homework2

Sharleen Price spp2122

2/19/2018

```
housing <- read.csv("NYChousing.csv", as.is=TRUE)
dim(housing)
```

```
## [1] 2506   22
```

```
apply(is.na(housing), 2, sum)
```

```
##                           UID                  PropertyName
##                             0                             0
##                           Lon                           Lat
##                            15                            15
##                       AgencyID                          Name
##                             0                             0
##                         Value                       Address
##                            52                             0
##                 Violations2010                    REACNumber
##                             0                          1873
##                       Borough                            CD
##                             0                             0
##             CityCouncilDistrict                   CensusTract
##                            10                             0
##                  BuildingCount                     UnitCount
##                             0                             0
##                      YearBuilt                         Owner
##                             0                             0
##                    Rental.Coop             OwnerProfitStatus
##                             0                             0
##       AffordabilityRestrictions StartAffordabilityRestrictions
##                             0                             5
```

    ii. The data has 2506 rows and 22 columns iii.This command goes through all the columns and checks if it has a value or if it is NA, it only counts it if its NA. Then it applies the sum function so that we can list the total amount of NA values in each category.

    iii.

```
housing <- na.omit(housing)
dim(housing)
```

```
## [1] 630  22
```

    v. I removed 1876 rows and it does agree with the result from iii.

vi.vii.viii.

```
housing$logValue <- log(housing$Value)
summary(housing$logValue)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.06   13.82   14.65   14.65   15.38   20.22
```

```
housing$logUnits <-log(housing$UnitCount)
housing$after1950 <-housing$YearBuilt>=1950
```

Part 2 i.

```
plot(housing$logUnits,housing$logValue, col = factor(housing$after1950), xlab = "Log of Units", ylab = 

legend("bottomright", legend = levels(factor(housing$after1950)), fill = unique(factor(housing$after1950
```
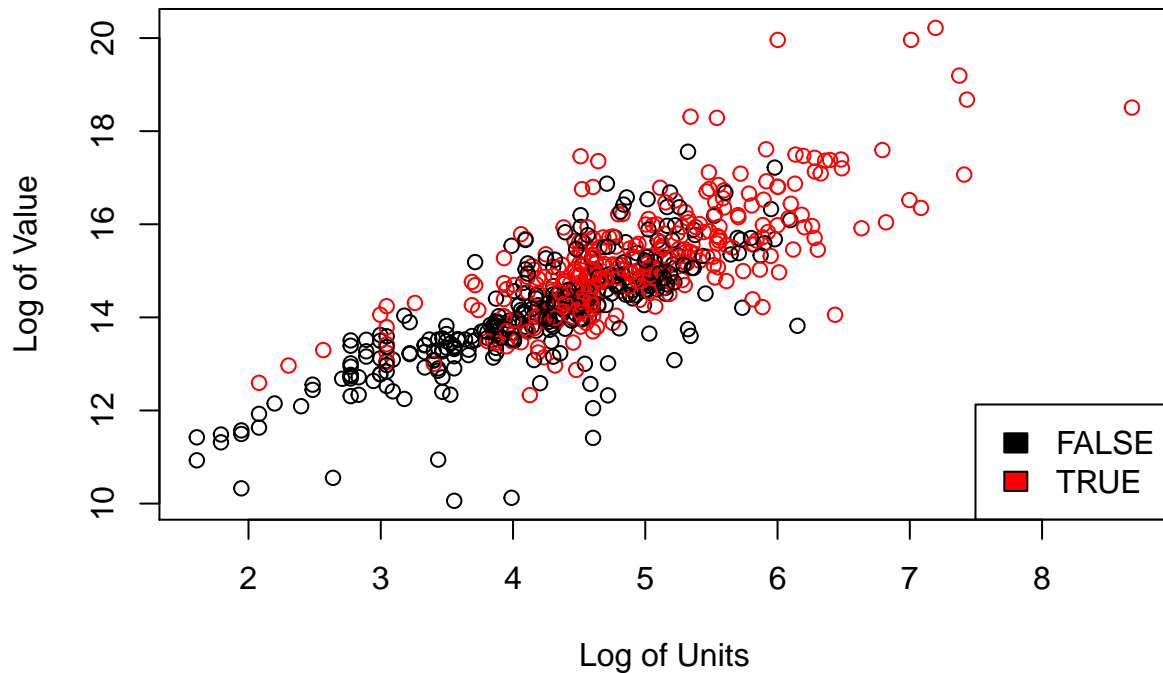


```
cor(housing$logUnits, housing$logValue)
```

```
## [1] 0.7988655
```

ii. The plot shows how the value of the apartments increased as the amount of units in the property increased. The colors show that after the 1950, bigger buildings with more units were built and thus also became more expensive.

