# HW 2

*Mike Lopez*

*September 2017*

***General instructions for homeworks***:

- Make a new R Markdown file (.Rmd) referring to thea assignment on the course Github page
- Change the heading to include your author name
- Save the R Markdown file (named as: [MikeID]-[Homework01].Rmd – e.g. "mlopez-Lab01.Rmd") to somewhere where you'll be able to access it later (zip drive, My Documents, Dropbox, etc)
- Your file should contain the code/commands to answer each question in its own code block, which will also produce plots that will be automatically embedded in the output file
- **Each answer must be supported by written statements (unless otherwise specified) as well as any code used**: In other words, if the answer is `24`, you should write "The answer is 24" (as opposed to just showing the code and output).
- Include the names of anyone you collaborated with at the top of the assignment

**Part I:**

Open Intro Chapter 1: 1.21, 1.36, 1.37, 1.38, 1.40, 1.42, 1.44, 1.49

## 1.37:

1-c, 2-a, 3-b

## 1.40:

a. The boxplot shows the outliers (not shown in the histogram), while the histogram does a slightly better job of identifying the bimodal nature of marathon times (two humps)
b. Best male times & best female times
c. The men tend to have lower finishing times, with a median around 2.2 hours, compared to 2.45 hours for females. Both distributions are skewed right. There's slightly more spread for the female times.
d. Marathon times have gotten slightly faster over time.

**Part II:**

5. Explain why average delay may not be best choice of metrics regarding delays, and propose a better one. Is your answer to (4) the same using a newer metric?

**Averages can be strongly impacted by outliers – perhaps median delay time would better represent the typical traveler**

6. Your answer to (4) identifies the specific date in January of 2013 that boasted the worst, on average, arrival delays. From a sampling perspective, explain why it could be incorrect to argue that this was the worst day as far as flying from NY to Boston.

**Flights could've been cancelled! This ties into the Wald article below – from a sampling perspective, always remember the observations that never got a chance to be observed**

7. Read the case of Abraham Wald (link at https://medium.com/@penguinpress/an-excerpt-from-how-not-to-be-wrong-by-jordan-ellenberg-664e708cfc3d), and tie into the measurement of airline delays.