

# STAT 450: REAL ESTATE

PETER HAN, XUECHUN LU, YUETONG LIU, YUTING WEN  
DEPARTMENT OF STATISTICS, UNIVERSITY OF BRITISH COLUMBIA



## ABSTRACT

- This study aims to build a predictive model that can accurately predict the 2020 Mill Rate for Residential, Light Industry and Business properties in Metro Vancouver.
- Basic and Advanced Linear Models were built to predict the future Mill Rate using variables selected from Exploratory Data Analysis.
- After assessing the goodness of fit and prediction power of each model and integrating them with the client’s preference, the transformed model was chosen to perform our 2020 prediction.

## INTRODUCTION

- **Objective:**  
Predict mill rates in 2020 in Metro Vancouver, and identify which explanatory variables are the most important in determining mill rates.  
  
Mill rate is adjusted based on the total assessment in each city so the municipal government can use tax earning (total assessment \* mill rate) to match their annual expense to balance the city’s budget.

- **Variables within the data:**

Variable	Type
Mill Rate	Continuous
Total Assessment	Continuous
Total number of properties	Continuous
Tax Class Code	Categorical
Municipality	Categorical

## METHODS

- **Exploratory Data Analysis (EDA):** visualize the main characteristics of our data and determine which variable is used in the linear models.
- **Measure:** MSPE (Mean Squared Prediction Error) measures Goodness of Fit and Prediction Power.
- **Models:** Ordinary Linear Models, Ridge and Lasso.
- **Cross Validation:** A 50-run of 10-fold cross-validation was used to compute the MSPE on the training set and the test set for each model.

