Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

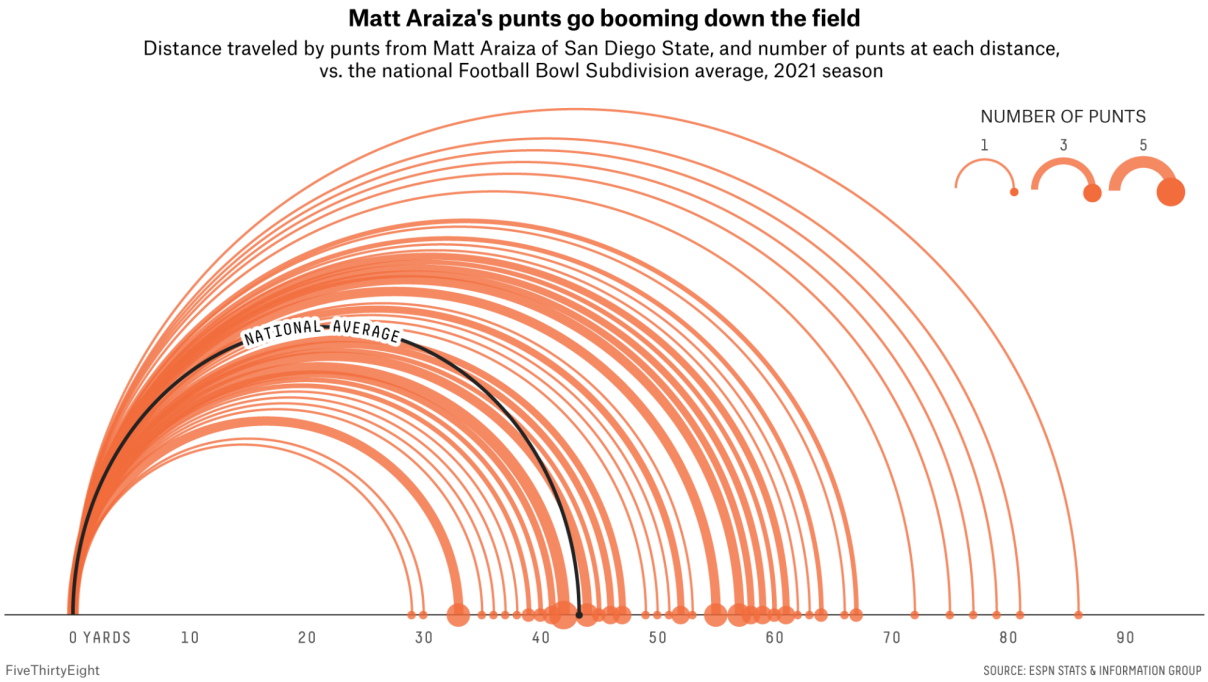
# Overview and Motivation

This project and the topic that we chose was inspired by our fandom of the National Basketball League and the many players that are a part of it. We also are interested in the statistics of our favorite players and are always interested in learning more about them. Our main goal of this project was to explore the performance of players over time in different statistics, and relate that performance to their salary. We were interested to find the relationship between these two things, and how closely they are correlated.

# Related Work

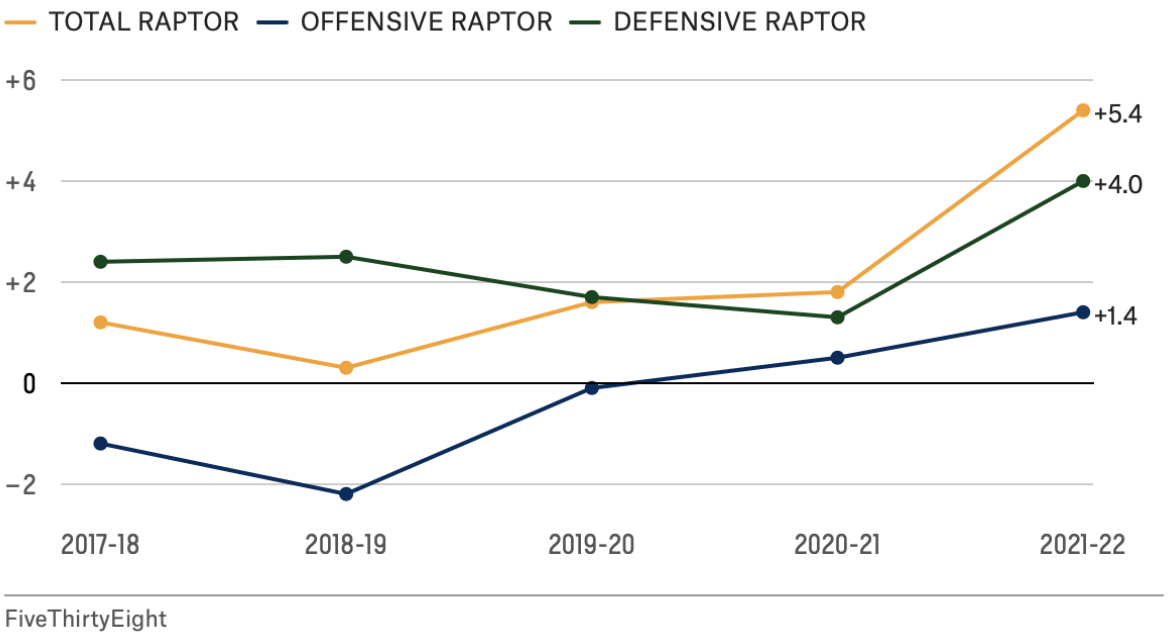
We had a few different motivations for this project. As fans, we follow the National Basketball League on social media and often see statistics and visualizations of these statistics on platforms such as Instagram and Twitter. There are also many websites that are dedicated to exploring statistics in many sports, not just the NBA. One of these websites is www.fivethirtyeight.com which is a website that uses data and statistics to make visualizations and tell data stories about many topics including Politics, Sports, and Science. The inspiration for our project came from the sports section of this website. Below is an example of a visualization they have that tells a data story.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021




