CENG 3516 Statistical Computing Midterm Homework

1. (30p) Please reach out to students dataset from https://userpage.fu-berlin.de/soga/200/2010_data_sets/students.csv This dataset gives a set of demographic and education information about ~9000 student and their salaries after graduation. Your main focus for the first part should be discovering **hidden patterns** on this dataset using **data visualization and summary statistics**. You can work/relate any of the variables but make sure you create advanced plots with **ggplot2**.
2. (10p) Find **confidence intervals** for nc.score, score1 and score2 evaluating the differences based on gender. Use **gglot2** to plot the corresponding confidence intervals to support your evaluations.
3. (10p) Find **confidence intervals** for salary evaluate the differences based on gender, religion and major. Use **gglot2** to plot the corresponding confidence intervals to support your evaluations.
4. (10p) Create **one sample hypothesis** for mean salary and score1 for women and test it in 95% confidence level.
5. (10p) Create **two sample hypothesis** for mean salary and score1 using gender as the grouping factor and test it in 95% confidence level.
6. (10p) Create a **two sample hypothesis** to compare mean salaries of the students graduated from “political sciences” and “economics and finance”. Test it in 95% confidence level.
7. (20p) Create working **functions** requested in RLecture4 page 7-8. Implement and demonstrate them using above dataset.

Details about submission:

1. You are supposed to submit a **statistical report** of your findings ideally via **RMarkdown** by 1th of May (17:00) .
2. Please post your.html files. The name of the files should be your name_surname_hwmidterm (like eralp_dogu_hwmidterm.html) through DYS.