

Week 7: Final Project Proposal

Title: TBD

Abstract: The ultimate goal of our final project is to gather new insights on the NBA—players, teams, franchises, etc. We want to find unique relationships/correlations within our data we wouldn't have found otherwise (i.e newspapers, tv, media). Some types of questions we may want to answer include, "Who scored the most points in 2016?", or "Which team has had the most growth between 2015 and 2020?" Given the extensive amount of data we have encountered, we want to be very specific and comprehensive. Our aim is to reach a wide audience, where anyone can understand our findings, regardless of their interest in basketball. To start, we're planning to use a Kaggle dataset; this dataset has information ranging from win percentage, ranking, player name, position, team conference, etc. In terms of design, we want to make our visualization very creative and comfortable for people to interpret. We want to tell a story with what we've found.

Week 8: Team Agreement & Detailed Project Plan

Team Agreement

- We will communicate via text, and have weekly in-person meetings on Thursdays at 4:30pm at the Smith Campus Center.
- Code will be written individually and collaboratively. All team members should be involved with the technical aspects of the project.
- Final design decisions will be discussed among all members; fair compromises should be made when necessary.
- Work hours should be split as evenly as possible (actual task output may differ based on an individual's ability / previous experience). This ensures not only fairness, but also learning opportunities for everyone. We will keep each other accountable so that one person does not work too much or too little.
- We will keep record of our progress as well as pending tasks on this page.
- Work will not always necessarily be done together in person, but good communication via text is expected in a timely manner. Work may be done remotely as long as collaboration and communication are done well.

Signatures: Anthony Morales, Sal Blanco

Date: 10/23/2022

Detailed Project Plan

- Basic Info
 - Project Title: TBD
 - Team members: Anthony Morales, Salvador Blanco
 - Team email addresses: anthonymorales@college.harvard.edu, sblanco@college.harvard.edu
 - Team Name: Los Cafeteros
- Background and Motivation
 - We are interested in this topic because of our love for sports; basketball has become a culturally defining topic in the past decade, and with that stature have come an influx of statistics which could be used to create valuable insights. Websites such as Kaggle and StatMuse have compiled years worth of data to provide NBA fans with answers to unique questions, such as “who is the best scorer on a Tuesday?” Our team would like to investigate this data and come up with some thought-provoking questions to share with the class.
- Related Work
 - Anthony has worked on a sports-related project in the past, investigating the correlation between team spending and team success in the Premier League (among other questions). This experience led us to think about how we could extrapolate this type of thinking to the NBA world. The NBA has changed drastically within the past couple of years, ranging from a dramatic increase in salary cap, to team valuations, and even “once in a generation talents” like 18 year old Victor Wembanyama (who is projected to add >\$500M dollars of value to the team that drafts him)! The cross-section of financial data, in-game statistics, and more is what we would like to play with in order to complete this project.
- Data
 - We are collecting data from nbastats.com using an open-source API (https://github.com/swar/nba_api), kaggle (<https://www.kaggle.com/datasets/wyattowalsh/basketball>), and other financial data which we are planning on gathering on our own.
- Data Cleanup
 - We do not expect there to be substantial data cleanup, while there most definitely will be some. For example, in our NBA games data from 1946-2021, we are going to select only regular season games in order to create a baseline for comparison across seasons. We are also considering calling on the NBA API in order to create a complete dataset leading up to the end of the 2021-2022 NBA season. Additionally, it would be beneficial to select seasons in which the NBA had 30 teams, since there are skewed numbers (such as the Boston Celtics

7-peat, and Wilt Chamberlain's career averages) which are a result of the era's competition or lack thereof.

