Exploring Factors Related COVID-19 Increase Rate in New York State

Diana(Yingtian) Liang

Introduction

This study aims to examine factors associated with the COVID-19 increase rate in New York State through Bayesian hierarchical modeling. We are trying to grasp population and mobility features that could better delineate the pandemic growth.

Data preparation

County-level daily new confirmed cases and 2019 census data are from https://usafactsstatic.blob.core.windows.net/public/data/covid-19. Community mobility data is from https://www.google.com/covid19/mobility/data_documentation.html?hl=en#about-this-data.

Location accuracy and the understanding of categorized places varies from region to region, so we only consider counties in metropolitan areas.

```
ny.data <- data.case %>% filter(state_name=='NY'&cum_cases>1&!is.na(county_pop)) %>%
  left_join(census.pop,by='county_fips') %>%
  left_join(census.ses,by='county_fips') %>%
  left_join(census.metro,by='county_fips') %>%
  mutate(logpop=log(county_pop),
         nyc_vicinity=county_fips%in%c(36005,36047,36081,36059,36061,36103,36119),
         logdensity=log(pop_dens))

ny.metro <- ny.data %>% filter(metro==1) %>%
  mutate(new_cases=if_else(new_cases>=0,new_cases,0),
         highrisk=county_fips%in%c(36005,36047,36061,36081,36103,36119,36085,36087,36071,36059)) #Bronx

ny.mobility <- global.mobility %>%
  filter(country_region_code=='US'&sub_region_1=='New York') %>%
  select(state_name=sub_region_1,
         county_name=sub_region_2,
         date,
         retail_recreation_change=retail_and_recreation_percent_change_from_baseline,
         grocery_pharmacy_change=grocery_and_pharmacy_percent_change_from_baseline,
         parks_change=parks_percent_change_from_baseline,
         transit_stations_change=transit_stations_percent_change_from_baseline,
         workplaces_change=workplaces_percent_change_from_baseline,
         residential_change=residential_percent_change_from_baseline)

dat <- ny.metro %>% select(-state_name) %>% left_join(ny.mobility,by=c('county_name','date'))
```

Model Structure

Let i index the county in New York State, let t denote the # of days elapsed since the first case was founded in that county, let y_{it} denote the observed # of confirmed cases in county i and epidemic days t, and let $\widetilde{y}_{it}$ denote per-capita confirmed rate, the model assumes that the expected daily death rate λ_{it} can be locally approximated by a curve proportional to a Gaussian kernel: $E(y_{it}) = \lambda_i(t) = \kappa_{it} \cdot exp\{-\frac{\omega_{it}}{2} \cdot (t-\mu_{it})^2\}$ $log\lambda_{it} = log\kappa_i - 0.5\omega_i(t-\mu_i)^2 \equiv \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2$, where: κ: maximum daily expected new cases, μ:the day on which the expected daily new cases achieves its maximum, ω: a steepness parameter, higher ω means the increase rate rises more rapidly as t approaches μ, and also falls more rapidly on the far side of μ. Specifically, the slope at the inflection point of the increase-rate curve is $\kappa\sqrt{\omega}$.

$$\begin{bmatrix} \omega \\ \mu \\ \kappa \end{bmatrix} = \begin{bmatrix} -2\beta_2 \\ \frac{\beta_1}{2\beta_2} \\ \exp\{\beta_0 - \frac{\beta_1^2}{2\beta_2}\} \end{bmatrix}$$

where λ_{it} denotes the expected value for the # of daily new confirmed cases y_{it}.

To deal with the potential over-dispersion issue, we choose to move on with a negative binomial model instead of a Poisson model.

To fit a hierarchical negative binomial regression model with mean λ and over-dispersion parameter r:

$$y_{it} \sim NegBin(\lambda_{it}, r) log\lambda_{it} = logN_i + \beta_{it,0} + \beta_{it,1}t + \beta_{it,2}t^2 (\beta_{it,0}, \beta_{it,1}, \beta_{it,2})^T = \mu + \Gamma x_{it} + \eta_i \eta_i \sim N(0, \Sigma)$$

, where $E(y_{it}) = \lambda_{it}$ and $Var(y_{it}) = \lambda_{it}(1 + \lambda_{it}/r)$.

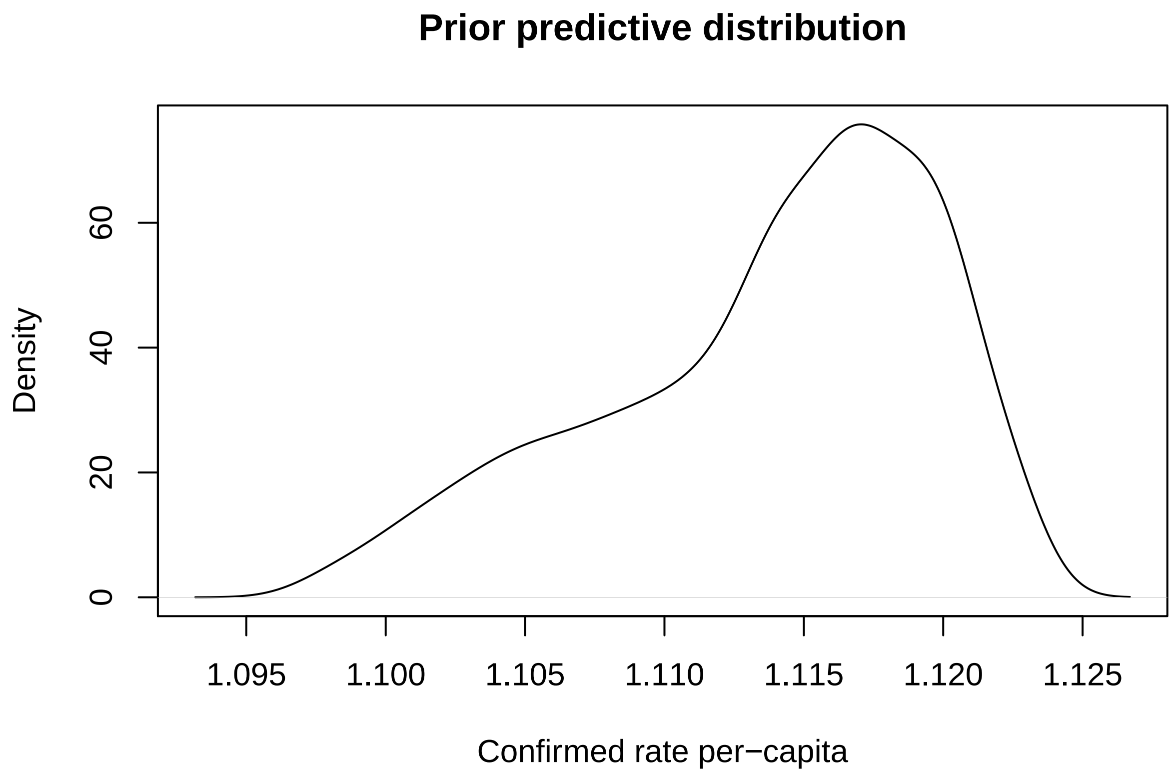
Here an offset is included for the logarithm of N_i, the population in county i, so that the linear predictor can be interpreted as the log per-capita new confirmed rate. Here x_{it} is a vector of census and mobility covariates. Census covariates vary across counties, and mobility covariates vary across days and across counties.

Baseline Model: County-Specific Time Effects

First fit a baseline model only including time and its quadratic term as predictors.

Prior Predictive Distribution

```
draws0 <- brm(new_cases~poly(day,2)+(poly(day,2)|county_fips)+offset(logpop),data=dat,family=negbinomia
              prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
                prior(normal(-0.1,0.05),class='b',coef='polyday22') +
                prior(normal(0.1,0.1),class='Intercept') +
                prior(lkj(2),class='cor') +
                prior(normal(0,0.1),class='sd'))

mu0.prior <- pp_expect(draws0)

plot(density(colMeans(mu0.prior)/dat$county_pop),xlab='Confirmed rate per-capita',main='Prior predictiv
```



We get daily new cases from the prior predictive distribution larger than the county population, which seems unreasonable. The prior predictive distribution is hard to control because we are using a simple model to delineate a complex situation, without controlling for observed data.

Posterior Distribution

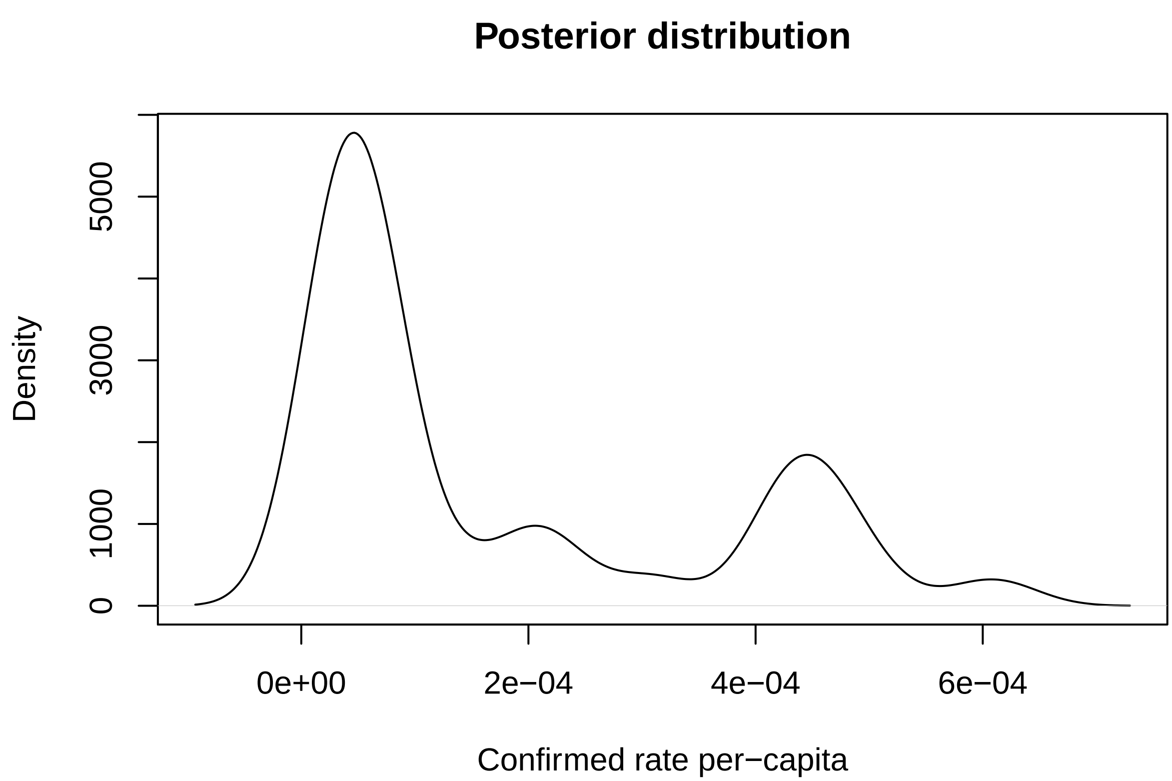
Then condition on the observed data to get posterior distribution.

```
post0 <- update(draws0,sample_prior='no')
post0
```

```
##  Family: negbinomial
##   Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + (poly(day, 2) | county_fips) + offset(logpop)
##    Data: dat (Number of observations: 2066)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 38)
##                            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)                  2.30      0.05     2.19     2.40 1.02      173
## sd(polyday21)                  0.19      0.11     0.01     0.40 1.00     1745
## sd(polyday22)                  0.31      0.12     0.05     0.53 1.00     1318
## cor(Intercept,polyday21)      -0.55      0.33    -0.94     0.31 1.00     2365
## cor(Intercept,polyday22)       0.76      0.21     0.22     0.97 1.00     1777
## cor(polyday21,polyday22)      -0.42      0.35    -0.90     0.42 1.00     2371
##                            Tail_ESS
```

```
## sd(Intercept)                448
## sd(polyday21)               1314
## sd(polyday22)                573
## cor(Intercept,polyday21)     2488
## cor(Intercept,polyday22)     1188
## cor(polyday21,polyday22)     2461
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    -0.53      0.10    -0.73    -0.33 1.00     2187     2420
## polyday21     0.28      0.10     0.08     0.47 1.00     5306     2666
## polyday22    -0.13      0.05    -0.22    -0.03 1.00     4926     3068
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape     1.01      0.04     0.94     1.08 1.00     4122     3090
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
mu0.post <- pp_expect(post0)

plot(density(colMeans(mu0.post)/dat$county_pop),xlab='Confirmed rate per-capita',main='Posterior distri
```



