Data Science XTern

setwd("~/OneDrive - Butler University/MA362")  
  
XTernData=read.table("XTernData.txt", sep = "\t", header = T)  
  
data.frame(XTernData)

## Location Distance\_from\_Porto\_Alegre  
## 1 Bru Burger 2.2  
## 2 Gomez Barbeque 5.7  
## 3 India Garden 2.1  
## 4 Fresco 1.8  
## 5 Bakersfield 2.0  
## 6 Salt on Mass 2.4  
## 7 Goodfellas 2.4  
## 8 Modita 2.8  
## 9 Bangkok 2.1  
## 10 Saffron Cafe 2.4  
## 11 Indy 500 3.5  
## 12 One Team Scavengar Hunt indianapolis 1.7  
## 13 The LUME 5.0  
## 14 Escape Room Indy 2.1  
## 15 Indianapolis Zoo 1.8  
## 16 Business Networking 1 day training 21.3  
## Distance\_from\_Winthrop\_Ave Distance\_from\_Cedar\_St  
## 1 4.6 18.1  
## 2 4.7 20.6  
## 3 5.0 18.1  
## 4 5.0 17.6  
## 5 4.6 18.1  
## 6 4.4 18.1  
## 7 4.4 18.1  
## 8 3.9 18.1  
## 9 4.8 18.2  
## 10 4.4 17.9  
## 11 8.3 13.8  
## 12 5.4 17.7  
## 13 2.9 13.6  
## 14 5.0 18.1  
## 15 7.6 19.4  
## 16 6.8 11.5  
## Distance\_from\_Visionary\_Way Distance\_from\_Mass\_Ave Distance\_from\_Meridian\_St  
## 1 20.7 0.2 5.1  
## 2 19.0 2.7 8.3  
## 3 21.1 0.3 5.1  
## 4 21.3 0.8 5.5  
## 5 20.8 0.1 5.3  
## 6 20.6 0.3 5.6  
## 7 20.7 0.3 5.6  
## 8 15.7 1.1 6.2  
## 9 21.0 0.2 5.1  
## 10 16.1 0.5 5.6  
## 11 19.7 5.0 13.6  
## 12 21.4 0.7 5.1  
## 13 15.0 5.0 19.1  
## 14 21.4 0.6 5.0  
## 15 23.5 3.2 7.6  
## 16 7.2 19.8 24.8

XTernData$Distance\_from\_Porto\_Alegre

## [1] 2.2 5.7 2.1 1.8 2.0 2.4 2.4 2.8 2.1 2.4 3.5 1.7 5.0 2.1 1.8  
## [16] 21.3

XTernData$Distance\_from\_Winthrop\_Ave

## [1] 4.6 4.7 5.0 5.0 4.6 4.4 4.4 3.9 4.8 4.4 8.3 5.4 2.9 5.0 7.6 6.8

XTernData$Distance\_from\_Cedar\_St

## [1] 18.1 20.6 18.1 17.6 18.1 18.1 18.1 18.1 18.2 17.9 13.8 17.7 13.6 18.1 19.4  
## [16] 11.5

XTernData$Distance\_from\_Visionary\_Way

## [1] 20.7 19.0 21.1 21.3 20.8 20.6 20.7 15.7 21.0 16.1 19.7 21.4 15.0 21.4 23.5  
## [16] 7.2

XTernData$Distance\_from\_Mass\_Ave

## [1] 0.2 2.7 0.3 0.8 0.1 0.3 0.3 1.1 0.2 0.5 5.0 0.7 5.0 0.6 3.2  
## [16] 19.8

XTernData$Distance\_from\_Meridian\_St

## [1] 5.1 8.3 5.1 5.5 5.3 5.6 5.6 6.2 5.1 5.6 13.6 5.1 19.1 5.0 7.6  
## [16] 24.8

attach(XTernData)  
summary(Distance\_from\_Porto\_Alegre)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.700 2.075 2.300 3.831 2.975 21.300

summary(Distance\_from\_Winthrop\_Ave)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.900 4.400 4.750 5.112 5.100 8.300

summary(Distance\_from\_Cedar\_St)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 11.50 17.68 18.10 17.31 18.10 20.60

summary(Distance\_from\_Visionary\_Way)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 7.20 18.27 20.70 19.07 21.15 23.50

