

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 \tilde{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 :

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

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

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.

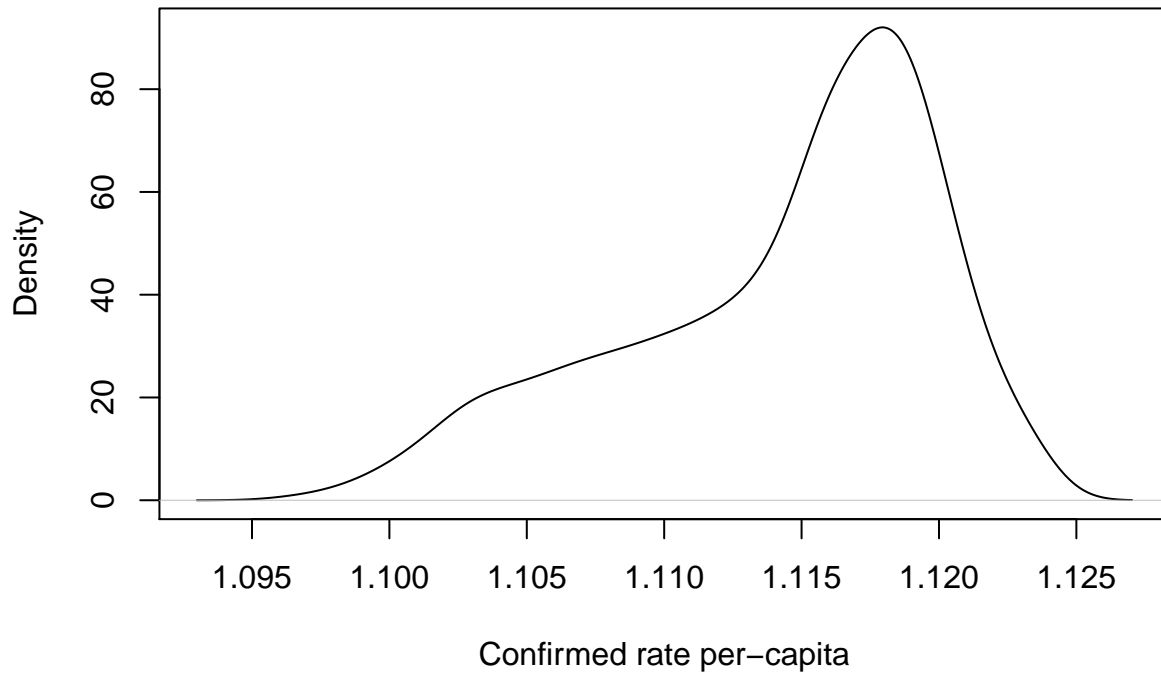
Prior Predictive Distribution

```
draws0 <- brm(new_cases ~ poly(day, 2) + (poly(day, 2) | county_fips) + offset(logpop), data=dat, family=negbinomial,
  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 predictive')
```

Prior predictive distribution



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

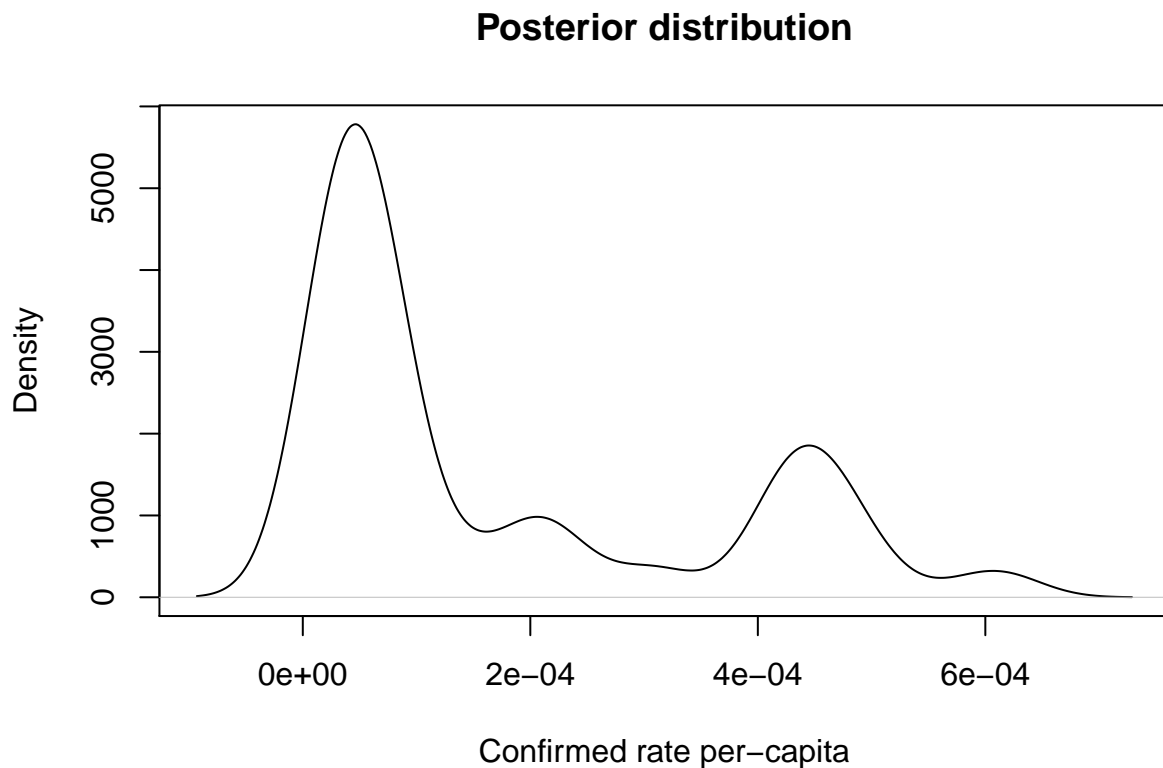
## Family: negbinomial
## Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + (poly(day, 2) | county_fips) + offset(logpop)
## Data: dat (Number of observations: 2066)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##           total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 38)
##
```

	Estimate	Est.Error	1-95% CI	u-95% CI	Rhat	Bulk_ESS
## sd(Intercept)	2.29	0.05	2.19	2.39	1.01	232
## sd(polyday21)	0.18	0.11	0.01	0.41	1.00	1500
## sd(polyday22)	0.31	0.12	0.07	0.53	1.00	2057
## cor(Intercept,polyday21)	-0.53	0.34	-0.94	0.38	1.00	2128
## cor(Intercept,polyday22)	0.77	0.18	0.25	0.97	1.00	2585
## cor(polyday21,polyday22)	-0.42	0.36	-0.91	0.46	1.00	2163
##						
	Tail_ESS					

```
## sd(Intercept)          476
## sd(polyday21)          809
## sd(polyday22)          981
## cor(Intercept,polyday21) 1980
## cor(Intercept,polyday22) 1758
## cor(polyday21,polyday22) 2059
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    -0.52     0.10   -0.72   -0.33 1.00    1986    2833
## polyday21     0.28     0.10    0.08    0.48 1.00    4700    2862
## polyday22    -0.13     0.05   -0.23   -0.03 1.00    4806    2948
##
## Family Specific Parameters:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape      1.01     0.03    0.95    1.08 1.00    3852    2881
##
## 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 distribution')
```



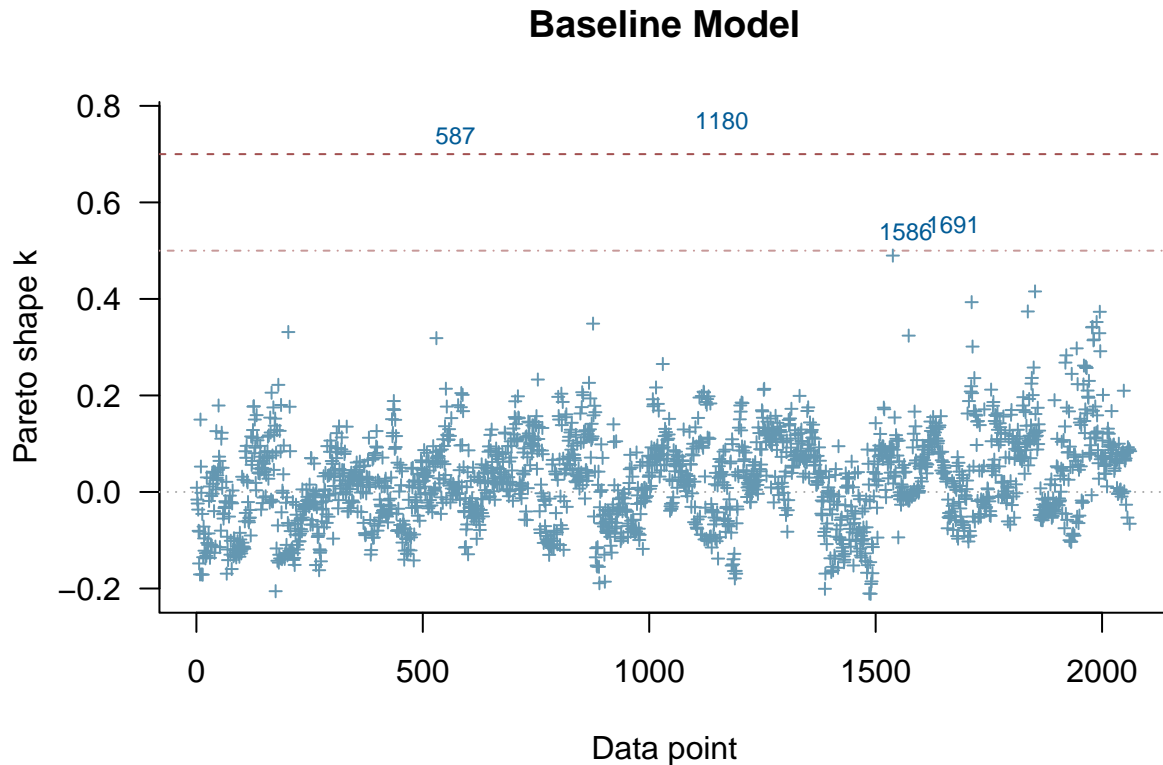
After controlling for the observed data, we get a more reasonable posterior distribution. Most of the counties have confirmed rate at around 0.01%, while some more vulnerable counties have higher rate at around 0.05%.

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.

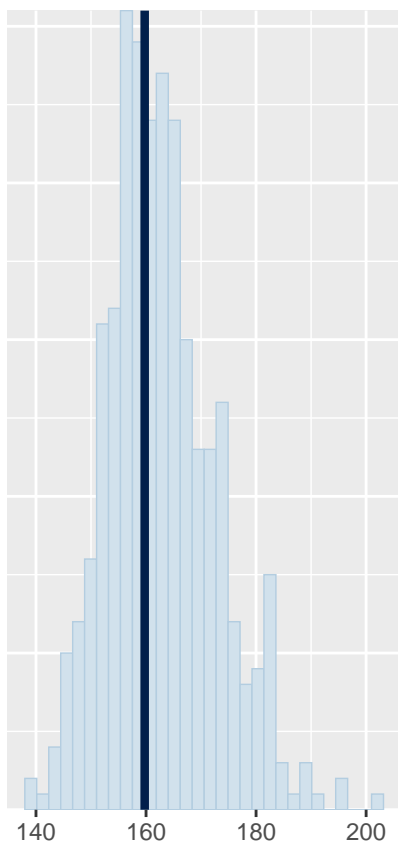
```
loo0 <- loo(post0)
plot(loo0,main="Baseline Model",label_points=T)
```



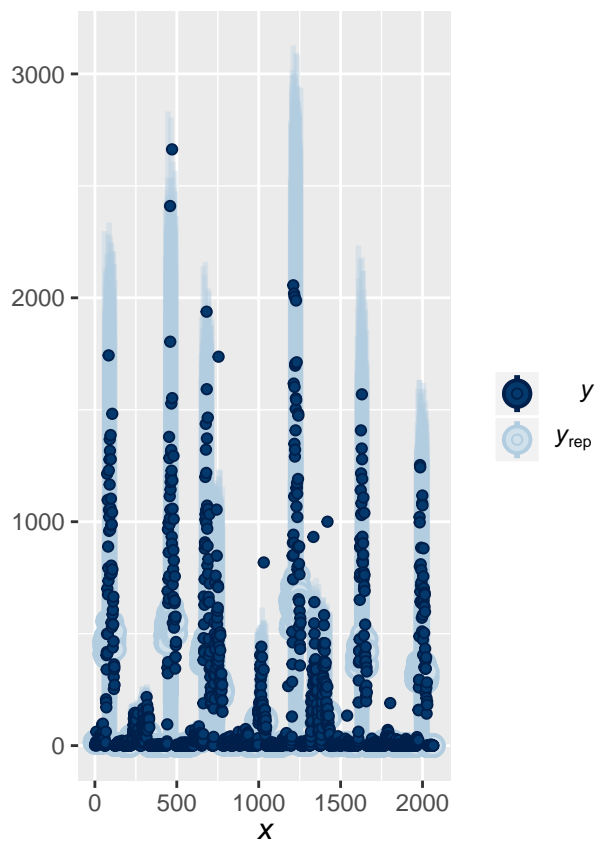
```
# numerical assessment of calibration
PPD0 <- posterior_predict(post0)
lower0 <- apply(PPD0,2,quantile,probs=.025)
upper0 <- apply(PPD0,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat$new_cases>lower0&dat$new_cases<upper0))
```

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

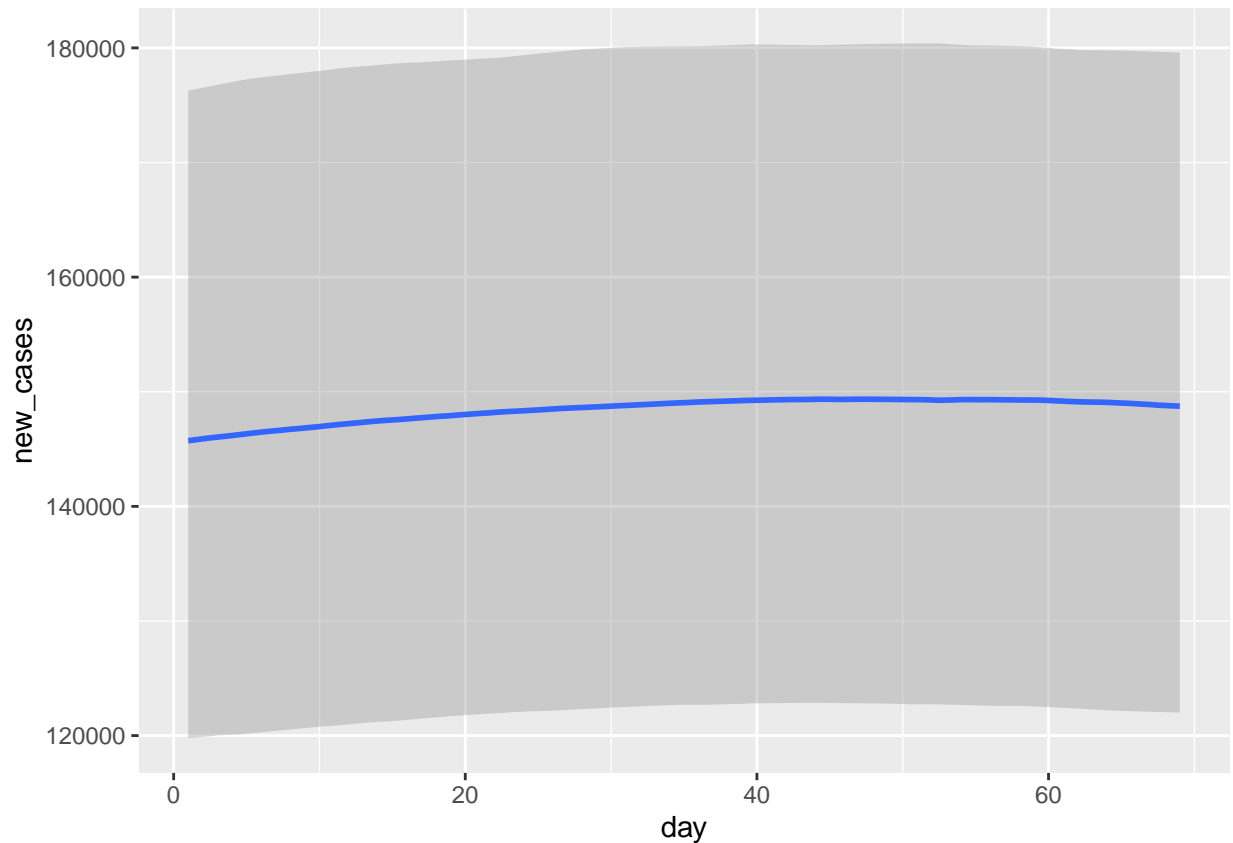
```
gridExtra::grid.arrange(
pp_check(post0,nsample=500,type='stat',stat='mean'),
pp_check(post0,nsample=500,type='intervals'),
ncol=2,nrow=1)
```



$T = \text{mean}$
 $T(y_{\text{rep}})$
 $T(y)$



