**Basic Info:**

Title: Investigating NFL Historical Data

Team Members: Michael Nwauche and Ethan Stavisky

Link to Repository: <https://github.com/mnwauche/data-viz-project-nwauche-stavisky>

**Background and Motivation:**

The primary motivation for this project is that we both love to watch and analyze football. Football certainly gets less attention when it comes to data analytics when compared to the NBA and MLB. This is likely due to the nature of football being such a team sport and that the offense does not play at the same time as the defense. However, there is an increasing amount of data out there pertaining to NFL performance. We are both interested in looking at this data and how it is being used to alter how NFL teams make decisions today. So, when we were discussing a possible topic for this group project, we both naturally thought about football. We think it would be really cool to create some visualizations based on this topic and this data that we are interested in. This was also somewhat influenced by our class project with NBA data. We both enjoyed creating that visualization because it related to sports and thought that we could do something similar for this project. This topic is not related to any research interest we have but, rather, simply a shared hobby.

**Objectives:**

Our primary objective is to create an interesting and highly interactive visualization. We hope to create something that people would be eager to play around with and investigate the features. We hope to learn more about the data we gather and insights about the NFL. In general, we expect to learn more about how to turn raw data into an effective visualization by utilizing user friendly features and visual designs.

**Data:**

We plan for all of our data to come from [pro-football-reference.com](https://www.pro-football-reference.com/years/2018/). This website has a tremendous amount of NFL data and will surely have almost everything we need. The data can be broken down by features such as year, team, offense or defense, rushing or passing etc. This website allows you to download the data into a CSV file which is a format we are very comfortable with. The data will be saved in provided format and will be modified to suit the needs of each individual visualization

For example: <https://www.pro-football-reference.com/years/2018/>

**Data Processing:** We don’t believe there will be too much typical data processing tasks like data cleanup and data aggregation for this project. Pro Football Reference has many features that will allow us to do some of these tasks from within the website before we download the data. We believe data processing should not take more than a couple hours.

**Must-Have Features:**

Our visualization will need to be highly interactive in order to support data exploration. This means that we must enable the user to filter data based on interactions with axes, marks, legends. We will also need a search function because of the large amount of data. The user will have a robust set of filtering options for the data. Features will include a time series that automatically adjusts zoom based on selected data (e.g. zoom in to show an entire

player's history, zoom out to show an entire teams history). We will also need to have a parallel coordinates plot in order to identify correlations between variables.

**Optional Features:**

Optional features that might be useful for data exploration would be a ranking function that orders data based on user defined weights and metrics.

**Project Schedule:**

Week 11 (October 28):

* Create Project Proposal
  + Ethan:
    - Basic Info, Background and Motivation, Objectives, Data, Project Schedule.
  + Michael
    - Data Processing, Must-Have Features, Optional Features, Visualization Design.

Week 12 (November 4):

* Prepare for Project Update (Due Wednesday)
  + Ethan
    - Summarize the data we have collected.
    - Explain how we have cleaned the data.
    - Discuss the data issues we have run into so far.
  + Michael
    - Provide initial analyses and visualizations that we have created to help understand our data.
  + Together
    - Maintain Process Book.

Week 13 (November 11):

* Create Project Prototype (Due next Monday)
  + Ethan
    - Finish collecting and cleaning data. Create data structures.
  + Michael
    - Implement ability to load up and visually display the data to some degree.
  + Together
    - Maintain Process Book.

Week 14 (November 18):

* Begin work on final project
  + Together
    - Implement ability to load up and display all data.
    - Begin to create all visualizations.
    - Begin to implement functionality and features.
    - Maintain Process Book

Week 15 (November 25):

* Finish work on final project
  + Together
    - Finish all visualizations.
    - Finish implementing all functionality and features.
    - Maintain Process Book.

Week 16 (December 2):

* Prepare for Project Presentation (Due Wednesday)
  + Together
    - Make final checks and edits on project
    - Discuss which parts of our project we want to spend the most time presenting about.
    - Plan how we will present the project.
    - Finish Process Book.

**Visualization Design:**

Our visualization design will include:

* Time series for NFL history, auto zooming and highlighting based on selected data.
* Stacked view of individual player careers: player careers will be plotted on a standard timescale with individual marks denoting games, the position will encode time and players performance based on adjustable metrics.
* Stacked view of seasons data: marks representing individual games, position encode time and team ranking.
* Stacked view of team history data: marks representing seasons, position will encode time and team ranking.
* Parallel coordinates for players
* Parallel coordinates for teams
* Dashboard for presenting details about selected data.