369 1.2

NGonyo

9/6/2021

#set the working directory  
setwd("~/Documents/MA369/Datasets")  
dir()

## [1] "Table1-2.DAT"

exerciseData=read.table("Table1-2.DAT", sep=" ",header=F)  
exerciseData

## V1 V2 V3 V4 V5 V6 V7  
## 1 NA NA 0.801 NA 121.41 NA 70.42  
## 2 NA NA 0.824 NA 127.70 NA 72.47  
## 3 NA NA 0.841 NA 129.20 NA 78.20  
## 4 NA NA 0.816 NA 131.80 NA 74.89  
## 5 NA NA 0.840 NA 135.10 NA 71.21  
## 6 NA NA 0.842 NA 131.50 NA 78.39  
## 7 NA NA 0.820 NA 126.70 NA 69.02  
## 8 NA NA 0.802 NA 115.10 NA 73.10  
## 9 NA NA 0.828 NA 130.80 NA 79.28  
## 10 NA NA 0.819 NA 124.60 NA 76.48  
## 11 NA NA 0.826 NA 118.31 NA 70.25  
## 12 NA NA 0.802 NA 114.20 NA 72.88  
## 13 NA NA 0.810 NA 120.30 NA 68.23  
## 14 NA NA 0.802 NA 115.70 NA 68.12  
## 15 NA NA 0.832 NA 117.51 NA 71.62  
## 16 NA NA 0.796 NA 109.81 NA 53.10  
## 17 NA NA 0.759 NA 109.10 NA 50.85  
## 18 NA NA 0.770 NA 115.10 NA 51.68  
## 19 NA NA 0.759 NA 118.31 NA 50.60  
## 20 NA NA 0.772 NA 112.60 NA 53.51  
## 21 NA NA 0.806 NA 116.20 NA 56.53  
## 22 NA NA 0.803 NA 118.00 NA 70.70  
## 23 NA NA 0.845 NA 131.00 NA 74.35  
## 24 NA NA 0.822 NA 125.70 NA 68.29  
## 25 NA NA 0.971 NA 126.10 NA 72.10  
## 26 NA NA 0.816 NA 125.80 NA 70.64  
## 27 NA NA 0.836 NA 125.50 NA 76.33  
## 28 NA NA 0.815 NA 127.80 NA 76.75  
## 29 NA NA 0.822 NA 130.50 NA 80.33  
## 30 NA NA 0.822 NA 127.90 NA 75.68  
## 31 NA NA 0.843 NA 123.90 NA 78.54  
## 32 NA NA 0.824 NA 124.10 NA 71.91  
## 33 NA NA 0.788 NA 120.80 NA 68.22  
## 34 NA NA 0.782 NA 107.40 NA 54.42  
## 35 NA NA 0.795 NA 120.70 NA 70.41  
## 36 NA NA 0.805 NA 121.91 NA 73.68  
## 37 NA NA 0.836 NA 122.31 NA 74.93  
## 38 NA NA 0.788 NA 110.60 NA 53.52  
## 39 NA NA 0.772 NA 103.51 NA 48.93  
## 40 NA NA 0.776 NA 110.71 NA 53.67  
## 41 NA NA 0.758 NA 113.80 NA 52.42

#Display how many rows and columns are in the dataset, ie the "dimensions"  
dim(exerciseData)

## [1] 41 7

#Check missing values  
is.na(exerciseData)

## V1 V2 V3 V4 V5 V6 V7  
## [1,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [2,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [3,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [4,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [5,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [6,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [7,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [8,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [9,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [10,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [11,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [12,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [13,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [14,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [15,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [16,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [17,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [18,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [19,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [20,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [21,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [22,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [23,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [24,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [25,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [26,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [27,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [28,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [29,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [30,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [31,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [32,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [33,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [34,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [35,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [36,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [37,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [38,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [39,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [40,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE  
## [41,] TRUE TRUE FALSE TRUE FALSE TRUE FALSE

#We can see the 3rd, 5th, and 7th rows are the only ones containing data, rest are spaces  
  
#Lets see how R read the data in the dataset, ie we need to check the "structure" of the dataset  
str(exerciseData)

## 'data.frame': 41 obs. of 7 variables:  
## $ V1: logi NA NA NA NA NA NA ...  
## $ V2: logi NA NA NA NA NA NA ...  
## $ V3: num 0.801 0.824 0.841 0.816 0.84 0.842 0.82 0.802 0.828 0.819 ...  
## $ V4: logi NA NA NA NA NA NA ...  
## $ V5: num 121 128 129 132 135 ...  
## $ V6: logi NA NA NA NA NA NA ...  
## $ V7: num 70.4 72.5 78.2 74.9 71.2 ...

