

Predicting NFL First-Year Quarterback Success

Browning Gentry

Question

- Every year, NFL teams attempt to select the best college football quarterbacks in the NFL draft
- Is it possible to predict which college quarterbacks will be successful?



Data Sources

- Data scraped from Sports-Reference.com
 - College and Pro statistics pulled separately
- Response – Players First Year Passer Rating
 - First year with a 10 game season
 - If no 10 game season, then first year
- Features:
 - Numeric – Play stats, height, weight
 - Categorical – Throws with, College Team, NFL Team, Home City/State



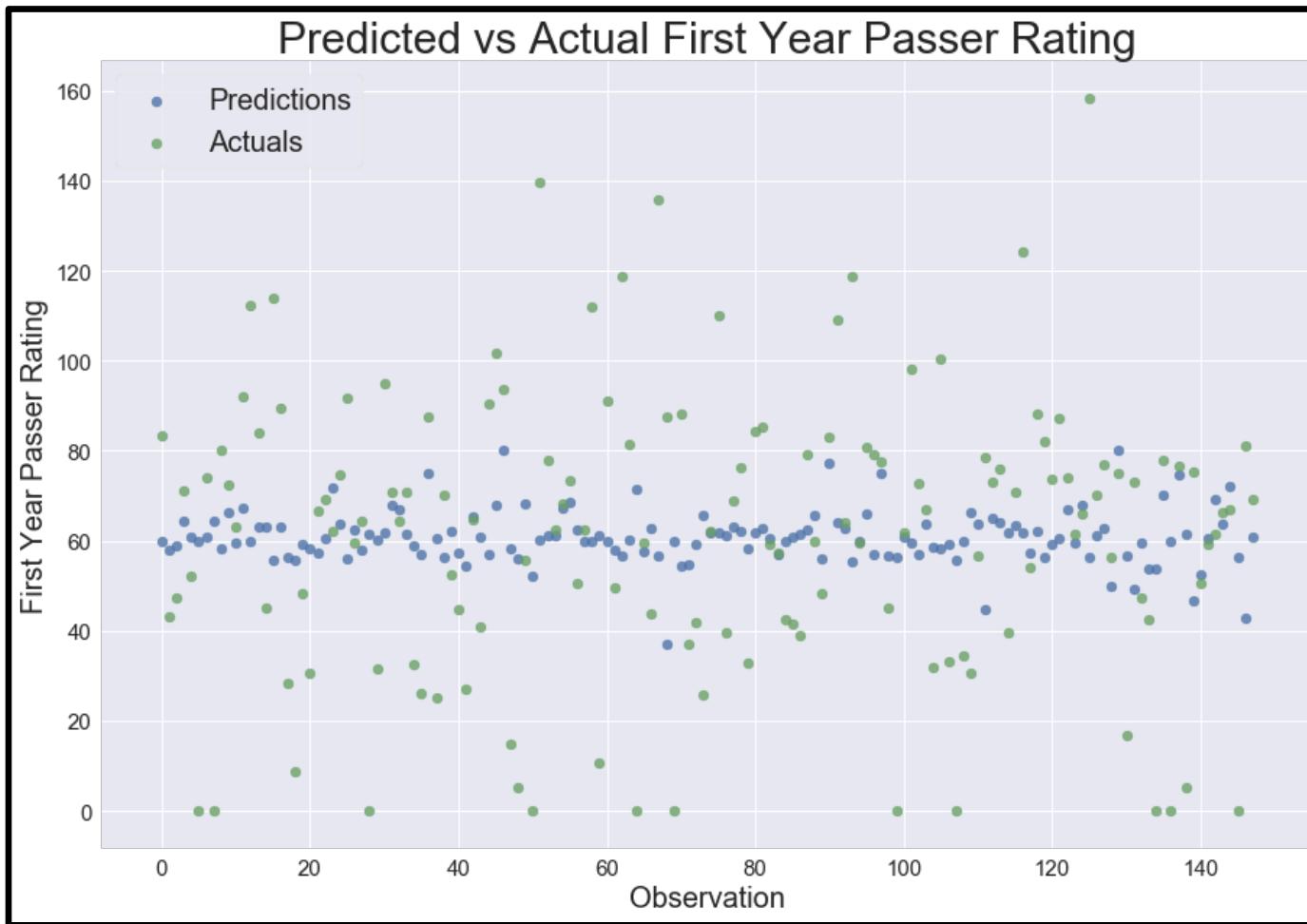
Data Cleaning

- Combined into a single DataFrame:
 - Player biography information
 - NFL first season passer rating
 - College performance statistics
- Dropped 6 quarterbacks that had duplicate names
- Since all college players are required to play 2 years before being drafted, calculated:
 - Graduate Year Statistics
 - 2nd to Last Year Statistics
 - Difference

