Fairness Analysis Figures and Tables

Ariane Stark and Dasuni Jayawardena

Data Editing

Important Values

#Wilcoxon rank sum between proportion of black constituent quartiles

```
all_the_data_pull_cutoff <- all_the_data_pull_cutoff %>%
  mutate(black_prop_quart=
           cut_number(all_the_data_pull_cutoff$Total_Black/all_the_data_pull_cutoff$Total_Pop,
             n=4, labels=FALSE))

quart_1_mae <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart==1) %>%
  pull(`1 week ahead MAE`)
quart_2_mae <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart==2) %>%
  pull(`1 week ahead MAE`)
quart_3_mae <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart==3) %>%
  pull(`1 week ahead MAE`)
quart_4_mae <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart==4) %>%
  pull(`1 week ahead MAE`)

# quart 1 and 2
wilcox.test(x=quart_1_mae,y=quart_2_mae)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  quart_1_mae and quart_2_mae
## W = 178316, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
#quart 1 and 3
wilcox.test(x=quart_1_mae,y=quart_3_mae)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  quart_1_mae and quart_3_mae
## W = 102559, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```r
#quart 1 and 4
wilcox.test(x=quart_1_mae,y=quart_4_mae)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  quart_1_mae and quart_4_mae
## W = 126617, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```r
#quart 3 and 4
wilcox.test(x=quart_3_mae,y=quart_4_mae)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  quart_3_mae and quart_4_mae
## W = 326137, p-value = 0.01052
## alternative hypothesis: true location shift is not equal to 0
```

Within Covid Case Quartile 1 Compare Across Race Quartiles

```r
case_quart_1_race_quart_1 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 1) %>%
              filter(black_prop_quart == 1) %>%
              pull(`1 week ahead MAE`)
case_quart_1_race_quart_2 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 1) %>%
              filter(black_prop_quart == 2) %>%
              pull(`1 week ahead MAE`)
case_quart_1_race_quart_3 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 1) %>%
              filter(black_prop_quart == 3) %>%
              pull(`1 week ahead MAE`)
case_quart_1_race_quart_4 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 1) %>%
              filter(black_prop_quart == 4) %>%
              pull(`1 week ahead MAE`)

#race quartile 1 and 2
wilcox.test(x=case_quart_1_race_quart_1,y=case_quart_1_race_quart_2)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_1_race_quart_1 and case_quart_1_race_quart_2
## W = 31550, p-value = 0.06056
## alternative hypothesis: true location shift is not equal to 0
```

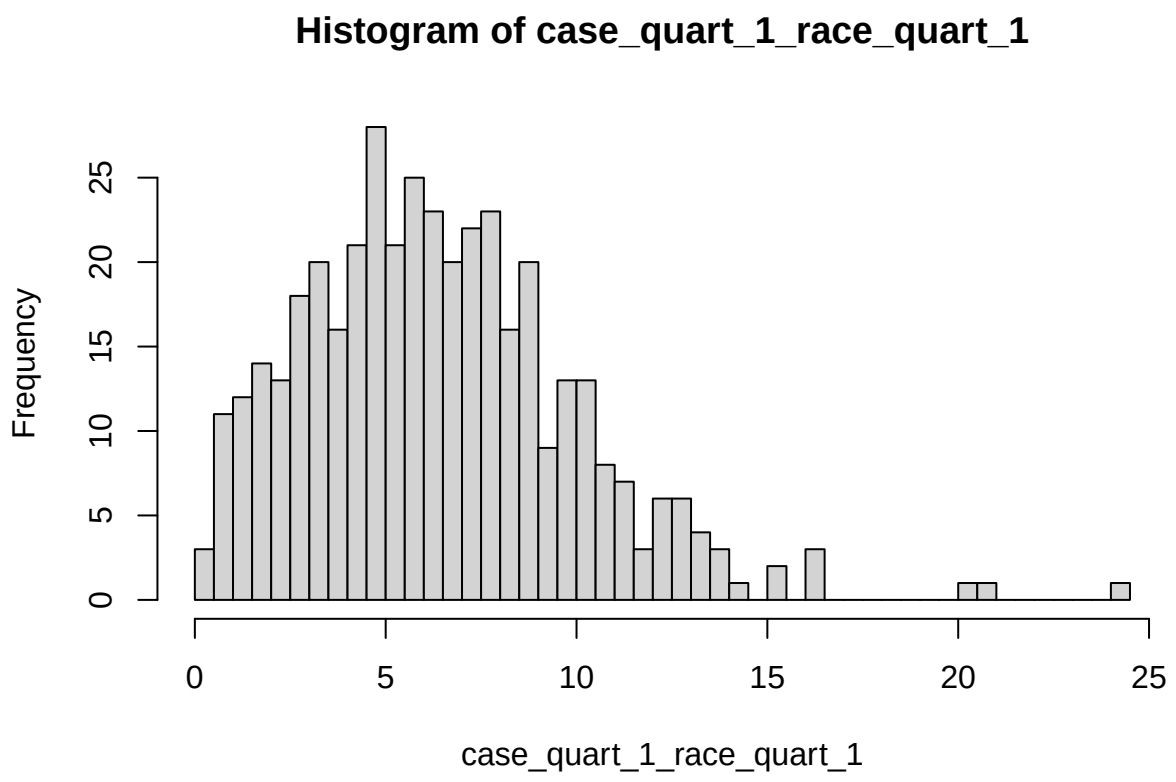
```r
#race quartile 1 and 3
wilcox.test(x=case_quart_1_race_quart_1,y=case_quart_1_race_quart_3)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_1_race_quart_1 and case_quart_1_race_quart_3
## W = 14316, p-value = 8.157e-05
## alternative hypothesis: true location shift is not equal to 0
```

