

Interuniversity Institute for Biostatistics
and statistical Bioinformatics

An introduction to R: Basic skills in R Studio and R Markdown

Developed by:

Rudradev Sengupta, Bernard Osang'ir, Leyla Kodalci, Julia Duda and Ziv Shkedy

Hasselt University

Updated: 08/25



Visit us on

Facebook

ER-BioStat

GitHub



<https://github.com/eR-Biostat>



@erbiostat

Overview

- R Studio/R markdown.
- How to produce output using R markdown ?
- How to develop output using R markdown ?
- Main concept:
 - we run the same analysis multiple times and show how to produce different type of output.
- Analysis:
 - simple plot + simple linear regression.

Rmd programs for today

- To run the analysis presented in the slides you need the following Rmd programs:

Dataset	Rmd file	Output
mtcars	R_course_UHasselt_demo_V0.Rmd	R code in a RMakdown file
mtcars	R_course_UHasselt_demo_V1.Rmd	Produce simple HTML/PDF output for a report
Airquality	R_course_UHasselt_2021(html)_V1.Rmd	Produce a HTML output (example how to work with R markdown)
Airquality	R_course_UHasselt_2021(pdf)_V1.Rmd	Produce HTML/PDF output
Old Faithful	R_course_UHasselt_2021(pdf)_V2.Rmd	A report for an example of a simple analysis (HTML/PDF)

See later in the slides how these files are used and which type of output can be produced.

Rmd programs for today

- Case study: The airquality data:
 - Slide 100-154.
- R & Rmd Programs:
 - [Case studies_Prog1_V1.Rmd](#).
 - [Airquality.R](#).

Data & Software

- Datasets:
 - All are part of R / R packages etc.
 - For some examples: some packages are needed to access the data.
- Install:
 - R Studio/R markdown.
 - Other R packages...
- To produce PDF files:
 - Need to install LaTex, Tex...

R Studio

What is R Studio ?

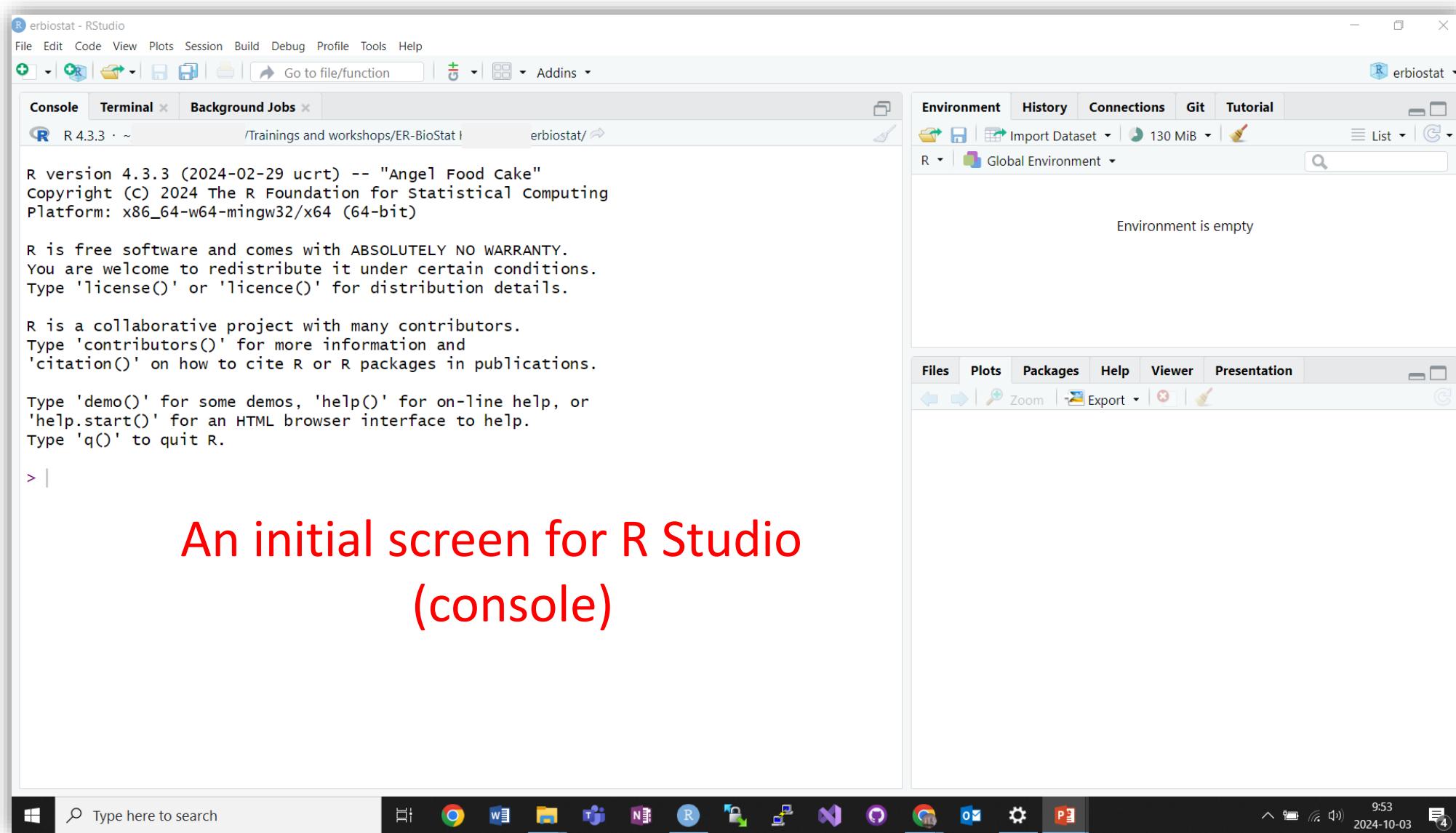
- R Studio is an integrated development environment (IDE) for R and Python.
 - It includes a **console, syntax-highlighting editor** that supports direct code execution.
 - Tools for plotting, history, debugging and workspace management.
- R Studio is available in **open source** and **commercial** editions and runs on the desktop (Windows, Mac, and Linux).

What is R Studio ?

- More information:

<https://rstudio.com/products/rstudio/>

R Studio



An initial screen for R Studio
(console)

The cars dataset in R

The data give the speed of cars and the distances taken to stop. Note that the data were recorded in the 1920s.

Two variables:

- Cars' speed.
- Cars' stopping distance.

```
> head(cars)
```

	speed	dist
--	-------	------

1	4	2
2	4	10
3	7	4
4	7	22
5	8	16
6	9	10

R Studio: example

erbiostat - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

R 4.3.3 · ~/Bernard OSANG'IR/Trainings and workshops/ER-BioStat Kenya 2024/erbiostat/

```
R version 4.3.3 (2024-02-29 ucrt) -- "Angel Food Cake"
Copyright (C) 2024 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.

> plot(cars$speed, cars$dist)
>
```

A plot in R Studio

- The **cars** dataset.
- Plot of the car's speed versus stopping distance.

Plot Zoom

cars\$dist

plot (x, y)

cars\$speed

Version

- 1.50.0
- 1.4-5
- 1.1-4.1
- 1.7-22
- 0.35
- 0.1-2
- 0.17-6
- 1.64.1

Type here to search

10:26
2024-10-03

R: example

RGui (64-bit)

File History Resize Windows

R Console

```
R version 4.3.3 (2024-02-29 ucrt) -- "Angel Food Cake"
Copyright (C) 2024 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.

[Previously saved workspace restored]

> plot(cars$speed, cars$dist)
> |
```

Untitled - R Editor

```
plot(cars$speed, cars$dist)
```

R Graphics: Device 2 (ACTIVE)

A scatter plot showing the relationship between car speed (x-axis) and stopping distance (y-axis). The x-axis is labeled 'cars\$speed' and ranges from 5 to 25. The y-axis is labeled 'cars\$dist' and ranges from 0 to 120. The plot shows a positive correlation, with data points scattered across the plot area, generally increasing from left to right.

- The **cars** dataset.
- Plot of the car's speed versus stopping distance.

plot (x, y)

A similar plot in R



General structure of R

- R Functions.
- R packages (written by developers) as a part of the software:
 - `glm()`
- R packages (written by developers) uploaded in CRAN, Bioconductor, Github....
- R packages (written by the user=you).

R functions

Example:

```
function (data)
```

```
> var (x)
```

The R function

data

A procedure that was programmed in R that uses data to produce output.

Calculate the sample variance.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

R packages

- A package in R: a **collection of functions** that can be used for analysis or data management.
- Example: the `nlme` package for linear mixed models.
- Some packages are included in R, and some can be installed from CRAN/Bioconductor.
- **CRAN**: website with R packages ready to be installed and used.

<https://cran.r-project.org/>

R Studio: packages

The screenshot shows the R Studio interface. The console window displays the R startup message and a command to plot cars data. A red arrow points from the text "Upload new a package to R Studio" to the "Install Packages" dialog box. A callout box contains the text: "If you try to run the Rmd programs, you will probably need to install some packages." The environment pane shows an empty global environment. Red arrows point to the "Plots" and "Packages" tabs in the bottom navigation bar. The package manager dialog box is open, showing the "User Library" with various R packages listed.

erbiostat - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal x Background Jobs x

R version 4.3.3 (2024-02-29 ucrt) -- "Angel Food Cake"
Copyright (c) 2024 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.

