

Homework 4 : Data Visualization

Submission: on Edufarm (Deadline: Tuesday, 3 April 2018, 13:00)

- 1) ไฟล์ MSWord หรือ Google Doc รวบรวมรายการของ html files และคำตอบในแต่ละข้อ (ชื่อไฟล์: studentID1_studentID2_answer)
- 2) 1 compressed file ของ html files (ถ้ามี) (ชื่อไฟล์: studentID1_studentID2_html)
 - การบันทึกผลลัพธ์ใน R/Pandas ให้เป็น html สามารถดูได้. ในหน้าที่ 3
 - การตั้งชื่อไฟล์ html สามารถใช้ชื่อข้อ เช่น problem2.html ว่าเป็นไฟล์ของข้อใด
- 3) 1 compressed file ของไฟล์ทุกไฟล์ที่ใช้ในการสร้างกราฟ เช่น Excel/Google Sheet/R/Pandas (ถ้ามี) (ชื่อไฟล์: studentID1_studentID2_graph)

จุดประสงค์ของการบ้าน

- นิสิตสามารถเลือกและสร้างกราฟได้ตรงตามจุดประสงค์ต่างๆ
- นิสิตสามารถใช้กราฟสำหรับวิเคราะห์ข้อมูลได้

ข้อกำหนดเบื้องต้น

- การบ้านนี้ เป็นการบ้านสำหรับนิสิต 2 คนทำส่งด้วยกัน หลังจากได้ผลลัพธ์เบื้องต้นของแต่ละกราฟ ขอให้ นิสิตในกลุ่ม วิเคราะห์และปรับปรุงกราฟของกลุ่มให้เหมาะสมมากขึ้น ก่อนส่ง
- ในการสร้างกราฟ พยายามเลือกใช้ interactive graph ให้มากที่สุด (เลือกใช้ Plotly ใน R/Pandas หรือ Google Chart) จากนั้นใส่ graph components เช่น title, แกน X, แกน Y เป็นต้น ปรับสีในกราฟให้เหมาะสม
- ในการส่งงาน สร้างตารางใน MSWord หรือ Google Doc สำหรับใส่ชื่อไฟล์หรือลิงค์ของกราฟ และคำตอบ ดังตัวอย่างด้านล่าง

ข้อ	html filename หรือ jsfiddle URL ของกราฟในข้อนี้	คำตอบ
1.
2.
...

คำถาม

ใช้ข้อมูลจากไฟล์ revenue_50_58.csv และ econ2558.csv สร้างกราฟตามจุดประสงค์ในแต่ละข้อให้เหมาะสม และตอบคำถามจากกราฟ (สามารถดูคำอธิบายข้อมูลได้ในตาราง หน้า 4) ดังนี้

1. แสดงแนวโน้มรายได้ของประชากรในภูมิภาคต่างๆ ระหว่างปี 2550-2558
 - คำถาม:
 - จากกราฟ ภาคใด มีแนวโน้มรายได้ลดลง
2. แสดงความแตกต่างของรายได้ของประชากรในภูมิภาคต่างๆ ไม่รวม ภูมิภาคกรุงเทพฯและปริมณฑล (Bangkok and suburb cities) กับรายได้เฉลี่ยของทั้งประเทศ ระหว่างปี 2550-2558
 - คำถาม:
 - 2.1 จากกราฟ ภาคใดมีรายได้มากที่สุด เมื่อเทียบกับรายได้เฉลี่ยของทั้งประเทศจากปีพ.ศ.เดียวกัน และเกิดขึ้นในปีพ.ศ.ใด
 - 2.2 จากกราฟ ภาคใดมีรายได้น้อยที่สุด เมื่อเทียบรายได้เฉลี่ยของทั้งประเทศจากปีพ.ศ.เดียวกัน และเกิดขึ้นในปีพ.ศ.ใด
3. แสดงสัดส่วนของจำนวนรถยนต์ของแต่ละภูมิภาค ในปี 2558 ในข้อนี้ ให้รวมภูมิภาคกรุงเทพฯและปริมณฑล
 - คำถาม:
 - 3.1 จากกราฟ ภูมิภาคใด มีสัดส่วนรถยนต์มากเป็นอันดับ 2 และคิดเป็นกี่เปอร์เซ็นต์
 - 3.2 จากกราฟ ภูมิภาคใด มีสัดส่วนรถยนต์เป็นอันดับสุดท้ายและคิดเป็นกี่เปอร์เซ็นต์
4. แสดงการกระจายตัวของรถยนต์ในแต่ละภูมิภาค ในปี 2558 ไม่รวม ภูมิภาคกรุงเทพฯและปริมณฑล
 - คำถาม:
 - จากกราฟ ภูมิภาคใด มีการกระจายตัวของรถยนต์น้อยที่สุด เพราะเหตุใด
5. แสดงปริมาณรถยนต์ของแต่ละจังหวัดใน 1 ภูมิภาค ในปี 2558
 - ให้เลือกภูมิดังนี้ กำหนดให้ X = รหัสนิติในกลุ่มที่มีค่าน้อยกว่า , Y = รหัสนิติในกลุ่มที่มีค่ามากกว่า

