชื้อแล้ว (Deaths) คำนวณจากจำนวนผู้เสียชีวิตสะสม/จำนวนผู้ติดเชื้อสะสม ณ วันนั้นๆ

กลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) โดยปัจจัยหลักของการที่กลุ่มเสี่ยงนั้นจะติดเชื้อ เกิดจากการที่กลุ่มเสี่ยงอยู่ใกล้ผู้ที่ติดเชื้อ ดังนั้นเมื่อมีผู้ติดเชื้อในระบบมากขึ้น รวมถึงโอกาสที่จะกลุ่มเสี่ยงจะติดเชื้อเพิ่มสูงขึ้น จะทำให้มีประชากรเปลี่ยนประเภทจากกลุ่มเสี่ยง (Susceptible) เป็นผู้ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) มากขึ้น

$$\frac{dS}{dt} = -\beta(I)S \tag{3.22}$$

กลุ่มติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สามารถเพิ่มขึ้นในระบบจากการที่ประชากรจากกลุ่มเสี่ยงติดเชื้อเพิ่มขึ้น และจะลดลงก็ต่อเมื่อผ่านระยะฟักตัว (Incubation period) กลายเป็นประชากรจำพวก กลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious)

$$\frac{dE}{dt} = \beta(I)S - \alpha E \tag{3.23}$$

กลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) สามารถเพิ่มขึ้นได้ในระบบจากการที่ประชากรจากกลุ่มติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) นั้นผ่านระยะการฟักตัวไปแล้ว และจะสามารถลดลงได้เมื่อผ่านระยะเวลาที่ติดเชื้อ (Infectious period) ไปแล้ว กลายเป็นประชากรกลุ่มที่รักษาจนหายแล้ว (Recovered) หรือ อาจกลายเป็นประชากรกลุ่มที่เสียชีวิตแล้วจากการติดเชื้อ (Deaths)

$$\frac{dI}{dt} = \alpha E - \gamma I \tag{3.24}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) สามารถเพิ่มขึ้นได้ในระบบ โดยการที่มีผู้ติดเชื้อที่รักษาจนหายแล้วในระบบมากขึ้น และจะไม่ลดลงเนื่องจากผู้ที่รักษาจนหายแล้วจะมีภูมิต้านทานโรค ทำให้ไม่สามารถเป็นซ้ำ และไม่สามารถแพร่เชื้อได้อีก

$$\frac{dR}{dt} = \gamma I \times (1 - \sigma) \tag{3.25}$$

กลุ่มที่เสียชีวิตจากโรคระบาด (Deaths) สามารถเพิ่มขึ้นได้ในระบบ โดยการที่มีผู้ติดเชื้อที่เสียชีวิตแล้วในระบบมากขึ้น และจะไม่ลดลงเนื่องจากผู้ที่เสียชีวิตแล้ว จะไม่สามารถแพร่เชื้อได้อีก

$$\frac{dD}{dt} = \gamma I \sigma \tag{3.26}$$

3.6 แพ็กเกจที่ใช้ใน R

แพ็กเกจที่ใช้ใน R เพื่อการวิเคราะห์ข้อมูลแบ่งเป็น 2 ส่วนใหญ่ คือ แพ็กเกจทั้งหมดที่ไม่ใช้ Dplyr และแพ็กเกจทั้งหมดที่ใช้ Dplyr ร่วมด้วย มีรายละเอียดดังต่อไปนี้

3.6.1 แพ็กเกจที่ใช้ร่วมกัน

เป็นแพ็กเกจที่ใช้ในการวิเคราะห์ข้อมูลทั้งแบบใช้ Dplyr และไม่ใช้

ตาราง 3.2 แพ็กเกจที่ใช้ร่วมกัน

ชื่อแพ็กเกจ	การใช้งาน
magrittr	ใช้อำนวยความสะดวกในการเขียนโค้ด และช่วยทำให้โค้ดอ่านง่ายขึ้นด้วยเครื่องมือที่เรียกว่า “Pipe operator” ที่เขียนเป็นสัญลักษณ์ “%>%” ใช้ในการส่งข้อมูลจากด้านซ้ายมือไปยังพารามิเตอร์แรกของฟังก์ชันทางด้านขวามือ
Lubridate	ใช้สำหรับจัดการข้อมูลประเภทวันที่และเวลาได้ สามารถแยกส่วนข้อมูลวันที่และเวลาและอัพเดตข้อมูลแต่ละส่วนเกี่ยวกับวันที่และเวลา รวมถึงเรียงลำดับวันเวลาได้
ggplot2	ใช้ในการสร้างกราฟิกจากชุดข้อมูล สามารถสร้างงานกราฟิกได้หลากหลาย และปรับแต่งองค์ประกอบต่าง ๆ ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 3.2 แพ็กเกจที่ใช้ร่วมกัน (ต่อ)

ชื่อ แพ็กเกจ	การใช้งาน
gridExtra	ใช้สำหรับจัดการกราฟิก โดยสามารถจัดการวัตถุประเภทกราฟิกให้อยู่ในตำแหน่งที่กำหนดได้

3.6.2 แพ็กเกจเพิ่มเติมแบบไม่ใช้แพ็กเกจ Dplyr

ตาราง 3.3 แพ็กเกจเพิ่มเติมแบบไม่ใช้แพ็กเกจ Dplyr

ชื่อ แพ็กเกจ	การใช้งาน
RODBC	ใช้สำหรับเชื่อมต่อระหว่าง R และฐานข้อมูล โดยมีคำสั่ง 2 ประเภทหลัก ได้แก่ คำสั่งที่ขึ้นต้นด้วย odbc* ใช้สำหรับเชื่อมต่อและสั่งงานฐานข้อมูลในระดับล่าง และคำสั่งที่ขึ้นต้นด้วย sql* ใช้สำหรับฐานข้อมูลที่มีโครงสร้างในระดับที่สูงขึ้น โดยสามารถอ่าน แก้ไข จัดการข้อมูลระหว่าง Data frame และตาราง SQL
reshape2	ใช้ในการแปลงรูปแบบข้อมูลระหว่างข้อมูลแบบกว้าง (Wide data format) และข้อมูลแบบยาว (Long data format)

ตาราง 3.3 แพ็กเกจเพิ่มเติมแบบไม่ใช้แพ็กเกจ Dplyr (ต่อ)

ชื่อ แพ็กเกจ	การใช้งาน
plyr	ใช้ในการแก้ไขปัญหาทั่วไป เช่น แบ่งส่วนข้อมูลให้มีขนาดเล็กลงเพื่อสามารถนำไปวิเคราะห์ได้ง่ายขึ้น จัดการกับข้อมูลแต่ละส่วนที่แบ่งได้ และทำการรวบรวมกลับเป็นกลุ่มข้อมูลดังเดิมได้
sqldf	ใช้เพื่อให้สามารถรันคำสั่ง SQL บน Data frame ได้ โดยไม่จำเป็นต้องอยู่ในรูปแบบของตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.6.3 แพ็กเกจเพิ่มเติมแบบใช้แพ็กเกจ Dplyr

ตาราง 3.4 แพ็กเกจเพิ่มเติมแบบใช้แพ็กเกจ Dplyr

ชื่อแพ็กเกจ	การใช้งาน
Dplyr	ใช้สำหรับจัดการกับข้อมูล โดยมุ่งเน้น Data frame และ Tibbles เป็นหลัก สามารถจัดการให้ข้อมูลอยู่ในรูปแบบที่ต้องการเพื่อนำไปใช้ในการวิเคราะห์ต่อไปได้ มีฟังก์ชันที่อำนวยความสะดวกในการจัดการกับข้อมูลต่าง ๆ เช่น การเพิ่มคอลัมน์ใหม่ การเลือกข้อมูล
tidyr	มีฟังก์ชันต่าง ๆ ในการจัดการข้อมูล โดยในที่นี้ใช้แพกเกจเพื่อการแปลงรูปแบบข้อมูลระหว่างข้อมูลแบบกว้าง (Wide data format) และข้อมูลแบบยาว (Long data format)
plyr	ใช้ในการแก้ไขปัญหาทั่วไป เช่น แบ่งส่วนข้อมูลให้มีขนาดเล็กลงเพื่อสามารถนำไปวิเคราะห์ได้ง่ายขึ้น จัดการกับข้อมูลแต่ละส่วนที่แบ่งได้ และทำการรวบรวมกลับเป็นกลุ่มข้อมูลดังเดิมได้
sqldf	ใช้เพื่อให้สามารถรันคำสั่ง SQL บน Data frame ได้ โดยไม่จำเป็นต้องอยู่ในรูปแบบของตาราง

3.7 การเชื่อมต่อฐานข้อมูล

การเชื่อมต่อฐานข้อมูลในที่นี้ใช้ฐานข้อมูล MySQL ในการเก็บข้อมูลเพื่อนำใช้ในการวิเคราะห์ โดยการนำเข้าข้อมูลมีอยู่ 2 แบบด้วยกัน คือ

3.7.1 แบบใช้ ODBC

ใช้ฟังก์ชัน odbcConnect ในแพ็กเกจ RODBC เพื่อทำการเชื่อมต่อระหว่าง R กับฐานข้อมูลโดยผ่านไดร์เวอร์ ODBC

โปรแกรม 3.1 การเชื่อมต่อฐานข้อมูลโดยผ่านไดร์เวอร์ ODBC

```
data <- odbcConnect("coronavirus19",
                    uid = "username",
                    pwd = "password")
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.7.2 แบบใช้แพ็กเกจ Dplyr

ใช้ฟังก์ชัน src_mysql ในแพ็กเกจ dplyr เพื่อทำการเชื่อมต่อระหว่าง R กับฐานข้อมูล

โปรแกรม 3.2 การเชื่อมต่อฐานข้อมูลโดยใช้ฟังก์ชัน src_mysql

