MM 217: Descriptive statistics

By this time, we are expecting that you will become more and more independent in terms of learning the commands in R and using them to solve given problems. Please use the **help** command as you deem fit. For this exercise, use any of the built-in data set such as **faithful** for example.

Problem 1

Use the R command **quantile** to calculate the 15^{th} and 87^{th} percentiles of the waiting periods.

Problem 2

Use the command **sapply** to calculate the mean (**mean**), standard deviation (**sd**), variance (**var**), minimum (**min**), maximum (**max**), median (**median**), range (**range**) and quantiles (**quantile**) of the waiting periods.

Problem 3

What do the commands **summary** and **fivenum** do? Use them to describe the waiting period data.

Problem 4

Load the library **Hmisc** and use the command **describe** to understand the waiting period data.

Problem 5

Write an R script to plot the relative frequencies and cumulative frequencies of the eruption duration.

Problem 6

Write an R script to give the stem-and-leaf plot of eruptions. (Hint: use the **help(stem)** command).

Problem 7

Here are a few more challenging problems. The data set **RainfallData.csv** contains the rain fall data from the year 1901 to 2016 for different regions of India (sourced from *data.gov.in* site). Plot the annual rain fall as a function of years for Tamilnadu and Kerala. Plot the data of annual rain fall versus year for the regions that are on top 10 in terms of annual rain fall in this period; name the data points with the relevant region. Plot the data of annual rain fall versus year for the regions that are on bottom 10 in terms of annual rain fall in this period; name the data points with the relevant region. Plot the cumulative rain fall in the year 1973 in Bihar as a function of time. Plot the histogram of rain fall in the year 2000 in Saurashtra & Kutch as a function of months.