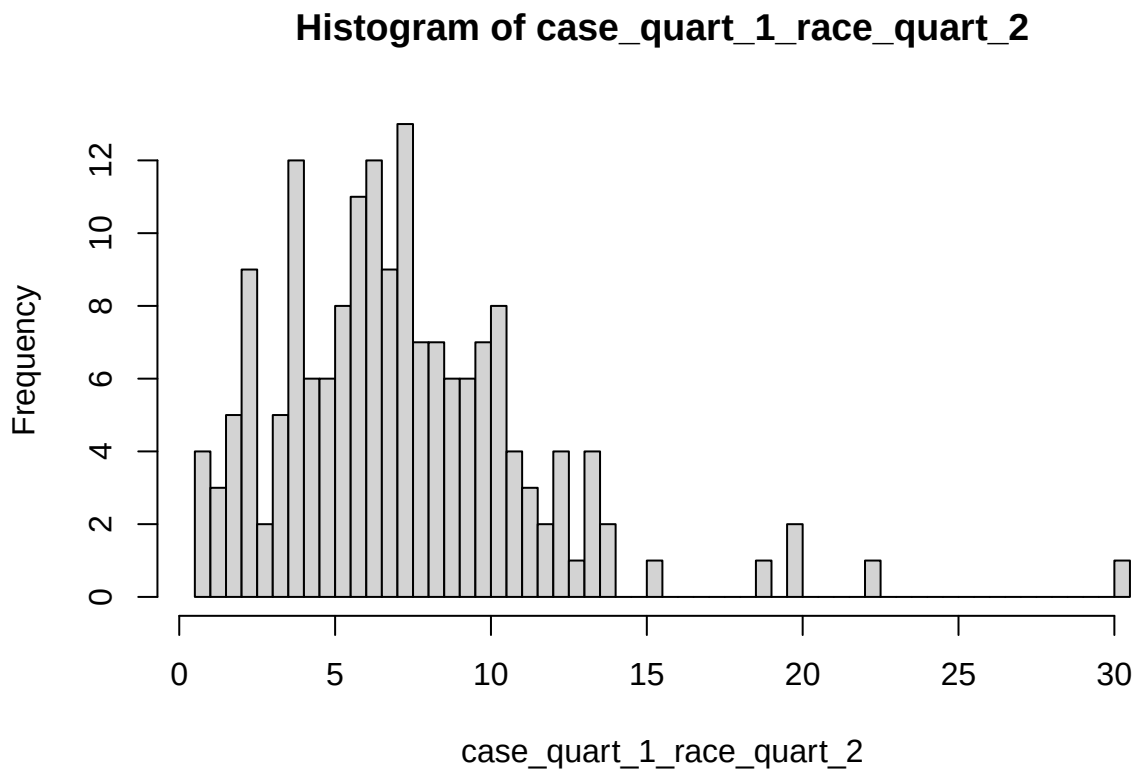
```r
#race quartile 1 and 4
wilcox.test(x=case_quart_1_race_quart_1,y=case_quart_1_race_quart_4)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_1_race_quart_1 and case_quart_1_race_quart_4
## W = 16146, p-value = 6.584e-05
## alternative hypothesis: true location shift is not equal to 0
```

```r
hist(case_quart_1_race_quart_1,breaks = 50)
```



```
hist(case_quart_1_race_quart_2,breaks = 50)
```



