ADV Data Analysis Group Project Notes

Mourtallah Faye sf2732

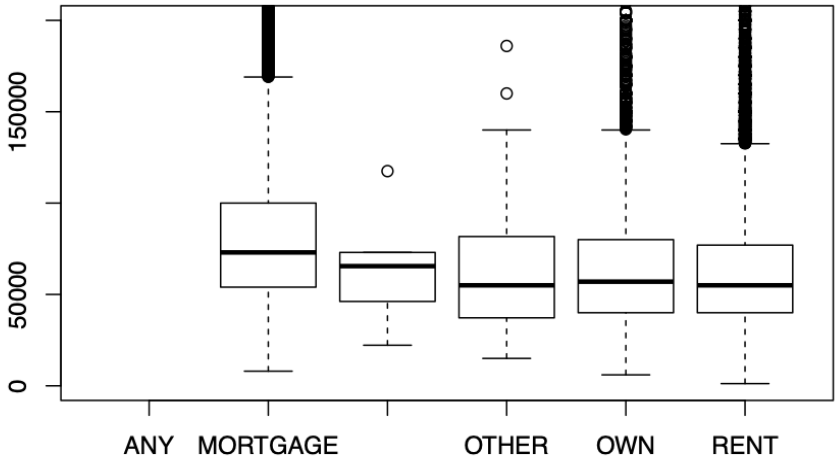
10/1/2018

Homeownership and Creditworthiness:

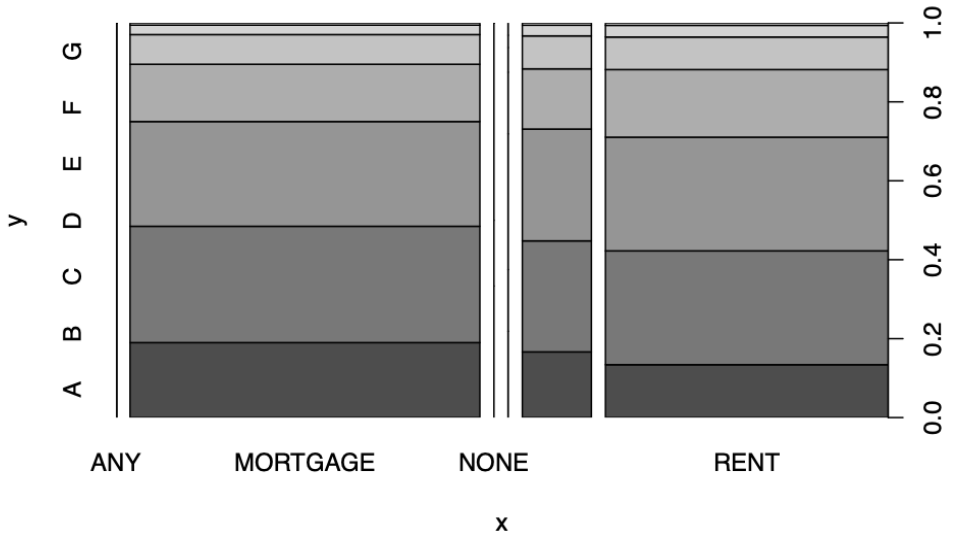
Mortgaging -> Loan Quality?

I've always thought that when it comes to credit and loans, it pays to be a winner. Homeownership is one of the hallmarks of credit worthiness, but I want to explore the relationship explicitly. Below are some cursory plots with some proxy variables. I propose multivariate regression to isolate variables that are correlated most directly with home ownership and creditworthiness, developing a model that teases out the relationship but also controls for the cofounding variables and multicollinearity that is bound to exist.

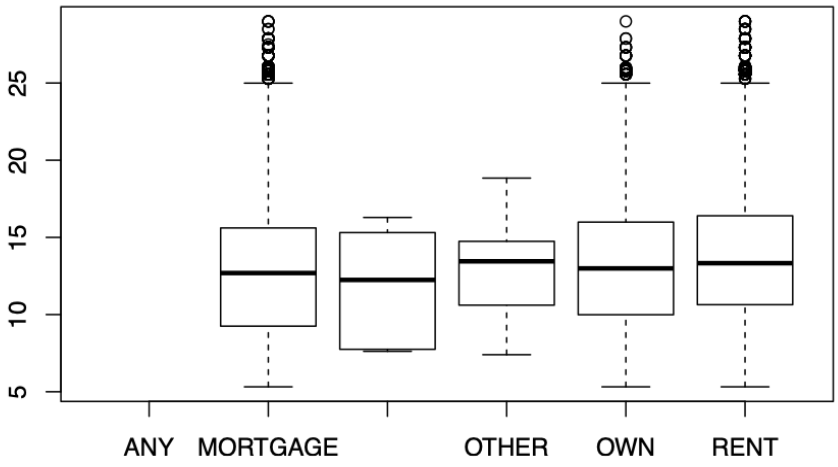
```
plot(sample$home_ownership, sample$annual_inc, ylim = c(0,2e+05))
```



