

# Super Bowl Data

By: Sam Jones and Cason Henderson



# What is our Data about

Superbowl  
Offensive  
Stats from  
1966-2021

23 columns  
729 rows

Passing stats  
and rushing  
stats were  
the most  
important

Team and  
player name  
least  
important

# Cleaning up data

How we found which columns to remove

Most were floats(Numbers) but some  
object/integers(Word)

Created a dictionary to fill in the missing data

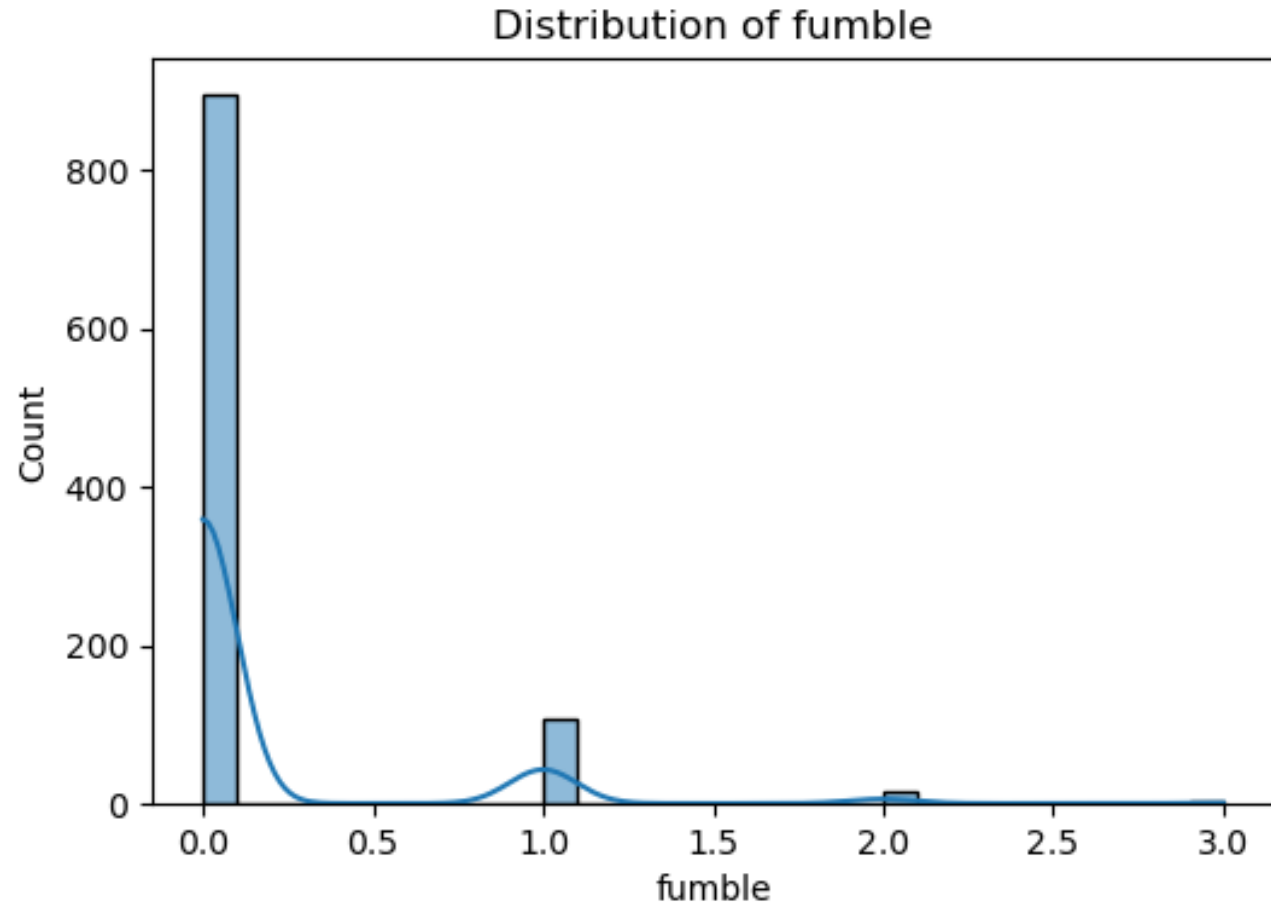
Dropped QBR, times\_sacked, sack\_yards,  
fumbles\_lost