```
my_db <- src_mysql( dbname = "databasename",
                    host = "localhost",
                    user = "username",
                    password = "password")
```


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4

การทดลองและผลการทดลอง

4.1 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ทั่วโลก

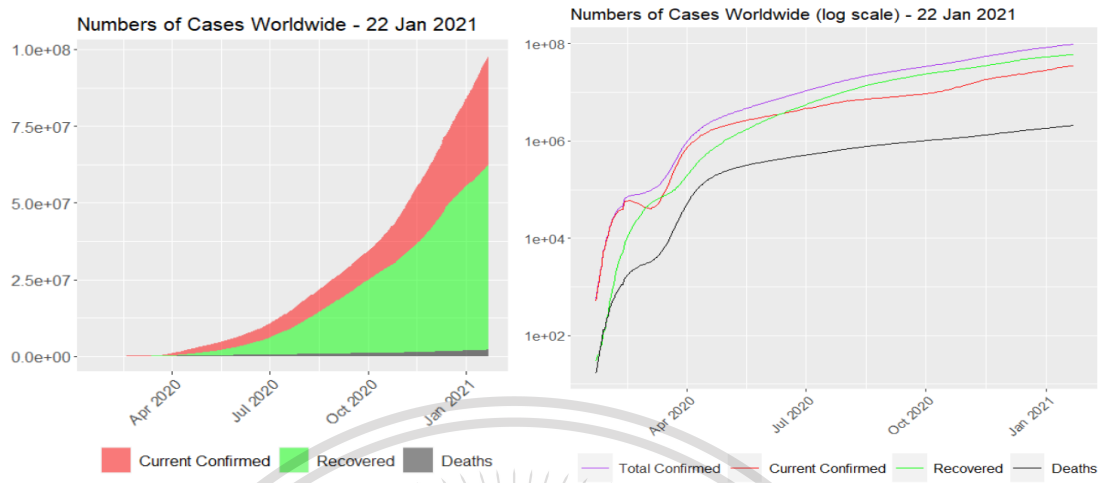
โรคไวรัสโคโรนา 2019 หรือที่รู้จักกันในชื่อว่า โควิด 19 หรือ COVID-19 ซึ่งย่อมาจาก Coronavirus Disease 2019 มีชื่ออย่างเป็นทางการว่า SARS-CoV-2 ไวรัสโคโรนาสายพันธุ์ที่กำลังแพร่กระจายทั่วโลกในขณะนี้เป็นสายพันธุ์ที่ไม่เคยพบมาก่อน โรคไวรัสโคโรนาสามารถติดเชื้อได้ทั้งในคนและสัตว์ เพราะไวรัสมีสารพันธุกรรม RNA ที่มีโอกาสกลายพันธุ์ได้สูง อีกทั้งยังมีลักษณะคล้ายคลึงกับโรคซาร์สที่เคยเกิดขึ้นมาก่อนแล้ว

ผู้ป่วยรายแรกเริ่มต้นในเดือนธันวาคม 2019 โดยพบครั้งแรกในเมืองอู่ฮั่น มณฑลหูเป่ย์ ประเทศจีน สันนิษฐานว่ามีการแพร่กระจายเชื้อมาจากสัตว์สู่คน องค์กรอนามัยโลกได้ประกาศเป็นภาวะฉุกเฉินทางสาธารณสุขระหว่างประเทศ เมื่อวันที่ 30 มกราคม 2020 เกิดการแพร่กระจายไปยังมณฑลอื่น ๆ ของประเทศจีนอย่างรวดเร็ว หลังจากนั้นในวันที่ 26 กุมภาพันธ์ 2020 ผู้ป่วยภายในประเทศจีนลดลง แต่ในประเทศอื่น ๆ กลับมีผู้ติดเชื้อเพิ่มขึ้นอย่างรวดเร็ว ส่งผลกระทบเป็นวงกว้างทางระบบสาธารณสุข เศรษฐกิจ นำไปสู่การเกิด New normal ที่ประชากรโลกต้องปรับตัวเพื่อรับมือการแพร่ระบาดของไวรัสโคโรนา 2019

ผู้ติดเชื้อส่วนใหญ่จะมีอาการทางระบบทางเดินหายใจและสามารถหายเองได้โดยไม่ต้องได้รับการรักษาเป็นพิเศษ แต่ในผู้สูงอายุและผู้ที่มีโรคประจำตัว เช่น โรคหัวใจ โรคเบาหวาน โรคระบบทางเดินหายใจ มีแนวโน้มที่จะมีอาการรุนแรง การแพร่กระจายของเชื้อคล้ายกับไข้หวัดใหญ่ โดยติดเชื้อผ่านทางละอองน้ำลายหรือน้ำมูกผ่านทางการไอ จาม ระยะการฟักตัวอยู่ที่ประมาณ 2 ถึง 14 วัน อาการที่พบเป็นส่วนมาก ได้แก่ มีไข้สูง ไอ จาม หายใจลำบาก การรับกลิ่นบกพร่อง และมีภาวะแทรกซ้อนเกี่ยวกับทางเดินหายใจ วิธีการป้องกันและการลดการแพร่เชื้อที่ดีที่สุดคือ การรับข้อมูลข่าวสารเกี่ยวกับโรคและการแพร่กระจายของเชื้อไวรัสโคโรนา ลดความเสี่ยงในการติดเชื้อโดยการล้างมือด้วยสบู่ หรือการใช้แอลกอฮอล์ล้างมือบ่อย ๆ ไม่ใช้มือสัมผัสใบหน้า และเมื่อพบว่ามีการเสี่ยงติดเชื้อควรป้องกันโดยการกักตัวอย่างน้อย 14 วันเพื่อสังเกตการณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.1.1 จำนวนผู้ติดเชื้อทั่วโลก



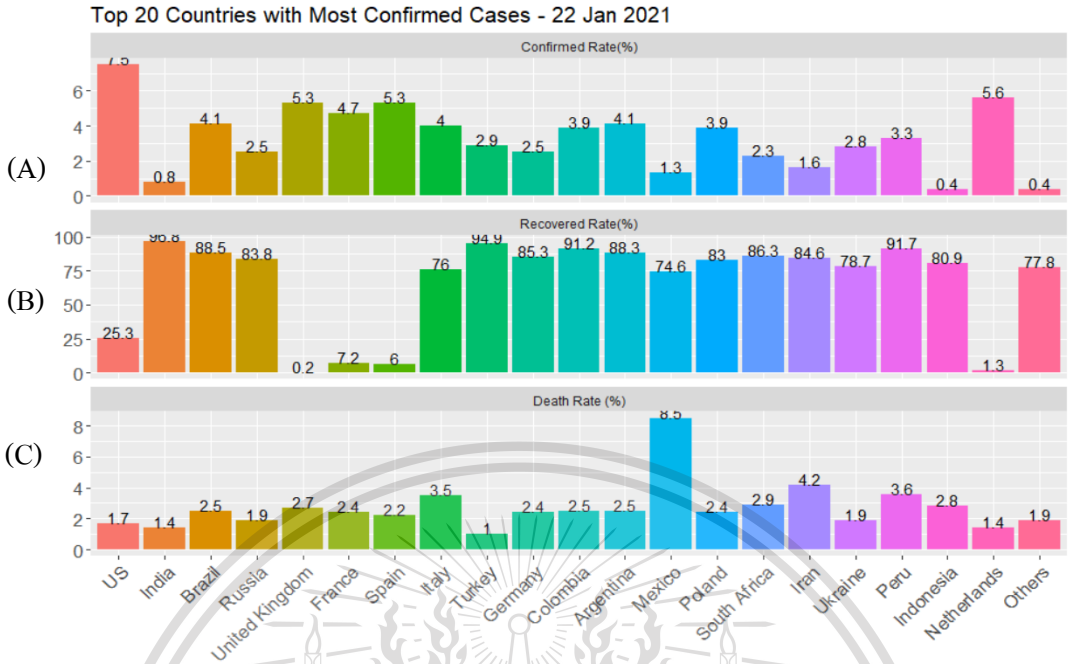
รูป 4.1 จำนวนผู้ติดเชื้อปัจจุบัน (สีแดง) จำนวนผู้รักษาหาย (สีเขียว) และจำนวนผู้เสียชีวิต (สีเทา) ทั่วโลกรูปแบบ Stack และจำนวนผู้ติดทั้งหมด (สีม่วง) จำนวนผู้ติดเชื้อปัจจุบัน (สีแดง) จำนวนผู้รักษาหาย (สีเขียว) และจำนวนผู้เสียชีวิต (สีดำ) ทั่วโลกในรูปแบบ Log scale

กราฟพื้นที่แบบ Stack แสดงจำนวนผู้ติดเชื้อปัจจุบัน ผู้รักษาหายและผู้เสียชีวิตทั่วโลก โดยมีแกน x เป็นแกนเวลานับตั้งแต่วันที่ 21 มกราคม 2020 จนถึงวันที่ 21 มกราคม 2021 แกน y เป็นจำนวนคน พื้นที่สีแดงแสดงถึงจำนวนผู้ติดเชื้อปัจจุบัน ณ เวลานั้น พื้นที่สีเขียวแสดงถึงจำนวนผู้รักษาหาย และพื้นที่สีดำแสดงถึงจำนวนผู้เสียชีวิต โดยกราฟแสดงจำนวนข้อมูลในรูปแบบ Log Scale

จากกราฟจำนวนผู้ติดเชื้อทั่วโลกในรูปแบบ Stack พบว่าจำนวนผู้ติดเชื้อมีการเติบโตแบบเอ็กซ์โพเนนเชียล (Exponential growth) คือ จำนวนผู้ติดเชื้อทั้งหมดเพิ่มสูงขึ้นตามระยะเวลา จากข้อมูลพบว่าในช่วงปลายเดือนมีนาคม 2020 นั้น ประชากรเกือบหนึ่งในสามของโลกต้องใช้ชีวิตอยู่ภายใต้ข้อจำกัดต่าง ๆ เพื่อป้องกันการแพร่กระจายของเชื้อไวรัส องค์การอนามัยโลกได้ออกมาเตือนถึงการขาดแคลนเวชภัณฑ์อย่างมาก จากกราฟสามารถสังเกตได้ว่าเดือนเมษายน 2020 มีผู้ติดเชื้อไวรัสโคโรนา 2019 จากข้อมูลพบว่าจำนวนผู้ติดเชื้อทั่วโลกเกินหนึ่งล้านคน ปลายเดือนเมษายน 2020 มีผู้ติดเชื้อทั่วโลกมีจำนวนเกิน 3.5 ล้านคนซึ่งเพิ่มสูงขึ้นอย่างรวดเร็ว เดือนกันยายน 2020 พบว่าผู้เสียชีวิตทั่วโลกทะลุ 1 ล้านคน และเมื่อเข้าสู่เดือน ตุลาคม 2020 มีผู้ติดเชื้อทั่วโลกกว่า 40 ล้านราย โดยผู้ติดเชื้อทั่วโลกมีจำนวนเพิ่มสูงขึ้นอย่างรวดเร็วและยังคงเพิ่มขึ้นต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.1.2 20 ประเทศที่มีจำนวนผู้ติดเชื้อมากที่สุดในโลก



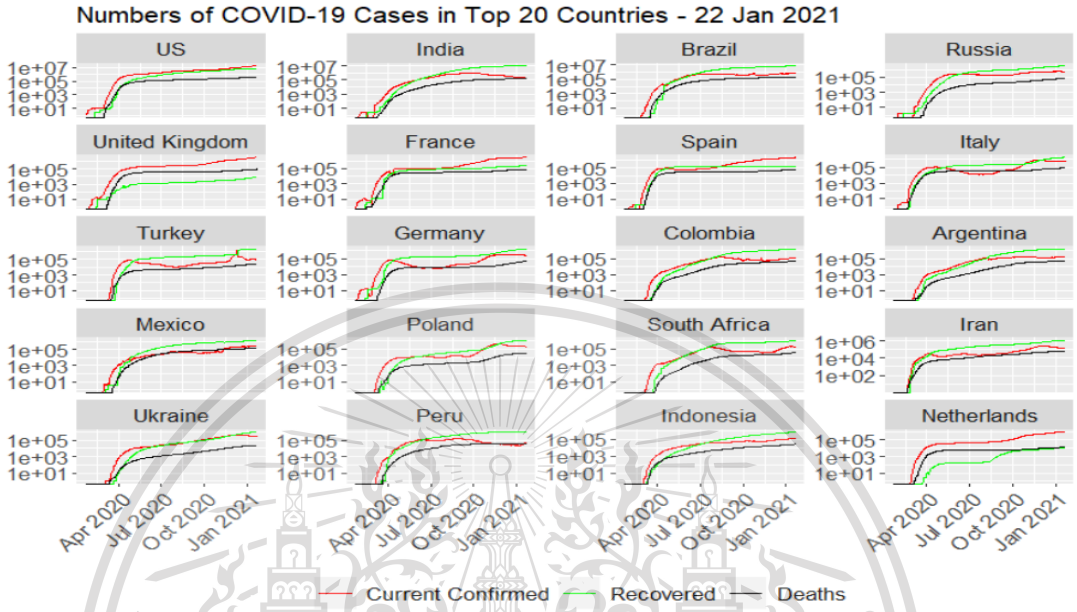
รูป 4.2 (A) อัตราการติดเชื้อ (B) อัตราการรักษาหาย และ (C) อัตราการเสียชีวิตของประเทศที่มีผู้ติดเชื้อไวรัสโคโรนาสูงสุด 20 ประเทศแรกและประเทศอื่น ๆ นอกเหนือจากประเทศที่ติดอันดับ

กราฟ (A) อัตราการติดเชื้อพบว่าสหรัฐอเมริกามีอัตราการติดเชื้อเทียบกับประชากรภายในประเทศ 6.8% เมื่อเทียบกับประเทศอื่น ๆ นอกเหนือจาก 20 อันดับแรกมีอัตราการติดเชื้อแค่ 0.4% กล่าวได้ว่าสหรัฐอเมริกามีการแพร่กระจายของโรคติดเชื้อไวรัสโคโรนา 2019 อย่างกว้างขวางภายในประเทศ นอกจากสหรัฐอเมริกาจะมีอัตราการติดเชื้อสูงแล้ว ยังพบว่าอัตราการเสียชีวิตและอัตราการรักษาหายต่ำ จึงส่งผลให้อัตราการติดเชื้อยังคงสูงอยู่ ในขณะเดียวกันสหราชอาณาจักรมีอัตราการรักษาหายต่ำอยู่ที่ 0.2% เท่านั้น ซึ่งมีอัตราการรักษาหายน้อยมาก ซึ่งหมายความว่าเมื่อมีผู้ติดเชื้อจะสามารถรักษาหายได้น้อยมากเช่นกัน อินเดียมีอัตราการรักษาหายสูงที่สุดในโลกเมื่อเทียบกับจำนวนผู้ติดเชื้อภายในประเทศ สรุปได้ว่าเมื่อมีการติดโรคติดเชื้อไวรัสโคโรนา 2019 ในอินเดียจะสามารถรักษาหายได้สูงมาก และมีอัตราการติดเชื้อ 0.8% เท่านั้นเมื่อเทียบกับประชากรภายในประเทศ ทั้งที่อินเดียมีจำนวนประชากรสูงเป็นอันดับ 2 ของโลก

กราฟ (B) อัตราการรักษาหาย พบว่าทุกประเทศมีอัตราการรักษาหายใกล้เคียงกัน ยกเว้นสหรัฐอเมริกา อังกฤษ ฝรั่งเศส สเปนและเนเธอร์แลนด์ที่มีอัตราการรักษาหายต่ำ แต่ทุกประเทศที่กล่าวมานั้นมีอัตราการเสียชีวิตใกล้เคียงกับประเทศอื่นที่มีอัตราการรักษาหายสูงกว่าทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กราฟ (C) อัตราเสียชีวิต พบว่าเม็กซิโกมีอัตราการเสียชีวิต 8.7% สูงกว่าประเทศอื่น ๆ ที่มีอัตราการเสียชีวิตไม่เกิน 5% นั่นหมายความว่าเมื่อมีผู้ติดเชื้อภายในเม็กซิโก จะมีโอกาสเสียชีวิตสูงกว่าเมื่อเทียบกับประเทศอื่น ๆ



รูป 4.3 จำนวนผู้ติดเชื้อปัจจุบัน (สีแดง) จำนวนผู้รักษาหาย (สีเขียว) และจำนวนผู้เสียชีวิต (สีเทา) ของ 20 ประเทศที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดทั่วโลก

จากกราฟแสดงจำนวนผู้ติดเชื้อปัจจุบัน จำนวนผู้รักษาหาย และจำนวนผู้เสียชีวิตของประเทศที่มีจำนวนผู้ติดเชื้อสูงสุด 20 ประเทศของโลก ประกอบด้วยสหรัฐอเมริกา อินเดีย บราซิล รัสเซีย อังกฤษ ฝรั่งเศส ตุรกี อิตาลี สเปน เยอรมนี โคลอมเบีย อาเจนตินา เม็กซิโก โปแลนด์ อิหร่าน แอฟริกาใต้ ยูเครน เปรู เนเธอแลนด์ และอินโดนีเซีย โดยกราฟแสดงจำนวนผู้ติดเชื้อ จำนวนผู้รักษาหายและจำนวนผู้เสียชีวิต แกน y เป็นจำนวนแบบ Log scale แกน x คือเดือนและปีแสดงระยะเวลาตั้งแต่เดือนมกราคม 2020 ถึงเดือนมกราคม 2021

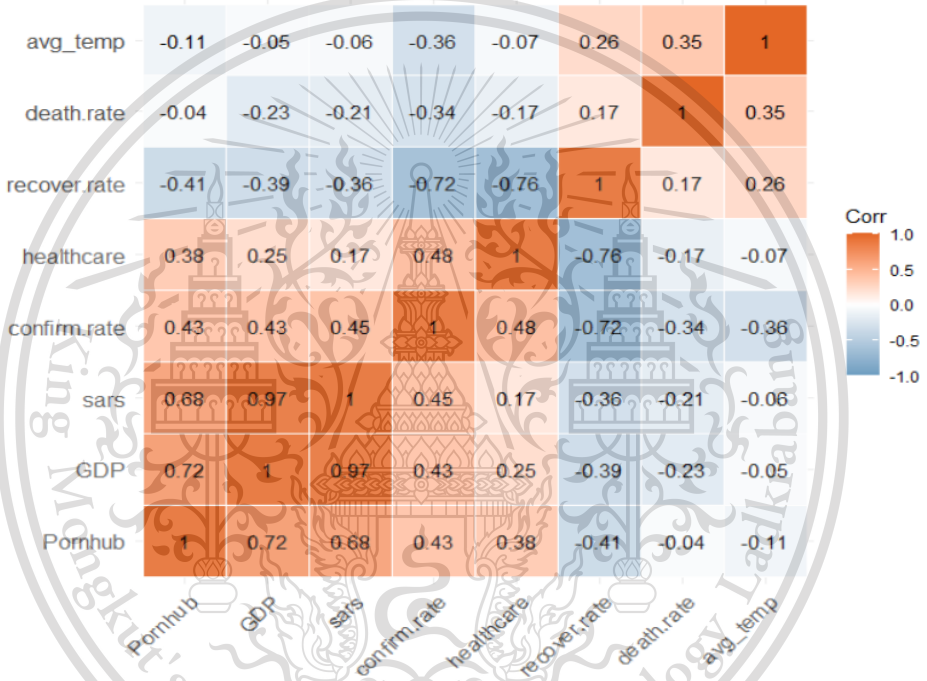
ประเทศที่มีผู้ติดเชื้อเป็นอันดับต้น ๆ ของโลก เริ่มมีจำนวนผู้ติดเชื้อเพิ่มขึ้นอย่างต่อเนื่องในช่วงเวลาที่ใกล้เคียงกัน คือ เพิ่มขึ้นสูงในช่วงเดือนเมษายน 2020 และเพิ่มขึ้นอย่างต่อเนื่องจนถึงชุดข้อมูลปัจจุบันคือเดือนมกราคม 2021 โดยในช่วงแรก ประเทศจีนที่มีผู้ติดเชื้อเป็นประเทศแรกของโลกสามารถควบคุมสถานการณ์การแพร่ระบาดภายในประเทศได้แล้ว จึงไม่ได้ติดอยู่ในอันดับต้น ๆ ทั้งที่จริงแล้วจีนนั้นมีผู้ติดเชื้อเป็นจำนวนมากในช่วงแรก หลังจากนั้นเริ่มมีการติดเชื้อสูงขึ้นอย่างรวดเร็วทำให้จำนวนผู้ติดเชื้อทั่วโลกรวมกันทะลุหนึ่งร้อยล้านคน และผู้เสียชีวิตทั่วโลกมากกว่า 2.3 ล้านคนภายในระยะเวลา 10 เดือน โดยประเทศที่มีจำนวนผู้ติดเชื้อมากที่สุดเป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

อันดับหนึ่งของโลก ได้แก่ สหรัฐอเมริกา ซึ่งมีผู้ติดเชื้อทั่วประเทศมากกว่า 26 ล้านคนซึ่งถือว่าเป็นจำนวนที่สูงมาก

จากกราฟซึ่งเป็นข้อมูล Log scale ทำให้ทราบว่าจำนวนผู้ติดเชื้อปัจจุบัน จำนวนผู้รักษาหาย และจำนวนผู้เสียชีวิตในประเทศนั้น ๆ มีทิศทางสอดคล้องกันทั้งหมด และมีแนวโน้มเพิ่มขึ้นในอนาคตอีกด้วย โดยสามารถสังเกตได้ว่าประเทศส่วนใหญ่มีจำนวนผู้รักษาหายสูงกว่าจำนวนผู้เสียชีวิต ยกเว้นในสหราชอาณาจักรและประเทศเนเธอร์แลนด์ซึ่งมีจำนวนผู้เสียชีวิตสูงกว่าจำนวนผู้รักษาหาย

4.1.3 ความสัมพันธ์ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรนา 2019



รูป 4.4 ค่า Correlation ของข้อมูลที่ใช้ในการวิเคราะห์

จากกราฟแสดงการหาค่า Correlation หรือค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูล ข้อมูลที่นำมาใช้ในการหาค่า ประกอบด้วยอัตราการรักษาหาย (recover.rate), อัตราการเสียชีวิต (death.rate), อุณหภูมิเฉลี่ยรายปี (avg_temp), อันดับสาธารณสุขโลก (healthcare), จำนวนผู้ติดเชื้อ (confirm.rate), จำนวนผู้ติดเชื้อโรคซาร์ส (sars), ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDP) และการเข้าชมเว็บไซต์ Pornhub (Pornhub) โดยหากค่า Correlation เข้าใกล้ 1 หมายถึง ข้อมูลนั้นมีความสัมพันธ์แบบแปรผันตรงกัน แต่หากข้อมูลมีค่า Correlation เข้าใกล้ -1 หมายถึง ข้อมูลนั้นมีความสัมพันธ์แบบแปรผกผันกัน โดยจากกราฟสามารถวิเคราะห์ได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

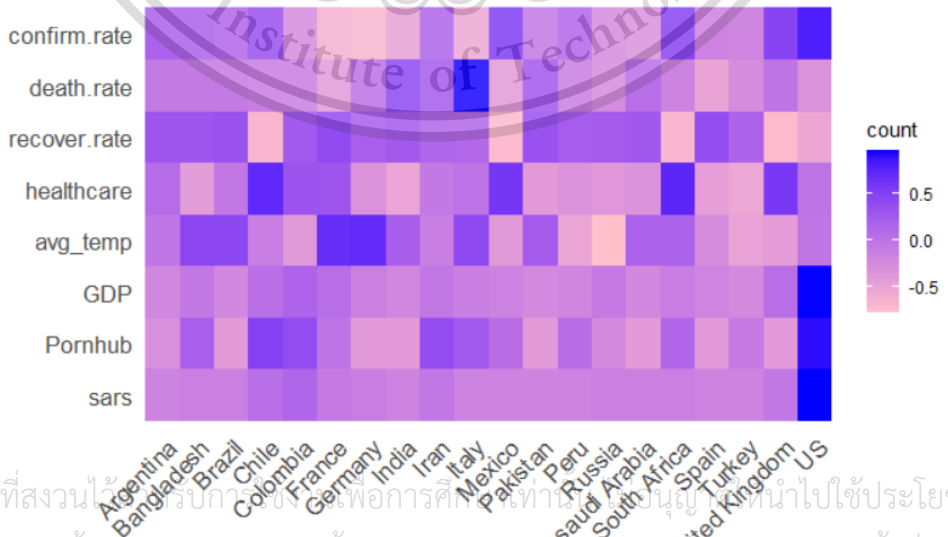
ปัจจัยที่ส่งผลต่ออัตราการติดเชื้อ พบว่าข้อมูลผลิตภัณฑ์มวลรวมในประเทศ การเข้าชมเว็บไซต์ Pornhub และอันดับสาธารณสุขโลก มีความสัมพันธ์แบบแปรผันตรงกันทั้งหมด แต่ค่าอุณหภูมิเฉลี่ยรายปีมีความสัมพันธ์แบบแปรผกผัน

ปัจจัยที่ส่งผลต่ออัตราการรักษาหาย พบว่าข้อมูลผลิตภัณฑ์มวลรวมในประเทศ การเข้าชมเว็บไซต์ Pornhub และอันดับสาธารณสุขโลก มีความสัมพันธ์แบบแปรผกผัน แต่ค่าอุณหภูมิเฉลี่ยรายปีนั้นไม่มีมีความสัมพันธ์กัน

ปัจจัยที่ส่งผลต่ออัตราการตาย พบว่าข้อมูลผลิตภัณฑ์มวลรวมในประเทศ การเข้าชมเว็บไซต์ Pornhub และอันดับสาธารณสุขโลกมีความสัมพันธ์แบบแปรผกผัน แต่ค่าอุณหภูมิเฉลี่ยรายปีมีความสัมพันธ์แบบแปรผันตรง

จากข้อมูลข้างต้นสามารถสรุปได้ว่า ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDP) การเข้าชมเว็บไซต์ Pornhub และค่าอุณหภูมิเฉลี่ยรายปีมีความสัมพันธ์ต่ออัตราการติดเชื้อ อัตราการรักษาหายและอัตราการตายน้อยมากหรือแทบไม่มีความสัมพันธ์กัน ส่วนอันดับสาธารณสุขโลกที่ได้จากการวิเคราะห์สถิติเกี่ยวกับคุณภาพโดยรวมของระบบสาธารณสุข โครงสร้างพื้นฐาน ความสามารถของบุคลากร ค่าใช้จ่ายในการรักษารายบุคคลต่อปี ความพร้อมด้านยา และความพร้อมของรัฐบาล รวมไปถึงสภาพแวดล้อม สถานที่ทางการสาธารณสุข ข้อกำหนดบทลงโทษสำหรับผู้กระทำความเสี่ยงที่ก่อให้เกิดโรค เมื่อเทียบกับอัตราการรักษาหายพบว่ามีความสัมพันธ์แบบแปรผกผันอย่างมากมีค่าสัมประสิทธิ์สหสัมพันธ์ -0.73 ทำให้สามารถวิเคราะห์ได้ว่าประเทศที่มีอันดับสาธารณสุขสูง จะมีความสามารถในการรักษาผู้ติดเชื้อภายในประเทศได้น้อยลง

4.1.4 ความถี่ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรน่า 2019ของ 20 ประเทศที่มีผู้ติดเชื้อสูงสุด



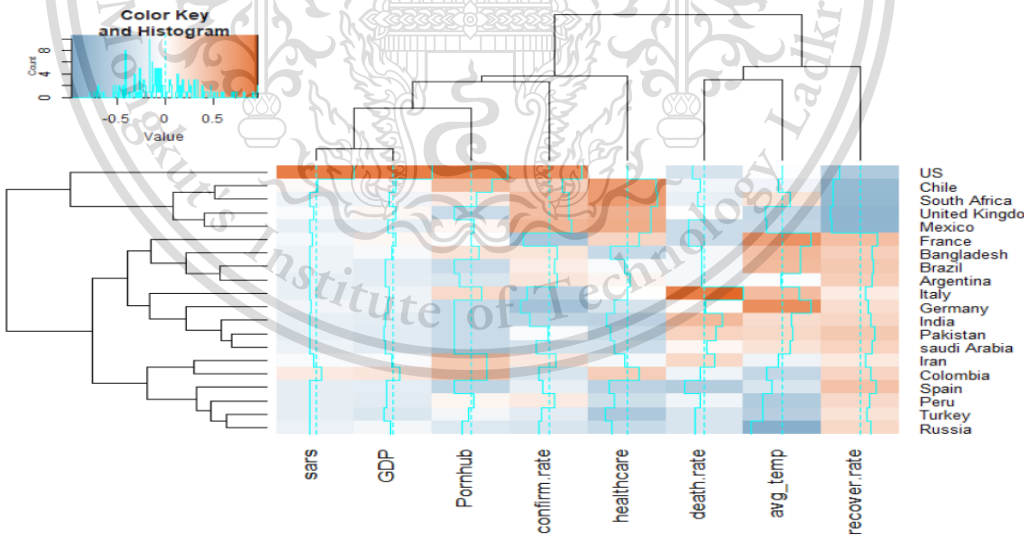
รูป 4.5 ความถี่ข้อมูลที่เกี่ยวข้องของประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากกราฟแสดงความถี่ของข้อมูล อัตราการติดเชื้อ (confirm.rate), อัตราการเสียชีวิต (death.rate), อัตราการรักษาหาย (recover.rate), อันดับสาธารณสุขโลก (healthcare), อุณหภูมิเฉลี่ยรายปี (avg_temp), ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDP), การเข้าชมเว็บไซต์ Pornhub (Pornhub) และจำนวนผู้ติดเชื้อโรคซาร์ส (sars) สามารถวิเคราะห์ได้ว่าประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับ มีข้อมูลที่สังเกตเห็นอย่างชัดเจน ดังนี้

1) ประเทศที่มีอันดับสาธารณสุขสูง คือ สหราชอาณาจักร แอฟริกาใต้ เม็กซิโก และชิลี แต่กลับมีอัตราการรักษาหายน้อยกว่าประเทศที่มีอันดับสาธารณสุขต่ำ
2) สหรัฐอเมริกามีข้อมูลผลิตภัณฑ์มวลรวมในประเทศสูงกว่าประเทศอื่น ๆ อย่างชัดเจนและมีอัตราผู้ติดเชื้อสูงด้วยเช่นกัน
3) สหรัฐอเมริกามีการเข้าชมเว็บไซต์ Pornhub ที่สูงกว่าประเทศอื่น ๆ อย่างชัดเจนและมีอัตราผู้ติดเชื้อสูงด้วยเช่นกัน
4) ไม่ว่าจะเป็นประเทศที่มีอุณหภูมิเฉลี่ยต่ำหรือสูง พบว่าไม่มีความสอดคล้องกับจำนวนอัตราผู้ติดเชื้อ เนื่องจากมีทั้งประเทศที่มีอุณหภูมิเฉลี่ยสูงแต่มีผู้ติดเชื้อต่ำและมีประเทศที่มีอุณหภูมิเฉลี่ยต่ำแต่มีผู้ติดเชื้อต่ำเช่นกัน

4.1.5 แบ่งกลุ่มความถี่ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรนา 2019 ของ 20 ประเทศที่มีผู้ติดเชื้อสูงสุดโดยใช้ Hierarchical clustering



รูป 4.6 ความถี่ข้อมูลที่เกี่ยวข้องของประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับแสดงในรูปแบบ Hierarchical clustering

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากกราฟแสดงความถี่ของข้อมูล แสดงอัตราการติดเชื้อ (confirm.rate), อัตราการเสียชีวิต (death.rate), อัตราการรักษาหาย (recover.rate), อันดับสาธารณสุขโลก (healthcare),

อุณหภูมิเฉลี่ยรายปี (avg_temp), ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDP), การเข้าชมเว็บไซต์ Pornhub (Pornhub) และจำนวนผู้ติดเชื้อโรคซาร์ส (sars) พร้อมทั้งแสดง Hierarchical clustering เพื่อจำแนกกลุ่มของข้อมูล

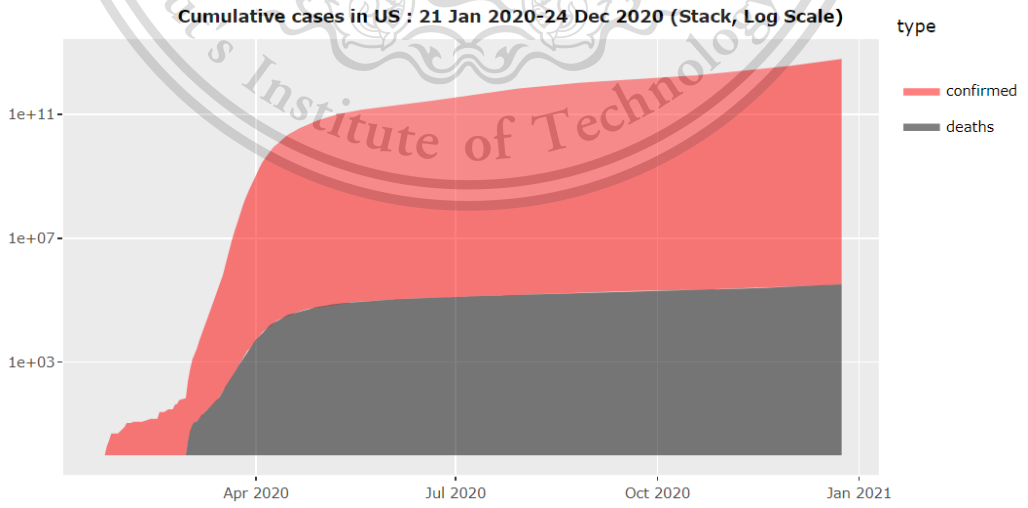
สามารถวิเคราะห์ได้ว่าจากการหา Hierarchical clustering ของประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับ พบว่าจากปัจจัยอัตราการรักษาสามารถแบ่งข้อมูลออกได้เป็น 2 กลุ่มได้อย่างชัดเจน คือ กลุ่มที่มีอัตราการรักษาหายเข้าใกล้ 1 กับกลุ่มที่มีอัตราการรักษาหายน้อยกว่า 0 นอกจากนี้ยังพบประเทศที่มีปัจจัยต่าง ๆ ที่เกี่ยวข้องใกล้เคียงกันในหลายประเทศ เช่น ตุรกีกับรัสเซีย ซึ่งมีความถี่ของปัจจัยต่าง ๆ ใกล้เคียงกันมาก

จากกราฟพบว่าจำนวนผู้ติดเชื้อโรคซาร์สกับข้อมูลผลิตภัณฑ์มวลรวมภายในประเทศที่มีผู้ติดเชื้อสูงสุด 20 อันดับนั้นมีความถี่ของข้อมูลใกล้เคียงกัน

4.2 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ศูนย์ควบคุมและป้องกันโรคในสหรัฐอเมริกา (CDC) ประกาศตรวจพบผู้ติดเชื้อในสหรัฐอเมริการายแรกในวันที่ 21 มกราคม 2020 ที่รัฐวอชิงตัน โดยผู้ติดเชื้อเดินทางกลับมาจากเมืองอู่ฮั่น ประเทศจีน หลังจากนั้นดอนัลด์ ทรัมป์ ประธานาธิบดีแห่งสหรัฐอเมริกาได้มีการประกาศภาวะฉุกเฉินด้านสาธารณสุขในสหรัฐอเมริกาในวันที่ 3 กุมภาพันธ์ 2020 โดยข้อมูลโรคติดเชื้อไวรัสโคโรนา 2020 ในสหรัฐอเมริกาที่นำมาใช้วิเคราะห์มีการรวบรวมตั้งแต่วันที่ 21 มกราคม 2020 จนถึงวันที่ 24 ธันวาคม 2020

4.2.1 จำนวนผู้ติดเชื้อและเสียชีวิตในสหรัฐอเมริกา



รูป 4.7 กราฟแสดงจำนวนผู้ติดเชื้อ (สีแดง) และจำนวนผู้เสียชีวิต (สีเทา) ในสหรัฐอเมริกา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กราฟพื้นที่แบบ Stack แสดงจำนวนผู้ติดเชื้อและผู้เสียชีวิตในสหรัฐอเมริกา โดยมีแกน x เป็นแกนเวลา แสดงเดือนและปี และแกน y เป็นจำนวนคน พื้นที่สีแดงแสดงถึงจำนวนผู้ติดเชื้อ และพื้นที่สีดำแสดงถึงจำนวนผู้เสียชีวิต โดยกราฟแสดงข้อมูลในรูปแบบ Log Scale

จากการวิเคราะห์พบว่าจำนวนผู้ติดเชื้อในสหรัฐอเมริกาช่วงแรกมีการเพิ่มขึ้นของจำนวนผู้ติดเชื้อแบบเชิงเส้น (Linear growth) ที่มีค่าความชันต่ำ แสดงให้เห็นว่าจำนวนผู้ติดเชื้อเพิ่มขึ้นอย่างช้า ๆ หลังจากนั้นพบว่าในเดือนมีนาคมมีการเพิ่มขึ้นของจำนวนผู้ติดเชื้อจากหลักสิบรายถึงหลักแสนรายภายในระยะเวลาเพียง 1 เดือน นอกจากนี้ยังพบผู้ติดเชื้อถึงหลักล้านรายครั้งแรกในเดือนเมษายนและยังคงเพิ่มขึ้นเรื่อย ๆ จนถึงเดือนพฤษภาคม ซึ่งสามารถสังเกตเห็นได้จากกราฟว่าในช่วงเดือนมีนาคมถึงเดือนพฤษภาคม จำนวนผู้ติดเชื้อเริ่มเพิ่มสูงขึ้นอย่างรวดเร็วและต่อเนื่องเป็นแบบเอ็กซ์โพเนนเชียล (Exponential growth) และหลังจากเดือนพฤษภาคมพบว่ายังคงมีการเพิ่มขึ้นของผู้ติดเชื้ออยู่ แต่เป็นการเพิ่มขึ้นเป็นแบบเส้นตรงและมีค่าความชันต่ำ แสดงให้เห็นว่าผู้ติดเชื้อรายวันมีจำนวนลดลง

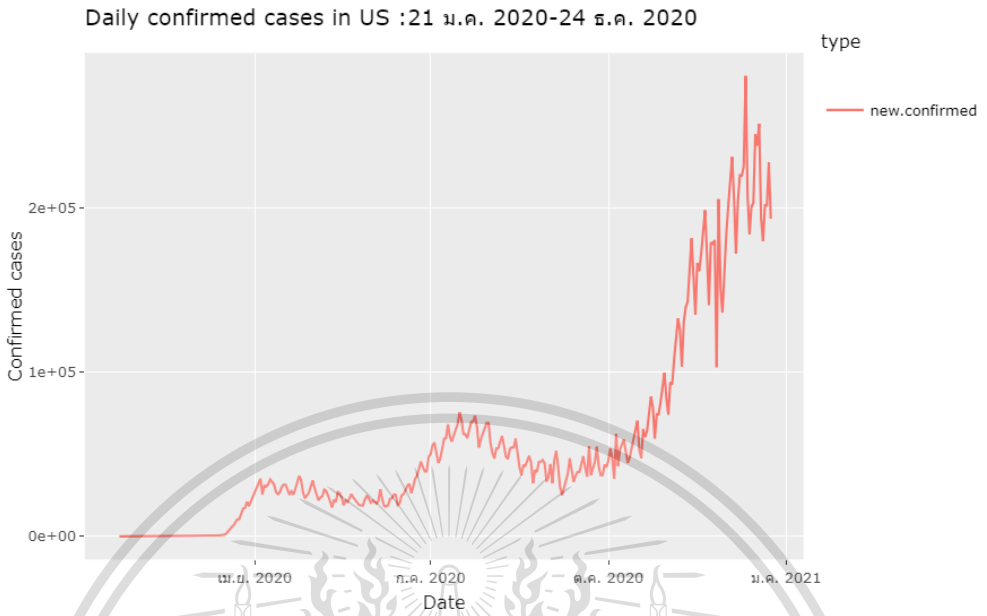
ในส่วนของจำนวนผู้เสียชีวิตในสหรัฐอเมริกาซึ่งแสดงเป็นพื้นที่สีดำในกราฟ พบว่ายังไม่มีผู้เสียชีวิตในช่วงแรก และเริ่มมีผู้เสียชีวิตหลังจากพบผู้ติดเชื้อมาในระยะหนึ่งแล้ว โดยมีการเพิ่มขึ้นของจำนวนผู้เสียชีวิตอย่างรวดเร็วและเพิ่มขึ้นในปริมาณมากอย่างต่อเนื่องแบบเอ็กซ์โพเนนเชียล จนกระทั่งหลังช่วงเดือนพฤษภาคมเป็นต้นไปพบว่ามีจำนวนผู้ติดเชื้อและผู้เสียชีวิตเพิ่มขึ้นเล็กน้อยแบบคงที่

จากข้อมูลพบว่าสหรัฐอเมริกามีการประกาศล็อกดาวน์ในช่วงกลางเดือนมีนาคม โดยมีการประกาศใช้มาตรการล็อกดาวน์ครั้งแรกที่รัฐ Puerto Rico ในวันที่ 15 มีนาคม และหลังจากนั้นก็มีการประกาศล็อกดาวน์ในรัฐอื่นตามมาในช่วงเดือนมีนาคมถึงเดือนเมษายน โดยมีช่วงเวลาประกาศใช้ที่แตกต่างกันในแต่ละรัฐ ทว่าไม่ได้มีประสิทธิภาพในการลดจำนวนผู้ติดเชื้อรายใหม่ในแต่ละวันมากนัก ดังจะเห็นได้จากมีการเพิ่มขึ้นของจำนวนผู้ติดเชื้ออย่างรวดเร็วแบบเอ็กซ์โพเนนเชียลในช่วงเวลาดังกล่าว ซึ่งแสดงให้เห็นว่าในสหรัฐอเมริกามีค่า Ro (Basic reproduction order) สูงมาก กล่าวคือ ผู้ที่ติดเชื้อในสหรัฐอเมริกามีโอกาสในการแพร่เชื้อสู่บุคคลอื่นสูง จึงทำให้เกิดการแพร่ระบาดของเชื้อในระยะเวลาอันสั้น

โดยรวมแล้วในสหรัฐอเมริกามีจำนวนผู้ติดเชื้อกว่า 18 ล้านรายและมีจำนวนผู้เสียชีวิตจากการติดเชื้อมากกว่า 329,237 ราย ซึ่งหากคิดเป็นอัตราส่วนแล้ว สามารถสรุปได้ว่าจำนวนผู้ติดเชื้อคิดเป็น 100 เท่าของจำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 ซึ่งแสดงให้เห็นว่ามีจำนวนผู้เสียชีวิตจากการติดเชื้อไม่สูงมากนักเมื่อเทียบกับจำนวนผู้ติดเชื้อทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.2 จำนวนผู้ติดเชื้อรายวันในสหรัฐอเมริกา



รูป 4.8 กราฟเส้นจำนวนผู้ติดเชื้อรายวันในสหรัฐอเมริกา

กราฟเส้นแสดงจำนวนผู้ติดเชื้อโคโรนา 2019 ในสหรัฐอเมริกา โดยแสดงข้อมูลตั้งแต่วันที่ 21 มกราคม 2020 จนถึงวันที่ 24 ธันวาคม 2020 แกน y แสดงจำนวนผู้ติดเชื้อในสหรัฐอเมริกาและแกน x แสดงวันที่ในรูปแบบเดือนและปี

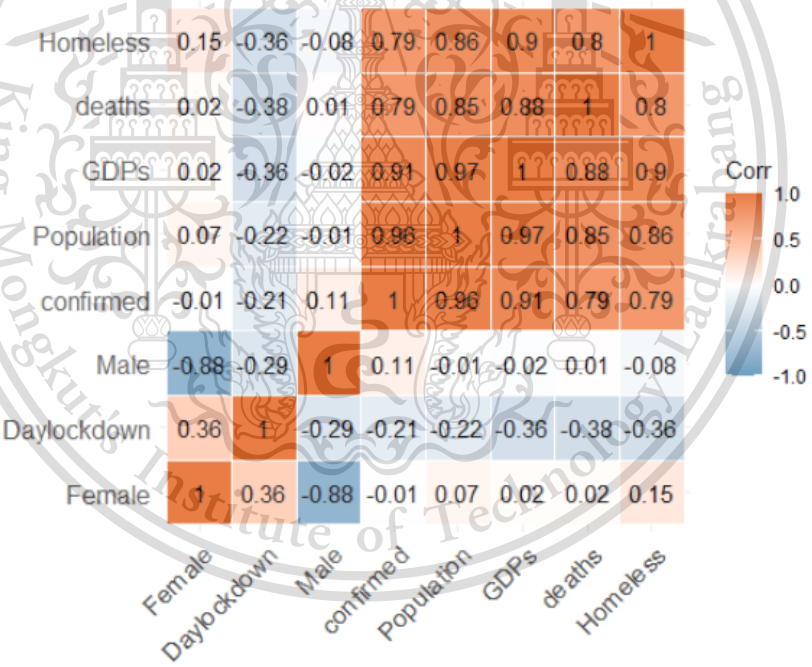
จากกราฟพบว่าในช่วงแรกมีจำนวนผู้ติดเชื้อคงที่หรือยังไม่มีผู้ติดเชื้อรายวันเพิ่มจนถึงในเดือนมีนาคมพบว่ามีการเปลี่ยนแปลงของจำนวนผู้ติดเชื้อรายใหม่อย่างเห็นได้ชัด จากข้อมูลพบว่าในช่วงต้นเดือนมีนาคมพบผู้ติดเชื้อรายวันเพิ่มขึ้นหลักร้อยคนต่อวัน จนกระทั่งสิ้นสุดเดือนมีนาคมพบว่ามียอดผู้ติดเชื้อรายวันจนถึงหลักหมื่นคน หลังจากนั้นยอดผู้ติดเชื้อรายวันยังคงเพิ่มขึ้นเรื่อย ๆ แบบคงที่โดยเฉลี่ยอยู่ที่หลักหมื่นรายต่อวันแม้จะมีการประกาศมาตรการล็อกดาวน์ในหลายพื้นที่ของสหรัฐอเมริกาแล้วก็ตาม

