

Data Storytelling

1. Ask the following questions and look for the answers using code and plots:
 1. Can you count something interesting?
 1. With all the statistical information available there is an almost endless number of things to count. The most interesting things I've found to count are consistency, correlation, and the outliers.
 2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
 1. Yes, I have found several trends within the data. For example, there is a strong trend for running backs who have a high number of rushing attempts to produce higher fantasy points. This trend is consistent across all positions, with a few anomalies. This leads me to question what causes these outliers and if there is a way to identify them. If you can identify the outliers this will give you an advantage over your peers.
 3. Can you make a bar plot or a histogram?
 1. Yes, I created a plot displaying the correlations of the positional stats that relate to the top performers for each position. For example, rushing yards and attempts have a high correlation to the top performing running backs.
 4. Can you compare two related quantities?
 1. Yes, I am comparing several related quantities for each position to find what stats relate most to overall fantasy point production.
 5. Can you make a scatterplot?
 1. Yes, I created scatterplots for multiple variables to visually display what stats consistently return higher fantasy point production. As well, these scatterplots point out any outliers in the data/stats.

6. Can you make a time-series plot?
 1. No, my data doesn't allow for any time-series integration. The data occurs once a week, so there is no benefit or use in creating a time-series plot.
2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?
 1. Based on the plots, I gathered insight into what specific stats are most important to a player's success (total fantasy points scored). As well, I am able to determine higher correlations for a high scoring player and the team/coach/scheme they play for so that I am not relying only on a single player in vacuum but am taking into account the environment or system they play within.
 2. My original hypothesis; The higher a players usage rate is (determined by attempts, percentage of snaps, percentage of team targets, or percentage of team attempts for the associated position) results in higher fantasy point production.
 3. This leads me to ask more questions about how a team or coach and their specific scheme relates to a certain player or a positions usage rate. How does the team/environment in which the player resides relate to a player's overall fantasy point production? Do players on teams with losing records perform worse than similar players on teams with winning records. Does a specific team/coach correlate to higher fantasy production for a certain position? How does the age of a player relate to their fantasy production? How high is the correlation between a quarterback and the skills position players on their team; i.e. does a high producing quarterback result in high producing wide receivers/tight ends/running backs?
3. Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and

plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

1. Predominantly it is easy to identify the top performers (top 5 players for each position) before a season begins. We will call these players our “studs,” and conversely, we’ll call the players who don’t perform to their expected level “duds.” The compelling story arises from identifying these “duds” and avoiding them, as well as finding the players who will outperform their expected production.
2. Next, we need to determine what beginning value a player has so we can compare this against their in-season production. For example, a player could be deemed a “dud” while still outperforming another player we deem a “stud.” There are a couple of options for us to consider here. The most common baseline for a season long fantasy league is a player’s average draft position (ADP). Another option would be a player’s salary value for a weekly/daily fantasy league such as DraftKings or FanDuel. When we consider the high rate of injury associated with playing in the NFL, the variability is much higher in a weekly/daily fantasy format and it would be nearly impossible to accurately predict how a single player would perform in a specific game several weeks into the season. With this in mind I will use ADP as my baseline for determining our “studs” and “duds.”
3. Now that we’ve determined what we’re using as our baseline value for each player for each season, using the last 5 years of data, we can use these trends and insights to find value and minimize risk.
4. A complete and compelling story in this case would consist of finding value based on a player’s ADP, minimizing the risk of selecting “duds” player, and finding the “studs” with most consistency over a 16 game season.