Graphical Exploration of Data in R

Dr. Beatrice

June 17, 2014

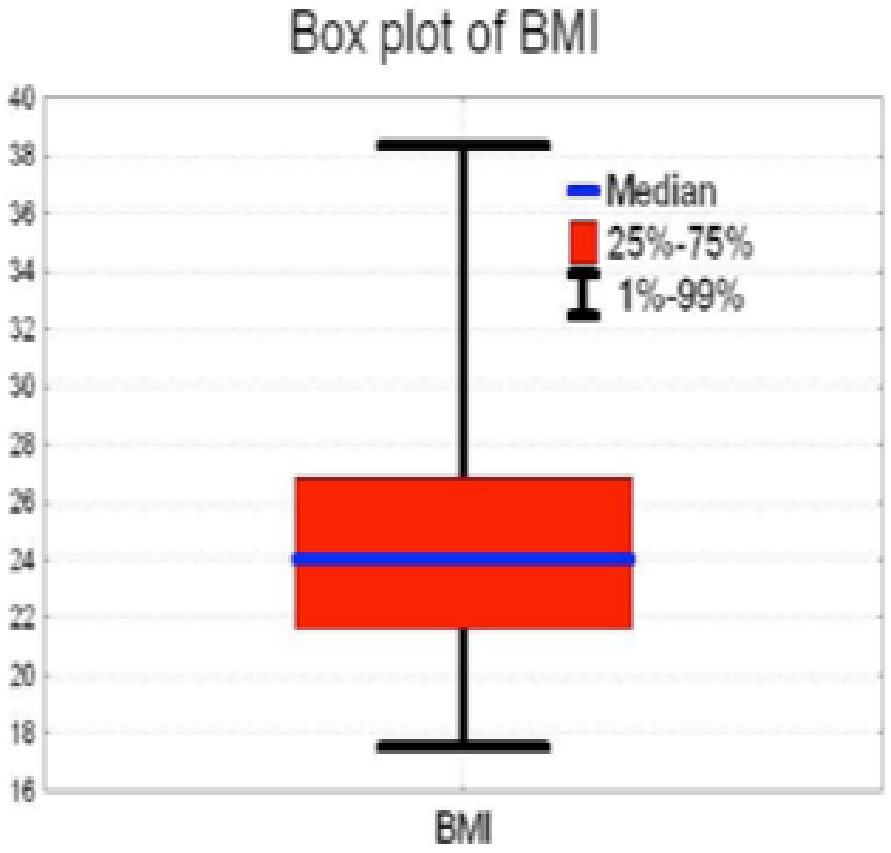
Why use graphics in data Analysis?