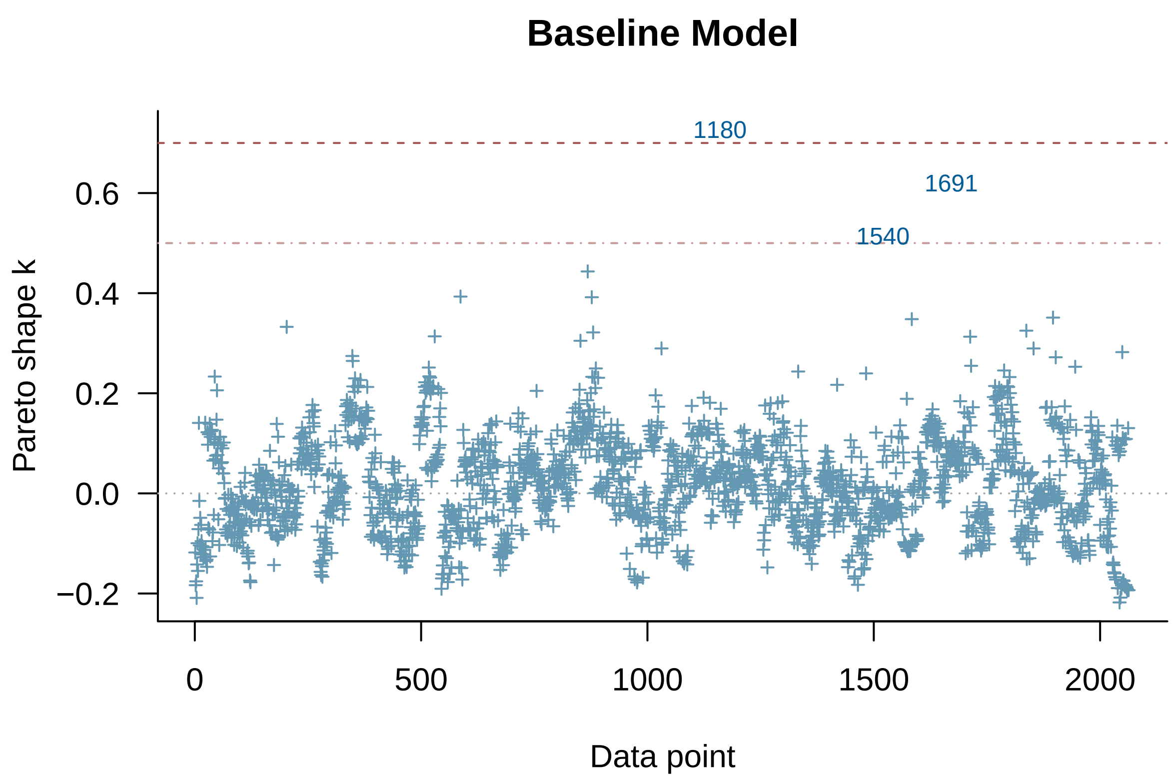
After controlling for the observed data, we get a more reasonable posterior distribution. Most of the counties have confirmed rate at around 0.01%, while some more vulnerable counties have higher rate at around 0.05%.

The group-level intercept has much larger effect than others, which means the expected baseline growth rate has large variability across counties.

Diagnostic

We use leave one out cross-validation to assess the model fit for the baseline model.

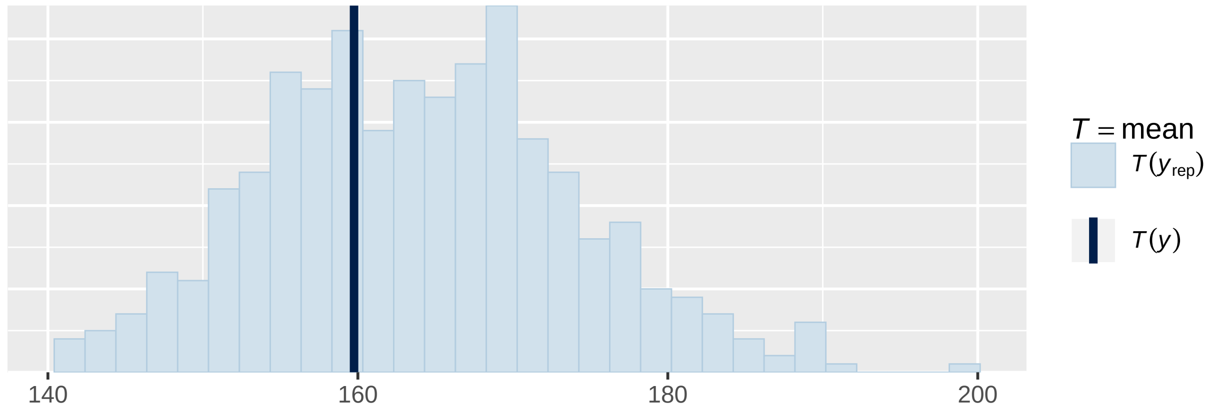
```
loo0 <- loo(post0)
plot(loo0,main="Baseline Model",label_points=T)
```

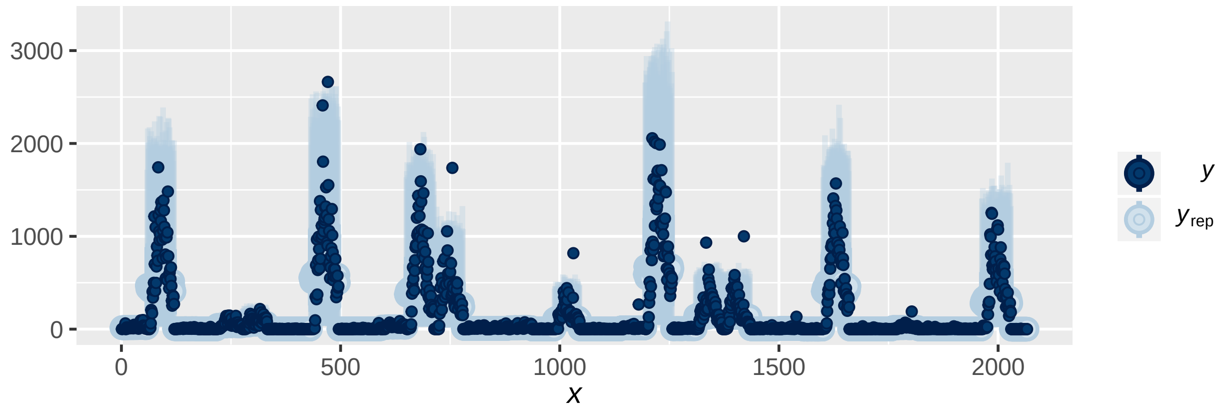



```
# numerical assessment of calibration
PPD0 <- posterior_predict(post0)
lower0 <- apply(PPD0,2,quantile,probs=.025)
upper0 <- apply(PPD0,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat$new_cases>lower0&dat$new_cases<upper0))
```

```
## The probability of observations fall into 95% of the posterior distribution is 0.8262343
```

```
grid.arrange(
pp_check(post0,nsample=500,type='stat',stat='mean'),
pp_check(post0,nsample=500,type='intervals'),
nrow=2)
```





