Big Data Basketball: Visualizing the Floor

Brian Ge, Connor Bitter, Ryan Plunkett

Project Proposal

As technology has improved, its effects have slowly found their way into the world of sports; teams in all leagues are hiring more and more analysts to investigate and make sense of new data as it becomes available to them. Compared to a sport like baseball, in which a discrete series of events can be initiated only when a pitch is delivered, the fluidity and complexity of basketball present unique challenges in analysis, as the ball and players are in motion nearly constantly, often independent of one another. However, with the advent of motion capture technology, new player-tracking data is becoming more prevalent, and it offers a unique opportunity to investigate basketball's core principles of defense, pacing, and spacing.

Inspired by Kirk Goldsberry's "Department of Defense" article (http://grantland.com/features/department-of-defense/), which was based on research conducted at Harvard and published at the Sloan Analytics Conference (http://www.sloansportsconference.com/wp-content/uploads/2015/02/SSAC15-RP-Finalist-Counterpoints2.pdf), our group is interested in visualizing the possessions of a basketball game, representing both the players and ball using icons and updating our graphic based on their motion in real-time.

In addition to this motion tracking, we also hope to visualize individual players' performances with the use of heatmaps and shot charts, perhaps identifying undervalued talent that has simply been overlooked when evaluated using only traditional metrics. We would also like to look at recent historical trends of shooting in general -- specifically the move from the mid-range game to three-pointers.

In order to obtain this data, we plan to contact the Harvard Sports Analysis Collective, which has partnered with the Boston Celtics for similar data in the past, and the Harvard Men's Basketball team, which records detailed statistics for each of their games. In addition to these datasets, we may augment our work using Basketball-Reference and KenPom, sites dedicated to analyzing both NBA and NCAA basketball.

Detailed Project Plan

- **Names** of all students who have **actively** worked on the project plan: Ryan Plunkett, Brian Ge, Connor Bitter
- Definition of **goals and tasks** of the final project (1-2 pages): Below
- A **description of your data** and where you will get the data from (at least concrete ideas on where to acquire the data):
  - We found a CSV on Kaggle (https://www.kaggle.com/dansbecker/nba-shot-logs/data) that includes information regarding a shooter's distance from the basket at the time of his shot, how many dribbles he had taken/how long he had held the ball, the nearest defender's name, the nearest defender's distance from the shooter, and a binary response variable indicating whether the shot is made or missed.
  - We also found another site (http://buckets.peterbeshai.com/app/#/playerView/201935_2015?playerSelector=true) that includes the actual positioning of the player on the court at the time of the shot, and we plan on contacting the site's creator to see if he could share the data with us.
- Visuals:
  - At least **3 sketches** of visualization ideas for your project: Below
  - A sketch of an **interaction storyboard**: Below
  - A sketch of your **webpage layout/storytelling**: Below
- A **project timeline** (with milestones when you are planning to finish which feature):
  - Sunday - Email Peter Beshai about data
  - Wednesday - Load whatever data we have and clean/wrangle
  - Thursday - Prepare for presentation in section with our slides
  - Saturday - Have basic visualizations done (probably no interactivity), make sure to structure correctly
  - Monday - Skeleton of storytelling website
- A **feature list** (with must-have, good-to-have, and optional items)
  - Must have:
    - Interactive Shot charts with percentages based only on distance to basket
    - Line chart of shooting percentage as a function of defender distance
    - Line chart of shooting percentage as a function of shot clock time remaining
    - Filter by team/player
  - Good to have:
    - Shot chart for different areas on the court, filter by defender distance? Specific defender? Team? Specific shooter?
    - Team vs team side by side comparison
  - Optional items:
    - Brushes/filters for parts of season

- - ■ Sample motion tracking video/gif for specific plays to help storytelling (no aggregation)
  - A **description of team roles:**
    - ○ Ryan - In charge of obtaining/aggregating data, wrangling data for filters, developing storyline for site and supporting with designs/coding
    - ○ Brian - In charge of defining goals for the project and deciding what we need/want/could have, website layout, coding visualizations
    - ○ Connor - In charge of drawing initial storyboards and sketches, implementing interactivity, overall design of website and adding front-end touches

Goals and Tasks

For our project, we have decided to focus on the 2014-2015 NBA Season in which the Golden State Warriors took the association by storm, showcasing a revolutionary new playing style that earned them the best record in the league, an NBA championship, and an MVP award for Stephen Curry. By looking at shot data across the league from that season, our goal is to put together a convincing narrative as to what the best teams did well and the worst teams did poorly in the shifting climate of the NBA. We believe that by visualizing and analyzing a variety of advanced metrics such as defender distance, time left on the shot clock, number of dribbles, etc. we can uncover a detailed story of how the league changed that year, and the best ways to adapt.

