

GSERM - Oslo

R (and \LaTeX)

Materials:

<https://github.com/PrisonRodeo/GSERM-Oslo-2018-git>

January 16, 2018

- R


- “R is a free software environment for statistical computing and graphics.”
- The R Project: <https://www.r-project.org/>
- Comprehensive R Archive Network (CRAN):
<https://cran.r-project.org/>

- RStudio

- A free, open-source GUI for R
- Website: <https://www.rstudio.com/>

R:

- Is an **object-oriented** language
- Is made up of:
 - Objects
 - Functions
 - Classes (of objects and functions)
- Is **Turing complete**
- Is **modular**
 - User-created **packages**
 - Organized into **task views**
(<https://cran.r-project.org/web/views/>)
- Runs on UNIX/Linux/OS-X/Windows



The image shows a screenshot of the R Console window on a macOS system. The title bar at the top reads "R Console". Below the title bar is a menu bar with standard macOS icons (stop, back, forward, search, etc.). The main content area displays the R version 3.3.1 (2016-06-21) -- "Bug in Your Hair" and copyright information. It also shows the R license text, which states that R is free software and comes with absolutely no warranty. The console output includes the R version, copyright, platform, and the R license text. The prompt is "> |".

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

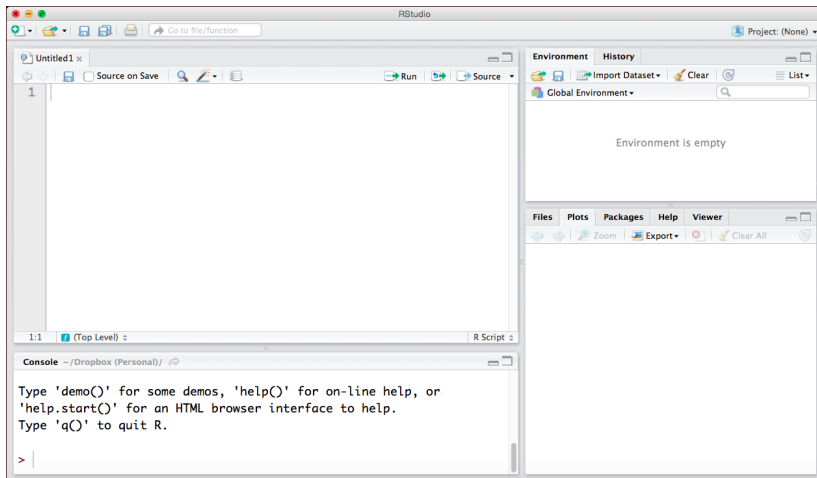
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.68 (7238) x86_64-apple-darwin13.4.0]

[Workspace restored from /Users/cuz101cl/.RData]
[History restored from /Users/cuz101cl/.Rapp.history]

> |
```



RStudio (annotated)

Source window:

- Click here to save your source code. Save often!
- Source on Save
- Run

This is the "Source" window.

- It's the place where you'll type the code that will then be sent to R.
- It's basically a text editor. You can open text files of any kind here if you want.
- Files that appear here end in (and should be saved with) the extension ".R" (as in "MyCode.R").

You'll spend most of your time working here.

Environment window:

This is the "Environment" window. It is where you can find all the various "objects" that you create, grouped by object type (data frames, lists, graphs, etc.). Environment is empty.

There's also a "History" tab above; switching to that will show what has transpired in the Console window recently.

Console window:

This is the "working directory." Anything you save will be saved here, unless you tell the program to save it somewhere else.

Console - /Dropbox (Personal) /

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

This is the "Console." When you run the code in the Source window, the results that aren't graphics appear here.

Files window:

This is a window that shows various other things. Those things are tabbed above ^ and include:

- Plots (graphs) that you have created
- Packages that are loaded
- Help results (obtained by typing "?XXX" in the Console window, e.g. "?table").

This:

```
> table(df$X)
```

... means “Type the phrase ‘table(df\$X)’ on the command line,” or – equivalently – “Type the phrase ‘table(df\$X)’ into your Source code, and then run it.”

More often, you'll see:

```
with(df, plot(Y~X,pch=19,col="red")) # draw a scatterplot
abline(h=0,lty=2) # add a horizontal line at zero
abline(v=0,lty=2) # add a vertical line at zero
text(df$X,df$Y,labels=df$names,pos=1) # add labels
```

... which means “Put this block of text into your Source code, and then run it.”

Note:

- R / RStudio ignores line breaks
- Anything to the right of a “#” is a comment

Very basic R examples...

(see `GSERM-Oslo-2018-R-Intro.R` in the github repo)

Help For Learning R(Studio)

In rough order of preference:

- Quick-R (<http://www.statmethods.net/>)
- The “Level-Zero” R Tutorial (doesn't integrate RStudio, but is otherwise very good)
- [Statistics with R](#)
- The [Do It Yourself Introduction to R](#)
- Also be sure to consult the AccessLex “Useful R Resources” guide (on GitHub).

Example Data: Infant Mortality

```
> url <- getURL("https://raw.githubusercontent.com/PrisonRodeo/
  GSERM-Oslo-2018-git/master/Data/CountryData2000.csv")
> Data <- read.csv(text = url) # read the "Country" data
> rm(url)
>
> # Summary statistics
>
> # install.packages("psych") <- Install psych package, if necessary
> library(psych)

> with(Data, describe(infantmortalityperK))
  vars    n mean    sd median trimmed   mad min max range skew kurtosis   se
1     1 179 43.83 40.39     29   38.38 34.26  2.9 167 164.1     1     0.06 3.02

> with(Data, describe(DPTpct))
  vars    n mean    sd median trimmed   mad min max range skew kurtosis   se
1     1 181 81.71 19.77     90   85.23 11.86  24  99   75 -1.31     0.57 1.47
```

OLS Regression

```
> IMDPT<-lm(infantmortalityperK~DTPpct,data=Data,na.action=na.exclude)
> summary.lm(IMDPT)
```

Call:

```
lm(formula = infantmortalityperK ~ DTPpct, data = Data)
```

Residuals:

Min	1Q	Median	3Q	Max
-56.801	-16.328	-5.105	11.777	86.590

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	173.2771	8.4893	20.41	<2e-16 ***
DTPpct	-1.5763	0.1009	-15.62	<2e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 26.19 on 175 degrees of freedom

(14 observations deleted due to missingness)

Multiple R-squared: 0.5824, Adjusted R-squared: 0.58

F-statistic: 244.1 on 1 and 175 DF, p-value: < 2.2e-16

Analysis of Variance

```
> anova(IMDPT)
```

```
Analysis of Variance Table
```

```
Response: infantmortalityperK
```

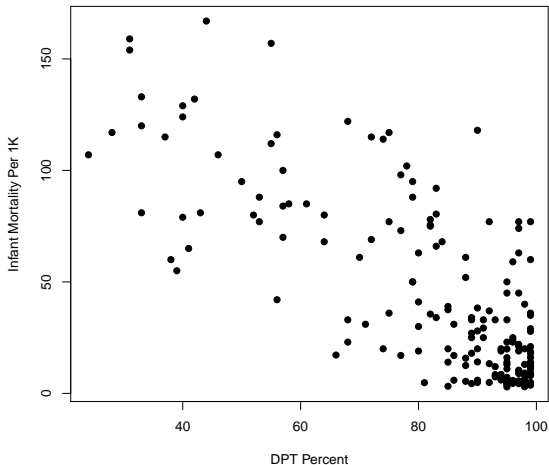
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
DPTpct	1	167423	167423	244.09	< 2.2e-16 ***
Residuals	175	120033	686		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A Basic Scatterplot

```
> with(Data,  
  plot(DPTpct,infantmortalityperK,pch=19,  
    xlab="DPT Percent",ylab="Infant Mortality Per 1K"))
```



R and L^AT_EX

- Sweave “...enables the embedding of R code within LaTeX documents to generate a PDF file that includes narrative and analysis, graphics, code, and the results of computations.”
- knitr is an R package that expands the capabilities of Sweave.
- Key point: **Integrates data analysis and document creation.**
 - Analysis and text are in the same document
 - Documents are *dynamic*: When the data changes, the document changes.