PRODIGY\_DS\_01.R

To create a bar chart or histogram for a categorical or continuous variable

#IMPORTING THE DATASET  
library(readxl)  
flavors\_of\_cacao <- read\_excel("flavors\_of\_cacao.xlsx")  
Flavors\_of\_cacao<-data.frame(flavors\_of\_cacao)  
#CHECKING THE DIMENSIONS  
dim(Flavors\_of\_cacao) #The data contains 1793 rows and 8 columns.

## [1] 1793 8

head(Flavors\_of\_cacao,2)

## Company Review\_Date Cocoa\_Percent Company\_Location Rating Bean\_Type  
## 1 A. Morin 2016 0.63 France 3.75 Â   
## 2 A. Morin 2015 0.70 France 2.75 Â   
## Broad\_Bean\_Origin Specific\_Bean\_Origin  
## 1 Sao Tome Agua Grande  
## 2 Togo Kpime

# CHECKING IF THERE ARE ANY MISSING VALUES IN THE DATAFRAME  
sum(is.na(Flavors\_of\_cacao))

## [1] 0

#There are no missing values  
#CHECKING THE STRUSTURE OF THE DATAFRAME  
str(Flavors\_of\_cacao)

## 'data.frame': 1793 obs. of 8 variables:  
## $ Company : chr "A. Morin" "A. Morin" "A. Morin" "A. Morin" ...  
## $ Review\_Date : num 2016 2015 2015 2015 2015 ...  
## $ Cocoa\_Percent : num 0.63 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 ...  
## $ Company\_Location : chr "France" "France" "France" "France" ...  
## $ Rating : num 3.75 2.75 3 3.5 3.5 2.75 3.5 3.5 3.75 4 ...  
## $ Bean\_Type : chr "Â " "Â " "Â " "Â " ...  
## $ Broad\_Bean\_Origin : chr "Sao Tome" "Togo" "Togo" "Togo" ...  
## $ Specific\_Bean\_Origin: chr "Agua Grande" "Kpime" "Atsane" "Akata" ...

#CHECKING THE SUMMARIES   
summary(Flavors\_of\_cacao$Cocoa\_Percent)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.420 0.700 0.700 0.717 0.750 1.000

table(Flavors\_of\_cacao$Rating)

##   
## 1 1.5 1.75 2 2.25 2.5 2.75 3 3.25 3.5 3.75 4 5   
## 4 10 3 32 14 126 259 341 303 392 210 97 2

table(Flavors\_of\_cacao$Review\_Date)

##   
## 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017   
## 72 77 93 123 111 164 194 184 247 285 219 24

table(Flavors\_of\_cacao$Bean\_Type)

##   
## Â  Amazon Amazon mix   
## 887 1 2   
## Amazon, ICS Beniano Blend   
## 2 3 41   
## Blend-Forastero,Criollo CCN51 Criollo   
## 1 1 153   
## Criollo (Amarru) Criollo (Ocumare 61) Criollo (Ocumare 67)   
## 2 2 1   
## Criollo (Ocumare 77) Criollo (Ocumare) Criollo (Porcelana)   
## 1 1 10   
## Criollo (Wild) Criollo, + Criollo, Forastero   
## 1 1 2   
## Criollo, Trinitario EET Forastero   
## 39 3 87   
## Forastero (Amelonado) Forastero (Arriba) Forastero (Arriba) ASS   
## 1 37 6   
## Forastero (Arriba) ASSS Forastero (Catongo) Forastero (Nacional)   
## 1 2 52   
## Forastero (Parazinho) Forastero(Arriba, CCN) Forastero, Trinitario   
## 8 1 1   
## Matina Nacional Nacional (Arriba)   
## 3 2 3   
## Trinitario Trinitario (85% Criollo) Trinitario (Amelonado)   
## 418 2 1   
## Trinitario (Scavina) Trinitario, Criollo Trinitario, Forastero   
## 1 9 2   
## Trinitario, Nacional Trinitario, TCGA   
## 1 1

