Estimating the difference of bicycle lanes use between sections

BDA3 Project

anonymous

0.1 Introduction

The Chilean government wants to justify the improvement of a cycle lane across the metropolitan region. This region has the particular characteristic that its demographics tends to separate itself through sections of its territory. It is of interest to determine if there is a difference of the use of bicycles between sections of the region, and if we can produce a model that determines the lane demand in a given point.

> **Objective**
>
> Determine if there is a difference of usage between sections of the city.

0.2 Data

As data is bounded to budget, there are few data points throughout the lane, taken by counting the number of bicycles passing at a certain time through an intersection of streets. The data presented shows the intersection (an id), the section, the date in which was taken the measurement, the time of day (HORA.INICIO), which quarter of hour is taken (Cuarto), and finally, the number of bicycles counted whether on the street or the cycle lane:

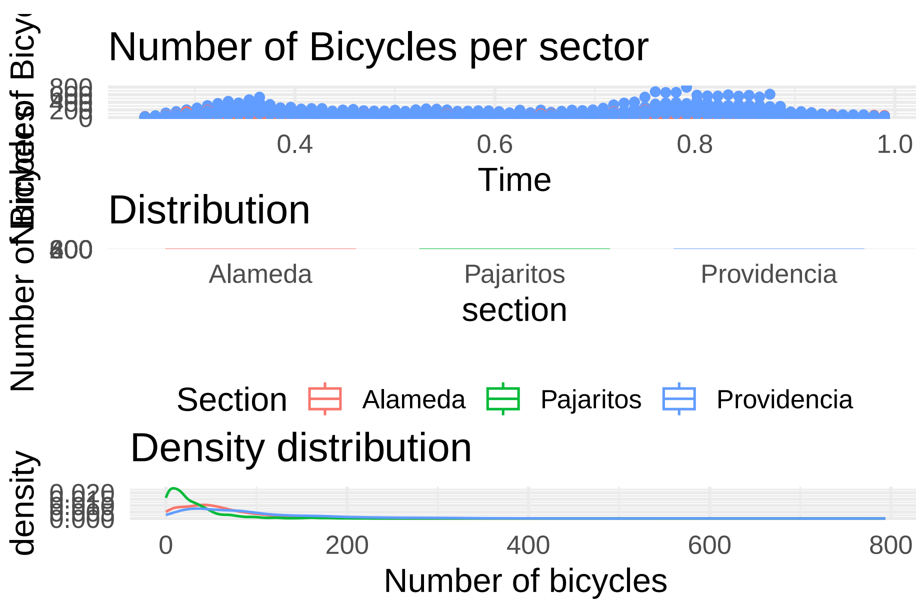
```r
library(bayesplot)
library(cmdstanr)

library(ggdist) # for stat_dotsinterval
library(posterior)
library(broom.mixed)
library(tidyverse)
library(colorspace)
library(brms)
```

```r
options(mc.cores = parallel::detectCores()-4)

# Set more readable themes with bigger font for plotting packages.
ggplot2::theme_set(theme_minimal(base_size = 14))
bayesplot::bayesplot_theme_set(theme_minimal(base_size = 14))

# This registers CmdStan as the backend for compiling cmdstan-chunks.
check_cmdstan_toolchain(fix = TRUE, quiet = TRUE)
register_knitr_engine(override = FALSE)

url<-"../../../extradrive1/00. Instituto Data Science UDD/FIC/Datos/Transfer/"

bbdd<-read_csv("/media/tom/extradrive1/Estudios/BDA/BicicleLane.csv")
head(bbdd)
```

```
# A tibble: 6 x 7
     PC Tramo     FECHA      HORA.INICIO Cuarto Street  lane
  <dbl> <chr>     <date>           <dbl>  <dbl>  <dbl> <dbl>
1     1 Pajaritos 2023-01-11       0.25     6.1      0     9
2     1 Pajaritos 2023-01-11       0.260    6.2      6    26
3     1 Pajaritos 2023-01-11       0.271    6.3     18    42
4     1 Pajaritos 2023-01-11       0.281    6.4     23    54
5     1 Pajaritos 2023-01-11       0.292    7.1     24    56
6     1 Pajaritos 2023-01-11       0.302    7.2     26    75
```

```r
bbdd$Bikes<-bbdd$Street+bbdd$lane


b1<-bbdd %>%
    ggplot(aes(x=HORA.INICIO,
               y=Bikes,
               col=as.factor(Tramo)))+
    geom_point()+
    ggtitle("Number of Bicycles per sector")+
    scale_y_continuous("Number of Bicycles")+
    scale_x_continuous("Time")+
    scale_color_discrete(name = "Section")+
    theme(legend.position="none")

b2<-bbdd %>%
    ggplot(aes(x=Tramo,
               y=Bikes,
               col=as.factor(Tramo)))+
```

```
    geom_boxplot()+
    ggtitle("Distribution")+
    scale_y_continuous("Number of Bicycles")+
    scale_x_discrete("section")+
    scale_color_discrete(name = "Section")+
    theme(legend.position="bottom")


b3<-bbdd %>%
    ggplot(aes(Bikes,
               col=as.factor(Tramo)))+
    geom_density()+ggtitle("Density distribution")+
    scale_x_continuous("Number of bicycles")+
    scale_color_discrete(name = "Section")+
    theme(legend.position = 'none')
ggpubr::ggarrange(b1,b2,b3,ncol=1)
```



The above graph shows the number of bicycles is somewhat similar between sections, yet `Providencia` seems to have a larger dispersion, specially at early and late values. Checking the data we can see that it starts at 0.25, but first `Cuarto` is 6.1, meaning 06:15 AM. We won‘t be changing this as it is of our interest to use the number as is and not a date.

The boxplot shows that `Providencia` has a lot of dispersion, although its median is similar to alameda and has a suspicion of difference with Pajaritos.

The density functions shows that Pajaritos has a great mass less than 100 or even less than the other lanes, which has considerable mass over this number.

0.3 Modelling

First we are going to set a generic seed so the project will be reproducible.

```
set.seed(1234)
```

Our first model is the generic normal model, but it has the particularity of been a mixed effects model. In this case, the model is a hierarchical model where each section has its own intercept, and is also regulated by the time in the intersections.

```
fit<-brm(Bikes~HORA.INICIO+Tramo+(1|Tramo)+(HORA.INICIO+Tramo|PC),
         data = bbdd,
         iter=3000,
         thin=10)
```

```
Warning: There were 15 divergent transitions after warmup. See
https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
to find out why this is a problem and how to eliminate them.
```

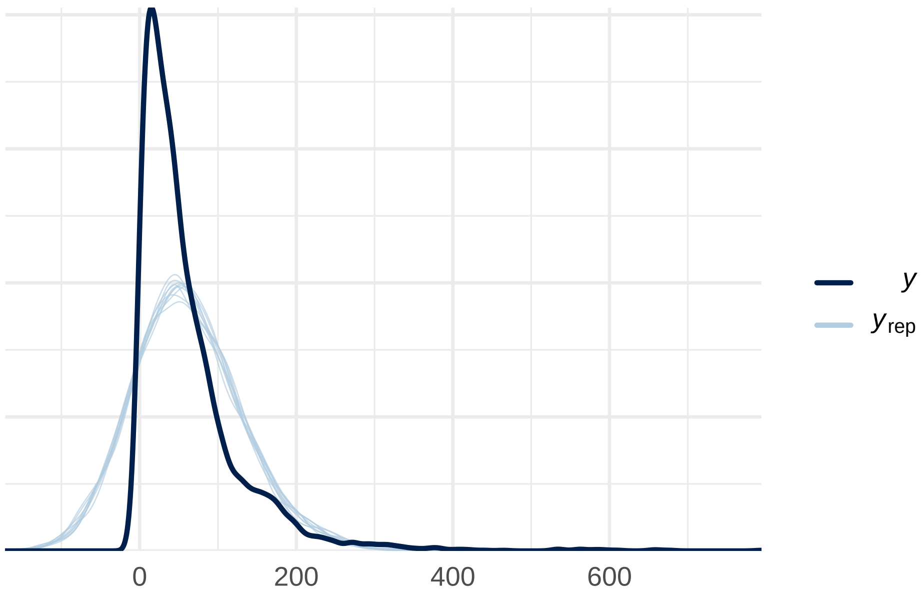
```
Warning: Examine the pairs() plot to diagnose sampling problems
```

```
Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and ta
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#tail-ess
```

