CS 171 Project Proposal

Christine Hwang, Joseph Song, Travis Yeh

Background and Motivation

Although the team shares an interest in basketball, we each have different perspectives that have allowed us to have a diverse view of how we would ideally consume basketball statistics. Joseph, a member of the Harvard Hoopsters club basketball team, enjoys keeping track of breakout players and trying to see how their performance in key games and seasons compares to their early years in the league. Travis, a fan of basketball that never felt particularly invested in a single team, focuses on the performance of teams as a whole. Christine, a diehard Lakers fan, really only cares about the Lakers and how the team and players are performing compared to previous years and league averages. Thus, we believe that our diverse backgrounds and goals when approaching basketball data allow us to work together and create a visualization that would be robust and useful.

Project Objectives

Basketball data is well-documented and available from many sources around the web. However, it is almost always presented in the standard box score format -- a massive chart, with a multitude of columns for each player or year. As a result, it's difficult to read and process the data that's given, and even harder to compare data across players, teams, and years. With our project, we hope to take the data that's available on player and team performance and translate it into a visualization that is elegant and easy to use. The box score is certainly an efficient and organized way to present data, but it fails to allow for much actual understanding besides the pure numbers. We hope to take the traditional box score and turn it into a visualization that facilitates data consumption for cases and purposes as diverse as those held by the members of our team, and allow easy comparison of teams and players against the rest of the league and their past performances.

Data

We will be scraping data from online resources. We can use R's Selenium package and Python's Beautiful Soup package to scrape from the tables that are available publicly

online to get the data that we need. We will start by scraping data from the CBSSports API http://www.cbssports.com/nba/players which allows for us to gain easy access to the basic statistics for each NBA player. Any additional data required can be gained from player profile pages such as http://espn.go.com/nba/player/_/id/110/kobe-bryant which holds their past few games and some more recent statistics on how they are doing currently. The rest of the more advanced statistics can be scraped using http://www.basketball-reference.com/ which holds statistics such as rankings of players amongst each other, per possession data, etc.

