CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

https://github.com/cmearl6/cpsc4030project.git

Background and Motivation

The National Basketball Association has been tracking and recording all different types of statistics and data for its players and teams since the league originated in 1946. As fans of basketball and specifically the NBA, we are very interested in the performance of teams and players, and the statistics that are associated with them. We are also interested in exploring the development and changes of the league since the first season. For example, the game is much different now in terms of the amount of points scored and three point attempts.

Project Objectives

There are many different ideas we have for visualizations and what we can learn from them. As I mentioned before, we are interested in seeing the changes of the league over time. Some questions we want to explore relating to these changes are as follows. How have salaries of individual players increased since the beginning of the league? The NBA is as popular as ever in this day in time, and it makes sense that the amount of money in the league has improved drastically. We think this would make for an interesting visualization in seeing the change in player salaries over time. It would also be interesting to see the distribution of salary across the players in the league. Do the top players today make more than the top players in the past compared to league average? Or is the wealth more evenly distributed among the players today? We would also like to do the same type of visualization for team salaries. Another question we

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651
https://github.com/cmearl6/cpsc4030project.git

would like to address is “How have statistics such as 3 point attempts, Rebounds, and Assists changed over time?” These were the first statistics that came to mind, but there are many more that we can expand to in our visualizations. We would also like to visualize how money impacts the performance of a team. Do teams with higher salaries win more on average? How closely does a player’s performance correlate with their salary over their career? We also have some geographical questions that we would like to explore. This would mean asking questions such as: “Are there states that produce more NBA talent than others?” and “Are there colleges that produce higher draft picks than others?” Finally, we want to explore how many more all star appearances on average first round draft picks have compared to second round picks, third round draft picks, etc.

Data

We collected our data from the following link: https://www.kaggle.com/wyattowalsh/basketball. This dataset has data from more than 60,000 games, the 30 teams in the NBA, and 4,500 players that have played since the inaugural season of the NBA. The data is stored using the SQLite relational database management system. We have converted these tables from SQLite to csv files. This dataset will be updated daily once the NBA season starts on October 19, 2021, but for simplicity’s sake we will not be using the data that is added in for this season. We also might need to pull in some extra data for the past salaries of players and teams.

CPSC 4030 Data Visualization Project Milestone 2
Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651
https://github.com/cmearl6/cpsc4030project.git

Data Processing

For data processing we only had to convert the .sqlite file to csv tables to make them easier to work with. Other than that, we do not anticipate any substantial data cleanup. The quantities that we will be deriving from our dataset will be mostly quantitative. The first quantitative attribute that we will use are dates for games, when a player played in the NBA, when stats were recorded, etc. We will also be using locations for things like where teams are located, where a player went to college, and where a player was born. This will help us in visualizations that involve maps. We can also use physical measurements such as height and weight of players. A big attribute we will utilize are counts and amounts. This will include stats like points scored, field goal percentage, and the amount of all star appearances a player has. For data processing, we will need to convert our csv files to json, which can be used with the D3 JavaScript library.

Visualization Design

For our visualization design, we will utilize a few different designs. As a general idea, we will use bar charts, line charts, spatial diagrams, and box plots. The line charts will be used for our visualizations that show how salaries of teams and players have changed over time. We could have two different line charts for team and player, or we could have one line chart with the option for the user to switch between player and team. Also, for the user to be able to dive in deeper to specific time periods, we could allow them to zoom in and out on the chart. Or, we

