STAT 8020 R Lab 16: Poisson Regression

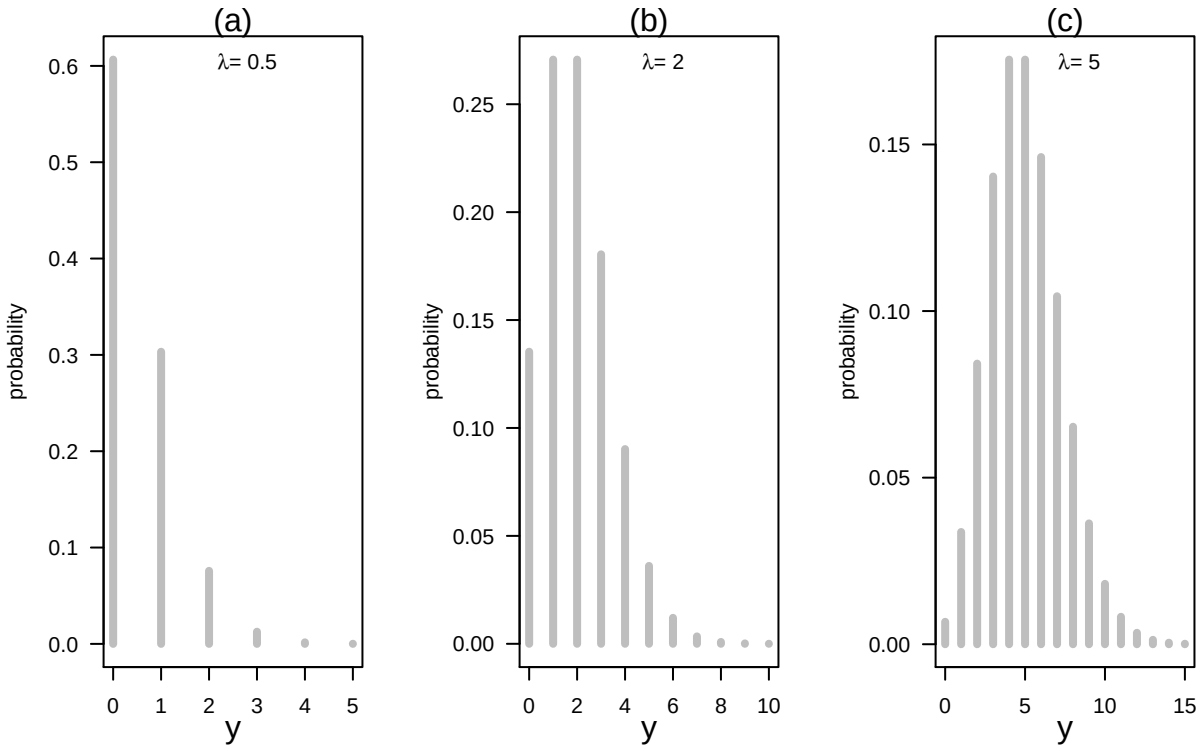
Whitney

October 27, 2020

Contents

Poisson Distribution

```
x1 <- 0:5; x2 <- 0:10; x3 <- 0:15
par(mfrow = c(1, 3))
plot(x1, dpois(x1, 0.5), type = "h", lwd = 4, col = "gray", las = 1,
     xlab = "", ylab = "probability")
mtext("y", side = 1, line = 2)
mtext("(a)")
legend("top", legend = expression(paste(lambda,  "= 0.5")), bty = "n")
plot(x2, dpois(x2, 2), type = "h", lwd = 4, col = "gray", las = 1,
     xlab = "", ylab = "probability")
mtext("y", side = 1, line = 2)
mtext("(b)")
legend("top", legend = expression(paste(lambda,  "= 2")), bty = "n")
plot(x3, dpois(x3, 5), type = "h", lwd = 4, col = "gray", las = 1,
     xlab = "", ylab = "probability")
mtext("y", side = 1, line = 2)
mtext("(c)")
legend("top", legend = expression(paste(lambda,  "= 5")), bty = "n")
```