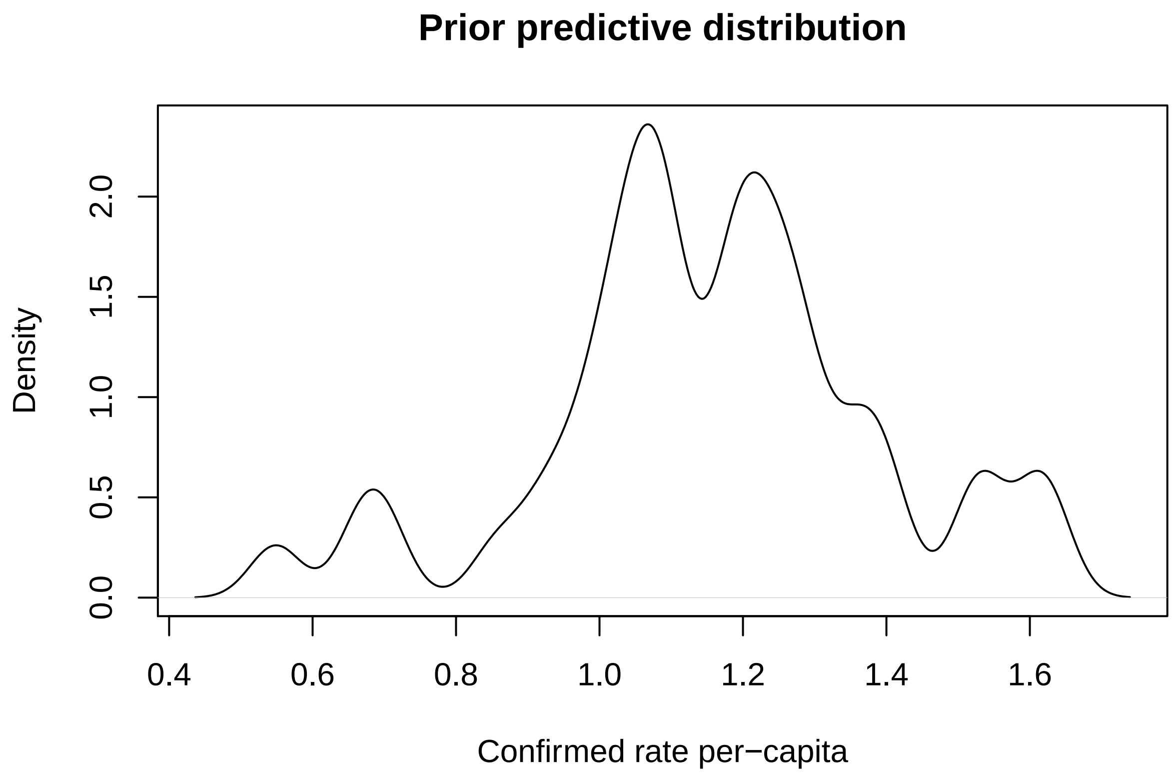
Two highly influenced data points were detected. 82.77% of the observed data points fall in 95% of the posterior distribution, which is not bad. Eyeballing the loo-based plots, the baseline model catches the mean fairly well.

Adding Census Covariates

Compared with other counties, NYC has extremely high new confirmed rate, so we include whether the county is in vicinity of NYC to see its effect. As we know, the elderly is more vulnerable to COVID-19 and the spread of virus is faster in more crowded regions, thus we include the percentage of elderly and the log density of population to see their effects.

Prior Predictive Distribution

```
draws1 <- brm(new_cases~poly(day,2)+nyc_vicinity+pct_pop_elder+logdensity+(poly(day,2)|county_fips)+off
           prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
             prior(normal(-0.1,0.05),class='b',coef='polyday22') +
             prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
             prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
             prior(normal(0.1,0.05),class='b',coef='logdensity') +
             prior(normal(0.1,0.1),class='Intercept') +
             prior(lkj(2),class='cor') +
             prior(normal(0,0.1),class='sd'))

mu1.prior <- pp_expect(draws1)

plot(density(colMeans(mu1.prior)/dat$county_pop),xlab='Confirmed rate per-capita',main='Prior predictiv
```



The model with census covariates generates a more wide spread prior predictive distribution, with both reasonable and unreasonable values.

Posterior Distribution

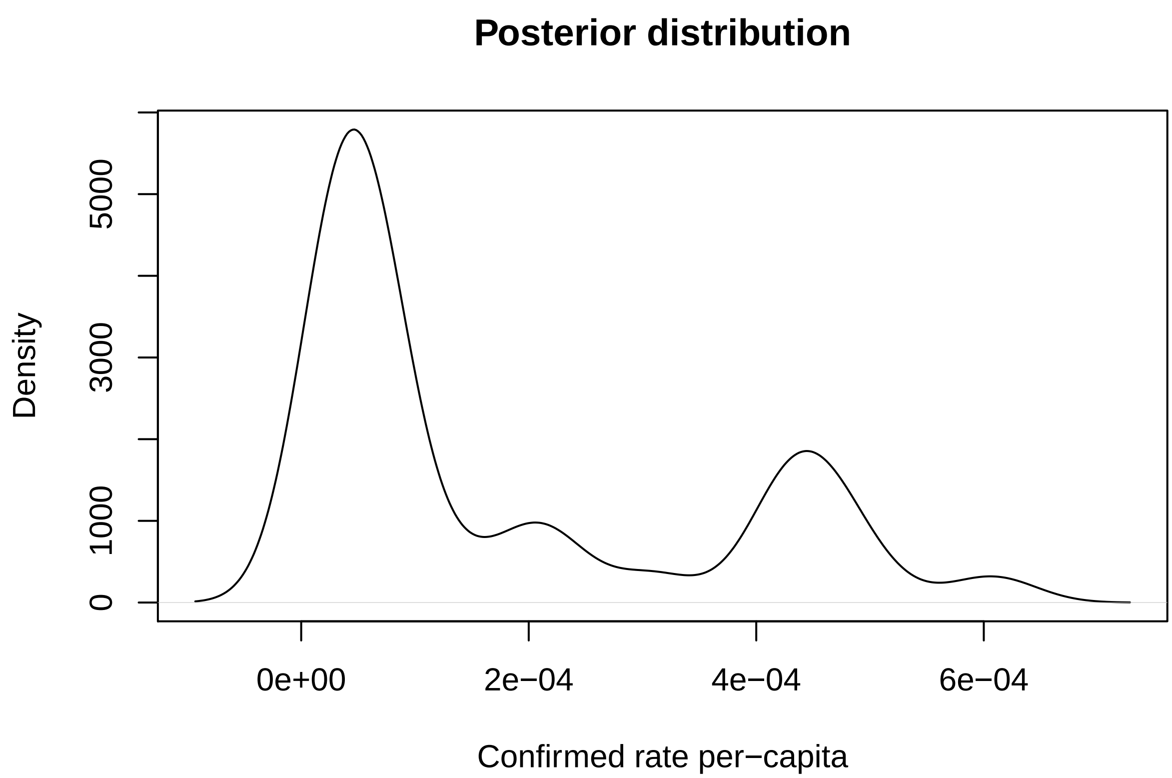
Then get posterior distribution conditioning on the observed data.

```
post1 <- update(draws1,sample_prior='no')
post1
```

```
##  Family: negbinomial
##   Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + nyc_vicinity + pct_pop_elder + logdensity + (poly(day, 2) | coun
##    Data: dat (Number of observations: 2066)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 38)
##                            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)                  2.29      0.05     2.19     2.39 1.01      405
## sd(polyday21)                  0.19      0.11     0.01     0.41 1.00     2075
## sd(polyday22)                  0.33      0.11     0.09     0.54 1.00     2443
## cor(Intercept,polyday21)      -0.55      0.35    -0.94     0.41 1.00     3419
## cor(Intercept,polyday22)       0.79      0.17     0.35     0.97 1.00     3033
## cor(polyday21,polyday22)      -0.44      0.36    -0.91     0.45 1.00     2815
##                            Tail_ESS
## sd(Intercept)                   840
```

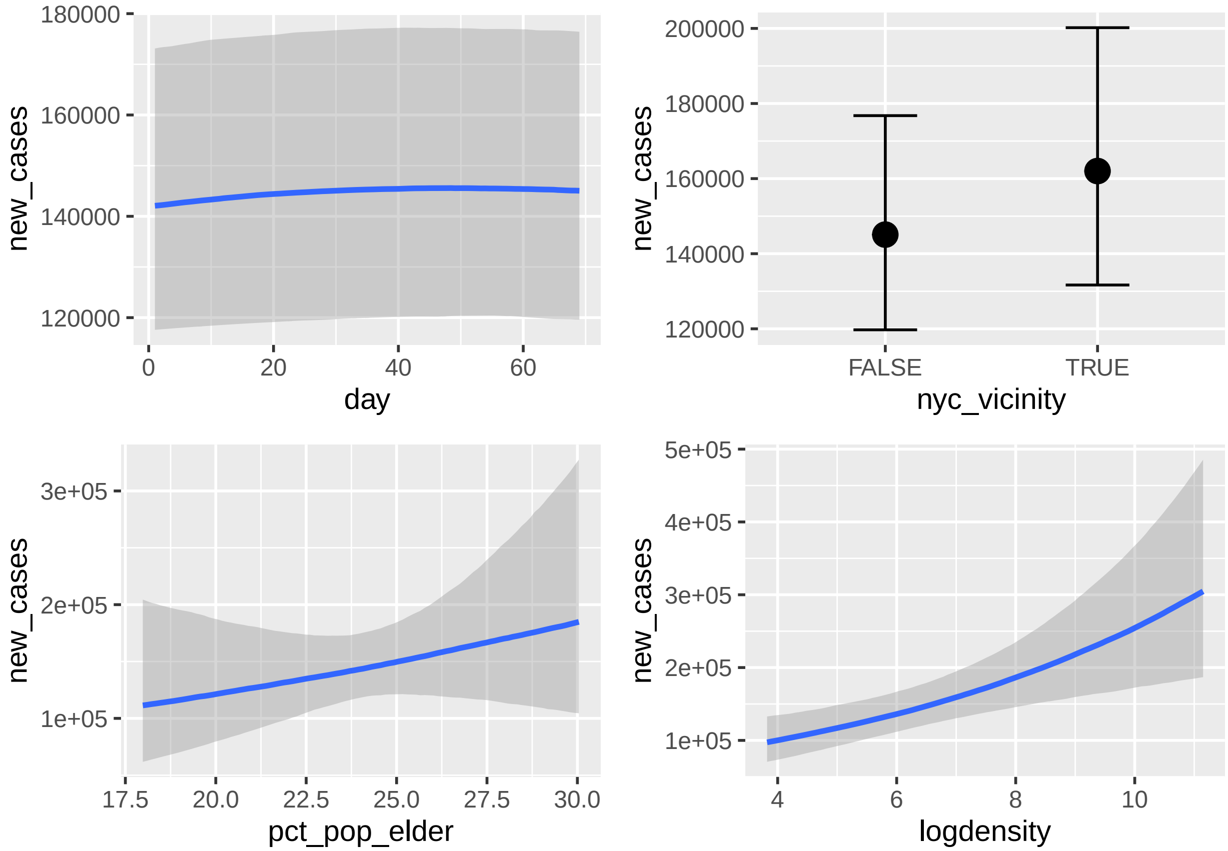
```
## sd(polyday21)               1306
## sd(polyday22)               1164
## cor(Intercept,polyday21)    2472
## cor(Intercept,polyday22)    1964
## cor(polyday21,polyday22)    2645
##
## Population-Level Effects:
##                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept              -2.56      1.20    -4.86    -0.23 1.00     1005     1817
## polyday21               0.28      0.10     0.07     0.48 1.00     6602     2499
## polyday22              -0.13      0.05    -0.23    -0.03 1.00     5861     2644
## nyc_vicinityTRUE        0.11      0.05     0.01     0.21 1.00     5923     3253
## pct_pop_elder           0.04      0.05    -0.05     0.13 1.00      776     1398
## logdensity              0.16      0.05     0.06     0.25 1.00     1355     2048
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape     1.01      0.03     0.95     1.08 1.00     4449     3016
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
mu1.post <- pp_expect(post1)

plot(density(colMeans(mu1.post)/dat$county_pop),xlab='Confirmed rate per-capita',main='Posterior distri
```



The posterior distribution seems reasonable after conditioning on observed data. It looks quite similar to that of the baseline model.

