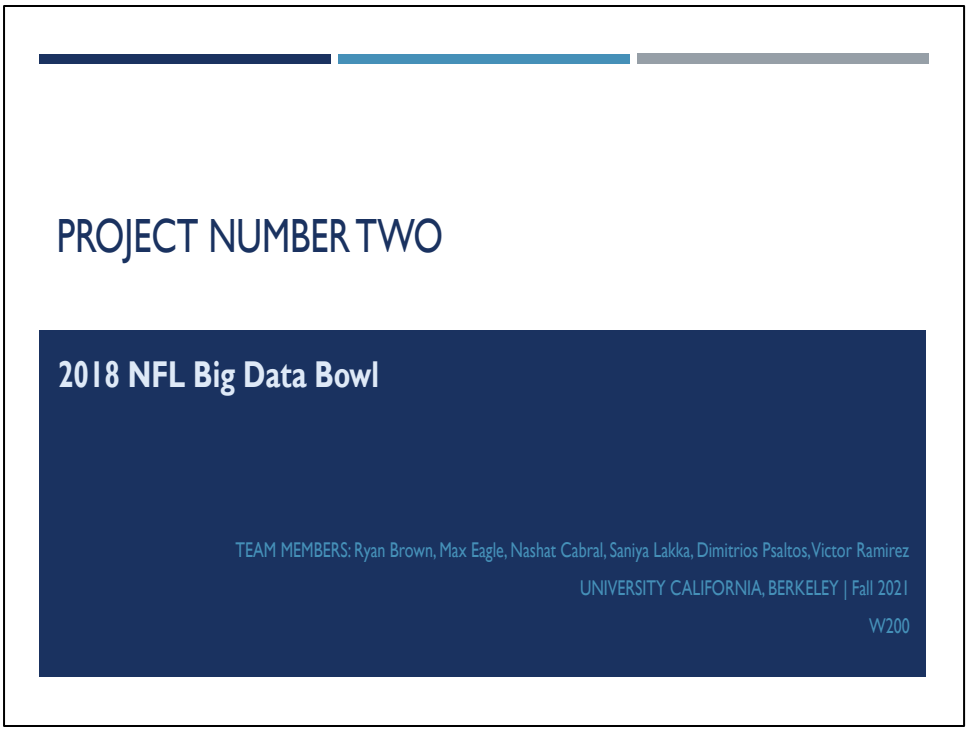
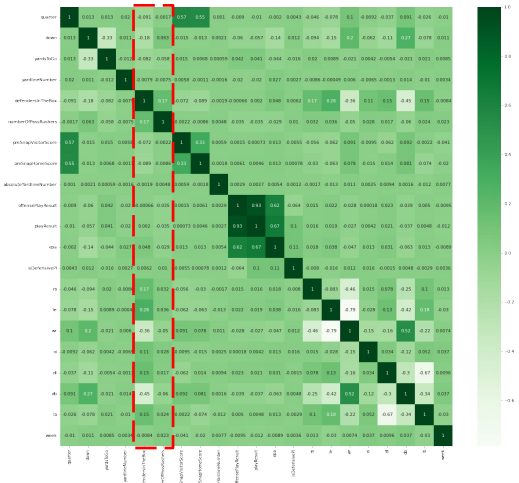



Thank you for joining us this evening for our Final Project presentation where we did an exploratory data analysis on the 2018 NFL Big Data Bowl dataset.

## Background, Purpose and Dataset Explanation

- NFL collects data from regular season games
- NFL 2018 Regular Season Dataset (3 files):
  - Plays - all plays from 2018 season (formations, field position, etc)
  - Games - all games from 2018 season (teams, etc.)
  - Players - all players for 2018 season (names, etc.)

**Goal: Can we understand the impact of defenders in the box and pass rushers on play outcome?**



The NFL has made an effort to start collecting data on games. In fact, these datasets are released to the public in the form of an annual competition called the NFL Big Data Bowl.