Density=exerciseData[1:40, 3]  
Machinedirection=exerciseData[1:40, 5]  
Crossdirection=exerciseData[1:40, 7]  
  
#Check to see each variable was assigned correctly   
Density

## [1] 0.801 0.824 0.841 0.816 0.840 0.842 0.820 0.802 0.828 0.819 0.826 0.802  
## [13] 0.810 0.802 0.832 0.796 0.759 0.770 0.759 0.772 0.806 0.803 0.845 0.822  
## [25] 0.971 0.816 0.836 0.815 0.822 0.822 0.843 0.824 0.788 0.782 0.795 0.805  
## [37] 0.836 0.788 0.772 0.776

Machinedirection

## [1] 121.41 127.70 129.20 131.80 135.10 131.50 126.70 115.10 130.80 124.60  
## [11] 118.31 114.20 120.30 115.70 117.51 109.81 109.10 115.10 118.31 112.60  
## [21] 116.20 118.00 131.00 125.70 126.10 125.80 125.50 127.80 130.50 127.90  
## [31] 123.90 124.10 120.80 107.40 120.70 121.91 122.31 110.60 103.51 110.71

Crossdirection

## [1] 70.42 72.47 78.20 74.89 71.21 78.39 69.02 73.10 79.28 76.48 70.25 72.88  
## [13] 68.23 68.12 71.62 53.10 50.85 51.68 50.60 53.51 56.53 70.70 74.35 68.29  
## [25] 72.10 70.64 76.33 76.75 80.33 75.68 78.54 71.91 68.22 54.42 70.41 73.68  
## [37] 74.93 53.52 48.93 53.67

#Sort data by Density variable  
exerciseData[order(Density),]

## V1 V2 V3 V4 V5 V6 V7  
## 17 NA NA 0.759 NA 109.10 NA 50.85  
## 19 NA NA 0.759 NA 118.31 NA 50.60  
## 18 NA NA 0.770 NA 115.10 NA 51.68  
## 20 NA NA 0.772 NA 112.60 NA 53.51  
## 39 NA NA 0.772 NA 103.51 NA 48.93  
## 40 NA NA 0.776 NA 110.71 NA 53.67  
## 34 NA NA 0.782 NA 107.40 NA 54.42  
## 33 NA NA 0.788 NA 120.80 NA 68.22  
## 38 NA NA 0.788 NA 110.60 NA 53.52  
## 35 NA NA 0.795 NA 120.70 NA 70.41  
## 16 NA NA 0.796 NA 109.81 NA 53.10  
## 1 NA NA 0.801 NA 121.41 NA 70.42  
## 8 NA NA 0.802 NA 115.10 NA 73.10  
## 12 NA NA 0.802 NA 114.20 NA 72.88  
## 14 NA NA 0.802 NA 115.70 NA 68.12  
## 22 NA NA 0.803 NA 118.00 NA 70.70  
## 36 NA NA 0.805 NA 121.91 NA 73.68  
## 21 NA NA 0.806 NA 116.20 NA 56.53  
## 13 NA NA 0.810 NA 120.30 NA 68.23  
## 28 NA NA 0.815 NA 127.80 NA 76.75  
## 4 NA NA 0.816 NA 131.80 NA 74.89  
## 26 NA NA 0.816 NA 125.80 NA 70.64  
## 10 NA NA 0.819 NA 124.60 NA 76.48  
## 7 NA NA 0.820 NA 126.70 NA 69.02  
## 24 NA NA 0.822 NA 125.70 NA 68.29  
## 29 NA NA 0.822 NA 130.50 NA 80.33  
## 30 NA NA 0.822 NA 127.90 NA 75.68  
## 2 NA NA 0.824 NA 127.70 NA 72.47  
## 32 NA NA 0.824 NA 124.10 NA 71.91  
## 11 NA NA 0.826 NA 118.31 NA 70.25  
## 9 NA NA 0.828 NA 130.80 NA 79.28  
## 15 NA NA 0.832 NA 117.51 NA 71.62  
## 27 NA NA 0.836 NA 125.50 NA 76.33  
## 37 NA NA 0.836 NA 122.31 NA 74.93  
## 5 NA NA 0.840 NA 135.10 NA 71.21  
## 3 NA NA 0.841 NA 129.20 NA 78.20  
## 6 NA NA 0.842 NA 131.50 NA 78.39  
## 31 NA NA 0.843 NA 123.90 NA 78.54  
## 23 NA NA 0.845 NA 131.00 NA 74.35  
## 25 NA NA 0.971 NA 126.10 NA 72.10