```
# conditional effect
conditional_effects(post0)
```



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.

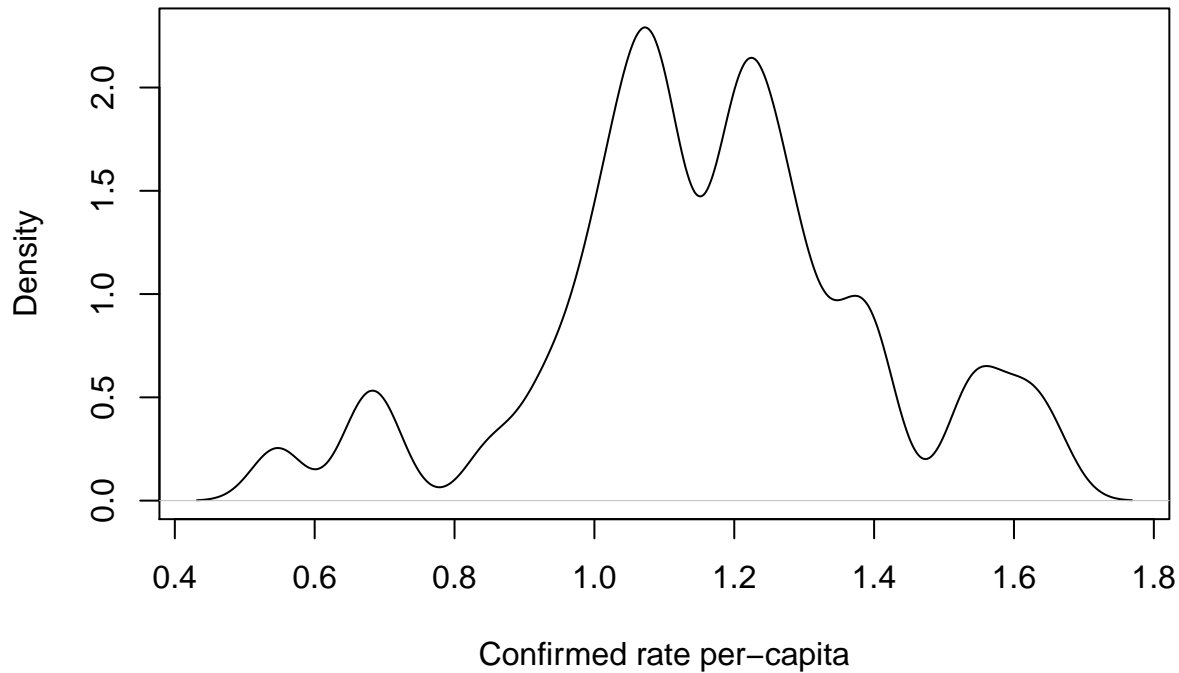
Prior Predictive Distribution

```
draws1 <- brm(new_cases~poly(day,2)+nyc_vicinity+pct_pop_elder+logdensity+(poly(day,2)|county_fips)+offset,
  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 predictive')
```

Prior predictive distribution



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
```

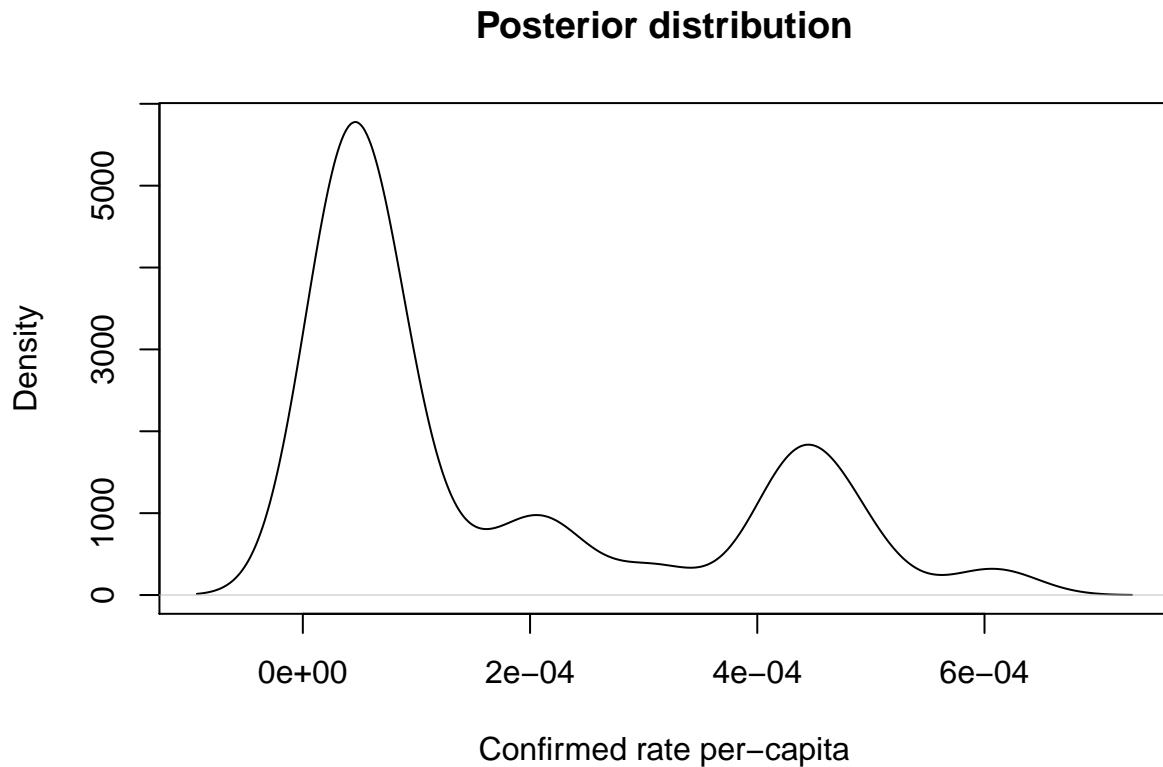
```
## Family: negbinomial
## Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + nyc_vicinity + pct_pop_elder + logdensity + (poly(day, 2) | county_fips)
## Data: dat (Number of observations: 2066)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##           total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 38)
##
```

	Estimate	Est.Error	1-95% CI	u-95% CI	Rhat	Bulk_ESS
## sd(Intercept)	2.29	0.05	2.19	2.40	1.01	291
## sd(polyday21)	0.19	0.11	0.01	0.40	1.00	1954
## sd(polyday22)	0.32	0.11	0.09	0.54	1.00	2044
## cor(Intercept,polyday21)	-0.56	0.32	-0.94	0.31	1.00	2885
## cor(Intercept,polyday22)	0.79	0.18	0.35	0.97	1.00	2649
## cor(polyday21,polyday22)	-0.45	0.35	-0.91	0.44	1.00	2766
##	Tail_ESS					
## sd(Intercept)	497					


```
## sd(polyday21)                1496
## sd(polyday22)                905
## cor(Intercept,polyday21)    2349
## cor(Intercept,polyday22)    1624
## cor(polyday21,polyday22)    2283
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      -2.57      1.19   -4.86   -0.18 1.00      781    1472
## polyday21       0.28      0.10    0.08    0.47 1.00     4880    3125
## polyday22      -0.13      0.05   -0.23   -0.03 1.00     4718    3110
## nyc_vicinityTRUE 0.11      0.05    0.01    0.21 1.00     3214    2694
## pct_pop_elder   0.04      0.05   -0.05    0.13 1.01      598    1208
## logdensity      0.15      0.05    0.06    0.25 1.00     1104    2233
##
## Family Specific Parameters:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape      1.01      0.04    0.94    1.08 1.00      3904     3043
##
## 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).
```

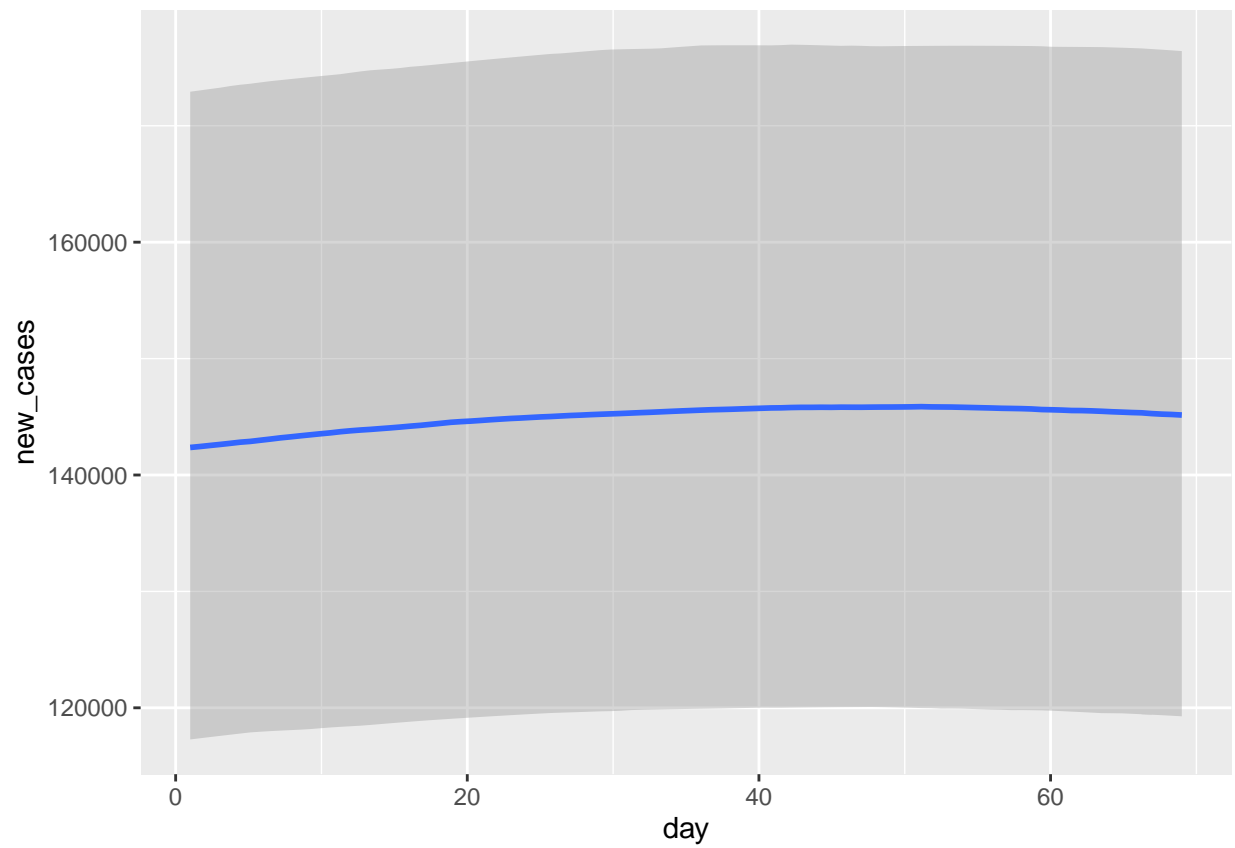
```
mu1.post <- pp_expect(post1)
```

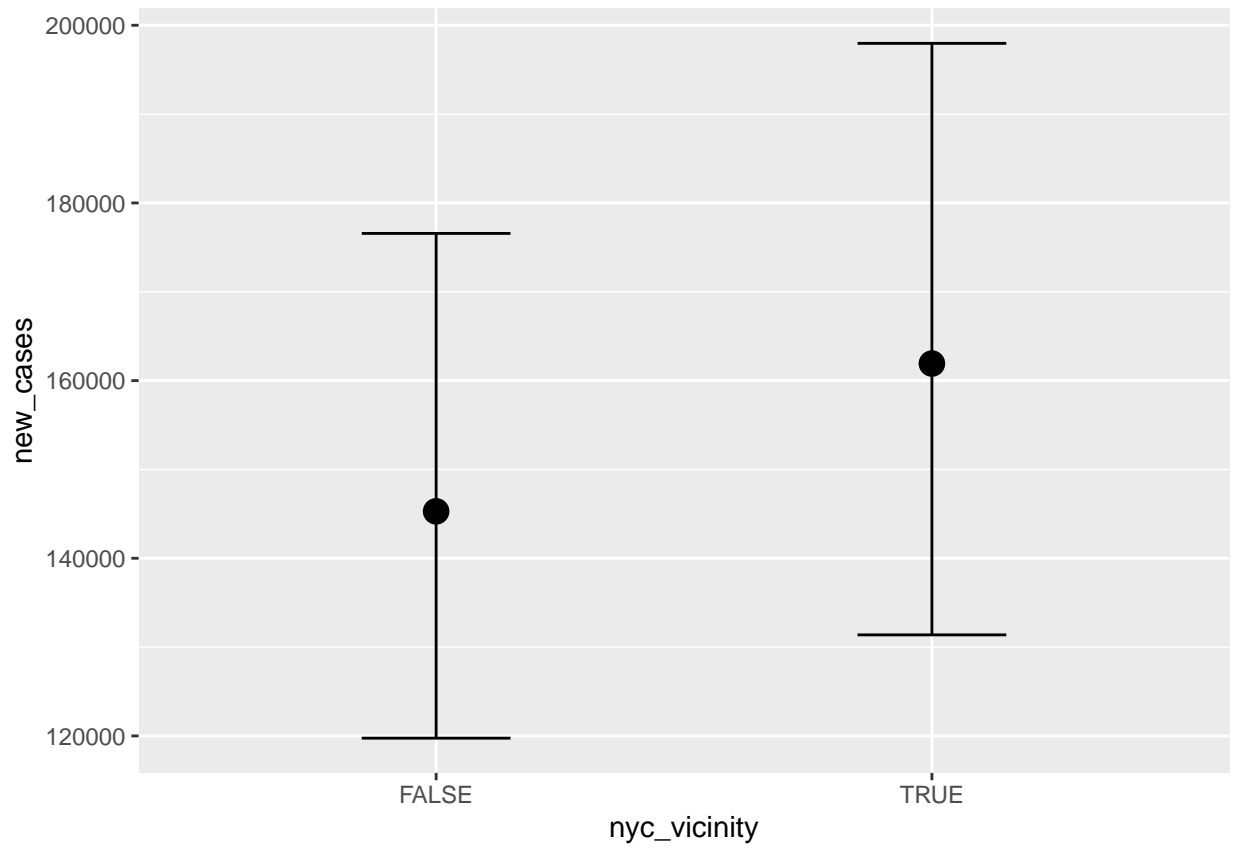
```
plot(density(colMeans(mu1.post)/dat$county_pop),xlab='Confirmed rate per-capita',main='Posterior distrib
```

