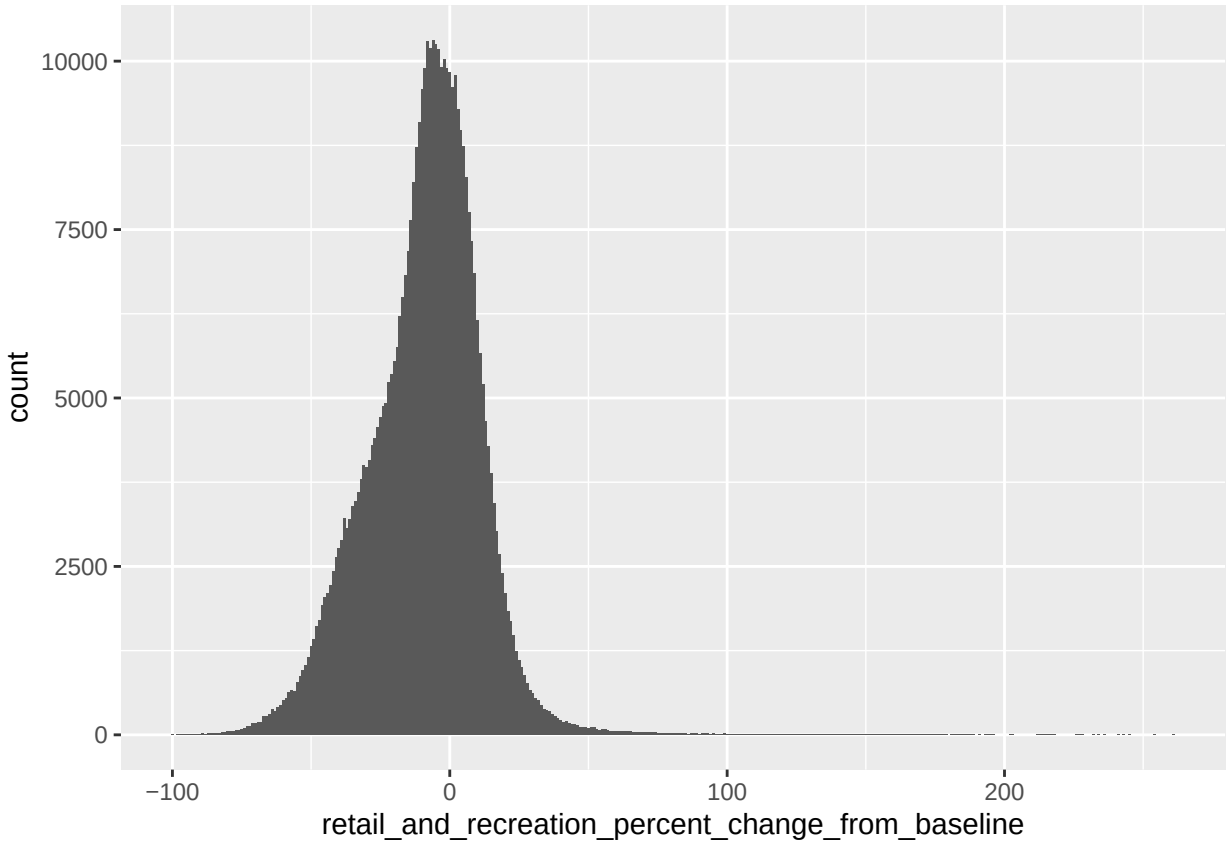
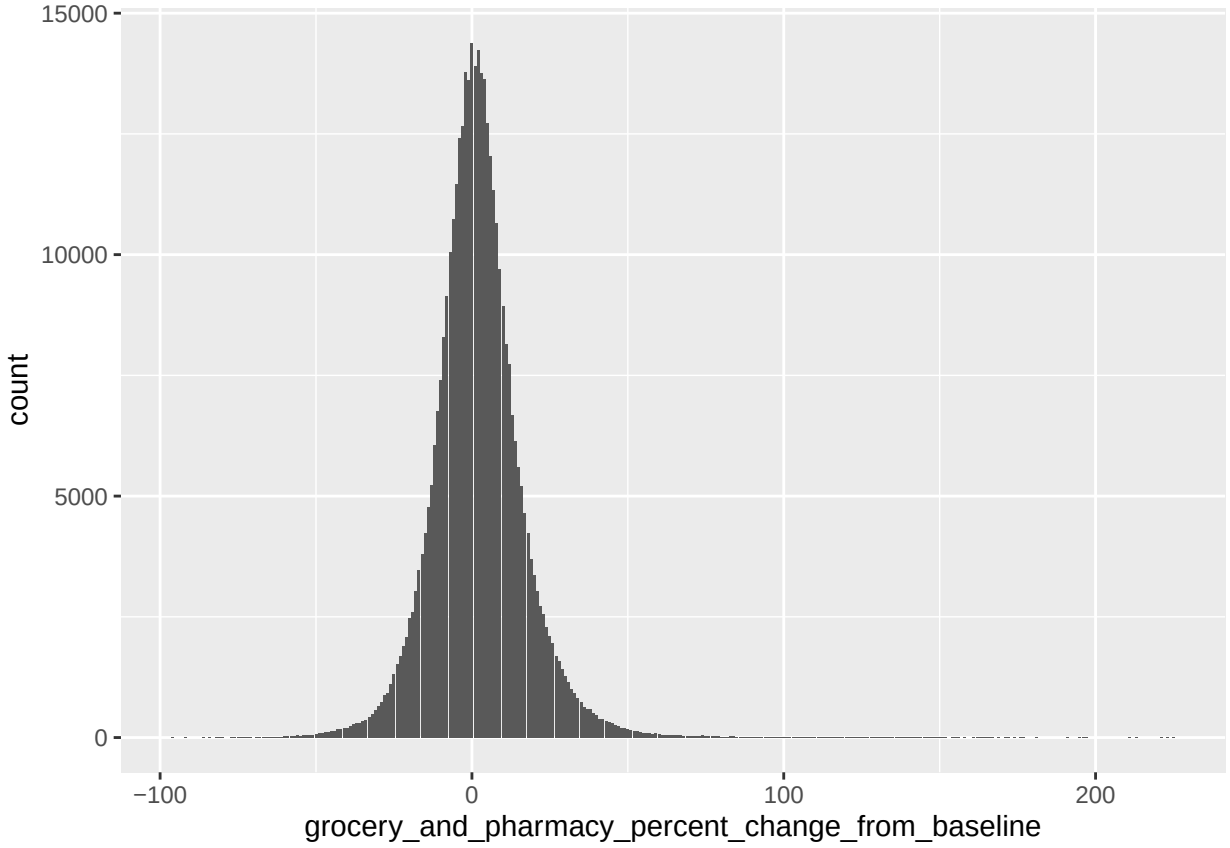