Week 9: Map

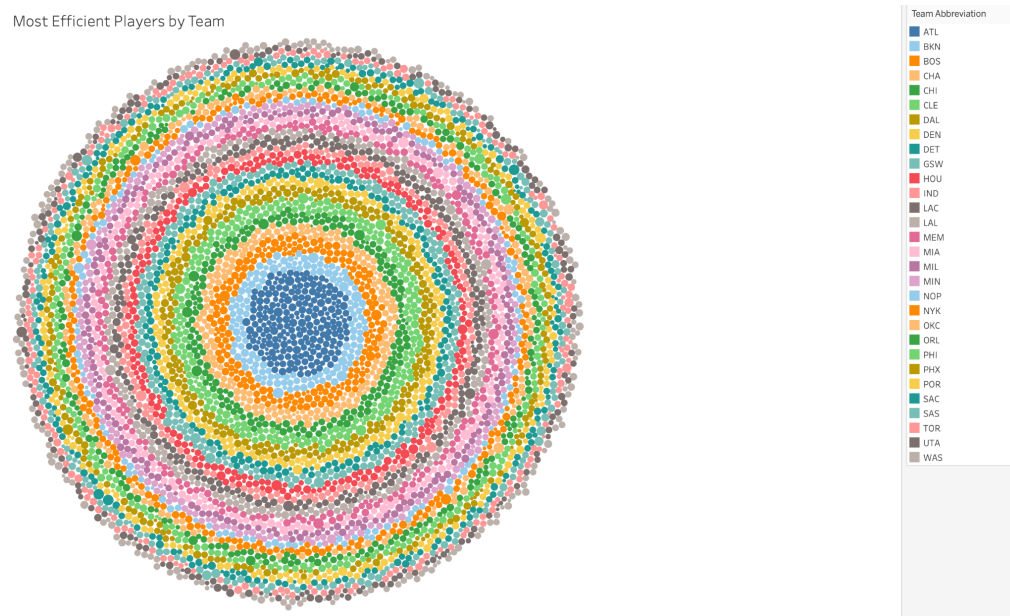
1. Anyone from any sports background should be able to understand and interpret our visualizations. We expect our audience to have a basic understanding of basketball and the NBA. For example, our audience should be familiar with terms such as points, assists, conferences, and playoffs. Our aim is to reach as wide of an audience as possible.
2. As mentioned above, our target audience isn't expected to know much about basketball or the NBA. Ideally, our target audience should have played basketball before. They should be somewhat interested in sports and should have a basic understanding of sports statistics. Moreover, they should also have basic visualization literacy. As we go through the design sprint, we think it's important we make it as simple as possible for our audience to comprehend our visualizations (i.e cognitive overload). However, we'll also include some more complex visualizations to demonstrate our mastery of course concepts and material.
3. Questions about our data:
 - a. Which [western] conference team ended the [2021] season with the highest number of [wins]?
 - b. Which team had the highest win percentage between [2016-2021]?
 - c. Which team has had the greatest number of #1 draft picks?
 - d. Which players have averaged the most amount of points per season?
 - e. Does the plus-minus statistic translate well to team success?
 - f. How has average efficiency changed over time?
 - g. Do the best assisting teams always score the most points? If not, do they have a top-scoring player?
 - h. How do defensive statistics translate to winning?
 - i. Has NBA defense become worse? (e.g. less BLKS, STLS; more PFs). If so, is this a result of a rule change?
 - j. Is there a metric which best predicts team success? (e.g. In recent years, DEFRTG vs. OFFRTG has been a huge predictor of NBA finals teams)
4. Our data is strictly quantitative with the exception of a couple of data columns which are categorical. These columns include player name, player nickname, team name, team name abbreviation, conference, arena name, city, Team Owner, NBA G-League affiliate, General Manager, and Head Coach. This is extremely beneficial to the questions we are trying to answer; the multitude of quantitative variables will allow us to easily see statistical correlations between performance metrics. In turn, this allows for easy comparison of teams from year to year. Paired alongside our categorical variables, we will be able to add more levels of detail and colored legends to provide a visually pleasing and readable experience to our audience.