หลังจากนั้นเกิดการเปลี่ยนแปลงของจำนวนผู้ติดเชื้อรายวันอย่างชัดเจนอีกครั้งในเดือนกรกฎาคม โดยพบผู้ติดเชื้อรายใหม่เพิ่มขึ้นจนแตะหลักครึ่งแสนหรือห้าหมื่นคนต่อวัน จนกระทั่งในช่วงเดือนสิงหาคมถึงกันยายนเริ่มมีจำนวนผู้ติดเชื้อรายวันลดลงเล็กน้อยและคงที่โดยเฉลี่ยอยู่ที่ 20,000 – 50,000 รายต่อวัน แสดงให้เห็นว่ายังคงมีการเพิ่มขึ้นของผู้ติดเชื้ออย่างต่อเนื่องในอัตราคงที่ ต่อมาจำนวนผู้ติดเชื้อรายวันกลับมาเพิ่มขึ้นอย่างมากอีกครั้งในช่วงเดือนตุลาคมจนถึงเดือนธันวาคม โดยในช่วงนี้พบจำนวนผู้ติดเชื้อรายใหม่ถึงหลักแสนคนเป็นครั้งแรก และยังคงเพิ่มขึ้นเรื่อย ๆ โดยเฉลี่ยแล้วในช่วงนี้มีผู้ติดเชื้อรายใหม่อยู่ที่ 100,000 – 200,000 รายต่อวัน แสดงให้เห็นว่ามีการเพิ่มขึ้นของจำนวนผู้ติดเชื้อรายวันแบบเอ็กซ์โพเนนเชียลหรือเท่าทวีคูณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สรุปโดยภาพรวมแล้วในสหรัฐอเมริกามีการพบผู้ติดเชื้อรายวันเพิ่มสูงขึ้นอย่างเห็นได้ชัดในช่วงเดือนมีนาคม เดือนสิงหาคม และเดือนตุลาคม โดยจำนวนผู้ติดเชื้อรายวันแรกเริ่มอยู่ที่หลักร้อยรายจนสิ้นสุดเดือนธันวาคมที่หลักแสนรายต่อวัน จากข้อมูลการตรวจสอบเกี่ยวกับโรคติดเชื้อไวรัสโคโรนาภายในสหรัฐอเมริกา พบว่ามีรายงานผู้เสียชีวิตในวันที่ 4 มีนาคม ซึ่งมีผลการทดสอบว่าเป็นผู้ป่วยติดเชื้อไวรัสโคโรนา หลังเดินทางกลับจากเที่ยวเรือสำราญ Grand Princess ที่เดินทางระหว่าง San Francisco และ Mexico และยังพบผู้อาศัยในรัฐ Sonoma ซึ่งมีผลตรวจพบเชื้อไวรัสโคโรนาและมีประวัติการเดินทางด้วยเรือสำราญ Grand Princess เที่ยวเดียวกันด้วยเช่นกัน โดยในเที่ยวการเดินทางนั้นมีผู้โดยสารทั้งหมดกว่า 3,500 ราย นอกจากนี้ยังเริ่มมีการลงพื้นที่ตรวจแบบเชิงรุกตั้งแต่เดือนมีนาคมเป็นต้นมา ซึ่งอาจทำให้มีโอกาสพบผู้ติดเชื้อที่แฝงอยู่มากขึ้น และอาจเป็นส่วนหนึ่งที่ทำให้มีผลจำนวนผู้ติดเชื้อเพิ่มสูงขึ้นอย่างรวดเร็วในช่วงเวลาดังกล่าว

4.2.3 ความสัมพันธ์ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรนา 2019 ของประเทศสหรัฐอเมริกา



รูป 4.9 กราฟ Correlation แสดงความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับโรคติดเชื้อไวรัสโคโรนา 2019 ใน 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา

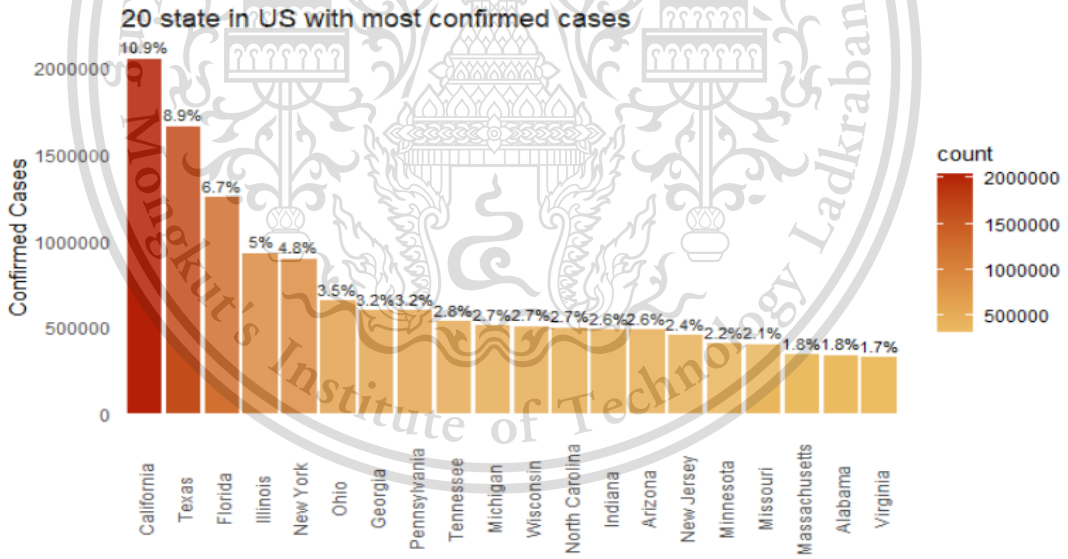
กราฟ Correlation แสดงความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับโรคติดเชื้อไวรัสโคโรนาในสหรัฐอเมริกา โดยใช้สีและตัวเลขเพื่อแสดงความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ได้แก่ ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDP), ข้อมูลจำนวนประชากร (Population), ข้อมูลประชากรไร้บ้าน (Homeless), จำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 (Deaths), จำนวนผู้ติดเชื้อไวรัสโค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โรนา 2019 (Confirmed), จำนวนประชากรเพศหญิง (Female), จำนวนประชากรเพศชาย (Male) และจำนวนวันที่ทำการล็อกดาวน์ (Daylockdown) โดยข้อมูลที่เป็นสีแดงหรือมีค่าเข้าใกล้ 1 แสดงถึงข้อมูลมีความสัมพันธ์กันในทิศทางเดียวกันอย่างมาก ข้อมูลที่มีสีฟ้าหรือมีค่าเข้าใกล้ -1 แสดงถึงข้อมูลมีความสัมพันธ์กันในทิศทางตรงกันข้ามอย่างมาก และข้อมูลที่เป็นสีขาวหรือมีค่าเป็น 0 แสดงว่าข้อมูลไม่มีความสัมพันธ์กันหรือแทบไม่มีความสัมพันธ์กันเลย

ผลจากการวิเคราะห์เมื่อเทียบข้อมูลกับจำนวนผู้ติดเชื้อพบว่าข้อมูลผลิตภัณฑ์มวลรวมในประเทศ จำนวนประชากรภายในรัฐ และจำนวนประชากรไร้บ้านมีความสัมพันธ์แบบแปรผันตรงกับจำนวนผู้ติดเชื้อเป็นอย่างมาก โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ที่ 0.79 – 0.91 และมีความสอดคล้องเป็นแบบแปรผันตรง ในส่วนของจำนวนวันที่ทำการล็อกดาวน์พบว่ามีความสอดคล้องกับจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตแบบแปรผกผัน โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตอยู่ที่ -0.21 และ -0.38 ตามลำดับ และเมื่อพิจารณาจำนวนประชากรเพศหญิงและเพศชายพบว่ามีความสอดคล้องกับจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตน้อยมากหรือแทบไม่มีความสอดคล้องกันเลย

4.2.4 จำนวนผู้ติดเชื้อสูงสุด 20 รัฐในสหรัฐอเมริกา



รูป 4.10 กราฟแท่งแสดงจำนวนผู้ติดเชื้อสูงสุด 20 รัฐในสหรัฐอเมริกา และแสดงร้อยละของจำนวนผู้ติดเชื้อในแต่ละรัฐเมื่อเทียบกับจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา

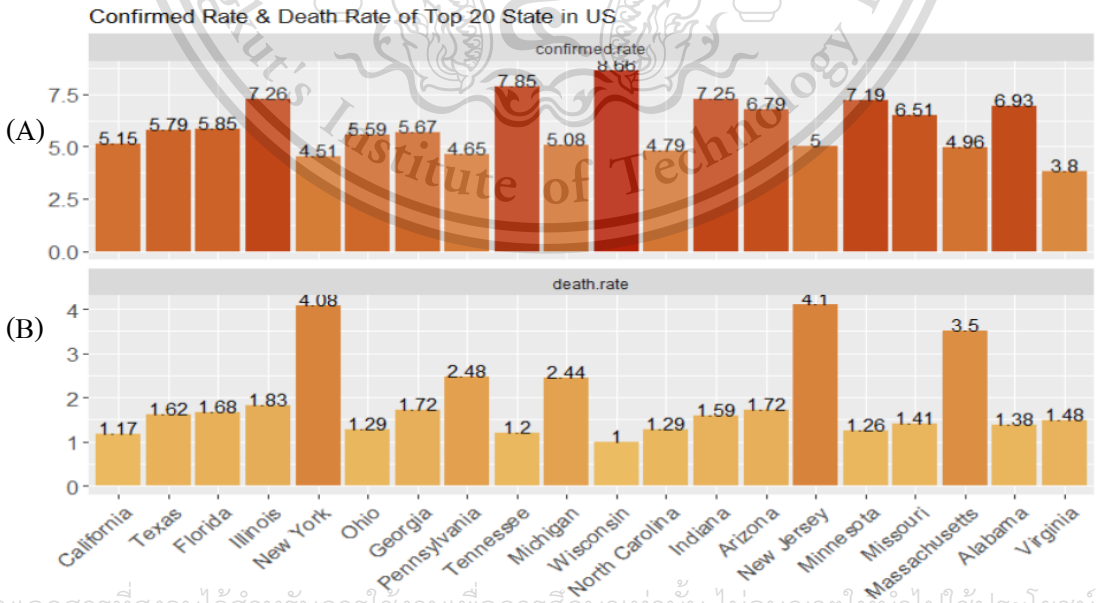
กราฟแท่งแสดงจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ของ 20 รัฐในสหรัฐอเมริกาที่มีจำนวนผู้ติดเชื้อสูงสุด โดยแกน x แสดงรายชื่อ 20 รัฐเรียงตามจำนวนผู้ติดเชื้อจากมากไปน้อย และแกน y แสดงจำนวนผู้ติดเชื้อ โดยแสดงร้อยละของจำนวนผู้ติดเชื้อในแต่ละรัฐเมื่อเทียบกับจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกาบนแกน x ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากการวิเคราะห์พบว่ารัฐ California เป็นรัฐที่มีจำนวนผู้ติดเชื้อสูงสุดในสหรัฐ-อเมริกา โดยมีจำนวนผู้ติดเชื้อทั้งหมด 2,047,919 ราย คิดเป็นร้อยละ 10.9 ของจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา รองลงมาคือรัฐ Texas และรัฐ Florida โดยมีจำนวนผู้ติดเชื้อ 1.6 ล้านราย และ 1.2 ล้านราย โดยคิดเป็นร้อยละ 8.9 และ 6.7 ตามลำดับ ซึ่งหมายความว่าจำนวนผู้ติดเชื้อรวมทั้ง 3 รัฐนี้ คิดเป็นร้อยละได้มากถึง 26.5 ของจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา และเมื่อรวมจำนวนผู้ติดเชื้อจาก 20 รัฐที่มีผู้ติดเชื้อสูงสุด พบว่ามากถึงร้อยละ 74.3 ของจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา ซึ่งหมายความว่า 20 รัฐที่มีจำนวนผู้ติดเชื้อสูงสุดนั้นมียอดรวมผู้ติดเชื้อมากกว่าครึ่งหนึ่งของผู้ติดเชื้อรวมทั้ง 50 รัฐในสหรัฐอเมริกา

จากการศึกษาพบว่าในรัฐ California มีประชากรมากที่สุด และมีพื้นที่ใหญ่เป็นอันดับที่ 3 ในสหรัฐอเมริกา นอกจากนี้ยังมีค่าผลิตภัณฑ์มวลรวมภายในรัฐสูงเป็นอันดับที่ 1 ในสหรัฐและเป็นอันดับที่ 6 ของโลก รัฐ California จึงถือได้ว่าเป็นหนึ่งในรัฐที่เป็นผู้นำกระแสต่าง ๆ รวมไปถึงเศรษฐกิจของโลก ซึ่งอาจเป็นเหตุผลหนึ่งที่ทำให้มีผู้คนแวะเวียนติดต่อในรัฐนี้ค่อนข้างมากและอาจทำให้เกิดการแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 ที่สูงมากกว่ารัฐอื่น ๆ ในสหรัฐอเมริกา ในส่วนของรัฐ Texas ที่มีผู้ติดเชื้อสูงเป็นอันดับที่ 2 ของสหรัฐอเมริกา พบว่ามีทั้งพื้นที่และจำนวนประชากรมากเป็นอันดับที่ 2 ของสหรัฐฯ รวมไปถึงรัฐ Florida ก็มีจำนวนประชากรสูงเป็นอันดับที่ 3 ของสหรัฐอเมริกาเช่นกัน จึงอาจกล่าวได้ว่าจำนวนประชากรในพื้นที่อาจเป็นหนึ่งในสาเหตุสำคัญที่ส่งผลต่อจำนวนผู้ติดเชื้อในรัฐนั้น ๆ

4.2.5 อัตราการติดเชื้อและอัตราการเสียชีวิตแบ่งตามรัฐของ 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา



รูป 4.11 แผนภูมิแท่งแสดงอัตราการติดเชื้อและอัตราการเสียชีวิตแบ่งตามรัฐของ 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

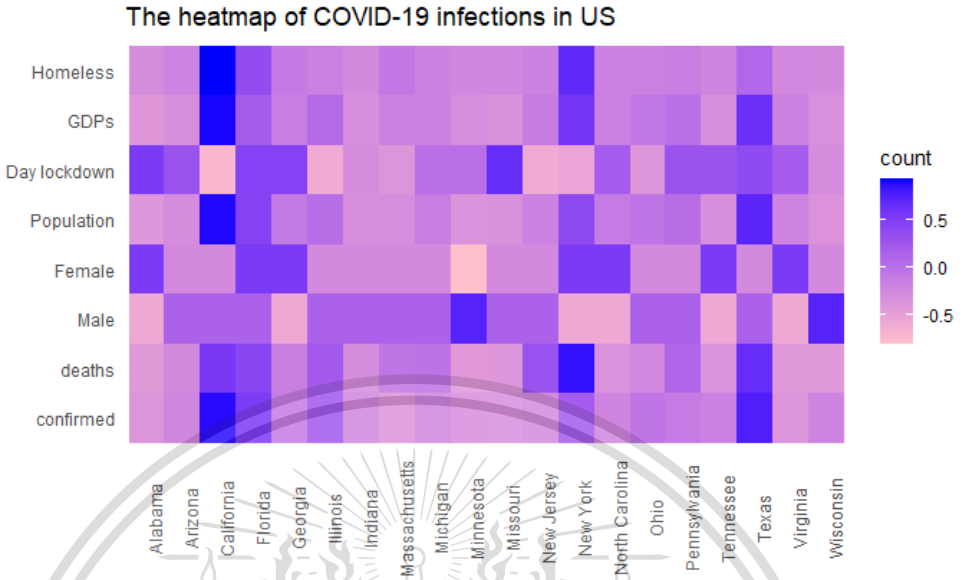
แผนภูมิแท่งแสดงอัตราการติดเชื้อและอัตราการเสียชีวิตในสหรัฐอเมริกาแบ่งตามรัฐ โดยแกน x เป็นรายชื่อรัฐที่มีจำนวนผู้ติดเชื้อสูงสุดจำนวน 20 รัฐ และแกน y เป็นอัตราการติดเชื้อและอัตราการเสียชีวิตคิดเป็นร้อยละตามลำดับ

กราฟ (A) อัตราการติดเชื้อคำนวณได้จากจำนวนผู้ติดเชื้อเทียบกับจำนวนประชากรในรัฐนั้น ๆ โดยจากการวิเคราะห์พบว่ารัฐที่มีอัตราการติดเชื้อสูงสุด ได้แก่ รัฐ Wisconsin, Tennessee, Illinois และ Minnessota คิดเป็นร้อยละ 8.66, 7.85, 7.26 และ 7.25 ตามลำดับ โดยเมื่อพิจารณาประกอบกับรูป 4.3.4 พบว่ารัฐ California เป็นรัฐที่มีจำนวนผู้ติดเชื้อสูงสุด แต่กลับมีอัตราการติดเชื้อ (confirmed.rate) คิดเป็นเพียงร้อยละ 5.15 เมื่อเทียบกับจำนวนประชากรในรัฐ ซึ่งแสดงว่ารัฐ California ไม่ได้มีจำนวนผู้ติดเชื้อสูงมากนักหากเทียบกับจำนวนประชากรทั้งหมดภายในรัฐ ในขณะที่รัฐ Wisconsin ที่มีจำนวนผู้ติดเชื้อรวมเป็นอันดับที่ 11 ในสหรัฐอเมริกากลับมีอัตราการติดเชื้อเทียบกับจำนวนประชากรในรัฐถึงร้อยละ 8.66 ซึ่งถือเป็นรัฐที่มีอัตราการติดเชื้อสูงสุดในสหรัฐอเมริกา จากข้อมูลข้างต้นแสดงให้เห็นว่าในรัฐ California มีประชากรมากกว่า แต่มีอัตราการติดเชื้อของประชากรในรัฐต่ำกว่า ซึ่งสามารถสรุปได้ว่าในรัฐ California มีการกระจายตัวของโรคติดเชื้อไวรัสโคโรนา 2019 ต่ำเมื่อเทียบกับรัฐอื่น ๆ ในขณะที่รัฐ Wisconsin มีจำนวนประชากรน้อยกว่า แต่มีอัตราการติดเชื้อไวรัสโคโรนาภายในรัฐค่อนข้างสูง ซึ่งสามารถสรุปได้ว่ามีการกระจายตัวของโรคติดเชื้อไวรัสโคโรนา 2019 สูงเมื่อเทียบกับรัฐอื่น ๆ

กราฟ (B) อัตราการเสียชีวิต (Crude Fatality Rate: CFR) คำนวณได้จากจำนวนผู้เสียชีวิตเมื่อเทียบกับจำนวนผู้ติดเชื้อในรัฐนั้น ๆ โดยจากการวิเคราะห์พบว่า 3 รัฐที่มีอัตราการเสียชีวิตสูงสุด ได้แก่ รัฐ New Jersey, New York และ Massachusetts ซึ่งคิดเป็นร้อยละ 4.1, 4.08 และ 3.5 ตามลำดับ โดยในรัฐเหล่านี้แม้จะมีอัตราการเสียชีวิตหรือค่า CFR สูง แต่กลับพบว่าไม่ได้มีอัตราการติดเชื้อสูงมากนักเมื่อเทียบกับรัฐอื่น ๆ ในสหรัฐอเมริกา (ร้อยละ 5, ร้อยละ 4.51 และร้อยละ 4.96 ตามลำดับ) จึงสามารถสรุปได้ว่าผู้ที่ติดเชื้อในรัฐดังกล่าว จะมีอัตราการเสียชีวิตที่สูง ในขณะที่รัฐที่มีจำนวนผู้ติดเชื้อสูงสุดอย่างรัฐ California กลับมีอัตราการเสียชีวิตเพียงร้อยละ 1.17 ซึ่งแสดงให้เห็นว่าประชากรในรัฐ California ที่ติดเชื้อ จะมีอัตราในการเสียชีวิตที่ต่ำกว่า รวมไปถึงรัฐ Wisconsin ที่ถึงแม้จะมีอัตราการติดเชื้อสูงสุดเมื่อเทียบกับรัฐอื่น ๆ แต่มีอัตราการเสียชีวิตเพียงร้อยละ 1 เท่านั้น ซึ่งหมายความว่าผู้ติดเชื้อในรัฐ Wisconsin มีอัตราการเสียชีวิตต่ำเช่นกัน จึงสามารถสรุปได้ว่าอัตราการติดเชื้อภายในรัฐไม่ส่งผลต่ออัตราการเสียชีวิตภายในรัฐ กล่าวคือรัฐที่มีจำนวนผู้ติดเชื้อสูง ไม่ได้หมายความว่าในรัฐนั้นจะมีจำนวนผู้เสียชีวิตสูงตามไปด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.6 ความถี่ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรนา 2019 ของสหรัฐอเมริกา



รูป 4.12 กราฟ Heatmap แสดงความถี่ของข้อมูลที่เกี่ยวข้องกับโรคติดเชื้อไวรัสโคโรนา 2019 ของ 20 รัฐที่มีผู้ติดเชื้อสูงสุดในสหรัฐอเมริกา

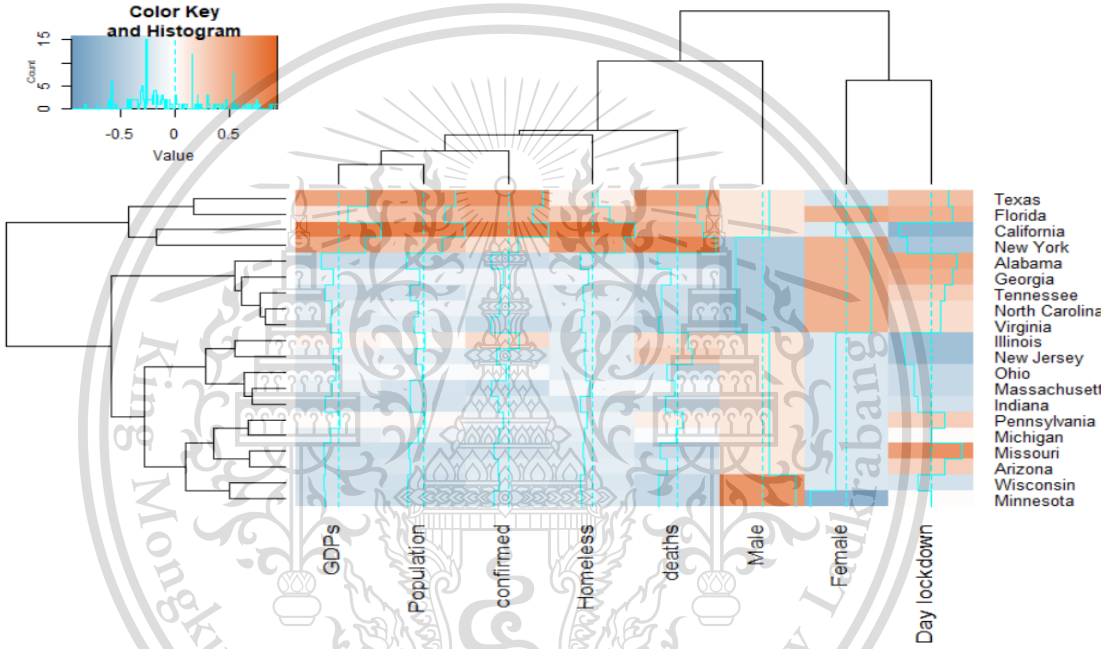
กราฟ Heatmap แสดงถึงความถี่ของข้อมูลผ่านคู่สี มีค่าตั้งแต่ -1 จนถึง 1 ข้อมูลที่มีปริมาณมากจะมีค่าเท่ากับ 1 หรือเข้าใกล้ 1 และปรากฏบนแผนภูมิ Heatmap เป็นสีน้ำเงิน ส่วนข้อมูลที่มีปริมาณน้อยจะมีค่าเท่ากับ -1 หรือเข้าใกล้ -1 และปรากฏบนแผนภูมิ Heatmap เป็นสีชมพู กราฟประกอบด้วยแกน x เป็นรายชื่อรัฐที่มีจำนวนผู้ติดเชื้อสูงสุดในสหรัฐอเมริกาจำนวน 20 รัฐ และแกน y เป็นข้อมูลที่นำมาวิเคราะห์ร่วมด้วย ได้แก่ ข้อมูลประชากรไร้บ้าน (Homeless), ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (GDPs), จำนวนวันที่ทำการล็อกดาวน์ (Daylockdown), ข้อมูลจำนวนประชากรในรัฐ (Population), จำนวนประชากรเพศหญิง (Female), จำนวนประชากรเพศชาย (Male), จำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 (Deaths) และจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 (Confirmed)

จากกราฟสามารถวิเคราะห์และสรุปผลได้ว่ารัฐที่มีผลิตภัณฑ์มวลรวมในประเทศ (GDP) สูง จำนวนประชากรไร้บ้านสูง และจำนวนประชากรในรัฐมากจะมีจำนวนผู้ติดเชื้อและผู้เสียชีวิตสูงมากขึ้นตามไปด้วย ดังจะเห็นได้จากรัฐ California ที่มีข้อมูลผลิตภัณฑ์มวลรวมในประเทศ จำนวนประชากรไร้บ้าน และจำนวนประชากรในรัฐสูงซึ่งข้อมูลเหล่านี้ปรากฏบนกราฟ Heatmap เป็นสีน้ำเงิน แสดงให้เห็นว่ามีข้อมูลในส่วนดังกล่าวเป็นจำนวนมาก และในขณะเดียวกันก็มีจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตสูงเช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ในขณะที่ข้อมูลจำนวนวันที่ทำการล็อกดาวน์ จำนวนประชากรเพศหญิง และจำนวนประชากรเพศชายแทบไม่มีความสัมพันธ์กับจำนวนผู้ติดเชื้อหรือผู้เสียชีวิต ดังจะเห็นได้จากรัฐ California ที่มีจำนวนวันที่ทำการล็อกดาวน์น้อย มีจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตค่อนข้างมาก และหากพิจารณารัฐ Texas ที่มีจำนวนวันที่ทำการล็อกดาวน์ค่อนข้างมาก กลับมีจำนวนผู้ติดเชื้อมากเช่นกัน สรุปโดยรวมพบว่ากราฟ Heatmap ที่แสดงปริมาณของปัจจัยที่ส่งผลต่อโรคติดเชื้อไวรัสโคโรนา 2019 มีข้อมูลที่สอดคล้องกับกราฟแสดงความสัมพันธ์ (รูป 4.3.3) ที่กล่าวถึงก่อนหน้านี้

4.2.7 ความถี่ของข้อมูลที่เกี่ยวข้องกับโรคไวรัสโคโรนา 2019 ของสหรัฐอเมริกา



รูป 4.13 ความถี่ข้อมูลที่เกี่ยวข้องของสหรัฐอเมริกาแสดงในรูปแบบ Hierarchical clustering

จากกราฟแสดงความถี่ของข้อมูลประชากรไร้บ้าน (Homeless), ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ(GDPs), จำนวนวันที่ทำการล็อกดาวน์ (Daylockdown), ข้อมูลจำนวนประชากรในรัฐ (Population), จำนวนประชากรเพศหญิง (Female), จำนวนประชากรเพศชาย (Male), จำนวนผู้เสียชีวิต (Deaths) และจำนวนผู้ติดเชื้อ
(Confirmed) พร้อมทั้งแสดง Hierarchical clustering เพื่อจำแนกกลุ่มของข้อมูล

จากการหา Hierarchical clustering ของประเทศสหรัฐอเมริกาพบว่าจากปัจจัยต่าง ๆ สามารถแบ่งข้อมูลออกเป็น 2 กลุ่มอย่างชัดเจนคือ กลุ่มที่ความถี่ของข้อมูลผลิตภัณฑ์มวลรวมในประเทศ ข้อมูลจำนวนประชากรในรัฐ จำนวนผู้ติดเชื้อ ข้อมูลประชากรไร้บ้าน และจำนวนผู้เสียชีวิต

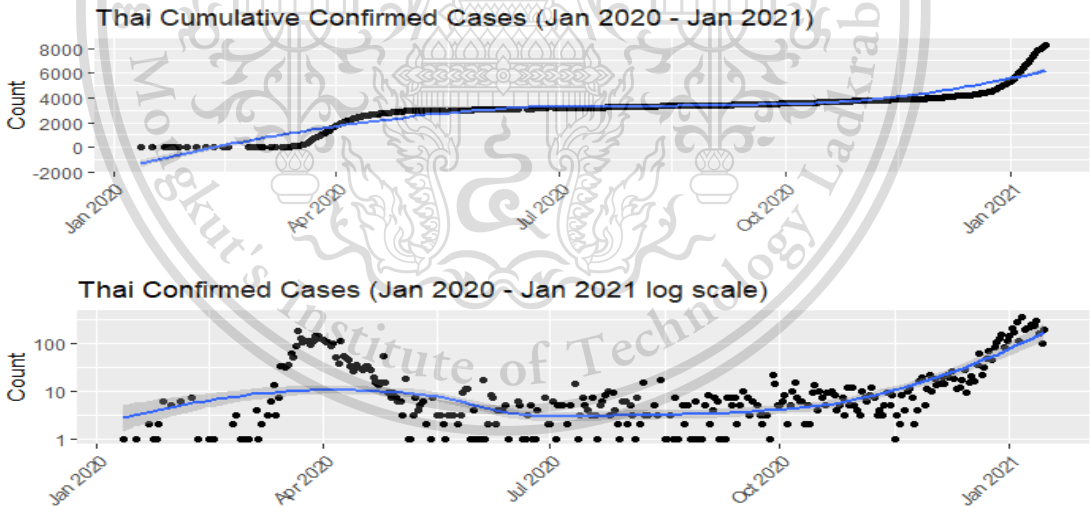
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

มีค่าเข้าใกล้ 1 ได้แก่ เท็กซัส ฟลอริดา แคลิฟอร์เนียและนิวยอร์ก กับรัฐอื่น ๆ ซึ่งปัจจัยที่กล่าวมาข้างต้นมีค่าน้อยกว่า 0

4.3 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย

ประเทศไทยมีการแพร่ระบาดไวรัสโคโรนา 2019 ครั้งแรกเมื่อวันที่ 12 มกราคม 2020 โดยเป็นประเทศที่มีผู้ป่วยยืนยันรายแรกนอกประเทศจีน ซึ่งเป็นนักท่องเที่ยวหญิงวัย 61 ปี สัญชาติจีน มีภูมิลำเนาอยู่ที่เมืองอู่ฮั่น ประเทศจีน ได้เดินทางออกจากเมืองอู่ฮั่นมายังท่าอากาศยานนานาชาติสุวรรณภูมิ พบว่ามีอาการเจ็บคอ มีไข้ หนาวสะท้าน และปวดหัว ศูนย์ปฏิบัติการภาวะฉุกเฉิน กรมควบคุมโรคได้ติดตามอาการผู้สัมผัสความเสี่ยงสูง 40 คน และรวบรวมข้อมูลสำหรับติดตามผู้สัมผัสความเสี่ยงต่ำอีก 145 คน ผลทดสอบหาโคโรนาไวรัสสายพันธุ์ใหม่เป็นบวก ต่อมาในวันที่ 31 มกราคม ชายไทยวัย 50 ปี ซึ่งขับแท็กซี่ในกรุงเทพมหานคร ได้รับผลตรวจว่าติดติดเชื้อไวรัสโคโรนา 2019 เนื่องจากรับผู้โดยสารชาวจีนจากเมืองอู่ฮั่นซึ่งมีอาการป่วยไปส่งโรงพยาบาล ถือว่า ชายคนนี้เป็นคนไทยรายแรกที่ติดไวรัสโคโรนาสายพันธุ์ใหม่ โดยไม่เคยมีประวัติเดินทางไปประเทศจีนมาก่อน นับได้ว่าเป็นจุดเริ่มต้นของการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย

4.3.1 จำนวนผู้ติดเชื้อในประเทศไทย



รูป 4.14 จำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

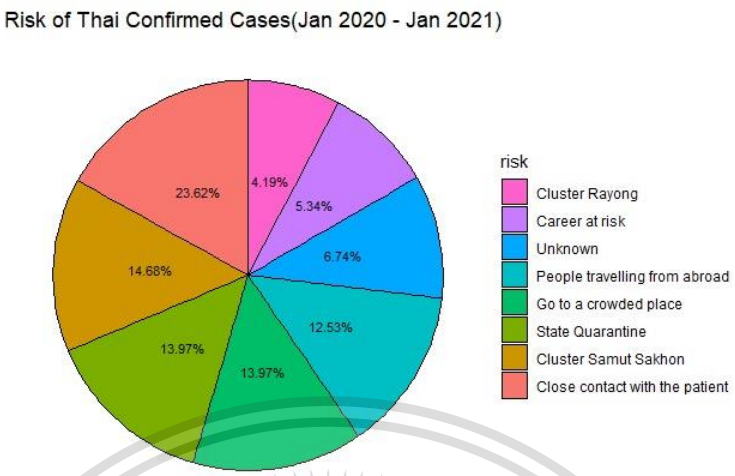
กราฟแสดงจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย ตั้งแต่วันที่ 1 มกราคม 2020 ถึงวันที่ 15 มกราคม 2021 ซึ่งกราฟด้านบนแสดงจำนวนผู้ติดเชื้อสะสม และกราฟด้านล่างเป็นกราฟ Log scale แสดงจำนวนผู้ติดในแต่ละวัน โดยแกน y คือจำนวนคน แกน x คือวันที่ เส้นสีน้ำเงินคือ Smoothed conditional means และแถบสีเทาคือ 95% Confidence Interval

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ช่วงเดือนมกราคมจนถึงกลางเดือนมีนาคมประเทศไทยพบผู้ติดเชื้อไวรัสโคโรนา 2019 น้อยมาก จำนวนผู้ติดเชื้อที่ตรวจพบยังไม่ถึง 100 คน ซึ่งในวันที่ 4 มกราคม 2020 ประเทศไทยได้เปิดศูนย์ปฏิบัติการภาวะฉุกเฉินหรือศูนย์อีโอซี (EOC) ขึ้นเพื่อรับมือกับสถานการณ์การแพร่ระบาดของไวรัสโคโรนา 2019 โดยเริ่มคัดกรองผู้โดยสารเที่ยวบินตรงจากเมืองอู่ฮั่น ประเทศจีนต้นตอการระบาด แต่เมื่อเข้าสู่ช่วงกลางเดือนมีนาคมจำนวนผู้ติดเชื้อรายใหม่ก็เติบโตขึ้นอย่างรวดเร็วแบบ Exponential ซึ่งทำให้จำนวนผู้ติดเชื้อสะสมเพิ่มแบบ Exponential ตามไปด้วย กล่าวคือ ในช่วงแรกการเติบโตของจำนวนผู้ติดเชื้อจะเกิดขึ้นอย่างช้า ๆ และจะเติบโตขึ้นเรื่อย ๆ เมื่อเวลาผ่านไป ช่วงเวลานั้นมีการตรวจพบผู้ป่วยติดเชื้อไวรัสโคโรนา 2019 จากการไปสนามมวยลุมพินี ทำให้เกิด Super spreader เป็นครั้งแรกในประเทศไทย ต้นเดือนเมษายนได้มีการประกาศพระราชกำหนดหรือ พ.ร.ก.ฉุกเฉิน (เคอร์ฟิว) โดยประกาศห้ามประชาชนออกนอกเคหสถานทั่วราชอาณาจักร สั่งห้ามไม่ให้คนต่างชาติและคนไทยเดินทางเข้าประเทศไทย เพื่อลดการแพร่ระบาดไวรัสโคโรนา 2019 เมื่อเข้าสู่ช่วงเดือนพฤษภาคมจนถึงเดือนพฤศจิกายนจะพบว่าจำนวนผู้ติดเชื้อไม่ได้เติบโตแบบ Exponential อีกต่อไป ยอดจำนวนผู้ติดเชื้อลดลงและเริ่มคงที่เติบโตในลักษณะ S-curve หรือการเติบโตแบบ Logistic กล่าวคือ กราฟมีอัตราการเติบโตเร็วในช่วงแรกและเริ่มช้าลงเรื่อย ๆ เมื่อเวลาผ่านไป บางวันไม่มีรายงานตรวจพบผู้ติดเชื้อ และไม่พบผู้ติดเชื้อต่อกันเป็นเวลาหลายวัน ประเทศไทยมีการประกาศเริ่มบังคับใช้มาตรการผ่อนปรน และสุดท้ายคือช่วงเดือนธันวาคมพบว่าจำนวนผู้ติดเชื้อกลับมาเติบโตแบบ Exponential อีกครั้ง มีการเปิดเผยว่าผลตรวจเชื้อโคโรนาของแรงงานต่างด้าวในจังหวัดสมุทรสาครพบผู้ติดเชื้อจำนวนมาก ศูนย์บริหารสถานการณ์แพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 หรือ ศบค. รายงานว่า พบผู้ติดเชื้อลามไป 22 จังหวัด และในเดือนมกราคม ศบค. สั่งห้ามเข้าออกพื้นที่ 5 จังหวัด ประกอบด้วย สมุทรสาคร ชลบุรี ระยอง จันทบุรี และ ตราด ยกเว้นเหตุจำเป็นหรือเหตุฉุกเฉิน และยอดผู้ติดเชื้อรายใหม่ยังเติบโตเพิ่มสูงขึ้นอย่างต่อเนื่องถึงวันที่ 15 มกราคม 2021 ในส่วนของจำนวนผู้เสียชีวิตในสหรัฐอเมริกาซึ่งแสดงเป็นพื้นที่สีดำในกราฟพบว่ายังไม่มีผู้เสียชีวิตในช่วงแรก และเริ่มมีผู้เสียชีวิตหลังจากพบผู้ติดเชื้อมาในระยะหนึ่งแล้ว โดยมีการเพิ่มขึ้นของจำนวนผู้เสียชีวิตอย่างรวดเร็วและเพิ่มขึ้นในปริมาณมากอย่างต่อเนื่องแบบเอ็กซ์โพเนนเชียล จนกระทั่งหลังช่วงเดือนพฤษภาคมเป็นต้นไปพบว่ามีจำนวนผู้ติดเชื้อและผู้เสียชีวิตเพิ่มขึ้นเล็กน้อยแบบคงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.3.2 จำนวนผู้ติดเชื้อในประเทศไทย



รูป 4.15 จำนวนกลุ่มความเสี่ยงจากการติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

จากกราฟวงกลมแสดงความเสี่ยงที่ผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย โดยแบ่งเป็น 8 กลุ่ม ระบุข้อมูลจากกรมควบคุมโรค กระทรวงสาธารณสุขแห่งประเทศไทย ซึ่งเป็นกลุ่มความเสี่ยง 8 อันดับที่พบมากที่สุดดังนี้

A) Close contact with the patient คือ กลุ่มที่สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้าหรือสัมผัสผู้ที่เดินทางมาจากต่างประเทศ คิดเป็นร้อยละ 23.62

B) Cluster Samut Sakhon คือ กลุ่มที่พบการติดเชื้อเป็นกลุ่มใหญ่ที่ตลาดกลางกุ้ง ตำบลมหาชัย อำเภอเมืองสมุทรสาคร จังหวัดสมุทรสาคร คิดเป็นร้อยละ 14.68

C) State Quarantine คือ กลุ่มผู้ที่เดินทางเข้ามาในราชอาณาจักรไทย ทั้งชาวไทย ชาวต่างชาติ และไม่มีอาการที่เข้าเกณฑ์สอบสวนโรคติดเชื้อไวรัสโคโรนา 2019 โดยใช้สถานประกอบการธุรกิจโรงแรม คิดเป็นร้อยละ 13.97

D) Go to a crowded place คือ กลุ่มที่ไปสถานที่ชุมชนหรือที่แออัด ได้แก่ สถานบันเทิง สนามมวย พิธีกรรมทางศาสนา สถานที่ท่องเที่ยวและคอนเสิร์ต คิดเป็นร้อยละ 13.97

E) People travelling from abroad คือ กลุ่มที่เดินทางมาจากต่างประเทศ ทั้งชาวต่างชาติหรือคนไทย คิดเป็นร้อยละ 12.53

F) Unknown คือ กลุ่มที่ไม่ระบุหรือไม่ทราบสาเหตุความเสี่ยงในการติดเชื้อ คิดเป็นร้อยละ 6.74

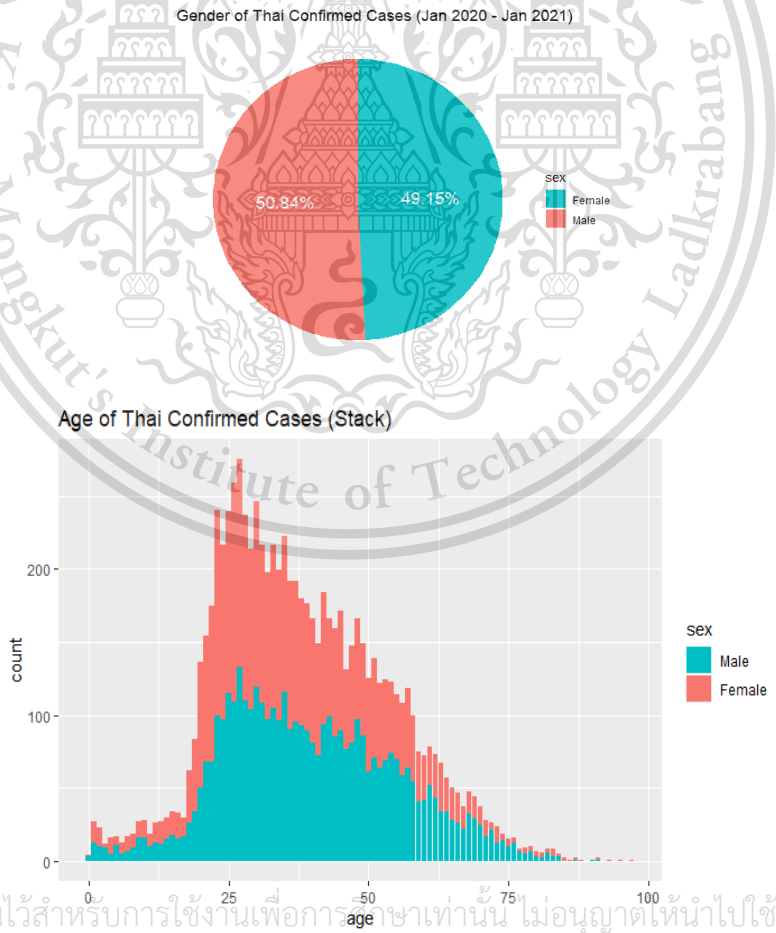
G) Career at risk คือ กลุ่มที่ทำอาชีพเสี่ยง ได้แก่ บุคลากรด้านการแพทย์ สาธารณสุข ทำงานในสถานที่แออัด หรือทำงานใกล้ชิดกับชาวต่างชาติ คิดเป็นร้อยละ 5.34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

H) Cluster Rayong คือ กลุ่มที่พบผู้ติดเชื้อจากบ่อนการพนัน จังหวัดระยอง คิดเป็นร้อยละ 4.19