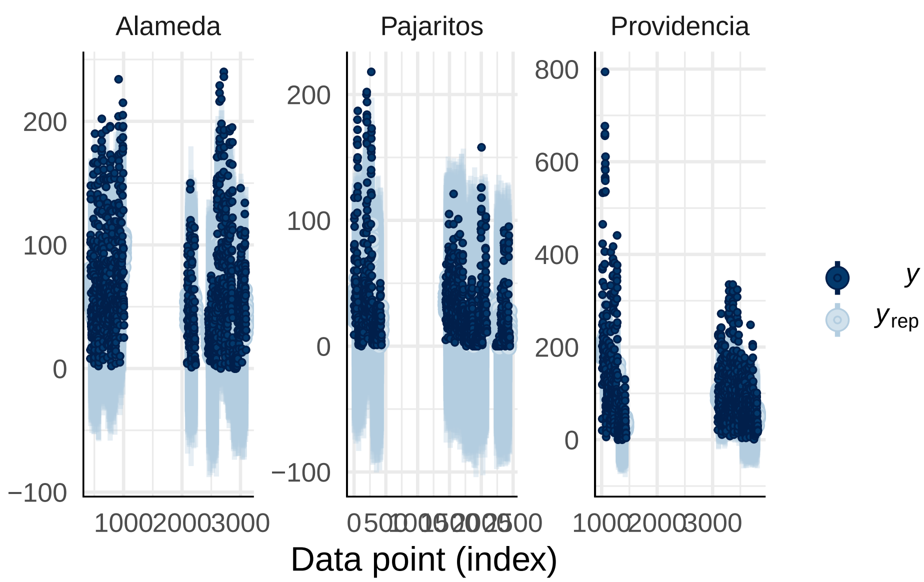
Given that brm is very chatty, I'm hidding its messages.

```
#plot(fit)
pp_check(fit)
```

```
Using 10 posterior draws for ppc type 'dens_overlay' by default.
```




```r
bayesplot::pp_check(fit,ndraws = 100,type = "intervals_grouped",group = "Tramo")
```



As expected, this model does not predict well the number of bikes, giving negative results as an option.

0.4 Log Normal Hierarchical model

Given the lack of proper fit, a transformation is required. A first approach is to apply the lognormal distribution. As there are 0 values, we are going to make a trick to fit. adding 1 to every value, the predicted values then would be the exponential result minus 1.

```
bbdd$BikesLog<-bbdd$Bikes+1
fit_log<-brm(BikesLog~HORA.INICIO +Tramo+(1|Tramo)+(HORA.INICIO +Tramo|PC),
             family = "lognormal",
             data=bbdd,
             iter=4000,
             thin=5)
```

```
Warning: There were 185 divergent transitions after warmup. See
https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
to find out why this is a problem and how to eliminate them.

Warning: There were 535 transitions after warmup that exceeded the maximum treedepth. Increa
https://mc-stan.org/misc/warnings.html#maximum-treedepth-exceeded

Warning: Examine the pairs() plot to diagnose sampling problems

Warning: The largest R-hat is 1.08, indicating chains have not mixed.
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#r-hat

Warning: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and median
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#bulk-ess

Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and ta
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#tail-ess
```

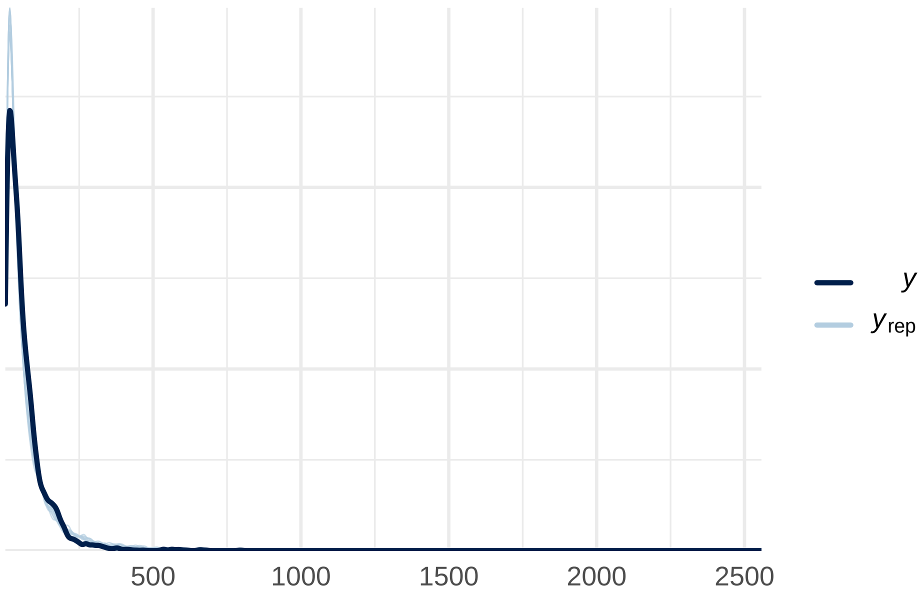
```
summary(fit_log)
```

```
Warning: Parts of the model have not converged (some Rhats are > 1.05). Be
careful when analysing the results! We recommend running more iterations and/or
setting stronger priors.

Warning: There were 185 divergent transitions after
warmup. Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: BikesLog ~ HORA.INICIO + Tramo + (1 | Tramo) + (HORA.INICIO + Tramo | PC)
   Data: bbdd (Number of observations: 3816)
```

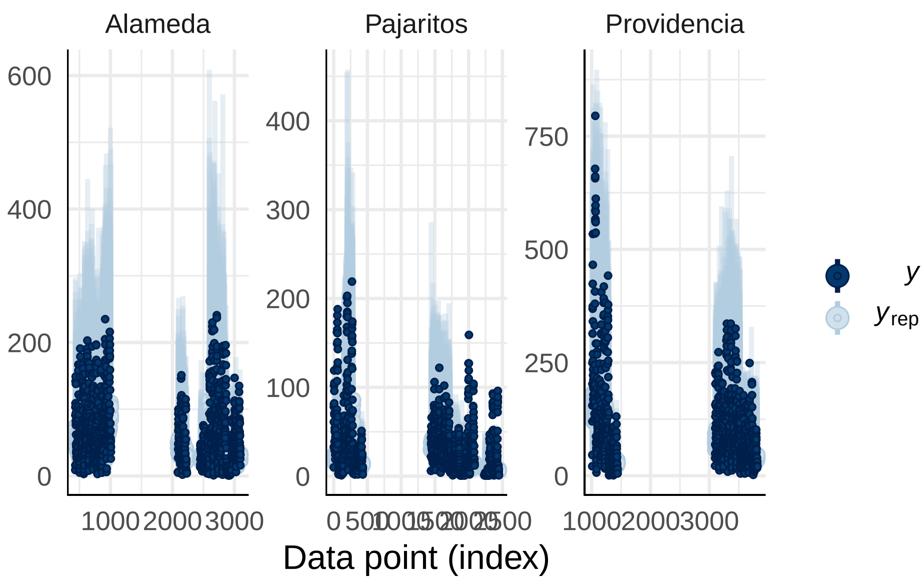
```
 Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 5;
        total post-warmup draws = 1600

Group-Level Effects:
~PC (Number of levels: 25)
                                     Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                            0.67      0.14     0.45     0.99 1.02
sd(HORA.INICIO)                          0.65      0.13     0.44     0.94 1.06
sd(TramoPajaritos)                       0.56      0.41     0.03     1.54 1.01
sd(TramoProvidencia)                     0.59      0.41     0.03     1.56 1.01
cor(Intercept,HORA.INICIO)              -0.47      0.23    -0.80     0.03 1.06
cor(Intercept,TramoPajaritos)           -0.18      0.43    -0.86     0.69 1.01
cor(HORA.INICIO,TramoPajaritos)          0.21      0.41    -0.67     0.87 1.02
cor(Intercept,TramoProvidencia)         -0.06      0.42    -0.79     0.78 1.00
cor(HORA.INICIO,TramoProvidencia)       -0.14      0.46    -0.84     0.74 1.02
cor(TramoPajaritos,TramoProvidencia)    -0.04      0.46    -0.80     0.80 1.04
                                     Bulk_ESS Tail_ESS
sd(Intercept)                             227     1419
sd(HORA.INICIO)                            54       40
sd(TramoPajaritos)                        465     1161
sd(TramoProvidencia)                      323     1259
cor(Intercept,HORA.INICIO)                 69       40
cor(Intercept,TramoPajaritos)             573     1405
cor(HORA.INICIO,TramoPajaritos)          1528     1656
cor(Intercept,TramoProvidencia)          1228     1427
cor(HORA.INICIO,TramoProvidencia)         205     1494
cor(TramoPajaritos,TramoProvidencia)      110     1207

~Tramo (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.83      1.66     0.06     6.57 1.08       39       17

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            3.75      1.59     0.58     7.12 1.04       86     1114
HORA.INICIO         -0.17      0.14    -0.46     0.11 1.05     1373     1273
TramoPajaritos      -0.80      2.17    -5.75     3.89 1.02      393     1293
TramoProvidencia     0.98      2.90    -4.26     8.62 1.08       46       15

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.88      0.01     0.86     0.90 1.01      809     1544

Draws were sampled 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).
```

```
#plot(fit_log)
pp_check(fit_log)
```

```
Using 10 posterior draws for ppc type 'dens_overlay' by default.
```




```
bayesplot::pp_check(fit_log,ndraws = 100,type = "intervals_grouped",group = "Tramo")
```



This model seems more appropriate and we could stop here but there is a couple of models that we should discuss. we could improve our model using priors although non informative, and as the number of bikes should follow a Poisson distribution, so fitting one of these to the mix and compare which one is better.