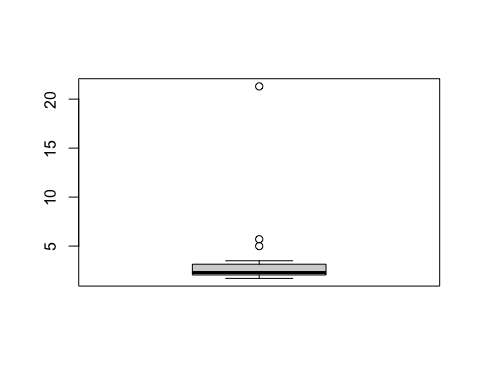
summary(Distance\_from\_Mass\_Ave)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.100 0.300 0.650 2.550 2.825 19.800

summary(Distance\_from\_Meridian\_St)

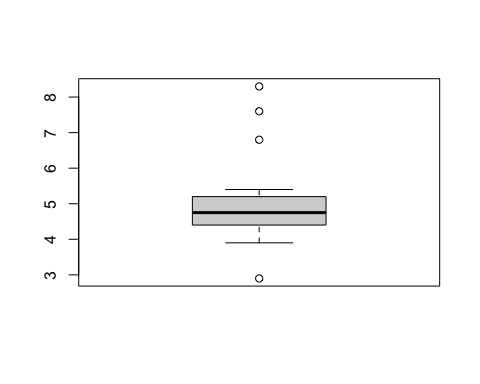
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5.000 5.100 5.600 8.287 7.775 24.800

#By analyzing the summary data we can get a sense of which locations have the smallest distance from our events. You can see that Porto Alegre and Mass Ave have the two smallest means. This tells us that both Porto Alegre and Mass Ave are going to be the closest to our events.   
  