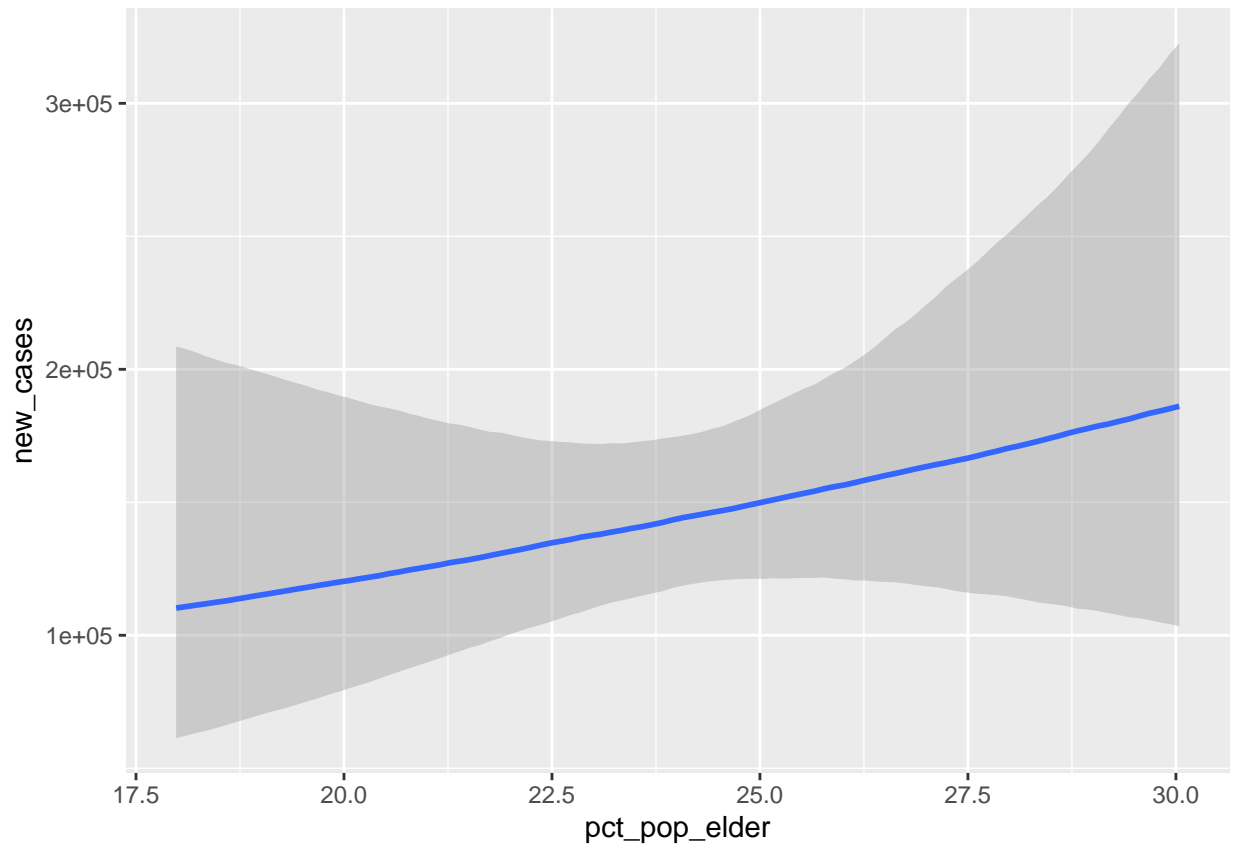


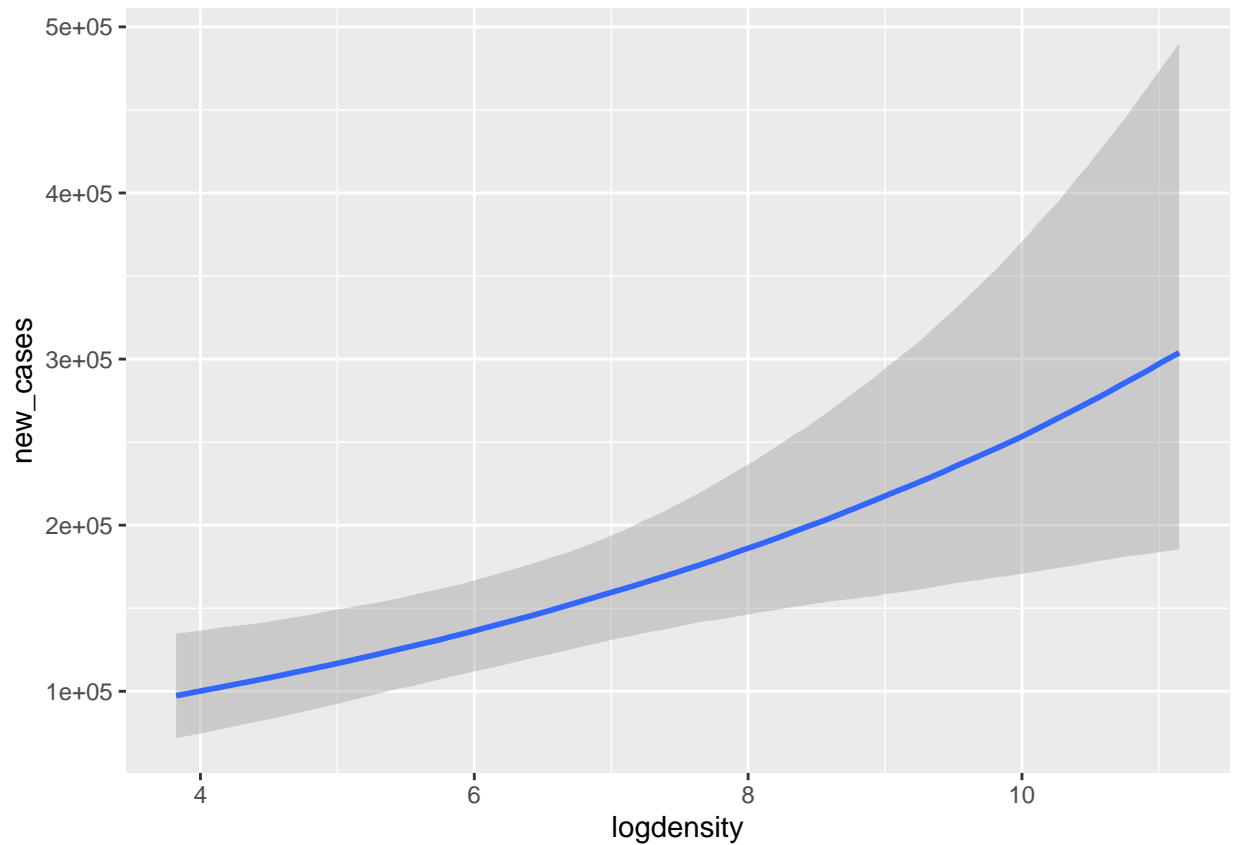
The posterior distribution seems reasonable after conditioning on observed data. It looks quite similar to that of the baseline model.

```
plot(conditional_effects(post1),ask=F)
```







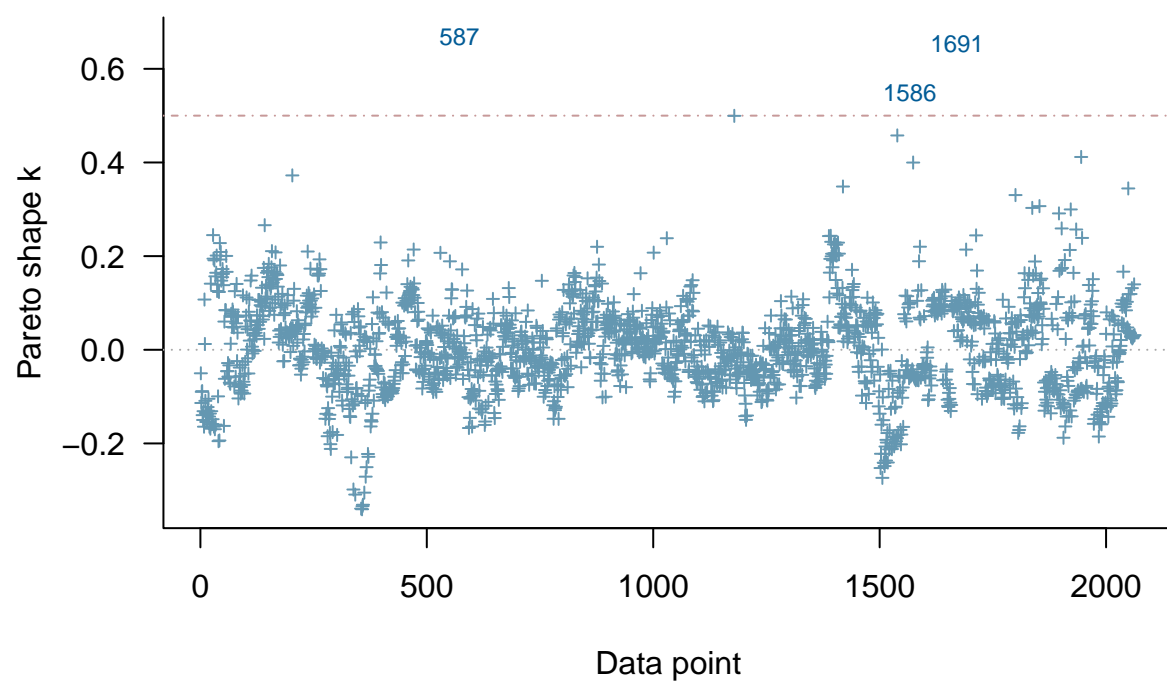


Holding everything else constant, we can see that the marginal effects of whether in vicinity of NYC, percentage of elderly, and population density are all positively associated with the daily new confirmed cases, as expected. The effect of the percentage of elderly is negligible with high uncertainty.

