# Week 4 Assignment – Visualization

By Zach Adair

Regis University

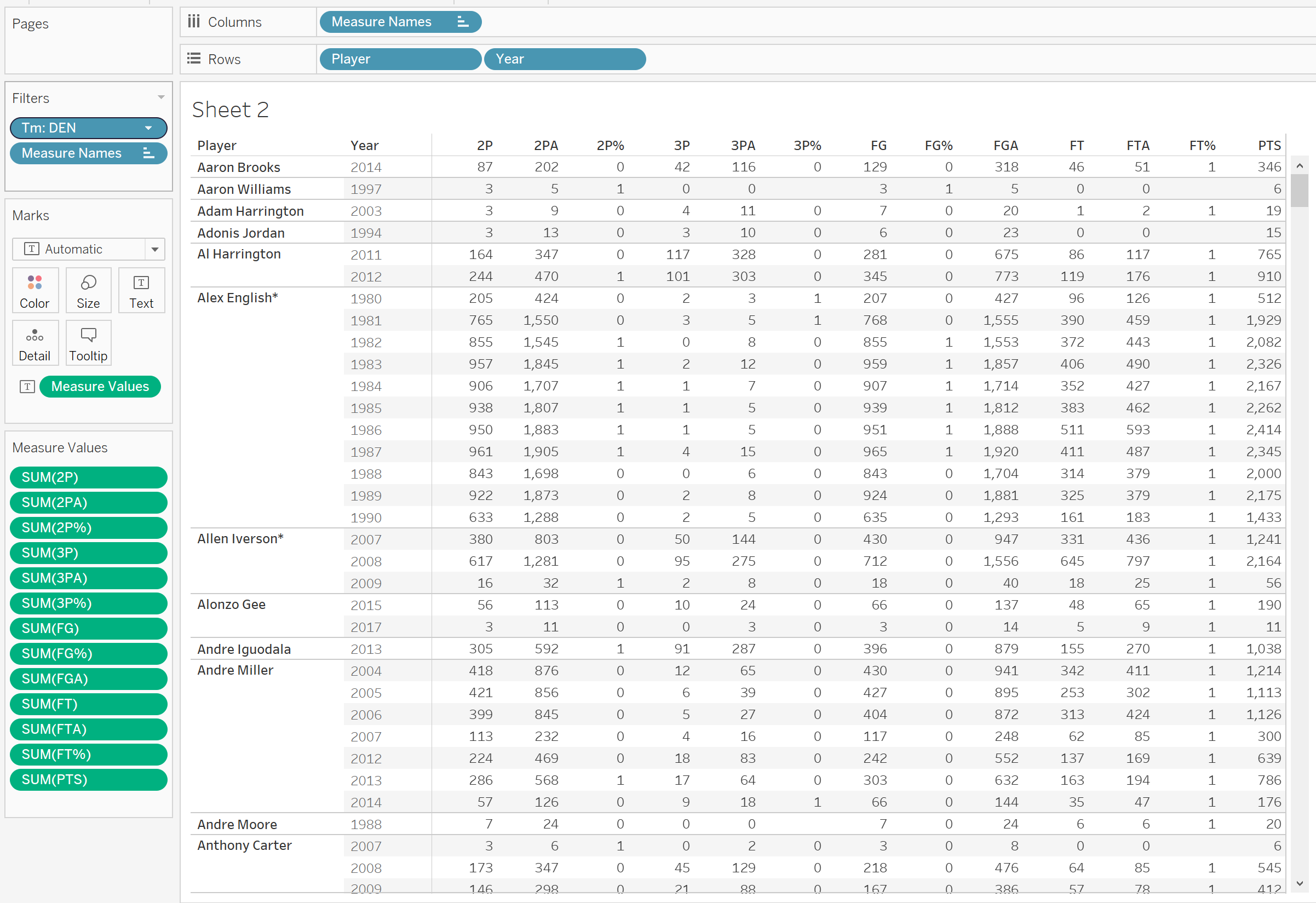
MSDS 650 – Data Analytics

## Introduction

This week the assignment calls for me to find and analyze a dataset using the visualization application tool Tableau. Tableau is a great tool for helping visualize what data represents and what information can be pulled from it. The dataset I retrieved is from the National Basketball Association and is the player statistics from 1950 to 2017. My goal in this analysis and walk through with Tableau is to discover a little bit more about some players I find historically relevant to the game of basketball and possibly bring some context to how good these players were during their time in the NBA.

