

R Wizarday Final Project: Tutorial for using R commander (Rcmdr) and graphing plugins RcmdrPlugin.IPSUR, RcmdrPlugin.HH and RcmdrPlugin.FactoMindeR



Aprial 13, 2016

Siyu liang

Contents

[Introduction of R commander and Rcmdr plugins 1](#_Toc447974185)

[Why Rcmdr is useful for me? 1](#_Toc447974186)

[Background of my data 1](#_Toc447974187)

[Installation for Windows users 2](#_Toc447974188)

[Run Rcmdr 3](#_Toc447974189)

[Rcmdr 4](#_Toc447974190)

[Importing data 4](#_Toc447974191)

[Add a column for a computed variable 4](#_Toc447974192)

[Save data set 5](#_Toc447974193)

[Calculate Mean and SD 5](#_Toc447974194)

[Create a new dataset 7](#_Toc447974195)

[Stack data set 9](#_Toc447974196)

[Merge data sets 9](#_Toc447974197)

[Graphing 9](#_Toc447974198)

[RcmdrPlugin.IPSUR for simply plots 9](#_Toc447974199)

[RcmdrPlugin.HH for correlation 11](#_Toc447974200)

[RcmdrPlugin.FactoMineR for clustering 12](#_Toc447974201)

# Introduction of R commander and Rcmdr plugins

R Commander (Rcmdr) is a basic graphical user interface for R. It is ideal for users who are not experienced in coding as it allows users to run a lots of statistic tests and creates graphics without typing a line of code. By making a number of packages into plugins for Rcmdr, it provides additional functions to users who require more specialized analysis, such as plugin BCA for business and customer analytics and plugin DoE for industrial design of experiments.

## Why Rcmdr is useful for me?

I have background in molecular plant and my research focus on the proof of two protein interaction using various experimental approaches. Thus, my data has minimum requirement for statistical test (ie. Student’ t-test) but the statistic tests from Rcmdr are too complex for my purpose. However, as a long-term Excel user, I do find Rcmdr has a lot of similar basic functions in data re-arrangement and computing new variables etc. Also, I can directly import my data in .xlsx format instead of converting it into .csv for a general format for R. In addition, I do find there are more graphing options provided in Rcmdr. Although my data is relatively simple, I still find four types of plots using RcmdrPlugins and this enriches my plot type from just bars and lines ☺. Because it is very user friendly, and provide more graphing options, it is useful to generate plots.

This tutorial introduces the basic functions in data management using Rcmdr and graphing using RcmdrPlugin.IPSUR, RcmdrPlugin.HH and RcmdrPlugin.FactoMindeR.

# Background of my data

The uploaded drt.xlsx file contains raw data for a 9-day drought assay collecting the pot weight of 7 samples (5 replicates) over the test period. In order to test the drought response of these samples, need to, firstly, compute drought index, mean, and standard deviations; then, plot graphs.

# Installation for Windows users

Download R from <http://cran.us.r-project.org/> (click on “Download R for Windows” > “base” > “Download R 3.2.4 for Windows”)

Install R. Leave all default settings in the installation options.

Download RStudio from <http://rstudio.org/download/desktop>and install it. Leave all default settings in the installation options.

Run RStudio. Go to the “Packages” tab and click on “Install”.

An “Install Packages” window pops up. Type “Rcmdr” for packages and select it (should be the first one) from the drop down list. Ensure that “Install dependencies” is checked to allow downloading and installing missing dependencies, then click “Install”.

The installation would take a while around 2-3 mins. If a new window pop up asking if want to install all the parts of the R Commander packages, click yes, then additional packages are installed.

Install RcmdrPlugin.IPSUR the same way as Rcmdr installation.

# Run Rcmdr

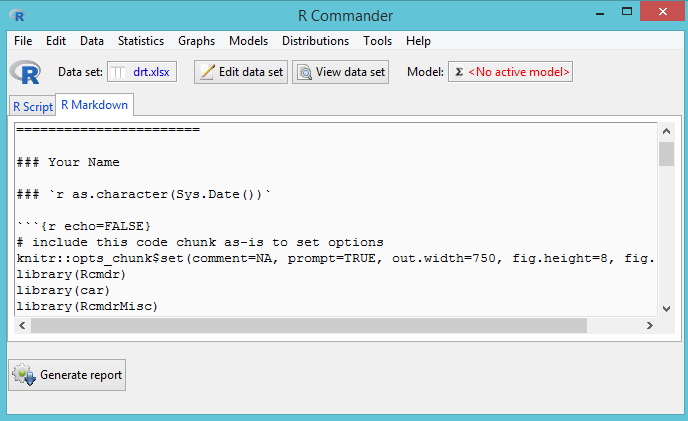
In RStudio, at the R console “ > “command prompt, type the following command and press the return key to launch Rcmdr:

Library(Rcmdr)

This step can be used to launch R commander each time. Or simply check the checkbox for Rcmdr in packages list.