This visualization shows the punts of San Diego State punter Matt Araiza. He is one of the best punters in NCAA Football history and this chart shows that. The marks for this visualization are lines and the channels would be the length and thickness of the lines. This inspired us to combine multiple attributes into our visualizations as this one has the length of the punts and the frequency.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

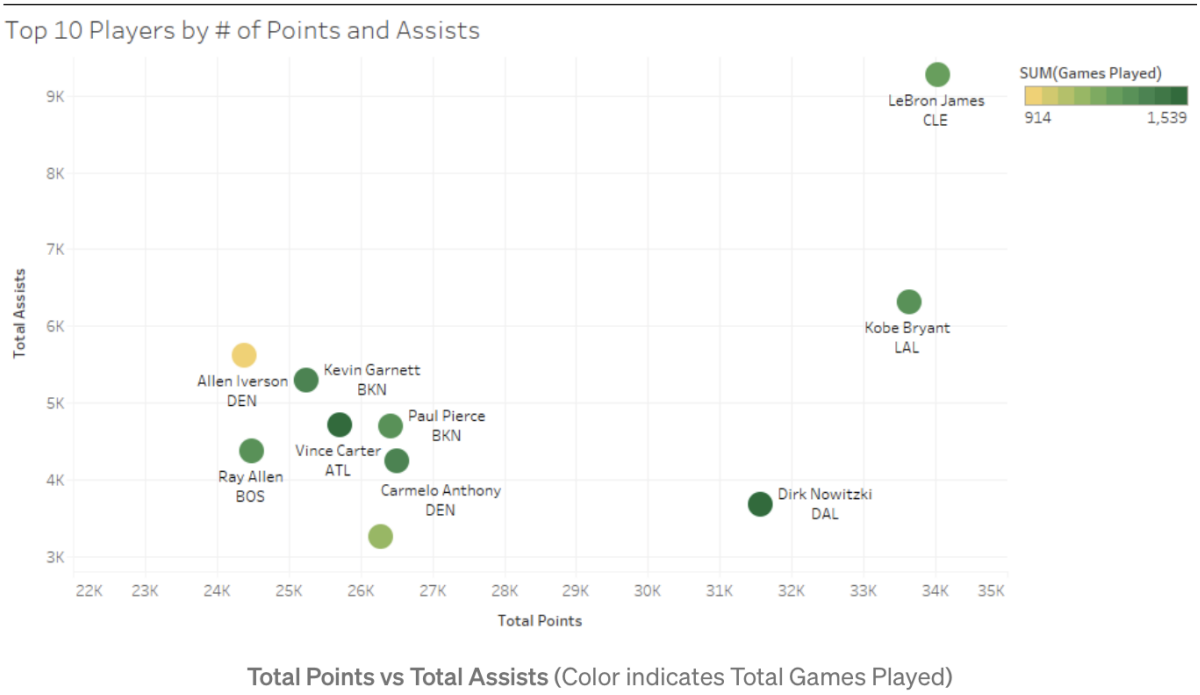



Here is another visualization that is from the NBA. This visualization shows the Offensive, Defensive, and Total Raptor rating for Royce O'Neale. This type of line chart is motivation for us to make line charts that show a player's statistics or ratings over time. It also gave us the idea to graph multiple different statistics in a chart, which is important for painting a complete picture of a player's performance.

The following website was also inspirational for us in brainstorming visualizations for our project: https://towardsdatascience.com/10-unique-visualizations-of-the-nba-b981cfdb78bf.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

## 5. Which players have the most Points and Assists?




This chart was very inspirational for us because it shows how you can use a chart to combine multiple different attributes. It uses horizontal and vertical position as channels and then it also uses color to represent a different attribute. Being able to have points as horizontal position, assists as vertical position, and then games played by color all in one chart is a very powerful visualization tool, and one that we wanted to replicate in our own project.

We were also inspired in class by the idea of showing all of the data in the initial chart and then giving the user the ability to drill down into different areas and create aggregates. We learned in class that it was important not to filter the data because then we are controlling what the user sees instead of letting them do their own learning from playing around with the charts.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

Instead of trying to tell the user what they should be seeing, we want to equip them with the tools to find trends and interesting statistics for themselves.

# Questions

One of the main questions we wanted to answer was “How have player performances changed over time?” This was our main question because we wanted to focus on the players of the National Basketball League and statistics such as points, rebounds, and assists. We also wanted to answer the questions of “How have player salaries changed over time?” and “How are the salaries of top players compared to the lower tier players in the NBA?” This relates to our objective of connecting performance to pay in the National Basketball League. An important question to answer with this project is “How does player performance in different statistical areas correlate to their salary?” Along with this, we wanted to compare the team's salary to the salaries of the players on the team. This made us ask the question, “How are teams salaries distributed across their players?” We originally wanted to ask a few different questions but they changed over the course of the project. We originally wanted to ask questions about how the NBA has changed over time, but we decided to focus more on current players and their statistics. This was a combination of being more interested in current players, and also it gave us a better opportunity to show all data points on our main chart and drill down on these visualizations. We also originally wanted to focus on locations across the countries such as colleges and hometowns of the players, but decided that this wouldn’t be very effective in supplementing our main questions.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

