

Statistics Bootcamp using R

DAY 1 INTRODUCTION TO STATISTICS IN BUSINESS

1.2 INTRODUCTION TO R

GU Zhan (Sam)
Institute of Systems Science
National University of Singapore

issgz@nus.edu.sg

Agenda

Day 1: Introduction to Statistics in Business

- Basic Vocabulary of Statistics & Data Types
- **Introduction to R**
- Data Collection & Summarization

Learning objectives

- Understand what R can do
- Understand R studio interface
- Run simple R commands
- Understand the common mistakes

What is R?

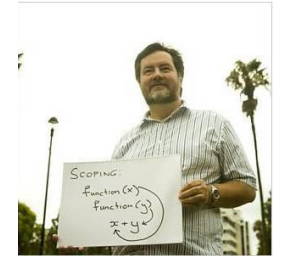
project considered in 1992

initial version released in
1995

stable beta version in
2000

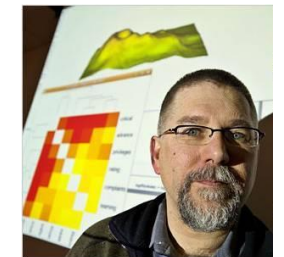


named after first names of
first 2 authors



was created at Auckland
New Zealand

was created by
Ross Ihaka &
Robert Gentleman



Source: <https://data-flair.training/blogs/r-interview-questions-and-answers/>

Why should we use R?