0.5 Log normal with priors

As we saw from the graph, the graph, the mean seems to be near 150 bikes every 15 minutes which log is 5.01. Giving some variance to the time relation, without incurring in more complex models as would be including autoregresive functions, we give another 150 to the time frame per 15 minutes. As a user of the lanes, I could think that Providencia and Alameda sections of the streets tend to be similar, so I would not suggest a difference between these parts, but I would think that there could be less bicycles in Pajaritos, which is far from the center of the city. Then the priors would be centered in 1 and 0.6.

```
fit_log_priors<-brm(BikesLog~HORA.INICIO +Tramo+(1|Tramo)+(HORA.INICIO +Tramo|PC),
             family = "lognormal",
             data =bbdd,
             iter = 4000,
             thin = 5,
               prior=c(
                 prior(normal(5.01,sigma), class="Intercept"),
                 # prior for mu_0, log of bikes+1 a guess of 149 bikes per 15 minutes
                 prior(normal(5,1000), class=b,coef=HORA.INICIO),
                 #prior for beta, Time related
```

```r
                    prior(normal(0.6,1000), class=b,coef=TramoPajaritos),
                    # small deviance from intercept
                    prior(normal(1,1000), class=b,coef=TramoProvidencia),
                    # no deviance from intercept
                    prior(inv_gamma(0.01,0.01), class="sigma"),
                    # prior for sigma, unkown.
                    prior(inv_gamma(0.01,0.01), class="sd")
    ),
    backend = "cmdstanr")
```

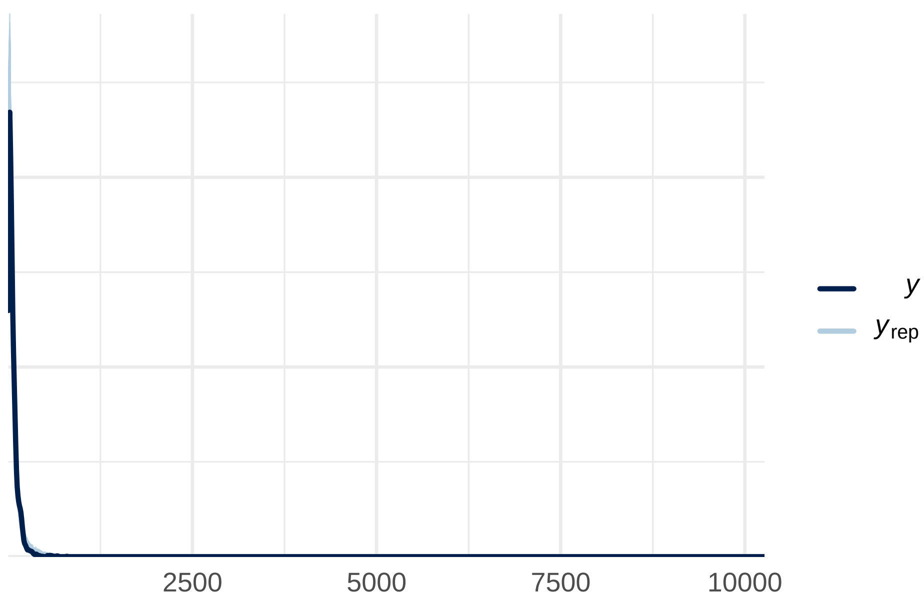
```r
summary(fit_log_priors)
```

```
Warning: There were 119 divergent transitions after
warmup. Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

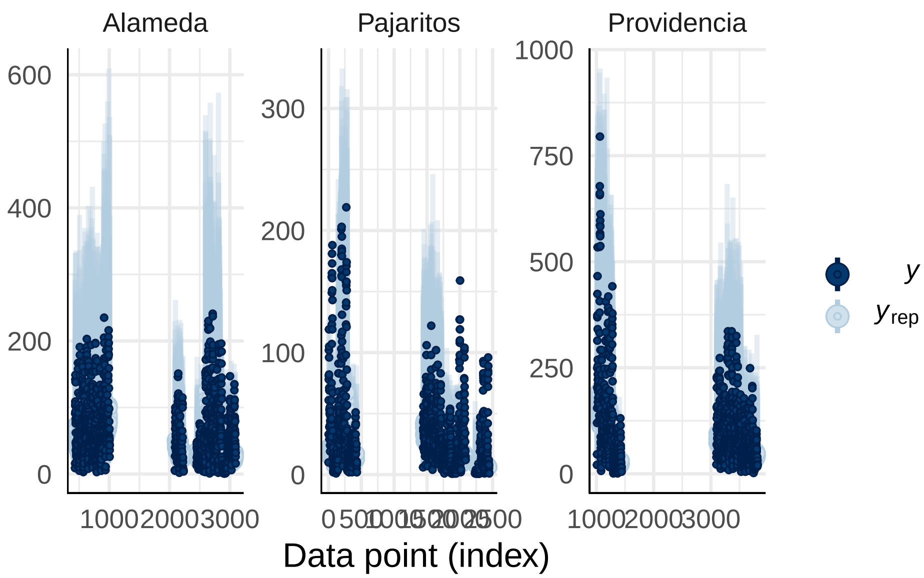
```
 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: BikesLog ~ HORA.INICIO + Tramo + (1 | Tramo) + (HORA.INICIO + Tramo | PC)
   Data: bbdd (Number of observations: 3816)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 5;
         total post-warmup draws = 1600

Group-Level Effects:
~PC (Number of levels: 25)
                                     Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                            0.66      0.12     0.47     0.93 1.01
sd(HORA.INICIO)                          0.64      0.12     0.45     0.90 1.00
sd(TramoPajaritos)                       0.25      0.32     0.01     1.14 1.00
sd(TramoProvidencia)                     0.25      0.31     0.01     1.04 1.00
cor(Intercept,HORA.INICIO)              -0.45      0.19    -0.77    -0.03 1.00
cor(Intercept,TramoPajaritos)           -0.08      0.44    -0.81     0.79 1.00
cor(HORA.INICIO,TramoPajaritos)          0.12      0.44    -0.76     0.84 1.00
cor(Intercept,TramoProvidencia)          0.01      0.43    -0.77     0.80 1.00
cor(HORA.INICIO,TramoProvidencia)       -0.06      0.45    -0.82     0.80 1.00
cor(TramoPajaritos,TramoProvidencia)    -0.01      0.45    -0.82     0.80 1.00
                                     Bulk_ESS Tail_ESS
sd(Intercept)                             884     1495
sd(HORA.INICIO)                          1366     1415
sd(TramoPajaritos)                        577      354
sd(TramoProvidencia)                      622      806
cor(Intercept,HORA.INICIO)                871     1392
cor(Intercept,TramoPajaritos)            1137     1375
cor(HORA.INICIO,TramoPajaritos)          1352     1615
```

```
cor(Intercept,TramoProvidencia)          1222     1178
cor(HORA.INICIO,TramoProvidencia)        1109     1231
cor(TramoPajaritos,TramoProvidencia)     1477     1448

~Tramo (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.38      0.63     0.01     2.23 1.05       91       83

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            4.07      0.50     3.24     5.30 1.02      835      496
HORA.INICIO         -0.17      0.15    -0.45     0.11 1.00     1311     1514
TramoPajaritos      -0.91      0.79    -2.60     0.79 1.03      705      220
TramoProvidencia     0.36      0.74    -1.23     1.90 1.03     1024      579

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.88      0.01     0.86     0.90 1.00     1647     1345

Draws were sampled using sample(hmc). 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).
```

```
#plot(fit_log_priors)
pp_check(fit_log_priors,ndraws = 100)
```



```
pp_check(fit_log_priors,ndraws = 100,type = "intervals_grouped",group = "Tramo")
```



As we can see, the it is interesting that there are more divergent results within the chains, although the pp_check suggest a good fit.

0.6 Discretization. A Poisson model

As stated before, counting the number of bikes suggests a Poisson model. Hence, we start fitting one with the same structure as before.