จากจำนวนผู้ติดเชื้อทั้งหมดในประเทศไทย ซึ่งสามอันดับแรกที่กล่าวไป รวมกันสูงถึงประมาณร้อยละ 50 จากกลุ่มความเสี่ยงทั้งหมดที่พบ จะเห็นได้ว่าสาเหตุหลักที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 มาจากจากการสัมผัสหรือใกล้ชิดกับผู้ป่วย และอยู่ในสถานที่ที่แออัดมีจำนวนคนมาก อีกหนึ่งความน่ากลัวของโรคคือความสามารถในการแพร่กระจาย ตัวชี้วัดหนึ่งที่ใช้กันคือค่า Ro หรือ Basic reproduction number สามารถบอกได้ว่าโดยเฉลี่ยแล้วผู้ติดเชื้อหนึ่งคนสามารถแพร่เชื้อให้ผู้ที่ยังไม่ติดเชื้อได้กี่คน ถ้า Ro มากกว่า 1 จะทำให้เกิดโรคระบาด คนติดเชื้อใหม่จะเพิ่มขึ้นเรื่อย ๆ แต่ถ้า Ro น้อยกว่า 1 จะไม่เกิดโรคระบาด คนติดเชื้อใหม่จะลดลงเรื่อย ๆ ดังนั้นจากกลุ่มความเสี่ยงสรุปได้ว่าการปฏิสัมพันธ์ของผู้ป่วยมีผลต่อค่า Ro ยิ่งผู้ป่วยพบปะผู้คนจำนวนมาก Ro จะยิ่งสูงขึ้น ดังนั้น หนึ่งในวิธีที่จะสามารถควบคุมได้เพื่อลดค่า Ro คือหลีกเลี่ยงการไปอยู่ในที่แออัดหรือมีคนจำนวนมาก หรือกักตัวอยู่บ้าน

4.3.3 อายุและเพศผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย



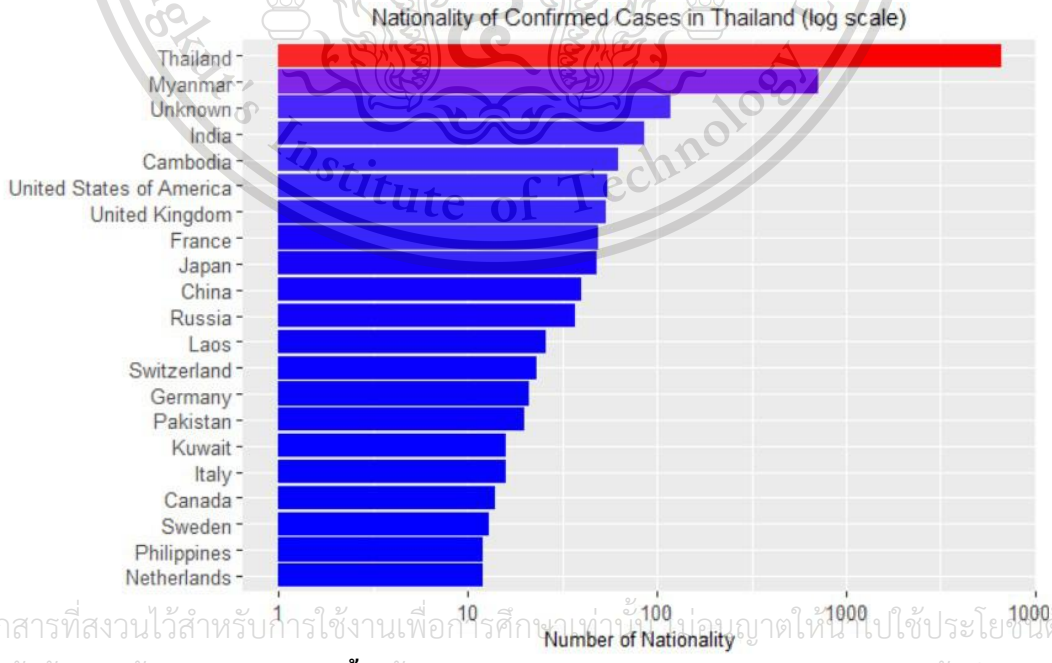
รูป 4.16 อายุ เพศชาย(สีฟ้า)และเพศหญิง(สีส้ม) ของผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากกราฟด้านซ้ายแสดงจำนวนของอายุผู้ติดเชื้อไวรัสโคโรนา 2019 โดยแบ่งตามเพศ กำหนดให้เป็นสีฟ้าเป็นเพศชาย และสีชมพูเป็นเพศหญิง ซึ่งเป็นกราฟแท่งแบบ Stack ส่วนพื้นที่สีขาวที่ขาดหายไปหมายถึงไม่พบจำนวนผู้ติดเชื้อในช่วงอายุนั้น ๆ แกน x คืออายุของผู้ติดเชื้อ ส่วนแกน y คือจำนวนผู้ติดเชื้อ

จำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 แบ่งตามเพศ โดยภาพรวมแล้วพบว่ามีจำนวนผู้ติดเชื้อเพศชายร้อยละ 50.84 และเพศหญิงร้อยละ 49.15 ซึ่งเป็นจำนวนที่ใกล้เคียงกันมาก โอกาสที่จะติดเชื้อในแต่ละเพศไม่ต่างกันมากนัก หากพิจารณาผู้ติดเชื้อแบ่งตามช่วงอายุ พบว่ามีผู้ติดเชื้อเกือบทุกช่วงของอายุ โดยผู้ติดเชื้อจำนวนมากอยู่ระหว่างช่วงอายุ 21 – 30 ปี และพบผู้ติดเชื้อจำนวนน้อยในช่วงอายุต่ำกว่า 20 ปีและสูงกว่า 60 ปี ซึ่งอายุ 25 ปีเป็นอายุที่พบผู้ติดเชื้อสูงที่สุด จะเห็นได้ว่าในช่วงอายุที่พบมากที่สุดคือช่วงอายุวัยทำงานและมีครอบครัว ซึ่งส่วนมากต้องออกไปทำงานข้างนอก ออกไปในสถานที่ที่เสี่ยงต่อการติดเชื้อ หรือออกไปพบผู้คนจำนวนมาก จึงมีโอกาสเสี่ยงที่จะได้รับการติดเชื้อไวรัสโคโรนา 2019 นอกจากนี้เมื่อกลุ่มช่วงอายุดังกล่าวทำงานหรือทำกิจกรรมที่เสี่ยงต่อการติดเชื้อเสร็จสิ้นก็กลับไปยังที่พักอาศัยของตน ซึ่งส่วนมากจะอาศัยอยู่กับครอบครัวอาจมีทั้งเด็กและผู้ใหญ่ภายในที่พักอาศัยทำให้เสี่ยงต่อการติดเชื้อไปด้วยทำให้ยากต่อการควบคุมโรคระบาดเป็นอย่างมาก หากโรคระบาดลุกลามจนควบคุมไม่ได้หนทางสุดท้าย คือต้องทำให้ประชากรเกิดภูมิคุ้มกันกลุ่ม (Herd immunity) เมื่อประชากรมีภูมิคุ้มกันจนถึงสัดส่วน 1-1/ Ro การระบาดของโรคจะยุติลง

4.3.4 สัญชาติผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย



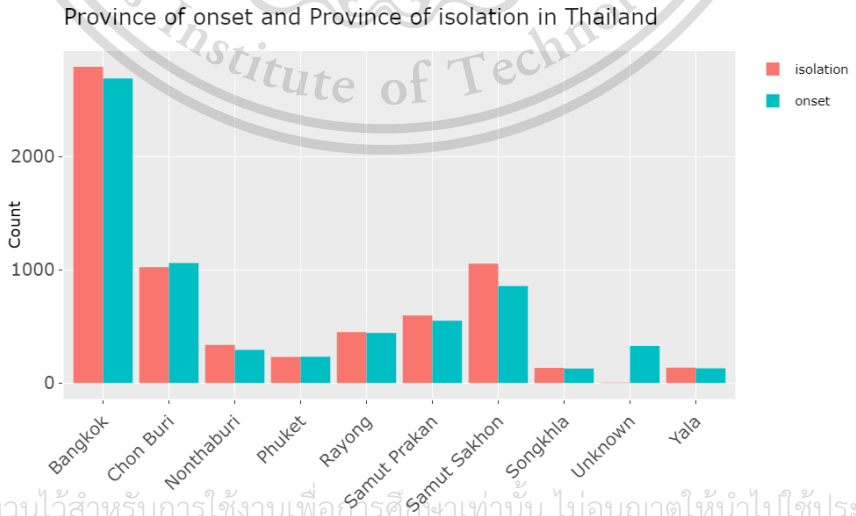
รูป 4.17 สัญชาติผู้ติดเชื้อจากการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากกราฟแสดงจำนวนสัญชาติของผู้ติดเชื้อโคโรนา 2019 แกน x คือจำนวนผู้ติดเชื้อ และแกน y คือ สัญชาติที่พบมากที่สุด 20 อันดับแรกในประเทศไทย โดยแสดงกราฟแบบ Log scale

ข้อมูลจำนวนของผู้ติดเชื้อโคโรนา 2019 แบ่งตามสัญชาติ ปรากฏว่าพบผู้ติดเชื้อสัญชาติไทยมากที่สุด อันดับสองเป็นผู้ติดเชื้อที่ไม่ระบุข้อมูลสัญชาติของตน และอันดับสามเป็นประเทศพม่า ซึ่งเป็นประเทศที่ติดกับประเทศไทย มีคนพม่าจำนวนมากเข้ามาทำงานในประเทศไทย โดยเฉพาะแรงงานต่างด้าวที่เข้าเมืองอย่างผิดกฎหมาย เริ่มจากกลุ่มคนไทยที่ทำงานในพม่า และลักลอบเข้าประเทศผ่านช่องทางธรรมชาติที่จังหวัดเชียงราย และกระจายไปยังจังหวัดต่าง ๆ ซึ่งในเดือนธันวาคมมีการแพร่ระบาดของไวรัสโคโรนา 2019 ระลอกใหม่ในประเทศไทย จากการตรวจพบว่าเจ้าของแผงในตลาดกลางกุ้งมหาชัย จังหวัดสมุทรสาคร และคนรอบข้างติดเชื้อไวรัสโคโรนา 2019 จนต้องปิดตลาด และหลังจากนำกลุ่มแรงงานต่างด้าวในตลาดกลางกุ้งมหาชัยมาตรวจคัดกรองพบว่ามีผู้ติดเชื้อที่เป็นแรงงานต่างด้าวกว่า 500 ราย และมีจำนวนเพิ่มขึ้นเรื่อย ๆ จนผู้ว่าราชการจังหวัดต้องมีคำสั่งล็อกดาวน์ จังหวัดสมุทรสาคร ก่อนจะพบว่ายอดการติดเชื้อไวรัสโคโรนา 2019 รอบใหม่ได้กระจายไปยังจังหวัดต่าง ๆ ทั่วประเทศในขณะนี้ กว่า 1,000 คน ส่วนสัญชาติอื่นที่พบส่วนใหญ่เป็นประเทศใกล้เคียงหรืออยู่ติดกับประเทศไทย ทำให้เดินทางเข้ามายังประเทศไทยได้ง่าย รวมถึงสัญชาติที่เป็นนักท่องเที่ยวเข้ามาในประเทศไทย ได้แก่ สัญชาติจีน เกาหลี ญี่ปุ่น อินเดีย รัสเซีย และสหรัฐอเมริกา อ้างอิงจากกองเศรษฐกิจการท่องเที่ยวและกีฬาแห่งประเทศไทย นอกจากนี้ยังพบสัญชาติที่ติดเชื้อเป็น 20 อันดับแรกของโลก ได้แก่ สหรัฐอเมริกา อังกฤษ รัสเซีย ปากีสถาน อิตาลี อินเดีย เยอรมนี ฝรั่งเศส และจีน

4.3.5 10 อันดับจังหวัดที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทย



รูป 4.18 จังหวัดที่รักษา (สีส้ม) และจังหวัดที่พบ (สีฟ้า) ของ 10 อันดับจังหวัดที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากกราฟแสดงจังหวัดที่พบผู้ติดเชื้อและจังหวัดที่รักษาตัวของผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย แกน x คือจำนวนผู้ติดเชื้อ แกน y คือจังหวัดที่พบผู้ติดเชื้อและจังหวัดที่รักษาตัว โดยแบ่งเป็นสีชมพูคือจังหวัดที่รักษาตัวของผู้ติดเชื้อ และสีฟ้าคือจังหวัดที่พบผู้ติดเชื้อ โดยแสดงตาม 10 อันดับจากจังหวัดที่พบผู้ติดเชื้อมากที่สุด

จังหวัดที่พบผู้ติดเชื้อและจังหวัดที่รักษาตัวของผู้ติดเชื้อไวรัสโคโรนา 2019 พบว่าจังหวัดที่พบผู้ติดเชื้อมากที่สุด คือ กรุงเทพมหานคร ซึ่งเป็นเมืองหลวงของประเทศไทย อีกทั้งยังเป็นจังหวัดที่พบแรงงานต่างด้าวมากที่สุดประเทศไทย (อ้างอิงข้อมูลจากกระทรวงแรงงานแห่งประเทศไทยปี 2019) และมีจำนวนประชากรสูงสุดในประเทศไทย (อ้างอิงจากสำนักบริหารการทะเบียน กรมการปกครองกระทรวงมหาดไทย) อันดับสองคือจังหวัดชลบุรี ซึ่งมีจำนวนแรงงานต่างด้าวมากที่สุดเป็นอันดับสี่ของประเทศไทย (อ้างอิงข้อมูลจากกระทรวงแรงงานแห่งประเทศไทยปี 2019) และมีจำนวนประชากรสูงเป็นอันดับเก้าของประเทศไทย นอกจากนี้ยังพบว่าจังหวัดชลบุรีมีนักท่องเที่ยวชาวต่างชาติมากเป็นอันดับสองรองจากกรุงเทพมหานคร (อ้างอิงจากกระทรวงการท่องเที่ยวแห่งประเทศไทยปี 2019) อันดับสามคือจังหวัดสมุทรสาครซึ่งมีจำนวนแรงงานต่างด้าวมากเป็นอันดับสองของประเทศรองจากกรุงเทพมหานคร ส่วนจังหวัดอื่นที่พบเป็นจังหวัดที่เป็น Cluster ผู้ติดเชื้อไวรัสโคโรนา 2019 และได้เดินทางไปยังจังหวัดต่าง ๆ ในประเทศไทย ทำให้เกิดการแพร่ระบาดไวรัสโคโรนา 2019 เกือบทั่วทั้งประเทศไทย เพื่อลดค่า Ro สำหรับผู้ที่เดินทางไปในพื้นที่เสี่ยงในช่วง 14 วันที่ผ่านมาควรปฏิบัติตัวโดยการกักตัวอยู่บ้านจนกว่าจะหาย หากมีอาการไม่สบายแม้เพียงเล็กน้อย เช่น ปวดหัว คัดจมูกไข้ไอ หรือหายใจติดขัดต้องไปพบแพทย์ทันที

4.4 ผลลัพธ์ที่ได้จาก SEIRD model of covid-19 in United States

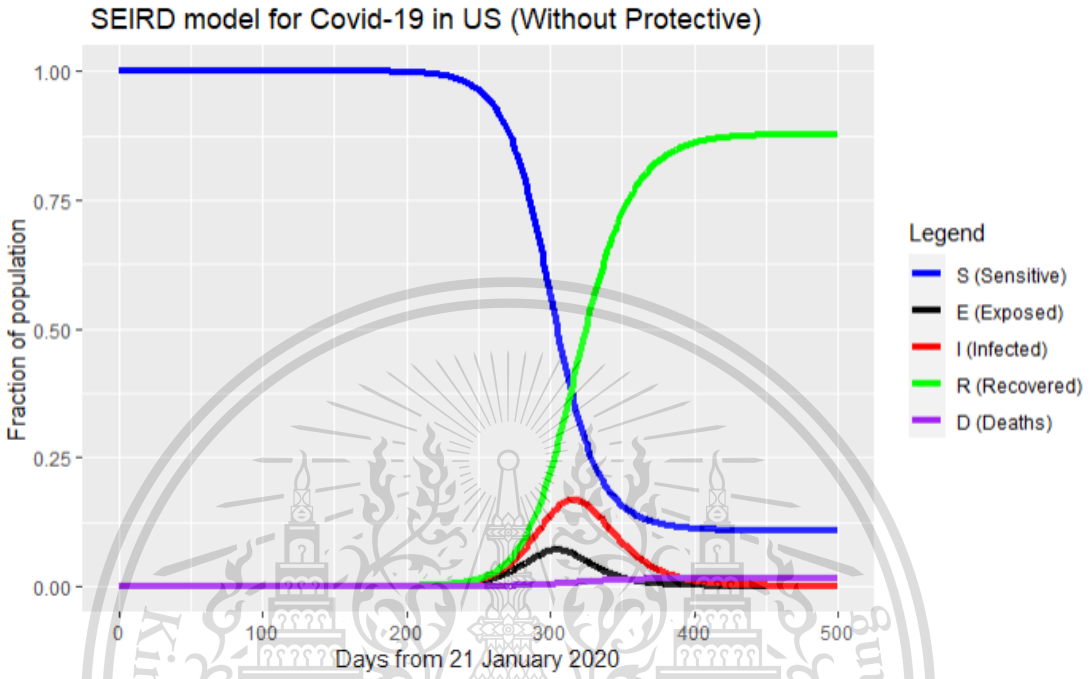
โมเดลนี้มีจุดเริ่มต้นจากวันที่มีเหตุการณ์ผู้ติดเชื้อรายแรกในสหรัฐอเมริกา (21 มกราคม 2020) โดยโมเดลนี้เป็นโมเดลอย่างง่าย และไม่มีปัจจัยการเกิดโรค แต่จะมีปัจจัยการเสียชีวิตของประชากร รวมถึงไม่มีการเข้าออกของประชากรระหว่างประเทศ ซึ่งจำนวนประชากรในระบบจะเท่าเดิมเสมอ นอกจากนั้นจะมีการกำหนดพารามิเตอร์ต่าง ๆ ดังนี้

1) จำนวนประชากรทั้งหมด (N) = 328.2 ล้านคน (จำนวนประชากรในสหรัฐอเมริกาทั้งหมด)
2) จำนวนผู้ติดเชื้อเริ่มต้น (I) = 1 คน
3) ระยะเวลาในการฟักตัวของเชื้อเฉลี่ย (D_E) = 5.2 วัน
4) ระยะเวลาการติดเชื้อเฉลี่ย (D_I) = 5 วัน
5) ระดับค่าการติดเชื้อ (Ro) = 2.5-3.3 (เลือกค่า Ro ที่สอดคล้องกับข้อมูลผู้ติดเชื้อในสหรัฐอเมริกามากที่สุด ซึ่งสอดคล้องกับหน่วยงานทั่วโลกที่มีค่า Ro อยู่ที่ระหว่าง 2 – 3 ซึ่งสหรัฐอเมริกา อ้างอิงจากองค์กรอนามัยโลก WHO)
6) อัตราการเสียชีวิตจากอาการติดเชื้อ (crude fatality Rate : CFR) = 0.01

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.4.1 ประชากรแต่ละกลุ่มของ SEIR model ในสหรัฐอเมริกาที่ยังไม่มีมาตรการป้องกัน

กำหนดให้ช่วงเวลาเริ่มต้น ทุกคนเป็นประชากรกลุ่มเสี่ยงที่สามารถติดเชื้อได้ทั้งหมด โดยมีจำนวนผู้ที่สามารถแพร่เชื้อเริ่มต้นเพียงคนเดียว



รูป 4.19 สัดส่วนประชากรแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ยังไม่มีมาตรการป้องกัน covid-19 กรณีนี้กำหนดให้ $\beta = 0.18$, $\gamma = 0.07$, $\alpha = 0.19$ และ $\sigma = 0.02$

กราฟเส้นแสดงจำนวนประชากรแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกา ที่ยังไม่มีมาตรการป้องกัน Covid-19 ได้แก่ สีน้ำเงินคือกลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) สีดำคือกลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สีแดงคือกลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) สีเขียวคือกลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) และสีม่วงคือกลุ่มที่เสียชีวิตจากการติดเชื้อ (Deaths) เมื่อระยะเวลาผ่านไป 500 วันหลังจากมีผู้ติดเชื้อรายแรกในสหรัฐอเมิกาวันที่ 21 มกราคม 2020 แกน x คือจำนวนวัน และแกน y คือสัดส่วนจำนวนประชากรในสหรัฐอเมริกา

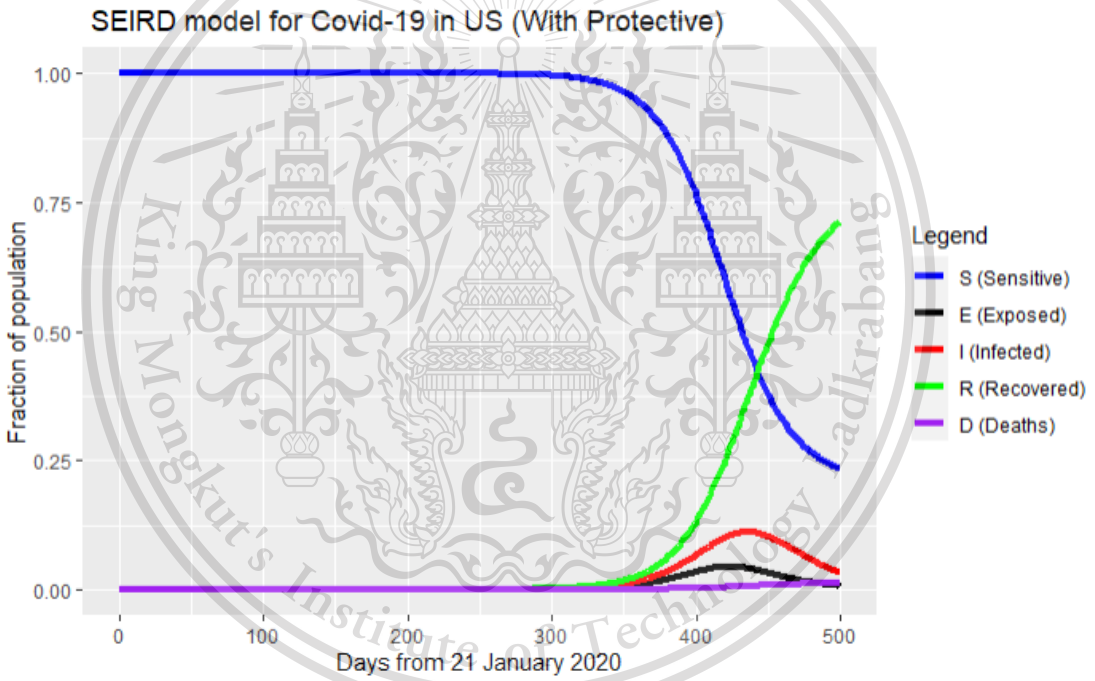
สังเกตได้ว่าหากในสถานการณ์ปกติที่ไร้มาตรการป้องกัน เช่น Social Distancing จะมีผู้ติดเชื้อที่พร้อมกันร้อยละ 15 ของจำนวนประชากรทั้งหมดในสหรัฐอเมริกาหรือประมาณ 49 ล้านคน ในช่วงที่มีจำนวนผู้ติดเชื้อสูงที่สุด คือช่วงเวลาประมาณ 320 วันหลังจากพบผู้ติดเชื้อรายแรกและจำนวนผู้เสียชีวิตจากการติดเชื้อจะค่อย ๆ เพิ่มสูงขึ้นเรื่อย ๆ เมื่อเวลาผ่านไป 400 วัน จำนวนผู้เสียชีวิตจะสูงกว่าจำนวนผู้ติดเชื้อ ซึ่งจะทำให้เกิดความสูญเสียเป็นวงกว้าง เนื่องจากระบบสาธารณสุขไม่สามารถรองรับผู้ติดเชื้อจำนวนมากในเวลาเดียวกันได้ด้วยทรัพยากรที่มีอยู่อย่างจำกัด อีกทั้งยารักษาหรือวัคซีนยังไม่สามารถเข้าถึงประชากรได้ทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.4.2 ผลลัพธ์ที่ได้จาก SEIRD model เมื่อ add social distancing factor

เมื่อเราเพิ่ม Social distancing factor ลงไปใน SEIRD model จะส่งผลให้ความรุนแรงในการแพร่ $\beta(I)$ ลดลง เนื่องจากประชากรมีการสัมผัสหรือเจอกันน้อยลง โดยแทนค่าตัวแปรของ Social distancing factor ด้วย ρ มีค่าอยู่ระหว่าง 0 – 1 หากมีค่าเป็น 0 หมายถึงประชากรทุกคนมีการเก็บตัวอยู่บ้าน และมีประสิทธิภาพในการทำ Social distancing สูงมาก และหากมีค่าเป็น 1 หมายถึงกรณีที่ไม่มีมาตรการป้องกันใด ๆ เลย โดยจะกำหนดให้ ρ มีค่า 0.8 เพื่อแสดงให้เห็นว่า หากมีมาตรการ Social distancing จะส่งผลต่อลักษณะการเติบโตของจำนวนผู้ติดเชื้ออย่างไร ซึ่งเมื่อมีตัวแปร ρ เข้ามา จะมีการแก้ไขในส่วนของความรุนแรงในการแพร่เชื้อ $\beta(I)$

$$\beta(I) = \rho \frac{R_0 I}{D_I N} \tag{4.2}$$



รูป 4.20 สัดส่วนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่มีมาตรการป้องกัน กรณีนี้กำหนดให้ $\beta = 0.18$, $\gamma = 0.07$, $\alpha = 0.18$, $\sigma = 0.02$ และ $\rho = 0.8$

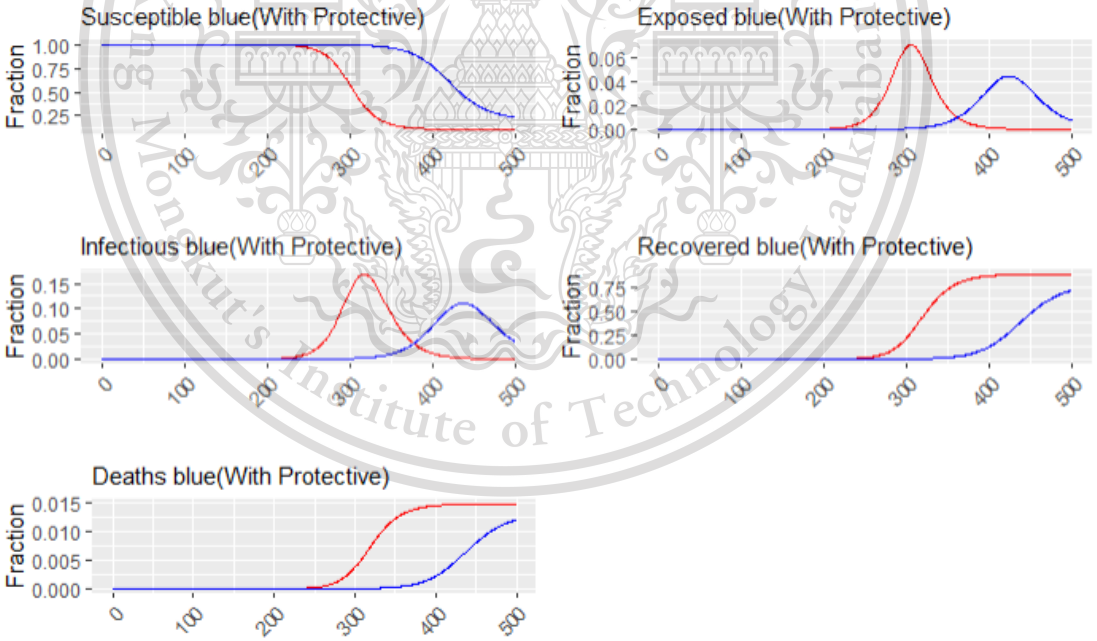
กราฟเส้นแสดงจำนวนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่เพิ่ม Social distancing factor โดย ρ มีค่า 0.8 ได้แก่ สีน้ำเงินคือกลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) สีดำคือกลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สีแดงคือกลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) สีเขียวคือกลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) และสีม่วงคือกลุ่มที่เสียชีวิตจากการติดเชื้อ (Deaths) เมื่อระยะเวลาผ่านไป 500 วันหลังจากมีผู้ติดเชื้อรายแรกใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สหรัฐอเมริกาเมื่อวันที่ 21 มกราคม 2020 แกน x คือวัน และแกน y คือสัดส่วนจำนวนประชากรในสหรัฐอเมริกา

สังเกตได้ว่าถ้าหากในสถานการณ์ปกติที่เพิ่มมาตรการป้องกัน Social distancing จะมีผู้ติดเชื้อในเวลาเดียวกันที่ร้อยละ 10 ของจำนวนประชากรทั้งหมดในสหรัฐอเมริกาหรือประมาณ 32 ล้านคน ในช่วงที่มีจำนวนผู้ติดเชื้อสูงที่สุด คือ ช่วงเวลาประมาณ 430 วันหลังจากพบผู้ติดเชื้อรายแรก และจำนวนผู้เสียชีวิตจากการติดเชื้อจะค่อย ๆ เพิ่มสูงขึ้นเรื่อย ๆ แต่ยังมีจำนวนผู้ติดเชื้อสูงกว่าจำนวนผู้เสียชีวิต หมายความว่าผู้ที่ติดเชื้อมีโอกาสเสียชีวิตจากโรคระบาดน้อยกว่าเมื่อมีมาตรการป้องกัน ซึ่งจะเห็นได้ว่าการมีมาตรการป้องกัน ทำให้ลดการแพร่ระบาดของไวรัสโคโรนา 2019 ได้มากเมื่อเทียบกับการไม่มีมาตรการป้องกันเลย ส่งผลให้เกิดความสูญเสียน้อยลง เนื่องจากระบบสาธารณสุขนั้นสามารถรองรับผู้ติดเชื้อได้เพียงพอ จากการที่จำนวนผู้ติดเชื้อในเวลาเดียวกันน้อยลง และประวิงเวลาจนกว่าจะคิดค้นยารักษาหรือประชากรได้รับวัคซีนอย่างทั่วถึง

4.4.3 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน



รูป 4.21 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน กรณีนี้กำหนดให้มีมาตรการป้องกัน $\rho = 0.8$ และไม่มีมาตรการป้องกัน $\rho = 1$

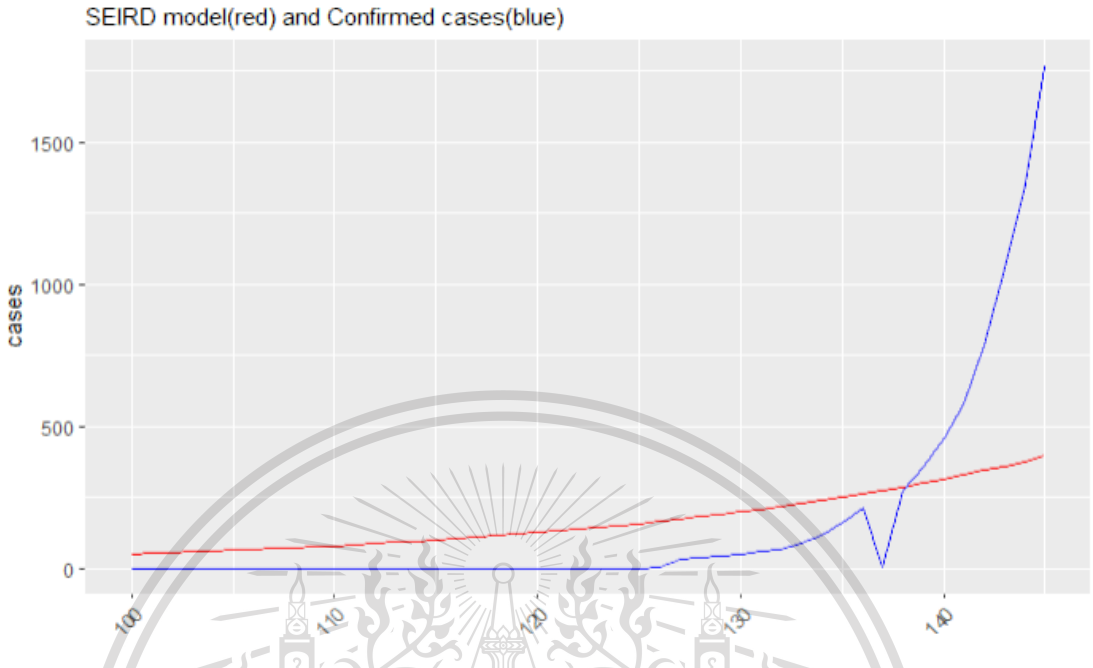
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กราฟเส้นแสดงแปรียบเทียบจำนวนประชากรในแต่ละกลุ่มของ SEIRD model ในสหรัฐอเมริกาที่ไม่มาตรการป้องกันและหลังเพิ่ม Social distancing factor โดย ρ มีค่า 0.8 โดยเส้นสีแดงคือไม่มีมาตรการป้องกัน และเส้นสีน้ำเงินคือมีมาตราการป้องกัน เมื่อระยะเวลาผ่านไป 500 วันหลังจากมีผู้ติดเชื้อรายแรกในสหรัฐอเมริกาเมื่อวันที่ 21 มกราคม 2020 แกน x คือวัน ส่วนแกน y คือสัดส่วนจำนวนประชากรในสหรัฐอเมริกา

สังเกตได้ว่าเมื่อมีการลดความรุนแรงจากการแพร่ I ลง จะสามารถทำให้จำนวนผู้ติดเชื้อเพิ่มด้วยอัตราที่ลดลงอย่างเห็นได้ชัด จากผู้ติดเชื้อ 49 ล้านคนภายในเวลา 320 วัน เมื่อเพิ่มมาตรการป้องกัน Social distancing หรือควบคุมให้ประชากรทุกคนมีการเก็บตัวอยู่บ้านเพียงแค่ร้อยละ 20 จะสามารถทำให้จำนวนผู้ติดเชื้อสูงสุดลดลงเหลือ 32 ล้านคนภายใน 430 วัน ในส่วนของจำนวนผู้เสียชีวิตจากการติดเชื้อหากไม่มีมาตรการป้องกัน พบว่าจำนวนผู้เสียชีวิตจากการติดเชื้อจะสูงขึ้นอย่างต่อเนื่อง แต่ถ้าหากมีมาตรการป้องกัน จำนวนผู้เสียชีวิตจะลดลงอย่างเห็นได้ชัดและมีโอกาสเสียชีวิตจากโรคระบาดน้อยกว่าเมื่อมีมาตรการป้องกัน ยิ่งเพิ่มมาตรการป้องกันให้มีประสิทธิภาพ จำนวนผู้ติดเชื้อ และจำนวนผู้เสียชีวิตจะลดลง โดยถ้าทุกคนร่วมมือกัน กักตัวอยู่บ้าน ก็จะสามารถผ่านวิกฤตโรคระบาดนี้ได้โดยที่สูญเสียน้อยที่สุด นอกจากนี้การแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกานั้นเป็นไวรัสที่ยังไม่สามารถรับมือได้อย่างทั่วถึง แม้ว่าจะมีการแจกจ่ายวัคซีนให้ประชาชนแล้ว เนื่องจากสหรัฐอเมริกามีพื้นที่ขนาดใหญ่ มีหลายรัฐ มีประชากรจำนวนมาก ทำให้มีตัวแปรหลายอย่างที่จะนำมาใช้ร่วมกันทุกรัฐในการจำลอง และการทำนายการแพร่ระบาดในสหรัฐอเมริกาทุกรัฐพร้อมกันให้แม่นยำนั้นเป็นไปได้ยาก จึงบอกได้เพียงแค่ว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.5.4 วัดประสิทธิภาพ SEIRD model โดยใช้ RMSE



รูป 4.22 วัดประสิทธิภาพ SEIRD model โดยใช้ RMSE

กราฟเส้นแสดงวัดประสิทธภาพ SEIRD model ด้วยวิธี RMSE แกน x คือวันที่และแกน y คือจำนวนผู้ที่ติดเชื้อ โดยเส้นสีน้ำเงินคือจำนวนผู้ติดเชื้อจริงในประเทศไทยและเส้นสีแดงคือ จำนวนผู้ติดเชื้อที่ได้จากการทำนาย SEIR model ในระยะเวลา 40 วัน

จากการใช้แบบจำลองทางคณิตศาสตร์ SEIRD model ในสหรัฐอเมริกาพบว่าแบบมีมาตรการป้องกันจะมีจำนวนผู้ติดเชื้อสูงสุดต่ำกว่าแบบไม่มีมาตรการป้องกันถึง 0.65 เท่าเมื่อเพิ่มมาตรการ Social distancing และเวลาที่พบผู้ติดเชื้อสูงสุดจะช้ากว่าแบบไม่มีมาตรการป้องกันประมาณ 100 วัน ส่วนจำนวนผู้เสียชีวิตจากการติดเชื้อแบบมีมาตรการป้องกันจะค่อย ๆ เพิ่มสูงขึ้นเรื่อย ๆ แต่ยังมีจำนวนผู้ติดเชื้อสูงกว่าจำนวนผู้เสียชีวิต นั่นหมายความว่าผู้ที่ติดเชื้อมีโอกาสเสียชีวิตจากโรคระบาดน้อยกว่าเมื่อมีมาตรการป้องกัน

เมื่อวัดประสิทธิภาพโมเดลด้วย RMSE เทียบกับผู้ติดเชื้อในสหรัฐอเมริกาจริงใน 40 วันแรก พบว่าค่า RMSE มีค่าเท่ากับ 299.79 นั่นหมายความว่าโดยเฉลี่ยโมเดลจะทำนายค่าผู้ติดเชื้อผิดไป +/- 299.79 คน จากการทำนายใน 40 วันแรกของการติดเชื้อ เมื่อเพิ่มระยะเวลาในการทำนายค่า RMSE จะเพิ่มขึ้นเรื่อย ๆ เนื่องจากยิ่งเวลามากขึ้นปัจจัยที่จะเข้ามามีผลต่อการติดเชื้อไวรัสโคโรนา 2019 ยิ่งมากขึ้น ไม่ว่าจะเป็นปัจจัยที่ส่งผลดี เช่น การได้รับวัคซีน หรือค้นพบยารักษาในอนาคต และปัจจัยที่ส่งผลเสีย เช่น การพบกลุ่มแพร่กระจายไวรัสกลุ่มใหม่ หรือการค้นพบไวรัสสายพันธุ์ใหม่ที่อาจมีผลร้ายแรงกว่าเดิม การจำลอง และทำนายสถานการณ์การแพร่ระบาดโดยใช้ SEIRD

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

model ให้แม่นยำนั้นเป็นไปได้ยาก จึงบอกได้เพียงว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

4.5 ผลลัพธ์ที่ได้จาก SEIR model of covid-19 in Thailand

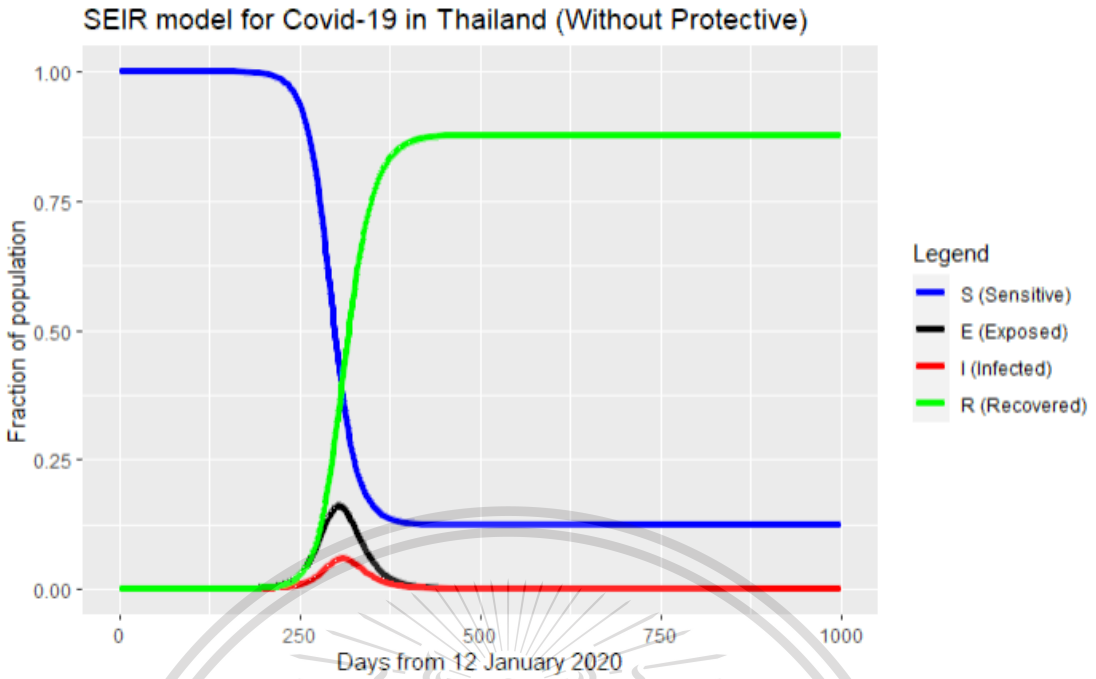
โมเดลนี้ใช้จุดเริ่มต้นจากวันที่มีเหตุการณ์ผู้ติดเชื้อรายแรกในประเทศไทย (12 มกราคม 2020) โดยโมเดลนี้เป็นโมเดลอย่างง่าย และไม่มีปัจจัยการเกิดโรค หรือการตายของประชากร รวมถึงไม่มีการเข้าออกของประชากรระหว่างประเทศ ซึ่งจำนวนประชากรในระบบจะเท่าเดิมเสมอ นอกจากนั้นจะมีการกำหนดพารามิเตอร์ต่าง ๆ ดังนี้

1) จำนวนประชากรทั้งหมด (N) = 69.63 ล้านคน (จำนวนประชากรในประเทศไทยทั้งหมด)
2) จำนวนผู้ติดเชื้อเริ่มต้น (I) = 1 คน
3) ระยะเวลาในการฟักตัวของเชื้อเฉลี่ย (D_E) = 5.2 วัน
4) ระยะเวลาการติดเชื้อเฉลี่ย (D_I) = 2.3 วัน
5) ระดับค่าการติดเชื้อ (Ro) = 1.5-2.3 (โดยเลือกค่า Ro ที่สอดคล้องกับข้อมูลผู้ติดเชื้อไทยมากที่สุด โดยสอดคล้องกับหน่วยงานทั่วโลกที่มีค่า Ro อยู่ที่ระหว่าง 2 – 3 ซึ่งประเทศไทยอ้างอิงจาก ศบค.)

4.5.1 ประชากรแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ยังไม่มีมาตรการป้องกัน

กำหนดให้ช่วงเวลาเริ่มต้น ทุกคนเป็นประชากรกลุ่มเสี่ยงที่สามารถติดเชื้อได้ทั้งหมด โดยมีจำนวนผู้ที่สามารถแพร่เชื้อเริ่มต้นเพียงคนเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.23 สัดส่วนประชากรแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ยังไม่มีมาตรการป้องกัน covid-19 กรณีนี้กำหนดให้ $\beta = 0.46$, $\gamma = 0.07$ และ $\alpha = 0.19$

กราฟเส้นแสดงจำนวนประชากรแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ยังไม่มีมาตรการป้องกัน covid-19 ได้แก่ สีน้ำเงินคือกลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) สีดำคือกลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สีแดงคือกลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) และสีเขียวคือกลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) เมื่อระยะเวลาผ่านไป 1,000 วันหลังจากมีผู้ติดเชื้อรายแรกในประเทศไทยเมื่อวันที่ 12 มกราคม 2020 แกน x คือจำนวนวัน และแกน y คือสัดส่วนจำนวนประชากรในประเทศไทย

สังเกตได้ว่าถ้าหากในสถานการณ์ปกติที่ไร้มาตรการป้องกันเช่น Social distancing จะมีผู้ติดเชื้อที่ในเวลาเดียวกันที่ร้อยละ 6 ของจำนวนประชากรทั้งหมดในประเทศไทยหรือประมาณ 4 ล้านคน ช่วงที่มีจำนวนผู้ติดเชื้อสูงที่สุดคือช่วงเวลาประมาณ 320 วันหลังจากพบผู้ติดเชื้อรายแรก ซึ่งจะทำให้เกิดความสูญเสียเป็นวงกว้าง เนื่องจากระบบสาธารณสุขไม่สามารถรองรับผู้ติดเชื้อจำนวนมากพร้อมกันได้ด้วยทรัพยากรที่มีอยู่อย่างจำกัด อีกทั้งยารักษาหรือวัคซีนยังไม่สามารถเข้าถึงประชากรได้ทั้งหมด

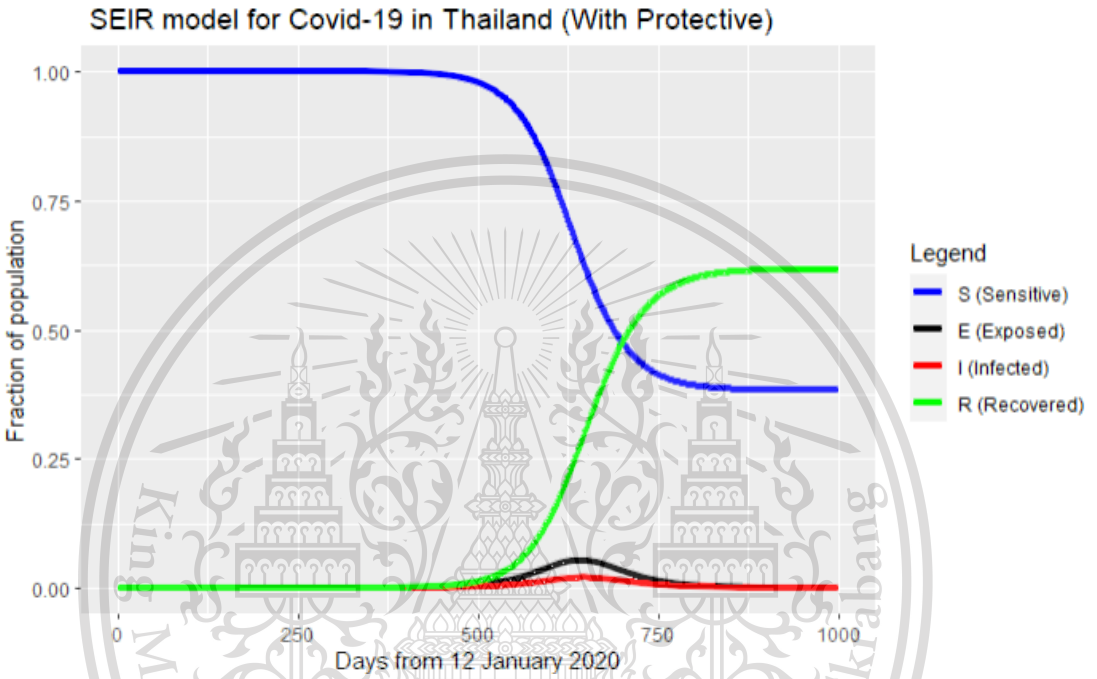
4.5.2 ผลลัพธ์ที่ได้จาก SEIR model เมื่อ add social distancing factor

