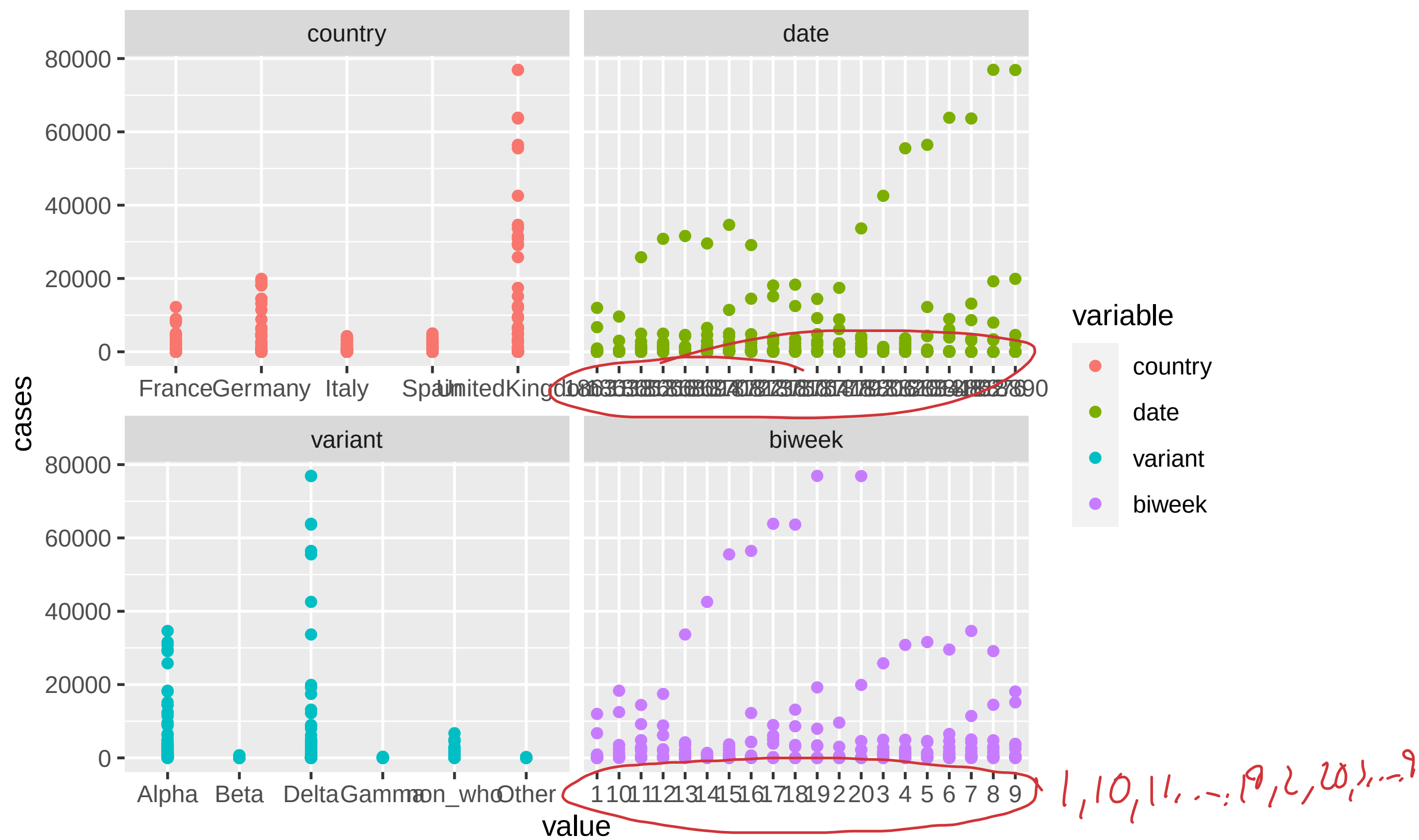
MTH3041 Bayesian Statistics, Philosophy and Practice CourseWork

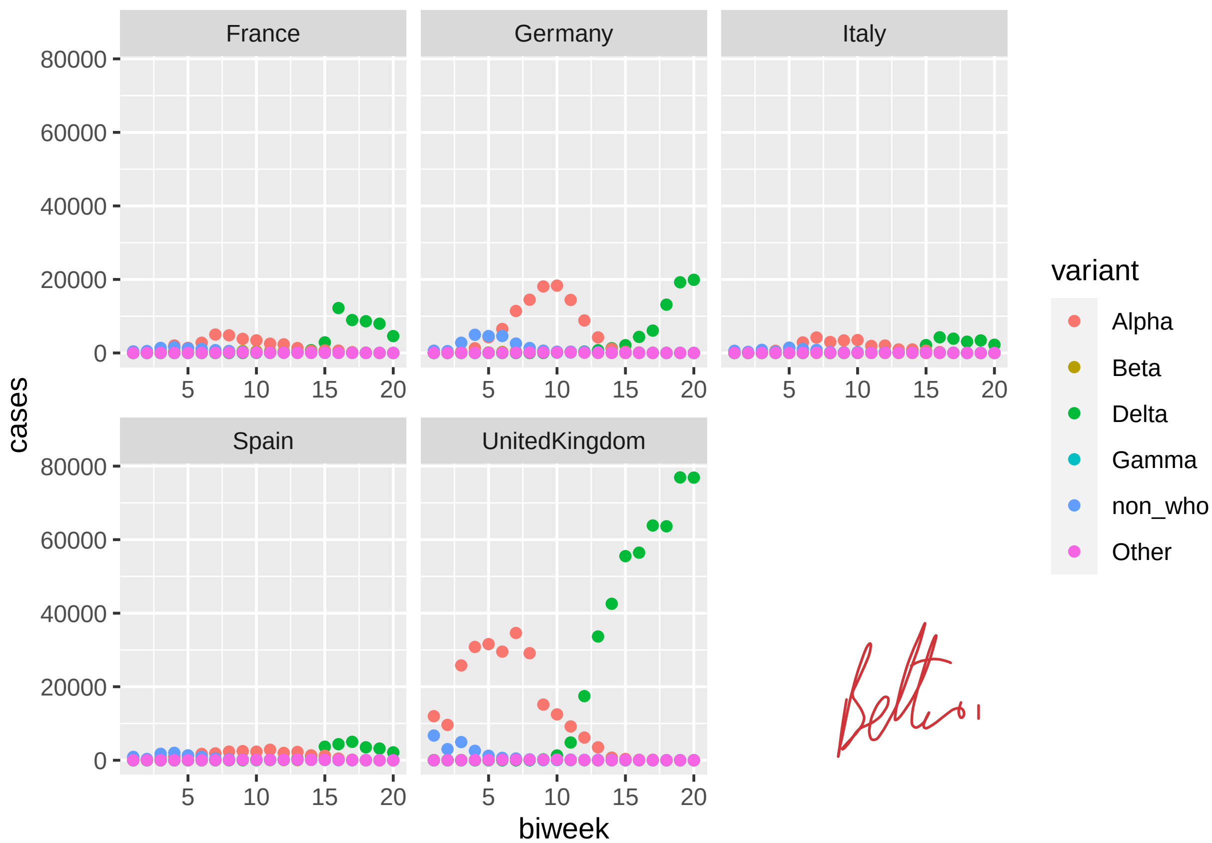
198728

1)

```r
variants_5Largest =
  rbind(variants_Germany, variants_UnitedKingdom, variants_France, variants_Italy, variants_Spain)
# view(variants_5Largest)
variants_5Largest %>% melt(id.vars="cases") %>%
  ggplot()+
  geom_point(aes(x=value, y=cases, colour=variable))+
  facet_wrap(~variable, scales="free_x")
```




```r
ggplot(variants_5Largest)+
  geom_point(aes(x = biweek, y = cases, colour = variant))+
  facet_wrap(~country, scales = "free_x")
```



Better,

```
ggplot(variants_5Largest)+
  geom_point(aes(x = date, y = cases, colour = variant))+
  facet_wrap(~country, scales = "free_x")
```

```
ggplot(variants_5Largest)+
  geom_point(aes(x = biweek, y = cases, colour = country))+
  facet_wrap(~variant, scales = "free_x")
```

These plots show that the data of 'date' and 'biweek' are the exact same data, since they are a one to one translation of each other, so there is no need to consider both of them in the model. So only 'biweek' will be consider as a variable in the model.
Cases can only be whole numbers so we are modelling counts. So let $y = cases$, so for a count distribution, can use poisson or negative binomial. But if poisson then $Mean[y] = Var[y]$. Which makes the likelihoods flat everywhere if there's no poisson where the mean and variance are equal that actually explains the data, so a poisson regression should not be used. So instead use a negative binomial regression, since it allows $Mean(y) \neq Var(y)$. So $y \mid \beta, x \sim NegBin(\lambda(\beta, x), \phi)$. So $P(y = K \mid \lambda, \phi) \propto (\frac{\mu}{\phi+\mu})^K (\frac{\phi}{\phi+\mu})^\phi$. So $E[y] = \lambda \geq 0$ and $Var[y] \neq \lambda$ but $Var[y] = \lambda(1 + \frac{\lambda}{\mu}) > \lambda$ for $\phi \geq 0$. So $log(\lambda(\beta, x)) = \sum_{i=0}^{p} \beta_i x_i$. So for posterior distribution: $\pi(\beta, \Sigma) = \pi(B, \sigma, \phi)$. So let $y_i = cases_i$ μ?