#Lets get a better idea of the numerical properties of each variable  
summary(Density)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.7590 0.7957 0.8155 0.8132 0.8265 0.9710

mean(Density)

## [1] 0.8132

sd(Density)

## [1] 0.03493965

summary(Machinedirection)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 103.5 115.5 121.7 121.1 127.0 135.1

mean(Machinedirection)

## [1] 121.1322

sd(Machinedirection)

## [1] 7.71345

summary(Crossdirection)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 48.93 65.22 70.95 68.11 74.90 80.33

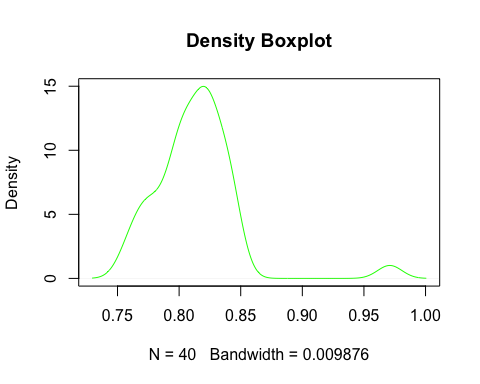
mean(Crossdirection)

## [1] 68.10575

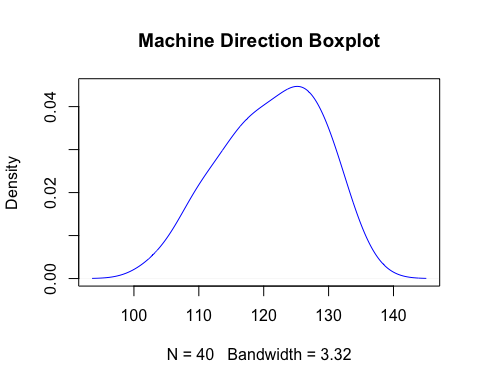
sd(Crossdirection)

## [1] 9.59998

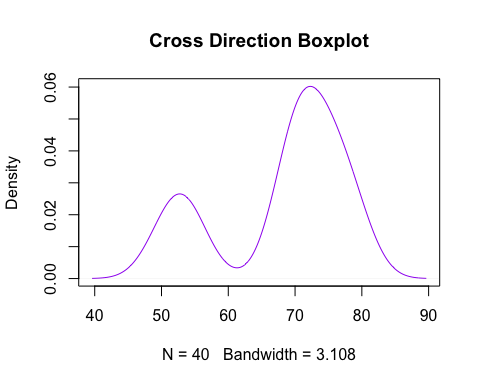
#Create density plot for each variable  
plot(density(Density), col="green", main="Density Boxplot")



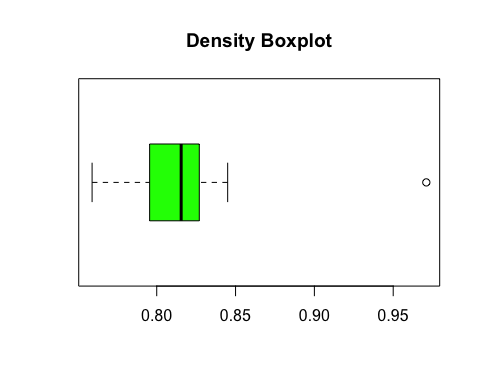
plot(density(Machinedirection), col="blue", main="Machine Direction Boxplot")