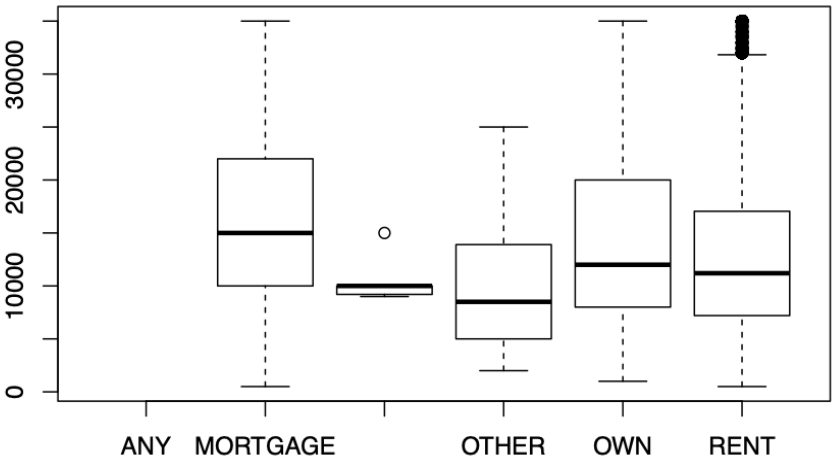

```
plot(sample$home_ownership , sample$grade)
```




```
plot(sample$home_ownership , sample$int_rate)
```




```
plot(sample$home_ownership , sample$loan_amnt)
```

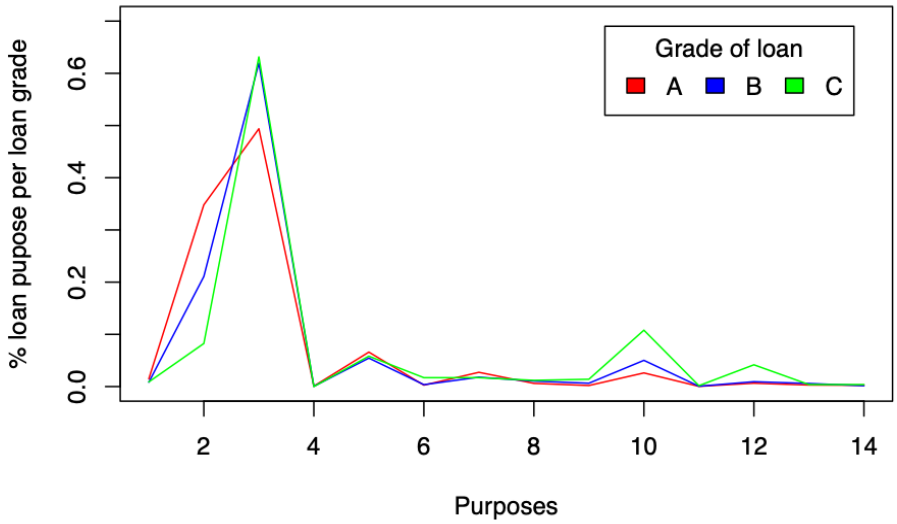


Loan Purpose and Grade?

Below I have plotted the proportions of loan purpose for each grade. With red being grade A, blue being C, and green being F. Along the x axis are the different purposes. This relationship would mean that either the banks see loans differently based on purpose and rank them accordingly, or that, and if, a certain type of person is more likely to get a certain grade loan, then they are more likely to be taking out a loan for so-and -so purpose. I would love to explore more about how these banks see debt and the risk with debt being delpoyed and different ways, or a profile of sorts detailing what debt deployment looks like at differnt levels of credit and loan health.

I intend to use logistic regression the odds ratio and the Chi-squared test to these ends.

```
plot(round(prop.table(table(sample$purpose, sample$grade)
                      ,margin = 2), 4)[,1],type = "l", col = 'red', ylim = c(0,.7), xlab = "Purposes", 
lines(round(prop.table(table(
  sample$purpose, sample$grade),margin = 2), 4)[,3], type = "l", col = "blue")
lines(round(prop.table(table(
  sample$purpose, sample$grade),margin = 2), 4)[,6], type = "l", col = "green")
legend("topright", inset=.05, title="Grade of loan",
    c("A","B","C"), fill=c('red','blue','green'), horiz=TRUE)
```




```
#axis(1, at = c(1:14), labels = levels(loanz$purpose))
```