- To understand data properties
- To communucate results

Graphs of single quantitative variables

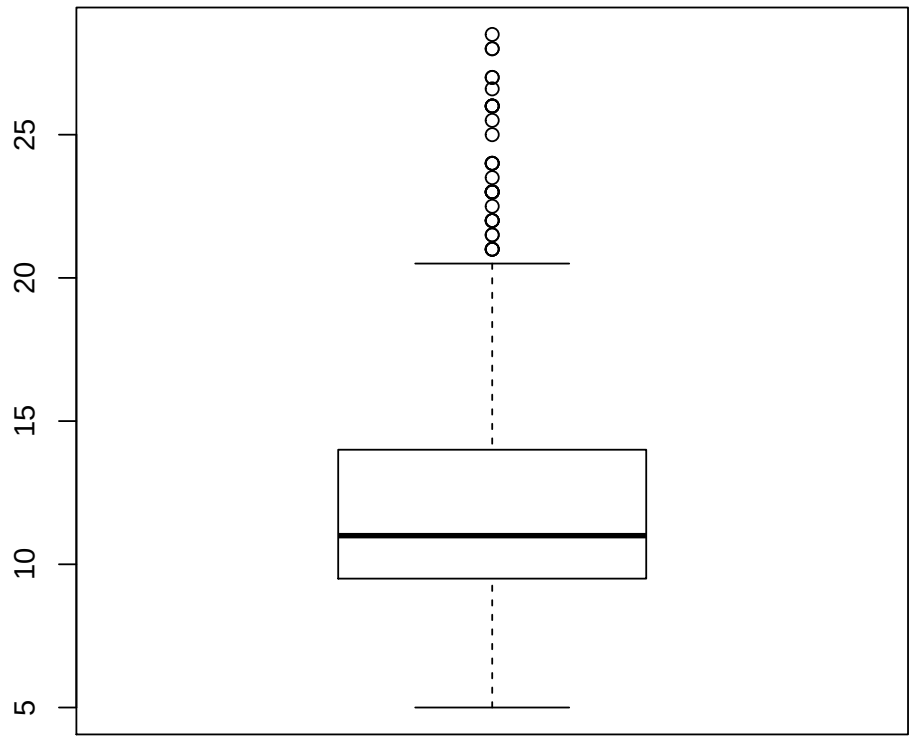
- **Box(-whisker)plot**

It is suitable for skewed data, where it may not be correct to show mean and standard deviations.



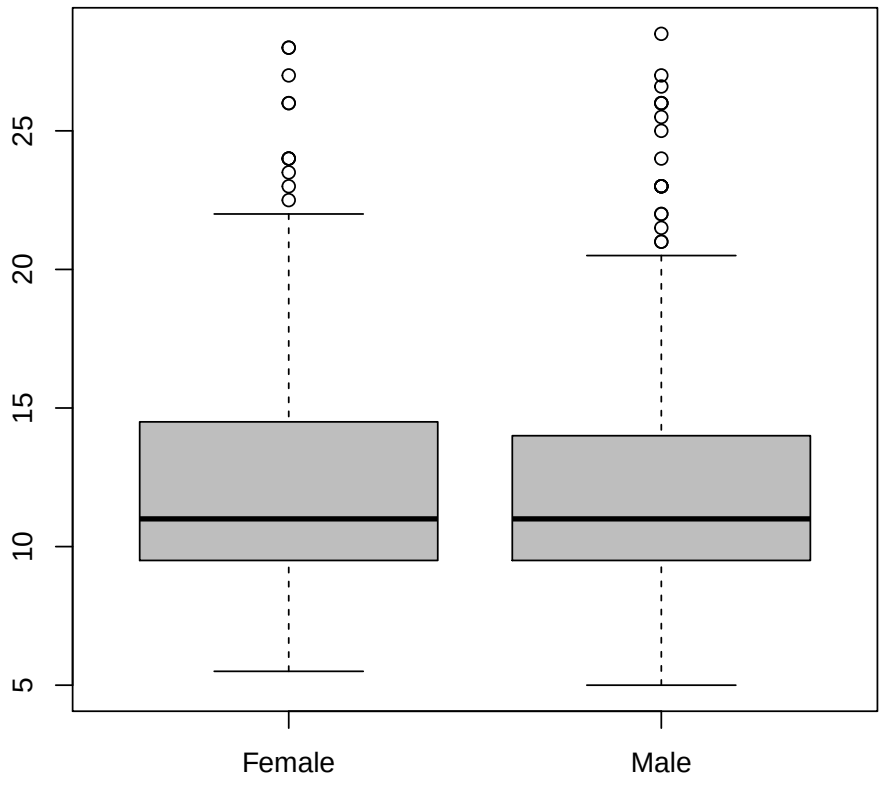
Boxplots

```
> boxplot(maltreat$weight)
>
```