For our final project, we decided to use one of these datasets from the 2018 NFL Regular Season. This dataset includes 3 files:

Plays.csv - contains data from every play of the season
- Formations
- Play outcome
- Game time

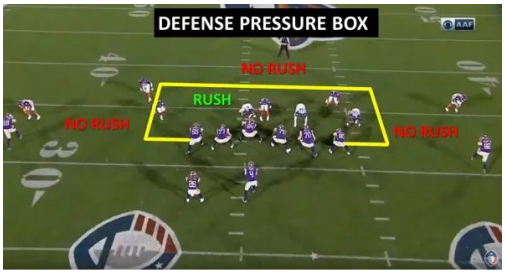
Games.csv - which contains the game data information
- Date of game
- Teams in game

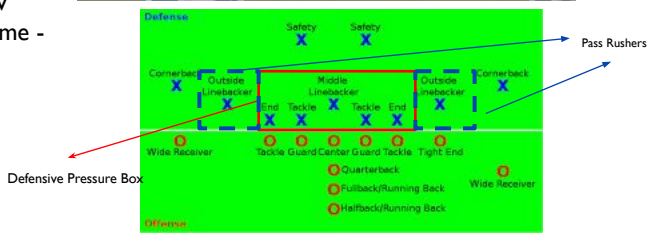
Players.csv - all players for the 2018 regular season.
- NFL player names
- Their birthdates

Our goal was to do an exploratory data analysis to understand the impact of the number of defenders in the box and pass rushers on play outcome.

## Definitions

- **Defenders in the Box** - (DIB) Defenders close to line of scrimmage
- **Pass Rushers** - Defenders who rush the quarterback
- **Passer Rating** - Metric based on passing attempts, completions, yards, touchdowns, and interceptions, which describes a quarterback's passing ability
- **EPA** - (Expected Points Added) Measure of how offensive team performs vs expectations (outcome - expectation)




https://www.espn.com/ncf/columns/davie/1427720.html
https://en.wikipedia.org/wiki/Pass_rush
https://en.wikipedia.org/wiki/Passer_rating
https://www.the33rdteam.com/epa-explained/

We wanted to cover a few definitions:

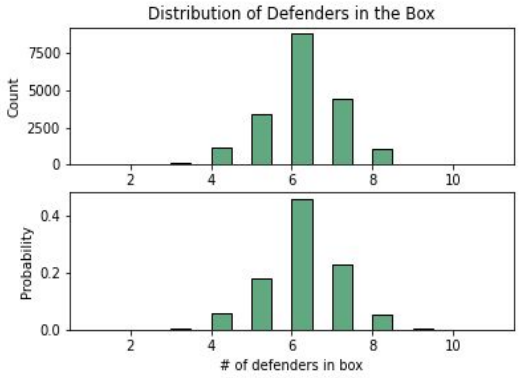
**Defenders in the box** are the defenders close to line of scrimmage who impact run

**Pass rushers** are defenders who rush the quarterback to prevent passes

**Passer rating** is a metric used to describe quarterback passing ability

**EPA or Expected Points Added** is a measure of how well a team performs vs expectations (outcome - expected points)

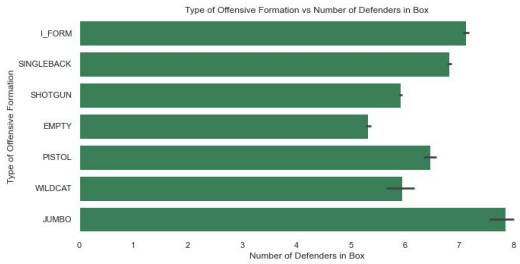
# Distribution of Defenders in the Box (DIB)

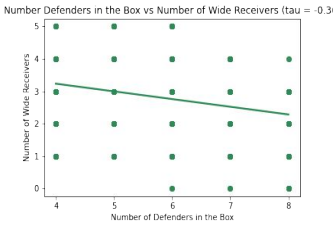



- DIB = 6 was most utilized during 2018 NFL season
- 98.5% of data explained by DIB = 4-8
  - Chose to focus on these formations in analysis