Sal Blanco: *Graph 1*



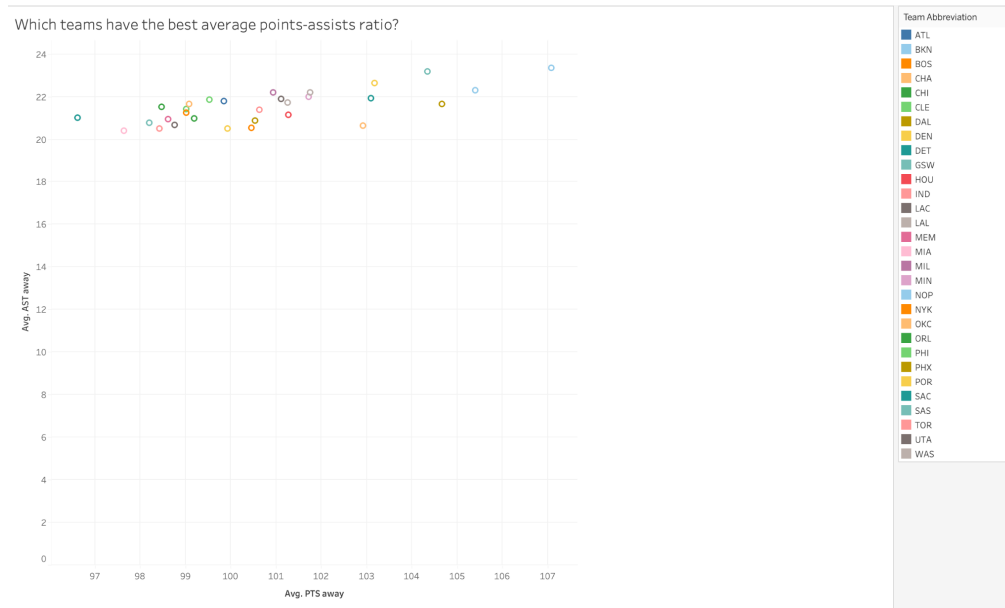
- This graph shows the teams with the highest average win percentage from 2004-2020.

Sal Blanco: *Graph 2*



- This graph shows the players with the highest field goal percentages categorized by the team they were playing for at the time.

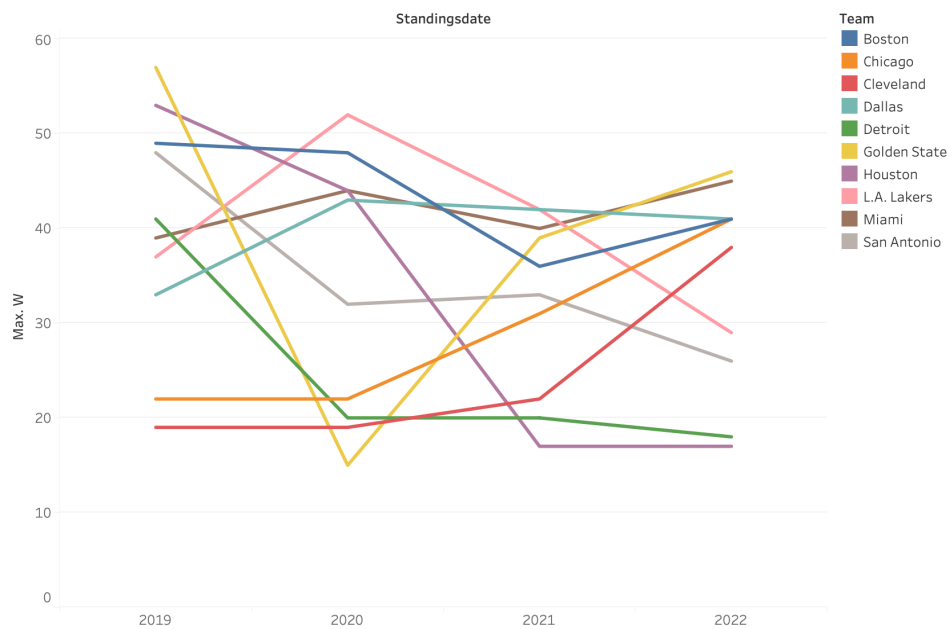
Sal Blanco: *Graph 3*



- This graph shows which teams have had the best average points to assists ratio when competing as the away team.

