Data Visualization Final Report: Visualizing Uncertainty in Fantasy Football Projections

Author: Greg Carmean

Abstract

This paper describes the application of uncertainty visualization principles to visualizing uncertainty in fantasy football projections. Uncertainty visualization is expanded from interval visualization to distribution visualization to give a fuller picture of potential outcomes. Several uncertainty distribution visualizations are considered for the problem space and the most appropriate are created in a dashboard for fantasy football players to use to evaluate their start/sit decisions.

Introduction

In the game of fantasy football, one of the most important decision is which player to start or sit any given week. Most fantasy football projections are shown as point estimates that do not capture the full extent of a players projected performance. Additionally, there is rarely consensus across the industry about how a player will perform exactly any given weeks, due to the small samples sizes inherent in a football season. Visualizing this uncertainty allows managers to make more informed decisions and gives them an edge.

One of problems with a point estimation projection of fantasy points scored is that fantasy football is scored such that a player gets six points for a touchdown, with causes significant variability in the outcome of the projection. A player may be projected to score seven points, when the projection consists of the player rushing for 40 yards, scoring four points, and then a 50% chance that they score a touchdown. While the mean projection for those scenarios is a seven-point finish, the player is much more likely either score four or ten points than the projected seven. This information would be useful for fantasy managers making start-sit decisions. Different uncertainty visualizations will be explored to best characterize the uncertainty behind each projection and

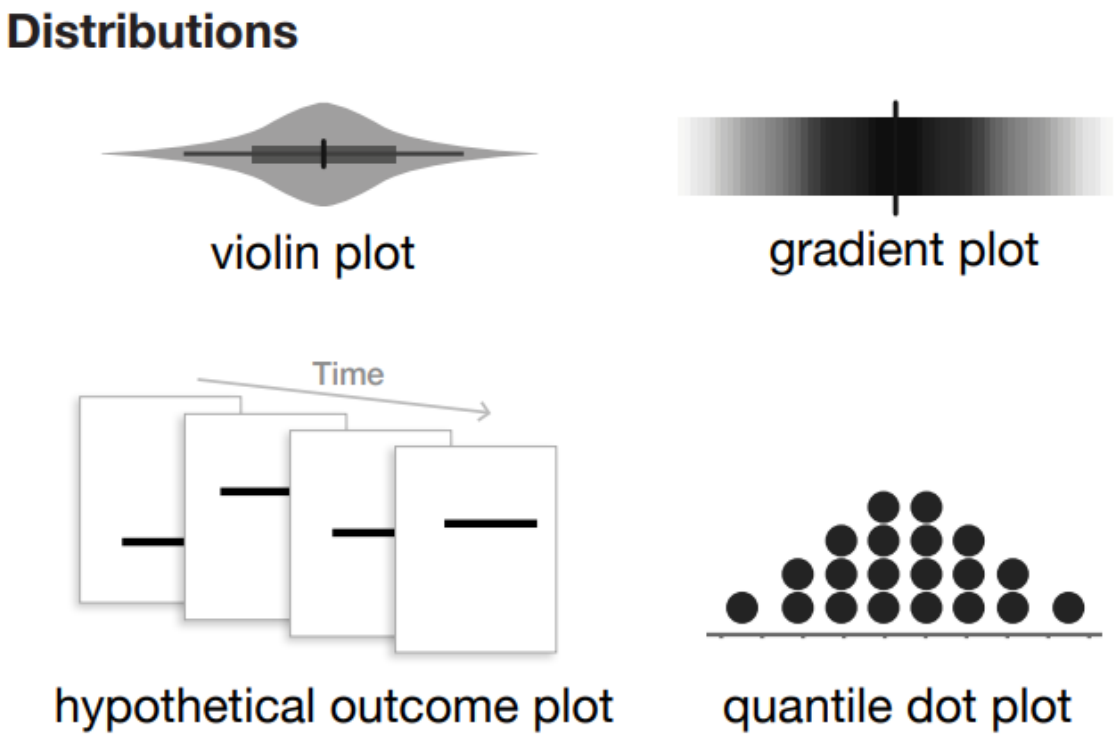
across positions generally. These will be combined into a dashboard that allows the user to examine aspects of the uncertainty in fantasy football projections that week.