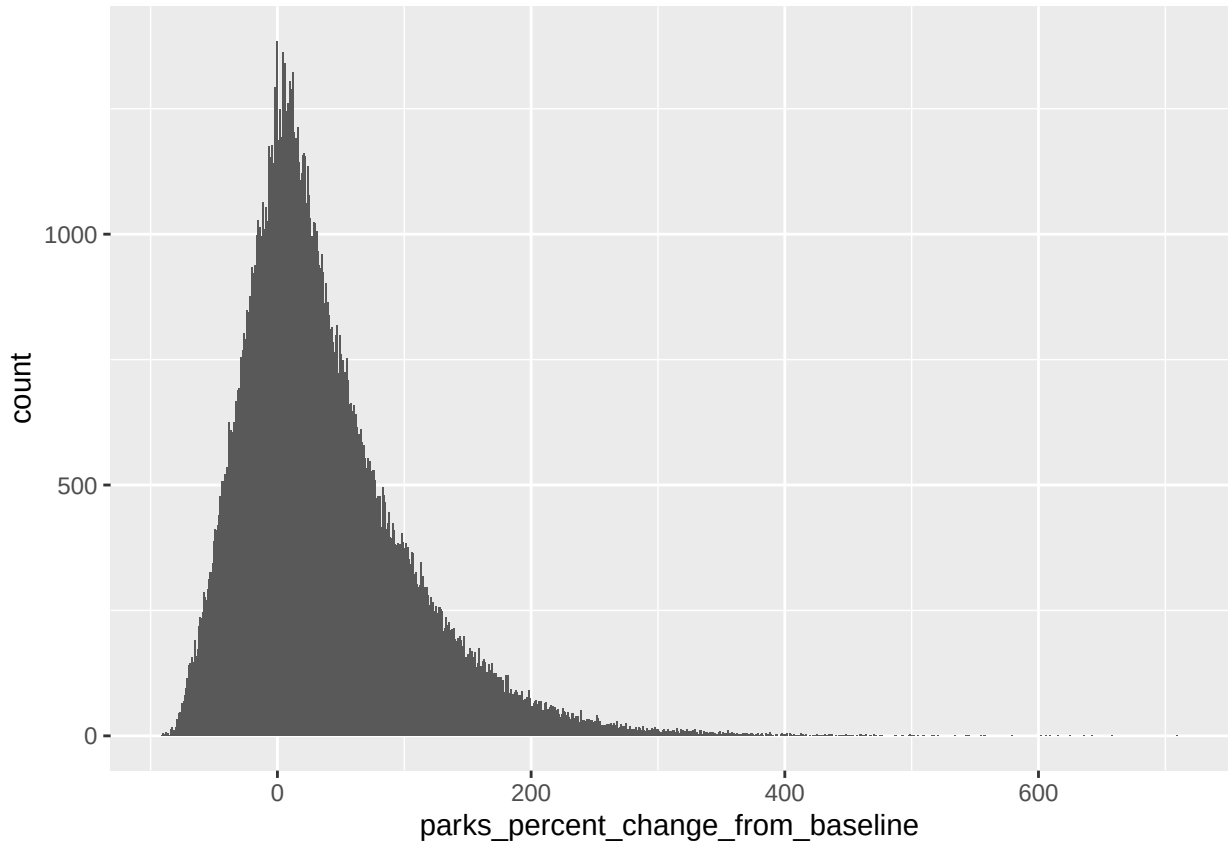
Google Movement Data Analysis

