NYC models

```
library(DT)
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.5.2
```

```
library(magrittr)
library(sf)
```

```
## Warning: package 'sf' was built under R version 3.5.2
```

```
## Linking to GEOS 3.7.2, GDAL 2.4.2, PROJ 5.2.0
```

```
library(pscl)
```

```
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
```

```
library(countreg)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'countreg'
```

```
## The following objects are masked from 'package:pscl':
##
##     hurdle, hurdle.control, hurdletest, zeroinfl, zeroinfl.control
```

```
library(GISTools)
```

```
## Loading required package: maptools
```

```
## Warning: package 'maptools' was built under R version 3.5.2
```

```
## Loading required package: sp
```

```
## Checking rgeos availability: TRUE
```

```
## Loading required package: RColorBrewer
```

```
## Loading required package: rgeos
```

```
## Warning: package 'rgeos' was built under R version 3.5.2
```

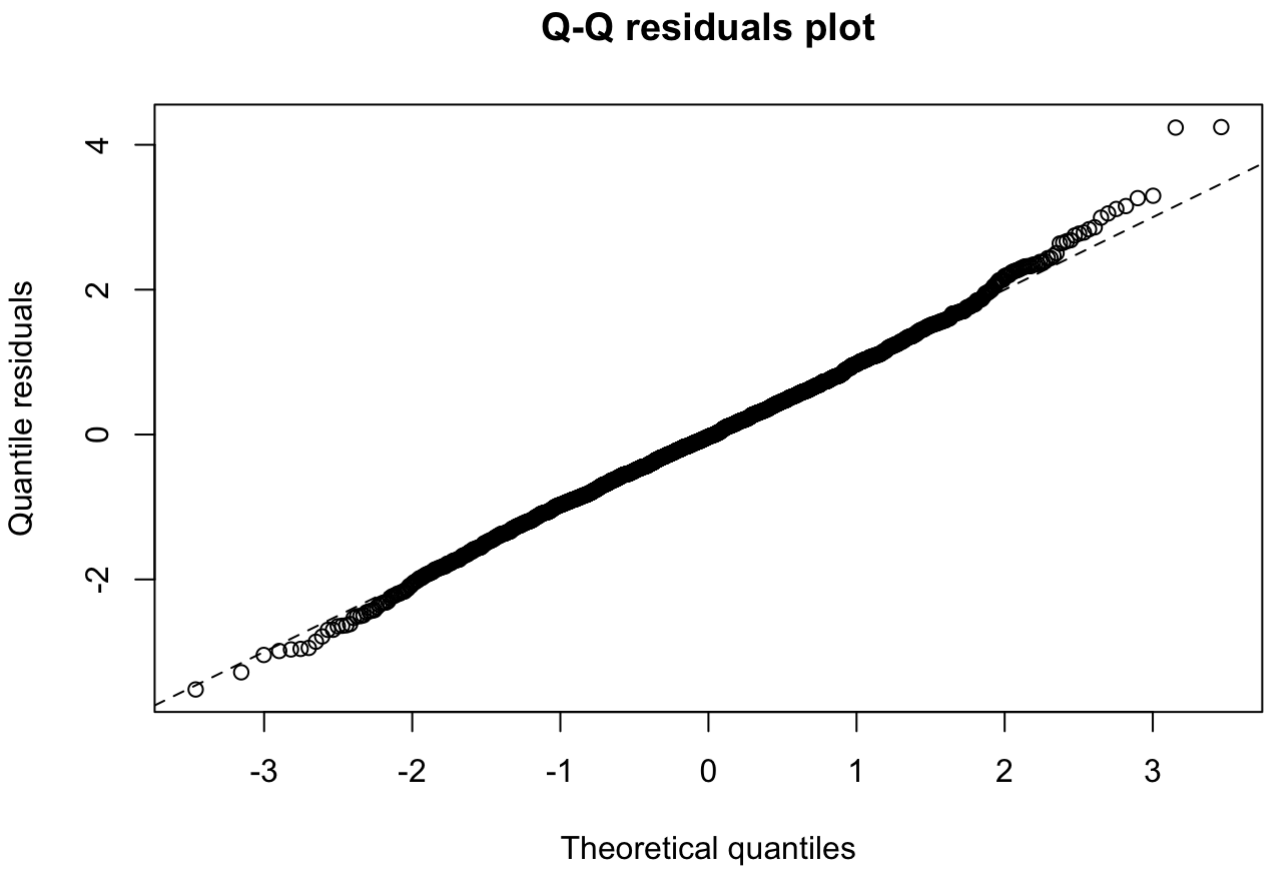
```
## rgeos version: 0.5-2, (SVN revision 621)
##  GEOS runtime version: 3.7.2-CAPI-1.11.2
##  Linking to sp version: 1.3-1
##  Polygon checking: TRUE
```

```
library(gtfsr)
library(sp)
library(stringi)
```

```
chicago_final<-fread("/Users/11kolop/Desktop/nyc_final.csv")
chicago_final<-chicago_final[,-c(1:2)]
dat.hom.chicago<-chicago_final[chicago_final$ofns_desc=="criminal homicide",]
dat.hom.chicago$transp<-as.numeric(as.character(dat.hom.chicago$transp))
pca <- princomp(na.omit(dat.hom.chicago)[,c(4:15)], cor = TRUE)

mod.zero.inflated.poisson<-zeroinfl(n ~ foreign_share2010 + share_black2010 + share_hisp
2010 + singleparent_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit
(dat.hom.chicago),dist="poisson")

mod.poisson<-glm(n ~ foreign_share2010 + share_black2010 + share_hisp2010 + singleparent
_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit(dat.hom.chicago),fa
mily="poisson")

mod.pca.zero.inflated.poisson<-zeroinfl(na.omit(dat.hom.chicago)$n~scale(na.omit(dat.ho
m.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)+pca$scores[,1] + pca$scores[,2]+pca$sco
res[,3] + pca$scores[,4],dist="poisson")

mod.pca.poisson<-glm(na.omit(dat.hom.chicago)$n~scale(na.omit(dat.hom.chicago)$transp/n
a.omit(dat.hom.chicago)$sqmi)+pca$scores[,1] + pca$scores[,2]+pca$scores[,3] + pca$score
s[,4],family="poisson")

mod.zero.inflated.nb<-zeroinfl(n ~ foreign_share2010 + share_black2010 + share_hisp2010
 + singleparent_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit(dat.