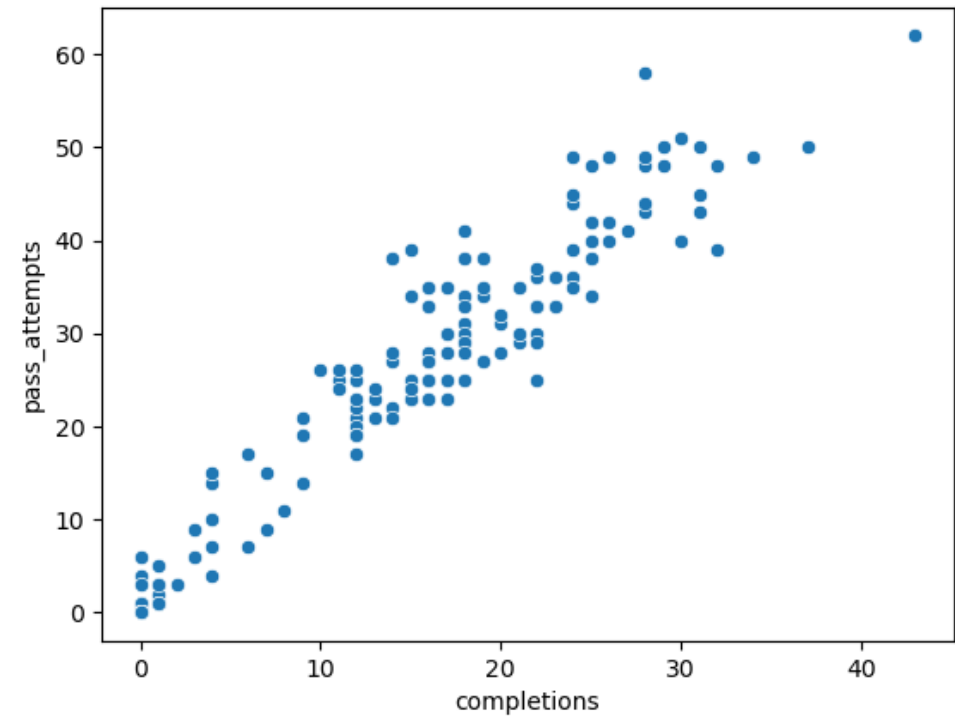
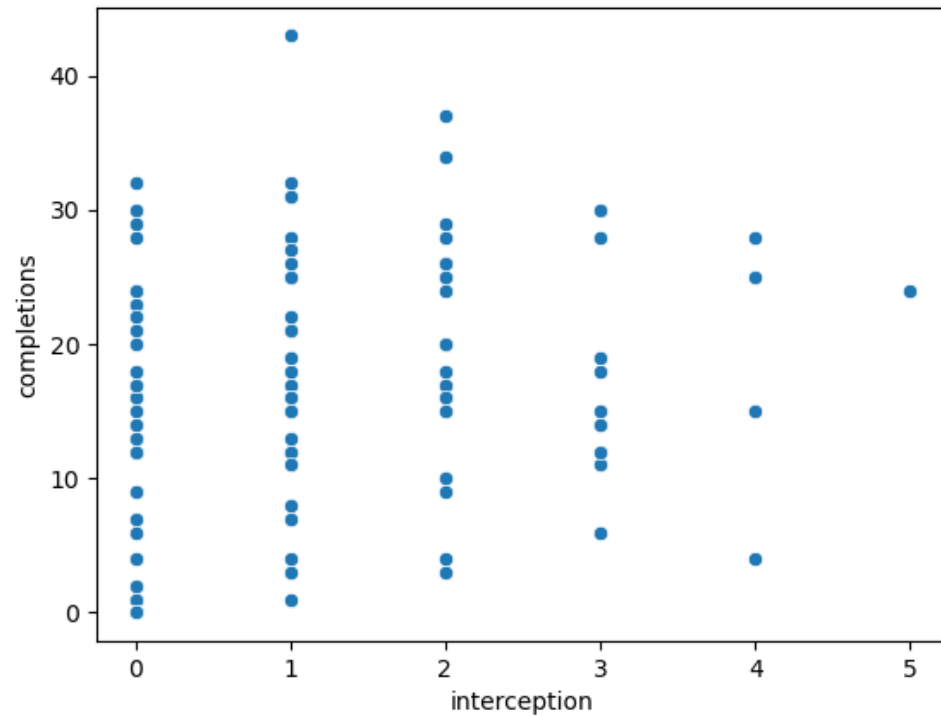
Displayed new data