Background

Uncertainty visualization is an active area of research with new concepts being developed to provide a more advanced look at uncertainty present in datasets. Padilla, Kay, & Hullman provide an overview of various state of the art uncertainty visualizations. The original error visualization that the most people are familiar with is the error bar, which is an interval visualization of error. Correll and Gleicher describe some of the disadvantages that have arisen with the widespread use of error bars. They are used inconsistently across visualizations so it can be difficult to immediately identify what an error bar represents, the standard deviation or the absolute error. Additionally, they can create an artificially firm division between things within the error interval and without the error interval when the true phenomenon is a spectrum. Another common plot used to convey uncertainty is the box plot, which is based off of the Tukey statistics.

Given the plethora of fantasy football projections available, the distribution uncertainty visualizations are relevant to the problem. Visualizing the distribution of the uncertainty adds extra information to the problem and gives the context needed for fantasy football players to make an informed choice. The distribution gives a sense of the projected floor and ceiling for a player which allows the fantasy football manager to make better decisions. Padilla, Kay, & Hullman provide an overview of various visualizations of the distribution of uncertainty. There are four distribution plots relevant to fantasy football use case, violin plots, gradient plots, hypothetical outcome plots, and quantile dot plots.

Figure 1: Examples of Different Distribution Visualizations



Adapted from Padilla, Kay, & Hullman (in press). Uncertainty Visualization. To appear in, Handbook of Computational Statistics and Data Science.

The violin plot was first proposed by Jerry L. Hintze & Ray D. Nelson in 1998. A violin plot is a continuous display of the underlying distribution of the data. It is an extension of the box plot which shows the frequency of occurrence of the values in the interval. The width of the shape demonstrates how a value is likely to be generated from that distribution. A gradient plot has a similar concept, except using the shading instead of width to denote frequency. The advantage of a gradient plot is that it is easier to make comparisons across distributions, whereas the violin plot shows more detail. Finally, the quantile dotplot is the discrete alternative to a violin plot. The data is discretized by a set number of points which are grouped by quantile. The number of dots in the quantile indicate the frequency of occurrence in that quantile. A related visualization first proposed by Hullman, Resnick and Adar is the hypothetical outcome plot (HOP). The HOP is an animation

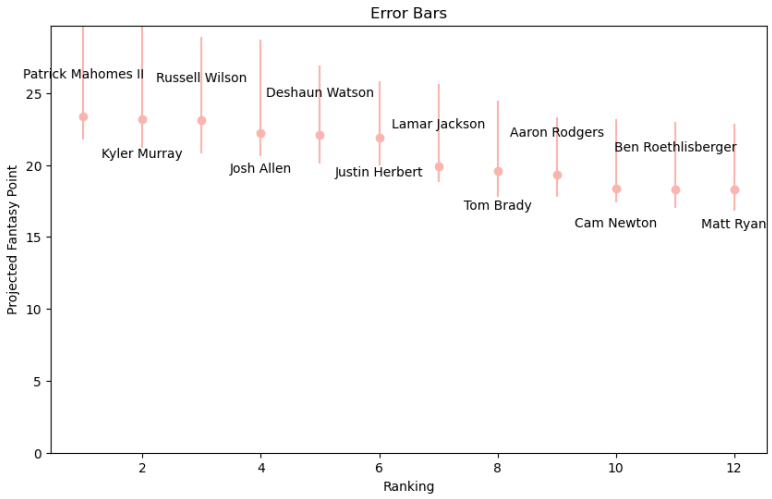
of random samples from the underlying distribution shown one at a time in a sequence. The advantages of hypothetical outcome plots are they cause the user to form a finite impression about the distribution and they are intuitive as there are no new marks introduced. However, the finite nature introduces some risk that the distribution is not fully represented by the HOP. Their research shows that it outperformed box plots and violin plots for making comparisons across uncertainty distributions.

Approach

Three visualizations were selected to be in an interactive dashboard to explore uncertainty in fantasy football projections; a visual of the average projections with error bars, a violin plot of specific players distribution, and a HOP to compare the expected performance of two players.