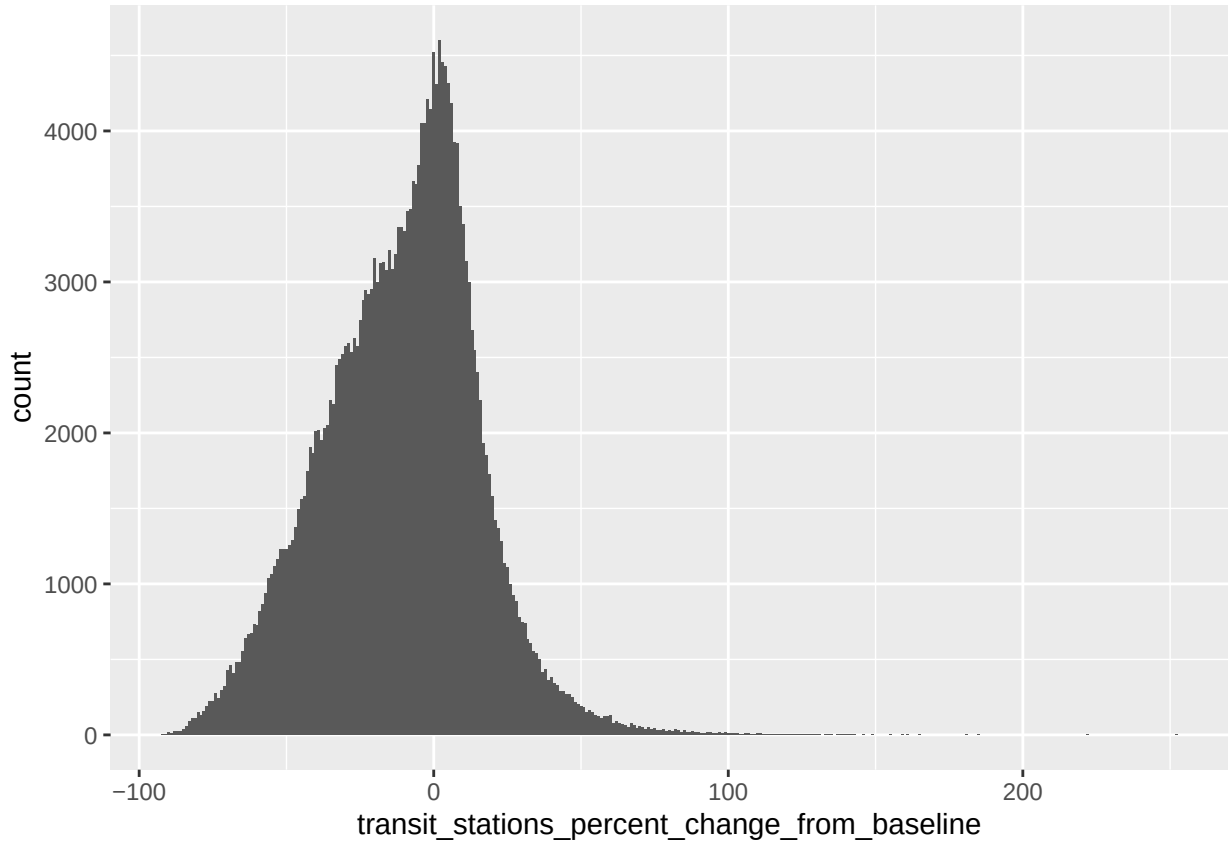
Albert Sun, Lily Zhu

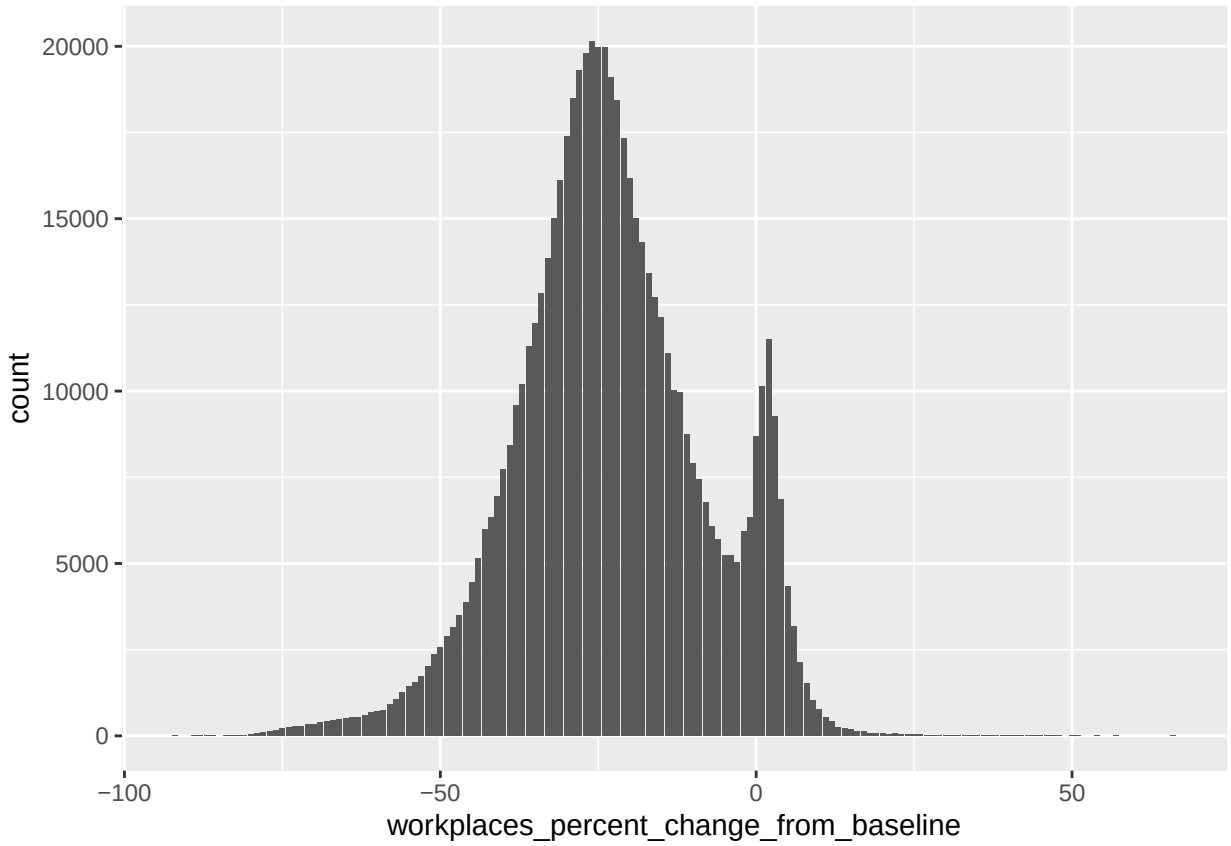
Distribution plots

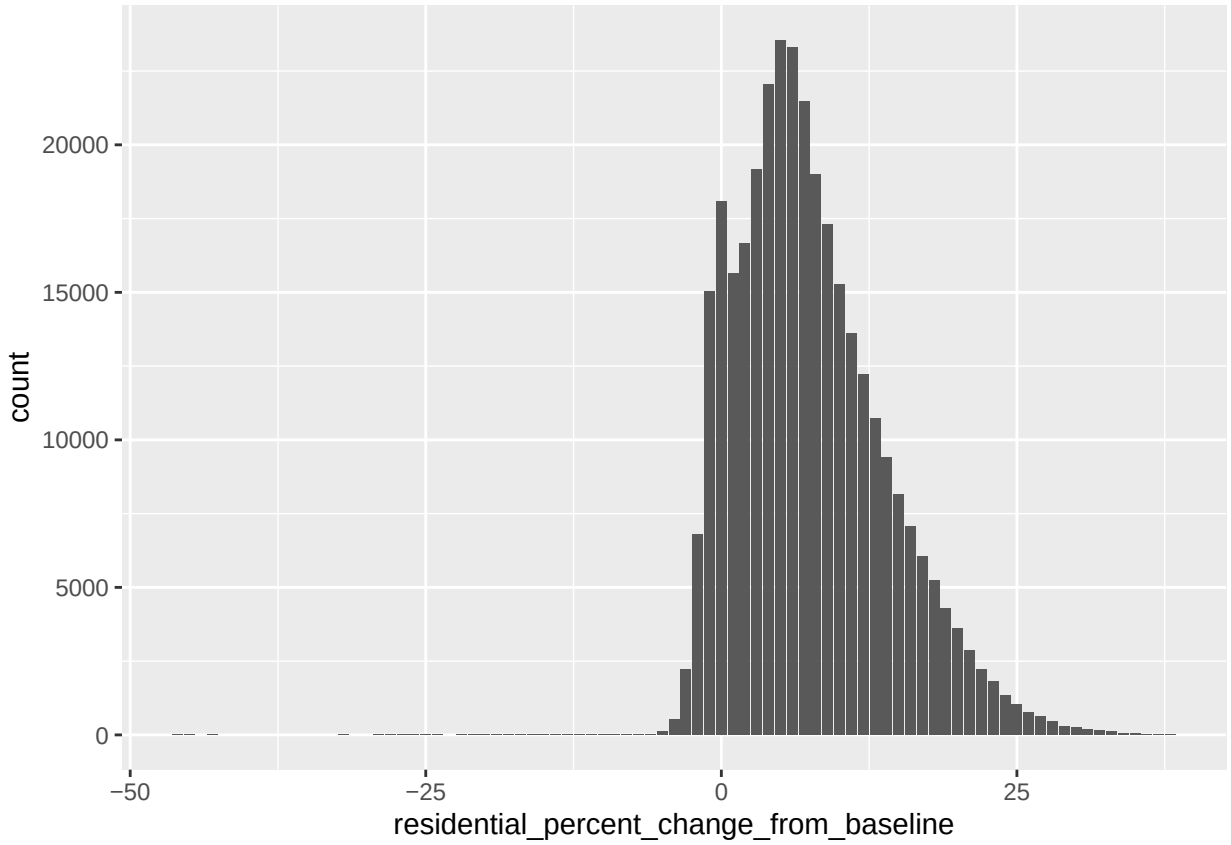





count
1000
500
0
0
200
400
600
parks_percent_change_from_baseline


count
4000
3000
2000
1000
0
−100
0
100
200
transit_stations_percent_change_from_baseline


20000
15000
10000
5000
0
count
−100
−50
0
50
workplaces_percent_change_from_baseline


20000
15000
10000
5000
0
count
−50
−25
0
25
residential_percent_change_from_baseline

Summary stats

```
##        mean median       sd iqr  min max
## 1 -9.100737     -7 20.51652  25 -100 261
```

```
##       mean median       sd iqr min max
## 1 2.047446      1 14.67781  15 -96 225
```