prior

```
nrow(variants_5Largest)
```

So $i = 1, ..., 600$. So $y_i \mid \underline{\beta} \sim NegBin(\lambda(\beta, x), \phi)$, where λ is the inverse link. So $log(\lambda(\beta, x)) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$. I believe cases will change drastically based on country because of the different population sizes and each one having a different government. So I wish to group by country, and so make country a grouping variable in my model. So because 'date' is just a replicate of 'biweek': x_1 is variant, x_2 is biweek.

```r
variants_5Largest$variant
```

So for grouping by country: $j = \{1, 2, 3, 4, 5\}$. So 1 is France, 2 is Germany, 3 is Italy, 4 is Spain, 5 is United Kingdom.

```r
sum(variants_5Largest$country == "France")
sum(variants_5Largest$country == "Germany")
sum(variants_5Largest$country == "Italy")
sum(variants_5Largest$country == "Spain")
sum(variants_5Largest$country == "UnitedKingdom")
```

So $n_1 = 120$, $n_2 = 120$, $n_3 = 120$, $n_4 = 120$, $n_5 = 120$. So $y_{ij} \mid \beta_j, x_{ij} \sim NegBin(\lambda(\beta, x), \phi)$. So $log(\lambda(\beta, x)) = \beta_{0j} + \beta_{1j}x_{variant,j} + \beta_{2j}x_{biweek,j}$ So model for β's are: $\beta_j \sim N(B, \Sigma)$. So prior $\pi(B, \Sigma)$. So $log(\lambda(\beta, x_{ij})) = b_0 + b_1x_{variant,ij} + b_2x_{biweek,ij} + \beta_{0j} + \beta_{1j}x_{variant,j} + \beta_{2j}x_{biweek,j}$. So $\beta_j \sim N(0, \Sigma)$. And need prior for $\pi(b, \Sigma)$. So $E[y] = \lambda = exp(b_0 + b_1x_{variant} + b_2x_{biweek} + \beta_0 + \beta_1x_{variant} + \beta_2x_{biweek}) = e^{b_0}e^{b_1x_{variant}}e^{b_2x_{biweek}}e^{\beta_0}e^{\beta_1x_{variant}}e^{\beta_2x_{biweek}}$. And $\Sigma = vec(\underline{\sigma}) \times \Omega_k \times vec(\underline{\sigma})$, where $\underline{\sigma} = \sigma_0, \sigma_{variant}, \sigma_{biweek}$. The cases in the data range from a lower bound of a count of 1, all the way up to 77000, with the vast majority of the counts being very small. So $e^{b_0} \in [1, 77000]$.

```r
log(c(1, 77000))
```

```
## [1]  0.00000 11.25156
```

```r
log(77000)/2
```

```
## [1] 5.62578
```

So set weakly informative prior on the intercept as $N(0, 4)$, and on the grouping as $N(0, 10)$. why?

```r
Intercept_Prior <- set_prior("normal(0,4)", class="Intercept")
sd_priors_Intercept <- set_prior("normal(0, 10)", class="sd", group="country", coef="Intercept")
```

In the data, biweek ranges from 1 to 20.

```r
exp(0.15*c(1, 20))
```

```
## [1]  1.161834 20.085537
```

Dont follow the logic.

So make b prior for biweek $N(0, 0.01)$ as well as for the grouping. And make the b priors as well as the groupings, for variant very weakly informative as $N(0, 100)$. oh, but why?

```r
biweek_priors <- set_prior("normal(0, 0.01)", class="b", coef = "biweek")
sd_priors_biweek <- set_prior("normal(0, 0.01)", class = "sd", group = "country", coef = "biweek")
variant_priors <- c(set_prior("normal(0, 100)", class="b", coef="variantBeta"),
    set_prior("normal(0, 100)", class="b", coef="variantDelta"),
    set_prior("normal(0, 100)", class="b", coef="variantGamma"),
    set_prior("normal(0, 100)", class="b", coef="variantnon_who"),
    set_prior("normal(0, 100)", class="b", coef="variantOther"))
sd_priors_variant<-c(set_prior("normal(0, 100)", class="sd", group="country", coef="variantBeta"),
    set_prior("normal(0, 100)", class="sd", group="country", coef="variantDelta"),
    set_prior("normal(0, 100)", class="sd", group="country", coef="variantGamma"),
    set_prior("normal(0, 100)", class="sd", group="country", coef="variantnon_who"),
    set_prior("normal(0, 100)", class="sd", group="country", coef="variantOther"))
```

So based on the variance formula:

```r
lambdas <- c(1, 77000)
varys <- function(phi){
  lambdas *(1+lambdas/phi)  }
sqrt(varys(800))
```

```
## [1]    1.000625 2736.466700
```

```r
2700/77000
```

```
## [1] 0.03506494
```

So, to make reasonable prediction error bars, for about 35% error bars, make the shape prior $N(400, 200)$.

Don't quite follow exact reasoning here.

```r
shape_prior <- set_prior("normal(400, 200)", class="shape")
```

```r
model_1<-brm(cases ~ biweek+ variant+ (biweek+ variant | country),
data = variants_5Largest, family = negbinomial(), prior=c(Intercept_Prior, sd_priors_Intercept,
biweek_priors, sd_priors_biweek, variant_priors, sd_priors_variant, shape_prior),
iter=6000, warmup = 5000, inits="0") # is my first model
```

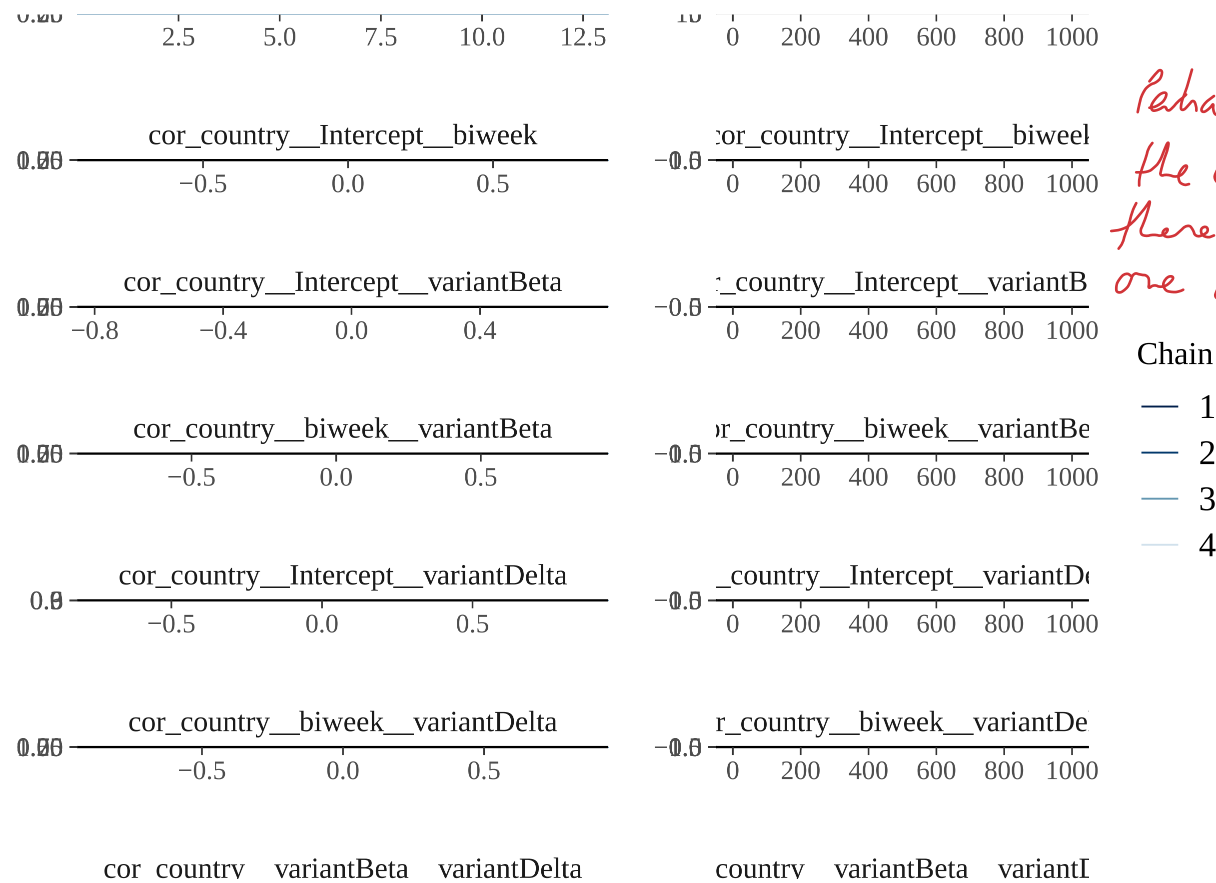
```r
summary_model_1 <- summary(model_1)
rbind(summary_model_1$random$country[-2:-4], summary_model_1$fixed[-2:-4],
      summary_model_1$spec_pars[-2:-4])
```

```
##                                    Estimate      Rhat  Bulk_ESS  Tail_ESS
## sd(Intercept)                   1.178018820 1.0184545  381.7273  146.1426
## sd(biweek)                      0.009359852 1.0016202 1638.5602 2034.2631
## sd(variantBeta)                 2.670555351 1.0005355 1118.5845 1460.6956
## sd(variantDelta)                1.030040643 1.0021051 1506.8777 1483.3983
## sd(variantGamma)                2.244572870 1.0051071  897.9396  441.3606
## sd(variantnon_who)              0.530122838 1.0049380  865.6571 1788.5258
## sd(variantOther)                0.982551827 1.0131778  338.0679  134.5648
## cor(Intercept,biweek)          -0.102476413 1.0027181 3737.3875 2415.4632
## cor(Intercept,variantBeta)     -0.035146039 1.0056547 1032.7487  225.8081
## cor(biweek,variantBeta)         0.064092421 1.0036660 1532.4334 2428.5260
## cor(Intercept,variantDelta)     0.144906464 1.0003813 2739.2121 2452.2025
## cor(biweek,variantDelta)       -0.090747299 1.0053205 1689.6774 2534.9643
## cor(variantBeta,variantDelta)  -0.039903989 1.0047781 1399.8894 1384.1790
## cor(Intercept,variantGamma)    -0.333899045 1.0014279 1478.5543  566.2504
## cor(biweek,variantGamma)        0.061799911 1.0004939 1879.2013 2392.8335
## cor(variantBeta,variantGamma)   0.011162502 1.0000485 2995.5186 2662.9233
## cor(variantDelta,variantGamma) -0.186605758 1.0011477 2505.2352 3164.8757
## cor(Intercept,variantnon_who)  -0.092212605 1.0037338 1653.0279  841.0911
## cor(biweek,variantnon_who)      0.057502097 0.9998067 3221.5491 2948.9471
## cor(variantBeta,variantnon_who) 0.054577559 1.0062807  962.5171  378.6696
## cor(variantDelta,variantnon_who) -0.058708971 1.0053763  760.5779  329.9412
## cor(variantGamma,variantnon_who) 0.090742221 1.0005533 2666.5365 3054.7605
## cor(Intercept,variantOther)    -0.214808410 1.0049661 2146.9193 2496.8637
## cor(biweek,variantOther)        0.080492145 1.0007705 2755.3530 2951.8334
```

