

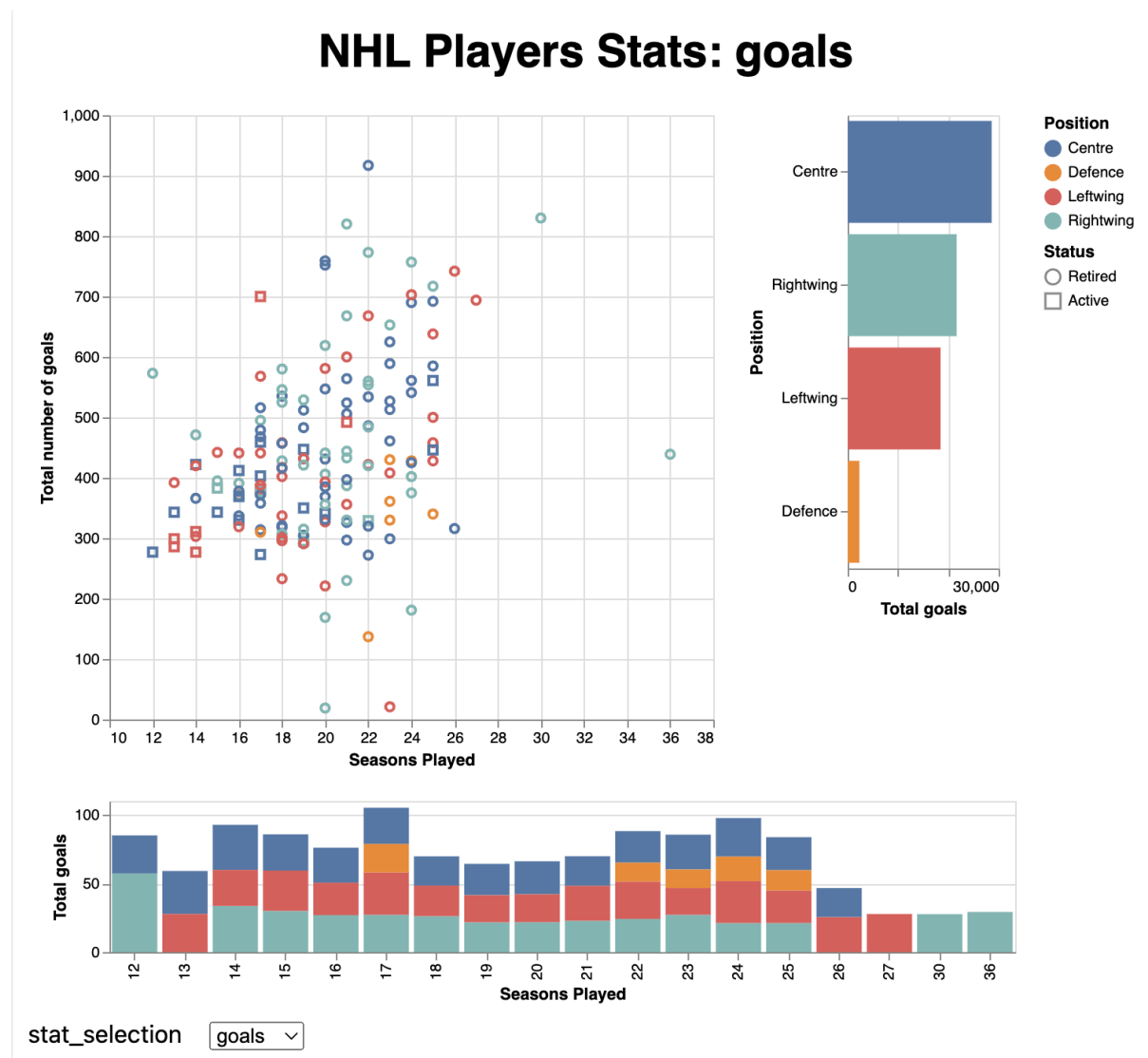
Comp40610 Visual Exploration Tool Design Document

Student Names: Oghenemalu Igboiye

Student number: 22208232

Title:

NHL Players Stats:



Dataset Overview:

The visualisation provided, presents an insight into the season_goals [dataset](#) provided by the HockeyReference.com, and downloaded from [Tidy Tuesday 2020 Hockey Goals](#). The original dataset provides details regarding various player stats of the top 250 players in the NHL from 1942-2010, with features such as goals, total_goals, assist, position, hand, rank and player. To get my desired output however, i utilised the Python Panda's library to derive a new version of the [dataset](#), by cleaning and making new features. Data cleaning was done to remove missing values from the dataset, and i cerated a new feature called `years_active` derived from calculating the difference between the year the player career began and the year they ended their career. Furthermore, In refining the dataset to ensure clarity, I transformed/renamed some of the categories in the `position` column. Categories such as "c" were expanded to "centre," "rw" was renamed to "rightwing," and similar adjustments were made for descriptive accuracy. These preprocessing steps were essential for shaping and creating a more insightful and user friendly data format for subsequent visual exploration and analysis

Scatter Plot: The scatter plot shows the relationship between the total number of goals scored or assists made for each player and the number of years active (i.e. seasons played). Individual players in the plot are represented by points, and have attributes as encoded by their channels, these includes colour coding each player based on their position played, also each point's shape is determined by the players current status (i.e. either active or retired). This approach, I believe, is fundamental to making the plot visually easy to summarise, based on the stat or metric selected in the `stat_selection` drop down menu. However, using scatter plot could be disadvantageous, especially in densely populated plots, where it could make it difficult to read or select specific players statistics. An alternative, might be to use a bar plot, with each player being represented by a unique bar, this inadvertently would result in way too many bars making the plot unreadable, and also make it difficult to encode both positions and status at the same time. Also since I am using total goals scored or assists made (i.e. sum), rather than just the goals or assists per season, this would reduce and control the density of the plot. Therefore scatter plots in this context, is a more visually appealing and intuitive way to plot a players goal or assist distribution, in comparison to other plots like bar plots, effectively boosting the users experience and ability to quickly summarise the data.

Horizontal Bar: The horizontal bar plot provides a comprehensive display of the total number of goals or assists made by players in various positions. Each bar is colour-coded representing the a specific position, the colour coding scheme is the

same as the scatter plots. This plot enables a quick visual comparison of contributions by roles/position in the NHL. The interactive bar chart uses cross-filtering to influence the data shown in the scatterplot and the vertical bar chart. When a bar (or position) is selected, the horizontal bar chart dynamically filters the dataset, ensuring only players in the specified position are revealed in the scatter plot, it also adjusts the vertical bar plot to displaying the total goals or assist for the selected position based on number of seasons played. The bar chart represents the best way to visualise categorical data against continuous ones, and due to the limited amount of positions, it is easily understandable by users. Furthermore, to enhance user experience, sorting for the horizontal bar chart is dynamic, based on the stat selected or other analysis being performed by the user.

Vertical Bar Chart: This stacked bar chart illustrates the relationship between the number of seasons played and the total number of stat selected (i.e. goals or assists) for number of seasons played. When a bar (or season) is selected, the bar chart interacts with the scatter plot and the horizontal bar chart, by showing players who only played the total amount of selected season and number of goals and assist made by position for the selected season length. While stacked bar charts are not easily understandable, i opted for this to show the total goals scored based on number of season played, furthermore, each bar is colour coded based on position and once selected the horizontal bar plot changes dynamically showing total goal or assist across all positions for the season selected.

Interaction Consideration:

The primary interaction approached utilised in this visualisation is cross-filtering, providing users with the ability to select or view different subset of the data in one chart, influencing the data presented in the other. The scatter plot, horizontal and vertical bar charts each enable selection based on different attributes such as position, season played, providing users with a thorough exploration of the dataset. I also implemented advanced features like the `stat_selector`, a drop down menu which allows users to dynamically change between what stats they want to see, (i.e. goals or assists) across all three charts. This dynamic change was also implemented on the titles, to ensure some of the titles changes based on stat selected, for instance the main title and y-axis title in the graph changes based on the stat selected.

This design implementations, aims to create a smooth trade-off between overview and detail, supporting the users overall experience and their ability to gain insights

into the total number of seasons played, goals scored or assist made, position and status of NHL players.