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651
https://github.com/cmearl6/cpsc4030project.git

could provide an input for start and end years and adjust the chart accordingly. To accompany this line chart, we think it would be interesting to have a box plot that shows the distribution of salary across the NBA's players. By clicking on different years, the user could see how the distribution of wealth in the NBA has changed over time. Line charts will also be used for our change in statistics over time. We think it would be a good idea for our design to have different lines for each position in basketball as well. This would allow us to combine another attribute and get a better idea of how the NBA has changed over time. For this chart we will give the user the option to change the stat that is being viewed on the chart, along with the ability to change the start and end date for the chart. It might also be interesting to have another line on the chart for total points scored in a game on average for each season. This would allow the user to view the change in total points scored over the years and also connect that to the other stats that are being displayed in the visualization. The final line chart that we will have will be used to visualize how closely a player's performance correlates with their salary over their career. The x-axis for this chart will be for the amount of seasons the player played and the y-axis will have the salary of the player. There are a few different ways we can approach this visualization to relate salary to performance. One option would be to mark on the x-axis when the player had an all star appearance, representing a good season. The other option would be to have a second y-axis representing a stat that is indicative of a good season such as points scored. We will make use of a bar chart for visualizing the relationship between team salaries and the performance of the team. We could do this by grouping the teams by top 10 salaries, middle 10 salaries, and lowest 10 salaries for each season and getting the average win percentage for each group.

CPSC 4030 Data Visualization Project Milestone 2
Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651
https://github.com/cmearl6/cpsc4030project.git

Finally, we plan on using maps to visualize where the most NBA players come from, and also what colleges generate the highest draft picks. We plan on using a bubble map with bigger bubbles for higher numbers of NBA players going to that school. For the second map we plan on having a color scale that shows where the highest draft picks have come from. We will do this by taking the average draft pick of each player from a school and use that average on a scale to compare to other schools. One design choice we could make would be to combine these two maps to have the size of the bubble and color be an indicator. We could also provide a zoom functionality that would allow the user to click on a state and zoom in to the individual schools in that state. Another zoom option would be to just let the user zoom normally with their mouse on the map.

Must-Have Features

Along with our design plan, we have a few features that we think we must-have. These features will focus on the user being able to interact with our visualizations that we create. One must-have feature is the ability for users to change the time period they are viewing on a chart. For example, if the visualization is showing the average points scored by a player over their career, we would like the user to be able to filter the start and end year for the data included in the visualization. Another must have feature is for the user to be able to see labels by hovering their mouse over a certain part of the visualization.

CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

https://github.com/cmearl6/cpsc4030project.git

Optional Features

Our optional features are features that we think would improve the project, but are not necessary to have. One optional feature would be to allow the user to change the statistic that they are viewing on a line graph. We would also like to look into being able to click on a state to zoom in and see data on a state level. In order to provide more customization, it might be cool to be able to combine stats on a line graph to show one line instead of multiple. For example, if the user wanted to view one line that represents points + rebounds + assists instead of three separate lines they can do that. Finally, we would like to figure out a way for the user to enter a year and have a box plot with the distribution of player salaries for that year.

Project Schedule

Week 1 (10/17 - 10/23)

- Create basic website pages using HTML, CSS, JS
- Connect dataset to website
- Start working on line chart visualizations

Week 2 (10/24 - 10/30)

- Finish line chart visualizations
- Start on map visualizations

Week 3 (10/31 - 11/6)

- Finish map visualization

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651
https://github.com/cmearl6/cpsc4030project.git

- Start on must have features for visualizations

Week 4 (11/7 - 11/13)

- Finish must have features
- Start on optional features

Week 5 (11/14 - 11/20)

- Keep working on optional features
- Wrap up visualizations with must-have features. Begin testing.

