DataExplorer

Jennifer Brosnahan

6/24/2020

Loading packages

```
library(DataExplorer)
```

```
## Warning: package 'DataExplorer' was built under R version 4.0.2
```

Importing data

Checking structure

```
str(choco)
```

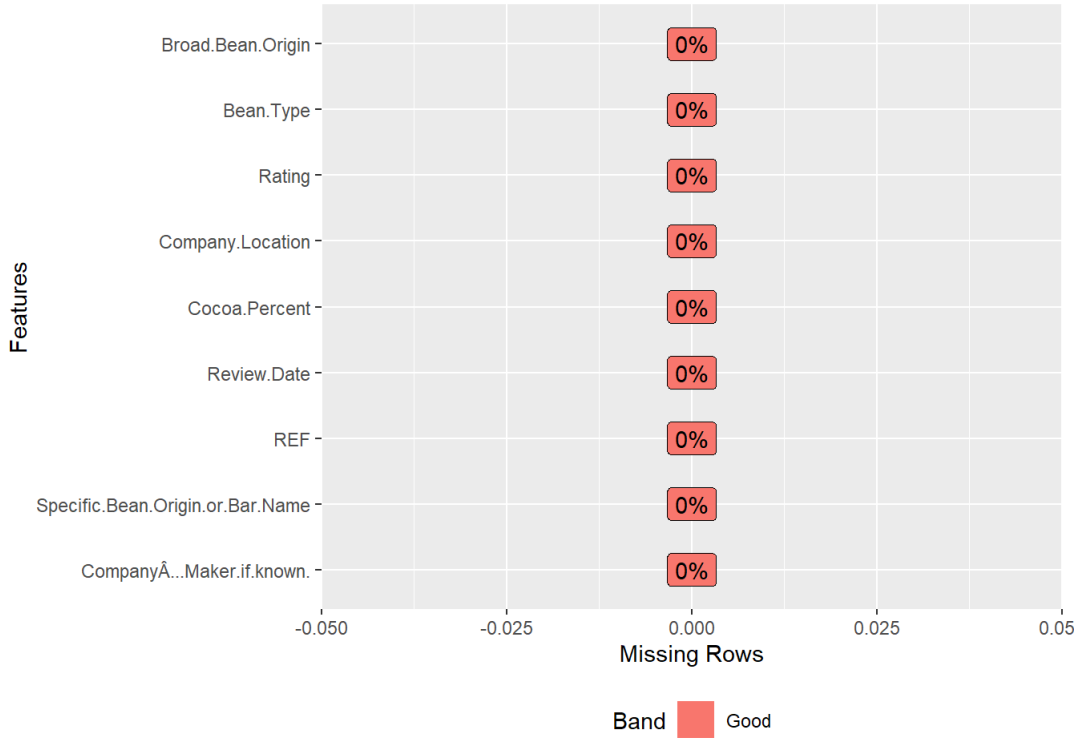
```
## 'data.frame':    1795 obs. of  9 variables:
##  $ CompanyÂ...Maker.if.known.      : chr  "A. Morin" "A. Morin" "A. Morin" "A. Morin" ...
##  $ Specific.Bean.Origin.or.Bar.Name: chr  "Agua Grande" "Kpime" "Atsane" "Akata" ...
##  $ REF                            : int  1876 1676 1676 1680 1704 1315 1315 1315 1319 1319 ...
##  $ Review.Date                    : int  2016 2015 2015 2015 2015 2014 2014 2014 2014 2014 ...
##  $ Cocoa.Percent                  : chr  "63%" "70%" "70%" "70%" ...
##  $ 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" ...
```

Cleaning data

```
choco$Cocoa.Percent = as.numeric(gsub('%','',choco$Cocoa.Percent))
choco$Review.Date = as.character(choco$Review.Date)
```

