Assignment2

1. Write a program to find the outliers using Box and Whiskerâs criterion discussed in the class.

find\_outlier<-function(v){  
 q1<-quantile(v,.25)  
 q3<-quantile(v,.75)   
 IQR<-q3-q1  
 Upper\_W<-q3+(1.5\*IQR)  
 Lower\_W<-q1-(1.5\*IQR)  
 return(v[v>Upper\_W | v<Lower\_W])  
}  
  
find\_outlier(c(-60,-3,2,3,4,5,6,11,50,100))

## [1] -60 50 100

2.Make two vectors:

X<-c(1,2,5,10,12)  
 Y<-c(2,5,1,0,12)

a.Find the values that are contained in both X and Y

intersect(X,Y)

## [1] 1 2 5 12

b.Find values that are in x but not in y and (vice versa)

#Values in X but not in Y  
setdiff(X,Y)

## [1] 10

#Values in Y but not in X  
setdiff(Y,X)

## [1] 0

1. Construct a vector that contains all values contained in either X or Y, and compare this vector to

c(X,Y)

## [1] 1 2 5 10 12 2 5 1 0 12

union(X,Y)

## [1] 1 2 5 10 12 0

c(X,Y) contains all values in both X and Y even if they are duplicate but union(X,Y) removes duplicates

Load USArrests data set.

data(USArrests)

3.Which states has most and least assault, (murder, and rape arrests.

#State with most assualts  
most\_assualt<-USArrests[USArrests$Assault==max(USArrests$Assault),]  
rownames(most\_assualt)

## [1] "North Carolina"

#State with least assualts  
least\_assualt<-USArrests[USArrests$Assault==min(USArrests$Assault),]  
rownames(least\_assualt)

## [1] "North Dakota"

#State with most murders  
most\_murder<-USArrests[USArrests$Murder==max(USArrests$Murder),]  
rownames(most\_murder)

## [1] "Georgia"

#State with least murders  
least\_murder<-USArrests[USArrests$Murder==min(USArrests$Murder),]  
rownames(least\_murder)

## [1] "North Dakota"

#State with most rapes  
most\_rape<-USArrests[USArrests$Rape==max(USArrests$Rape),]  
rownames(most\_rape)

## [1] "Nevada"

#State with least rapes  
least\_rape<-USArrests[USArrests$Rape==min(USArrests$Rape),]  
rownames(least\_rape)

## [1] "North Dakota"

1. Which states are in the bottom 25% of murder and in the top 25% of the murder.

#Bottom 25%  
q25<-quantile(USArrests$Murder,0.25)  
subset(USArrests, Murder < q25, select= Murder)

## Murder  
## Connecticut 3.3  
## Idaho 2.6  
## Iowa 2.2  
## Maine 2.1  
## Minnesota 2.7  
## New Hampshire 2.1  
## North Dakota 0.8  
## Rhode Island 3.4  
## South Dakota 3.8  
## Utah 3.2  
## Vermont 2.2  
## Washington 4.0  
## Wisconsin 2.6

#Top 25%  
q75<-quantile(USArrests$Murder,0.75)  
subset(USArrests, Murder > q75, select= Murder)

## Murder  
## Alabama 13.2  
## Florida 15.4  
## Georgia 17.4  
## Louisiana 15.4  
## Maryland 11.3  
## Michigan 12.1  
## Mississippi 16.1  
## Nevada 12.2  
## New Mexico 11.4  
## North Carolina 13.0  
## South Carolina 14.4  
## Tennessee 13.2  
## Texas 12.7

1. The following function calculates the mean and standard deviation of a numeric vector Fn1<-function(x){ Mean1<- mena(x) Sd1<-sd(x) return(mean = mean1, sd = sd1) } Modify the function so that:
2. the default is to use rnorm to generate 30 random normal numbers, and return the standard deviation,

Fn1<-function(x=rnorm(30)){  
 Sd1<-sd(x,na.rm=TRUE)  
 return(sd = Sd1)  
}  
  
Fn1()

## [1] 1.064058

Fn1(c(1,2,3,4))

## [1] 1.290994

1. if there are missing values, the mean and standard deviation are calculated for the remaining values.

Fn1<-function(x){  
 Mean1<- mean(x,na.rm=TRUE)  
 Sd1<-sd(x,na.rm=TRUE)  
 return(c(mean = Mean1, sd = Sd1))  
}  
  
Fn1(c(1,2,3,4,5,6))

## mean sd   
## 3.500000 1.870829