hom.chicago),dist="negbin")

mod.pca.zero.inflated.nb<-zeroinfl(na.omit(dat.hom.chicago)$n~scale(na.omit(dat.hom.chic
ago)$transp/na.omit(dat.hom.chicago)$sqmi)+pca$scores[,1] + pca$scores[,2]+pca$scores[,3
] + pca$scores[,4],dist="negbin")

BIC(mod.poisson,mod.zero.inflated.poisson,mod.pca.poisson,mod.pca.zero.inflated.poisson,
mod.zero.inflated.nb,mod.pca.zero.inflated.nb)
```

```
##                                  df      BIC
## mod.poisson                       7 5791.418
## mod.zero.inflated.poisson        14 5683.125
## mod.pca.poisson                   6 5682.764
## mod.pca.zero.inflated.poisson    12 5598.517
## mod.zero.inflated.nb             15 5488.244
## mod.pca.zero.inflated.nb         13 5446.777
```

```
qqrplot(mod.pca.zero.inflated.nb)
```

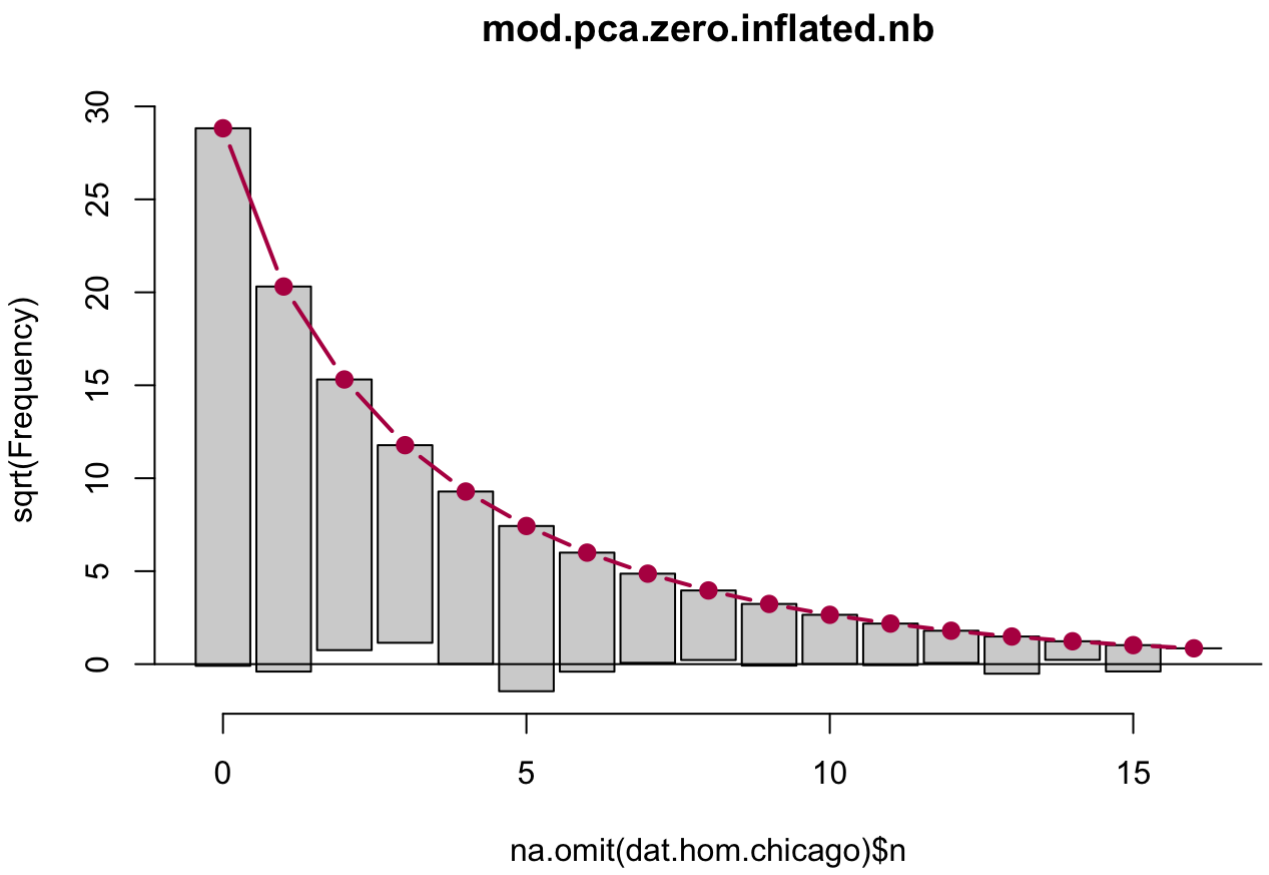



```
summary(mod.pca.zero.inflated.nb)
```

```
##
## Call:
## zeroinfl(formula = na.omit(dat.hom.chicago)$n ~ scale(na.omit(dat.hom.chicago)$trans
p/na.omit(dat.hom.chicago)$sqmi) +
##     pca$scores[, 1] + pca$scores[, 2] + pca$scores[, 3] + pca$scores[,
##     4], dist = "negbin")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.4134 -0.6588 -0.4250  0.4495 11.6150
##
## Count model coefficients (negbin with log link):
##                                                                          Estimate
## (Intercept)                                                               0.11268
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi) -0.12925
## pca$scores[, 1]                                                          -0.41596
## pca$scores[, 2]                                                           0.16675
## pca$scores[, 3]                                                           0.04247
## pca$scores[, 4]                                                           0.08168
## Log(theta)                                                                1.04538
##                                                                          Std. Error
## (Intercept)                                                                 0.05153
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)    0.03163
## pca$scores[, 1]                                                             0.02387
## pca$scores[, 2]                                                             0.02308
## pca$scores[, 3]                                                             0.02083
## pca$scores[, 4]                                                             0.03843
## Log(theta)                                                                  0.12151
##                                                                          z value
## (Intercept)                                                                2.187
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)  -4.086
## pca$scores[, 1]                                                          -17.427
## pca$scores[, 2]                                                            7.224
## pca$scores[, 3]                                                            2.039
## pca$scores[, 4]                                                            2.125
## Log(theta)                                                                 8.603
##                                                                          Pr(>|z|)
## (Intercept)                                                                0.0288
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi) 4.39e-05