เมื่อเราเพิ่ม Social distancing factor ลงไปใน SEIR model จะส่งผลให้ความรุนแรงในการแพร่ $\beta(I)$ ลดลง เนื่องจากประชากรมีการสัมผัสหรือเจอกันน้อยลง โดยแทนตัวแปรของ Social distancing factor ด้วย ρ โดยจะมีค่าอยู่ระหว่าง 0 – 1 โดยถ้ามีค่าเป็น 0 หมายถึงประชากรทุกคนมีการเก็บตัวอยู่บ้าน และมีประสิทธิภาพในการทำ Social distancing สูงมาก และ 1 หมายถึงกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ไม่มีมาตรการป้องกันใด ๆ เลย โดยจะกำหนดให้ ρ มีค่า 0.65 เพื่อแสดงให้เห็นว่า หากมีมาตรการ Social distancing ลักษณะการเติบโตของจำนวนผู้ติดเชื้อจะเป็นอย่างไร ซึ่งเมื่อมีตัวแปร ρ เข้ามาจะมีการแก้ไขในส่วนของความรุนแรงในการแพร่เชื้อ $\beta(I)$

$$\beta(I) = \rho\frac{R_0 I}{D_I N} \tag{4.1}$$



รูป 4.24 สัดส่วนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่มาตรการป้องกัน กรณีนี้กำหนดให้ $\beta = 0.46, \gamma = 0.07, \alpha = 0.19$ และ $\rho = 0.65$

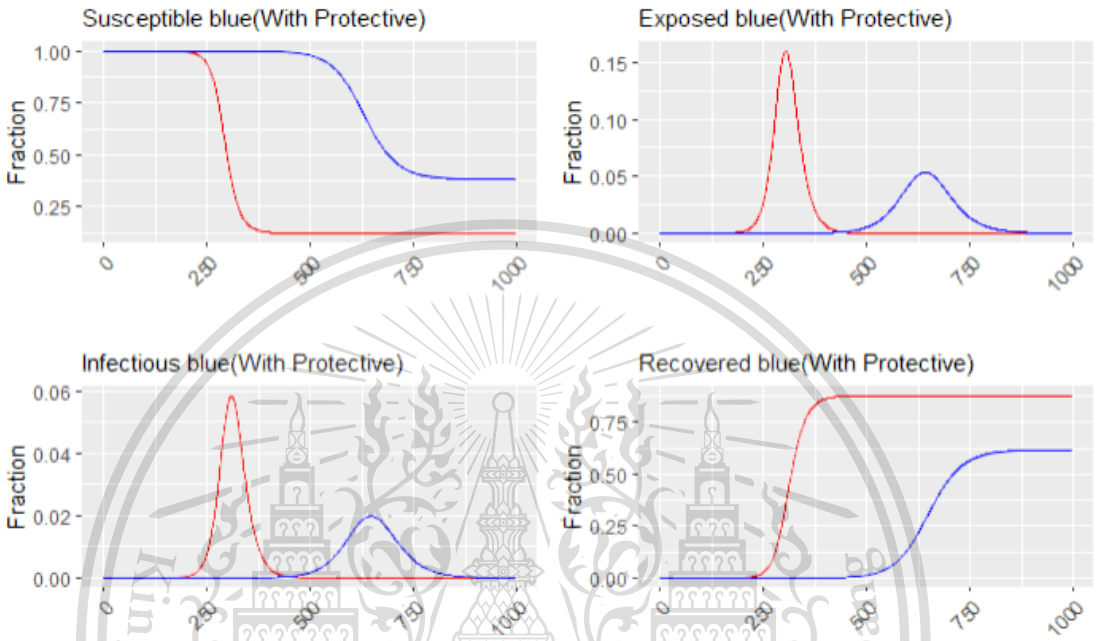
กราฟเส้นแสดงจำนวนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่เพิ่ม Social distancing factor โดย ρ มีค่า 0.65 ได้แก่ สีน้ำเงินคือกลุ่มเสี่ยงที่มีโอกาสติดเชื้อได้ (Susceptible) สีดำคือกลุ่มที่ติดเชื้อที่อยู่ในระยะฟักตัว (Exposed) สีแดงคือกลุ่มที่ติดเชื้อที่สามารถแพร่เชื้อได้ (Infectious) และสีเขียวคือกลุ่มที่หายจากการติดเชื้อแล้ว (Recovered) เมื่อระยะเวลาผ่านไป 1000 วันหลังจากมีผู้ติดเชื้อรายแรกในประเทศไทยเมื่อวันที่ 12 มกราคม 2020 แกน x คือจำนวนวัน และแกน y คือสัดส่วนจำนวนประชากรในประเทศไทย

สังเกตได้ว่าถ้าหากในสถานการณ์ปกติที่เพิ่มมาตรการป้องกัน Social distancing จะมีผู้ติดเชื้อในช่วงเวลาเดียวกันที่ร้อยละ 1.9 ของจำนวนประชากรทั้งหมดในประเทศไทย ในช่วงที่มีจำนวนผู้ติดเชื้อสูงที่สุด คือ ช่วงเวลาประมาณ 630 วันหลังจากพบผู้ติดเชื้อรายแรก ซึ่งจะเห็นได้ว่าการมีมาตรการป้องกันจะสามารถลดการแพร่ระบาดของไวรัสโคโรนา 2019 ได้มากเมื่อเทียบกับการไม่มีมาตรการป้องกันเลย ทำให้เกิดความสูญเสียน้อยลง เนื่องจากระบบสาธารณสุขนั้นจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สามารถรองรับผู้ติดเชื้อพร้อมกันได้ และประวิงเวลาจนกว่าจะคิดค้นยารักษาหรือประชากรได้รับวัคซีนอย่างทั่วถึงทั้งประเทศ

4.5.3 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน



รูป 4.25 เปรียบเทียบสัดส่วนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ไม่มีมาตรการป้องกันและมีมาตรการป้องกัน กรณีนี้กำหนดให้มีมาตรการป้องกัน $\rho = 0.65$ (เส้นสีน้ำเงิน) และไม่มีมาตรการป้องกัน $\rho = 1$ (เส้นสีแดง)

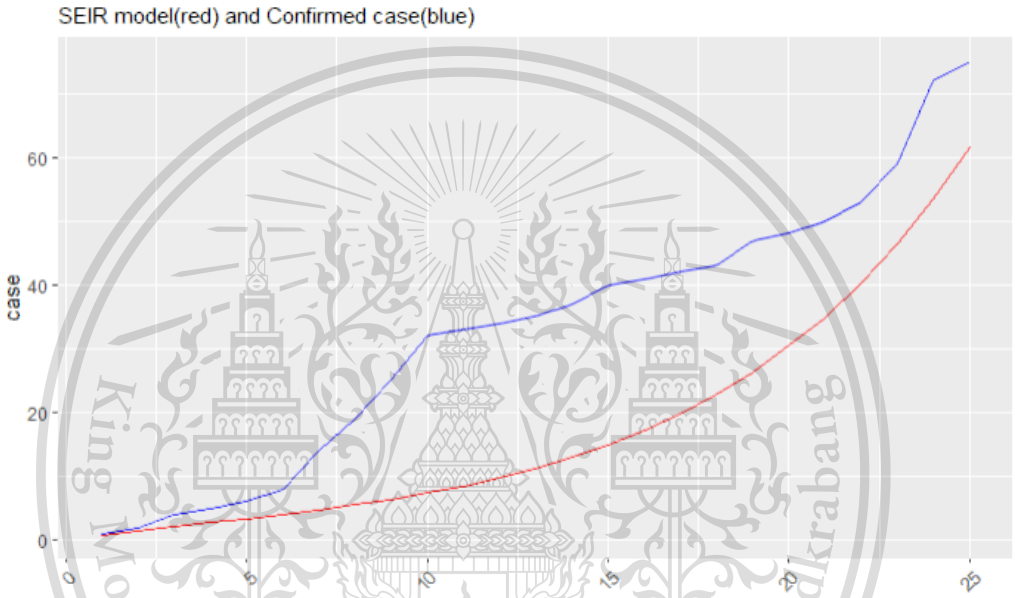
กราฟเส้นแสดงเปรียบเทียบจำนวนประชากรในแต่ละกลุ่มของ SEIR model ในประเทศไทยที่ไม่มาตรการป้องกันและที่เพิ่ม Social distancing factor โดย ρ มีค่า 0.65 โดยเส้นสีแดงคือไม่มีมาตรการป้องกัน และเส้นสีน้ำเงินคือมีมาตราการป้องกัน เมื่อระยะเวลาผ่านไป 1,000 วันหลังจากมีผู้ติดเชื้อรายแรกในประเทศไทยเมื่อวันที่ 12 มกราคม 2020 แกน x คือวัน ส่วนแกน y คือ สัดส่วนจำนวนประชากรในประเทศไทย

สังเกตได้ว่าเมื่อมีการลดความรุนแรงจากการแพร่ I ลง จะสามารถทำให้กราฟจำนวนผู้ติดเชื้อนั้นเพิ่มขึ้นด้วยอัตราที่น้อยลงอย่างเห็นได้ชัด จากผู้ติดเชื้อ 4 ล้านคนภายในเวลา 320 วัน เมื่อเพิ่มมาตรการป้องกัน Social distancing โดยควบคุมให้ประชากรทุกคนมีการกักเก็บตัวอยู่บ้านเพียงร้อยละ 35 จะสามารถทำให้จำนวนผู้ติดเชื้อสูงสุดลดลงเหลือ 1 ล้านคนภายใน 630 วัน ยิ่งเพิ่มมาตรการป้องกันให้มีประสิทธิภาพ จำนวนผู้ติดเชื้อก็จะยิ่งลดลง โดยถ้าทุกคนร่วมมือกัน กักตัวอยู่บ้าน ก็จะสามารถผ่านวิกฤตโรคระบาดนี้ได้โดยที่สูญเสียน้อยที่สุด นอกจากนี้การแพร่ระบาดไวรัส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โคโรนา 2019 ในประเทศไทยนั้นเป็นไวรัสที่ยังไม่สามารถรับมือได้อย่างทั่วถึงทั้งประเทศ การตรวจหาเชื้อไวรัสส่วนใหญ่จะตรวจเฉพาะกลุ่มเสี่ยง หรือหากมีความต้องการตรวจด้วยตัวเองก็สามารถทำได้ แต่จะมีค่าใช้จ่ายสูง ทำให้มีประชากรจำนวนมากไม่ได้รับการตรวจ อีกทั้งยังขาดตัวแปรหลายอย่างในการจำลอง และการทำนายสถานการณ์การแพร่ระบาดให้แม่นยำนั้นเป็นไปได้ยาก จึงระบุได้เพียงว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

4.5.4 วัดประสิทธิภาพ SEIR model โดยใช้ RMSE



รูป 4.26 วัดประสิทธิภาพ SEIR model โดยใช้ RMSE

กราฟเส้นแสดงวัดประสิทธภาพ SEIR model ด้วยวิธี RMSE แกน x คือวันที่และแกน y คือจำนวนผู้ที่ติดเชื้อ โดยเส้นสีน้ำเงินคือจำนวนผู้ติดเชื้อจริงในประเทศไทยและเส้นสีแดงคือจำนวนผู้ติดเชื้อที่ได้จากการทำนาย SEIR model ในระยะเวลา 25 วันเริ่มจากวันเจอผู้ติดเชื้อรายแรก

จากการใช้แบบจำลองทางคณิตศาสตร์ SEIR model ในประเทศไทยพบว่าแบบมีมาตรการป้องกันจะมีจำนวนผู้ติดเชื้อสูงสุดต่ำกว่าแบบไม่มีมาตรการป้องกันถึง 4 เท่าเมื่อเพิ่มมาตรการ Social distancing และเวลาที่พบผู้ติดเชื้อสูงสุดจะช้ากว่าแบบไม่มีมาตรการป้องกันประมาณ 310 วัน

เมื่อวัดประสิทธิภาพโมเดลด้วย RMSE เทียบกับผู้ติดเชื้อในประเทศไทยจริงใน 25 วันแรก พบว่าค่า RMSE มีค่าเท่ากับ 17.29 นั่นหมายความว่าโดยเฉลี่ยโมเดลจะทำนายค่าผู้ติดเชื้อผิดไป +/- 17.29 คน จากการทำนายใน 25 วันแรกของการติดเชื้อ เมื่อเพิ่มระยะเวลาในการทำนาย ค่า RMSE จะเพิ่มขึ้นเรื่อย ๆ เนื่องจากยิ่งเวลามากขึ้นปัจจัยที่จะเข้ามามีผลต่อการติดเชื้อไวรัสโคโรนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2019 ยิ่งมากขึ้น ไม่ว่าจะเป็นปัจจัยที่ส่งผลดี เช่น การได้รับวัคซีน หรือค้นพบยารักษาในอนาคต และปัจจัยที่ส่งผลเสีย เช่น การพบกลุ่มแพร่กระจายไวรัสกลุ่มใหม่ หรือการค้นพบไวรัสสายพันธุ์ใหม่ที่อาจมีผลร้ายแรงกว่าเดิม การจำลองและการทำนายสถานการณ์การแพร่ระบาดโดยใช้ SEIR model ให้แม่นยำนั้นเป็นไปได้ยาก จึงบอกได้เพียงว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

4.6 การวัดประสิทธิภาพโดยใช้ Profiling

4.6.1 การวัดประสิทธิภาพเพื่อเปรียบเทียบระหว่างการใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr

การวัดประสิทธิภาพโดยใช้ Profiling เป็นการวัดประสิทธิภาพในด้านหน่วยความจำและเวลาที่ใช้ในการประมวลผล โดยในที่นี้ใช้วิธีการ Profiling เพื่อเปรียบเทียบการใช้หน่วยความจำและการใช้เวลาในการประมวลผลระหว่างแบบใช้แพ็กเกจ Dplyr และแบบไม่ใช้แพ็กเกจ Dplyr โดยมีการทดลอง ดังต่อไปนี้

1) ทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด
2) ทดสอบการเชื่อมต่อฐานข้อมูลระหว่างฟังก์ชัน odbcConnect และฟังก์ชัน src_mysql
3) ทดสอบฟังก์ชัน cleandata()
4) ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล
5) ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว
6) ทดสอบการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

4.6.1.1 ทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

นำโค้ดทั้งหมดตั้งแต่การเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด โดยแบ่งโค้ดเป็น 2 ส่วน คือ โค้ดแบบใช้แพ็กเกจ Dplyr และโค้ดแบบไม่ใช้แพ็กเกจ Dplyr เพื่อนำมาเปรียบเทียบการใช้หน่วยความจำ และเวลาที่ใช้ในการประมวลผล จากนั้นจึงนำมาวัดประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 4.1 ผลการทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr

รายการที่	รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
		ไม่ใช้ Dplyr	ใช้ Dplyr	ไม่ใช้ Dplyr	ใช้ Dplyr
1	ทดสอบการเชื่อมฐานข้อมูลระหว่างฟังก์ชัน odbcConnect และฟังก์ชัน src_mysql	0.7	1.7	90	30
2	ทดสอบฟังก์ชัน cleandata()	71	46	970	700
3	ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล	90.6	61.1	1,200	920
4	ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว	10.9	47.3	50	70
5	ทดสอบการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุด	2.3	1.4	40	50
รวม		175.5	157.5	2,350	1,770

จากตาราง 4.1 เมื่อเปรียบเทียบหน่วยความจำระหว่างใช้ Dplyr พบว่าแบบใช้แพ็กเกจ Dplyr ใช้หน่วยความจำน้อยกว่าแบบไม่ใช้แพ็กเกจ โดยคิดเป็นอัตราส่วนใช้แพ็กเกจกับไม่ใช้แพ็กเกจได้เป็น 1 : 1.11 หมายความว่าการใช้แพ็กเกจนั้นใช้หน่วยความจำน้อยกว่า 1.11 เท่า เมื่อเปรียบเทียบเวลาที่ใช้ พบว่าแบบใช้แพ็กเกจ Dplyr ใช้เวลาในการประมวลผลน้อยกว่าแบบไม่ใช้แพ็กเกจ โดยคิดเป็นอัตราส่วนใช้แพ็กเกจกับไม่ใช้แพ็กเกจได้เป็น 1 : 1.33 หมายความว่าการใช้แพ็กเกจนั้นใช้เวลาน้อยกว่า 1.3 เท่า

หากพิจารณาเพียงรายการทดสอบที่ 1-2 ซึ่งเป็นการเชื่อมต่อฐานข้อมูลและการทำ Data cleaning พบว่าการใช้แพ็กเกจ Dplyr มีการใช้หน่วยความจำน้อยกว่า 1.5 เท่า และใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เวลาน้อยกว่า 1.45 เท่าซึ่งแสดงให้เห็นได้ชัดว่าแพ็กเกจ Dplyr นั้นใช้หน่วยความจำและเวลาน้อยกว่าแบบไม่ใช้แพ็กเกจ

4.6.2 การวัดประสิทธิภาพของแพ็กเกจ Dplyr เมื่อมีปัจจัยอื่นเข้ามาเกี่ยวข้อง

จากการวัดประสิทธิภาพเพื่อเปรียบเทียบระหว่างการใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr แล้ว ผลจากการทดลองแสดงให้เห็นว่าโดยภาพรวมแล้วการใช้แพ็กเกจ Dplyr มีการใช้หน่วยความจำและเวลาในการประมวลผลน้อยกว่า (อ้างอิงจากตาราง 4.1) และเหมาะสมในการนำมาใช้จัดการข้อมูลต่อไป หากแต่อาจมีปัจจัยอื่น ๆ ที่ส่งผลต่อประสิทธิภาพในการประมวลผลของแพ็กเกจ Dplyr

ด้วยเหตุนี้ เราจึงทำการวัดประสิทธิภาพของแพ็กเกจ Dplyr เพิ่มเติม เพื่อศึกษาปัจจัยด้านต่าง ๆ ที่อาจส่งผลต่อการทำงานของแพ็กเกจ Dplyr อันได้แก่ ปัจจัยด้าน Environment ปัจจัยด้านการเชื่อมต่อฐานข้อมูล และปัจจัยด้านขนาดของข้อมูลที่นำไปประมวลผล เพื่อเปรียบเทียบความแตกต่างในการใช้หน่วยความจำและเวลาที่ใช้ของแพ็กเกจ Dplyr เมื่อมีปัจจัยดังกล่าวเข้ามาเกี่ยวข้อง โดยมีรายการการทดสอบ ดังนี้

1) ทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด
2) ทดสอบการเชื่อมต่อฐานข้อมูล
3) ทดสอบฟังก์ชัน cleandata()
4) ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล
5) ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว
6) ทดสอบการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

ในส่วนของปัจจัยที่นำมาร่วมในการวัดประสิทธิภาพมีรายละเอียด ดังนี้

1) ปัจจัยด้าน Environment

เปรียบเทียบการประมวลผลของแพ็กเกจ Dplyr ใน Environment ที่ต่างกัน 2 แบบ โดยมีรายละเอียดของแต่ละ Environment ดังตาราง

ตาราง 4.2 รายละเอียด Environment ที่ใช้ในการเปรียบเทียบเพื่อวัดประสิทธิภาพ

รายละเอียด Environment	Environment 1	Environment 2
OS	Ubuntu 20.04.2 LTS	Windows 10
Processor	AMD Ryzen ThreadRipper CPU 3.5 GHz 12-core/24-Threads	Intel Core i7-8750H CPU 2.20 GHz 6-core/12-Threads
RAM	64.00 GB	8.00 GB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2) ปัจจัยด้านขนาดของข้อมูล

เปรียบเทียบการประมวลผลของแพ็กเกจ Dplyr ระหว่างชุดข้อมูลขนาดเล็กและชุดข้อมูลขนาดใหญ่ โดยมีรายละเอียดดังต่อไปนี้

I) **ชุดข้อมูลขนาดเล็ก**

ชุดข้อมูลขนาดเล็กประกอบด้วยข้อมูลผู้ติดเชื้อ ผู้รักษาหาย และผู้เสียชีวิต โดยมีระยะเวลาตั้งแต่วันที่ 1 เมษายน 2020 ถึง 31 สิงหาคม 2020 (รวมระยะเวลา 5 เดือน) แต่ละไฟล์ข้อมูลมีจำนวน 157 คอลัมน์ 267 แถว รวมขนาดไฟล์อยู่ที่ 0.66 MiB และหลังทำการ Clean data ไฟล์มีขนาด 28,917 แถว 6 คอลัมน์

J) **ชุดข้อมูลขนาดใหญ่**

ชุดข้อมูลขนาดใหญ่ประกอบด้วยข้อมูลผู้ติดเชื้อ ผู้รักษาหาย และผู้เสียชีวิต โดยมีระยะเวลาตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021 (รวมระยะเวลา 1 ปี) แต่ละไฟล์ข้อมูลมีจำนวน 371 คอลัมน์ 267 แถว รวมขนาดไฟล์อยู่ที่ 1.48 MiB และหลังทำการ Clean data ไฟล์มีขนาด 70,831 แถว 6 คอลัมน์

3) ปัจจัยด้านการเชื่อมต่อฐานข้อมูล

การเชื่อมต่อฐานข้อมูล Mysql โดยใช้ฟังก์ชัน src_mysql จากแพ็กเกจ Dplyr โดยเปรียบเทียบการเชื่อมต่อทั้งหมด 2 รูปแบบ คือ การเชื่อมต่อฐานข้อมูลแบบ Local database และการเชื่อมต่อฐานข้อมูลบนระบบคลาวด์

4.6.2.1 ทดสอบการเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

นำโค้ดการใช้แพ็กเกจ Dplyr มาวัดประสิทธิภาพผ่าน Environment ที่แตกต่างกัน โดยแบ่งออกเป็น 5 การทดสอบ ดังนี้

การทดสอบที่ 1 ทดสอบการเชื่อมต่อฐานข้อมูล

การทดสอบที่ 2 ทดสอบฟังก์ชัน cleandata()

การทดสอบที่ 3 ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล

การทดสอบที่ 4 ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว

การทดสอบที่ 5 ทดสอบการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 4.3 ผลการทดสอบด้านเวลาที่ใช้ในการประมวลผลตั้งแต่การเชื่อมต่อฐานข้อมูลไปจนถึงการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

การทดสอบที่	เวลาที่ใช้ในการประมวลผล (ms)											
	Environment 1						Environment 2					
	Local database			Database บนคลาวด์			Local database			Database บนคลาวด์		
	เล็ก	ใหญ่	รวม	เล็ก	ใหญ่	รวม	เล็ก	ใหญ่	รวม	เล็ก	ใหญ่	รวม
1	48	58	106	74	84	158	38	30	68	786	762	1,548
2	446	1,014	1,460	480	1,020	1,500	384	966	1,350	348	932	1,280
3	378	1,080	1,458	382	1,116	1,498	650	1,174	1,824	716	1,800	2,516
4	38	132	170	50	106	156	44	110	154	48	130	178
5	40	20	60	24	36	60	38	32	70	32	34	66
รวม	950	2,304	3,254	1,010	2,352	3,362	1,154	2,312	3,466	1,930	3,658	5,588

จากผลการทดสอบพบว่าเวลาที่ใช้ในการประมวลผลใน Environment 1 น้อยกว่าใน Environment 2 ทั้งในการเชื่อมต่อฐานข้อมูลบน Local database และบนคลาวด์ โดยการเชื่อมต่อฐานข้อมูลบน Local database ภายใต้ Environment 1 ใช้เวลาน้อยกว่า Environment 2 ประมาณ 1.07 เท่าซึ่งไม่แตกต่างกันมากนัก และแบบเชื่อมต่อฐานข้อมูลบนคลาวด์ Environment 1 ใช้เวลาน้อยกว่าประมาณ 1.66 เท่า

ในส่วนของการเชื่อมต่อฐานข้อมูลโดยภาพรวม สามารถสรุปได้ว่าการเชื่อมต่อฐานข้อมูลบน Local database นั้นใช้เวลาน้อยกว่าการเชื่อมต่อฐานข้อมูลบนคลาวด์ โดยหากเปรียบเทียบส่วน Environment 1 พบว่าการเชื่อมต่อฐานข้อมูลผ่าน Local database ใช้เวลาน้อยกว่าการเชื่อมต่อฐานข้อมูลบนคลาวด์ประมาณ 1.03 เท่า และในส่วน Environment 2 การเชื่อมต่อฐานข้อมูลผ่าน Local database ใช้เวลาน้อยกว่าประมาณ 1.61 เท่า แสดงให้เห็นว่าทั้งใน Environment 2 การเชื่อมต่อฐานข้อมูลค่อนข้างส่งผลต่อเวลาที่ใช้ในการประมวลผล ซึ่งหากสังเกตจากผลการทดสอบที่ 1 ซึ่งเป็นการทดสอบด้านการเชื่อมต่อฐานข้อมูล จะพบว่าการเชื่อมต่อ Database บนคลาวด์ใช้เวลามากกว่าการเชื่อมต่อฐานข้อมูลแบบ Local database ซึ่งค่อนข้างแตกต่างกันอย่างเห็นได้ชัด

เมื่อพิจารณาปัจจัยด้านขนาดของข้อมูล พบว่าโดยรวมแล้วข้อมูลขนาดใหญ่จะใช้เวลาในการประมวลผลมากกว่าในทุก Environment และทุกรูปแบบการเชื่อมต่อฐานข้อมูล คิดอัตราส่วนเวลาที่ใช้ในการประมวลผลระหว่างข้อมูลขนาดเล็กและข้อมูลขนาดใหญ่โดยเฉลี่ยได้เป็น 1 : 2 ซึ่งหมายความว่าข้อมูลขนาดใหญ่ใช้เวลาในการประมวลผลมากกว่าข้อมูลขนาดเล็กถึง 2 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โดยสรุปแล้ว การวัดประสิทธิภาพด้านเวลาที่ใช้ในการประมวลผลพบว่าปัจจัยด้านการเชื่อมต่อฐานข้อมูลและปัจจัยด้านขนาดของข้อมูลมีความเกี่ยวข้องกับเวลาที่ใช้ในการประมวลผล โดยในส่วนของการเชื่อมต่อฐานข้อมูลพบว่าการเชื่อมต่อฐานข้อมูลบน Local database ใช้เวลาน้อยกว่าการเชื่อมต่อฐานข้อมูลแบบคลาวด์อย่างเห็นได้ชัดใน Environment 2 และในส่วนของขนาดข้อมูลพบว่าข้อมูลขนาดเล็กใช้เวลาในการประมวลผลน้อยกว่าข้อมูลขนาดใหญ่ แต่หากสังเกตจากเวลาที่ใช้ในแต่ละการทดสอบแล้ว พบว่าใน Environment 1 และ Environment 2 ใช้เวลาแตกต่างกันในการทดสอบที่ 1 (การเชื่อมต่อฐานข้อมูล) เท่านั้น ส่วนการทดสอบอื่น ๆ พบว่าใน Environment 1 และ Environment 2 ใช้เวลาในการประมวลผลใกล้เคียงกัน ดังนั้น อาจสรุปได้ว่าความแตกต่างด้าน Environment อาจไม่ส่งผลต่อเวลาที่ใช้ในการประมวลผลมากนัก

ในส่วนของผลแต่ละการทดสอบ พบว่าในการทดลองที่ 1 ซึ่งเป็นการทดสอบการเชื่อมต่อฐานข้อมูล พบว่าการเชื่อมต่อฐานข้อมูลแบบคลาวด์ใน Environment 2 ใช้เวลามากกว่าอย่างเห็นได้ชัด แสดงให้เห็นว่าปัจจัยด้านการเชื่อมต่อฐานข้อมูลที่แตกต่างค่อนข้างส่งผลต่อเวลาที่ใช้ในการประมวลผล และในส่วนการทดสอบที่ 2 การทดสอบฟังก์ชัน cleandata() และการทดสอบที่ 3 การทดสอบการรวมตารางและเพิ่มแถวข้อมูล พบว่าในส่วนของการทดสอบดังกล่าวใช้เวลาในการประมวลผลที่มากกว่าการทดสอบอื่น ๆ โดยการทดสอบที่ 2 มีการทำความสะอาดข้อมูลโดยผ่านหลายขั้นตอน เช่น การกลับตาราง การเปลี่ยนรูปแบบข้อมูล และในการทดสอบที่ 3 มีฟังก์ชันที่ใช้ในการรวมตารางข้อมูลขนาดใหญ่เข้าด้วยกันและกลับตารางข้อมูล ซึ่งกระบวนการทั้งหมดนี้อาจเป็นสาเหตุที่ทำให้การทดสอบทั้งสองใช้เวลาในการประมวลผลมากกว่าการทดสอบอื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

จากการทดลอง 6 ข้อตามสมมติฐานที่กล่าวไว้ สามารถสรุปผลได้ดังนี้

5.1.1 การวิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ทั่วโลก

จากการวิเคราะห์ข้อมูลเกี่ยวกับจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ทั่วโลก ในระยะเวลา 12 เดือนนับตั้งแต่วันที่ 21 มกราคม 2020 จนถึงวันที่ 21 มกราคม 2021 สามารถสรุปผลได้ดังต่อไปนี้

1) **การวิเคราะห์ 20 ประเทศที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุด** พบว่าประเทศที่มีจำนวนผู้ติดเชื้อมากที่สุดในโลกอันดับหนึ่งคือสหรัฐอเมริกาโดยมีร้อยละ 23.6 ของจำนวนผู้ติดเชื้อทั้งหมดทั่วโลก พบว่ามีร้อยละมากกว่าประเทศอื่นที่ไม่ติดอันดับ 20 ประเทศรวมกันซึ่งมีเพียงร้อยละ 17 ของจำนวนผู้ติดเชื้อทั้งหมดทั่วโลก ตามรูปที่ 4.2
2) **ปัจจัยที่มีผลต่อการแพร่ระบาดไวรัสโคโรนา 2019** ข้อมูลผลิตภัณฑ์มวลรวมในประเทศ(GDP) การเข้าชมเว็บไซต์ Pornhub และค่าอุณหภูมิเฉลี่ยรายปีส่งผลต่ออัตราการติดเชื้อ อัตราการรักษาหายและอัตราการตายน้อยมากหรือแทบไม่ส่งผลเลยส่วนอันดับสาธารณสุขโลกที่ได้จากการวิเคราะห์สถิติเกี่ยวกับคุณภาพโดยรวมของระบบสาธารณสุข เมื่อเทียบกับอัตราการรักษาหายมีความสัมพันธ์แบบแปรผกผันอย่างมากมีค่าสัมประสิทธิ์สหสัมพันธ์ -0.73 ทำให้สามารถสรุปได้ว่าอันดับสาธารณสุขยิ่งมีอันดับสูงแต่กลับสามารถรักษาผู้ติดเชื้อภายในประเทศได้น้อยมาก ตามรูปที่ 4.4

5.1.2 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

จากการวิเคราะห์ข้อมูลเกี่ยวกับจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ทั่วโลก ในระยะเวลา 12 เดือนตั้งแต่วันที่ 21 เดือนมกราคม 2020 ถึง 21 เดือนมกราคม 2021 สามารถสรุปผลได้ดังต่อไปนี้

1) **การวิเคราะห์ 20 รัฐที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในสหรัฐอเมริกา** พบว่ารัฐที่มีจำนวนผู้ติดเชื้อมากที่สุดในสหรัฐอเมริกาอันดับหนึ่งคือ California โดยมีร้อยละ 10.9 ของจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา รองลงมาคือรัฐ Texas และรัฐ Florida โดยคิดเป็นร้อยละ 8.9 และ 6.7 ตามลำดับ ตามรูปที่ 4.10 จำนวนผู้ติดเชื้อรวมทั้ง 3 รัฐนี้ คิดเป็นร้อยละได้มากถึง 26.5 ของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา และเมื่อรวมจำนวนผู้ติดเชื้อจาก 20 รัฐที่มีผู้ติดเชื้อสูงสุด พบว่ามากถึงร้อยละ 74.3 ของจำนวนผู้ติดเชื้อทั้งหมดในสหรัฐอเมริกา ซึ่งหมายความว่า 20 รัฐที่มีจำนวนผู้ติดเชื้อสูงสุดนั้นมียอดรวมผู้ติดเชื้อมากกว่าครึ่งหนึ่งของผู้ติดเชื้อรวมทั้ง 50 รัฐในสหรัฐอเมริกา

2) **ปัจจัยที่มีผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา**พบว่า ผลิตภัณฑ์มวลรวมในประเทศ (GDP) สูง จำนวนประชากรไร้บ้านสูง และจำนวนประชากรในรัฐมากจะมีจำนวนผู้ติดเชื้อและผู้เสียชีวิตสูงมากขึ้นตามไปด้วย ในขณะที่ข้อมูลจำนวนวันที่ทำการล็อกดาวน์ จำนวนประชากรเพศหญิง และจำนวนประชากรเพศชายแทบไม่มีความสัมพันธ์กับจำนวนผู้ติดเชื้อหรือผู้เสียชีวิต ตามรูปที่ 4.12