To do this, we have a variety of tasks to accomplish. We have a dataset that contains almost all of the relevant information, and have reached out to someone who has worked on a similar project that might have access to even more detailed data. We need to clean and wrangle this data to make it usable for our project, and then we have to decide which metrics we would like to focus on.

Next, we have to design our visualizations with our narrative in mind. To do this, we will first need to analyze the data that we have access to and see what kind of narrative this will end up being; though we know that we're hoping to discover which of these advanced metrics contribute the most to winning, the story itself has the potential to change a lot depending on which metrics these end up being. To figure out the story and really finalize a structure for the website, we need to take a close look at the data to see which ideas we have for visualizations will end up being the most meaningful.

Once we have a concrete idea of what the story will be, we can start the process of designing the visualizations themselves. This includes the static look of the visualizations, keeping in mind all of the principles we've learned in class, along with interaction storyboards and what kinds of filtering and focusing mechanisms will be most useful to tell our story. We will likely have options to filter based on players, teams, shot distance, part of game (start of first quarter, end of half, last two minutes, etc.), and whichever more advanced metrics we find affect shooting percentage in a meaningful way.

With the visualizations designed and story in our minds, we also need to go about laying out the overall structure of the website, including non-visualization elements such as textual elements, images, videos, and other analysis that can help inform how we want users to interact with the visualizations, and the trends we they will see. Once we have this layout, we can start to lay down the groundwork for our codebase, structuring the html and javascript files in ways that will allow us to easily integrate all of the different parts of the website. Doing this in a smart and consistent way will be crucial to how smoothly the implementation phase of this project goes.

Once we have laid down the groundwork, it will then be time to implement the visualizations we have settled on as well. This will probably take a good amount of time, as many of our proposed visualizations are definitely non-trivial. At the same time, we will be putting the rest of the website into code as well, such as textual elements and style guidelines. Finally, if we have finished all of these steps and find ourselves with time leftover, we can look into smaller, difficult tasks that aren't crucial to the website that would be cool to have.

These are the rough goals and tasks that we imagine will be a part of our project; at the end of the day, we hope to be able to construct a convincing picture for how the game of basketball was being played in the revolutionary 2014-2015 season, what was working and what wasn't, and what aspects teams should start and have started focusing on in the modern game.

Storytelling

Below is a basic outline of our website (you should zoom in!). First, we'll explain the goals described above. Then, we'll get a bit deeper into each of the insights we find from the data. Lastly, we'll use the shot chart to bring all of these insights together in an interactive, informative way.
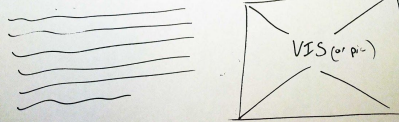
Users will get the most out of our website if they are well-acquainted with basketball, with knowledge of its rules, strategy, and terminology. We're okay with limiting our audience due to the steep learning curve of the material -- those who would be interested in our insights would likely already have a good feel for the game, so we don't need to take on the burden of teaching others.
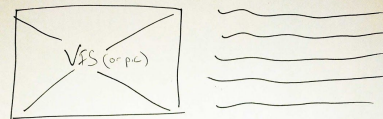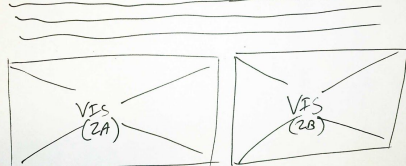
How Golden State changed the league in
2014-2015

Full Screen
Picture
(GSW
(Champions)
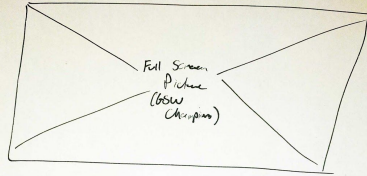
73 Wins — How?

_____   _____
_____   _____
_____
_____

Insight 1: X changed Y

_____
_____   VIS (or pic)
_____
_____
_____

Insight 2: A changed B

VIS (or pic)   _____
                      _____
                      _____

What about 3-pointers?