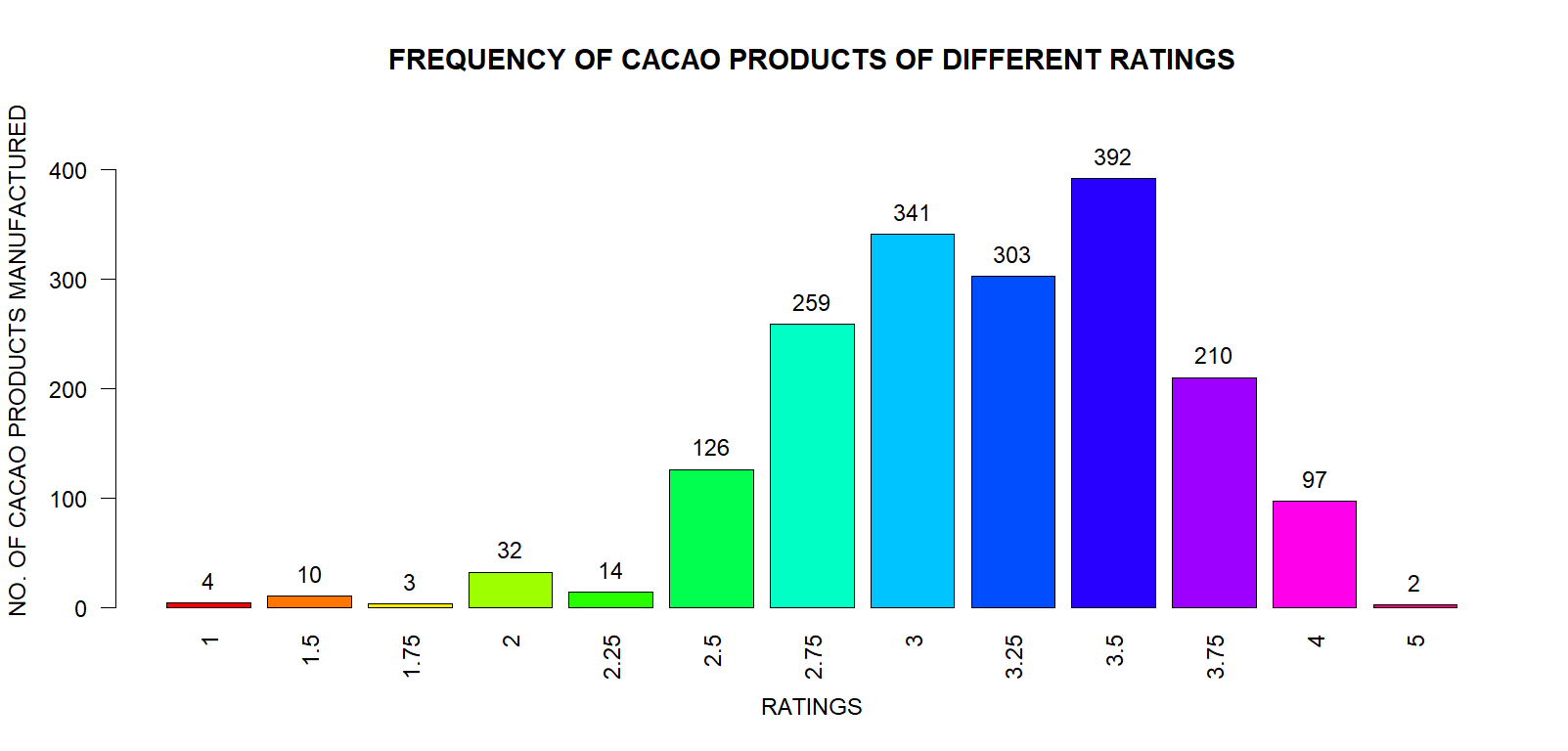
table(Flavors\_of\_cacao$Company\_Location)