Flying-Bomb Hits on London During World War II [Clarke, 1946; Feller, 1950]

```
count <- c(229, 211, 93, 35, 7, 1)
grids <- 576
hits <- 537
lambda <- hits / grids
count_expected <- c(grids * dpois(0:4, lambda = lambda), grids * ppois(4, lambda = lambda, lower.tail =
round(count_expected, 1)
```

```
## [1] 226.7 211.4  98.5  30.6   7.1   1.6
```

US Landfalling Hurriances

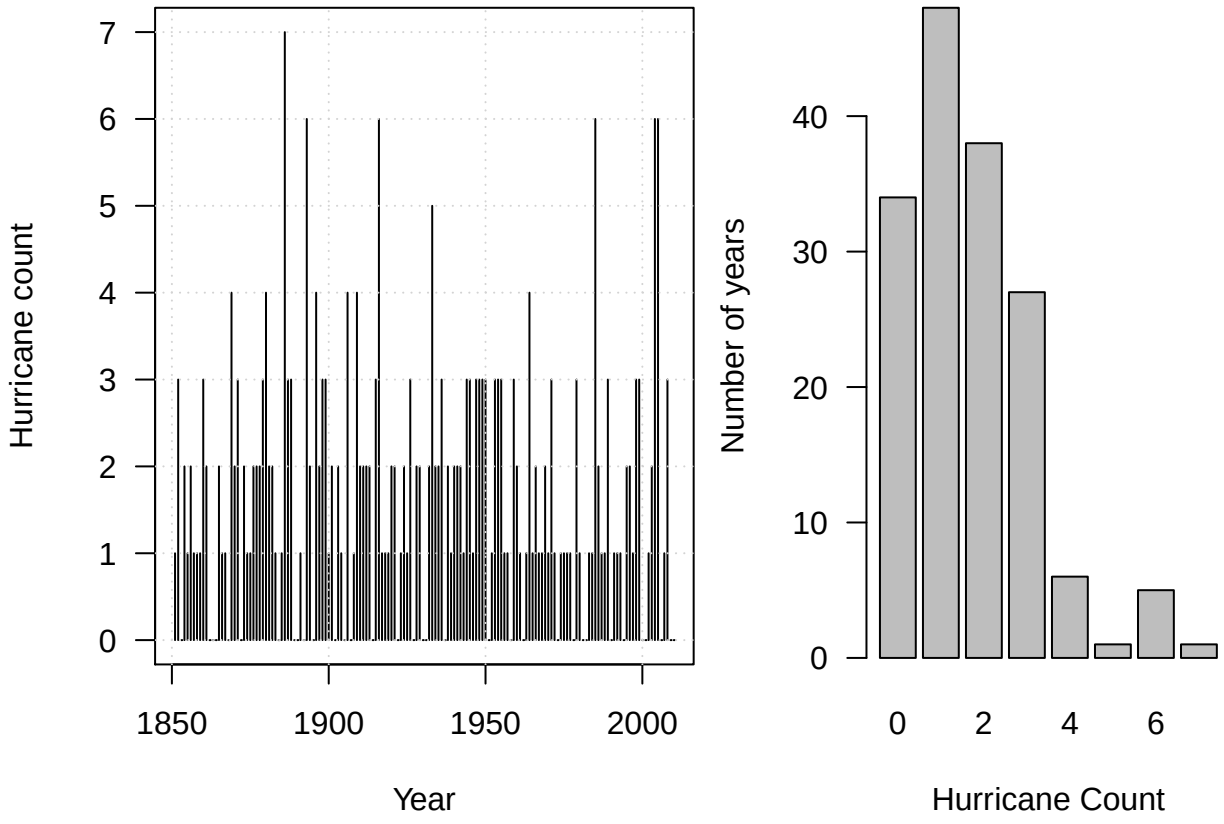
Load the hurriance count

```
con = "http://myweb.fsu.edu/jelsner/Book/Chap07/US.txt"
hurricanes = read.table(con, header = T)
head(hurricanes)
```

```
##   Year All MUS G FL E
## 1 1851   1   1 0  1 0
## 2 1852   3   1 1  2 0
## 3 1853   0   0 0  0 0
## 4 1854   2   1 1  0 1
## 5 1855   1   1 1  0 0
## 6 1856   2   1 1  1 0
```

```
par(las = 1, mar = c(4.6, 3.9, 0.8, 0.6))
layout(matrix(c(1, 2), 1, 2, byrow = TRUE), widths = c(0.57, 0.43))
plot(hurricanes$Year, hurricanes$All, type = "h", xlab = "Year", ylab = "Hurricane count")
```