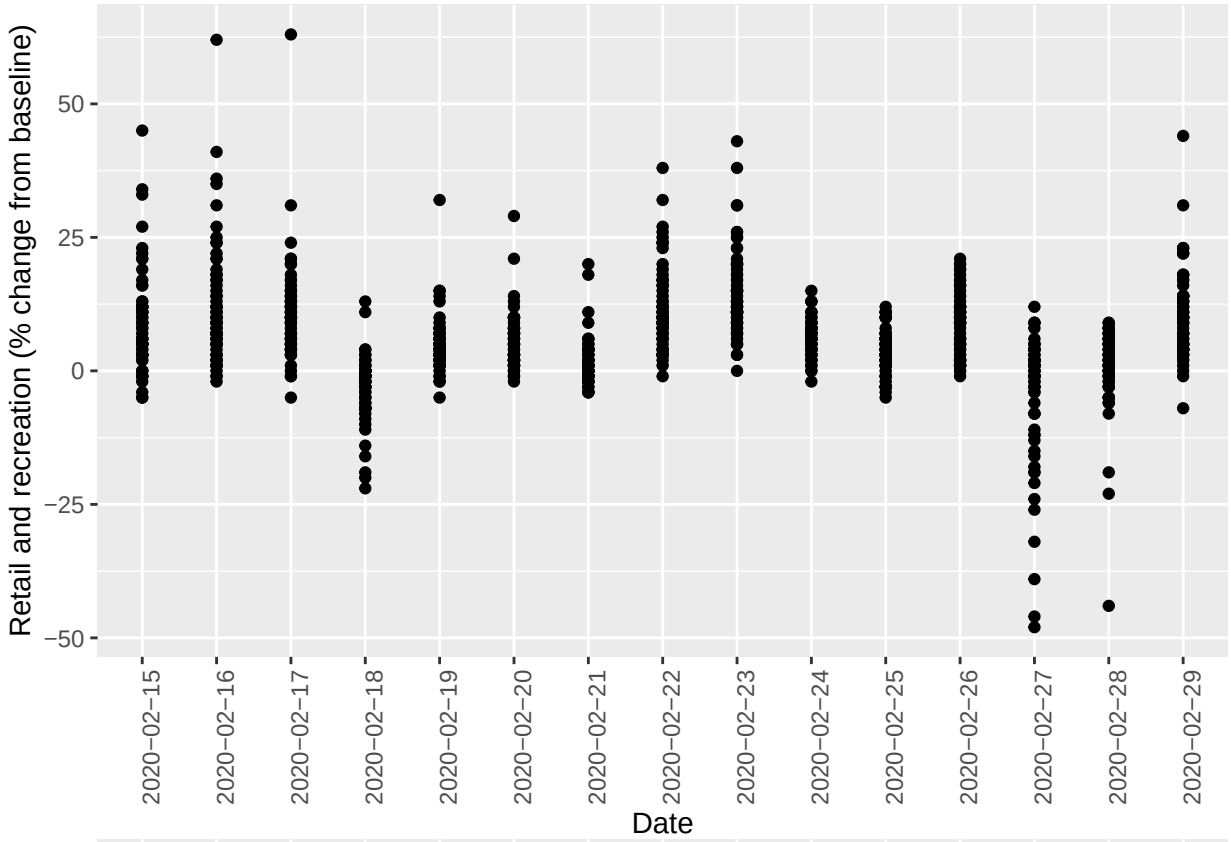
```
##       mean median       sd iqr min max
## 1 37.33111     23 66.49339  76 -91 709
```

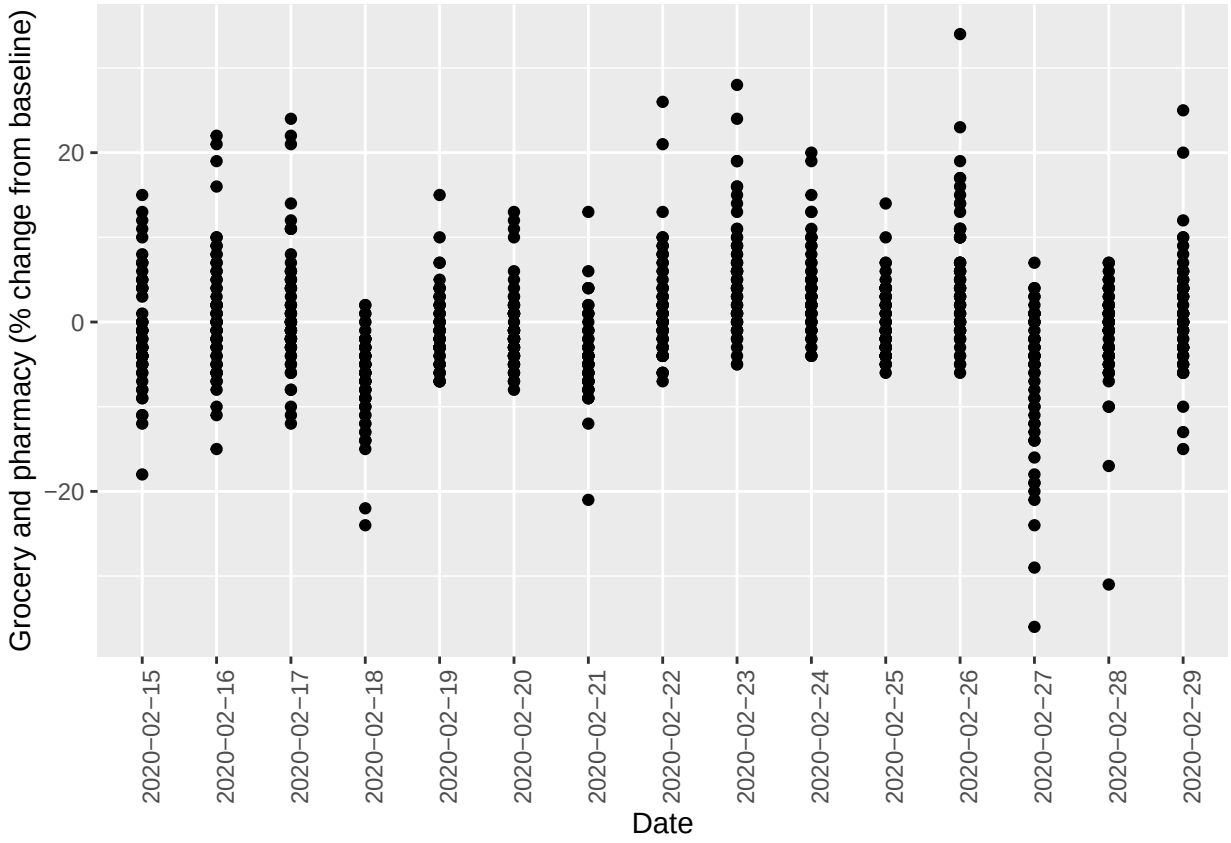
```
##        mean median       sd iqr min max
## 1 -10.91589     -9 26.18124  35 -92 252
```