## RESULTS

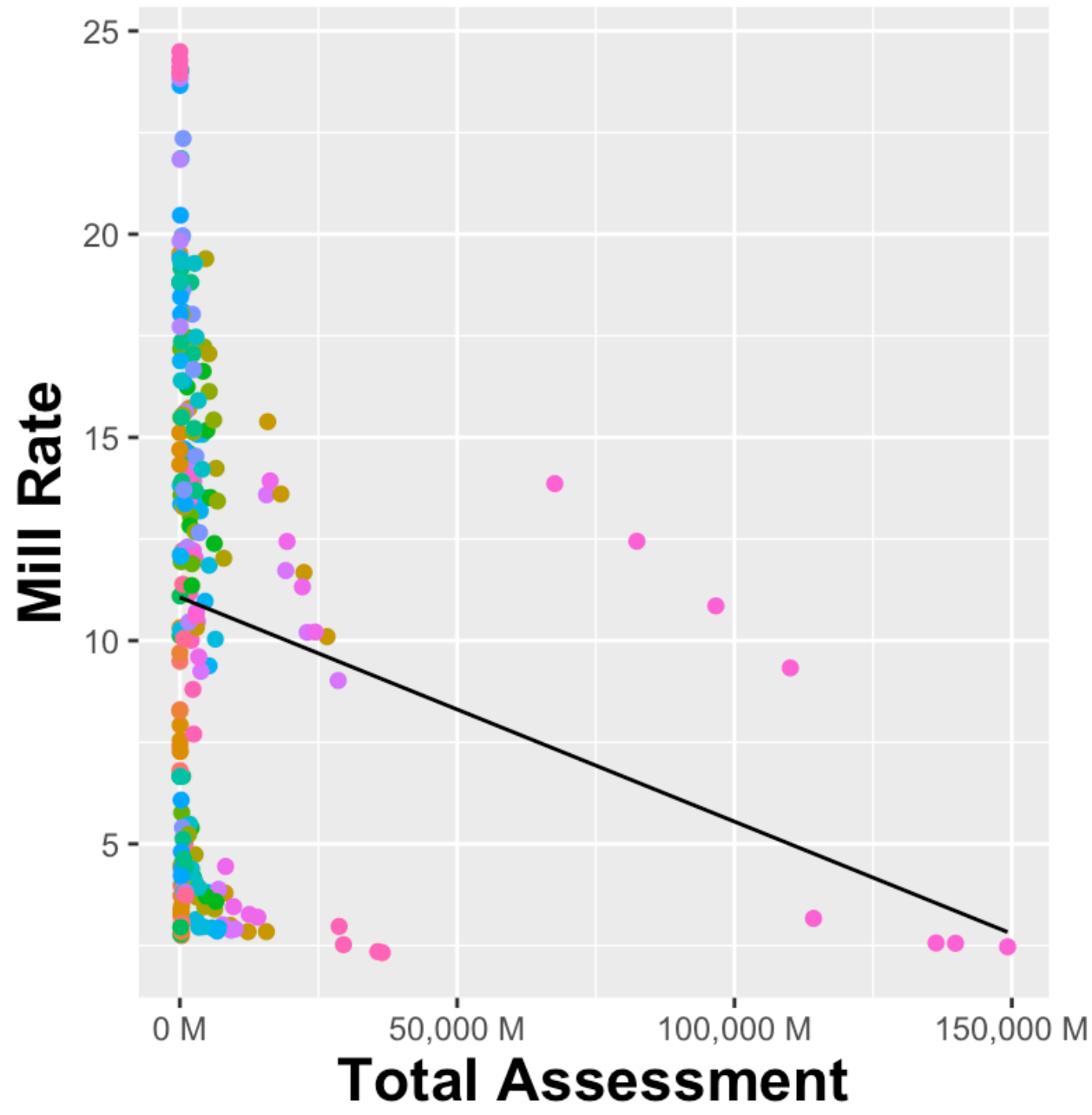


Figure 1: Mill rate v.s. Total Assessment

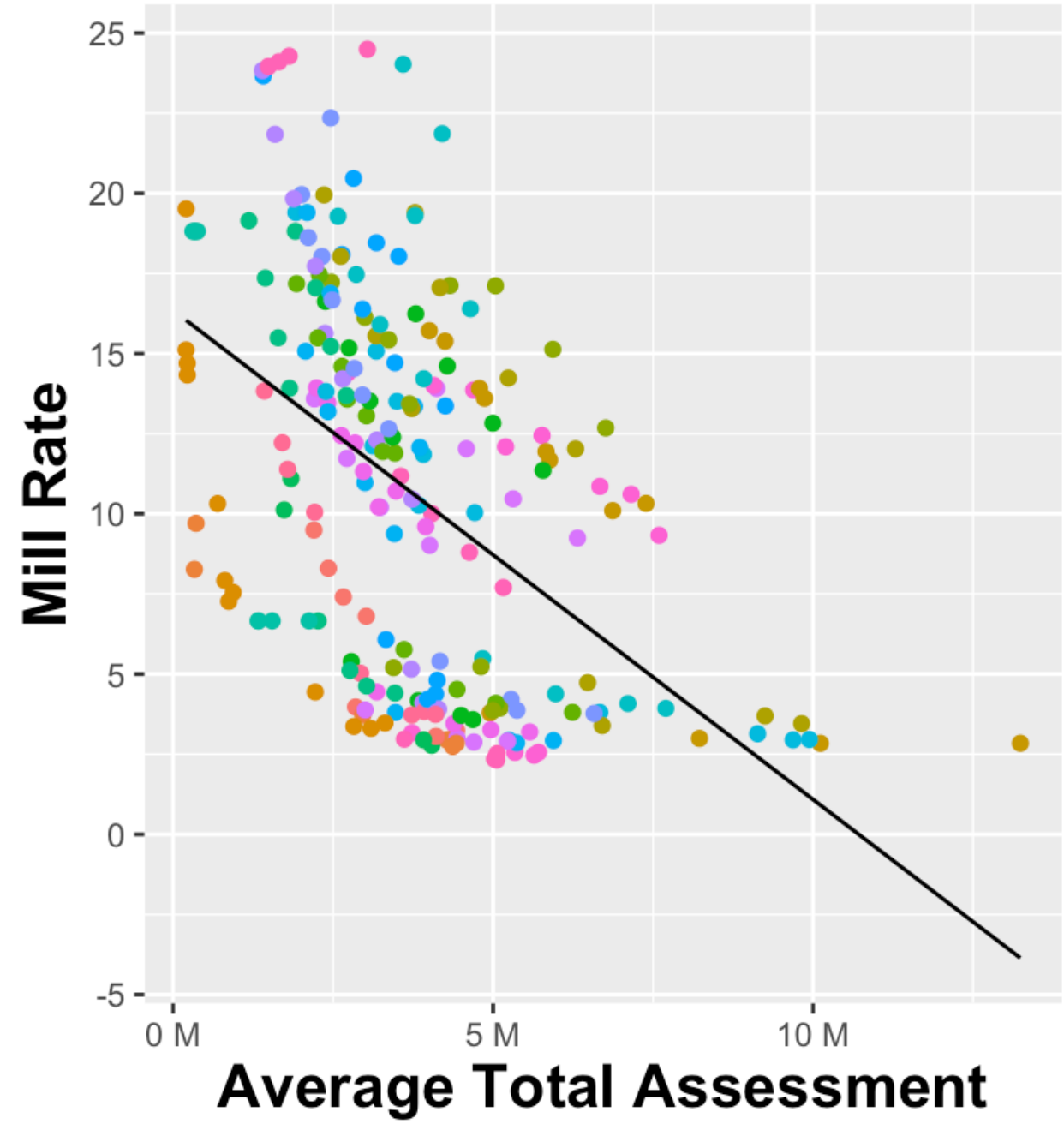


Figure 2: Mill rate v.s. Average Total Assessment

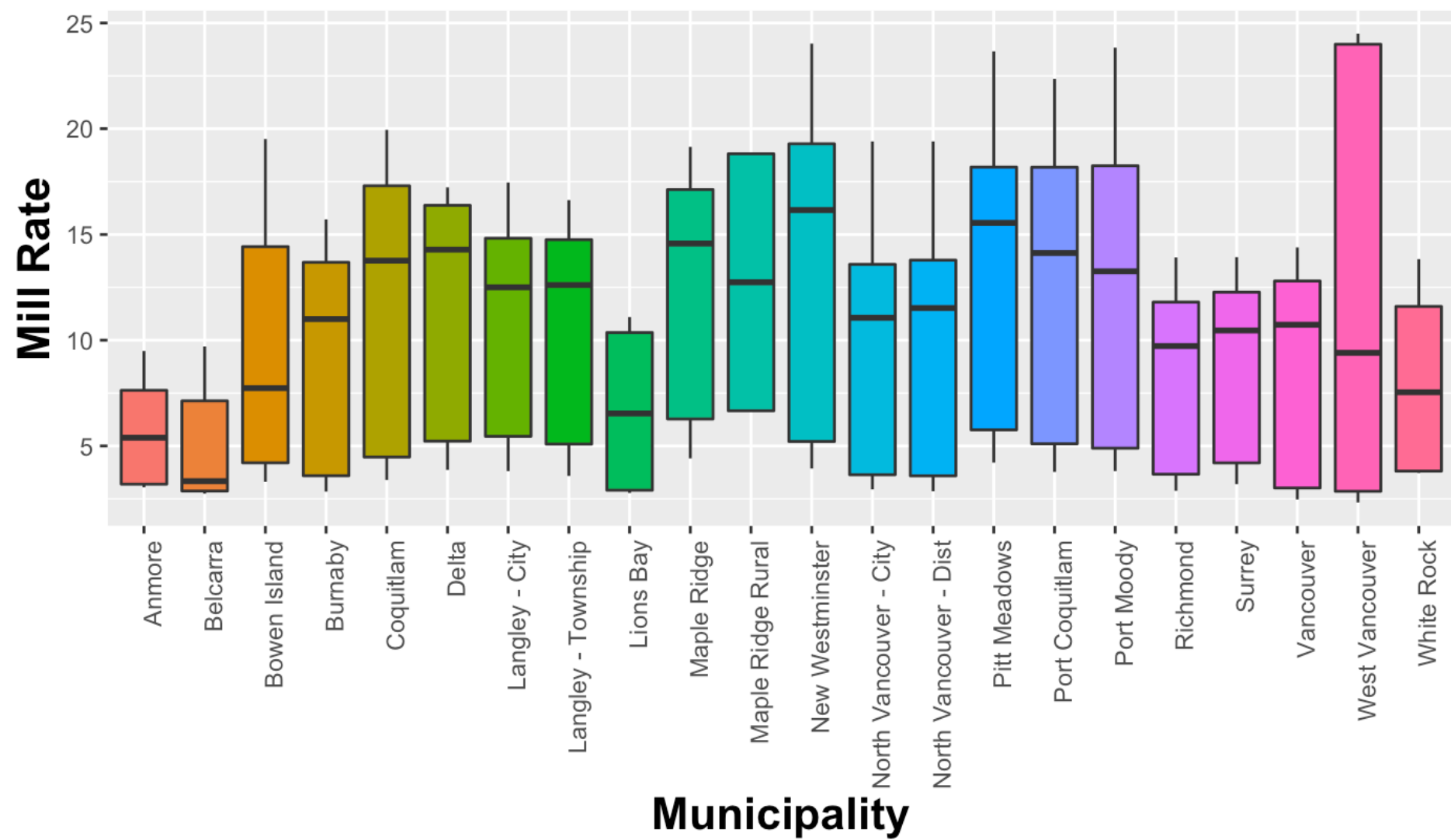


Figure 3: Mill rate across Municipalities