Checking for NAs

```
plot_missing(choco)
```

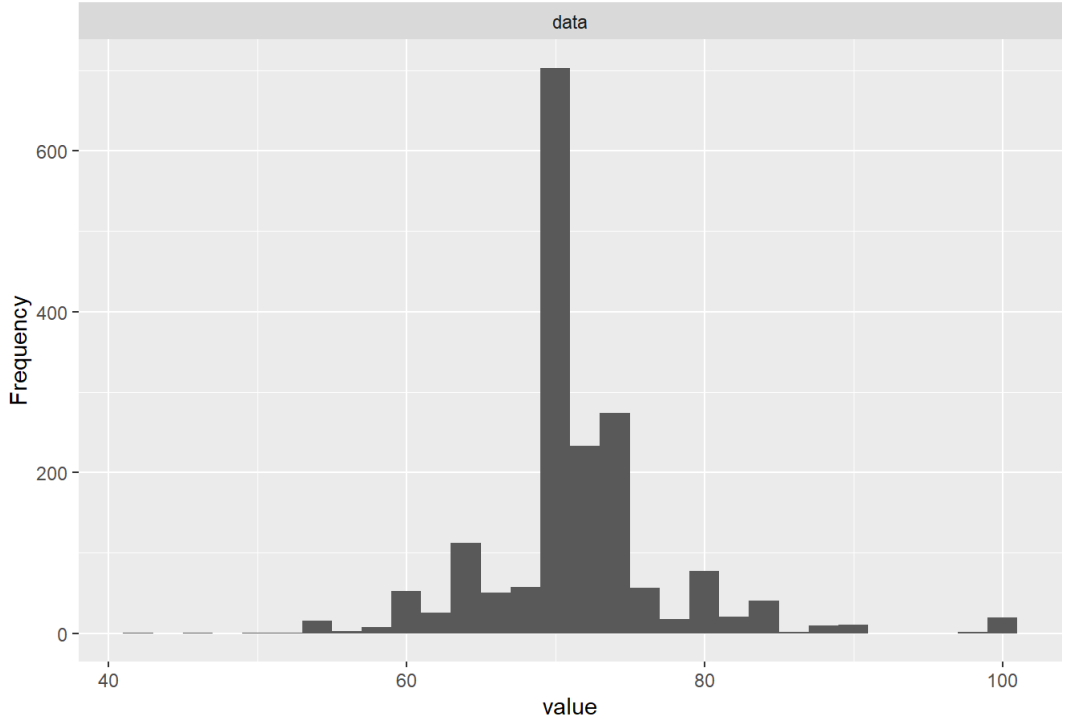


Nice visual of structure

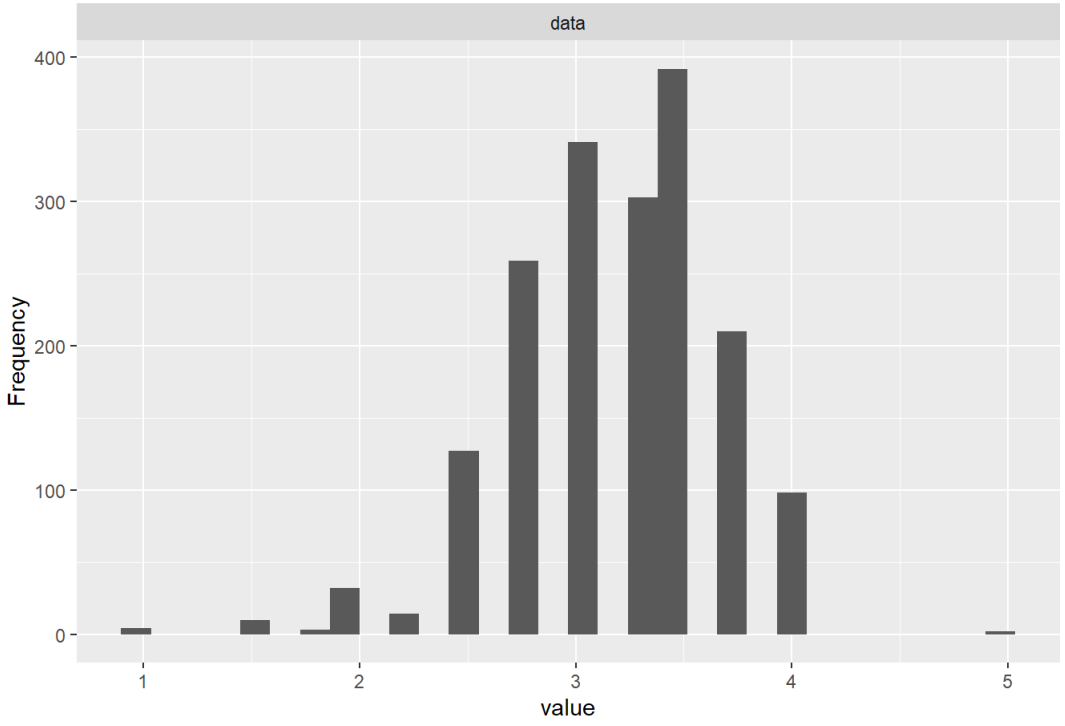
```
plot_str(choco)
```

Univariate analysis using histograms and density plots

