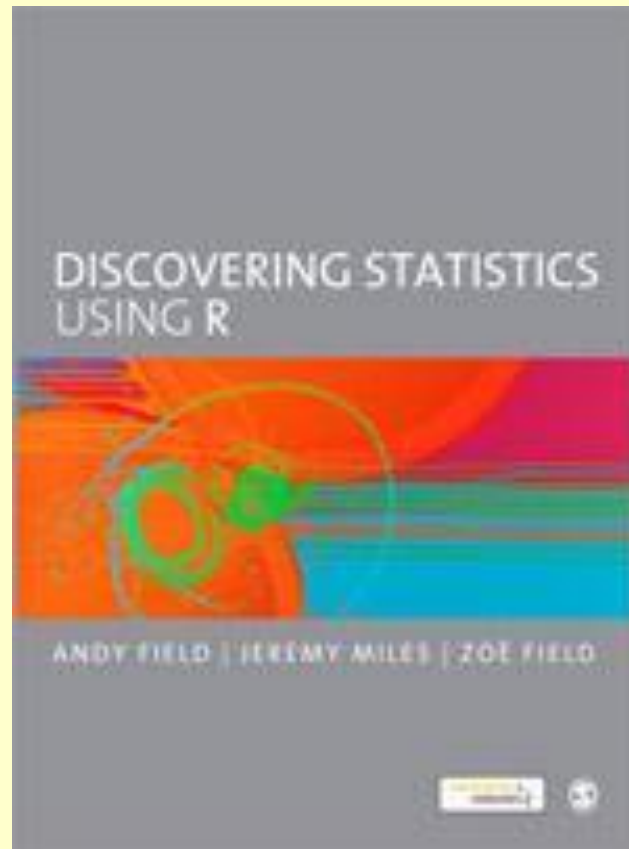


# Graphical Exploration I:

## R Commander



# Exploring Data With Graphs

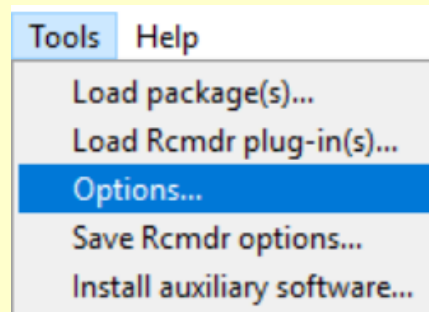


## Chapter 4

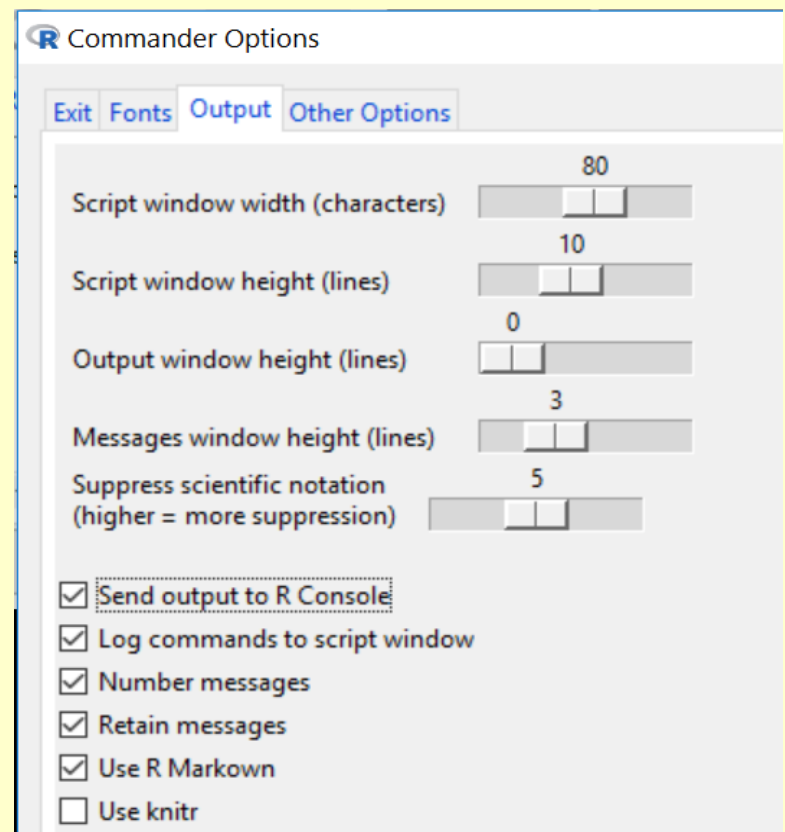
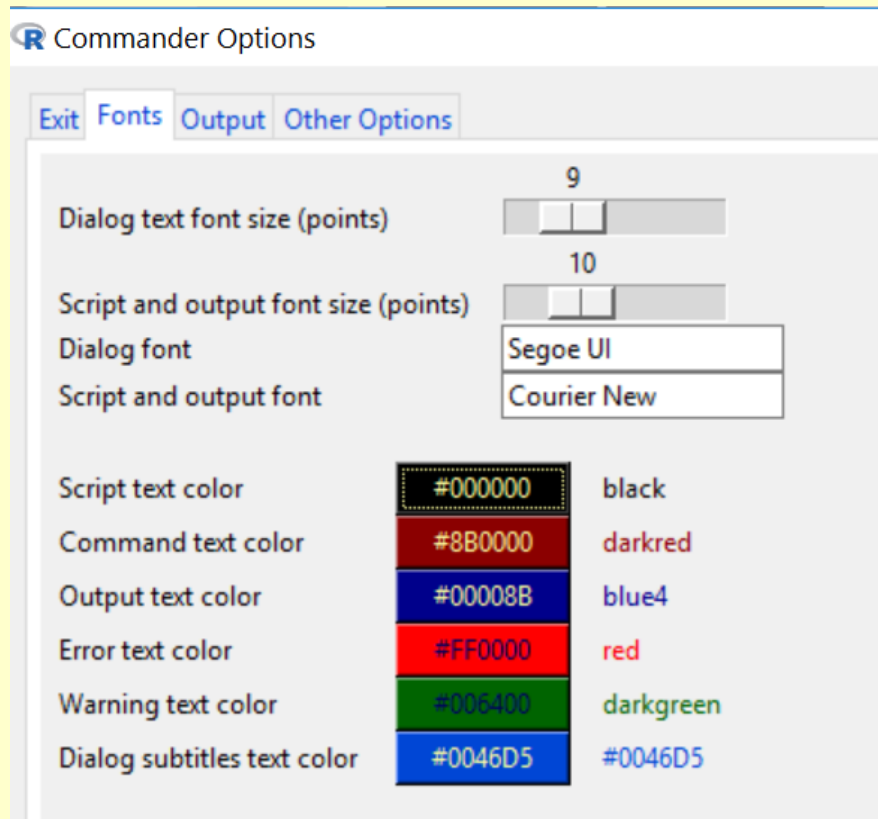
- **Aim:** Provide overview of Rcmdr graphs
- **The Basics:**
  - Histograms
  - Density plots
  - Boxplots
  - Scatterplots
  - Line graphs

# Options in Rcmdr

Rcmdr provides tool options

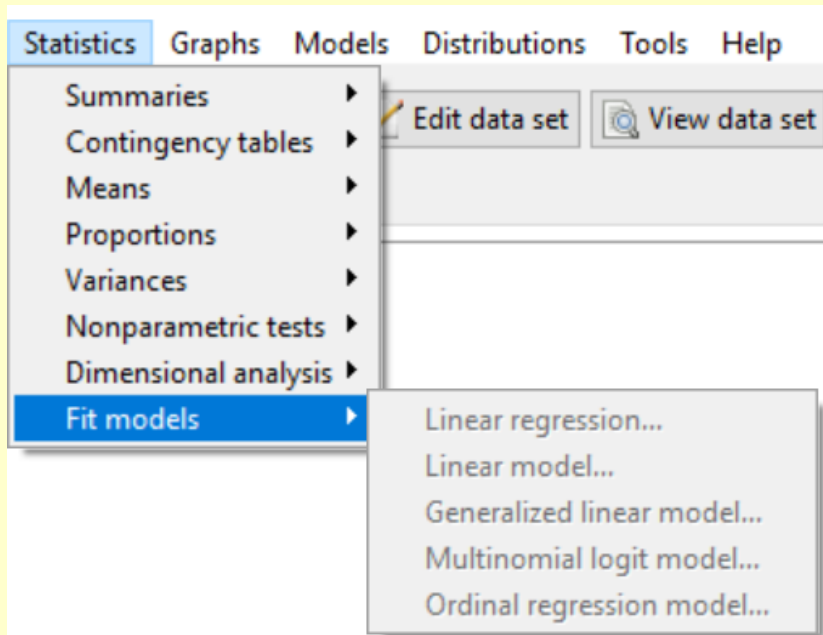


Font type, size, color  
Output size

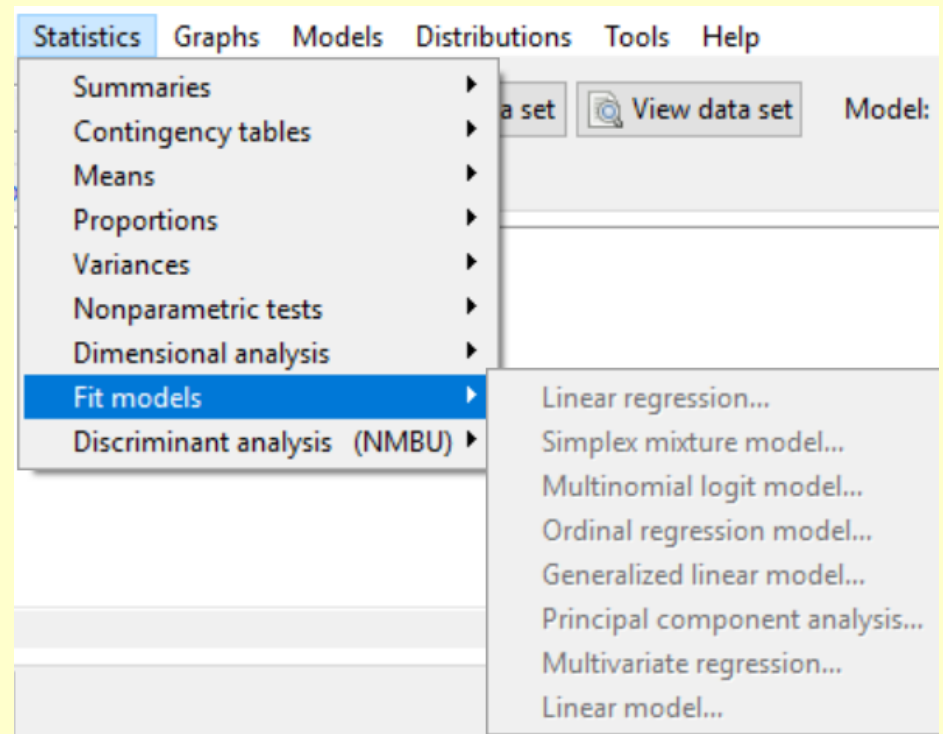


# Rcmdr Toolkits

## Rcmdr Statistics Menu



## Expanded Functionality after installing package 'RcmdrPlugin.NMBU'



# Rcmdr Plug-ins

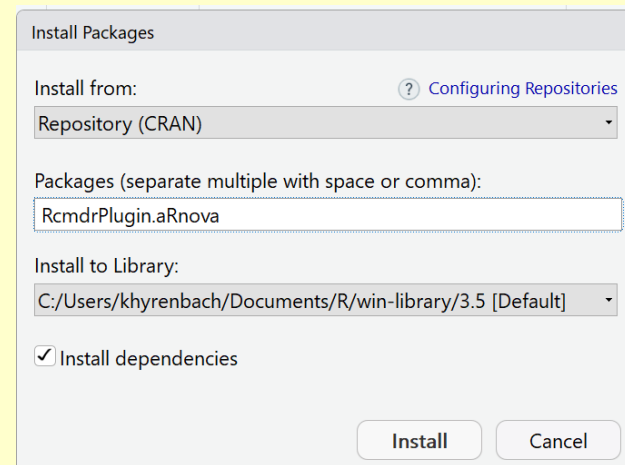
Install package:  
'RcmdrPlugin.NMBU'