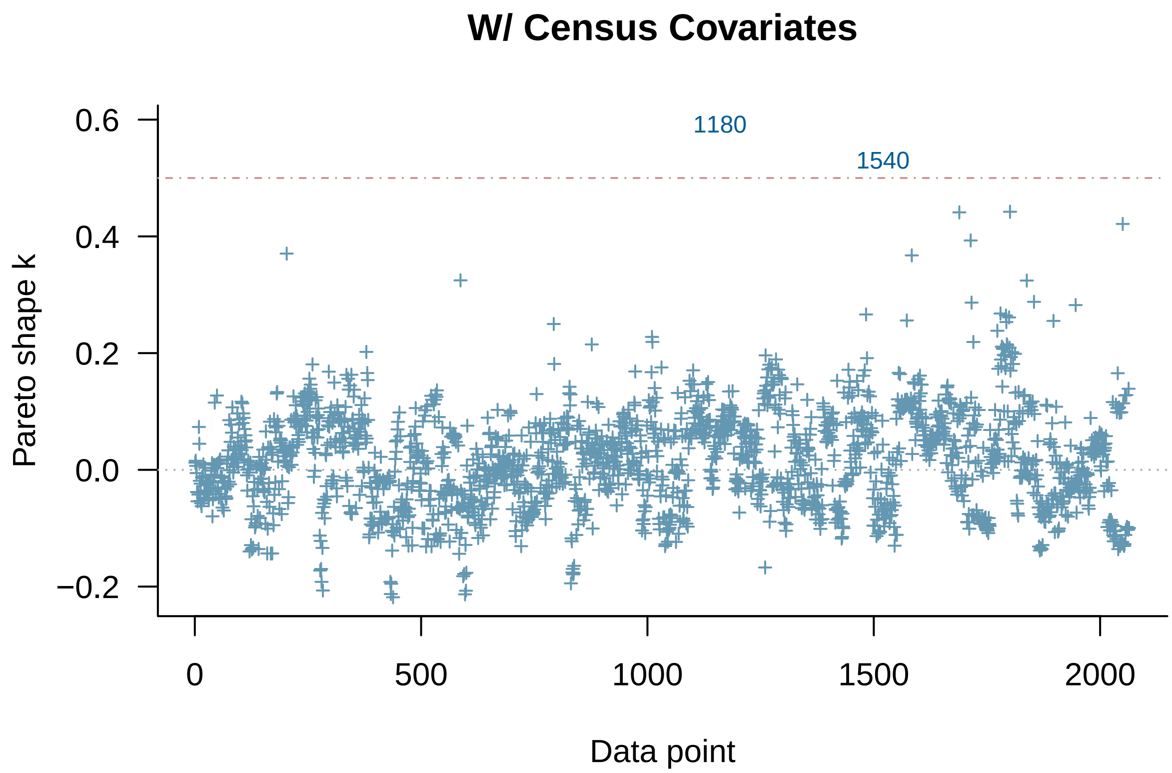
```
# conditional effects
con1.1 <- plot(conditional_effects(post1,'day'),plot=F,ask=F)$day
con1.2 <- plot(conditional_effects(post1,'nyc_vicinity'),plot=F,ask=F)$nyc_vicinity
con1.3 <- plot(conditional_effects(post1,'pct_pop_elder'),plot=F,ask=F)$pct_pop_elder
con1.4 <- plot(conditional_effects(post1,'logdensity'),plot=F,ask=F)$logdensity
grid.arrange(con1.1,con1.2,con1.3,con1.4,nrow=2)
```



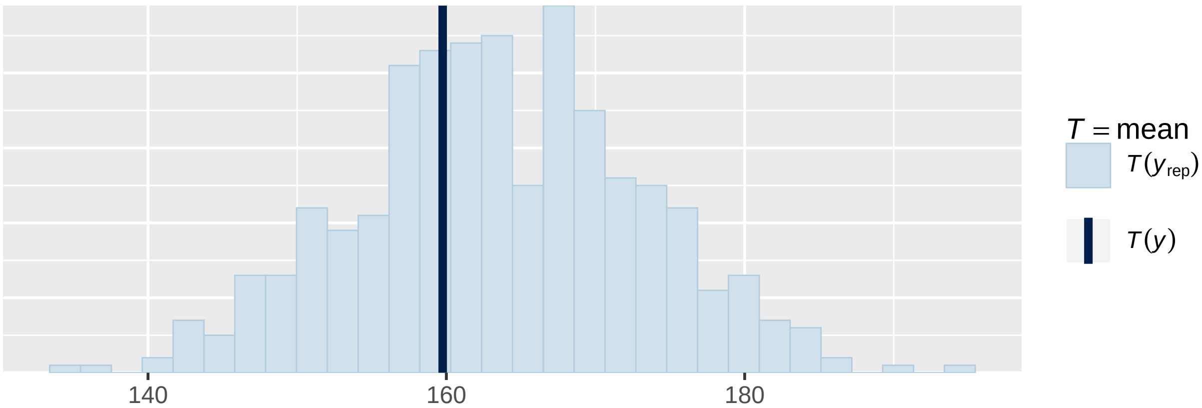
Holding everything else constant, we can see that the marginal effects of whether in vicinity of NYC, percentage of elderly, and population density are all positively associated with the daily new confirmed cases, as expected. The effect of the percentage of elderly is negligible with high uncertainty.

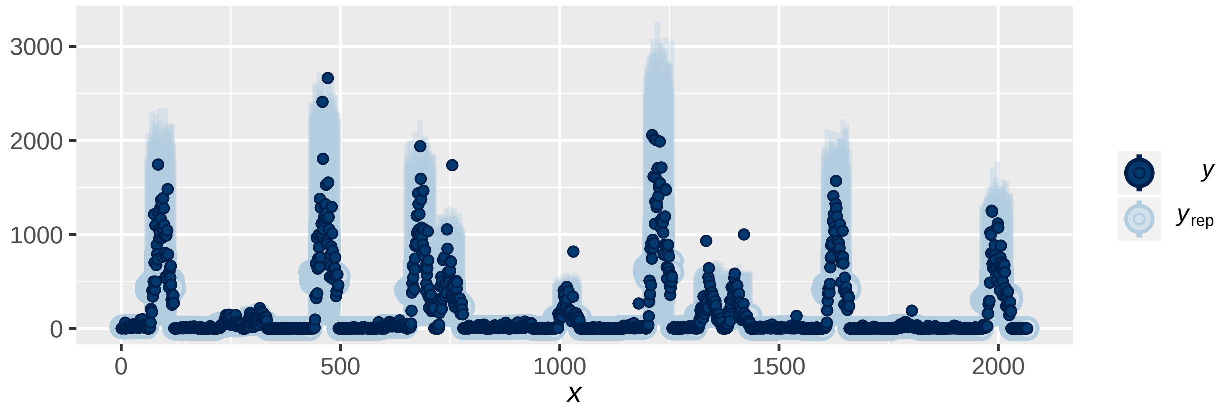
Diagnostics

```
loo1 <- loo(post1)
plot(loo1,main="W/ Census Covariates",label_points=T)
```




```
gridExtra::grid.arrange(
pp_check(post1,nsample=500,type='stat',stat='mean'),
pp_check(post1,nsample=500,type='intervals'),
nrow=2)
```






```
# numerical assessment of calibration
PPD1 <- posterior_predict(post1)
lower1 <- apply(PPD1,2,quantile,probs=.025)
upper1 <- apply(PPD1,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat$new_cases>lower1&dat$new_cases<upper1))
```

```
## The probability of observations fall into 95% of the posterior distribution is 0.8272023
```

```
# model comparison
loo_compare(loo0,loo1)
```

```
##       elpd_diff se_diff
## post1  0.0       0.0
## post0 -1.7       0.3
```

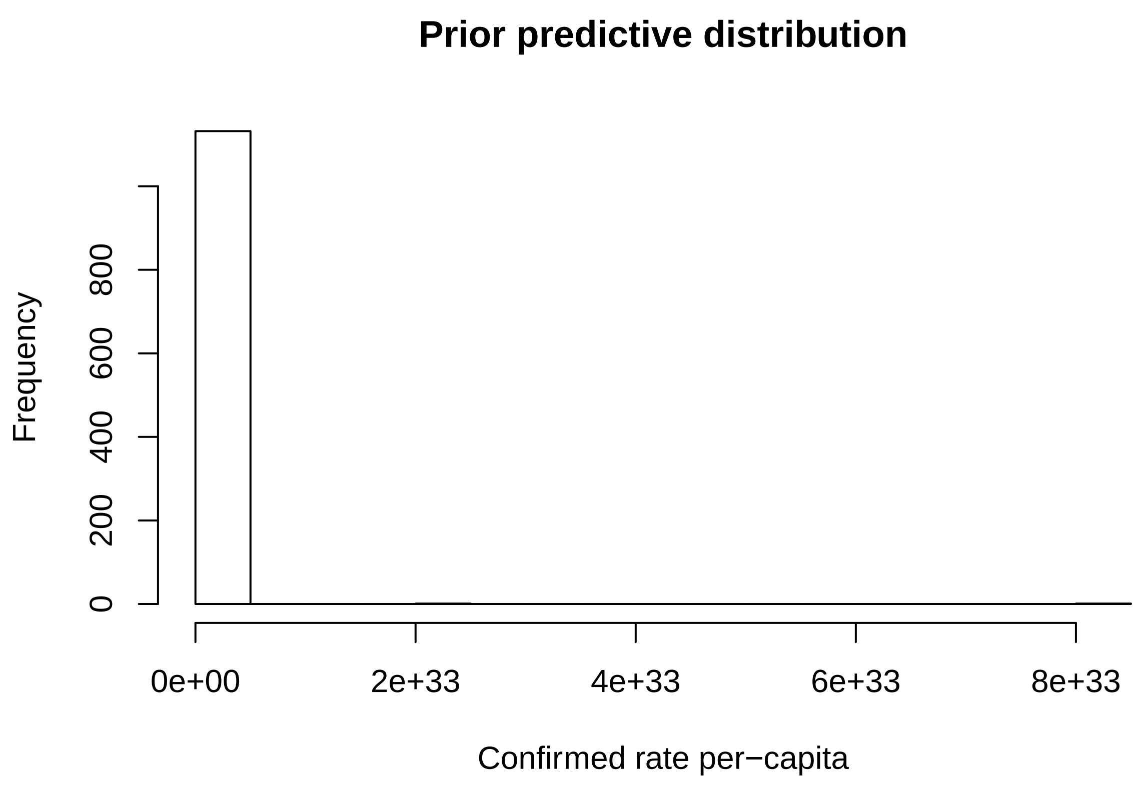
One highly influenced observation was detected. 82.72% of the observed data points fall in 95% of the posterior distribution, which is not bad. Eyeballing the loo-based plots, the model with census covariates catches the mean fairly well. The ELPD suggests that the model with census covariates is preferred over the baseline model.

