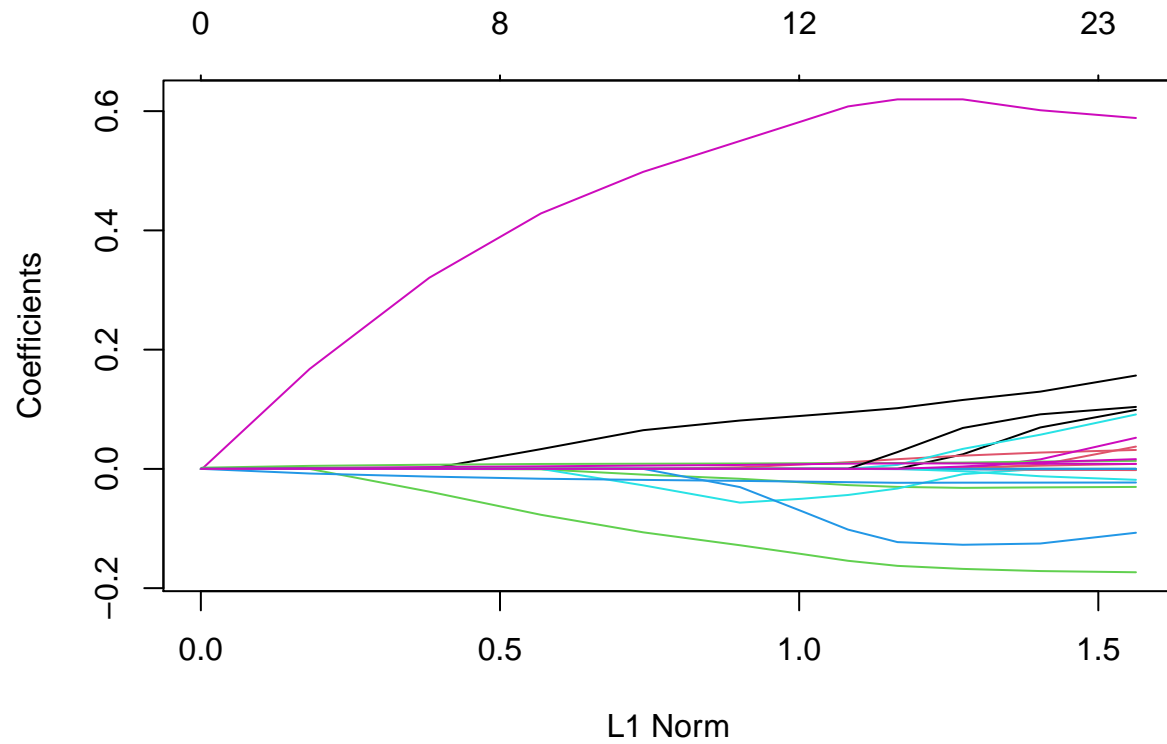