Boxplots

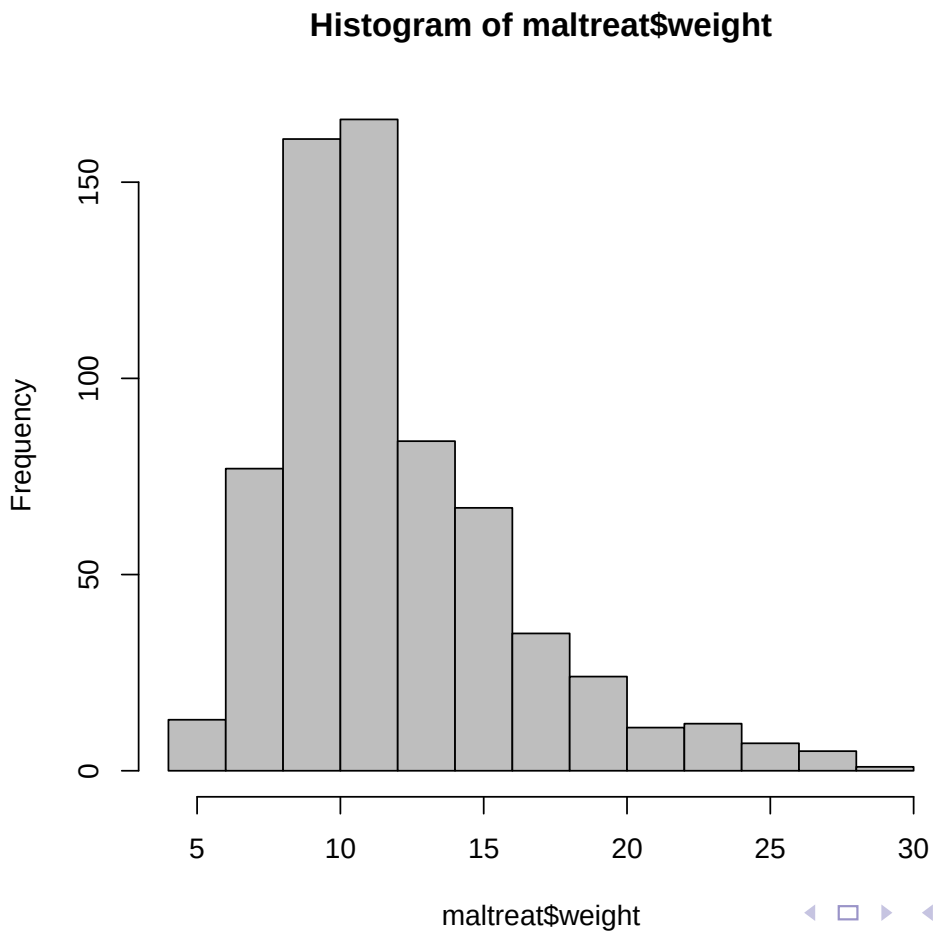
```
> boxplot(maltreat$weight~maltreat$sex,
+         col="grey")
>
```



Histograms

Explore the assumption of normality.

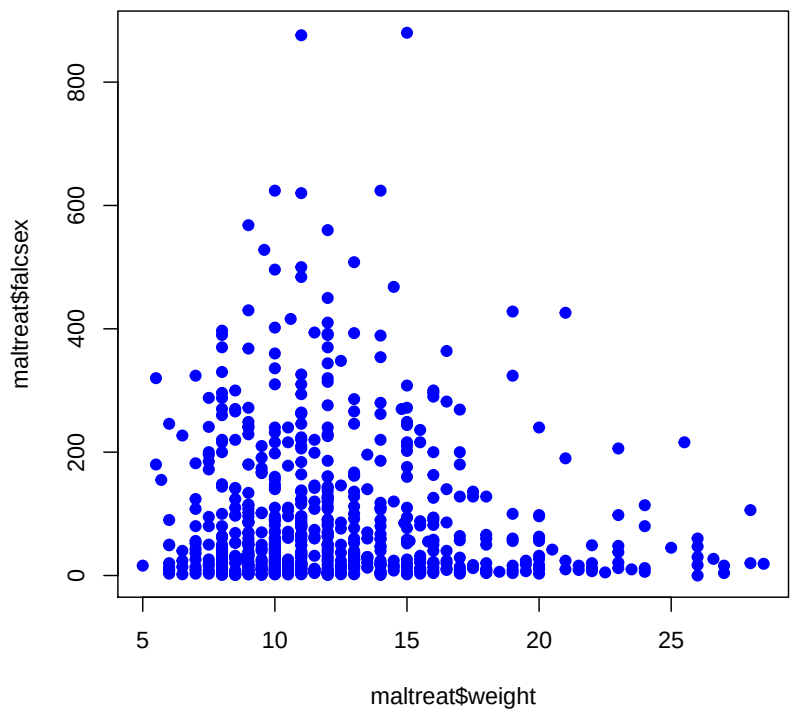
```
> hist(maltreat$weight, col="grey")
>
```



Scatterplots

- Visualize relationship between two quantitative variables.

