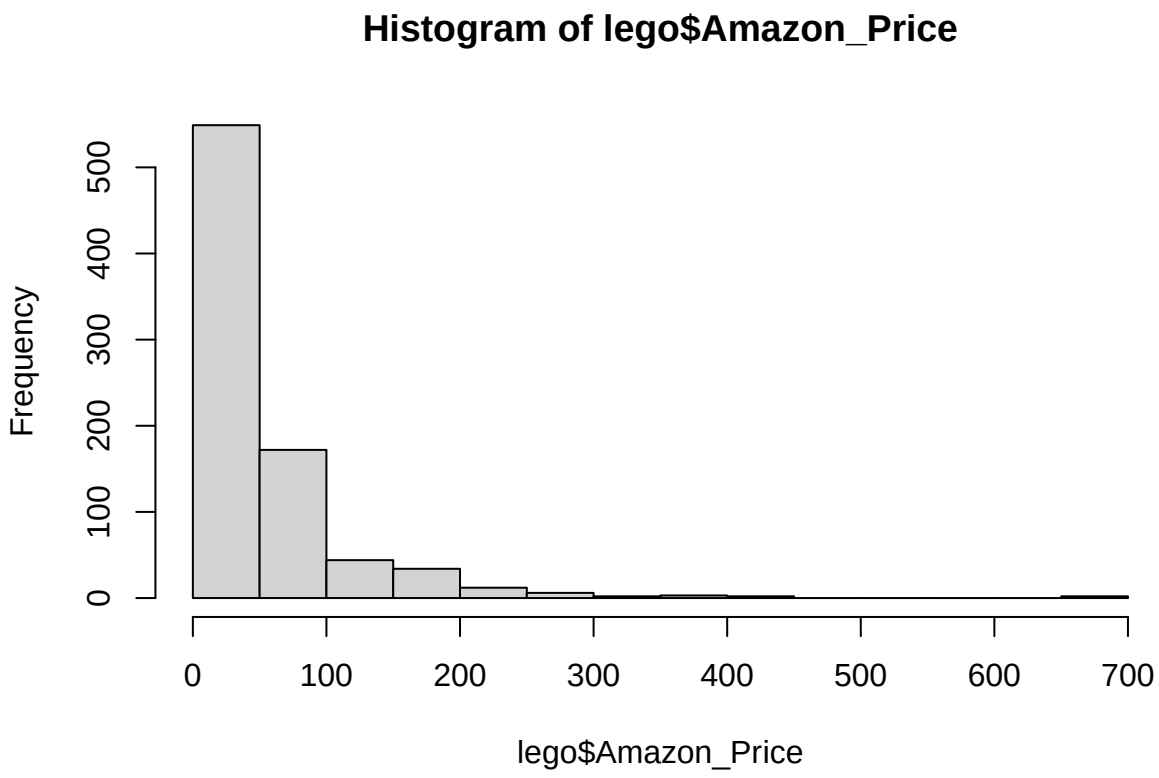
Supplement for Lecture 2: Four Step Process

Read in Datasets

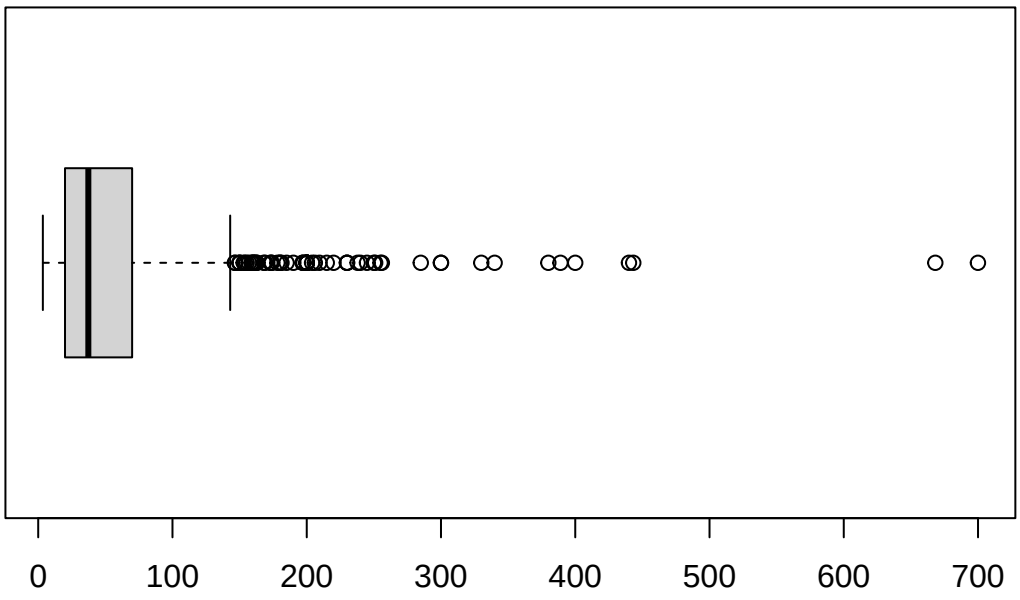
```
lego <- read_csv("lego.csv")
```

Visuals of Amazon Price

```
hist(lego$Amazon_Price)
```




```
boxplot(lego$Amazon_Price, horizontal = TRUE)
```



Estimating Constant for Models

```
#We know that we don't have Amazon Prices for Certain Lego Sets
Amazon_lego = ...

xbar = mean(Amazon_lego$Amazon_Price)
m = median(Amazon_lego$Amazon_Price)

xbar
m
```

Assessing Model Fit

```
xbar_resid = ...
m_resid = ...

xbar_sse=sum(...)
xbar_sae=sum(...)

m_sse=sum(...)
m_sae=sum(...)

data.frame(Estimator=c("Mean","Median"),Sum_Squared_Errors=c(xbar_sse,m_sse), Sum_Absolute_Errors=c(xba
```

Use of Mosaic Package

```
mean(Amazon_Price~Theme,data=Amazon_lego) #Average Price Per Theme
tally(~Theme,data=Amazon_lego) #Number of Observations For Each Theme
sd(Amazon_Price~Theme,data=Amazon_lego) #SD of Price Per Theme
boxplot(Amazon_Price~Theme,data=Amazon_lego,las=2) #Side-by-Side Boxplots
```

Two-Sample t-Test for Difference in Means

```
lego_2_theme = subset(...)

t.test(...) #Conduct t-Test

boxplot(...) #Evaluate Assumptions
tally(...) #Sample size > 30, Enough Data for CLT
```