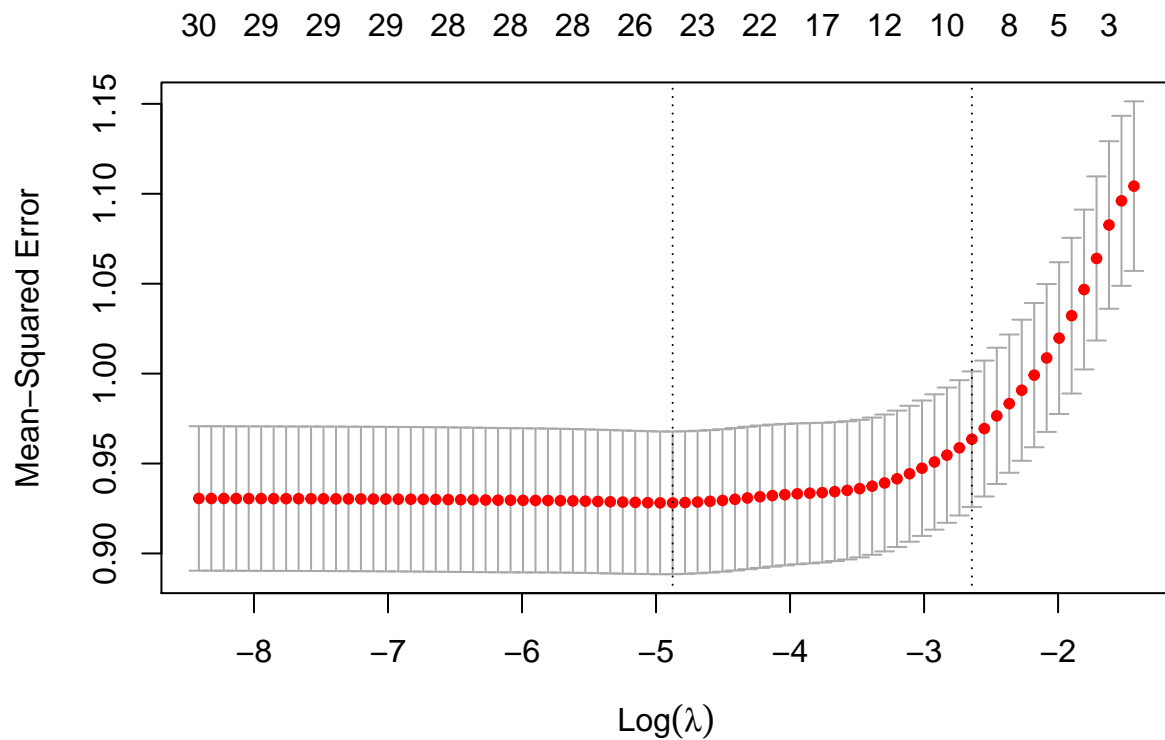