_____
_____

VIS   VIS
(2A)   (2B)

Distance of Defender

_____   VIS
_____   (3A/B)
_____
_____
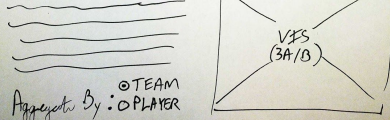
Aggregate By: ⊙ TEAM
              ⊙ PLAYER

Putting it together: Interactive Shot Chart
_____

VIS
(#4)

Summary

_____
_____   Picture
_____
_____
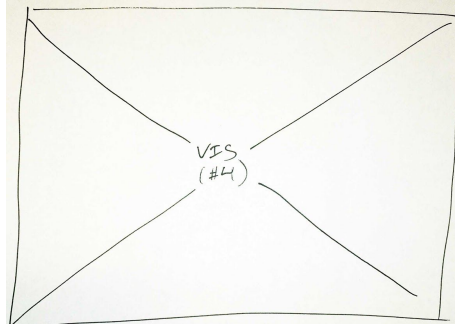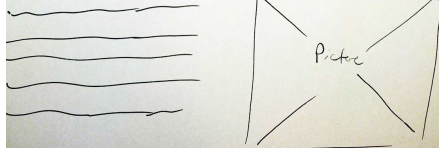
Sketches

These sketches are based on theoretical data for different possible insights. Our actual visualizations will depends on what our data shows.

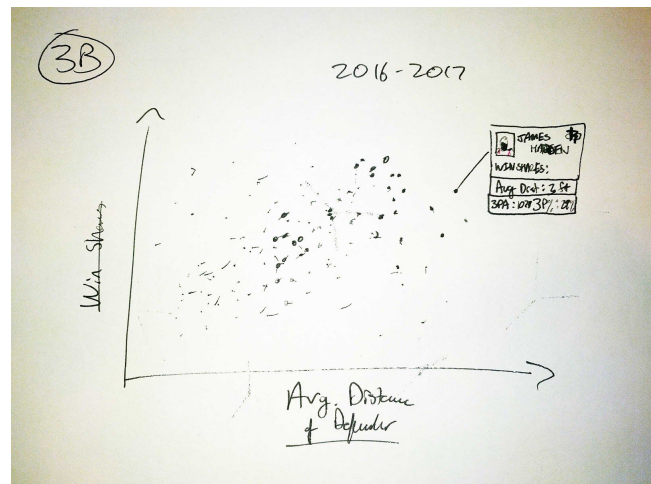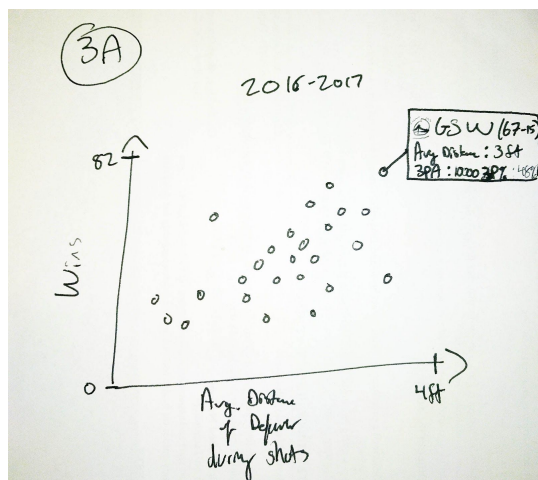*Sketch 1: What makes a good shot?*
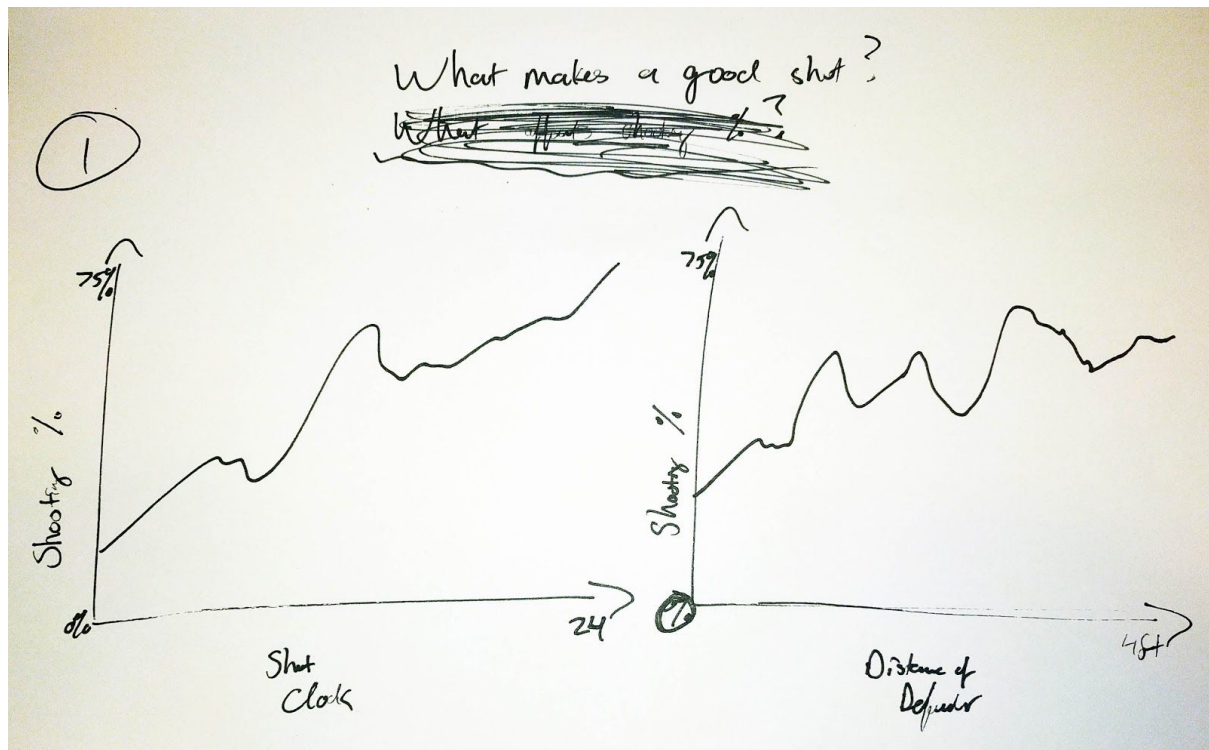
This is two visualizations of a single season's overall shooting percentage as a function of either: 1) time remaining on the shot clock or 2) distance of defender.