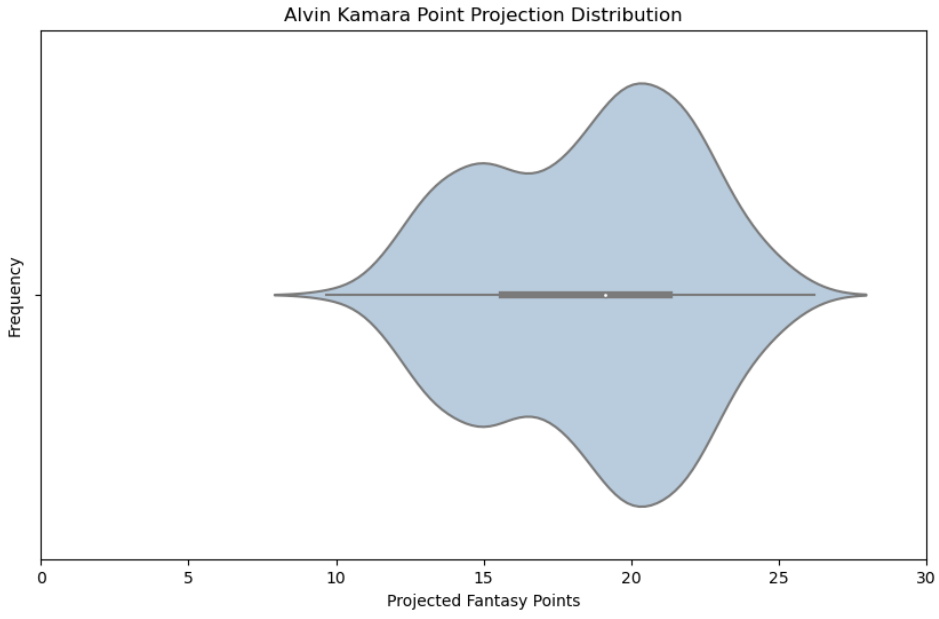
The first visualization plot of the average projected fantasy points across positions with error bars which represent the high and low projection for that player respectively. An example is shown in Figure 2. This visualization also the user to compare projected performance quickly across many different players of the same position.

Figure 2: Error Bars Example



The next visualization available for the user is a violin plot which allows the user to see the full range of the generated distribution for the player. This gives the full distribution of their actual fantasy points. The violin plot gives the best overall summary of all of a players reasonably understood outcomes. This implementation of the violin plot from seaborn includes the box plot in the center of the plot for comparison purposes, while showing the most frequent values around the plot. This shows the extra information which indicates how, such as in the example shown below, the average value is not always the most likely to occur. For fantasy football projections, the quantile dot plot discretization is less accurate because of the continuous nature of rushing, receiving and passing yards. These are all continuous which makes a continuous representation the best fit. This implementation uses kernel density estimation to create the continuous distribution.