R Commander Plug-in for  
University Level Applied Statistics

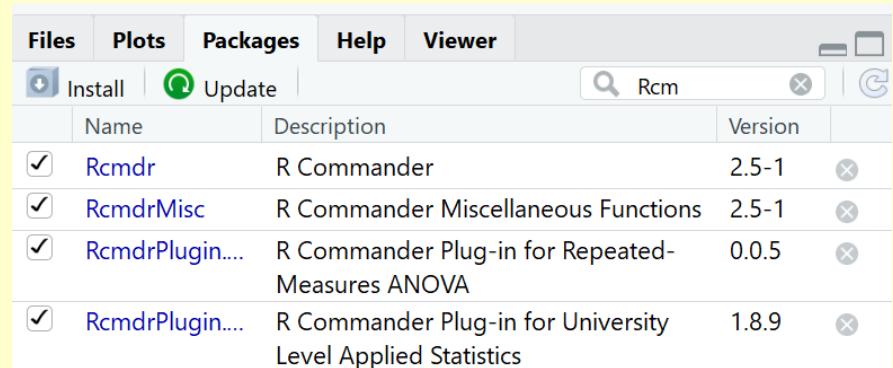


Documentation for package  
'RcmdrPlugin.NMBU' version 1.8.9

1) Install package and  
dependencies from CRAN

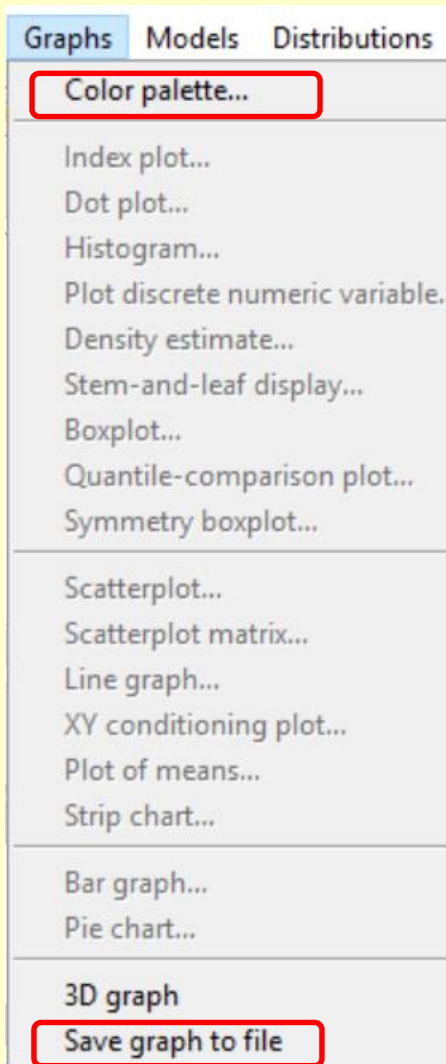


2) Activate package



# Graphs in Rcmdr

## Standard Rcmdr Menu



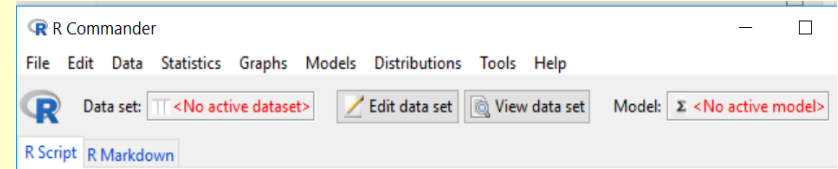
Set Color Palette

Greyed out buttons indicate they are not available:

Why not?

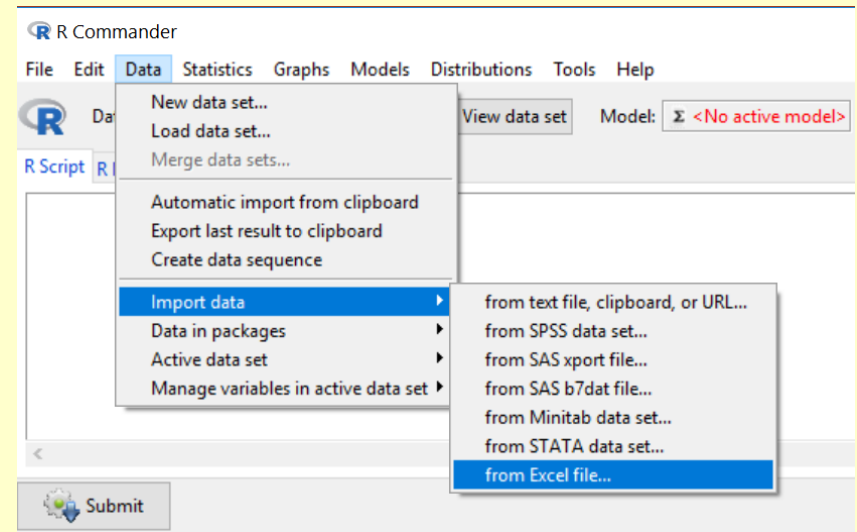
No data uploaded

Save Graph to File

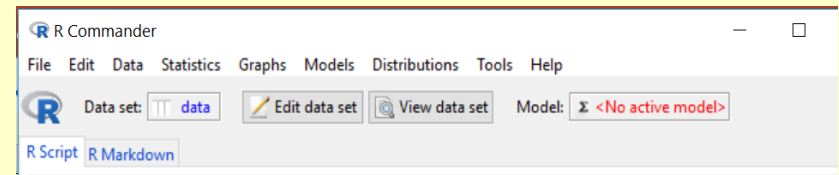


## To Load a Dataset

1) Import xls file:



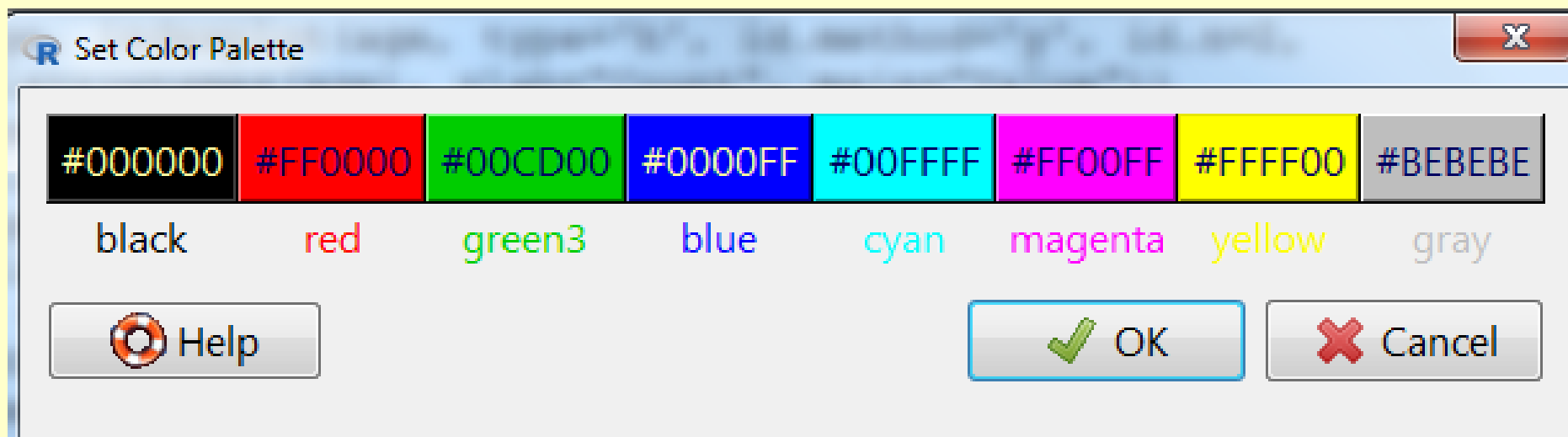
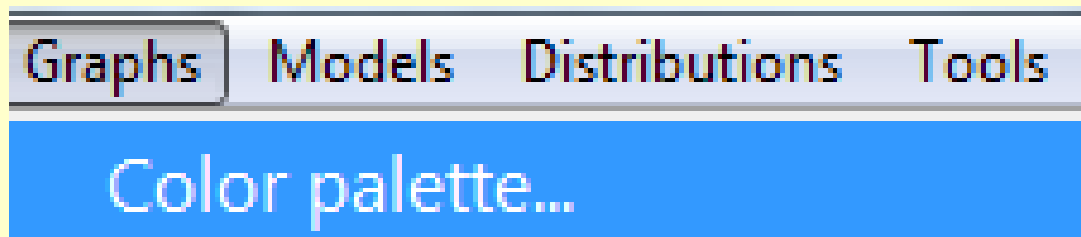
2) Select dataset:



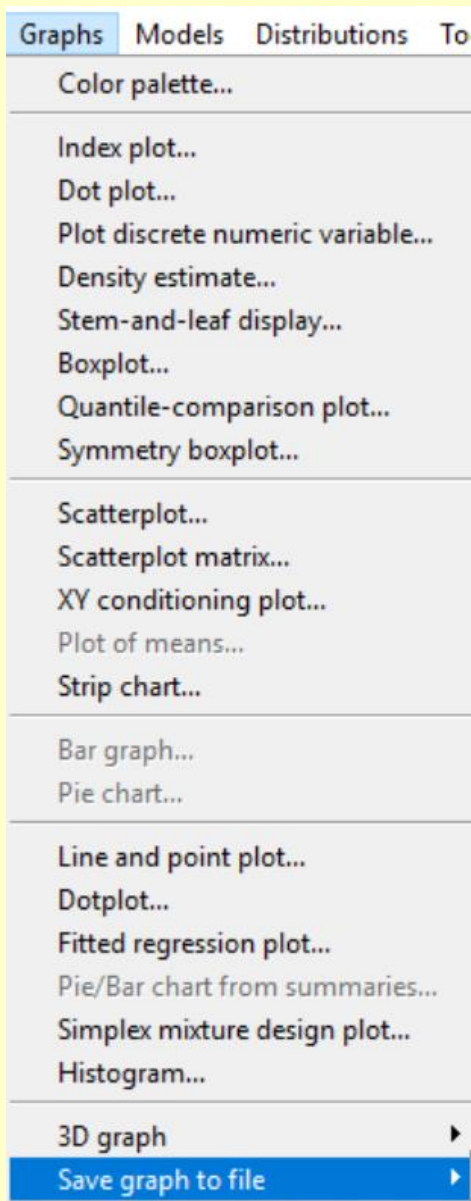
# Graphs in Rcmdr - Palette

## Graphs Tab Menu

Change the color options used in the figures



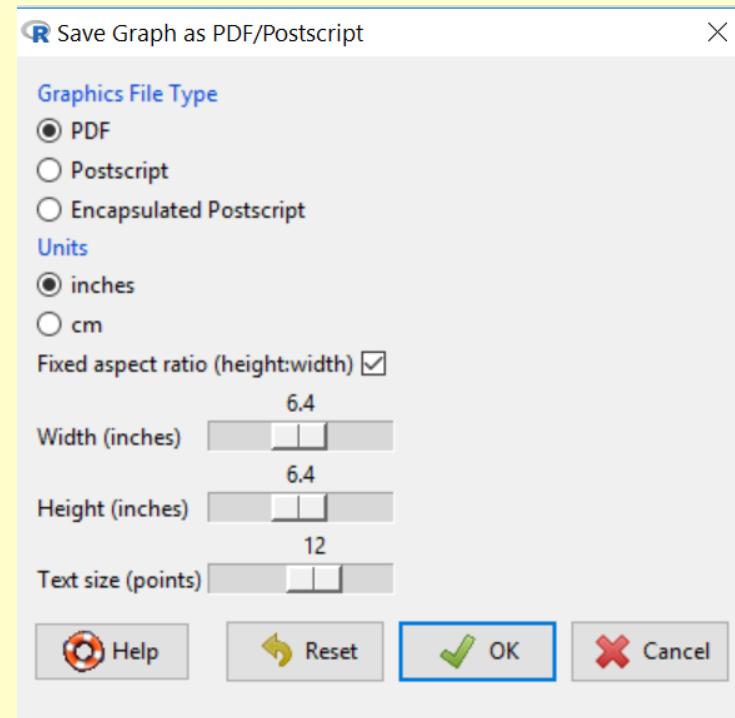
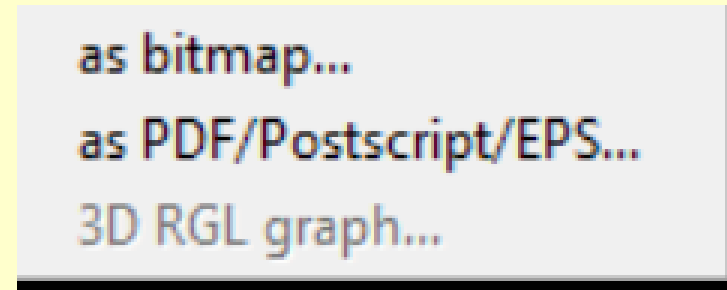
# Graphs in Rcmdr - Save Figure