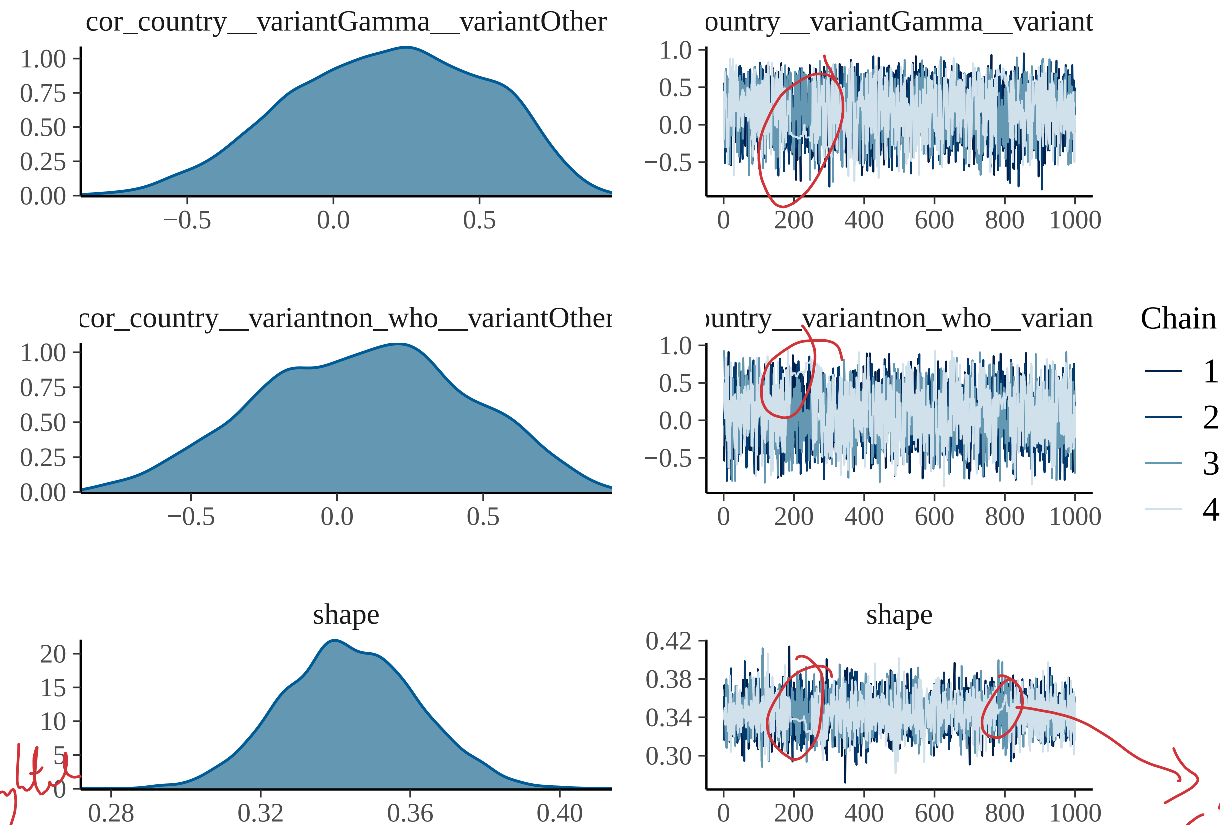
```
## cor(variantBeta,variantOther)      0.116344047 1.0021083 3130.2481 2717.8934
## cor(variantDelta,variantOther)    -0.080059835 1.0074768  536.7563  186.4523
## cor(variantGamma,variantOther)     0.183455828 1.0002016 1972.8309 2944.3245
## cor(variantnon_who,variantOther)   0.090909612 1.0049336 1132.0486 1376.4759
## Intercept                          7.977300958 1.0020156 1009.0879 1398.0636
## biweek                            -0.003250502 1.0038026 4048.0442 2745.6956
## variantBeta                       -3.960871695 1.0026749 1275.3823 1652.1023
## variantDelta                       0.039212875 1.0041501 1225.3079 1520.4159
## variantGamma                      -4.113235710 1.0068501  688.6339  418.1550
## variantnon_who                    -1.593045561 1.0042711 1409.3378  864.7554
## variantOther                      -4.230312564 1.0114443  403.2153  133.7641
## shape                              0.344008562 1.0018202 4528.0343 2806.9581
```

All the Rhat's are 1, and all the effective sample size's for the Bulk and the Tail are very big. Which is an extremely good sign.

```
plot(model_1, N=33, ask=FALSE)
```



Perhaps it states the obvious that these traceplots are difficult to interpret as displayed?