Within Covid Case Quartile 2 Compare Across Race Quartiles

```
case_quart_2_race_quart_1 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 2) %>%
              filter(black_prop_quart == 1) %>%
              pull(`1 week ahead MAE`)
case_quart_2_race_quart_2 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 2) %>%
              filter(black_prop_quart == 2) %>%
              pull(`1 week ahead MAE`)
case_quart_2_race_quart_3 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 2) %>%
              filter(black_prop_quart == 3) %>%
              pull(`1 week ahead MAE`)
case_quart_2_race_quart_4 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 2) %>%
              filter(black_prop_quart == 4) %>%
              pull(`1 week ahead MAE`)

#race quartile 1 and 2
wilcox.test(x=case_quart_2_race_quart_1,y=case_quart_2_race_quart_2)
```

```
##
```

```
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_2_race_quart_1 and case_quart_2_race_quart_2
## W = 21654, p-value = 0.03436
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 3
wilcox.test(x=case_quart_2_race_quart_1,y=case_quart_2_race_quart_3)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_2_race_quart_1 and case_quart_2_race_quart_3
## W = 9685, p-value = 3.063e-08
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 4
wilcox.test(x=case_quart_2_race_quart_1,y=case_quart_2_race_quart_4)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_2_race_quart_1 and case_quart_2_race_quart_4
## W = 17708, p-value = 0.003986
## alternative hypothesis: true location shift is not equal to 0
```

Within Covid Case Quartile 3 Compare Across Race Quartiles

```
case_quart_3_race_quart_1 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 3) %>%
              filter(black_prop_quart == 1) %>%
              pull(`1 week ahead MAE`)
case_quart_3_race_quart_2 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 3) %>%
              filter(black_prop_quart == 2) %>%
              pull(`1 week ahead MAE`)
case_quart_3_race_quart_3 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 3) %>%
              filter(black_prop_quart == 3) %>%
              pull(`1 week ahead MAE`)
case_quart_3_race_quart_4 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 3) %>%
              filter(black_prop_quart == 4) %>%
              pull(`1 week ahead MAE`)

#race quartile 1 and 2
wilcox.test(x=case_quart_3_race_quart_1,y=case_quart_3_race_quart_2)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_3_race_quart_1 and case_quart_3_race_quart_2
## W = 12882, p-value = 0.01097
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 3
wilcox.test(x=case_quart_3_race_quart_1,y=case_quart_3_race_quart_3)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_3_race_quart_1 and case_quart_3_race_quart_3
## W = 11604, p-value = 0.002071
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 4
wilcox.test(x=case_quart_3_race_quart_1,y=case_quart_3_race_quart_4)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_3_race_quart_1 and case_quart_3_race_quart_4
## W = 10542, p-value = 0.05698
## alternative hypothesis: true location shift is not equal to 0
```

Within Covid Case Quartile 4 Compare Across Race Quartiles

```
case_quart_4_race_quart_1 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 4) %>%
              filter(black_prop_quart == 1) %>%
              pull(`1 week ahead MAE`)
case_quart_4_race_quart_2 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 4) %>%
              filter(black_prop_quart == 2) %>%
              pull(`1 week ahead MAE`)
case_quart_4_race_quart_3 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 4) %>%
              filter(black_prop_quart == 3) %>%
              pull(`1 week ahead MAE`)
case_quart_4_race_quart_4 <- all_the_data_pull_cutoff %>%
              filter(`Case Quartile` == 4) %>%
              filter(black_prop_quart == 4) %>%
              pull(`1 week ahead MAE`)

#race quartile 1 and 2
wilcox.test(x=case_quart_4_race_quart_1,y=case_quart_4_race_quart_2)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_4_race_quart_1 and case_quart_4_race_quart_2
## W = 2195.5, p-value = 0.6978
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 3
wilcox.test(x=case_quart_4_race_quart_1,y=case_quart_4_race_quart_3)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_4_race_quart_1 and case_quart_4_race_quart_3
## W = 4235, p-value = 0.2756
## alternative hypothesis: true location shift is not equal to 0
```

```
#race quartile 1 and 4
wilcox.test(x=case_quart_4_race_quart_1,y=case_quart_4_race_quart_4)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  case_quart_4_race_quart_1 and case_quart_4_race_quart_4
## W = 3159, p-value = 0.01546
## alternative hypothesis: true location shift is not equal to 0
```