# Univariate Analysis - Distributions

---

- Not very useful
- Uneven player distribution
- Helps moving forward

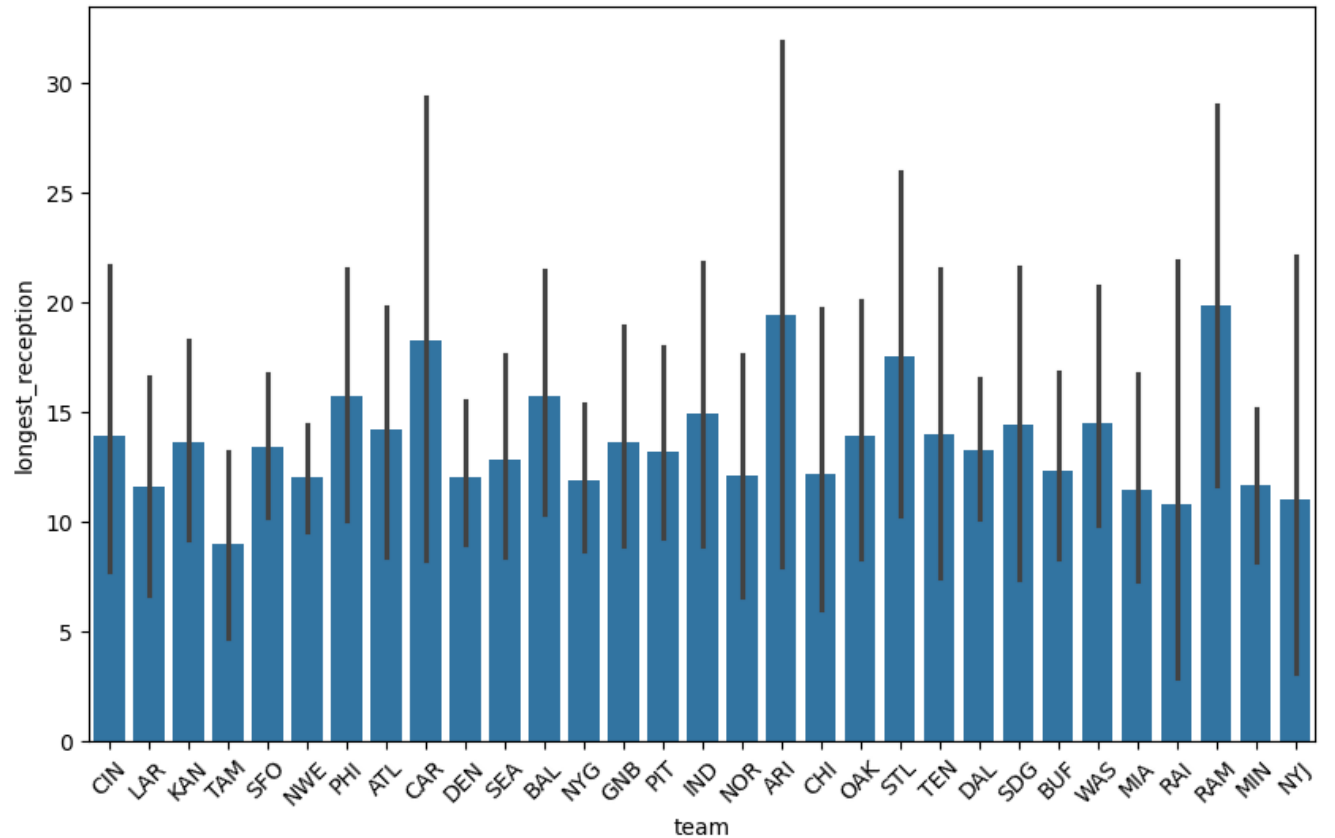




# Scatter Plots

- Very helpful for majority numerical data
- Display patterns and direct relationships
- Highlights outliers