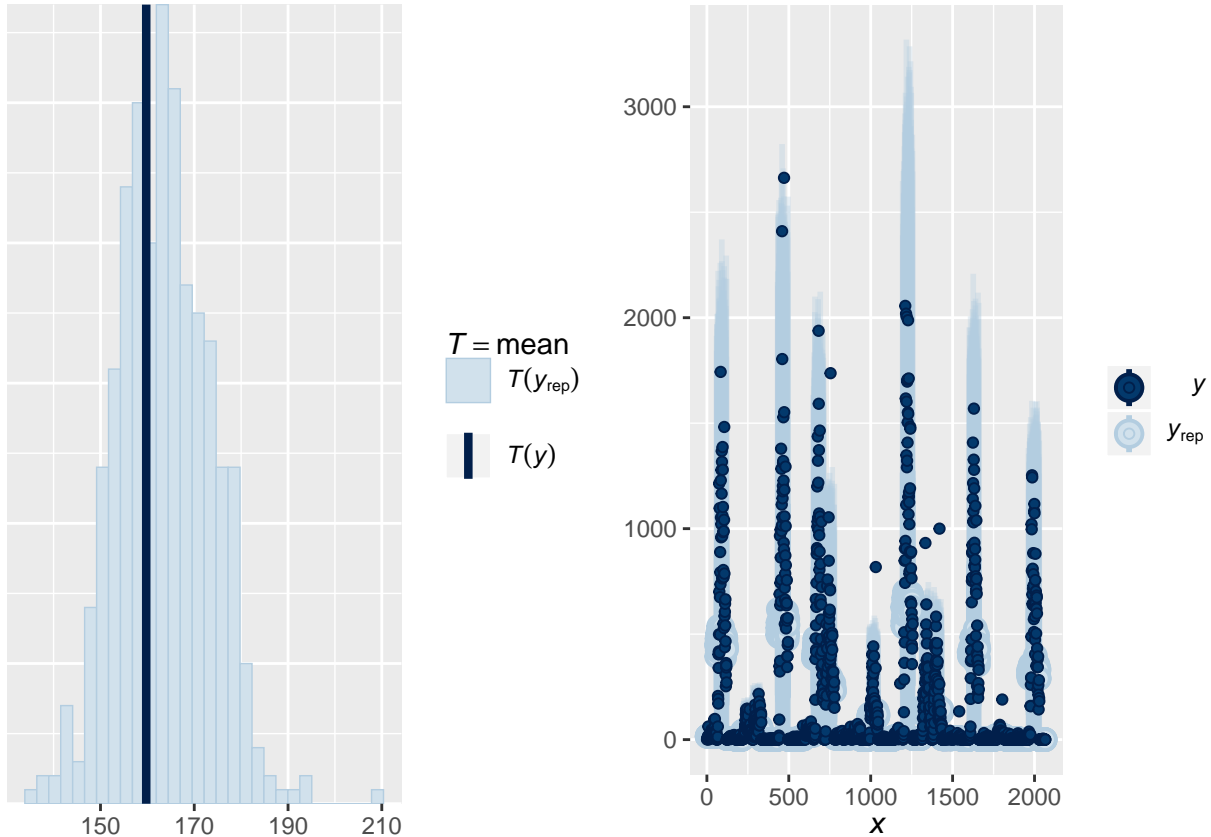
Diagnostics

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

W/ Census Covariates



```
gridExtra::grid.arrange(  
  pp_check(post1, nsample=500, type='stat', stat='mean'),  
  pp_check(post1, nsample=500, type='intervals'),  
  ncol=2, nrow=1)
```



```
# 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 0.8262343
```

```
# model comparison
```

```
loo_compare(loo0,loo1)
```

```
##      elpd_diff se_diff
## post1  0.0      0.0
## post0 -1.2      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.

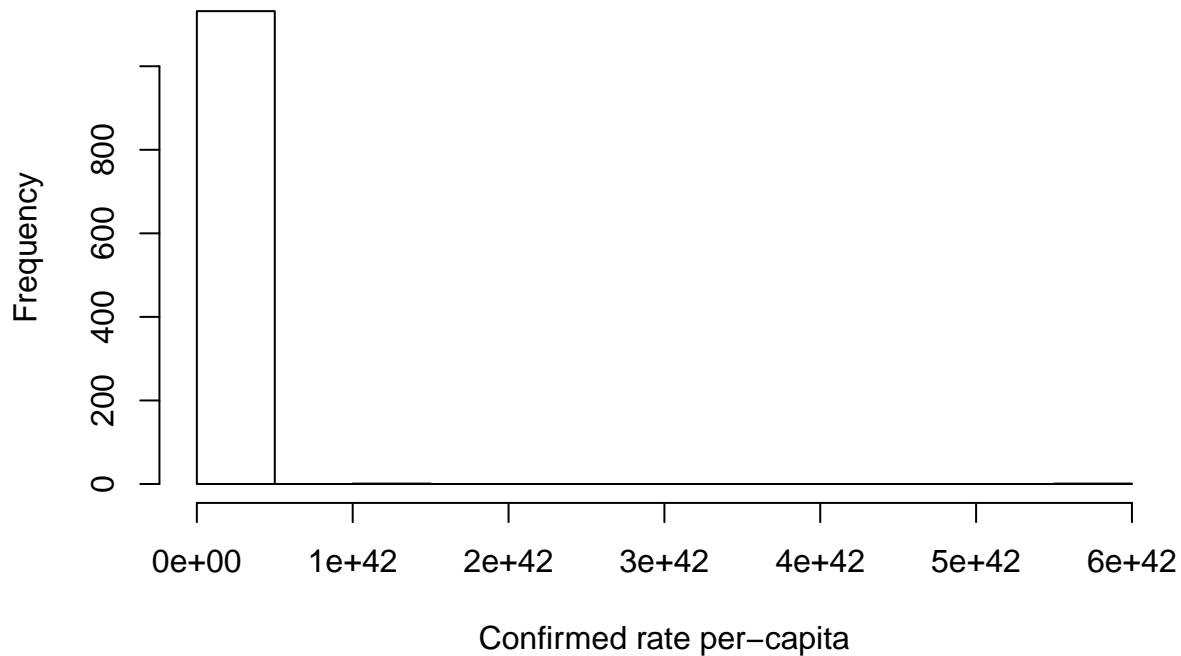
Prior Predictive Distribution

```
draws2 <- brm(new_cases~poly(day,2)+transit_stations_change+grocery_pharmacy_change+parks_change+nyc_vicinityTRUE,
  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
```

Prior predictive distribution



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.

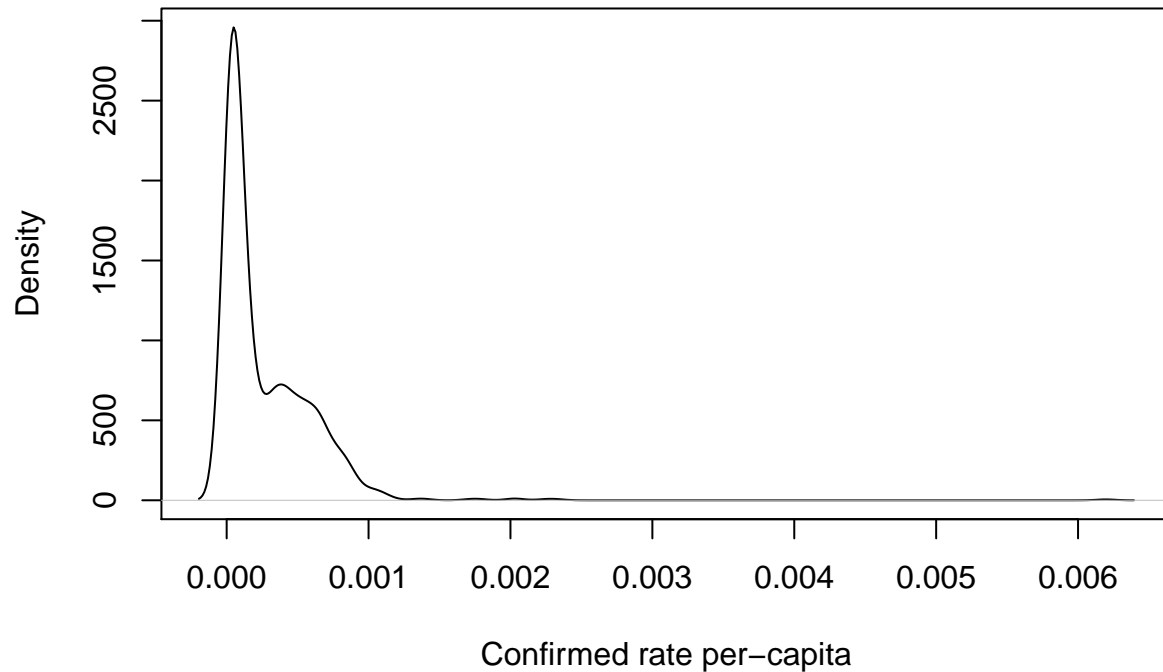

```
post2 <- update(draws2,sample_prior='no')
post2
```

```
## 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 1-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)          2.07      0.05    1.98    2.17 1.01      541
## sd(polyday21)           0.14      0.09    0.01    0.34 1.00     2666
## sd(polyday22)           0.20      0.10    0.01    0.40 1.00     2174
## cor(Intercept,polyday21) -0.40      0.37   -0.90    0.49 1.00     5770
## cor(Intercept,polyday22)  0.58      0.31   -0.25    0.94 1.00     4238
## cor(polyday21,polyday22) -0.24      0.40   -0.87    0.60 1.00     4527
##           Tail_ESS
## sd(Intercept)          1176
## sd(polyday21)           2255
## sd(polyday22)           1517
## cor(Intercept,polyday21)  2985
## cor(Intercept,polyday22)  2174
## cor(polyday21,polyday22)  2585
##
## Population-Level Effects:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
## Intercept             -6.05      1.15   -8.37   -3.88 1.00     2546
## polyday21               0.26      0.10    0.06    0.46 1.00     7573
## polyday22              -0.12      0.05   -0.22   -0.02 1.00     9064
## transit_stations_change -0.07      0.00   -0.08   -0.07 1.00     5807
## grocery_pharmacy_change -0.00      0.00   -0.01    0.00 1.00     6381
## parks_change            0.01      0.00    0.00    0.01 1.00     3854
## nyc_vicinityTRUE         0.11      0.05    0.01    0.21 1.00     8220
## pct_pop_elder            0.04      0.05   -0.05    0.13 1.00     1982
## logdensity              0.15      0.05    0.06    0.25 1.00     3342
##           Tail_ESS
## Intercept              2651
## polyday21              2622
## polyday22              2907
## transit_stations_change 3502
## grocery_pharmacy_change 3645
## parks_change           3550
## nyc_vicinityTRUE        3162
## pct_pop_elder           2526
## logdensity              3097
##
## Family Specific Parameters:
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape          2.00      0.10    1.81    2.19 1.00     6204     3004
##
## 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)

