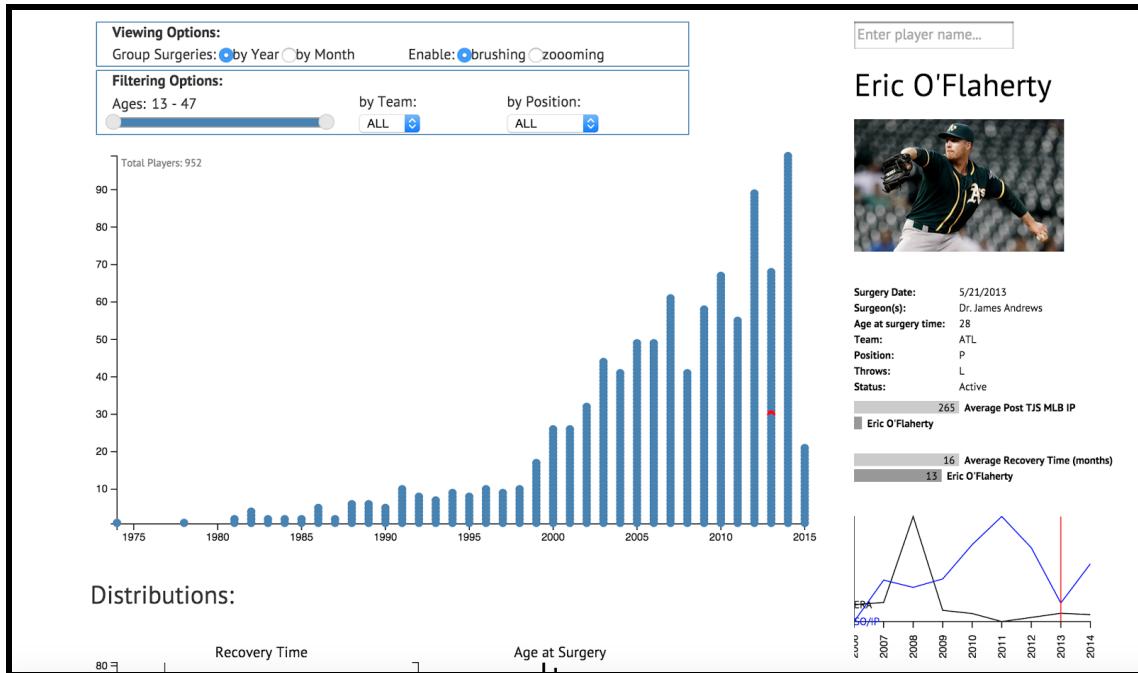


Tommy John Visualization

Process Book

Matt Timoney, Sean O'Neill, Emily Lawton

{mtimoney, seanoneill, emilylawton}@college.harvard.edu



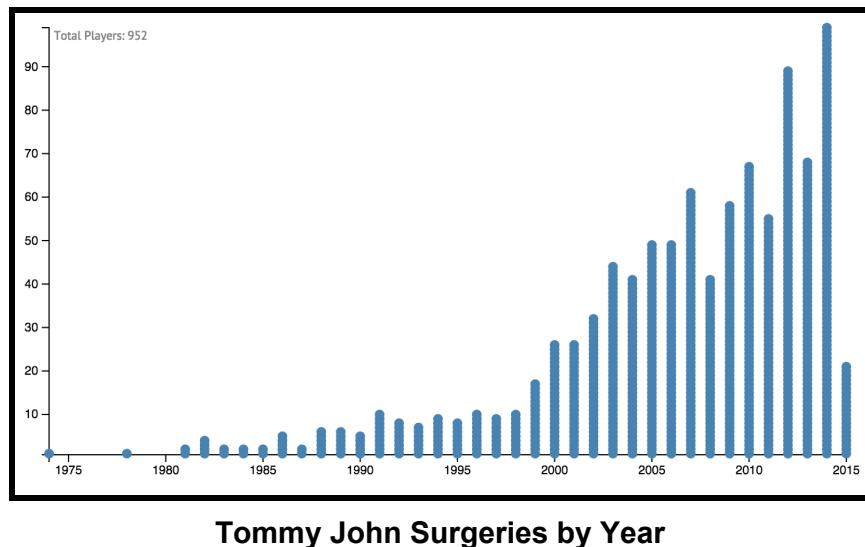
Overview and Motivation:

Los Angeles Dodgers lefty pitcher Tommy John's career came to a startling halt in 1974 when a nagging pain in his pitching arm forced him off the mound. Many thought this was the devastating end to a long and successful career for the 31 year old pitcher already with 124 major league wins. Tommy John was not convinced. He called up Dr. Frank Jobe, the Los Angeles Dodgers team physician. Due to the less sophisticated nature of surgery prior to the 1970s, the majority of injured baseball players avoided surgery at all costs and instead attempted to play through their pain. Understanding that without surgery he would never pitch again, Tommy John was not willing to settle. At John's insistence, Jobe proposed an idea for a new operation. According to Dr. Jobe the odds were stacked up 100 to 1 against Tommy John returning to pitch after the surgery. On September 25, 1974, Dr. Jobe performed the first

ulnar collateral ligament reconstruction for Tommy John, replacing a ligament in the medial elbow with a tendon from elsewhere in the body.