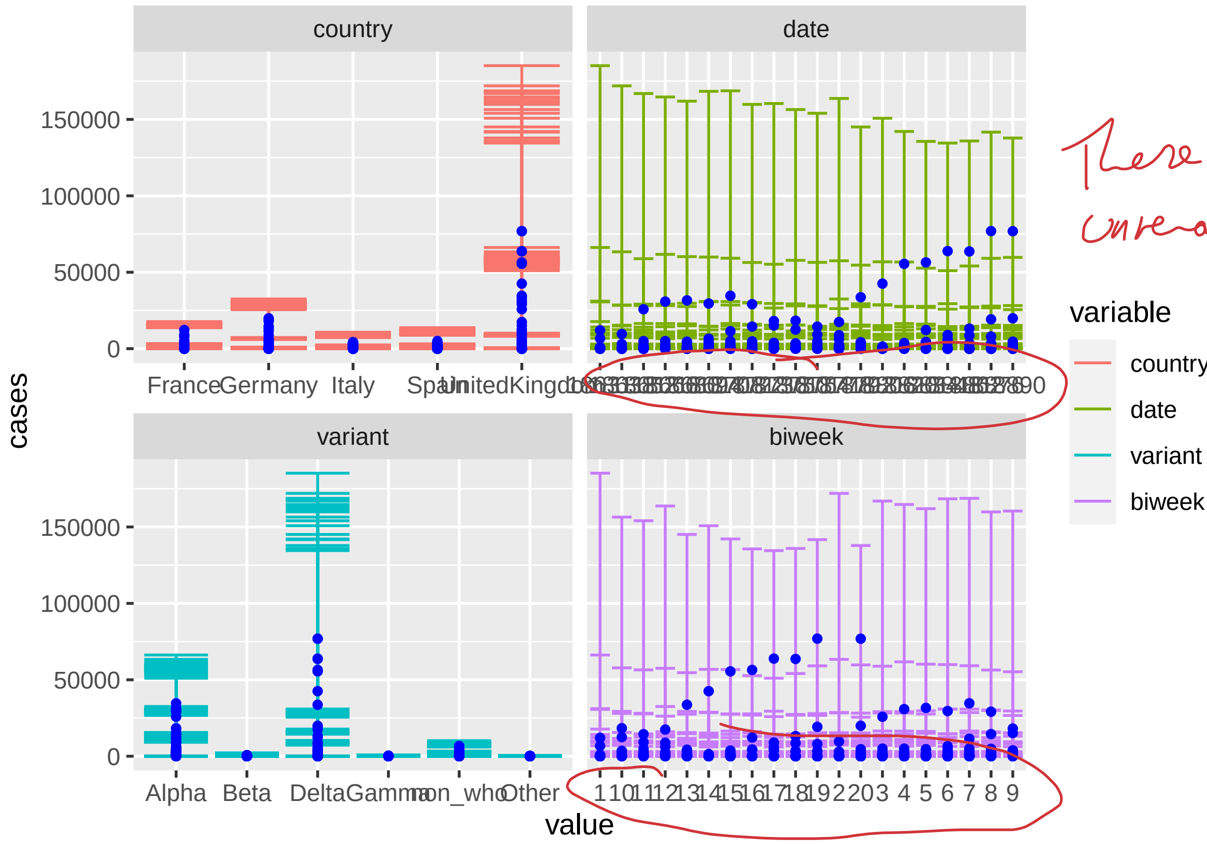
see highlighted stuck areas...

Not true

None of the trace plots got stuck anywhere and they are all very spiky, so they have all converged. This, together with the analysis of the summary, means the model has converged very well. The intercept has moved from the prior by an amount that is fairly believable. The shape mean and mode is about 0.34, which may leed to big error bars. Both biweek and variant have moved away from my prior, by a fairly believable amount, which is a good sign. The model is reasonably small which is good.

what does "fairly believable" mean

```
case_preds <- predict(model_1, newdata = variants_5Largest)
preds <- cbind(variants_5Largest, as.tibble(case_preds)) %>%
  dplyr::select(-Est.Error) %>%
  melt(id.vars = c("Estimate", "Q2.5", "Q97.5", "cases")) %>%
  ggplot()+ geom_errorbar(aes(x=value, ymin=Q2.5, ymax=Q97.5, colour = variable))+
  geom_point(aes(x=value, y=cases), colour="blue", pch=16)+
  facet_wrap(~variable, scale="free_x")
preds
```

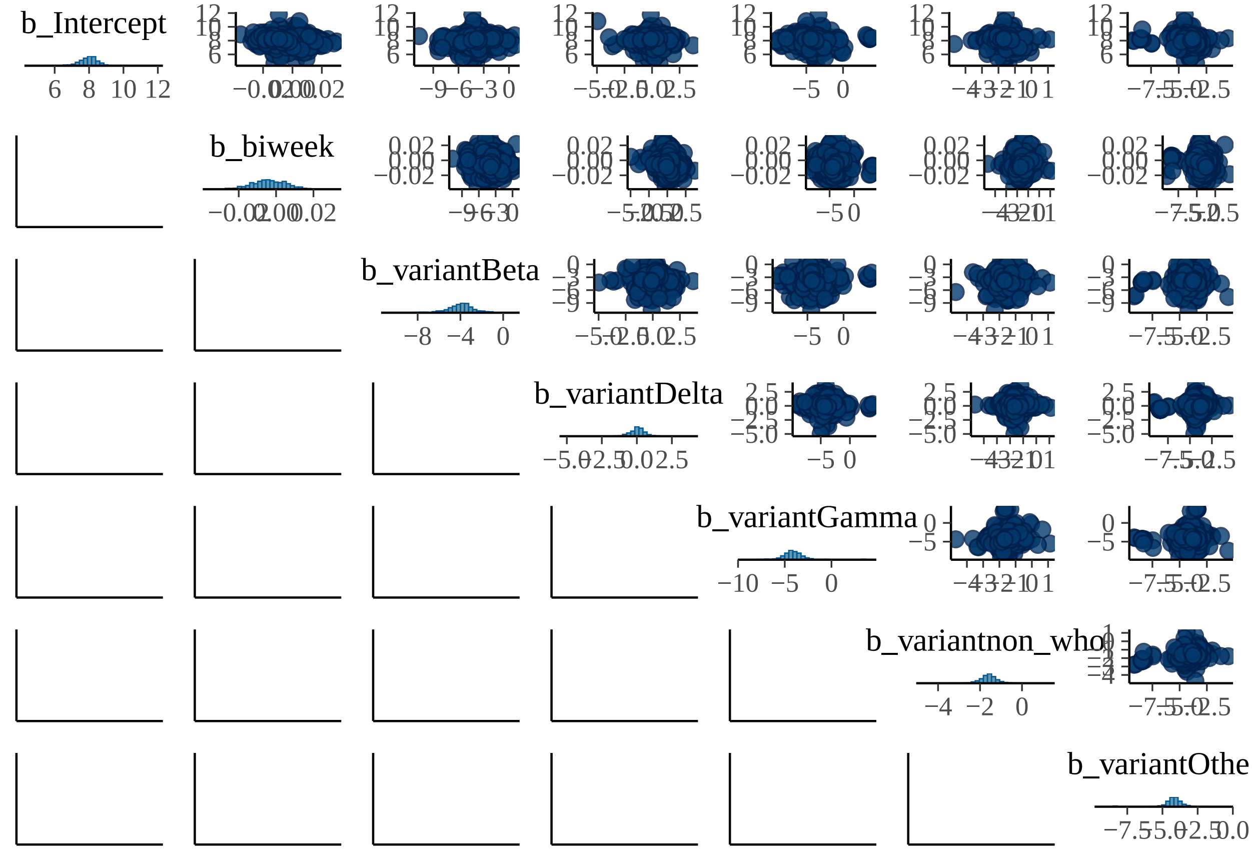


There are unreadable.

The error bars seem reasonably skillful and none of the data points are outside the error bars. Which is good considering how wide ranging the data is.

You can't assess this from these plots.

```
samples <- posterior_samples(model_1, subset=seq(from=1,to=4000,by=4))
mcmc_pairs(samples[,1:7])
```



looking at the mcmc plots for biweek, it is easy to see, there are no correlations between biweek and variant.

Not easy given size of plots.

```
sum(samples$b_biweek<0)/length(samples$b_biweek)
```

```
## [1] 0.65
```

Is the $P(biweek < 0)$ which shows signs, that the mean of biweek is not zero, which means biweek has moved away from my prior.

```
ranef(model_1)
```

```
## $country
## , , Intercept
##
##                  Estimate Est.Error        Q2.5     Q97.5
## France         -0.2540797 0.6264088 -1.43964485 1.1074297
## Germany         0.4995146 0.6277490 -0.62585988 1.8473218
## Italy          -0.6177521 0.6340276 -1.80778609 0.6697288
## Spain          -0.4510390 0.6393902 -1.76744573 0.8316282
## UnitedKingdom   1.2165054 0.6758361  0.03026119 2.6522811
##
## , , biweek
##
##                    Estimate   Est.Error        Q2.5      Q97.5
## France          0.001773964 0.009568604 -0.01638736 0.02555707
```

```
## Germany       -0.001094240 0.009046730 -0.02248043 0.01633117
## Italy         -0.000387160 0.011053218 -0.02409864 0.02283731
## Spain          0.004574718 0.010458604 -0.01177012 0.03084174
## UnitedKingdom -0.007861679 0.015035949 -0.04769505 0.01194004
##
## , , variantBeta
##
##                 Estimate Est.Error       Q2.5    Q97.5
## France         1.9603371  1.311773 -0.5109685 4.821087
## Germany        0.3147633  1.320850 -2.2914071 3.166180
## Italy         -1.4778838  1.310647 -4.1576524 1.271872
## Spain          0.8417316  1.323846 -1.8105061 3.646232
## UnitedKingdom -1.1954779  1.328160 -3.9132751 1.464348
##
## , , variantDelta
##
##                  Estimate Est.Error       Q2.5     Q97.5
## France         0.02748791 0.6513216 -1.2796636 1.3547947
## Germany       -0.12138325 0.6564354 -1.4924087 1.1614889
## Italy         -0.28113631 0.6526608 -1.6086322 0.9761512
## Spain         -0.33553560 0.6741141 -1.8151605 0.8659412
## UnitedKingdom  0.91160712 0.7637863 -0.2308664 2.6114289
##
## , , variantGamma
##
##                 Estimate Est.Error       Q2.5        Q97.5
## France         0.4519899  1.339945 -1.9646651  2.845881590
## Germany       -0.4758315  1.334619 -3.0039714  1.904619459
## Italy          1.5416428  1.343741 -0.8702836  4.134217127
## Spain          0.6833346  1.362420 -1.7469203  3.123152556
## UnitedKingdom -2.3780860  1.367580 -4.9228634 -0.006374396
##
## , , variantnon_who
##
##                  Estimate Est.Error       Q2.5     Q97.5
## France         0.02309038 0.4421880 -0.9131979 0.9667116
## Germany        0.10012445 0.4272180 -0.6228816 1.0779445
## Italy          0.08032578 0.4547156 -0.7534626 1.1684189
## Spain          0.14640297 0.4607050 -0.6293169 1.2434922
## UnitedKingdom -0.31029755 0.5026857 -1.4749951 0.4413056
##
## , , variantOther
##
##                 Estimate Est.Error       Q2.5     Q97.5
## France         0.2922804 0.8211968 -0.7992016 2.0648593
## Germany       -0.1006695 0.7522104 -1.3606147 1.5425474
## Italy          0.2246743 0.8088114 -0.9235219 2.0173403
## Spain          0.5587096 0.8764608 -0.4174780 2.3477649
## UnitedKingdom -0.5891482 0.8029012 -2.0342347 0.9112287
```

