IE6600: Computaton and Data Visualisation
Homework 4

Prof. Mohammad Dehghani

2/19/2021

Assignment Guidelines

1. Students need to complete the assignment in **groups**.
2. All the assignments are required to be done in RStudio.
3. Provide necessary comments using '#' for better understanding of your script.
4. The code should **follow tidyverse style guide** (https://style.tidyverse.org/index.html) The tidyverse style guide has style standards for naming objects, indentation and how to write long lines of codes to name a few
5. **If you take help from any external sources, please mention that in the reference. Violating academic integrity policies may include zero credit on the work.**
6. The assignment report needs to include the following sections:

- **Problem statement:** A brief about your understanding on the assignment questions (maximum 3 lines)

- **Result:** What were your finding after creating the code and running it in R. This section may include:

  - Graphs / charts / plots
  - Final data frame for your result
  - Results obtained

- **Conclusion:** What were the statistical inferences and observations from the results obtained.

7. Assignment report has to be included in the .rmd file.

Deliverables:

1. **Please submit a **.rmd* file which includes your code and can be knit into a PDF(recommended) ****

   i. ** Please contact the TAs in office hours to learn how to knit the files into PDF

2. The above mentioned file has to be labeled as: 'HW # - IE 6600 – Sec # - <Student Name>'
3. Submit your HW deliverable via CANVAS
4. **Deadline:** 25th **February 23:59**

All visualizations have to be complete in terms of legends, titles, labels etc. Points will be deducted for incomplete submissions.

Visualizations included in the pdf are only for your purpose and not the final versions. You are expected to

Task 1

Generate the density plot as shown in the figure below.

- Hint: Create a function to transform the x axis using a log10 to better visualize the data spread.
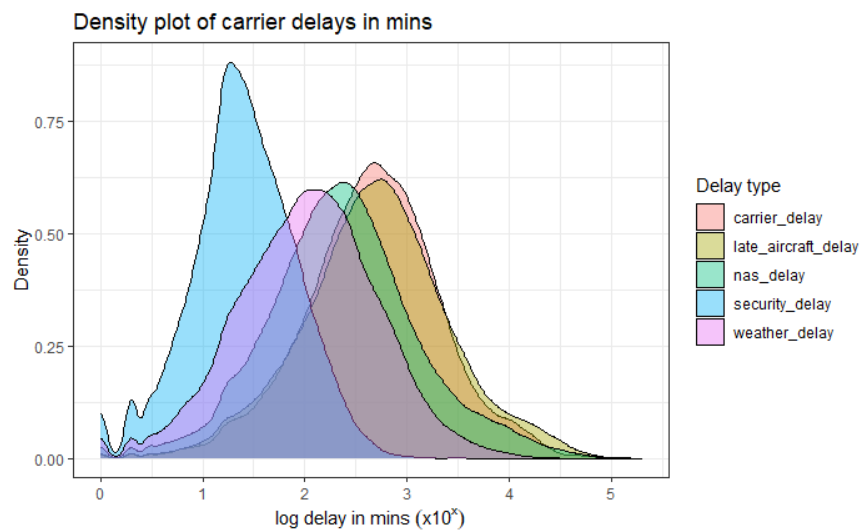
Task 2

Generate correlation plots for arr_flights, arr_del15, arr_cancelled, arr_diverted, arr_delay, carrier_delay, weather_delay, nas_delay, security_delay and late_aircraft_delay. The below image is just for your reference, you are expected to create a plot with labels properly aligned and not overlapping.