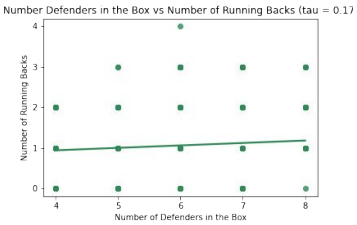
During the 2018 NFL season, the majority of defensive formations included 6 defenders in the box. About 99% of the dataset is explained by between 4-8 defenders in the box. For this reason, we chose to clean the data by excluding plays where less than 4 or more than 8 defenders were in the box.

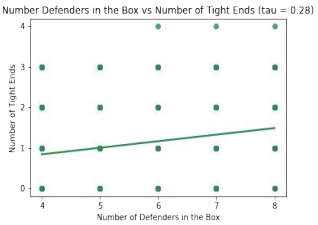
## When would defenses change their formation?



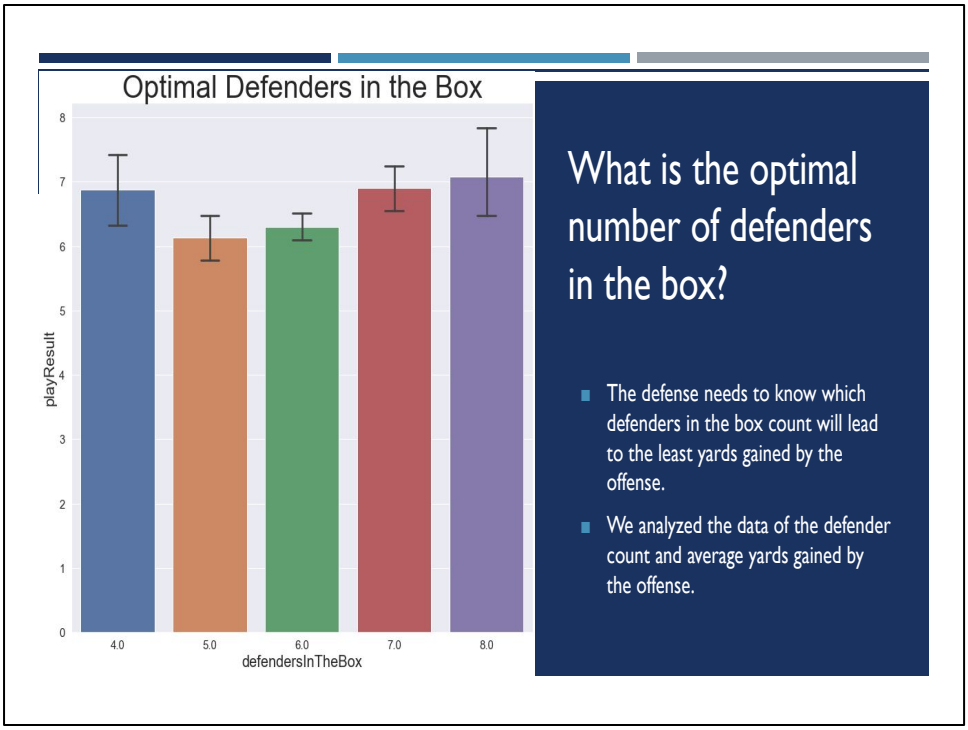




- Passing formations (Empty, Shotgun) have fewer DIB
- Running plays (I-Form, Jumbo) have more DIB