Adding Population-Level Mobility Effects

Mobility might be another influential factor associated with the spread of the pandemic. Thus, we further include several mobility covariates to the model in order to see their impacts. Here we consider the percent change in visits to places like grocery stores and pharmacy, parks, and public transport hubs within each county. Changes for each day are compared to a baseline value for that day of the week.

Prior Predictive Distribution

```
draws2 <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc_vi
          prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
            prior(normal(-0.1,0.05),class='b',coef='polyday22') +
            prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
            prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
            prior(normal(0.1,0.05),class='b',coef='logdensity') +
            prior(normal(0.1,0.05),class='b',coef='transit_stations_change') +
            prior(normal(0.1,0.05),class='b',coef='grocery_pharmacy_change') +
            prior(normal(0.1,0.05),class='b',coef='parks_change') +
            prior(normal(0.1,0.1),class='Intercept') +
            prior(lkj(2),class='cor') +
            prior(normal(0,0.1),class='sd'))

mu2.prior <- pp_expect(draws2)

hist((colMeans(mu2.prior)/dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_change
```



Under weakly informative priors, the prior predictive distribution seems quite widespread, with some extremely large values which even exceed the population of the whole world.

Posterior Distribution

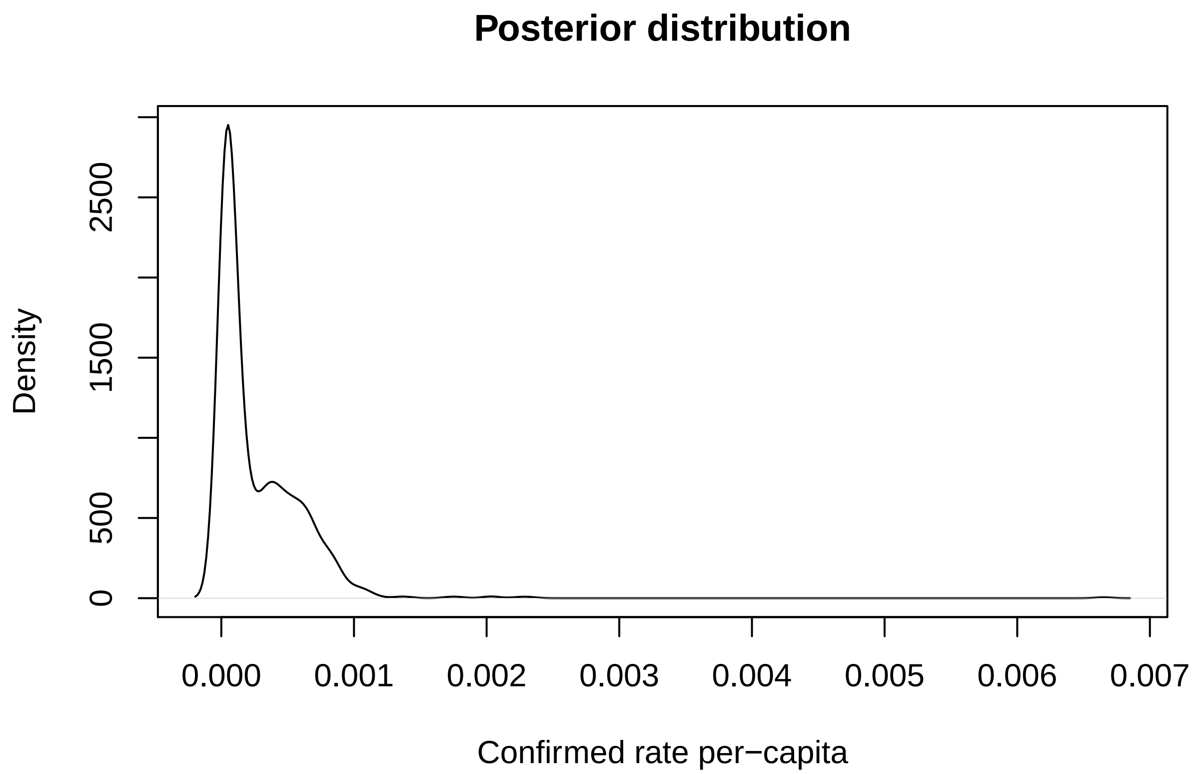
Then get posterior distribution conditioning on observed data.

```r
post2 <- update(draws2,sample_prior='no')
post2
```

```
##  Family: negbinomial 
##   Links: mu = log; shape = identity 
## Formula: new_cases ~ poly(day, 2) + transit_stations_change + grocery_pharmacy_change + parks_change
##    Data: dat (Number of observations: 1134) 
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
## 
## Group-Level Effects: 
## ~county_fips (Number of levels: 26) 
##                           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)                 2.07      0.05     1.97     2.18 1.01      578
## sd(polyday21)                 0.14      0.09     0.01     0.33 1.00     2914
## sd(polyday22)                 0.20      0.11     0.01     0.41 1.00     2529
## cor(Intercept,polyday21)     -0.41      0.37    -0.90     0.45 1.00     6064
## cor(Intercept,polyday22)      0.58      0.31    -0.30     0.94 1.00     4783
## cor(polyday21,polyday22)     -0.24      0.39    -0.85     0.59 1.00     4680
##                           Tail_ESS
## sd(Intercept)                 1232
## sd(polyday21)                 2134
## sd(polyday22)                 1773
## cor(Intercept,polyday21)      3169
## cor(Intercept,polyday22)      2500
## cor(polyday21,polyday22)      2992
## 
## Population-Level Effects: 
##                          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                   -6.05      1.19    -8.31    -3.74 1.00     2740
## polyday21                    0.26      0.10     0.07     0.46 1.00     9007
## polyday22                   -0.12      0.05    -0.22    -0.02 1.00    10278
## transit_stations_change     -0.07      0.00    -0.08    -0.07 1.00     6366
## grocery_pharmacy_change     -0.00      0.00    -0.01     0.00 1.00     5636
## parks_change                 0.01      0.00     0.00     0.01 1.00     4291
## nyc_vicinityTRUE             0.11      0.05     0.01     0.21 1.00     8746
## pct_pop_elder                0.04      0.05    -0.06     0.13 1.00     2152
## logdensity                   0.15      0.05     0.06     0.25 1.00     3122
##                          Tail_ESS
## Intercept                    2762
## polyday21                    2358
## polyday22                    2753
## transit_stations_change      3045
## grocery_pharmacy_change      3533
## parks_change                 3538
## nyc_vicinityTRUE             2251
## pct_pop_elder                2520
## logdensity                   3144
## 
## Family Specific Parameters: 
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape     2.00      0.10     1.81     2.20 1.00     7386     2661
## 
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
```

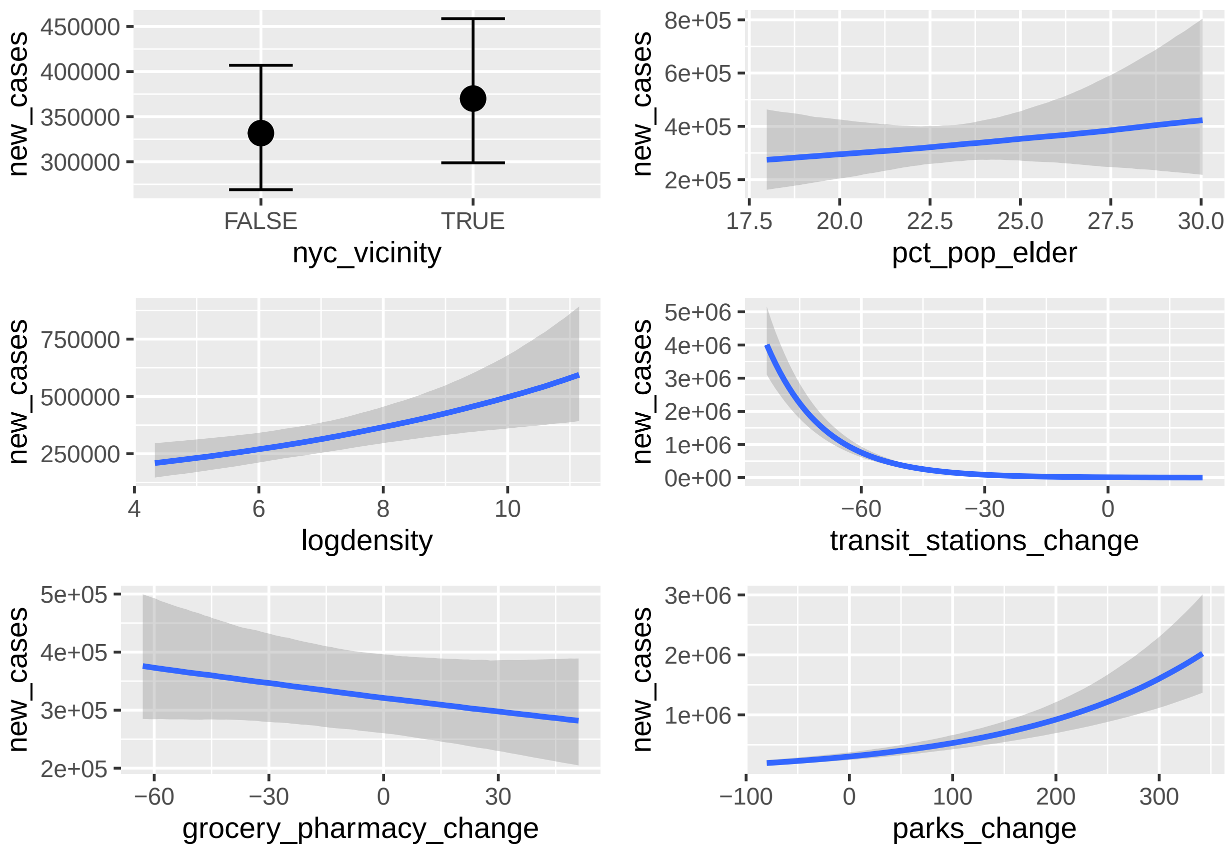
```
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
mu2.post <- pp_expect(post2)

plot(density(colMeans(mu2.post)/dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_c
```



Conditioning on observed data, the posterior distribution is largely narrowed down to a smaller range. Most counties have the confirmed rate less than 1%, with several counties more than 6%.

