

Deducer & RStudio

Mandy Vogel

May 31, 2015

Table of Contents

Deducer

- Installation

Ready to Use

- Run Deducer

- Load and Install Packages

Data

- Loading Data

- Data Viewer

- Examples

RStudio

- Installation

- Features

Table of Contents

Deducer

Installation

Ready to Use

Run Deducer

Load and Install Packages

Data

Loading Data

Data Viewer

Examples

RStudio

Installation

Features

Why Deducer?

Deducer is designed to be a free easy-to-use alternative to proprietary data analysis software such as SPSS, JMP, and Minitab. It has a menu system to perform common data manipulation and analysis tasks, and an excel-like spreadsheet in which to view and edit data frames. The goal of the project is two fold.

- Provide an intuitive graphical user interface (GUI) for R, encouraging non-technical users to learn and perform analyses without programming getting in their way.
- Increase the efficiency of expert R users when performing common tasks by replacing hundreds of keystrokes with a few mouse clicks. Also, as much as possible the GUI should not get in their way if they just want to do some programming.
- It may lower the entry threshold.

Why Not?

- Deducer is java-dependend and therefore sometimes not stable (although it has been a long time since I had problems, but I work very rarely with the deducer package; maybe it is more stable these days)
- R is designed for text based interactions, the full functionality is not available through menus
- the course will be based on typing the commands but maybe the Deducer GUI helps to overcome your inhibition to use R

Table of Contents

Deducer

Installation

Ready to Use

Run Deducer

Load and Install Packages

Data

Loading Data

Data Viewer

Examples

RStudio

Installation

Features

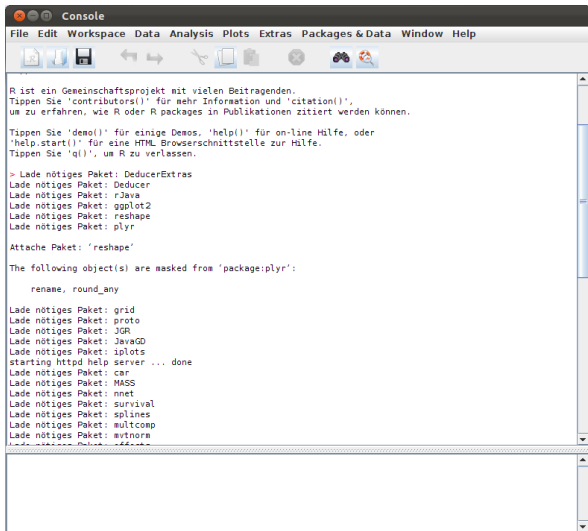
Installation

- there are instructions how to install at
<http://www.deducer.org/pmwiki/index.php?n=Main.DownloadingAn>

Run

- from within R:
 - run R
 - type `library(JGR)`
 - followed by `JGR()`
- there is also a script created during the installation; the path is shown when you start Deducer via R (e.g.
`/R/i686-pc-linux-gnu-library/2.14/JGR/scripts/run`

Load Packages I



```
Console
File Edit Workspace Data Analysis Plots Extras Packages & Data Window Help

R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

> Lade nötiges Paket: DeducerExtras
Lade nötiges Paket: Deducer
Lade nötiges Paket: rJava
Lade nötiges Paket: ggplot2
Lade nötiges Paket: reshape
Lade nötiges Paket: plyr

Attache Paket: 'reshape'

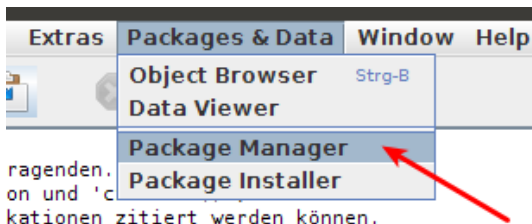
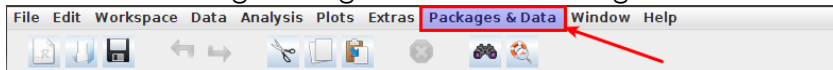
The following object(s) are masked from 'package:plyr':

  rename, round_any

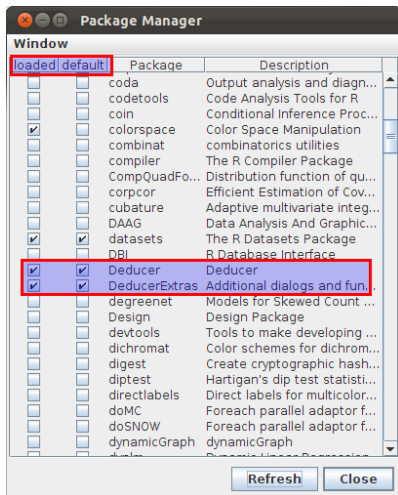
Lade nötiges Paket: grid
Lade nötiges Paket: proto
Lade nötiges Paket: JGR
Lade nötiges Paket: JavaGD
Lade nötiges Paket: iplots
starting httpd help server ... done
Lade nötiges Paket: car
Lade nötiges Paket: MASS
Lade nötiges Paket: nnet
Lade nötiges Paket: survival
Lade nötiges Paket: splines
Lade nötiges Paket: multcomp
Lade nötiges Paket: mvtnorm
Lade nötiges Paket: rffast
```