##   
## Amsterdam Argentina Australia Austria   
## 4 9 49 26   
## Belgium Bolivia Brazil Canada   
## 40 2 17 124   
## Chile Colombia Costa Rica Czech Republic   
## 2 23 9 1   
## Denmark Domincan Republic Ecuador Eucador   
## 15 5 54 1   
## Fiji Finland France Germany   
## 4 2 156 35   
## Ghana Grenada Guatemala Honduras   
## 1 3 10 6   
## Hungary Iceland India Ireland   
## 22 3 1 4   
## Israel Italy Japan Lithuania   
## 9 63 17 6   
## Madagascar Martinique Mexico Netherlands   
## 17 1 4 4   
## New Zealand Niacragua Nicaragua Peru   
## 17 1 5 17   
## Philippines Poland Portugal Puerto Rico   
## 1 8 3 4   
## Russia Sao Tome Scotland Singapore   
## 1 4 10 3   
## South Africa South Korea Spain St. Lucia   
## 3 5 25 2   
## Suriname Sweden Switzerland U.K.   
## 1 5 38 96   
## U.S.A. Venezuela Vietnam Wales   
## 763 20 11 1

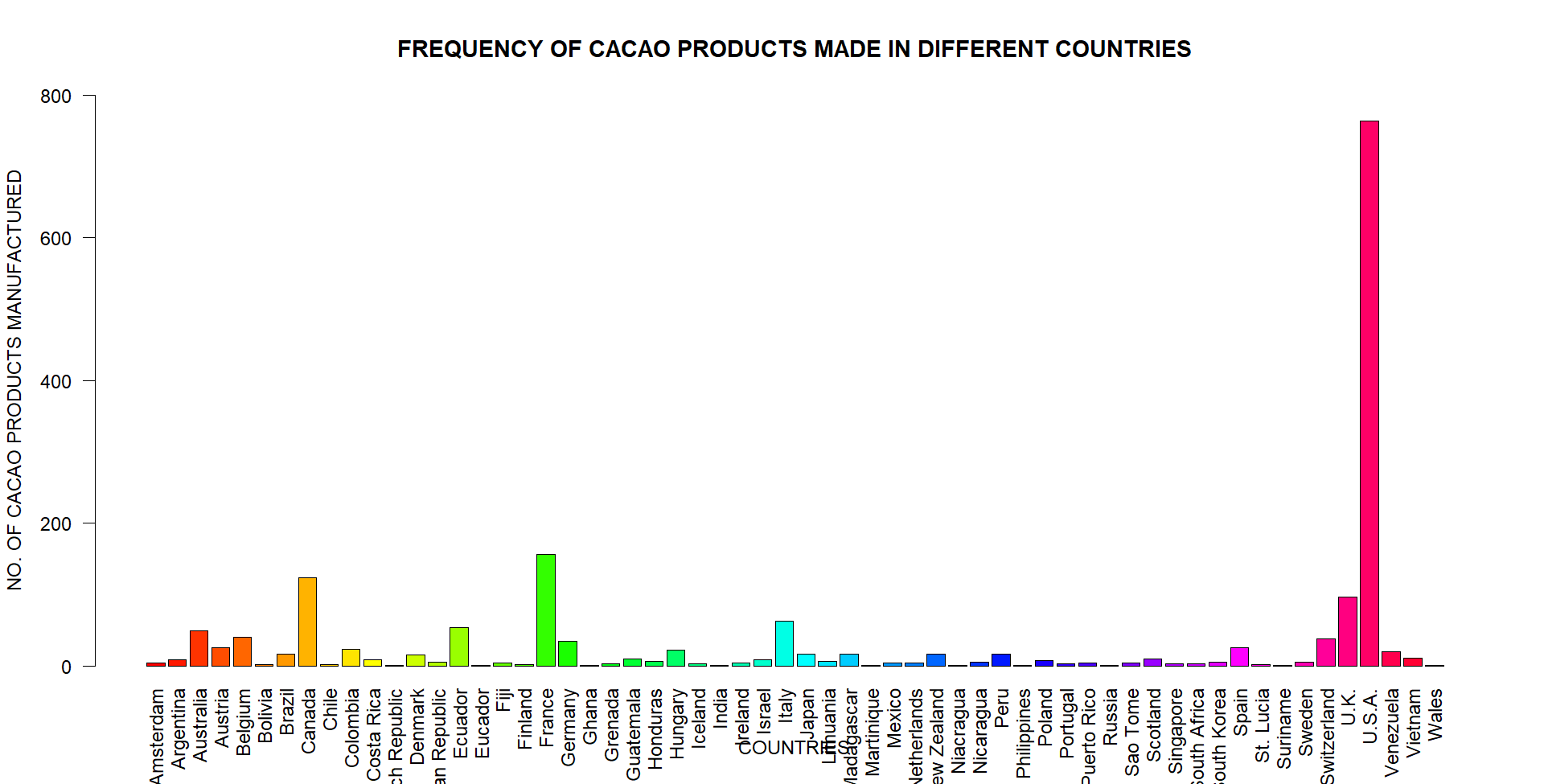
table(Flavors\_of\_cacao$Broad\_Bean\_Origin)