After sitting out the entirety of the 1975 baseball season, Tommy John proved to all those who thought his career was over and returned to pitch for another fifteen years, earning another 164 major league wins to bring him up to 288, the seventh highest among all left-handed pitchers in major league history.

Now commonly known as Tommy John surgery (TJS), ulnar collateral ligament reconstruction now has a complete recovery rate of 85-92%. It has been reported that a third of all current major league pitchers have had the surgery.



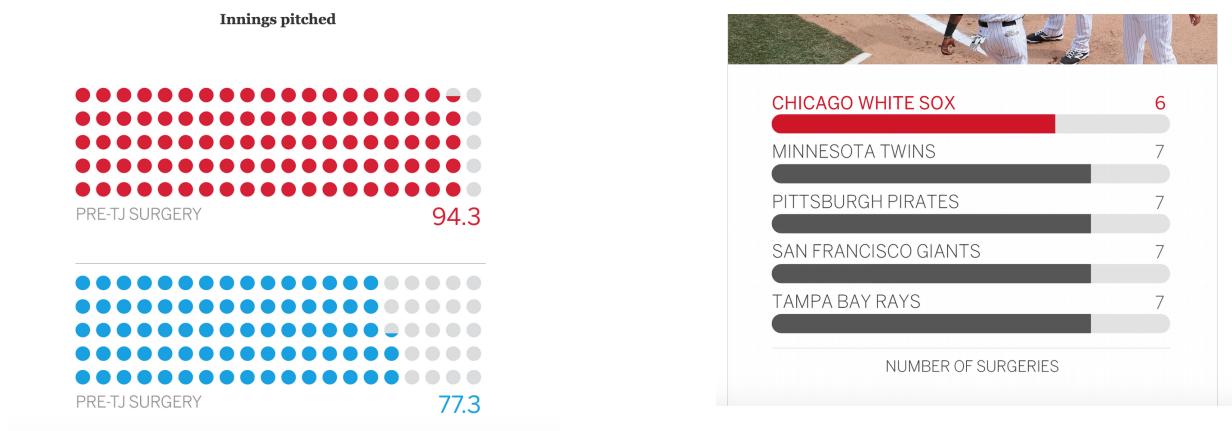
In the last couple years, interest in Tommy John surgery has spiked. Nearly 100 baseball players decided to undergo the surgery in 2014 alone. Because it is such a big phenomenon in baseball today, our goals in completing this project are to raise awareness about Tommy John surgery and related trends, investigate the question of what is causing the recent surges in interest, and begin to understand whether this attraction is justified.

Related Work:

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

http://espn.go.com/mlb/story/_/id/12648769/what-missed-tommy-john-surgery

Though this article was published on April 9, shortly after we selected the topic for our project, it addresses the recent increase in *repeat* Tommy John surgeries via a series of simple visualizations. A close spin-off of the content we intend to visualize in our project, this article certainly serves as inspiration. However, we were not impressed by the design choices made here.



Other literature:

- <http://bleacherreport.com/articles/1675444-tommy-john-surgery-the-realities-and-myths-of-sports-most-famous-operation>
- http://espn.go.com/mlb/story/_/id/10831700/is-mlb-midst-tommy-john-epidemic

Questions:

Questions we initially wanted to have answered upon beginning the project include:

- What ages are most highly represented among players undergoing TJS?
 - Is the average age of players undergoing TJS changing over time?
- Is improved pitching performance correlated with recovery from TJS?
 - One possible explanation for the surge in interest is the belief that pitchers can throw harder in the aftermath of recovery from Tommy John surgery. In fact, surgeons have been asked by the parents of young pitchers to perform the surgery on their uninjured sons in the hopes of increasing their child's performance and future career prospects. We in turn intend to investigate whether improved pitching performance (measured along several different evaluation metrics) is correlated with recovery from TJS. This visualization project thus has the potential to dispel the potential misconception that Tommy John surgery directly causes improved performance.

- Note: This ended up being a critical question that we considered as we worked on the final iteration of our design. This information is viewable as part of the playerVis view.
- What baseball teams do the players that undergo TJS play for?
 - Are there any geographical trends that can be unearthed?
 - Note: We ultimately decided that visualizing the surgeries graphically (as represented by what team the player is associated with) does not provide very useful information. Our visualization does enable the user to view the frequency of TJS instances on a team by team basis (over time). This is viewable as a part of the countVis view.
- How is player salary impacted as a result of opting to undergo TJS?
 - Note: In the end, we did not see this as particularly valuable and did not have easy access to salary information for many of the players included in our dataset.

