

MLB Player Salary Relation with Team Success Analysis

Samantha Delfino¹

Melis Akinci²

Benjamin Wyant³

Rachel Bringas⁴

Northeastern University

ABSTRACT

The MLB is one of the most popular sports leagues in the US and we are creating a visualization displaying how much money different teams spend on players each year vs. how well they perform. Our visualization will allow the end user to make their own decisions on whether a franchise spent a reasonable sum of money on their total player salaries based on their win percentage. This project is important because it provides background and further analysis on team spending that will help team leadership make important decisions, optimizing budget for performance.

<https://github.com/DS4200-S22/final-project-mlb-team-spending>

Keywords: MLB. Salary.

Index Terms: MLB player salary team performance—Visualization

1 INTRODUCTION

Major League Baseball (MLB) is one of America's oldest pastimes and is an organization that generates billions of dollars in revenue every year. However, the amount of money teams spend and generate is not evenly distributed among the 30 organizations. Unlike other leagues the MLB does not have a salary cap, which means teams in larger markets with wealthier owners can spend significantly more money than other organizations. Due to the limited salary restrictions, some questions arise. For the teams that spend a lot of money: is it worth spending the money to be good? Did the spending even improve team performance or revenue generated? For the teams that are less likely to spend money on big name players: how well did you perform relative to the big spenders? Would spending money on some "big name" (popular/high performing) players be worth it?

These are the questions we hope to provide answers to with our visualizations. Our object is to provide graphics to help inform the average fan or potentially even front offices of these organizations as to what might be the "sweet spot" for the annual salary cap. In addition, the goal of our project will be to determine how "successful" these organizations are relative to their total salary cap each year.

2 RELATED WORK

Analyzing data mining methods in sports analytics: a case study in NHL player salary prediction—This is a case study of player salary

in the NHL and being able to predict player salary. There are some valuable visualizations and explanations for data mining that would be useful to our project as they were working with data mining of player salaries and relational data. This might also provide more ideas for our project so that we can apply to MLB player salaries.

State of the Art of Sports Data Visualization—This paper discusses the advances and challenges of visualizations of sports data. The report discusses three types of sports data: box score data, tracking data, and meta-data, of which the commentary on box score data will be most useful to us. The report comments on how analysis of sports data can result in better team performance, which we hope our visualizations will do. This report can inform our project by letting us know the areas in which we may run into challenges, so we can avoid those areas. It can also help us with which visualizations will most accurately convey our information.

Data Always Tells a Story: Salary Disparity in Major League Baseball—This blog post uses Tableau to represent a large amount of baseball data. He talks a lot about using Tableau to discover and explore baseball data, however, in the conclusions section, he dives into trends in ticket sales and attendance. One visualization that is interesting to me was the Average Salary of Pitchers v. Position Players. While this is only a blog post, it will help guide us through our data visualization process.

Salary Inequity, Team Success, League Policies, and the Superstar Effect—A key research paper that focuses on all aspects related to what we are researching in our project except in the sport of hockey. A team's salary distribution was used and graphed with the data from the Collective Bargaining Agreement made between the owners and players from the NHL, leading to a strong visualization that shows the relationship between a team's salary distribution and their winning percentage. This will be useful in our project as we are attempting to do the same thing except for the MLB.

Latest Space Cartography: Visual Analysis of Vector Space Embedding—This literature from EuroVis 2019 is unrelated to our topic but is a great example of the power of visualization to transform ideas which may be hard to grasp conceptually or through words alone into digestible concepts. This piece is about the visualization of vector space embeddings. This is something that is extremely important as it is used in machine learning to map high dimensional spaces into data that can be used by unsupervised machine learning algorithms. One case where this visualization is useful is to help understand how Natural Language Processing is able to be achieved.

¹
Delfino.s@northeastern.
edu

²
Akinci.m@northnortheastern.
edu

³
Wyant.b@northeastern.
edu

⁴
Bringas.r@northeastern.
edu

3 USE CASE

For baseball teams, there is a constant balance between creating a winning team and staying within the budget that has been predefined for the year. Front office departments would benefit from understanding if there is a relationship between how much is spent by a team and the success of said team. This will help build their plan for player selection and payroll for the year. It is ultimately a baseline that teams can use when starting to plan for the next year. There is also a salary tax put up each year, where if a team spends a certain amount of money on a player's salary, they pay a salary tax. Understanding whether it is more valuable to stay under the salary line or go over would define the path that teams will go to building their team for that year.

The visualizations would show the relationship between team player salary and the success of the team. For each team, there are many different variables where one salary number might work for them but might not work for another team. These visualizations would consider team revenue, inflation for the year, difficulty of team schedule, among other factors. This will allow teams to customize the number for them.

4 DATA

- Rods Sports Business Data—Team payroll for each team for each year from 1988 through 2020 seasons. This data set has salaries listed on Opening Day and End of Season. Data will need to be rotated so each year is no longer a row but a column so it would be Team-Year to be able to match the format of FanGraphs data.
- Baseball Reference-- This shows the wins of each team. Available before 1988 but since the above data is only from 1988, will only pull starting in 1988. This data set will need to be transposed and the unique identifiers for each team will need to be made the same as all data sets.
- FanGraphs—This shows the teams batting, pitching, and fielding of each team for each season. These values will be able to be used as a measurement of success aside from a number of wins.
- Rods Sports Business Data provides different variations of datasets from different sources. There are backups within google drive.
- Retrosheets—Provides game logs for all games from 1988. This can be used to aggregate the games to find the number of wins. This would be a backup as the number of wins is already aggregated.
- Baseball Prospectus—This would be a backup for the FanGraphs data but would be required to be aggregated by teams and can only be downloaded by each year so data sets would have to be unioned.

5 IMPLEMENTATION PLAN & PRELIMINARY WORK

We hope to show a visualization which represents the success of a team and their spending habits. We plan on using Tableau to present the visualization; however, we are open to using other methods as we continue to learn them in class. Another element of the visualization that we hope to showcase is the possibility to overlay money spent and total revenue or a measure of success. Another element that we hope to add is a line on the graph showing the luxury tax as well as the most desirable salary payout would be for the teams in terms of their total salary payout.