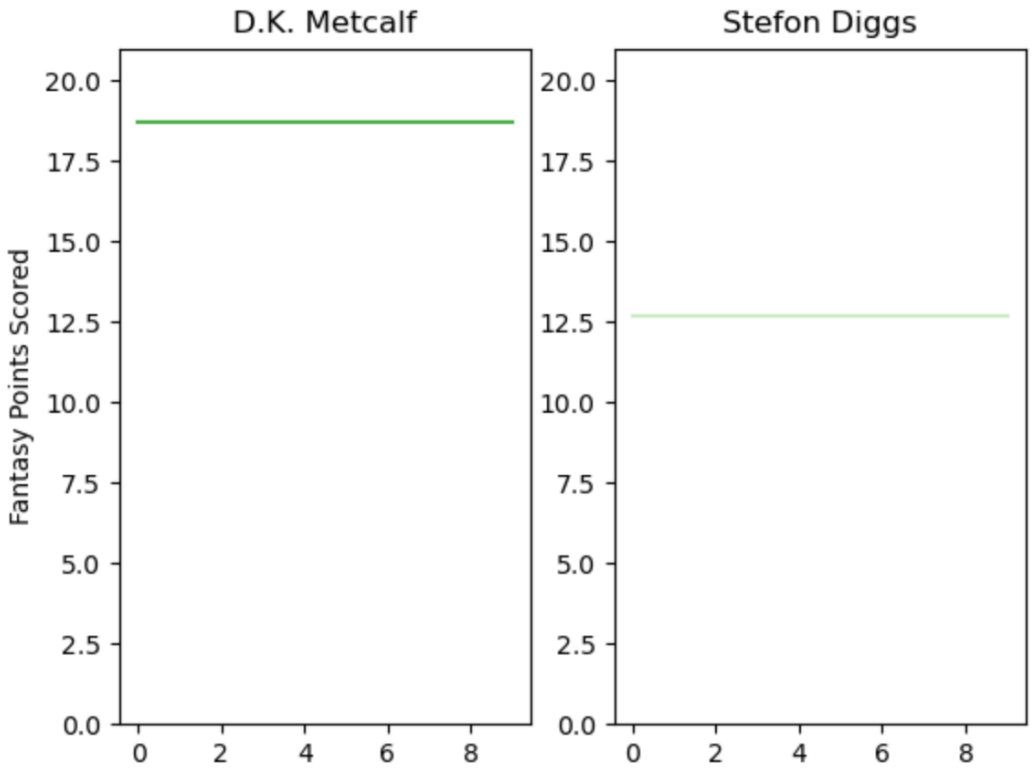
Figure 3: Violin Plot Example



The final visualization available for the user is the hypothetical outcome plot. One hundred random draws are done one at a time from the relevant distributions to compare the performance of two

players. This visualization is hypothesized to be a very good fit for the problem because, at the end of the day, the fantasy player will produce a finite amount of points. The hypothetical outcome plot helps the user think about what the outcome is from a player in a way a continuous comparison across player distributions like a gradient plot would. The hypothetical outcome plot is created by making a plot for every draw from each player distribution, and animating the draws, clearing the figure before each new draw. An example of what the user sees every second is shown in Figure 4.

Figure 4: Hypothetical Outcome Plot Example



Before diving into the results, where the distribution of fantasy points comes from should be discussed. The data is pulled from FantasyPros weekly fantasy point projections. FantasyPros is a website that aggregates and ranks fantasy football analysists on their performance to get the best

industry projections. Every week they aggregate projections for every players across the various platforms and put out the players average projection as well as their high and low projected performance. The players stats and fantasy points are projected. This is the data used to create the uncertainty visualizations. However, these projections still use averages to generate their fantasy point projections. Each player various statistics is modeled using a truncated normal distribution, where the high and low of each projection is used to truncate the results. This makes the resulting distribution representative of the range of realistic results without considering extreme scenarios. For all stats the mean of the normal distribution is the average projected stat. For stats related to yards the standard deviation is half of the difference between the high and low, as these stats are consistent between players. Touchdowns have more variance so the standard deviation for touchdown is double the distance between the high and low projection. The statistics are rounded to mirror real game events. Therefore, a touchdown chance of 0.4 gives 0 touchdowns and a projection of 0.6 gives 1 touchdown. An area for improvement is fine tuning the model to create the projection distributions. A distribution is created for every statistic and then the distributions are sampled randomly to create the final fantasy points distribution with inform the plots.

Results

The result is a dashboard that allows the user to explore fantasy football projections through the three visualizations. The users can specify projections from a specific week, their preferred scoring settings, the position they want to inspect, and the players to visualize in the violin plot and Start/Sit Simulator, which is the implementation of the hypothetical outcome plots created for this paper. This dashboard was created using the streamlit package in python.

Figure 5: Dashboard Overview

The sidebar options allow the user to select the week, position and scoring system they want to see projections for. The mean projection with error bars and violin plot are shown side by side which their controls shown above. The hypothetical outcome plots with the ability to select each of the players is shown belove. Figures 6 and 7 show a zoomed in look at the top and bottom of the dashboard.

