Capstone Project: NFL Analysis of Years Played

Eric Young

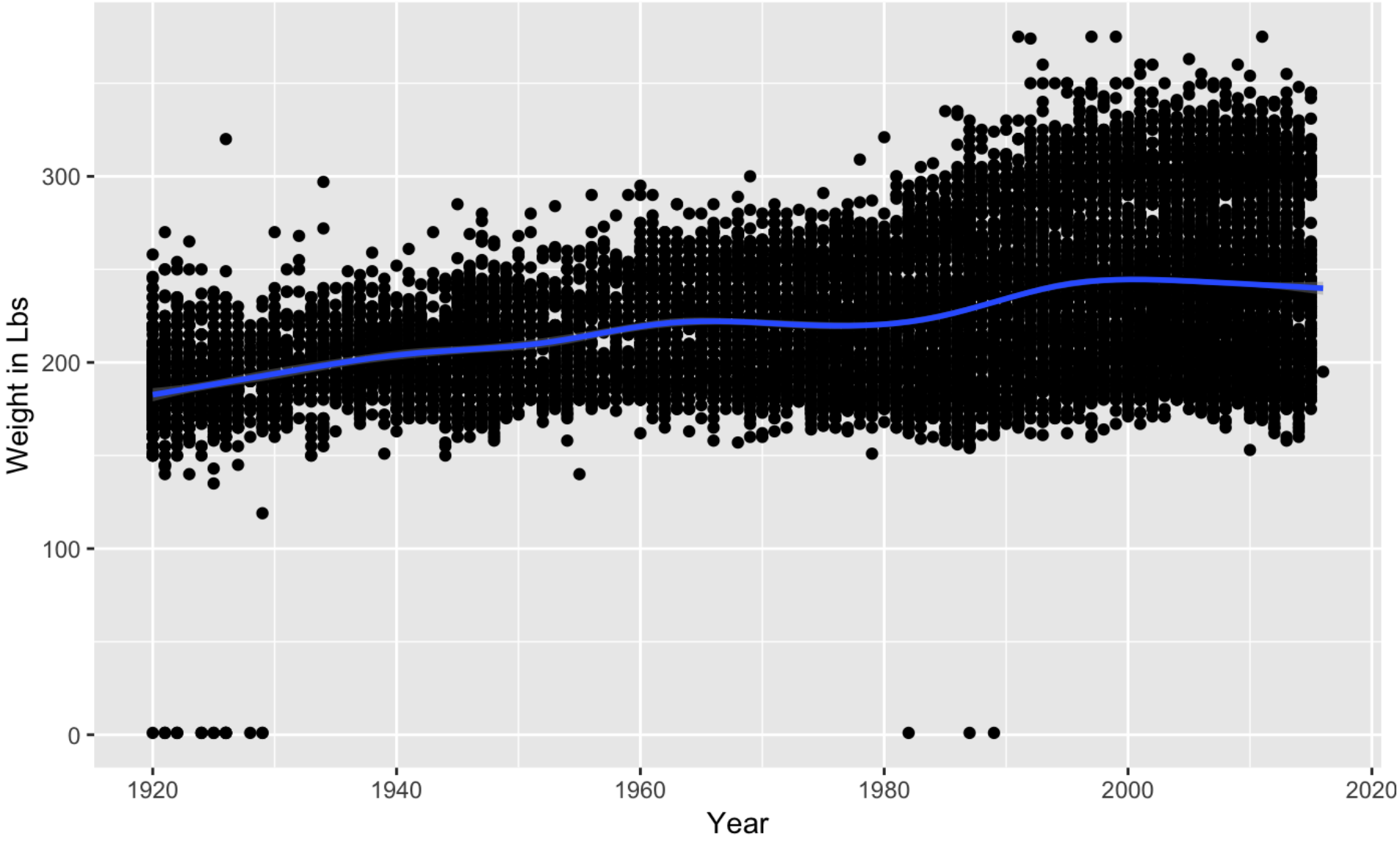
Foundations of Data Science Workshop

December 3, 2019

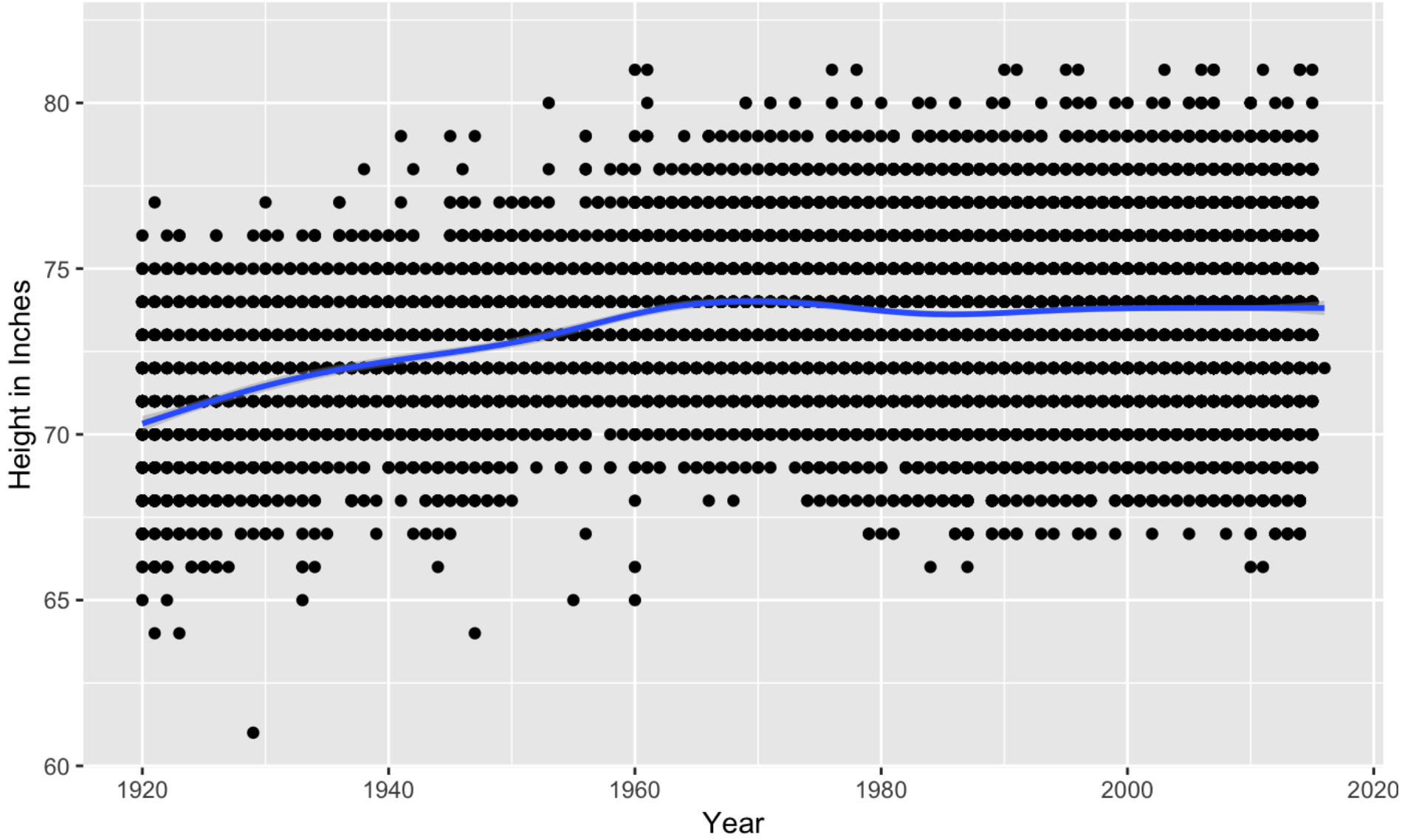
Business Case

- The goal of this project is to predict which players from NCAA College Football will be drafted into the NFL
- The outcome will whether the player is drafted or not
- Prediction starts from Freshman - Senior year
- Data comes from two different data sets one for NFL and other for NCAA