Sketch 2: Historical 3-point Shooting Trends
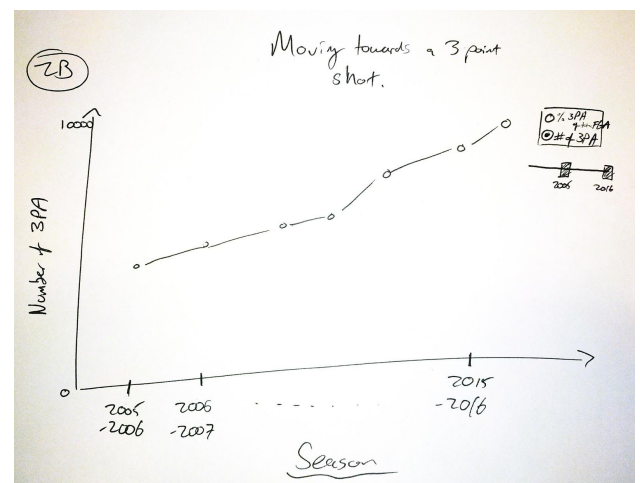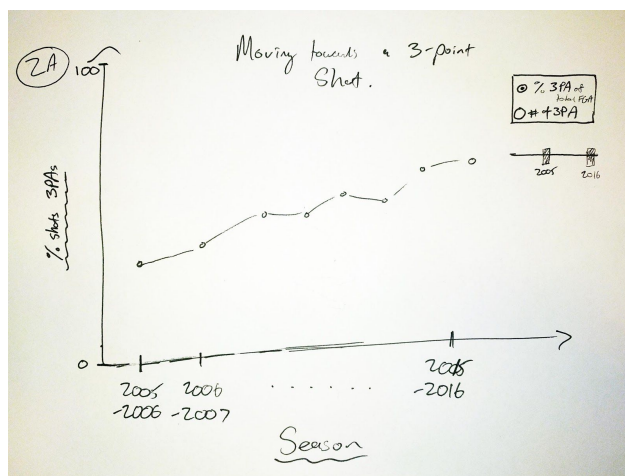
These visualizations show historical trends related to 3-point shooting.

*Sketch 3: How does do defenders affect winnings?*

These visualizations show how the average distance of a defender affects winning, aggregated by either 1) team (measured in wins) or 2) individual player (measured in win shares).