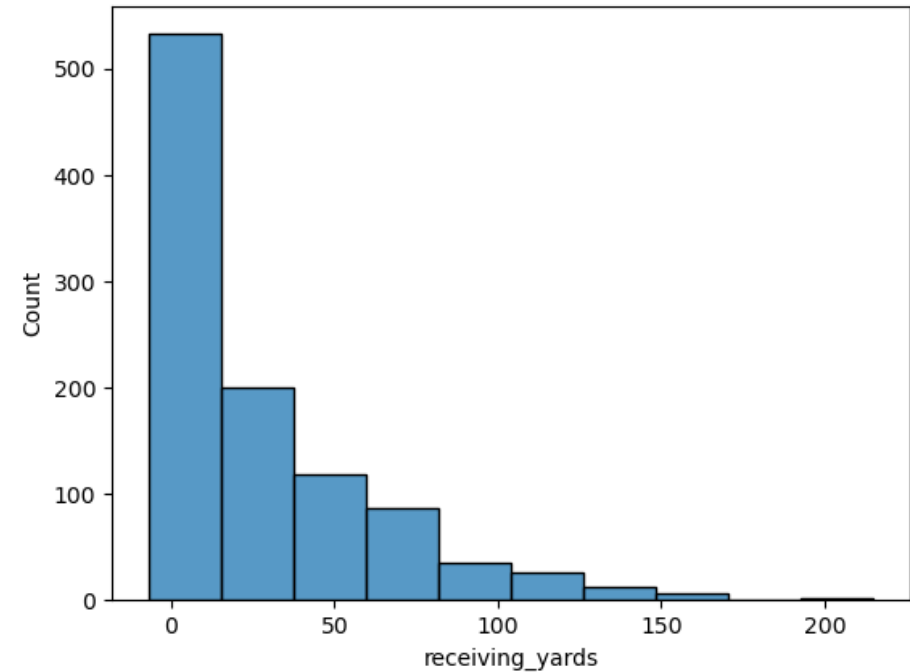
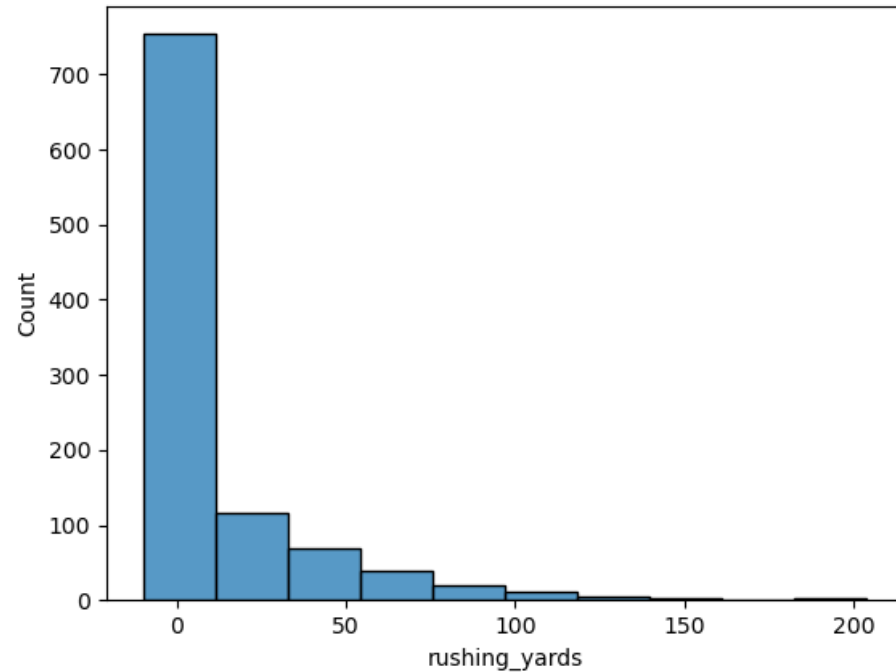
# Barplot



- Able to implement categorical data
- Compare categorical to numerical
- Easiest graph to understand
- Supports passing data