5.1.3 วิเคราะห์ข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในประเทศไทย

จากการวิเคราะห์ข้อมูลเกี่ยวกับจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย ในระยะเวลา 12 เดือนตั้งแต่วันที่ 12 เดือนมกราคม 2020 ถึง 12 เดือนมกราคม 2021 สามารถสรุปผลได้ดังต่อไปนี้

1) **การวิเคราะห์ 10 จังหวัดที่มีจำนวนผู้ติดเชื้อไวรัสโคโรนา** 2019 สูงสุดในประเทศไทย พบว่าจังหวัดที่มีจำนวนผู้ติดเชื้อมากที่สุดอันดับหนึ่งคือกรุงเทพมหานคร อันดับที่สองคือจังหวัดชลบุรี และอันดับที่สามคือจังหวัดสมุทรสาคร ซึ่งสามจังหวัดแรกที่กล่าวมารวมจำนวนผู้ติดเชื้อประมาณร้อยละ 41 ของจำนวนผู้ติดเชื้อทั้งหมดในประเทศไทย นอกจากจากนี้ยังพบว่าสามจังหวัดนี้มีจำนวนนักท่องเที่ยวชาวต่างชาติและแรงงานต่างด้าวมากเป็นอันดับต้นๆของประเทศไทย ตามรูปที่ 4.18

2) **จำนวนกลุ่มความเสี่ยง**ที่พบจากการติดเชื้อตามรูปที่ 4.15 สามอันดับแรกที่ได้แก่ กลุ่ม Close contact with the patient กลุ่ม Cluster Samut Sakhon และกลุ่ม State Quarantine รวมกันสูงถึงประมาณร้อยละ 50 จากกลุ่มความเสี่ยงทั้งหมดที่พบ จะเห็นได้ว่าสาเหตุหลักๆที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 มาจากจากการสัมผัสหรือใกล้ชิดกับผู้ป่วย และอยู่ในสถานที่ที่แออัดมีจำนวนคนมากซึ่งเป็นปัจจัยหลักในการแพร่ระบากไวรัสโคโรนา 2019

3) **เพศและอายุ** จำนวนเพศที่พบของผู้ติดเชื้อพบว่ามีโอกาสเกิดการติดเชื้อใกล้เคียงกันหรือแทบไม่ส่งผลต่อจำนวนการติดเชื้อ ตามรูปที่ 4.16 ส่วนอายุของผู้ติดเชื้อพบว่าส่วนใหญ่อยู่ในกลุ่มอายุ 21-30 ปี จะเห็นได้ว่าในช่วงอายุที่พบมากที่สุดคือช่วงอายุวัยทำงานและมีครอบครัว ซึ่งต้องออกไปทำงานข้างนอก ออกไปใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สถานที่ที่เสี่ยงต่อการติดเชื้อ และออกไปพบผู้คนจำนวนมากจึงมีโอกาสเสี่ยงที่จะได้รับการติดเชื้อไวรัสโคโรนา 2019

4) **สัญชาติ** ที่พบในจำนวนผู้ติดเชื้อในประเทศไทย อันดับหนึ่งคือสัญชาติไทย อันดับสองคือคือสัญชาติเมียนม่าร์ และอันดับสามคือไม่ระบุสัญชาติ พบว่าสัญชาติที่พบผู้ติดเชื้อส่วนใหญ่ในประเทศไทยเป็นประเทศเพื่อนบ้านที่เข้ามาทำงานและเป็นนักท่องเที่ยวชาวต่างชาติ ตามรูปที่ 4.17

5.1.4 ผลลัพธ์ที่ได้จาก SEIR model of covid-19 in Thailand

จากการใช้แบบจำลองทางคณิตศาสตร์ SEIR model ในประเทศไทยพบว่าแบบมีมาตรการป้องกันจะมีจำนวนผู้ติดเชื้อสูงสุดต่ำกว่าแบบไม่มีมาตรการป้องกันถึง 4 เท่า เมื่อเพิ่มมาตรการ social distancing และเวลาที่พบผู้ติดเชื้อสูงสุดจะช้ากว่าแบบไม่มีมาตรการป้องกันประมาณ 310 วัน เมื่อเพิ่มมาตรการ social distancing

เมื่อวัดประสิทธิภาพโมเดลด้วย RMSE เทียบกับผู้ติดเชื้อในประเทศไทยจริงใน 7 วันแรก พบว่าค่า RMSE มีค่าเท่ากับ 6.8 นั่นหมายความว่าโดยเฉลี่ยโมเดลจะทำนายค่าผู้ติดเชื้อผิดไป +/- 6.8 คน จากการทำนายใน 7 วันแรกของการติดเชื้อ เมื่อเพิ่มระยะเวลาในการทำนาย ค่า RMSE จะเพิ่มขึ้นเรื่อย ๆเนื่องจากยิ่งเวลามากขึ้นปัจจัยที่จะเข้ามามีผลต่อการติดเชื้อไวรัสโคโรนา 2019 ยิ่งมากขึ้น ไม่ว่าจะเป็นปัจจัยที่ส่งผลดี เช่น การได้รับวัคซีน หรือค้นพบยารักษาในอนาคต และปัจจัยที่ส่งผลเสีย เช่น การพบกลุ่มแพร่กระจายไวรัสกลุ่มใหม่ หรือการค้นพบไวรัสสายพันธุ์ใหม่ที่อาจมีผลร้ายแรงกว่าเดิม การจำลอง และทำนายสถานการณ์การแพร่ระบาดโดยใช้ SEIR model ให้แม่นยำนั้นเป็นไปได้ยาก จึงบอกได้เพียงแค่ว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

5.1.5 ผลลัพธ์ที่ได้จาก SEIRD model of covid-19 in United States

จากการใช้แบบจำลองทางคณิตศาสตร์ SEIRD model ในสหรัฐอเมริกาพบว่าแบบมีมาตรการป้องกันจะมีจำนวนผู้ติดเชื้อสูงสุดต่ำกว่าแบบไม่มีมาตรการป้องกันถึง 0.65 เท่า เมื่อเพิ่มมาตรการ social distancing และเวลาที่พบผู้ติดเชื้อสูงสุดจะช้ากว่าแบบไม่มีมาตรการป้องกันประมาณ 100 วัน เมื่อเพิ่มมาตรการ social distancing ส่วนจำนวนผู้เสียชีวิตจากการติดเชื้อแบบมีมาตรการป้องกันจะค่อยๆเพิ่มสูงขึ้นเรื่อย ๆ แต่ยังมีจำนวนผู้ติดเชื้อสูงกว่าจำนวนผู้เสียชีวิตนั่นหมายความว่าผู้ที่ติดเชื้อมีโอกาสเสียชีวิตจากโรคระบาดน้อยกว่าเมื่อมีมาตรการป้องกัน

เมื่อวัดประสิทธิภาพโมเดลด้วย RMSE เทียบกับผู้ติดเชื้อในสหรัฐอเมริกาจริงใน 7 วันแรก พบว่าค่า RMSE มีค่าเท่ากับ 0.48 นั่นหมายความว่าโดยเฉลี่ยโมเดลจะทำนายค่าผู้ติดเชื้อผิดไป +/- 0.48 คน จากการทำนายใน 7 วันแรกของการติดเชื้อ เมื่อเพิ่มระยะเวลาในการทำนาย ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

RMSE จะเพิ่มขึ้นเรื่อย ๆเนื่องจากยิ่งเวลามากขึ้นปัจจัยที่จะเข้ามามีผลต่อการติดเชื้อไวรัสโคโรนา 2019 ยิ่งมากขึ้น ไม่ว่าจะเป็นปัจจัยที่ส่งผลดี เช่น การได้รับวัคซีน หรือค้นพบยารักษาในอนาคต และปัจจัยที่ส่งผลเสีย เช่น การพบกลุ่มแพร่กระจายไวรัสกลุ่มใหม่ หรือการค้นพบไวรัสสายพันธุ์ใหม่ที่อาจมีผลร้ายแรงกว่าเดิม การจำลอง และทำนายสถานการณ์การแพร่ระบาดโดยใช้ SEIRD model ให้แม่นยำนั้นเป็นไปได้ยาก จึงบอกได้เพียงแค่ว่าหากมีหรือไม่มีมาตรการป้องกันจะส่งผลอย่างไร เพื่อเป็นประโยชน์ในการรับมือกับการแพร่ระบาดไวรัสโคโรนา 2019 ที่จะเกิดขึ้นในอนาคต

5.1.6 การวัดประสิทธิภาพโดยใช้ Profiling

1) **การวัดประสิทธิภาพเพื่อเปรียบเทียบระหว่างการใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr** จากการวัดประสิทธิภาพโดยใช้ Profiling สามารถสรุปผลได้ว่า การวิเคราะห์ข้อมูลโดยใช้แพ็กเกจ Dplyr ดีกว่าแบบไม่ใช้แพ็กเกจ Dplyr เนื่องจากแบบใช้แพ็กเกจ Dplyr มีการใช้หน่วยความจำและใช้เวลาในการประมวลผลต่ำกว่า นอกจากนี้เมื่อพิจารณาส่วนของ Flame Graph ของการทดลองแบบใช้แพ็กเกจ Dplyr พบว่าเวลาที่ใช้ส่วนมากเกิดจากการเรียกใช้ฟังก์ชัน “%>%” หรือ “%<>%” ซึ่งเป็น Pipeline Operation จึงทำให้ในบางกรณีแพ็กเกจ Dplyr อาจใช้พื้นที่หรือเวลามากกว่าจากการใช้ Pipeline Operation ตามตารางที่ 4.1 ซึ่งอาจเพิ่มประสิทธิภาพในการใช้งาน Pipeline ได้ด้วยการใช้ %T>% เพื่อลดเวลาในการประมวลผล
2) **การวัดประสิทธิภาพของแพ็กเกจ Dplyr เมื่อมีปัจจัยอื่นเข้ามาเกี่ยวข้อง** ปัจจัยด้านการเชื่อมต่อฐานข้อมูล และปัจจัยด้านขนาดของข้อมูลมีความเกี่ยวข้องกับเวลาที่ใช้ในการประมวลผล โดยส่วนของการเชื่อมต่อฐานข้อมูลพบว่าการเชื่อมต่อฐานข้อมูลบน Local database ใช้เวลาน้อยกว่าการเชื่อมต่อฐานข้อมูลแบบคลาวด์ และในส่วนของขนาดข้อมูลพบว่าข้อมูลขนาดเล็กใช้เวลาในการประมวลผลน้อยกว่าข้อมูลขนาดใหญ่ ในส่วน Environment พบว่าทั้งใน Environment 1 และ Environment 2 ซึ่งแตกต่างกันในเรื่อง OS มีการใช้เวลาประมวลผลที่ค่อนข้างใกล้เคียงกัน ดังนั้น Environment ที่แตกต่างไม่ส่งผลต่อการประมวลผลมากนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

5.2 ปัญหาและอุปสรรค

5.2.1 การวิเคราะห์ข้อมูล

1) ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลโรคติดเชื้อไวรัสโคโรนา 2019 ซึ่งเกี่ยวข้องกับเหตุการณ์ที่เกิดขึ้นในปัจจุบันส่งผลให้การค้นหาข้อมูลที่ต้องการนำมาวิเคราะห์ร่วมกันนั้นเป็นไปได้ยาก เนื่องจากข้อมูลบางชนิดเช่น ผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ต้องรอระยะเวลาให้ครบปีเท่านั้นจึงจะสามารถนำมาวิเคราะห์ได้
2) เนื่องจากข้อมูลที่นำมาวิเคราะห์ บางชุดเป็นข้อมูลที่มีหน่วยต่างกัน จึงทำให้ยากต่อการทำ Visualization อย่าง Correlation ซึ่งจำเป็นต้องมีแกนเป็นหน่วยเดียวกัน เช่นหากแกนเวลาเป็นปี ทุกชุดข้อมูลจะต้องมีข้อมูลเป็นปีเหมือนกัน จึงจะสามารถนำมาหาความสัมพันธ์กันได้
3) เครื่องมือที่ใช้ในการวิเคราะห์คือ Rstudio มีแพ็กเกจให้เลือกใช้เป็นจำนวนมาก ชื่อฟังก์ชันบางแพ็กเกจซ้ำกันเมื่อเรียกใช้พร้อมกันทำให้มีปัญหาเนื่องจากมี Argument ต่างกัน เช่นฟังก์ชัน rename() ในแพ็กเกจ tidyverse และ plyr

5.2.2 การวัดประสิทธิภาพโดยใช้ Profiling

1) เนื่องจาก Profiler ทำงานเป็นแบบสุ่มตัวอย่าง (Sampling) จึงทำให้การใช้หน่วยความจำและเวลาที่ใช้ มีค่าต่างกันในแต่ละครั้งที่ทำการประมวลผล ทำให้ไม่สามารถทราบค่าที่แน่นอนได้ จึงต้องทำการ Profiling ภายในลูปเดียวกันเพื่อเปรียบเทียบเป็นอัตราส่วนแทน
2) การใช้หน่วยความจำที่ปรากฏในผลการ Profiling มีการทำงานแบบ Non-deter ministic behavior ของ Garbage Collection ทำให้ข้อมูลการใช้งานหน่วยความจำมีการทำงานแบบสุ่ม
3) โค้ดแบบใช้แพ็กเกจ Dplyr และแบบไม่ใช้แพ็กเกจ Dplyr มีบางฟังก์ชันที่มีชื่อเหมือนกัน แต่มี Argument ต่างกัน ทำให้ไม่สามารถทดลอง Profiling ด้วยกันได้ และส่งผลทำให้ไม่สามารถเทียบสัดส่วนได้โดยตรง ข้อมูลที่ใช้ในการเปรียบเทียบจึงอาจมีความคาดเคลื่อนเล็กน้อย
4) ฟังก์ชัน Magrittr มี Pipeline Operation ซึ่งส่งผลต่อค่าการใช้หน่วยความจำและเวลาในการประมวลผลของแพ็กเกจ Dplyr ทำให้ค่าดังกล่าวสูงกว่าที่ควรเป็น โดยที่ค่านั้นไม่ได้เกิดจากการเรียกใช้ฟังก์ชันของแพ็กเกจ Dplyr

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

5.2.3 ข้อจำกัดของ R

1) มีปัญหากับการประมวลผลข้อมูลที่มีขนาดใหญ่มาก ๆ เนื่องจากข้อมูลใน R จะถูกจัดเก็บในหน่วยความจำของคอมพิวเตอร์ ซึ่งส่งผลต่อการจัดการหน่วยความจำและอาจทำให้โปรแกรมต่าง ๆ มีการประมวลผลที่ช้าลง
2) ไม่สามารถใช้บนเว็บบราวเซอร์ได้ เนื่องจากมีปัญหาในด้านความปลอดภัย

5.3 แนวทางการพัฒนาต่อ

1) การนำเอาปัจจัยที่เกี่ยวข้องอื่นๆมาคิดเพิ่มในแบบจำลอง model ทางคณิตศาสตร์เพื่อ Predictive Analytics แนวโน้มการแพร่ระบาดของโรคติดต่อไวรัสโคโรนา 2019 ในอนาคตให้แม่นยำยิ่งขึ้นนี้
2) การทำ Application รายงานสถานที่เสี่ยงต่อการแพร่ระบาดไวรัสโคโรนา 2019 ที่ใกล้กับตำแหน่งที่เราอยู่
3) การทำ Predictive Analytics ว่าติดเชื้อไวรัสโคโรนา 2019 หรือไม่จากอาการที่เป็นอยู่และพัฒนาเป็นระบบ AI

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

Brian Ripley. 2020. **ODBC Connectivity.** [Online]. Available : https://rdrr.io/cran/RODBC/f/inst/doc/RODBC.pdf.

Vitalie Spinu., Garrett Grolemund., Hadley Wickham. 2020. **lubridate.** [Online]. Available : https://lubridate.tidyverse.org.

Vitalie Spinu. **Make Dealing with Dates a Little Easier**. 2020. [Online]. Available : https://cran.r-project.org/web/packages/lubridate/lubridate.pdf.

Wickham, H. **An Introduction to reshape2**. 2007. [Online]. Available : https://seananderson.ca/2013/10/19/reshape.

Hadley Wickham. **Flexibly Reshape Data: A Reboot of the Reshape Package.** 2020. [Online]. Available : https://cran.r-project.org/web/packages/reshape2/reshape2.pdf.

Hadley Wickham. **Tools for Splitting, Applying and Combining Data.** 2020. [Online]. Available : https://cran.r-project.org/web/packages/plyr/plyr.pdf.

Luke Tierney., Riad Jarjour. **Examining R Profiling Data: The proftools Package**. 2016. [Online]. Available : https://cran.r-project.org/web/packages/proftools/vignettes/proftools.pdf.

Hadley Wickham. **Advanced R 23 Measuring performance**. 2019. [Online]. Available : https://adv-r.hadley.nz/perf-measure.html.

Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE). **Novel Coronavirus (COVID-19) Cases Data**. 2020. [Online]. Available : https://data.humdata.org/dataset/ novel-coronavirus-2019-ncov-cases.

IMF คืออะไร ทำไมถึงมีบทบาทเป็นเจ้าหนี้ของโลก. 2018. [Online]. Available : https://money.kapook.com/view123084.html.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

คมปทิต สกุลหวง. **จาก Great Depression สู่ Great Lockdown: IMF คาดเศรษฐกิจโลกหดตัว 3% ในปี 2020 จากพิษโควิด-19 GDP ไทยติดลบเหมือนหลายประเทศ.** 2020. [Online]. Available : https://thestandard.co/great-depression-to-great-lockdown.

World Economic Outlook (WEO). **World Economic Outlook (October 2020).** 2020. [Online]. Available : https://www.imf.org/external/datamapper/datasets/WEO.

International Monetary Fund. **IMF's Balance of Payments and International Investment Position Manual (BPM6)**. 2020. [Online]. Available : https://www.imf.org/en/Data.

List of cities by average temperature. 2020. [Online]. Available : https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature.

Department of Economic and Social Affairs Population Dynamics. **Population Division World Population Prospects 2019.** 2019. [Online]. Available : https://population.un.org/wpp.

Tanu N Prabhu. **Population by Country – 2020**. 2020. [Online]. Available : https://www.kaggle.com/tanuprabhu/population-by-country-2020.

World Bank Group. **Age dependency ratio, young (% of working-age population)**. 2019. [Online]. Available : https://data.worldbank.org/indicator/SP.POP.DPND.YG.

Hfocus. **ไทยติดอันดับ 6 ระบบสุขภาพดีที่สุดในโลก ไต้หวันคว้าที่ 1**. 2019. [Online]. Available : https://www.hfocus.org/content/2019/09/17663.

Sophie Ireland. **Revealed: Countries With The Best Health Care Systems, 2019**. 2019. [Online]. Available : https://ceoworld.biz/2019/08/05/revealed-countries-with-the-best-health-care-systems-2019.

Engkarat Techapanurak. **Bayesian Neural Network (ตอนที่ 3): อะไรคือ Deep Learning และอะไรคือ Neural Network - ฉบับมือใหม่**. 2018. [Online]. Available : https://medium.com/@dopplerz/bayesian-neural-network-cce9227ab599.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Keng Surapong. **Activation Function คืออะไร ใน Artificial Neural Network, Sigmoid Function คืออะไร-Activation Function ep.1**. 2019. [Online]. Available : https://www.bualabs.com/archives/1261.

Sigmoid function. 2020. [Online]. Available : https://en.wikipedia.org/wiki/Sigmoid_function.

CORALINE CO. LTD. **สถิติเบื้องต้นง่ายๆ ที่จะทำให้คุณเข้าใจการวิเคราะห์มากขึ้น (ตอนที่ 2)**. 2017. [Online]. Available : https://medium.com/@info_46914/1f94b6664ede.

สถิติและการวิเคราะห์ข้อมูลทางสุขภาพ. **บทที่ 7 สหสัมพันธ์ (Correlation).** 2019. [Online]. Available : http://intraserver.nurse.cmu.ac.th/mis/download/course/lec_567730_lesson_07.pdf.

Sarayut Khamkhiao. **ทำ Heatmap ด้วย Crazy Egg**. 2018. [Online]. Available : https://engineering.thinknet.co.th/7506aaddbd3c.

Karnawat Wongudom. **[DataVisualization] ทำความรู้จักกับ Tree Map และ Heat Map.** 2020. [Online]. Available : https://siamchamnankit.co.th/17861452f85c.

KASIDIS SATANGMONGKOL. **ความสัมพันธ์ระหว่าง Covariance, Correlation และ R-Squared.** 2019. [Online]. Available : https://datarockie.com/2019/09/03/covariance-correlation-rsquared.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาคผนวก ก

การออกแบบฐานข้อมูล

การออกแบบฐานข้อมูล

ภายในฐานข้อมูลประกอบด้วยตารางทั้งหมด 38 ตาราง ดังนี้

ตาราง ก.1 ข้อมูล D1: แสดงข้อมูลผู้ติดเชื้อไวรัสโคโรนา 2019 ที่ได้รับการยืนยัน

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	Province.State	TEXT(65535)	ชื่อเมือง, รัฐ	
2	PK	Country.Region	TEXT(65535)	ชื่อประเทศ	
3	-	Lat	DOUBLE(22)	เส้นละติจูด	
4	-	Long	DOUBLE(22)	เส้นลองจิจูด	
5	-	22/01/2020-22/01/2021	INT(11)	วันที่ (ตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021)	

ตาราง ก.2 ข้อมูล D2: แสดงข้อมูลผู้ที่เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	Province.State	TEXT(65535)	ชื่อเมือง, รัฐ	
2	PK	Country.Region	TEXT(65535)	ชื่อประเทศ	
3	-	Lat	DOUBLE(22)	เส้นละติจูด	
4	-	Long	DOUBLE(22)	เส้นลองจิจูด	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ก.2 ข้อมูล D2: แสดงข้อมูลผู้ที่เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
5	-	22/01/2020-22/01/2021	INT(11)	วันที่ (ตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021)	

ตารางที่ ก.3 ข้อมูล D3: ตารางแสดงข้อมูลผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา 2019

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	Province.State	TEXT(65535)	ชื่อเมือง, รัฐ	
2	PK	Country.Region	TEXT(65535)	ชื่อประเทศ	
3	-	Lat	DOUBLE(22)	เส้นละติจูด	
4	-	Long	DOUBLE(22)	เส้นลองจิจูด	
5	-	22/01/2020-22/01/2021	INT(11)	วันที่ (ตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021)	

ตารางที่ ก.4 ข้อมูล D4: แสดงข้อมูลอุณหภูมิทั่วโลก ประจำปี 2020

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	MyUnknownColumn	INT(11)	ลำดับ	
2	PK,FK	Country	TEXT(65535)	ชื่อประเทศ	city
3	-	City	TEXT(65535)	ชื่อเมือง, รัฐ	
4	-	Jan - Dec	DOUBLE(22)	เดือน(ตั้งแต่เดือนมกราคมถึงเดือนธันวาคม ปี 2020)	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.4 ข้อมูล D4: แสดงข้อมูลอุณหภูมิทั่วโลก ประจำปี 2020 (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
5	-	Avg_Year	TEXT(65535)	อุณหภูมิเฉลี่ยของปี 2020	5
6	-	Continent	TEXT(65535)	ชื่อทวีป	6

ตารางที่ ก.5 ข้อมูล D5: แสดงจำนวนประชากรของแต่ละประเทศทั่วโลกประจำปี 2020

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	Country	TEXT(65535)	ชื่อประเทศ	
2	-	Population(2020)	INT(11)	จำนวนประชากรภายในประเทศ	

ตารางที่ ก.6 ข้อมูล D6 แสดงข้อมูลการแพร่ระบาดโรคซาร์ส

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	Date	TEXT(65535)	วันที่	
2	PK	country	TEXT(65535)	ชื่อประเทศ	
3	-	Cumulative_number	INT(11)	จำนวนผู้ติดเชื้อสะสม	
4	-	Number_deaths	INT(11)	จำนวนผู้เสียชีวิต	
5	-	Number_recover	INT(11)	จำนวนผู้รักษาหาย	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.7 ข้อมูล D7:แสดงข้อมูลผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ประจำปี 2019

ฟิลด์	คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	code	TEXT(65535)	รหัสประเทศ	
2	-	rank	INT(11)	อันดับ GDP ของประเทศ	
3	PK	country	TEXT(65535)	ชื่อประเทศ	
4	-	GDP (millions of US dollars)	INT(11)	ค่าผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP)	

ตารางที่ ก.8 ข้อมูล D8: แสดงข้อมูลด้านสาธารณสุข ประจำปี 2019

ฟิลด์	คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	rank	INT(11)	อันดับด้านสาธารณสุขของประเทศ	
2	PK	country	TEXT(65535)	ชื่อประเทศ	
3	-	healthCareIndex	DOUBLE(22)	คะแนนภาพรวมทางด้านสาธารณสุข	
4	-	infrastructure	DOUBLE(22)	คะแนนด้านโครงสร้างพื้นฐาน	
5	-	professional	DOUBLE(22)	คะแนนด้านบุคลากรผู้เชี่ยวชาญ	
6	-	cost	DOUBLE(22)	คะแนนเกี่ยวกับค่าใช้จ่ายด้านสาธารณสุข	
7	-	medAvail	DOUBLE(22)	คะแนนการเข้าถึงยา	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.8 ข้อมูล D8: แสดงข้อมูลด้านสาธารณสุข ประจำปี 2019 (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
8	-	govReadiness	DOUBLE(22)	คะแนนความพร้อมของรัฐบาล	

ตารางที่ ก.9 ข้อมูล D9: แสดงข้อมูลการเข้าชมเว็บไซต์ PornHub ประจำปี 2019