Load Packages II

choose Package Manager from menu Packages & Data



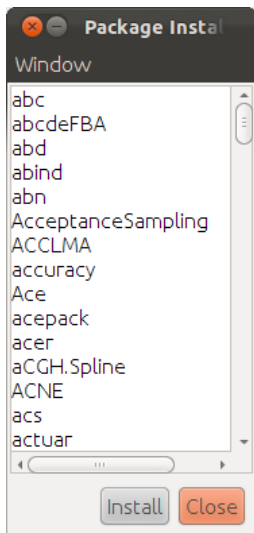
Package Manager



Now you can choose the packages you want to load for the current session and those you want to load by default each time.

The packages Deducer and DeducerExtra should be chosen as default.

Package Installer



The package installer can also be found in the menu Packages & Data

Additional Deducer Packages

R-Forge offers a central platform for the development of R packages, R-related software and further projects. There are three additional packages for Deducer. You can install them by typing:

```
install.packages(c("DeducerRichOutput",  
                  "DeducerANOVA",  
                  "DeducerPSY220"),  
                repos="http://R-Forge.R-Project.org")
```

DeducerRichOutput

without DeducerRichOutput

```
> one.sample.test(variables=d(Dose),
+ data=x,
+ test=t.test,
+ alternative="two.sided")
```

One Sample t-test

	mean of x	95% CI Lower	95% CI Upper	t	df	p-value
Dose	4.625833	4.502193	4.749474	74.01293	131	9.489402e-109
HA:	two.sided					
H0:	mean = 0					

with DeducerRichOutput

```
> one.sample.test(variables=d(Dose),
+ data=x,
+ test=t.test,
+ alternative="two.sided")
```

One-Sample Test

Method: One Sample t-test

	<i>mean of x</i>	<i>95% CI Lower</i>	<i>95% CI Upper</i>	<i>t</i>	<i>df</i>	<i>p-value</i>
Dose	4.63	4.50	4.75	74.01	131.00	<0.001

Notes:

HA: two.sided
H0: mean=0

Table of Contents

Deducer

Installation

Ready to Use

Run Deducer

Load and Install Packages

Data

Loading Data

Data Viewer

Examples

RStudio

Installation

Features

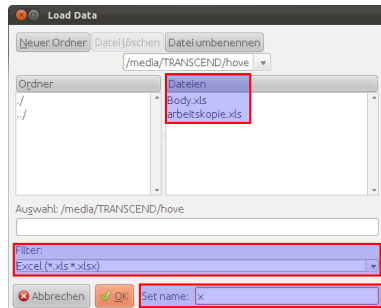
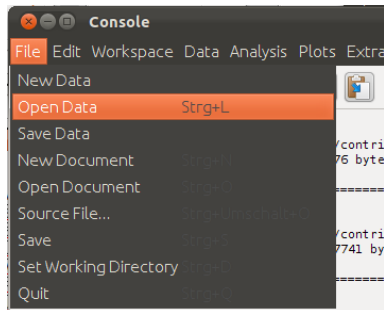
data types Deducer can handle

