Project #3 - Interactive Visualization

Jacob M. Lundeen

3/3/2021

Purpose

A number of software tools have been designed to help users visually explore data and to quickly create visualizations or dashboards. Popular off-the-shelf tools include Tableau, Qlikview, Spotfire, Microsoft Power BI, MicroStrategy, Birst, and Logi among many others. The purpose of this assignment is to use existing software tools to formulate and answer a series of analytical questions about a specific data set of your choice. After writing a number of analytical questions or hypotheses, you must create a Tableau dashboard designed to present some of the answers of your hypotheses.

The Data

For this project, the data set I am going to use comes from Pro Football Focus (PFF). It is their 2021 NFL Draft Statistics report, basically a collection of traditional and modern statistics they collected and derived for the 2021 draft class. In particular, I will be focusing on the coverage sheet: this sheet lists all potential draft prospects that played at least a single snap in coverage over their college career. So it will contain variables like number of interceptions, but also more modern statistics likes forced incompletion rate and completion percentage allowed. The data set is 24 x 1174, with 3 character variables and 21 numerical. Lets look at the overall descriptive statistics first and then we can break down the variables more in-depth.

| Name | Piped data |
|------------------------|------------|
| Number of rows | 1174 |
| Number of columns | 24 |
| Column type frequency: | |
| character | 3 |
| numeric | 21 |
| Group variables | None |

Table 1: Data summary

Variable type: character

| skim_variable | \min | \max | empty | n_unique | whitespace | |
|---------------|--------|--------|-------|----------|------------|--|
| Name | 7 | 25 | 0 | 1170 | 0 | |
| Team | 3 | 10 | 0 | 135 | 0 | |
| POS | 1 | 2 | 0 | 4 | 0 | |

Variable type: numeric

| skim_variable | mean | sd | p0 | p25 | p50 | p75 | p100 |
|--------------------|--------|---------------------|-------|--------|--------|--------|--------|
| PFF Coverage Grade | 62.58 | 11.24 | 26.00 | 55.82 | 63.20 | 69.60 | 93.70 |
| Coverage Snaps | 202.04 | 78.80 | 53.00 | 136.00 | 188.00 | 257.00 | 525.00 |

| skim_variable | mean | sd | p0 | p25 | p50 | p75 | p100 |
|--------------------------------------|--------|---------------------|-------|--------|--------|--------|--------|
| Press Coverage Snaps | 25.53 | 41.28 | 0.00 | 1.00 | 5.00 | 38.00 | 344.00 |
| Targets Allowed | 23.92 | 12.16 | 2.00 | 15.00 | 22.00 | 31.00 | 74.00 |
| Receptions Allowed | 15.33 | 7.70 | 0.00 | 10.00 | 14.00 | 20.00 | 53.00 |
| Completin % Allowed | 0.66 | 0.15 | 0.00 | 0.56 | 0.65 | 0.76 | 1.00 |
| Yards Allowed | 185.17 | 105.30 | 0.00 | 110.00 | 168.00 | 237.00 | 696.00 |
| TDs Allowed | 1.34 | 1.39 | 0.00 | 0.00 | 1.00 | 2.00 | 8.00 |
| First Downs Allowed | 8.55 | 4.71 | 0.00 | 5.00 | 8.00 | 11.00 | 30.00 |
| INTs | 0.70 | 0.99 | 0.00 | 0.00 | 0.00 | 1.00 | 6.00 |
| Dropped INTs | 0.38 | 0.65 | 0.00 | 0.00 | 0.00 | 1.00 | 4.00 |
| Forced Incompletions | 2.29 | 2.26 | 0.00 | 1.00 | 2.00 | 3.00 | 17.00 |
| Forced Incompletion $\%$ | 0.09 | 0.08 | 0.00 | 0.04 | 0.08 | 0.13 | 1.00 |
| Defensive Stops | 4.15 | 2.89 | 0.00 | 2.00 | 4.00 | 6.00 | 17.00 |
| Tackles | 13.64 | 6.76 | 0.00 | 8.25 | 12.00 | 18.00 | 44.00 |
| Assists | 2.15 | 1.83 | 0.00 | 1.00 | 2.00 | 3.00 | 12.00 |
| Missed Tackles | 2.57 | 1.97 | 0.00 | 1.00 | 2.00 | 4.00 | 11.00 |
| Passer Rating When Targeted | 93.55 | 29.62 | 0.00 | 73.82 | 94.50 | 114.60 | 158.30 |
| Average Depth of Target | 9.38 | 4.53 | -1.29 | 5.57 | 9.64 | 12.56 | 26.75 |
| Yards per Reception | 12.28 | 4.20 | 0.00 | 9.48 | 11.65 | 14.43 | 36.33 |
| Dropped Targets + Overthrown Targets | 5.05 | 3.48 | 0.00 | 2.00 | 4.00 | 7.00 | 22.00 |

Character Variables

The first three variables are character variables and all three are nominal data. 'Name' is self explanatory, it is the name of the player.

'POS' is the position of the player, here it is Cornerback (CB), Safety (S), Edge Defender (ED) or Linebacker (LB). The inclusion of ED here is interesting because in the NFL ED's do not drop into coverage except on rare occasions. In college this is a little different, but more importantly EDs change positions to LB in the NFL frequently. So knowing how well a college ED performed in coverage is relevant to teams.