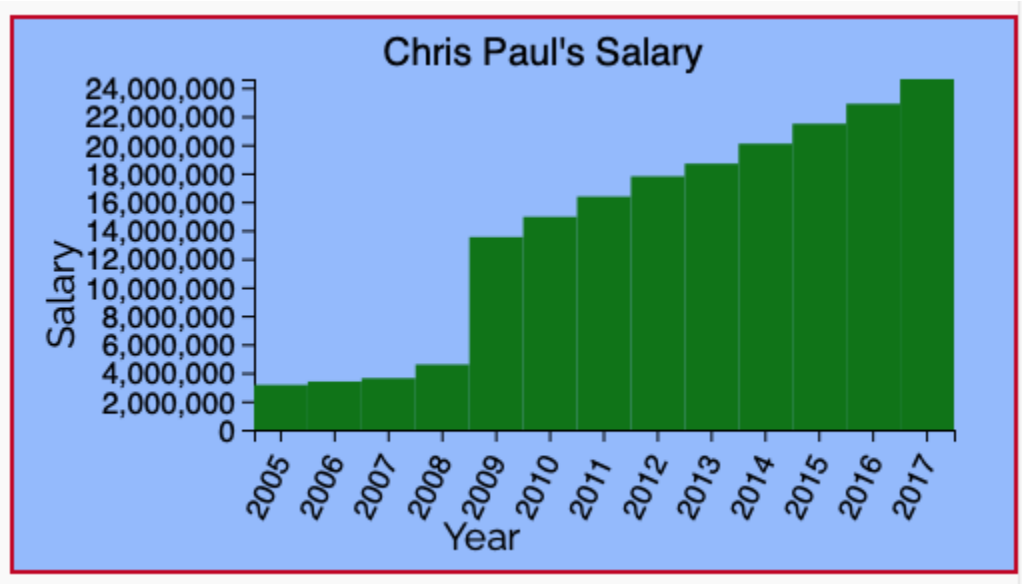
# Data

The source of our data was www.kaggle.com. We started with the following dataset https://www.kaggle.com/wyattowalsh/basketball. This was used for player attributes such as the team they are on, whether they are active or not, and their career averages. This dataset was the source of our main scatter plot. We also had to bring in the following dataset for some visualizations that we wanted to make https://www.kaggle.com/drgilermo/nba-players-stats. This was used for the stats by season for each player. Specifically, this was used for the line chart that shows each player's points per game, rebounds per game, and assists per game. For this dataset we had to do some manipulation to get a player's points per game. To do this we had to divide the points they scored that season by the number of games they played. We thought points per game would be better because some players may score less points over a season if they play less games due to injury or other factors which would be misleading. We also used this dataset https://www.kaggle.com/whitefero/nba-player-salary-19902017 to show the players salaries over time. We also had to do some data manipulation on this dataset to get the salaries for teams over time, which we could not find a dataset for. Instead, we had to go through this dataset and use excel formulas to add up each player on a team's salary each year. This gave us the team's salary over time. These datasets also required some synchronization of team abbreviations and team names. Some tables used different abbreviations for the same team which caused errors in visualization. Also, different tables had different names for players with "Jr.", nicknames, etc.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

# Exploratory Data Analysis

Our original charts we looked at were basic charts with a single player's salary over time. We were able to see that salaries generally increase over time for a player. This makes sense because we also saw that team salaries increased over time. We also made visualizations that allowed us to look at the statistics of a player. We found that players generally have increasing stats over the course of their career.




In the above chart, you can see an example of a player's salary continuing to increase as their career progresses. Another takeaway that we had from this was the amount of years they have played. Generally speaking the best players have the longest careers, which makes sense. You can also see that the highest scoring players have generally been playing for a long time, which makes sense because teams are willing to pay them and keep them on the roster.

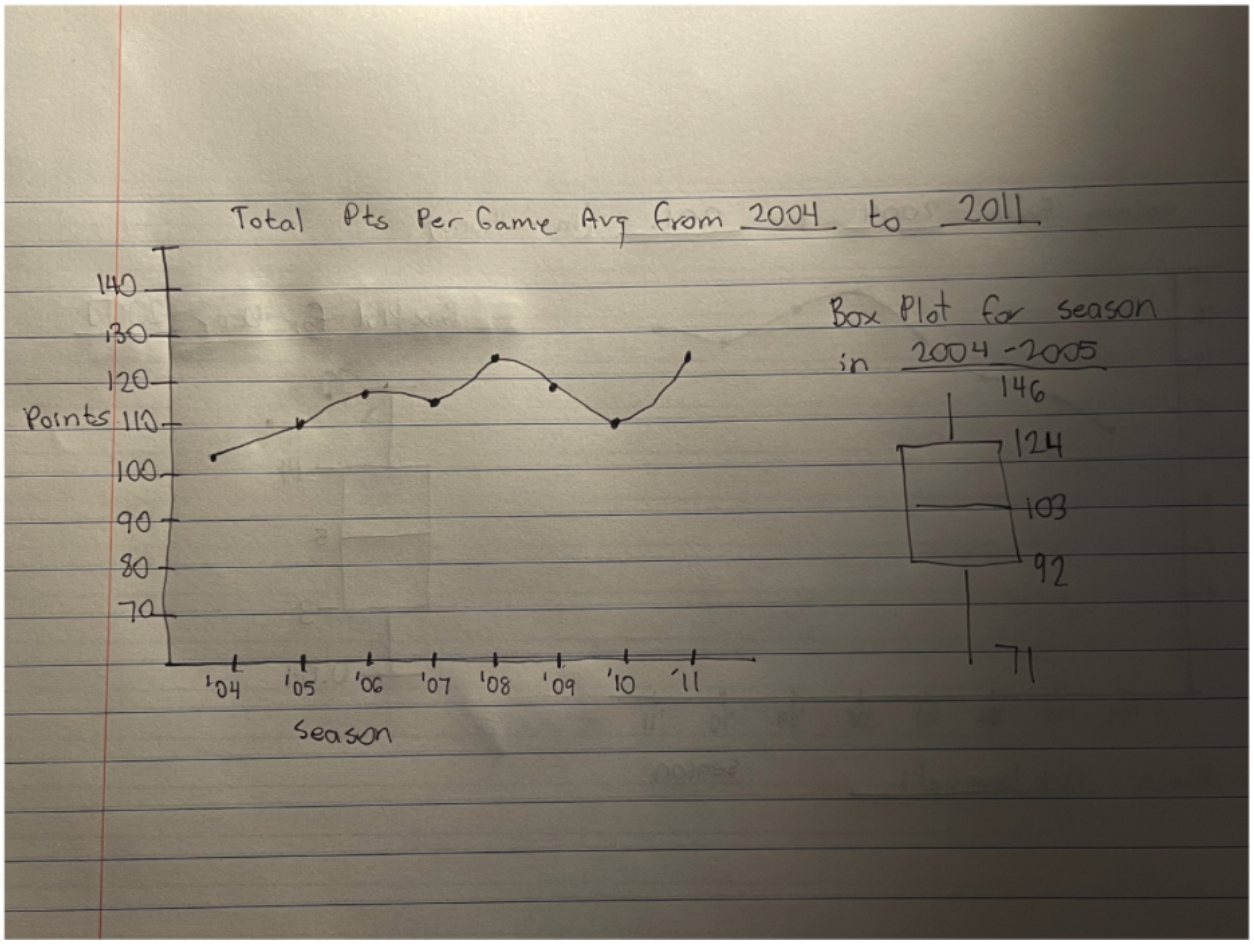
Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021