Summary Statistics Case Quartile

```
case_quart_1_summary <- all_the_data_pull_cutoff %>%
  filter(`Case Quartile` == 1) %>%
  pull(`1 week ahead MAE`)

case_quart_4_summary <- all_the_data_pull_cutoff %>%
  filter(`Case Quartile` == 4) %>%
  pull(`1 week ahead MAE`)
```

Summary Statistics By Race Quartile

```
race_quart_1_summary <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart == 1) %>%
  pull(`1 week ahead MAE`) %>%
  summary() %>%
  round(digits = 3)

race_quart_1_summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.118   5.912  10.279  17.038  17.890 574.088
```

```
race_quart_4_summary <- all_the_data_pull_cutoff %>%
  filter(black_prop_quart == 4) %>%
  pull(`1 week ahead MAE`) %>%
  summary() %>%
  round(digits = 3)

race_quart_4_summary
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    1.882   14.721   29.500   86.246   81.926 2559.765
```

Median MAE and RMAE tables

```
##              Prop. Black Quartile
## Case Quartile lowest 25%     Q2     Q3 highest 25% Overall
##   lowest 25%       6.059  6.559  7.794       7.632   6.647
##   Q2              14.103 14.882 17.706      15.147  14.912
##   Q3              25.853 29.088 29.603      28.912  28.515
##   highest 25%     79.059 73.368 91.324     111.779  93.353
##   Overall         10.279 19.588 35.618      29.500  20.735
```

MAE table generated by LaTex

	Prop. Black Quartile				
Case Quartile	lowest 25%	Q2	Q3	highest 25%	Overall
lowest 25%	6.059	6.559	7.794	7.632	6.647
Q2	14.103	14.882	17.706	15.147	14.912
Q3	25.853	29.088	29.603	28.912	28.515
highest 25%	79.059	73.368	91.324	111.779	93.353
Overall	10.279	19.588	35.618	29.500	20.735

Table 1: Pink for statistically significant at $\alpha = 0.05$ and red for statistically significant at $\alpha = 0.00333$ (from the Bonferroni Correction over 15 tests) when comparing to Quartile 1 of Prop. Black. across the case quartiles and overall

```
##              Prop. Black Quartile
## Case Quartile lowest 25%    Q2    Q3 highest 25% Overall
##   lowest 25%       0.898 0.901 0.910       0.885   0.898
##   Q2               0.906 0.887 0.902       0.887   0.897
##   Q3               0.911 0.899 0.905       0.879   0.898
##   highest 25%      0.923 0.890 0.889       0.894   0.891
##   Overall          0.903 0.896 0.897       0.887   0.896
```

RMAE Table from LaTex

	Prop. Black Quartile				
Case Quartile	lowest 25%	Q2	Q3	highest 25%	Overall
lowest 25%	0.898	0.901	0.910	0.885	0.898
Q2	0.906	0.887	0.902	0.887	0.897
Q3	0.911	0.899	0.905	0.879	0.898
highest 25%	0.923	0.890	0.889	0.894	0.891
Overall	0.903	0.896	0.897	0.887	0.896