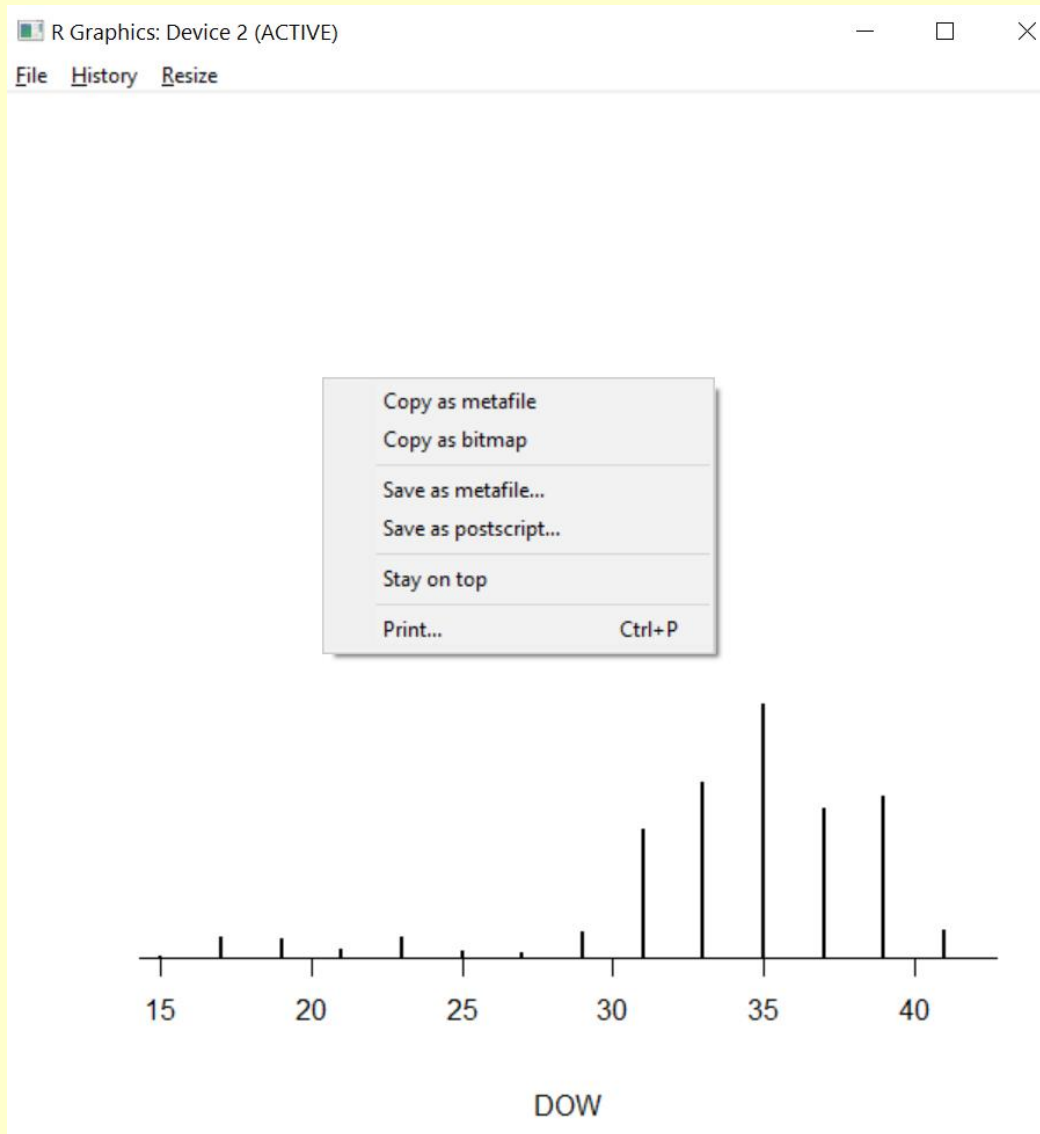
## Graphs Tab Menu

Change format of exported files

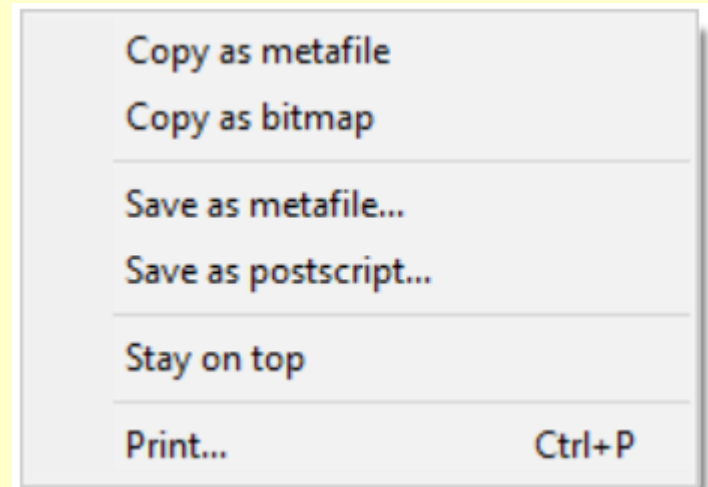
Select size of figure and text



# Graphs in Rcmdr - Save Figure

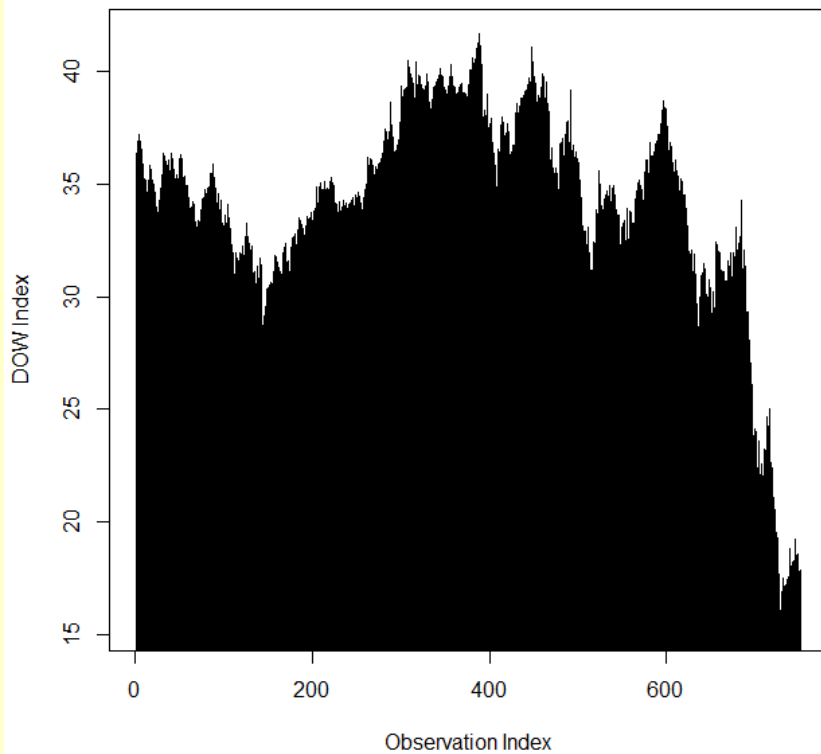


Copy and Save File

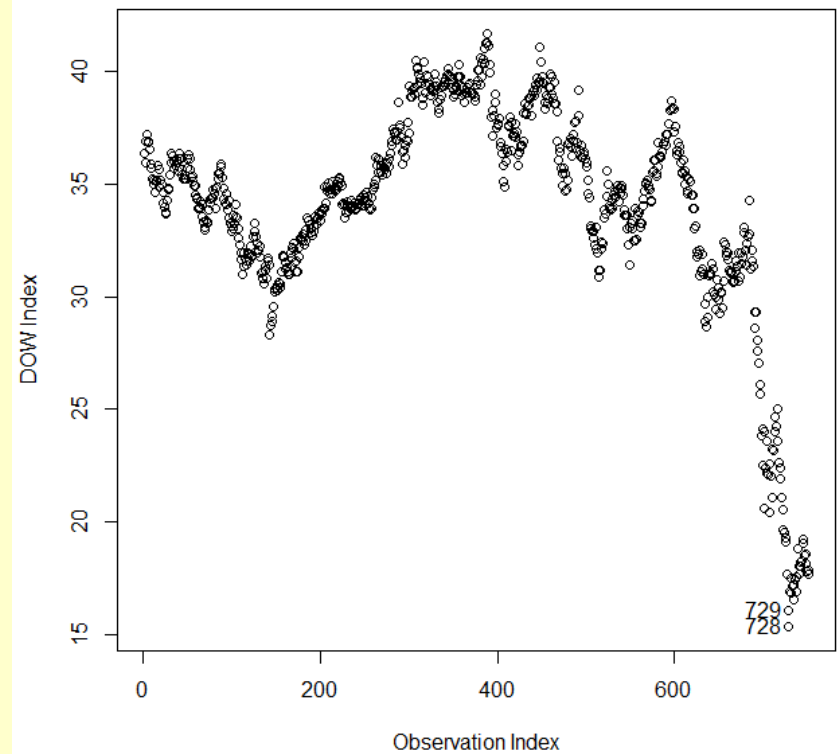


# Graphs in Rcmdr - Index Plot

Time Series of DOW Index

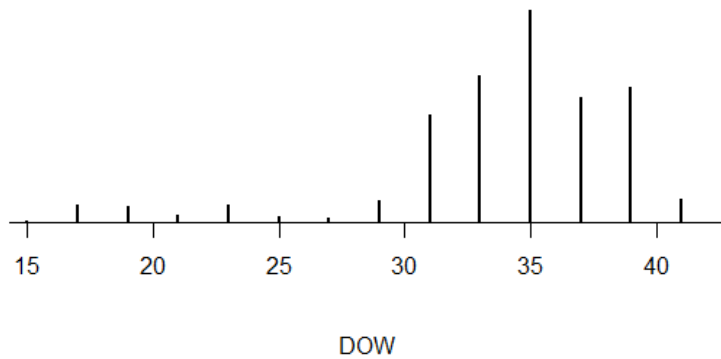


Time Series of DOW Index



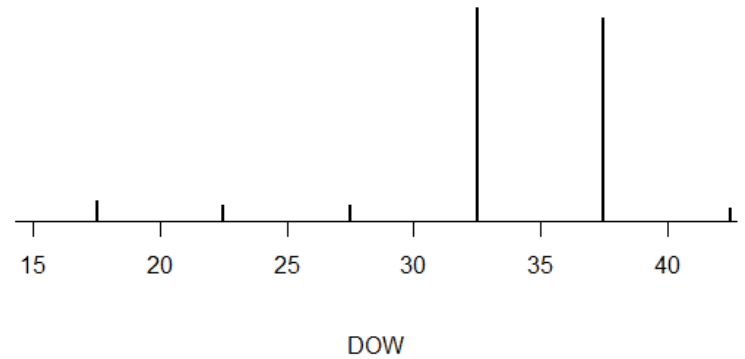
Plot all data values (as spikes or dots) sequentially