A Open source
programming language

C R groups are noted for its
energetic contributions

B Serves as a glue
language

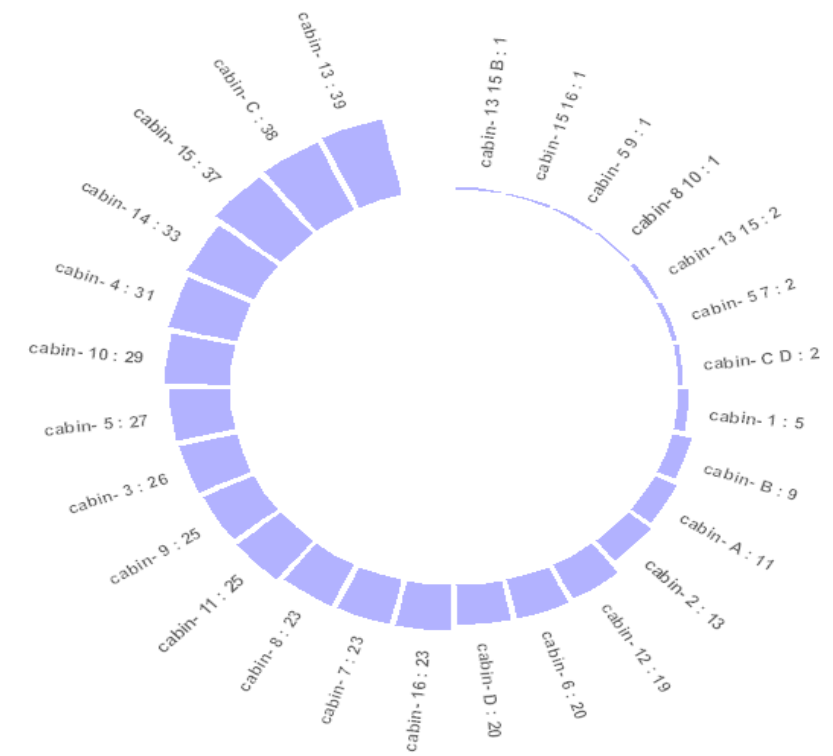
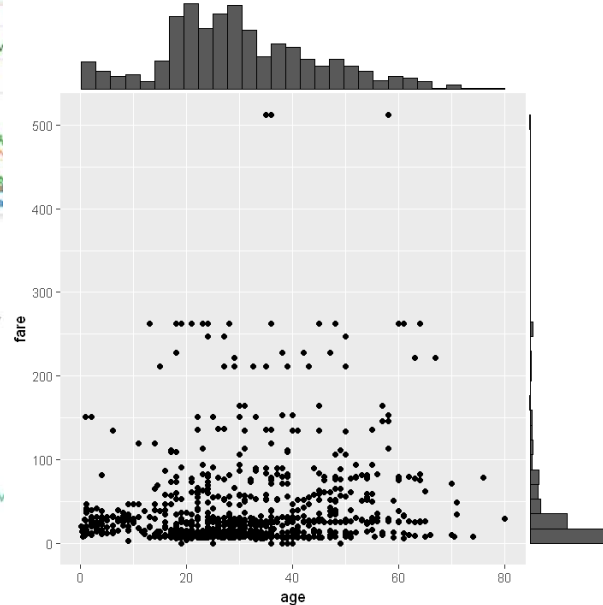
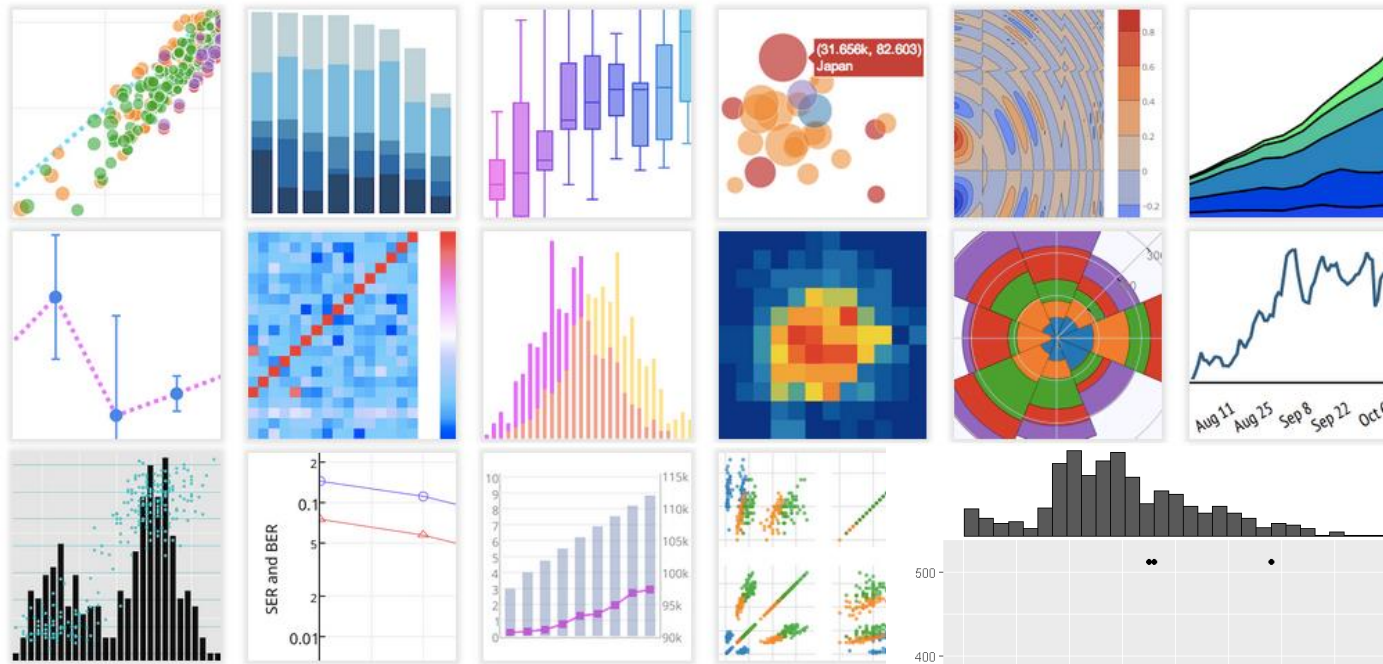
D R has gotten faster over
time

Source: <https://data-flair.training/blogs/r-interview-questions-and-answers/>

Other equivalent software : **Commercial:** SAS, SPSS, JMP etc.

Open Source: Python etc.

R has great graphic capabilities



How applicable is R?

A R is used in wide range of industry such as finance, e-commerce, semiconductors, manufacturing and many more

B Considered as an alternate execution of science

C R is one of the most prevalent language though Python is also gaining popularity