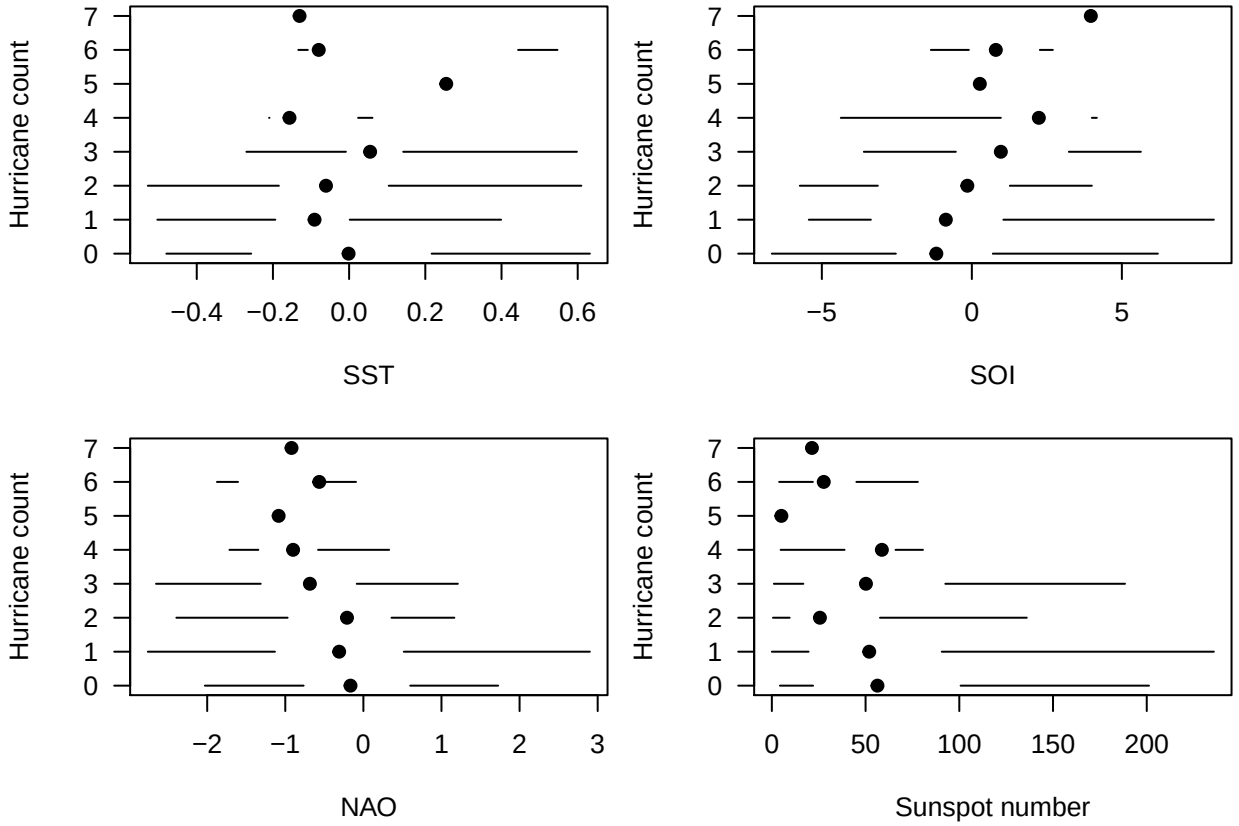
```
grid()
barplot(table(hurricanes$All), xlab = "Hurricane Count", ylab = "Number of years", main = "")
```