The random effects also back up this hypothesis that biweek has no correlation to variant. And the effect of biweek, and the effect of variant, has moved away from my prior's, of them having no effect, to them having an effect on my model. So on this first model, I will do a sensitivity step and analysis. The sensitivity step will be to remove the variable: 'variant' from the β variable's of the model. So the second model will be the

This model doesn't remove Variant as you said above?

sensitivity model. So $log(\lambda(\beta, x_{ij})) = b_0 + b_1 x_{variant,ij} + b_2 x_{biweek,ij} + \beta_{0j} + \beta_{1j} x_{variant,j}$ is my sensitivity model.

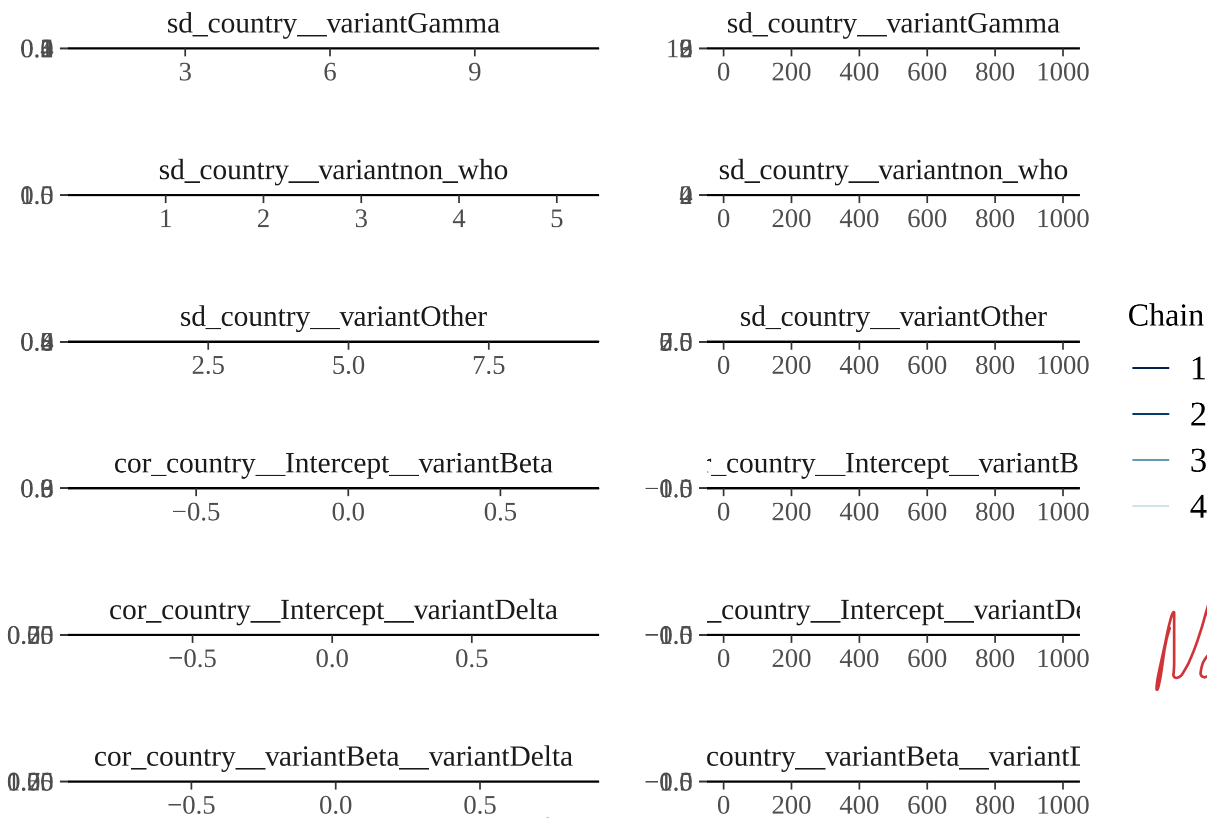
```r
model_2<-brm(cases ~ biweek+ variant+ (variant | country), data=variants_5Largest,
family = negbinomial(), prior=c(Intercept_Prior, sd_priors_Intercept,
biweek_priors, variant_priors, sd_priors_variant, shape_prior), iter=6000,
warmup = 5000, inits="0") # is my sensitivity model
```

```r
summary_model_2 <- summary(model_2)
rbind(summary_model_2$random$country[-2:-4], summary_model_2$fixed[-2:-4],
      summary_model_2$spec_pars[-2:-4])
```

```
##                                      Estimate      Rhat  Bulk_ESS  Tail_ESS
## sd(Intercept)                     1.129644686 1.0027252  859.5246 1004.9130
## sd(variantBeta)                   2.912246574 1.0032548 1391.8405 1527.5365
## sd(variantDelta)                  0.950384187 1.0009566 1135.6366  732.3082
## sd(variantGamma)                  2.314306503 1.0049157  733.8799  464.9261
## sd(variantnon_who)                0.513019589 1.0031862 1403.3087 1811.0667
## sd(variantOther)                  0.890686967 1.0005169 1134.2783 1536.3338
## cor(Intercept,variantBeta)       -0.046126442 1.0010381 1809.8329 2409.0861
## cor(Intercept,variantDelta)       0.148110032 1.0026593 2350.2085 1362.4162
## cor(variantBeta,variantDelta)    -0.036331784 1.0004665 3318.8057 2739.4096
## cor(Intercept,variantGamma)      -0.370825317 1.0010041 1742.9217 1430.2355
## cor(variantBeta,variantGamma)     0.026792097 1.0032211 3032.9879 2879.0052
## cor(variantDelta,variantGamma)   -0.214025560 1.0020114 1856.2066 2165.3566
## cor(Intercept,variantnon_who)    -0.110063870 1.0018321 3542.5879 2565.2510
## cor(variantBeta,variantnon_who)   0.052547120 1.0008143 3420.3867 2729.7107
## cor(variantDelta,variantnon_who) -0.062650310 1.0029544 3000.2901 2798.2735
## cor(variantGamma,variantnon_who)  0.102767249 1.0008724 2221.0581 2175.1269
## cor(Intercept,variantOther)      -0.255302086 1.0006715 2790.0191 2860.3425
## cor(variantBeta,variantOther)     0.142137903 1.0009563 3490.6545 1976.8580
## cor(variantDelta,variantOther)   -0.093027629 1.0017800 2940.8790 3354.8698
## cor(variantGamma,variantOther)    0.210310627 0.9997915 2005.5474 1540.7964
## cor(variantnon_who,variantOther)  0.089731054 1.0004221 1960.2584 1703.2765
## Intercept                         7.949438149 1.0063149 1011.9016  950.5655
## biweek                           -0.003429046 1.0019325 4483.9699 2876.9214
## variantBeta                      -3.901050328 1.0090140  863.9308 1177.2538
## variantDelta                      0.070179353 1.0041852 1203.1023 1685.5330
## variantGamma                     -4.142103160 1.0085216  824.1618  523.4327
## variantnon_who                   -1.576407966 1.0006785 1825.1836 1543.3800
## variantOther                     -4.150766218 1.0012704 1528.1004 1204.9333
## shape                             0.343486989 1.0010569 3512.1798  903.2146
```

All the Rhat's are 1, and all the effective sample size's for the Bulk and the Tail are very big. Which is an extremely good sign.

```r
plot(model_2, N=26, ask=FALSE)
```



No plots!

Marks (having read to end of Q1).

Modelling: I don't follow the logic of having a static variant effect but the same temporal pattern for each variant. Mathematics mostly correct up to rogue m's. NB well used. Conversion to brm. Sensitivity features its model choice. 11/14.