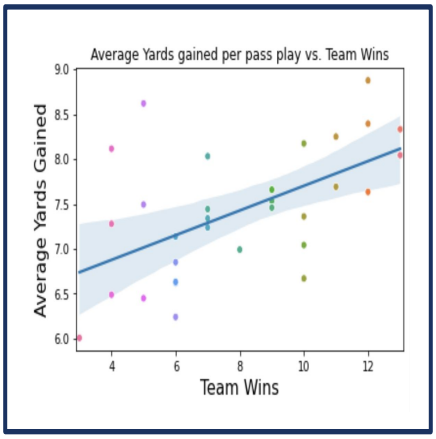




What causes a defense to change their formation (ie. adding more or less defenders to box)? While many game situations impact what players are on the field and their formation, everything comes back to the offensive formation. Based on the offensive formation, teams will change their defensive positioning. Running formations like Jumbo or I-formation are generally played with more defenders in the box vs a passing formation like empty backfield or shotgun, which will use less DIB. We can see that there is a slight positive trend between the number of defenders in the box and the number of running backs and tight ends, and a slight negative trend in the number of wide receivers and the number of players in the box.




Here we have a bar chart visualizing the data examined.
The chart contains the defenders in the box counts on the x axis and the average yards gained on the y axis.
As you can see that on average the best performing defenders in the box count is between 5 to 6 defenders. The recommendation would be to play 5 to 6 defenders in the box.

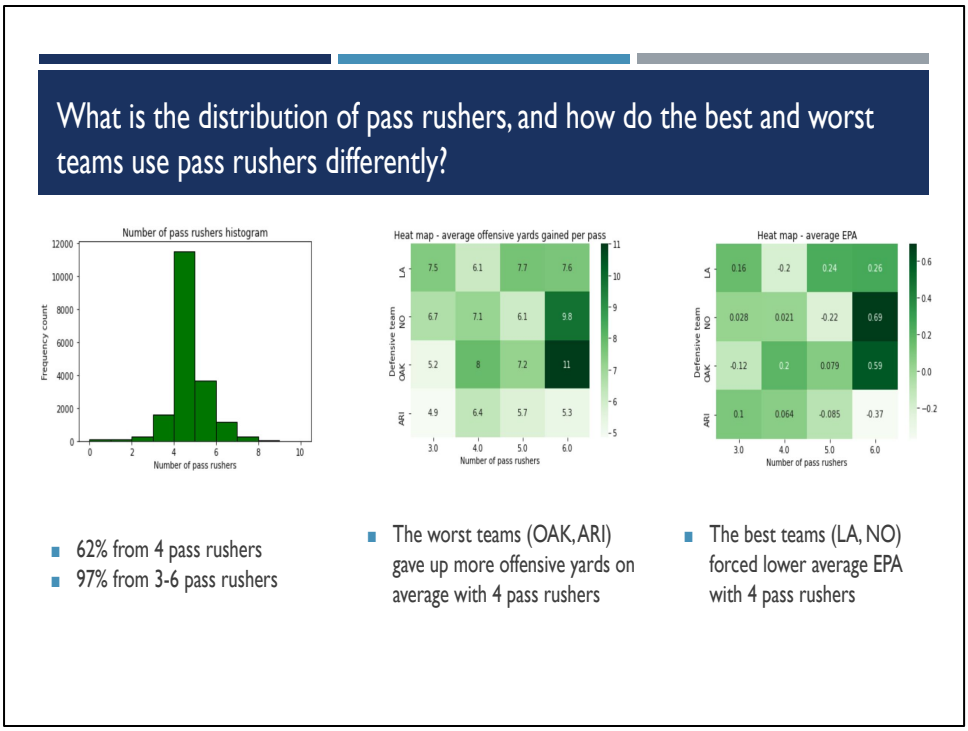
## WHAT IS THE AVERAGE YARDS PER PASS PLAY PER TEAM WINS





- Higher performing teams had more average yards gained.

- This graph shows the average yards gained against team wins. Each data point correlates to a team name. From this chart we can see that higher winning teams are averaging higher yards gained.