## Text Table

I’m not counting this text table as one of my visualizations but I like to help me determine what kind of visualizations I would like to do. With this data set it appears the easiest ones to use would be bar charts and line charts. With individual players I can create a heat map to highlight the high and low points in their career or when I compare multiple players careers against one another.



Looking at this data it should be pretty easy to find trends in the data based on points and the types of ways they scored their points, the other visualizations can help easily show data and gather information for specific players.

## Bar Chart

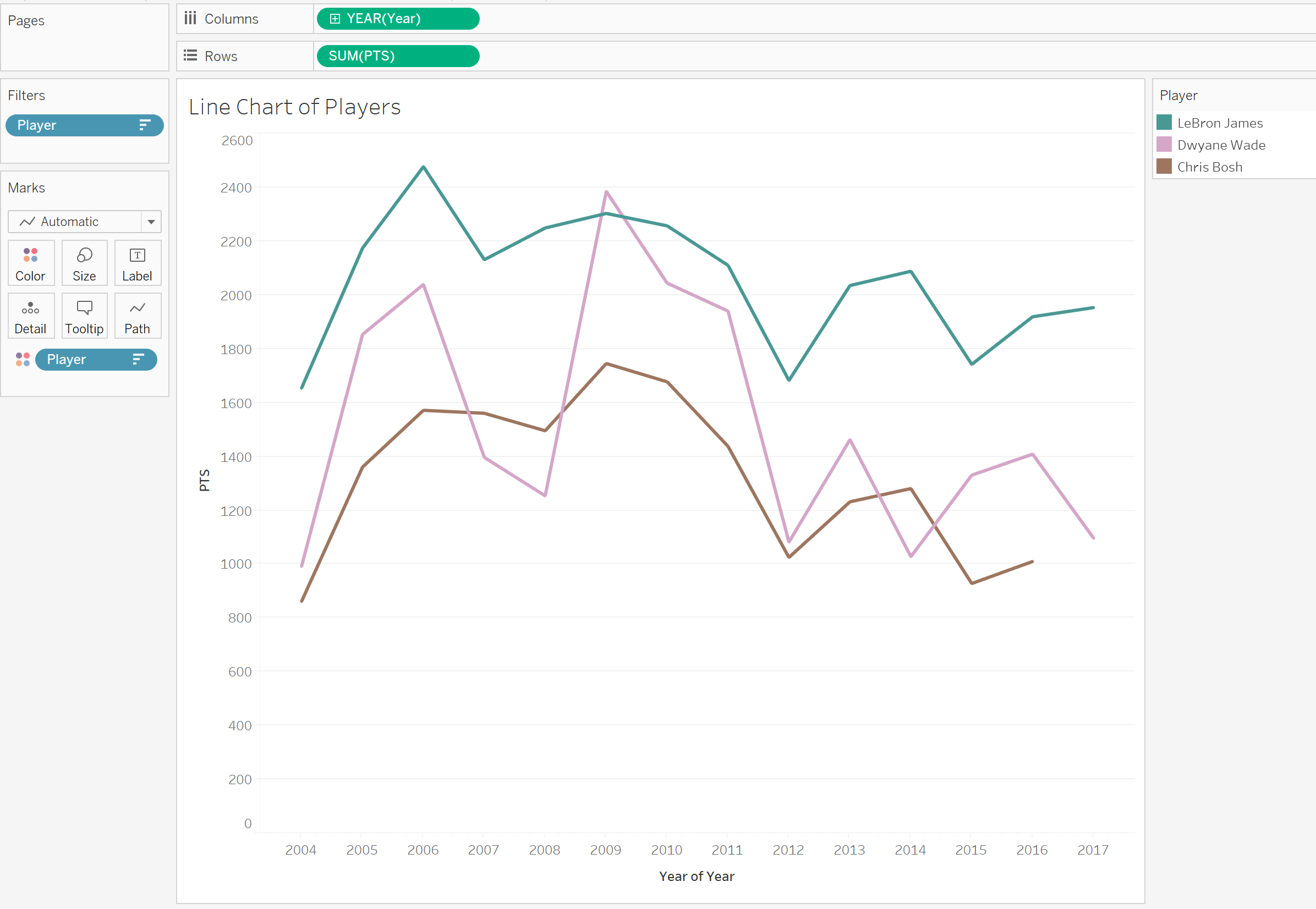
This bar chart is designed to examine the volume of points by players within their careers.



From our bar chart we can see the career points for players, sorted from highest to lowest. This is a great representation of the most points scored by an individual in the NBA. The visualization helps show how much of a difference from one volume of points scored by a player to another. It is easier to notice the differences by a visual representation than by just seeing the two numbers next to one another.