6 IMPLEMENTATION PLAN REVISION

The visualizations we will need for the final project are a bracket visualization and a line graph. For the bracket visualization, we will also need to have charts that show the details on demand, such as team name, the amount spent, number of wins, and then cost of wins. These details on-demand can also be shown for multiple teams if an end-user selects multiple teams. For the bracket visualization, the teams will be color-coded based on their spending- teams that have spent more money are darker, and teams that have spent less are lighter in color. The second graph is a line chart and shows the success score vs. the years. The success score is the win percentage of each team. The teams shown on the line chart will be the teams that were highlighted on the bracket graph. The line chart will also include a line of best fit, which shows the trend of percentage wins over time per team. The interactions we will be implementing are brushing and linking, highlighting, and details-on-demand. The details on demand will come from the bracket graph, allowing a user to select a certain team to see more information on that team for that year. The brushing will be when different brushes are brushed over on the bracket, it will be linked to the line graph on the right to show the certain brushed-over teams. When the teams are brushed over, the teams will highlight to show that they have been selected and will also coincide with the lines on the line chart. The first visualization that we would like to accomplish is the bracket graph for each year. That gives us the best overview of the team's performance that year, and the details on demand for that graph will give us a better idea of the team's spending in that year. Next, we would like to assemble the line graph with the brushing and linking. Our "nice-to-have" would be adding team logos, to better show what team has participated in a certain year's postseason. Another nice thing to have are the best fit lines on the line chart. For our API, we plan on using D3 to create the outline of the bracket and the line chart. We will also use D3 for brushing and linking and the details on demand. We will use python for the actual analytics of our visualizations, such as computing the cost of a win. Our project does not require any server-side components.

7 VISUALIZATION DESIGN

Our final visualization has several components. The first graph shows the Percent Wins v. Years for each individual team. The percent wins are calculated by finding the ratio of wins/games and converting it into a percentage. The second graph is the Opening Day Salary (OD_Salary) v. Years for each of the respective teams. The color of the line on the first and second line graph is in the team color and when you hover over each of the points in both graphs, you get the year along with the Percent Win value and OD_Salary, respectively. The varying colors for each team will further allow users to identify which team they are analyzing and being able to connect the Percent Win Value with the OD_Salary side by side allows users to easily compare the trends between salary and wins over the years. These two graphs help in answering the question: does spending more to create a team result in winning?

The interactions only further the users understanding of the trends by providing more specific details and values. In addition to the two line graphs, there is a postseason bracket which shows the progression of teams that won their division and the wildcard spot in the American and National League towards the World Series Championship. This allows the user to see how each team

compares with each other and their spending habits. We hope to add linking and brushing between the 3 visualizations so that users will further be able to compare and analyze the teams and their success in relation to spending. This will be done by the user selecting the year in the drop-down menu for the bracket which will display the postseason bracket for the respective year. The bracket will also display a color gradient where the darker the color, the more money is spent by the team in that given year. When brushing over the teams in the bracket, the Percent Wins v. Year and OD_Salary v. Year for those teams will be shown in the two line graphs.

REFERENCES

- [1] BaseballProspectus. (2022). *Hitting Leaders*. Retrieved from <https://www.baseballprospectus.com/leaderboards/hitting/>
- [2] BaseballReference. (2022, 2 23). <https://www.baseball-reference.com/leagues/majors/index.shtml>. Retrieved from Baseball Reference: <https://www.baseball-reference.com/leagues/majors/index.shtml>
- [3] C. Perin, R. V. (10 July 2018). State of the Art of Sports Data Visualization. *Wiley Online Library*. doi:<https://doi.org/10.1111/cgf.13447>
- [4] Cyrenne, P. (20 February 2017). SALARY INEQUALITY, TEAM SUCCESS, LEAGUE POLICIES, AND THE SUPERSTAR EFFECT. *Wiley Online Library*.
- [5] FanGraphs. (2020). *Baseball Graphs*. Retrieved from <https://www.fangraphs.com/leaders.aspx?pos=all&stats=bat&lg=all&qual=0&type=1&season=2020&month=0&season1=1988&ind=1&team=0,ts&rost=0&age=0&filter=&players=0&startdate=&enddate=>
- [6] Mincev, S. (2021, March). Analyzing data mining methods in sports analytics: a case study in NHL player salary prediction. *NOVA Information Management School*. Retrieved from <https://run.unl.pt/bitstream/10362/117866/1/TGI0404.pdf>
- [7] RetroSheet. (2021, 12 02). *Play-b*. Retrieved from Play-by-Play Data Files (Event Files): <https://www.retrosheet.org/game.htm>
- [8] Riddett, K. (2018, 09 25). *Data Always Tells a Story: Salary Disparity in Major League Baseball*. Retrieved from interworks: <https://interworks.com/blog/kriddett/2018/09/25/data-always-tells-a-story-salary-disparity-in-major-league-baseball/>
- [9] Rodey For's Sports Economics. (2020, 12 19). *Rodney Fort's Sports Business Data*. Retrieved from MLBPAYROLL: <https://drive.google.com/drive/folders/10J3MM7oAPT0Rk4DaDjV4FQH6NLEO3BHw>
- [10] Y. Liu, E. J. (2019). Latent Space Cartography: Visual Analysis of Vector Space Embeddings. Paul G. Allen School of Computer Science & Engineering, University of Washington. doi: <https://idl.cs.washington.edu/files/2019-LatentSpaceCartography-EuroVis.pdf>

APPENDIX A

1 GROUP PURPOSE

The purpose of our group's formation is to analyze financial trends in the MLB for our Data Visualization class here at Northeastern University. We would like to specifically look at team spending as it correlates to team performance, to determine if there is an ideal spending amount.

2 GROUP GOALS

The following are our group's project, process, and quality goals:

- Create minimalist yet eye catching visualizations for our complicated datasets
- Ensure clear readability and understanding for all potential users
- Finish all week's work 2 days prior to assignment deadline
- Check over each member's work to ensure high-quality output

3 GROUP MEMBER ROLES/RESPONSIBILITIES

- Melis Akinci: group leader
- Sam Delfino: meeting facilitator
- Ben Wyant: project manager
- Sunny Kang: communications
- Rachel Bringas: documentation coordinator

4 GROUND RULES

Our group will meet weekly on Thursdays at 6:15pm as well as on Sunday evenings at 5:30pm to discuss the progress and milestones of the project. All meetings will be over Zoom unless otherwise discussed. Our Thursday meetings will be for work delegation, and our Sunday meetings will be for working together on the project. We will all work to create an inclusive and positive work environment through collaboration. This will allow us to learn from each other and highlight each team members' strengths. Each week during the meeting, we will discuss the overall updates to the projects along with any issues we are facing.

Our group agrees to clearly communicate with each other during all steps of the project. If a group member needs help, it is crucial that they reach out sooner rather than later to get help. It is expected that each member will get their work done and put in sufficient work. If a member cannot make a meeting, it is expected that they will give sufficient (at least 24h) notice.

Our group will support one another in their work. If a group member needs help in any aspect, the other group members will do what they can to support the group member and aid in problem solving.