Sketch 4: Interactive Shot Map
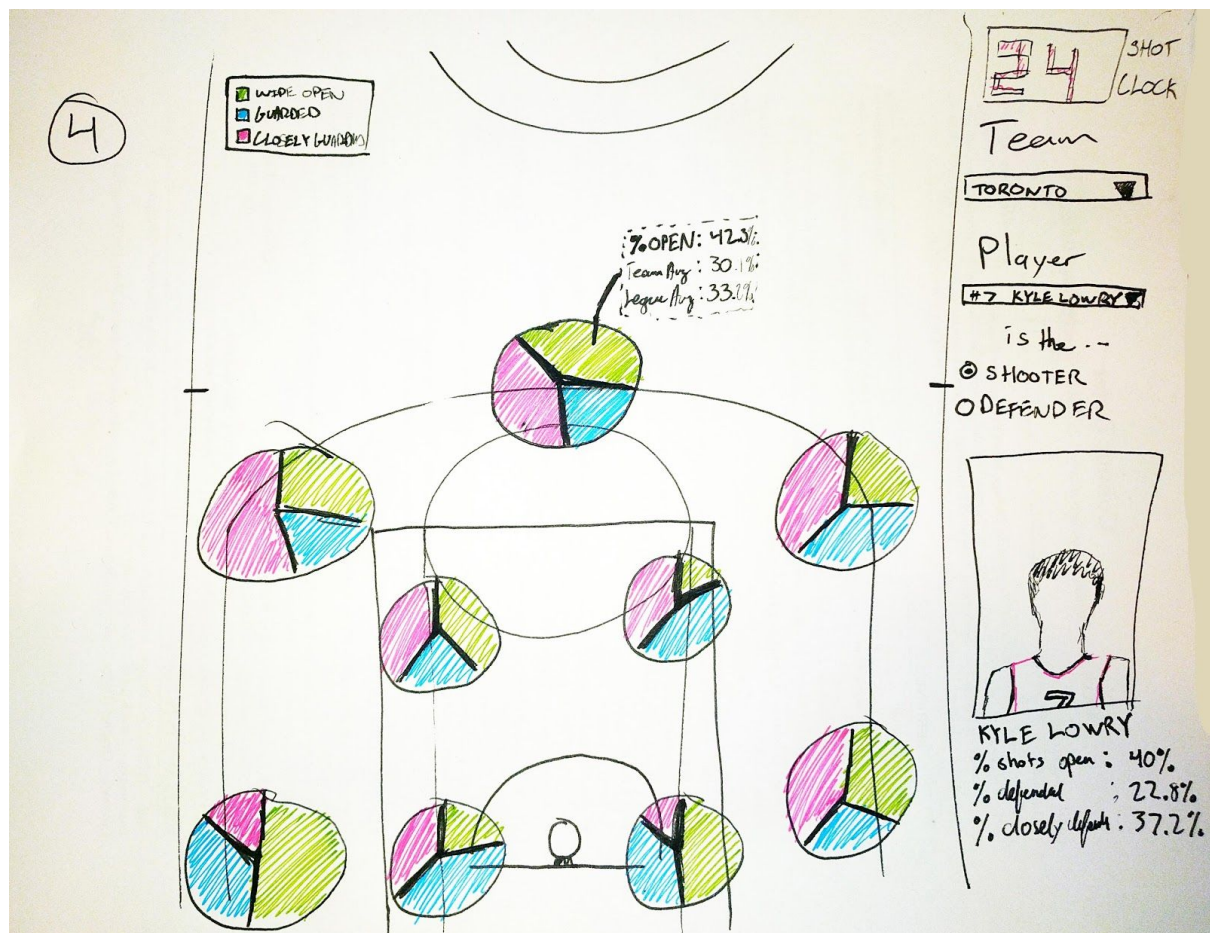
Interaction Storyboard