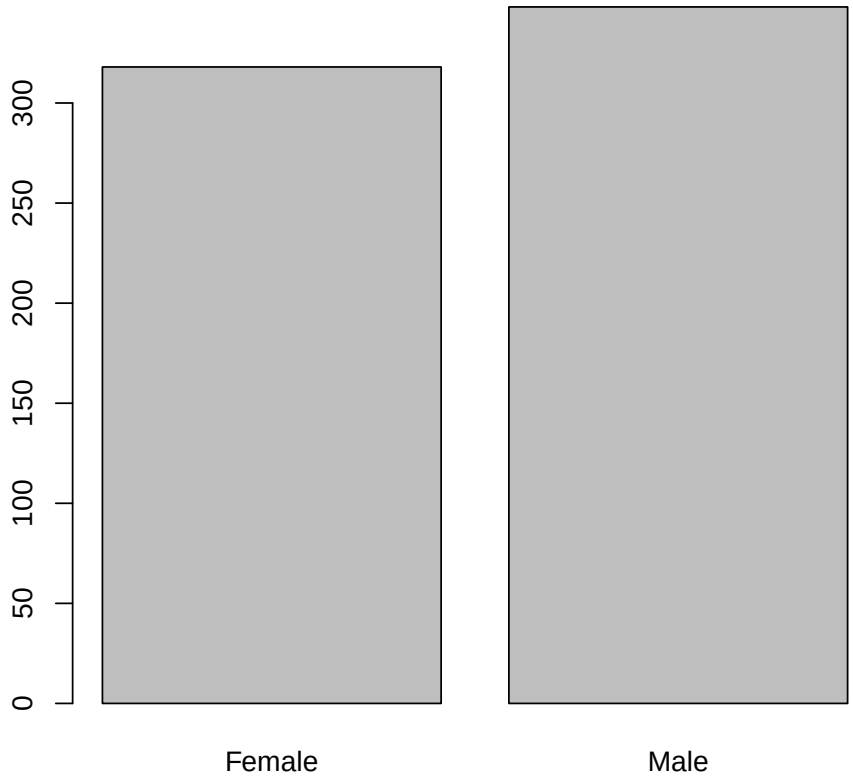
```
> plot(maltreat$weight,maltreat$falcsex, pch=19,
+      col="blue")
>
```



Graphs on single qualitative variables

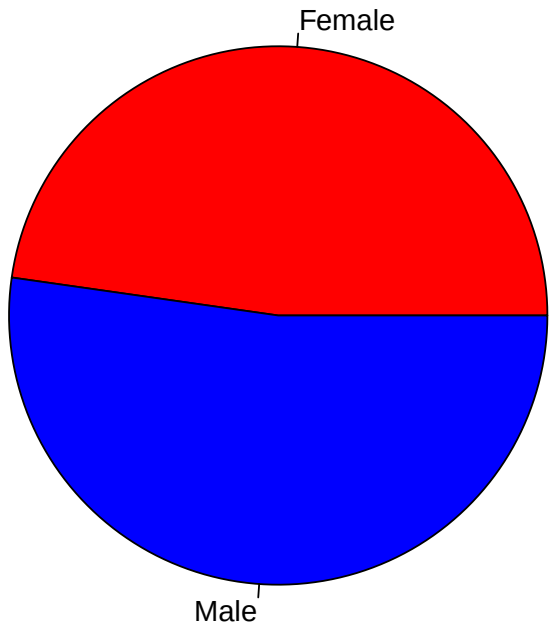
- **Bar plots**

```
> barplot(table(maltreat$sex))
```