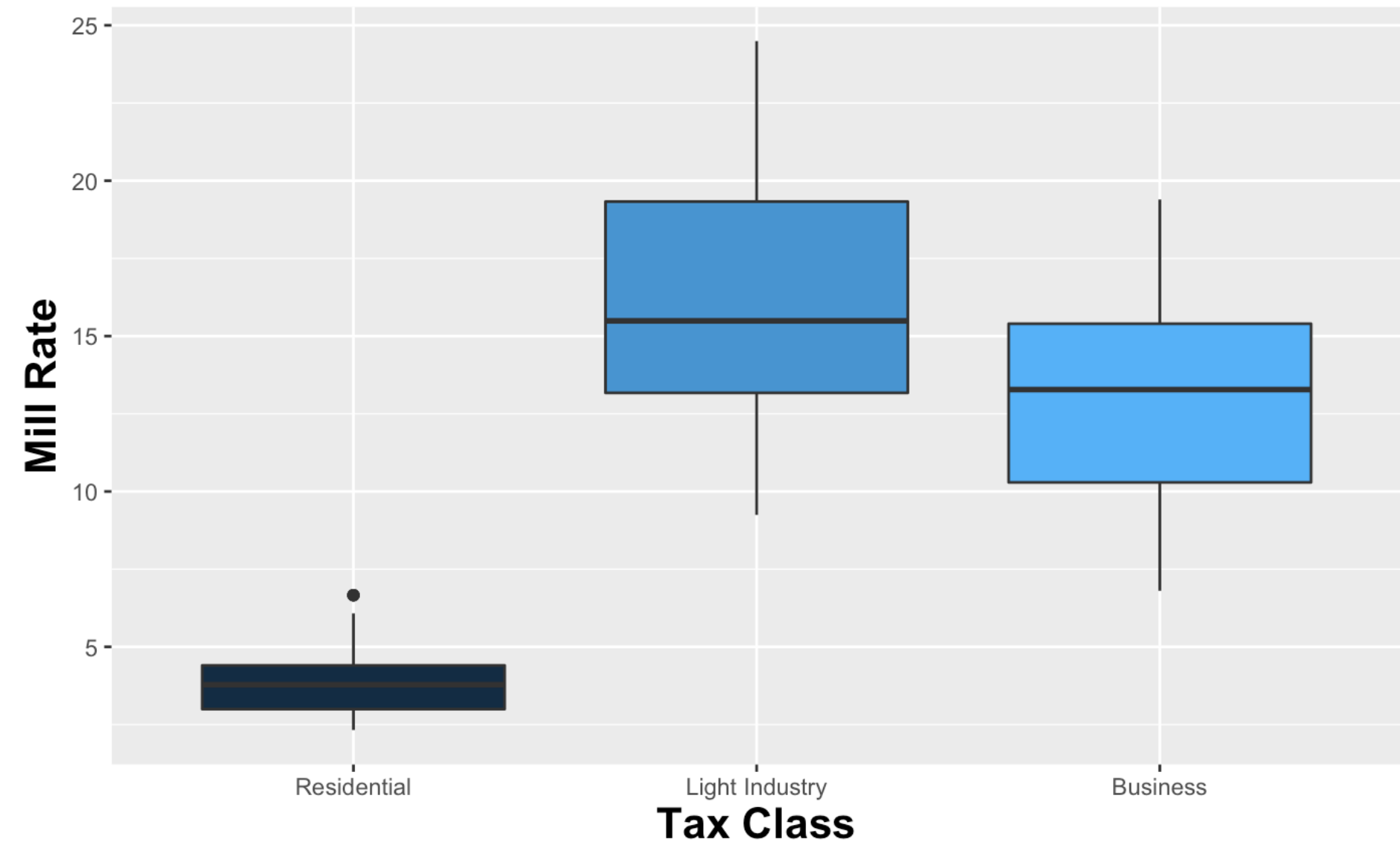


Figure 4: Mill rate across Tax Classes

### EDA - the most relevant features:

- There was no clear trend between the mill rate and total assessment (**Figure 1**). Hence total assessment across all municipalities was transformed by taking total assessment divided by the number of properties of each municipality and tax class. This had reduced the outliers in our data set (**Figure 2**). It showed a moderately strong correlation. Also, there were unequal mean and variance across tax classes and municipalities (**Figure 3, 4**). The p-values from **Kruskal Wallis test** had also supported that there were nonidentical distributions of mill rates across municipalities (0.00675) and tax classes ( $\approx 0$ ).
- Therefore, Average Total Assessment, Municipality, and Tax Class were included in the Transformed Ordinary Linear Model.