## Line Chart

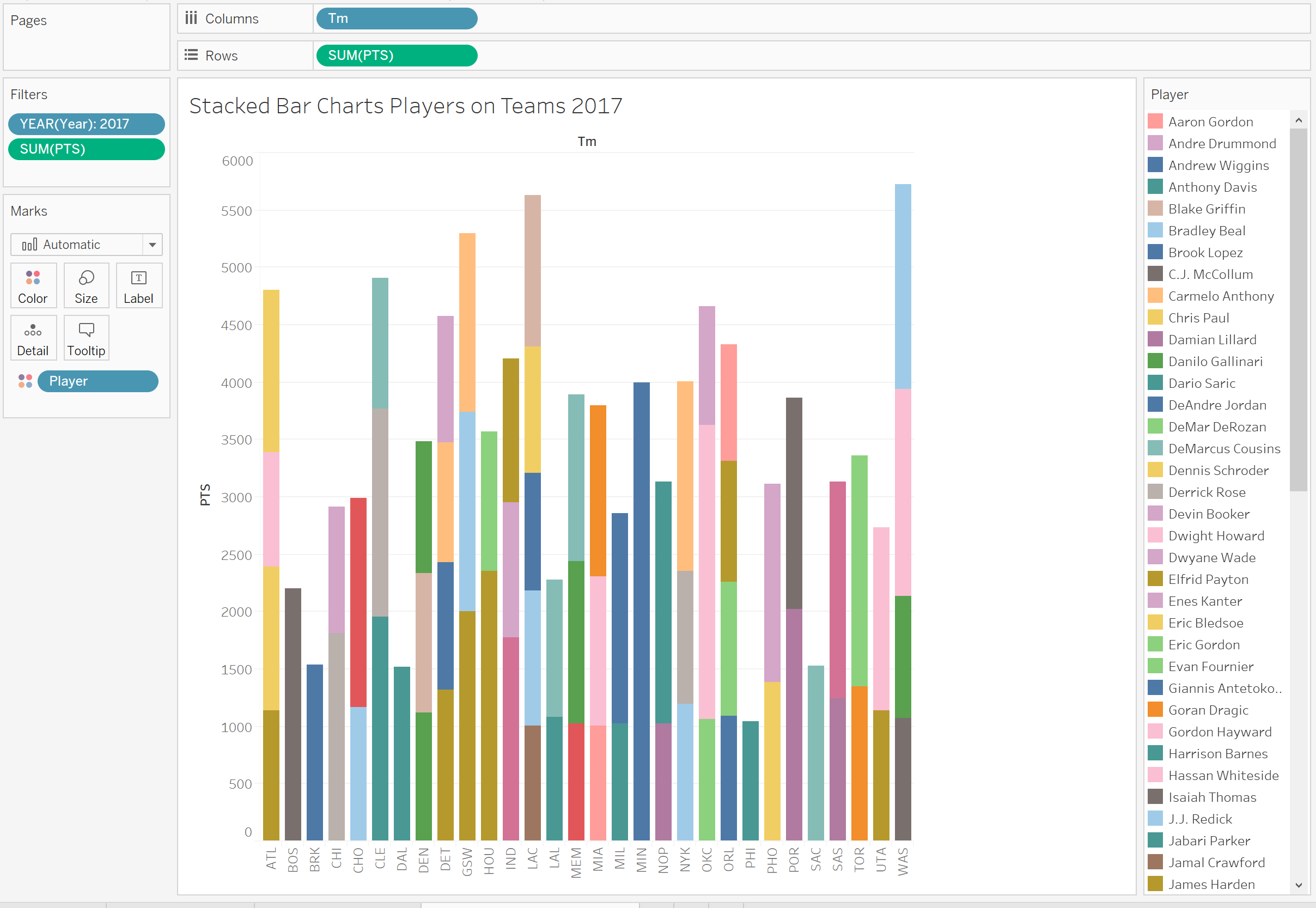
This line chart is designed to represent the points scored by Lebron James, Dwayne Wade, and Chris Bosh and the number of points they scored over their career from year to year. They all came into the league at the same time so we should be able to spot where these players hit their peak performance and where they dropped off towards the end of their career.



