Assignment 5

Max Troeger

2025-03-28

1. Derived Dataset - Staten Island

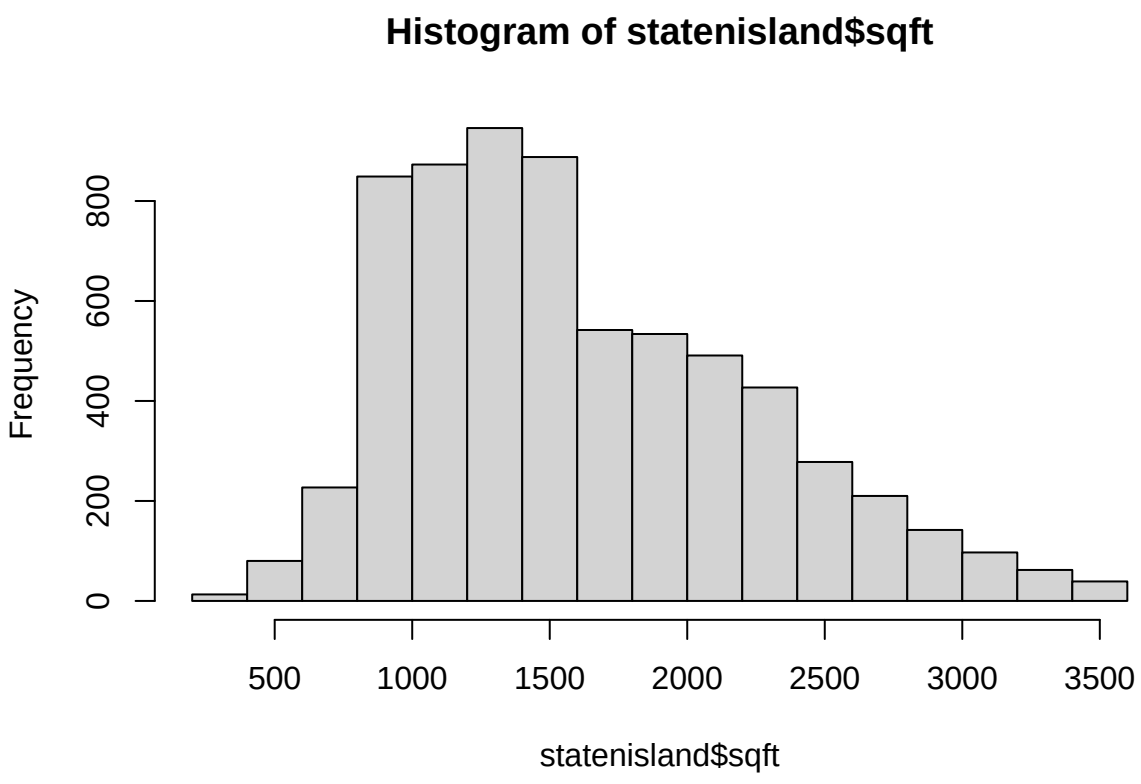
a. For the first derived dataset I chose Staten Island. I'm looking for the interaction among the gross square footage, total number of units, and sale price terms. Because many sale prices are 0 or ridiculously high/low, I'll need to constrain the prices to eliminate outliers.

```
statenisland <- subset(nycdata,BOROUGH=="STATEN ISLAND")
statenisland$sqft <- as.numeric(statenisland$`GROSS SQUARE FEET`)
```

```
## Warning: NAs introduced by coercion
```

```
statenisland <- subset(statenisland,sqft!=0)

quartiles <- quantile(statenisland$sqft, probs=c(.25, .75), na.rm = TRUE)
IQR <- IQR(statenisland$sqft)
Lower <- quartiles[1] - 1.5*IQR
Upper <- quartiles[2] + 1.5*IQR
statenisland <- subset(statenisland, statenisland$sqft > Lower & statenisland$sqft < Upper)
hist(statenisland$sqft)
```



b. c.

this is a test[1]

