

# Final Project

```
library(dplyr)
library(lubridate)
library(ggplot2)
library(tree)
midterm_polls = read.csv("C:/Users/mitch/Documents/school/499/data/mid_polls.csv")

approve_polls = read.csv("C:/Users/mitch/Documents/school/499/data/approv_polls.csv")

approve_polls$enddate = mdy(approve_polls$enddate)
midterm_polls$enddate = mdy(midterm_polls$enddate)

approve_mean = approve_polls %>% group_by(enddate) %>% summarise(appr = mean(approve))

mid_appr = left_join(midterm_polls,approve_mean, by = "enddate")

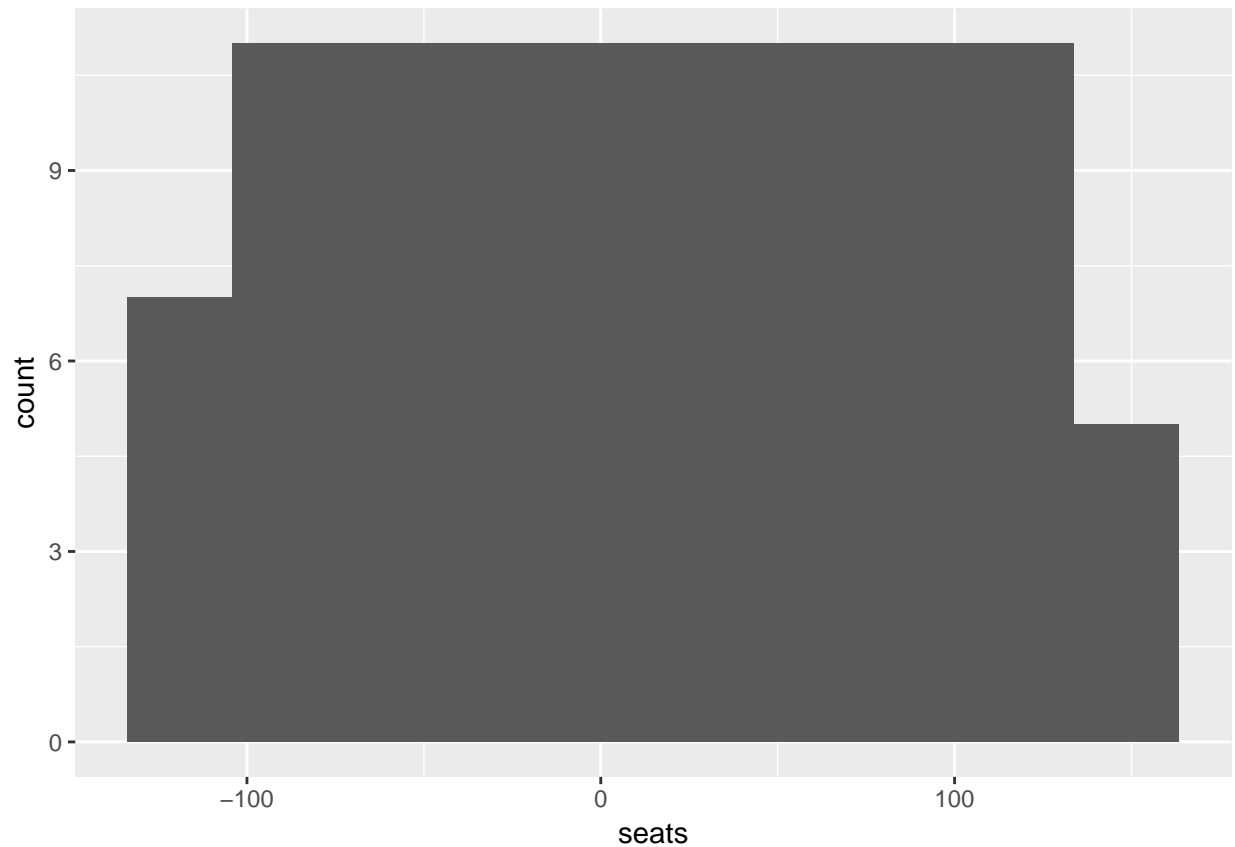
mid_appr = mid_appr %>% mutate(demwin = as.numeric(dem - rep), demwin2 = as.numeric(dem>rep))

mod1 = lm(demwin ~ appr + samplesize + enddate,data = mid_appr)

x = ymd("2018-04-19")
enddate = rep(x,100)
appr = 1:100
samplesize = rep(1000,100)

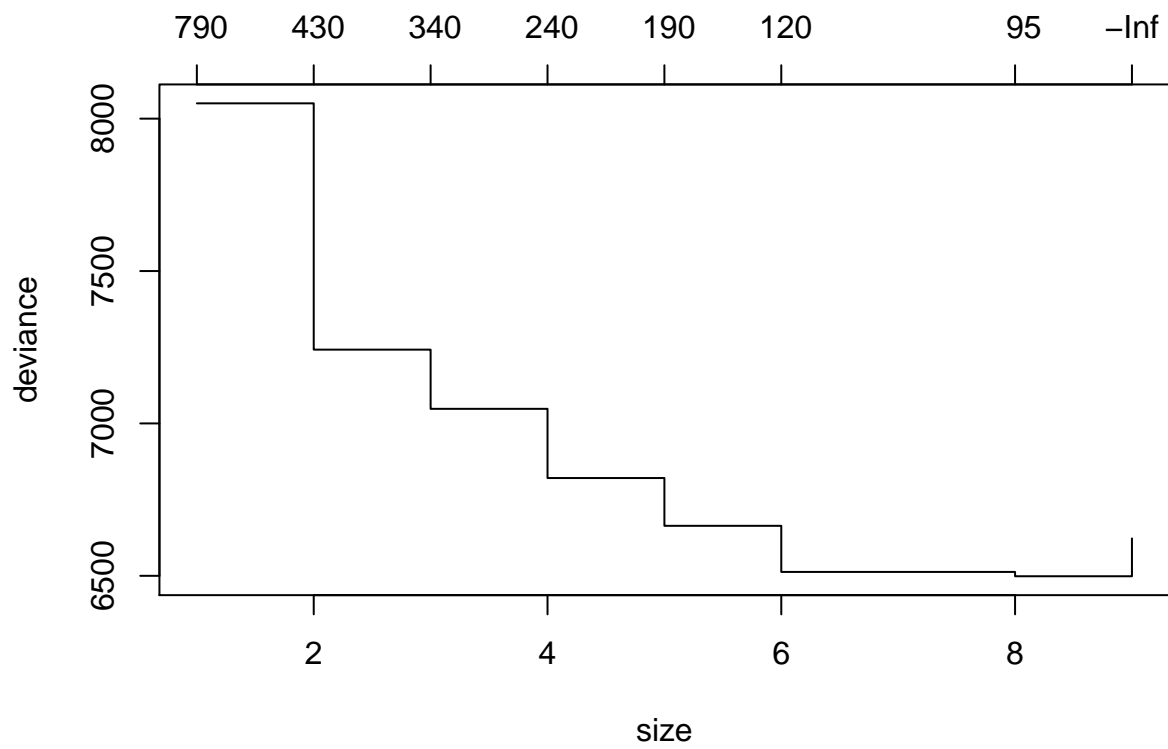
df = data.frame(appr,samplesize,enddate)

predictions = data.frame(predict(mod1,df))
predictions = predictions %>% mutate(probs = predict.mod1..df./100, seats = 468*probs)
ggplot(predictions, aes(x = seats)) + geom_histogram(bins = 10)
```