As noted above, in the time that has passed since the creation of our project proposal, the main questions we wanted to answer evolved slightly. This was also a product of growing increasingly familiar with trends in the data we had access to. Important questions we ended up focusing on include:

- What ages are most highly represented among players undergoing TJS? How is this changing over time?
- Is improved pitching performance correlated with recovery from TJS?
 - Specifically, we intend to look at player ERA (earned run average) along with number of strikeouts per inning pitched as a function of time.
- What baseball teams do the players that undergo TJS play for?
- Are the majority of players that undergo TJS pitchers?
- How has recovery time changed over the years since the surgery was first introduced?

Data:

Tommy John Surgery Data:

<https://docs.google.com/a/college.harvard.edu/spreadsheets/d/1gQujXQQGOVNaiuwSN680Hq-FDVscwvN-3AazykOBON0/edit#gid=0>

We obtained the vast majority of our data from the above dataset. We wrote a simple python script to convert the above .csv file to JSON and subsequently cleaned up the data (renamed columns, formatted dates, etc.). The spreadsheet is maintained by Jon Roegele, baseball analyst and writer for The Hardball Times (<http://www.hardballtimes.com/>). This continuously updated Google Spreadsheet includes data (TJ Surgery Date, Team, Position, Throws,

Recovery Time, etc.) on nearly 1,000 baseball players that underwent Tommy John surgery, beginning with Tommy John himself in 1974.

For a detailed look at how we cleaned up the data, see the `formatData()` function in `index.html`.

Pitching Performance Data:

<http://www.seanlahman.com/baseball-archive/statistics/>

We wrote another python script to extract the relevant pitching statistics for the `playerVis` from Sean Lahman's Baseball Database, which included pitching statistics from 1871 to 2014. The metrics that we extracted included player name, yearly ERA, number of strikeouts per year, and number of innings pitched per year. See `getPitchingData.py` for more details.

While we had hoped that the `mlbamid` number associated with each player could be used as a key to find the pitching statistics of each pitcher that underwent the surgery, this id was not used in Sean Lahman's database. Thus, extracting the relevant performance statistics was not as straightforward as we initially hoped. As an alternative, we created a list of {player name, id} pairs from the `master.csv` file in the Lahman's database and used that list to extract the pitching statistics we wanted (mentioned above). We were then able to match up the names from the `master.csv` file and names in the Tommy John Surgery spreadsheet.

Pictures:

<http://mlb.mlb.com/images/players>

The pictures incorporated in the `playerVis` were scraped from the above site. These pictures reside in the `img/players/` directory in our repository on github. See `getPictures.py` for more details.

Exploratory Data Analysis:

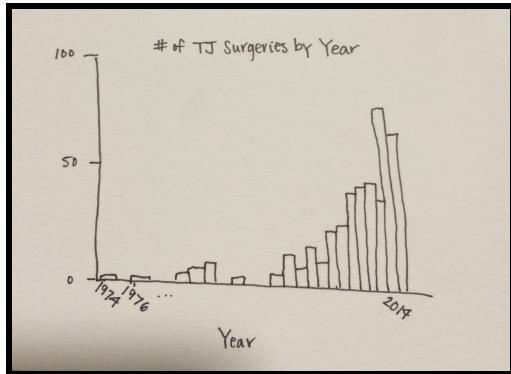
Because two of the three of our project members have played baseball for almost the entirety of their lives, we drew a lot of insights for developing our visualizations based on intuition that these players built up over the years - intuition that was gained through playing baseball and through keeping up with news in the baseball world.

In addition, we drew inspiration and were able to make hypotheses from a few articles in the news that commented on the phenomenon of increased Tommy John surgery popularity.

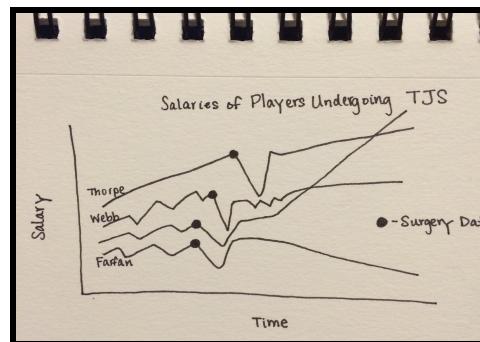
Many of these articles addressed the common belief that Tommy John surgery is an enhancement procedure rather than a reparative one.