On Thursdays, we will review the project milestone as a group, and delegate parts based on interest and strength. On Sundays, we will conduct our discussions by first having each member discuss their work on the project milestone up until this point, and what work they still have left to do before the due date. We will talk about any questions, comments, or concerns we have as a group, and maintain a space to work collaboratively and bounce ideas off one another.

We will handle dissenting views among members with respect and kindness. Rather than fighting back at them, we should assess their points and ask follow-up questions to dig deeper into the issue to ensure that we truly understand why they feel this way about it. Furthermore, we will ask the dissenting members for solutions to the issue, allowing us as a team to think creatively. Rather than pushing them away and criticizing them for dissenting, we hope to express our gratitude towards them and approach it in a positive manner since it will help us grow as group members and the project.

It is important that we do not let differing opinions hinder us as a group.

We will hold each other accountable for these rules and task completions at our weekly meetings where we will discuss the project and weekly progress. Rather than being upset if someone is unable to complete their work for that week, we will collaboratively come up with a solution and work together to solve any issues and come up with a new game plan.

We are expecting equal amounts of participation from each other and a willingness to learn and improve throughout the project. A moderate level of commitment is expected since we are all in several other classes and have other obligations such as jobs and clubs. Because of this, we hope to split up the work in a reasonable manner so that we can manage our individual workloads.

5 POTENTIAL BARRIERS AND COPING STRATEGIES

Some potential barriers that we may face are finding a consistent time to meet for the project, however, we plan to meet during weekends so that we do not conflict with classes and club meetings. If there are any other obligations that materialize, we will directly communicate them with the team so that we can handle it. We are all involved in other activities and understand that things come up and are willing to work with the other members of our team if they need help. In the past, we have all dealt with group members, not communicating with each other and providing honest updates regarding the project. This has caused lots of unwanted stress and difficulties. Because of this, we have all agreed to be transparent about our work and struggles for us to successfully collaborate as a team.

6 GROUP CHARTER REVISION

We have all been abiding by our agreed upon guidelines. Our group roles are working well for us and we have no issues with those. We don't think we have any issues that we need to troubleshoot. One thing that I (Rachel) have seen my group members contribute is putting in the time to have a well-formatted, clean project. One thing that I (Sam) have seen my group members contribute is to be inclusive with everyone's ideas into the project. One thing that I (Melis) have seen my group members contribute is checking over each other's work and truly collaborating with each other. We have been able to work well and answer each other's questions to ensure that we understand the assignment. One thing that I (Ben) have seen my group members contribute is the hard work and dedication to producing a good visualization and project from each of my group members. They have all put in extra work to ensure that they understand the goals of the project.

APPENDIX B: DATA EXPLORATION

1 DATA REVIEW

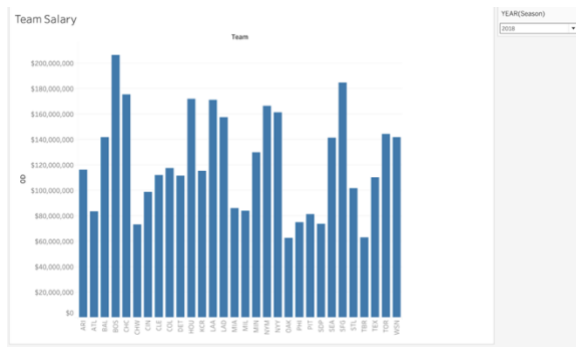
The data types present in our dataset are tables with items. The attributes of these items are both categorical and ordered. An example of the categorical data is the team, name, age, and season. The ordered data is quantitative data is the payroll each year, the team wins and game statistics. When performing a review of our data sets, we can see that there seems to be a large gap between the top 5 teams and their payrolls from the remainder of the teams. Most of the statistical analyses of our data shows that they are about normally distributed, however, there has been a consistent increase throughout the years.

- There are some missing values in a couple of columns, specifically in the FanGraphs data. However, the missing values are not included in columns that are essential for us to use, so we don't foresee these missing values as a problem
- There is no one dataset that has all the values that we want to look at, so there is a need to combine the datasets to be able to create a full analysis. To combine these datasets, we will use python to manipulate and join the datasets together.
- There are some team names that were changed over time. Therefore, their identification in the datasets changed. To address this, if the team has not moved cities, we will use the most up to date name for the team, and adjust any data as needed. If the team has moved cities, we will keep the names as-is.
 - There have been some teams that have been created as well as ended between 1988 and 2018, so there won't be a continuous analysis for every team in this timeframe.

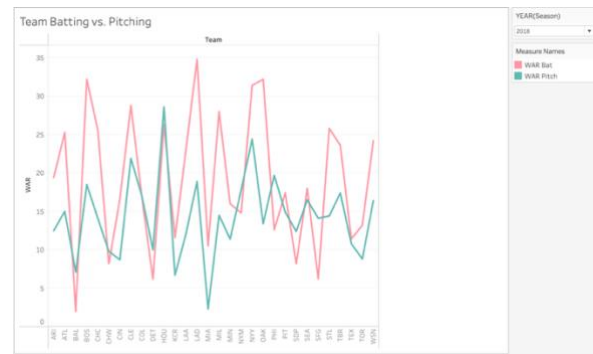
2 INSIGHTS

The total Opening Day salary (total amount of money spent on player salary for the year reported by opening day of year recorded) each year steadily increased, except from 2017 to 2018, the total salary decreased. Over time, Opening Day salary has increased, from an average of about \$15 million in 1988 to an average of \$125 million in 2018. The amount of money spent by each team is constantly changing, but until salaries are adjusted for inflation, it is unclear how drastic team spending changes year-over-year. Currently, without adjustments, each team has increased their spending from 1988 to 2018, some steadier than others. One further error that we have identified is the need to adjust for inflation. We will solve this by creating a calculation that will adjust each team's salary for each year to the 2018-dollar value. This will help us stay consistent and allow us to analyze the data more accurately.