In the line chart we find that each of the players were relatively productive and consistent throughout their careers. Lebron James is consistently performing the better than Dwayne Wade and Chris Bosh in points scored in a season, but Dwayne did have one year where his numbers spiked and he actually was had the best scorer in 2009. Each of the three players had a spike in their performance after their first year and there really wasn’t a dip in any performance until 2012, which was a strike shortened season, but the downward trend of their numbers stayed down in that area for the three players, even though there was a small bump once they want back to normal length season in 2013.

## Stacked Bar Charts

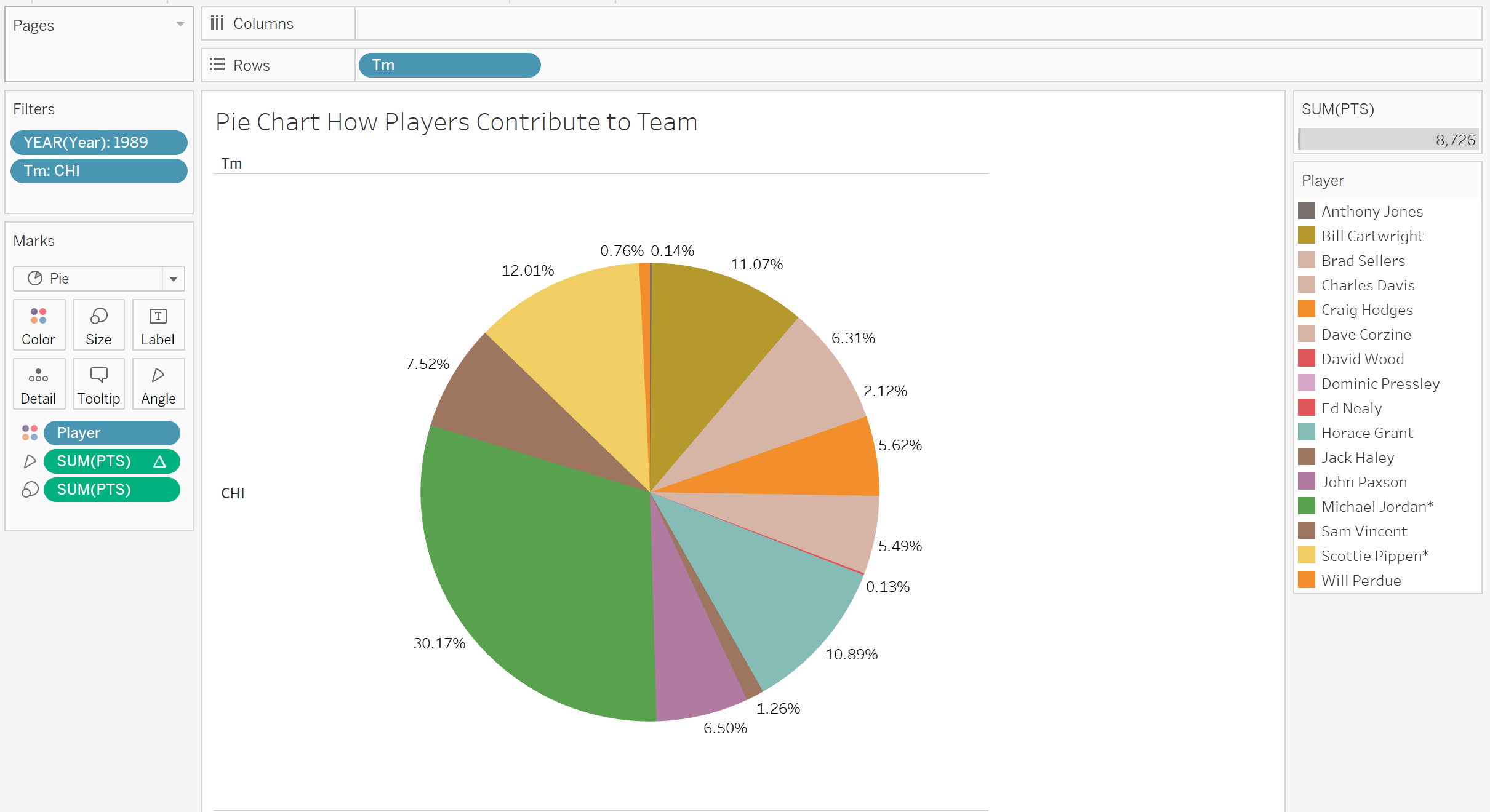
From what I’ve seen already in this data, it appears the most productive scorers year after year are the ones who scores at least 1000 points a year. So in my stacked bar chart I created a filter to only show me the scorers who got at least 1000 points and created a stacked bar chart by teams with the players on those teams that scored at least 1000 points in the 2017 season.