'Team' is the name of the school the player plays for. There are 135 unique schools.

Numeric Variables

There are 21 numeric variables, 14 discrete and 7 continuous. Most of the discrete variables are traditional counting statistics for the NFL: Coverage Snaps, Interceptions (INTs), Tackles, Missed Tackles and others. A few of the continuous variables warrant further explanation.

'PFF Coverage Grade' is a metric created by PFF where they assign a numerical grade (on a scale from -2 to +2, by 0.5 increments where 0 is average or "expected") to every play a player participates in over their entire career. These play grades are summed up, weighted by certain factors and normalized to a 0-100 scale (I am not privy to their methodology so I cannot provide further granularity). The grades break down as such:

- 90-100: Elite
- 80-89: High Quality
- 70-79: Above Average
- 60-69: Average
- 50-59: Below Average
- <49: Poor

'Completion % Allowed' is the ratio of receptions allowed over the number of times targeted. As an example, if a player is targeted in coverage 50 times and allowed 10 receptions, that would be a Completion Percentage Allowed of 20%.

'Forced Incompletion %' is the rate at which the defender is directly responsible for forcing an incomplete pass. If a pass is incomplete, it can be due to a variety of reasons: the ball was tipped at the line of scrimmage, the ball was overthrown and uncatchable, or the defender knocked the ball away before the target could secure the ball. This variable speaks to how active the defender is in actually defending passes versus getting lucky with passes that are simply uncatchable.

Questions

- 1. How well does a player's coverage grade relate to their allowed completion percentage?
- 2. How well does a player's coverage grade relate to their forced incompletion percentage?
- 3. Which position produced the most interceptions?
- 4. Which position was best at defensive stops?
- 5. What school produced the highest average coverage grade across all positions?

Visualization Design

For my interactive visualizations, I wanted to highlight how PFF's coverage grades related to some of the other metrics and give the user the ability to look at their favorite position, player or school and see who they might like to see their favorite NFL team draft this year.

First, I did a visualization for coverage grade versus completion percentage allowed. 'POS', 'Team', and 'Name' are set as filters, 'POS' is set to color so the user can see the separation immediately and the number of coverage snaps are set to size. The tooltip includes: 'Name', 'Team', 'POS', 'PFF Coverage Grade', 'Completion % Allowed' and 'Coverage Snaps'. See Figure 1.



Figure 1: Interactive Visualization 1

My second chart is similar to the first, but this time it is PFF Coverage Grade versus Forced Incompletion Percentage. All the design choices are the same, as are the available filters. Please see Figure 2.

My third chart is a heat map showing the average coverage grade across all positions by school. The filters are the same, but our design differs. The PFF coverage grade is averaged across all positions for each school and that is used in the size and color marks. The map is labeled by school and the tooltip includes: 'Team', 'Avg. PFF Coverage Grade', 'Avg. Completion % Allowed', and 'Total Coverage Snaps'. This will present the user with a look at how schools compare to each other in defending the passing game. Please see Figure 3.

Finally we have the dashboard. Here, all three charts act as filters on each other. So if you click on a data point in the first chart, it will bring up that specific player in the second chart and that player's school in the third chart (heat map). If you click on a school in the heat map, it will filter the players from that school on the first two charts. Four global filters are added on the right, 'POS' (which I combined with the legend that shows the color assignment for 'POS'), 'Player Coverage Snaps', 'Team', and 'Name'. The 'Name' filter, for whatever reason, is more of a search box. I couldn't get it to provide a selection pane like with 'Team' and 'Pos'. This might be because of the 1174 players. I put the heat map on the right because if it was underneath the other two charts, it would look squished. Please see Figure 4.



Figure 2: Interactive Visualization 2



Figure 3: Interactive Visualization 3



Figure 4: Interactive Visualization 4

Discussion

This dashboard directly answers questions 1, 2 and 5.

Question 1

We can see from the trend line, we have a negative correlation between coverage grade and completion percentage allowed. This makes sense, however, in that a player would tend to be playing better if they were forcing opposing quarterbacks to have a lower completion rate when targeting them.

Question 2

Here, we have a positive correlation between coverage grade and forced incompletion percentage. Again, this makes sense. A defender is generally performing well if they are forcing incomplete passes through their direct actions. Of interesting note, both trend lines for these charts have a p-value below 0.0001, but with completion percentage allowed the R value is lower (0.083) then with forced incompletion rate (0.128). If you increase the minimum number of coverage snaps you can get better R values, but not by much (0.125 and 0.166 respectively). Neither of these are high indicating that either there is a non-linear relationship and/or there are missing variables that correlate with the coverage grade.

Question 5

The heat map answers our question here, and honestly it wasn't the answer I was expecting with West Virginia coming in on top. With the filters we can break it down by position. We see Alabama has the top grade when only looking at CBs, which is not surprising considering they have had multiple players drafted highly each and every year.

Tableau Public

Below is the link to my published dashboard:

 $https://public.tableau.com/profile/jacob.m.lundeen \#!/vizhome/Lundeen_Project_03/Dashboard1?publish=yes$