We also found that the salary of the teams have been going up over time. This makes sense as the popularity of the NBA has gone up.

# Design Evolution

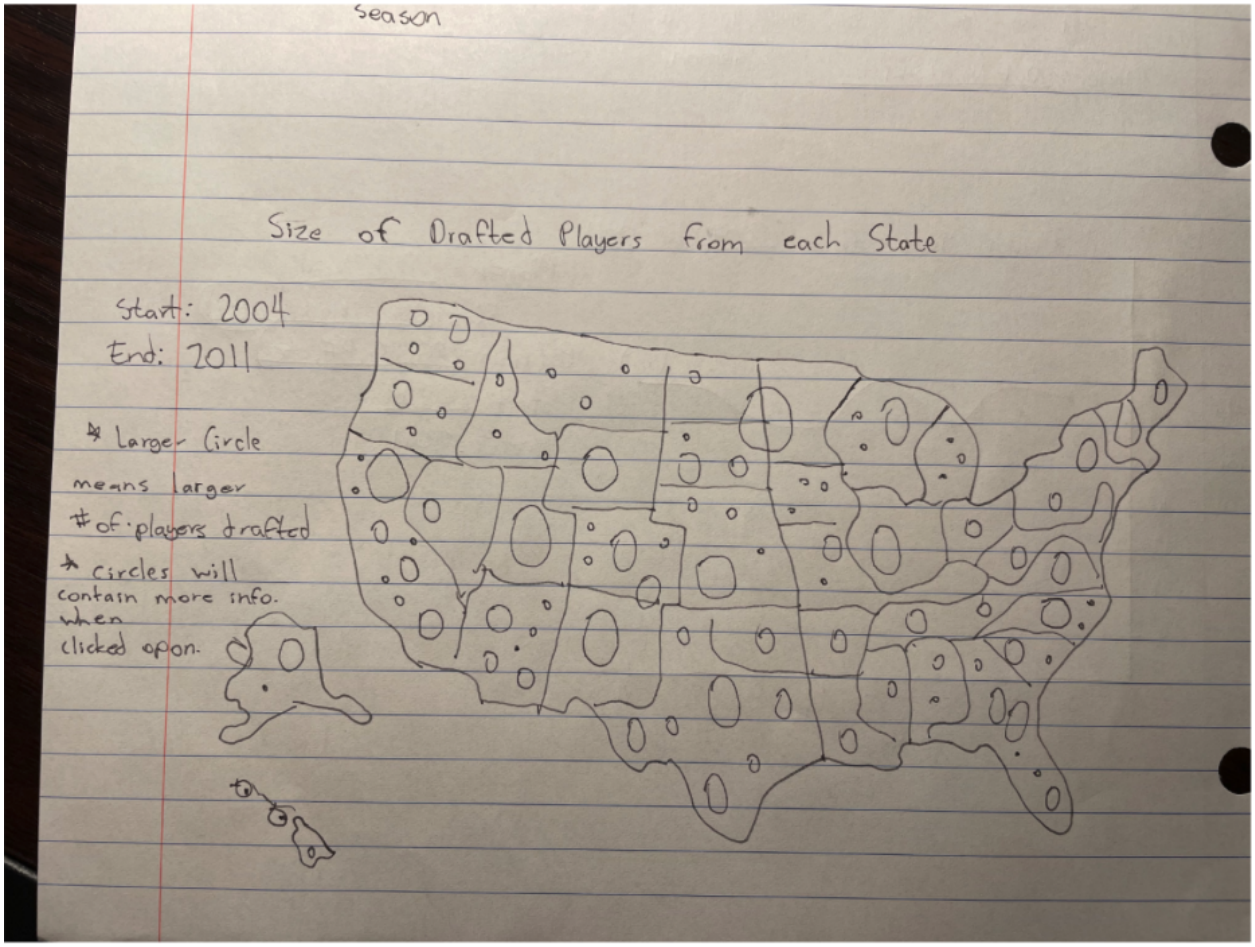
Our initial design was to have different visualizations that were independent of each other and visualized many different questions. We wanted to see the change in statistics over time, so we were planning on having line charts in order to do this.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