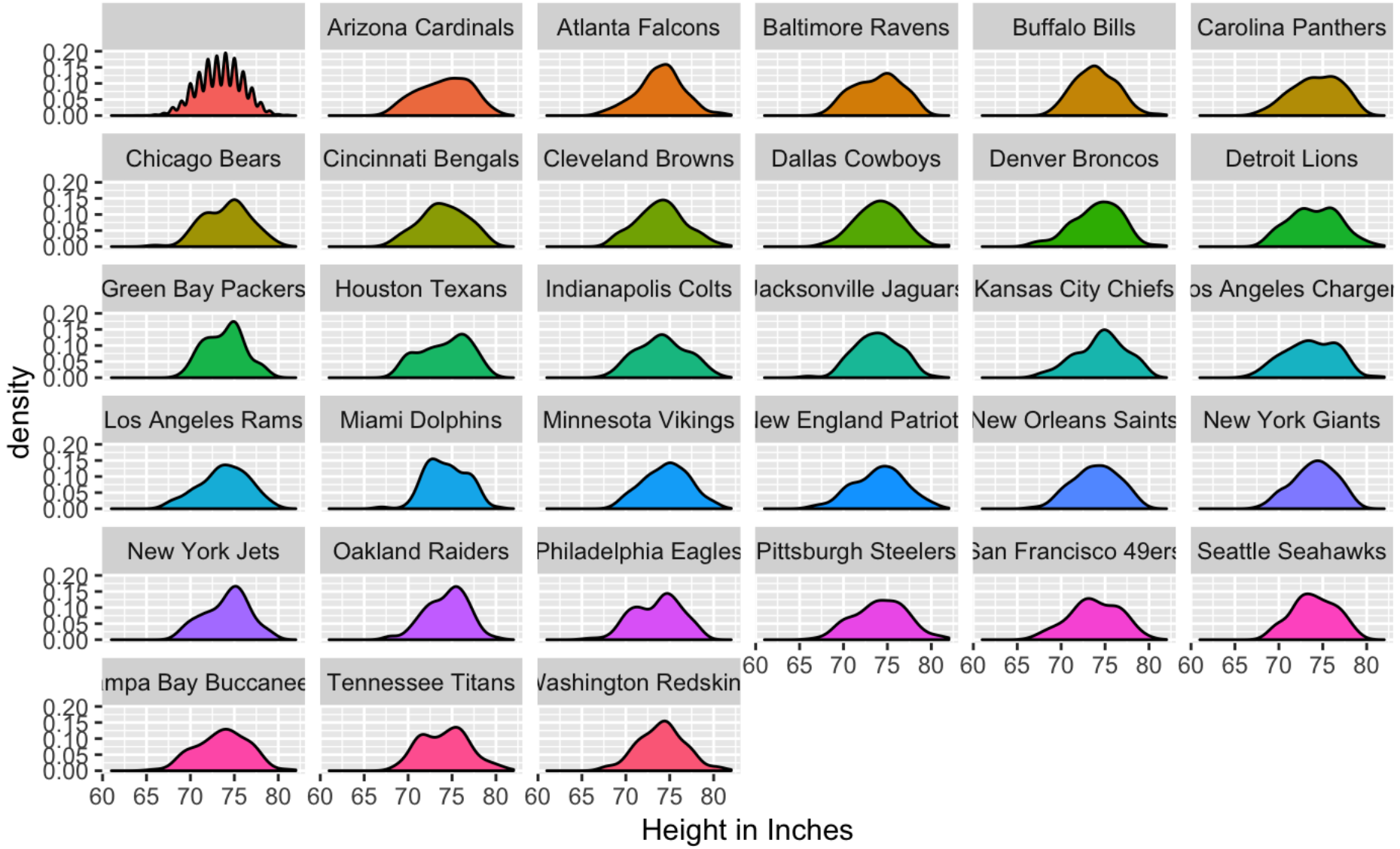
Are Players Bigger Now Than Before?