Table 2: text

Figures

Figure 1

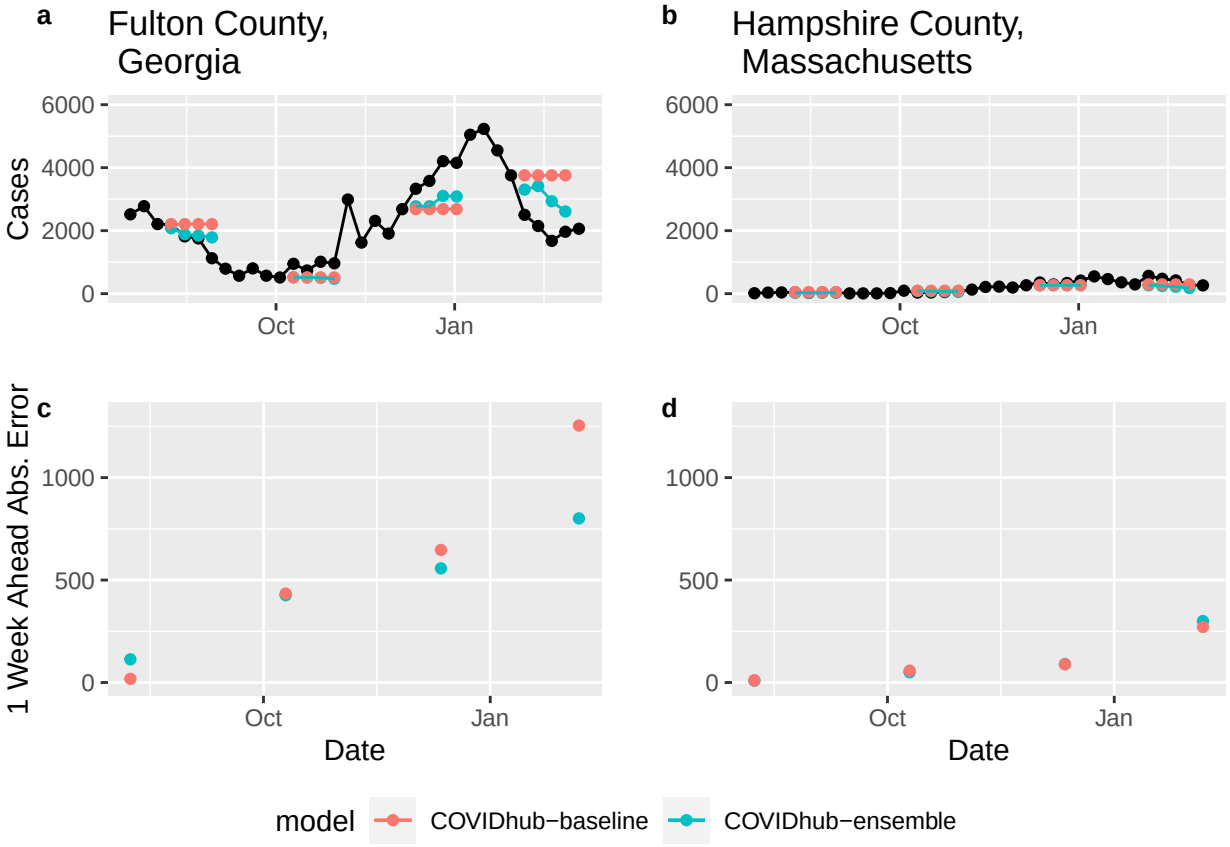

a
Fulton County,
Georgia
b
Hampshire County,
Massachusetts
Cases
6000
4000
2000
0
Oct
Jan
1 Week Ahead Abs. Error
c
d
1000
500
0
Date
model
COVIDhub−baseline
COVIDhub−ensemble

