

Midterm project

MA615

October 5, 2018

This project requires you to acquire, clean, organize, explore, visualize, and report on a dataset. Each team will build **two datasets** for exploring the relationship between weather conditions and attendance at sporting events. The focus is on attendance at Red Sox and Celtics games.

There are many websites that are dedicated to baseball and basketball and to the Redsox and the Celtics. For the Red Sox you will find: [the official Red Sox Site](#), [Fenway Park](#), [Baseball Reference](#), and [ESPN](#) – to name a few.

For the Celtics, there is also a wide variety of sites: [the official Celtics site](#), a site dedicated to [Celtics statistics](#), the [TD Garden](#) site, the [ESPN](#) site, and the [basketball reference](#). Again there are many to choose from.

Weather and climate site are much easier to find. You already know how to access NOAA weather buoy data. Of course, there are many other sites that make weather data available. The [National Center for Environmental Information](#) has makes a variety of data products availabl including [Storm events](#) and [Severe weather](#).

Please note: you are not restricted to the sites listed above. They have been included to make sure you have enough material to complete the assignment.

The assignment is to assemble a dataset that can be used to explore the relationship between the weather and attendance at Red Sox and Celtics games. **Attendance data is more difficult to find than you might think.** Here are two starting points:

- [Red Sox Attendance 2017](#)
- [Attendance Milwaukee Bucks at Boston Celtics Box Score, October 18, 2017](#)

Build a dataset that covers the **2012 seasons through the 2017 seasons**. **Clean** and organize the data so that it is ready for analysis. **Explore** the datasets and produce a report about your datasets and how you built them. Make sure your work is reproducible. Be sure to include **plots** and other graphics.

In addition to your report, **prepare a shiny application** that a user could use to explore the relationship between weather and attendance to baseball and basketball games.

Weather may not be the principal determinant of attendance. If your dataset sheds light on other factors affecting attendance, you should note them.

Rubric

Refer to the rubric in the syllabus for general guidance, keeping in mind that the assignment requires you to demonstrate your ability to acquire, clean, organize, explor, visualize and report. Use the assignment to deepen your understanding of R and to demonstrate your skills.

PLEASE NOTE: 10% of the grade on this assignment will reflect the originality of the work submitted.

Submissions

You must submit individual assignments, although each of the individuals on the team may submit the same github url. In addition to the github URL, please submit individually your estimate of your work contribution to the assignment as a percentage of the total work done by your team. Also, please list up to three features of the delivered assignment where you made a significant contribution.