Design Evolution:

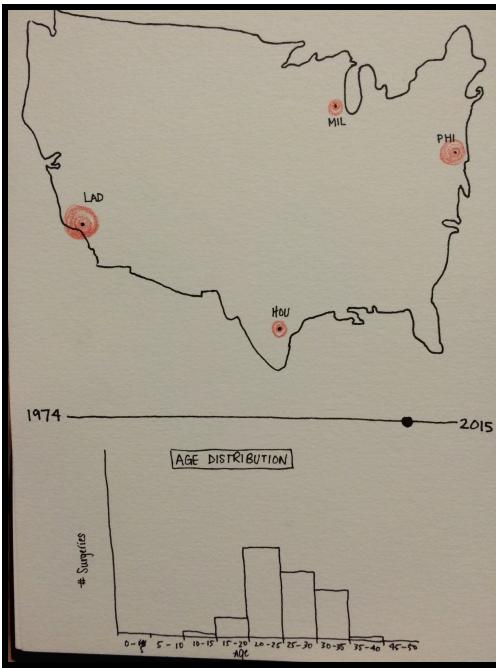
Some of the initial sketches we made for our visualization are included below:



Vis #1: A simple bar chart showing the number of Tommy John surgeries performed from 1974 to the current year. This sketch served as the basis for our countVis view.



Vis #2: A line chart showing the change in salary of players undergoing TJS (with the time at which the surgery was performed clearly indicated). While we did not end up visualizing salary data of players undergoing TJS, several components from this design still served as inspiration in our final design choices. We decided to use a line chart to instead visualize pitching performance statistics. On this chart too, we clearly indicated the time at which the surgery was performed so that viewers can see if the surgery had an impact on pitching stats.



Vis #3/4: A map showing the geographical location of the teams of players undergoing TJS over time along with accompanying histogram displaying the age distribution of patients for the given year. As explained above, we decided against visualizing the geographical location of the teams of TJS patients, we did carry the bar chart showing the distribution of ages of TJS patients through to the final iteration of our design.

Design Studio (April 14):

We collaborated during the design studio with **Cynthia Meng** and **Hugh Zabriskie**.

The feedback we received during the design studio was generally positive. The comments were both fair and constructive. Cynthia and Hugh were not familiar with Tommy John surgery and its popularity among baseball pitchers - attesting to the fact that the general public (one user-type we are concerned with in developing our visualization) is not aware of this phenomenon.

Specific ideas that were brought up during our discussion were as follows:

- Being able to interact with countVis (our main visual that shows the number of surgeries performed as a function of time) and select individual surgery instances was well received. This addition to what would otherwise be a simple bar chart enables viewers to explore specific information (recovery time, surgery date, surgeons, etc.) for individual players.
- We discussed the pros and cons of representing our visualization as a story versus as a single compact set of views that does not require any scrolling.

- Note: As we worked on developing our visualizations, it was the case that an interesting narrative began to take form. We decided to incorporate both a story format and a compact view by placing all of the views in the middle of our visualization site and enabling the user to scroll down further to read up on our own interpretation of what the visualizations exhibit.
- One last thought that Cynthia and Hugh brought up was that we could extend our project to not focus solely on Tommy John surgeries but to also consider other injuries to baseball players as well. This would involve changing our playerVis to include a timeline of other injuries and surgery instances.
 - Note: Given the timeframe we had to complete the project, we kept the scope limited to Tommy John surgeries.

For more details, see `design_studio_feedback.md`.

Meeting with Hendrik (April 14):

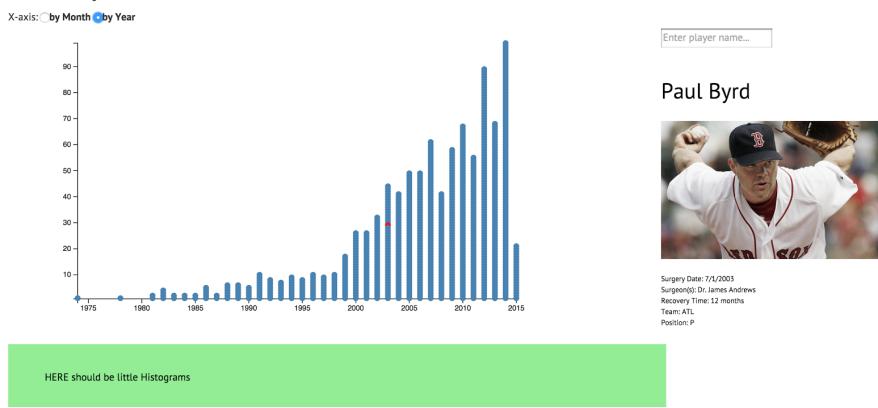
Our meeting with Hendrik following the design studio was constructive. We received feedback on our initial design choices and made several revisions to the features we included in our proposal:

- We discussed our listed must-have features and optional features, ultimately deciding that visualizing player salary over time should be an optional feature and that visualizing player performance before and after surgery should have higher priority.
- When talking about visualizing the geographical location of players undergoing TJS over time, we decided that unless we see very distinctive trends in basic graphs a map is not necessary.
- During our conversation, we brought up how we intend to make our visualization interactive. These ideas include brushing functionality on the `countVis` (which shows the number of Tommy John surgeries performed over time) and the ability to search for specific players.
- Continuing the conversation from our design studio, we discussed the pros of presenting our visualizations as a story. As mentioned in our design studio feedback, the direction of our story will continue to become more apparent as our visualizations develop.