The first image is of the general interaction is less of a storyboard. The second image is a storyboard of our interactive shot chart.
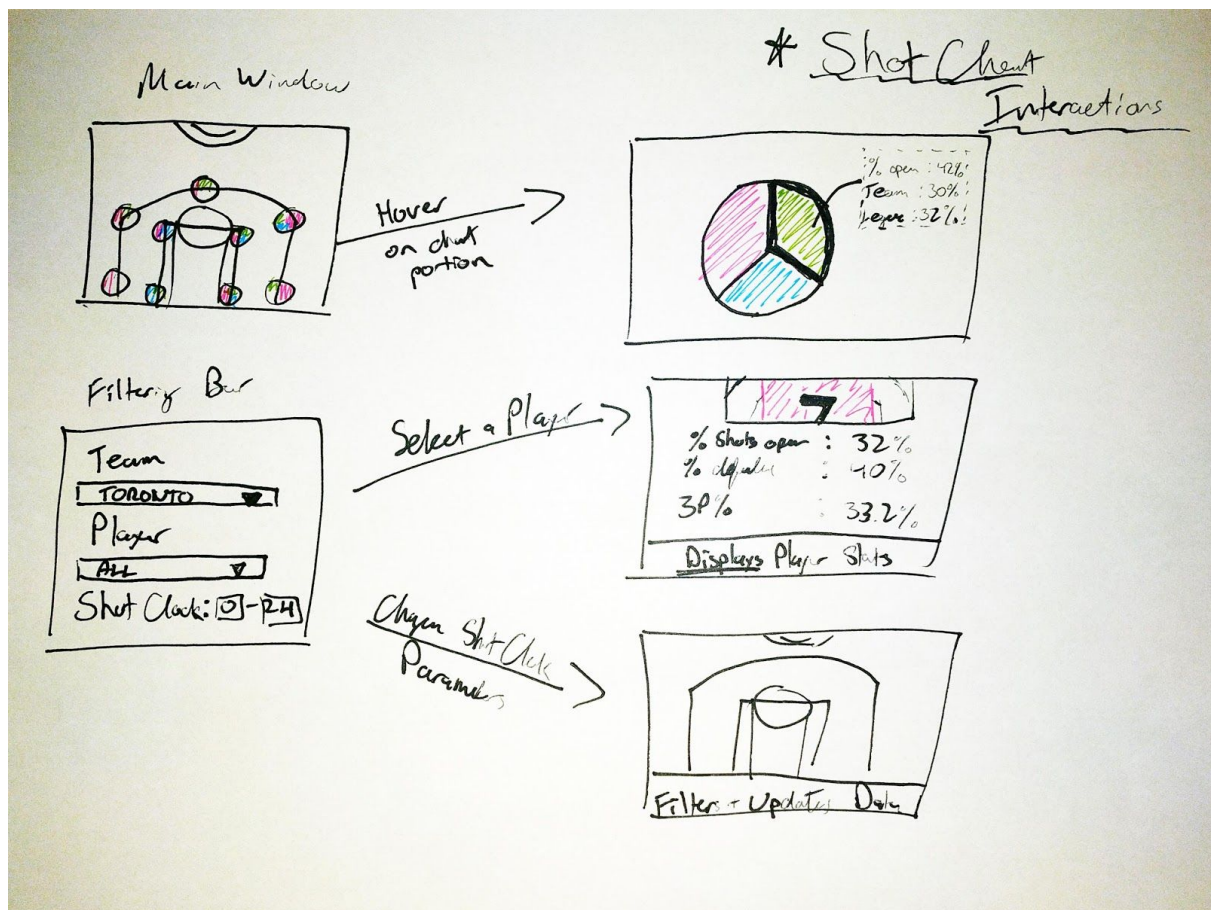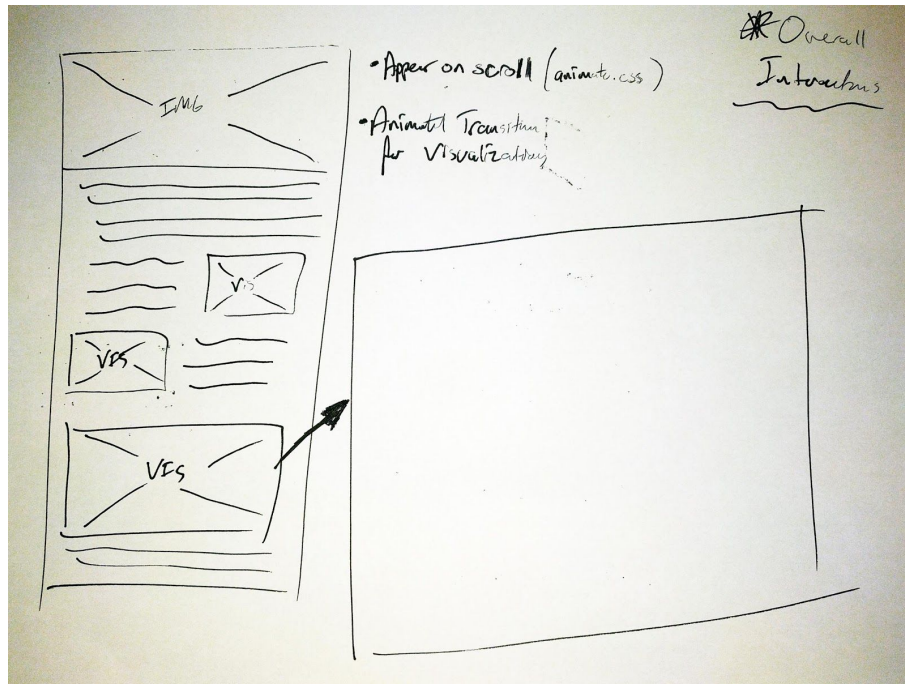
PROTOTYPE V1

Names: Ryan Plunkett, Connor Bitter, Brian Ge

Data: Code for scraping and cleaning in wrangle_data.js. Successfully merged datasets but still working on efficiently storing them. Now just using 10% sample.