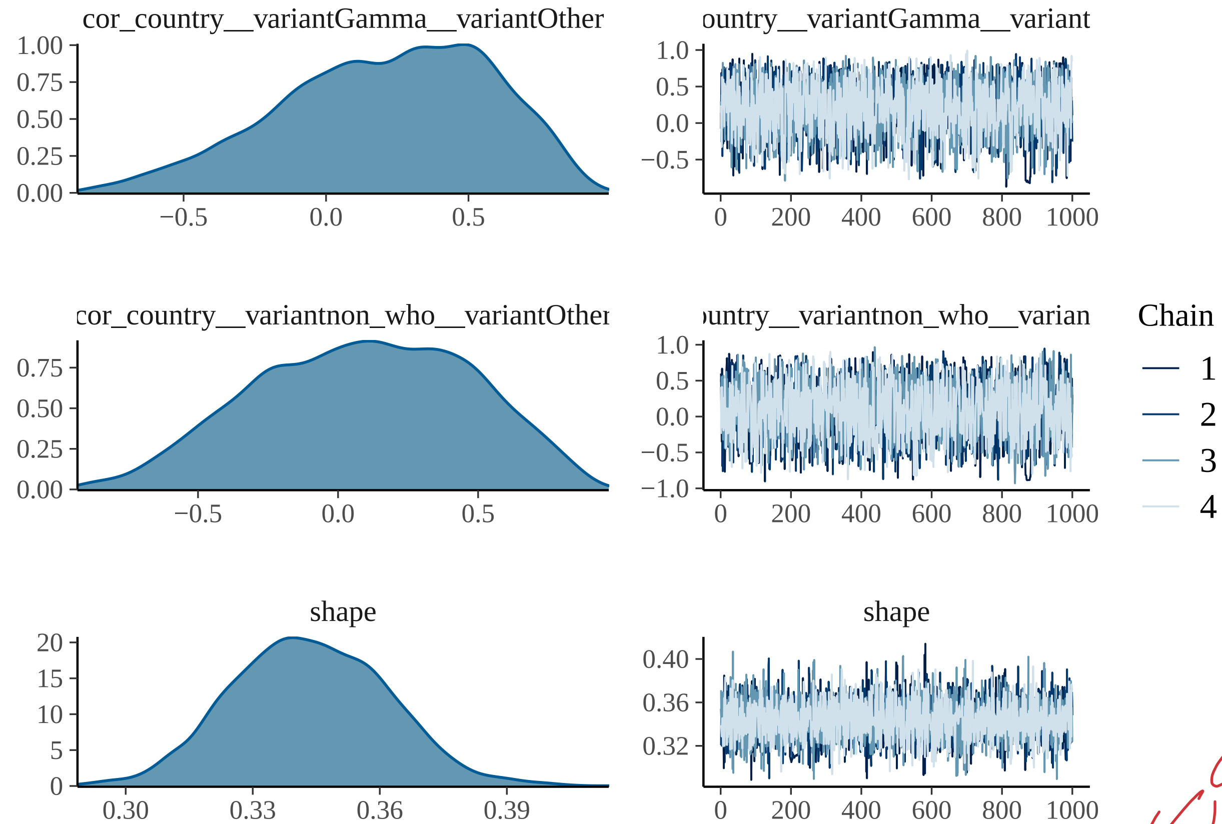
Priors: Subjective (even a couple of weakly informative) priors given. Didn't get logic of the prior on ϕ of $N(0,10)$ for group interests given $N(0,4)$ on main interest. Investigation into alternatives not looked at. 11/14

Convergence: Mixing, trace, Rhats, Neff all covered. 2 problems: 1 erroneous claim that no chains got stuck append below a chain that got stuck, and > 50% of traceplots were included AND invisible! 9/14.

Validation: In sample validation plots uninterpretable (claim data within error bars not immediately verifiable from plots. Sensitivity to grouping on binned looked at but

Couldn't assess based on plots. 5/14

Critique: Looked at priors moving was good. Conditions in chains looked at. Quite loose with language ("fairly believable", "pretty insensitive") 4/7.



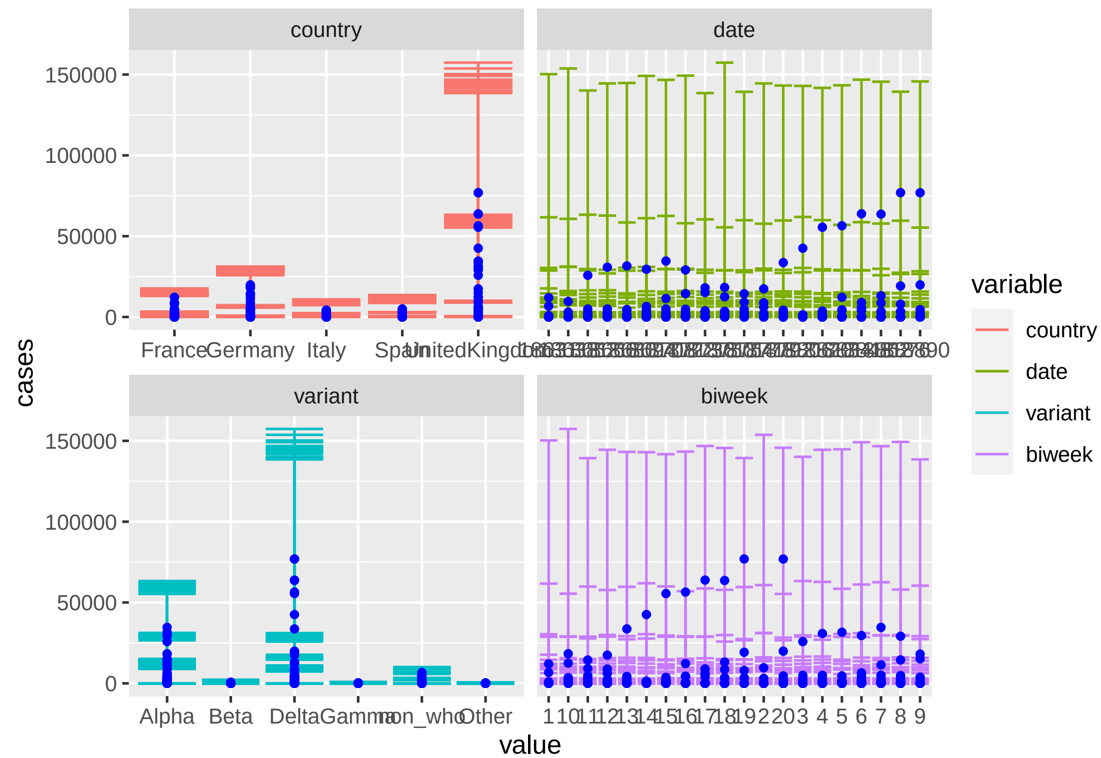
Only 3 visible.

None of the trace plots got stuck anywhere and they are all very spiky, so they have all converged. This, together with the analysis of the summary, means the model has converged very well. The intercept and variant variables have not changed in this sensitivity step.

```
case_preds <- predict(model_2, newdata = variants_5Largest)
preds <- cbind(variants_5Largest, as.tibble(case_preds)) %>%
  dplyr::select(-Est.Error) %>%
  melt(id.vars = c("Estimate", "Q2.5", "Q97.5", "cases")) %>%
  ggplot()+ geom_errorbar(aes(x=value, ymin=Q2.5, ymax=Q97.5, colour = variable))+
  geom_point(aes(x=value, y=cases), colour="blue", pch=16)+
  facet_wrap(~variable, scale="free_x")
preds
```

Presentation: Blank plots and illegible verification plots are main problem. Code is fine, mathematics should be better spaced (like individual lines for key equations.) 4/7.

Q1: 44/70



Again how can this be interpreted?

The predictions and error bars have not changed by much at all, in this sensitivity step. So my conclusions seem pretty insensitive to this main choice I have made. So my final model is my first model.

2a)

```
pred_UnitedKingdom_Delta_90 <- predict(model_1, newdata = filter(variants_5Largest,
  variant== 'Delta', country == "UnitedKingdom"),
  allow_new_levels=TRUE, prob = c(0.05, 0.95))
max(pred_UnitedKingdom_Delta_90[,"Estimate"])
```

what about biweeks 21:26.

```
## [1] 28768.01
```

Is the estimate for the peak number of sequenced Delta cases in the UK in 2021.

```
max_value <- which.max(pred_UnitedKingdom_Delta_90[,"Estimate"])
c(pred_UnitedKingdom_Delta_90[max_value,"Q5"], pred_UnitedKingdom_Delta_90[max_value,"Q95"])
```

```
##       Q5      Q95
##     10.0 129293.2
```

Missing rest of 2021

5

Is the 90% prediction interval for the peak number of sequenced Delta cases in the UK in 2021.