Data Processing

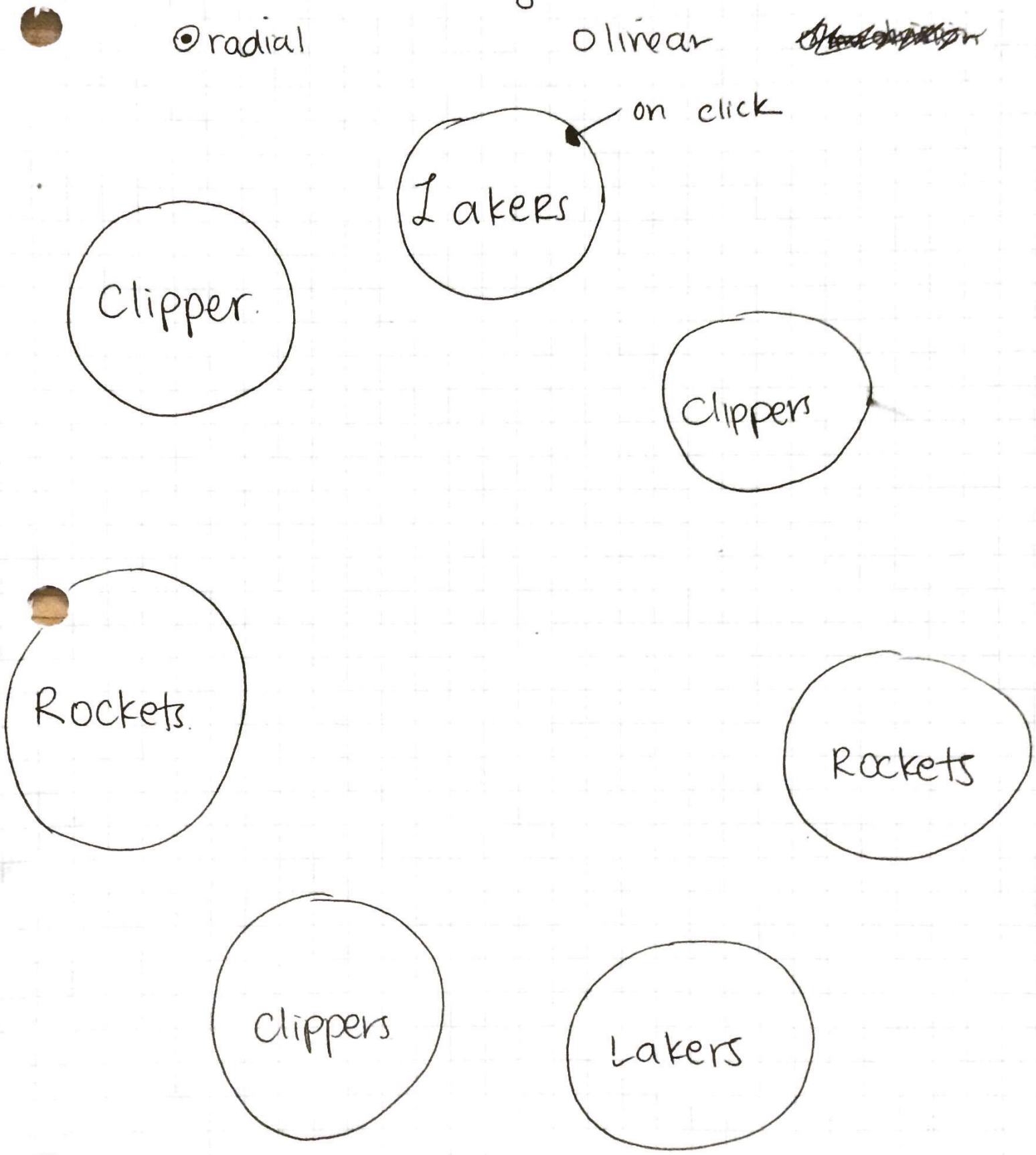
We do not expect to do any substantial data cleanup except for casting and pivoting tables to achieve the data that we are looking for. We want to be able to compare team performances, player performances, player to team performances, and change over time for all of these separate groups. Data processing should not be necessary besides for basic aggregation and pivoting that can be done in R's reshape2 or python's pandas.