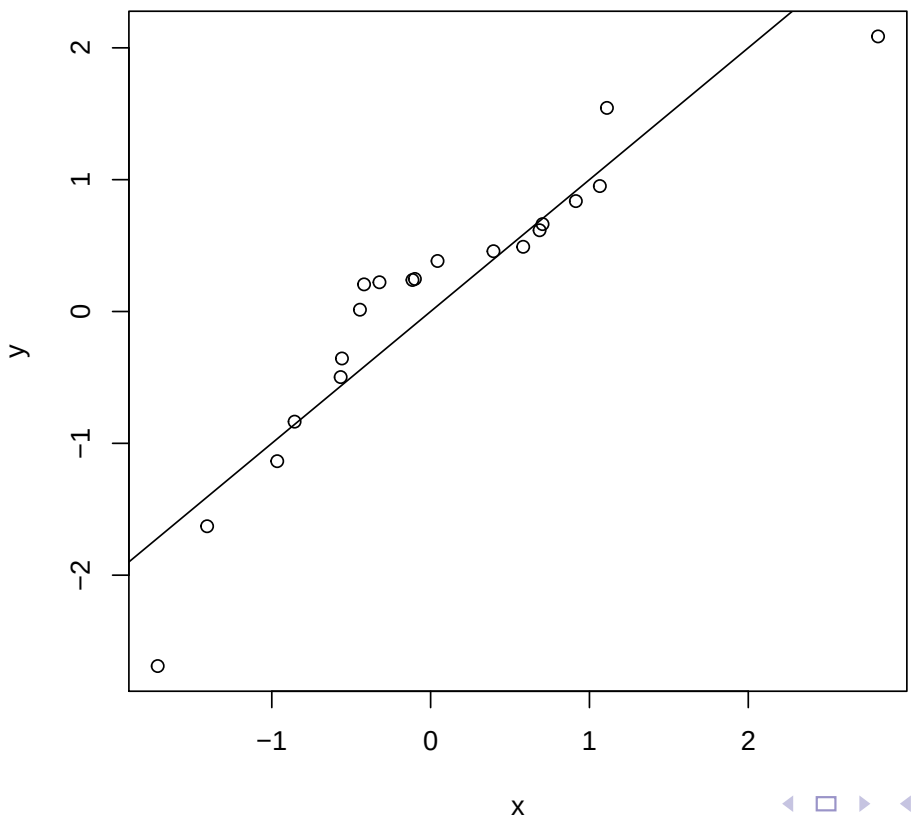
Pie charts

```
> pie(table(maltreat$sex), col=c("red","blue"),
+     main="Pie chart of gender")
```



QQ Plots

```
> x<-rnorm(20)
> y<-rnorm(20)
> qqplot(x,y)
> abline(c(0,1))
```



Graphical workflow

Graphical workflow

- Start with a rough plot

Graphical workflow

- Start with a rough plot
- Tweak it to make it expository

Graphical workflow

- Start with a rough plot
- Tweak it to make it expository
- Save the file

Graphical workflow

- Start with a rough plot
- Tweak it to make it expository
- Save the file
- Include it in presentations/manuscript

Useful Graphical Parameters

par function gives the parameters for the plot window.

- pch: the plotting symbol (default is open circle)
- lty: the line type (default is solid line), can be dashed, dotted, etc.
- lwd: the line width, specified as an integer multiple
- col: the plotting color, specified as a number, string, or hex code; the colorsfunction gives you a vector of colors by name. colors() gives a list of all the available colours.
- las: the orientation of the axis labels on the plot

Useful Graphical Parameters

- bg: the background color
- mar: the margin size
- mtext: add arbitrary text to the margins (inner or outer) of the plot
- mfrow: number of plots per row, column (plots are filled row-wise)
- mfcol: number of plots per row, column (plots are filled column-wise)

Resources

R graph gallery
Cookbook for R
How to display data badly