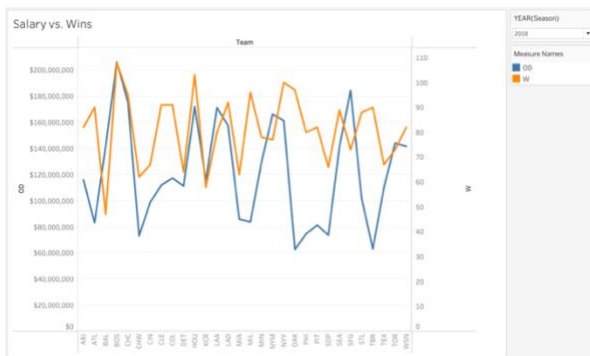
3 SCREENSHOTS



The data we are looking at here is the Team vs. the Opening Day salary for the year 2018. From this visualization, we can determine which teams have the highest and lowest salaries. Some of the teams with the highest salaries include the Boston Red Sox and the San Francisco Giants. Some of the teams with the lowest salaries include the Tampa Bay Rays and the Oakland A's. There is no trend displayed in this visualization other than the difference between team spending in the 2018 season. The visual encodings used in this visualization are align region and rectilinear axis orientation. Align region is used to categorize each team and compare the total team salary against each team. A rectilinear axis is used to show the total Opening Day salary on the y-axis and each team on the x-axis.



This visualization displays the pitching WAR (wins above replacement) for each team. Pitching WAR measures how well a pitcher pitches. It considers all aspects of pitching and determines how many more wins a player is worth than a replacement-level pitcher. Batting WAR measures how well a batter bats. It considers all aspects of batting and determines how many more wins a player is worth than a replacement-level batting. This visualization shows the average WAR per team. For the 2018 season, Team Batting vs. Pitching visualization is demonstrating whether a team is good at batting or pitching, or both, or neither. The visual encodings used in this visualization are align region and rectilinear axis orientation. Align region is used to categorize the teams to be able to compare each team's batting and pitching WAR. The rectilinear axis orientation is used to show the team categories and their WAR. Since the WAR is on the same scale for batting and pitching, the y-axes were synchronized.



In this visualization, we are looking at the team's name, along with their Opening Day Salary and Wins. This visualization gives us a better look at how salary correlates to wins. Overall, there seems to be a trend between more wins and a higher salary. For example, the Boston Red Sox has the highest opening day salary and has the highest number of wins. The visual encodings used in this visualization are align region and rectilinear axis orientation. Align region is used to categorize the teams to be able to compare each team's salary and wins against other teams. The rectilinear axis orientation is used to show the team categories, their total Opening Day salary, and their total wins, providing reference values for the lines graphed in the visualization.

4 DATA SNIPPET

Season	Team	GS	IP	K%	BB%	HR%	BA%	LOB%	GB%	HR%	WPA	ERA	WHIP	FIP	WAR	St	W	L	SV	WAR	St	W	L	SV	WAR	St	W	L	SV
2018	ARI	2028	6107	176	880	858	79	81%	23.7%	0.162	0.298	0.228	0.31	0.387	0.338			88	7.1	40.3	70.2	18.4	82	80	73				
2018	ATL	2377	6251	178	707	717	80	8.2%	25.8%	0.161	0.302	0.257	0.324	0.417	0.319			96	12.8	-18.5	66.8	25.3	90	72	40	71			
2018	BAL	2313	6034	188	622	558	81	7.0%	23.4%	0.152	0.288	0.239	0.324	0.381	0.399			87	-8.8	-104.1	-75.1	2	47	115	28	65			
2018	BOB	2345	6302	208	870	829	125	8.0%	19.8%	0.185	0.308	0.258	0.339	0.458	0.34			111	-5.1	74.1	36.9	32.2	108	54	48	69			
2018	CHC	2480	6388	187	781	722	86	9.0%	21.8%	0.152	0.313	0.258	0.333	0.41	0.321			100	2.2	3.4	44	25.6	95	68	46	76			
2018	CHW	2348	6071	182	698	628	98	7.0%	26.3%	0.16	0.304	0.241	0.302	0.401	0.304			92	-7.2	-87.5	-62.8	8.2	82	100	34	71			

Season	Team	GS	IP	K%	BB%	HR%	BA%	LOB%	GB%	HR%	WPA	ERA	WHIP	FIP	WAR	St	W	L	SV	WAR	St	W	L	SV	WAR	St	W	L	SV
2018	ARI	182	1483	8.81	3.21	1.07	0.289	75.7%	47.3%	13.7%	92.1	3.73	3.91	3.81	12.5	118038416	118038416	4144223	2500000										
2018	ATL	182	1486.2	8.79	3.82	0.85	0.278	74.1%	43.3%	11.2%	93.7	3.75	3.99	4.17	15	8338841	8338841	287262	750000										
2018	BAL	182	1421	7.57	3.7	1.47	0.31	70.8%	40.9%	14.4%	93	3.19	4.58	4.75	7.1	147814116	147814116	523584	2300000										
2018	BOB	182	1488.2	9.61	3.16	1.09	0.294	75.8%	42.4%	11.9%	93.6	3.75	3.82	3.92	18.5	20824788	20824788	6445243	2137500										
2018	CHC	182	1478.1	8.13	3.79	0.96	0.284	78.2%	45.0%	11.9%	92.4	3.85	4.14	4.29	14.1	175310202	175310202	6982682	3837500										
2018	CHW	182	1427	7.89	4.09	1.23	0.292	69.4%	40.8%	11.7%	93.3	4.85	4.73	4.88	9.8	73003333	73003333	3187549	1000000										

5 TASK ANALYSIS

Task ID	Domain Task	Analyze Task	Search Task	Analytic Task
Owner	I want to use these visualizations to see if we are spending too much on player salaries and if cutting the budget would positively or negatively affect team success.	Consume → Discover	Explore	Identify

General Manager	I want to use a visualization to see if the players and overall team are performing differently based on various budgets and spending patterns.	Consume → Discover	Explore	Identify
CFO	I want a visualization tool to show how our organization can best go about maximizing total revenue based on which players generate the most revenue and which	Consume → Present	Explore	Compare

	contracts are most effective for a salary to increased revenue ratio			
--	--	--	--	--

The primary consumer of our visualization will be the general manager of the team. The type of consumption our visualization will be primarily used for is for discovery. This is because the general manager will use the visualization to make discoveries about team performance as it relates to team spending, to make decisions about the future of the team. The general manager can use this data to decide if the team needs to spend more or less to reach peak performance, which they can then use to decide which players will remain on the team and which players will not.

APPENDIX C: INTERVIEW

1 END-USER PERSONAS

The end-user will be the general manager of the team; the general manager is the individual that assembles a baseball team. We envision that a general manager has used visualizations in the past but has never created any themselves. Therefore, they are comfortable with the terminology and can understand more complicated visualizations. The specific task that a general manager would need a visualization for is using the visualization to create a baseline to set a budget for their team for the team they are creating that year. It can help the general manager determine how they want to sign players for the following year and determine if it's worth it to spend more on certain players. Currently, without this specific visualization, the general manager makes decisions on players by looking at budget spreadsheets and player statistics. This visualization will help enhance the general manager's decision-making process; they might already use certain visualizations but not in this way. Overall, the visualization will help the general manager pick a better team by helping them find the "sweet spot" of spending to achieve optimal team performance and will also aid in player contract negotiation by setting a budget per player.