From the data we see that out of 30 teams all but three of them have at least 2 players score 1000 points in the 2017 season with the 1000 points scorers for Washington have the highest volume of points by their 1000 point scorers, the LA Clippers having the most individuals score 1000 points with five players, and Philadelphia having the lowest total and tie for the lowest number of players score over a 1000 points with just one. The stacked bar chart helps show the amount of high performers these teams have as well as how reliant these teams are towards those players. I originally did this graph with no filter on the amount of points a player can score but quickly changed that when I found the graph to be very busy and cluttered with 12-20 players on each team making the graph hard to read and diagnosis meaningful information from, which is an issue an analyst can have with a stacked bar chart. I know from my experience that using a stacked bar chart it is important to not have such a cluttered visualization and make sure it is readable for the audience you are creating it for.

## Pie Chart

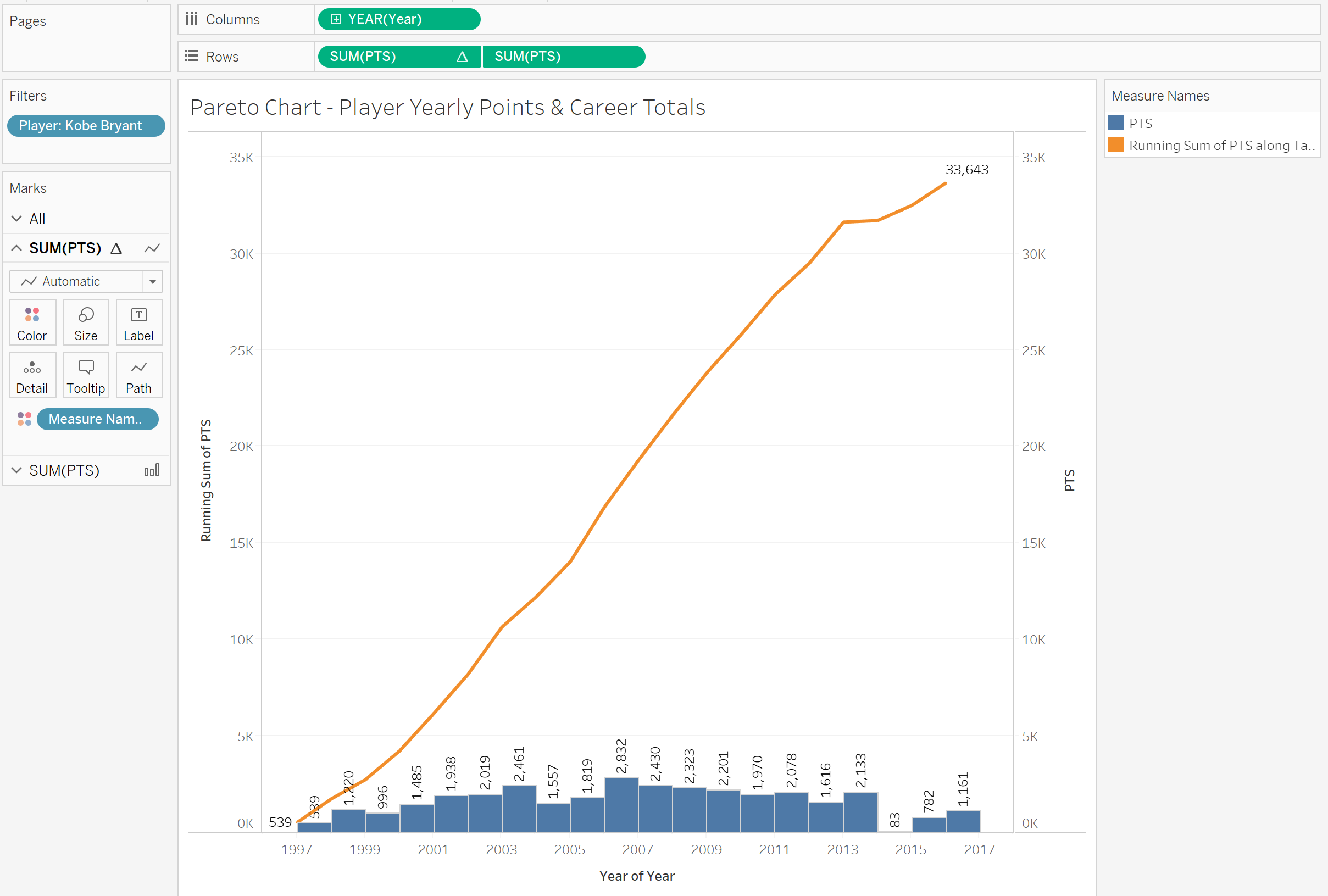
My next visualization I am going to use to analyze the NBA dataset is a pie chart, a pie chart is great to use when you need to describe components that represent parts of a whole. In the pie chart below I filtered by the year 1989, one team the Chicago Bulls, and again were looking at the total points scored by the team and the percentage those points came from by player.