##   
## Â  Africa, Carribean, C. Am.   
## 73 1   
## Australia Belize   
## 3 49   
## Bolivia Brazil   
## 57 58   
## Burma Cameroon   
## 1 1   
## Carribean Carribean(DR/Jam/Tri)   
## 8 1   
## Central and S. America Colombia   
## 4 40   
## Colombia, Ecuador Congo   
## 3 10   
## Cost Rica, Ven Costa Rica   
## 1 38   
## Cuba Dom. Rep., Madagascar   
## 11 1   
## Domincan Republic Dominican Rep., Bali   
## 25 1   
## Dominican Republic DR, Ecuador, Peru   
## 141 1   
## Ecuador Ecuador, Costa Rica   
## 193 1   
## Ecuador, Mad., PNG El Salvador   
## 1 2   
## Fiji Gabon   
## 8 1   
## Ghana Ghana & Madagascar   
## 33 1   
## Ghana, Domin. Rep Ghana, Panama, Ecuador   
## 2 1   
## Gre., PNG, Haw., Haiti, Mad Grenada   
## 1 19   
## Guat., D.R., Peru, Mad., PNG Guatemala   
## 1 28   
## Haiti Hawaii   
## 9 28   
## Honduras India   
## 15 4   
## Indonesia Indonesia, Ghana   
## 16 1   
## Ivory Coast Jamaica   
## 5 20   
## Liberia Mad., Java, PNG   
## 3 1   
## Madagascar Madagascar & Ecuador   
## 145 1   
## Malaysia Martinique   
## 3 1   
## Mexico Nicaragua   
## 30 60   
## Nigeria Panama   
## 1 7   
## Papua New Guinea Peru   
## 42 165   
## Peru(SMartin,Pangoa,nacional) Peru, Belize   
## 1 1   
## Peru, Dom. Rep Peru, Ecuador   
## 1 1   
## Peru, Ecuador, Venezuela Peru, Mad., Dom. Rep.   
## 1 1   
## Peru, Madagascar Philippines   
## 1 5   
## PNG, Vanuatu, Mad Principe   
## 1 1   
## Puerto Rico Samoa   
## 4 2   
## Sao Tome Sao Tome & Principe   
## 10 7   
## Solomon Islands South America   
## 4 3   
## South America, Africa Sri Lanka   
## 1 2   
## St. Lucia Suriname   
## 8 1   
## Tanzania Tobago   
## 34 2   
## Togo Trinidad   
## 3 33   
## Trinidad-Tobago Trinidad, Ecuador   
## 1 1   
## Trinidad, Tobago Uganda   
## 2 8   
## Vanuatu Ven, Trinidad, Ecuador   
## 7 1   
## Ven., Indonesia, Ecuad. Ven., Trinidad, Mad.   
## 1 1   
## Ven.,Ecu.,Peru,Nic. Venez,Africa,Brasil,Peru,Mex   
## 1 1   
## Venezuela Venezuela, Carribean   
## 214 1   
## Venezuela, Dom. Rep. Venezuela, Ghana   
## 1 1   
## Venezuela, Java Venezuela, Trinidad   
## 1 1   
## Venezuela/ Ghana Vietnam   
## 1 38   
## West Africa   
## 6