2 INTERVIEW SCRIPT

What does your day-to-day work consist of?

What tools do you need to be successful?

Can you explain how you would use a visualization to assist in your day-to-day work?

Please describe three pain points in your current process.

Please describe three positive aspects of your current process

What data would you like to work with?

What is the goal of this visualization?

Are there any characteristics you would like to have in your visualization?

3 INTERVIEW NOTES

What does your day-to-day consist of?

- Manage player contracts, trades, and roster throughout each season
- Overall goal → do we want to give them an offer/stay on the team

What tools do you need to be successful?

- Budget spreadsheets for each player, statistics of their success

Can you explain how you would use the visualization to assist in your day-to-day work?

- It will help with making decisions on the value of players when making deals with the player themselves, or with other teams, or for determining how valuable a player truly is
- Is it worth signing a player for a given value (should we pay them more or less?)

Please describe three pain points in your current process.

- Massive budget sheets are difficult to read and understand what the data is showing
- Difficult to combine player analysis/performance with corresponding salaries
- Want to maximize profit and efficient spending of the team

Please describe three positive aspects of your current process.

- Are getting large budget spreadsheets and can analyze basic correlation between budget and performance
- High level of analysis on player development and performance
- Want to integrate this w budget

What data would you like to work with?

- Budget, year, spending, performance metrics (rate of wins out of 1.0), batting v. pitching, home runs/runs, WAR
- Across teams

What is the goal of this visualization?

- Want to keep improving the process in order to return the most in the investment
- Know if we are maximizing the budget to the fullest

Are there any characteristics you would like to have in your visualization?

- Year to year comparison to see any trends
- Team performance v. budget → over years

7 INTERVIEW RESULTS

What does your day-to-day consist of?

- Our day-to-day tasks consist of managing player contracts, trades, and the team roster throughout each season. One overarching goal of our team is to evaluate whether we want to keep the player on our roster and continue to keep them on our payroll.

What tools do you need to be successful?

- In our job, we use spreadsheets that break down our budget and the spending habits of our team. In addition to these, we have several spreadsheets that have statistics on both the overall team success and player success that we use to help gauge the monetary value of our team and players. This allows us to make informed decisions about where we should focus our spending.

Can you explain how you would use the visualization to assist in your day-to-day work?

- This visualization will help in our day-to-day work by helping us make decisions on the value of players, which is crucial when making deals with other teams or with the players themselves.

Please describe three pain points in your current process.

- One pain point in our current process is that we are given massive amounts of data with the budget sheets, and they are difficult to read and understand. Another pain point is that it is difficult to analyze player performance with their salaries, as that data is not centralized in one place. Lastly, our main pain point is that we do not believe that we are maximizing profit and efficient team spending.

Please describe three positive aspects of your current process.

- One of the positive aspects of our current process is that we have received large budget spreadsheets already, so we have a lot of data regarding our budget. We are able to use the budget sheets to analyze the basic correlation between budget and player performance. Another positive aspect of our current process is that we have a high level of analysis on player development and performance and are hoping to get our budget analysis to that level. Finally, the current process we have is streamlined after years of operating, so we are able to know what we are looking for in the budget spreadsheets.

What data would you like to work with?

- We would like to work with all the current budgeting data we have on file, including year-over-year spending, yearly revenue, and yearly spending. In regard to team performance data, we would like to use performance metrics, such as rate of wins (in decimal form out of 1.0), basic batting statistics that show team success at the plate, basic pitching statistics that show team success on the mound, how many runs and homeruns the team gets each year to show offensive success and Wins Above Replacement (WAR) to get an overall

rating of how well the team did that year. We want to look at the team level so we can see how the overall budget set for the team correlates with team success.

What is the goal of this visualization?

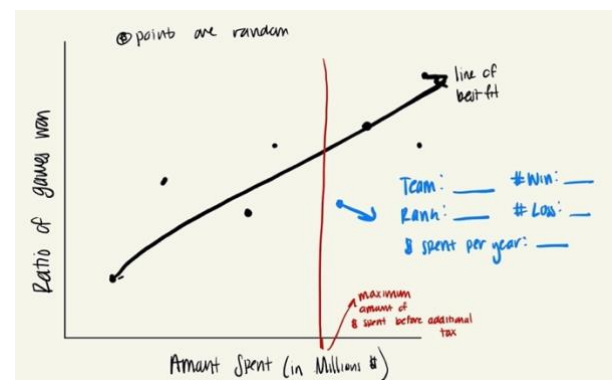
- The goal of this visualization is to allow our organization to improve the process in which we go about measuring which salaries create the most return on investment. In addition to this, we also want to maximize the total budget in order to generate the most revenue and have the most team success with the smallest amount of spending necessary.

Are there any characteristics you would like to have in your visualization?

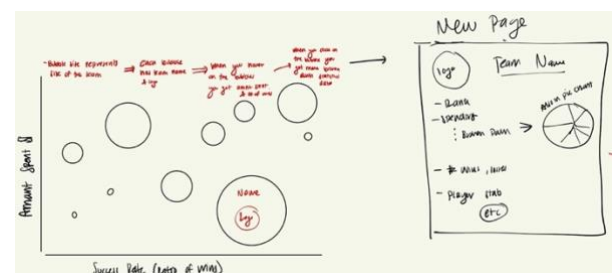
- We would like to see year-to-year comparisons of any trends as well as the team performance vs. the budget over the years.

APPENDIX D: INTERVIEW

1 INDIVIDUAL SKETCHES



This graph shows the relationship between the amount spent by each team and the ratio of wins to losses. When the user hovers over a point (shown in blue), additional data regarding the point is represented including the Team name, the number of wins and losses, their rank, and how much was spent in that given year. Furthermore, a mark used is color where each of the different teams will have a distinct color in order to differentiate from each other and so users are able to further examine the graph. This visualization addresses the General Manager (GM)'s task of examining the team's performance in relation to their spending habits. Moreover, the GM would be able to assess whether their investment is truly leading to better performance. (Sketch #1-Melis)