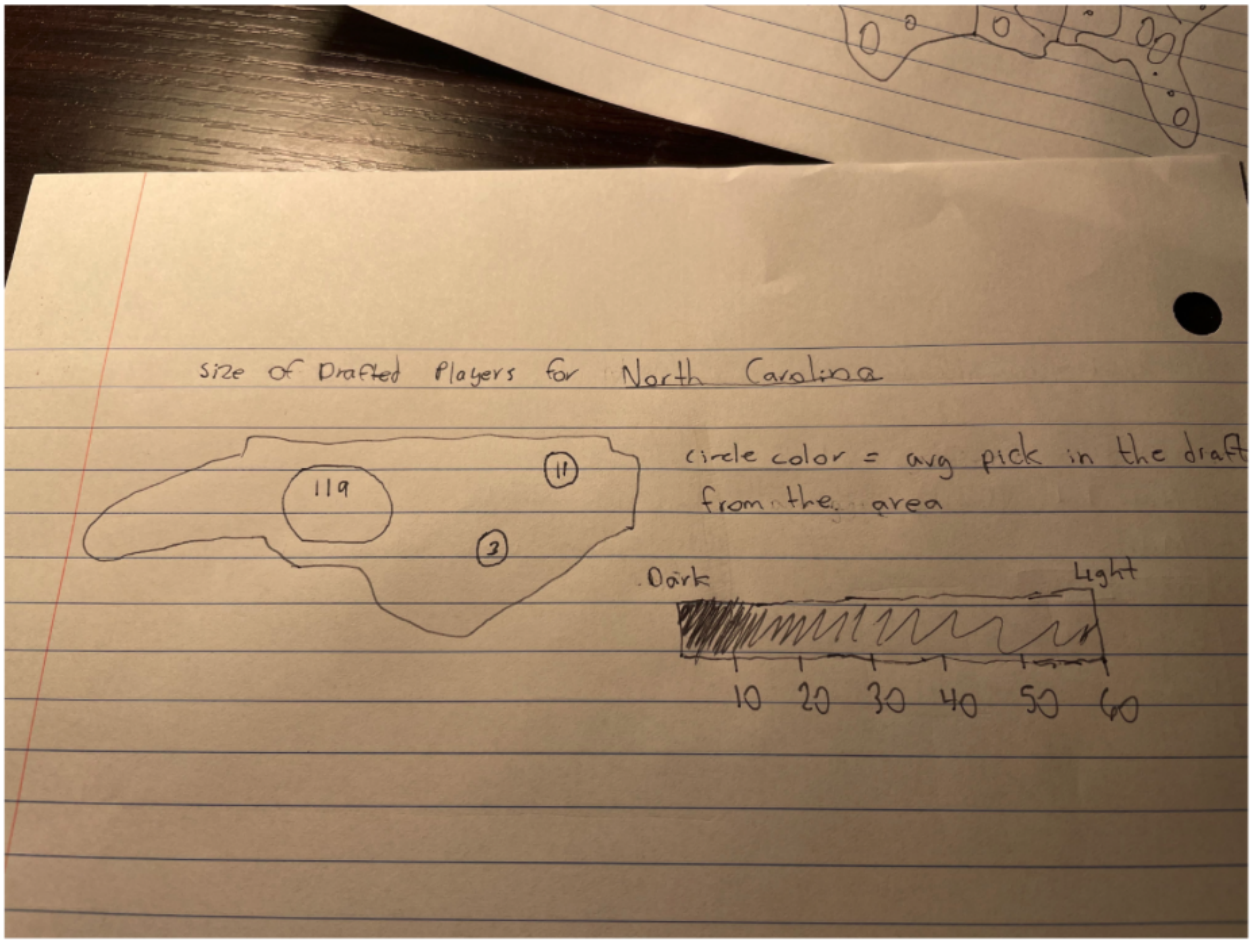



We also made other charts similar to this to show individual statistics and salaries. We also were planning on making a map similar to the following images.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