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	Rank	INT(11)	อันดับประเทศที่เข้าชมเว็บไซต์ Pornhub	
2	PK	country	TEXT(65535)	ชื่อประเทศ	
3	-	PornHubIndex(%)	DOUBLE(22)	ร้อยละคะแนนอันดับการเข้าชมเว็บไซต์ PornHub	

ตารางที่ ก.10 ข้อมูล D10: แสดงข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	MyUnknownColumn	INT (11)	จำนวนแถว	
2	-	date	TEXT(65535)	วันที่	
3	PK	state	TEXT(65535)	ชื่อรัฐ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.10 ข้อมูล D10: แสดงข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
4	-	fips	INT (11)	มาตรฐานของรัฐบาลสหรัฐฯ ด้านเทคโนโลยีสารสนเทศและระบบความปลอดภัยคอมพิวเตอร์	
5	-	cases	INT (11)	จำนวนผู้ติดเชื้อ	
6	-	deaths	INT (11)	จำนวนผู้เสียชีวิต	

ตารางที่ ก.11 ข้อมูล D11: แสดงข้อมูลจำนวนประชากร ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	MyUnknownColumn	INT(11)	จำนวนแถว	
2	PK	State	TEXT(65535)	ชื่อรัฐ	
3	-	Population	INT(11)	จำนวนประชากร	

ตารางที่ ก.12 ข้อมูล D12: แสดงข้อมูลจำนวนแต่ละเพศ ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	MyUnknownColumn	INT (11)	จำนวนแถว	
2	-	State	TEXT(65535)	ชื่อรัฐ	
3	PK	Male	DOUBLE(22)	ร้อยละเพศชาย	
4	-	Female	DOUBLE(22)	ร้อยละเพศหญิง	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.12 ข้อมูล D12: แสดงข้อมูลจำนวนแต่ละเพศ ในสหรัฐอเมริกา (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
5	-	Unnamed: 3	INT (11)	ร้อยละโดยรวม	

ตารางที่ ก.13 ข้อมูล D13: แสดงข้อมูลจำนวนวัน lockdown ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	MyUnknownColumn	INT(11)	จำนวนแถว	
2	PK	State	TEXT(65535)	ชื่อรัฐ	
3	-	Month	INT(11)	เดือนที่ lockdown	
4	-	Day	INT(11)	วันที่ lockdown	
5	-	Date	TEXT(65535)	วันที่	
6	-	Type	INT(11)	ชนิดข้อประเภทที่ lockdown	
7	-	Day lockdown	INT(11)	จำนวนวันที่ lockdown	

ตารางที่ ก.14 ข้อมูล D14 แสดงข้อมูลผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ประจำปี 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	State	TEXT(65535)	ชื่อรัฐ	
2	PK	GDPs	INT(11)	แสดงข้อมูลผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP)	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.14 ข้อมูล D14 แสดงข้อมูลผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ประจำปี 2019 ในสหรัฐอเมริกา (ต่อ)

ฟิลด์	คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
3	-	GDP Growth in 2018	TEXT(65535)	แสดงข้อมูลการเติบโตผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP)	

ตารางที่ ก.15 ข้อมูล D15: แสดงข้อมูลจำนวนประชากรที่ไร้บ้าน ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	State	TEXT(65535)	ชื่อรัฐ	
2	-	Number of CoCs	INT(11)	จำนวนผู้ติดเชื้อ	
3	-	Homeless	INT(11)	จำนวนผู้ไร้บ้าน	

ตารางที่ ก.16 ข้อมูล D16: แสดงข้อมูลจำนวนประชากร ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	State	TEXT(65535)	ชื่อรัฐ	
2	-	GDPs	INT(11)	ผลิตภัณฑ์มวลรวมในประเทศ	
3	-	GDP Growth in 2018	DOUBLE(22)	การเติบโตผลิตภัณฑ์มวลรวมในประเทศ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.17 ข้อมูล D17: แสดงข้อมูลคะแนนระบบสาธารณสุข ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	State	TEXT(65535)	ชื่อรัฐ	
2	-	Health	INT (11)	คะแนนสาธารณสุขแต่ละรัฐ	

ตารางที่ ก.18 ข้อมูล D18: แสดงข้อมูลการระบาดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	No.	INT(11)	ลำดับ	
2	-	announce_date	TEXT(65535)	วันที่พบผู้ติดเชื้อ	
3	-	Notification_date	TEXT(65535)	วันที่รักษาหาย	
4	-	sex	TEXT(65535)	เพศ	
5	-	age	INT(11)	อายุ	
6	-	nationality	TEXT(65535)	สัญชาติผู้ติดเชื้อ	
7	-	province_of_isolation	TEXT(65535)	จังหวัดที่รักษาผู้ติดเชื้อ	
8	-	province_of_onset	TEXT(65535)	จังหวัดที่พบผู้ติดเชื้อ	
9	-	risk	TEXT(65535)	ความเสี่ยงจากการติดเชื้อ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.19 ข้อมูล dx1: แสดงข้อมูลการแพร่ระบาดไวรัสโคโรนา 2019 ทั่วโลก

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	country	TEXT(65535)	ชื่อประเทศ	
2	-	date	INT(11)	วันที่ (ตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021)	
3	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	
4	-	deaths	INT(11)	จำนวนผู้เสียชีวิต	
5	-	recovered	INT(11)	จำนวนผู้รักษาหาย	

ตารางที่ ก.20 ข้อมูล dx2: แสดงข้อมูล 20 อันดับประเทศที่ติดเชื้อไวรัสโคโรนา 2019 สูงสุด

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	country	TEXT(65535)	ชื่อประเทศ 20 อันดับที่ติดเชื้อสูงสุด	
2	-	date	INT(11)	วันที่ (ตั้งแต่วันที่ 22 มกราคม 2020 ถึง 22 มกราคม 2021)	
3	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	
4	-	deaths	INT(11)	จำนวนผู้เสียชีวิต	
5	-	recovered	INT(11)	จำนวนผู้รักษาหาย	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.21 ข้อมูล dx3: แสดงข้อมูลปัจจัยที่อาจส่งผลต่อการแพร่ระบาดไวรัสโคโรนา 2019

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	country	TEXT(65535)	ชื่อประเทศ 20 อันดับที่ติดเชื้อสูงสุด	
2	-	confirmed.rate	DOUBLE(22)	อัตราการติดเชื้อ	
3	-	deaths.rate	DOUBLE(22)	อัตราการเสียชีวิต	
4	-	recovered.rate	DOUBLE(22)	อัตราการรักษาหาย	
5	-	healthcare	DOUBLE(22)	คะแนนสาธารณสุข	
6	-	avg_temp	DOUBLE(22)	ค่าเฉบี่ยอุณหภูมิปี 2020	
7	-	GDP	DOUBLE(22)	ค่าผลรวมผลิตภัณฑ์มวลรวมภายในประเทศ (GDP)	
8	-	Pornhub	DOUBLE(22)	ร้อยละคะแนนอันดับการเข้าชมเว็บไซต์ PornHub	
9	-	sars	DOUBLE(22)	จำนวนผู้ติดเชื้อโรคซาร์ส	

ตารางที่ ก.22 ข้อมูล dx4: แสดงข้อมูลจำนวนการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	date	INT(11)	วันที่	
2	PK	state	TEXT(65535)	ชื่อรัฐ	
3	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.22 ข้อมูล dx4: แสดงข้อมูลจำนวนการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
4	-	deaths	INT(11)	จำนวนผู้เสียชีวิต	
5	-	new.confirmed	INT(11)	จำนวนผู้ติดเชื้อราย ใหม่	
6	-	new.deaths	INT(11)	จำนวนผู้เสียชีวิตราย ใหม่	

ตารางที่ ก.23 ข้อมูล dx5: แสดงข้อมูลอัตราการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	date	INT(11)	วันที่	
2	PK	state	TEXT(65535)	ชื่อรัฐ	
3	-	confirmed.rate	DOUBLE(22)	อัตราการติดเชื้อ	
4	-	deaths.rate	DOUBLE(22)	อัตราการเสียชีวิต	

ตารางที่ ก.24 ข้อมูล dx6: แสดงข้อมูล 20 รัฐที่ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	state	TEXT(65535)	ชื่อรัฐ 20 รัฐที่ติดเชื้อ สูงสุด	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	
3	-	deaths	INT(11)	จำนวนผู้เสียชีวิต	
4	-	confirmed.rate	DOUBLE(22)	อัตราการติดเชื้อ	
5	-	deaths.rate	DOUBLE(22)	อัตราการเสียชีวิต	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.25 ข้อมูล dx7: แสดงข้อมูลปัจจัยที่อาจส่งผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	PK	state	TEXT(65535)	ชื่อรัฐ 20 รัฐที่ติดเชื้อสูงสุด	
2	-	Homeless	INT(11)	จำนวนคนไร้บ้าน	
3	-	GDPs	DOUBLE(22)		
4	-	Day lockdown	INT(11)	จำนวนวันที่ lockdown	
5	-	Population	INT(11)	จำนวนประชากร	
6	-	Female	INT(11)	จำนวนประชากรเพศหญิง	

ตารางที่ ก.25 ข้อมูล dx7: แสดงข้อมูลปัจจัยที่อาจส่งผลต่อการแพร่ระบาดไวรัสโคโรนา 2019 ในสหรัฐอเมริกา (ต่อ)

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	male	INT(11)	จำนวนประชากรเพศชาย	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	
3	-	deaths	INT(11)	จำนวนผู้เสียชีวิต	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.26 ข้อมูล dx8: แสดงข้อมูลจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	announce_date	INT(11)	วันที่พบผู้ติดเชื้อ	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อ	

ตารางที่ ก.27 ข้อมูล dx9: แสดงข้อมูลความเสี่ยงจากการติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	announce_date	INT(11)	วันที่พบผู้ติดเชื้อ	
2	-	risk	TEXT(65535)	ความเสี่ยงจากการติดเชื้อ	

ตารางที่ ก.28 ข้อมูล dx10: แสดงข้อมูลจังหวัดที่พบผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	province	TEXT(65535)	จังหวัดที่พบผู้ติดเชื้อ	
2	-	onset	INT(11)	จำนวนผู้ติดเชื้อที่พบในจังหวัดนั้นๆ	

ตารางที่ ก.29 ข้อมูล dx11: แสดงข้อมูลจังหวัดที่รักษาผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	province	TEXT(65535)	จังหวัดที่รักษาผู้ติดเชื้อ	
2	-	isolation	INT(11)	จำนวนผู้ติดเชื้อที่รักษาในจังหวัดนั้นๆ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.30 ข้อมูล dx12: แสดงข้อมูลเพศผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	sex	TEXT(65535)	เพศผู้ติดเชื้อ	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อแต่ละเพศ	

ตารางที่ ก.31 ข้อมูล dx13: แสดงข้อมูลอายุผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	age	INT(11)	อายุผู้ติดเชื้อ	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อแต่ละอายุ	

ตารางที่ ก.32 ข้อมูล dx14: แสดงข้อมูลสัญชาติผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	nationality	INT(11)	สัญชาติผู้ติดเชื้อ	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อแต่ละสัญชาติ	

ตารางที่ ก.33 ข้อมูล dx15: แสดงข้อมูลจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 แต่ละวันในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	announce_date	INT(11)	วันที่พบผู้ติดเชื้อ	
2	-	new.confirmed	INT(11)	จำนวนผู้ติดเชื้อรายใหม่	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.34 ข้อมูล dx16: แสดงข้อมูลจำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สะสมในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	announce_date	INT(11)	วันที่พบผู้ติดเชื้อ	
2	-	confirmed_cumulative	INT(11)	จำนวนผู้ติดเชื้อสะสม	

ตารางที่ ก.35 ข้อมูล dx17: แสดงข้อมูลจำนวนกลุ่มความเสี่ยงจากผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	announce_date	INT(11)	วันที่พบผู้ติดเชื้อ	
2	-	risk	TEXT(65535)	ความเสี่ยงจากการติดเชื้อ	
3	-	confirmed_risk	INT(11)	จำนวนผู้ติดเชื้อในแต่ละกลุ่มความเสี่ยง	

ตารางที่ ก.36 ข้อมูล dx18: แสดงข้อมูล 10 จังหวัดผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	province	TEXT(65535)	10 จังหวัดผู้ติดเชื้อสูงสุด	
2	-	isolation	INT(11)	จำนวนผู้ติดเชื้อที่รักษาในจังหวัดนั้นๆ	
3	-	onset	INT(11)	จำนวนผู้ติดเชื้อที่พบในจังหวัดนั้นๆ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.37 ข้อมูล dx19: แสดงข้อมูลเพศและอายุผู้ติดเชื้อไวรัสโคโรนา 2019 ในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	-	sex	TEXT(65535)	เพศผู้ติดเชื้อ	
2	-	confirmed_sex	INT(11)	จำนวนผู้ติดเชื้อแต่ละเพศ	
3	-	confirmed_age	INT(11)	จำนวนผู้ติดเชื้อแต่ละอายุ	

ตารางที่ ก.38 ข้อมูล dx20: แสดงข้อมูล 10 อันดับสัญชาติผู้ติดเชื้อไวรัสโคโรนา 2019 สูงสุดในประเทศไทย

ฟิลด์	คีย์	ชื่อแอดทริบิวต์	ชนิดตัวแปร	รายละเอียด	ตารางที่อ้างถึง
1	FK	nationality	INT(11)	10 อันดับสัญชาติผู้ติดเชื้อ	
2	-	confirmed	INT(11)	จำนวนผู้ติดเชื้อแต่ละสัญชาติ	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาคผนวก ข

การวัดประสิทธิภาพโดยใช้ Profiling

ภาคผนวก ข.1 ผลการทดสอบเพื่อเปรียบเทียบประสิทธิภาพระหว่างใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr

ข.1.1 ทดสอบการเชื่อมต่อฐานข้อมูลระหว่างฟังก์ชัน odbcConnect และฟังก์ชัน src_mysql

เปรียบเทียบฟังก์ชันที่ใช้ในการเชื่อมต่อฐานข้อมูล ดังนี้

1) แบบไม่ใช้แพ็กเกจ Dplyr ใช้ฟังก์ชัน odbcConnect ในแพ็กเกจ ODBC เป็นฟังก์ชันในการเชื่อมต่อฐานข้อมูล
2) แบบใช้แพ็กเกจ Dplyr ใช้ฟังก์ชัน src_mysql ในแพ็กเกจ Dplyr เป็นฟังก์ชันในการเชื่อมต่อฐานข้อมูล

ตาราง ข.1 เปรียบเทียบหน่วยความจำและเวลาที่ใช้โดยรวมระหว่างฟังก์ชัน src_mysql และฟังก์ชัน odbcConnect

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ในการประมวลผล (ms)
	obdc_db()	**0.7**	**90**
3	• my_db1 <- odbc_db()	0.7	90
	srcmysql_db()	**1.7**	**30**
9	• My_db2 <- srcmysql_db()	1.7	30

จากตาราง ข.1 พบว่า ฟังก์ชัน odbcConnect ใช้หน่วยความจำน้อยกว่า แต่ฟังก์ชัน src_mysql ใช้เวลาในการประมวลผลน้อยกว่า

โดยในส่วนของหน่วยความจำที่ใช้ระหว่างฟังก์ชัน odbcConnect และฟังก์ชัน src_mysql สรุปได้ว่า ฟังก์ชัน odbcConnect มีการใช้หน่วยความจำน้อยกว่าฟังก์ชัน src_mysql โดยคิดเป็นอัตราส่วน 1 : 2.43 แสดงให้เห็นว่าฟังก์ชัน odbcConnect ใช้เวลาน้อยกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น และไม่อนุญาตให้นำไฟล์นี้ไปเผยแพร่ต่อในทุกรูปแบบ
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ในส่วนของเวลาที่ใช้ในการประมวลผล การใช้ฟังก์ชัน src_mysql มีการใช้เวลาที่น้อยกว่าฟังก์ชัน odbcConnect คิดเป็นอัตราส่วน 1 : 3 ซึ่งแสดงให้เห็นว่าใช้เวลาน้อยกว่าถึง 3 เท่า

ตาราง ข.2 สัดส่วนการใช้หน่วยความจำและเวลาในแต่ละบรรทัดของฟังก์ชัน odbc_db()

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
4	db1 <- odbcConnect("coronavirus19",uid = "username",pwd = "****")	0.1	12.5	10	12.5
5	raw.confirmed <- sqlQuery(db1, "SELECT * FROM covid19_confirmed_global")	0.2	25	20	25
6	raw.deaths <- sqlQuery(db1, "SELECT * FROM covid19_deaths_global")	0.3	37.5	30	37.5
รวม		0.8	100	80	100

ตาราง ข.3 สัดส่วนการใช้หน่วยความจำและเวลาในแต่ละบรรทัดของฟังก์ชัน srcmysql_db()

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้(%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้(%)
11	df_conf <- tbl(db2, sql("select * from covid19_confirmed_global"))	1.5	~48.39	10	~33.33
13	df_deaths <- tbl(db2, sql("select * from covid19_deaths_global"))	0.8	~25.81	10	~33.33
15	df_recover <- tbl(db2, sql("select * from covid19_recovered_global"))	0.8	~25.81	10	~33.33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.3 สัดส่วนการใช้หน่วยความจำและเวลาในแต่ละบรรทัดของฟังก์ชัน srcmysql_db() (ต่อ)

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้(%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้(%)
รวม		3.1	100	30	100

จากตาราง ข.3 จะเห็นได้ว่ามีการใช้เวลาในการเชื่อมต่อฐานข้อมูลโดยใช้ฟังก์ชัน odbcConnect ประมาณ 10 มิลลิวินาที ซึ่งคิดเป็น 12.5% ของระยะเวลาที่ใช้ทั้งหมดในการเชื่อมต่อแบบ odbc ในขณะที่พิจารณาตาราง 4.4 จะพบว่าการเชื่อมต่อโดยใช้ฟังก์ชัน src_mysql() นั้นใช้เวลาในการเชื่อมต่อน้อยมาก

ข.1.2 ทดสอบฟังก์ชัน cleandata()

การทดสอบฟังก์ชัน cleandata() เป็นการทดสอบฟังก์ชันที่ใช้ในกระบวนการ Data Cleaning หรือการทำให้ข้อมูลอยู่ในรูปแบบที่พร้อมนำไปใช้ต่อ โดยใช้ข้อมูลจำนวน 3 ชุด ได้แก่ ชุดข้อมูลผู้ติดเชื้อไวรัสโคโรนา ชุดข้อมูลผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา และชุดข้อมูลผู้ที่หายจากโรคติดเชื้อไวรัสโคโรนา โดยมีฟังก์ชันที่ทดสอบ ดังนี้

1) แบบไม่ใช้แพ็กเกจ Dplyr ฟังก์ชันที่ใช้ คือ nd_cleandata() เป็นฟังก์ชันในการclean data แบบไม่ใช้แพ็กเกจ Dplyr
2) แบบใช้แพ็กเกจ dplyr ฟังก์ชันที่ใช้ คือ d_cleandata() เป็นฟังก์ชันในการ clean data แบบไม่ใช้แพ็กเกจ Dplyr

ตาราง ข.4 เปรียบเทียบหน่วยความจำและเวลาที่ใช้โดยรวมของฟังก์ชัน cleandata()

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้(MB)	เวลาที่ใช้ในการประมวลผล (ms)
	nd_cleandata()	**71**	**970**
2	dataConfirmed1 <- raw.confirmed %>% nd_cleanData()%>% rename(c("count"="confirmed"))	24.0	350

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.4 เปรียบเทียบหน่วยความจำและเวลาที่ใช้โดยรวมของฟังก์ชัน cleandata()

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้(MB)	เวลาที่ใช้ในการประมวลผล (ms)
4	dataRecovered1 <- raw.recovered %>% nd_cleanData() %>% rename(c("count"="recovered"))	23.7	310
	d_cleandata()	**46**	**700**
2	dataConfirmed2 <- raw.confirmed %>% d_cleanData() %>% rename(confirmed=count)	15.5	190
3	dataDeaths2 <- raw.deaths %>% d_cleanData() %>% rename(deaths=count)	15.7	230
4	dataRecovered2 <- raw.recovered %>% d_cleanData() %>% rename(recovered=count)	14.8	280

จากตาราง ข.4 พบว่า การใช้แพ็กเกจ Dplyr ใช้หน่วยความจำน้อยกว่าและใช้เวลาในการประมวลผลน้อยกว่าในส่วนของหน่วยความจำที่ใช้ การใช้แพ็กเกจ Dplyr ใช้หน่วยความจำน้อยกว่าการไม่ใช้ โดยเทียบเป็นอัตราส่วนได้ 1 : 1.54 แสดงให้เห็นว่า การใช้แพ็กเกจ Dplyr ใช้หน่วยความจำน้อยกว่าประมาณ 1.54 เท่า

ในส่วนของเวลาที่ใช้ในการประมวลผล พบว่า การใช้แพ็กเกจ Dplyr มีการใช้เวลาที่น้อยกว่าการไม่ใช้ Dplyr โดยเทียบเป็นอัตราส่วนการใช้ต่อการไม่ใช้ได้เป็น 1 : 1.38 ซึ่งแสดงให้เห็นว่า การใช้แพ็กเกจ Dplyr ในการทำ Data Cleaning นั้นใช้เวลาน้อยกว่าแบบไม่ใช้แพ็กเกจ Dplyr ประมาณ 1.38 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.5 สัดส่วนการใช้หน่วยความจำและเวลาในแต่ละบรรทัดของฟังก์ชัน cleandata แบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้(%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
7	raw.confirmed <- melt(raw.confirmed,id = c("country")) %>% rename(c("variable" = "date"))%>% rename(c("value" = "count"))	1.7	~9.77	10	~3.26
8	raw.confirmed <- aggregate(. ~ country + date , data = raw.confirmed, sum, na.rm=TRUE)%>% as.data.frame()	13.9	~79.86	160	~51.61
9	raw.confirmed$date <- as.Date(raw.confirmed$date,tryFormats = c("%m/%d/%Y", "%Y-%m-%d"))	1.8	~10.34	140	~45.16
รวม		17.4	100	310	100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.6 สัดส่วนการใช้หน่วยความจำและเวลาที่ใช้ในแต่ละบรรทัดของฟังก์ชัน cleandata แบบใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้(%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
6	df_conf1 <- df_conf %<>% select(-c(Province.State, Lat, Long)) %>% rename(country = Country.Region)	0.1	~0.65	10	~4.76
8	df_conf %<>% gather(key=date, value=count, -country)	0.8	~5.19	10	~4.76
10	df_conf %<>% mutate(date = date %>% mdy())	4.6	~29.87	10	~4.76
12	df_conf %<>% group_by(country, date) %>% summarise(count=sum(count, na.rm=T)) %>% as.data.frame()	9.9	~64.29	190	~90.48
รวม		15.4	100	210	100

จากตาราง ข.5 จะเห็นได้ว่า แบบไม่ใช้แพ็กเกจ Dplyr นั้นมีการใช้เวลามากในการประมวลผลบรรทัดที่ 8 และ 9 ซึ่งเป็นฟังก์ชัน aggregate ที่ใช้ในการจัดกลุ่มข้อมูล (คล้ายการ group by) และฟังก์ชัน as.Date ที่ใช้ในการแปลงรูปแบบวันที่ตามลำดับ ซึ่งใช้เวลารวมประมาณ 200 ms