> plot(cars\$speed, cars\$dist)

> |

Upload new a package to R Studio

If you try to run the Rmd programs, you will probably need to install some packages.

erbiostat

Environment History Connections Git Tutorial

Import 161 MiB List

Global Environment

Environment is empty

File Plots Packages Help Viewer Presentation

Install Update Name Description Vers...

User Library

Name	Description	Vers...
a4Core	Automated Affymetrix Array Analysis Core Package	1.50.0
abind	Combine Multidimensional Arrays	1.4-5
additivity...	Additivity Tests in the Two Way Anova with Single Sub-Class Numbers	1.1-4.1
ade4	Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences	1.7-22
admisc	Adrian Dusa's Miscellaneous	0.35
alluvial	Alluvial Diagrams	0.1-2
analogue	Analogue and Weighted Averaging Methods for Paleoenvironmental	0.17-6

10:41 2024-10-03

Example 1

The mtcars data

Part 1: R and R Studio

- Example of a data analysis: the `mtcars` data.
 - Use `help(mtcars)` to get information about the data.
-
- ✓ Running the analysis in R.
 - ✓ Running the analysis in R-Studio.

The mtcars data: analysis in R

- Use `help(mtcars)` in R.

A screenshot of an R session in RGui (64-bit). The R Console window shows the following text:

```
R version 4.3.3 (2024-02-29 ucrt)
Copyright (C) 2024 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 details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information.
Type 'citation()' on how to cite R or R packages.
Type 'demo()' for some demos, 'help()'
'help.start()' for an HTML browser interface,
Type 'q()' to quit R.

[Previously saved workspace restored]

> help(mtcars)
starting httpd help server ... done
>

> |
```

The R Help window displays the documentation for the mtcars dataset. The title is "R: Motor Trend Car Road Tests". The "Description" section states: "The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)." The "Format" section describes it as "A data frame with 32 observations on 11 (numeric) variables." Below this, a list of variables is shown:

- [, 1] mpg Miles/(US) gallon ←
- [, 2] cyl Number of cylinders
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (1000 lbs) ←
- [, 7] qsec 1/4 mile time
- [, 8] vs Engine (0 = V-shaped, 1 = straight)
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [,10] gear Number of forward gears
- [,11] carb Number of carburetors

A red arrow points to the first item in the list, and another red arrow points to the sixth item in the list.

Description of the data in the help system.

The mtcars data in R

```
> head(mtcars)
```

		mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda	RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda	RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun	710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet	4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet	Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant		18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

- 
- Variables in the data.
 - For our example: **mpg** and **wt**.

Analysis

- Plot mpg Vs. weight.
- Calculate the mean weight.
- Fit a simple regression model for mpg on Weight.

To run the analysis in the example, use the program in R Studio:

[R_course_UHasselt_demo_V0.Rmd](#)

Analysis in basic R & output

RGui (64-bit)

File History Resize Windows

R Console

```
> plot(mtcars$wt, mtcars$mpg)
> mean(mtcars$mpg)
[1] 20.09062
> fit.lm<-lm(mtcars$mpg~mtcars$wt)
> summary(fit.lm)

Call:
lm(formula = mtcars$mpg ~ mtcars$wt)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5432 -2.3647 -0.1252  1.4096  6.8727 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 37.2851    1.8776 19.858 < 2e-16 ***
mtcars$wt   -5.3445    0.5591 -9.559 1.29e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared:  0.7528, Adjusted R-squared:  0.7528 
F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10
```

see next slide

The estimated model

Untitled - R Editor

```
summary(mtcars)
plot(mtcars$wt, mtcars$mpg)
mean(mtcars$mpg)
fit.lm<-lm(mtcars$mpg~mtcars$wt)
summary(fit.lm)
```

R Graphics: Device 2 (ACTIVE)

mtcars\$mpg

mtcars\$wt

11:25 2024-10-03

The R code for the analysis

```
summary(mtcars)
plot(mtcars$wt, mtcars$mpg)
mean(mtcars$mpg)
fit.lm<-lm(mtcars$mpg~mtcars$wt) —————> fit a simple linear regression
model
summary(fit.lm)
```

The regression output

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

mtcars\$mpg
response

mtcars\$wt
predictor

The mtcars data: an analysis in R Studio

The screenshot shows the R Studio interface with the following components:

- Source Editor:** Displays the R Markdown file `R_Course_UHasselt_demo_V0.Rmd`. The code is annotated with a red box around lines 18-22, which are highlighted in yellow. A red arrow points from the text "The R code" to this box.
- Console:** Shows the output of the R code, including the regression results for the relationship between weight (wt) and miles per gallon (mpg).

```
(Intercept) 37.2851 1.8776 19.858 < 2e-16 ***
mtcars$wt -5.3445 0.5591 -9.559 1.29e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446
F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10
```
- Environment:** Shows the global environment with an object named `fit.lm`.
- Packages:** Shows the user library with various packages listed.

The R code

The mtcars data : an analysis in R Studio

The screenshot illustrates the workflow for analyzing the mtcars dataset in R Studio. The interface is divided into several panels:

- Code Editor (Source tab):** Shows the R code for setting up the environment and performing a linear regression analysis on the mtcars dataset. A red arrow points from this panel to the "Run" menu.
- Run Menu:** A context menu is open over the code editor, highlighting the "Run Selected Line(s)" option. Other options include "Run Current Chunk", "Run All Chunks Above", and "Run All". A red arrow points from the menu to the "Run" button in the toolbar.
- Console:** Displays the output of the R code, including the summary statistics of the mtcars dataset, the results of the lm() function, and the significance codes.
- Output Panel:** Shows the results of the regression analysis, including the residual standard error, R-squared values, F-statistic, and p-value.
- Environment Tab:** Shows the global environment, listing the fit.lm object as a list of 12 items.
- Packages Tab:** Shows the user library with various installed packages like a4Core, abind, ade4, etc.

Annotations in red text:

- The R code**: Points to the code editor area.
- The output**: Points to the console output area.

Analysis in R Studio: the output

R erbiostat - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

R_Course_UHasselt_demo_V0.Rmd*

Go to file/function Addins

Source Visual

```
1 ---  
2 title: "Demo 1"  
3 author: "Rudradev Sengupta, Bernard Osang'ir & Ziv Sh  
4 date: "01/10/24"  
5 output:  
6 html_document: default  
7 word_document: default  
8 pdf_document: default  
9 params:  
10 snapshot: "lubridate::ymd_hms(\"2015-01-01 12:30:00")"  
11 start: "lubridate::ymd(\"2015-01-01\")"  
12 ---  
13  
14 ``{r setup, include=FALSE}  
15 knitr::opts_chunk$set(echo = TRUE)  
16 ``  
17  
18 summary(mtcars)  
19 plot(mtcars$wt,mtcars$mpg)  
20 mean(mtcars$mpg)  
21 fit.lm<-lm(mtcars$mpg~mtcars$wt)  
summary(fit.lm)
```

code

22:16 (Top Level) ↴

Console Terminal Background Jobs

R 4.3.3 · ~/Bernard OSANG'IR/Trainings and workshops/ER-BioStat Kenya 2024/erbiostat/

Min	1Q	Median	3Q	Max
-4.5432	-2.3647	-0.1252	1.4096	6.8727

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.2851	1.8776	19.858	< 2e-16 ***
mtcars\$wt	-5.3445	0.5591	-9.559	1.29e-10 ***

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

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446
F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

> |

Plot Zoom

mtcars\$mpg

mtcars\$wt

The estimated model

Environment History Connections Git Tutorial

Import Dataset 186 MB List

Data fit.lm List of 12

Files Plots Packages Help Viewer Presentation

Zoom

Graphical window

12:14 2024-10-03

R Studio

- So far, R Studio (in this setting): similar to R BUT....
- R Studio + R markdown: a different level of output.

R markdown

R markdown: what?

- Markdown allows you to write a file format independent document using an **easy-to-read** and **easy-to-write** plain text format.
- Instead of marking up text so that is easy for a computer to read
 - e.g. HTML: <html><body>Name</body></html>
- The goal is to mark down the text so that it is easy and human-readable (instead of machine-readable):
 - e.g. ****Name****
- The most easy way to understand it: R markdown allows us to write free text and run R together in the same file.

R markdown: what?

- Markdown is a specific Markup language which is structured very loosely
=> any file format can be generated using pandoc.
- Pandoc: R function to convert documents To other formats.
- From one Markdown document you can generate different file formats:
 - **HTML**
 - **PDF**
 - **Docx**
 - **slideshows**
 - **rtf, etc.**
- The downside is that there is slightly less control over formatting.

R markdown: what?

- Extension of Markdown via R:
 - Allowing *R code* and its *results* to be merged with Markdown.
 - Ensuring that *R Markdown documents are fully reproducible*.
 - Enabling extra modifications to the original markdown specification.
- Provides a *unified authoring framework for data science*, combining your code, its results, etc.
- In practice: just by changing the dataset, the entire analysis can be rerun, and the new report can be produced.
- Integrates a number of R packages and external tools.

R markdown: what?