Week 6 (10/31 - 12/3)

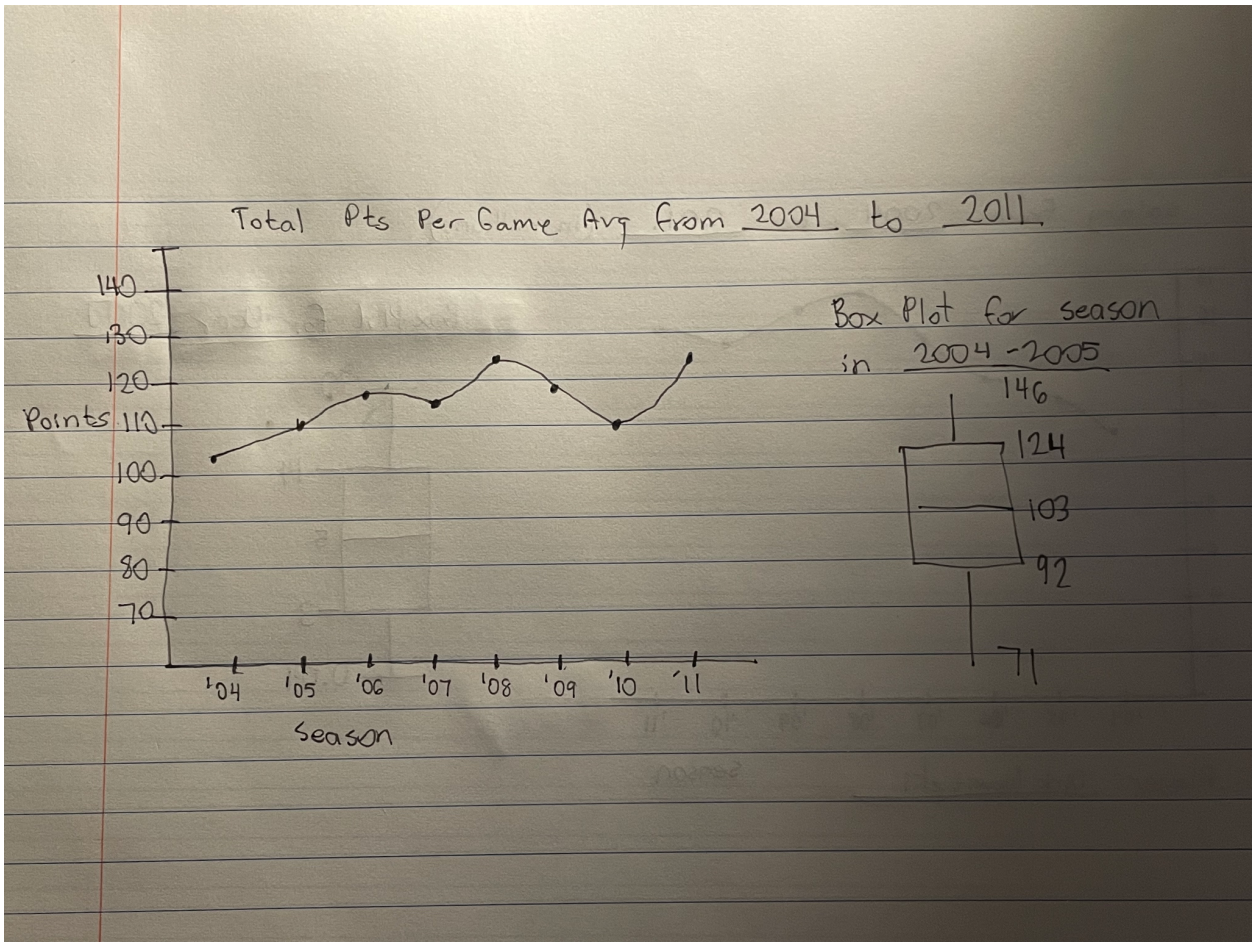
- Finalize any optional features we were able to add.
- Finish testing for any bugs and finalize the website.

CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

https://github.com/cmearl6/cpsc4030project.git

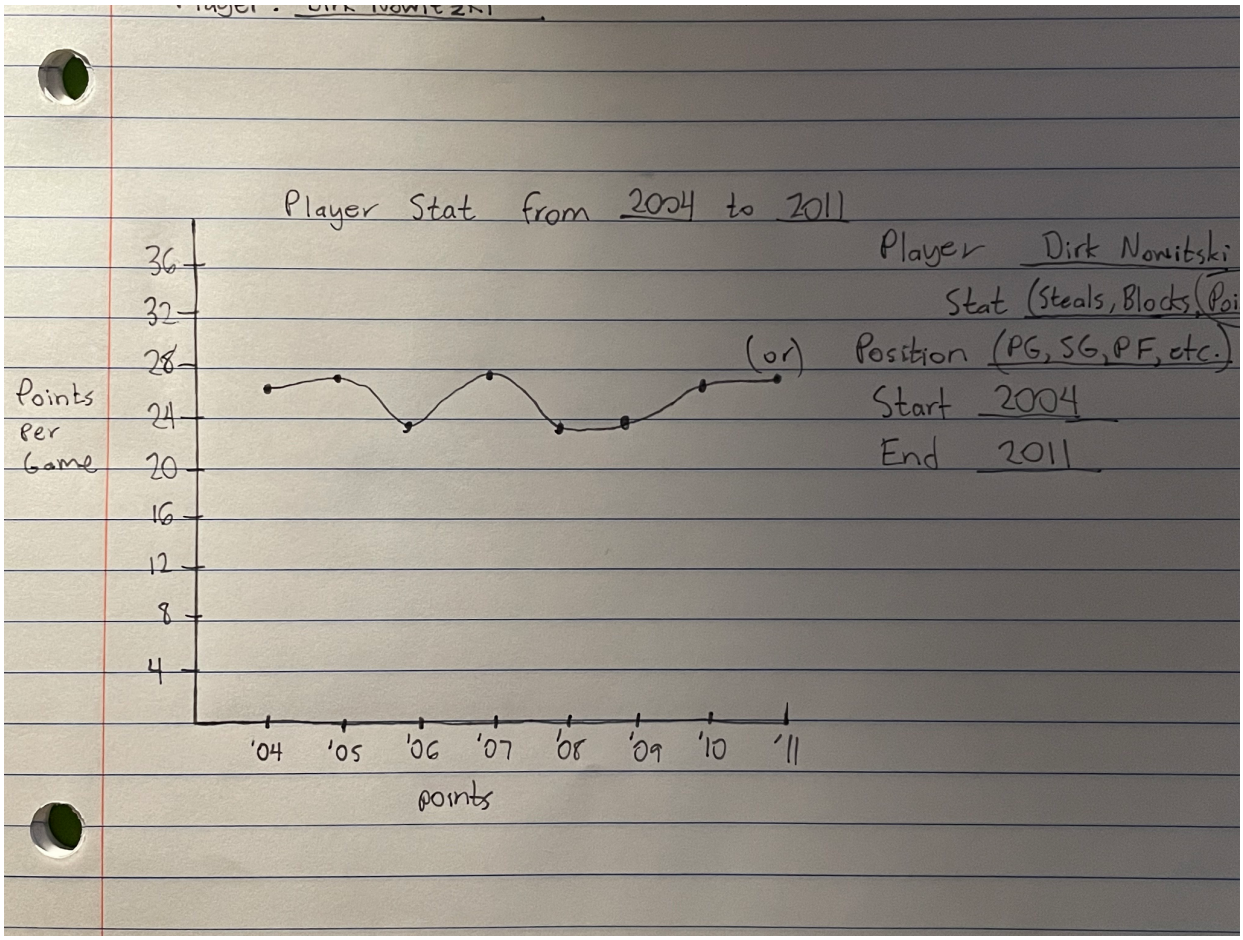
Sketches



CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

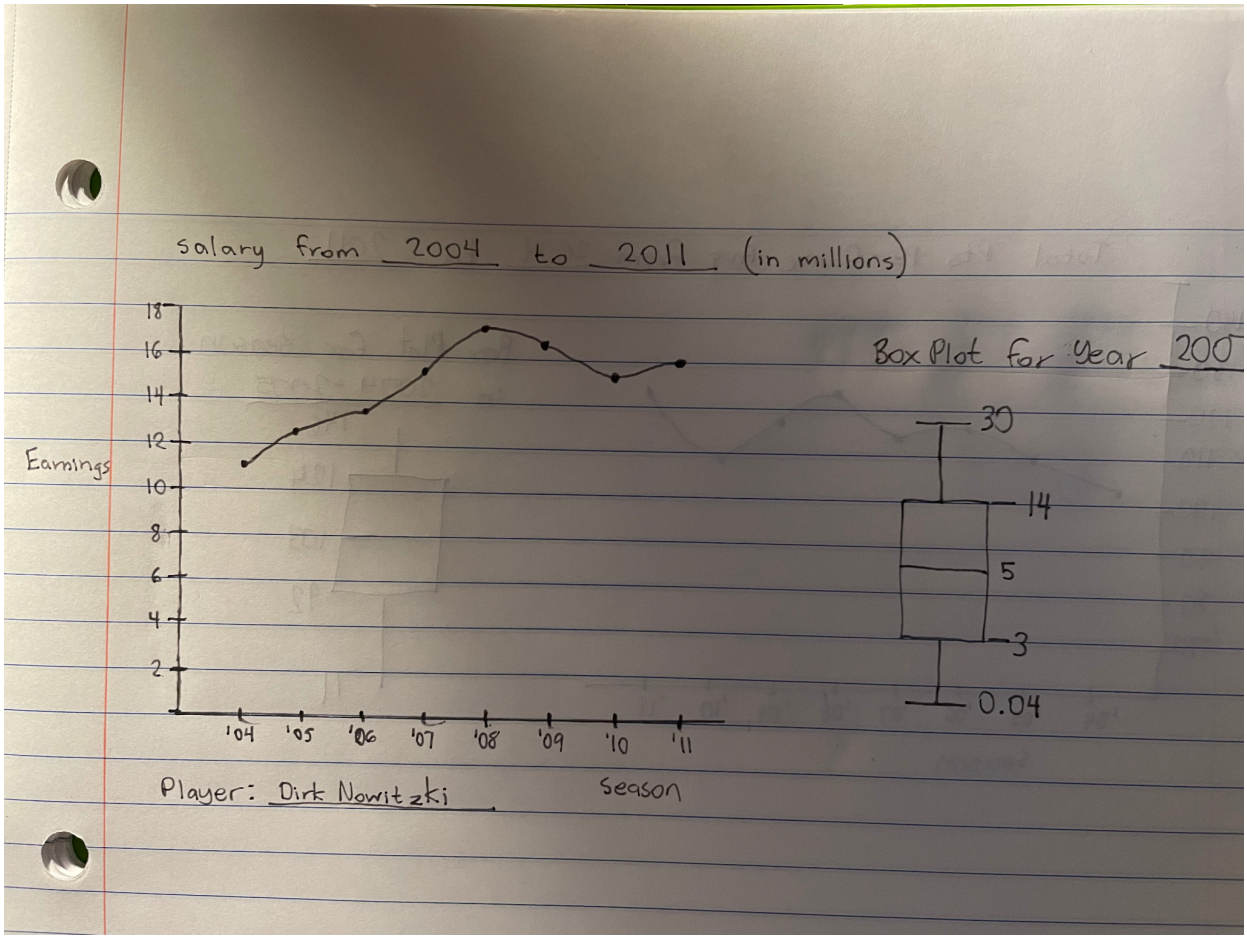
https://github.com/cmearl6/cpsc4030project.git



CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

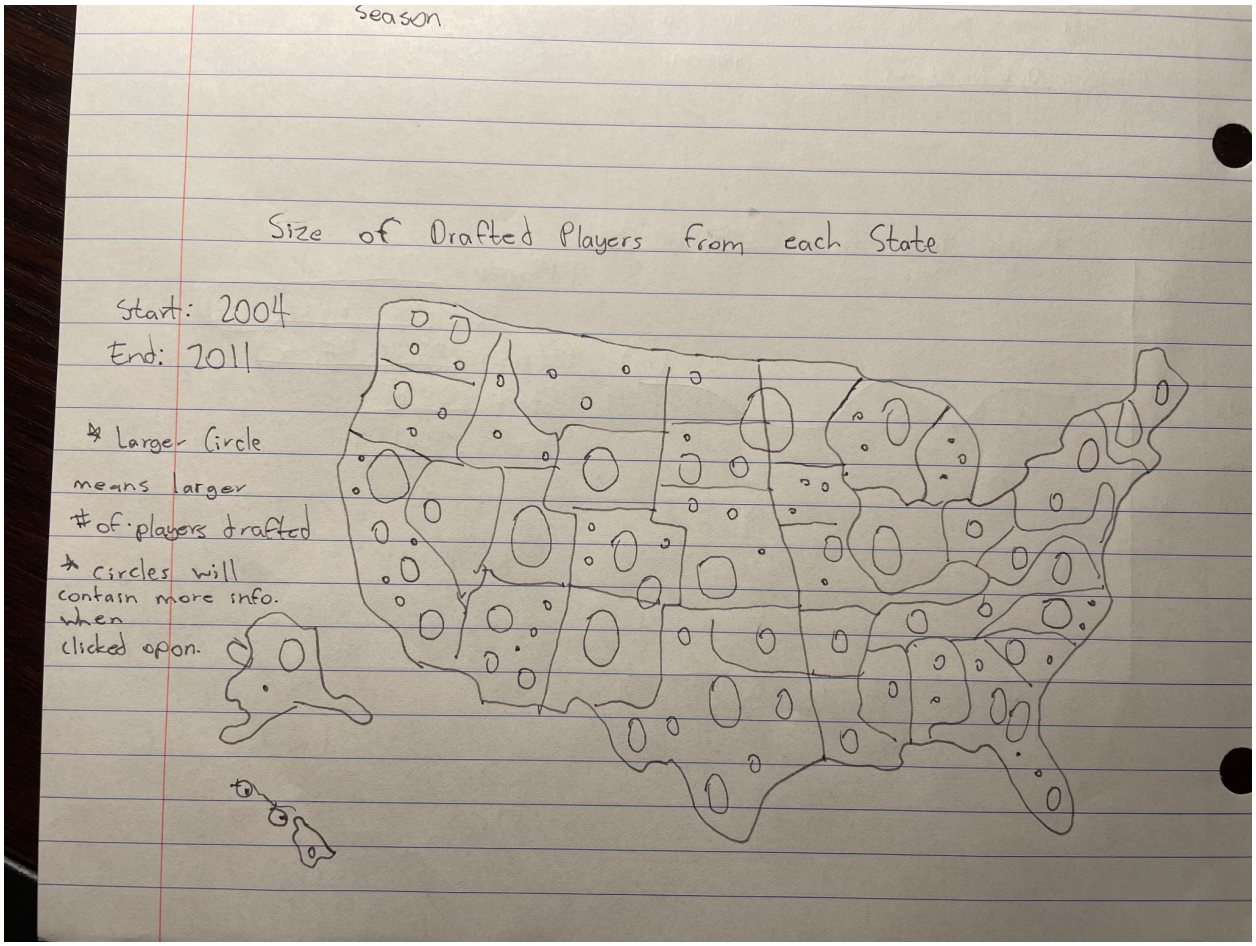
https://github.com/cmearl6/cpsc4030project.git



CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

https://github.com/cmearl6/cpsc4030project.git



CPSC 4030 Data Visualization Project Milestone 2

Cole Earl | Raj Patel
cmearl@g.clemson.edu | rajkump@g.clemson.edu
C62728872 | C65897651

https://github.com/cmearl6/cpsc4030project.git