Visualizations: See website and below sketches.
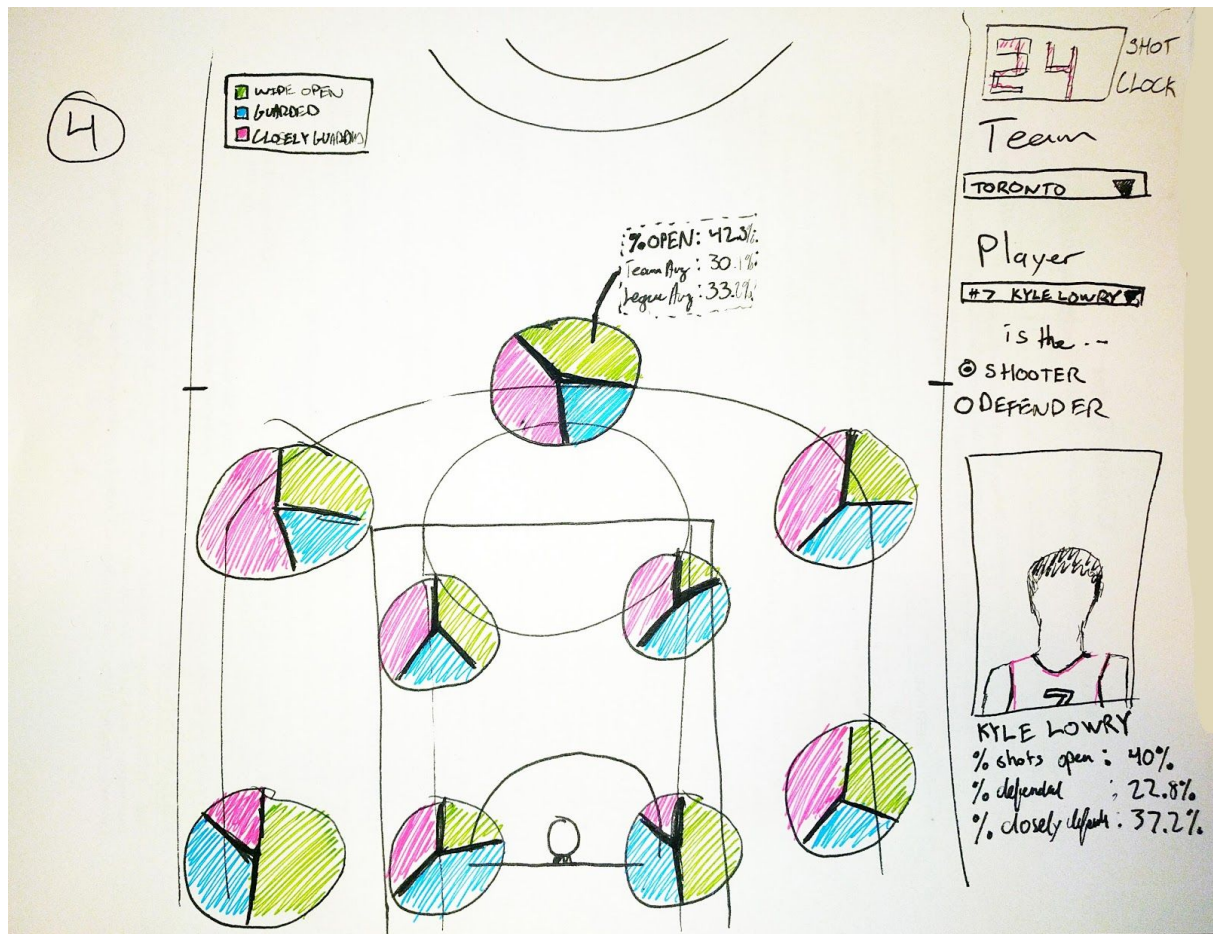
Design and structure: See site.