- A lot of online materials.
- RMarkdown Cheat Sheet: *Help > Cheatsheets > R Markdown Cheat Sheet* (<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>)
- RMarkdown Reference Guide: *Help > Cheatsheets > R Markdown Reference Guide*
- Both cheatsheets are also available at <http://rstudio.com/cheatsheets>
- *Help > Markdown Quick Reference*

The R markdown program: components

The screenshot shows the RStudio interface with an R Markdown file open. The file contains the following content:

```
---  
title: "Demo"  
author: "Rudradev Sengupta"  
date: "16 July 2019"  
output: html_document  
---  
```{r setup, include=FALSE}  
knitr::opts_chunk$set(echo = TRUE)
```  
## R Markdown  
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.  
when you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
```{r cars}  
summary(cars)
```  
## Including Plots  
You can also embed plots, for example:  
```{r pressure, echo=FALSE}  
plot(pressure)
```  
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

Annotations in the screenshot:

- A bracket on the left labeled "YAML Header" points to the YAML header section.
- A bracket on the left labeled "Formatted Text" points to the free text and section headers.
- A bracket on the right labeled "Code Chunks" points to the R code blocks.
- A red arrow points from the text "R code" to the first R code block.
- Two red boxes highlight specific code blocks: one around `summary(cars)` and another around `plot(pressure)`.

Programname.Rmd

There are principally three sections to an R Markdown document:

- YAML header surrounded by `---`
- Code chunks surrounded by `````
- Free text mixed with simple text formatting like `#heading` and `_italics_`

Starting point

- Available materials to try out:
 - <https://teams.microsoft.com/#/files/Reference%20Material?threadId=19:05ec12c79df7460ca9cdfbd8b620f16a@thread.skype&ctx=channel&context=Rmarkdown%2520Help> – templates by Stefan to create pdf/html/.. documents
 - <https://teams.microsoft.com/#/files/Reference%20Material?threadId=19:05ec12c79df7460ca9cdfbd8b620f16a@thread.skype&ctx=channel&context=Code%2520Repository%252FAssignment%25206> – solutions for Assignment 6, by different groups in US

Starting point

- Available materials to try out:
 - https://rmarkdown.rstudio.com/articles_intro.html
 - materials from RStudio

Example 1 (continue)

The mtcars data

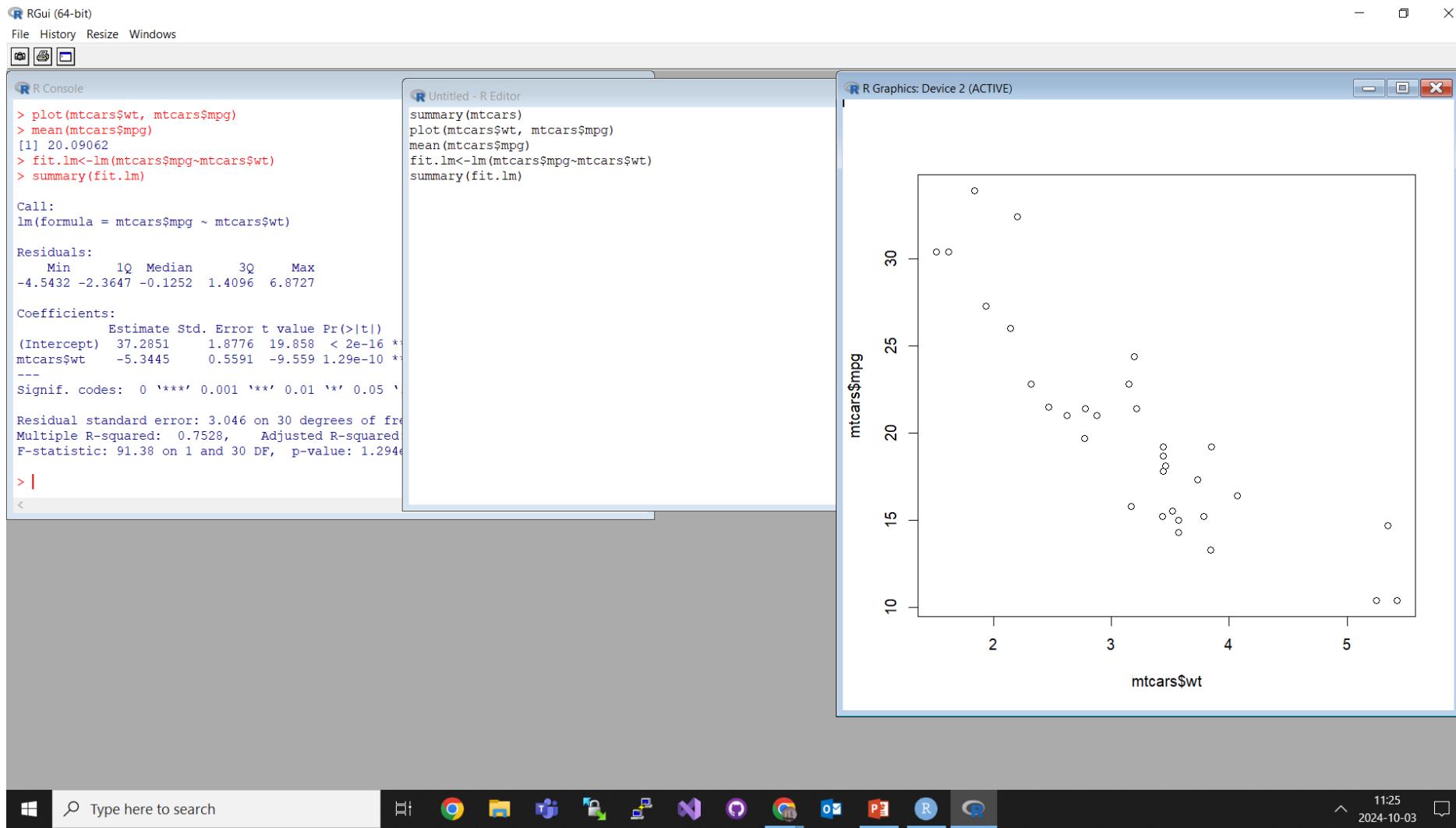
Part 1: Analysis with R and R Studio

Analysis

- Plot mpg Vs. weight.
- Calculate the mean weight.
- Fit a regression model for Mpg on Weight.
- R code: [see next page](#).

Analysis in R

- The same slides as 18-22.



Analysis in R Studio

The screenshot shows the R Studio interface with the following components:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New Project, Open, Save, Run, Knit, and Addins.
- Source Editor:** Displays the R Markdown file `R_Course_UHasselt_demo_V0.Rmd*`. The code includes YAML front matter and R code for summarizing the mtcars dataset and fitting a linear model.
- Console:** Shows the output of the R session, including the summary of `mtcars`, a scatter plot of `wt` vs `mpg`, the mean of `mpg`, the fitted linear model `fit.lm`, and its summary statistics.
- Environment:** Shows the global environment with an object named `fit.lm`.
- Packages:** Shows the user library with packages like `a4Core`, `abind`, `additivityT...`, `ade4`, `admisc`, `alluvial`, `analogue`, and `Annotatio...`.
- Bottom Bar:** Windows taskbar with various application icons.

Analysis in R Studio

The screenshot shows the R Studio interface with several panels:

- Source Editor:** Displays R code for a document named "R_Course_UHasselt_demo_V0.Rmd". A red arrow points from the "Run" button in the toolbar to a context menu that is open over the code. The menu includes options like "Run Selected Line(s)", "Run Current Chunk", and "Run All".
- Console:** Shows the R environment and command-line output. It includes the R version (R 4.3.3), working directory (~/Bernard OSANG'IR/Trainings and workshops/ER-BioStat Kenya 2024/erbiostat), and statistical results for a linear model (lm) fit.
- Environment:** Shows the global environment with objects like "fit.lm" and their types (e.g., "List of 12").
- Packages:** Shows the user library with installed packages such as a4Core, abind, additivityT..., ade4, admisc, alluvial, and analogue.

Annotations in red text:

- "The R code" is placed over the Source Editor area.
- "The output" is placed over the Console output area.

Analysis in R Studio

R erbiostat - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

R_Course_UHasselt_demo_V0.Rmd*

Source Visual

```
1 ---  
2 title: "Demo 1"  
3 author: "Rudradev Sengupta, Bernard Osang'ir & ziv Shmilovich"  
4 date: "01/10/24"  
5 output:  
6 html_document: default  
7 word_document: default  
8 pdf_document: default  
9 params:  
10 snapshot: "lubridate::ymd_hms(\"2015-01-01 12:30:00\")"  
11 start: "lubridate::ymd(\"2015-01-01\")"  
12 ---  
13  
14 ``{r setup, include=FALSE}  
15 knitr::opts_chunk$set(echo = TRUE)  
16 ---  
17  
18 summary(mtcars)  
19 plot(mtcars$wt, mtcars$mpg)  
20 mean(mtcars$mpg)  
21 fit.lm<-lm(mtcars$mpg~mtcars$wt)  
22 summary(fit.lm)
```

22:16 (Top Level) code

Console Terminal Background Jobs

R 4.3.3 · ~/Bernard OSANG'IR/Trainings and workshops/ER-BioStat Kenya 2024/erbiostat/

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -4.5432 | -2.3647 | -0.1252 | 1.4096 | 6.8727 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 37.2851 | 1.8776 | 19.858 | < 2e-16 *** |
| mtcars\$wt | -5.3445 | 0.5591 | -9.559 | 1.29e-10 *** |

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

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446
F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

> |

Plot Zoom

Scatter plot showing the relationship between mtcars\$mpg (Y-axis, ranging from 10 to 30) and mtcars\$wt (X-axis, ranging from 2 to 5). The data points show a negative correlation, where lower weight is associated with higher fuel efficiency.

The estimated model

Graphical window

Scatter plot showing the same data as above, with a red arrow pointing to the 'Zoom' button in the RStudio toolbar. The plot is labeled 'Graphical window'.

Part 2: Analysis with R markdown

To run the analysis in the example, use the program in R Studio:

R_course_UHasselt_demo_V1.Rmd

Different types of Output

- Two types of output:
 - HTML.
 - PDF.
 - DOCX
- Text and R code in the same document.

Running R in Markdown

- General form:

```
```{r}  
R code
...
```
```

Code chunks for
the analysis.

- Example of a scatterplot:

```
```{r}  
plot(x, y)
```
```

Use the function `plot()` to
produce a scatterplot.

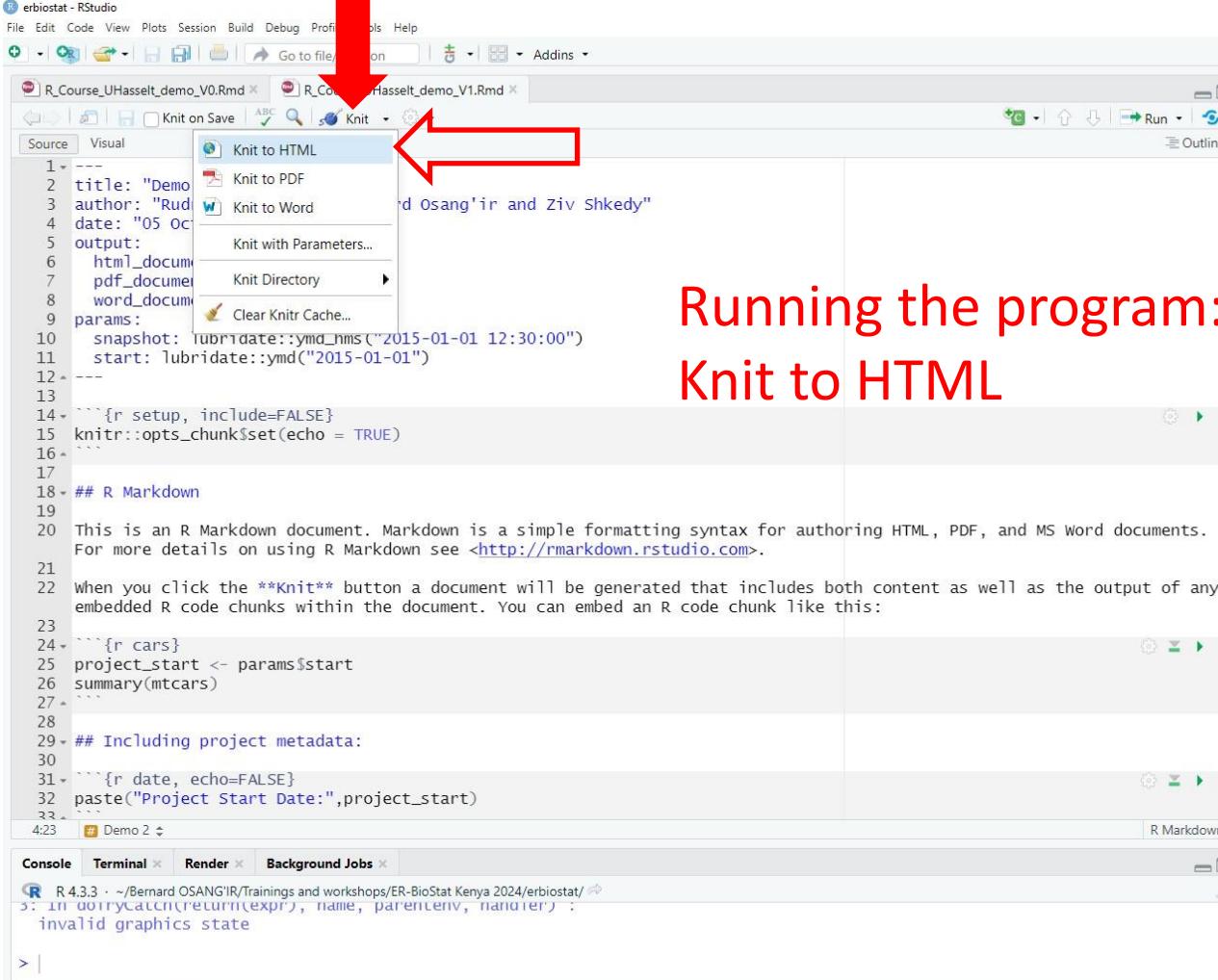
The Rmd program

Title and info.

The screenshot shows the RStudio interface with the following details:

- YAML Header:** Lines 1-17 show the YAML header, which defines the document's title, author, date, output type (HTML), and parameters. A red arrow points to the "title: Demo 2" line, and a red brace labeled "YAML Header" groups these lines.
- R code:** Lines 18-33 show an R code chunk that reads the `mtcars` dataset and prints its summary. A red brace labeled "R code" groups these lines.
- Example of free text:** Lines 20-22 provide instructions on using R Markdown, mentioning Markdown syntax and the Knit button. This text is highlighted with a blue box.
- Environment:** The right sidebar shows the environment pane with the global environment and a list of 12 objects, including `fit.lm`.
- Console:** The bottom-left pane shows the R console output for the `summary(mtcars)` command.

Knit to HTML (how to run the program)



A screenshot of the RStudio interface. The main window shows an R Markdown file named 'R_Course_UHasselt_demo_V0.Rmd'. A red arrow points from the text 'Knit to HTML' in the slide to the 'Knit' button in the toolbar, which has a dropdown menu open. The dropdown menu includes options like 'Knit to HTML' (selected), 'Knit to PDF', 'Knit to Word', 'Knit with Parameters...', 'Knit Directory', and 'Clear Knitr Cache...'. To the right of the code editor is the 'Environment' tab of the global environment pane, showing a list of objects including 'fit.lm'. At the bottom, there's a taskbar with tabs for 'Console', 'Terminal', 'Render', and 'Background Jobs', and a system tray.

Running the program:
Knit to HTML

```
1 ---  
2 title: "Demo"  
3 author: "Rudolf Hasselt"  
4 date: "05 Oct 2015"  
5 output:  
6 html_document  
7 pdf_document  
8 word_document  
9 params:  
10 snapshot: lubridate::ymd_hms("2015-01-01 12:30:00")  
11 start: lubridate::ymd("2015-01-01")  
12 ---  
13  
14 ``{r setup, include=FALSE}  
15 knitr::opts_chunk$set(echo = TRUE)  
16 ``-  
17  
18 ## R Markdown  
19  
20 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents.  
For more details on using R Markdown see <http://rmarkdown.rstudio.com>.  
21  
22 When you click the **Knit** button a document will be generated that includes both content as well as the output of any  
embedded R code chunks within the document. You can embed an R code chunk like this:  
23  
24 ``{r cars}  
25 project_start <- params$start  
26 summary(mtcars)  
27 ``-  
28  
29 ## Including project metadata:  
30  
31 ``{r date, echo=FALSE}  
32 paste("Project Start Date:",project_start)  
33 ``-
```

Output: HTML

The screenshot illustrates the R Markdown workflow in RStudio. On the left, the R Markdown source code for "Demo 2" is shown, with a red circle highlighting the line "# R Markdown". A red arrow points from this line to the generated HTML output on the right. The HTML output displays the document title "Demo 2", authors "Rudradev Sengupta, Bernard Osang'ir and Ziv Shkedy", and the date "05 October 2024". The "R Markdown" section is also highlighted with a red box. A red arrow points from the "See page 43" text at the top right to this box. The "The output" section is also highlighted with a red box, and a red arrow points from the "See page 43" text to it. The output includes a summary of the mtcars dataset and sections for "Including project metadata:" and "Including Plots". The status bar at the bottom right shows the time as 18:51 and the date as 2024-10-03.

See page 43

Demo 2

Rudradev Sengupta, Bernard Osang'ir and Ziv Shkedy

05 October 2024

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
project_start <- params$start
summary(mtcars)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
```{r cars}
project_start <- params$start
summary(mtcars)
```

## Including project metadata:

```
[1] "Project Start Date: lubridate::ymd(\"2015-01-01\")"
```

## # Including Plots

The output

18:51  
2024-10-03

# The HTML output

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

HD\_Unsupervised\_NBA\_2017.R Example\_prog\_2016.R HD\_Unsupervised\_Examp

Source Visual

## Including Plots

You can also embed plots, for example:

```
plot(mtcars$wt,mtcars$mpg)
```

Note that the `echo = FALSE` parameter was added to the code that generated the plot.