- Hint: Use corrplot.mixed() to complete this task
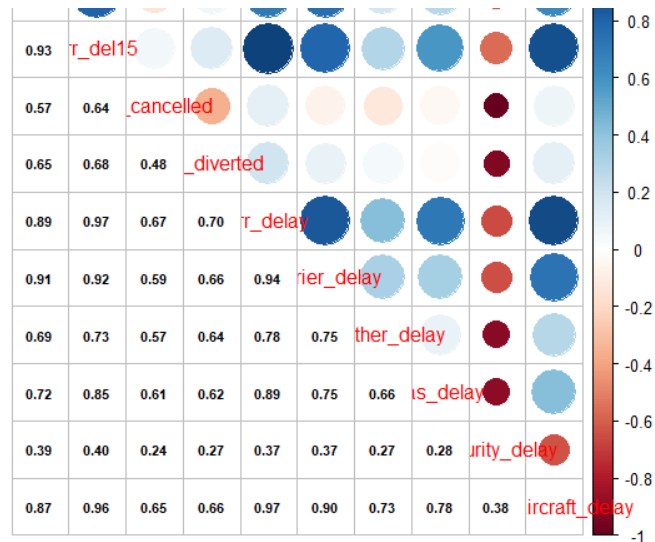
Task 3

From wages_jobs.csv generate a heat map as shown below. The variable Difference is defined as the difference between number of male employees and the number of female employees. A negative value indicate more number of female than male employees. In addition to the plot critique the below visualization:
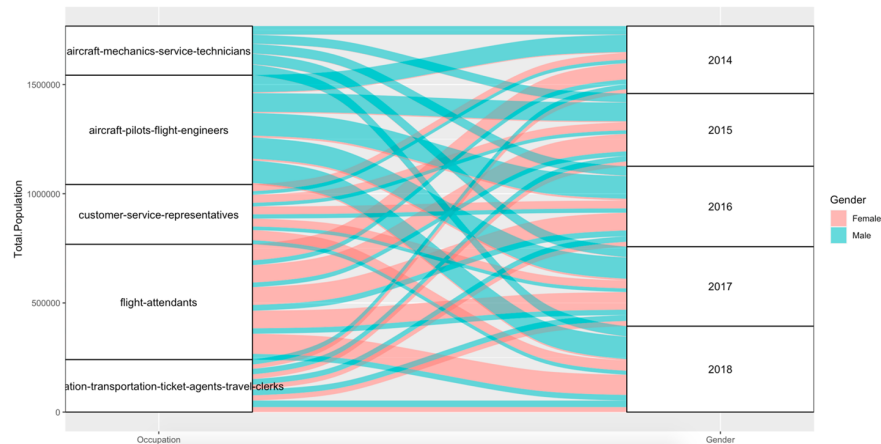
- Transform the x axis using a log10 to better visualize the data spread.
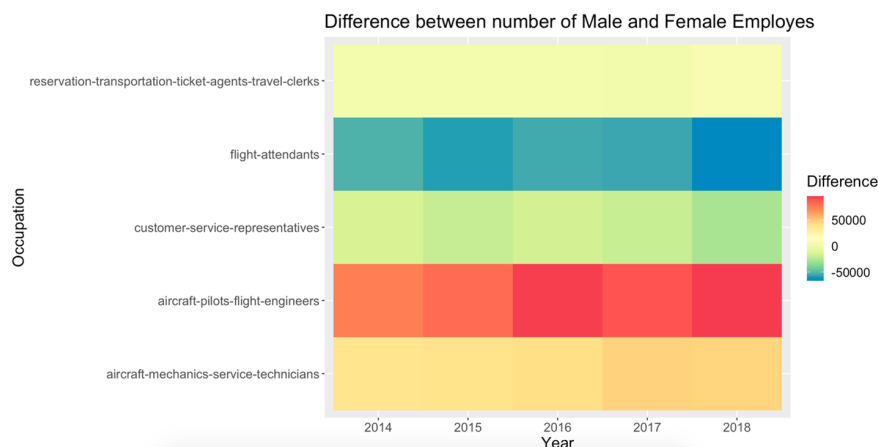
Task 4

From wages_jobs.csv generate the following alluvial chart.

- Read documentation of geom_alluvial and geom_stratum to solve this task

Task 5

From wages_jobs.csv generate a stacked bar plot for the year 2018 with Occupation and Average Wage as the axis and Gender as the color

Task 6

From occupations.csv generate the following tree map. The area of each rectangle is proportional to the number of people working in that Detailed Occupation.