# Graphs in Rcmdr - Dot Plot



Raw Data

OR



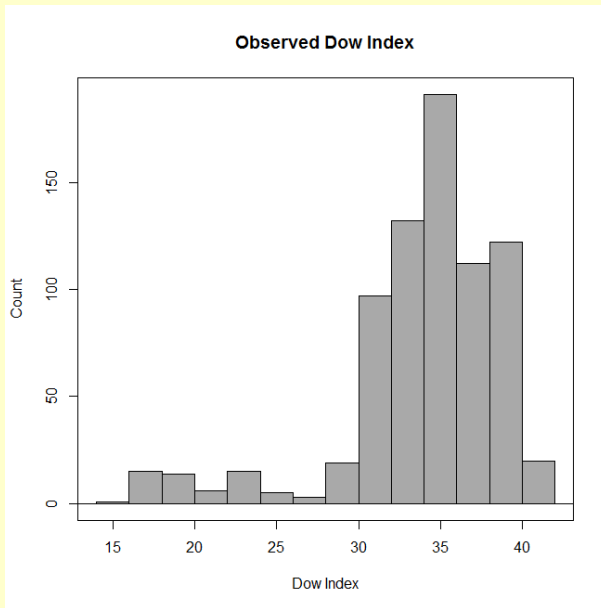
Binned Data  
(bin = 5)

# Graphs in Rcmdr - Histogram

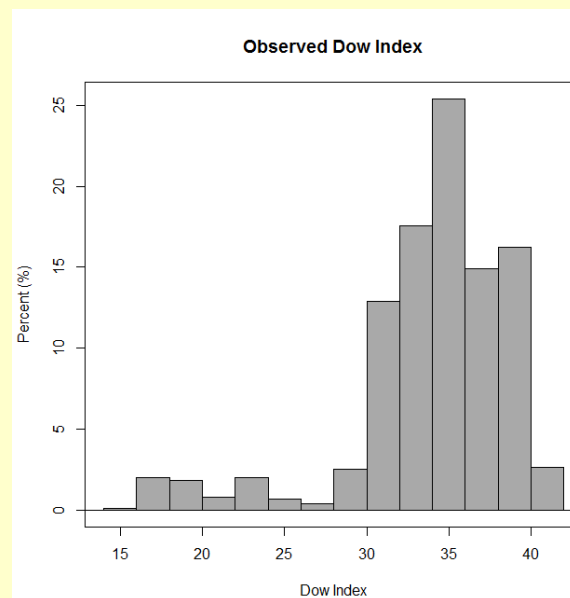
Plot all data values (binned) in histograms

Define the number of bins (or use "auto" option)

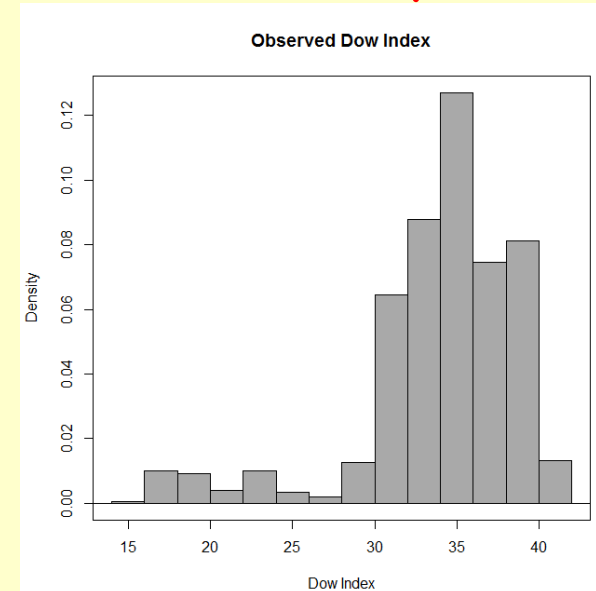
Count



Percent



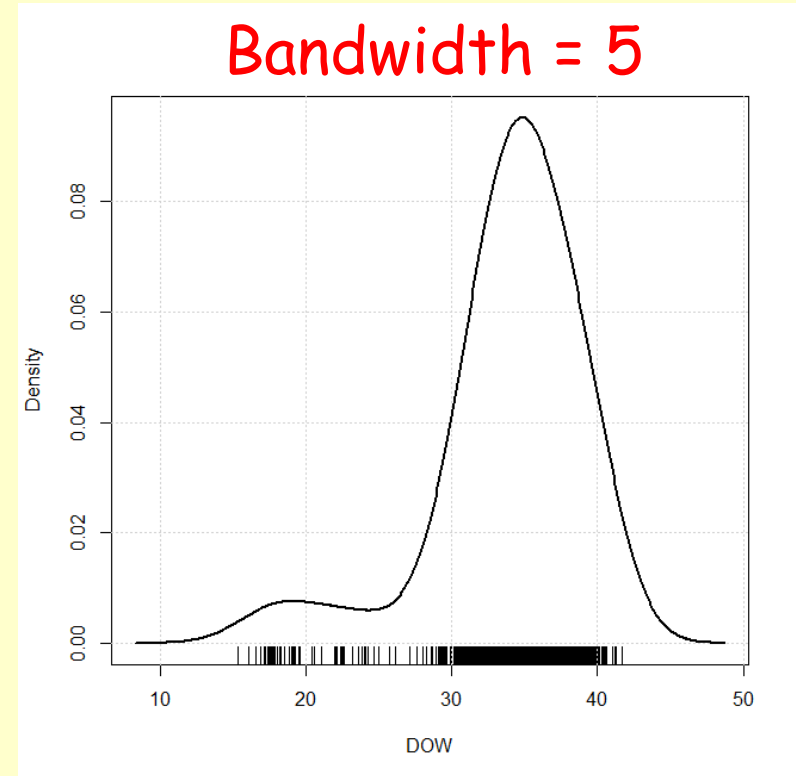
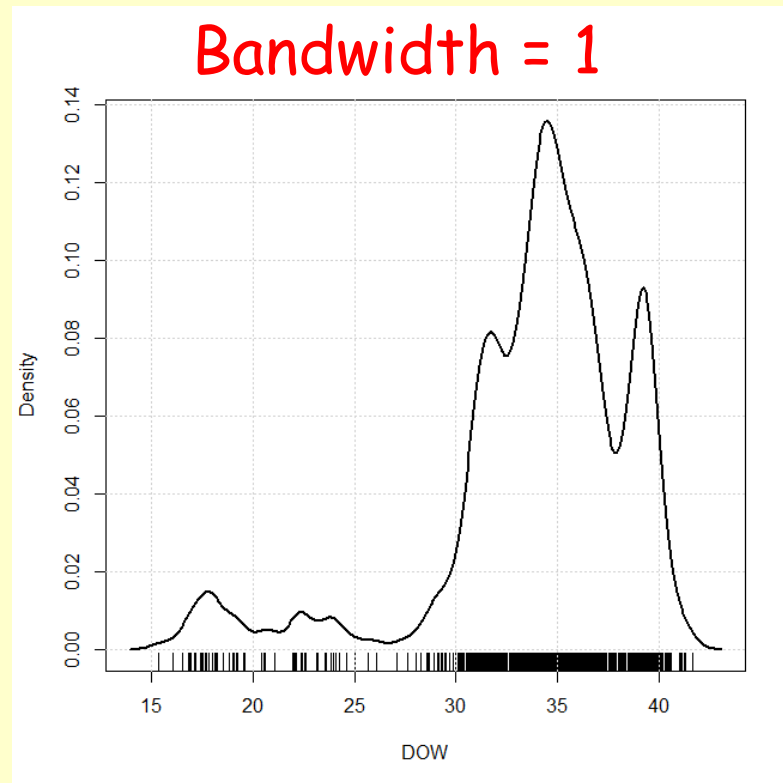
Density



Frequency

# Graphs in Rcmdr - Density

Plot a non-parametric estimate of the data, based on different methods and various smoothing parameters




**The Rug = shows location of observed X values**

# Graphs in Rcmdr - Histogram

Plot all data values (binned) in histograms

Add a density distribution fit - for reference:

 Histogram

Variable (pick one)

DOW  
FTSE  
NASDAQ  
Time

Number of bins: <auto>

Axis Scaling

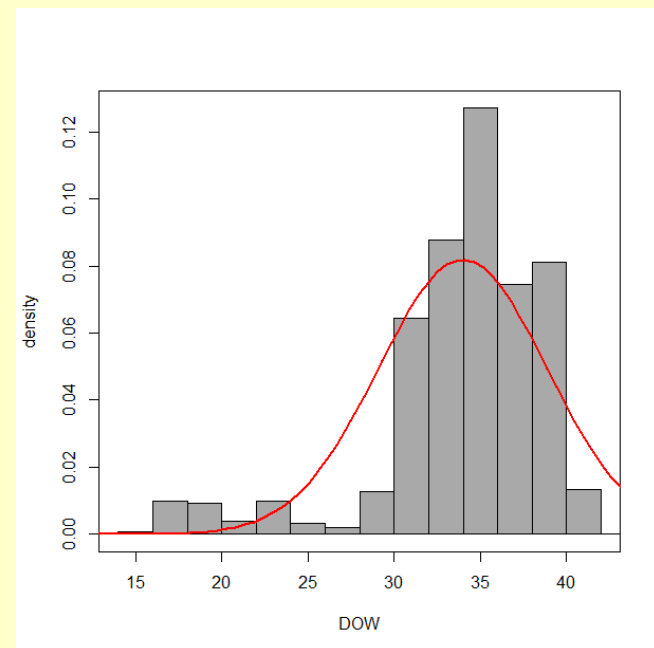
Frequency counts

Percentages

Densities

Discrete variable

Density fit: normal

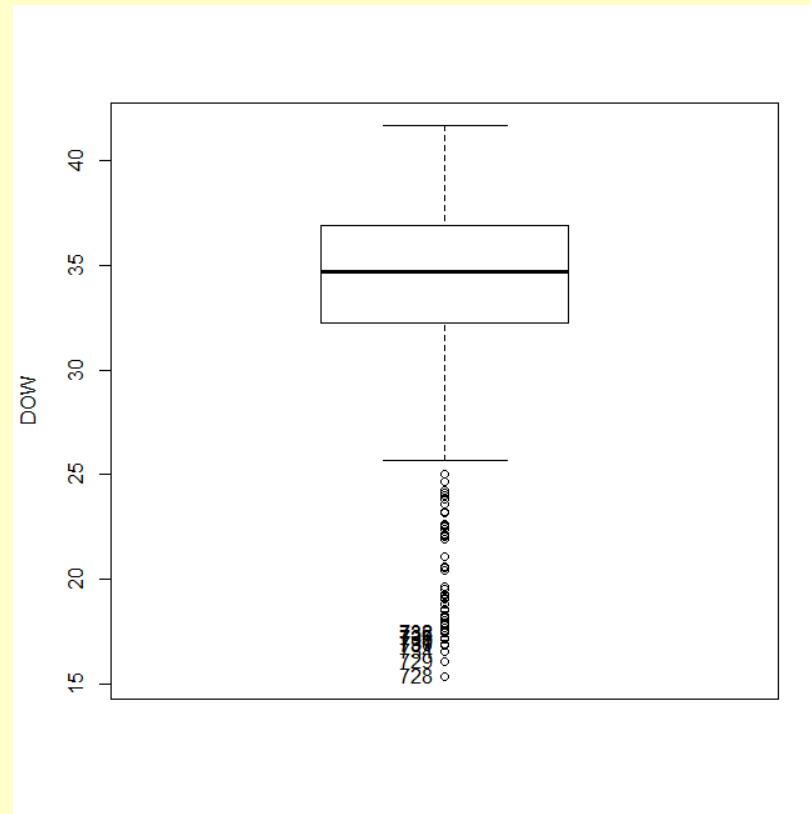


Normal Distribution  
(same mean and SD)

# Graphs in Rcmdr - Box Plots

Plots quantiles of the distribution: 5, 25, 50, 75, 95

Identifies outliers: values too far from the median



Identify Points: Automatic or Interactively

# Graphs in Rcmdr - Quantiles

The quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from populations with a common distribution. It shows the quantiles of the first data set against the quantiles of the second data set.