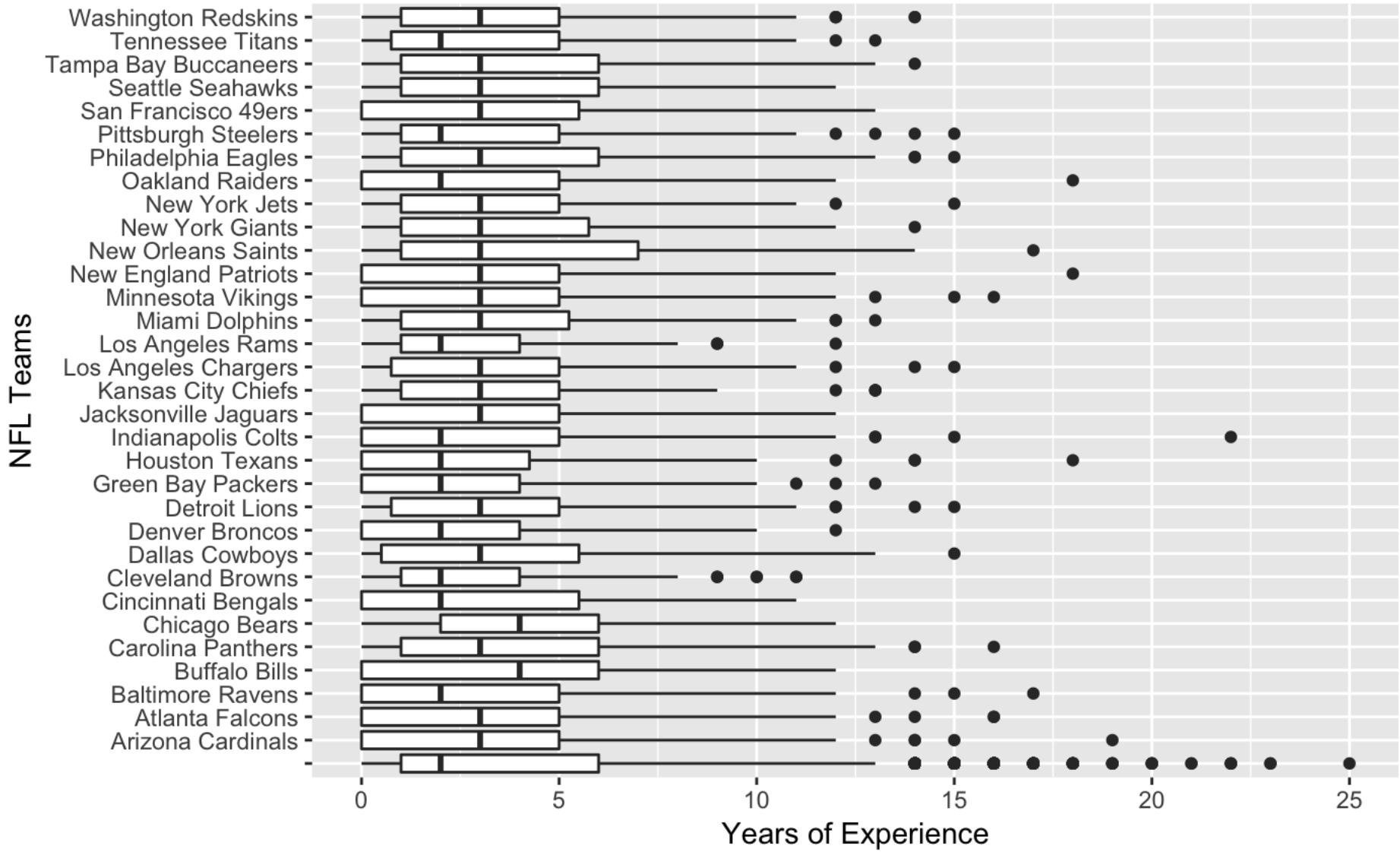
Are Players Bigger Now Than Before?