Anthony Morales: *Graph 4*

Top 10 NBA Teams With Most Wins From 2019-2022

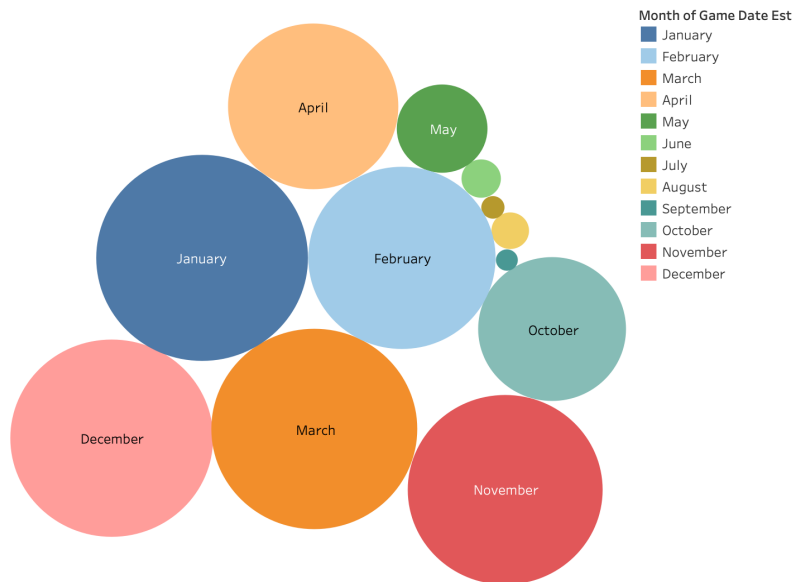


The trend of maximum of W for Standingsdate Year. Color shows details about Team. Details are shown for Conference. The view is filtered on Standingsdate Year, Conference and Team. The Standingsdate Year filter keeps 2019, 2020, 2021 and 2022. The Conference filter keeps East and West. The Team filter has multiple members selected.

- This graph shows the top 10 teams with the greatest number of wins in each season from 2019-2022.

Anthony Morales: *Graph 5*

When Do We See the Greatest Home Advantage?

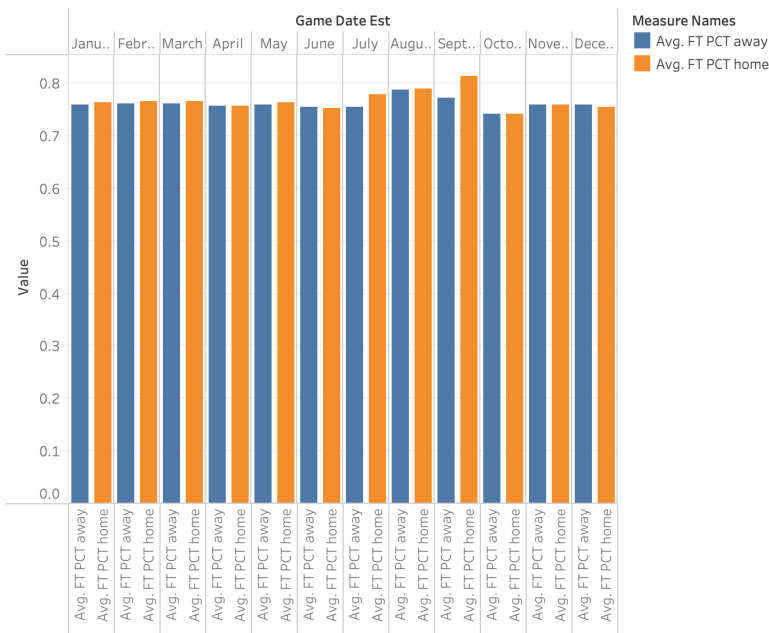


Game Date Est Month. Color shows details about Game Date Est Month. Size shows sum of PTS home. The marks are labeled by Game Date Est Month.

- This graph shows the disparity between months when it comes to how many points have been scored at home between 2004-2022.

Anthony Morales: *Graph 6*

Do Players Perform Worse On Away Territory?



Avg. FT PCT away and Avg. FT PCT home for each Game Date Est Month. Color shows details about Avg. FT PCT away and Avg. FT PCT home.

- This graph shows the non-existent difference between shooting free-throws on away territory as opposed to home. Most averages throughout the years have remained equal. This means that no matter where you are, you are just as likely to convert a free-throw.

Some of these graphs answer questions similar to those we initially proposed. For example, Graph 1 corresponds to question b, and Graph 3 corresponds to question g. However, we were not able to answer all of the questions because some require outside research that is not included within our dataset. For example, we would have to add who won the NBA championship in that given year, who made the playoffs, what the playoff bracket looked like and what were the outcomes of each round. Additionally, we would have to add other metrics such as DEFRTG and OFFRTG which, in the past couple of years, have become very useful when tracking team performances. Additionally, we are planning on adding more data using the NBA API. We ran into some technical difficulties when importing our data; one of our sources is only compatible with SQLite editors. As a result, we are missing some data which could help us answer the questions listed above. What makes some questions better than others is that they provide insight on topics which have been left unanswered. For example, coming up with a new metric to better understand “what it takes to win” in the league would be extremely valuable in NBA research. While we are not trying to reinvent the wheel, it would be nice to come up with some visualizations which are never before seen rather than investigating correlations which have been seen previously.