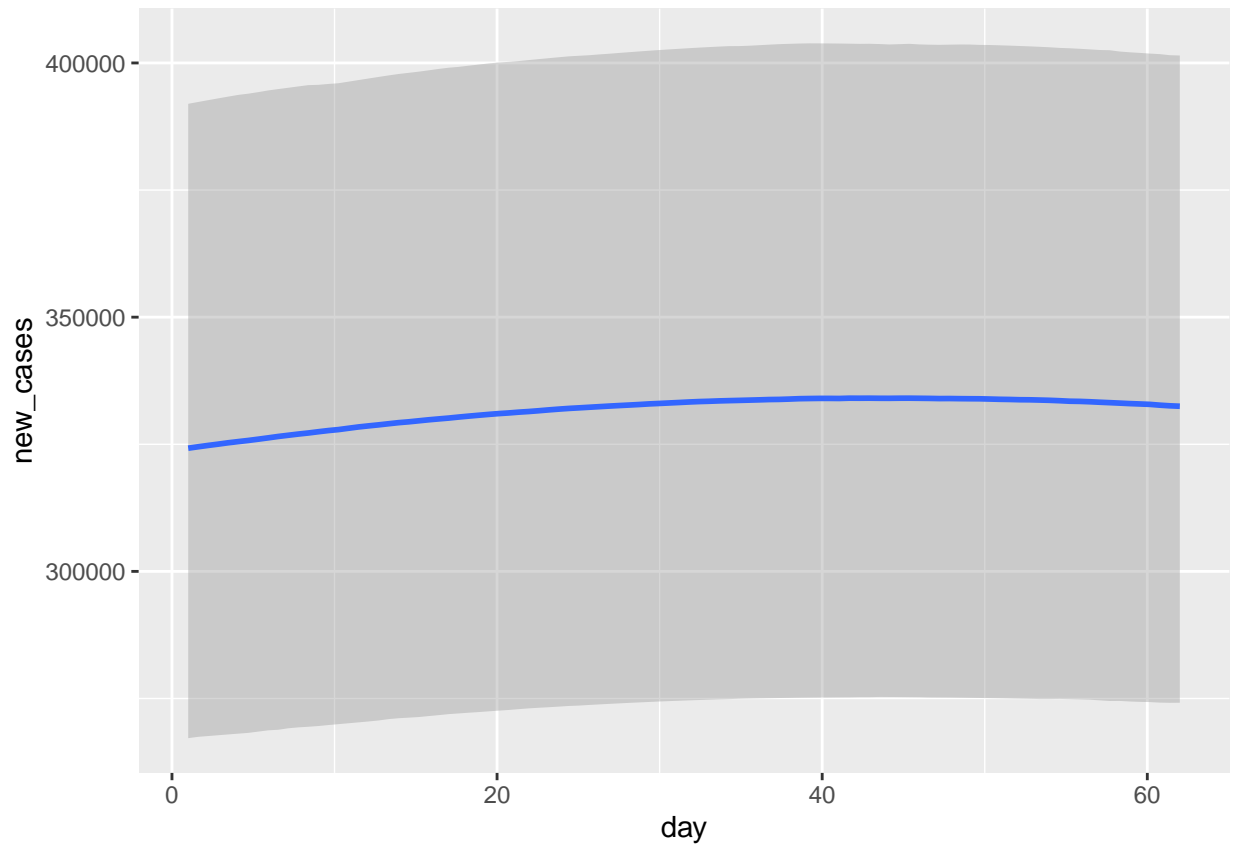
plot(density(colMeans(mu2.post))/dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_
```

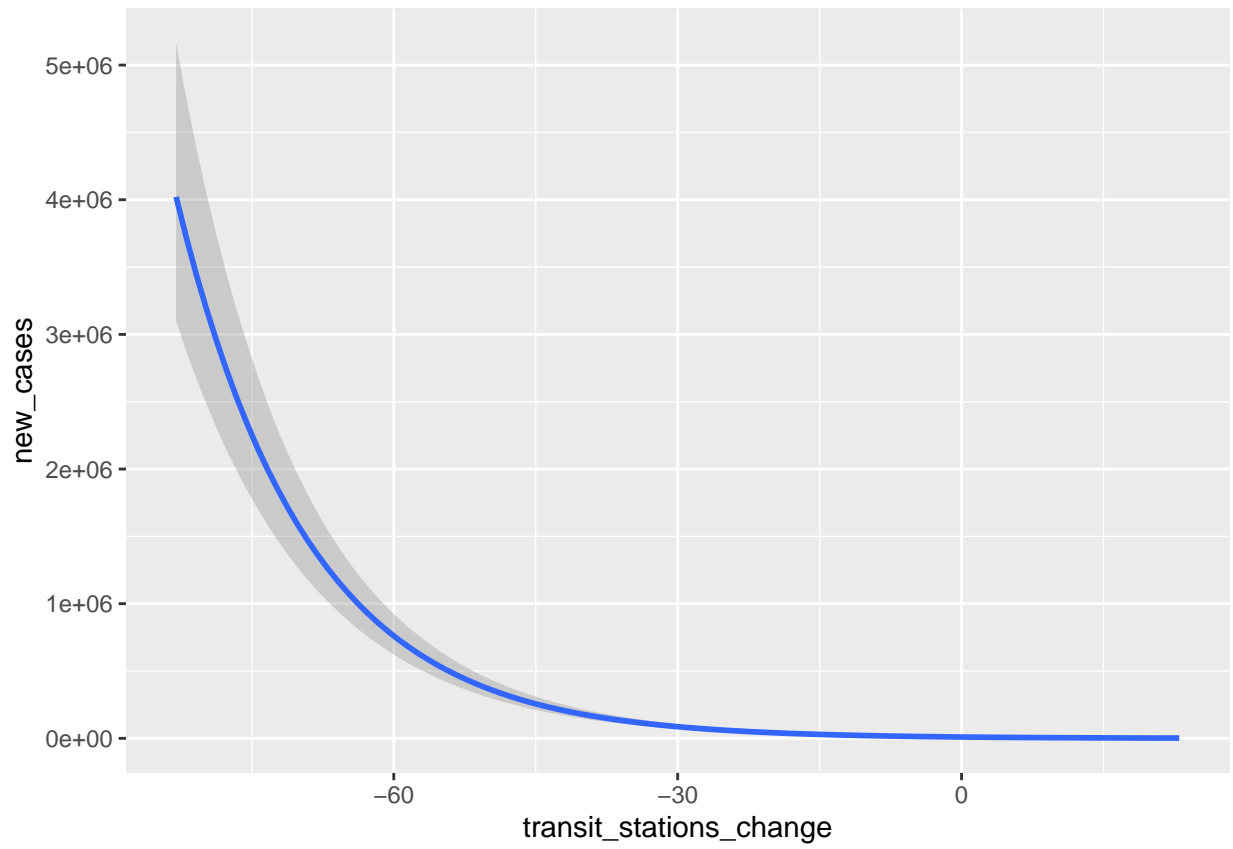
Posterior distribution

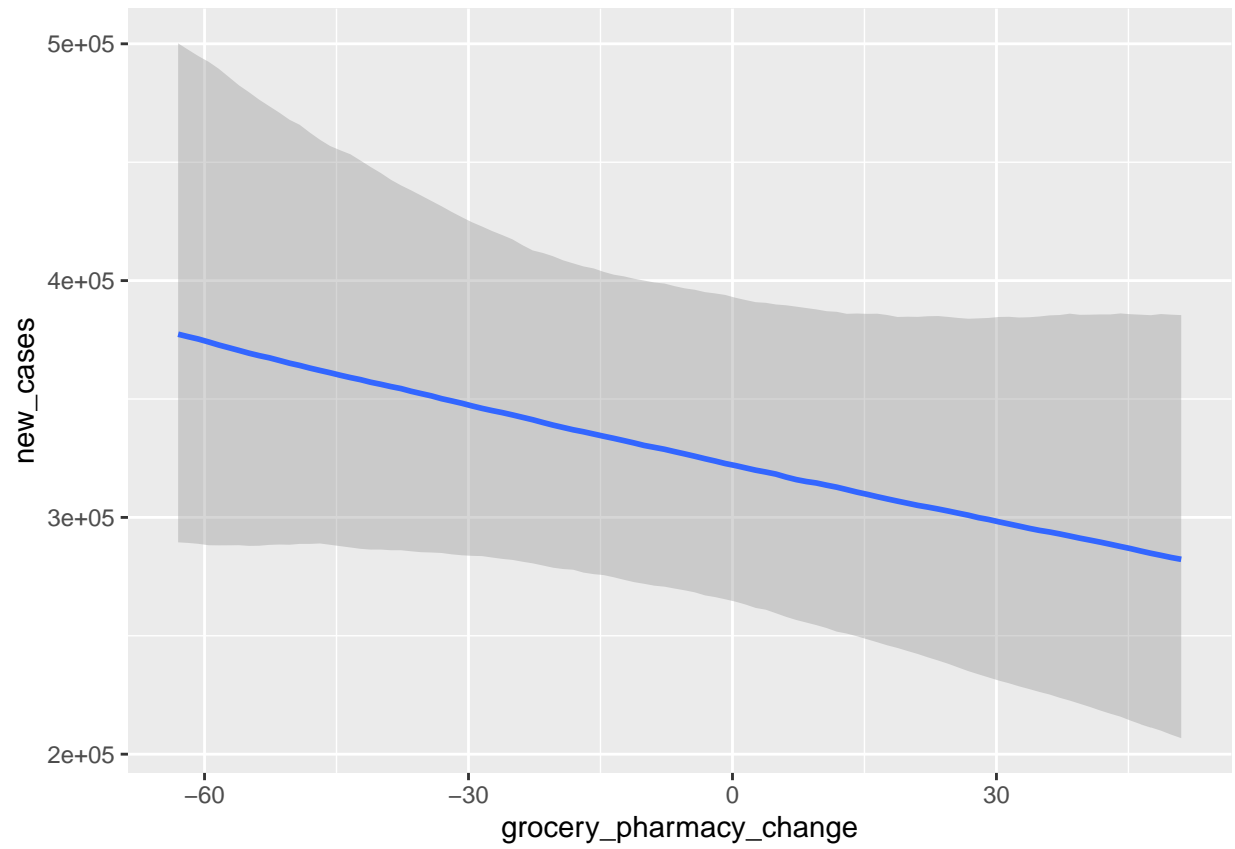


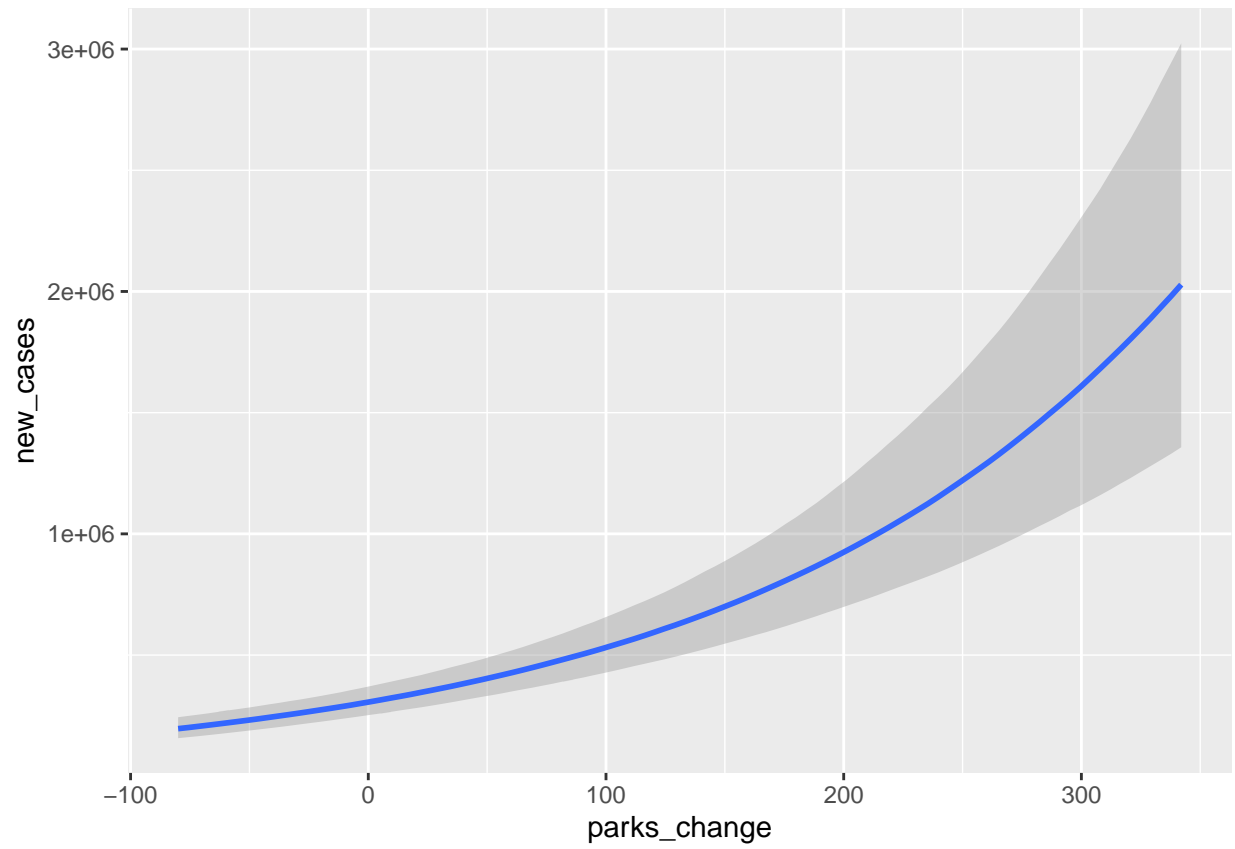
Conditioning on observed data, the posterior distribution is largely narrowed down to a smaller range. Most counties have the confirmed rate less than 1%, with several counties more than 6%.

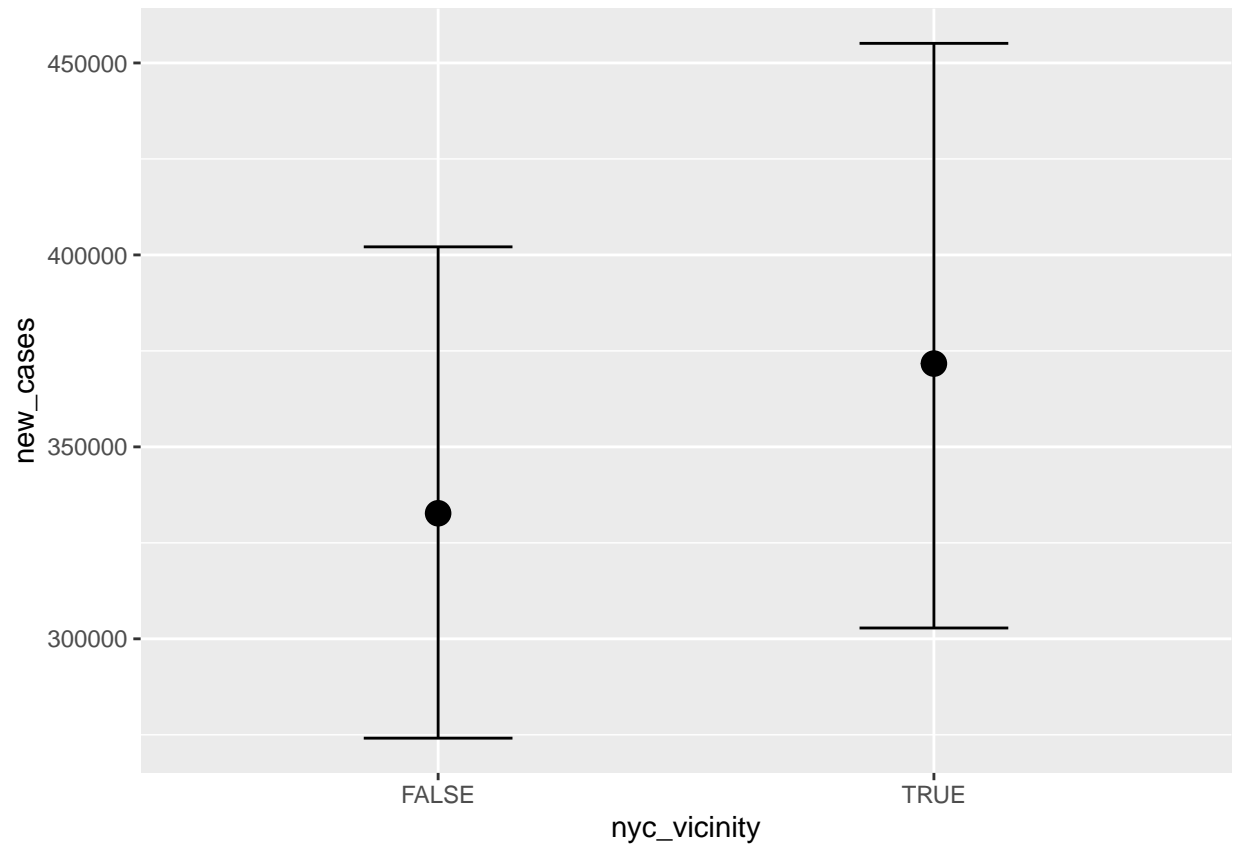
```
plot(conditional_effects(post2),ask=F)
```

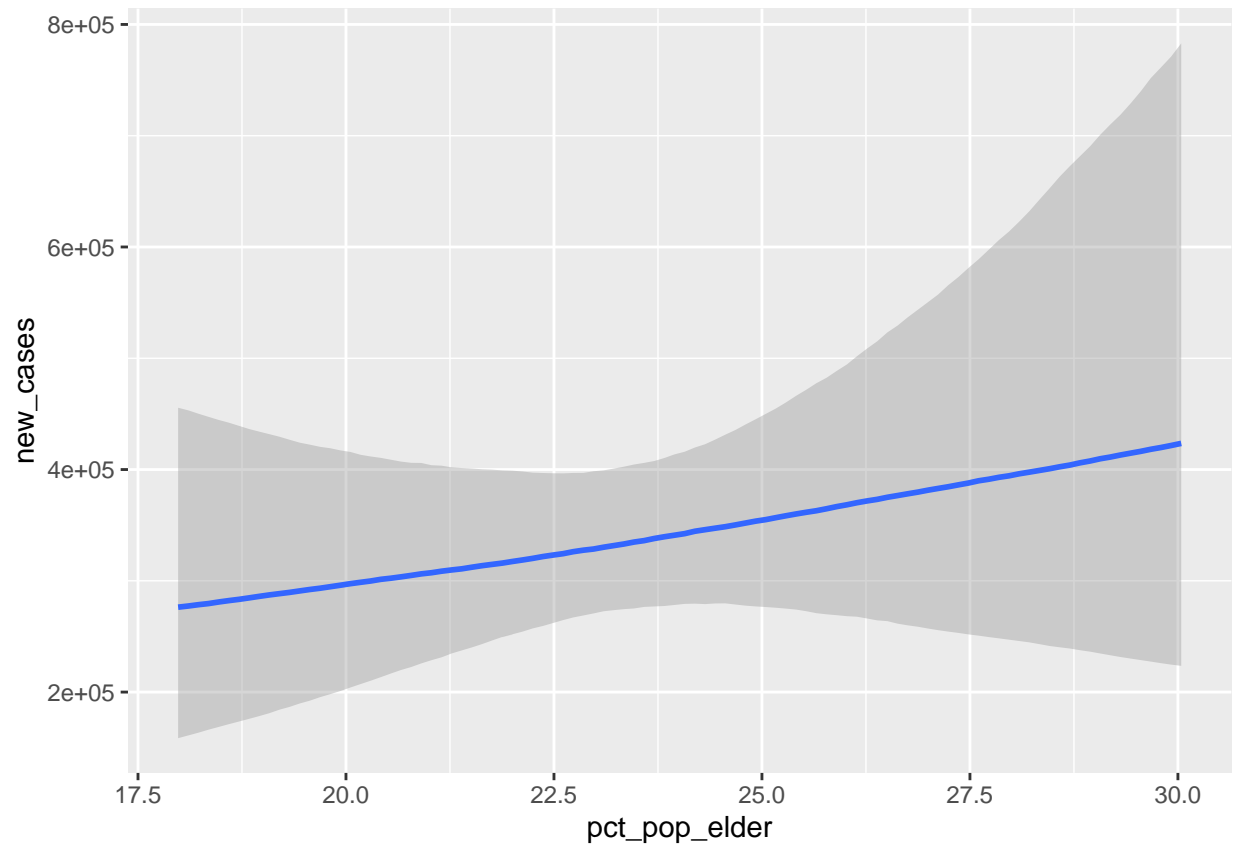


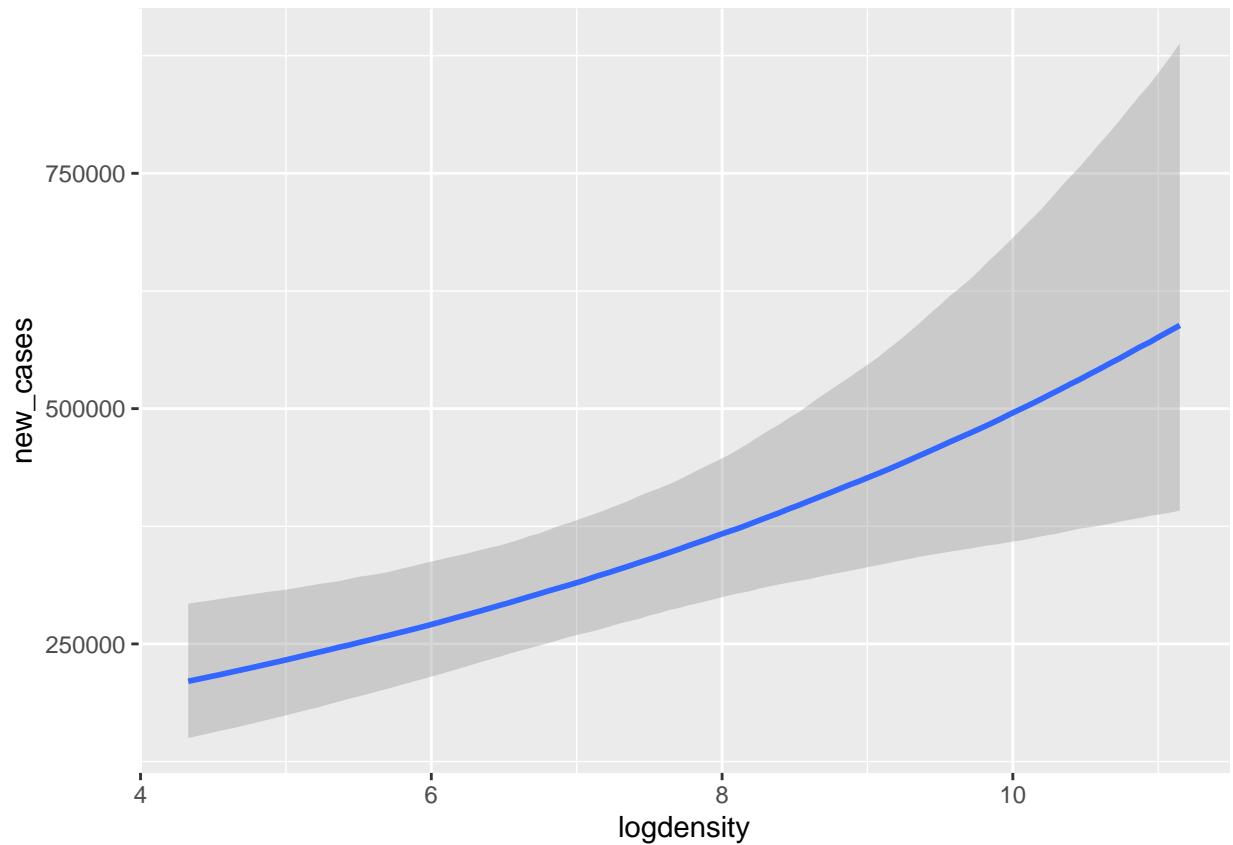










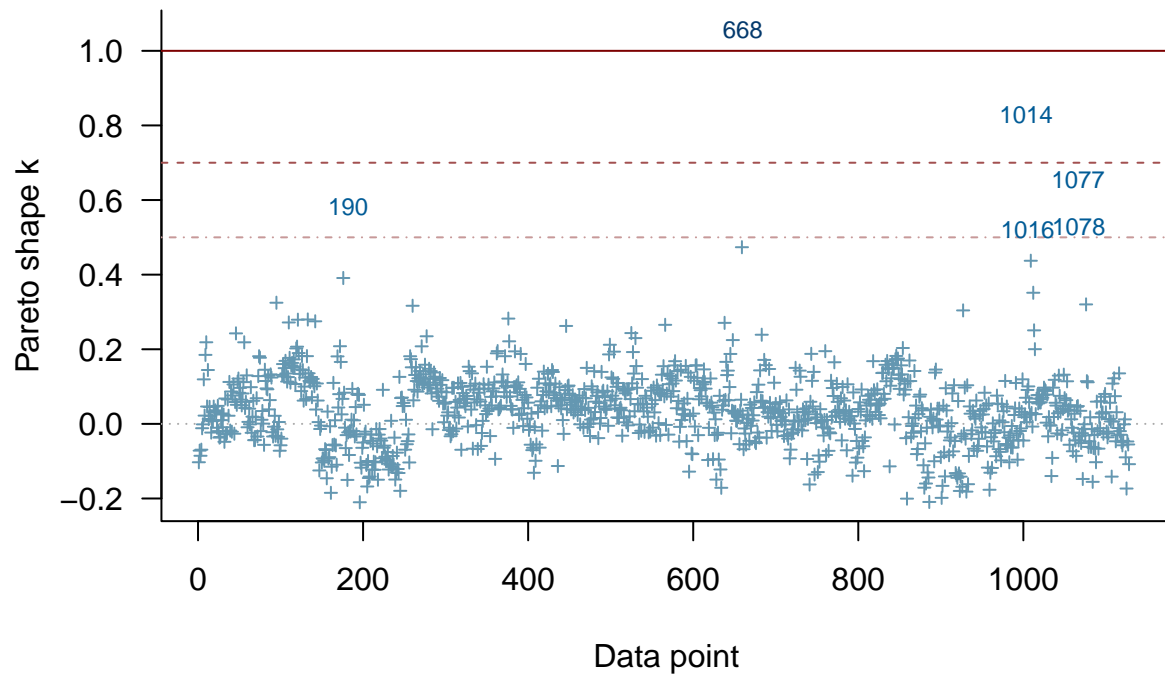


After adding mobility trends covariates, the effects of other covariates don't change much. The marginal effect of mobility trends for transportation seems to be negatively associated with the new cases, which is out of our expectation that new cases will decrease when people take less public transportation.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

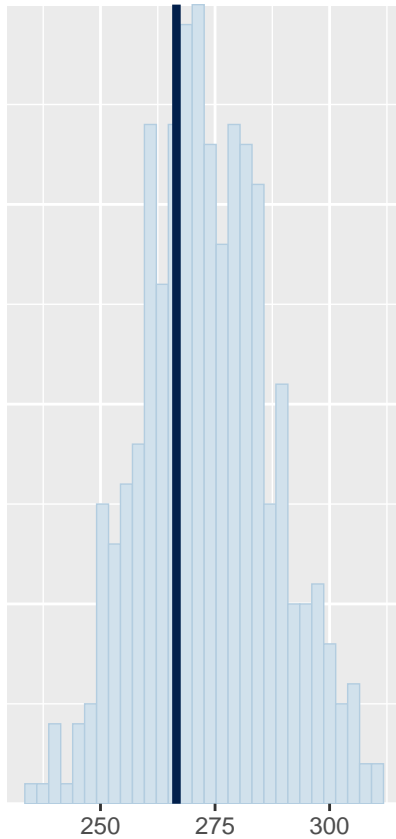
Diagnostics

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

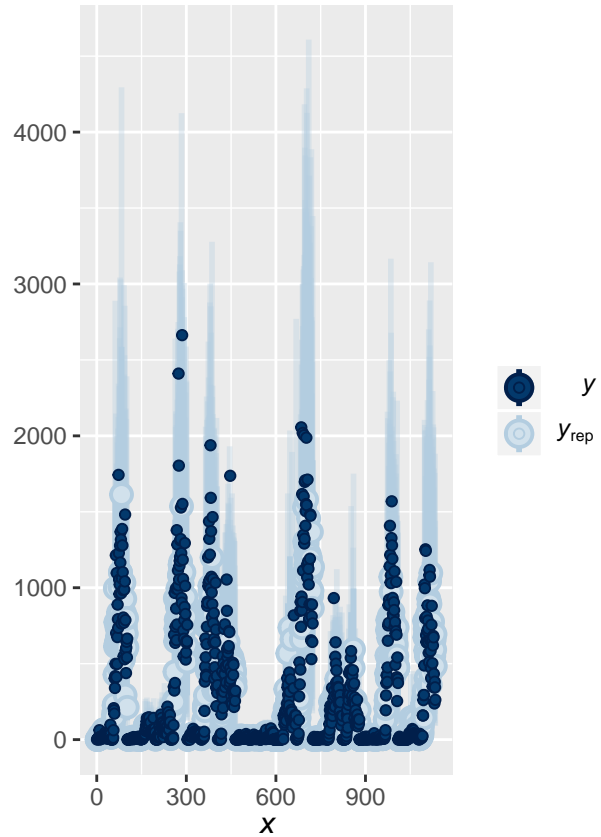
W/ Census & Mobility Covariates