D R helps in data importing and cleaning

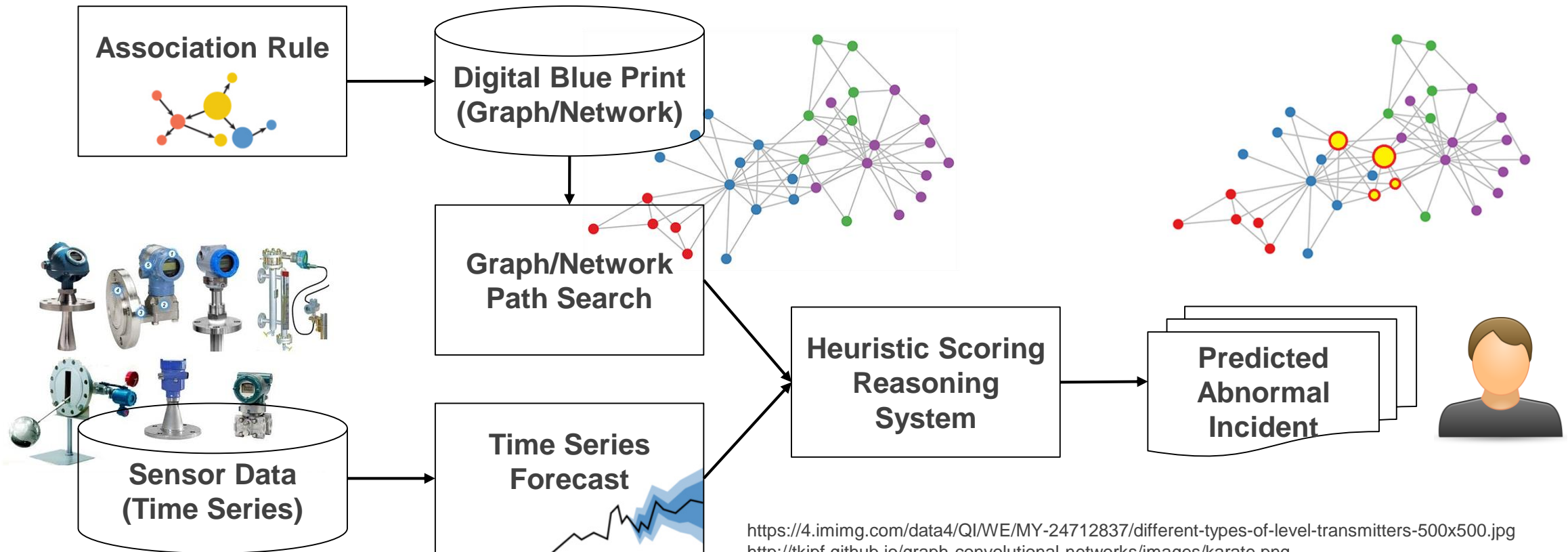
E R is one of the best open source tools for data science today

Source: <https://data-flair.training/blogs/r-interview-questions-and-answers/>

How applicable is R?

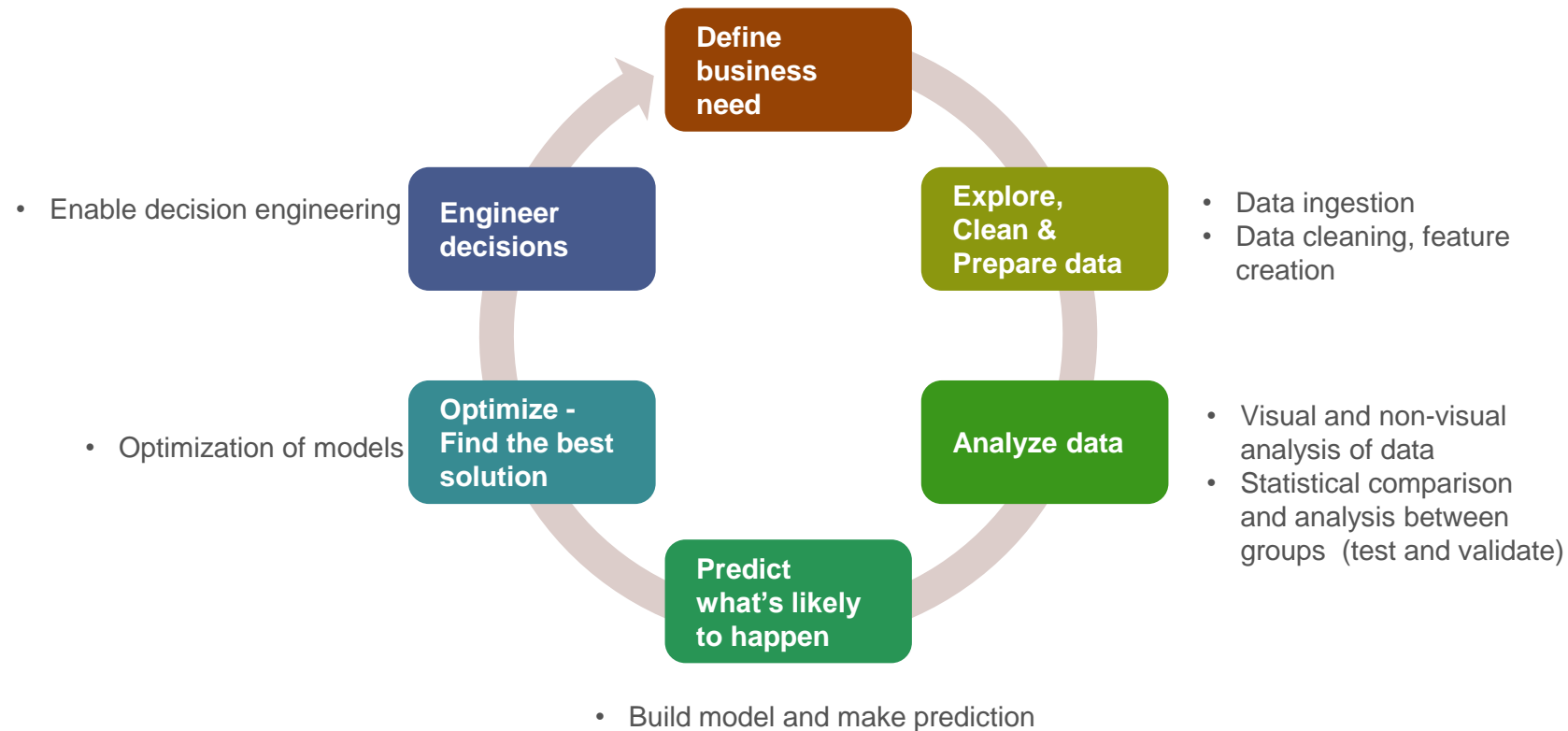
Use case: Plant Abnormal Situation Prediction (Predictive Maintenance)