Week 10: Sketch, Decide, Storyboarding

Title:

- What Factors Can Be Mostly Attributed to a Team's Success?
- How to Create a Successful Team in the NBA
- What Does It Take To Create a Championship-Winning Team in the NBA?
- A Deep Dive Into the Success Behind the San Antonio Spurs

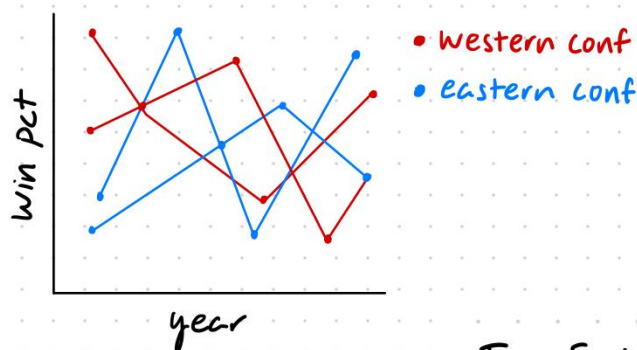
Abstract: The ultimate goal of our final project is to demonstrate what it takes to be a successful team in the NBA. We want to thoroughly study teams (i.e San Antonio Spurs) that have done well in the past with the goal of finding what factors contribute to their success the most. By doing so, we'll be able to determine certain thresholds that deem a team “successful”. Of course, we want to find unique relationships/correlations within our data we wouldn't have found otherwise (i.e newspapers, tv, media). Some types of questions we may want to answer include, "Does team ownership/management affect performance", or "How many all-star players are needed for a winning team?". Given the extensive amount of data we have encountered, we want to be very specific and comprehensive. Our aim is to reach a wide audience, where anyone can understand our findings, regardless of their interest in basketball. To start, we're planning to use a number of Kaggle datasets; these datasets have information ranging from win percentage, ranking, player name, position, team conference, etc. In terms of design, we want to make our visualization very creative and comfortable for people to interpret. We want to tell a story with what we've found.

Questions about our data:

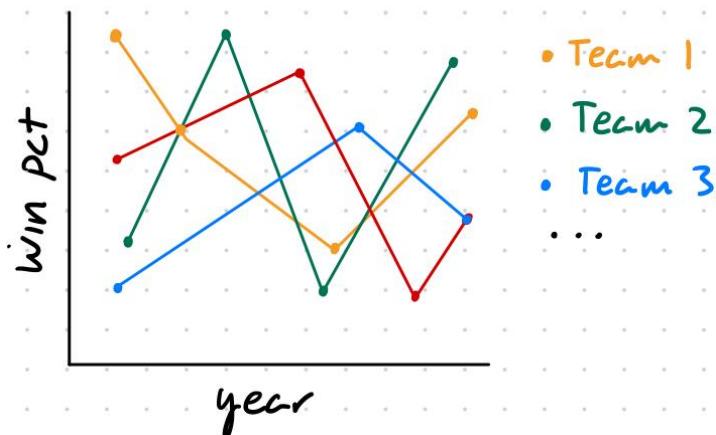
- a. Which [western] conference team ended the [2021] season with the highest number of [wins]?
- b. Which team had the highest win percentage between [2016-2021]?
- c. Which team has had the greatest number of #1 draft picks?
- d. Which **teams**/players have averaged the most amount of points per season?
- e. Does the plus-minus statistic translate well to team success?
- f. How has average efficiency changed over time?
- g. Do the best assisting teams always score the most points? If not, do they have a top-scoring player?
- h. How do defensive statistics translate to winning?
- i. Has NBA defense become worse? (e.g. less BLKS, STLS; more PFs). If so, is this a result of a rule change?
- j. Is there a metric which best predicts team success? (e.g. In recent years, DEFRTG vs. OFFRTG has been a huge predictor of NBA finals teams)
- k. Does team ownership/management affect performance?
- l. How many all-star players are needed for a winning team?
- m. Are higher-paying teams more successful?
- n. What are the mean 3-point and free-throw percentages/attempts in the league?
- o. Are the most elite teams being affected by playing away?
- p. Where are the best draft picks getting drafted from?