By a quantile, we mean the percent of data points below the given value. That is, the 30% quantile is the point at which 30% percent of the data fall below and 70% fall above that value. The median is the 50% quantile.

If the two sets come from a population with the same distribution, the points will fall along a 45 degree line. The greater the departure from this reference line, the stronger the evidence that the two data sets come from populations with different distributions.

# Graphs in Rcmdr - Quantiles

Graphically compares an observed (empirical) distribution (points) with a chosen theoretical expectation (line)

R Quantile-Comparison (QQ) Plot

Data Options

Plot Options

Distribution

Normal

t df =

Chi-square df =

F Numerator df =  Denominator df =

Other Specify:  Parameters:

Identify Points

Automatically

Interactively with mouse

Do not identify

Number of points to identify

Help Reset

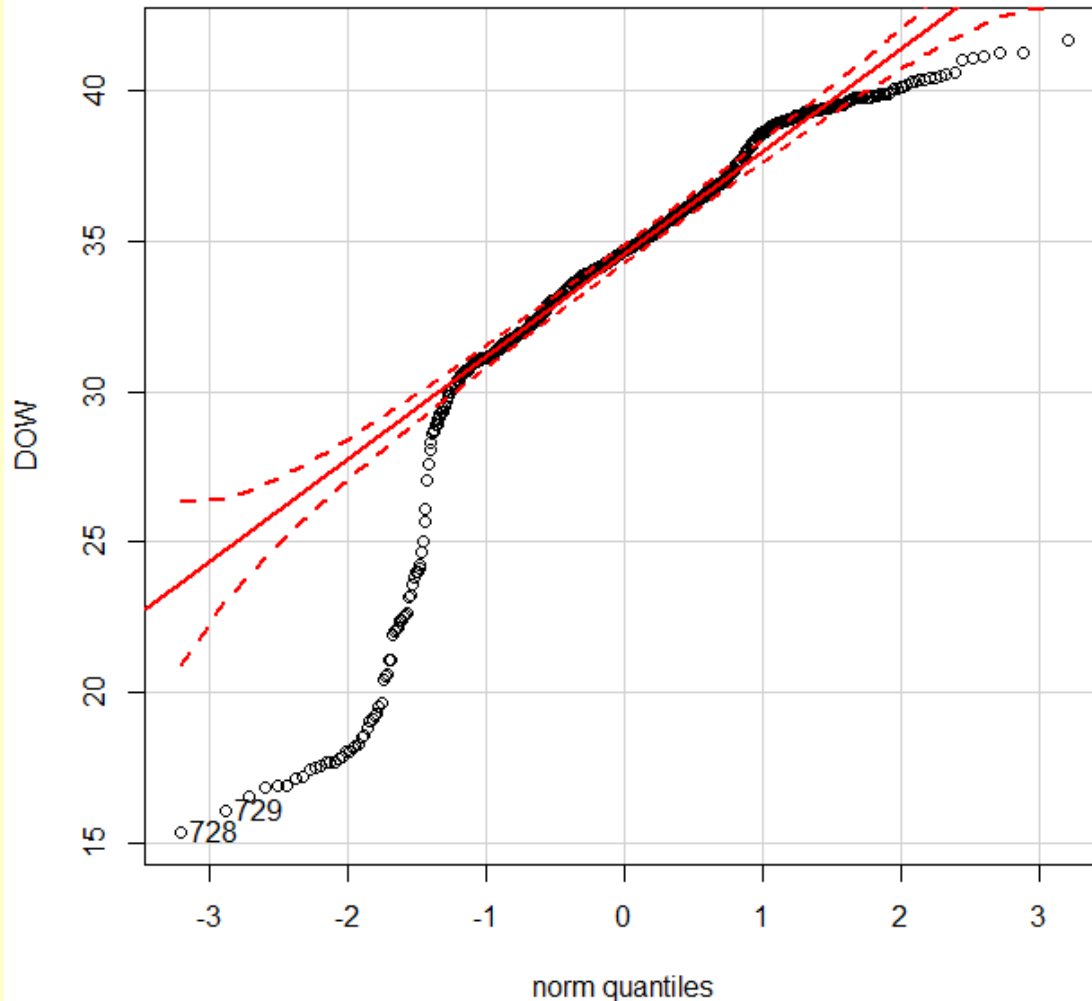
Normal Distribution is the Default

Identifies Max / Min as Default

Identify Points: Automatic or Interactively

# Graphs in Rcmdr - Quantiles

## Normal Distribution



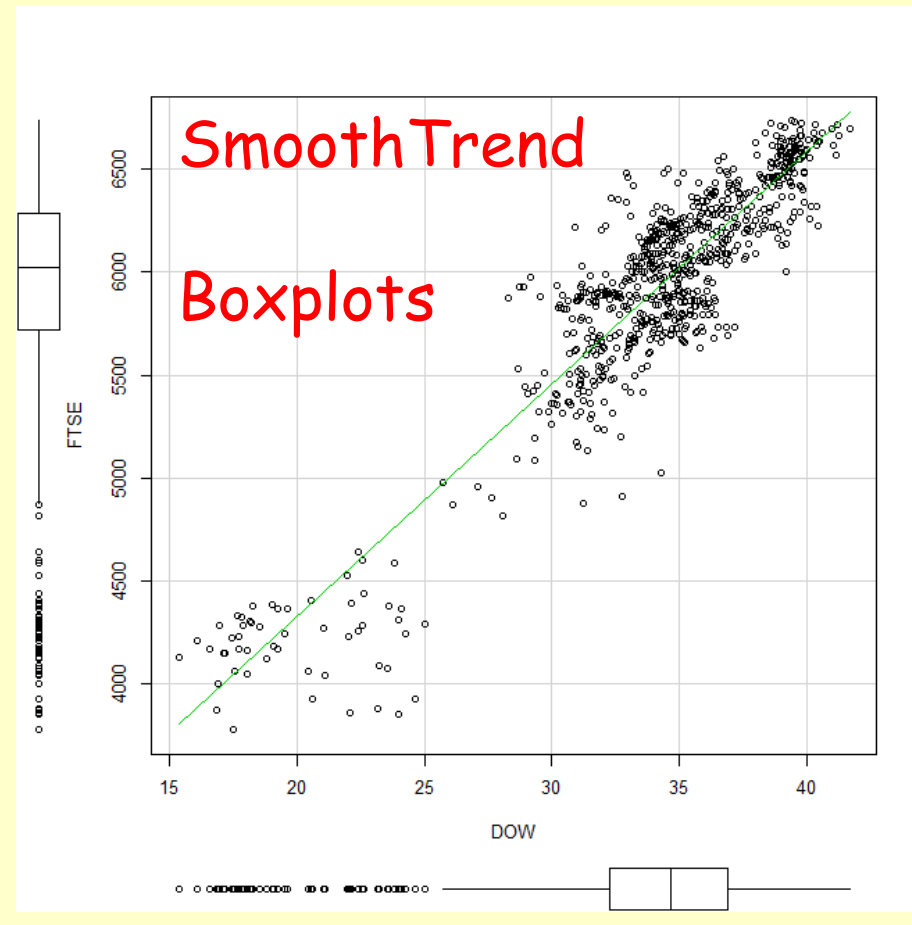
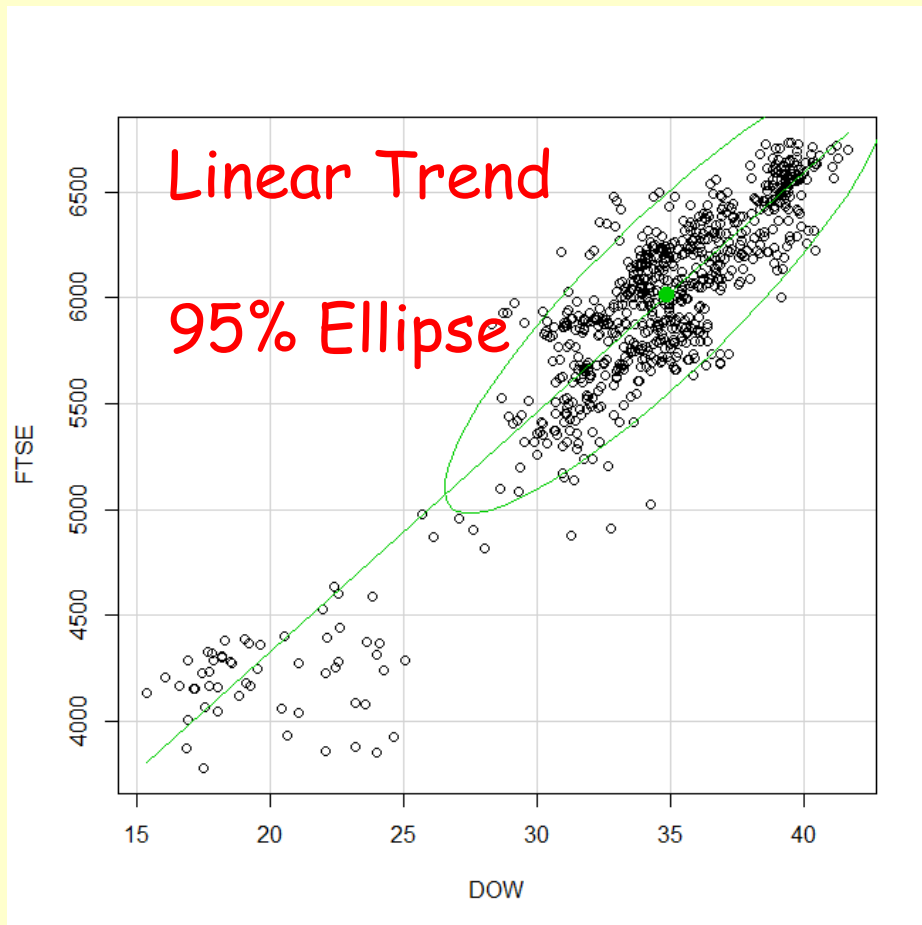
The solid red line is the expected pattern a normal distribution with the same mean and SD and the sampled data.

Points outside of the dashed line envelope suggest significant deviations

# Graphs in Rcmdr - Scatterplots


Plots two samples on the same coordinates

Can add a linear relationship or a smooth trend



# Graphs in Rcmdr - Symmetry

Plots boxplots of data after several transformations  
Can add a linear relationship or a smooth trend

 Symmetry Boxplot

Variable (pick one)