Sub-systems: Time Series Forecast, Association Rule, Graph Search, Heuristic Score



<https://4.imimg.com/data4/QI/WE/MY-24712837/different-types-of-level-transmitters-500x500.jpg>
<http://tkipf.github.io/graph-convolutional-networks/images/karate.png>
<https://www.microsoft.com>

R is a vital part of the analytics process life-cycle



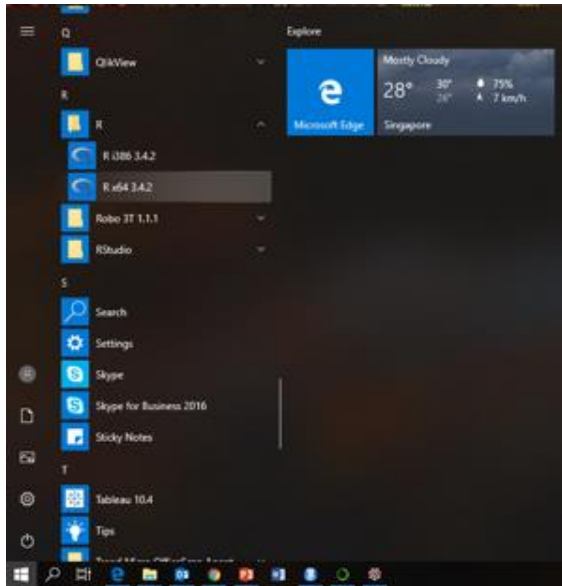
Obtaining R

- Freely available from the Comprehensive R Archive Network (CRAN)

- Downloadable at <http://CRAN.R-project.org/mirrors.html>

- Runs on Windows, Mac and Linux


- To access the installed R, for windows user, go to *Start Menu*, select '*All apps*', go to folder '*R*', click '*R x64 3.x.x*'



** It is possible to run R on USB stick, when you don't have administrative right to the computer.*

RGui (64-bit)