Anthony Morales: *Sketch 1, Sketch 2*

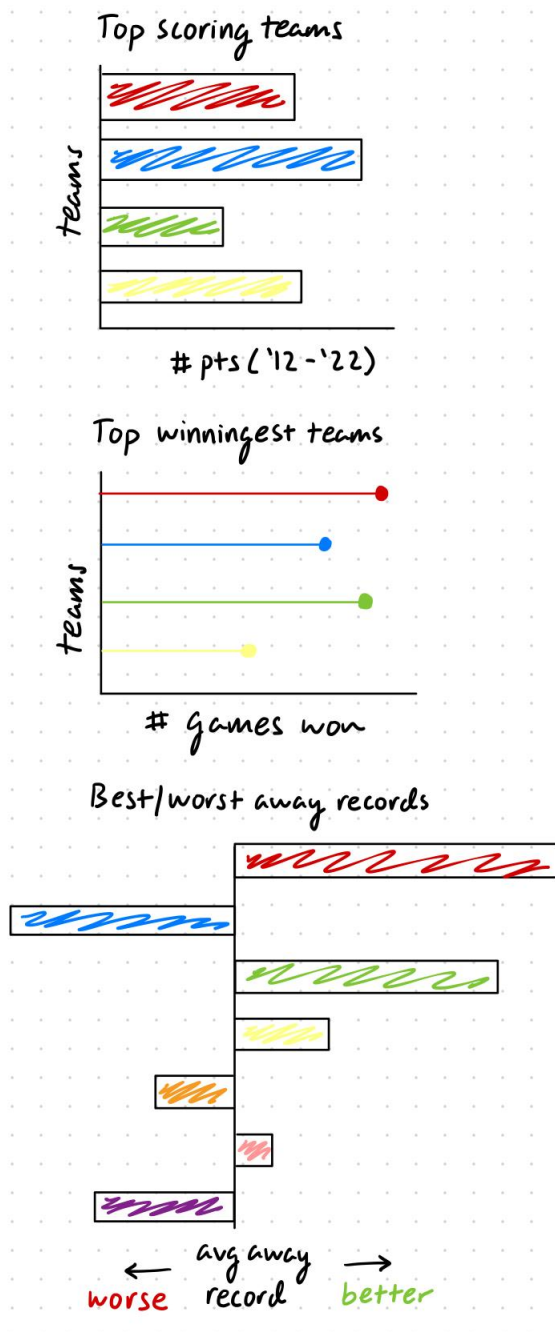
Top 10 teams w/highest win % / (i.e. most successful)



Top 5 teams w/highest win % (w.conf)?

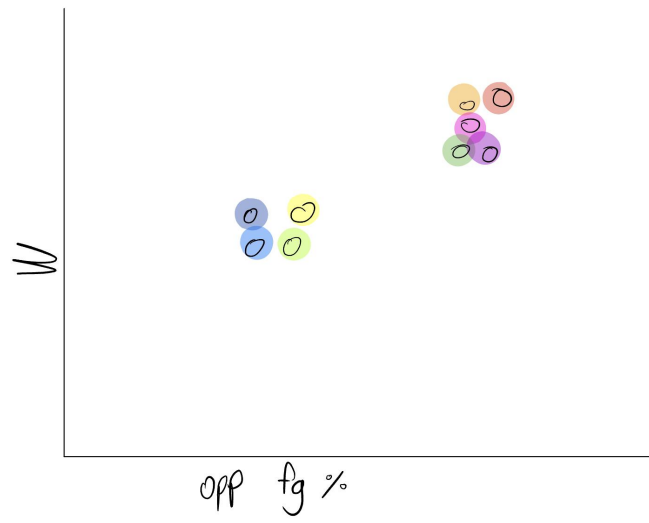


Anthony Morales: *Sketch 3, Sketch 4, Sketch 5*



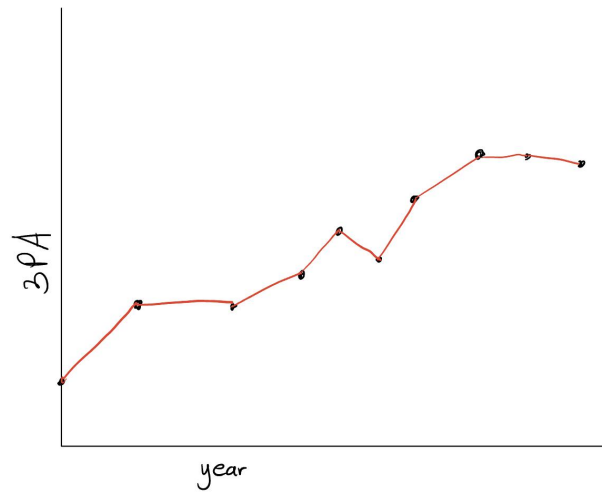
Sal Blanco: *Sketch 6*

How does opponent field goal percentage translate to winning? In other words, do better defensive teams win more games?



