

Supplement for Lecture 2: Four Step Process

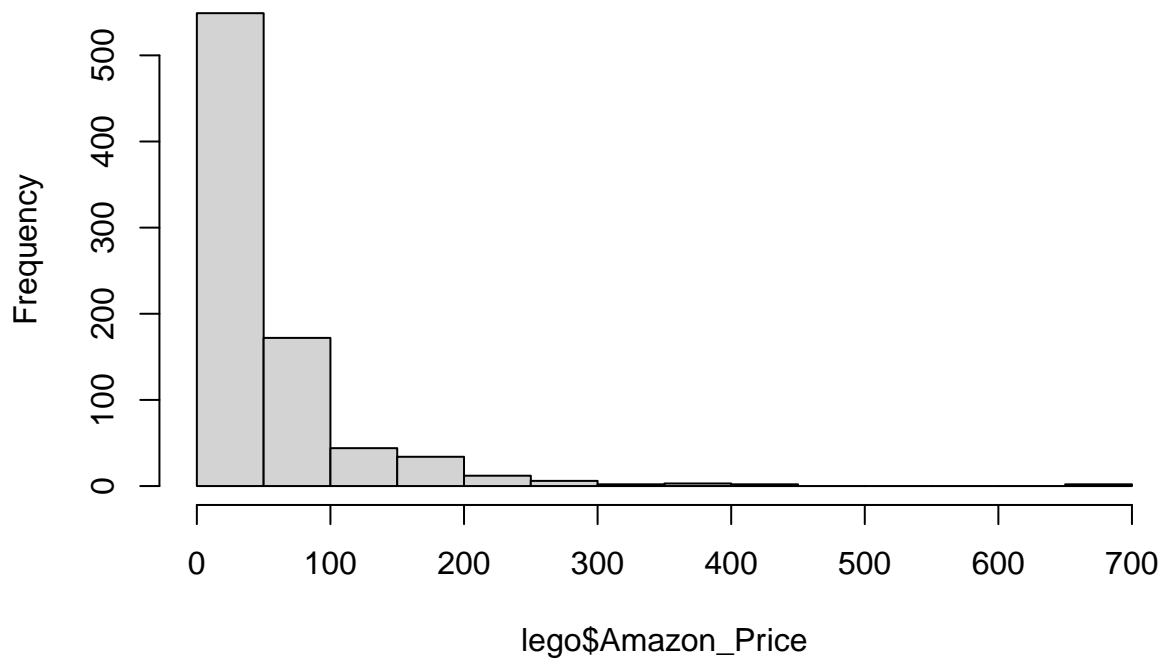
Read in Datasets

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

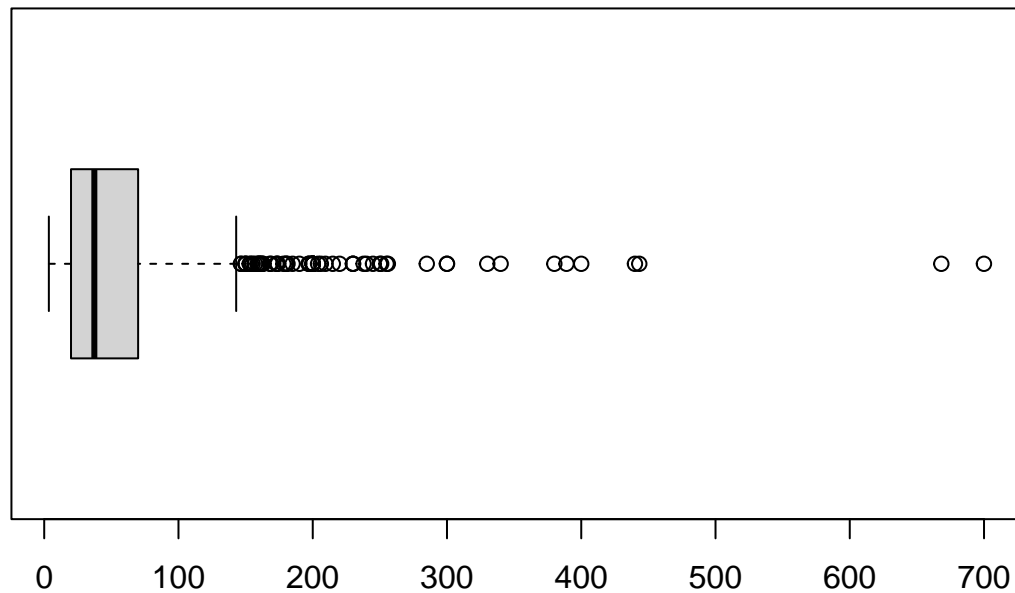
Visuals of Amazon Price

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

Histogram of 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 = subset(lego, is.na(Amazon_Price)==F)
```

```
xbar = mean(Amazon_lego$Amazon_Price)
```

```
m = median(Amazon_lego$Amazon_Price)
```

```
xbar
```

```
## [1] 57.8232
```

```
m
```

```
## [1] 37.325
```

Assessing Model Fit

```
xbar_resid = Amazon_lego$Amazon_Price - xbar
```

```
m_resid = Amazon_lego$Amazon_Price - m
```

```
xbar_sse=sum(xbar_resid^2)
```

```
xbar_sae=sum(abs(xbar_resid))
```

```
m_sse=sum(m_resid^2)
```

```
m_sae=sum(abs(m_resid))
```

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