In the pie chart we see the percentages of total points scored by individual players on the Chicago Bulls in 1989. Michael Jordan was overwhelming the biggest points contributor this year tallying over 30% of the total points scored in the year with the next highest contributor not even outputting half of Michael’s point total at 12.01% by Scottie Pippen. Pie charts can be a great representation of what goes into certain aspects of data. It divides up the whole and shows where the major and minor contributions which make up the whole come from and shows, in this situation, who are the most impactful players to the overall scoreboard.

## Pareto Chart

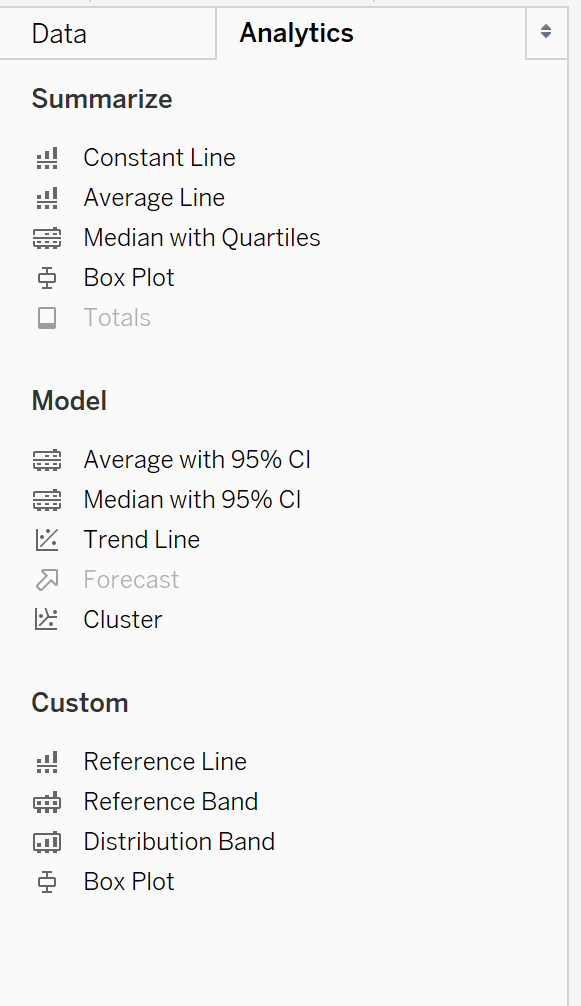
My final visualization is a pareto chart, which combines a line chart and a bar chart. In the graph below I picked one specific player, Kobe Bryant, and tracked his total points scored over his career. The bar chart tallying his yearly output of points, and the line chart giving a running total of his career points scored.



In my pareto chart there a couple things that must be done to give the visualization the look that you see above. First I had to create a dual axis chart which puts both measures on the same x-y plane, this can be done by just right clicking and choosing the option “dual axis”. The second thing I needed to do is set up synchronized axis as well where both running total of points scored and points per year were measured separately but on the same scale overall. This allows us to imagine if we took our final point of our running total and compared that to all our bars, we can visually represent them if stacked together would reach that final total points tip which helps represent what the total number of points scored for the entire career. A pareto chart is terrific at simultaneously showing the total number of points scored in a season by a player throughout their career while also showing the year by year output which contributed to that final tally, while also representing the years of excellence that player had and also show where they struggled and where they perhaps started to fade in performance as their career came to an end.

## Summary

Tableau is a great tool for allowing an analyst to visually represent a data set and give its audience a visual of what information can be pulled from the data. In my current job I get to use Tableau a lot so I know quite a bit about its capabilities and how useful of a tool it is for an organization driven to make decisions based on analytical research. The five visualizations I used are common in descriptive style analytics but there is also capabilities within tableau that allow a user to deploy diagnostic style analytics and even predictive analytics under the analytic pane.



With all these capabilities Tableau is rapidly coming a common tool to use in most analytic houses and is terrific for taking a dataset and making it easily digestible information for users who are not as analytically savvy.