X เป็นเลขคี่, Y เป็นเลขคี่	ภาคเหนือ
X เป็นเลขคี่, Y เป็นเลขคู่	ภาคตะวันออกเฉียงเหนือ
X เป็นเลขคู่, Y เป็นเลขคี่	ภาคกลาง (ไม่รวมจังหวัดในภูมิภาคกรุงเทพฯและปริมณฑล)
X เป็นเลขคู่, Y เป็นเลขคู่	ภาคใต้
 - คำถาม:
 - จากกราฟ อธิบายปริมาณรถยนต์ในภูมิภาคที่เลือกนี้

6. จากปี 2558 แสดงความสัมพันธ์ระหว่างรายได้และจำนวนรถยนต์ของแต่ละจังหวัด, ระหว่างรายจ่ายและจำนวนรถยนต์ของแต่ละจังหวัด, ระหว่างหนี้สินและจำนวนรถยนต์ของแต่ละจังหวัด

- คำถาม:

- 6.1 จากกราฟ จังหวัดที่มีรายได้มาก จะมีจำนวนรถยนต์ที่มากขึ้นตามไปด้วยหรือไม่
- 6.2 จากกราฟ จังหวัดที่มีรายจ่ายมาก จะมีจำนวนรถยนต์ที่มากขึ้นตามไปด้วยหรือไม่
- 6.3 จากกราฟ จังหวัดที่มีหนี้สินมาก จะมีจำนวนรถยนต์ที่มากขึ้นตามไปด้วยหรือไม่

การบันทึกผลลัพธ์ให้เป็น html file

R

- ติดตั้ง `htmlwidgets` package
- ในโปรแกรม R เรียก `htmlwidgets` library และหลังจากสร้างกราฟเรียบร้อยแล้ว ให้เพิ่มส่วนบันทึกไฟล์ ดังตัวอย่างต่อไปนี้

```
library(plotly)
library(htmlwidgets)

...

p <- plot_ly( ... )

htmlwidgets::saveWidget(widget=p, "problem1.html", selfcontained = FALSE)
```

Pandas

ในคำสั่ง `plot` เติมพารามิเตอร์ เป็นชื่อไฟล์ที่ต้องการบันทึกเป็น html

```
import plotly as ply

...

ply.offline.plot(fig, filename="problem2.html")
```

คำอธิบายของข้อมูล

revenue_50_58.csv		econ2558.csv	
ชื่อ Column	ข้อมูลใน Column นั้นๆ	ชื่อ Column	ข้อมูลใน Column นั้นๆ
Province	จังหวัด	Province	จังหวัด
Region	ภูมิภาค	Region	ภูมิภาค
2550, 2552, ..., 2558	รายได้เฉลี่ยของประชากรในแต่ละจังหวัดในปีพ.ศ.นั้น (หน่วย = บาท)	Revenue	รายได้เฉลี่ยของประชากรในแต่ละจังหวัด (หน่วย = บาท)
		Expense	ค่าใช้จ่ายเฉลี่ยของประชากรในแต่ละจังหวัด (หน่วย = บาท)
		Debt	หนี้สินเฉลี่ยของประชากรในแต่ละจังหวัด (หน่วย = บาท)
		NumVehicles	จำนวนรถยนต์ที่จดทะเบียนในจังหวัดนั้น (หน่วย = คัน)

ที่มาของข้อมูล

สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

กรมการขนส่งทางบก กระทรวงคมนาคม