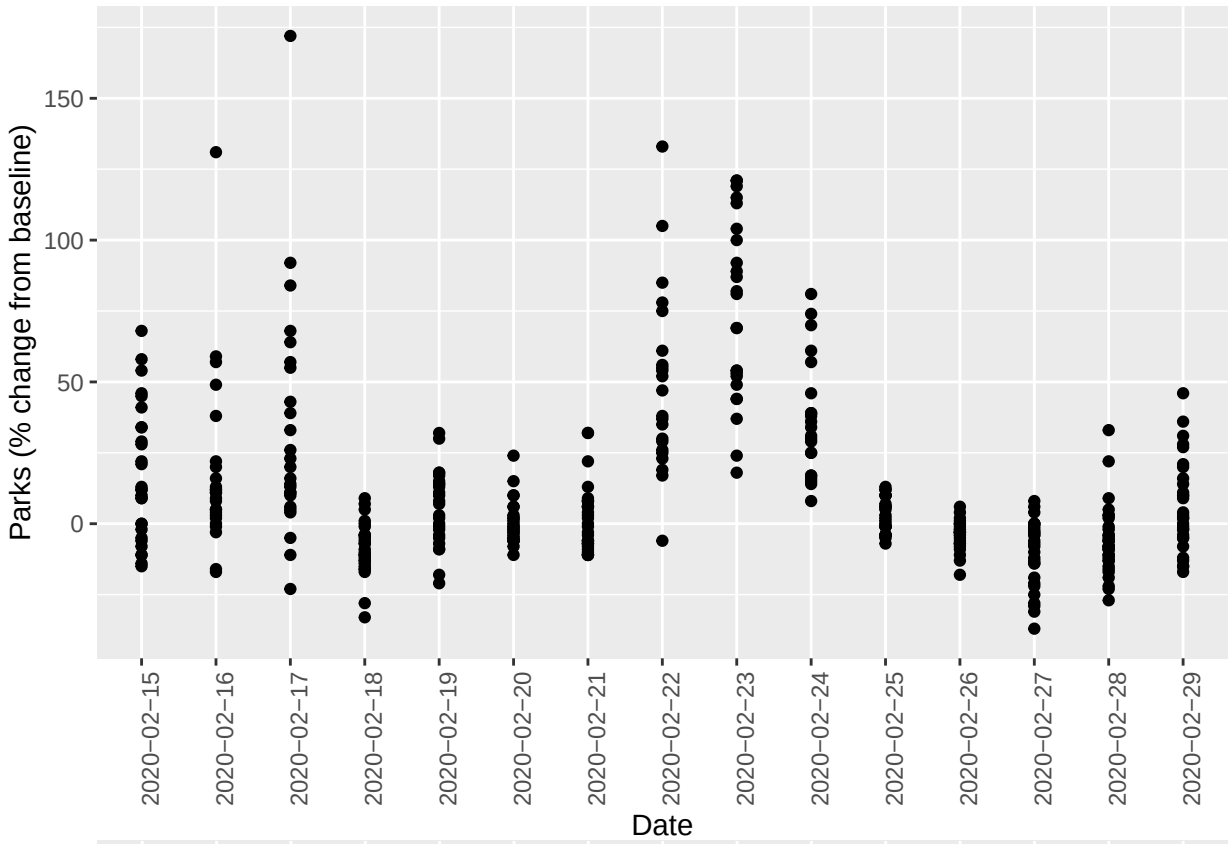
```
##       mean median       sd iqr min max
## 1 -23.0098    -24 14.96862  18 -92  66
```