```
fit_pois<-brm(Bikes~HORA.INICIO+Tramo+(1|Tramo)+(HORA.INICIO+Tramo|PC),
              family = "poisson",
              data =bbdd,iter = 4000,
              thin=5,
  backend = "cmdstanr")
```

In this case, there is a 6\% of divergence which is interesting. Also there is a computational warning that I need to address in future studies.

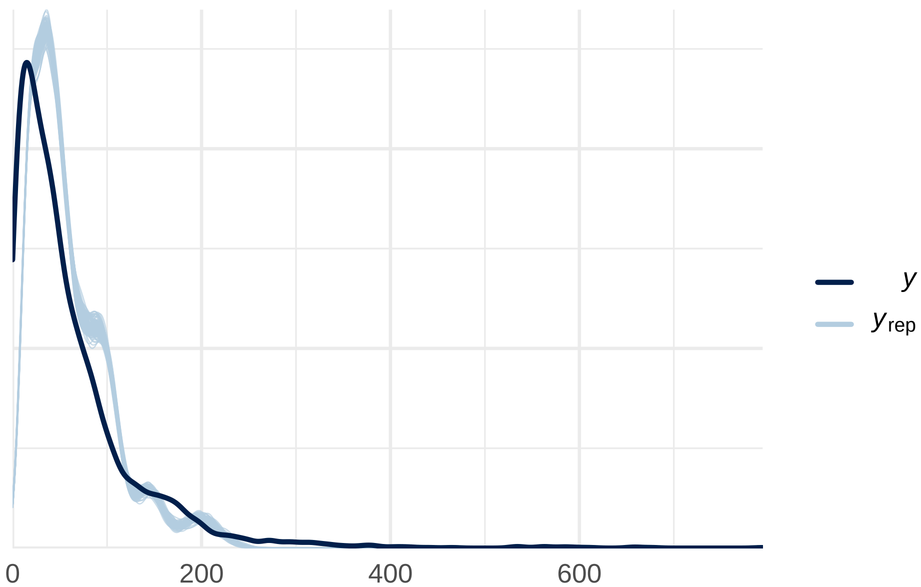
```
summary(fit_pois)
```

```
Warning: Parts of the model have not converged (some Rhats are > 1.05). Be
careful when analysing the results! We recommend running more iterations and/or
setting stronger priors.
```

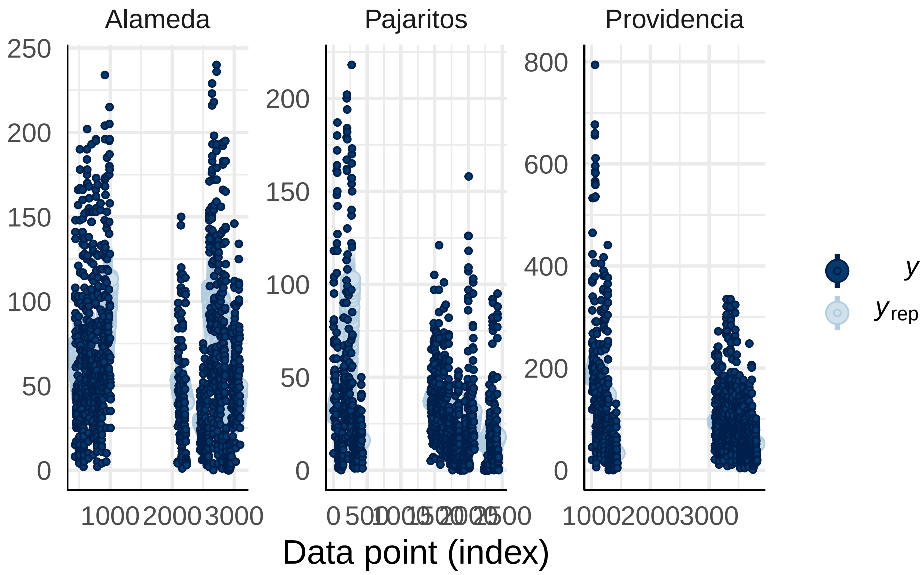
```
Warning: There were 77 divergent transitions after
warmup. Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

 Family: poisson
  Links: mu = log
Formula: Bikes ~ HORA.INICIO + Tramo + (1 | Tramo) + (HORA.INICIO + Tramo | PC)
   Data: bbdd (Number of observations: 3816)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 5;
         total post-warmup draws = 1600

Group-Level Effects:
~PC (Number of levels: 25)
                                     Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                            0.56      0.11     0.39     0.81 1.03
sd(HORA.INICIO)                          0.58      0.10     0.43     0.82 1.05
sd(TramoPajaritos)                       0.35      0.28     0.02     1.05 1.03
sd(TramoProvidencia)                     0.44      0.34     0.03     1.27 1.04
cor(Intercept,HORA.INICIO)              -0.36      0.21    -0.71     0.07 1.03
cor(Intercept,TramoPajaritos)           -0.13      0.41    -0.81     0.72 1.01
cor(HORA.INICIO,TramoPajaritos)          0.11      0.40    -0.69     0.78 1.01
cor(Intercept,TramoProvidencia)          0.02      0.43    -0.72     0.84 1.03
cor(HORA.INICIO,TramoProvidencia)        0.05      0.45    -0.80     0.82 1.02
cor(TramoPajaritos,TramoProvidencia)     0.02      0.44    -0.79     0.84 1.02
                                     Bulk_ESS Tail_ESS
sd(Intercept)                             137      379
sd(HORA.INICIO)                            83      112
sd(TramoPajaritos)                        133      520
sd(TramoProvidencia)                      110      318
cor(Intercept,HORA.INICIO)                 84      199
cor(Intercept,TramoPajaritos)             293      650
cor(HORA.INICIO,TramoPajaritos)           283      464
cor(Intercept,TramoProvidencia)           193      376
cor(HORA.INICIO,TramoProvidencia)         327      349
cor(TramoPajaritos,TramoProvidencia)      318      623

~Tramo (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.28      1.07     0.06     3.99 1.04      164      468

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            4.08      1.16     1.60     6.80 1.04      215      407
HORA.INICIO          0.13      0.11    -0.09     0.34 1.04       71      250
TramoPajaritos      -1.14      1.81    -5.51     2.26 1.03      278      460
TramoProvidencia     0.35      1.84    -3.65     4.49 1.03      229      160
```

```
Draws were sampled using sample(hmc). 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).
```

```
#plot(fit_pois)
pp_check(fit_pois,ndraws = 100)
```




```
pp_check(fit_pois,ndraws = 100,type = "intervals_grouped",group = "Tramo")
```



It is very interesting that there are parts that did not converge. The suggestion of stronger priors is exiting but as this is the first attempt I will leave it as is.

The posterior predictive check suggests that there is overdispersion in the data. This makes us think that the first graph also suggests it. Given that the boxplots suggests that the median is under 50 bicycles and the densities suggest a large concentrations of zero, a zero inflated model could be a solution. Another solution is fitting a negative binomial model instead of a poisson.

```
fit_zero_pois<-brm(Bikes~HORA.INICIO+Tramo+(1|Tramo)+(HORA.INICIO+Tramo|PC),
             family = "zero_inflated_poisson",
             data =bbdd,
             iter = 4000,
             thin=5,
  backend = "cmdstanr")
```

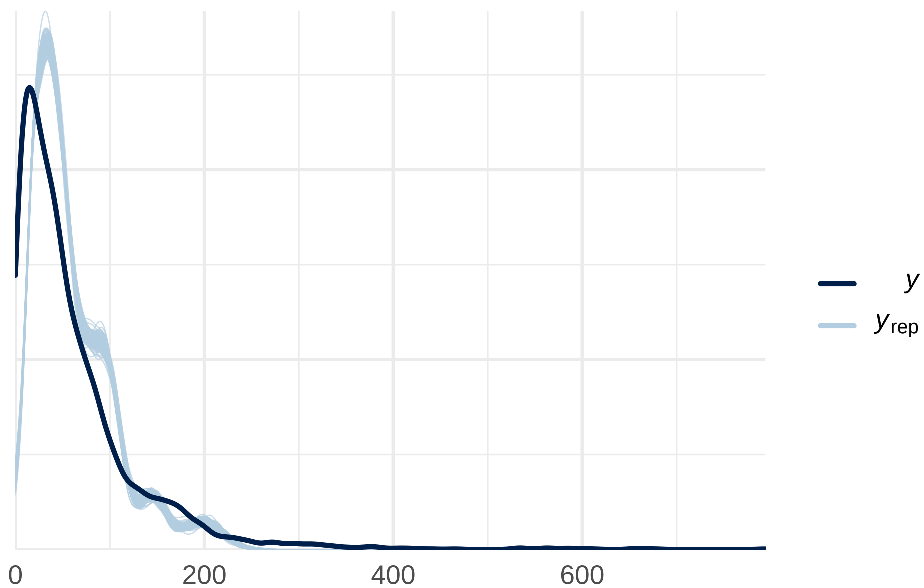
```
summary(fit_zero_pois)
```

```
Warning: Parts of the model have not converged (some Rhats are > 1.05). Be
careful when analysing the results! We recommend running more iterations and/or
setting stronger priors.