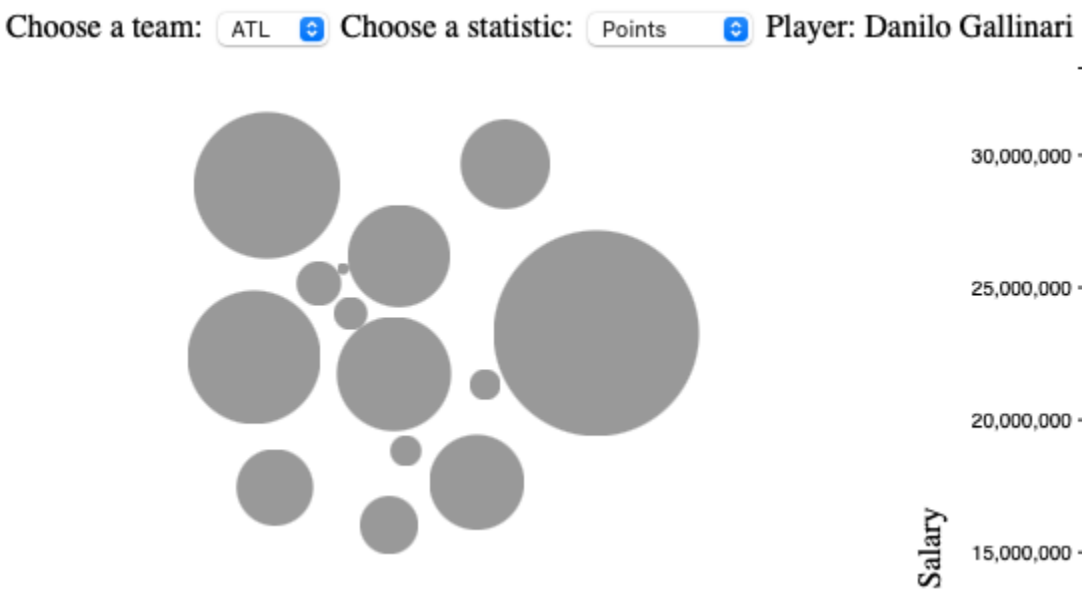


Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021



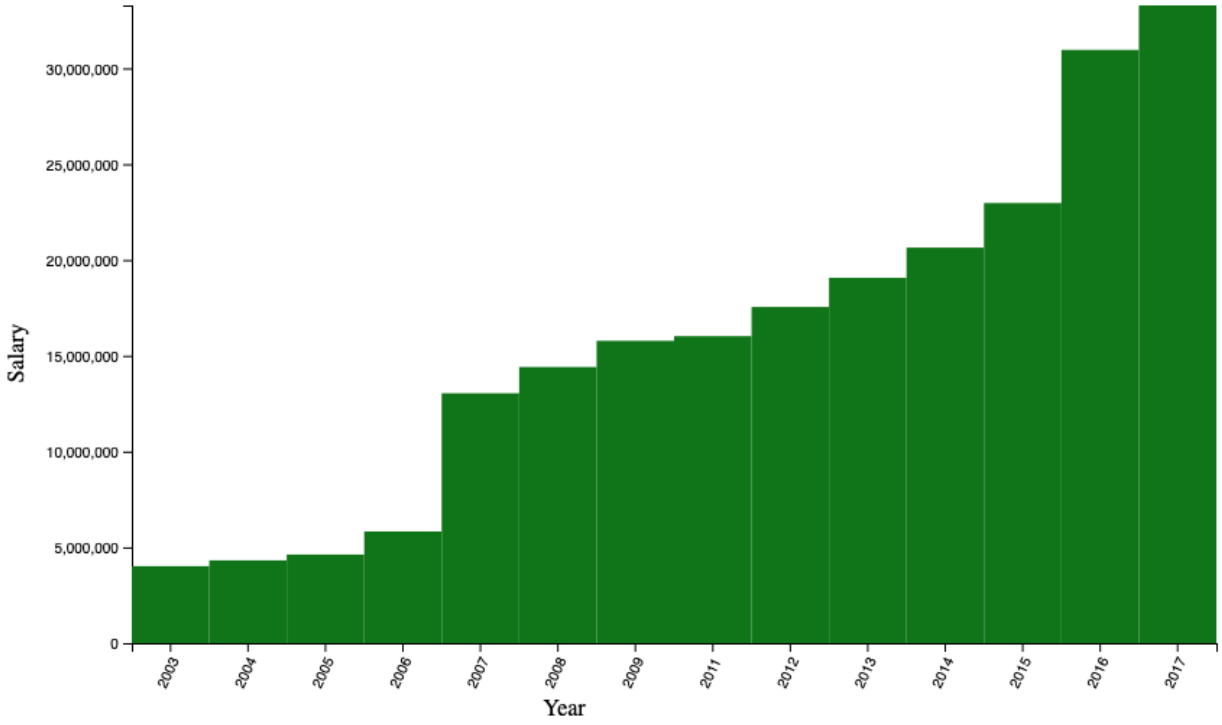

This was intended to show the amount of players drafted from each state and each college. As you can see, these maps and line charts were not closely related to each other so that they could easily have interactions. Because of this, we decided to make one main visualization and then have supplemental visualizations to go along with the main one. The following is our first main chart we tried out.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021




This was a force chart that showed all the players on a certain team as a bubble. The channel was the size of the bubble, which represents the statistic shown on the dropdown. We were able to change each statistic and it would change the size of the bubble. You could also change the team with the dropdown. The problem with this visualization was that it was automatically filtering the players by team instead of showing all data points. This relates back to what we talked about earlier with not filtering data for the user. Also, the position was not significant in representing any data, it was just random. We decided we needed to use position as a channel and make it represent an attribute. It also was not clear what the size of the bubble represented. This all factored into our next design of our main chart.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