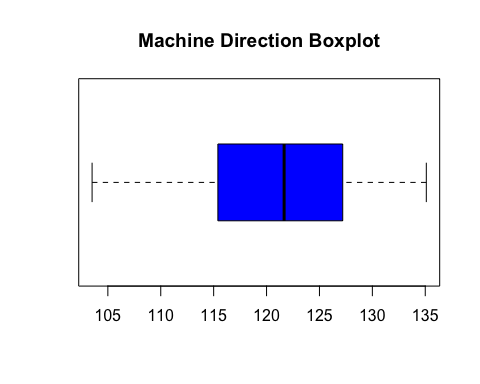
plot(density(Crossdirection), col="purple", main="Cross Direction Boxplot")



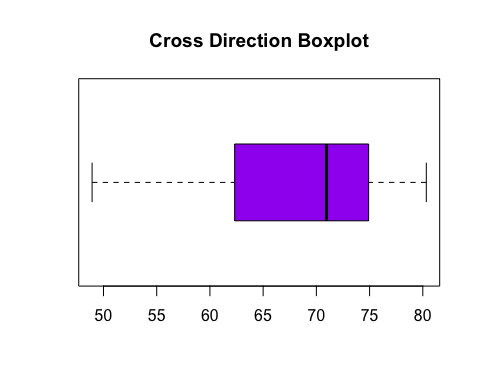
boxplot(Density, horizontal = T, col="green", main="Density Boxplot")



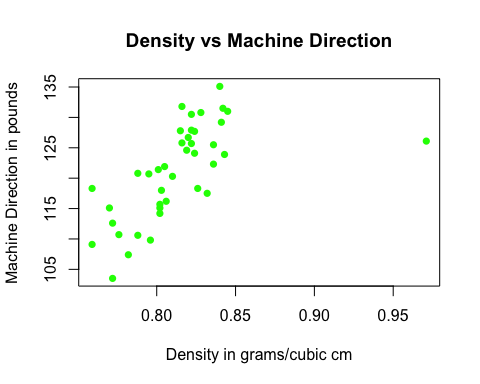
boxplot(Machinedirection, horizontal = T, col="blue", main="Machine Direction Boxplot")



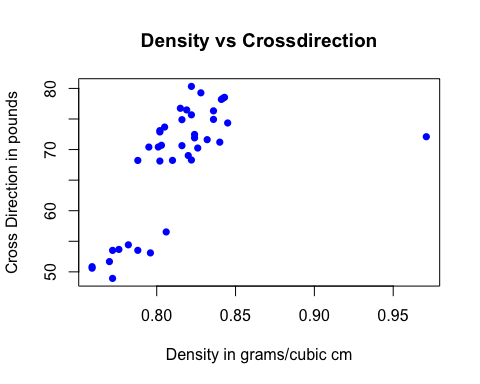
boxplot(Crossdirection, horizontal = T, col="purple", main="Cross Direction Boxplot")



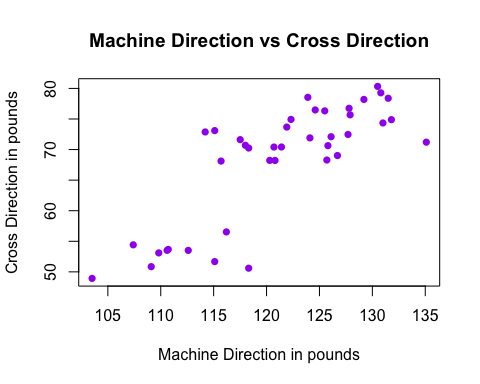
#We can see there is an outlier present in the Density variable, shown as the dot on the right side of the boxplot.  
  
plot(Density, Machinedirection, pch=16, col="green", xlab="Density in grams/cubic cm", ylab="Machine Direction in pounds", main="Density vs Machine Direction")



plot(Density, Crossdirection, pch=16, col="blue", xlab="Density in grams/cubic cm", ylab="Cross Direction in pounds", main="Density vs Crossdirection")



plot(Machinedirection, Crossdirection, pch=16, col="purple", xlab="Machine Direction in pounds", ylab="Cross Direction in pounds", main="Machine Direction vs Cross Direction")



#In these first 2 plots we can clearly see the outlier on the right side of the plots. Ignoring this outlier, it seems that each variable is positively correlated to each other, ie as one variable increases as do the other two. This relationship is most clear when graphing density vs crossdirection, as well as density vs machinedirection