iii.

```
Manhattan <- housing[housing$Borough == "Manhattan",]
cor(Manhattan$logUnits, Manhattan$logValue)
```

```
## [1] 0.8710823
```

```
Brooklyn <- housing[housing$Borough == "Brooklyn",]
cor(Brooklyn$logUnits, Brooklyn$logValue)
```

```
## [1] 0.8053241
```

```
after<- housing[housing$after1950,]
cor(after$logUnits, after$logValue)
```
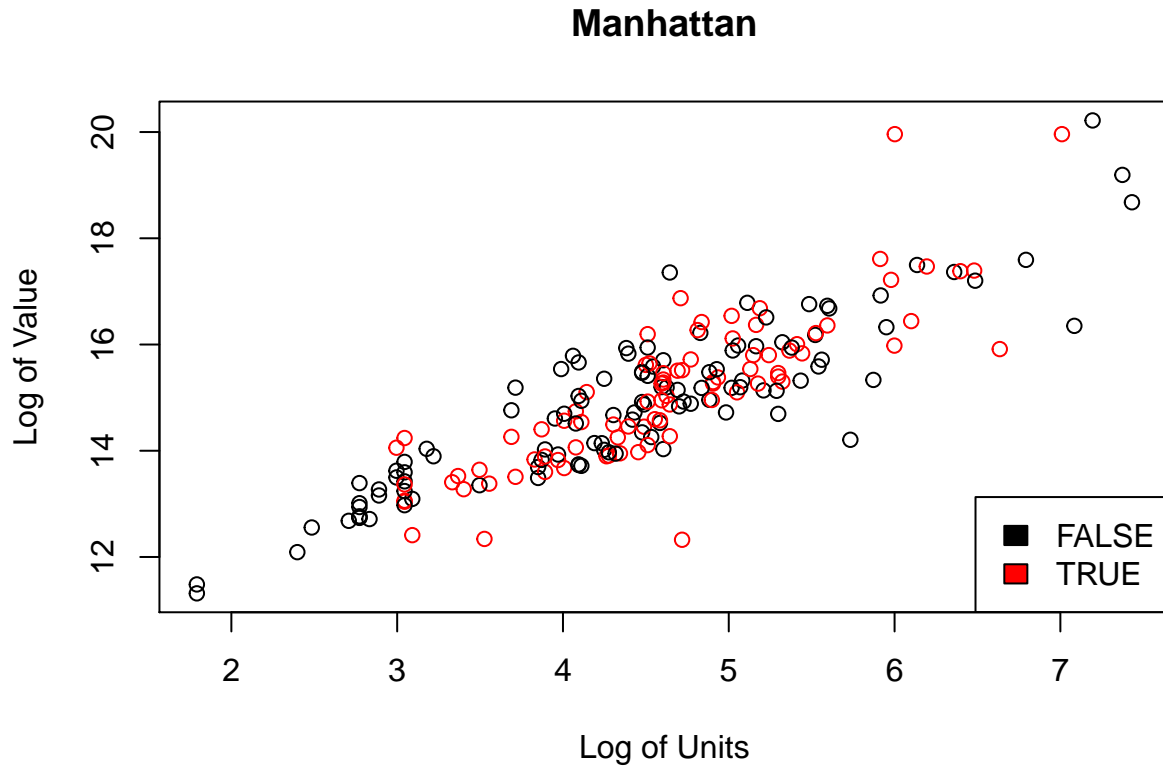
```
## [1] 0.746731
```

```
before<-housing[housing$YearBuilt<1950,]
cor(before$logUnits, before$logValue)
```

```
## [1] 0.7720285
```

iv.

```
plot(Manhattan$logUnits, Manhattan$logValue, col = factor(housing$after1950),main = "Manhattan",xlab = "
legend("bottomright", legend = levels(factor(housing$after1950)), fill = unique(factor(housing$after1950
```

Manhattan


```
plot(Brooklyn$logUnits, Brooklyn$logValue, col = factor(housing$after1950),main = "Brooklyn", xlab = "L
legend("bottomright", legend = levels(factor(housing$after1950)), fill = unique(factor(housing$after195
```