The load data menu can handle the following data types

File Type	Extension
R workspace	*.rda and *.rdata
R object	*.robj
Comma seperated	*.csv
Text file	*.txt
SPSS	*.sav
SAS export	*.xpt
DBase	*.dbf
Stata	*.dta
Systat	*.sys and *.syd
ARFF	*.arff
Epiinfo	*.rec
Minitab	*.mtp
S data dump	*.s3
Excel	*.xls,*.xlsx

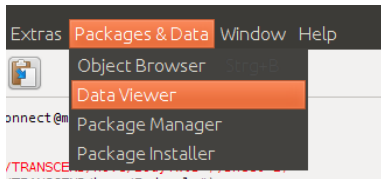
Loading Data

It is pretty easy: via the menu **File** → **Open Data**



Open the Data Viewer

The data viewer provides an easy to use, spreadsheet-like environment to view and edit data. Copy and pasting is supported, and is compatible with Excel 2003-2007, so data can be moved from Excel to R by simply copying it to the data viewer. Contextual menus are used to insert, delete and copy rows and columns.



The Data Viewer - Data View

Data Viewer

File Edit Help

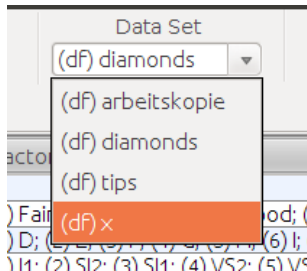
Data Set
(df) diamonds

Data View Variable View

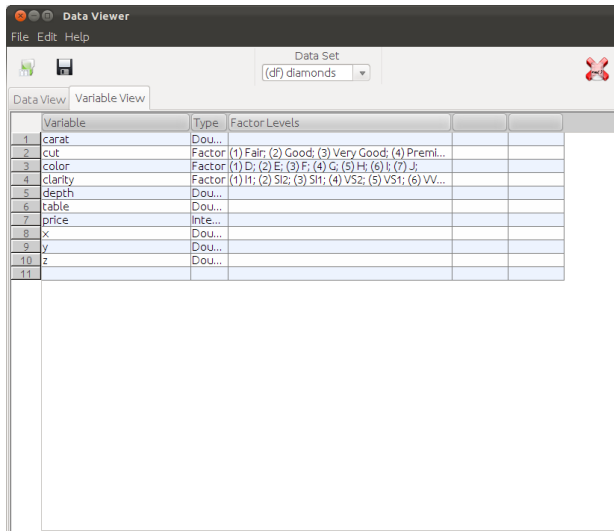
	carat	cut	color	clarity	depth	table	price	x	y	z
7	0.24	Very Good	I	VVS1	62.3	57.0	336	3.95	3.98	2.41
8	0.26	Very Good	H	SI1	61.9	55.0	337	4.07	4.11	2.51
9	0.22	Fair	E	VS2	65.1	61.0	337	3.87	3.78	2.43
10	0.23	Very Good	H	VS1	59.4	61.0	338	4.0	4.05	2.31
11	0.3	Good	J	SI1	64.0	55.0	339	4.25	4.28	2.71
12	0.23	Ideal	J	VS1	62.8	56.0	340	3.93	3.9	2.41
13	0.22	Premium	F	SI1	60.4	61.0	342	3.88	3.84	2.31
14	0.31	Ideal	J	SI2	62.2	54.0	344	4.35	4.37	2.71
15	0.2	Premium	E	SI2	60.2	62.0	345	3.79	3.75	2.21
16	0.32	Premium	E	I1	60.9	58.0	345	4.38	4.42	2.61
17	0.3	Ideal	I	SI2	62.0	54.0	348	4.31	4.34	2.61
18	0.3	Good	J	SI1	63.4	54.0	351	4.23	4.29	2.71
19	0.3	Good	J	SI1	63.8	56.0	351	4.23	4.26	2.71
20	0.3	Very Good	J	SI1	62.7	59.0	351	4.21	4.27	2.61
21	0.3	Good	I	SI2	63.3	56.0	351	4.26	4.3	2.71
22	0.23	Very Good	E	VS2	63.8	55.0	352	3.85	3.92	2.41
23	0.23	Very Good	H	VS1	61.0	57.0	353	3.94	3.96	2.41
24	0.31	Very Good	J	SI1	59.4	62.0	353	4.39	4.43	2.61
25	0.31	Very Good	J	SI1	58.1	62.0	353	4.44	4.47	2.51
26	0.23	Very Good	G	VVS2	60.4	58.0	354	3.97	4.01	2.41
27	0.24	Premium	I	VS1	62.5	57.0	355	3.97	3.94	2.41
28	0.3	Very Good	J	VS2	62.2	57.0	357	4.28	4.3	2.61
29	0.23	Very Good	D	VS2	60.5	61.0	357	3.96	3.97	2.41
30	0.23	Very Good	F	VS1	60.9	57.0	357	3.96	3.99	2.41
31	0.23	Very Good	F	VS1	60.0	57.0	402	4.0	4.03	2.41
32	0.23	Very Good	F	VS1	59.8	57.0	402	4.04	4.06	2.41
33	0.23	Very Good	E	VS1	60.7	59.0	402	3.97	4.01	2.41
34	0.23	Very Good	E	VS1	59.5	58.0	402	4.01	4.06	2.41
35	0.23	Very Good	D	VS1	61.9	58.0	402	3.92	3.96	2.41