File Edit View Misc Packages Windows Help



R Console

```
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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

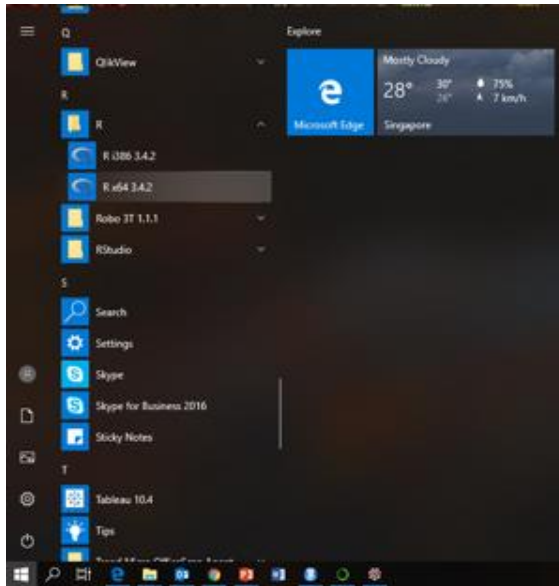
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

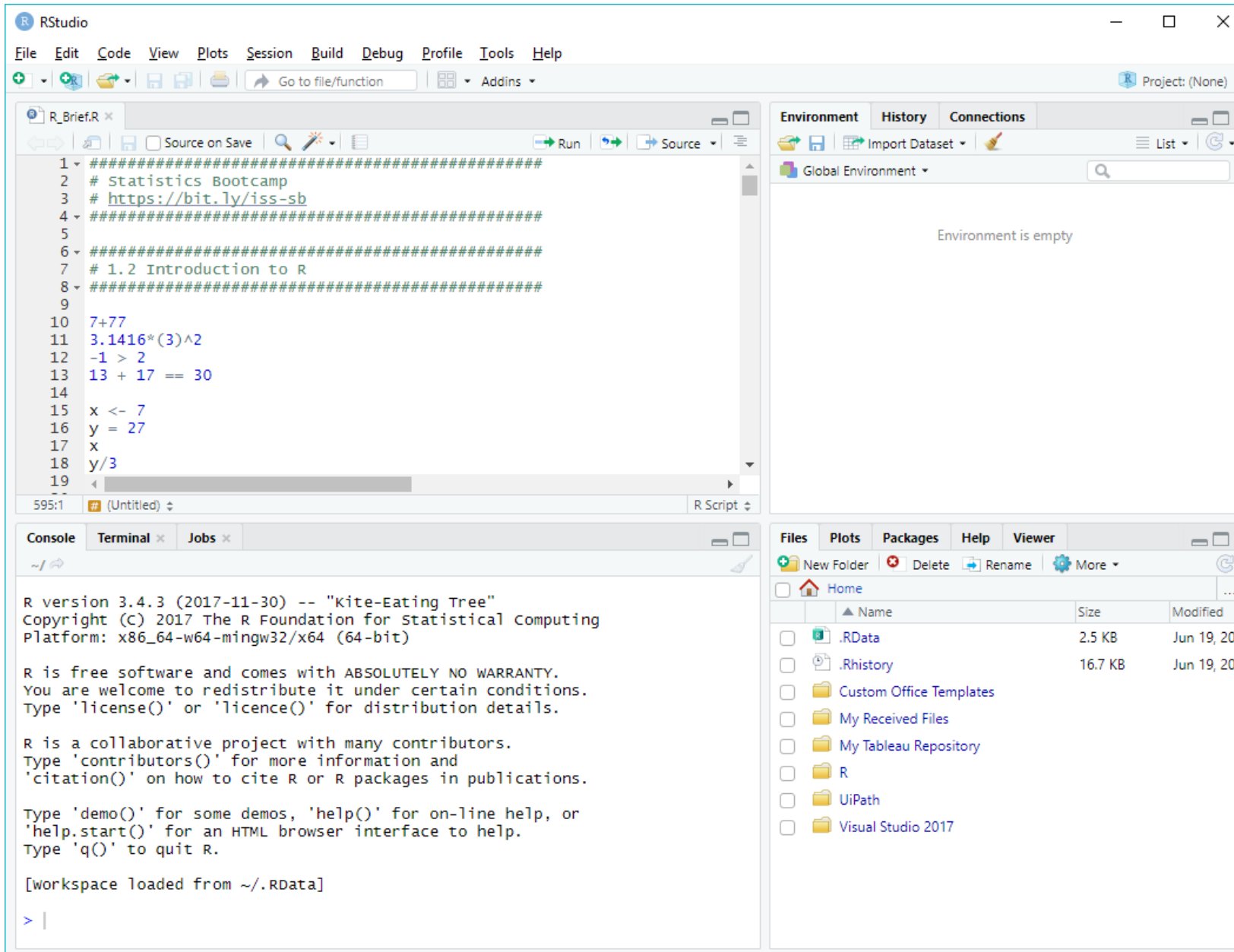
> |
```

Installing R studio

- An integrated development environment (IDE) built for R
- Downloadable at <http://www.rstudio.com>
- Runs on Windows, Mac and Linux, or over the web using RStudio Server
- To access the installed R, for windows user, go to *Start Menu*, select '*All apps*', go to folder '*RStudio*', click '*RStudio*'



** It is possible to run R and R studio on USB stick. You need to install R before installing R studio.*



The screenshot shows the RStudio IDE interface. The main editor window displays an R script file named 'R_Brief.R' with the following content:

```
1 #####
2 # Statistics Bootcamp
3 # https://bit.ly/iss-sb
4 #####
5
6 #####
7 # 1.2 Introduction to R
8 #####
9
10 7+77
11 3.1416*(3)^2
12 -1 > 2
13 13 + 17 == 30
14
15 x <- 7
16 y = 27
17 x
18 y/3
19 <
```