Load the environmetal variables

```
load("annual.RData")
data <- data.frame(All = hurricanes$All, SOI = annual$soi, NAO = annual$nao,
                   SST = annual$sst, SSN = annual$ssn)
data <- data[-(1:15),]

H <- hurricanes

par(mfrow = c(2, 2), mar = c(4.5, 4, 1, 0.6))
plot(range(annual$sst, na.rm = TRUE), c(0, 7), type = "n", ylab = "Hurricane count", xlab = "SST",
     las = 1)
for(i in 0:7){
  points(fivenum(annual$sst[H$All == i])[3], i, pch = 19)
  lines(c(fivenum(annual$sst[H$All == i])[1], fivenum(annual$sst[H$All == i])[2]), c(i, i))
  lines(c(fivenum(annual$sst[H$All == i])[4], fivenum(annual$sst[H$All == i])[5]), c(i, i))
}
plot(range(annual$soi, na.rm = TRUE), c(0, 7), type = "n", ylab = "Hurricane count", xlab = "SOI",
     las = 1)

for(i in 0:7){
  points(fivenum(annual$soi[H$All == i])[3], i, pch=19)
  lines(c(fivenum(annual$soi[H$All == i])[1], fivenum(annual$soi[H$All == i])[2]), c(i, i))
  lines(c(fivenum(annual$soi[H$All == i])[4], fivenum(annual$soi[H$All == i])[5]), c(i, i))
```

```
}
plot(range(annual$nao, na.rm = TRUE), c(0, 7), type = "n", ylab = "Hurricane count", xlab = "NAO",
     las = 1)

for(i in 0:7){
  points(fivenum(annual$nao[H$All == i])[3], i, pch=19)
  lines(c(fivenum(annual$nao[H$All == i])[1], fivenum(annual$nao[H$All == i])[2]), c(i, i))
  lines(c(fivenum(annual$nao[H$All == i])[4], fivenum(annual$nao[H$All == i])[5]), c(i, i))
}
plot(range(annual$ssn, na.rm = TRUE), c(0, 7), type = "n", ylab = "Hurricane count",
     xlab = "Sunspot number", las = 1)

for(i in 0:7){
  points(fivenum(annual$ssn[H$All == i])[3], i, pch = 19)
  lines(c(fivenum(annual$ssn[H$All == i])[1], fivenum(annual$ssn[H$All == i])[2]), c(i, i))
  lines(c(fivenum(annual$ssn[H$All == i])[4], fivenum(annual$ssn[H$All == i])[5]), c(i, i))
}
```



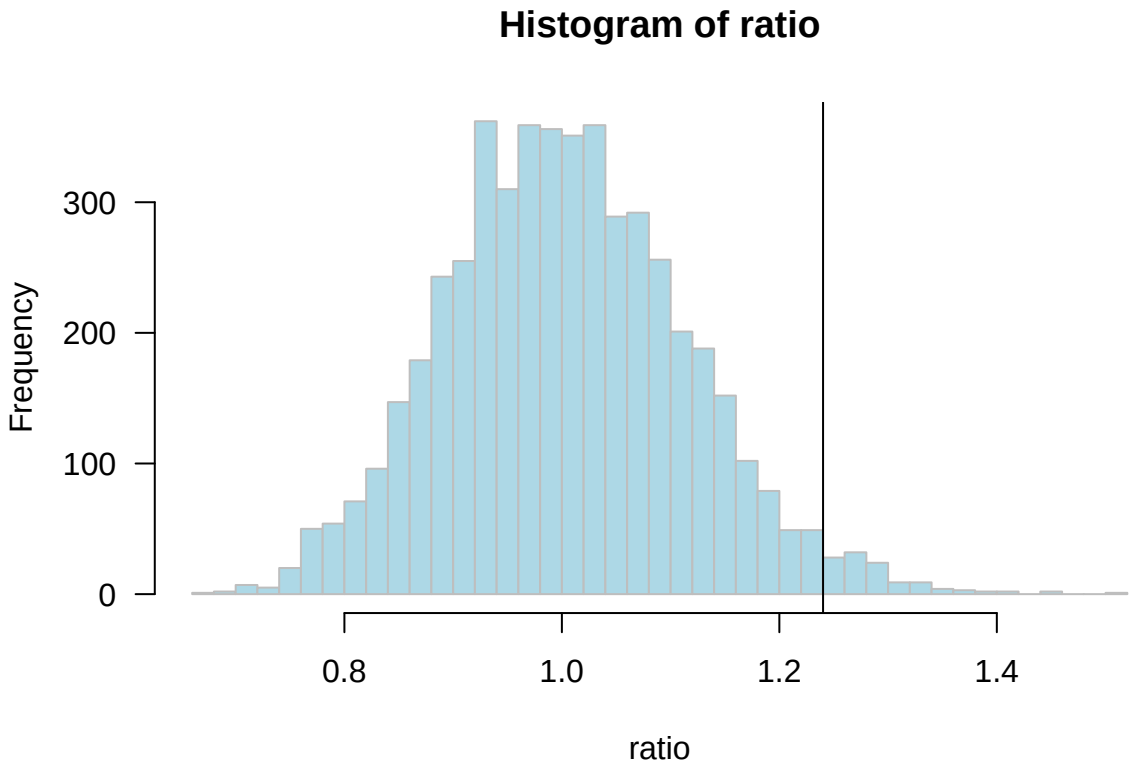
Monte Carlo Simulation

```
rate = mean(H$All)
var(H$All)/rate
```

```
## [1] 1.240201
```

```
n <- length(H$All)

set.seed(1234)
```

```
ratio = numeric()
m = 5000
for (i in 1:m) {
    h = rpois(n = n, lambda = rate)
    ratio[i] = var(h) / mean(h)
}

hist(ratio, 50, las = 1, col = "lightblue", border = "gray")
abline(v = var(H$All)/rate)
```




```
sum(ratio > var(H$All)/rate) / m
```

```
## [1] 0.0232
```