## pca$scores[, 1]                                                           < 2e-16
## pca$scores[, 2]                                                          5.04e-13
## pca$scores[, 3]                                                            0.0415
## pca$scores[, 4]                                                            0.0336
## Log(theta)                                                                < 2e-16
##
## (Intercept)                                                              *
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi) ***
## pca$scores[, 1]                                                          ***
## pca$scores[, 2]                                                          ***
## pca$scores[, 3]                                                          *
## pca$scores[, 4]                                                          *
## Log(theta)                                                               ***
##
```

```
## Zero-inflation model coefficients (binomial with logit link):
##                                                                          Estimate
## (Intercept)                                                               -3.5242
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)      -1.2248
## pca$scores[, 1]                                                            0.8777
## pca$scores[, 2]                                                           -0.3319
## pca$scores[, 3]                                                            0.3272
## pca$scores[, 4]                                                            0.4728
##                                                                          Std. Error
## (Intercept)                                                                 1.0647
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)        0.6833
## pca$scores[, 1]                                                             0.2383
## pca$scores[, 2]                                                             0.2743
## pca$scores[, 3]                                                             0.2997
## pca$scores[, 4]                                                             0.3951
##                                                                          z value
## (Intercept)                                                               -3.310
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)      -1.793
## pca$scores[, 1]                                                            3.683
## pca$scores[, 2]                                                           -1.210
## pca$scores[, 3]                                                            1.092
## pca$scores[, 4]                                                            1.197
##                                                                          Pr(>|z|)
## (Intercept)                                                              0.000933
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)     0.073030
## pca$scores[, 1]                                                          0.000231
## pca$scores[, 2]                                                          0.226263
## pca$scores[, 3]                                                          0.274906
## pca$scores[, 4]                                                          0.231438
##
## (Intercept)                                                              ***
## scale(na.omit(dat.hom.chicago)$transp/na.omit(dat.hom.chicago)$sqmi)     .
## pca$scores[, 1]                                                          ***
## pca$scores[, 2]
## pca$scores[, 3]
## pca$scores[, 4]
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 2.8445
## Number of iterations in BFGS optimization: 42
## Log-likelihood: -2674 on 13 Df
```

```
rootogram(mod.pca.zero.inflated.nb)
```


mod.pca.zero.inflated.nb
sqrt(Frequency)
0
5
10
15
20
25
30
0
5
10
15
na.omit(dat.hom.chicago)$n