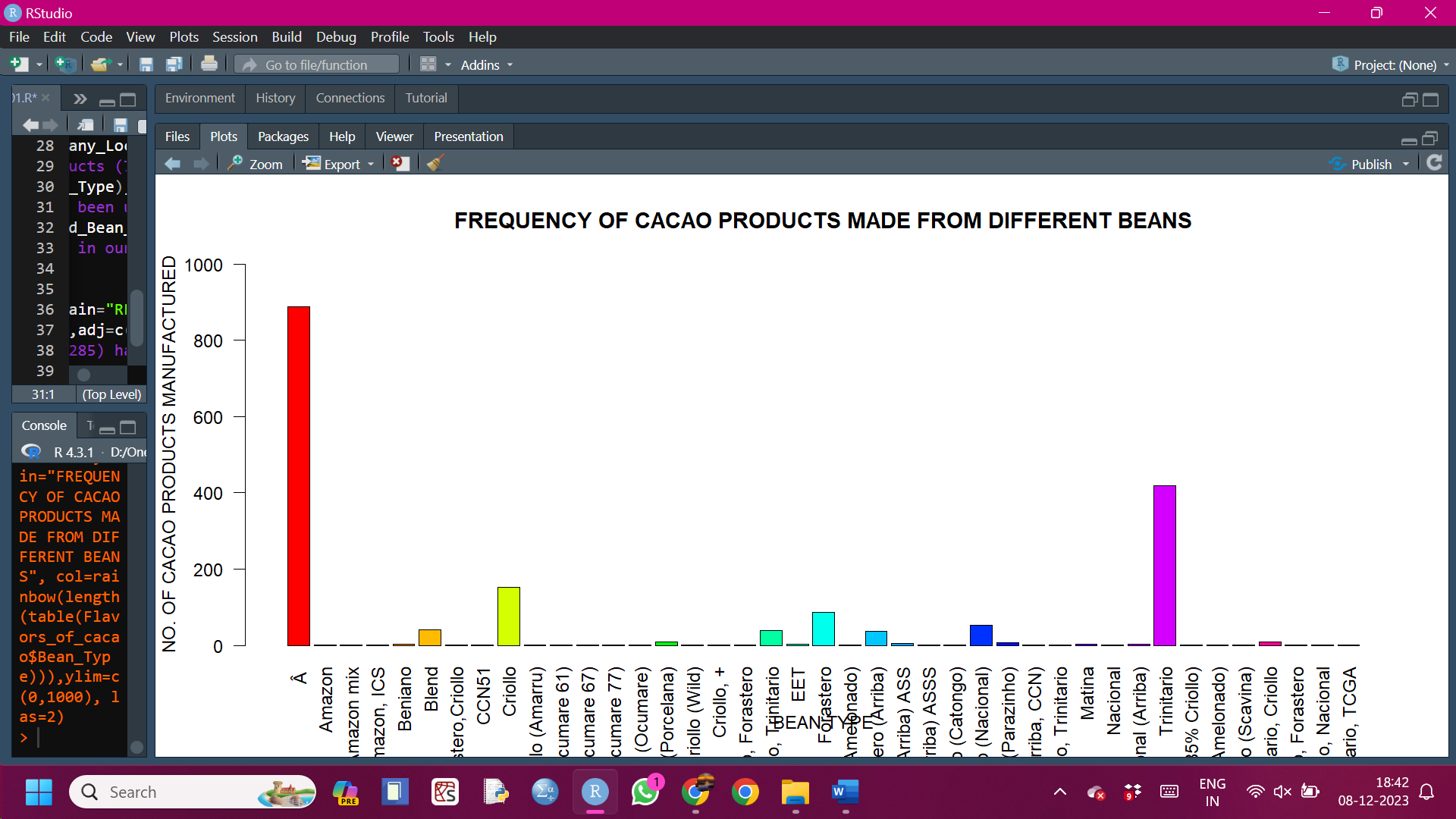
#With the summary statistic used above, we can see the max, min and average of Cocoa\_Percent and Rating as well as the number of chocolates in categories of Bean\_Type,Company\_Location,Broad\_Bean\_Origin and Review\_Date  
  