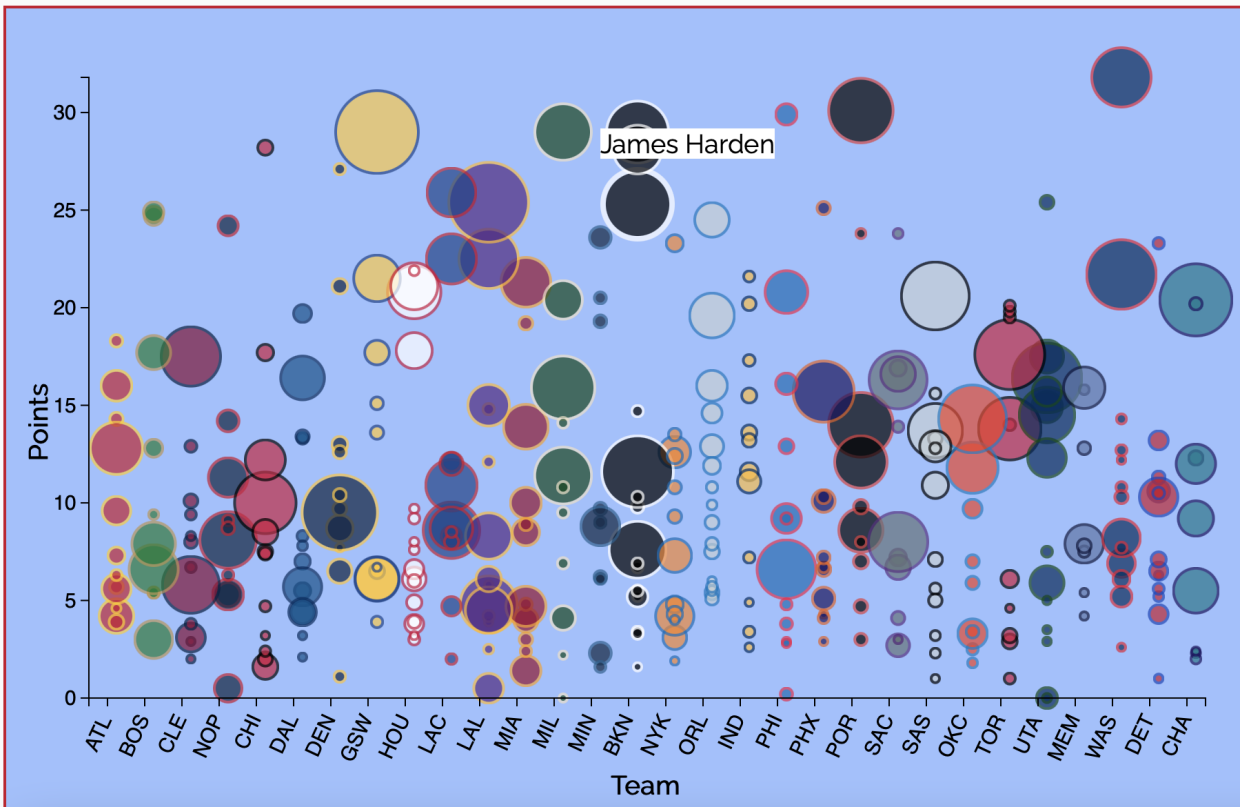



This was the initial bar chart we had to go along with our force chart. When you clicked on a bubble it would change the salary that is shown in the bar chart. This was a good design choice because it effectively shows the increase in salary over time. We needed to add a title though to show who's salary you are looking at. We believe our final design choice of using the scatter plot to show the players is the most effective way to show all NBA players across the league. This allowed us to show three different attributes on one chart, which we're team, salary, and either points, rebounds, or statistics. This design choice also allows the user to view all players in the league and also view them within their team. It also allows the user to drill down into individual players statistics over time and compare them to their salary over time. The user will be able to play around with different statistics and different players, along with being able to view the salaries of entire teams in the NBA.

Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

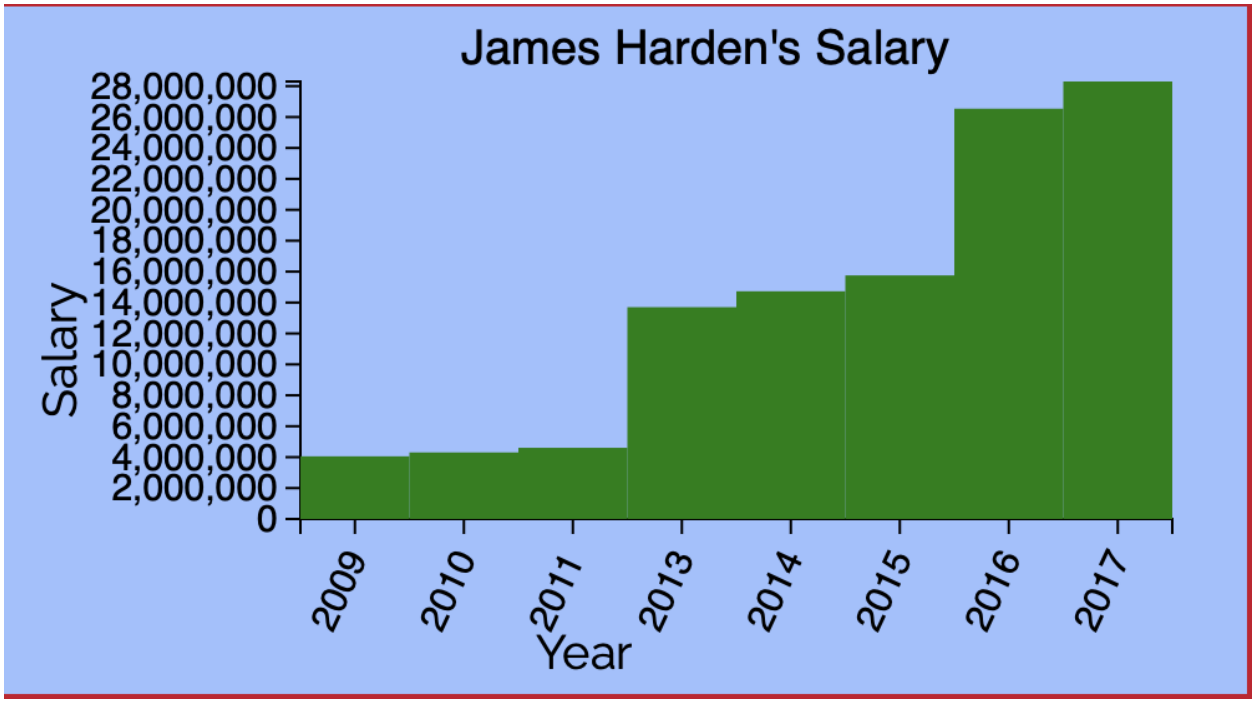
# Implementation

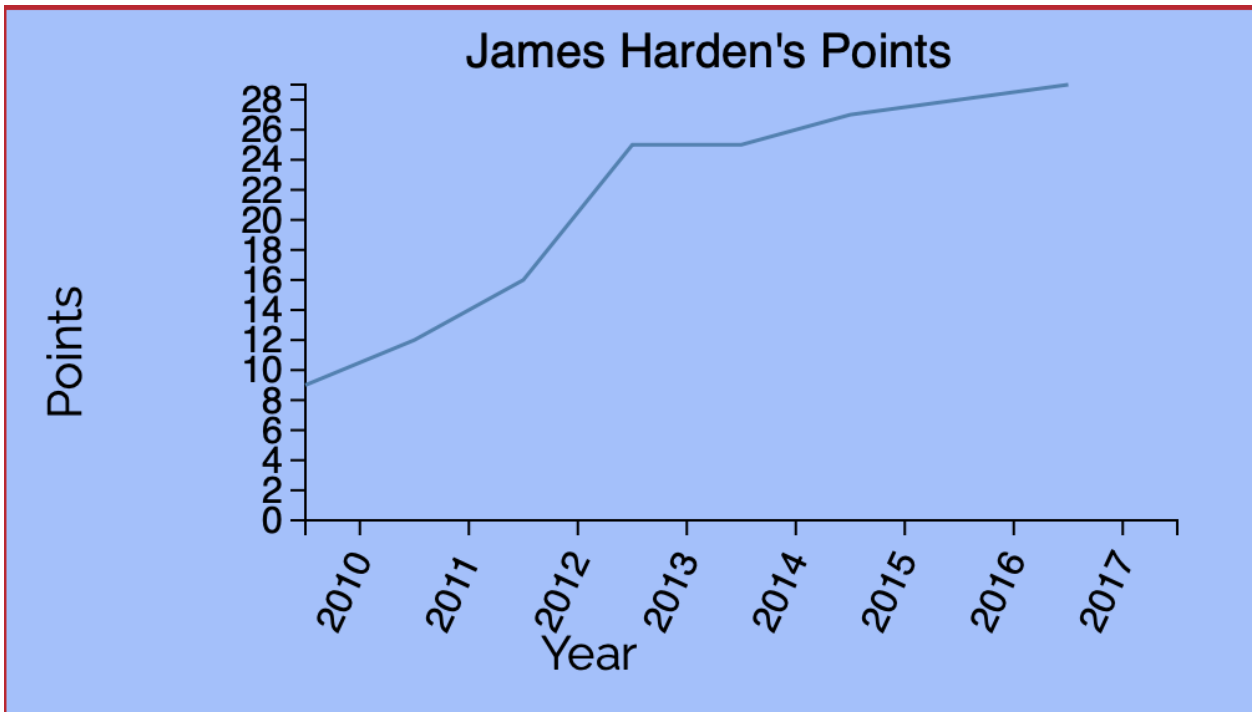
The following scatter plot is the main chart of our visualization. This is intended to show all the players in the NBA grouped by their team. The size of the bubble is a channel that represents their current salary, and the vertical position is a channel used to show their points per game.