2b)

```r
UnitedKingdom_Delta_Pred_90 <- variants_5Largest %>%
  filter(variant== 'Delta', country == "UnitedKingdom") %>%
  mutate(Cases = pred_UnitedKingdom_Delta_90[,"Estimate"], Q5 = pred_UnitedKingdom_Delta_90[,"Q5"],
         Q95=pred_UnitedKingdom_Delta_90[,"Q95"])
UnitedKingdom_Delta_Pred_90[max_value,"date"]
```

```
## # A tibble: 1 x 1
##   date
##   <date>
## 1 2021-01-04
```

Delta didn't exist in January.

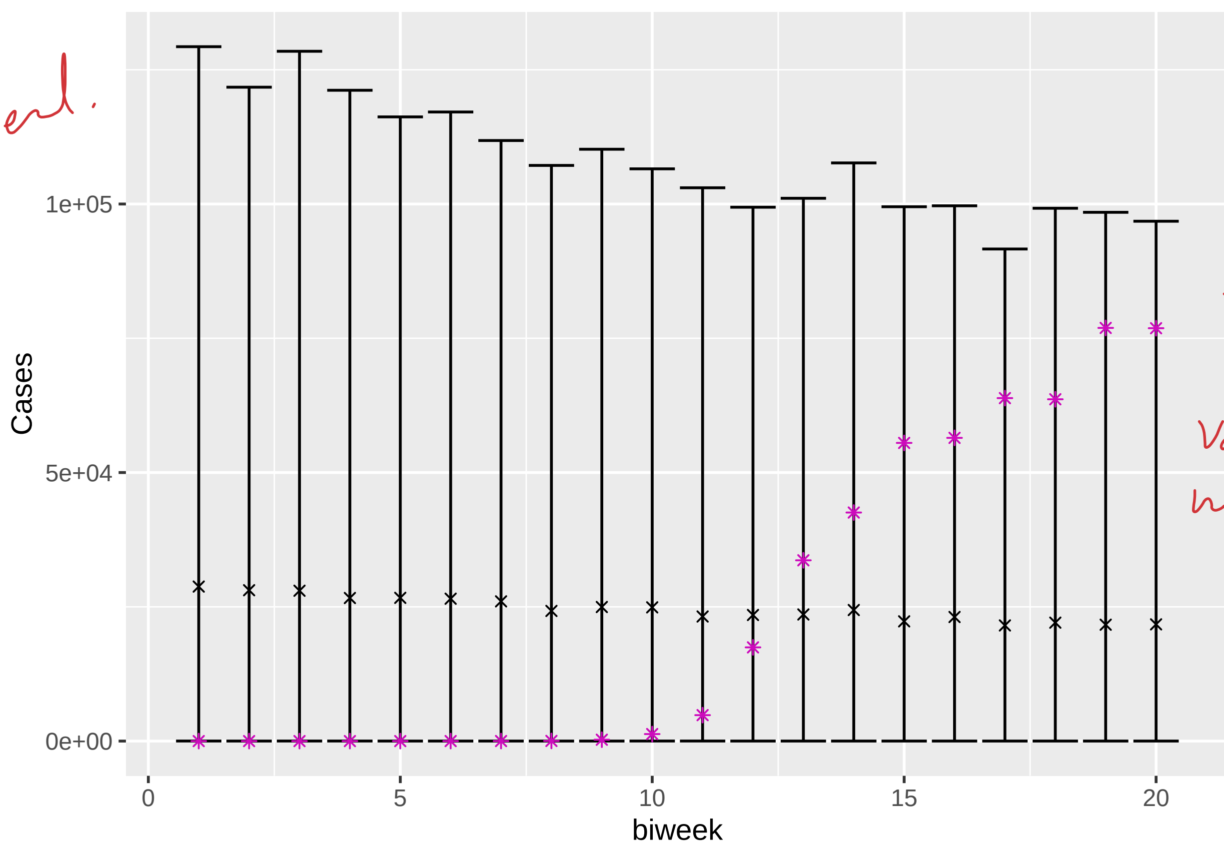
3

Is the expected date of the peak.

2c)

```r
ggplot(UnitedKingdom_Delta_Pred_90) +
  geom_errorbar(aes(x=biweek, ymin=Q5, ymax=Q95)) +
  geom_point(aes(x=biweek, y=Cases), col=1, pch=4)+
  geom_point(aes(x=biweek, y=cases), col=6,pch=8)
```

legend.



Plots like this would have improved validation work in Q1.

4

The graph shows the model has an extremely low efficacy, shown by how by the error bars are at the beginning of the year. The model is predicting the highest number of cases at the very start of the year, for the delta variant in the UK. And predicts only a decreases in the number of its cases, as the date increases.

12/15

Where as the data clearly shows a trend, that at the date increases, the number of cases for the delta variant in the UK increases. With the highest number of cases recorded, being for the most recent date. So the model predicts the complete opposite of what the true data shows, showing an extremely low efficacy. So there was no need to try to go beyond biweek 20, since the predicted estimate's from the model, only decreases as biweek increases. Your either need to show that mathematically or demonstrate it by going beyond 20.

3a)

```
Ess_Model_1 <- rbind(summary_model_1$random$country[-1:-5], summary_model_1$fixed[-1:-5],
                     summary_model_1$spec_pars[-1:-5])
Ess_Bulk_Model_1 <- Ess_Model_1[-2]
Bulk_Ess <- min(Ess_Bulk_Model_1)
pred_samples_France_Beta <- predict(model_1, newdata = filter(variants_5Largest,
     variant== 'Beta', country == "France"), allow_new_levels=TRUE, summary = FALSE)
pred_samples_France_non_who <- predict(model_1, newdata = filter(variants_5Largest,
     variant== 'non_who', country == "France"), allow_new_levels=TRUE, summary = FALSE)
n_France_Beta <- ncol(pred_samples_France_Beta)
prob_France_Beta_more_non_who <- numeric(n_France_Beta)
sd_France_Beta_more_non_who <- numeric(n_France_Beta)
for (i in 1:n_France_Beta) {
  prob_France_Beta_more_non_who[i]<-mean(pred_samples_France_Beta[,i]>pred_samples_France_non_who[,i])
  sd_France_Beta_more_non_who[i] <- sd(pred_samples_France_Beta[,i]>pred_samples_France_non_who[,i])
}
MC_error_France_Beta_more_non_who <- sd_France_Beta_more_non_who/Bulk_Ess
MC_error_France_Beta_more_non_who
```

```
##  [1] 0.001470364 0.001467010 0.001472628 0.001472417 0.001467200 0.001466625
##  [7] 0.001468844 0.001467389 0.001468219 0.001467106 0.001469871 0.001469788
## [13] 0.001470445 0.001473242 0.001469364 0.001470119 0.001472130 0.001467946
## [19] 0.001465838 0.001469871
```

```
MC_error_France_Beta_more_non_who<0.01
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE
```

Estimates of dominance probabilities all have Monte Carlo error less than 0.01.

```
prob_France_Beta_more_non_who
```

```
##  [1] 0.44550 0.43600 0.45300 0.45225 0.43650 0.43500 0.44100 0.43700 0.43925
## [10] 0.43625 0.44400 0.44375 0.44575 0.45525 0.44250 0.44475 0.45125 0.43850
## [19] 0.43300 0.44400
```

```
prob_France_Beta_more_non_who>0.5
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
prob_France_Beta_more_non_who_True_or_False <- as.numeric(prob_France_Beta_more_non_who>0.5)
as.numeric(Position(function(x) x>0, prob_France_Beta_more_non_who_True_or_False))
```

```
## [1] NA
```

Is the first biweek, my model predicts; the Beta variant first dominates the original Covid-19 strain, in France, if the output is 'NA', then the model predicts that Beta variant never dominates the original Covid-19 strain. So my model never predicts that Beta will dominate the original strain in France.

5

3b)

```r
pred_samples_Italy_Delta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Delta', country == "Italy"), allow_new_levels=TRUE, summary = FALSE)
pred_samples_Italy_Gamma <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Gamma', country == "Italy"), allow_new_levels=TRUE, summary = FALSE)
n_Italy_Delta <- ncol(pred_samples_Italy_Delta)
prob_Italy_Delta_more_Gamma <- numeric(n_Italy_Delta)
sd_Italy_Delta_more_Gamma <- numeric(n_Italy_Delta)
for (i in 1:n_Italy_Delta) {
  prob_Italy_Delta_more_Gamma[i]<-mean(pred_samples_Italy_Delta[,i]>pred_samples_Italy_Gamma[,i])
  sd_Italy_Delta_more_Gamma[i] <- sd(pred_samples_Italy_Delta[,i]>pred_samples_Italy_Gamma[,i])
}
MC_error_Italy_Delta_more_Gamma <- sd_Italy_Delta_more_Gamma/Bulk_Ess
MC_error_Italy_Delta_more_Gamma
```

```
##  [1] 0.001255263 0.001286089 0.001302054 0.001311853 0.001301655 0.001296826
##  [7] 0.001302453 0.001300856 0.001293972 0.001302453 0.001295606 0.001293562
## [13] 0.001321320 0.001318326 0.001309532 0.001291917 0.001295606 0.001297636
## [19] 0.001304041 0.001300054
```

```r
MC_error_Italy_Delta_more_Gamma<0.01
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE
```

Estimates of dominance probabilities all have Monte Carlo error less than 0.01.

```r
prob_Italy_Delta_more_Gamma
```

```
##  [1] 0.76450 0.74700 0.73725 0.73100 0.73750 0.74050 0.73700 0.73800 0.74225
## [10] 0.73700 0.74125 0.74250 0.72475 0.72675 0.73250 0.74350 0.74125 0.74000
## [19] 0.73600 0.73850
```

```r
prob_Italy_Delta_more_Gamma>0.5
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE
```

```
prob_Italy_Delta_more_Gamma_True_or_False <- as.numeric(prob_Italy_Delta_more_Gamma>0.5)
as.numeric(Position(function(x) x>0, prob_Italy_Delta_more_Gamma_True_or_False))
```

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

Is the first biweek, my model predicts; the Delta variant dominates the Gamma variant, in Italy, if the output is 'NA', then the model predicts that Delta variant never dominates the Gamma variant. So my model predicts that Gamma variant will never dominate the Delta variant in Italy. 4

not NA so date?