# Histograms

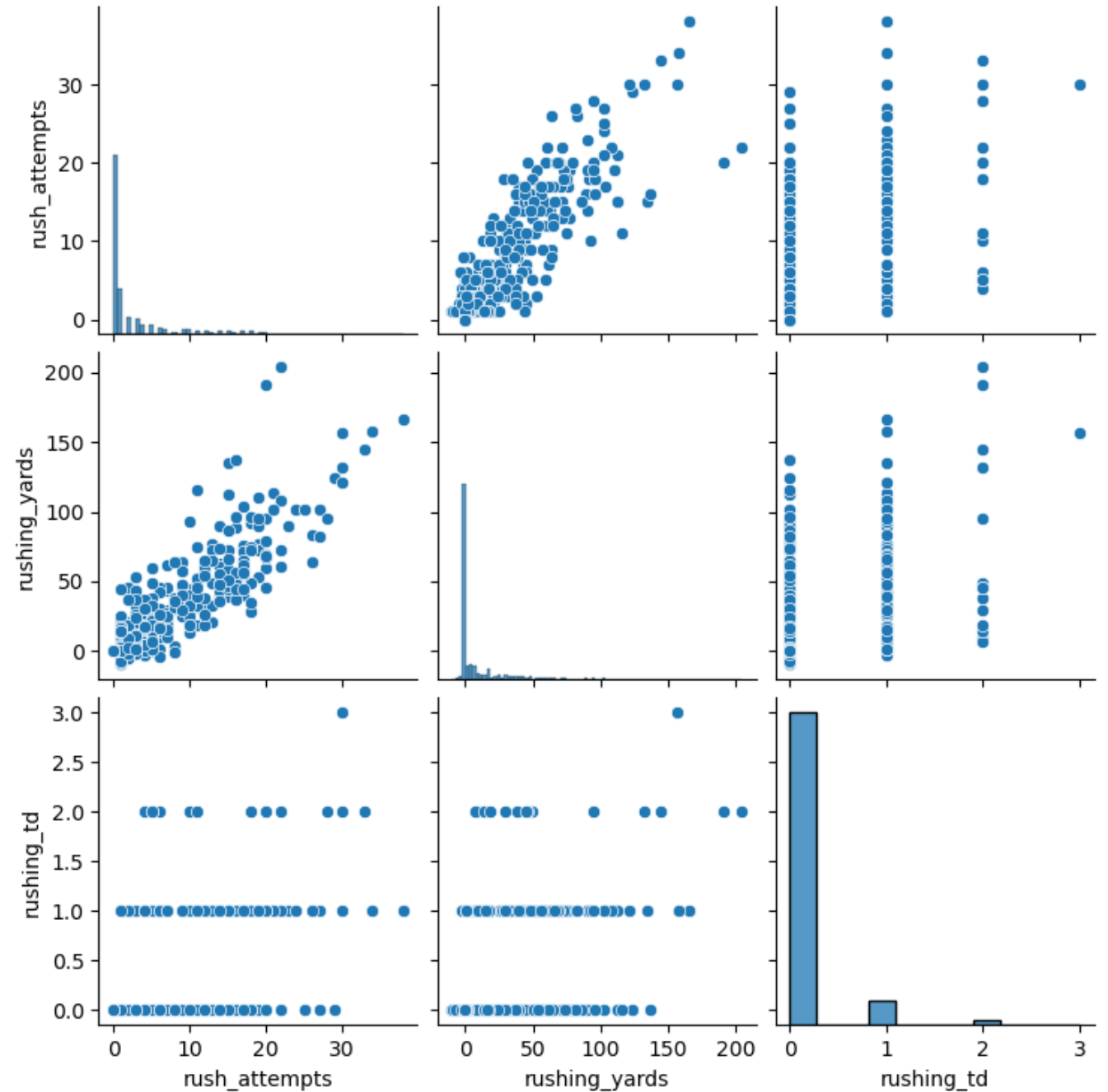
- Disproportionate
- Not very useful for our focus
- Similar to distribution graphs