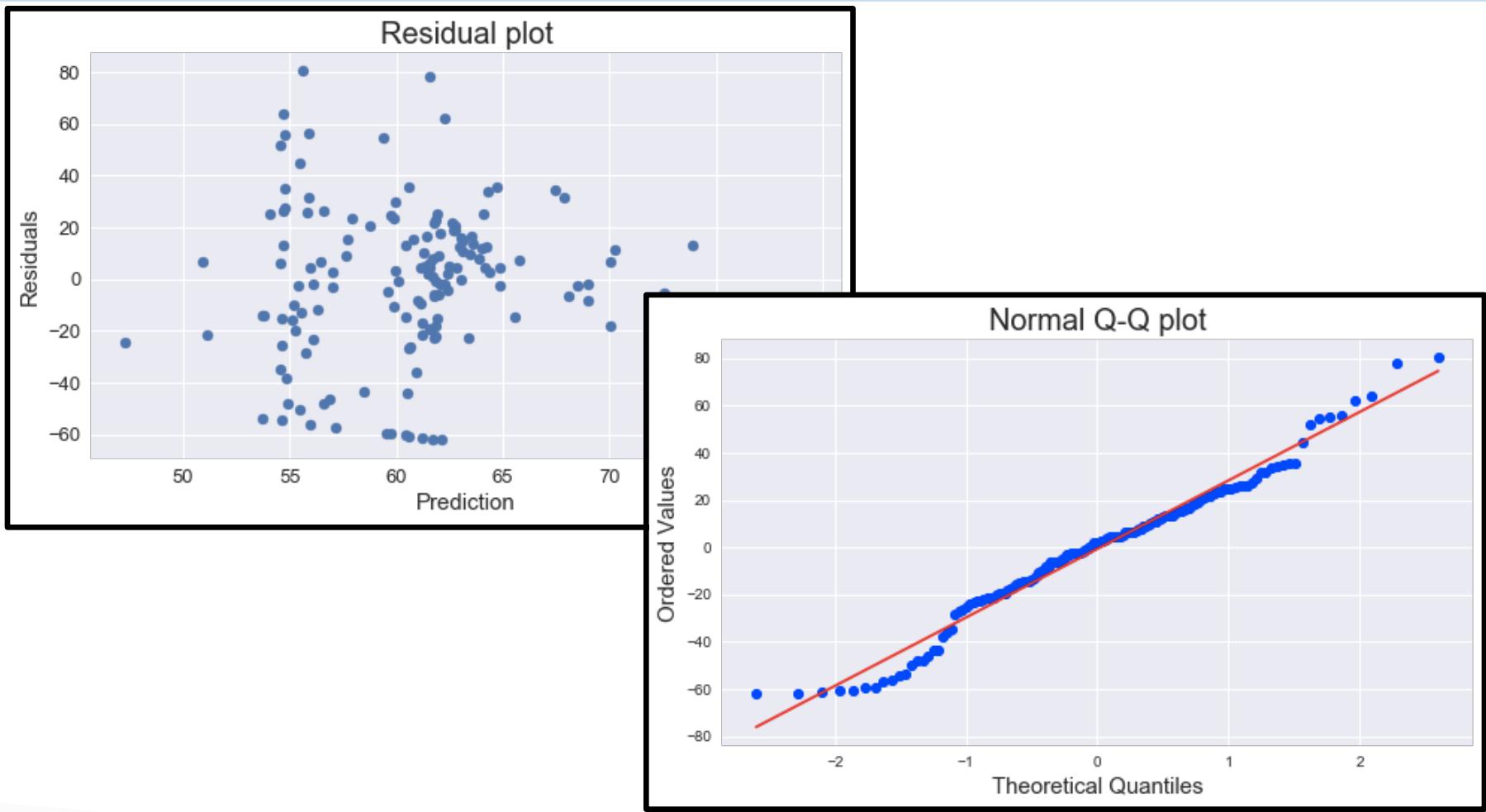
Linear Regression

- Roughly 500 observations
- 39 Numeric Features, 6 Categorical
- Initial Linear Regression model trained on all data showed R^2 value of $\sim .96$
- Once Train/Test Split was introduced, R^2 on the test set dropped to $\sim .02$, indicating the model was overfit