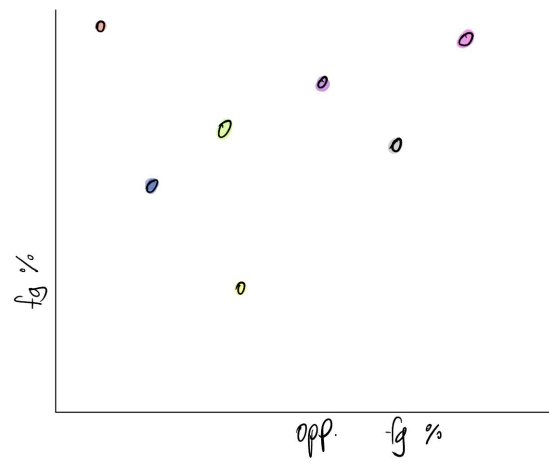
Sal Blanco: *Sketch 7*

How have 3 point attempts changed over the course of time? What is the standard today?



Sal Blanco: *Sketch 8*

What teams have the best offense and defense? (how does this translate to winning)



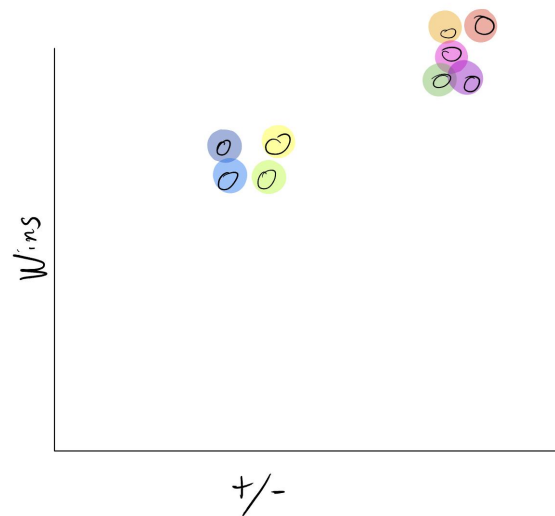
Sal Blanco: *Sketch 9*

Wins by Western Conference Team (shown by size of circle)



Sal Blanco: *Sketch 10*

Is a team's plus/minus statistic indicative of their amount of wins? (the +/- stat is used to highlight an individual player's performance and how many points they add or subtract from their team's lead when they are on the court; recently, many people have denounced this stat, stating it is not very telling of someone's overall performance)



Sketch ID	Question ID	Author
1, 2	b.	AM
3	d.	AM
4	a.	AM
5	o.	AM
6	j.	SB
7	n.	SB
8	h.	SB
9	a.	SB
10	e.	SB

As seen above, we've selected sketches 1, 2, 5, 8, and 9. This is because we believe these sketches are going to be the 'foundation' of our project. From these sketches, we can then branch out and answer more specific questions that our audience may have. We might

plan on implementing one or two of these sketches, but the main goal here for us is to get some ideas down to have as a reference.

Anthony Morales

- Insight 1: The San Antonio Spurs have been the team with the highest win percentage since 2004 (~65%, as shown in *Graph 1*)
- Insight 2: The San Antonio Spurs have also been one of the teams with the highest points-assists ratio (Seen in *Graph 2*)
- Insight 3: The most successful teams in the last 10 years have been Western conference teams.

Sal Blanco

- Insight 1: The San Antonio Spurs have had the best FG% and OPP FG% since the 2000 season.
- Insight 2: In recent years, teams with the highest 2P% have either:
 - Ended their season at the top of their respective conference
 - Made the NBA Finals
 - Won the NBA Championship
- This is in contrast to 3P%, which has no correlation to team success

Main Insight: Because of their massive success, the San Antonio Spurs can be used as an example of what a true championship winning team looks like.