Warning: There were 84 divergent transitions after
warmup. Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

 Family: zero_inflated_poisson
```

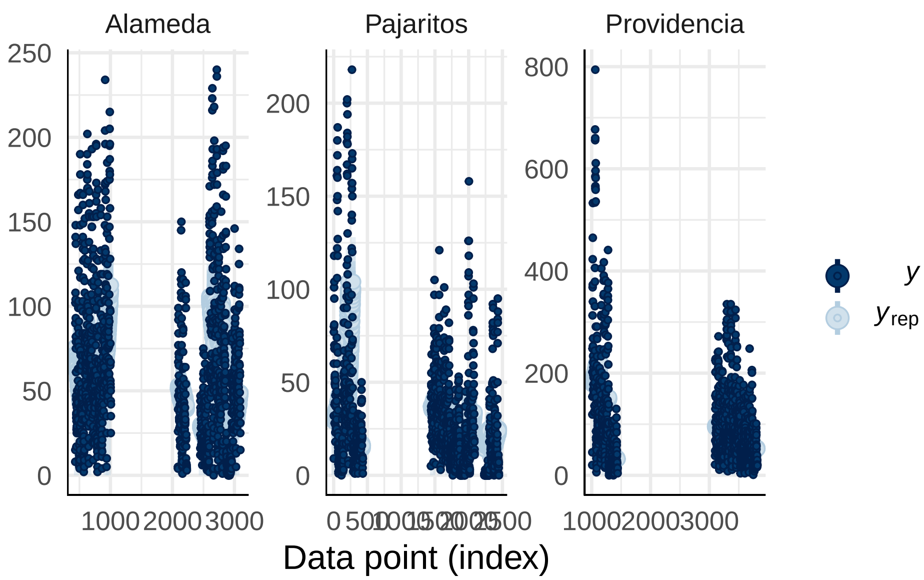
```
  Links: mu = log; zi = identity
Formula: Bikes ~ HORA.INICIO + Tramo + (1 | Tramo) + (HORA.INICIO + Tramo | PC)
   Data: bbdd (Number of observations: 3816)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 5;
         total post-warmup draws = 1600

Group-Level Effects:
~PC (Number of levels: 25)
                                     Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                            0.54      0.11     0.37     0.78 1.02
sd(HORA.INICIO)                          0.59      0.09     0.44     0.78 1.03
sd(TramoPajaritos)                       0.38      0.29     0.02     1.09 1.10
sd(TramoProvidencia)                     0.56      0.39     0.03     1.51 1.04
cor(Intercept,HORA.INICIO)              -0.34      0.23    -0.70     0.18 1.11
cor(Intercept,TramoPajaritos)           -0.16      0.44    -0.83     0.80 1.03
cor(HORA.INICIO,TramoPajaritos)          0.04      0.42    -0.77     0.76 1.03
cor(Intercept,TramoProvidencia)         -0.07      0.44    -0.80     0.78 1.01
cor(HORA.INICIO,TramoProvidencia)        0.10      0.44    -0.79     0.84 1.00
cor(TramoPajaritos,TramoProvidencia)    -0.02      0.44    -0.81     0.79 1.04
                                     Bulk_ESS Tail_ESS
sd(Intercept)                             146      297
sd(HORA.INICIO)                           131      335
sd(TramoPajaritos)                         43      130
sd(TramoProvidencia)                      106      197
cor(Intercept,HORA.INICIO)                 28       96
cor(Intercept,TramoPajaritos)             138      186
cor(HORA.INICIO,TramoPajaritos)           235      419
cor(Intercept,TramoProvidencia)           184      363
cor(HORA.INICIO,TramoProvidencia)         300      253
cor(TramoPajaritos,TramoProvidencia)      227      501

~Tramo (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.80      1.37     0.06     5.82 1.02      130       99

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            4.05      1.78     0.25     7.69 1.06       80       81
HORA.INICIO          0.11      0.12    -0.12     0.33 1.05       72      159
TramoPajaritos      -1.18      2.65    -7.51     4.12 1.05       98      123
TramoProvidencia     0.43      2.55    -4.77     5.96 1.05       89       59

Family Specific Parameters:
   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
zi     0.02      0.00     0.01     0.02 1.00      433      944
```

```
Draws were sampled using sample(hmc). 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).
```

```
#plot(fit_zero_pois)
pp_check(fit_zero_pois,ndraws = 100)
```




```
pp_check(fit_zero_pois,ndraws = 100,type = "intervals_grouped",group = "Tramo")
```



As with the previous poisson, there are a lot of issues in this model, which suggests that it is not a good fit.

0.7 Negative Binomial Model

```
fit_neg_bin<-brm(Bikes~HORA.INICIO+Tramo+(1|Tramo)+(HORA.INICIO+Tramo|PC),
                   family = "negbinomial2",
                   data =bbdd,iter = 4000,
                   thin=5,
                   backend = "cmdstanr")
```

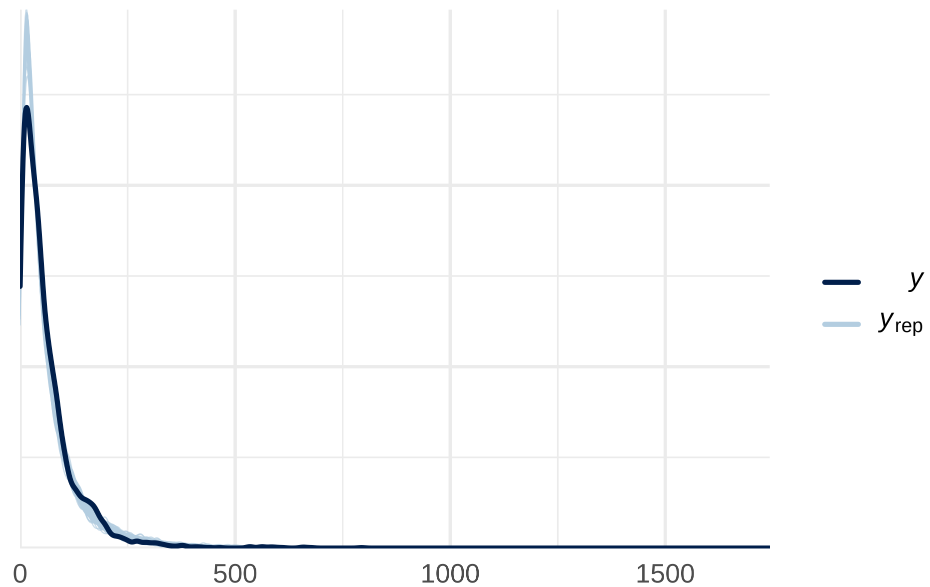
```
summary(fit_neg_bin)
```

```
Warning: There were 136 divergent transitions after
warmup. Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