Check mark RcmdrPlugins.IPSUR to run it.

Rcmdr GUI contains 2 windows, R Script and R Markdown. One can run line of code in R Script window. R Markdown records all the commands run on Rcdmr. The output shows up as console in RStudio. Clicking “Generate report” in R Markdown window will allow creating a report that shows all the commands and output.



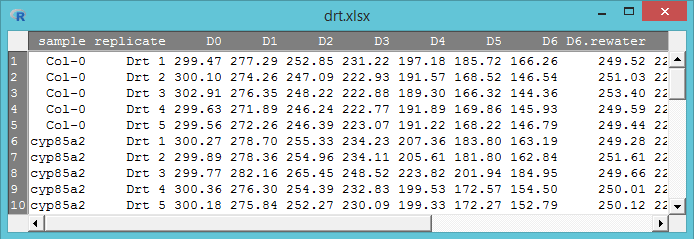
# Rcmdr

## Importing data

Make sure the name of the dataset contains no space otherwise Rcmdr can’t read it as a valid name.

Data -> import data -> from Excel file -> type dataset name ie. drt.xlsx -> select the file in a pop-up “open” window

View data set can show the content of the data set in a new window



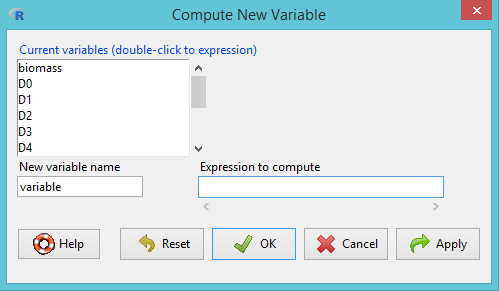
## Add a column for a computed variable

Data -> Manage variable in active dataset -> compute new variable -> in a pop-up window edit the following information

New variable name: DrtIndex (requires no space)

Expression to compute: total water loss / biomass (select by double-click expression)

Click ok.



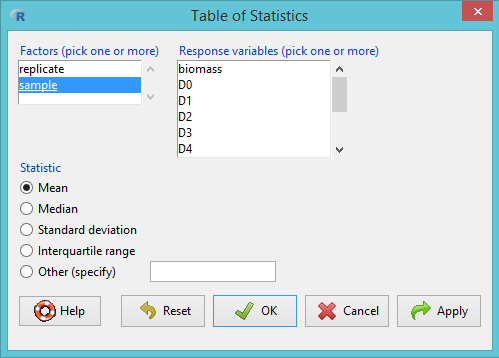
## Save data set

Data -> active data set -> save active data set -> drt\_DI (also can save as the same name and override the original data)

## Calculate Mean and SD

Import drt\_DI.xlsx. Once it is imported, tab “data set” shows the current active dataset is drt\_DI. When click on drt\_DI, a new window pops up showing all the active datasets that are imported. Can change the current active dataset by selecting any one on the list.

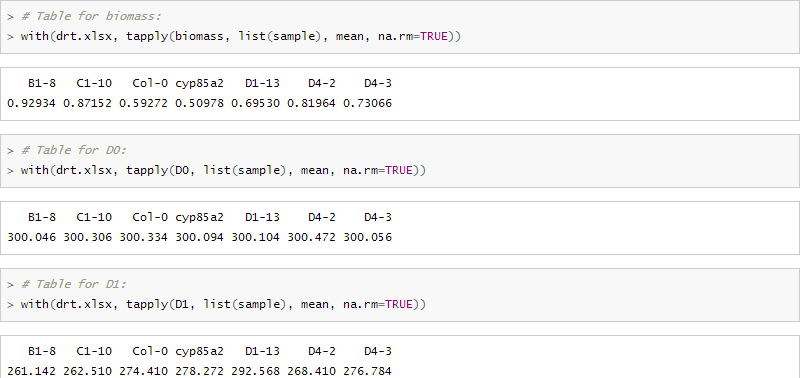
Statistics -> Summaries -> table of statistics -> select sample and all response variables -> mean



The result will show in RStudio console window. Alternatively, click “generate report” at the bottom of R markdown.