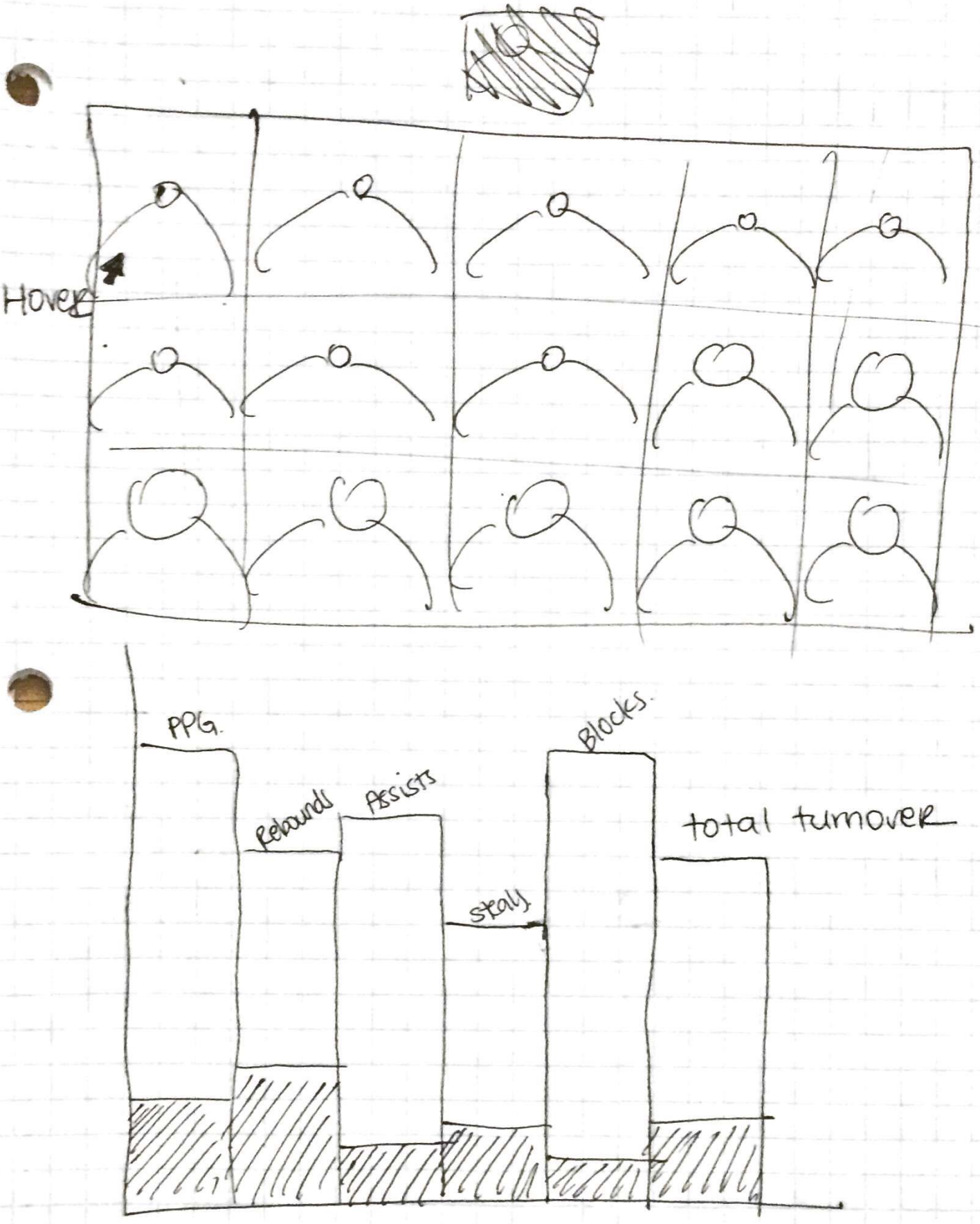
Visualization

We want to be able to display our data using a myriad of the different techniques that we learned through the previous problem sets and new examples that we saw in the d3 website. We want to implement the table from homework 1, circle foci and radial views from homework 2, and even possibly a brush (given able to find the correct data to implement something like this). Some of the new visualizations that we wanted to try are implementing much more transitions, and thus we want to attempt to create the show reel that was on the d3 website, where we are able to draw the line upon selection as a cool transition to engage the user on what is happening through a given player's career. We also want to have much cleaner transitions between the different views, where upon click of a team's logo would expand out and then contract in to the team page. Another example of this would be upon selection of the player logos to implement a similar sort of transition. An important part of this project is to help us see relative difference in stats compared to an individual and their teammates. An example of how we will show this, is on our team players page, we will have a default bar chart that shows total PPG, turnovers, rebounds, etc for the entire team. Then, when we hover over a team player image, it will draw a bar chart on top of the existing bar chart, and this will help us see the proportion this specific player helped contribute to team in this statistic.

