CS 3300 Project 2

Brendan Quinn (bjq6), Andrew Peterson (alp98), Shan Jiang (sj438)

April 28, 2016

1 Mapping Data to Visual Elements

1.) Charlie O'Connor (Broadstreethockey.com Writer) - manually tracked zone entries
2.) ESPN - Player headshots and positions

Charlie is an active participant of the booming online community known as "hockey twitter". He and other writers at Broad Street Hockey are constantly expanding the field of data-driven journalism with each article they write. This year, Charlie manually tracked every Philadelphia Flyers hockey game. He documented every zone entry (offensively and defensively), every player on-ice for each entry, the type of entry (controlled/uncontrolled), and the shots following each entry. He compiled several datasets, separately tracking offensive and defensive zone entries (opponent entries), along with individual player statistics. Because Charlie tracked every player on-ice for each entry, we were able to create an algorithm which keeps track of how every player on the team performed (in terms of zone entries) with (or without) every other player on the team. This is important, as it will be elaborated later, because this allows us to see which players statistically compliment others in terms of zone entries and shots generated and shots suppressed. We then calculated the number of shots actually generated after each zone entry, knowing that we could then compare how pairs of players performed against their expectation. Furthermore, we kept track of how a player's performance changed over the season by analyzing game-by-game zone entry and shot data and tracking their cumulative shot and entry totals.

Had we had access to a time-on-ice dataset (how long every player spent paired with every other player), we would have been able to normalize the shot rates to a per-60 minute basis. This would have allowed us to see the rates in which players generate zone entries and shots. Nevertheless, the raw numbers still give the user a clear picture of how certain players generate entries and shots over the entire season.

The outcome of our data analytics was data.json.

data.json

Indexed by each player's number, data.json holds a player's:

- name - string, used to identify the player on his player card
- number - int, used to index the player
- position - string, used to map the player to a position on the rink
- gameCumulative - JSON, stores a list of how the player did in each game
- controlled_entry_with - JSON, stores number of controlled entries with each player
- controlled_entry_without - JSON, stores number of controlled entries w/o each player
- controlled_entry_with_against - JSON, stores number of controlled entries against the Flyers with each player and this player
- controlled_entry_without_against - JSON, stores number of controlled entries against the Flyers with this player but w/o each other player
- for_fenwick_with - JSON, stores number of fenwicks for the Flyers with each player and this player
- for_wenwick_without - JSON, stores number of fenwicks for the Flyers with this player but without each other player
- against_fenwick_with - JSON, stores number of fenwicks against the Flyers with each player and this player
- against_wenwick_without - JSON, stores number of fenwicks against the Flyers with this player but without each other player
- uncontrolled_entry_with - JSON, stores number of zone entries for the Flyers with this player and with each other player
- uncontrolled_entry_without - JSON, stores number of uncontrolled zone entries for the Flyers with this player and without each other player
- uncontrolled_entry_with_against - JSON, stores number of uncontrolled zone entries against the Flyers w/ this player and w/ each other player
- uncontrolled_entry_without_against - JSON, stores number of uncontrolled zone entries against the Flyers w/ this player and w/o each other player

/src/ folder

- /src/overlays/ - Stores overlays for the Flyers Title gif, underlay behind each player's player card, and the Flyers logo
- /src/player_faces/ - Stores the headshot transparency png for each player indexed by player number

2 Mapping Data to Visual Elements

2.1 Player Selection

As the main focus of the visualization is to show how much better a player does with or without other players, we decided to that a major focus should be on player selection. For each player we decided to make an ESPN-esque player card to give basic player information:

Example ESPN Player Card for Rakdo Gudas 3

Along with each player card, we map the typical starting position of each player onto the ice rink for better comparison of how plays rely on the different positions of their team members.

There is one button (represented by SVG "rect" element) corresponding to each of the 27 players. We arrange the buttons into two sets of selection panels along the two sides of the ice rink. The corresponding player card, the starting position on the ice rink, and corresponding point on the actual and expected WoWY's are displayed when the users hover on the button. When a click is made, we change the opacity of the button/add the border around the button/make the text bold to indicate a selection of the first player.