Linear Regression

```
lmFull <- lm(All ~ ., data = data)
predict(lmFull, newdata = data.frame(SOI = -3, NAO = 3, SST = 0, SSN = 250))
```

```
##         1
## -0.318065
```

```
step(lmFull)
```

```
## Start:  AIC=102.86
## All ~ SOI + NAO + SST + SSN
##
##        Df Sum of Sq    RSS    AIC
## - SST   1    1.5048 276.60 101.65
## <none>              275.10 102.86
## - SSN   1    5.3888 280.48 103.67
```

```
## - NAO   1   12.2586 287.35 107.18
## - SOI   1   15.7312 290.83 108.92
##
## Step:  AIC=101.65
## All ~ SOI + NAO + SSN
##
##        Df Sum of Sq    RSS    AIC
## <none>              276.60 101.65
## - SSN   1    4.6879 281.29 102.08
## - NAO   1   14.5246 291.12 107.07
## - SOI   1   15.6850 292.29 107.64
```

```
##
## Call:
## lm(formula = All ~ SOI + NAO + SSN, data = data)
##
## Coefficients:
## (Intercept)          SOI          NAO          SSN
##    1.859176     0.113710    -0.312752    -0.003634
```

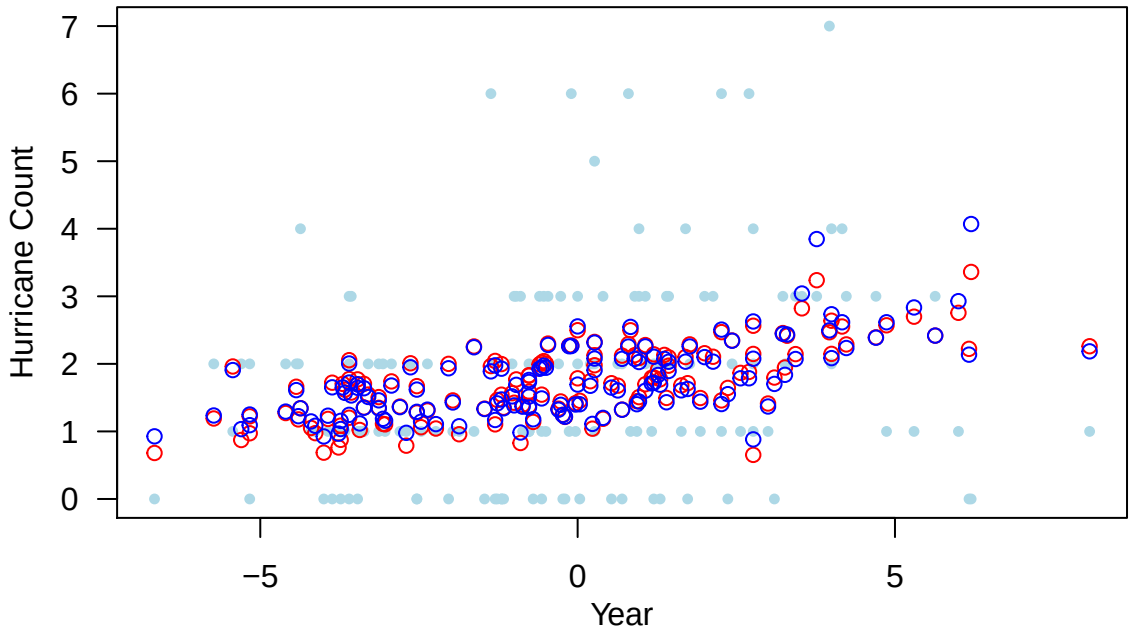
Poisson Regression

```
PoiFull <- glm(All ~ ., data = data, family = "poisson")
summary(PoiFull)
```

```
##
## Call:
## glm(formula = All ~ ., family = "poisson", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8530  -0.8984  -0.1376   0.6027   2.4720
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.595288   0.103342   5.760 8.39e-09 ***
## SOI          0.061863   0.021319   2.902  0.00371 **
## NAO         -0.166595   0.064427  -2.586  0.00972 **
## SST          0.228972   0.255289   0.897  0.36977
## SSN         -0.002306   0.001372  -1.681  0.09284 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 197.89  on 144  degrees of freedom
## Residual deviance: 174.81  on 140  degrees of freedom
## AIC: 479.64
##
## Number of Fisher Scoring iterations: 5
```

```
plot(data$SOI, hurricanes$All[-(1:15)], pch = 16, cex = 0.75, col = "lightblue",
     xlab = "", ylab = "", las = 1)