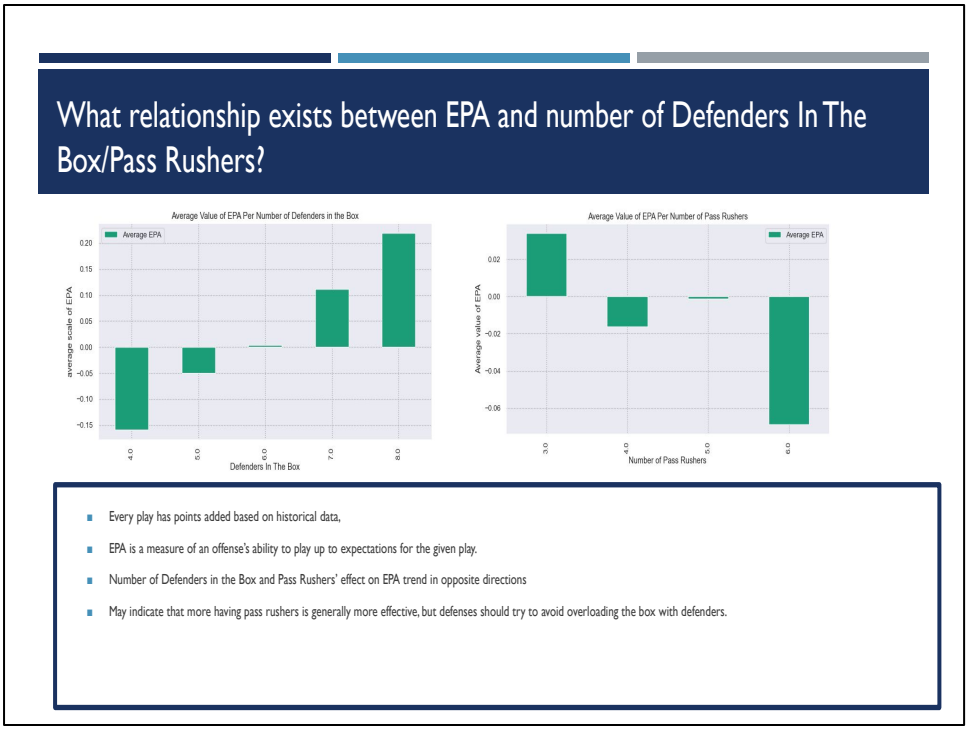



The first chart shows a histogram of number of pass rushers for all games and plays. The number of pass rushers has a nice bell shaped distribution with a mean of 4.2 and a median of 4. Note that 4 pass rushers explains 62% of the data and that 3, 4, 5, or 6 pass rushers explains 97% of the data. Consequently, the following charts focus on this subset of the data with respect to number of pass rushers.

NOTE: in the 2018 season, LA and NO were the best teams, and OAK and ARI were the worst teams.

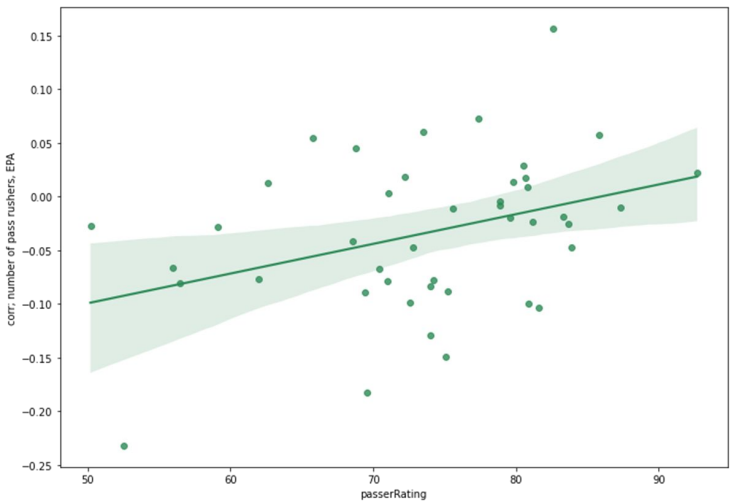
The second chart shows a heat map of average offensive yards gained per pass by defensive team and number of pass rushers. While on defense, the worst teams gave up more offensive yards on average with 4 pass rushers ***relative*** to their defensive performance with other numbers of pass rushers.