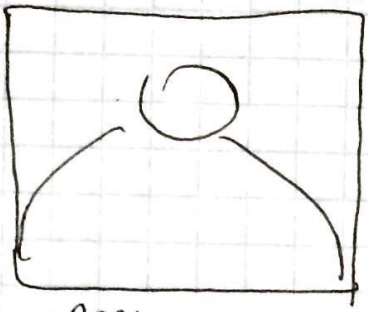
Home Page

⊙ radial ○ linear ~~○ [illegible]~~

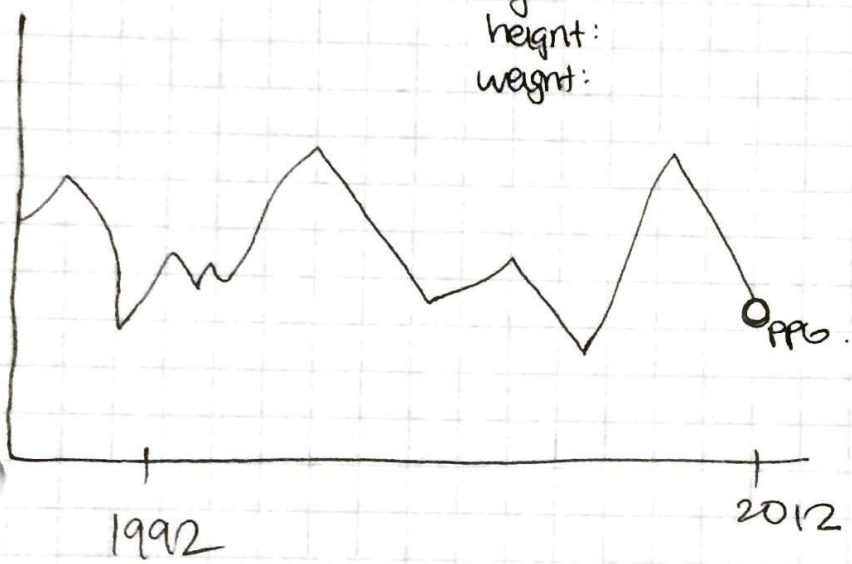



Hover
PPG.
Rebounds
Assists
steals
Blocks.
total turnover

Individual Player



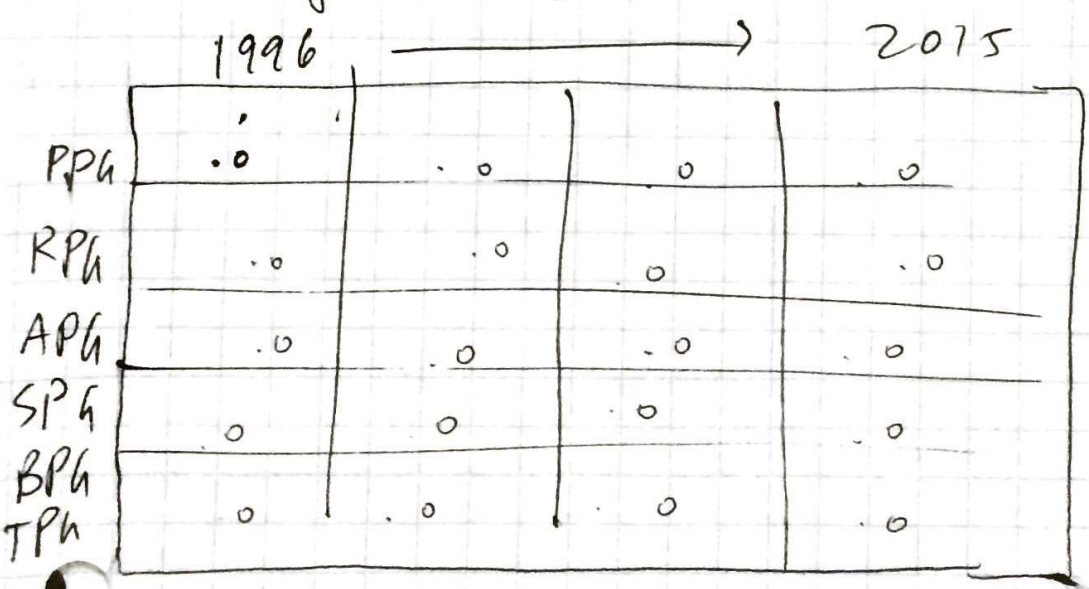
Age:
height:
weight:



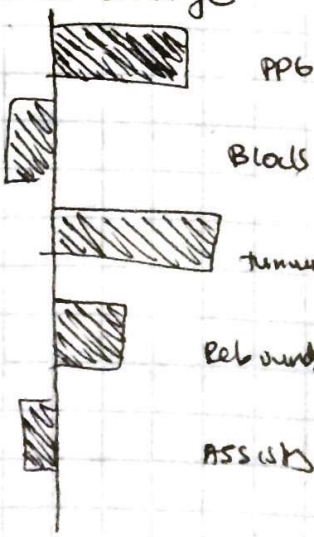
Radio Buttons

- (•) PPG
- ( ) Blocks
- ( ) Turnover
- ( ) Rebounds

~~Current year stats~~



current year perf vs averge



West
East
1993
2018

Must-Have Features

We have many must-have features. We want to first begin with a homepage of the teams with their logos, where upon click they are able to enter a team view where they can see a roster and a chart that is able to show an individual player's performance relative to the team. This is so that the user is able to see how important a given player is to the overall team's performance, and this will be implemented by showing a bar chart that shows how much of the overall team performance a given player is responsible for. The next feature is to be able to enter a player view where we are able to see a table of their basic stats (points, rebounds, assists, etc.) throughout the career and give the user the option to sort ascending and descending based off a certain statistic. This is to give the user an easy way to actually access the numerical values of these statistics instead of having to rely on visualizations to make comparisons. Below this, we want to implement a change over time so that the user can see the player progression as they went along in their careers. We want to implement a line chart where we will draw a single statistic from a player over time and see how they did in terms of points per game, or rebounds per game, etc. throughout their career.

Optional Features

Some higher-hanging fruits that we will attempt to achieve are animations to see which teams managed to make the playoffs and their seed in those playoffs, separated by conference, on a yearly basis from the history of the NBA. Another would be aggregated +/- data so that we can give the user the ability to pick their own teams out of current players and "simulate" the score if these lineups were to play, providing a prediction to the result of the game. Another consideration would be make a heatmap where we could track a team's performance throughout their existence, where the colors are populated by the number of wins in that year.

Project Schedule

Week of April 5th: Data Acquisition and homepage

Travis: Clean up data into csv and import into javascript/d3. Update as needed as Joe scrapes more data. Create a radial layout and linear of team names (ie the homepage)

Joseph: Set up the data scraper using python and R (beginning of the week). Start off by scraping 100 players so that other members can begin processing. Continue to scrape throughout the week as Christine and Travis work on implementing design.

Christine: Scrape the player images and team images for the team page. Update as needed as Joe scrapes more data.

Week of April 12th: Team Page

Travis: Responsible for team page (ie hover over team player's image and show a fluctuating bar chart to show relative difference in players compared to overall team averages). Have pictures imported and hover feature implemented.

Joseph: Work on show reel for individual page. (Given more time because this is a new concept). Create radio button to restart show reel depending on PPG, Blocks, Turnover, and Rebounds

Christine: Begin to analyze current year's individual stats to stats over the years. Implement bar chart that shows delta in PPG, Blocks, turnover, rebound, assists. Begin to construct table with player's stat.

Week of April 19th: Finalize mandatory features:

Travis: If not completed with previous task, then complete. Begin to work on the clicking function. For example, when we are on the team's general page with all player's pictures, implement feature that takes user to the page with individual statistics when a player is clicked. Also work on transitions between pages.

Joseph: Begin to work on CSS of the page, including color, font, style to improve visualization. Also begin optional implementation

Christine: Implement the filter and sorting for the table. Create features similar to table.html in pset 1.

Week of April 26th : Begin to implement optional features

Travis: Work on creating an animation where playoff matchups can be seen over time

Joseph: Work on the heatmap that tracks team's performance, which density populated by number of wins per year.

Christine: Assist Joe with CSS and optional feature. Begin to work on optional feature of the possibility of adding a simulation feature to the website.

Week of May 3rd : Everybody finalize details and be ready to submit by May 5th.