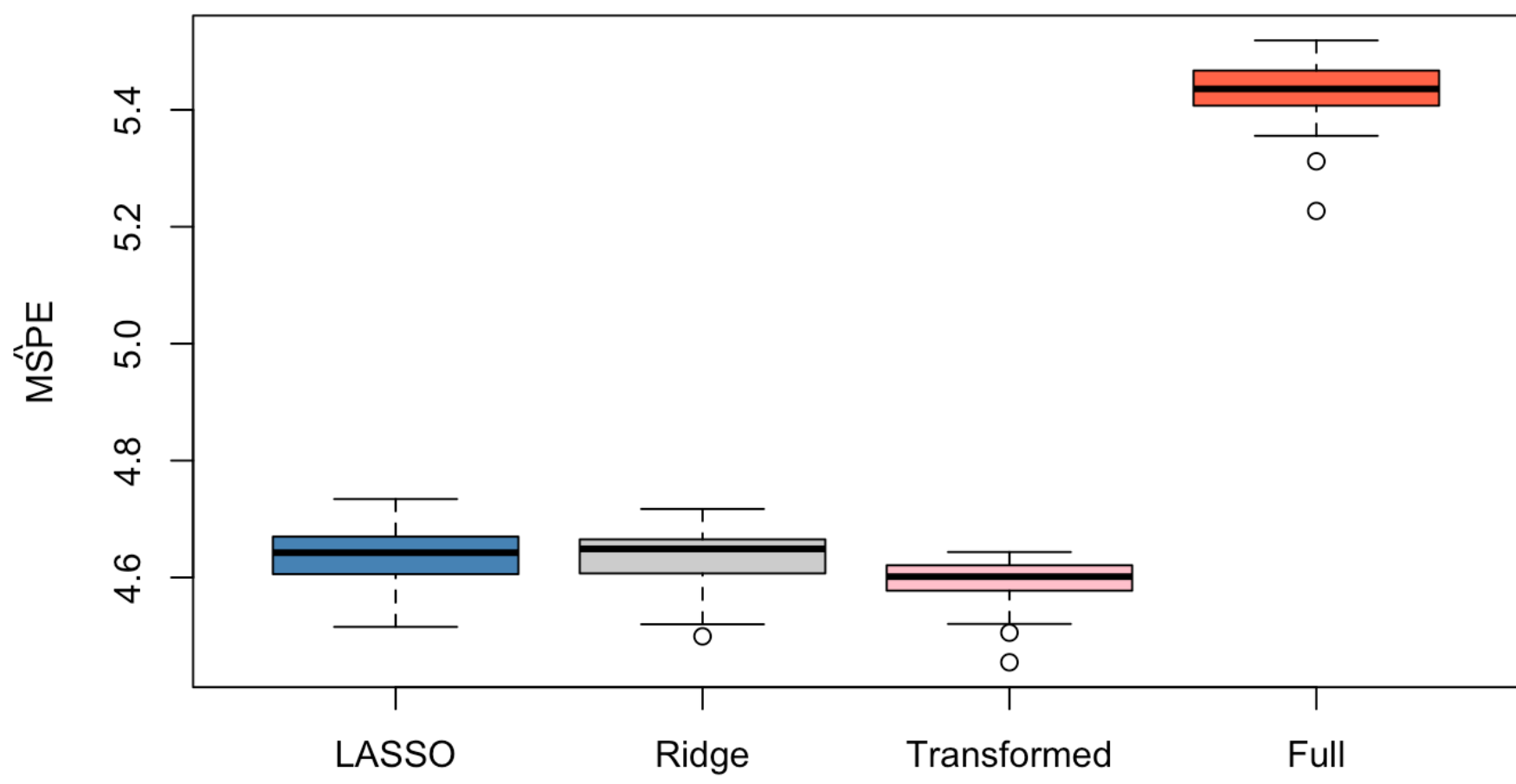


Figure 5: Goodness of Fit

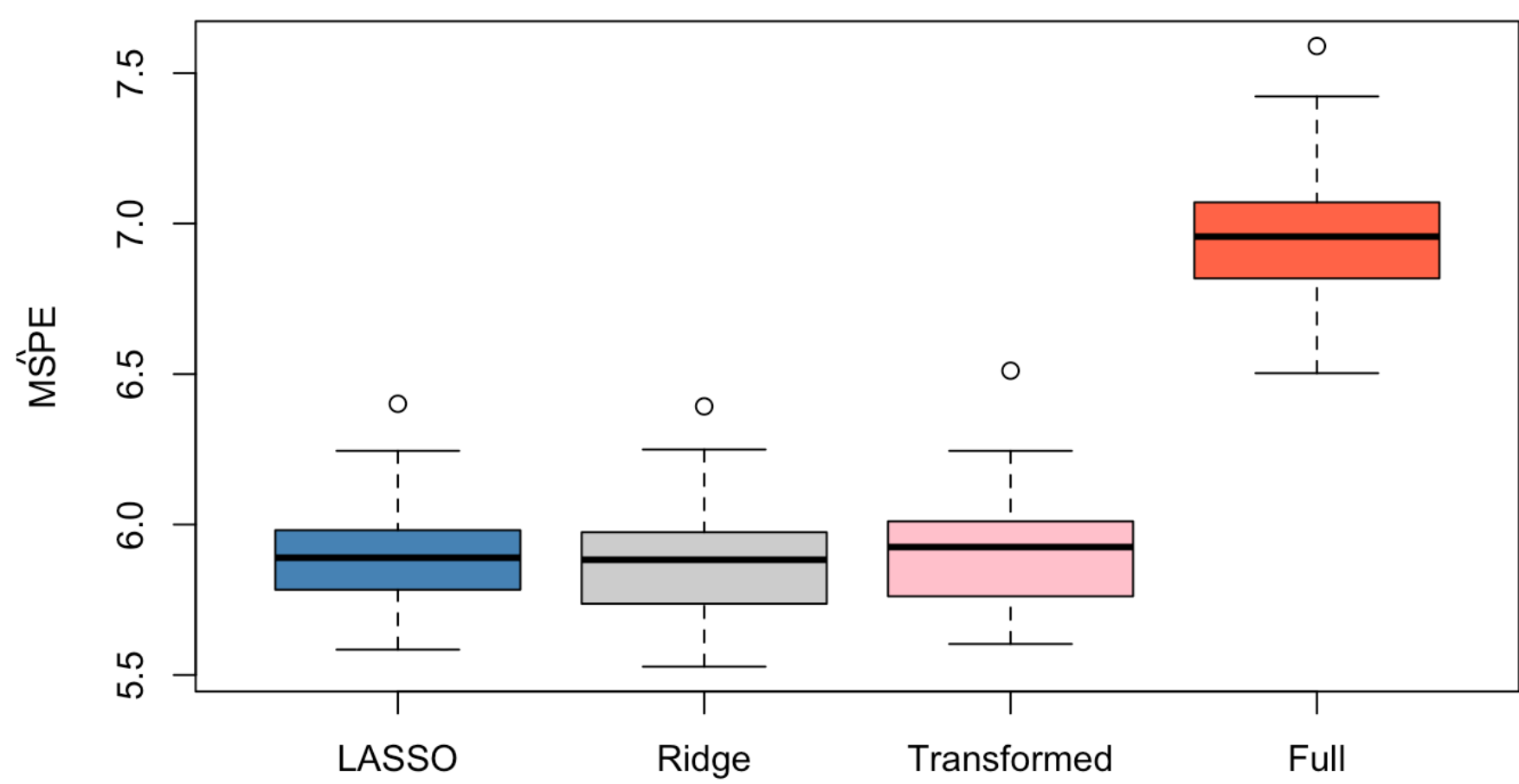


Figure 6: Prediction Power

## RESULTS

### Models:

- Ordinary Linear Regression, Transformed Linear Regression, LASSO and Ridge Regression were fitted using a 50-run 10-fold cross validation.

### Measures:

- **Figure 5** showed models shared similar performance; MSPEs across these models were around 4.6, but *OLR transformed* performed slightly better. *OLR full* performed the worst, around 5.4.
- **Figure 6** showed models shared similar performance; MSPEs across these models were close to 6. *OLR full* was the worst with MSPE around 6.9.

## CONCLUSION

- **EDA:** Average Assessment Total, Tax Class and Municipality were used to fit the transformed model.
  - After transforming Total Assessment to Average Total Assessment, the effect of outliers was reduced.
  - The correlation analyses had shown that the transformed assessment total is the only continuous variable that had a relatively strong correlation with the mill rate.
  - Kruskal-Wallis had suggested that mill rates in each municipality and tax class were significantly different.
- **Models:** Transformed OLR, Ridge and Lasso Regression were able to make good predictions based on MSPE on the training sets and test sets from cross-validation.
- **Final Conclusion:** Since the client preferred a simpler model, the transformed model was chosen to make the 2020 prediction.

## ACKNOWLEDGEMENTS

- Our Client, **Harry Xu**, for providing us this opportunity to apply statistical knowledge in real world setting.
- Our professor and supervisor, **Gabriela Cohen Freue**, **Sara Mostafavi**, **Gian Carlo Diluvi**, **Mallory Flynn**, **Sophia Li**, **Vittotio Romaniello**, and **Miguel Biron** for helping us throughout the course