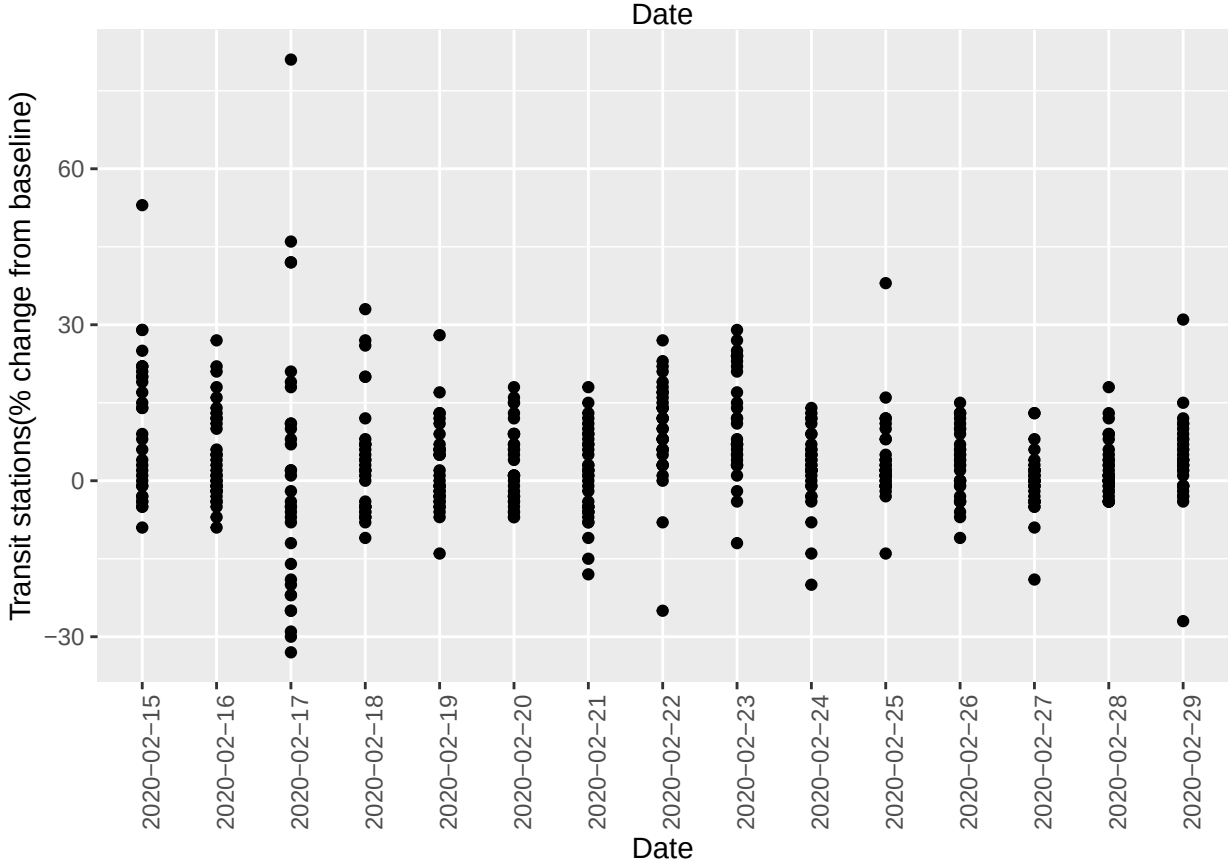
```
##       mean median       sd iqr min max
## 1 7.505763      7 6.300294   8 -46  38
```