Brooklyn

v.
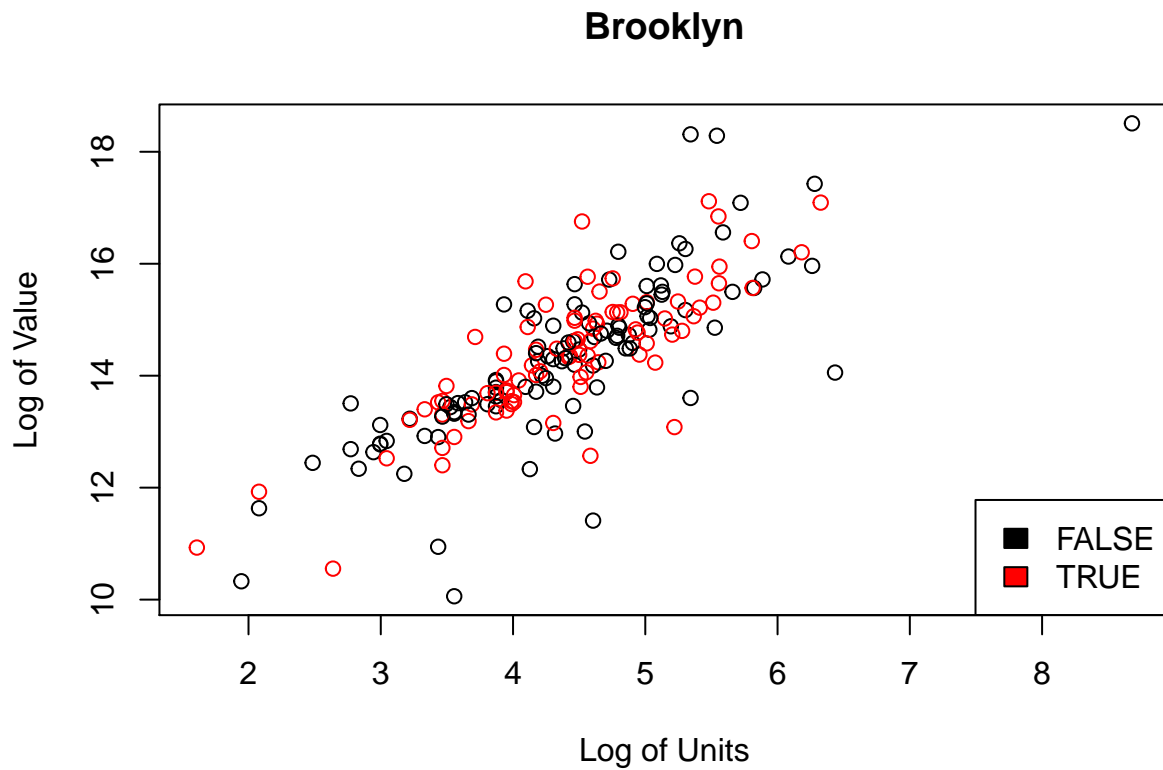
```r
manhat.props <- c()
    for (props in 1:nrow(housing)) {
      if (housing$Borough[props] == "Manhattan") {
        manhat.props <- c(manhat.props, props)
      }
}
    med.value <- c()
    for (props in manhat.props) {
      med.value <- c(med.value, housing$Value[props])
    }
    med.value <- median(med.value, na.rm = TRUE)

    med.value
```
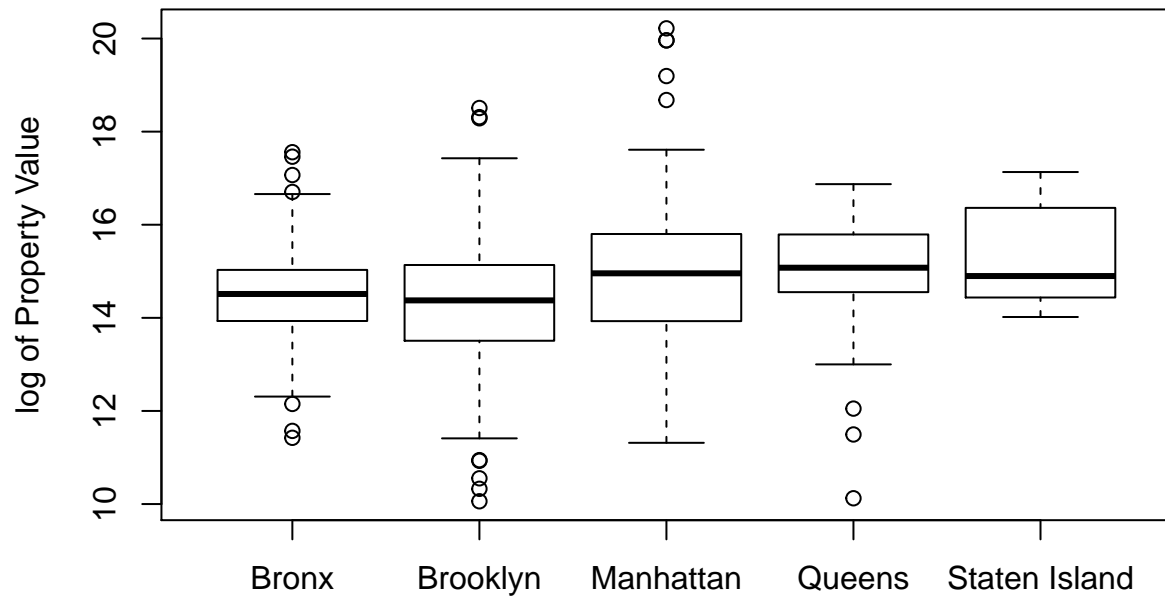
```
## [1] 3129300
```

```r
    median(housing$Value[housing$Borough=="Manhattan"],na.rm = TRUE)
```

```
## [1] 3129300
```

vi.

```r
boxplot(housing$logValue~housing$Borough,main="Housing by Borough",ylab="log of Property Value", data=h
```

Housing by Borough

vii.

```r
tapply(housing$Value,housing$Borough,median)
```

```
##        Bronx     Brooklyn    Manhattan       Queens Staten Island
##      2008260      1749465      3129300      3529800       2952900
```