```
## Estimator Sum_Squared_Errors Sum_Absolute_Errors
## 1 Mean 3622919 34330.20
## 2 Median 3969984 30441.84
```

Use of Mosaic Package

```
mean(Amazon_Price~Theme,data=Amazon_lego) #Average Price Per Theme
```

```
## Architecture Batman BrickHeadz City
## 71.85545 40.57786 46.35000 51.97899
## Classic Creator 3-in-1 Creator Expert DC
## 24.88667 40.65657 178.39667 79.72000
## Disney DOTS DUPL0 Friends
## 38.14206 10.48000 40.24698 42.26448
## Harry Potter Hidden Side Ideas Juniors
## 69.33176 44.44421 124.87000 41.66333
## Jurassic World LEGO Art LEGO Frozen 2 LEGO Super Mario
## 69.70250 119.97667 33.82000 33.12900
## Marvel Minecraft Minions NINJAGO
## 51.69756 52.11731 31.31000 55.36123
## Overwatch Powered UP Powerpuff Girls Speed Champions
## 40.47875 267.70667 28.26500 34.85000
## Spider-Man Star Wars Stranger Things Technic
## 39.39000 80.70656 199.95000 103.44639
## THE LEGO MOVIE 2 Trolls World Tour Unikitty!
## 51.52762 25.89333 27.50400
```

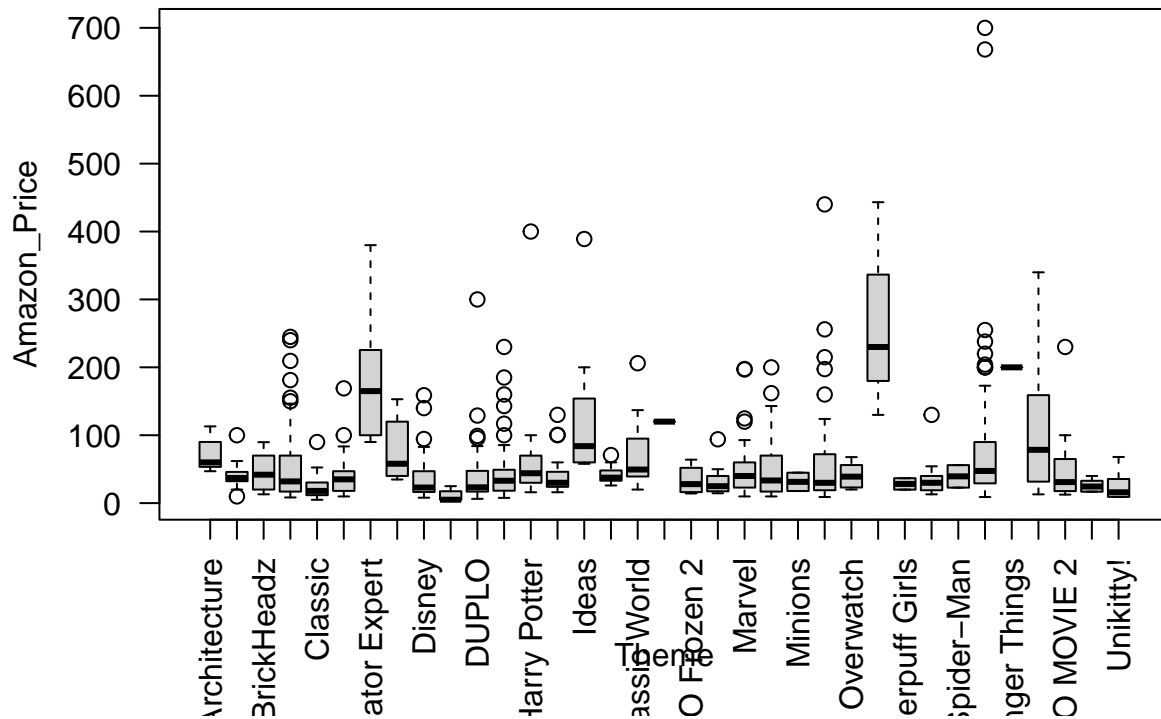
```
tally(~Theme,data=Amazon_lego) #Number of Observations For Each Theme
```

```
## Theme
## Architecture Batman BrickHeadz City
## 11 14 14 89
## Classic Creator 3-in-1 Creator Expert DC
## 15 35 12 9
## Disney DOTS DUPL0 Friends
## 34 15 43 87
## Harry Potter Hidden Side Ideas Juniors
## 17 19 10 12
## Jurassic World LEGO Art LEGO Frozen 2 LEGO Super Mario
## 16 3 8 10
## Marvel Minecraft Minions NINJAGO
## 41 26 2 65
## Overwatch Powered UP Powerpuff Girls Speed Champions
## 8 3 2 17
## Spider-Man Star Wars Stranger Things Technic
## 2 96 1 36
## THE LEGO MOVIE 2 Trolls World Tour Unikitty! <NA>
## 21 6 5 22
```