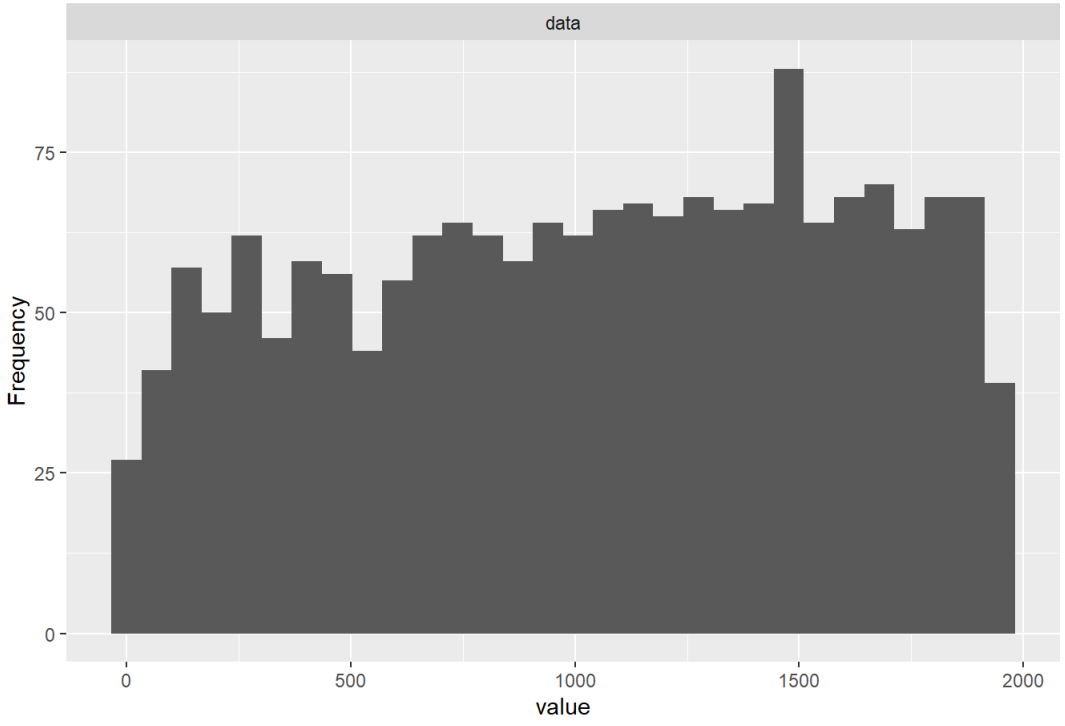
```
plot_histogram(choco$Cocoa.Percent,
               geom_histogram_args = list(bins = 30L))
```



```
plot_histogram(choco$Rating,
               geom_histogram_args = list(bins = 30L))
```