3c)

```
pred_samples_France_Delta <- predict(model_1, newdata = filter(variants_5Largest,
           variant== 'Delta', country == "France"),  allow_new_levels=TRUE, summary = FALSE)
variant_names <- c("Alpha", "Beta", "Gamma", "non_who", "Other")
variant_NotDelta_length <- length(variant_names)
France_Delta_dominates <- numeric(variant_NotDelta_length)
for (j in 1:variant_NotDelta_length) {
  pred_samples_France_NotDelta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== variant_names[j], country == "France"), allow_new_levels=TRUE, summary = FALSE)
  n_France_Delta <- ncol(pred_samples_France_Delta)
  prob_France_Delta_MoreThan <- numeric(n_France_Delta)
  for (i in 1:n_France_Delta) {
    prob_France_Delta_MoreThan[i] <-
      mean(pred_samples_France_Delta[,i]>pred_samples_France_NotDelta[,i])
  }
  prob_France_Delta_MoreThan_True_or_False <- as.numeric(prob_France_Delta_MoreThan>0.5)
  France_Delta_dominates[j]<-
    as.numeric(Position(function(x) x>0, prob_France_Delta_MoreThan_True_or_False))
}
max(France_Delta_dominates)
```

```
## [1] 2
```

Is the first biweek, my model predicts; the Delta variant first dominates all the other Covid-19 variants, in France, if the output is 'NA', then the model predicts that Delta variant never dominates all the other Covid-19 variants, in France.

```
pred_samples_Germany_Delta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Delta', country == "Germany"),  allow_new_levels=TRUE, summary = FALSE)
variant_names <- c("Alpha", "Beta", "Gamma", "non_who", "Other")
variant_NotDelta_length <- length(variant_names)
Germany_Delta_dominates <- numeric(variant_NotDelta_length)
for (j in 1:variant_NotDelta_length) {
  pred_samples_Germany_NotDelta <- predict(model_1, newdata = filter(variants_5Largest,
      variant== variant_names[j], country == "Germany"), allow_new_levels=TRUE, summary = FALSE)
  n_Germany_Delta <- ncol(pred_samples_Germany_Delta)
  prob_Germany_Delta_MoreThan <- numeric(n_Germany_Delta)
  for (i in 1:n_Germany_Delta) {
    prob_Germany_Delta_MoreThan[i] <-
      mean(pred_samples_Germany_Delta[,i]>pred_samples_Germany_NotDelta[,i])
```

```
  }
  prob_Germany_Delta_MoreThan_True_or_False <- as.numeric(prob_Germany_Delta_MoreThan>0.5)
  Germany_Delta_dominates[j] <-
    as.numeric(Position(function(x) x>0, prob_Germany_Delta_MoreThan_True_or_False))
}
max(Germany_Delta_dominates)
```

```
## [1] 11
```

Is the first biweek, my model predicts; the Delta variant first dominates all the other Covid-19 variants, in Germany, if the output is 'NA', then the model predicts that Delta variant never dominates all the other Covid-19 variants, in Germany.

```
pred_samples_Italy_Delta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Delta', country == "Italy"),  allow_new_levels=TRUE, summary = FALSE)
variant_names <- c("Alpha", "Beta", "Gamma", "non_who", "Other")
variant_NotDelta_length <- length(variant_names)
Italy_Delta_dominates <- numeric(variant_NotDelta_length)
for (j in 1:variant_NotDelta_length) {
  pred_samples_Italy_NotDelta <- predict(model_1, newdata = filter(variants_5Largest,
      variant== variant_names[j], country == "Italy"), allow_new_levels=TRUE, summary = FALSE)
  n_Italy_Delta <- ncol(pred_samples_Italy_Delta)
  prob_Italy_Delta_MoreThan <- numeric(n_Italy_Delta)
  for (i in 1:n_Italy_Delta) {
    prob_Italy_Delta_MoreThan[i] <-
      mean(pred_samples_Italy_Delta[,i]>pred_samples_Italy_NotDelta[,i])
  }
  prob_Italy_Delta_MoreThan_True_or_False <- as.numeric(prob_Italy_Delta_MoreThan>0.5)
  Italy_Delta_dominates[j] <-
    as.numeric(Position(function(x) x>0, prob_Italy_Delta_MoreThan_True_or_False))
}
max(Italy_Delta_dominates)
```

```
## [1] NA
```

Is the first biweek, my model predicts; the Delta variant first dominates all the other Covid-19 variants, in Italy, if the output is 'NA', then the model predicts that Delta variant never dominates all the other Covid-19 variants, in Italy.

```
pred_samples_Spain_Delta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Delta', country == "Spain"),  allow_new_levels=TRUE, summary = FALSE)
variant_names <- c("Alpha", "Beta", "Gamma", "non_who", "Other")
variant_NotDelta_length <- length(variant_names)
Spain_Delta_dominates <- numeric(variant_NotDelta_length)
for (j in 1:variant_NotDelta_length) {
  pred_samples_Spain_NotDelta <- predict(model_1, newdata = filter(variants_5Largest,
      variant== variant_names[j], country == "Spain"), allow_new_levels=TRUE, summary = FALSE)
  n_Spain_Delta <- ncol(pred_samples_Spain_Delta)
  prob_Spain_Delta_MoreThan <- numeric(n_Spain_Delta)
  for (i in 1:n_Spain_Delta) {
    prob_Spain_Delta_MoreThan[i] <-
      mean(pred_samples_Spain_Delta[,i]>pred_samples_Spain_NotDelta[,i])
```

```
  }
  prob_Spain_Delta_MoreThan_True_or_False <- as.numeric(prob_Spain_Delta_MoreThan>0.5)
  Spain_Delta_dominates[j] <-
    as.numeric(Position(function(x) x>0, prob_Spain_Delta_MoreThan_True_or_False))
}
max(Spain_Delta_dominates)
```

```
## [1] NA
```

Is the first biweek, my model predicts; the Delta variant first dominates all the other Covid-19 variants, in Spain, if the output is 'NA', then the model predicts that Delta variant never dominates all the other Covid-19 variants, in Spain.

```
pred_samples_UnitedKingdom_Delta <- predict(model_1, newdata = filter(variants_5Largest,
   variant== 'Delta', country == "UnitedKingdom"),  allow_new_levels=TRUE, summary = FALSE)
variant_names <- c("Alpha", "Beta", "Gamma", "non_who", "Other")
variant_NotDelta_length <- length(variant_names)
UnitedKingdom_Delta_dominates <- numeric(variant_NotDelta_length)
for (j in 1:variant_NotDelta_length) {
  pred_samples_UnitedKingdom_NotDelta <- predict(model_1, newdata = filter(variants_5Largest,
      variant==variant_names[j],country=="UnitedKingdom"),allow_new_levels=TRUE,summary = FALSE)
  n_UnitedKingdom_Delta <- ncol(pred_samples_UnitedKingdom_Delta)
  prob_UnitedKingdom_Delta_MoreThan <- numeric(n_UnitedKingdom_Delta)
  for (i in 1:n_UnitedKingdom_Delta) {
    prob_UnitedKingdom_Delta_MoreThan[i] <-
      mean(pred_samples_UnitedKingdom_Delta[,i]>pred_samples_UnitedKingdom_NotDelta[,i])
  }
  prob_UnitedKingdom_Delta_MoreThan_True_or_False <-
    as.numeric(prob_UnitedKingdom_Delta_MoreThan>0.5)
  UnitedKingdom_Delta_dominates[j] <-
    as.numeric(Position(function(x) x>0, prob_UnitedKingdom_Delta_MoreThan_True_or_False))
}
max(UnitedKingdom_Delta_dominates)
```

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

Is the first biweek, my model predicts; the Delta variant first dominates all the other Covid-19 variants, in the United Kingdom, if the output is 'NA', then the model predicts that Delta variant never dominates all the other Covid-19 variants, in the United Kingdom.

Code is fine but no dates. 4.

13/15