Project Milestone (April 17):

The current state of our visualization is as follows:

Tommy John Vis

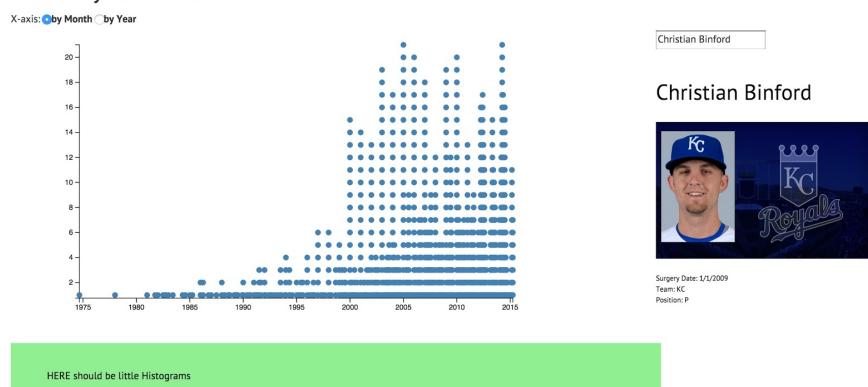


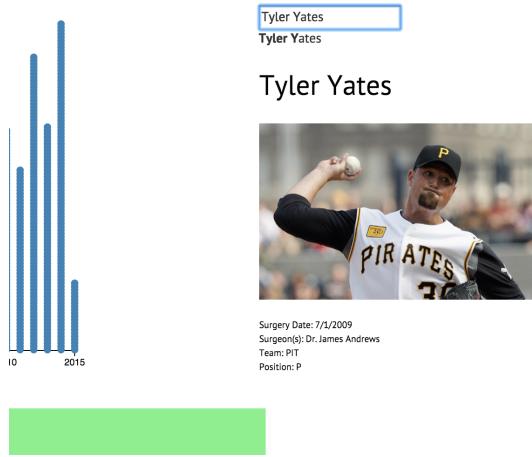
The above screenshot shows `countVis` (which visualizes the number of Tommy John surgeries over time, satisfying one of our must-have features) on the left and `playerVis` (which shows the player name and picture along with various information regarding his surgery) on the right. The x-axis of the graph on the left can be changed from month to year by switching the radio button and vice versa. Hovering over a surgery occurrence causes the node to turn red and the player's name and picture along with other information is shown at right.

Just from this initial visualization, we can already see that there has indeed been a huge spike in the number of Tommy John surgery occurrences in the past five years. Why the dramatic increase? Hopefully we may have an answer to this question at the completion of our project.

We have also begun implementing a slider that enables viewers to select a range of ages to filter the graph by.

Tommy John Vis





These next two screenshots demonstrate the player search functionality. Viewers may enter in a player's name to bring up their picture and corresponding Tommy John surgery information. This is useful if a viewer has a specific player in mind they want to check up on.

As we met each of the goals we set out to achieve by Milestone 1, we are satisfied with the progress we made thus far on our project. Comments on this iteration of our design are as follows:

- We feel as though the ability to hover over surgery instances to view individual player information is a strong asset.
- The spacing in `countVis` is distracting, and we are not very fond of our decision to use circles to represent surgery occurrences.
- With our decision to represent surgery instances with circles in `countVis`, brushing may not be as intuitive as it is with an area layout.

Back to the drawing board:

- When users search for players or hover over specific surgery occurrences in `countVis`, show the picture of the player and the player name in the backdrop of `countVis` itself.
- Enable users to search not only for players but also for teams.
- Possibly have surgery occurrences be nodes that can be grouped by team, age, or a performance metric.

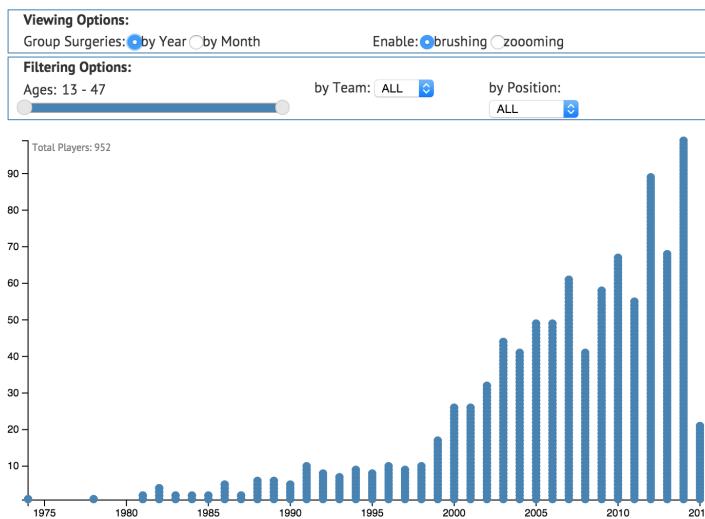
Meeting with Hendrik (April 23):