# Bivariate Analysis – Pairplot

---

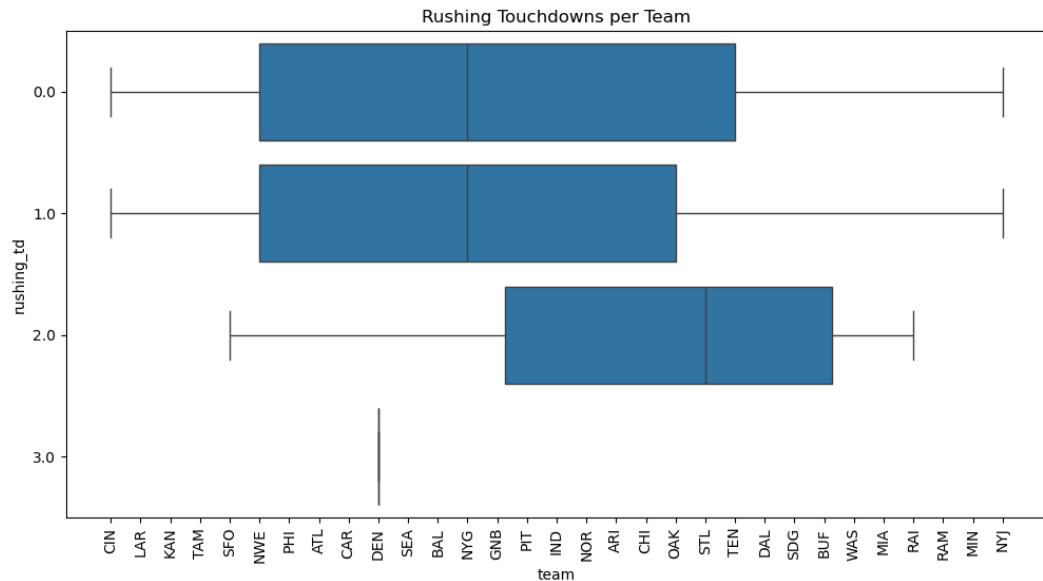
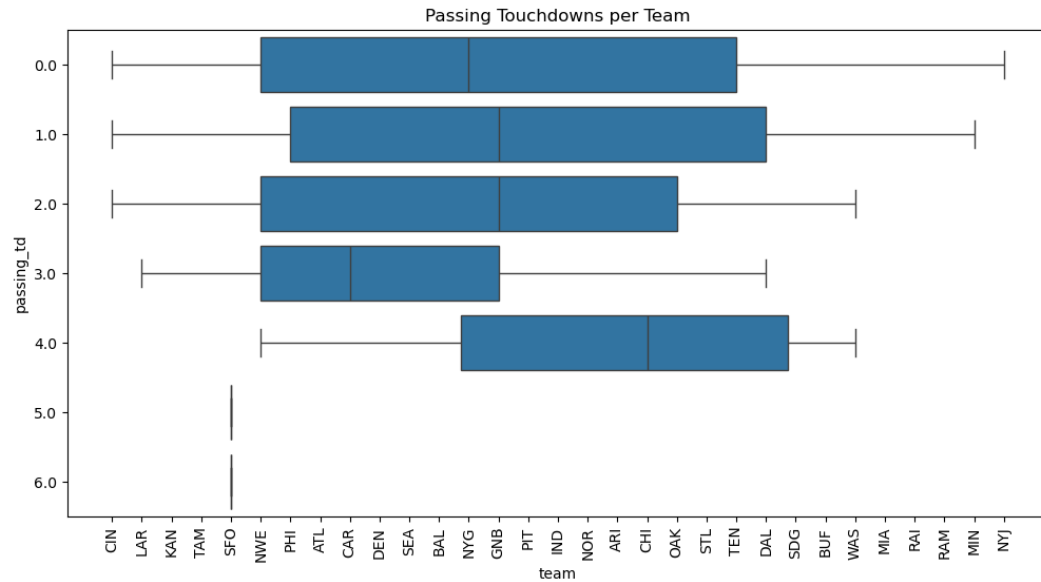
- Compares 3 variables together
- Overview of rushing statistics
- Scatter>Bar





# Box and Whisker Plots

- Correlation between multiple variables
- Compare categorical to numerical again



# Key Takeaways

- Defense playing good in Superbowl
- Passing game is more involved the rushing game
- Surprised to see higher passing touchdowns than rushing touchdowns

# Summary and Next steps

- 
- Can be used for Sports wagering
  - See defensive stats next
  - Would go with position-based data instead of all offense

Questions

