IM5211701 Big Data Analytics and Application Take-Home Exam

Instructor: Chao-Lung Yang

This exam is an individual take-home exam. Each student needs to work solely to finish the R coding and write a short report to describe your result. Once again, you are not supposed to work together on this take-home exam. However, you can ask other people or search some solutions on the internet, especially for R coding. If somebody including internet resource help you work on this exam, please specify them on report to show your courtesy (list all your reference). Without doing this, I will consider it is cheating.

Two files should be submitted through Blackboard system. They are:

1. R code (please write down the question number before your code, and write sown the comments to avoid the confusing)
2. MS Word report (10-page limit, font size 12, single line)

Deadline: 11/18 13:30 PM

Data Description

You can download two files from Blackboard system (see attachment). You have to use R to read the files, plot the data, and answer some questions.

flowdata.csv

flowdata.csv contains the data about traffic flow of a certain road in Taipei city. The csv file has 5 fields. They are described in the following table.

Field name	Meaning	Unit
Datetime	Date Time	
weekday	Weekday of the day	1: Monday, 2: Tuesday, 3: Wednesday, 4: Thursday, 5: Friday, 6: Saturday, 7: Sunday
avg_traffic_flow	Average car traffic flow rate	Kilometer per hour, -1 means no data
avg_small_flow	Average car traffic flow rate for small sedan	~~Kilometer per hour, -1~~ 0 means no data
avg_big_flow	Average car traffic flow rate for bigger car	~~Kilometer per hour, -1~~ 0 means no data

raindata.csv

raindata.csv contains the rainfall data associated with the road specified in the flowdata. The csv file has 3 fields. They are described in the following table.

Field name	Meaning	Unit
Datetime	Date Time	
Weekday	Weekday of the day	1: Monday, 2: Tuesday, 3: Wednesday, 4: Thursday, 5: Friday, 6: Saturday, 7: Sunday
Rainfall	Average rain per hour	Millimeter 0: no rain at that hour

Questions:

1. (15%) Please plot the **avg_traffic_flow** data by R to compare traffic flow by weekday, and hour. The x-axis of the plot should be 'hour' from 1am to 24pm, and y-axis should be avg_traffic_flow (kilometer/hour). There should be seven curves in the plot. Each curve identifies the avg_traffic_flow of a particular weekday such as Monday. So, we can compare the traffic flow by different weekday. **Please note that you have to specify how you calculate the average avg_traffic_flow for a particular hour of a certain weekday in the report. Without addressing this in the report will cost you 50% of the grade of this question!!**

2. (15%) Can you generate the similar plot mentioned in the question 1 to compare the traffic flow of **small sedan**, and **bigger car**? Do you see any interesting result?

3. (20%) Do you find any pattern related to traffic data? For example, which weekday or which hour has busiest traffic and smooth traffic. If there are two peak rush hour on that road such as 7am ~9am (morning) and 5pm~7pm (afternoon), can you find any pattern on the traffic data? **Please specify how you calculate the average avg_traffic_flow for peak rush hour in the report.**

4. (20%) Similarly, please plot the average rainfall data by R to compare rainfall by weekday, and hour. The x-axis of the plot should be 'hour' from 1am to 24pm, and y-axis should be rainfall (mm). There should be seven curves in the plot. Each curve identifies the average rainfall of a particular weekday such as Monday. So, we can compare the rainfall by different weekday. **Please note that you have to**

specify how do you calculate the average rainfall for a particular hour by a certain weekday. Without addressing this in the report will cost you 50% of the grade of this question!!

Hint: you can see the rainfall is zero for a non-raining day. So, the average rainfall should only count the hour when it rains!

5. (30%) By compare the traffic data and rainfall data, can you find the correlation between the traffic and rainfall data? If yes, please show your evidence to specify how the rainfall is correlated to traffic. If no, please address your reason. Write down your comments and conclude your finding in the report.