STAT 210
Applied Statistics and Data Analysis:
Homework 2

Due on Oct. 1/2020

This homework is partly based on problems from the book *Data Analysis and Graphics Using R – An Example-Based Approach, 3rd Ed.* by J. Maindonald and W.J. Brown, Cambridge Univ. Press, 2010. You can find the problems on the file `HW2B&M.pdf`.

To access some of the data sets, you need to download the package `DAAG` associated with this book. If you are using `RStudio`, select the `Packages` tab on the panel on the right and then select the `Install` tab. Type `DAAG` on the pop-up window and click `install`. After installing, you need to load the package using `library(DAAG)`.

Question 1

Answer question 1 in the file `HW2B&M.pdf`.

Question 2

Answer question 2 in the file `HW2B&M.pdf`. Comment on the differences between the two graphs and how they would affect the conclusions you would draw from them regarding the risks involved in launching the shuttle.

The Challenger Disaster is a well-known case that has been analyzed from several points of view. A particularly interesting analysis from the optic of statistical graphics can be found in E. Tufte's book *Visual Explanations*, Graphics Press (1997).

Question 3

For this question, we will use the file `Cars93` in the `MASS` package.

(i) Use `str` to see the structure of the data set.
(ii) Do a matrix of plots with the variables `MPG.highway, MPG.city, EngineSize, Horsepower`, and `Weight`. Use a solid dot for the plots and color the dots by `Origin`.
(iii) Do a boxplot of `Price` as a function of cylinders. Add the points to the boxplots and color them by `AirBags`. Add a legend for the color of the points on the top left corner.
(iv) Comment on the relation between these three variables. Can you explain the odd shapes of the plots for `Cylinders` equal to 5 and `rotary`?

Question 4

The file `data_q4.csv` has four simulated samples of size 20 coming from the following distributions

- Standard normal, $N(0,1)$
- Uniform in [0,1], $U[0,1]$
- Chi-square with 2 degrees of freedom, χ^2_2
- t with 2 degrees of freedom, t_2

You have to identify which is which using quantile plots. Since you will need to draw quantile plots with respect to distributions other than the normal, it will be convenient to use a new function named `qqPlot` in the package `car`. You will need to install this package following a similar procedure as described above for `DAAG`. This function has syntax

```
qqPlot(x, distribution = 't', df = 2)
```

for plotting a quantile graph of vector `x` with respect to the t distribution with two degrees of freedom. The default distribution for `qqPlot` is the normal distribution. You can find more details in the help for `qqPlot`. By default, this function draws confidence bands which I find in many cases of little use, and in some cases misleading. If you don't want them in your graph, add `envelope = FALSE` in your call.

Explain clearly the reasons for your choices.