2. Derived Dataset - Manhattan

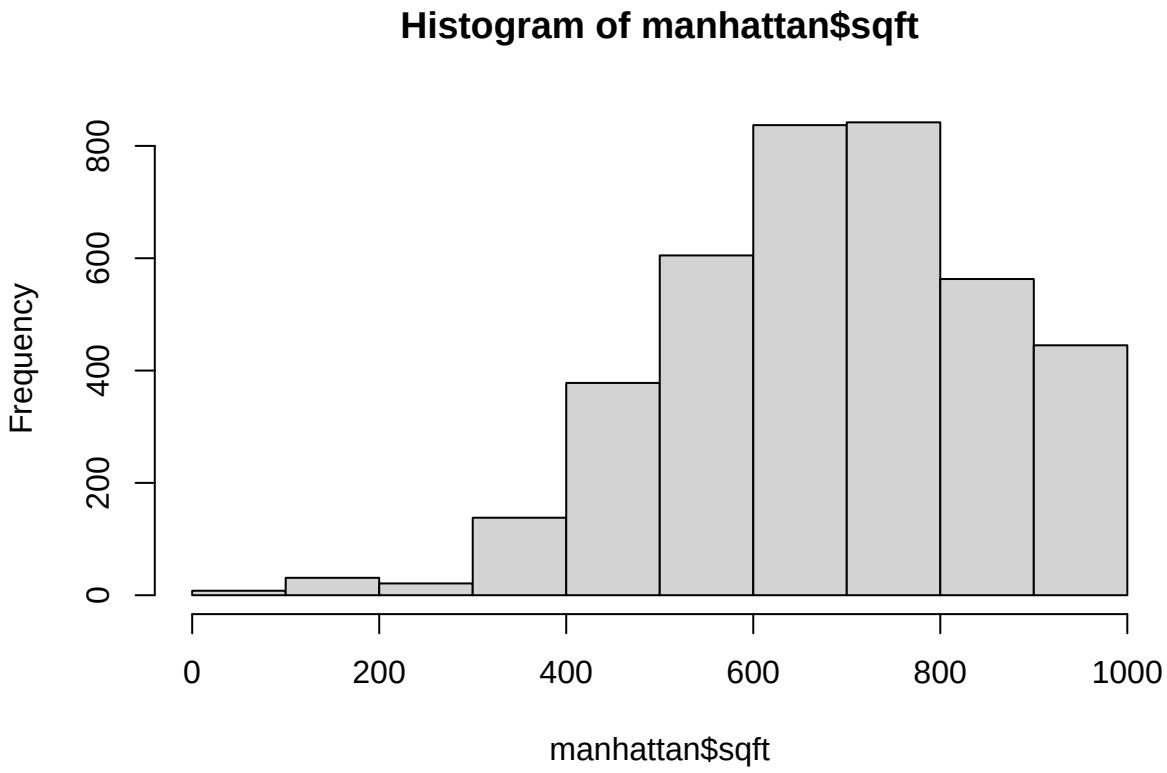
For the second derived dataset I chose Manhattan. Although units in Manhattan include entries of the form `BOROUGH=["1","MANHATTAN"]`, choosing just the former gives a sufficiently large sample size for our purposes (~96,000 observations). Again, I'm looking for the interaction among the gross square footage, total number of units, and sale price terms with the same comment about outliers.

```
manhattan <- subset(nycdata,BOROUGH=="1")
manhattan$sqft <- as.numeric(manhattan$`GROSS SQUARE FEET`)
```

```
## Warning: NAs introduced by coercion
```

```
manhattan <- subset(manhattan,sqft!=0)

quartiles <- quantile(manhattan$sqft, probs=c(.25, .75), na.rm = TRUE)
Lower <- quartiles[1] - 1.5*IQR
Upper <- quartiles[2] + 1.5*IQR
manhattan <- subset(manhattan, manhattan$sqft > Lower & manhattan$sqft < Upper)
hist(manhattan$sqft)
```



3. Conclusions

[1] footnote hehe