```
```{r,echo = FALSE}
plot(mtcars$wt,mtcars$mpg)
````
```

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.

Go to file/function Addins

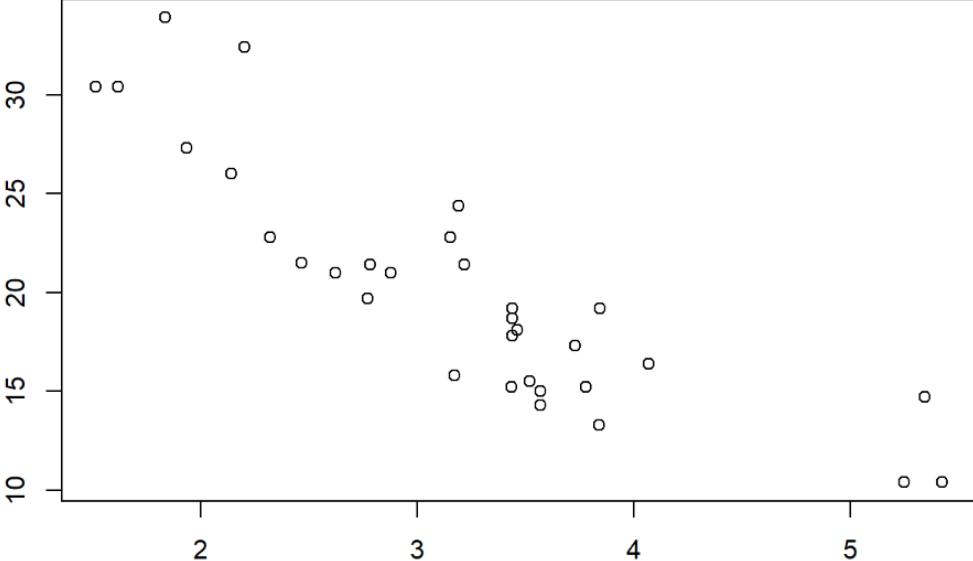
C:/Ziv\_Temp\_2023/Wprkshop\_Kenya/Shortcourse/Rmarkdown1/Rmds/R\_Course\_UHasselt\_demo\_V1.html

R\_Course\_UHasselt\_demo\_V1.html Open in Browser Find Publish

## Including Plots

You can also embed plots, for example:

```
plot(mtcars$wt,mtcars$mpg)
```



A data frame with 15 observations on 2 variables.

Windows Taskbar: Type here to search, File Explorer, Google Chrome, Microsoft Edge, R, 8:23, ENG, 9/05/2025, 49

# The HTML output

The screenshot shows the RStudio interface with an R Markdown file open in the left pane and its generated HTML output in the right pane.

**Left Pane (RStudio Editor):**

- File:** HD\_Unsupervised\_NBA\_2017.R, Example\_prog\_2016.R, HD\_Unsupervised\_Examp
- Tools:** Go to file/function, Addins
- Code View:** Source (selected), Visual
- Code:**

```
49
50
51 ## Short Analysis
52
53 ``{r}
54 mean(mtcars$mpg)
55 fit.lm<-lm(mtcars$mpg~mtcars$wt)
56 `
57
58 ## Output for the regression model
59
60 ``{r}
61 summary(fit.lm)
62 `
63
```
- Console:** R 4.3.2 · C:/Ziv\_Temp\_2023/Wprkshop\_Kenya/Shortcourse/Rmarkdown1/Rmds/
- Output:**

```
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.
```

**Right Pane (Generated HTML Output):**

- Title:** C:/Ziv\_Temp\_2023/Wprkshop\_Kenya/Shortcourse/Rmarkdown1/Rmds/R\_Course\_UHasselt\_demo\_V1.html
- Buttons:** Open in Browser, Find
- Content:**

## Short Analysis

```
mean(mtcars$mpg)
```

```
[1] 20.09062
```

```
fit.lm<-lm(mtcars$mpg~mtcars$wt)
```

## Output for the regression model

```
summary(fit.lm)
```

```
##
Call:
lm(formula = mtcars$mpg ~ mtcars$wt)
##
Residuals:
Min 1Q Median 3Q Max
-4.5432 -2.3647 -0.1252 1.4096 6.8727
##
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.2851 1.8776 19.858 < 2e-16 ***
mtcars$wt -5.3445 0.5591 -9.559 1.29e-10 ***

```

A data frame with 15 observations on 2 variables.

# The Rmd output

- Output in **ONE** document.
- Text and R output.
- R code can be included.
- Type of document: HTML/PDF.
- The data analyst designs the output.

# Knit to PDF

The screenshot shows the RStudio interface. In the top-left corner, there is a red arrow pointing down at the 'Knit' button in the toolbar. A red box highlights the 'Knit' dropdown menu, which is open and displays several options: 'Knit to HTML', 'Knit to PDF' (which is highlighted in blue), 'Knit to Word', 'Knit with Parameters...', 'Clear Knitr Cache...', and 'Knit with Parameters...'. The main workspace shows an R Markdown file with code and output. The right side of the screen features the Environment pane displaying the Global Environment, and the Packages pane showing the installed packages.

```
R_Course_UHasselt_demo_V0.Rmd | R_Course_UHasselt_demo_V1.Rmd
```

```
5 out
6 h Knit to HTML
7 p Knit to PDF
8 par Knit to Word
9 si id_hms("2015-01-01 12:30:00")
10 s id_hm("2015-01-01")
11 s Knit with Parameters...
12 ...
13 Clear Knitr Cache...
14 knitr::opts_chunk$set(echo = TRUE)
15
16
17 ## R Markdown
18
19 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
20
21 when you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
22
23 ``{r cars}
24 project_start <- params$start
25 summary(mtcars)
26
27
22:1 R Markdown
```

Console R Markdown x

```
~/
#> mtcars$mpg ~ mtcars$wt
```

Residuals:

| Min     | Q1      | Median  | Q3     | Max    |
|---------|---------|---------|--------|--------|
| -4.5432 | -2.3647 | -0.1252 | 1.4096 | 6.8727 |

Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )     |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 37.2851  | 1.8776     | 19.858  | < 2e-16 ***  |
| mtcars\$wt  | -5.3445  | 0.5591     | -9.559  | 1.29e-10 *** |

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

Residual standard error: 3.046 on 30 degrees of freedom  
Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446  
F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

10:18  
29/09/2020

# The PDF output

A screenshot of the RStudio interface demonstrating the generation of a PDF document from R Markdown code.

**Left Panel (RStudio Environment):**

- Code Editor:** Shows the R Markdown code for "R\_Course\_UHasselt\_demo\_V0.Rmd". It includes code chunks for setting output to HTML and PDF, specifying a project start date, and embedding an R code chunk for summarizing the mtcars dataset.
- Console:** Displays the results of the R code execution, including the summary statistics for the mtcars dataset.

**Right Panel (PDF Preview):**

- Title:** The title "Demo 2" is highlighted with a red box and an annotation pointing to it from the top right: "See page 43 for the title".
- Author:** "Rudradev Sengupta and Ziv Shkedy"
- Date:** "29 September 2020"
- Content:** A section titled "R Markdown" provides an overview of the syntax and how R code chunks are embedded. It also shows the generated R code for summarizing the mtcars dataset.
- Footnote:** A small note at the bottom of the PDF page indicates "1".

# The PDF output

The screenshot shows the RStudio interface with a PDF output window open. The PDF window displays an R Markdown document titled "Demo 2". The document includes code for generating a plot of mpg vs. wt from the mtcars dataset. The plot is embedded in the PDF, showing a scatter of points with a regression line.

**Code in R\_Course\_UHasselt\_demo\_V0.Rmd:**

```
1 ---
2 title: "Demo 2"
3 author: "Rudradev Sengupta"
4 date: "29 September 2020"
5 output:
6 pdf_document: default
7 html_document: default
8 params:
9 snapshot: lubridate::ymd()
10 start: lubridate::ymd()
11 ---
12
13 ``{r setup, include=FALSE}
14 knitr::opts_chunk$set(
15
16
17 }## R Markdown
18
19 This is an R Markdown document. It provides a simple framework for documents. For more details
20 see http://yihui.name/knitr/.
21 When you click the **Knit** button, a document will be generated that includes both the content of the
22 R code chunks and the output of those chunks.
1:1 Demo 2
```

**Console Output:**

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 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 more information.

R is a collaborative project with many contributors.
Type 'contributors()' for more information.
Type 'citation()' on how to cite R or its package.

Type 'demo()' for some demos,
'help.start()' for an HTML browser-based help system,
Type 'q()' to quit R.

[workspace loaded from ~/RDat]
> |
```

**PDF Content:**

Including project metadata:

```
Median :0.0000 Median :4.0000 Median :2.000
Mean :0.4062 Mean :3.688 Mean :2.812
3rd Qu.:1.0000 3rd Qu.:4.0000 3rd Qu.:4.800
Max. :1.0000 Max. :5.0000 Max. :8.000
```

Including Plots

You can also embed plots, for example:

```
plot(mtcars$wt, mtcars$mpg)
```

**Plots:**

A scatter plot showing the relationship between weight (wt) and miles per gallon (mpg) for the mtcars dataset. The x-axis represents weight (wt) ranging from approximately 1.6 to 5.4. The y-axis represents miles per gallon (mpg) ranging from approximately 10 to 35. The plot shows a negative correlation, with mpg decreasing as weight increases. A regression line is drawn through the data points.

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

**Project Explorer:**

- Project: (None)
- Files:
  - bles
  - les
  - les
  - bles
- Version:
  - 1.4-5
  - anova with Single Sub- 1.1-4
  - SSH 1.1
  - 0.2.1
  - atic Likelihood Ratio 2015.6.25
  - roduced Since R- 1.1.4
  - 0.1-3
  - 1.69.0-1
  - 2.0.1
  - Package Repository 1.30.4
  - ther Bioconductor 2.12.0
  - conductor packages 3.9.0
  - 1.0-6
  - ocuments with R 0.16
  - Angelo Canti for S 1.3-23

# The PDF output

The screenshot shows the RStudio interface with a PDF output window open. The PDF contains R code and its output. A red text overlay 'The regression output' is placed over the middle section of the PDF content.

