

Stat 437 HW2

Your Name (Your student ID)

General rule

Please show your work and submit your computer codes in order to get points. Providing correct answers without supporting details does not receive full credits. This HW covers:

- Advanced Visualizations via `ggplot2`: adjusting legends, fonts, orientation, and math expressions
- Visualizing networks as graphs
- Interactive visualization

For an assignment or project, you DO NOT have to submit your answers or reports using typesetting software. However, your answers must be well organized and well legible for grading. Please upload your answers in a document to the course space. Specifically, if you are not able to knit a `.Rmd/.rmd` file into an output file such as a `.pdf`, `.doc`, `.docx` or `.html` file that contains your codes, outputs from your codes, your interpretations on the outputs, and your answers in text (possibly with math expressions), please organize your codes, their outputs and your answers in a document in the format given below:

Problem or task or question ...

Codes ...

Outputs ...

Your interpretations ...

It is absolutely not OK to just submit your codes only. This will result in a considerable loss of points on your assignments or projects.

Problem 1

Please refer to the NYC flight data `nycflights13` that has been discussed in the lecture notes and whose manual can be found at <https://cran.r-project.org/web/packages/nycflights13/index.html>. We will use `flights`, a tibble from `nycflights13`.

You are interested in looking into the average `arr_delay` for 4 different `month` 12, 1, 7 and 8, for 3 different `carrier` “UA”, “AA” and “DL”, and for `distance` that are greater than 700 miles, since you suspect that colder months and longer distances may result in longer average arrival delays. Note that you need to extract observations from `flights`, and that you are required to use `dplyr` for this purpose.

The following tasks and questions are based on the extracted observations.

(1.a) For each combination of the values of `carrier` and `month`, obtain the average `arr_delay` and obtain the average `distance`. Plot the average `arr_delay` against the average `distance`, use `carrier` as facet; add a title “Base plot” and center the title in the plot. This will be your base plot, say, as object `p`. Show the plot `p`.

(1.b) Modify `p` as follows to get a plot `p1`: connect the points for each `carrier` via one type of dashed line; code the 3 levels of `carrier` as α_1 , $\beta_{1,2}$ and $\gamma^{[0]}$, and display them in the strip texts; change the legend title into “My ζ ” (this legend is induced when you connect points for each `carrier` by a type of line), and put the legend in horizontal direction at the bottom of the plot; add a title “With math expressions” and center the title in the plot. Show the plot `p1`.

(1.c) Modify `p1` as follows to get a plot `p2`: set the font size of strip text to be 12 and rotate the strip texts counterclockwise by 15 degrees; set the font size of the x-axis text to be 10 and rotate the x-axis text clockwise by 30 degrees; set the x-axis label as “ $\hat{\mu}$ for mean arrival delay”; add a title “With front and text adjustments” and center the title in the plot. Show the plot `p2`

Problem 2

This problem requires you to visualize the binary relationship between members of a karate club as an undirected graph. Please install the R library `igraphdata`, from which you can obtain the data set `karate` and work on it. Create a graph for `karate`. Once you obtain the graph, you will see that each vertex is annotated by a number or letter. What do the numbers or letters refer to? Do you see subgraphs of the graph? If so, what do these subgraphs mean?

Problem 3

This problem requires to to create an interactive plot using `plotly`. If you want to display properly the plot in your HW answers, you may well need to set your HW document as an html file (instead of doc, docx or pdf file) when you compile your R codes.

Please use the `mpg` data set we have discussed in the lectures. Create an interactive, scatter plot between “highway miles per gallon” `hwy` (on the y-axis) and “engine displacement in litres” `displ` (on the x-axis) with the `color` aesthetic designated by “number of cylinders” `cyl`, and set the x-axis label as “engine displacement in litres” and y-axis label as “highway miles per gallon”. You need to check the object type for `cyl` and set it correctly when creating the plot. Add the title “# of cylinders” to the legend and adjust the vertical position of the legend, if you can. For the last, you may look through <https://plotly.com/r/legend/> for help.