boxplot(Distance\_from\_Porto\_Alegre)$out



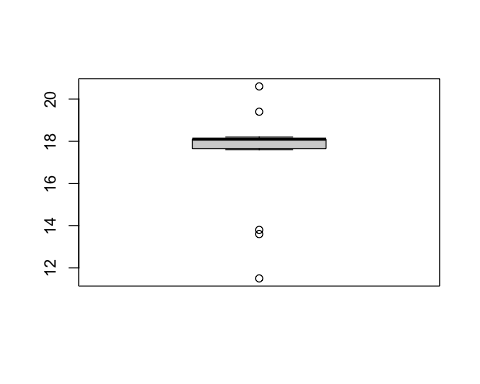
## [1] 5.7 5.0 21.3

boxplot(Distance\_from\_Winthrop\_Ave)$out



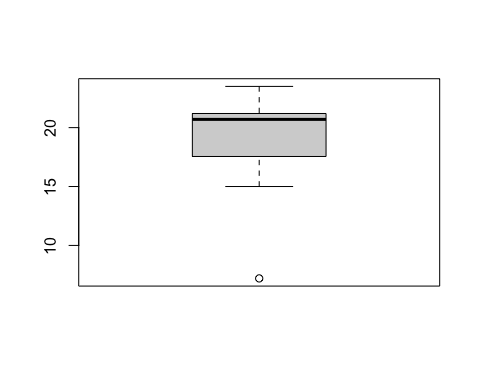
## [1] 8.3 2.9 7.6 6.8

boxplot(Distance\_from\_Cedar\_St)$out



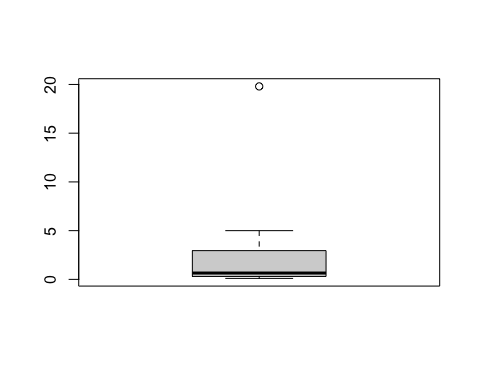
## [1] 20.6 13.8 13.6 19.4 11.5

boxplot(Distance\_from\_Visionary\_Way)$out



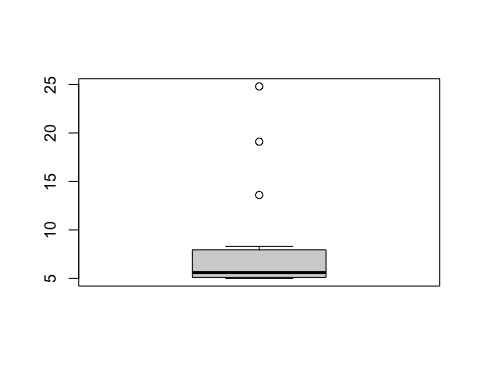
## [1] 7.2

boxplot(Distance\_from\_Mass\_Ave)$out



## [1] 19.8

boxplot(Distance\_from\_Meridian\_St)$out



## [1] 13.6 19.1 24.8

#I took a look at the boxplots next to see if any of the locations had outliers that would possibly be skewing the summary data from above. I found that Porto Alegre, Mass Ave, and Meridian St had upper bound outliers. Meaning most of the events took place close to the location, however there were a few outliers farther away from the location causing the mean to be slightly higher than the median. Winthrop Ave and Cedar St had both upper and lower outliers. Meaning there were a few events farther and a few events closer to the location of housing than the average location of the events. This should not affect their means too drastically, as you can see the boxplots still look normal. Lastly, Visionary Way had a couple lower bound outliers. Meaning most of the events take place further away, but there were some events pretty close to the housing location. This could have skewed the mean slightly lower than the median.   
  
  
mean(Distance\_from\_Porto\_Alegre)

## [1] 3.83125

mean(Distance\_from\_Winthrop\_Ave)

## [1] 5.1125

mean(Distance\_from\_Cedar\_St)

## [1] 17.3125

mean(Distance\_from\_Visionary\_Way)

## [1] 19.075

mean(Distance\_from\_Mass\_Ave)

## [1] 2.55

mean(Distance\_from\_Meridian\_St)

## [1] 8.2875

#The mean distance in miles from the housing location to the event location tells us which housing location is closest to all the events. The mean distance from smallest to largest is Mass Ave (2.55 miles), Porto Alegre (3.83 miles), Winthrop Ave (5.11 miles), Meridian St (8.29 miles), Cedar St (17.31 miles), and Visionary Way (19.08 miles). This tells us that Mass Ave is the closest housing location to all the event locations.   
  
sd(Distance\_from\_Porto\_Alegre)

## [1] 4.796765

sd(Distance\_from\_Winthrop\_Ave)

## [1] 1.366199

sd(Distance\_from\_Cedar\_St)

## [1] 2.320596

sd(Distance\_from\_Visionary\_Way)

## [1] 3.932684

sd(Distance\_from\_Mass\_Ave)

## [1] 4.89108

sd(Distance\_from\_Meridian\_St)

## [1] 5.84487

#Next I analyzed the standard deviation of each location to determine how dispersed the event locations are from the true mean distance of the events. A lower standard deviation is what we are looking for because this means the event locations are all about the same distance away (clustered around the mean). In order from smallest to largest: Winthrop Ave (1.366), Cedar St (2.231), Visionary Way (3.933), Porto Alegre (4.800), Mass Ave (4.891), and Meridian St (5.845). This shows us that yes Winthrop Ave, Cedar St, and Visionary way all have a significantly larger mean that Mass Ave, they are more centralized around that mean. Mass Ave has a slightly larger standard deviation than Porto Alegre, but not enough to say Porto Alegre is the better option.   
  
#With all of this data in mind, it is clear to see that the Mass Ave location for housing is the realistic option in terms of distance in miles from all of the events happening during the summer. The Mass Ave location has a mean distance of 2.55 miles from all event locations, the standard deviation is only slightly higher than the next location. This makes the housing location super convenient and accessible.