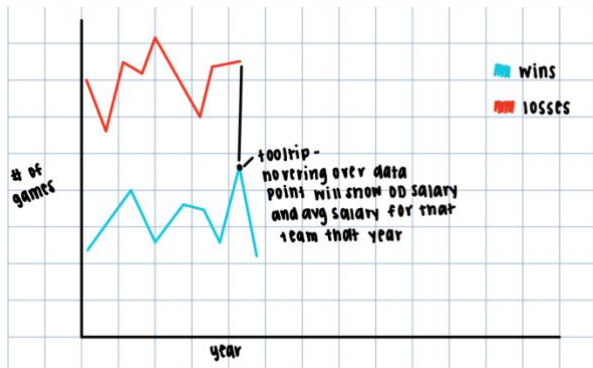


This visualization displays the relationship between the Success Rate of a team and the amount spent. The size of each of the bubbles which contain the logo and team name is based on the size of the team and their success. I chose to use the mark size since it is easy to identify and differentiate amongst the graphs. When you hover over a given bubble, additional statistical

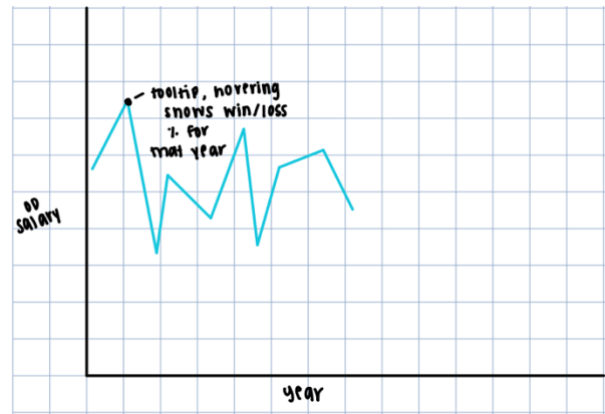
information is displayed such as the team's name, rank, number of wins, and the spending separated into categories in a pie chart for users to digest easily. (Sketch #2- Melis)



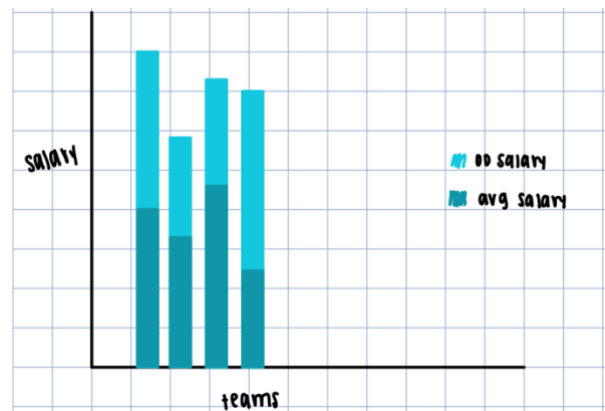
This visualization uses the map of the US to showcase the various teams across the country. The size of the logos would be dependent on their spending score rate which is calculated by dividing the ratio of wins by the total money spent. The mark and channel of size is used in this visualization. Furthermore, when the logo is hovered over, it will display additional data such as the amount spent per year, team name, and the number of wins and losses. Another mark and channel which is used in this visualization is the use of color. The spending score of the team is separated by color where green represents a good spending score (top 1/3), yellow a moderate score (middle 1/3) and red is a bad spending score (bottom 1/3). This allows individuals to easily distinguish which teams have good and bad spending scores. The GM would be able to utilize this graph to see if their spending investment towards success is truly benefiting them and which teams are able to maximize their spending habits. (Sketch #3- Melis)



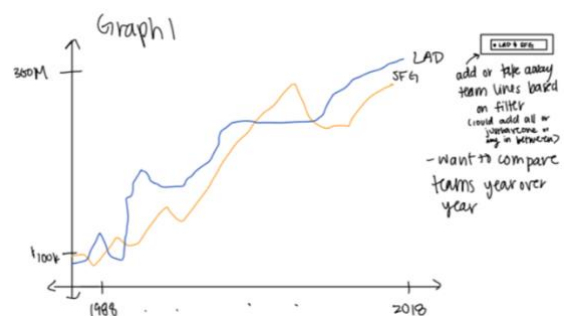
This graph shows the number of wins and losses for one team over the years. I chose to represent this data as a line graph to show the trends in wins and losses over the years. I chose red to represent losses as it has a more negative connotation and blue to represent wins for a positive connotation. This graph would also include a tooltip; when a user hovers over a data point for a year (on either wins or losses), it will let the user know the OD salary and the average salary for that year. This visualization addresses the GM's task of looking at team performance and using their performance metrics to look at salary. (Sketch #4- Rachel)



This graph shows the OD salary for one team over the years. I chose to represent this data also in a line graph to show the salary trends over the years. This graph would also include a tooltip; hovering over a year would show the wins and losses for that year. This visualization would most benefit the GM, as it would aid in looking at how much a team is spending vs. their performance. (Sketch #5- Rachel)

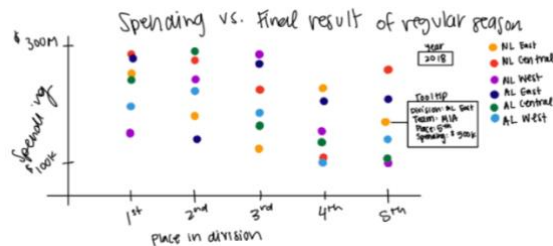


This graph shows all of the teams, their OD salary, and their average salary. I represented these points as a bar graph to best compare the teams to one another and did a stacked bar chart to show the average salary compared to the OD salary. This graph would most benefit the CFO, as the CFO for a certain team could see how their average salary compares to their OD salary, to make decisions about changing or keeping the average salary. (Sketch #6- Rachel)



This graph shows the team spending over time. This graph allows for the flexibility to look at just one team, compare multiple teams' spending, or compare all of the teams' spending over time. I showed the trend of spending as a line graph, to emphasize that

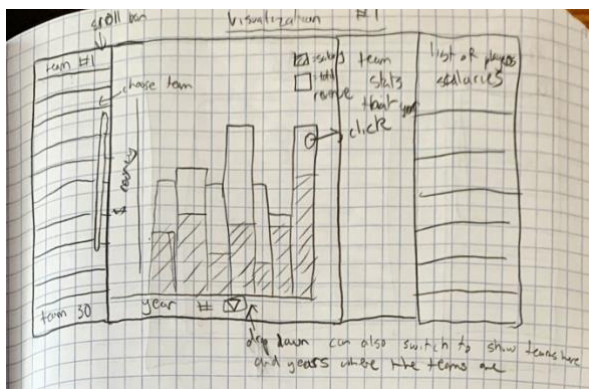
this is meant to be looked at as a trend. This graph would benefit the Owner to see where their team is, spending-wise, compared to other teams. They would be able to see if they have been consistently spending more or less than other teams. This would help the owner decide whether they want to keep on this trend or move their budget in the future. (Sketch #7- Sam)