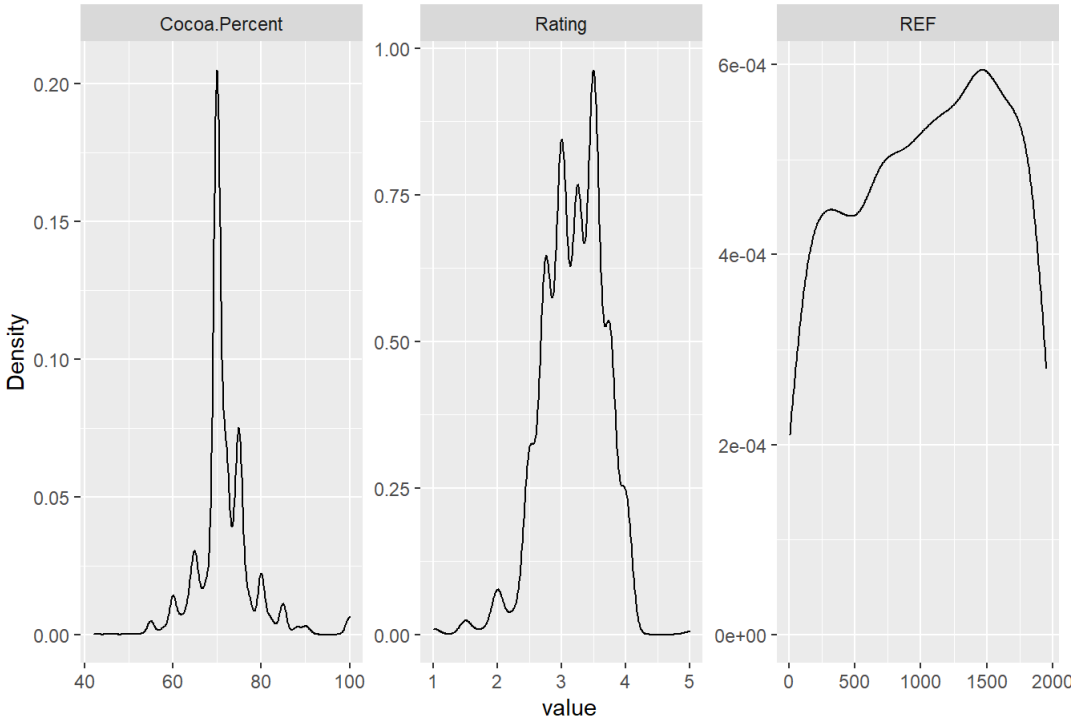



```
plot_histogram(choco$REF,
               geom_histogram_args = list(bins = 30L))
```