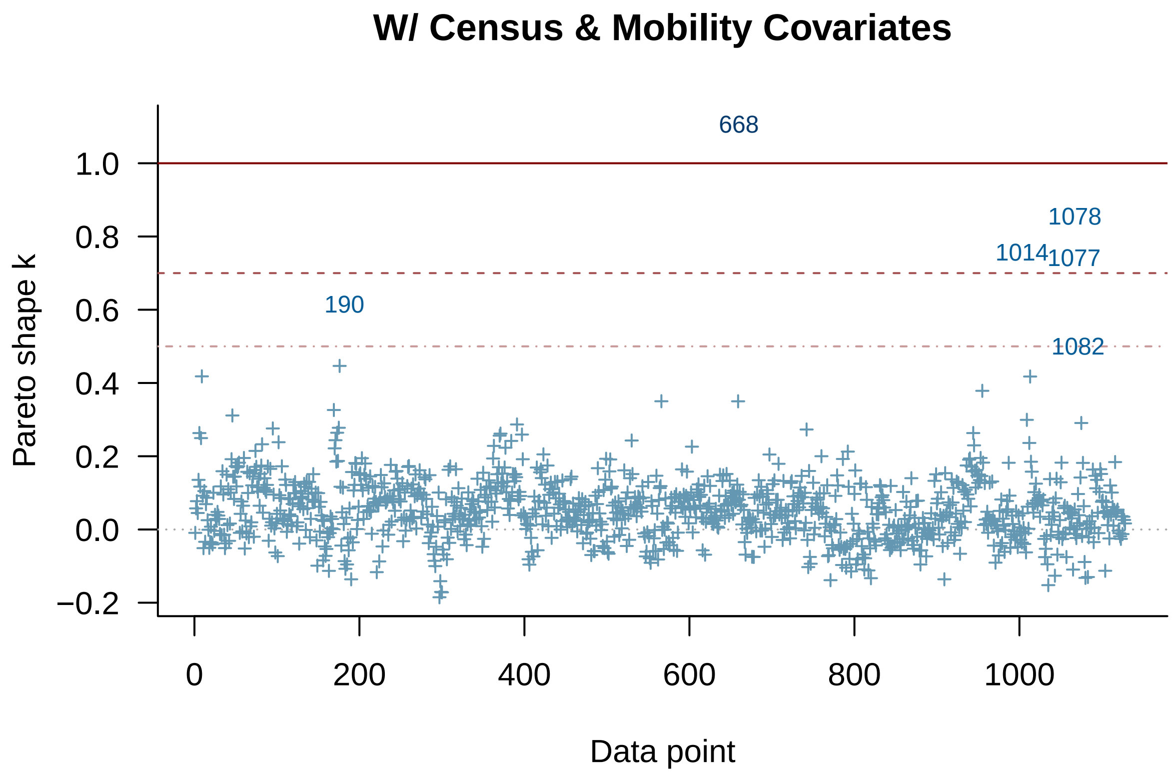
```
# conditional effects
con2.1 <- plot(conditional_effects(post2,'nyc_vicinity'),plot=F,ask=F)$nyc_vicinity
con2.2 <- plot(conditional_effects(post2,'pct_pop_elder'),plot=F,ask=F)$pct_pop_elder
con2.3 <- plot(conditional_effects(post2,'logdensity'),plot=F,ask=F)$logdensity
con2.4 <- plot(conditional_effects(post2,'transit_stations_change'),plot=F,ask=F)$transit_stations_chang
con2.5 <- plot(conditional_effects(post2,'grocery_pharmacy_change'),plot=F,ask=F)$grocery_pharmacy_chan
con2.6 <- plot(conditional_effects(post2,'parks_change'),plot=F,ask=F)$parks_change
grid.arrange(con2.1,con2.2,con2.3,con2.4,con2.5,con2.6,nrow=3,ncol=2)
```



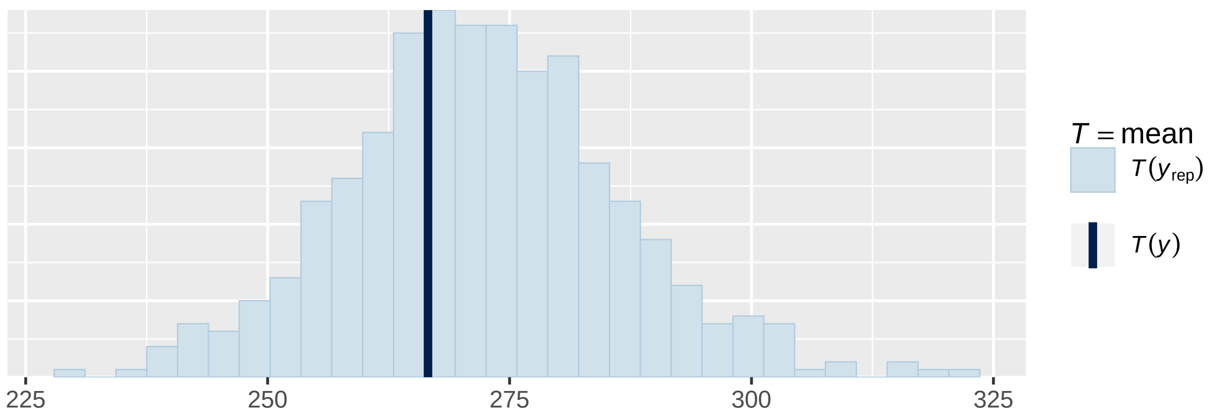
After adding mobility trends covariates, the effects of other covariates don't change much. The marginal effect of mobility trends for transportation seems to be negatively associated with the new cases, which is out of our expectation that new cases will decrease when people take less public transportation. The marginal effects of mobility trends for grocery and pharmacy and mobility trends for parks are very small.

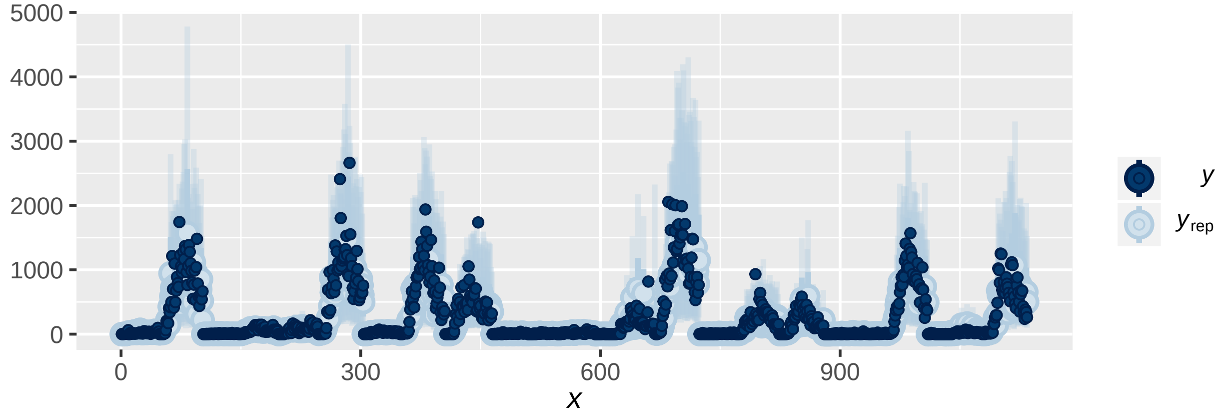
Diagnostics

```
loo2 <- loo(post2)
plot(loo2,main="W/ Census & Mobility Covariates",label_points=T)
```




```
grid.arrange(
pp_check(post2,nsample=500,type='stat',stat='mean'),
pp_check(post2,nsample=500,type='intervals'),
nrow=2)
```






```
# numerical assessment of calibration
PPD2 <- posterior_predict(post2)
lower2 <- apply(PPD2,2,quantile,probs=.025)
upper2 <- apply(PPD2,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_change,parks_change,nyc_vic
```

```
## The probability of observations fall into 95% of the posterior distribution is 0.9021164
```

```
# model comparison
loo1.1 <- loo(post1,newdata=dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_chang

loo_compare(loo1.1,loo2)
```

```
##       elpd_diff se_diff
## post2    0.0       0.0
## post1 -379.6      33.6
```

Three highly influenced observations were detected. Eyeballing the loo-based plots, the model with census and mobility effects predicts a slightly deviated mean. The ELPD suggests that the model with mobility covariates is preferred.

Adding County-Specific Mobility Effects

As the effect of mobility trends might differ across counties, we further consider county-specific mobility effects.

Transit stations only

Only include county-specific mobility trends for public transportation hub.

```
# get posterior distribution
post.trans <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc
          prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
            prior(normal(-0.1,0.05),class='b',coef='polyday22') +
            prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
            prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
            prior(normal(0.1,0.05),class='b',coef='logdensity') +
            prior(normal(0.1,0.05),class='b',coef='transit_stations_change') +
            prior(normal(0.1,0.05),class='b',coef='grocery_pharmacy_change') +
            prior(normal(0.1,0.05),class='b',coef='parks_change') +
            prior(normal(0.1,0.1),class='Intercept') +
            prior(lkj(2),class='cor') +
            prior(normal(0,0.1),class='sd'))
# model comparison
loo.trans <- loo(post.trans)
loo_compare(loo.trans,loo2)
```

```
##              elpd_diff se_diff
## post.trans   0.0       0.0
## post2      -74.3      15.9
```

The ELPD suggests that the model with county-specific transportation mobility effect is preferred.

Add grocery & pharmacy

Further include county-specific mobility trends for grocery and pharmacy.

```
# get posterior distribution
post.tg <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc_vi
          prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
            prior(normal(-0.1,0.05),class='b',coef='polyday22') +
            prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
            prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
            prior(normal(0.1,0.05),class='b',coef='logdensity') +
            prior(normal(0.1,0.05),class='b',coef='transit_stations_change') +
            prior(normal(0.1,0.05),class='b',coef='grocery_pharmacy_change') +
            prior(normal(0.1,0.05),class='b',coef='parks_change') +
            prior(normal(0.1,0.1),class='Intercept') +
            prior(lkj(2),class='cor') +
            prior(normal(0,0.1),class='sd'))
# model comparison
loo.tg <- loo(post.tg)
loo_compare(loo.trans,loo.tg)
```

```
##              elpd_diff se_diff
## post.tg      0.0       0.0
## post.trans -2.7       2.9
```

According to the ELPD, the model with county-specific grocery and pharmacy mobility change is not better than the previous model.