Figure 2

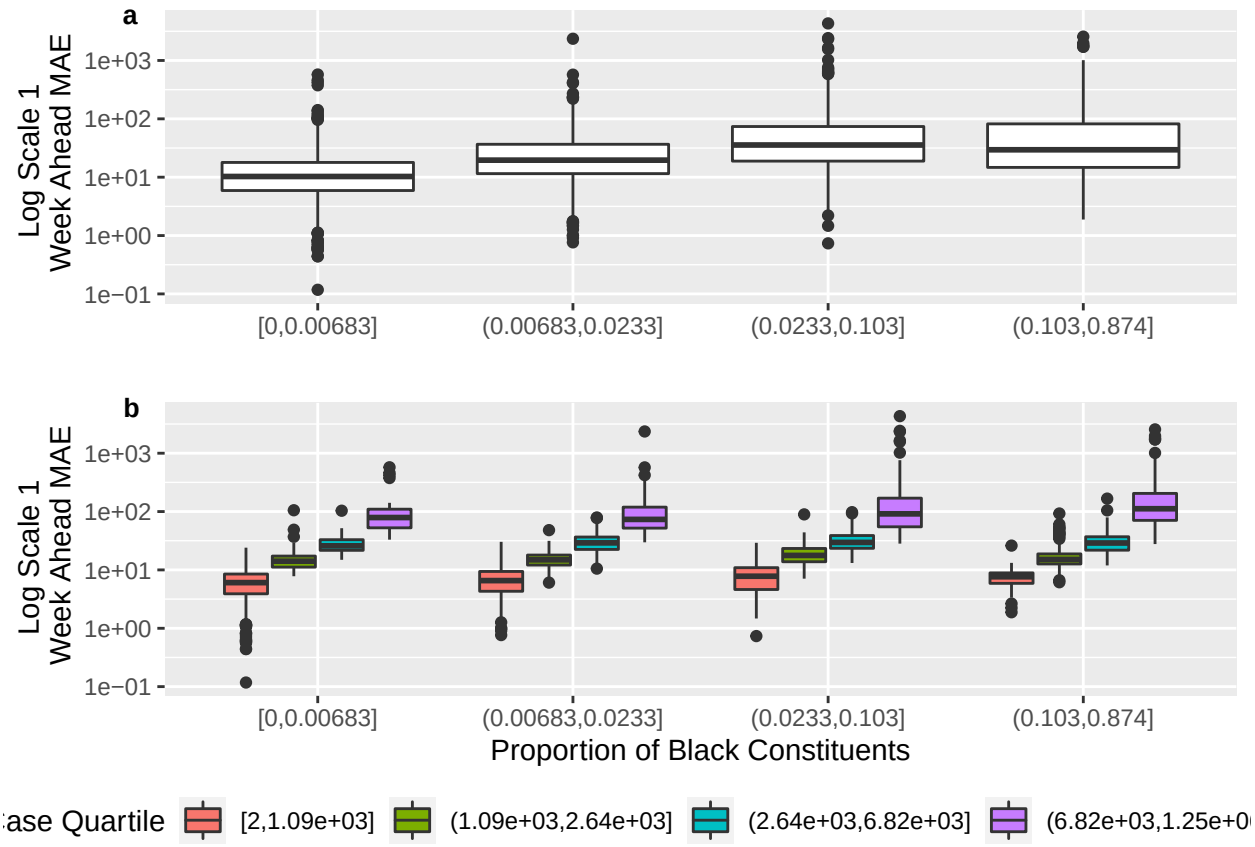

a
b
Log Scale 1
Week Ahead MAE
1e+03
1e+02
1e+01
1e+00
1e−01
[0,0.00683]
(0.00683,0.0233]
(0.0233,0.103]
(0.103,0.874]
Proportion of Black Constituents
ase Quartile
[2,1.09e+03]
(1.09e+03,2.64e+03]
(2.64e+03,6.82e+03]
(6.82e+03,1.25e+0