The Data Viewer - Data View 2

- a right click on the row or column headers
 - allows one to insert, copy and delete columns and rows
 - sort by one column
- you can also edit the data
- in the drop down menu Data Set you can choose the data frame



The Data Viewer - Variable View



	Variable	Type	Factor Levels		
1	carat	Dou...			
2	cut	Factor	(1) Fair; (2) Good; (3) Very Good; (4) Premi...		
3	color	Factor	(1) D; (2) E; (3) F; (4) G; (5) H; (6) I; (7) J;		
4	clarity	Factor	(1) I1; (2) SI2; (3) SI1; (4) VS2; (5) VS1; (6) VV...		
5	depth	Dou...			
6	table	Dou...			
7	price	Inte...			
8	x	Dou...			
9	y	Dou...			
10	z	Dou...			
11					

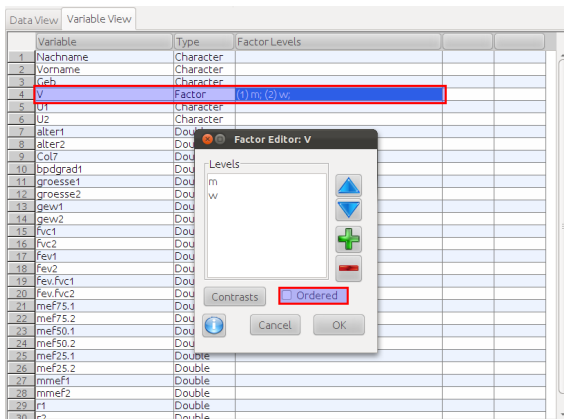
The Data Viewer - Variable View 2

In the variable view The `variable` column represents the variable name. The `type` column determines the storage type.

- the properties of each variable in the data frame can be edited
- the type column determines the storage type; variables can be stored as
 - Strings (character)
 - Doubles (Numeric)
 - Integers
 - Logicals (yes/no) or
 - Factors
- The levels of Factors are displayed in the 'Factor Levels' column, and can be edited by clicking on the appropriate cell, which brings up the Factor Editor

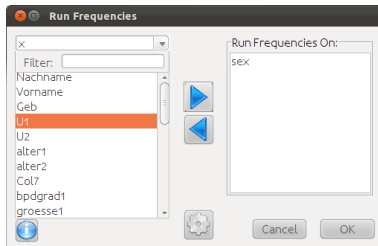
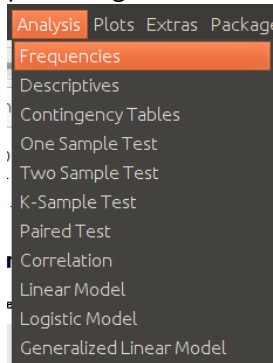
The Data Viewer - Variable View 3

The levels of Factors are displayed in the 'Factor Levels' column, and can be edited by clicking on the appropriate cell, which brings up the Factor Editor.



Frequencies

Frequency tables provide descriptive information for categorical and ordinal variables. They display the number of cases that fall into each category of a specific variable, as well as calculate percentages.