Add parks

Then try adding parks mobility instead of grocery and pharmacy at county-specific level.

```
# get posterior distribution
post.tp <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc_vi
           prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
             prior(normal(-0.1,0.05),class='b',coef='polyday22') +
             prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
             prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
             prior(normal(0.1,0.05),class='b',coef='logdensity') +
             prior(normal(0.1,0.05),class='b',coef='transit_stations_change') +
             prior(normal(0.1,0.05),class='b',coef='grocery_pharmacy_change') +
             prior(normal(0.1,0.05),class='b',coef='parks_change') +
             prior(normal(0.1,0.1),class='Intercept') +
             prior(lkj(2),class='cor') +
             prior(normal(0,0.1),class='sd'))
# model comparison
loo.tp <- loo(post.tp)
loo_compare(loo.trans,loo.tp)
```

```
##            elpd_diff se_diff
## post.trans  0.0       0.0
## post.tp    -0.6       4.3
```

According to the ELPD, the model with both county-specific grocery and pharmacy mobility change and county-specific transportation mobility effect is not better than the one with only county-specific transportation mobility effect.

Add all mobility covariates

Then include 3 county-specific mobility covariates.

```
post3 <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc_vic
           prior=prior(normal(0.2,0.1),class='b',coef='polyday21') +
             prior(normal(-0.1,0.05),class='b',coef='polyday22') +
             prior(normal(0.1,0.05),class='b',coef='nyc_vicinityTRUE') +
             prior(normal(0.1,0.05),class='b',coef='pct_pop_elder') +
             prior(normal(0.1,0.05),class='b',coef='logdensity') +
             prior(normal(0.1,0.05),class='b',coef='transit_stations_change') +
             prior(normal(0.1,0.05),class='b',coef='grocery_pharmacy_change') +
             prior(normal(0.1,0.05),class='b',coef='parks_change') +
             prior(normal(0.1,0.1),class='Intercept') +
             prior(lkj(2),class='cor') +
             prior(normal(0,0.1),class='sd'))

# diagnostics
loo3 <- loo(post3)

# model comparison
loo_compare(loo.trans,loo3)
```

```
##            elpd_diff se_diff
## post3       0.0       0.0
## post.trans -5.8       5.8
```

According to the ELPD, the model with all three county-specific mobility effects is not significantly better than the one with only transportation effect.

Final model

After examining the above models, the model including population-level census, grocery & pharmacy mobility and parks mobility effects as well as county-specific transportation mobility effect seems to be the best.

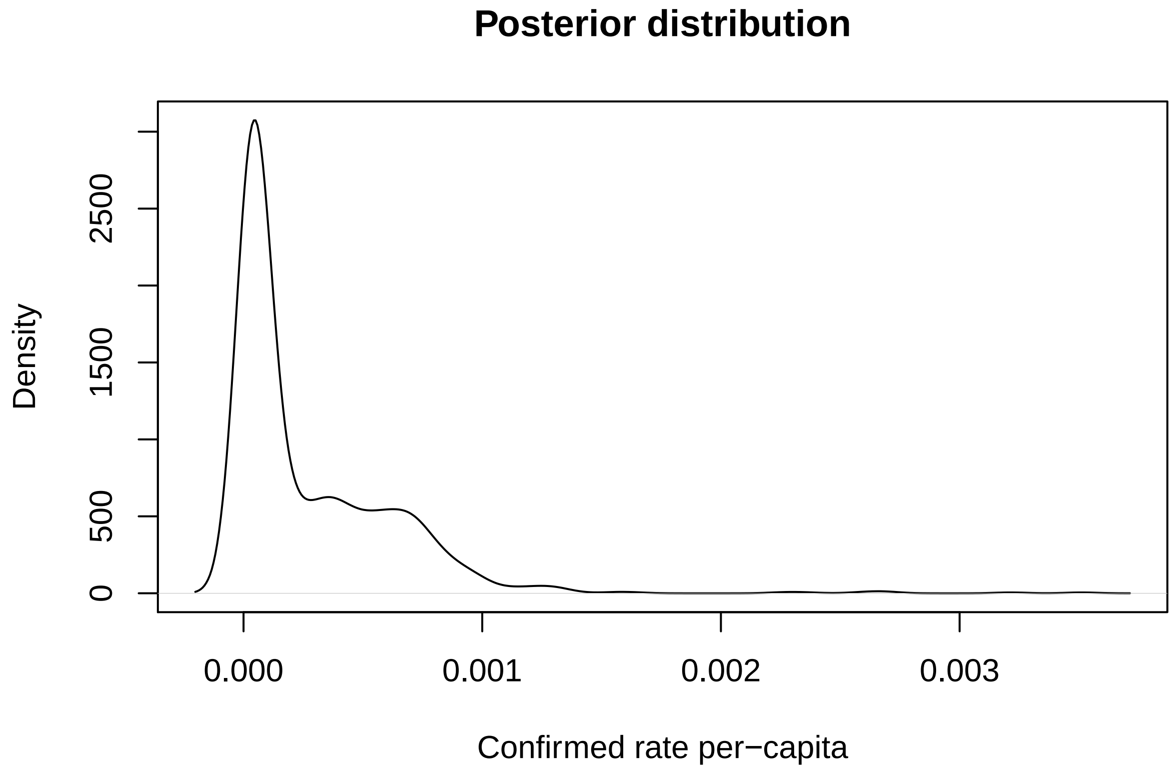
Posterior Distribution

```
post.trans
```

```
##  Family: negbinomial
##   Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + transit_stations_change + grocery_pharmacy_change + parks_change
##    Data: dat (Number of observations: 1134)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 26)
##                                         Estimate Est.Error l-95% CI u-95% CI
## sd(Intercept)                               0.42      0.07     0.28     0.56
## sd(polyday21)                               0.08      0.06     0.00     0.23
## sd(polyday22)                               0.09      0.06     0.00     0.24
## sd(transit_stations_change)                 0.19      0.02     0.15     0.24
## cor(Intercept,polyday21)                   -0.00      0.38    -0.71     0.70
## cor(Intercept,polyday22)                    0.04      0.37    -0.65     0.71
## cor(polyday21,polyday22)                    0.00      0.37    -0.69     0.71
## cor(Intercept,transit_stations_change)     -0.03      0.20    -0.42     0.35
## cor(polyday21,transit_stations_change)      0.05      0.38    -0.70     0.73
## cor(polyday22,transit_stations_change)     -0.10      0.38    -0.77     0.67
##                                         Rhat Bulk_ESS Tail_ESS
## sd(Intercept)                           1.00     1592     2041
## sd(polyday21)                           1.00     2111     1396
## sd(polyday22)                           1.00     2221     1520
## sd(transit_stations_change)             1.00     1340     1832
## cor(Intercept,polyday21)                1.00     3249     2295
## cor(Intercept,polyday22)                1.00     3703     2773
## cor(polyday21,polyday22)                1.00     3070     2753
## cor(Intercept,transit_stations_change) 1.00      486     1041
## cor(polyday21,transit_stations_change) 1.01      182      512
## cor(polyday22,transit_stations_change) 1.02      222      504
##
## Population-Level Effects:
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                 -14.08      0.99   -16.00   -12.18 1.00     2239
## polyday21                   0.25      0.10     0.05     0.45 1.00     3744
## polyday22                  -0.12      0.05    -0.21    -0.02 1.00     3361
## transit_stations_change    -0.26      0.00    -0.26    -0.25 1.00      824
## grocery_pharmacy_change    -0.00      0.00    -0.00     0.00 1.00     5664
## parks_change                0.01      0.00     0.00     0.01 1.00     4584
## nyc_vicinityTRUE            0.08      0.05    -0.02     0.17 1.00     3438
## pct_pop_elder               0.08      0.04     0.00     0.15 1.00     2536
```

```
## logdensity                  -0.03      0.04    -0.11     0.05 1.00     1925
##                          Tail_ESS
## Intercept                    2798
## polyday21                    2878
## polyday22                    3085
## transit_stations_change      1834
## grocery_pharmacy_change      2996
## parks_change                 3349
## nyc_vicinityTRUE             2965
## pct_pop_elder                2881
## logdensity                   2781
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape     2.30      0.12     2.07     2.53 1.00     3087     2686
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

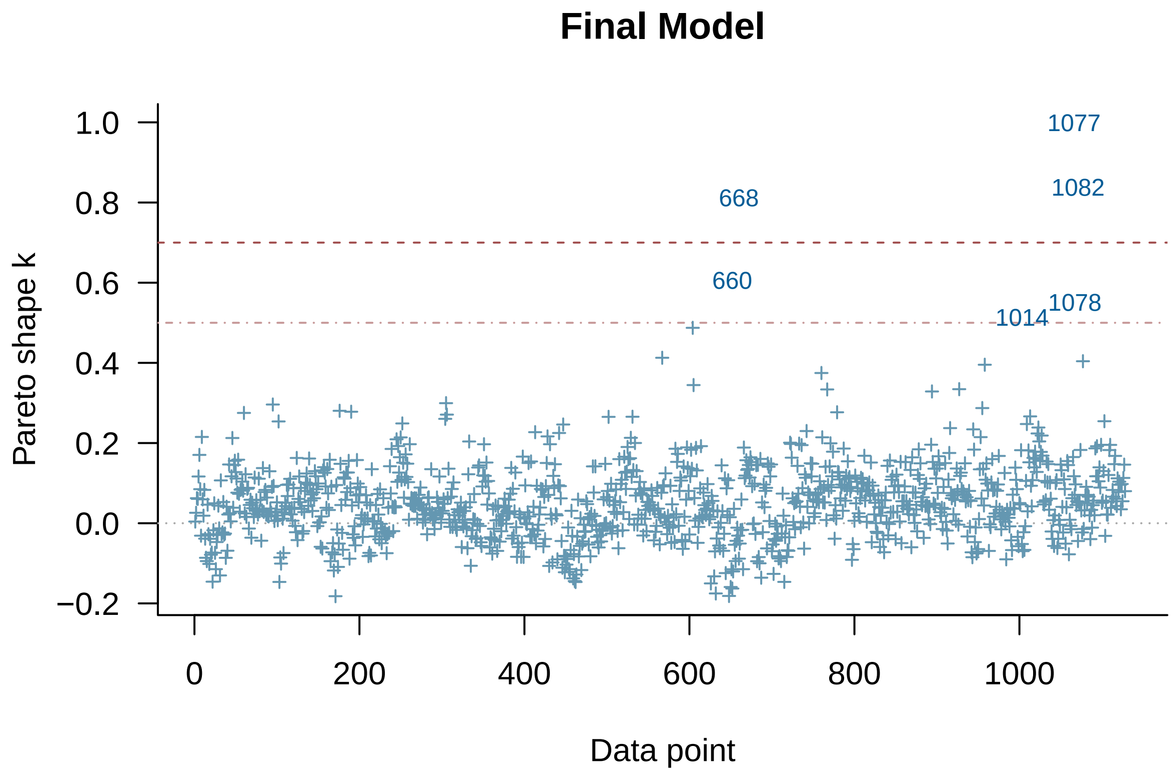
```
mu.trans <- pp_expect(post.trans)