The Environment pane on the right shows 'Global Environment' and 'Environment is empty'. The Console pane at the bottom displays the R startup message and workspace information:

```
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (c) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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.

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.

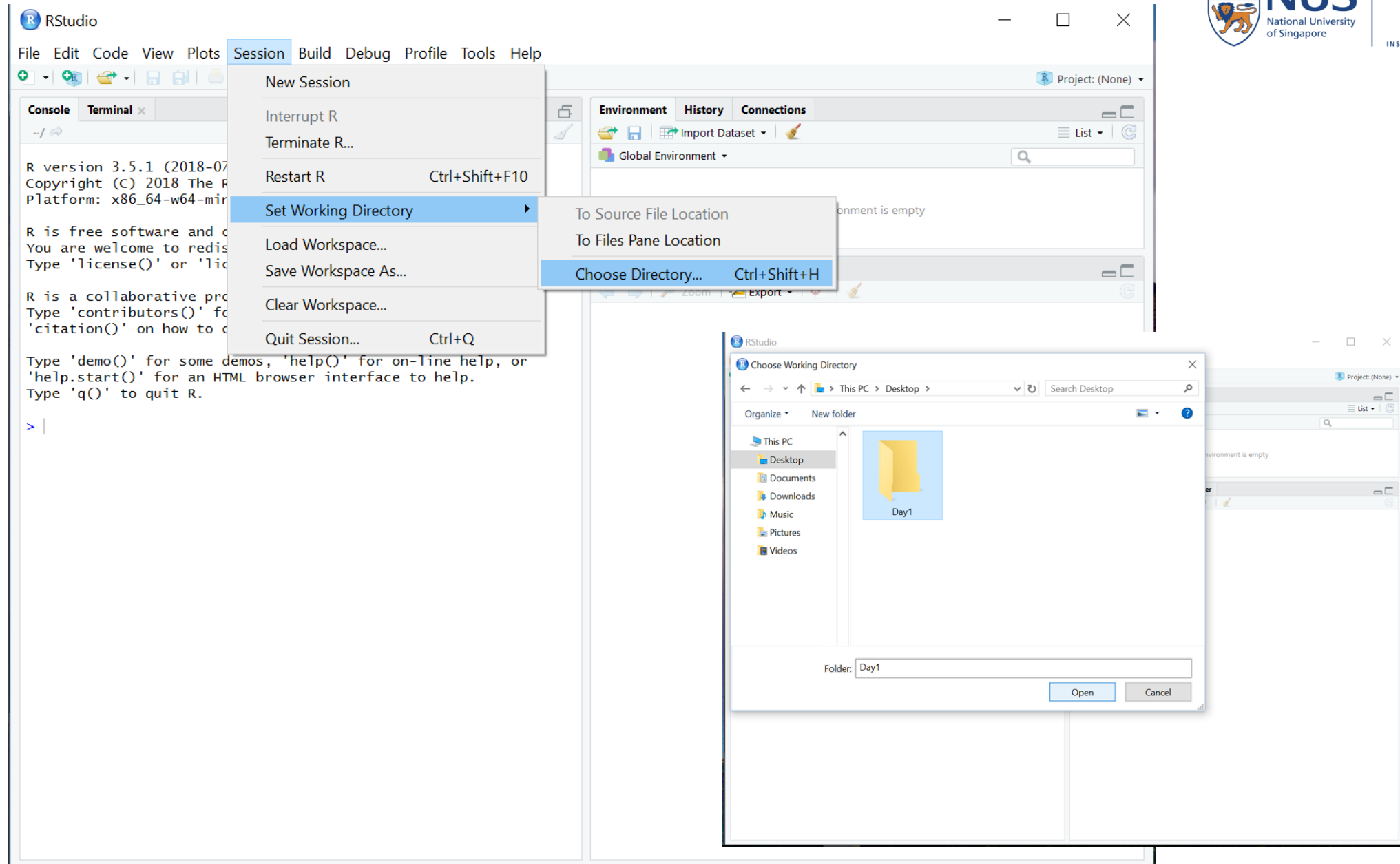
[workspace loaded from ~/.RData]
> |
```

The Files pane on the bottom right shows a list of files in the Home directory:

Name	Size	Modified
.RData	2.5 KB	Jun 19, 2017
.Rhistory	16.7 KB	Jun 19, 2017
Custom Office Templates		
My Received Files		
My Tableau Repository		
R		
UiPath		
Visual Studio 2017		