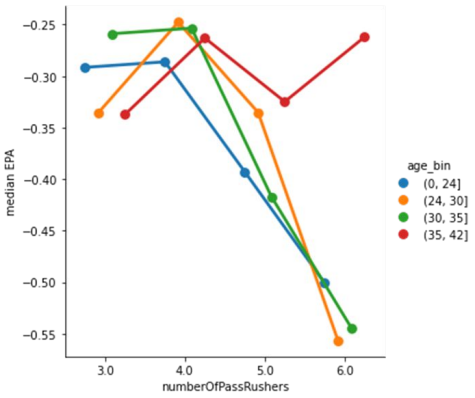
The third chart shows a heat map of average EPA by defensive team and number of pass rushers. While on defense, the best teams had lower average EPA with 4 pass rushers ***compared to*** the worst teams using 4 pass rushers. Having a negative or near zero value is desirable since that means the offensive result didn't meet or exceed expectations for yardage, so the best teams defensive maneuvers mitigated expected offensive results well.




- Another question we seek to answer in the analysis is What relationship exists between EPA and number of Defenders In The Box/Pass Rushers?
- While EPA is generally seen as an offensive statistic, in these visuals we'll use it here to determine a defense's success in limiting its opposing offenses.
- So for a defense, less EPA is more.
- Here we see two graphs visualizing the relationship of EPA and the number of defenders in the box and the number of pass rushers.
- Generally speaking across the entire dataset, we see a trend in opposite directions in the relationships between EPA and defenders in the box and EPA and the number of pass rushers
- We can see that, less defenders in the box is more effective than more at stopping an offense and more pass rushers is more effective than less.
    - However, in the analysis we determined that having more pass rushers increased the variance of EPA. We noticed that as we add pass rushers we see an increase in in the variance of EPA, meaning that while it generally is effective at stopping offenses, it increases the likelihood of risk of giving up points to an offense.

## HOW DOES THE NUMBER OF PASS RUSHERS AFFECT EPA IN PASSING SITUATIONS WHEN CONTROLLING FOR SKILL, EXPERIENCE?




To evaluate the efficacy of additional pass rushers, we really wanted to hone in on number of pass rushers vs QB experience level in purely passing plays. To do so, we filtered the dataset to passing situations based on pre-snap score, time remaining in the game or half, and down and distance. Third and 10, for instance, is usually a passing play, such that both the offense and defense know it, whereas 2nd and 2 is not ordinarily a passing play, unless there's 50 seconds in the game and you're down by 7.

To evaluate experience, we looked at two different proxies: passer rating and age. Passer rating is defined by the number of completions, passing yards, touchdowns, and interceptions over total attempts for a quarterback, with a higher number being better.

The chart on the left looks at the correlation between number of pass rushers and EPA by passer rating. A positive correlation suggests better relative offensive outcomes when additional pass rushers are used, whereas a negative correlation suggests the inverse. As you can see, there is a distinct positive trend, which for the best passer ratings we've seen, tapers off just above zero. This suggests additional pass rushers have a negative effect on relative offensive outcome for nearly every quarterback, but the better the quarterback, the more resistant they are to this association.

We see this same theme echoed in the chart on the right. When we look at median EPA by number of pass rushers and age grouping, we can see that every age

grouping is affected similarly, with the exception of the oldest age bracket, 35 and up. This again suggests that nearly every grouping is negatively impacted, but the most experienced quarterbacks are more resilient.

# KEY FINDINGS

- Higher performing teams had more average yards gained.
- 6 defenders in box was most common during the 2018 NFL season, and offensive formation determined the number of defenders in box.
- Best performing defenders in the box count is between 5 to 6 defenders.
- With 4 pass rushers, the worst teams gave up more offensive yards; the best teams forced lower EPA.
- Generally, fewer defenders in the box is more effective than more at stopping an offense, and more pass rushers is more effective than fewer but is more prone to risk.
- Additional pass rushers negatively affect most offenses on average, but QBs with higher passer ratings, as well as quarterbacks aged 35 and older, seem to be more resilient

# QUESTIONS?