STAT 406: HW3

- All computer code should be written using the language R. Type ALL your code into one PLAIN Text format file. Plain text format is available by default in R. Please do not use Microsoft Word .doc format or .rtf format of .pdf format. Inside your plain text file, make sure you identify each problem in a comment placed at the beginning of the problem. The file name should match your name as in 'JohnDoe.R'. Submit your R code file online (under Assignments) at or before the due date, and hand in a hard copy of the code with any additional material to your lab.
- I recommend that before submitting your homework, you also create a new directory and run your R code, to make sure that it is self-contained and runs as you intended.

1. Download the college board data from

   `https://collegescorecard.ed.gov/data/`

   We will work with the 2009 dataset (MERGED2009_10_PP.csv) and 2014 dataset (MERGED2014_15_PP.csv), and the following variables: the school's identifier (UNITID), its name (INSTNM), the control of the institution (variable CONTROL); that is whether the institution is public (coded as 1), private not-for-profit (coded as 2), or private for-profit (coded as 3), and the average annual cost of attending that school (variable COSTT4_A).

   - For each school compute the rate of change in the average annual cost of attendance between 2009 to 2014. Work only with the schools for which the variable COSTT4_A is not missing in neither of the years.
   - Make boxplots that present the distributions of the rate of change for public schools, private not-for-profit schools, and private for-profit schools. Put the three boxplots on the same figure.

2. The file consumer_complaints.txt is a dataset containing complaints received by the Federal Bureau of Consumer Financial Protection (CFPB). The file can be downloaded from Canvas or from the link

   `http://www.stat.lsa.umich.edu/~yvesa/consumer_complaints.txt`

   We will work with the following variables: the date the complaint is received by CFPB date_received, and whether the company (against which the complaint is filed) responded timely to the complaint timely_response.

   (a) For each month of the years 2012 to 2016, compute the proportion of complaints timely addressed by each of the following company: JPMorgan Chase & Co., Bank of America, Wells Fargo & Company, Citibank. In this task, the R function strsplit can be useful to split a date in order to extract the year and the month.

   (b) For each of these four companies, you should obtain a series giving the proportion of timely addressed complaints over the given months. Plot these four series on the same graph.