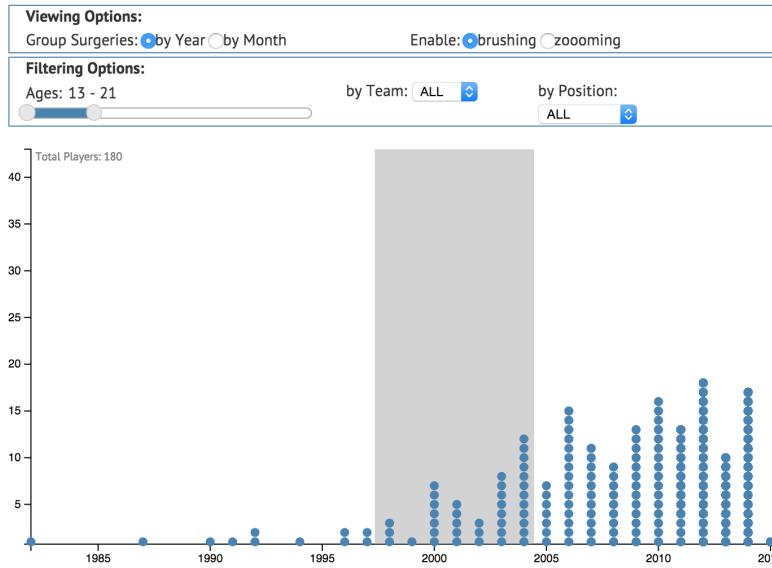
- Make search box intuitive

- The user must simply type in a player's name. One suggestion is provided as they type. We made this decision in order to keep visible the playerVis which appears directly below the search box. Once the user has come across the player whose TJS information they would like to view, they must press Enter.
- split aggregate & individuals views
 - We split the aggregate and individual views by visualizing all aggregate data on the left side of the site and the individual view on the right-hand side of the site.
- minimize transition time of histogram changes due to brushing
 - Because transitions made for distracting node movements, we decided to drastically minimize transitions
- make a line graph of average recovery time over the years
 - Instead of making a line graph of average recovery time over the years, we created a histogram that responded to brushing that showed the distribution of recovery time.
- show distribution over ages
- compare early years to later years (possibly implement multiple brushing)
- select two individuals and show their statistics/features side by side
 - comparison plots

Implementation:

countVis View





Our main view, countVis, (see countvis.js) is our main visualization. It depicts the frequency of Tommy John Surgery instances across time (1974 to present). The basic intention of this design is to show the drastic popularity in Tommy John surgery instances over time.

Surgery instances are represented as nodes (circles). Hovering over circles updates the playerVis view. Clicking a node locks the playerVis on that specific player.

Interactive elements of this view include grouping surgeries by year or by month. Other interactive elements as shown above include the ability to filter the data to show only the surgery instances for a specific age range. The viewer may also select a team from the drop down menu or show only the frequency of surgery instances for position players or pitchers.

We also implemented both brushing and zooming on this view. Brushing filters the data to include only surgery instances in a particular time period and subsequently updates the age and recovery time histograms below countVis. Zooming enables the user to get a close up look at the frequency of surgery instances. It should be noted that the brushing and zooming functionalities cannot be experienced simultaneously. This is why we chose to utilize radio buttons for these features - only one of the two options may be selected at a time.

This view is thus able to answer questions stated earlier such as which baseball teams do the pitchers that undergo TJS play for and are the majority of players that undergo TJS pitchers as opposed to position players.

playerVis View

Eric O'Flaherty

Eric O'Flaherty

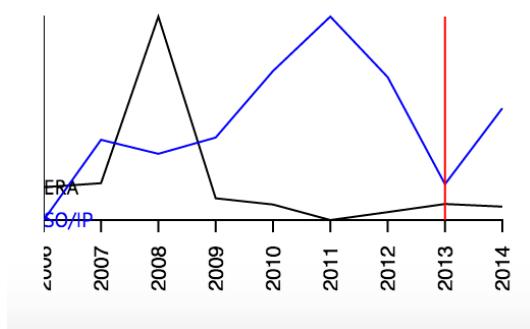
Eric O'Flaherty



Surgery Date: 5/21/2013
Surgeon(s): Dr. James Andrews
Age at surgery time: 28
Team: ATL
Position: P
Throws: L
Status: Active

265 | Average Post TJS MLB IP
Eric O'Flaherty

16 | Average Recovery Time (months)
13 | Eric O'Flaherty



The second view, playerVis (see playervis.js) visualizes information pertaining to individual players who underwent Tommy John surgery. The search box at the top of the view enables users to search for players by typing in a player name and pressing 'Enter' to lock the view on that particular player. One suggestion is given via the search box as the user types a player name.