Storytelling: Explain more about Warriors' 2014-2015 season and ongoing 3-point barrage in NBA.
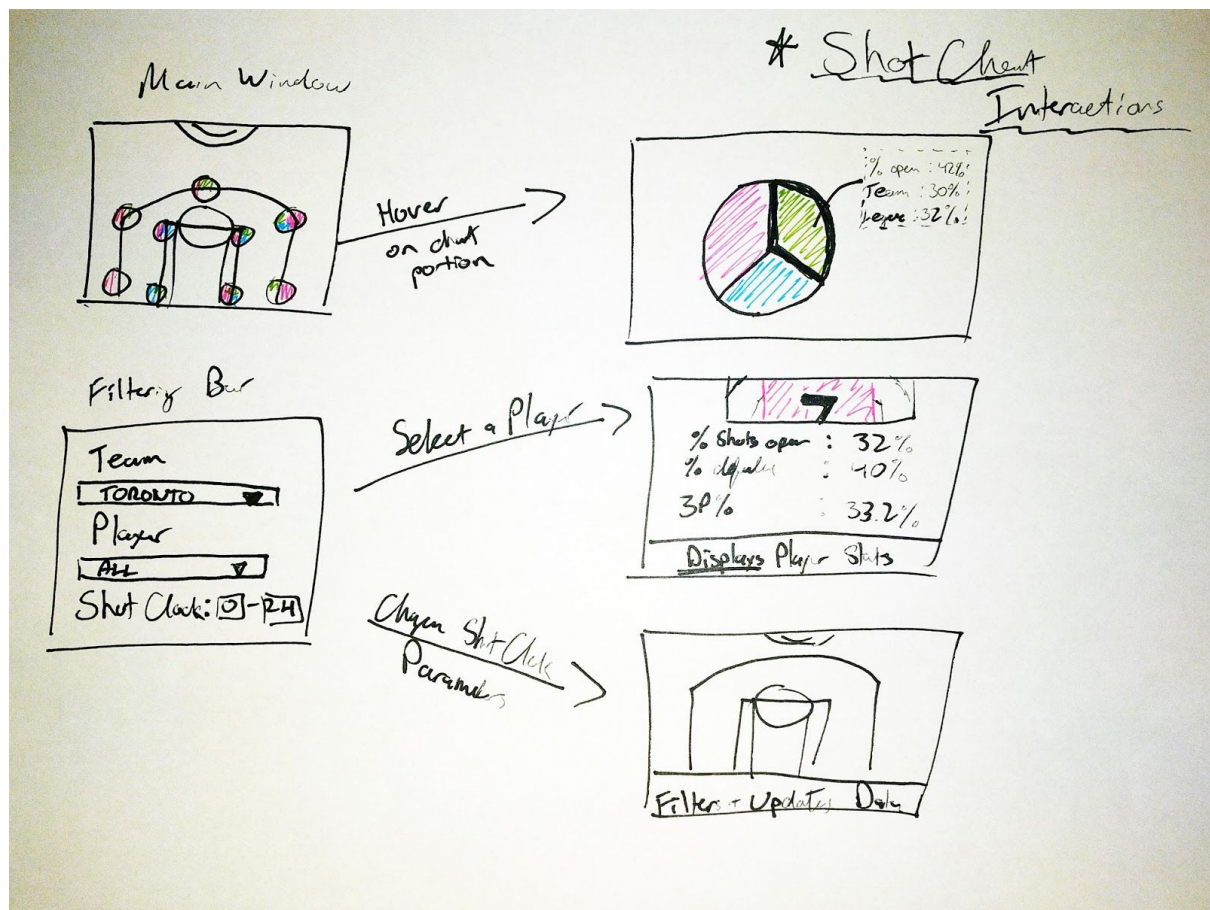
Innovative: Shot chart?

Interactivity: See storyboard below.

Process book: Updated.

Visualization:

Interaction Storyboard:

Main Window

Hover on chart portion

* Shot Chart Interactions

% open : 42%
Team : 30%
player : 32%

Filtering Bar

Team
TORONTO ▼
Player
ALL ▼
Shot Clock: [0]-[24]

Select a Player →

% Shots open : 32%
% default : 40%
3P% : 33.2%
Displays Player Shots

Chosen Shot Clock Parameters →

Filter + Updates Data

PROTOTYPE V2

Names: Ryan Plunkett, Connor Bitter, Brian Ge

Work thus far:

Connor - Formatted website layout, implemented line charts showing FG% as a function of closest defender distance

Brian - Wrote code to scrape and store data, refactored and debugged code for website visualizations

Ryan - Helped with scraping data, wrote visualization code for interactive shot chart and bar chart showing average closest defender distance

Our project required a significant amount of data wrangling, so we're not as far as we would have liked on the website itself. However, we're pretty excited about the shot chart visualization, as we all contributed to it and feel like it shows exactly what we want. Since our data is now ready to go, we feel like our only limiting factors are time and creativity. We think we'll be able to make some really neat bar charts and scatter plots in the next couple weeks showing team and player ranks in several statistics and also uncovering the importance of different metrics as they contribute to winning.

TODO: Finish by Sunday Night

- Make the shot chart half court: Brian
- Fix tooltips: Connor
- Best fit line: Ryan
- Defender distance vs wins: Ryan
- Group bar chart for shot chart: Brian
- Add Line chart data to Scatter plot and see if there is correlation: Ryan
- Add details view to scatter plot: Connor
    - Maybe more to be done?
- Add footer with citations, names, download button for the dataset zip: Ryan
- Translate team names to abbr. for defensive vis: Brian
- Vis about defense (defender distance, shot clock info): Brian
- Add more to shot chart bar charts: Connor, if we have time
- Overall aesthetic/scrolling: Connor
- Add legend to shot chart: Ryan
- Text: Everyone

Project is done.