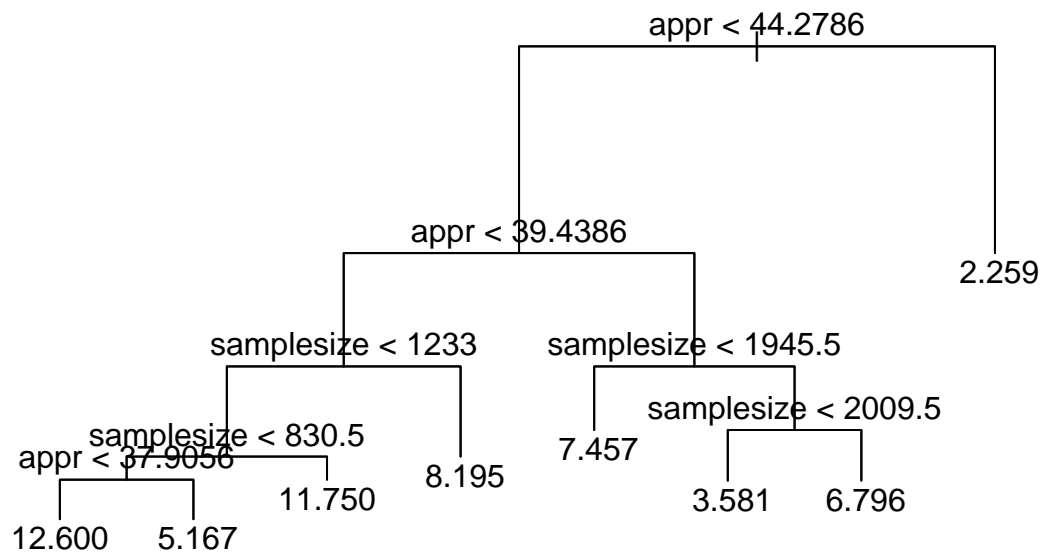


```
library(randomForest)
OJ = mid_appr %>% select(demwin,appr,samplesize,enddate)
#OJ$demwin2 = as.factor(OJ$demwin2)
set.seed(1)
train <- sample(1:nrow(OJ), 800)
OJ.train <- OJ[train, ]
OJ.test <- OJ[-train, ]
mod2 <- tree(demwin ~ .-demwin, data = OJ.train)

cv.oj <- cv.tree(mod2)
plot(cv.oj)
```



```
mod2prune <- prune.tree(mod2, best = 8)
plot(mod2prune)
text(mod2prune, pretty = 0)
```



```

predictions2 = data.frame(predict(mod2prune,df))
predictions2 = predictions2 %>% mutate(probs = predict.mod2prune..df./100, seats = 468*probs)
ggplot(predictions2, aes(x = seats)) + geom_histogram()

```