FAVORITE - This graph compares how each team finished in their division and their spending. I represented each team with dots, color coded by each division, so it could also compare division-by-division. This can be filtered by year. It answers the question for the general manager, is our spending adequate for our performance? This would allow the general manager to see if they are spending too much money on poor performance, or vice versa. It also compares each team, so the general manager can see where they stand compared to other teams. (Sketch #8- Sam)

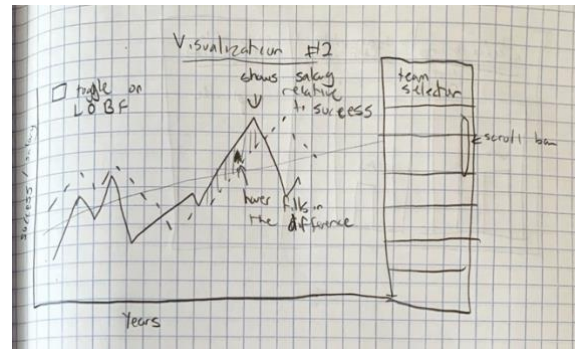


FAVORITE- This graph shows whether the high-spending or low-spending teams advance in the postseason (and win it all) or if they lose out. It would show whether more spending actually equals better performance. The scale of spending is portrayed with a gradient scale, so the darker the color, the more spent, the lighter the color, the less spent. The advancement throughout the postseason is shown via the postseason bracket. It can be filtered by year. This would most benefit the general manager because it would allow the general manager to see if their spending on players is worth the desired result, which for every team is winning the World Series (postseason). (Sketch #9- Sam)



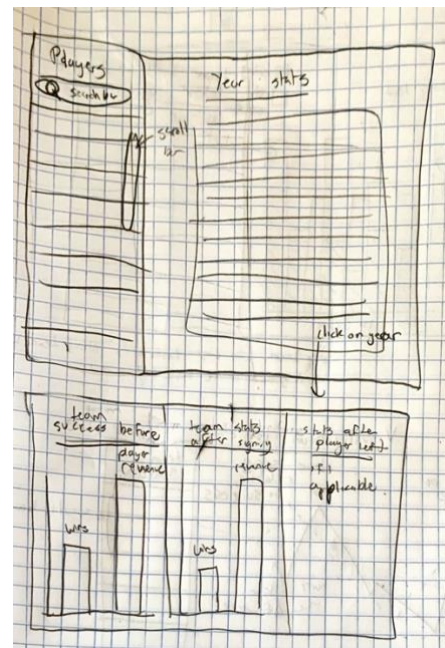
This graph shows the revenues generated each year for a team as well as the salaries for the players on the team. This allows you to

see how much revenue was generated relative to how much players were paid so you could see if more money spent on players means more revenue generated. This would benefit the COO because it would allow them to realize whether from a financial standpoint, spending more money on players generated a higher net income. Clicking on a bar would show the team stats for that year as well as the salaries for that year. Additionally, there is a scroll bar along the left side where you can switch between the teams. (Sketch #10- Ben)



FAVORITE - This visualization has a team selector on the right side to swap between teams. It shows a team's salaries throughout the years as well as the number of wins the team got that year.

This can be used to show whether increased or decreased spending on players correlated with the number of wins a team was able to achieve in a season. This could be utilized by a GM to determine whether spending more on players is worth it solely from the perspective of whether spending on many high caliber players truly increases team success. (Sketch #11- Ben)



This rough sketch is a way to show how effective players are and whether a GM should spend money on the player. There would be a scroll bar with each MLB player and then next to it that player's stats would show up. This would also be linked to another view that attempts to show a player's effectiveness at generating wins and increasing revenue, however that may be calculated. This

could be used by a GM and CEO to determine from both a number of wins and a financial standpoint whether signing a certain player at a given price is worth it or not. (Sketch #12- Ben)

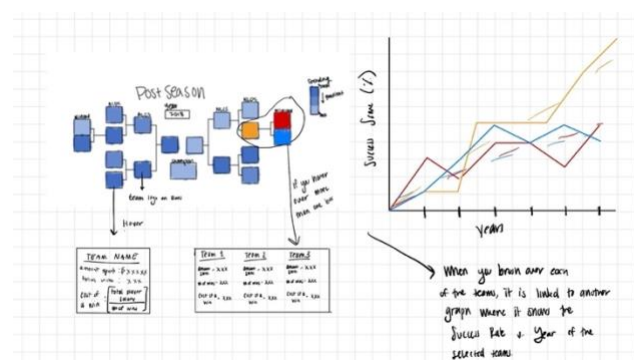
2 GROUP SKETCH SELECTION

We picked visualization #8 (Sam's) as one of our favorite visualizations. This visualization shows where the team finished in their division (measured by how they ranked within their division by total games percent of games won—shown out of 1,000) and how much they spent on their team. Each dot represents a team and is color coded by their division. This visualization would support the task of determining if it takes more money to win a division (the desired outcome of every team). This visual encoding is effective because it would allow the end user to see if they should be spending more or less money to win their division.

We picked visualization #9 (Sam's) as one of our favorite visualizations. This visualization shows the postseason games, and shows whether the higher-spending or lower-spending teams advance in the postseason. This visualization is another way to determine if team spending relates to team success. It is effective because it shows a bracket, which many people in the sports industry are familiar with, and uses shading to identify higher and lower-spending teams.

We picked visualization #11 (Ben's) as one of our favorite visualizations. This visualization shows the relationship between team success and salary over the years. Team success in this case is quantified by a win-loss ratio or percentage. There is also a best-fit line that can be toggled to show the trend of team success or team spending over the years. This visualization would support the task of determining if team spending on players relates to team success. This visual encoding is effective because it would allow the end-user to visualize how team spending correlates to their performance in terms of games won.

Hand-Drawn Design

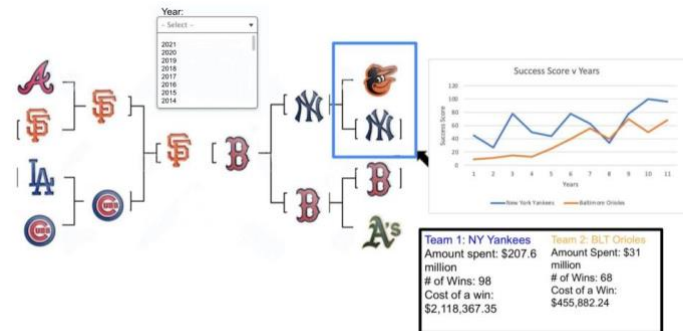


The coordinated visualization tool we designed will allow users to examine the success of a team in relation to their spending amounts. This is done using bracket formation which many users are very familiar with due to sports tournaments making it easy and simple for them to understand the patterns. The bracket will have the team logo displayed in each of the boxes and the background color of the box will show the amount spent from the color gradient such that the darker the blue is, the more money that team spent. Furthermore, when you hover over a box, a hover box will pop up, showcasing the total amount spent by that team in the selected year, the total amount of wings, and the cost of a win. The cost of a win is calculated by dividing the total player salary by the number of wins thus producing a value. The user is

able to adjust the bracket based on what year they hope to examine. We selected a variety of elements from each of the visualizations in our top three. We selected the bracket format from visualization #9, the color schematic, and the use of color to identify the teams along with the spending habits from visualization #8. Furthermore, we selected the formula to determine the cost of a win from visualization #11 which is an interesting statistic for users to examine.

