

# 2021 ASA Data Challenge Expo

## Helping Communities During the COVID-19 Pandemic

Walter Yu, Code for America (Sacramento Brigade)

Primary Sponsor: Statistical Computing, ASA Section

Problem Statement: Help individuals, businesses, communities, or governments respond to the COVID-19 pandemic

Source Code Repository: [Github](#)

### INTRODUCTION

This project aims to help disadvantaged communities during the COVID-19 pandemic by answering the questions listed below through analysis of core and supplemental datasets.

### METHODOLOGY

Explore the relationship between socioeconomic features of the U.S. population and disadvantaged communities.

Identify disadvantaged communities based on their median household income. These communities are likely be more impacted by the COVID-19 pandemic and in need of public services.

### FINDINGS

Education attainment, commute mode and housing cost had the best relationship with median household income. Disadvantaged communities would most benefit from assistance with access to education, transit and affordable housing.

### RESULTS

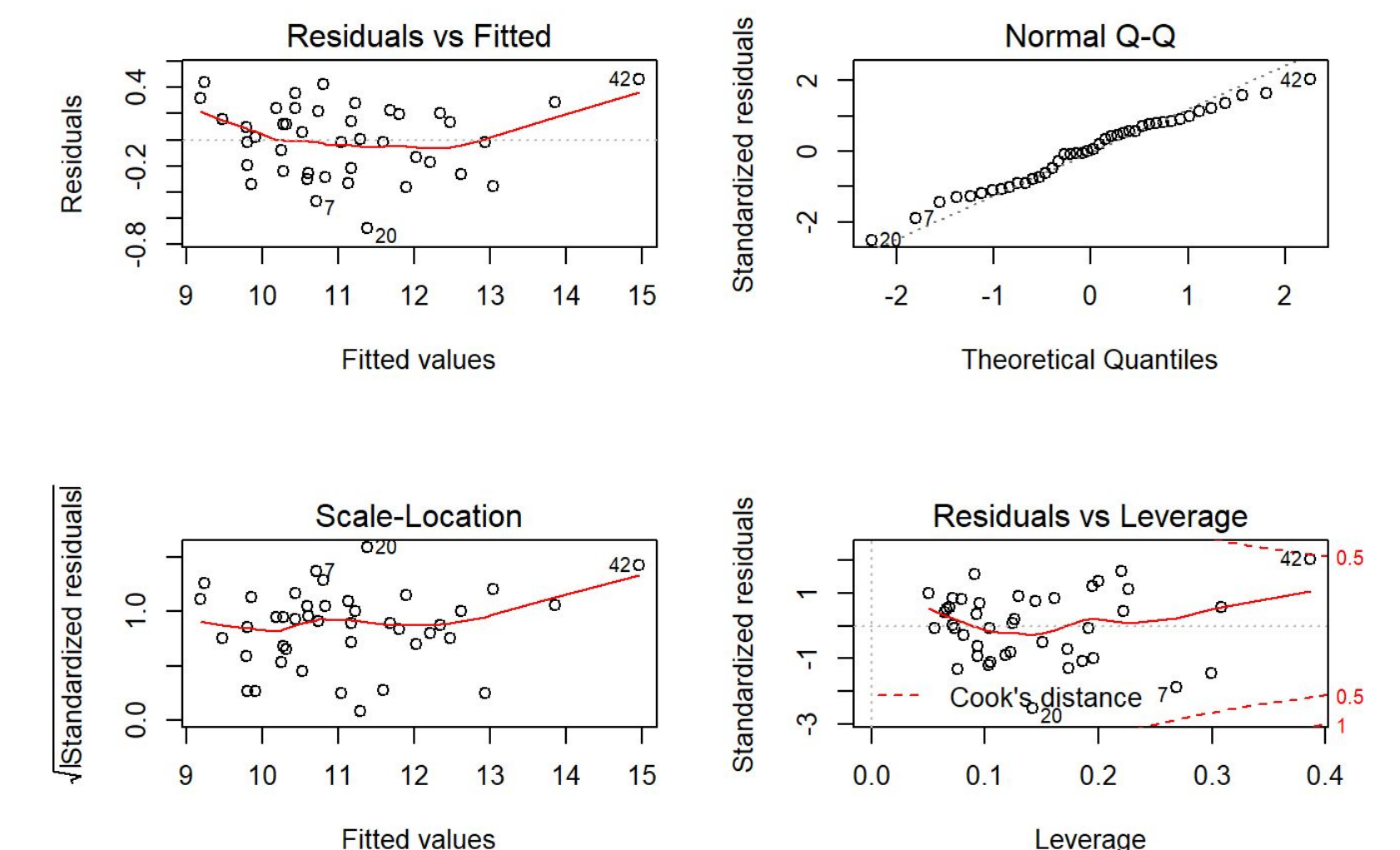
#### FINAL LINEAR MODEL SUMMARY AND DIAGNOSTIC PLOT

Coefficients:

|                                    | Estimate | Std. Error | t value | Pr(> t )     |
|------------------------------------|----------|------------|---------|--------------|
| (Intercept)                        | 1.03065  | 0.83491    | 1.234   | 0.2250       |
| log(as.numeric(housing_more_3000)) | -0.54663 | 0.09685    | -5.644  | 2.08e-06 *** |
| log(as.numeric(commute_car))       | -0.12776 | 0.06731    | -1.898  | 0.0657 .     |
| log(as.numeric(commute_walk))      | -0.12168 | 0.06091    | -1.998  | 0.0534 .     |
| log(as.numeric(edu_doctorate))     | -0.10887 | 0.10591    | -1.028  | 0.3108       |
| log(as.numeric(edu_bachelor))      | 1.52892  | 0.12788    | 11.956  | 4.28e-14 *** |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2918 on 36 degrees of freedom  
Multiple R-squared: 0.9534, Adjusted R-squared: 0.947  
F-statistic: 147.4 on 5 and 36 DF, p-value: < 2.2e-16



### TOOLS



### CONCLUSION

The linear regression models generally had good fits, and combining data features into the same model generated a better fit than the individual models.

**References & Acknowledgements:**  
**Poster Template:** [TRI Australia](#)  
**Brigade Project:** [Code for Sacramento](#)