Mishek's Models

```
knitr::opts_chunk$set(message = FALSE, echo = FALSE, warning=FALSE)
```

Lasso





```
## [1] 0.8217405
```

```
##          (Intercept)                popChange
##          -2.847141e+00            -3.829419e-06
##      GQ_ESTIMATES_2019             birthRate
##              0.000000e+00             0.000000e+00
##              deathRate             increaseRate
##              1.734056e-02            -2.365893e-02
##      intMigrationRate             domMigrationRate
##              -3.128053e-03             0.000000e+00
##      netMigrationRate                proTrump
##              -1.378885e-03             0.000000e+00
##      regionNortheast                regionSouth
##              4.658435e-01            -4.207370e-02
##      regionWest                    govPartyRepublican
##              1.235590e-01             0.000000e+00
##      metro_areametro                metro_areanon-metro
##              1.221535e-02            -1.837732e-01
##      main_econgoverment                main_econmanufacturing
##              -1.363412e-01             5.365983e-02
##      main_econmining                  main_econnonsspecialized
##              5.009907e-02             1.324115e-01
##      main_econrecreation                lon
##              0.000000e+00             1.557805e-02
##              lat                    per_capita_income
```

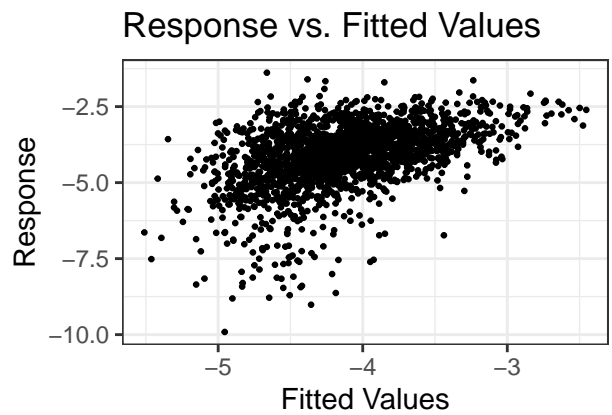
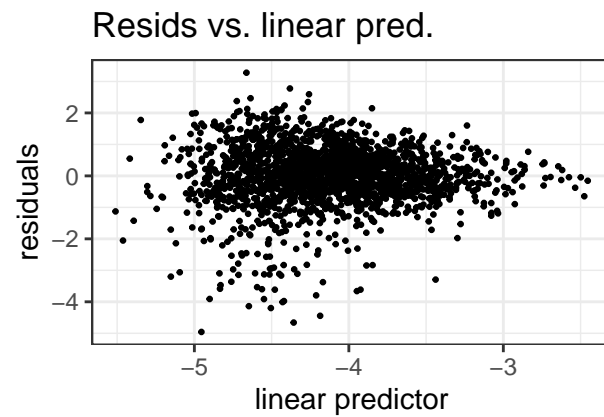
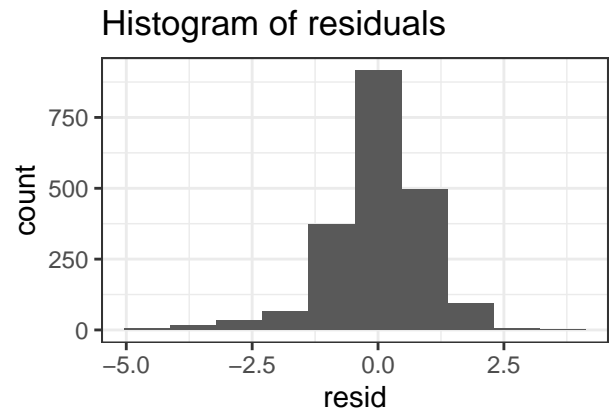
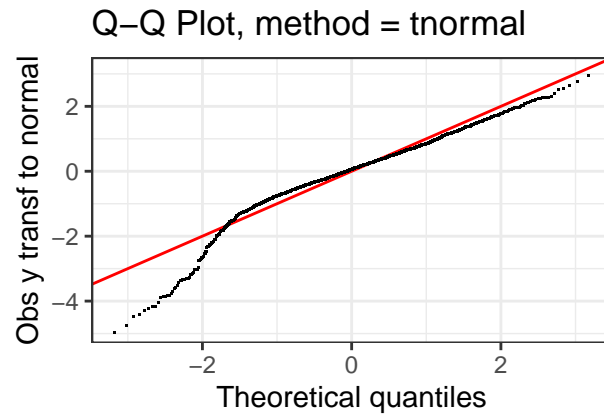
```
##                -2.411929e-02                2.250908e-05
##                percent_below_poverty          percent_minorities
##                0.000000e+00                8.692482e-03
## percentile_rank_social_vulnerability          percent_uninsured
##                1.673974e-01                -2.414282e-04
##                total_population                area_sqmi
##                4.277032e-08                1.239967e-05
##                percent_smokers                percent_fair_or_poor_health
##                -1.005915e-02                1.025401e-02
```

Percent_smokers, percent_uninsured, percent_below_poverty, proTrump, birthrate, GQ estimates, and percent_fair_or_poor_health were the variables that were removed by lasso selection. This led to a cv error of 0.87.

```
#GAM
```