APPENDIX E: DIGITAL SKETCH

3 DIGITAL SKETCH



The domain tasks a user needs to complete are to determine how successful one team is truly able to be based on the amount of money spent on player salaries. Using postseason success is a great way to measure how well a team performed. The goal of any MLB team is to win the World Series and how far a team makes it in the postseason is a clear-cut measurement of success. From there, a user can see how much money each postseason team spent represented by a color gradient where a darker hue means more money spent.

The user will interact with the visualization in a few manners. First, the user will see a postseason bracket for the year chosen in the drop-down menu, where the background color of the boxes of each bracket will be the hue representing the amount of money spent. The darkest hue will be determined by the team in the league which spent the most money and the lightest hue will be determined by the team with the least amount spent on player salaries. Taking inspiration from Sam's visualization #8 which represents dots as team colors, we will overlay the team logos on the boxes in the brackets to demonstrate which teams are the ones succeeding. The visualization will also include a tooltip that generates a pop-up box when hovering over a team in the bracket. In the hover box will be the team's name, total payroll, total wins, and cost of a single win based on total payroll divided by the number of wins. Additionally, if multiple teams are brushed over, the box will include each team's information compared side by side. When brushing over teams the connected visualization will also show multiple teams success rates throughout the years, as a line graph. When hovering over one team the chart will only show a single team.

APPENDIX F: USABILITY TESTING

1 PREPARATION

Our visualization tool showcases MLB data throughout the years. We display this data to allow a user to visualize each team's success (which is measured by both percentage of games won as well as postseason success) in relation to the total payroll for that team. Major League Baseball is a sport with loads of data, and we showcase each team's opening day salary as well as the corresponding percentage of wins for that given season. We display percent wins vs. total number of wins due to the number of games played in a season not being consistent. Additionally, we display the postseason bracket for the selected year because how far a team advances in the postseason are how most fans and organizations measure true success.

Three specific tasks we have determined will be useful for a subject to complete with our visualization tool are to:

1. Analyze the trends of teams spending habits over time
 - We are testing for this because we want to determine whether the visualization makes it easy for the user to compare multiple team's spending habits against each other
 - The outcomes we will be looking for are whether a user can correctly identify which teams spend more money and are increasing or decreasing spending as well as ask them a series of questions about how effective the visualization was in getting the data across in a pleasing manner so that we can understand how to improve our visualization.
2. Compare the spending habits of a given team relative to their performance
 - The goal of this task is to determine whether a user can understand if it is worth it for a team to spend big on players and whether paying up generates more success in terms of both postseason success as well as winning percentage.
 - Outcomes we will be looking for are whether the user can clearly identify which teams had the best ratio of money spent on opening day salary to overall success. If the user can clearly do this, then we will ask whether they enjoyed interacting with the visualization and how we can improve it, so the user experience is more pleasing.
3. Compare the average spending to the spending of particular teams over time
 - The purpose of this task is to determine whether our visualization makes it clear which teams spend more than the league average and which spend less than the average when it comes to player salaries.
 - The outcomes we will be looking for are whether the user can quickly see if a team spent more, less, or about the same amount as the league average. We can measure this through a simple questionnaire where we provide the user a few teams of our choice then ask the user to determine whether they spent more, less, or about the same amount as the league average throughout the years. If they can answer these questions quickly and

correctly then our visualization has proved effective in getting the data across.

2 RESULTS

There were no major issues found through usability testing, however there were a few minor issues found. One thing to note is that neither group had completed projects since we were still working on adding final touches to their projects and visualizations. One thing they liked about our project was the brushing and linking between the two line graphs at the top of our page. They said that it was easy to use, understand, and interesting to compare the two. Another aspect of our tool that they liked was the bracket visualization, however, they said that adding a legend to explain what the bracket was would also be useful along with the spending amount that each color represents. In addition to the scale, they suggested adjusting our color scale so that we had a wider range of colors, making it easier to identify the different spending values. They also suggested a few changes to the project which would allow users who aren't as familiar with baseball to comprehend the project easier such as using the full name of the teams rather than their abbreviations.

1. First, can you analyze the trends of teams' spending habits over time?
 - Yes, they were able to analyze the trend of teams' spending habits over time. This was done using the two line graphs on the top and view each of the teams spending over time. The metrics they were analyzing showed the spending habits of the team over years in a line graph. They said that it was easy for them to analyze the trends and understand what the visualization was showing. We do not need to make a change to this design since they were able to complete this task.
2. Second, could you compare the spending habits of a given team relative to their performance?
 - They were able to compare each team's spending habits relative to their performance over time. They like that we had the two graphs next to each other which made it easy to compare the two graphs. Another aspect that they liked was the brushing and linking which allowed them to view specific ranges of points on each graph. We do not need to make a change to this design.
3. Lastly, please compare the average spending to the spending of teams over time
 - They liked that we had the average line in our visualization which made it easy to compare the team's spending to the average spending and analyze any trends. Our test results did not indicate that a change to our design is necessary since it was easy to use and compare.

One element that we will add to our visualization tool is a tooltip on the bracket that will show the overall spending of the given team, which will further aid the understanding of the bracket and its coloring, giving specific values of the spending. In addition, we will add a legend by the bracket visualization which will show the range of colors that are used in the bracket along with the range of values in the bracket, to give the user a better understanding about a team's spending. Furthermore, we added a gold circle around the winning postseason team to further aid the visualization and add clarity for the user. In our line graphs, we

are adjusting the drop-down menu from the abbreviation of the teams to the full name of the teams, so users can recognize each team. This was added because not all users may be familiar with the abbreviations for the teams. Another element that we are adding is a tooltip on both line graphs to give specific information about each of the graphs and their data points. We hope to have the tooltips show the salary value and percent win for the respective graphs to be able to better compare the salary value and the percent wins for different teams and years.