```
sd(Amazon_Price~Theme,data=Amazon_lego) #SD of Price Per Theme
```

##	Architecture	Batman	BrickHeadz	City
##	23.31210302	21.45698960	26.80881371	51.57962666
##	Classic	Creator 3-in-1	Creator Expert	DC
##	22.50993611	32.31908514	92.68584499	47.19911519
##	Disney	DOTS	DUPLO	Friends
##	35.22640786	7.66767147	49.13191391	37.97985665
##	Harry Potter	Hidden Side	Ideas	Juniors
##	89.63316994	31.82733858	103.70226409	13.15945103
##	Jurassic World	LEGO Art	LEGO Frozen 2	LEGO Super Mario
##	48.76308105	0.02309401	20.57715724	24.27327174
##	Marvel	Minecraft	Minions	NINJAGO
##	43.80120471	48.91803885	19.12016736	69.80713423
##	Overwatch	Powered UP	Powerpuff Girls	Speed Champions
##	18.92740073	160.01155469	11.91474926	27.99127051
##	Spider-Man	Star Wars	Stranger Things	Technic
##	23.47594514	104.93632443	NA	90.02498425
##	THE LEGO MOVIE 2	Trolls World Tour	Unikitty!	
##	50.66333950	9.13094446	25.06838806	

```
boxplot(Amazon_Price~Theme,data=Amazon_lego,las=2) #Side-by-Side Boxplots
```



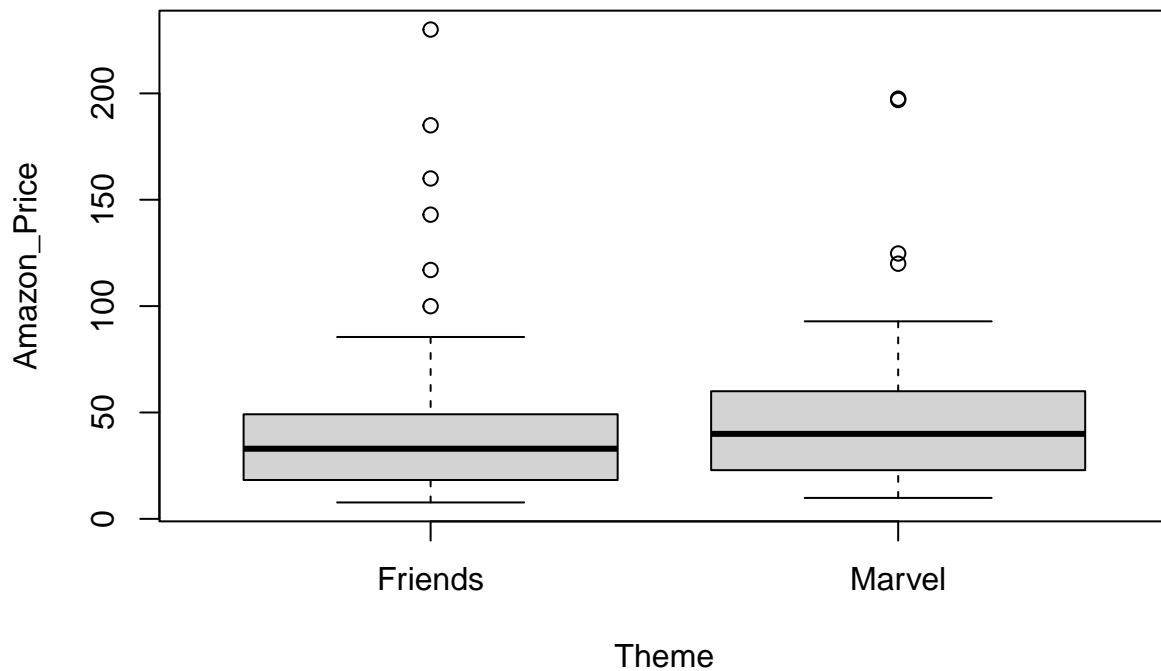
Two-Sample t-Test for Difference in Means

```
lego_2_theme = subset(Amazon_lego, Theme == 'Friends' | Theme == 'Marvel')  
#lego_2_theme = subset(Amazon_lego, Theme %in% c('Friends', 'Marvel'))
```

```
t.test(Amazon_Price~Theme, data= lego_2_theme) #Conduct t-Test
```

```
##  
## Welch Two Sample t-test  
##  
## data: Amazon_Price by Theme  
## t = -1.1849, df = 69.32, p-value = 0.2401  
## alternative hypothesis: true difference in means between group Friends and group Marvel is not equal  
## 95 percent confidence interval:  
## -25.313078 6.446922  
## sample estimates:  
## mean in group Friends mean in group Marvel  
## 42.26448 51.69756
```

```
boxplot(Amazon_Price~Theme, data=lego_2_theme) #Evaluate Assumptions
```



```
tally(~Theme, data=lego_2_theme) #Sample size > 30, Enough Data for CLT
```

```
## Theme  
## Friends Marvel  
## 87 41
```