Figure 3: 1 and 4 week MAE and RMAE for Racial Demographics

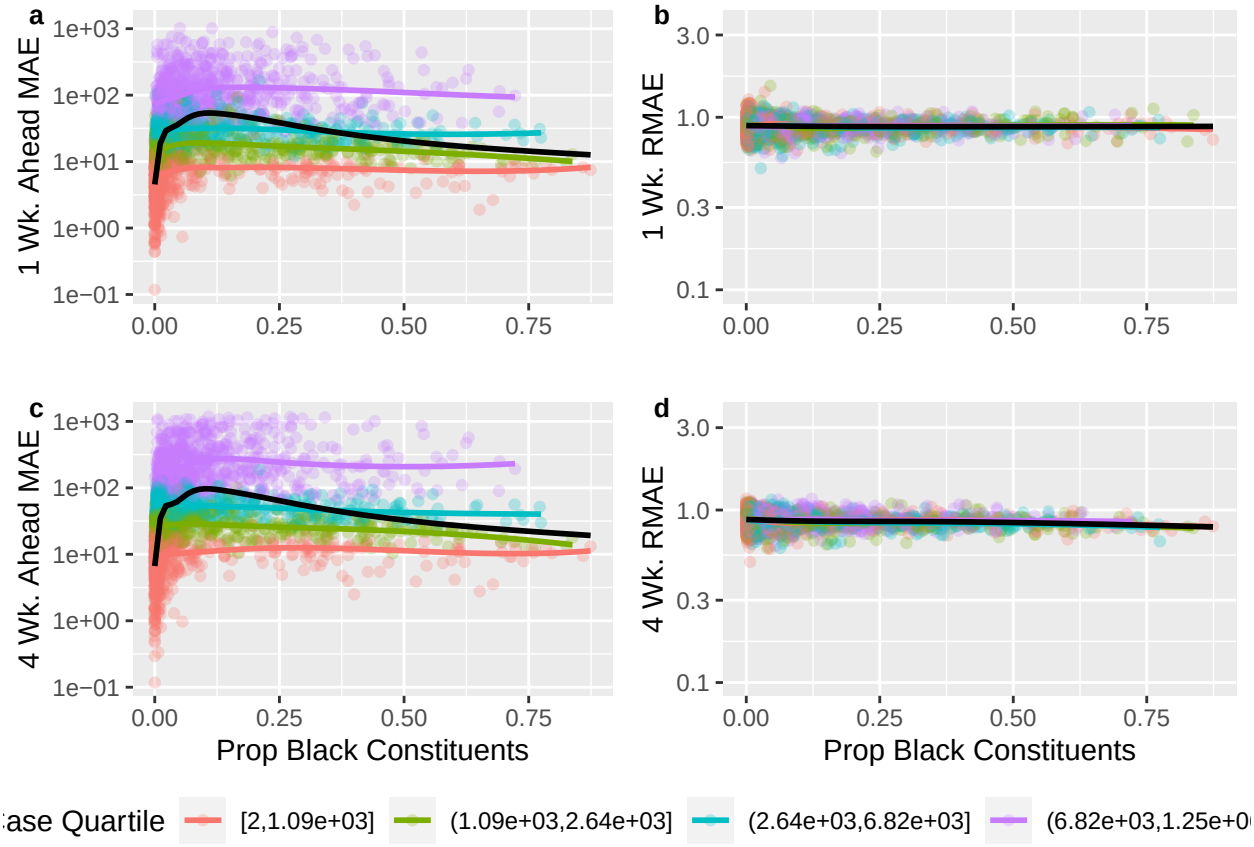


Figure 3 option 2: 1 and 4 week MAE and RMAE for Racial Demographics

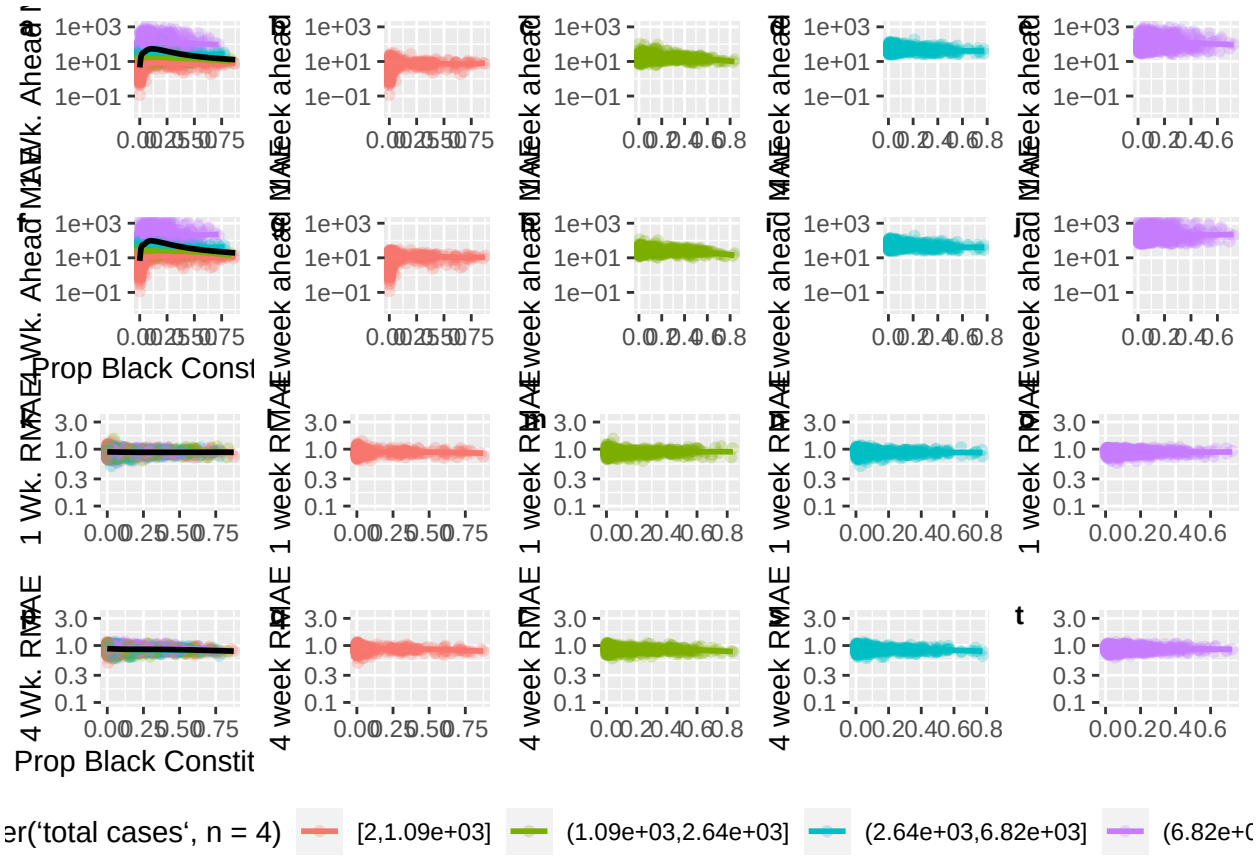
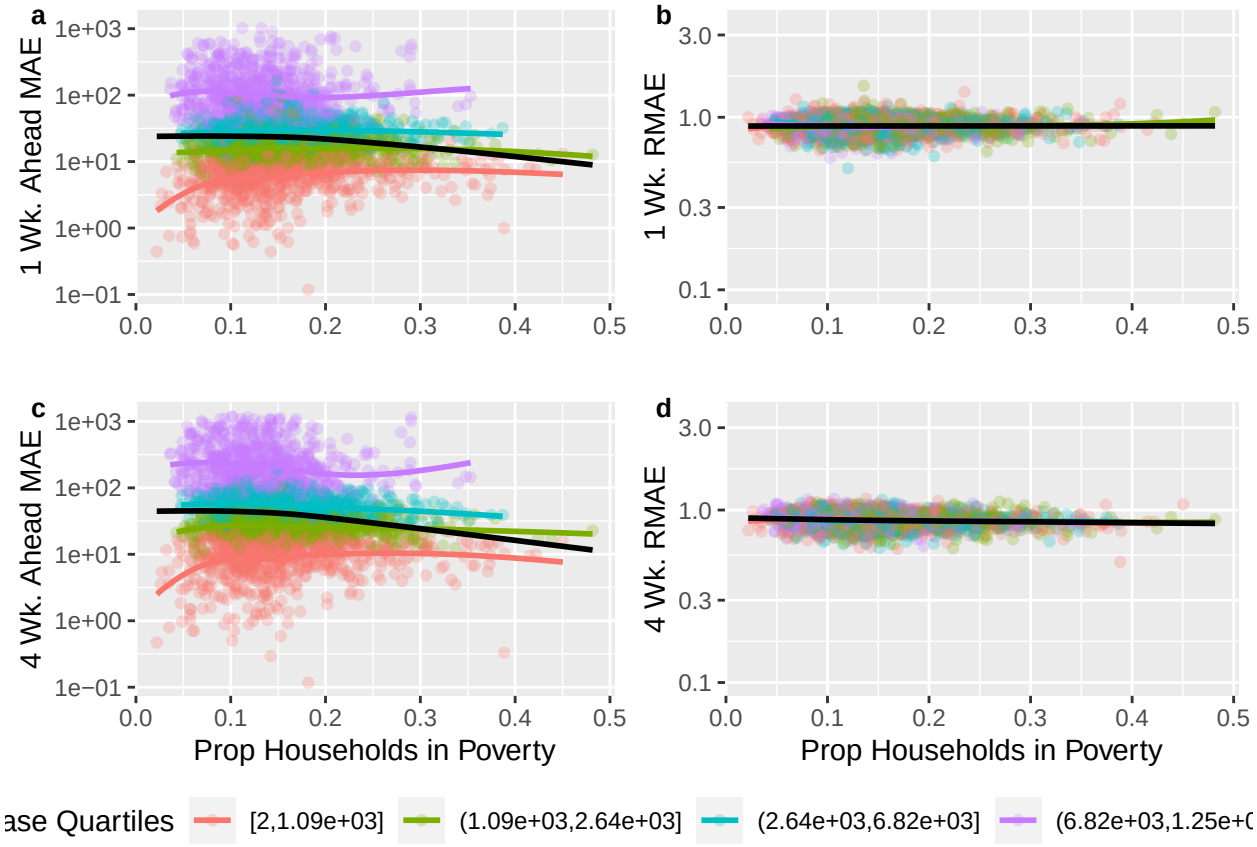