**PDF Content:**

```
mtcars$wt

Short Analysis

mean(mtcars$mpg)
[1] 20.09062
fit.lm<-lm(mtcars$mpg~mtcars$wt)

Output for the regression model

summary(fit.lm)

Call:
lm(formula = mtcars$mpg ~ mtcars$wt)

Residuals:
Min 1Q Median 3Q Max
-4.5432 -2.3647 -0.1252 1.4096 6.8727

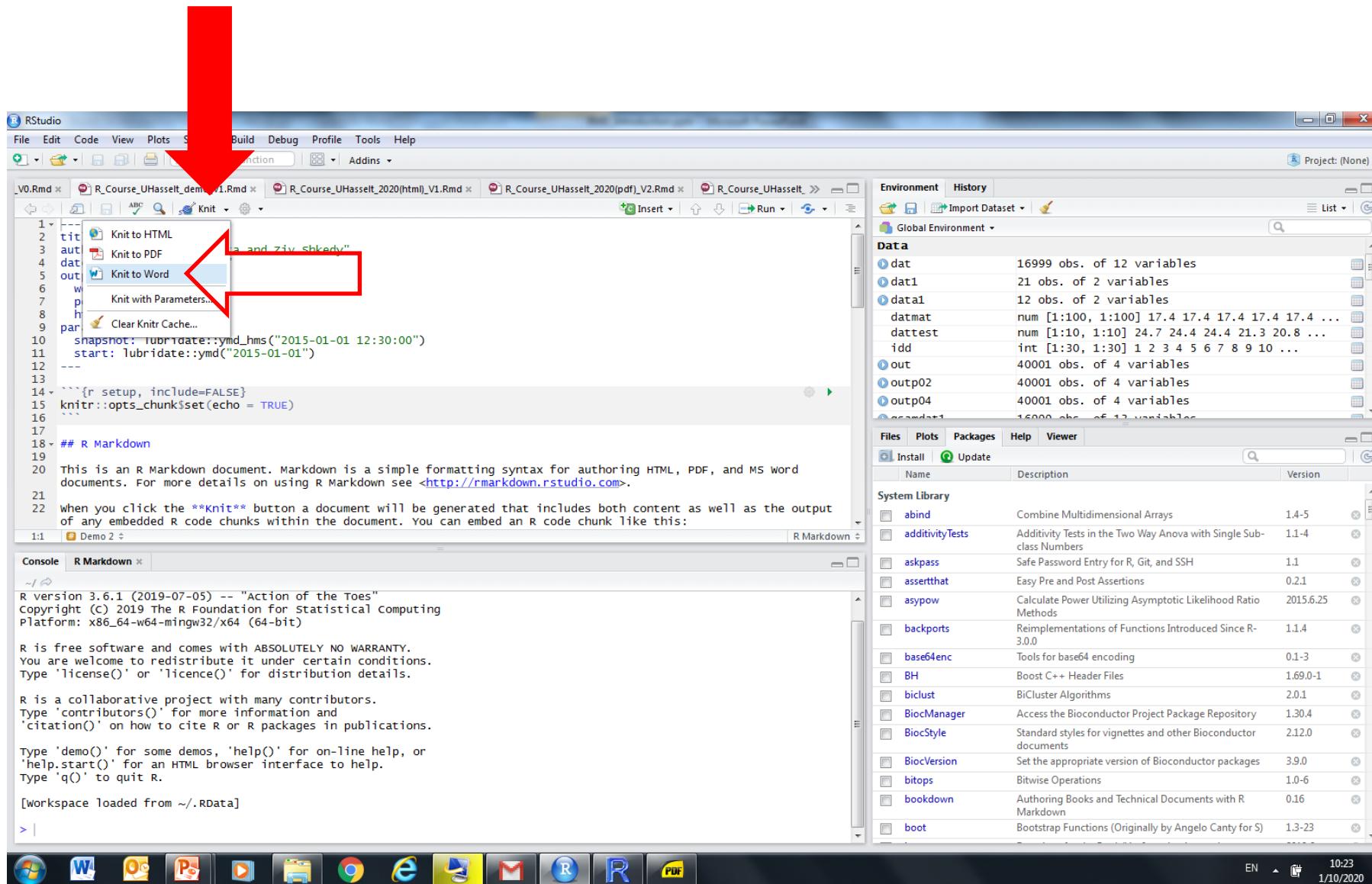
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.2851 1.8776 19.858 < 2e-16 ***
mtcars$wt -5.3445 0.5591 -9.559 1.29e-10 ***

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

3
```

**Important:** To produce the PDF file, you will need to install [LaTeX or Tex](#)

# Knit to word



# The word doc output

A screenshot of Microsoft Word showing an R Markdown document. The ribbon at the top has tabs like File, Home, Insert, Page Layout, References, Mailings, Review, and View. The Home tab is selected. The ribbon contains various icons for text, font, and paragraph styling. Below the ribbon is a toolbar with Cut, Copy, Paste, Format Painter, and Clipboard buttons. The main content area shows a title 'Demo 2', author information 'Rudradev Sengupta and Ziv Shkedy', and a date '29 September 2020'. A section titled 'R Markdown' follows. Below it is a code chunk: 'project\_start <- params\$start summary(mtcars)'. The output of this code is a table of statistics for the mtcars dataset. To the right of the text area is a plot generated by R code: 'plot(mtcars\$wt, mtcars\$mpg)'. The plot is a scatter plot of weight (wt) versus miles per gallon (mpg) for the mtcars dataset. The x-axis is labeled 'mtcars\$wt' and ranges from 1 to 5. The y-axis is labeled 'mtcars\$mpg' and ranges from 10 to 30. The plot shows a negative correlation, with points generally decreasing as weight increases. At the bottom of the page, there are navigation icons for back, forward, search, and file operations. The status bar at the bottom shows 'Page: 1 of 4', 'Words: 435', 'English (U.S.)', '90%', 'EN', and the date '10/10/2020'.

## Example 2

Output development

# Focus

- How to develop an output document:
  - HTML.
  - PDF/HTML.
- Sections/Subsections.
- R code as a part of the text or not?

# The program

- A simple user guide for a document that includes:
  - Free text.
  - R code.
  - Graphical displays.
- To produce the document:
  - run the programs in R Studio.

# Development of a **HTML** document

- The output:
  - HTML document.

To run the analysis in the example, use the program in R Studio:

[`R\_course\_UHasselt\_2021\(html\)\_V1.Rmd`](#)

# The program

R\_course\_UHasselt\_2020(html)\_V1.Rmd

YAML Header

```
1 ---
2 title: "Basic skills in R Markdown"
3 author:
4 - name: Ziv Shkedy
5 affiliation: Hasselt University, Belgium
6 subtitle: The HTML file (2020)
7
8 #institute: UHasselt
9 date: September, 29, 2020
10 output:
11 rmdformats::readthedown:
12 highlight: kate
13 use_bookdown: TRUE
14 vignette: >
15 %>%vignetteEngine{knitr::rmarkdown}
16 editor_options:
17 chunk_output_type: console
18 categories: ["R"]
19 bibliography: bibliography.bib
20 ---
21
22
23 <script type="text/javascript"
24 src="http://cdn.mathjax.org/mathjax/latest/MathJax.js?config=TeX-AMS-MML_HTMLorMML">
1:1 Basic Skills in R Markdown
```

R version 3.6.1 (2019-07-05) -- "Action of the Toes"  
Copyright (C) 2019 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]

> |

Environment History

Data

- dat 16999 obs. of 12 variables
- dat1 21 obs. of 2 variables
- data1 12 obs. of 2 variables
- datmat num [1:100, 1:100] 17.4 17.4 17.4 17.4 ...
- dattest num [1:10, 1:10] 24.7 24.4 24.4 21.3 20.8 ...
- idd int [1:30, 1:30] 1 2 3 4 5 6 7 8 9 10 ...
- out 40001 obs. of 4 variables
- outp02 40001 obs. of 4 variables
- outp04 40001 obs. of 4 variables
- scamdat1 16000 obs. of 12 variables

Files Plots Packages Help Viewer

System Library

- abind Combine Multidimensional Arrays 1.4-5
- additivityTests Additivity Tests in the Two Way Anova with Single Sub-class Numbers 1.1-4
- askpass Safe Password Entry for R, Git, and SSH 1.1
- assertthat Easy Pre and Post Assertions 0.2.1
- asypow Calculate Power Utilizing Asymptotic Likelihood Ratio Methods 2015.6.25
- backports Reimplementations of Functions Introduced Since R-3.0.0 1.1.4
- base64enc Tools for base64 encoding 0.1-3
- BH Boost C++ Header Files 1.69.0-1
- biclust BiCluster Algorithms 2.0.1
- BiocManager Access the Bioconductor Project Package Repository 1.30.4
- BiocStyle Standard styles for vignettes and other Bioconductor documents 2.12.0
- BiocVersion Set the appropriate version of Bioconductor packages 3.9.0
- bitops Bitwise Operations 1.0-6
- bookdown Authoring Books and Technical Documents with R Markdown 0.16
- boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-23

EN 9:04 6/10/2020

# The HTML output

The screenshot shows a Microsoft PowerPoint window with the title bar "RMD Introduction.pptx - Microsoft PowerPoint". The slide content is titled "Basic Skills in R Markdown". The slide contains the following sections:

- 1 Introduction**
- 2 Sections and subsections**
- 3 Including R code**
- 4 Items**
- 5 Use R as a part of your text**
- 6 How to add a link to your document**
- 7 How to create a math formula**
- 8 Just do it**

The slide also includes two code snippets:

```
Warning: package 'ggplot2' was built under R version 3.6.3
```

```
Warning: package 'mvtnorm' was built under R version 3.6.2
```

A red callout box on the right side of the slide contains the text:

**Very basic guide how  
to produce a  
markdown document**

The left side of the image shows the PowerPoint ribbon and the slide navigation pane with slides 57, 58, 59, and 60. Slides 57 and 60 show screenshots of RStudio. Slide 59 is highlighted with a yellow border.