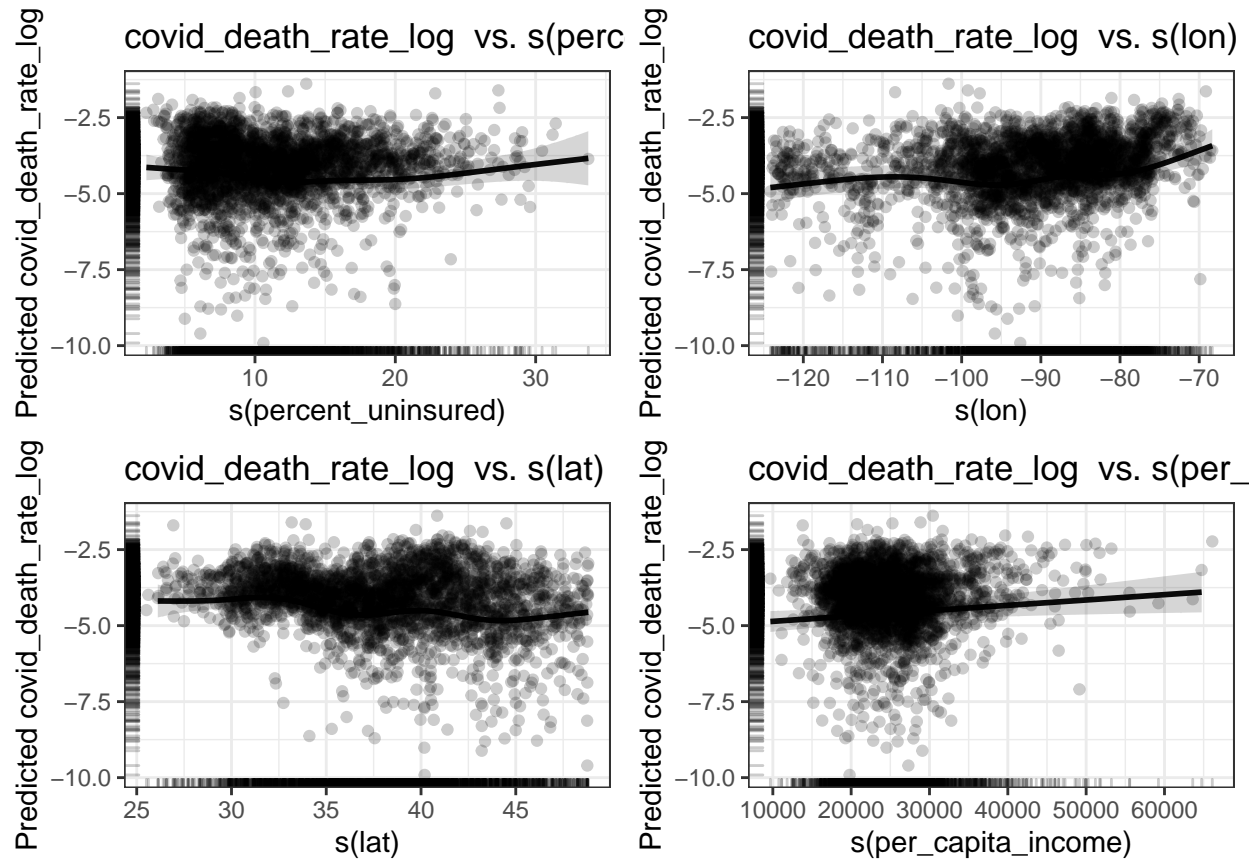
```
## [1] 0.7931617
```

```
##
## Method: REML   Optimizer: outer newton
## full convergence after 11 iterations.
## Gradient range [-0.0003607182,0.001193478]
## (score 2768.851 & scale 0.8786752).
## Hessian positive definite, eigenvalue range [4.928059e-05,989.5207].
## Model rank = 116 / 116
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                k'   edf k-index p-value
## s(density)      9.00 1.00    1.00  0.45
## s(area_sqmi)    9.00 1.00    0.99  0.40
## s(total_population) 9.00 2.69    1.00  0.42
## s(percentile_rank_social_vulnerability) 9.00 1.45    1.05  0.96
## s(percent_minorities) 9.00 1.00    0.99  0.24
## s(per_capita_income) 9.00 1.00    1.03  0.89
## s(lat)          9.00 7.25    1.01  0.71
## s(lon)          9.00 6.49    0.97  0.15
## s(netMigrationRate) 9.00 1.00    1.01  0.62
## s(increaseRate)  9.00 1.00    1.01  0.70
## s(deathRate)     9.00 4.76    0.99  0.28
## s(popChange)     9.00 1.00    0.98  0.26
```



Generally, the gam appears to perform better than the lasso.

```
## [1] 0.7736487
```



However, the forward/backward subset selection does not perform as well as the lasso selection.