mtext("Hurricane Count", side = 2, line = 2)
mtext("Year", side = 1, line = 2)
```

```r
points(data$SOI, predict(lmFull), col = "red")
points(data$SOI, predict(PoiFull, type = "response"), col = "blue")
```



Another Example

```r
library(faraway)
data(gala)
gala <- gala[, -2]
PoiFit <- glm(Species ~ ., family = poisson, gala)
summary(PoiFit)
```

```
##
## Call:
## glm(formula = Species ~ ., family = poisson, data = gala)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8.2752  -4.4966  -0.9443   1.9168  10.1849
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.155e+00  5.175e-02  60.963  < 2e-16 ***
## Area        -5.799e-04  2.627e-05 -22.074  < 2e-16 ***
## Elevation    3.541e-03  8.741e-05  40.507  < 2e-16 ***
## Nearest      8.826e-03  1.821e-03   4.846 1.26e-06 ***
## Scruz       -5.709e-03  6.256e-04  -9.126  < 2e-16 ***
## Adjacent    -6.630e-04  2.933e-05 -22.608  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3510.73  on 29  degrees of freedom
```

```
## Residual deviance:  716.85  on 24  degrees of freedom
## AIC: 889.68
##
## Number of Fisher Scoring iterations: 5
```

```
step(PoiFit)
```

```
## Start:  AIC=889.68
## Species ~ Area + Elevation + Nearest + Scruz + Adjacent
##
##             Df Deviance     AIC
## <none>           716.85  889.68
## - Nearest    1   739.41  910.24
## - Scruz      1   813.62  984.45
## - Area       1  1204.35 1375.18
## - Adjacent   1  1341.45 1512.29
## - Elevation  1  2389.57 2560.40
```

```
##
## Call:  glm(formula = Species ~ Area + Elevation + Nearest + Scruz +
##     Adjacent, family = poisson, data = gala)
##
## Coefficients:
## (Intercept)         Area    Elevation      Nearest        Scruz     Adjacent
##   3.1548079   -0.0005799    0.0035406    0.0088256   -0.0057094   -0.0006630
##
## Degrees of Freedom: 29 Total (i.e. Null);  24 Residual
## Null Deviance:       3511
## Residual Deviance: 716.8     AIC: 889.7
```