Frequencies Output

With DeducerRichOutput

Frequencies

Frequencies (sex)

	<i>Value</i>	<i># of Cases</i>	<i>%</i>	<i>Cumulative %</i>
1	m	16	42.10	42.10
2	w	22	57.90	100.00

Case Summary (sex)

	<i>Valid</i>	<i>Missing</i>	<i>Total</i>	<i>% Missing</i>
1	38.00	0.00	38.00	0.00

Frequencies (bpdgrad1)

	<i>Value</i>	<i># of Cases</i>	<i>%</i>	<i>Cumulative %</i>
1	0	17	44.70	44.70
2	1	21	55.30	100.00

Case Summary (bpdgrad1)

	<i>Valid</i>	<i>Missing</i>	<i>Total</i>	<i>% Missing</i>
1	38.00	0.00	38.00	0.00

Without DeducerRichOutput

```
> frequencies(x[,c("sex","bpdgrad1")], r.digits = 1)
```

```
$sex
```

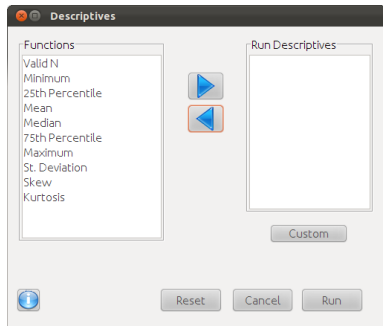
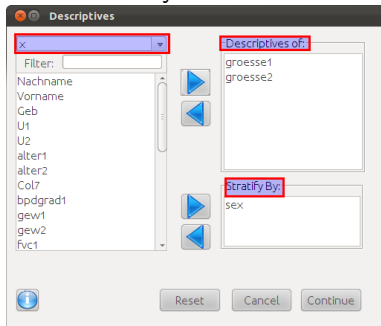
```
-----
..                               Frequencies                               ..
..                               ..                                       ..
..  Value # of Cases      % Cumulative %                               ..
1    m           16      42.1         42.1
2    w           22      57.9        100.0
..                               ..                                       ..
..                               Case Summary                               ..
..                               ..                                       ..
..                               Valid Missing Total                               ..
# of cases    38      0      38
..                               ..                                       ..
..                               ..                                       ..
-----
```

```
$bpdgrad1
```

```
-----
..                               Frequencies                               ..
..                               ..                                       ..
..  Value # of Cases      % Cumulative %                               ..
1    0           17      44.7         44.7
2    1           21      55.3        100.0
..                               ..                                       ..
..                               Case Summary                               ..
..                               ..                                       ..
..                               Valid Missing Total                               ..
# of cases    38      0      38
..                               ..                                       ..
..                               ..                                       ..
-----
```

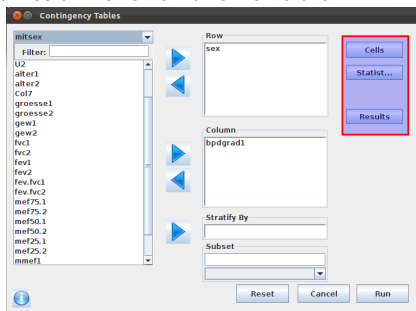
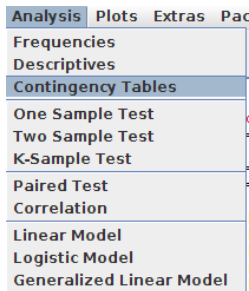
Descriptives

Calculates descriptive statistics for a set of variables. Possibly stratified by another set of variables.



Contingency Tables

Contingency tables (sometimes called crosstabs) are used to summarize and analyze the joint distribution of two variables, possibly stratified by a third. A table of observation counts will be created for each combination of the variables in the row list and each variable in the column list. If a stratum variable is specified, separate tables are created for each level of the variable.



Contingency Tables - Cells

In addition to observation counts, there are a number of additional cell values that can be displayed.