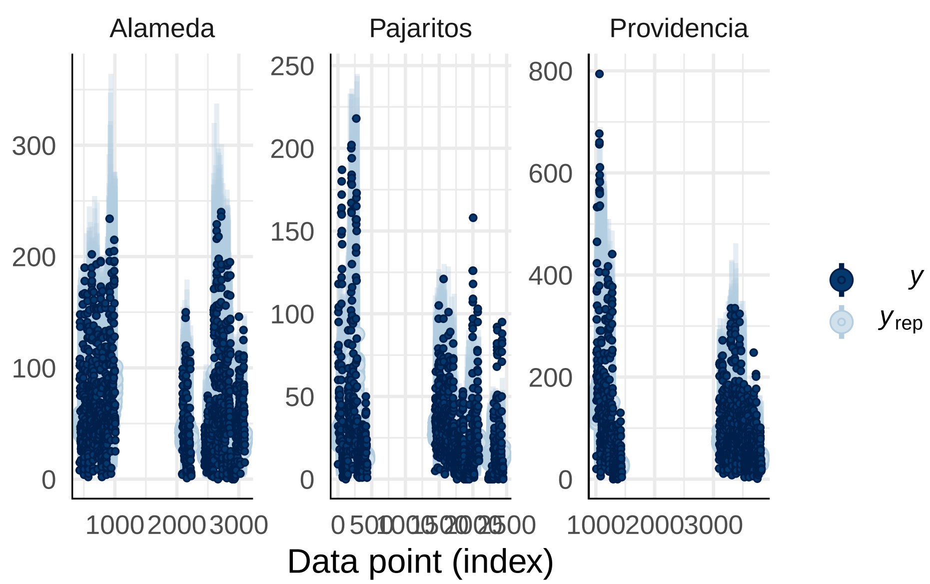
```
 Family: negbinomial2
  Links: mu = log; sigma = identity
Formula: Bikes ~ HORA.INICIO + Tramo + (1 | Tramo) + (HORA.INICIO + Tramo | PC)
   Data: bbdd (Number of observations: 3816)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 5;
         total post-warmup draws = 1600

Group-Level Effects:
```

```
~PC (Number of levels: 25)
                                    Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                           0.54      0.13     0.35     0.84 1.00
sd(HORA.INICIO)                         0.55      0.11     0.37     0.79 1.01
sd(TramoPajaritos)                      0.38      0.29     0.02     1.12 1.00
sd(TramoProvidencia)                    0.55      0.37     0.03     1.35 1.00
cor(Intercept,HORA.INICIO)             -0.36      0.24    -0.78     0.16 1.00
cor(Intercept,TramoPajaritos)          -0.15      0.43    -0.85     0.74 1.00
cor(HORA.INICIO,TramoPajaritos)         0.07      0.41    -0.74     0.81 1.00
cor(Intercept,TramoProvidencia)        -0.08      0.43    -0.82     0.77 1.00
cor(HORA.INICIO,TramoProvidencia)       0.09      0.44    -0.76     0.86 1.00
cor(TramoPajaritos,TramoProvidencia)   -0.00      0.46    -0.80     0.82 1.00
                                    Bulk_ESS Tail_ESS
sd(Intercept)                           1094     1554
sd(HORA.INICIO)                          621     1154
sd(TramoPajaritos)                      1008     1458
sd(TramoProvidencia)                     938     1397
cor(Intercept,HORA.INICIO)              1220     1426
cor(Intercept,TramoPajaritos)           1443     1426
cor(HORA.INICIO,TramoPajaritos)         1492     1615
cor(Intercept,TramoProvidencia)         1385     1567
cor(HORA.INICIO,TramoProvidencia)       1499     1349
cor(TramoPajaritos,TramoProvidencia)     544      294

~Tramo (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     1.64      1.25     0.08     4.49 1.02      382      143

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            3.97      1.64     0.26     7.53 1.01     1108     1105
HORA.INICIO          0.12      0.12    -0.12     0.38 1.00     1513     1493
TramoPajaritos      -0.85      2.37    -6.14     4.43 1.01      911      886
TramoProvidencia     0.56      2.63    -4.66     7.78 1.01      328       61

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.61      0.01     0.58     0.64 1.00     1865     1615

Draws were sampled using sample(hmc). 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
#plot(fit_neg_bin)
pp_check(fit_neg_bin,ndraws = 100)
```




```
pp_check(fit_neg_bin,ndraws = 100,
         type = "intervals_grouped",group = "Tramo")
```



Although in this case there is a 9% divergence the maximum treedepth is the warning that more suspicions gives me. The Rhats in the model suggests a good fit.

1 Cross validation

For comparison, we use a leave-one-out comparison, to check which of the models can claim the best fit.

```
loo::loo_compare(loo::loo(fit),
                 loo::loo(fit_log),
                 loo::loo(fit_log_priors),
                 loo::loo(fit_pois),
                 loo::loo(fit_zero_pois),
                 loo::loo(fit_neg_bin))
```

```
Warning: Found 1 observations with a pareto_k > 0.7 in model 'fit'. It is
recommended to set 'moment_match = TRUE' in order to perform moment matching for
problematic observations.
```

```
Warning: Found 59 observations with a pareto_k > 0.7 in model 'fit_pois'. It is
recommended to set 'moment_match = TRUE' in order to perform moment matching for
problematic observations.
```

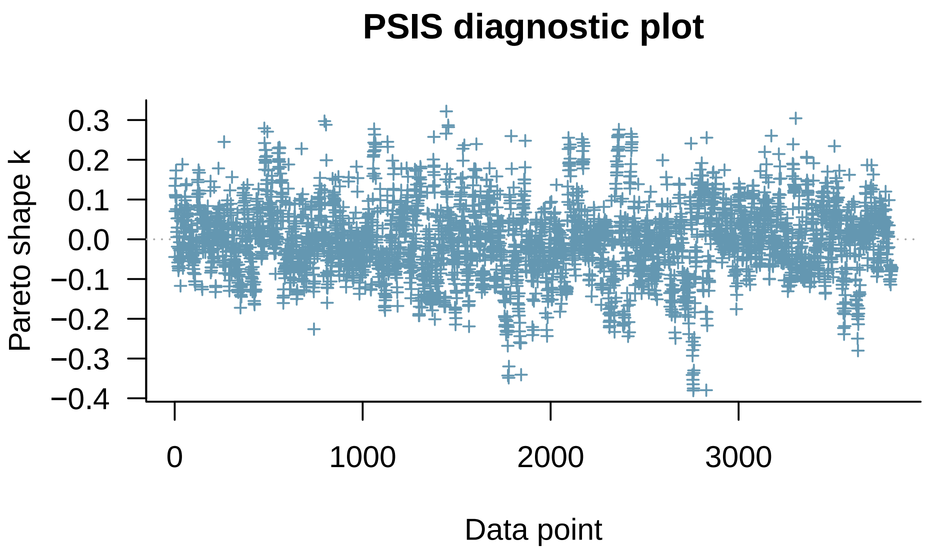
```
Warning: Found 57 observations with a pareto_k > 0.7 in model 'fit_zero_pois'.
It is recommended to set 'moment_match = TRUE' in order to perform moment
matching for problematic observations.
```

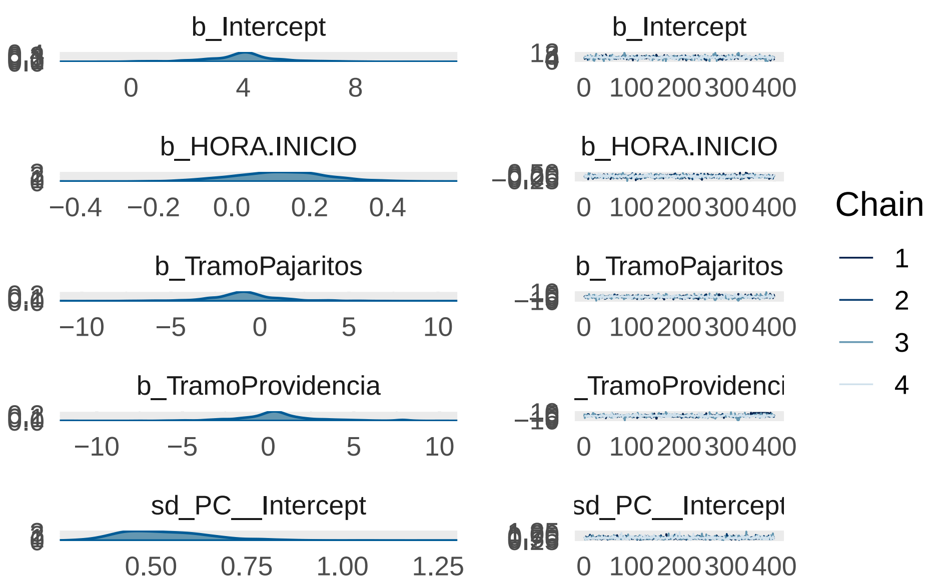
```
Warning: Not all models have the same y variable. ('yhash' attributes do not
match)
```