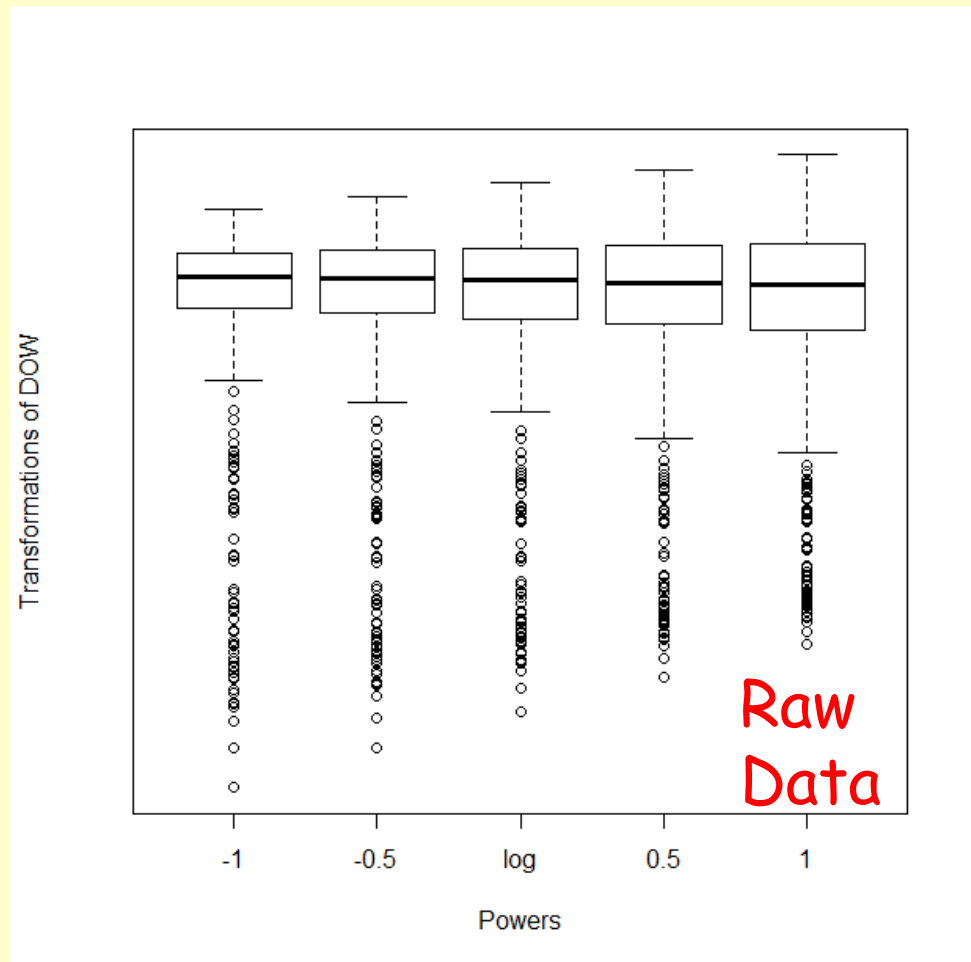
- DOW**
- FTSE
- NASDAQ
- Time

Transformation Family

- Box-Cox**
- Box-Cox with negatives
- Yeo-Johnson

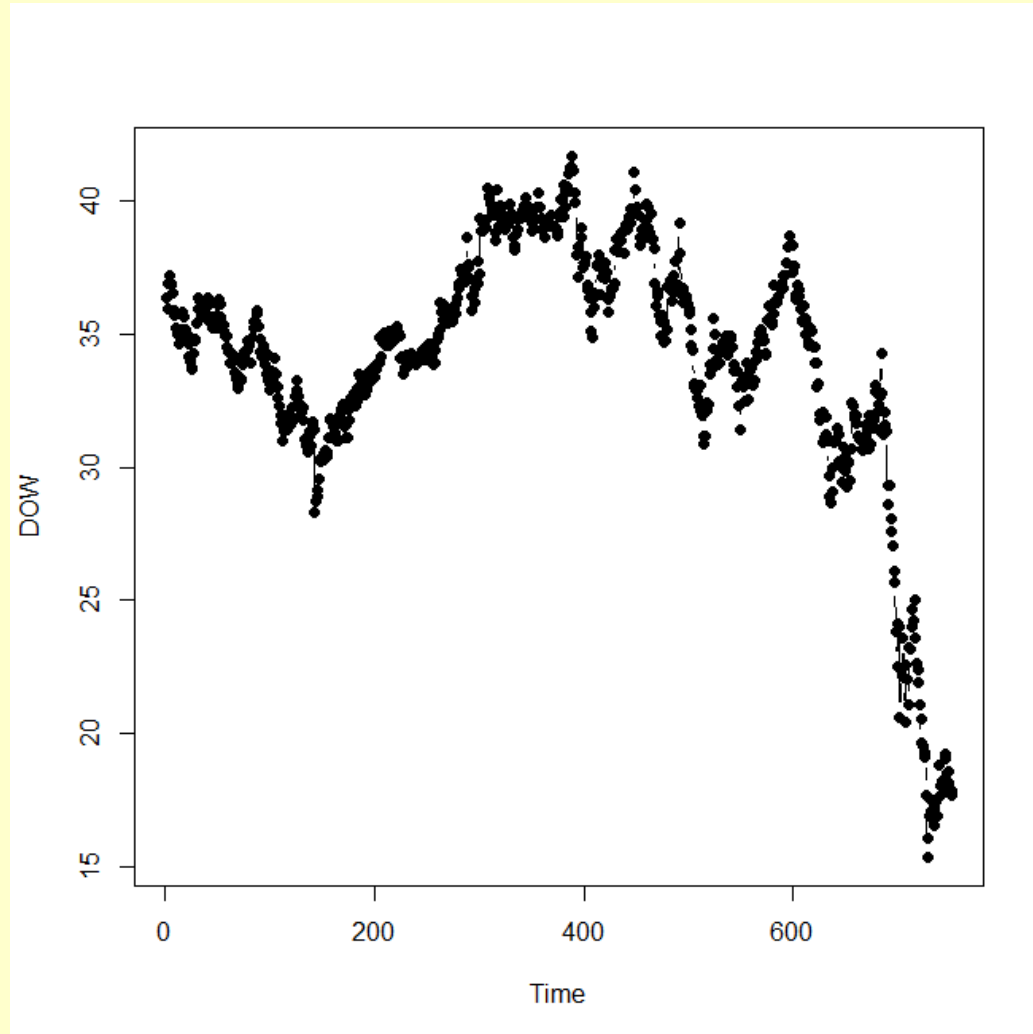
Start (optional)

Powers



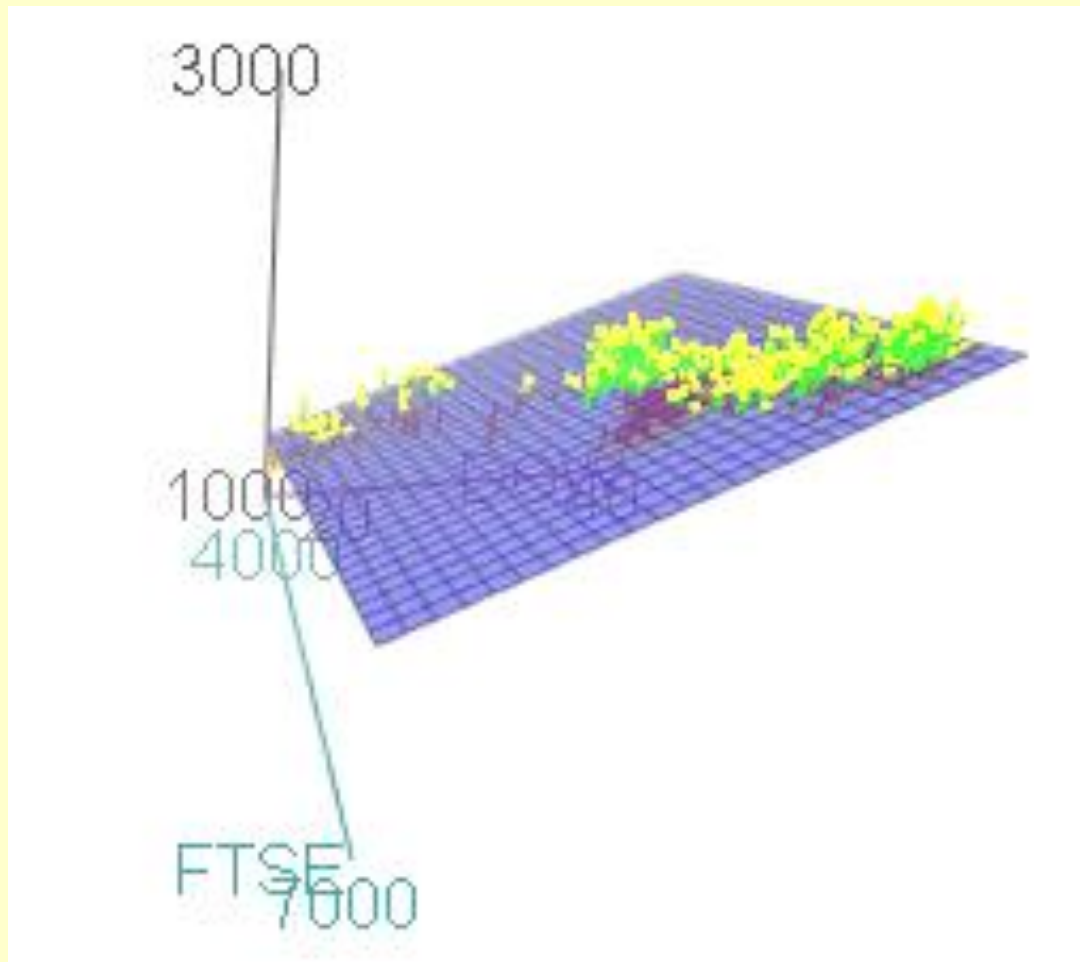
# Graphs in Rcmdr - Line Plots

Plots data points linked with a line



# Graphs in Rcmdr - 3D Scatterplots

Plots three variables (one response and two drivers)  
and adds a response surface and the deviations



# Graphs in Rcmdr - Regression Fit

R Fitted regression plot

X: regressor variable (pick one)  
DOW  
FTSE  
NASDAQ  
Time

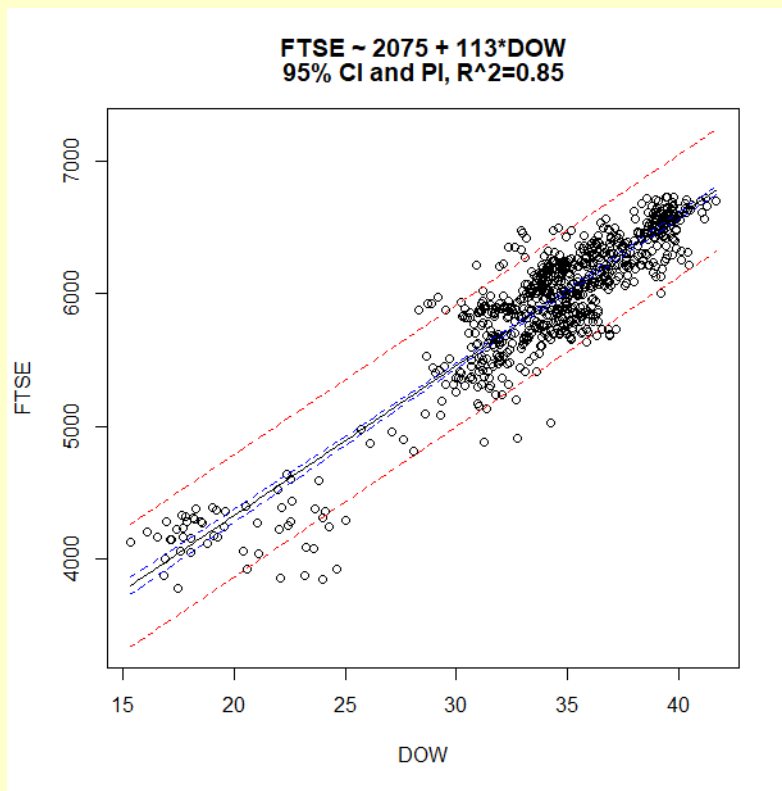
Y: response variable (pick one)  
DOW  
FTSE  
NASDAQ  
Time

X transf.: none

Confidence Level  
.95

Plots two variables (response and driver), adds best-fit line and desired C.L.

Allows to transform X variable



Plots C.I. and P.I.