#ANALYZING THE DATA USING GRAPHICAL METHODS  
  
#BARPLOTS  
b1<-barplot(table(Flavors\_of\_cacao$Rating),xlab="RATINGS", ylab="NO. OF CACAO PRODUCTS MANUFACTURED",main="FREQUENCY OF CACAO PRODUCTS OF DIFFERENT RATINGS", col=rainbow(length(table(Flavors\_of\_cacao$Rating))),las=2,ylim=c(0,450))  
text(b1, table(Flavors\_of\_cacao$Rating), round(table(Flavors\_of\_cacao$Rating), 1),cex=1,pos=3)



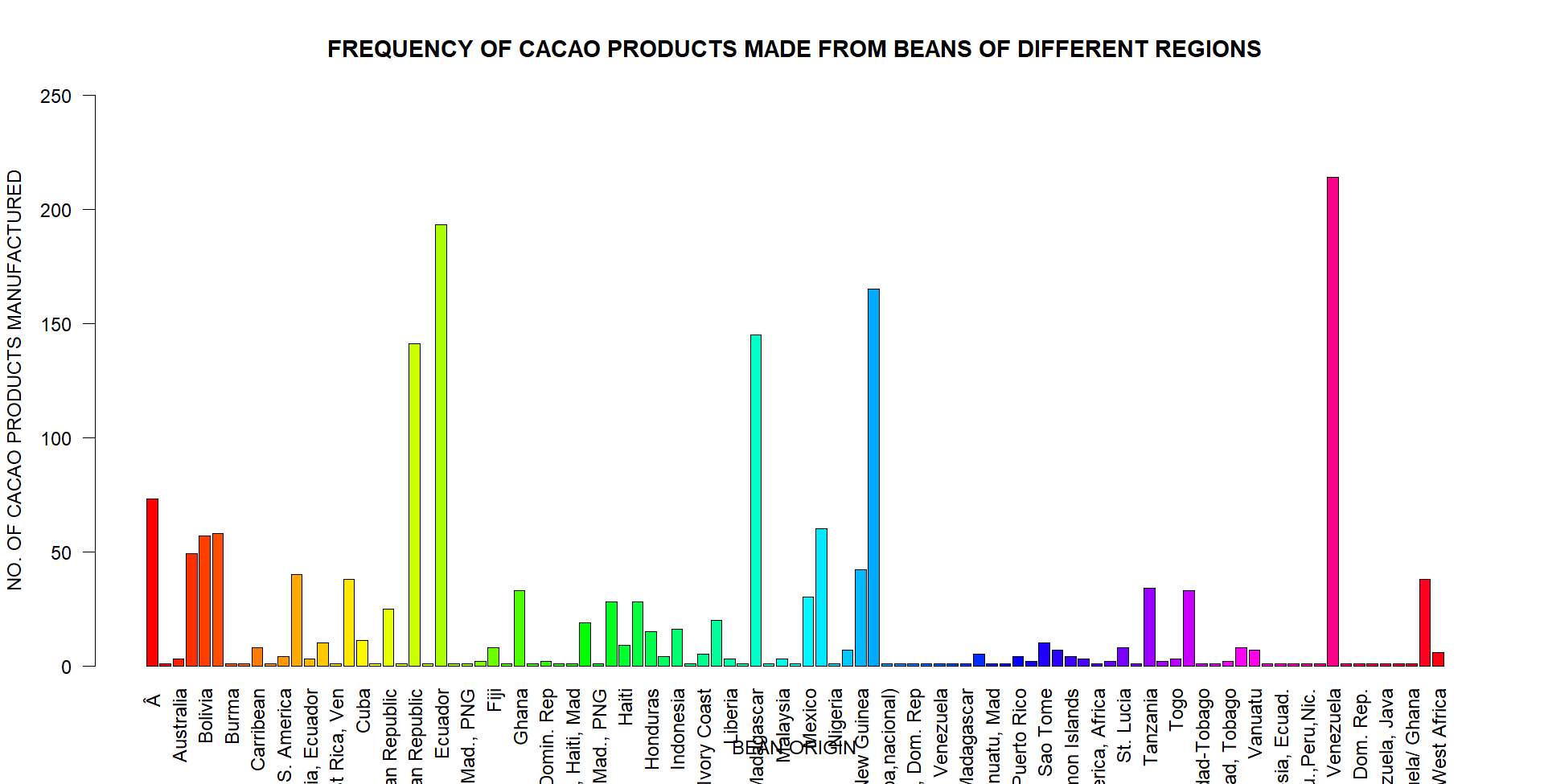
#We see that a majority of cacao products (392) have a rating of 3.5, whereas only 2 products were rated 5.  
b2<-barplot(table(Flavors\_of\_cacao$Company\_Location),xlab="COUNTRIES", ylab="NO. OF CACAO PRODUCTS MANUFACTURED",main="FREQUENCY OF CACAO PRODUCTS MADE IN DIFFERENT COUNTRIES", col=rainbow(length(table(Flavors\_of\_cacao$Company\_Location))),las=2,ylim=c(0,800))