ElasticNet Regularization



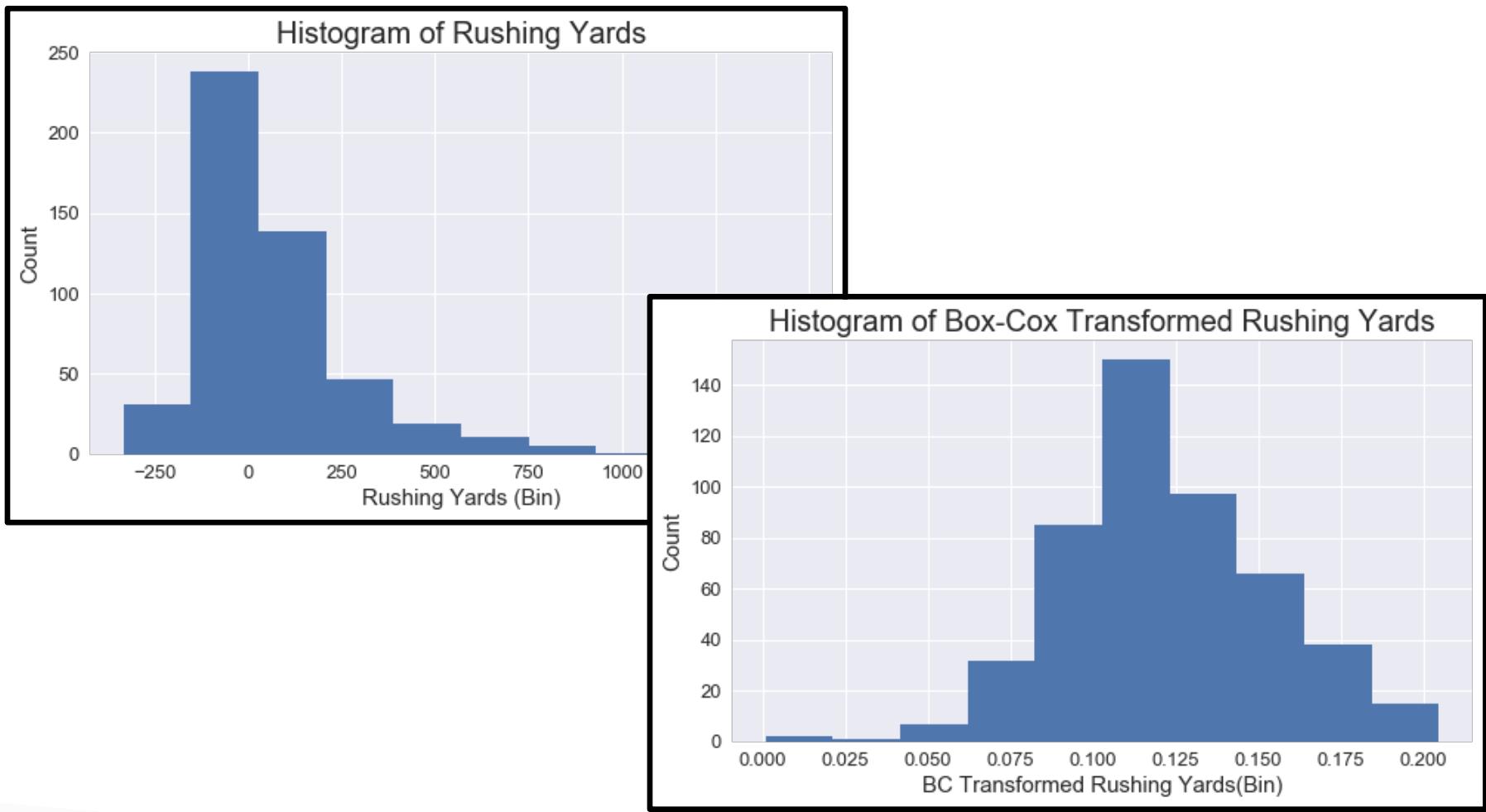
ElasticNet Regularization Validation



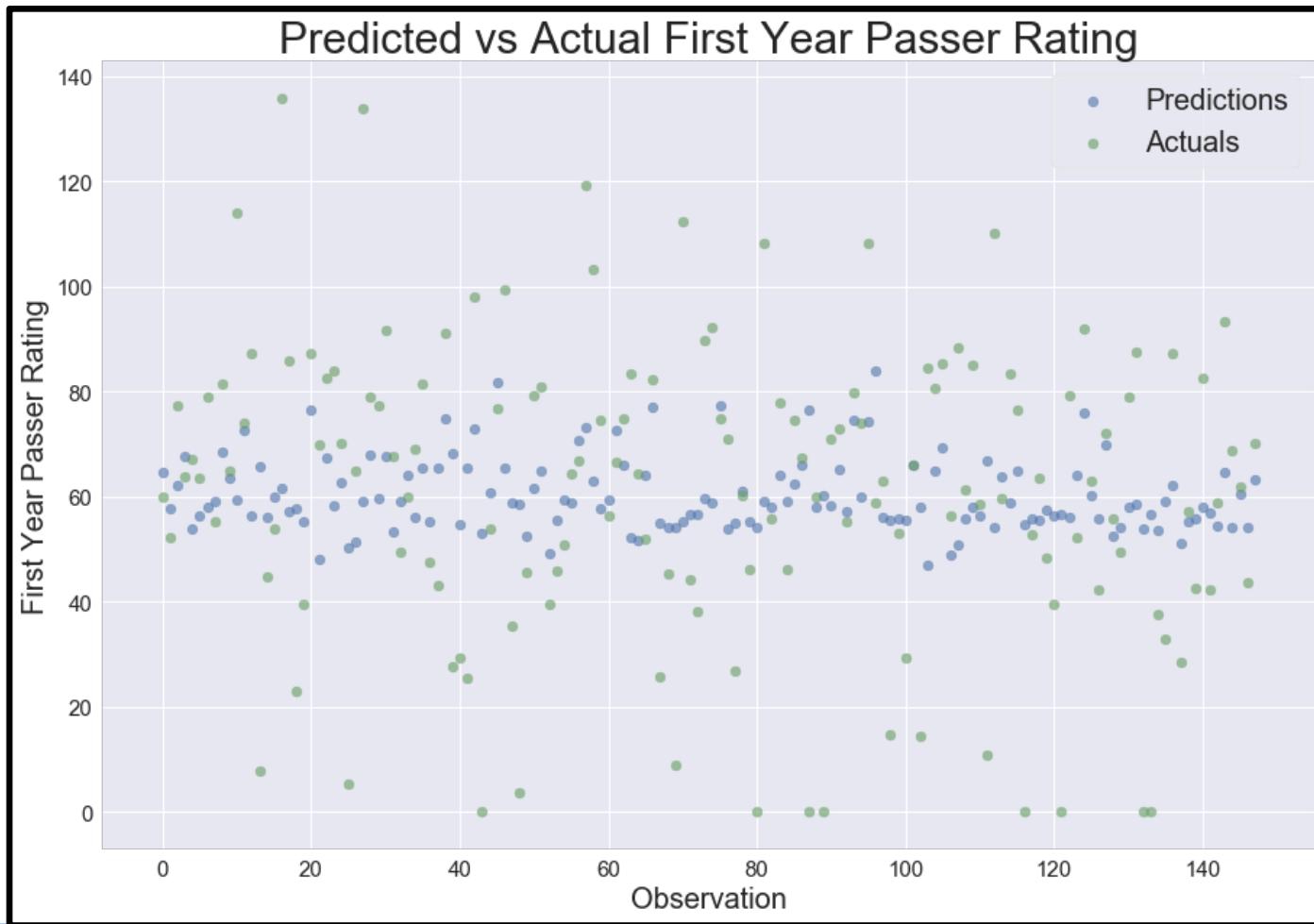
Box-Cox Transformations

- Many negative values in features - , used MinMaxScaler() to scale values between 0 and 1 prior to Box Cox transformation
- Initially attempted to transform all features, but resulted in an very low R^2 – even on the entire dataset
- In second iteration, selected features manually
 - Found that certain features, when transformed, were causing R^2 to be low when fit

Box-Cox Transformations



ElasticNet Regularization with BoxCox Transformed Features



Polynomial Features

- Used second-degree polynomial features to improve the predictive ability and introduce feature interactions
- Initially did not run because of the large number of features – had to remove some categorical features
 - Home State, Home City
- Still resulted in low R^2

Conclusions

- Model has low predictive power
- Generally predicts values around the mean
- Need to be able to predict “outliers” to sufficiently answer the question – NFL teams are not looking for average quarterbacks

Next Steps

- Psychological Data Features
 - News Articles – Positive vs. Negative, Conflict, etc.
 - Wunderlic Test/Player Assessment Tool
- Consider NFL team data – can a quarterback really overcome the weaknesses of the team?
- Injury Data
- High School Data
- Different Response Variables
 - Games Started in Career
 - Career Average Passer Rating