Calculates R-squared

# PSA-1: Histograms

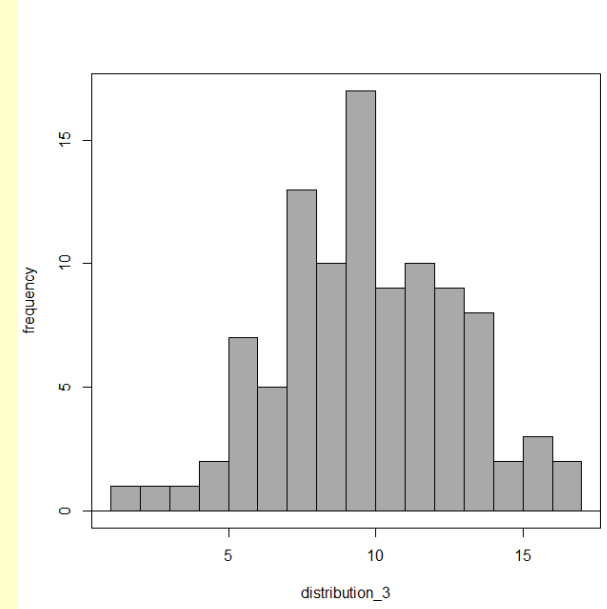
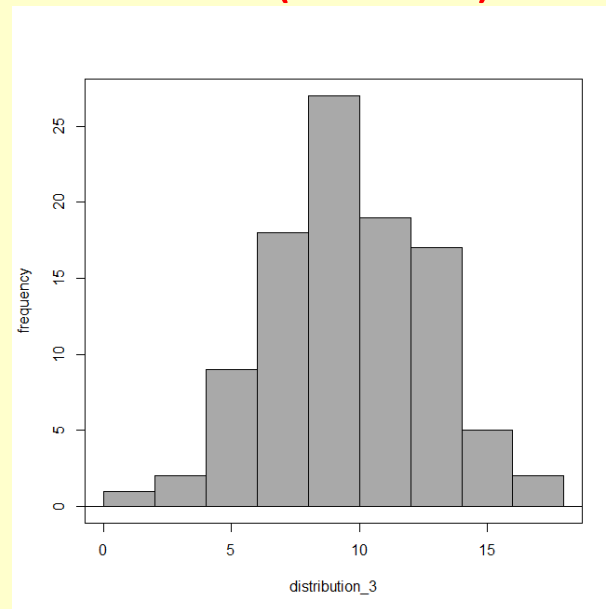
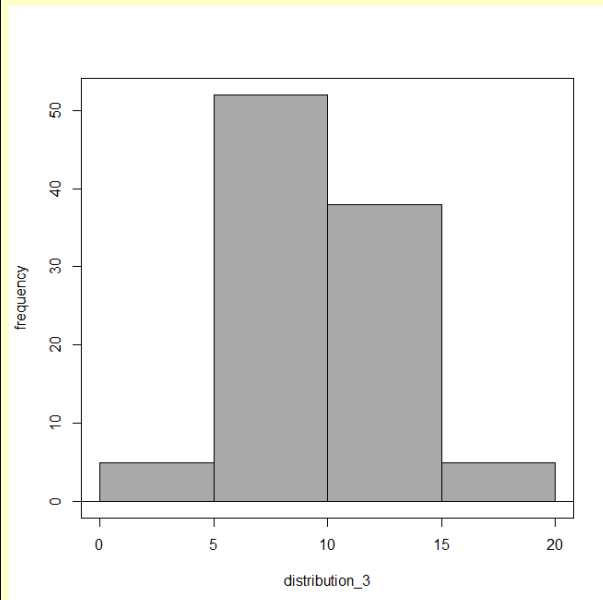
- Number of bins affects:  
    "shape" of distribution & number / location of modes
- Even when plotting a Normal Distribution:

	<u>mean</u>	<u>sd</u>	<u>5%</u>	<u>50%</u>	<u>95%</u>	<u>n</u>
dist_3	9.705097	3.033430	5.113512	9.761769	14.40838	100

5 bins

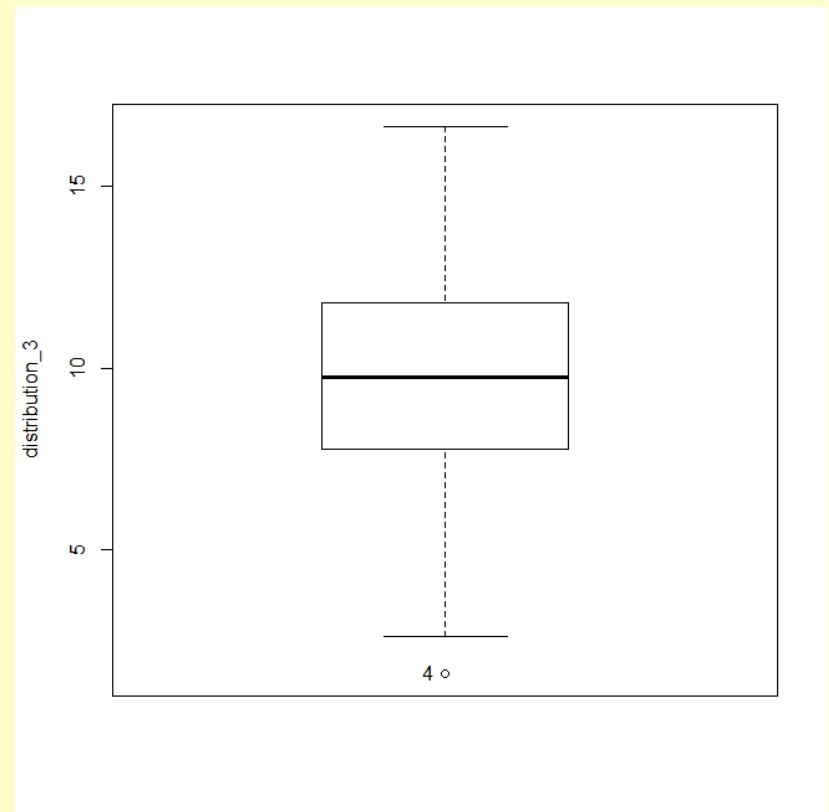
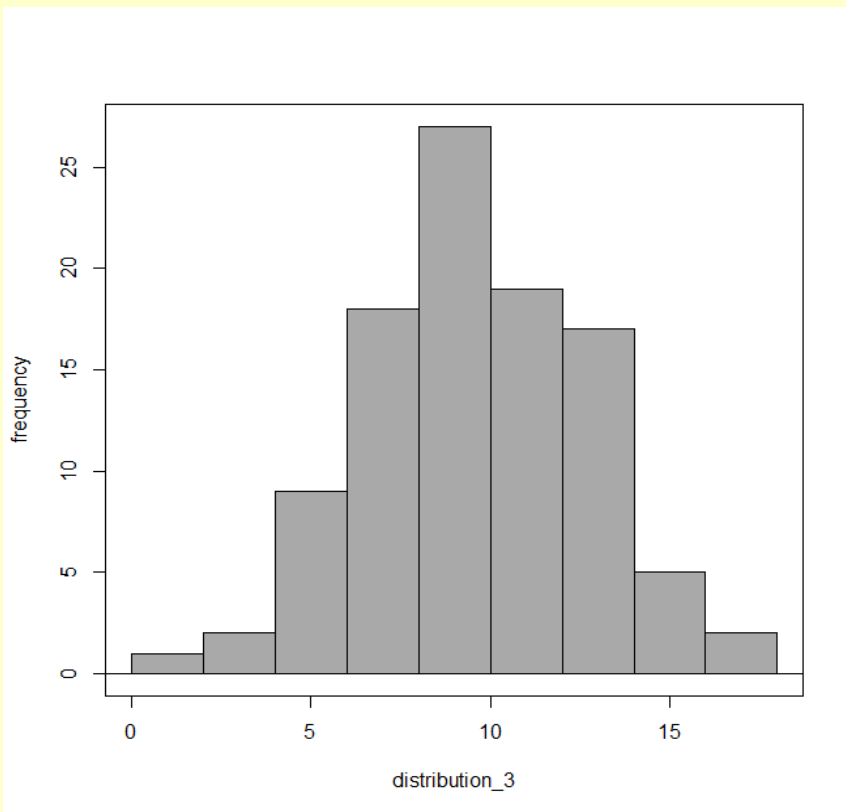
9 bins (default)

20 bins



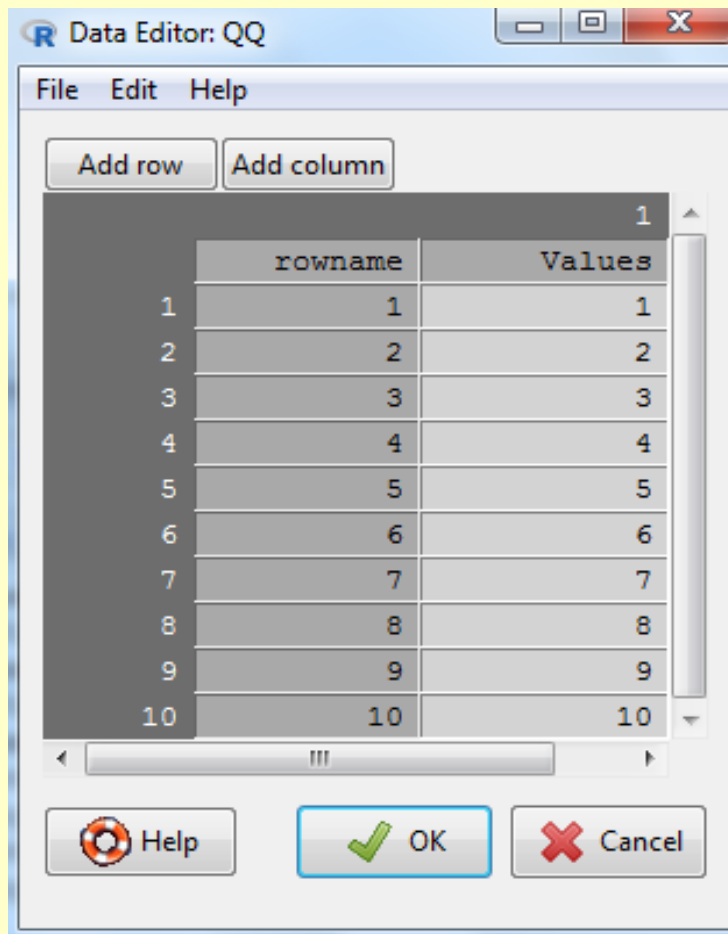
# PSA-2: Boxplots

- "Shape" of distribution does not depend on binning
- Highlights the percentiles and outliers



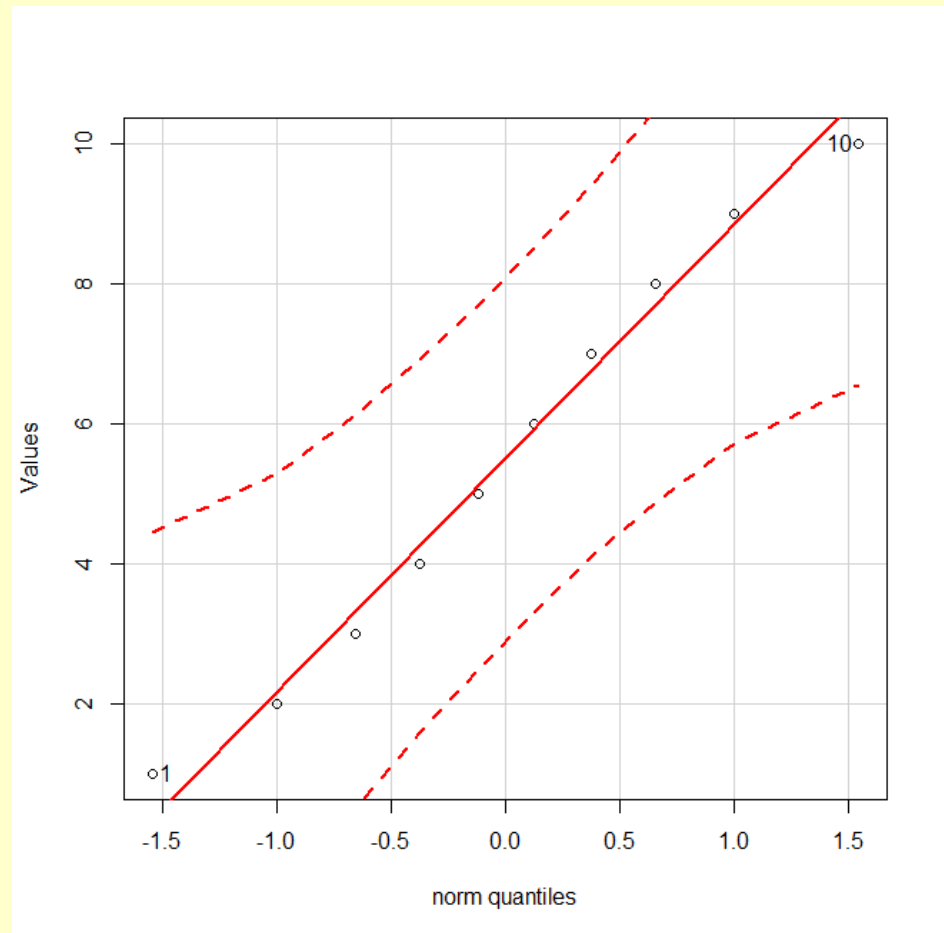
# PSA-3: Q-Q Plots

**NOTE:** Each value in the distribution is plotted.  
What are the x and the y coordinates?  
X axis are quartiles of "normal distribution"