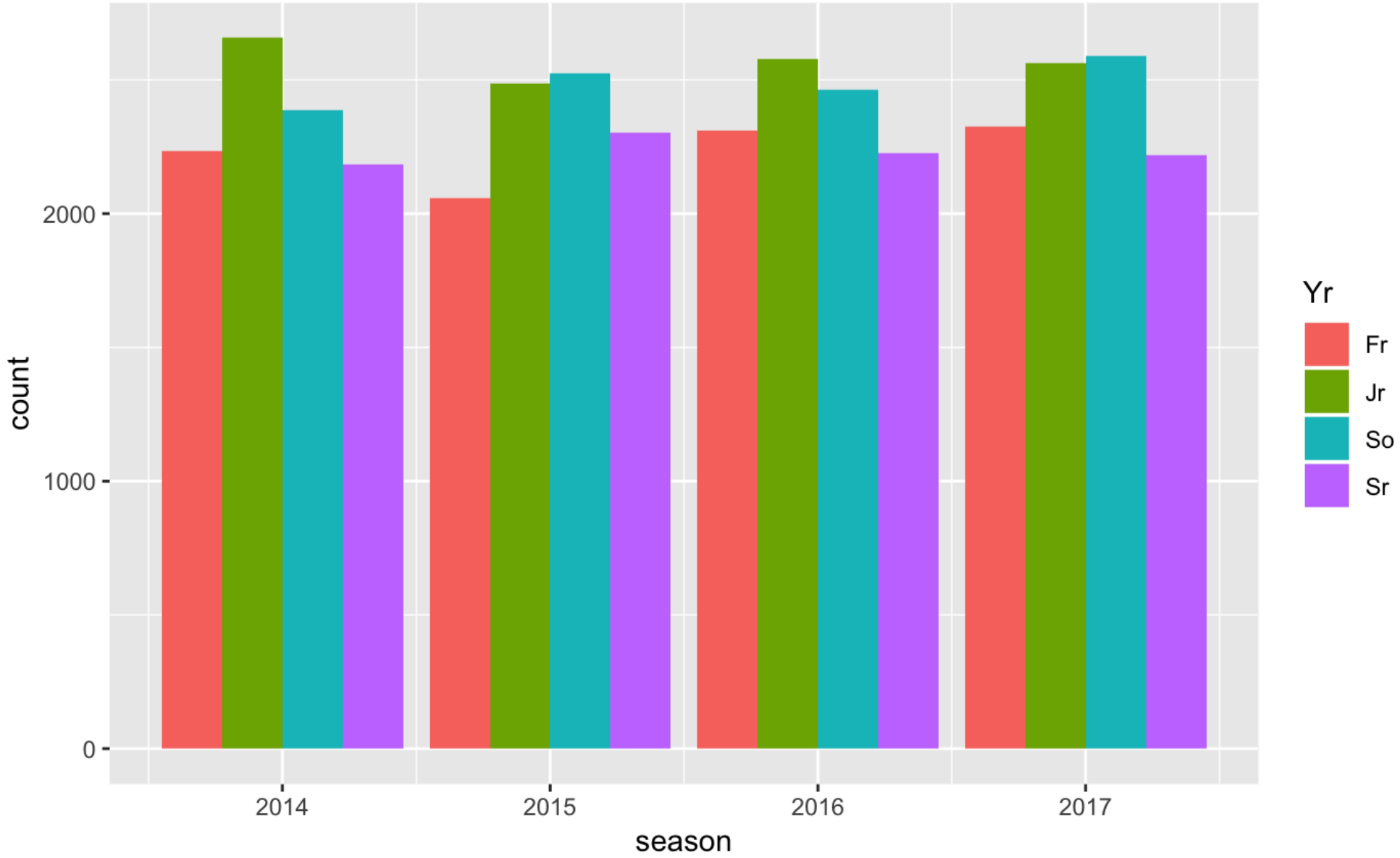
Height Difference Between NFL Teams



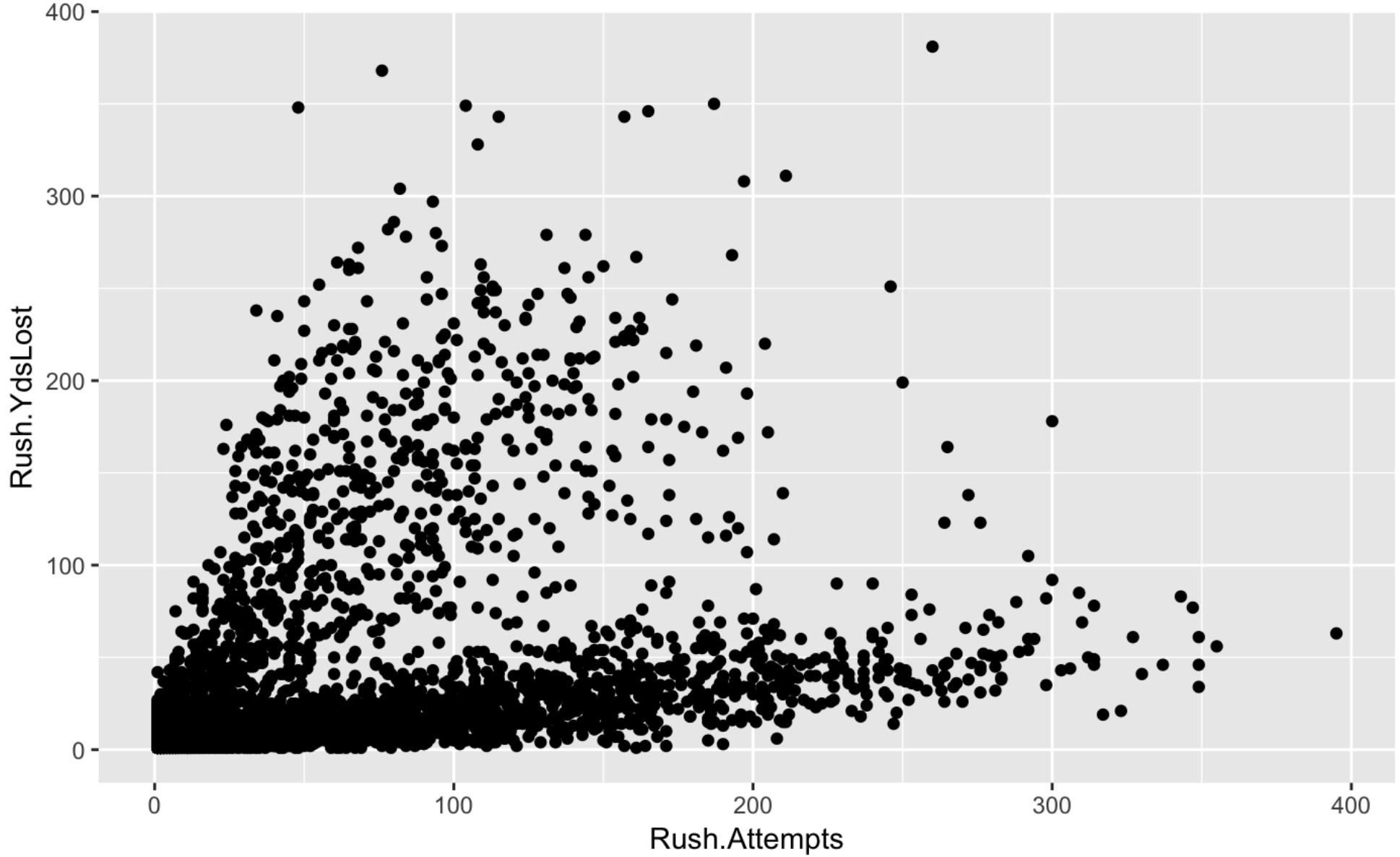
Players Years of Experience by Team



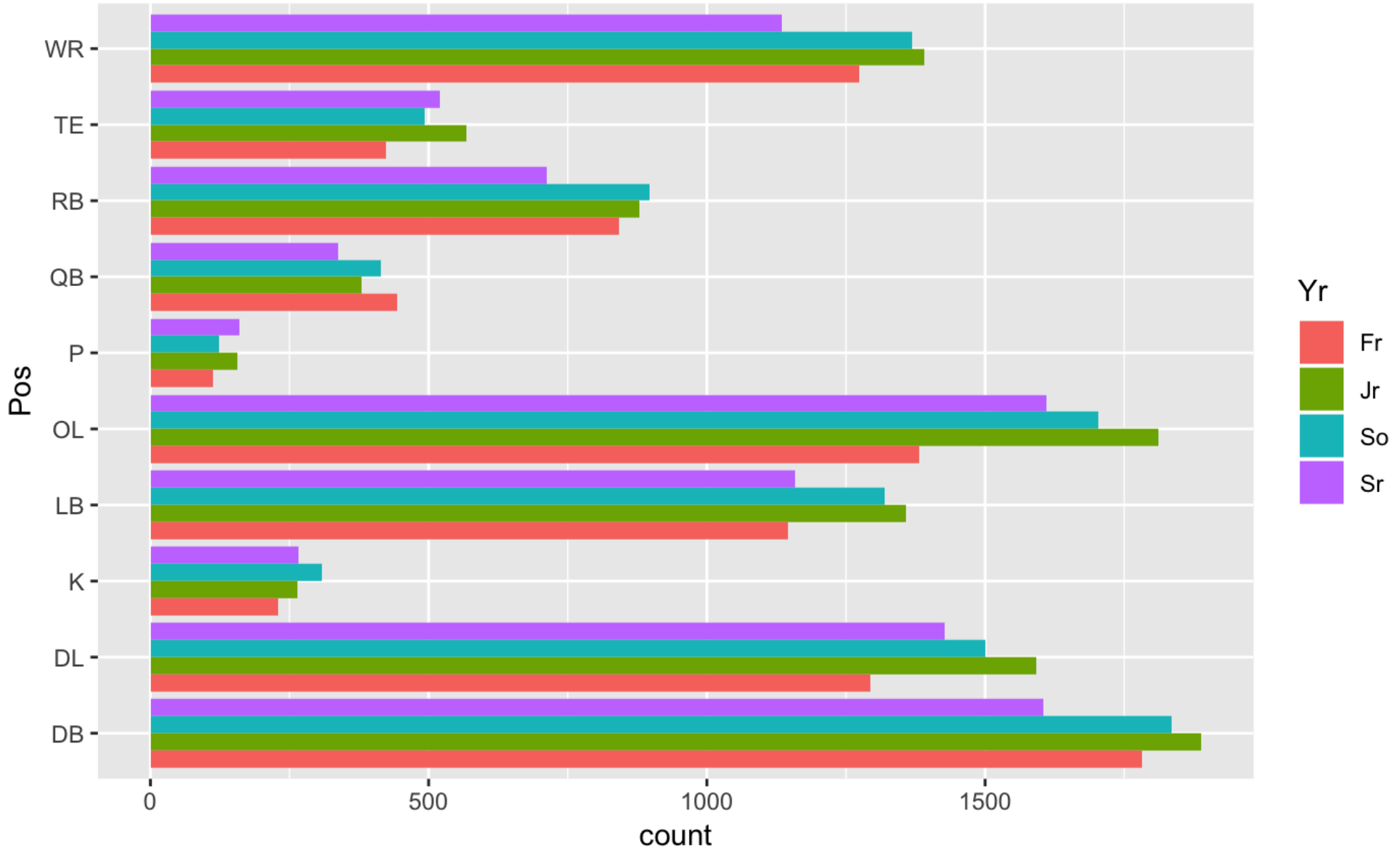
Distribution of Players by Year



Rushing Attempts by Rushing Yards Lost



Players Position by Year



Colleges with at least one Player Drafted

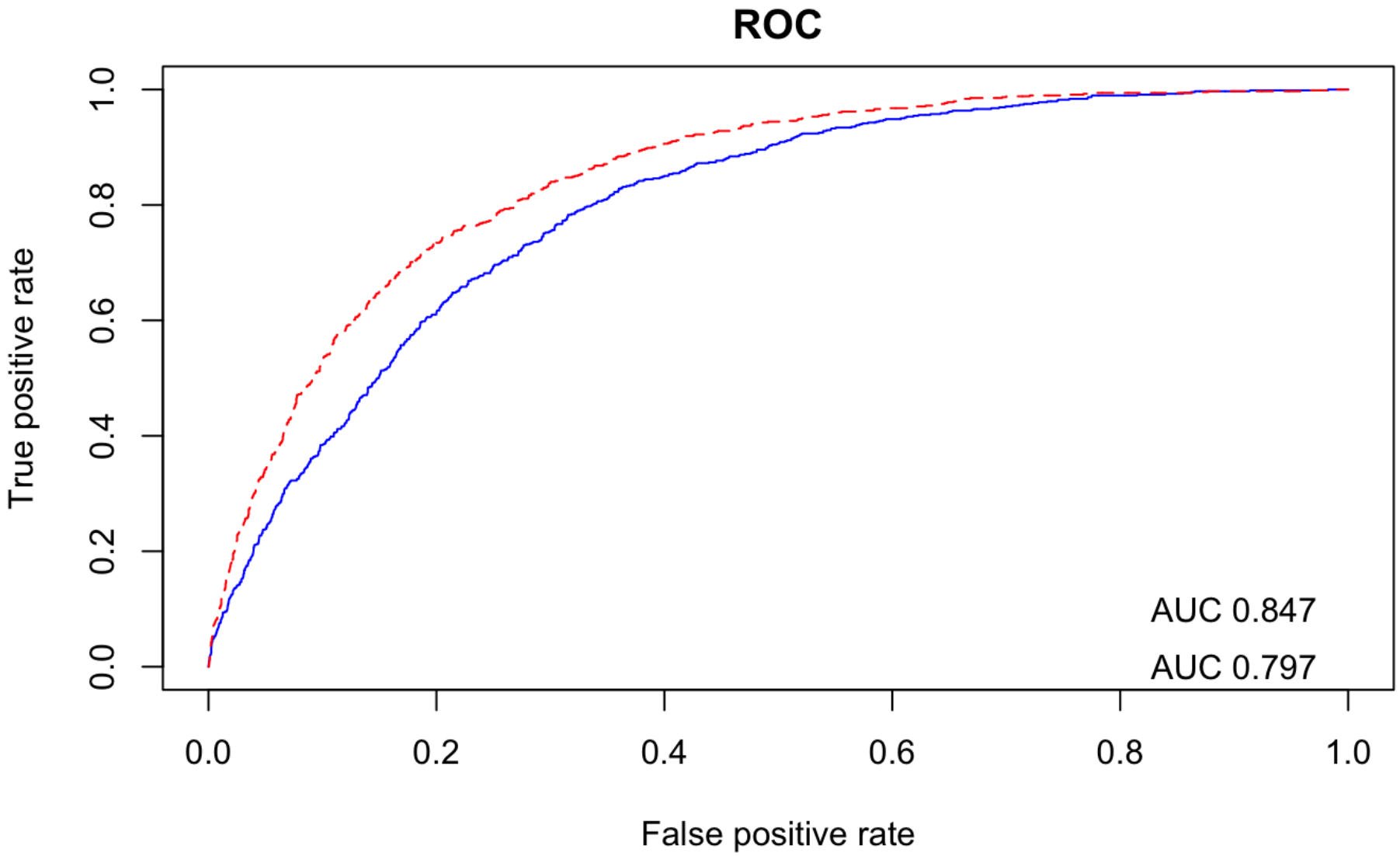
Linear Regression Models for Experience

- Linear Model 1: College