Set working directory

- This step is often needed when you start RStudio
- Not setting the correct working directory is a common mistake among new learners



How R works

1. Type expressions at the R prompt
2. Expressions evaluated by the R interpreter
3. Then computed values are printed below the prompt
4. Some examples:

R prompt ←

```
## expressions  
> 7+77  
[1] 84
```

computed output ←


```
> 3.1416*(3)^2  
[1] 28.2744
```

comment ←

```
## logical values  
> -1 > 2  
[1] FALSE
```

```
> 13 + 17 == 30  
[1] TRUE
```

File Edit Code View Plots Session Build Debug Profile Tools Help



Addins

Project: (None)

Source

Console Terminal x

```

~/
> 7+77
[1] 84
> 3.1416*(3)^2
[1] 28.2744
> -1 > 2
[1] FALSE
> 13 + 17 == 30
[1] TRUE
> |

```

Environment History Connections

Import Dataset
List
Global Environment

Environment is empty

Files Plots Packages Help Viewer

Zoom Export

- Variable: store value (number, string, etc) for access later
- Assignment operator: `<-` or `=`

```
> x <- 7  
> y = 27  
> x  
[1] 7
```

```
> y/3  
[1] 9
```

```
> a = 'Learning statistics'  
> B <- 'Learning R'
```

- `+`, `-`, `/`, `<-`, `=`, `%%`, `==`, these are called operators:







```
> z = 11  
> z = z + 1  
> z  
[1] 12
```


```
> z / 2  
[1] 6
```

```
> z + (5-3)  
[1] 14
```

```
> z %% 3  
[1] 0
```

File Edit Code View Plots Session Build Debug Profile Tools Help


Addins

Project: (None)

Source



Console Terminal x

```

~/
> x <- 7
> y = 27
> x
[1] 7
> y/3
[1] 9
> a = 'Learning statistics'
> B <- 'Learning R'
>
> z = 11
> z = z + 1
> z
[1] 12
> z / 2
[1] 6
> z + (5-3)
[1] 14
> z %% 3
[1] 0
> |

```

Environment History Connections





Import Dataset
List

Global Environment

Values

a	"Learning statistics"
B	"Learning R"
x	7
y	27
z	12

Files Plots Packages Help Viewer



Zoom
Export

Function

- A piece of code written to carry out a specified task
- It can or can not accept arguments
- It can or can not return one or more values

```
> sum(2, 4, 5)
```

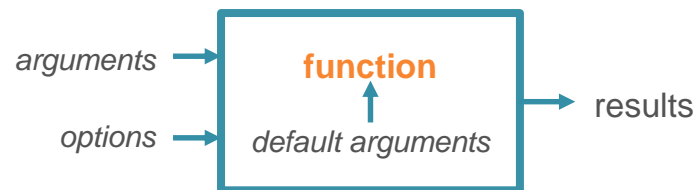
```
[1] 11
```

```
> sqrt(121)
```

```
[1] 11
```

- Some arguments have names. For example to display the help page for 'stats' package, do:

```
> help(package="stats")
```



R script

- For most of the time, we run R interactively. Enter commands at the prompt and observe the result.
- At times, we may want to run a groups of R commands at one go. In this case, we write R commands in a plain text file (with “.R” extension) and run the script.
- To run a script file named ‘info.R’, type:

```
> source("info.R")
```

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Project: (None)

health-check-Rstudio.R x
Untitled1* x
info.R x

Source on Save
Source

```

1 print("Welcome to NUS ISS")
2 print("You are attending Statistics Bootcamp Using R")
3 print(Sys.Date())

```

1:1 (Top Level)
R Script

Console Terminal x

/media/sf_vm_shared_folder/SB/Day1/

```

> source("info.R")
[1] "Welcome to NUS ISS"
[1] "You are attending Statistics Bootcamp Using R"
[1] "2019-01-09"
>

```

Environment History Connections

Import Dataset
List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

/ > media > sf_vm_shared_folder > SB > Day1

	Name	Size	Modified
<input type="checkbox"/>	filmGrossing.R	3 KB	Oct 16, 2018,
<input checked="" type="checkbox"/>	info.R	102 B	Jul 17, 2018, 3
<input type="checkbox"/>	PUB.csv	275 B	Oct 17, 2018,
<input type="checkbox"/>	PUB.xlsx	10 KB	Oct 22, 2018,
<input type="checkbox"/>	PUBc.csv	273 B	Oct 22, 2018,