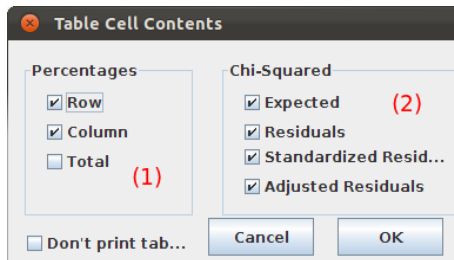
1. Percentages

- 1.1 Row - Percentage in cell out of observations within each row
- 1.2 Column - Percentage in cell out of observations within each column
- 1.3 Total - Percentage in cell

2. χ^2 -test

- 2.1 Expected - The expected count of the cell if there were no relationship between the two variables
- 2.2 Residuals - The observed count minus the expected count.
- 2.3 Standardized residuals - The residuals standardized such that (if the two variables were independent) they have mean 0 and standard deviation 1. These residuals are useful in determining which cells of a contingency table contribute most to a significant χ^2 test.
- 2.4 Adjusted residuals - These adjust the residuals by the row and column totals.

Contingency Tables - Cells



Contingency Tables - Stats

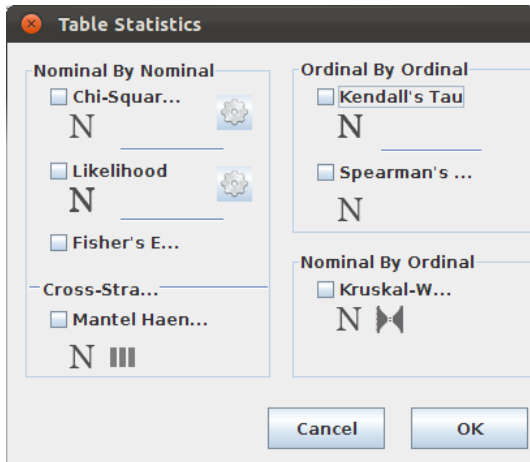


Table of Contents

Deducer

Installation

Ready to Use

Run Deducer

Load and Install Packages

Data

Loading Data

Data Viewer

Examples

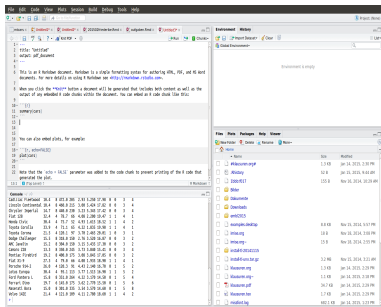
RStudio

Installation

Features

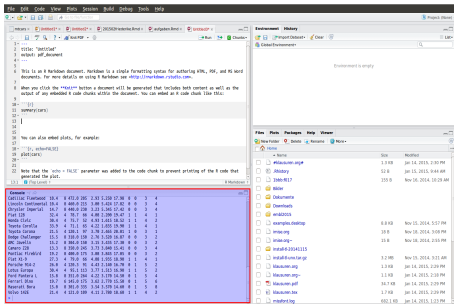
Getting R-Studio

RStudio is a free and open source integrated development environment (IDE) for R. You can run it on your desktop (Windows, Mac, or Linux) or even over the web using RStudio Server. Available at <http://rstudio.org/> (install R first)



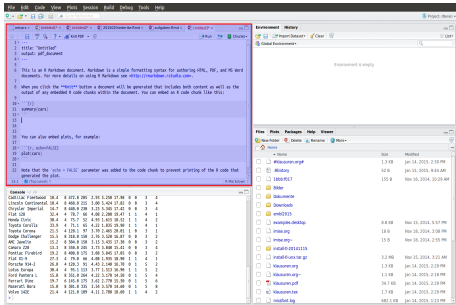
RStudio - Features

- integration of the R console



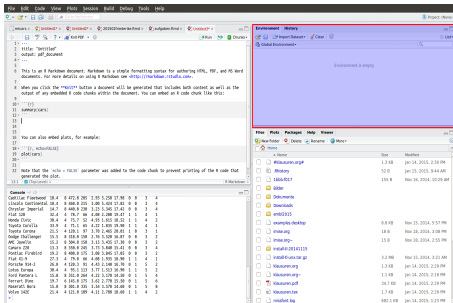
RStudio - Features

- code execution (Ctrl + Enter)
- different file types (mark down, c++, html, r)
- data viewer



RStudio - Features

- object browser
- history browser



Why RStudio?

- under the hood RStudio is a customized mozilla firefox so it is stable on Windows, Linux, and Mac operating systems
- it is optimized for text based interactions with R
- it provides rich facilities to integrate documentation into analyses (reproducible research), results can be exported to (la)-tex (and further to pdf), html, and even MS Word
- RStudio is highly supported by industry