- Linear Model 2: Weight..lbs. + College
- Linear Model 3: Height..inches. + Weight..lbs. + I(Weight..lbs.^2) + Position
- Linear Model 4: Height..inches. + Weight..lbs. * Position
- According to SSE the best model was weight and college with an SSE of 198610.6. However, when I used AIC it preferred the model height and weight by position with a value of 90677.07.

Linear and Logistic Models for drafting

- Linear Model 1: Yr + Pos + GP + GS
- Linear Model 2: Yr + Pos + GP + Rush.Attempts + Rush.Net.Yards + Rush.YdsGained
- According to RSE model 1 is better with RSE of 0.2487. The Adjusted R-squared is 0.06974 compared to model 2’s Adjusted R-squared of 0.06529.
- Logistic Model 1: GP + Pos + Yr + team_name
- Logistic Model 2: Pos + GP + GS + season + team_name
- From the ROC plot on the next slide we see that logistic model 2 is the better model as the curve pulls more towards the top left corner.

ROC Curve Comparison



Conclusion

- Logistic Regression was most successful in predicting players being drafted into NFL.
  - The information would be useful for players and agents
- Linear Regression model on experience did not fit the data well
- Linear Regression model for drafting was better than experience model but was not as accurate as the Logistic Regression models.
- The most important variables were games played, games started, and year.