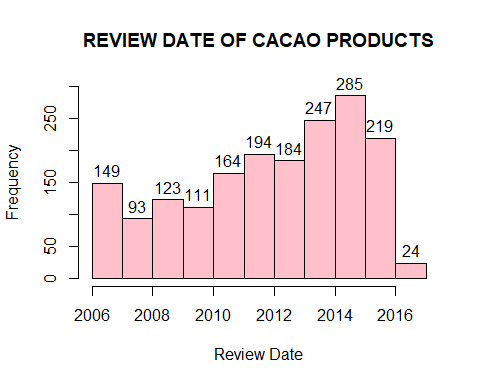
#We observe that majority of cacao products (763) are manufactured in USA  
b3<-barplot(table(Flavors\_of\_cacao$Bean\_Type),xlab="BEAN TYPE", ylab="NO. OF CACAO PRODUCTS MANUFACTURED",main="FREQUENCY OF CACAO PRODUCTS MADE FROM DIFFERENT BEANS", col=rainbow(length(table(Flavors\_of\_cacao$Bean\_Type))),ylim=c(0,1000), las=2)



#We observe that ^A and Trinitario have been used in majority of cacao products.  
b4<-barplot(table(Flavors\_of\_cacao$Broad\_Bean\_Origin),xlab="BEAN ORIGIN", ylab="NO. OF CACAO PRODUCTS MANUFACTURED",main="FREQUENCY OF CACAO PRODUCTS MADE FROM BEANS OF DIFFERENT REGIONS", col=rainbow(length(table(Flavors\_of\_cacao$Broad\_Bean\_Origin))),ylim=c(0,250), las=2)



#We observe that a majority of products in our sample were made from beans obtained from countries of South America and Africa, such as Venezuela, Dominican Republic, Peru and Madagascar, which makes sense as these countries have optimal climate conditions for growing cacao beans.  
  
#HISTOGRAMS  
h1<-hist(Flavors\_of\_cacao$Review\_Date,main="REVIEW DATE OF CACAO PRODUCTS", xlab="Review Date", col="pink", ylim=c(0,300))  
text(h1$mids,h1$counts,labels=h1$counts,adj=c(0.5,-0.5))



#We observe that majority of products (285) have been reviewed in 2014-15 whereas only 24 products have been recently reviewed.