Storyboard:

<https://jamboard.google.com/d/1Xz06YfWFEznNceX7UuluV6rDqNbJo-OZ25ID30QE2U8/edit?usp=sharing>



Week 11:

Meeting with Ethan: *Tips and Reminders*

- Make sure code structure is cohesive and well-designed
 - Use a previous homework assignment as a starting point
 - Consider Bootstrap
- Use drop shadows
 - Box-shadow CSS property
- Implement story scrolling
 - Consider animate-on-scroll JS library
- Finalize colors and fonts
 - Colors
 - Red: #D91E36
 - Blue: #384E77
 - White: #FFFFFF
 - Fonts
 - Bebas Neue
 - font-family: -apple-system, BlinkMacSystemFont, "Segoe UI", Roboto, Oxygen, Ubuntu, Cantarell, "Fira Sans", "Droid Sans", "Helvetica Neue", Arial, sans-serif, "Apple Color Emoji", "Segoe UI Emoji", "Segoe UI Symbol";
- Final implementation must include a creative visualization
 - Cannot be line graph, bar graph, or scatterplot
- Set up GitHub repository
 - Deploy project with github pages

Articles of Interest:

- <https://bleacherreport.com/articles/1202990-7-ways-gregg-popovich-and-san-antonio-spurs-changed-the-nba>
- <https://www.slamonline.com/nba/san-antonio-spurs-culture/>
- <https://time.com/5125421/gregg-popovich-san-antonio-spurs-success/>

GitHub Repository: <https://github.com/anthonymorales6/CS-171-Final-Project>

(Designs and sketches in .zip file)

Week 12: Think Aloud Study

Tester Name: Michael Sacco and Andrew Nickerson

Tester Email: michaelsacco@college.harvard.edu, anickerson@college.harvard.edu

General Observations from the think-aloud study:

- Need more interactivity
- Focusing on one team is effective; may need more explanation on topics included in the visualizations
- Storytelling aspect draws audience

What does the tester like about your data story?

- Michael and Andrew both highlighted the story behind the San Antonio Spurs and how that made our project much more relatable
- Focusing on one team allows for the audience to become more engaged with the presentation as they are gaining a deeper understanding of the Spurs
- Scatter plots were effective as it is easy to see which are the best teams per statistic

What improvements does the tester point out?

- Both Andrew and Michael stated that more interactivity is necessary
- They brought up the idea of drop down boxes to split up winning records
- They also brought up the idea of window zooms to filter through specific seasons/eras
- We must include introductory text/images in order to set the scene/tone for our deep dive into the Spurs

Was the intended key message clear to the tester? Why or why not?

- Yes; Michael and Andrew both stated that the storytelling aspect of our project allows the point to get across very clearly; since we are focusing solely on the San Antonio Spurs, the audience knows that we want to highlight what makes them successful and how teams can emulate that

Did the tester get your next steps or call to action? Why or why not?

- Yes; as stated before, since we focused on the Spurs and presented what contributed most to their success, it was evident that we are pointing out what changes other NBA teams can make in their front office/system in order to reach this same dynastic level of success

Based on the results of this study, we must improve the interactivity of our data story. By allowing the audience to click for further information, filter specific seasons, and use drop down boxes for new insights, we will allow them to answer their own questions about our data. This

would greatly benefit the receptiveness towards our visualization, since it would allow people who are not familiar with NBA history to explore our project on their own. Additionally, it would be useful to spread out our visualizations across multiple pages; by adding text boxes to explain our graphs, we could convey information which would not be explicit on our plots. Also, some people may not recognize what some statistics mean. Including informational text would bridge that gap.

Following this study, we have realized that more variety/interactivity in our graphs is effective. We will implement new line charts to display 2PFG% by team, win percentage by team, and include filtering for home and away records. Additionally, a scrolling layout for our webpage would add to the storytelling aspect of our visualization since we will be able to include more text, images, and graphs without it becoming too cluttered. Michael and Andrew both stated that our message gets across effectively, and that it was easy to understand where we were going with based on the presented graphs.

We will implement the following:

1. Scrolling layout for webpage - Anthony; complete
2. Drop down box to filter home/away records - Anthony
3. More tooltips for line charts in order to highlight which team we are focusing on - Sal
4. Refine tooltips for scatterplot to show specific statistic (e.g. OPP FG%) and highlight which team we are focusing on - Sal
5. Include images of team logo, players, coaches to further improve storytelling aspect - Sal
6. Make webpage color scheme fit the Spurs (difficult to do because team colors are black/gray/white) - Anthony