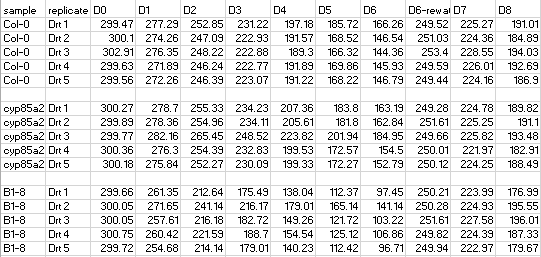
Repeat to calculate SD

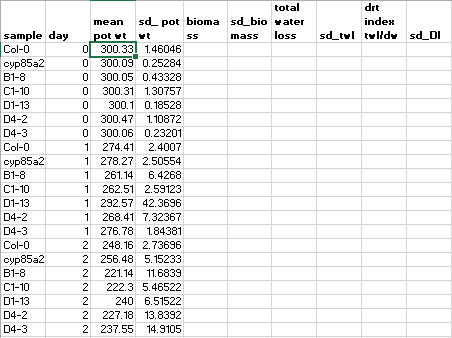
A report is generated in an e-Brower with mean by replicates and SD for each sample.



## Create a new dataset

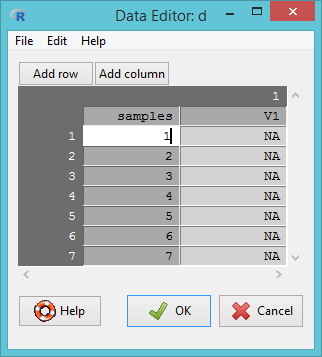
The original drt.xlsx will be used for statistic test. For graphing, I need a data set that shows the mean for pot weight of each day and create a new variable named “day”. Basically, I need to convert the D0~D8 from row into column.





Data -> new data set -> name it as drt\_day -> pop up a window with 2-by-2 table

Add 7 rows and change the first row name as “sample”



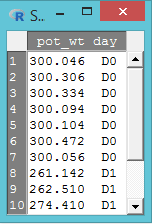
Copy sample names from the report and click on “1” in row “sample” then paste. Each sample name fills in each cell of the row.

Generate a 7-by-21 table filled with mean and sd for each day of pot weight.

Generate another new data set that contains the mean and SD for the rest of variables and name as drt\_rest.

## Stack data set

Data -> active data set -> stack variables in active data set -> in a pop up window, select D0-D8. Name it as stackedDay.



## Merge data sets

Data -> Merge data sets -> select drt\_day and drt\_rest to merge.

Under “Edit”, the cells, rows, and columns can be cut and pasted as an Excel table if needed.

Save the Merged data as drt\_2.xlsx

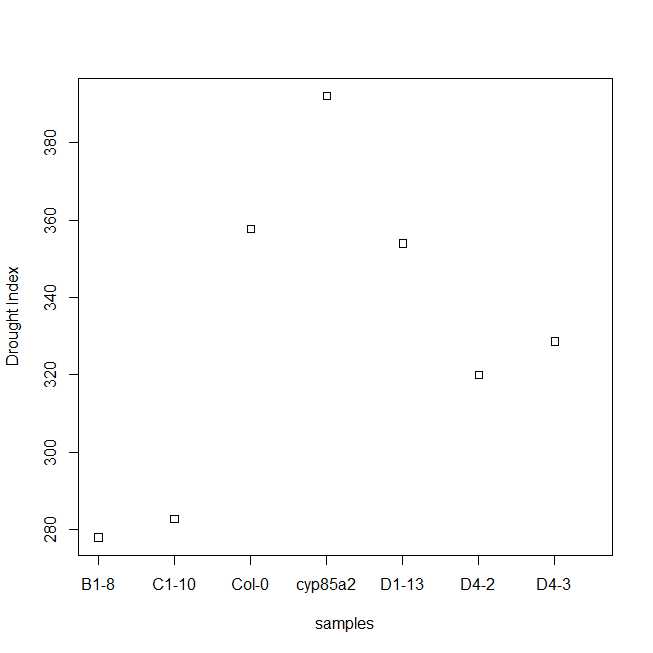
# Graphing

## RcmdrPlugin.IPSUR for simply plots

active set is drt\_2

Graph -> scatter (IPSUR) -> Select “sample” and “DrtIndex”