```
gridExtra::grid.arrange(  
  pp_check(post2, nsample=500, type='stat', stat='mean'),  
  pp_check(post2, nsample=500, type='intervals'),  
  ncol=2, nrow=1)
```



$T = \text{mean}$
 $T(y_{\text{rep}})$
 $T(y)$



```
# 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_chan

loo_compare(loo1.1,loo2)

##      elpd_diff se_diff
## post2      0.0      0.0
## post1 -380.6     33.5
```

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_vicinityTRUE,
  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       -73.1     15.7
```

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_vicinityTRUE,
  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  -3.9      3.0
```

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_vicinityTRUE,
               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    -0.5         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_vicinityTRUE,
              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 -7.1         5.9
```

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
```

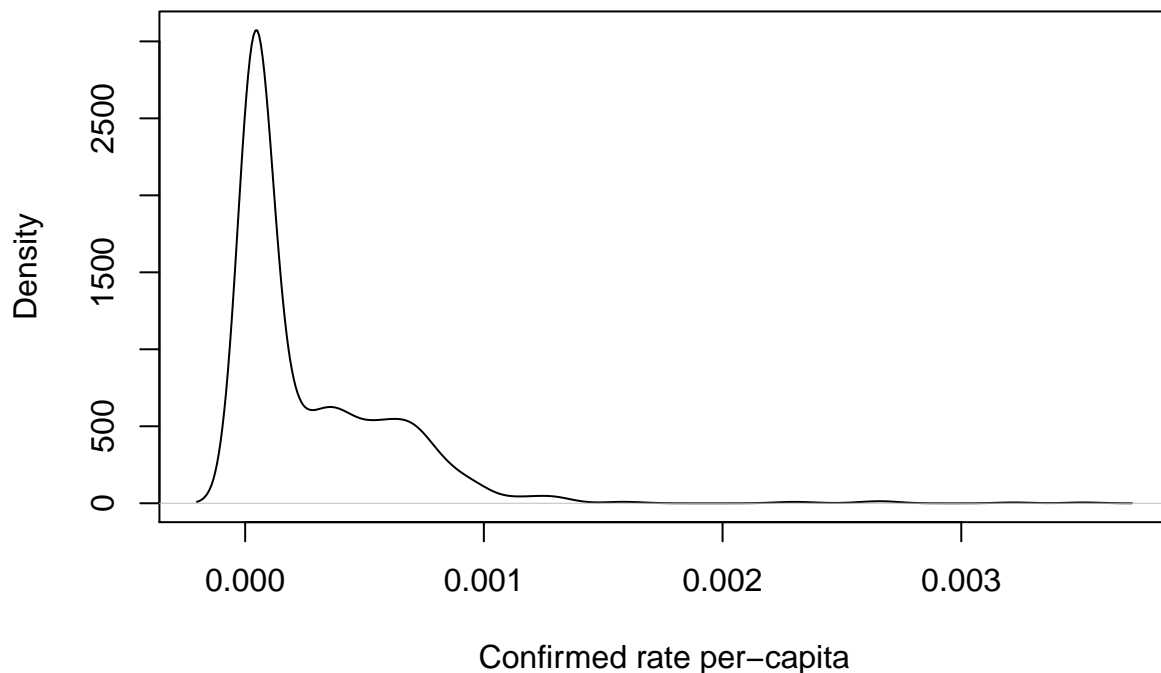
```
## Family: negbinomial
## Links: mu = log; shape = identity
## Formula: new_cases ~ poly(day, 2) + transit_stations_change + grocery_pharmacy_change + parks_change
## Data: dat (Number of observations: 1134)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Group-Level Effects:
## ~county_fips (Number of levels: 26)
##
##          Estimate Est.Error 1-95% 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)       0.08      0.06    0.00    0.24
## sd(transit_stations_change) 0.19    0.02    0.15    0.24
## cor(Intercept,polyday21) -0.01    0.37   -0.71    0.69
## cor(Intercept,polyday22)  0.05    0.37   -0.67    0.72
## cor(polyday21,polyday22) -0.00    0.38   -0.71    0.70
## cor(Intercept,transit_stations_change) -0.06    0.20   -0.44    0.32
## cor(polyday21,transit_stations_change)  0.04    0.39   -0.70    0.74
## cor(polyday22,transit_stations_change) -0.07    0.39   -0.74    0.68
##
##          Rhat Bulk_ESS Tail_ESS
## sd(Intercept)      1.00    1389    1806
## sd(polyday21)       1.00    2105    1364
## sd(polyday22)       1.00    1867    1096
## sd(transit_stations_change) 1.00    1147    2026
## cor(Intercept,polyday21) 1.00    3369    2391
## cor(Intercept,polyday22) 1.00    3214    2283
## cor(polyday21,polyday22) 1.00    3054    2875
## cor(Intercept,transit_stations_change) 1.01     436     932
## cor(polyday21,transit_stations_change) 1.02     125     346
## cor(polyday22,transit_stations_change) 1.01     186     603
##
## Population-Level Effects:
##          Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
## Intercept      -14.07      1.02   -16.03   -12.11 1.00     2497
## polyday21         0.25      0.10     0.05     0.45 1.00     3383
## polyday22        -0.12      0.05    -0.21    -0.02 1.00     2982
## transit_stations_change -0.26      0.00    -0.26    -0.25 1.00      967
## grocery_pharmacy_change -0.00      0.00    -0.00     0.00 1.00     6992
## parks_change       0.01      0.00     0.00     0.01 1.00     4348
## nyc_vicinityTRUE     0.08      0.05    -0.02     0.17 1.00     2890
## pct_pop_elder       0.08      0.04     0.00     0.15 1.00     2518
```