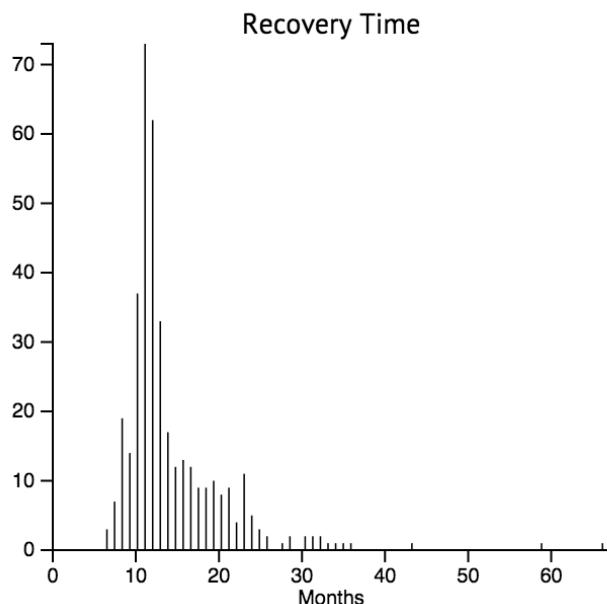
If a player is selected from the search box, the player's node is highlighted in red in the countVis view. This allows the user to view when the player underwent surgery as compared to the group.

In addition, this view displays a variety of information specific to the selected player (including picture, surgeon(s), age at time of surgery, team, position, throwing arm, and active status).

Moving down the view, the first set of bars shows the number of post TJS MLB innings pitched as compared with the average across all other players that underwent TJS. The second set of bars shows the recovery time (in months) of the selected player as compared with the average recovery time across all other players that underwent TJS. Note that these bars are only visible if the recovery time data is actually available for the selected player. This is why the playerVis view may appear to have more information for some players as compared to others.

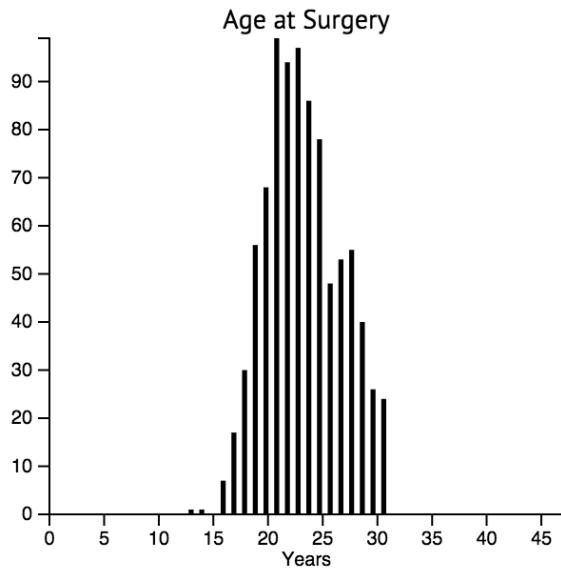
The last (and potentially most important) component of the playerVis view is the line chart showing two performance metrics (ERA and SO/IP) over time. ERA (Earned Run Average) represents how many earned runs a pitcher allows per 9 innings. SO/IP can be viewed as a measure of how electric a pitcher is (something that would only be high if a pitcher's arm is in good shape). The selected player's ERA can be traced over time by the black line, and the player's SO/IP can be traced over time by following the blue line. If the range of years of pitching data we have for the selected player includes the year of the surgery, that surgery year is denoted by a vertical red line. This allows the viewer to easily compare the pitching performance statistics pre and post procedure.

recoveryVis View



The next view, recoveryVis, (see recoveryvis.js) clearly shows a distribution of the recovery time of players that underwent TJS. This histogram responds to the brushing and filtering options that are selected in countVis. recoveryVis is thus able to answer the question of how recovery time has changed over the years since the surgery was first introduced.

ageVis View



The last view, ageVis, (see agevis.js) clearly shows a distribution of the ages of players that underwent TJS. This histogram responds to the brushing and filtering options that are selected in ageVis. This view is thus able to answer the question of what ages are most highly represented among players undergoing TJS and how this is changing over time.

Evaluation:

Overall, we learned a great deal both from working on developing our visualizations and from the visualizations themselves.

When evaluating our visualizations, it is important to look back at the questions we were attempting to answer and address whether we accomplished what we sought out to discover:

- Is improved pitching performance correlated with recovery from TJS?
 - As described on our site, there is a great deal of speculation in the baseball world surrounding the idea that Tommy John surgery may aid pitchers. In fact,

some people view it partially as an enhancement procedure instead of a simply reparative one.