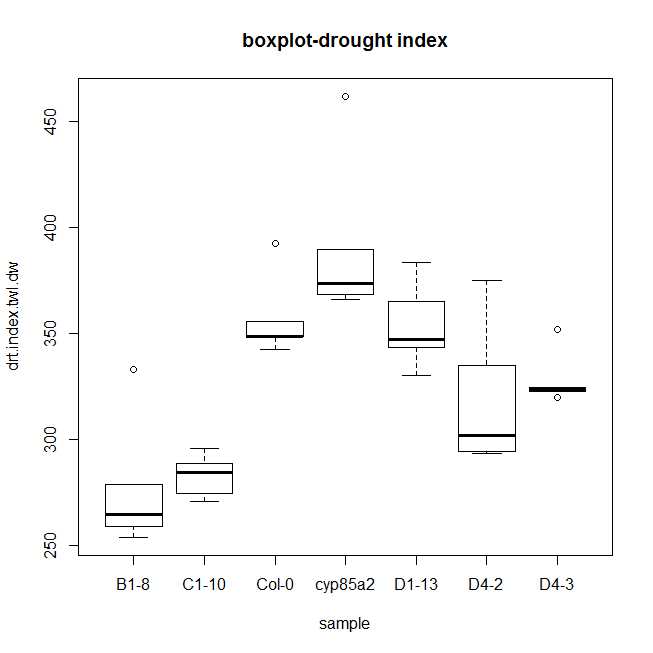
This shows the mean of drought index for each sample.



Change active set to drt.

Graph ->Box (IPSUR) -> Select “sample” and “DrtIndex”

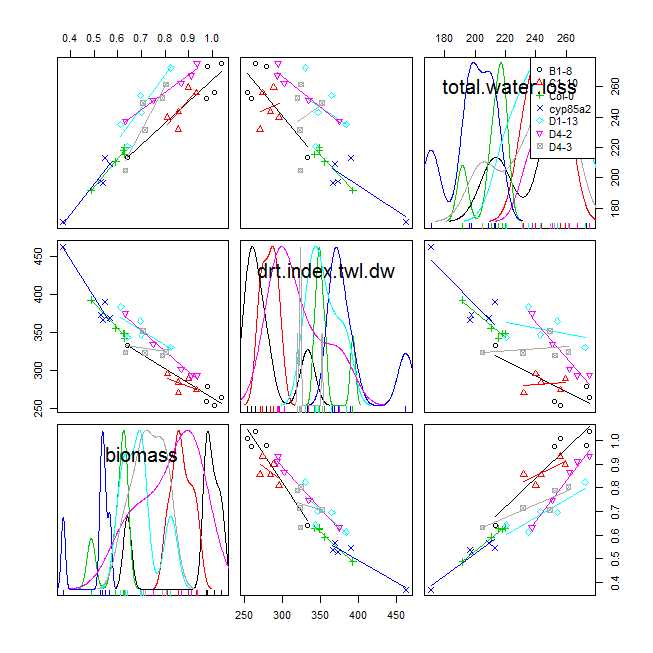
This shows the drought index for each sample with 5 replicates and the mean of drought index as the black bar.



## RcmdrPlugin.HH for correlation

Graph ->ScatterPlot Matrix (HH) -> Select “biomass”, “drtIndex”, and “total water loss”, grouping by “samples”

This plot matrix shows the correlation for a paired variables. The different colour/shape represents sample with 5 replicates for each sample.



## RcmdrPlugin.FactoMineR for clustering

FactoMineR tab -> PCA -> select “sample” as supplementary factor and keep everything else unchanged (so other parameters will choose all elements by default)

This plot shows 2 axes. The opposite side of one axis means the opposite response to drought condition. This plot clusters B1-8, C1-10 and D4-2 has similar response (more resistant to drought) whereas Col-0 and cyp85a2 shows the opposite response (more sensitive to drought).

This plot would be more powerful to look at many unrelated events and find similarity/differences in behaviour.