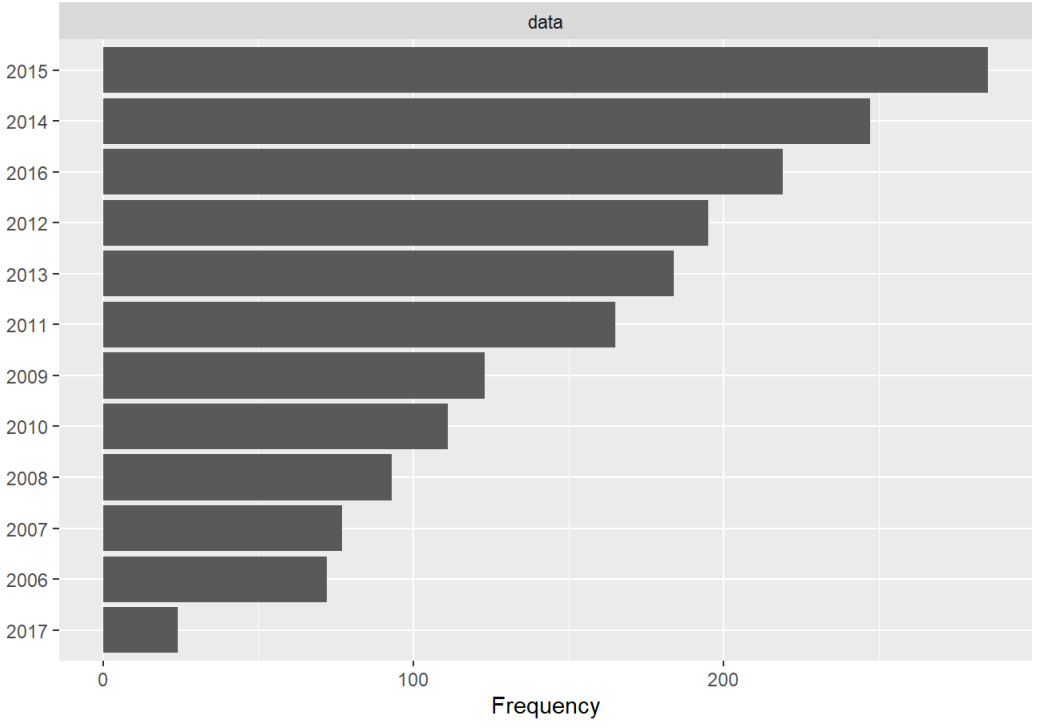


Density plot

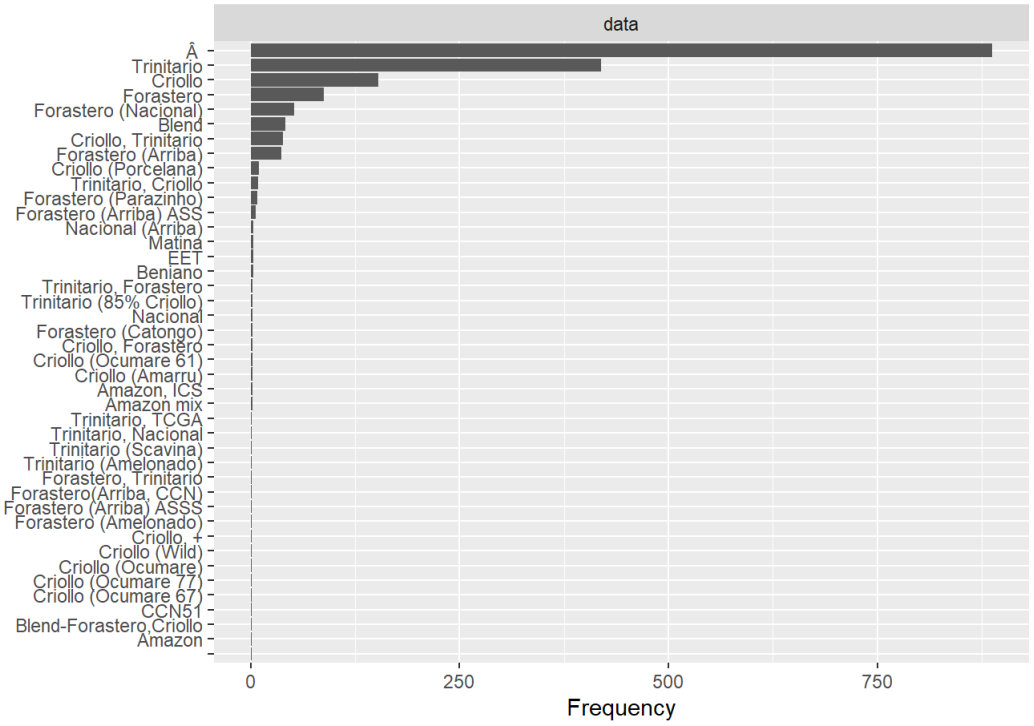
```
plot_density(choco)
```




```
plot_bar(choco$Review.Date)
```




```
plot_bar(choco$Bean.Type)
```



Bivariate and multivariate

analysis of correlation matrix

```
plot_correlation(choco, type = 'continuous','Review.Data')
```