```
## logdensity          -0.03      0.04    -0.12     0.05 1.00     1933
##                               Tail_ESS
## Intercept          2682
## polyday21          2630
## polyday22          2873
## transit_stations_change 2156
## grocery_pharmacy_change 3196
## parks_change        3199
## nyc_vicinityTRUE     2824
## pct_pop_elder        2834
## logdensity          2440
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## shape      2.30      0.11     2.08     2.54 1.00     2615     2593
##
## 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_
```

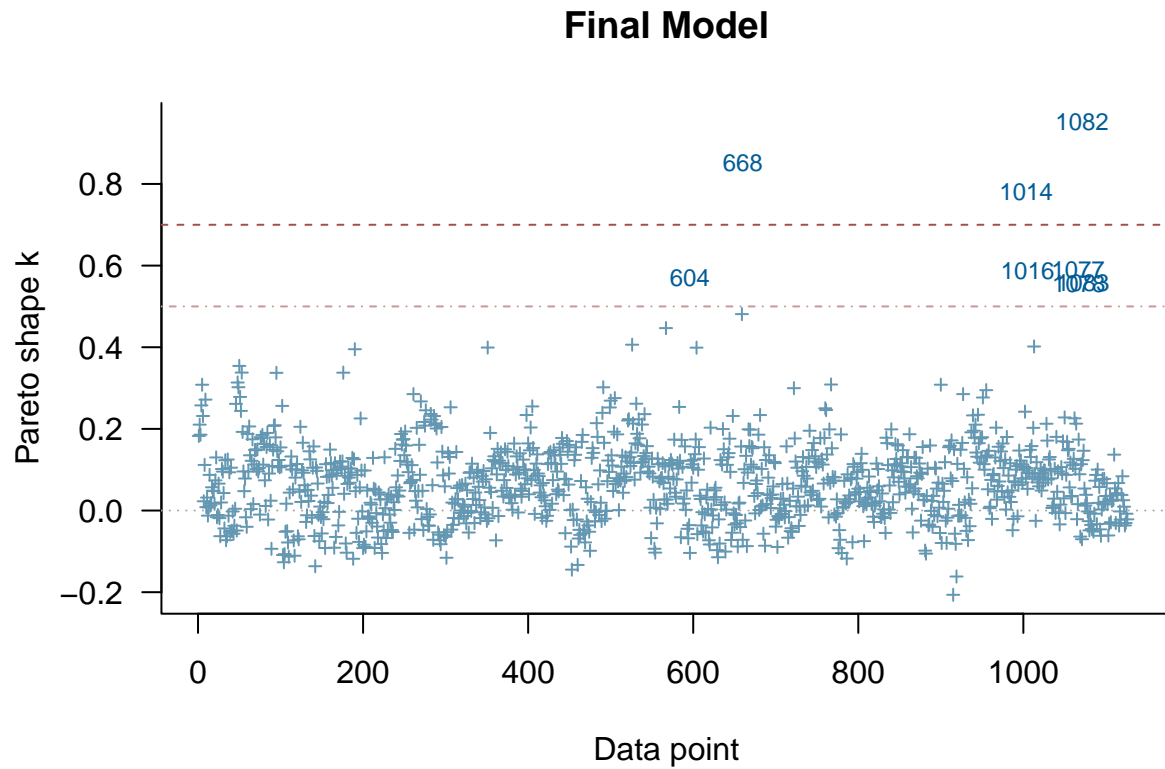
Posterior distribution



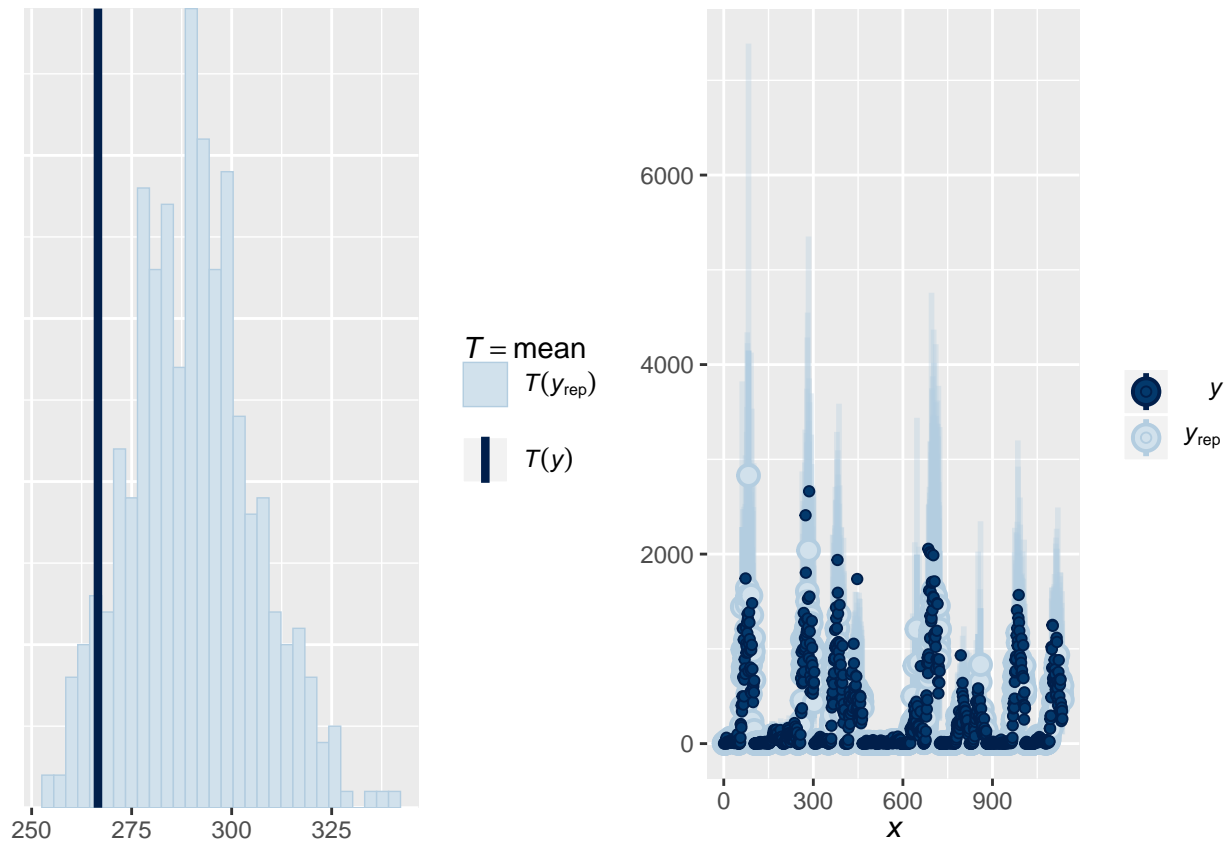
With county-specific transportation mobility trends, the range of posterior distribution is smaller. Probably the effect of some highly influential counties like NYC is smoothed by the group-specific effect.

Diagnostics

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



```
gridExtra::grid.arrange(  
  pp_check(post.trans,nsample=500,type='stat',stat='mean'),  
  pp_check(post.trans,nsample=500,type='intervals'),  
  ncol=2,nrow=1)
```

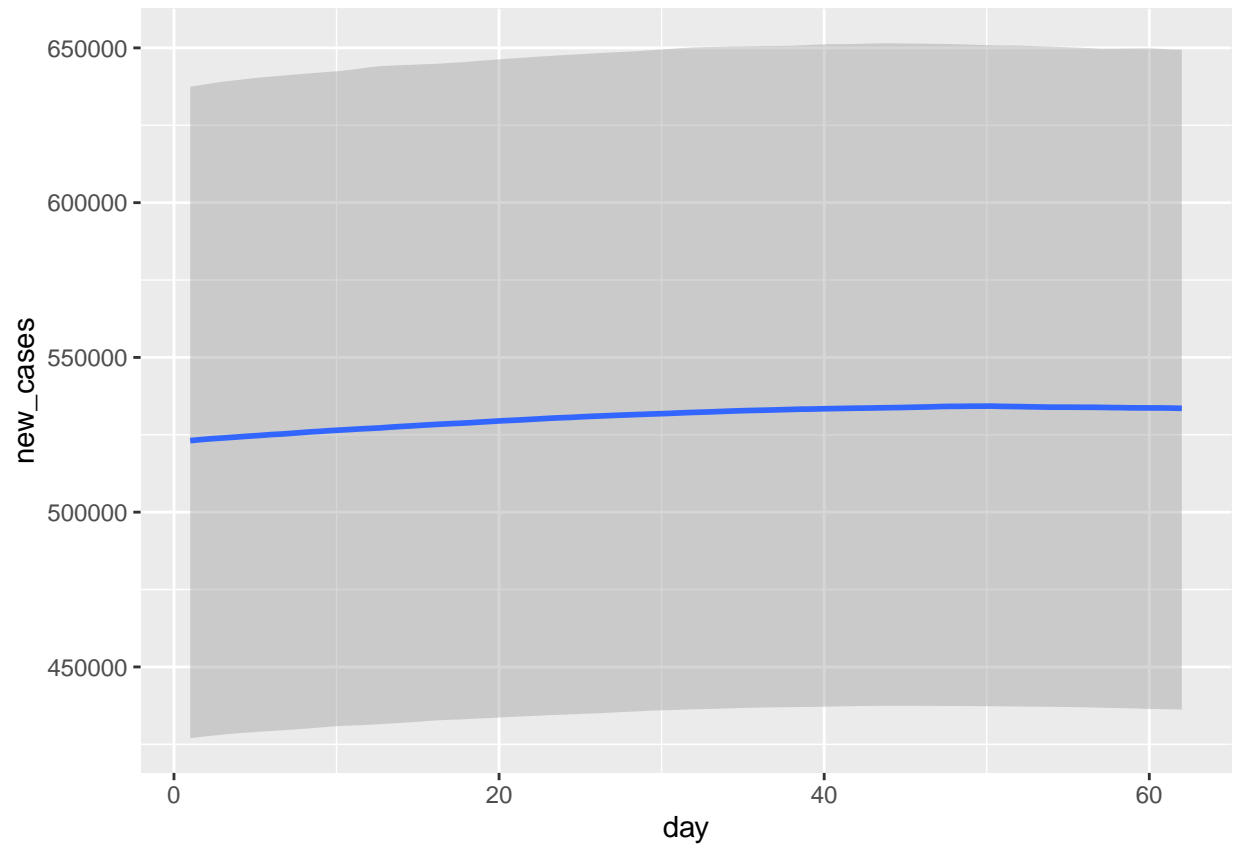
```
# numerical assessment of calibration
PPD.trans <- posterior_predict(post.trans)
lower.trans <- apply(PPD.trans,2,quantile,probs=.025)
upper.trans <- apply(PPD.trans,2,quantile,probs=.975)
cat('The probability of observations fall into 95% of the posterior distribution is',
mean(dat[complete.cases(select(dat,transit_stations_change,grocery_pharmacy_change,pa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

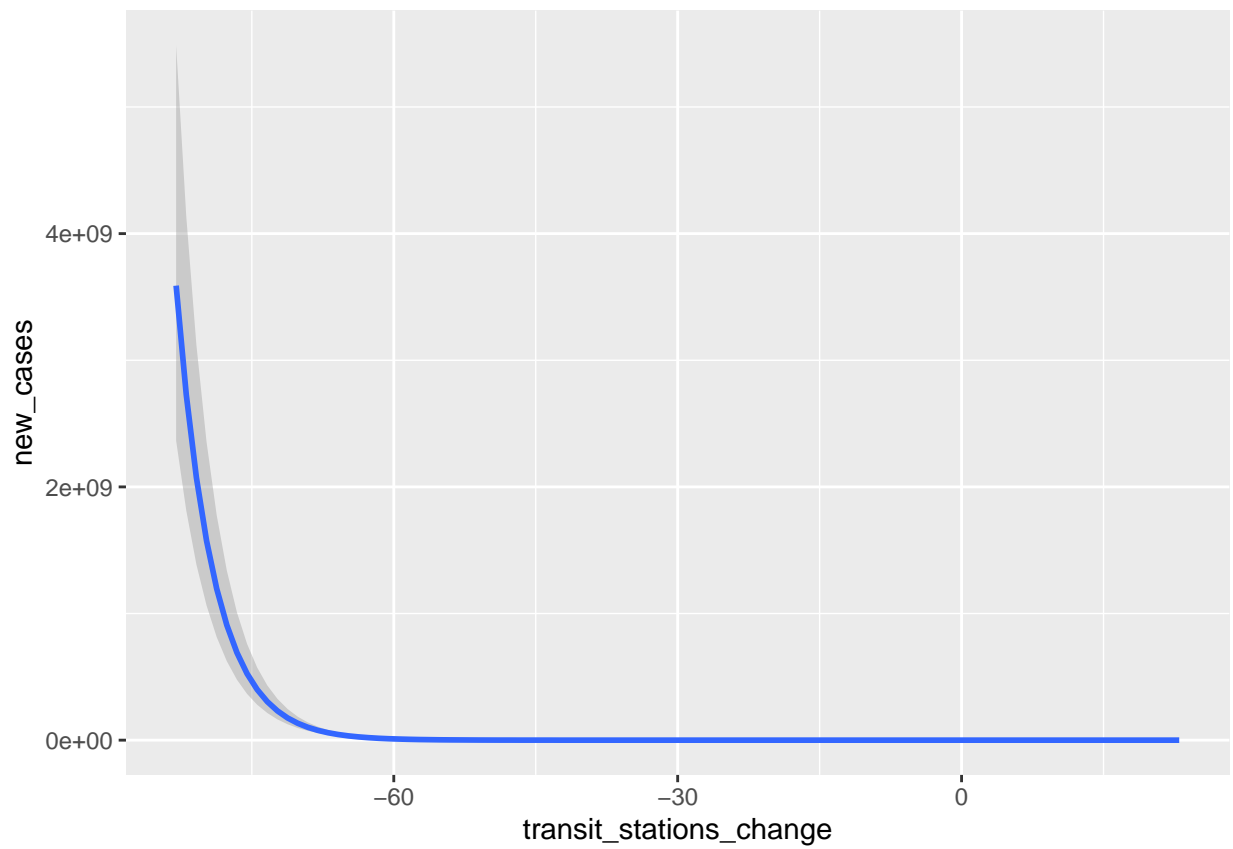
The probability of observations fall into 95% of the posterior distribution is 0.9003527

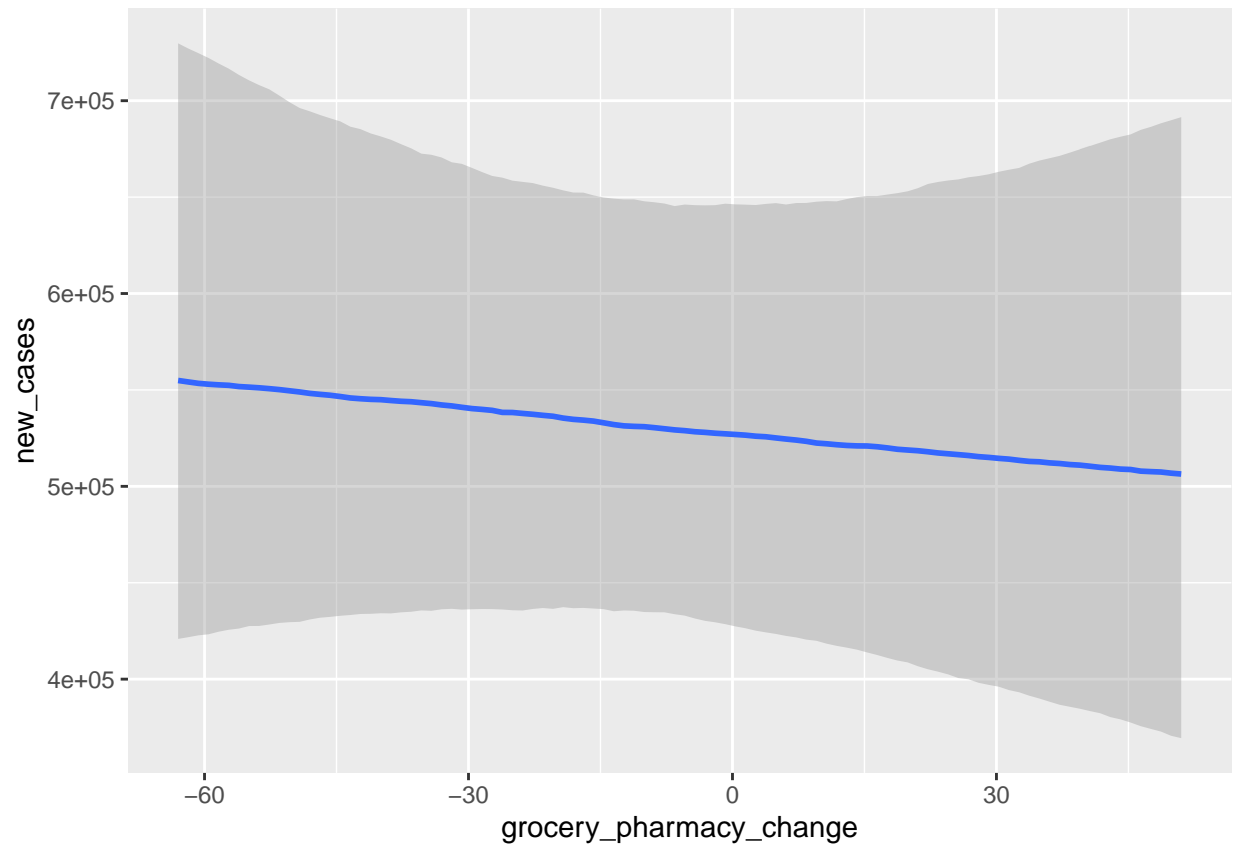
Three highly influenced observations were detected. Eyeballing the loo-based plots, the final model overestimates the mean.