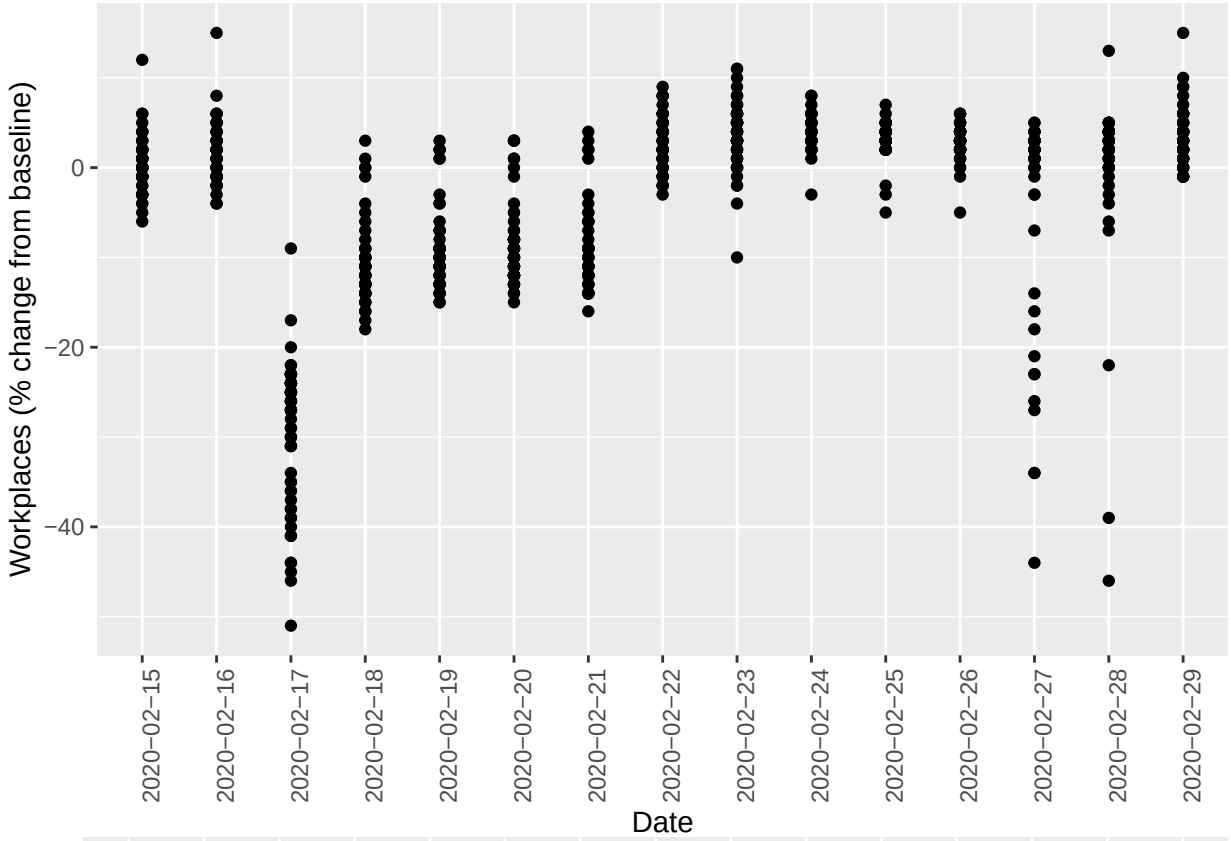
NY

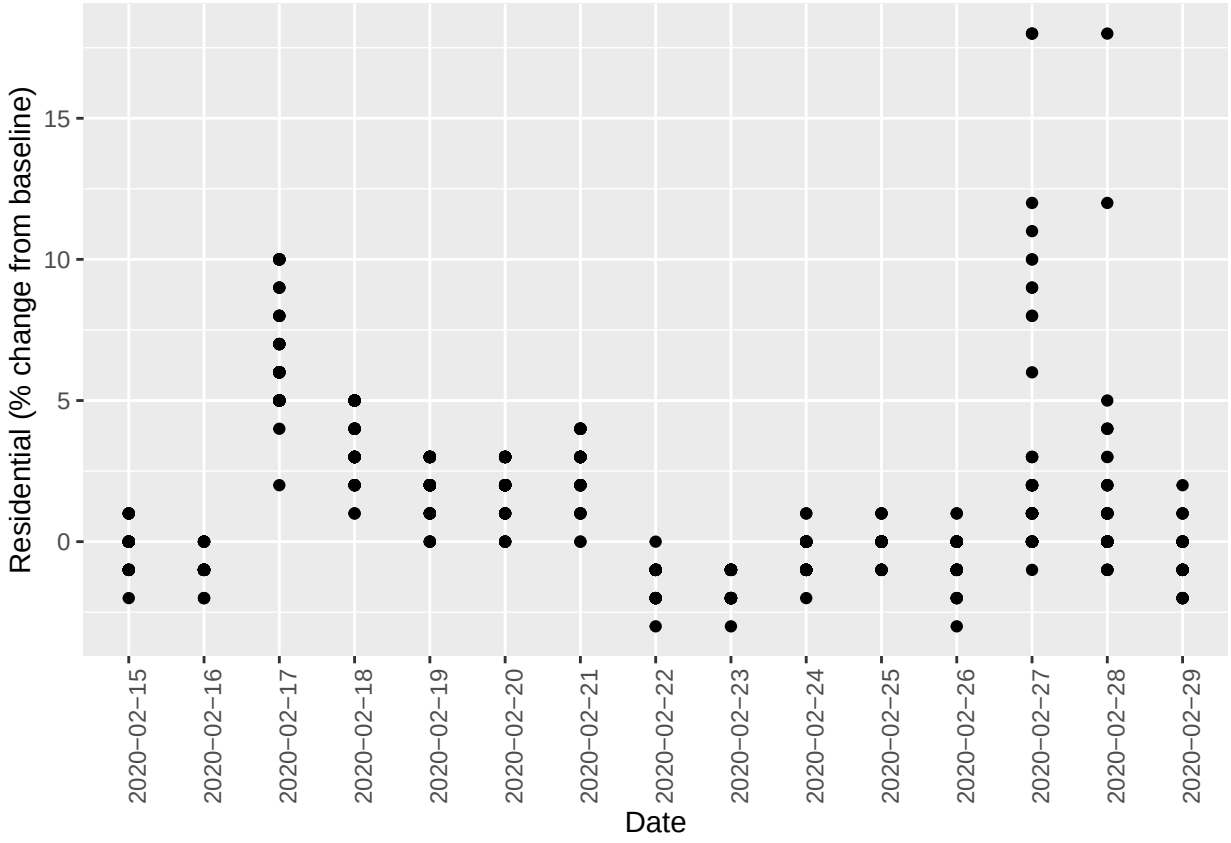

Retail and recreation (% change from baseline)
50
25
0
−25
−50
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date


Grocery and pharmacy (% change from baseline)
20
0
−20
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date


Parks (% change from baseline)
150
100
50
0
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date


Transit stations(% change from baseline)
60
30
0
−30
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date


Workplaces (% change from baseline)
0
−20
−40
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date


Residential (% change from baseline)
15
10
5
0
2020−02−15
2020−02−16
2020−02−17
2020−02−18
2020−02−19
2020−02−20
2020−02−21
2020−02−22
2020−02−23
2020−02−24
2020−02−25
2020−02−26
2020−02−27
2020−02−28
2020−02−29
Date

Durham

NC

```
## # A tibble: 99 x 8
## # Groups:   sub_region_2 [99]
##    sub_region_2 na.rm retail_avg grocery_avg parks_avg transit_avg
##    <chr>        <lgl>      <dbl>       <dbl>     <dbl>       <dbl>
##  1 ""           TRUE      -14.9       -1.01      45.4      -24.5
##  2 "Alamance C~ TRUE      -13.0        0.648     32.6       -2.04
##  3 "Alexander ~ TRUE       -7.44      11.7       NaN        NaN
##  4 "Alleghany ~ TRUE      -12.5       -9.75      NaN        NaN
##  5 "Anson Coun~ TRUE       -7.99      10.5       NaN        NaN
##  6 "Ashe Count~ TRUE        4.41       8.2       NaN        NaN
##  7 "Avery Coun~ TRUE      -12.6       13.1       NaN        NaN
##  8 "Beaufort C~ TRUE       -8.22       3.65       17.8      NaN
##  9 "Bertie Cou~ TRUE       18.8        7.02      NaN        NaN
## 10 "Bladen Cou~ TRUE       -3.61       5.86      -24.7      NaN
## # ... with 89 more rows, and 2 more variables: workplaces_avg <dbl>,
## #   residential_avg <dbl>
```

EDA

Linear Regression Code