# The program

An example how to create section/subsection/subsubsection...

Free text

RStudio interface showing R Markdown code and output:

Code Editor (R\_Course\_UHasselt\_demo\_V0.Rmd):

```
49 This document provides basic tools to produce a html file using R markdown. The best way to use this document is
50 to run the file in R studio and then read the <tt>.Rmd</tt> file to see how the output was created. The file can
be used to produce a very basic html document an you can add later more components to you document.
51 # Sections and subsections
52 This is a an example of a R markdown file that produces htnl output. This is a section in the document.
53 ## Subsection
54 This text apears in a subsection
55 ### Subsubsection
56 This text is a part of a subsection.
57
58 # Including R code
59
60 ## Print R code and output
61
62 This is an example how to include R code and output in the document. we use the <tt>airquality</tt> data as an
63
64
65
66
67
68
69
70 This is an example how to include R code and output in the document. we use the <tt>airquality</tt> data as an
```

Console (R Markdown):

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 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.

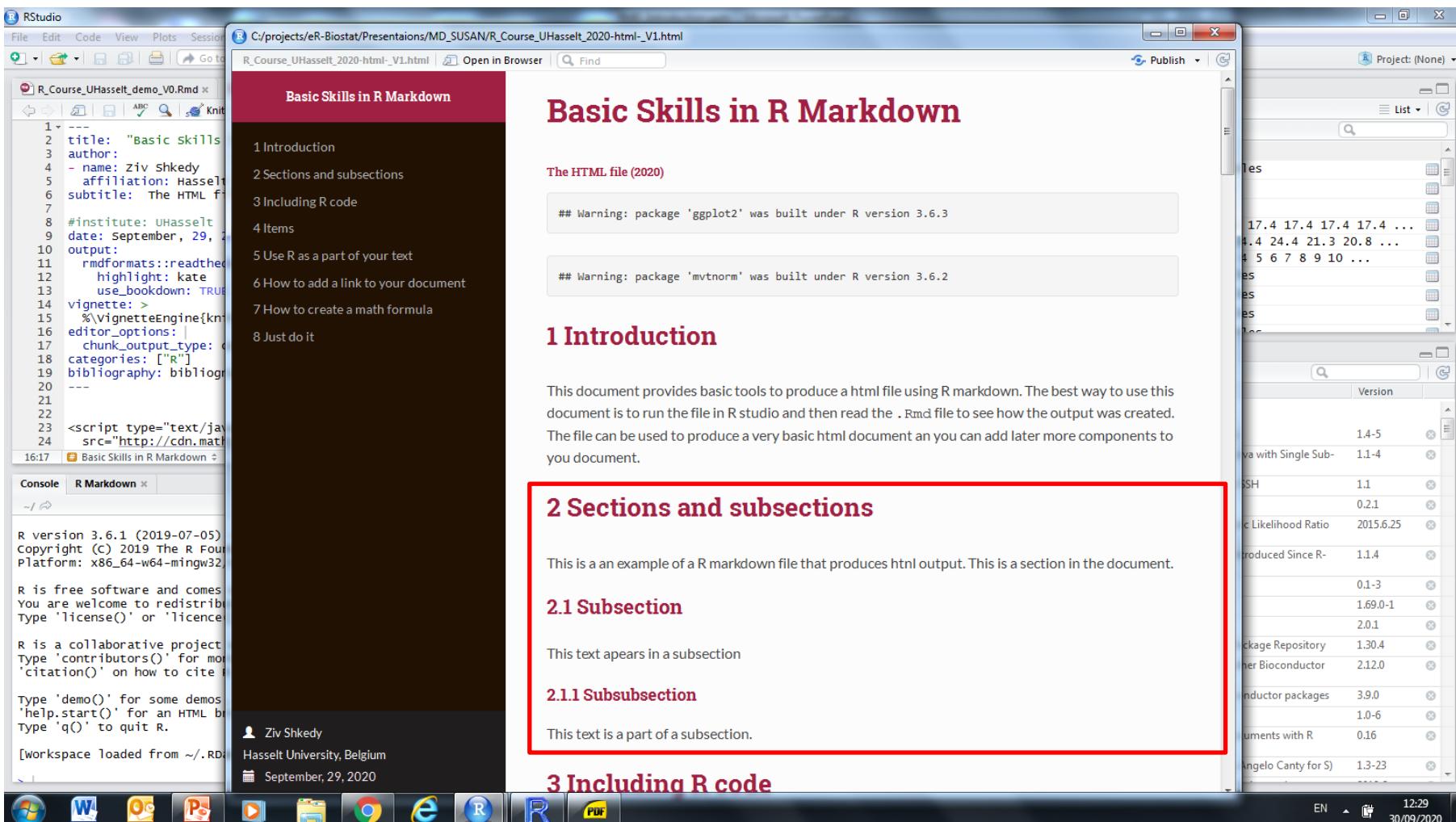
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[workspace loaded from ~/RData]
```

Environment:

| additivityTests | Additivity Tests in the Two Way Anova with Single Sub-class Numbers | 1.1-4     |
|-----------------|---------------------------------------------------------------------|-----------|
| askpass         | Safe Password Entry for R, Git, and SSH                             | 1.1       |
| assertthat      | Easy Pre and Post Assertions                                        | 0.2.1     |
| asypow          | Calculate Power Utilizing Asymptotic Likelihood Ratio Methods       | 2015.6.25 |
| backports       | Reimplementations of Functions Introduced Since R-3.0.0             | 1.1.4     |
| base64enc       | Tools for base64 encoding                                           | 0.1-3     |
| BH              | Boost C++ Header Files                                              | 1.69.0-1  |
| bclust          | BiCluster Algorithms                                                | 2.0.1     |
| BiocManager     | Access the Bioconductor Project Package Repository                  | 1.30.4    |
| BiocStyle       | Standard styles for vignettes and other Bioconductor documents      | 2.12.0    |
| BiocVersion     | Set the appropriate version of Bioconductor packages                | 3.9.0     |
| bitops          | Bitwise Operations                                                  | 1.0-6     |
| bookdown        | Authoring Books and Technical Documents with R Markdown             | 0.16      |
| boot            | Bootstrap Functions (Originally by Angelo Canty for S)              | 1.3-23    |

# The output



# R code in the output

The screenshot shows the RStudio interface. On the left, the R Markdown editor displays the following R code:

```
63
64
65
66 # Including R code
67
68 ## Print R code and output
69
70 This is an example how to include R code and output in the document. We use the <tt>airquality</tt> data as an
example.
71
72
73 ``{r}
74 x<-na.omit(airquality$ozone)
75 print(x)
76 ``
77
78 ## Do not print the R code but print the output
79
80 If we do not want to print the R code, but we want to see the output use the option <tt>echo=FALSE</tt>:
81
82 ``{r,echo=FALSE}
83 x<-na.omit(airquality$ozone)
84 print(x)
85
```

The right side of the interface shows the RStudio environment pane, which includes a system library browser. A red bracket on the right margin of the code editor points towards the environment pane, and a red callout box contains the text:

How to include R code and output

The RStudio toolbar at the bottom includes icons for various applications like Word, Excel, and R.

# The output

The screenshot shows the RStudio interface with a browser window displaying the output of an R Markdown document. A red box highlights the code block and its corresponding output.

**RStudio Environment:**

- File Edit Code View Plots Session Help**
- Project: (None)**
- Console R Markdown**
- Workspace loaded from ~/.RData**
- Ziv Shkedy**
- Hasselt University, Belgium**
- September, 29, 2020**

**Browser Window Content:**

**C:/projects/eR-Biostat/Presentations/MD\_SUSAN/R\_Course\_UHasselt\_2020-html\_V1.html**

This is an example of a R Markdown file that produces HTML output. This is a section in the document.

## 2.1 Subsection

This text appears in a subsection

### 2.1.1 Subsubsection

This text is a part of a subsection.

## 3 Including R code

### 3.1 Print R code and output

This is an example how to include R code and output in the document. We use the airquality data as an example.

```
```{r}
x<-na.omit(airquality$Ozone)
print(x)
```



```

#> #> [1] 41 36 12 18 28 23 19 8 7 16 11 14 18 14 34 6 30 11 1 11 4 32
#> #> [24] 45 115 37 29 71 39 23 21 37 20 12 13 135 49 32 64 40 77 97 97 85 10
#> #> [47] 7 48 35 61 79 63 16 80 108 20 52 82 50 64 59 39 9 16 78 35 66 122
#> #> [70] 110 44 28 65 22 59 23 31 44 21 9 45 168 73 76 118 84 85 96 78 73 91
#> #> [93] 32 20 23 21 24 44 21 28 9 13 46 18 13 24 16 13 23 36 7 14 30 14
#> #> [116] 20
#> #> attr(,"na.action")
#> #> [1] 5 10 25 26 27 32 33 34 35 36 37 39 42 43 45 46 52 53 54 55 56 57
#> #> [25] 60 61 65 72 75 83 84 102 103 107 115 119 150
#> #> attr(,"class")
#> #> [1] "omit"