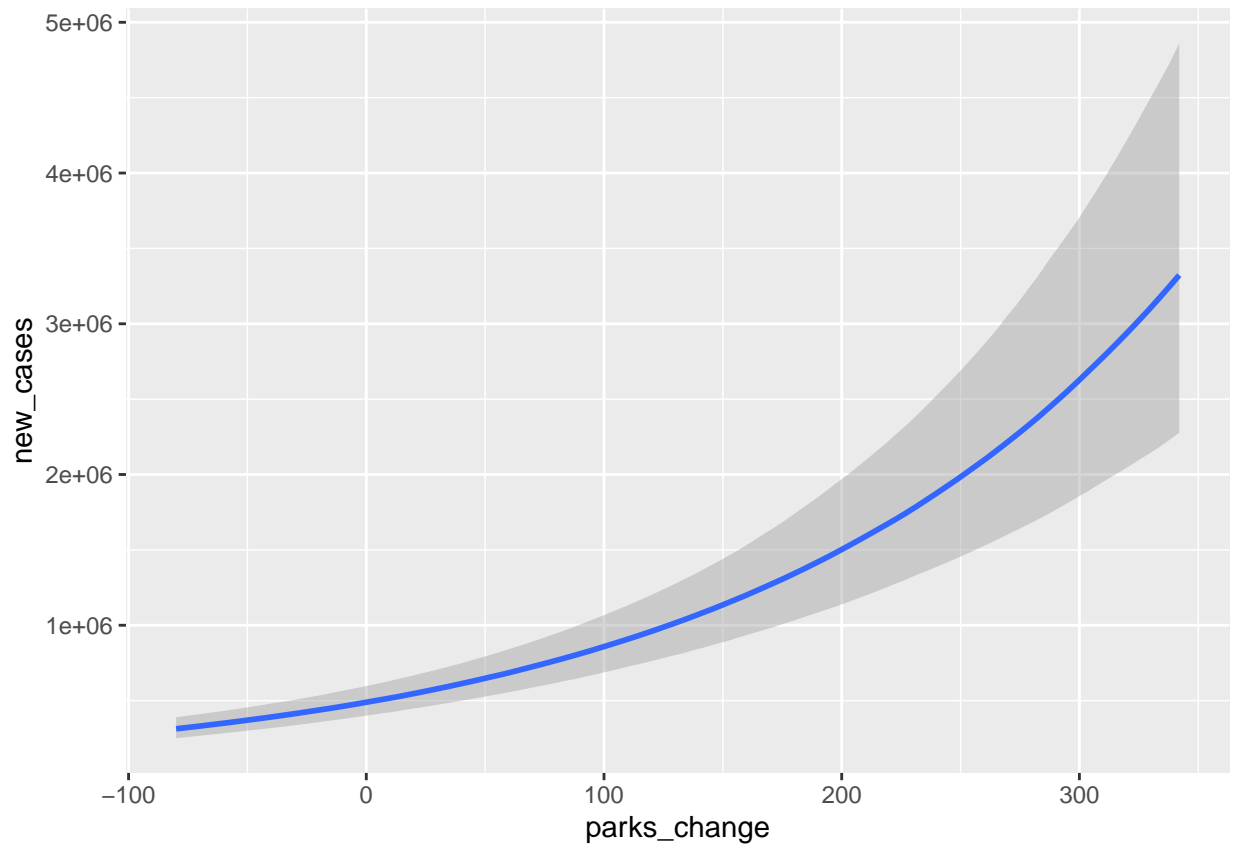
Conditional Effects

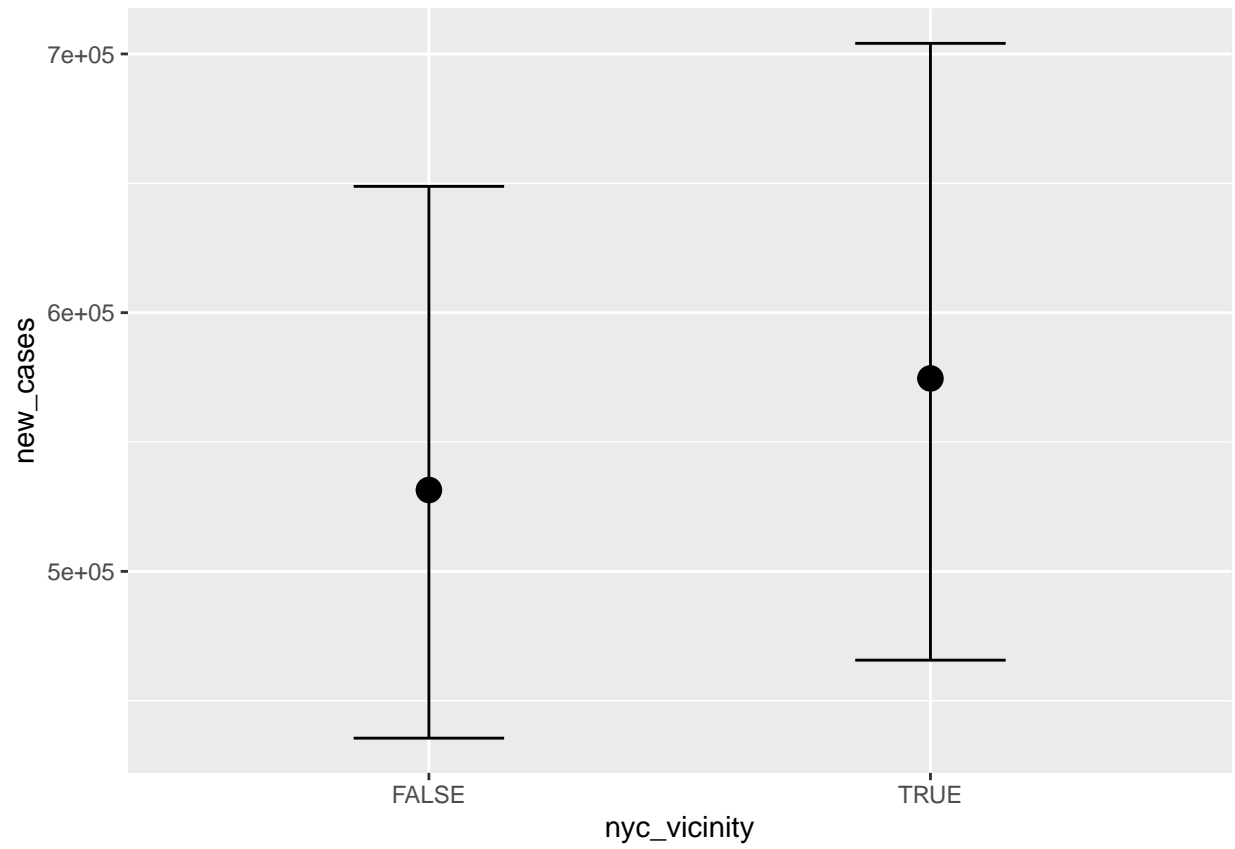
```
plot(conditional_effects(post.trans),ask=F)
```

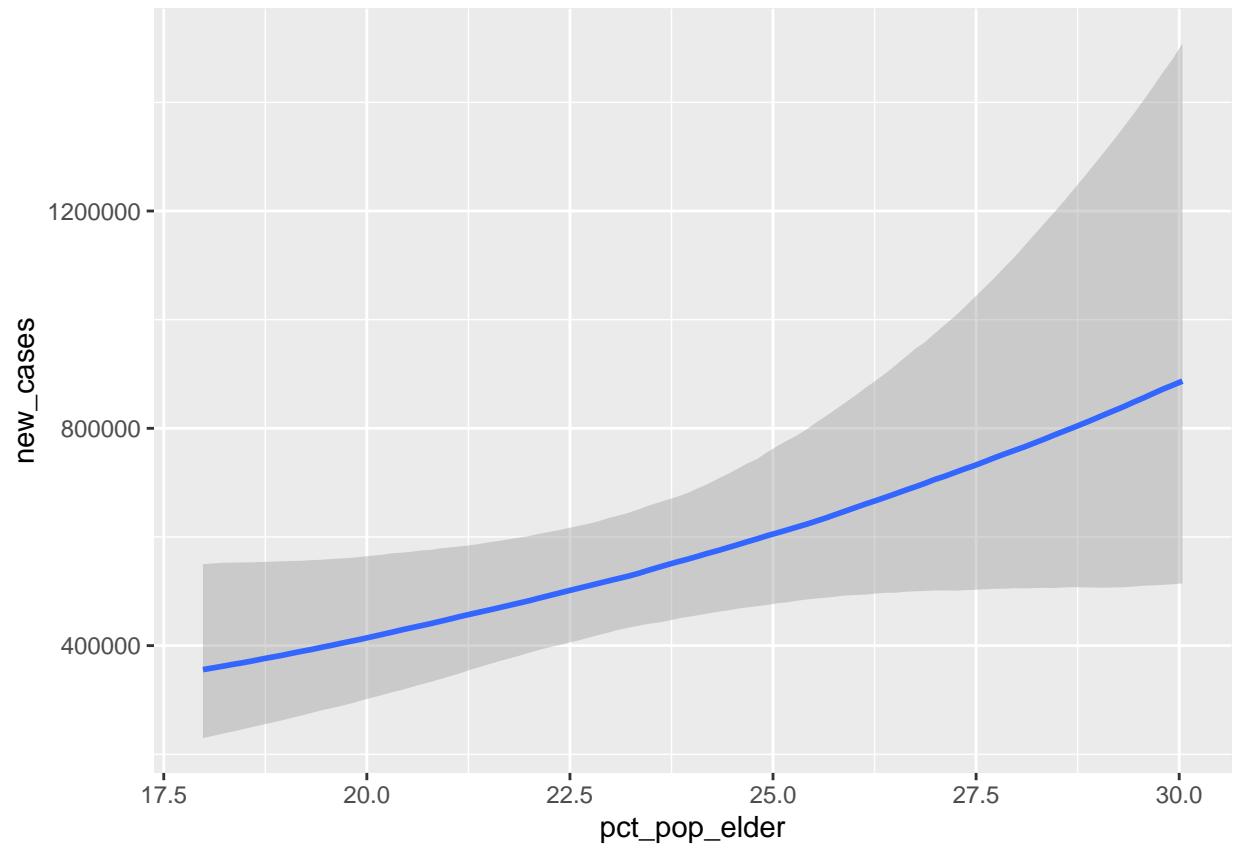


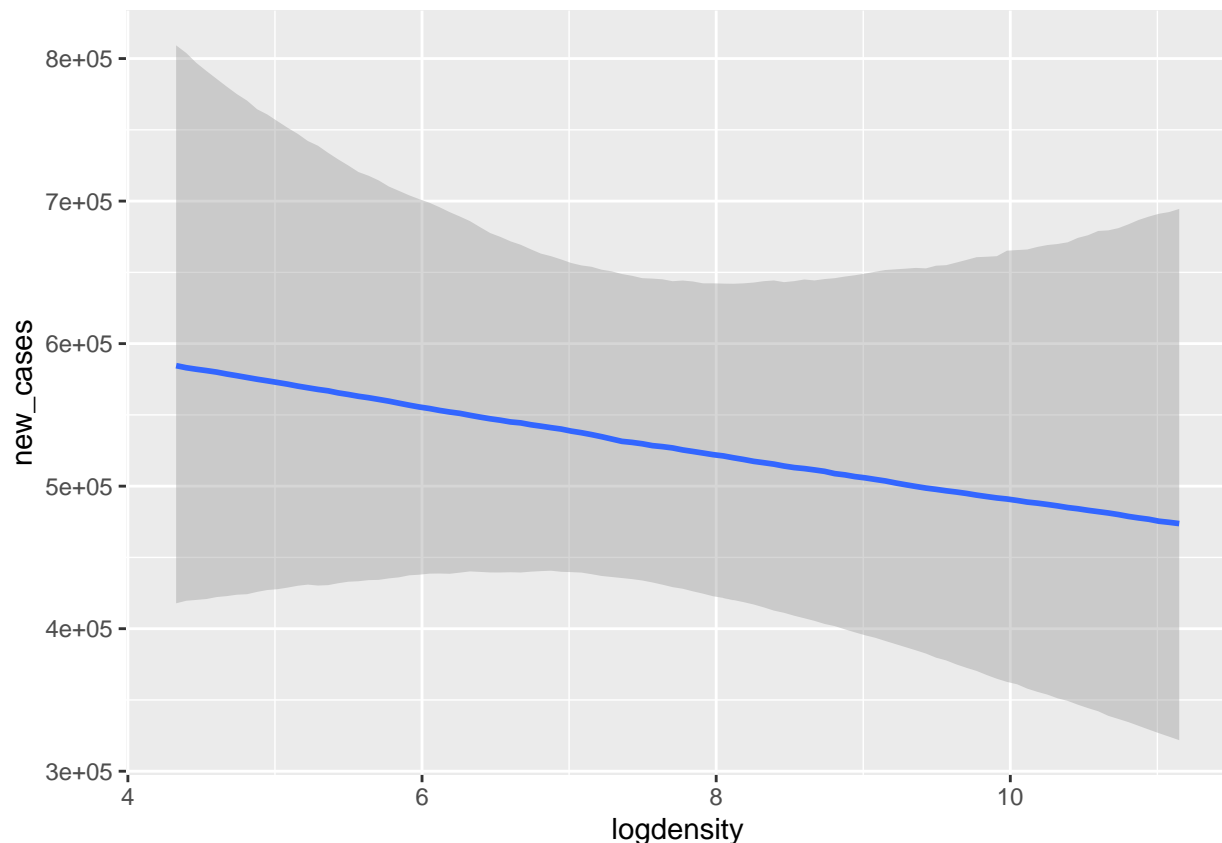












Unexpectedly, the conditional effect of population density turns negative after including county-specific transportation mobility effect. The mobility change effects of transportation and grocery are negatively associated with the daily new cases. It might be because NYC is having a large effect on the 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.