plot(density(colMeans(mu.trans)/dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_c
```



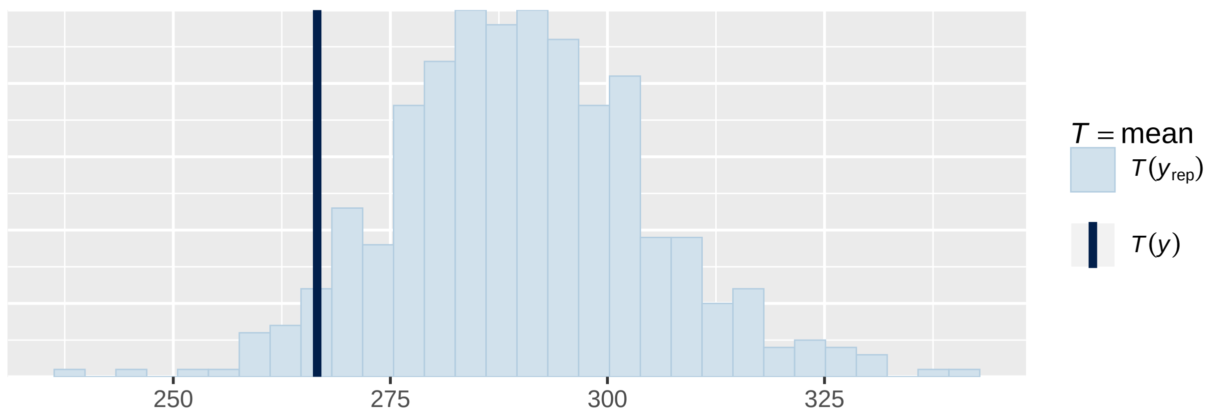
With county-specific transportation mobility trends, the range of posterior distribution is smaller. Probably the effect of some highly influential counties like NYC is smoothened by the group-specific effect.

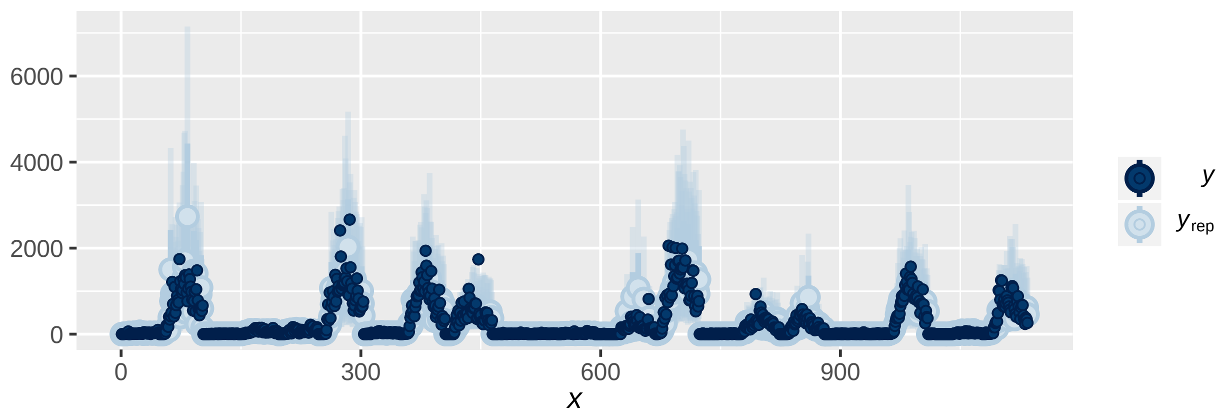
Diagnostics

```
plot(loo.trans,main='Final Model',label_points=T)
```




```
grid.arrange(
pp_check(post.trans,nsample=500,type='stat',stat='mean'),
pp_check(post.trans,nsample=500,type='intervals'),
nrow=2)
```





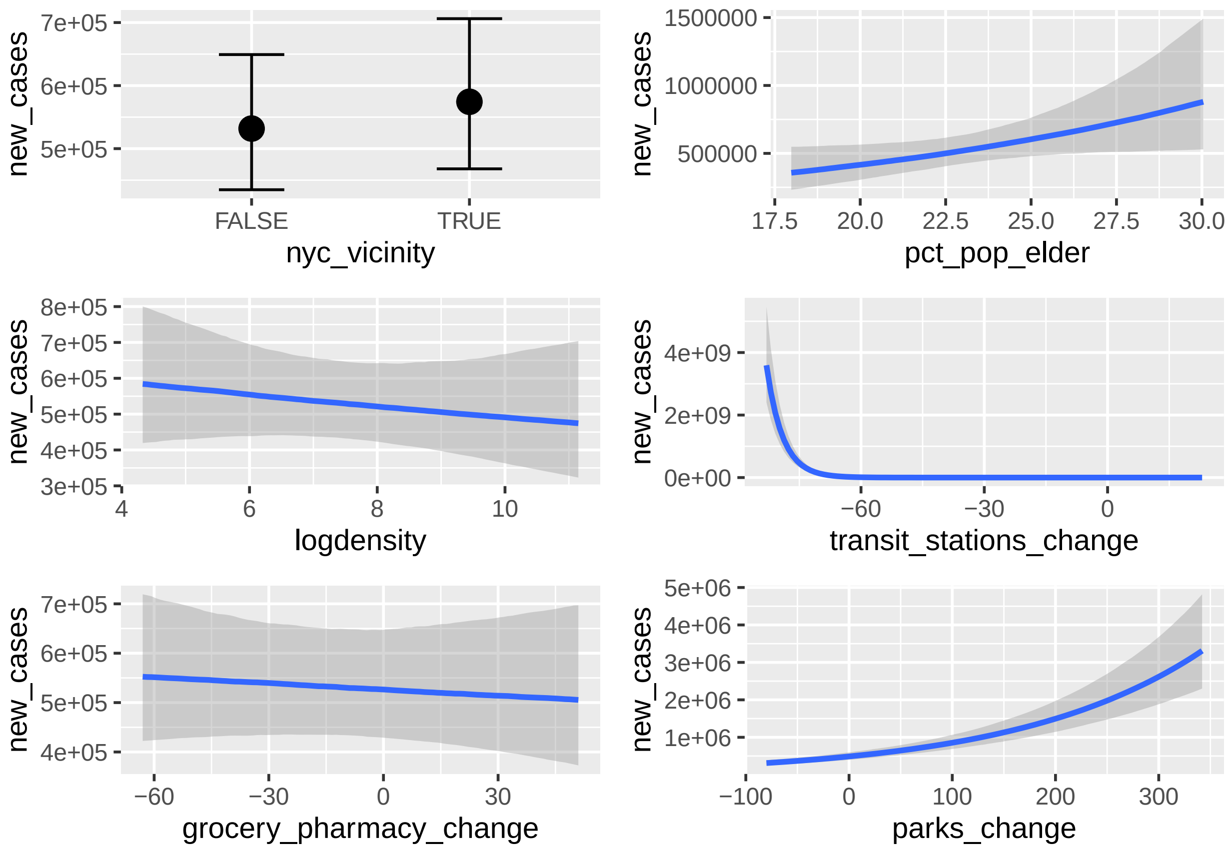

```
# numerical assessment of calibration
PPD.trans <- posterior_predict(post.trans)
lower.trans <- apply(PPD.trans,2,quantile,probs=.025)
upper.trans <- apply(PPD.trans,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_change,parks_change,nyc_vici
```

```
## The probability of observations fall into 95% of the posterior distribution is 0.9012346
```

Three highly influenced observations were detected. Eyeballing the loo-based plots, the final model overestimates the mean.

Conditional Effects

```
# conditional effects
con.t1 <- plot(conditional_effects(post.trans,'nyc_vicinity'),plot=F,ask=F)$nyc_vicinity
con.t2 <- plot(conditional_effects(post.trans,'pct_pop_elder'),plot=F,ask=F)$pct_pop_elder
con.t3 <- plot(conditional_effects(post.trans,'logdensity'),plot=F,ask=F)$logdensity
con.t4 <- plot(conditional_effects(post.trans,'transit_stations_change'),plot=F,ask=F)$transit_stations_
con.t5 <- plot(conditional_effects(post.trans,'grocery_pharmacy_change'),plot=F,ask=F)$grocery_pharmacy_
con.t6 <- plot(conditional_effects(post.trans,'parks_change'),plot=F,ask=F)$parks_change
grid.arrange(con.t1,con.t2,con.t3,con.t4,con.t5,con.t6,nrow=3,ncol=2)
```



Unexpectedly, the conditional effect of population density turns negative after including county-specific transportation mobility effect. The mobility change effects of transportation and grocery are negatively associated with the daily new cases. It might be because NYC is having a large effect on the results: people in NYC take much less public transportation under the quarantine policy, and the population density in NYC is very high but it also had the strongest policy reaction.

Conclusion

To sum up, besides time, census data reflects population features of the region and could be used to explain the growth trend of COVID-19 pandemic. Mobility can also be influential factors to the spread of pandemic. According to our negative binomial hierarchical models, whether in vicinity of NYC, percentage of elderly, population density, transportation mobility change, grocery & pharmacy mobility change, and parks mobility change are associated with the growth rate of COVID-19. Specifically, time effects and transportation mobility effects differ across counties. Without conditioning on mobility effects, holding everything else constant, the marginal effects of whether in vicinity of NYC, percentage of elderly, and population density are all positively associated with the daily new confirmed cases, as expected. However, after taking county-specific transportation mobility factor into account, the effect of log density turns negative, probably due to highly influential counties like NYC.