After the selection of the first player, both the player card and the starting position of the selected player become fixed there. The WoWY graphs are updated so that only the default data of the selected player is displayed. Additionally, two more buttons (Best NZ pair Best Actual pair) are created. Clicking on these buttons trigger blinking effects on the selection button. When the selected player pairs with the one whose corresponding button blinks, it leads to the best performance. In the next step, users can choose other players to see how the performance of the first selected player is altered when paired with or without other players. The With/Without panel is created upon selecting the second players. Clicking on the button appends corresponding data points to the WoWY graphs. Note that users are able to select multiple players in the second step, we change the opacity of the button /add border around the button that is most recently clicked to differentiate it from previously selected players(change in opacity but without border).

Unselecting the first selected player restore everything to default. Unselecting the players in the second step will remove the data points associated with the players from both WoWY graphs, and also changes the opacity of the button to default.

2.2 With or Without You Graphs

The WoWY graph another major component of our data visualization. There are two main WoWY graphs, one is the expected WoWY, and the other one is the actual WoWY. Both graphs show the information about how a particular player performs with every player on the team also without every player on the team. We use the linear scale to map the data. The xScale() maps the expected/actual number of shots the particular player have for every other player to the x-axis; the yScale() maps the expected/actual number of shots the player have against every other player to the y-axis. The best player will aim to maximize the controlled entries offensively, while minimizing the controlled entries against. We use circles in blue together with the player's number to represent the player's expected/actual shots for/against himself(the default data point for the player). We use circles in green to represent the player's expected/actual shots for/against when he plays with a particular player. Similarly, circles in red are used to represent the player's shots when he plays without a particular player. Clicking on the with/without buttons will generate these data points on both graphs accordingly. The generated points are connected to the default data point by grey lines.

Hovering on the data points will generate a tooltip indicating the percent difference between the pairing and the selected player's default ratio of for/against. A positive change in percentage(in green color) indicates how much better the selected player's performance is when he pairs with/without other players. A negative change in percentage(in red color) refers to how much worse the player's performance is when he pairs with/without other players compared to his default performance.

Each of the WoWY graphs are divided into four sections where the top and right corner is considered a "good zone". Typically, data points fall into this section are players with relatively better performance (more goals, less goals against). On the other hand, the bottom and left corner is a "bad zone" where the players tend to have less goals and more goals against. The line along the diagonal is an indication of the average performance.

2.3 Bar Charts

Upon player selection, a bar chart is created for that player's Entry breakdown. The top bar corresponds to the raw number of entries For the Flyers (by the Flyer's into the other team's zone) completed comprising of two individual bars - controlled and uncontrolled. The bottom bar corresponds to the raw number of entries Against the Flyers (by the other team into the Flyer's zone) completed comprising of two individual bars - controlled and uncontrolled. Zone entry is usually a good indicator of performance in games - if there are a higher number of zone entries For than zone entries Against, each player and consequently the team is expected to do better. Further broken down, controlled entries are more heavily weighted than uncontrolled entries. Therefore the player with a higher weighted number of controlled/uncontrolled entries ratio of For/Against, is more likely to perform well as defined previously.

2.4 Cumulative Shot Differential Plots

Upon selecting the main player, the cumulative shot differential plot is created right below the WoWY graphs. The plots shows the selected player's performance over time to-date. The light blue line is the player's cumulative NZ(expected) Fenwick over the season (expected Fenwick = number of expected Fenwicks-For minus number of expected Fenwicks-Against To-Date), and the green line represents the player's cumulative actual Fenwick over the season (actual Fenwick = number of actual Fenwicks-For minus number of actual Fenwicks-Against To-Date). Hovering on the lines creates a bisector at current game number which allows the users to see the actual values. Linear scales are used to map the data. xScale() maps the game number to the x-axis, and the yScale() maps the cumulative expected/actual Fenwicks to the y-axis.

3 Meaning

This visualization is able to responsively tell the user which players work well together on the ice. By analyzing their historic performances together and without each other, a user could predict with some degree of certainty whether or not a given pair should put on the ice. Imagine a coach, pulling one player out of the game after he has in a brawl on the ice. There are only so many players he can put in. The coach would quickly be able to choose what players he wants take out of the game based, which to keep on the ice, and which players to add based on whose skills complement each other.

This visualization is interesting as it is able to pinpoint key pairs who when they play together, combined play significantly better than either of them play individually. On the flip side, it also shows which players should not be on the ice together based on how poorly they perform together compared to how they perform individually.

Of course correlation does not equal causation, but this data has historically had a statistically significant probability of predicting the outcomes of games as shown by the Cumulative Shot Differential Plot. Here we are able to see exactly how close each predicted number of fenwicks is to the actual number of fenwicks. As such, we are able to predict how well each player will perform in a given game based on who he will be playing With or Without.