IST 387 HW 6

Copyright 2021, Jeffrey Stanton, Jeffrey Saltz, and Jasmina Tacheva

```
# J.J. Q2 this homework by myself, with help from the book and the professor.
# J.J. Q3 this homework with help from the book and the professor and classmate Jasmina.
# J.J. Q4 this homework with help from the book and the professor and your peers any code.
```

Remember of things to practice from previous weeks:
Descriptive statistic means ()
Frequency table (table()function ()
Transposed: Ask if for help with a command

The website Data visualization is important because because many people can make sense of data more easily when it is presented in graphic form. As a data scientist, you will have to present complex data to decision makers who may not have the data sophistication for them. From your experience with Excel and other tools, you know that there are a variety of ways to present numbers in charts and graphs. In this homework, you will:

The main procedure for data visualization is R in a strict **gg**ame. Written by computer scientist **Hadley Wickham**, the **"ggplot2 grammar"** tool kartz visualizations in layers. This method provides immense flexibility, but takes a bit of practice to master.

Step 1: Make a copy of the data

1. Copy the **adhd.csv** file from the course website to your **/ist387/datasets** directory.
2. Create a new dataset called **TbCases**. This file contains a so-called **multivariate time series**, a sequence of measurements on 23 Tuberculosis-related (TB) variables captured repeatedly over time (1980-2013). Familiarize yourself with the nature of the 23 variables by consulting the dataset's codebook which can be found here: https://ist387.ist.psu.edu/~j.stanton/datasets/TB_data_dictionary_2013-02-08.xls

```
URL="http://ist387.ist.psu.edu/~j.stanton/datasets/data/tbcases.csv"
TbCases, data.frame(read.csv(URL))
TbCases
```

```
                                                
```