Appendix

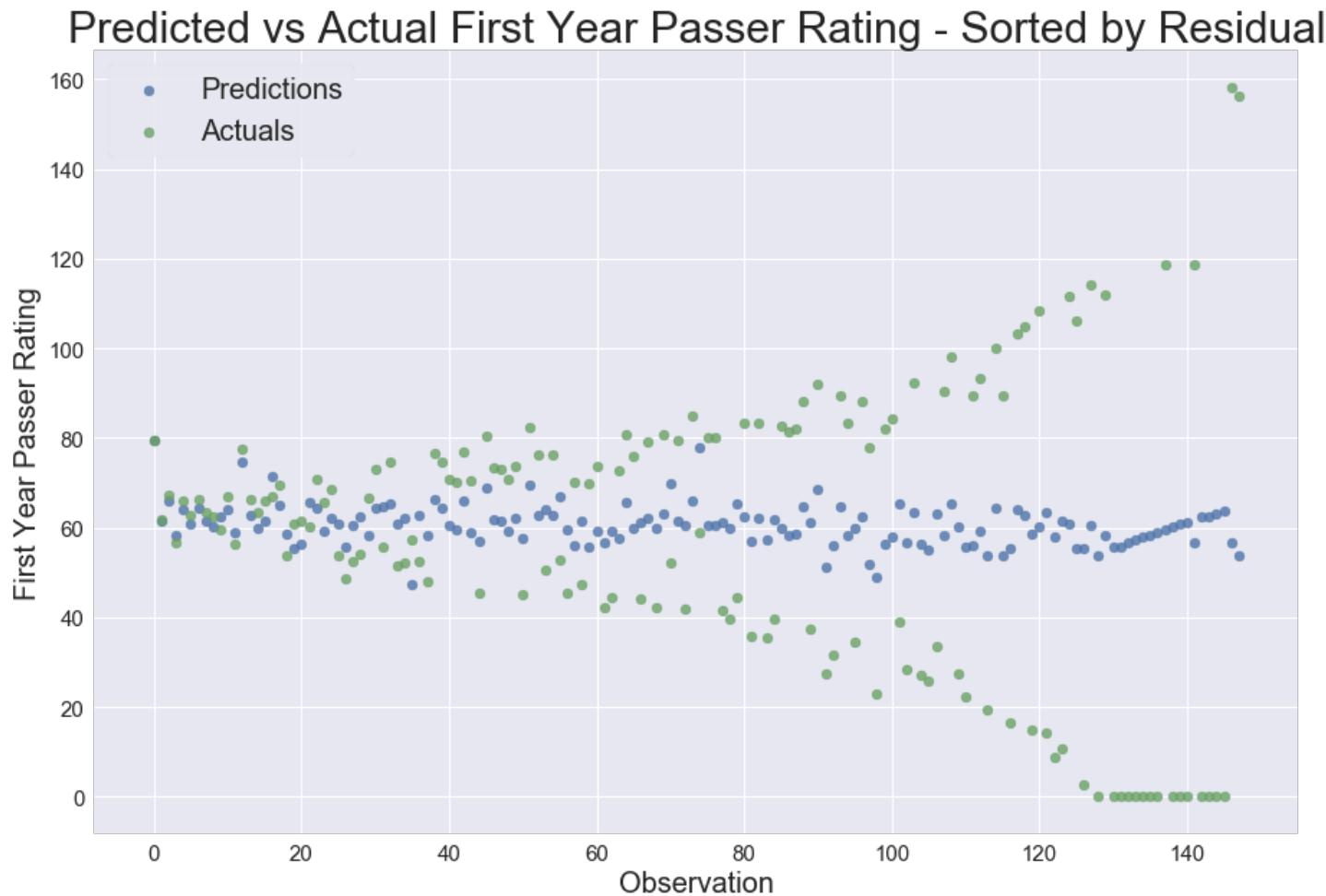


Image Links

- <https://theultimatesportsblog.com/2017/11/23/what-every-nfl-team-has-to-be-thankful-for-2/>
- <https://frma.toluna.com/opinions/3616777/%C3%87a-vous-est-d%C3%A9j%C3%A0-arriv%C3%A9-de-vouloir-absolument-poser-une>
- <http://bleacherreport.com/articles/2662128-the-education-of-lamar-jackson-how-louisville-qb-went-from-project-to-superstar>
- <http://wyomingpublicmedia.org/post/uw-quarterback-josh-allen-attends-manning-passing-academy>
- http://soonersports.com/fls/31000/microsites/15fb_bakerheisman/index-working.html
- <http://weratethings.com/2015/10/20/rate-tom-brady-and-the-patriots/>
- <http://www.neosportsinsiders.com/johnny-manziel-defends-actions-twitter/>