This chart serves a couple different purposes. The first purpose is to allow the user to see how a certain statistic is related to the amount of money a player is paid. You can change the statistic between points, rebounds, and assists and see the change in the chart. It also allows the user to

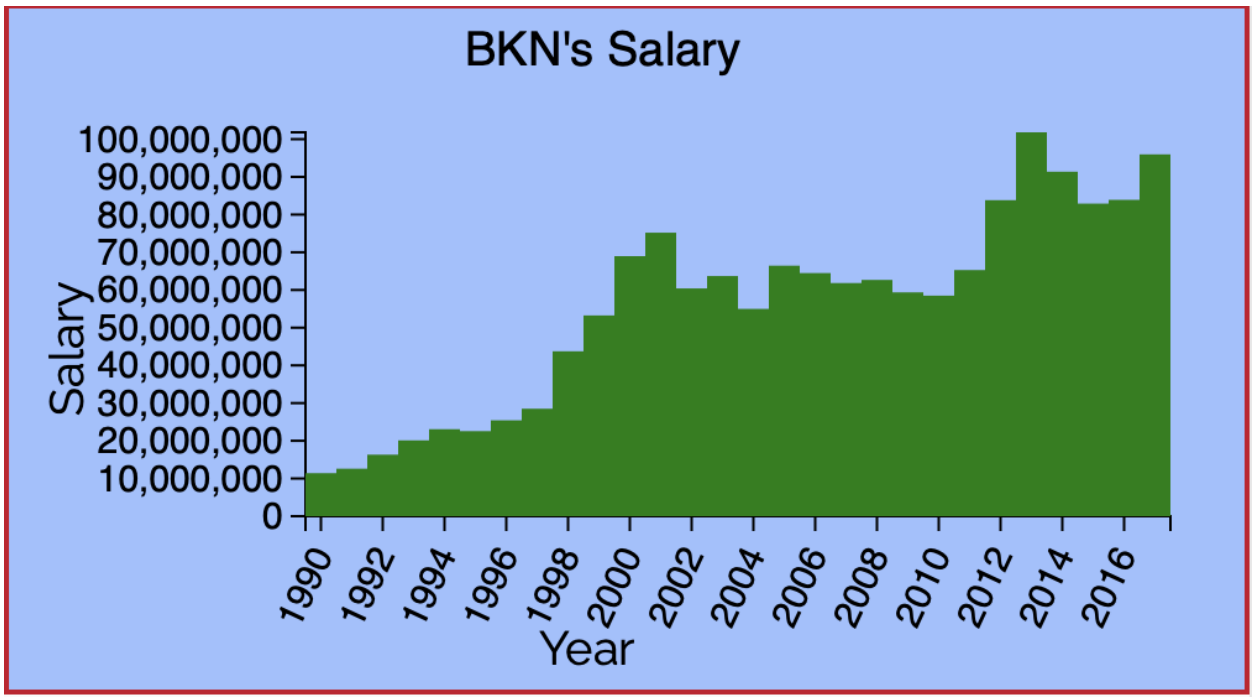
drill down into a specific player or team. If you click on a player's bubble, you will see the bar chart and line chart on the right change. These charts are shown below.



Final Project Report
CPSC 4030 001
Cole Earl, Tyler Piel, Raj Patel
December 5, 2021

These charts work together to show the increase or decrease in a player's salary or points. If you change the scatter plot to assists or rebounds and click on the player's bubble, the line chart will show you that statistic. If you are to click on the team's tick mark on the bottom of the scatter plot, the bar chart will change to show the team's salary over time. This is shown below.




# Evaluation

Throughout the design of our project, we learned quite a bit about the players of the National Basketball League and trends within the NBA. The first thing that we learned is that NBA salaries mostly increase over time. This was answered with our bar chart that shows salary over time. We also found that players who have the highest statistics of points, rebounds, and assists have the highest salaries in the league. This connection suggests that there is a correlation between the highest salaries and team-leading statistics. This is shown on the scatter plot where

the biggest bubbles tend to be positioned higher up on the graph. You can also see through our visualizations that as a player's performance increases throughout their career, they earn more money. This is shown through the combination of our bar chart and line chart. Another interesting finding is that the NBA seems to value points scored over any other statistic. You can see on the scatter plot that on the points visualization the larger bubbles are at the top a lot more than on the rebounds and assists visualizations. As far as exploring the team's salaries, we found that teams with higher salaries seem to have better players. You can see that larger market teams have more players on the top of the graph for statistics, especially points. In order to improve our visualization, we think that we could add a feature to compare players on the same chart. We also could add other metrics such as all star appearances for players. Another interesting addition would be to add a chart that shows a teams performance and how this relates to player performance and salaries.