- Our findings are that for pitchers that are able to record significant innings post surgery, their post surgery stats are anywhere from far worse to far better than pre surgery. The outcome seems to just depend on the individual. For pitchers who are not able to record a significant number of innings following the surgery, on the other hand, the procedure seems to generally have a negative impact on pitching ability.
- Tommy John surgery is rarely undertaken unless the athlete is trying to continue playing baseball post-surgery. Therefore, when a pitcher is not able to play much after surgery, we can deem the surgery and rehab process a failure.
- As shown by the six pitchers that we chose to analyze in the analysis on our site, the results of pitching performance after surgery are highly varied. The metrics we chose to quantify pitcher success were Earned Run Average (ERA) which is how many earned runs a pitcher allows per 9 innings and strikeouts per inning (K/IP) which is a good measure of how electric a pitcher is (something that would only be high if a pitcher's arm is in good shape).
- Scott Erickson and Jorge De La Rosa are both good examples of pitchers that were able to continue to play post surgery but were never able to return to their peak performance. We know this because our visualization shows that all of Erickson's ERAs following the surgery are higher than his before the surgery. Additionally, his strikeouts per inning declines substantially post procedure. Similarly, the year after De La Rosa's surgery he has his worst ERA and strikeouts per inning in the last 10 years.
- Carl Pavano and Neal Cotts, on the other hand, both seem to have benefitted from the surgery, at least temporarily. We see Pavano's K/IP numbers steadily declining until he gets the surgery and they spike back up. Cott's ERA is steadily increasing until he gets the procedure and his ERA drops. In both cases, the numbers appear to return to pre-surgery levels after a few years but the immediate results of the surgeries is enhanced pitching ability.
- Graphs depicting these trends are displayed in the analysis portion of our site.
- What baseball teams do the players that undergo TJS play for?
 - This question may be answered by selecting particular teams from the drop down menu in countVis. From this view, it is clear that the Atlanta Braves have had many more players undergo TJS than most all of the other teams featured in our visualization (47 players).
- Are the majority of players that undergo TJS pitchers?
 - This question may be quickly answered by selecting either PITCHERS or POSITION players from the drop down menu in countVis. It is clear from this view that the vast majority of players that undergo TJS are indeed pitchers.
- How has recovery time changed over the years since the surgery was first introduced?
 - The information displayed by our visualization are quite indicative of the fact that rehabilitation programs following Tommy John Surgery are improving. Over

time, the average recovery time stays relatively constant at around 16 months. However, what does change drastically is the spread of the distribution of recovery times. The distribution tightens up dramatically as time goes on, indicating to us that the rehab programs followed by athletes undergoing TJS are becoming more uniform and successful over time. The prevalence of outliers on the long end (indicative of athletes that did not rehab properly, suffering setbacks in their journey back to baseball) and on the short end (indicative of athletes that are trying to rush to get back to peak playing form) drops severely over time.

- Graphs depicting these trends are displayed in the analysis portion of our site.
- What ages are most highly represented among players undergoing TJS? How is this changing over time?
 - This question may be answered by playing with brushing in countVis and observing how the ageVis is updated. It is apparent from this histogram that the mode of surgery ages has decreased over time. This may be attributed to the belief that Tommy John surgery may enhance performance as younger pitchers may be seeking out the surgery when it is not completely necessary. This trend may also indicate that pitchers are suffering from overuse at a younger age as time goes on.

Given the short time period with which we had to work on this project, we are satisfied with how well our visualization works and the narrative that can be derived from the visualizations. Given more time, however, we would incorporate additional functionality and iterate upon our design even more. A non-exhaustive list of changes we would likely make are as follows:

- Allow for multiple brushes on countVis to compare the data of multiple selected time intervals.
- Work harder to derive data, instead of simply juxtaposing the data from two selections.
- Show the average throwing velocity of each pitcher undergoing TJS before and after surgery.
- Visualize TJS recovery time as a function of the surgeon who performed the surgery.
- To produce a more robust visualization, search for a more complete dataset (even though the one we found appears to be most the most complete TJS dataset currently available).

Conclusion

In closing, our visualization does not support the claim that Tommy John surgery enhances pitcher performance as many would like to believe. As shown in our analysis, the impact of Tommy John surgery on baseball players is much more complicated than can be expressed with a simple *does improve* or *does not improve* performance. From our visualizations, we argue that rehabilitation programs that athletes follow post Tommy John surgery are becoming more uniform and successful over time. Our visualization sheds light on this interesting phenomenon of the drastic rise in Tommy John surgery instances over the last

decade among baseball players and helps seek to unearth trends that have cropped up along with the increased popularity.

Thank you, and we hope that you enjoy interacting with our visualization!