ในขณะที่หากพิจารณาตาราง ข.6 จะพบว่าแบบใช้แพ็กเกจ Dplyr การจัดกลุ่มข้อมูลโดยใช้ฟังก์ชัน group by (บรรทัดที่ 12) และการแปลงรูปแบบวันที่โดยใช้ฟังก์ชัน mdy (บรรทัดที่ 10) โดยรวมแล้วใช้เวลา 190 ms ซึ่งเร็วกว่าแบบไม่ใช้แพ็กเกจประมาณ 1.05 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ข.1.3 ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล

ทดสอบการการรวมตารางระหว่างข้อมูลผู้ติดเชื้อ ข้อมูลผู้เสียหาย ข้อมูลผู้รักษาหาย และการเพิ่มแถวข้อมูล ได้แก่ ข้อมูลรวมทั่วโลก ซึ่งเป็นการรวมข้อมูลผู้ติดเชื้อไวรัสโคโรน่า 2019 จากทุกประเทศทั่วโลกตั้งแต่เดือนเมษายนถึงเดือนสิงหาคม โดยมีฟังก์ชันที่ใช้ ดังนี้

1) แบบไม่ใช้แพ็กเกจ Dplyr ฟังก์ชันที่ใช้ คือ nd_dataTotal() เป็นฟังก์ชันในการคำนวณข้อมูลทั่วโลกแบบไม่ใช้แพ็กเกจ Dplyr
2) แบบใช้แพ็กเกจ Dplyr ฟังก์ชันที่ใช้ คือ d_dataTotal() เป็นฟังก์ชันในการคำนวณข้อมูลทั่วโลกแบบใช้แพ็กเกจ Dplyr

ตาราง ข.7 เปรียบเทียบหน่วยความจำและเวลาที่ใช้โดยรวมของฟังก์ชันการรวมตารางและการเพิ่มแถวข้อมูล

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ในการประมวลผล (ms)
	nd_dataTotal()	**90.6**	**1200**
2	• nd_dataTotal(dataConfirmed1, dataDeaths1, dataRecovered1)	90.6	1200
	d_dataTotal()	**61.1**	**920**
12	• d_dataTotal(dataConfirmed2, dataDeaths2, dataRecovered2)	61.1	920

จากตาราง ข.7 พบว่าการคำนวณข้อมูลทั่วโลกโดยใช้แพ็กเกจ Dplyr มีการใช้หน่วยความจำที่น้อยกว่าและใช้เวลาในการประมวลผลน้อยกว่า

ในด้านการใช้หน่วยความจำ เมื่อเทียบระหว่างแบบใช้แพ็กเกจ Dplyr และแบบไม่ใช้แพ็กเกจ เทียบเป็นอัตราส่วนได้ 1 : 1.48 ตามลำดับ คือ แบบใช้แพ็กเกจ Dplyr มีการใช้พื้นที่น้อยกว่าประมาณ 1.48 เท่า

ในส่วนของเวลาที่ใช้ในการผระมวลผล การใช้แพ็กเกจ Dplyr มีการใช้เวลาที่น้อยกว่าแบบไม่ใช้แพ็กเกจ คิดเป็นอัตราส่วนได้ 1 : 1.3 ซึ่งแสดงให้เห็นว่าการใช้แพ็กเกจ Dplyr ใช้เวลาในการประมวลผลน้อยกว่าประมาณ 1.3 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.8 สัดส่วนการใช้หน่วยความจำและเวลาในการประมวลผลแต่ละบรรทัดของฟังก์ชันการรวมตารางและการเพิ่มแถวข้อมูลแบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
3	data_total <- dataConfirmed1 %>% merge(dataDeaths1,all = T) %>% merge(dataRecovered1,all = T)	59.6	~67.27	1030	~78.03
6	dataWorld <- aggregate(. ~ country + date , data = data_total1, sum, na.rm=TRUE)	22.5	~25.40	160	~12.12
8	data_total <- sqldf("SELECT * FROM data_total group by country, date order by country")	6.5	~7.34	130	~9.85
รวม		88.6	100	1320	100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.9 สัดส่วนการใช้หน่วยความจำและเวลาในการรันแต่ละบรรทัดของฟังก์ชันการรวมตารางและการเพิ่มแถวข้อมูลแบบใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
13	data <- dataConfirmed2 %>% merge(dataDeaths2,all = T) %>% merge(dataRecovered2,all = T)	56.3	~93.21	810	~97.59
14-18	data.world <- data %>% group_by(date) %>% summarise(country='World', confirmed = sum(confirmed, na.rm=T), deaths = sum(deaths, na.rm=T), recovered = sum(recovered, na.rm=T))	1.7	~2.81	10	~1.20
19	data %<>% rbind(data.world)	2.4	~3.97	10	~1.20
รวม		60.4	100	830	100

จากตาราง ข.8 และ ข.9 สามารถสรุปได้ว่า ฟังก์ชันที่ใช้เวลาในการประมวลผลทั้งแบบใช้แพ็กเกจ Dplyr และไม่ใช้แพ็กเกจ Dplyr มีส่วนของฟังก์ชัน merge ที่ใช้ในการรวมตารางข้อมูลซึ่งใช้พื้นที่และเวลาสูงสุด โดยแบบไม่ใช้แพ็กเกจจะมีการใช้พื้นที่และเวลาในฟังก์ชัน aggregate รองลงมาจากฟังก์ชัน merge ซึ่งส่งผลให้แบบไม่ใช้แพ็กเกจ Dplyr มีใช้พื้นที่และเวลาโดยรวมมากกว่าแบบที่ใช้แพ็กเกจ Dplyr

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ข.1.4 ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว

ฟังก์ชันที่ใช้

1) แบบไม่ใช้แพ็กเกจ Dplyrฟังก์ชันที่ใช้ คือ nd_long() เป็นฟังก์ชันในการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว แบบไม่ใช้แพ็กเกจ Dplyr
2) แบบไม่ใช้แพ็กเกจ Dplyrฟังก์ชันที่ใช้ คือ nd_long() เป็นฟังก์ชันในการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาวแบบไม่ใช้แพ็กเกจ Dplyr

ตาราง ข.10 เปรียบเทียบหน่วยความจำและ เวลาที่ใช้โดยรวมในการแปลงตารางจากแบบกว้างเป็นแบบยาว

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ในการประมวลผล (ms)
	nd_long()	**10.9**	**50**
2	• data_rate_nd <- nd_long()	10.9	50
	d_long()	**47.3**	**70**
2	• data_rate_d <- d_long()	47.3	70

จากตาราง ข.10 พบว่า แบบใช้แพ็กเกจ Dplyr ใช้หน่วยความจำมากกว่า และใช้เวลาในการประมวลผลมากกว่าคิดเป็นอัตราส่วนได้ 4.34 : 1 ซึ่งแสดงให้เห็นว่า การใช้แพ็กเกจ Dplyr มีการใช้พื้นที่มากกว่า 4.34 เท่า

ในส่วนของเวลาที่ใช้ แบบใช้แพ็กเกจใช้เวลามากกว่าแบบไม่ใช้แพ็กเกจคิดเป็นอัตราส่วน 70 : 50 หรือ 1.4 : 1 จึงสรุปได้ว่าการคำนวณอัตราต่าง ๆ โดยใช้แพ็กเกจ Dplyr ใช้เวลามากกว่าแบบไม่ไม่ใช้แพ็กเกจประมาณ 1.5 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.11 สัดส่วนการใช้หน่วยความจำและเวลาในการประมวลผลแต่ละบรรทัดของฟังก์ชันการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาวแบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
14	rates.long <- melt(rates.long,id = c("country","date"))	11.7	100	10	100
รวม		11.7	100	10	100

ตาราง ข.12 สัดส่วนการใช้หน่วยความจำและเวลาในการรันแต่ละบรรทัดของฟังก์ชันการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาวแบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
4-7	data.long <- data %>% select(c(country, date, confirmed, current.confirmed, recovered, deaths)) %>% gather(key=type, value=count, -c(country, date))	2.4	~5.12	20	~28.57

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.12 สัดส่วนการใช้หน่วยความจำและเวลาในการรันแต่ละบรรทัดของฟังก์ชันการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาวแบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
8-12	data.long %<>% mutate(type=recode_factor(type, confirmed='Total Confirmed', current.confirmed='Current Confirmed', recovered='Recovered', deaths='Deaths'))	26.3	~56.08	20	~28.57
13-15	rates.long <- data %>% select(c(country, date, rate.upper, rate.lower, rate.daily)) %>% gather(key=type, value=count, -c(country, date))	3.4	~7.25	10	~14.28
16	rates.long %<>% mutate(type=recode_factor(type, rate.daily='Daily', rate.upper='Upper bound'))	14.8	~31.56	20	~28.57
รวม		46.9	100	70	100

จากตาราง ข.11 แสดงให้เห็นว่าฟังก์ชันต่าง ๆ ในแต่ละบรรทัดแทบไม่ใช้เวลาในการประมวลผลหรือใช้เวลาน้อยมาก จึงมีเพียงบางบรรทัดเท่านั้น คือ ฟังก์ชัน melt ที่ใช้เวลา 10 ms

ในส่วนของการใช้แพ็กเกจ Dplyr จากตาราง ข.12 สามารถสรุปได้ว่ามีการใช้เวลามากในฟังก์ชันหลายส่วน โดยเฉพาะในบรรทัดที่มีการส่งผ่านข้อมูลด้วย Pipeline Operation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ข.1.5 ทดสอบการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

ฟังก์ชันที่ใช้

1) แบบไม่ใช้แพกเกจ dplyr ฟังก์ชันที่ใช้ คือ nd_top20() เป็นฟังก์ชันในการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบไม่ใช้แพ็กเกจ Dplyr
2) แบบใช้แพกเกจ dplyr ฟังก์ชันที่ใช้ คือ d_top20() เป็นฟังก์ชันในการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบใช้แพ็กเกจ Dplyr

ตาราง ข.13 เปรียบเทียบหน่วยความจำและเวลาที่ใช้โดยรวมของฟังก์ชันการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

บรรทัดที่	ฟังก์ชันที่ใช้	หน่วยความจำที่ใช้ (MB)	เวลาที่ใช้ในการประมวลผล (ms)
	nd_top20()	**2.3**	**40**
3	• data1 <- data %>% nd_top20()	2.3	40
	d_top20()	**1.4**	**50**
2	• data2 <- data %>% d_top20()	1.4	50

จากตาราง ข.13 พบว่า การใช้แพ็กเกจ Dplyr ใช้หน่วยความจำที่น้อยกว่า แต่ใช้เวลาในการประมวลผลมากกว่าโดยในส่วนของหน่วยความจำ แบบใช้แพ็กเกจ Dplyr มีการใช้หน่วยความจำน้อยกว่า โดยคิดเป็นอัตราส่วน 1 : 1.64 แสดงให้เห็นว่าแบบใช้แพ็กเกจ ใช้หน่วยความจำน้อยกว่าประมาณ 1.64 เท่า ในส่วนของเวลาที่ใช้ในการประมวลผลแบบใช้แพ็กเกจ Dplyr มีการใช้เวลาที่มากกว่า คิดเป็นอัตราส่วนได้ 1.25 : 1 ซึ่งแสดงให้เห็นว่าใช้เวลามากกว่า 1.25 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.14 สัดส่วนการใช้หน่วยความจำและเวลาในการประมวลผลแต่ละบรรทัดของฟังก์ชันการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบไม่ใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
20-24	data.latest <- ddply(data.latest, .(country, date), summarise, confirmed = sum(confirmed), new.confirmed = sum(new.confirmed), current.confirmed = sum(current.confirmed), recovered = sum(recovered), deaths = sum(deaths), new.deaths = sum(new.deaths)) %>% mutate(death.rate=(100 * deaths/confirmed) %>% round(1))	1.9	100	10	100
รวม		1.9	100	10	100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.15 สัดส่วนการใช้หน่วยความจำและเวลาในการรันแต่ละบรรทัดของฟังก์ชันการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบใช้แพ็กเกจ Dplyr

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
4-7	data.latest.all <- data %>% filter(date == max(date)) %>% select(country, date,confirmed, new.confirmed, current.confirmed, recovered, deaths, new.deaths, death.rate=rate.lower) %>% mutate(ranking = dense_rank(desc(confirmed)))	0.8	80	10	~33.33
18-22	data.latest %<>% group_by(country) %>% summarise(confirmed=sum(confir med), new.confirmed=sum(new.confirme d), current.confirmed=sum(current.con firmed), recovered=sum(recovered), deaths=sum(deaths), new.deaths=sum(new.deaths)) %>% mutate(death.rate=(100 * deaths/confirmed) %>% round(1))	0.1	10	10	~33.33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.15 สัดส่วนการใช้หน่วยความจำและเวลาในการรันแต่ละบรรทัดของฟังก์ชันการดึงข้อมูล 20 อันดับประเทศที่มีการติดเชื้อสูงสุดแบบใช้แพ็กเกจ Dplyr (ต่อ)

บรรทัดที่	โค้ดที่ใช้ประมวลผล	หน่วยความจำที่ใช้ (MB)	สัดส่วนหน่วยความจำที่ใช้ (%)	เวลาที่ใช้ (ms)	สัดส่วนเวลาที่ใช้ (%)
23-24	data.latest %<>% select(c(country, confirmed, deaths, death.rate, new.confirmed, new.deaths, current.confirmed))	0.1	10	10	~33.33
รวม		1	100	30	100

จากตาราง ข.14 แสดงให้เห็นว่าโค้ดที่ไม่ได้ใช้แพ็กเกจ Dplyr นั้น มีฟังก์ชันที่ใช้หน่วยความจำและเวลาเพียงฟังก์ชันเดียว คือ ฟังก์ชัน melt ที่ใช้เวลารวมทั้งสิ้น 10 ms ในขณะที่พิจารณาตาราง ข.15 พบว่าแบบใช้แพ็กเกจ Dplyr มีการใช้พื้นที่และเวลาในหลายฟังก์ชัน โดยเฉพาะในบรรทัดที่มีการส่งผ่านข้อมูลด้วย Pipeline Operation

ภาคผนวก ข.2 การวัดประสิทธิภาพของแพ็กเกจ Dplyr เมื่อมีปัจจัยอื่นเข้ามาเกี่ยวข้อง

ข.2.1 ทดสอบการเชื่อมต่อฐานข้อมูล

ฟังก์ชันที่ใช้ในการทดสอบ

1) datafrommysql() ฟังก์ชันเชื่อมต่อฐานข้อมูลผ่าน Local database เพื่อดึงชุดข้อมูลขนาดเล็ก
2) datafromserver() ฟังก์ชันเชื่อมต่อฐานข้อมูลผ่านระบบคลาวด์เพื่อดึงชุดข้อมูลขนาดเล็ก
3) datafrommysql.big() ฟังก์ชันเชื่อมต่อฐานข้อมูลผ่าน Local database เพื่อดึงชุดข้อมูลขนาดใหญ่
4) datafromserver.big() ฟังก์ชันเชื่อมต่อฐานข้อมูลผ่านระบบคลาวด์เพื่อดึงชุดข้อมูลขนาดใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.16 ผลการทดสอบการเชื่อมต่อฐานข้อมูลภายใต้ปัจจัยต่าง ๆ

รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
	Environment 1	Environment 2	Environment 1	Environment 2
datafrommysql()	2.72	2.56	48	38
datafrommysql.big()	5.88	6.12	58	30
datafromserver()	3.34	3.08	642	786
datafromserver.big()	6.5	6.16	768	762

ผลจากการทดสอบในส่วนของหน่วยความจำ เมื่อพิจารณาปัจจัยด้าน Environment พบว่าใน Environment 1 กับ Environment 2 โดยเฉลี่ยแล้วมีการใช้หน่วยความจำใกล้เคียงกัน โดยการเชื่อมต่อฐานข้อมูลแบบ Local database ใช้หน่วยความจำน้อยกว่าประมาณ 1.1 เท่าเมื่อเทียบกับการเชื่อมต่อฐานข้อมูลบนคลาวด์ ซึ่งหมายความว่าการเชื่อมต่อแทบไม่ส่งผลต่อการพื้นที่หน่วยความจำ และในส่วนปัจจัยด้านขนาดของข้อมูล พบว่าข้อมูลขนาดใหญ่มีการใช้พื้นที่มากกว่าข้อมูลขนาดเล็กประมาณ 2.05 เท่า

ข.2.2 ทดสอบฟังก์ชัน cleandata()

ใช้ฟังก์ชัน Cleandata() เป็นฟังก์ชันสำหรับทดสอบการจัดการข้อมูลของ Dplyr โดยนำไปทดสอบกับข้อมูลต่างขนาด และเชื่อมต่อฐานข้อมูลจากต่างแหล่ง โดยมีกลุ่มข้อมูลทดสอบ ดังนี้

1) กลุ่มข้อมูลขนาดเล็กจากฐานข้อมูล Local database ได้แก่ small_confirmed, small_deaths และ small_recover
2) กลุ่มข้อมูลขนาดใหญ่จากฐานข้อมูล Local database ได้แก่ big_confirmed, big_deaths และ big_recover
3) กลุ่มข้อมูลขนาดเล็กจากฐานข้อมูลคลาวด์ ได้แก่ small_confirmed.server, small_deaths.server และ small_recover.server
4) กลุ่มข้อมูลขนาดใหญ่จากฐานข้อมูลคลาวด์ ได้แก่ big_confirmed.server, big_deaths.server และ big_recover.server

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง ข.17 ผลการทดสอบฟังก์ชัน Clean data ภายใต้ปัจจัยต่าง ๆ

รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
	Environment 1	Environment 2	Environment 1	Environment 2
small_confirmed, small_deaths small_recover	44.1	34.86	446	384
big_confirmed big_deaths big_recover	94.46	91.2	1,014	966
small_confirmed.server small_deaths.server small_recover.server	40.84	33.8	480	348
big_confirmed.server big_deaths.server big_recover.server	99.96	86.6	1,020	932

จากตาราง ข.17 พบว่าการใช้พื้นที่หน่วยความจำใน Environment 1 และ Environment 2 มีความใกล้เคียงกันทั้งในด้านการเชื่อมต่อฐานข้อมูลและด้านขนาดของข้อมูล หากสังเกตในส่วนของการเชื่อมต่อฐานข้อมูลพบว่าข้อมูลขนาดเดียวกันแต่ใช้การเชื่อมต่อฐานข้อมูลต่างกัน จะมีการใช้หน่วยความจำที่ใกล้เคียงกัน ส่วนในด้านขนาดของข้อมูลพบว่าการเชื่อมต่อแบบเดียวกันแต่ทดสอบด้วยข้อมูลที่มีขนาดแตกต่างกัน พบว่าข้อมูลที่มีขนาดใหญ่จะใช้พื้นที่หน่วยความจำมากกว่าข้อมูลขนาดเล็กอยู่ที่ 2.43 เท่า ขนาดของข้อมูลส่งผลต่อการใช้หน่วยความจำ และเมื่อพิจารณาเวลาที่ใช้ในการประมวลผลพบว่าทุก Environment และทุกการเชื่อมต่อใช้เวลาในการประมวลผลที่ใกล้เคียงกัน แต่ข้อมูลที่มีขนาดใหญ่จะใช้เวลาในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ประมวลผลที่มากกว่าประมาณ 2.3 เท่า ซึ่งแสดงให้เห็นว่าในส่วนของฟังก์ชัน Cleandata ขนาดของข้อมูลส่งผลต่อการใช้หน่วยความจำและการประมวลผล

ข.2.3 ทดสอบการรวมตารางและการเพิ่มแถวข้อมูล

ใช้ฟังก์ชัน MergeData() ในการวัดประสิทธิภาพของแพ็กเกจ Dplyr ในการรวมชุดข้อมูล โดยมีรายละเอียดชุดข้อมูล ดังนี้

1) small_merge เป็นตัวแปรแทนการรวมชุดข้อมูลขนาดเล็กที่ดึงมาจากฐานข้อมูลแบบ Local database
2) big_merge เป็นตัวแปรแทนการรวมชุดข้อมูลขนาดใหญ่ที่ดึงมาจากฐานข้อมูลแบบ Local database
3) small_merge.server เป็นตัวแปรแทนการรวมชุดข้อมูลขนาดเล็กที่ดึงมาจากฐานข้อมูลบนคลาวด์
4) big_merge.server เป็นตัวแปรแทนการรวมชุดข้อมูลขนาดใหญ่ที่ดึงมาจากฐานข้อมูลแบบคลาวด์

ตาราง ข.18 ผลการทดสอบการรวมตารางและการเพิ่มแถวข้อมูลภายใต้ปัจจัยต่าง ๆ

รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
	Environment 1	Environment 2	Environment 1	Environment 2
small_merge	60.2	56.44	378	650
big_merge	163.18	150.16	1,080	1,174
small_merge.server	59.46	62.54	382	716
big_merge.server	159.52	141.26	1,116	1,800

จากตาราง ข.18 ในส่วนของหน่วยความจำมีการใช้พื้นที่ที่ใกล้เคียงกันในทั้ง 2 Environment เช่นเดียวกับการเชื่อมต่อฐานข้อมูลที่มีการใช้พื้นที่หน่วยความจำใกล้เคียงกัน แต่ในส่วนของขนาดข้อมูลพบว่าข้อมูลขนาดใหญ่ใช้พื้นที่ในการประมวลผลมากกว่าข้อมูลขนาดเล็ก โดยแตกต่างกันอยู่ประมาณ 2.43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เท่า และในส่วนเวลาที่ใช้ในการประมวลผลพบว่าใน Environment 1 จะใช้เวลาในการประมวลผลน้อยกว่าโดยเฉลี่ยอยู่ที่ประมาณ 1.47 เท่า ส่วนรูปแบบการเชื่อมต่อพบว่าทุกรูปแบบการเชื่อมต่อพบว่าใช้เวลาประมวลผลใกล้เคียงกัน และเมื่อพิจารณาปัจจัยด้านขนาดข้อมูล พบว่าข้อมูลขนาดใหญ่ใช้เวลาประมวลผลมากกว่าประมาณ 2.39 เท่า จึงสรุปได้ว่าปัจจัยด้าน Environment ส่งผลต่อเวลาในการประมวลผลและขนาดของข้อมูลส่งผลต่อหน่วยความจำและเวลาที่ใช้ในการประมวลผลในการรวมตารางและเพิ่มแถวข้อมูล

ข.2.4 ทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาว

ใช้ฟังก์ชัน ChangeDataFormat() เพื่อวัดประสิทธิภาพของแพ็กเกจ Dplyr ในการเปลี่ยนรูปแบบของข้อมูลจากแบบกว้าง (Wide format) เป็นแบบยาว (Long format) โดยกำหนดตัวแปรแทนข้อมูล ดังนี้

1) small_dataFormat เป็นตัวแปรแทนข้อมูลชุดเล็กที่ดึงมาจากฐานข้อมูลแบบ Local Database
2) big_dataFormat เป็นตัวแปรแทนข้อมูลชุดใหญ่ที่ดึงมาจากฐานข้อมูลแบบ Local Database
3) small_dataFormat.server เป็นตัวแปรแทนข้อมูลชุดเล็กที่ดึงมาจากฐานข้อมูลบนคลาวด์
4) big_dataFormat.server เป็นตัวแปรแทนข้อมูลชุดใหญ่ที่ดึงมาจากฐานข้อมูลบนคลาวด์

ตาราง ข.19 ผลการทดสอบการแปลงตารางจากแบบกว้างเป็นแบบยาวภายใต้ปัจจัยต่าง ๆ

รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
	Environment 1	Environment 2	Environment 1	Environment 2
small_dataFormat	27.26	28.96	38	44
big_dataFormat	67.24	67.04	132	110
small_dataFormat.server	29.48	28.46	50	48
big_dataFormat.server	65.28	64.46	106	130

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากตาราง ข.19 สรุปได้ว่าหน่วยความจำที่ใช้และเวลาที่ใช้ในการประมวลผลค่อนข้างใกล้เคียงกันในทุก Environment และทุกรูปแบบการเชื่อมต่อฐานข้อมูล โดย แต่หากพิจารณาด้านขนาดของข้อมูลพบว่าข้อมูลขนาดใหญ่ใช้พื้นที่หน่วยความจำและเวลาในการประมวลผลที่มากกว่า

ข.2.5 ทดสอบการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

ใช้ฟังก์ชัน CalcTop20() เพื่อวัดประสิทธิภาพในการคัดกรองและเรียงลำดับข้อมูล

1) small_top20 เป็นตัวแปรแทนข้อมูลชุดเล็กที่ดึงมาจากฐานข้อมูลแบบ Local database
2) big_top20 เป็นตัวแปรแทนข้อมูลชุดใหญ่ที่ดึงมาจากฐานข้อมูลแบบ Local database
3) small_top20.server เป็นตัวแปรแทนข้อมูลชุดเล็กที่ดึงมาจากฐานข้อมูลบนคลาวด์
4) big_top20.server เป็นตัวแปรแทนข้อมูลชุดใหญ่ที่ดึงมาจากฐานข้อมูลบนคลาวด์

ตาราง ข.20 ผลการทดสอบการคำนวณ 20 อันดับประเทศที่มีการติดเชื้อสูงสุด

รายการการทดสอบ	หน่วยความจำที่ใช้ (MB)		เวลาที่ใช้ในการประมวลผล (ms)	
	Environment 1	Environment 2	Environment 1	Environment 2
small_top20	1.4	0.88	40	38
big_top20	1.94	1.74	20	32
small_top20.server	1.04	2.18	24	32
big_top20.server	2.02	1.88	26	34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากตาราง ข.20 พบว่าในทุก Envronment ทุกรูปแบบการเชื่อมต่อ และทุกขนาดข้อมูลมีหน่วยความจำที่ใช้และเวลาที่ใช้ในการประมวลผลที่ใกล้เคียงกัน จึงสามารถสรุปได้ว่าความแตกต่างด้าน Environment รูปแบบการเชื่อมต่อฐานข้อมูล และขนาดของข้อมูลส่งผลน้อยหรือแทบไม่ส่งผลต่อหน่วยความจำที่ใช้และเวลาที่ใช้ในการประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.