Getting help

- General help (Menu > Help > R Help)
- To get a help on a particular function, type

```
> help(sum)
> ?sum
> ?sqrt
```

- Run the examples of an R function if provided

```
> example(sum)
```

File Edit Code View Plots Session Build Debug Profile Tools Help

+ Add New
+ Add New
Go to file/function
Addins
Project: (None)

Source

Console Terminal x

/media/sf_vm_shared_folder/SB/Day1/

```

> help(sum)
> ?sum
> ?sqrt
> example(sum)

sum> ## Pass a vector to sum, and it will add the elements together.
sum> sum(1:5)
[1] 15

sum> ## Pass several numbers to sum, and it also adds the elements.
sum> sum(1, 2, 3, 4, 5)
[1] 15

sum> ## In fact, you can pass vectors into several arguments, and everything gets added.
sum> sum(1:2, 3:5)
[1] 15

sum> ## If there are missing values, the sum is unknown, i.e., also missing, ....
sum> sum(1:5, NA)
[1] NA

```

Environment History Connections

Files Plots Packages Help Viewer

R: Miscellaneous Mathematical Functions Find in Topic

MathFun {base} R Documentation

Miscellaneous Mathematical Functions

Description

abs(x) computes the absolute value of x, sqrt(x) computes the (principal) square root of x, \sqrt{x} .

The naming follows the standard for computer languages such as C or Fortran.

Usage

```
abs(x)
sqrt(x)
```

Arguments

Common mistakes in R programming

	Example	
	✓	✗
Using the wrong case	<code>help()</code>	<code>Help()</code>
Forgetting to use quote marks	<code>install.packages("lattice")</code>	<code>install.packages(lattice)</code>
Forgetting to include the parentheses in a function call	<code>help()</code>	<code>help</code>
Using the \ in a pathname on Windows	<code>c:/mydata</code> <code>c:\\mydata</code>	<code>c:\mydata</code>
Using a function from a package that's not loaded	<code>install.packages("lattice")</code> <code>contourplot()</code>	<code>contourplot()</code>

Useful R functions

Function	Purpose
<code>library()</code>	Load add-on packages.
<code>install.packages()</code>	Download and install packages from CRAN-like repositories or from local files.
<code>library(help="package_name")</code>	Show documentation for a package.
<code>sum()</code>	Compute the sum of values.
<code>sqrt()</code>	Compute the square root.
<code>rep()</code>	Replicate a value.
<code>help()</code>	Access the primary interface to the R help systems.
<code>??function</code>	To find out which package carries this function.
<code>example()</code>	Run the examples of an R function
<code>source()</code>	Submit a script to the current R session

Useful tips

Tip	Purpose
Use arrow keys	To navigate between previously typed in commands
Control L	Clears the console
library(help="package_name")	Show documentation for a package.
Typing inside quotes	Both quotes are automatically added
Setting working directory	See screenshot on page 23
Keyboard shortcuts	Alt + Shift + K
Open a Terminal Window	Shift + Alt + r
If you see a "+" instead of a ">"	Your command is incomplete. Check your syntax.

Other resources

<http://www.r-project.org>

<http://www.r-bloggers.com>

<http://www.rstudio.com>

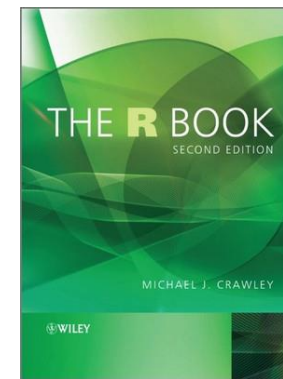
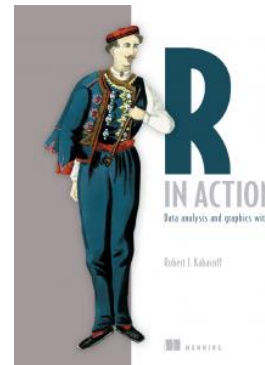
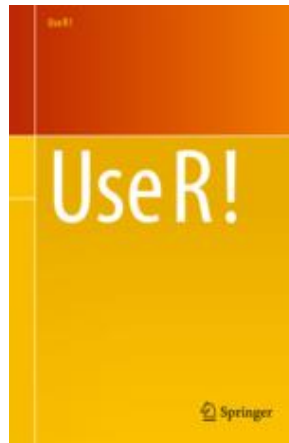
<http://stackoverflow.com>

<http://stats.stackexchange.com>

<http://quora.com>

<http://www.google.com>

<https://github.com/telescopeuser/S-SB-Workshop>



End of Lecture Notes