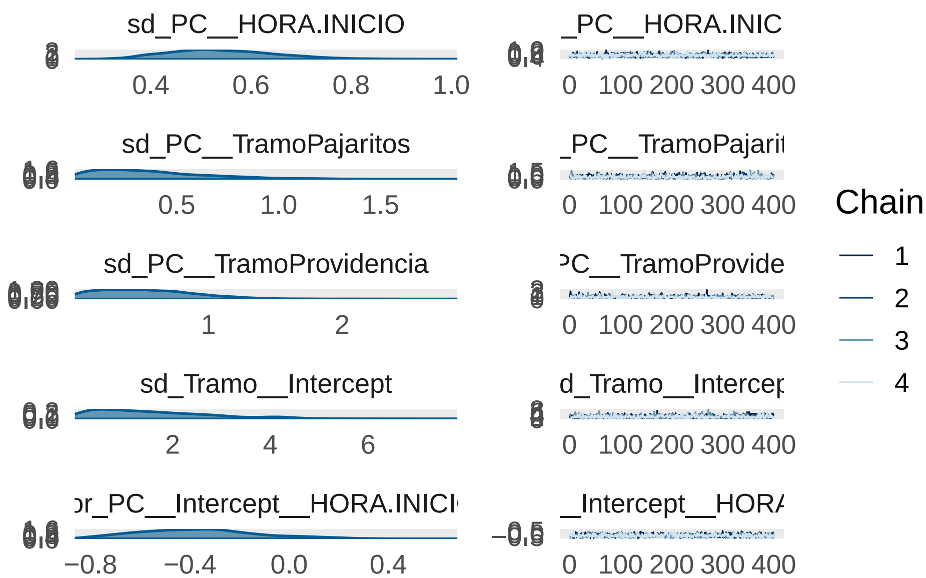
```
               elpd_diff se_diff
fit_neg_bin         0.0       0.0
fit_log_priors   -106.8      21.2
fit_log          -107.4      21.2
fit             -2150.6     112.7
fit_zero_pois  -48878.3    1572.4
fit_pois       -50174.1    1589.2
```

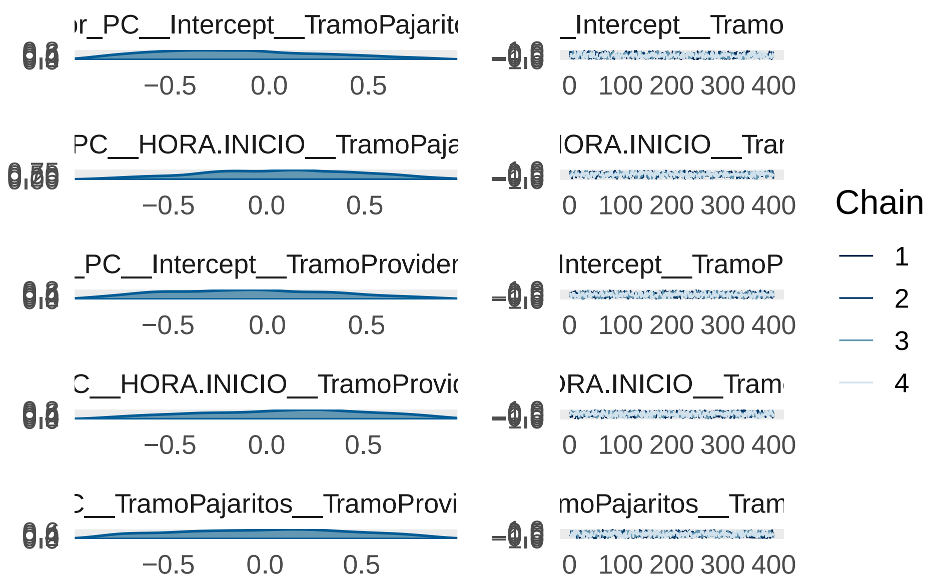
Even though several models have their issues with the pareto, the best one is the only one that presents no such message. Hence the best model is the last one, negative binomial.

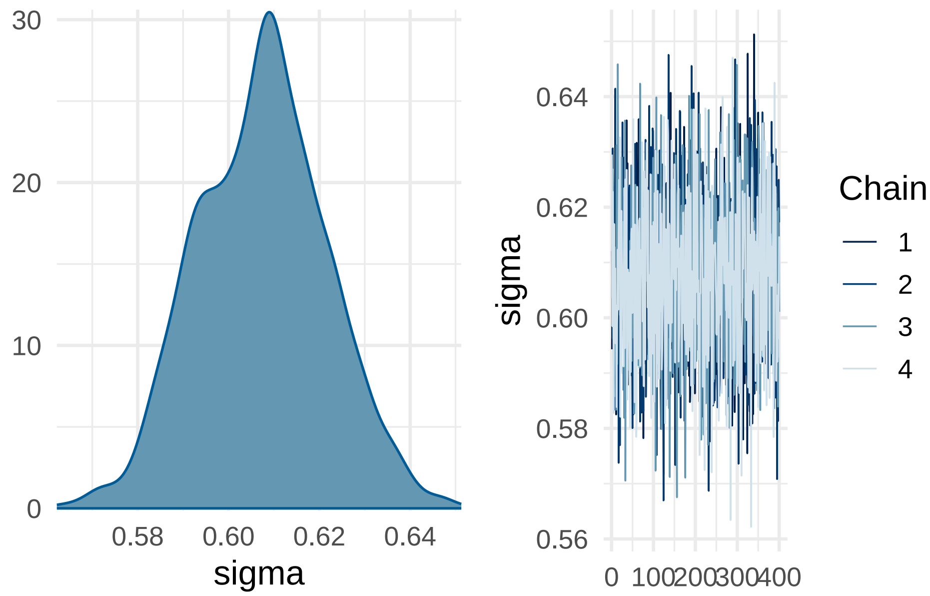
```
plot(loo::loo(fit_neg_bin))
plot(fit_neg_bin)
```


PSIS diagnostic plot
Pareto shape k
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
0
1000
2000
3000
Data point


b_Intercept
0
4
8
b_Intercept
0 100 200 300 400
b_HORA.INICIO
−0.4 −0.2 0.0 0.2 0.4
b_HORA.INICIO
0 100 200 300 400
b_TramoPajaritos
−10 −5 0 5 10
b_TramoPajaritos
0 100 200 300 400
b_TramoProvidencia
−10 −5 0 5 10
_TramoProvidenci
0 100 200 300 400
sd_PC__Intercept
0.50 0.75 1.00 1.25
sd_PC__Intercept
0 100 200 300 400
Chain
1
2
3
4


sd_PC__HORA.INICIO
_PC__HORA.INIC
sd_PC__TramoPajaritos
_PC__TramoPajarit
sd_PC__TramoProvidencia
PC__TramoProvide
sd_Tramo__Intercept
d_Tramo__Intercep
or_PC__Intercept__HORA.INICI
_Intercept__HOR/
0 100 200 300 400
Chain
1
2
3
4


or_PC__Intercept__TramoPajarit
_Intercept__Tramo
PC__HORA.INICIO__TramoPaja
IORA.INICIO__Trar
_PC__Intercept__TramoProvider
Intercept__TramoP
C__HORA.INICIO__TramoProvi
RA.INICIO__Tram
C__TramoPajaritos__TramoProvi
moPajaritos__Tram
0 100 200 300 400
Chain
1
2
3
4



2 Hypothesis

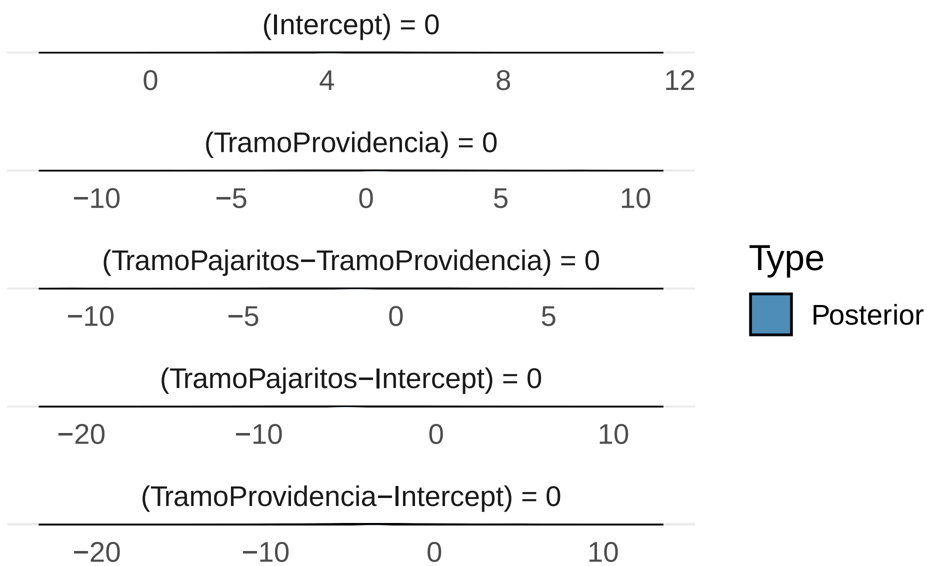
As for the main objective, now having a model we can test the hypothesis over the intercept (given by alphabetical order to the section Alameda).