Figure 6: Top Half of Dashboard

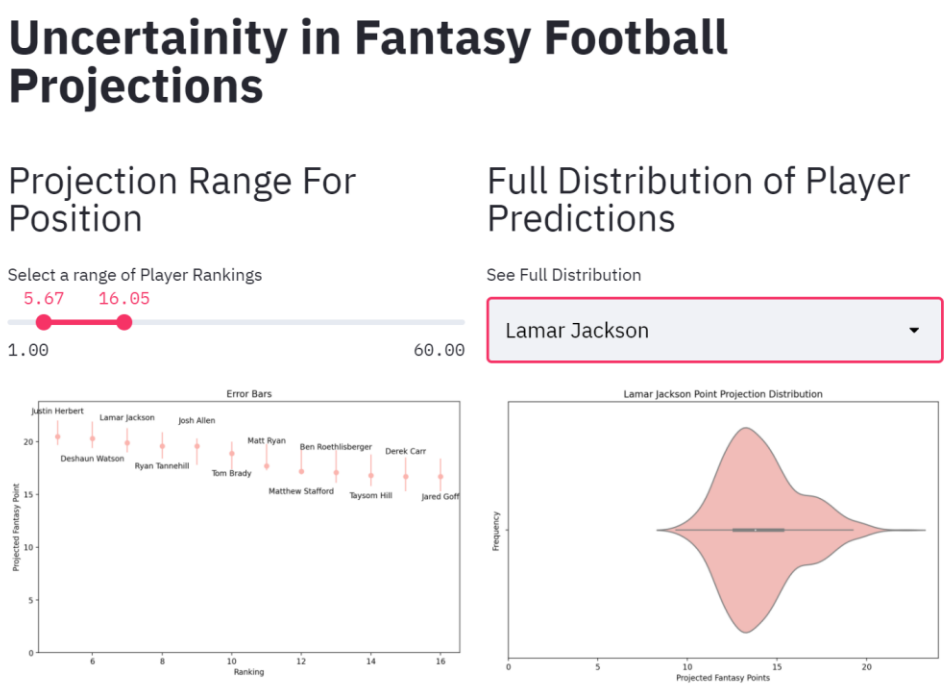
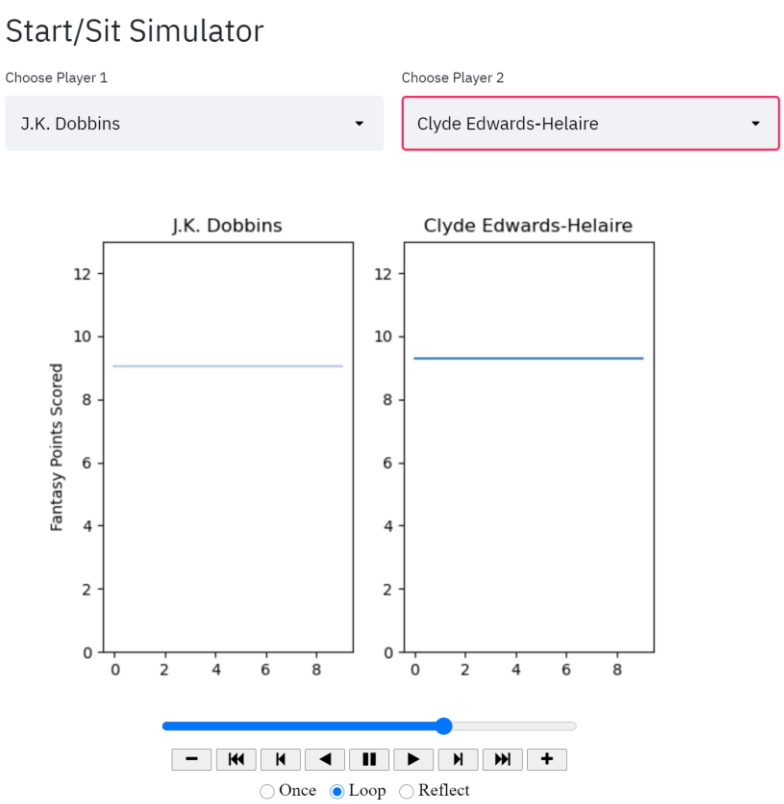


Figure 7: Start/Sit Simulator



Conclusions

The construction and testing of the dashboard shows that uncertainty visualizations can be helpful in making start/sit decisions in fantasy football. The violin plot enhances the box plot by making the frequency of occurrence evident. This is particularly useful when considering players who score a variety of different ways. The frequency shows the impact of scoring a touchdown or catching a long pass has on the total points scored, which is not well captured by the mean. Additionally, the hypothetical outcome plots are a highly effective start/sit decision tool. Observing multiple draws from the distribution, even for players whose mean projection are very similar, yields a decisive conclusion about which players should be started.

References

Hullman J, Resnick P, Adar E (2015) Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. PLoS ONE 10(11): e0142444. https://doi.org/10.1371/journal.pone.0142444

Jerry L. Hintze & Ray D. Nelson (1998) Violin Plots: A Box Plot-Density Trace Synergism, The American Statistician, 52:2, 181-184, DOI: 10.1080/00031305.1998.10480559

Lewinson, E. (2020, February 24). Violin plots explained. Retrieved November 17, 2020, from https://towardsdatascience.com/violin-plots-explained-fb1d115e023d

M. Correll and M. Gleicher, "Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2142-2151, 31 Dec. 2014, doi: 10.1109/TVCG.2014.2346298.

Padilla, Kay, & Hullman (in press). Uncertainty Visualization. To appear in, Handbook of Computational Statistics and Data Science.