The screenshot shows the R Data Editor window titled "Data Editor: QQ". It contains a table with 10 rows and 3 columns. The columns are labeled "rowname", "Values", and an unlabeled column with values 1 through 10. The data is as follows:

	rowname	Values	
			1
1	1	1	
2	2	2	
3	3	3	
4	4	4	
5	5	5	
6	6	6	
7	7	7	
8	8	8	
9	9	9	
10	10	10	



# PSA-3: Q-Q Plots

**NOTE:** What are the normalized quantiles?

```
Rcmdr>  
summary(QQ)
```

```
values
```

```
Min. : 1.00  
1st Qu.: 3.25  
Median : 5.50  
Mean : 5.50  
3rd Qu.: 7.75  
Max. : 10.00
```

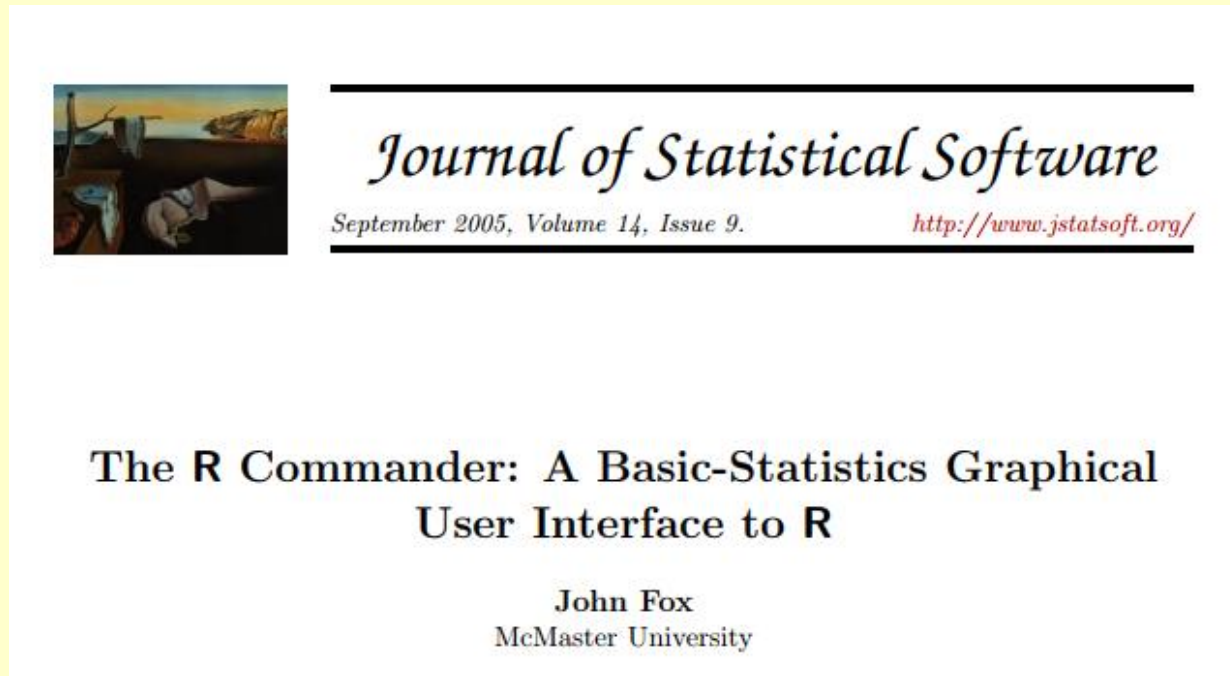
```
Rcmdr> numSummary  
(QQ[, "values", drop=FALSE],  
 statistics=c("IQR", "quantiles"),  
 quantiles=c(.05, 0.95))
```

```
IQR: 4.5  
5%: 1.45  
95%: 9.55  
N: 10
```

$$\text{IQR} = (3^{\text{rd}} \text{ Qu}) - (1^{\text{st}} \text{ Qu}): 7.75 - 3.25 = 4.5$$

# Graphs in Rcmdr

Package Rcmdr provides basic graphing tools (Fox 2005):

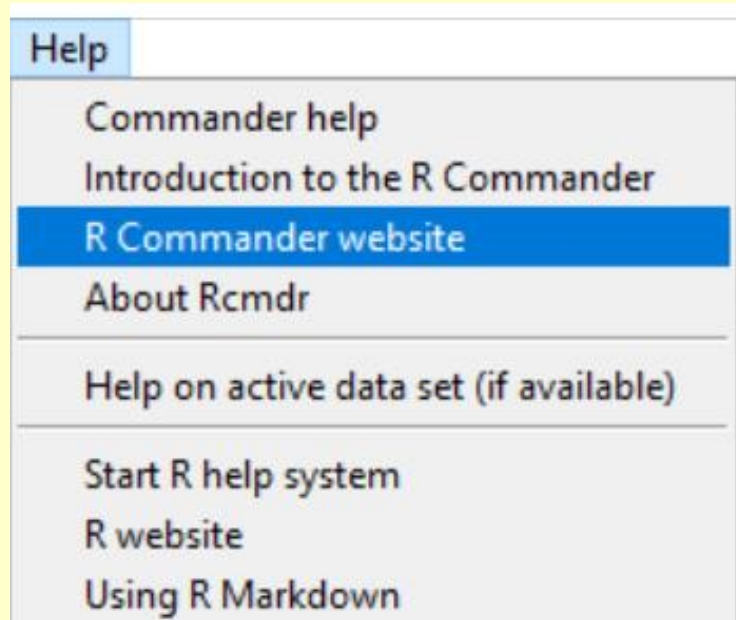


- Index plot
- Histogram
- Density plot
- Box plot
- Q-Q plots
- Scatterplots
- Line plots
- 3D Scatterplots

**Remember:** The help files for the current version of the Rcmdr package are available on the CRAN website at <http://CRAN.R-project.org/doc/packages/Rcmdr.pdf>.

# Help in Rcmdr

Package Rcmdr provides online help:



**Current Version: 2.6-x**

**John Fox**

---

**Please Read the R Commander Installation Notes!**

---



(click on the image for a larger view)

The image shows a screenshot of the Rcmdr interface. It features a scatter plot in the top-left pane, a data table in the bottom-right pane, and several smaller plots in the bottom-left pane. The interface has a blue and white color scheme.

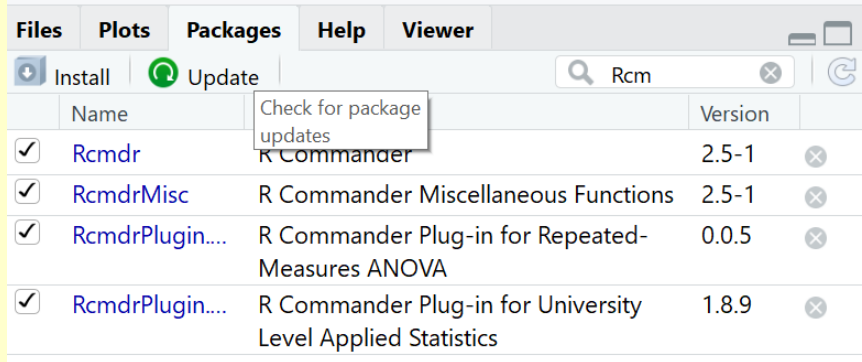
**Remember:**

Newer versions will be posted. Make sure you update regularly.

**Check your Version:**

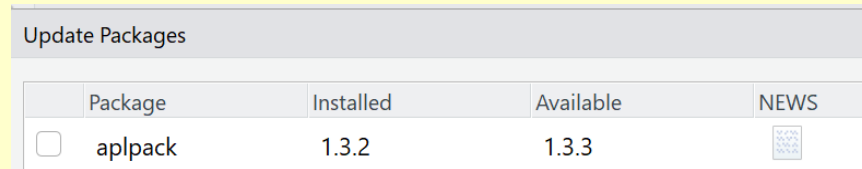
# Updating Rcmdr

## Update Packages in Rstudio:




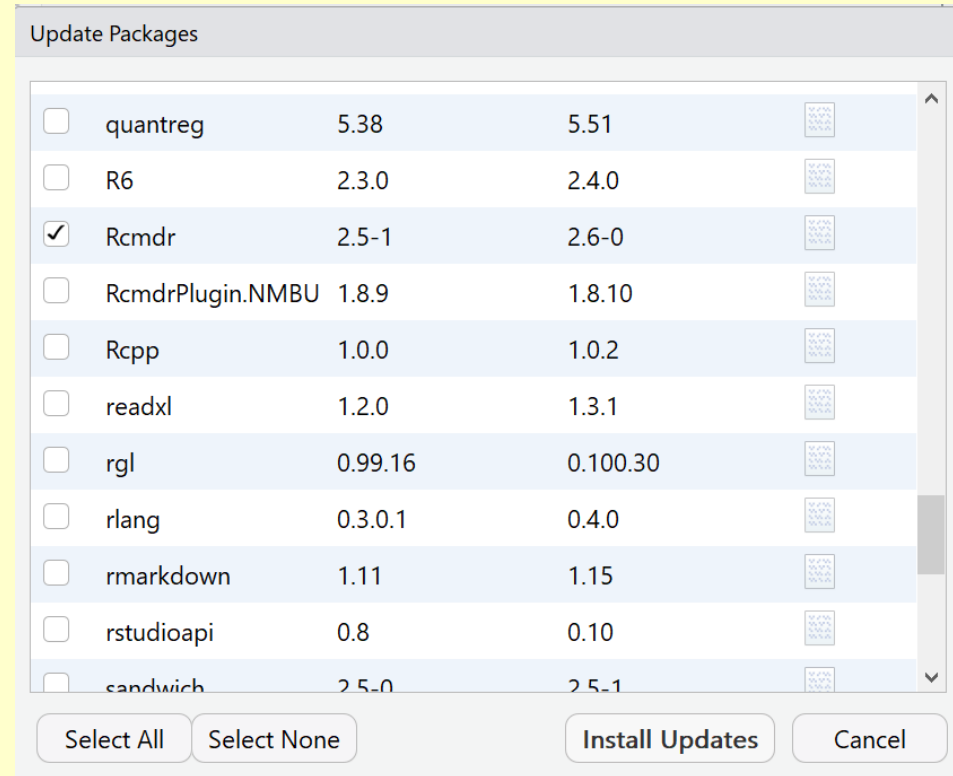
The screenshot shows the RStudio Packages pane with the 'Update' button highlighted. A tooltip for 'Check for package updates' is visible over the 'Update' button. The table below lists the installed and available versions for Rcmdr and its plugins.

Name	Version
Rcmdr	2.5-1
RcmdrMisc	2.5-1
RcmdrPlugin...	0.0.5
RcmdrPlugin...	1.8.9



The screenshot shows the 'Update Packages' dialog box with a table listing the 'aplpack' package. The table has columns for Package, Installed, Available, and NEWS.

Package	Installed	Available	NEWS
aplpack	1.3.2	1.3.3	



The screenshot shows the 'Update Packages' dialog box with a list of packages. The table below lists the installed and available versions for each package.

Package	Installed	Available
quantreg	5.38	5.51
R6	2.3.0	2.4.0
Rcmdr	2.5-1	2.6-0
RcmdrPlugin.NMBU	1.8.9	1.8.10
Rcpp	1.0.0	1.0.2
readxl	1.2.0	1.3.1
rgl	0.99.16	0.100.30
rlang	0.3.0.1	0.4.0
rmarkdown	1.11	1.15
rstudioapi	0.8	0.10
sandwich	2.5-0	2.5-1

## Update Options:

Compare Installed / Available versions.

Read about new features in NEWS.

Update packages manually or select "all" updates available.