Figure 4: 1 and 4 week MAE and RMAE for Poverty



Tables

Table 1

	R squared
1 wk Ahead MAE as a funct. of Prop. Black and Prop. Pov	0.0309
1 wk Ahead MAE as a funct. of Total Population	0.8712
1 wk Ahead MAE as a funct. of Total Cases	0.8862
1 wk Ahead MAE as a funct. of Cases and Minority Prop.	0.8863
4 wk Ahead MAE as a funct. of Prop. Black and Prop. Pov	0.0239
4 wk Ahead MAE as a funct. of Total Population	0.9108
4 wk Ahead MAE as a funct. of Total Cases	0.9543
4 wk Ahead MAE as a funct. of Cases and Minority Prop.	0.9554
## Regression Equations	

```
summary(lm(One_Week_MAE~Tot_Cases,data = regressionData))
```

```
##
## Call:
```

```
## lm(formula = One_Week_MAE ~ Tot_Cases, data = regressionData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -752.72   -9.19   -5.67    1.31 1238.92
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.330e+01  1.015e+00   13.11   <2e-16 ***
## Tot_Cases   4.046e-03  2.598e-05  155.74   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.53 on 3115 degrees of freedom
## Multiple R-squared:  0.8862, Adjusted R-squared:  0.8862
## F-statistic: 2.425e+04 on 1 and 3115 DF,  p-value: < 2.2e-16
```