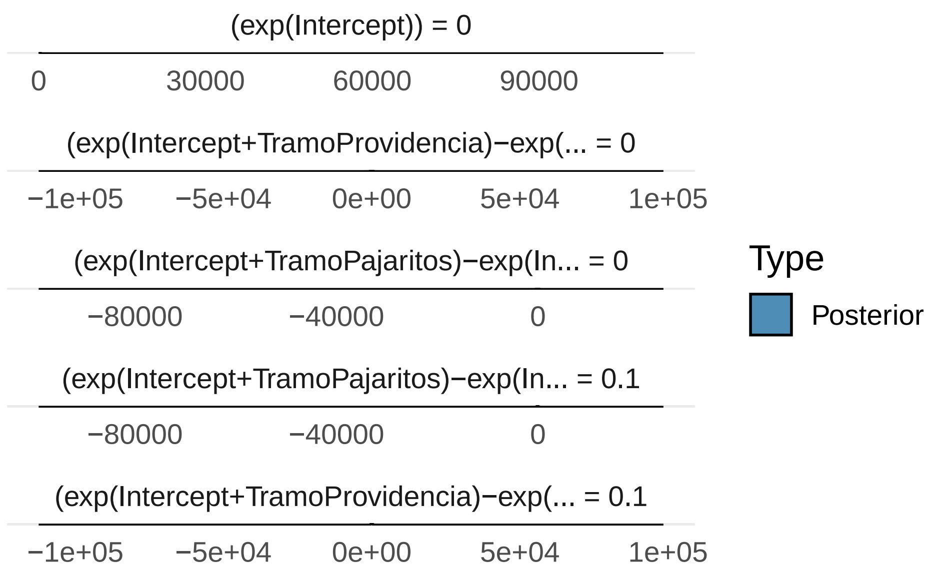
```
## Hypothesis testing
hypNegBin = hypothesis(
  fit_neg_bin,
  hypothesis = c(
    # Look only at Diet1 which is coded as Intercept in our dummy coding
    "Intercept = 0",
    "TramoProvidencia = 0",
    "TramoPajaritos - TramoProvidencia = 0",
    "TramoPajaritos - Intercept = 0",
    "TramoProvidencia - Intercept = 0",
    # as the negative model is log(mu), then we will look at the exp values:
    "exp(Intercept) = 0",
    "exp(Intercept+TramoProvidencia) -exp(Intercept)= 0",
    "exp(Intercept+TramoPajaritos) - exp(Intercept+TramoProvidencia) = 0",
    "exp(Intercept+TramoPajaritos) - exp(Intercept) = 0",
    "exp(Intercept+TramoProvidencia) - exp(Intercept) = 0"
  ))
hypNegBin
```

```
Hypothesis Tests for class b:
```

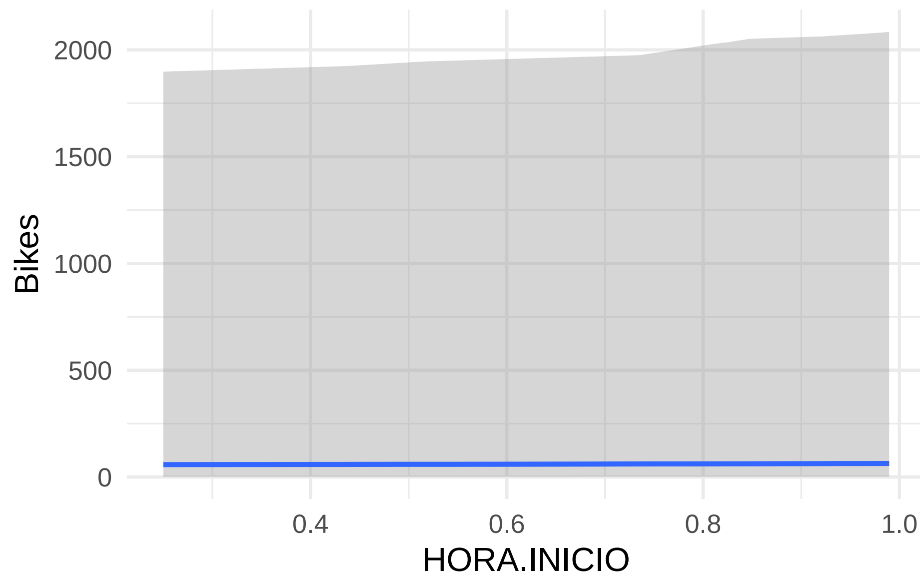
```
                Hypothesis Estimate Est.Error  CI.Lower CI.Upper Evid.Ratio
1           (Intercept) = 0     3.97      1.64      0.26     7.53         NA
2    (TramoProvidencia) = 0     0.56      2.63     -4.66     7.78         NA
3  (TramoPajaritos-T... = 0    -1.42      2.66     -9.69     3.72         NA
4  (TramoPajaritos-I... = 0    -4.83      3.73    -12.93     3.62         NA
5  (TramoProvidencia... = 0    -3.41      3.98    -11.27     5.11         NA
6      (exp(Intercept)) = 0   521.42   5268.98      1.30  1863.40         NA
7  (exp(Intercept+Tr... = 0  1056.39   9606.82  -1578.49 36301.53         NA
8  (exp(Intercept+Tr... = 0 -1405.44   8066.90 -38624.45   683.38         NA
9  (exp(Intercept+Tr... = 0  -349.05   5394.68  -1582.93   718.52         NA
10 (exp(Intercept+Tr... = 0  1056.39   9606.82  -1578.49 36301.53         NA
   Post.Prob Star
1         NA    *
2         NA
3         NA
4         NA
5         NA
6         NA    *
7         NA
8         NA
9         NA
10        NA
---
'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
'*': For one-sided hypotheses, the posterior probability exceeds 95%;
for two-sided hypotheses, the value tested against lies outside the 95%-CI.
Posterior probabilities of point hypotheses assume equal prior probabilities.
```

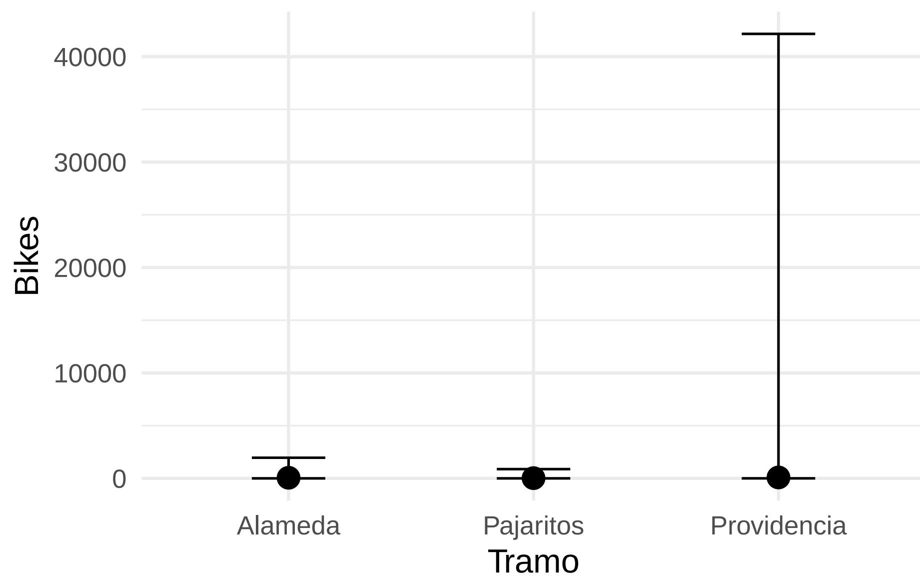
```
plot(hypNegBin)
```






```
conditional_effects(fit_neg_bin)
```





As we can see, there is no difference between sections of the city, even though we suspected that there was.

2.1 Conclusion

We fit several models to realworld data, expecting to have conclusions over whether there is a difference between the use in hours over different sections of the city. As a result we obtained that even though there is a larger dispersion on one of these, there is no real difference under the best model fitted. This can be improved if we could get other variables to incorporate to the model, such as vehicules and other means of transportations, weather conditions and some more envirormental variables that could affect the usage of bikes.