```


```

# Development of a **PDF/HTML** document

- The output:
  - PDF/HTML document (the user can choose).

To run the analysis in the example, use the program in R Studio:

[`R\_course\_UHasselt\_2021\(pdf\)\_V1.Rmd`](#)

# The program

R\_course\_UHasselt\_2021(pdf)\_V1.Rmd

A screenshot of the RStudio interface. On the left, the code editor shows an R Markdown file named 'R\_course\_UHasselt\_2021(pdf)\_V1.Rmd'. A red arrow points from the text 'Produce a PDF output' to the 'Knit' button in the toolbar above the editor. The 'Knit' button has a dropdown menu open, showing options like 'Knit to HTML', 'Knit to PDF', and 'Knit to Word'. The 'Knit to PDF' option is highlighted. Below the editor, the R console displays the standard R startup message. To the right, the Global Environment pane lists various R objects and their details, and the Packages pane shows a list of installed packages.

Produce a PDF output

R Course UHasselt demo V0.Rmd x R Course UHasselt demo V1.Rmd x R Course UHasselt 2020(html)\_V1.Rmd x R Course UHasselt 2020(pdf)\_V1.Rmd x

File Edit Code View Plots Session Build Debug Profile Tools Help

Knit

1 to 20)

16999 obs. of 12 variables

21 obs. of 2 variables

12 obs. of 2 variables

num [1:100, 1:100] 17.4 17.4 17.4 17.4 17.4 ...

int [1:30, 1:30] 1 2 3 4 5 6 7 8 9 10 ...

40001 obs. of 4 variables

40001 obs. of 4 variables

40001 obs. of 4 variables

16000 obs. of 12 variables

abind Additivity Tests in the Two Way Anova with Single Sub-class Numbers

askpass Safe Password Entry for R, Git, and SSH

assertthat Easy Pre and Post Assertions

asypow Calculate Power Utilizing Asymptotic Likelihood Ratio Methods

backports Reimplementations of Functions Introduced Since R-3.0.0

base64enc Tools for base64 encoding

BH Boost C++ Header Files

biclust BiCluster Algorithms

BiocManager Access the Bioconductor Project Package Repository

BiocStyle Standard styles for vignettes and other Bioconductor documents

BiocVersion Set the appropriate version of Bioconductor packages

bitops Bitwise Operations

bookdown Authoring Books and Technical Documents with R Markdown

boot Bootstrap Functions (Originally by Angelo Canty for S)

R Markdown

R

EN 12:39  
30/09/2020

# The PDF output

The image shows a Windows desktop with three main windows open:

- RStudio:** On the left, the RStudio interface is visible. It shows an R Markdown file named `_Course_UHasselt_demo_V0.Rmd` with the following code:

```
1 ---
2 title: 'Basic skills in R Markdown: the pdf file'
3 output:
4 pdf_document: default
5 html_document: default
6 word_document: default
7 subtitle: ziv shkedy (2020)
8 layout: page
9 toc: yes
10 ---
11 ---
12 output: html_document
13 use_bookdown: TRUE
14 ---
15
16 ```{r, echo=FALSE}
17 library(e1071)
18 library(lattice)
19 library(ggplot2)
20 library(mvtnorm)
21
22
23 \newpage
24
16:18 [Chunk 1]
```

- PDF Viewer:** In the center, a PDF viewer window titled "R\_Course\_UHasselt\_2020-pdf\_V1.pdf" displays the generated PDF document. The title is "Basic Skills in R Markdown: the pdf file" by Ziv Shkedy (2020). The PDF contains a table of contents with various sections and their page numbers.
- R Help Browser:** On the right, the R Help browser is open, showing a list of packages and their versions. The list includes:

  - 16999 obs. of 12 variables
  - 21 obs. of 2 variables
  - 12 obs. of 2 variables
  - num [1:100, 1:100] 17.4 17.4 17.4 17.4 17.4 ...
  - num [1:10, 1:10] 24.7 24.4 24.4 21.3 20.8 ...
  - int [1:30, 1:30] 1 2 3 4 5 6 7 8 9 10 ...
  - 40001 obs. of 4 variables
  - 40001 obs. of 4 variables
  - 40001 obs. of 4 variables
  - 16000 obs. of 12 variables
  - Combine Multidimensional Arrays 1.4-5
  - Additivity Tests in the Two Way Anova with Single Sub-class Numbers 1.1-4
  - Safe Password Entry for R, Git, and SSH 1.1
  - Easy Pre and Post Assertions 0.2.1
  - Calculate Power Utilizing Asymptotic Likelihood Ratio Methods 2015.6.25
  - Reimplementations of Functions Introduced Since R-3.0.0 1.1.4
  - Tools for base64 encoding 0.1-3
  - Boost C++ Header Files 1.69.0-1
  - BiCluster Algorithms 2.0.1
  - Access the Bioconductor Project Package Repository 1.30.4
  - Standard styles for vignettes and other Bioconductor documents 2.12.0
  - Set the appropriate version of Bioconductor packages 3.9.0
  - Bitwise Operations 1.0-6
  - Authoring Books and Technical Documents with R Markdown 0.16
  - Bootstrap Functions (Originally by Angelo Canty for S) 1.3-23

Table of contents:  
see next slide

# The program

A screenshot of the RStudio interface. On the left, the R Markdown editor shows a code block with the line `toc: yes` highlighted by a red arrow. The code also includes library imports like `library(e1071)`, `library(lattice)`, `library(ggplot2)`, and `library(mvtnorm)`. The right side of the interface shows the Environment pane with a list of objects: `dat`, `dat1`, `data1`, `datmat`, `dattest`, `idd`, `out`, `outp02`, `outp04`, and `param1`. Below it is the System Library pane listing various R packages. The bottom of the screen shows the Windows taskbar with icons for Word, Excel, and R.

toc: yes Produce the table of contents

```
1 ---
2 title: 'Basic Skills in R Markdown: the pdf file'
3 output:
4 pdf_document: default
5 html_document: default
6 word_document: default
7 subtitle: zivishkedy (2020)
8 layout: page
9 toc: yes
10 ---
11 ---
12 output: html_document
13 use_bookdown: TRUE
14 ---
15
16 ```{r, echo=FALSE}
17 library(e1071)
18 library(lattice)
19 library(ggplot2)
20 library(mvtnorm)
21 ````|
22
23 \newpage
24
21:4 | (Top Level) ▾ R Markdown ▾
```

R version 3.6.1 (2019-07-05) -- "Action of the Toes"  
Copyright (C) 2019 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]  
> |

Environment History  
Import Dataset Addins Project: (None)  
Global Environment  
Data  
dat 16999 obs. of 12 variables  
dat1 21 obs. of 2 variables  
data1 12 obs. of 2 variables  
datmat num [1:100, 1:100] 17.4 17.4 17.4 17.4 17.4 ...  
dattest num [1:10, 1:10] 24.7 24.4 24.4 21.3 20.8 ...  
idd int [1:30, 1:30] 1 2 3 4 5 6 7 8 9 10 ...  
out 40001 obs. of 4 variables  
outp02 40001 obs. of 4 variables  
outp04 40001 obs. of 4 variables  
param1 16000 obs. of 12 variables  
Files Plots Packages Help Viewer  
Install Update  
Name Description Version  
System Library  
abind Combine Multidimensional Arrays 1.4-5  
additivityTests Additivity Tests in the Two Way Anova with Single Sub-class Numbers 1.1-4  
askpass Safe Password Entry for R, Git, and SSH 1.1  
assertthat Easy Pre and Post Assertions 0.2.1  
asypow Calculate Power Utilizing Asymptotic Likelihood Ratio Methods 2015.6.25  
backports Reimplementations of Functions Introduced Since R-3.0.0 1.1.4  
base64enc Tools for base64 encoding 0.1-3  
BH Boost C++ Header Files 1.69.0-1  
bicluster BiCluster Algorithms 2.0.1  
BiocManager Access the Bioconductor Project Package Repository 1.30.4  
BiocStyle Standard styles for vignettes and other Bioconductor documents 2.12.0  
BiocVersion Set the appropriate version of Bioconductor packages 3.9.0  
bitops Bitwise Operations 1.0-6  
bookdown Authoring Books and Technical Documents with R Markdown 0.16  
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-23

# The program

A section that explains how to include a figure in the document.

90 var(x)  
91  
92 ## Graphical displays in the document  
93 A histogram for the ozone level can be produced using the function <t  
"histogram"></tt>:  
94  
95 ``{r}  
96 Ozone.R<-data.frame(x)  
97 qplot(x, data = Ozone.R, geom = "histogram", binwidth = 0.1)  
98  
99 To add a caption to the figure we use <tt>[r figchp1,fig.cap="Sepal length (III)"]</tt>.  
100  
101  
102  
103 we can refer to the figure from the text in the document. For example, Figure \@ref(fig:figchp2) presents a  
histogram that was produced using the function <tt>qplot()</tt> function.  
104  
105 ``{r figchp1,fig.cap="Sepal length (III)"}  
106 Ozone.R<-data.frame(x)  
107 qplot(x, data = Ozone.R, geom = "histogram", binwidth = 0.1)  
108  
109  
110  
111  
20:17 [ Chunk 1 ]

R version 3.6.1 (2019-07-05) -- "Action of the Toes"  
Copyright (c) 2019 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]

> |

99 obs. of 12 variables  
obs. of 2 variables  
obs. of 2 variables  
[1:100, 1:100] 17.4 17.4 17.4 17.4 17.4 ...  
[1:10, 1:10] 24.7 24.4 24.4 21.3 20.8 ...  
int [1:30, 1:30] 1 2 3 4 5 6 7 8 9 10 ...  
idd  
out  
outp02  
outp04  
scandat1  
40001 obs. of 4 variables  
40001 obs. of 4 variables  
40001 obs. of 4 variables  
16000 obs. of 12 variables

Files Plots Packages Help Viewer

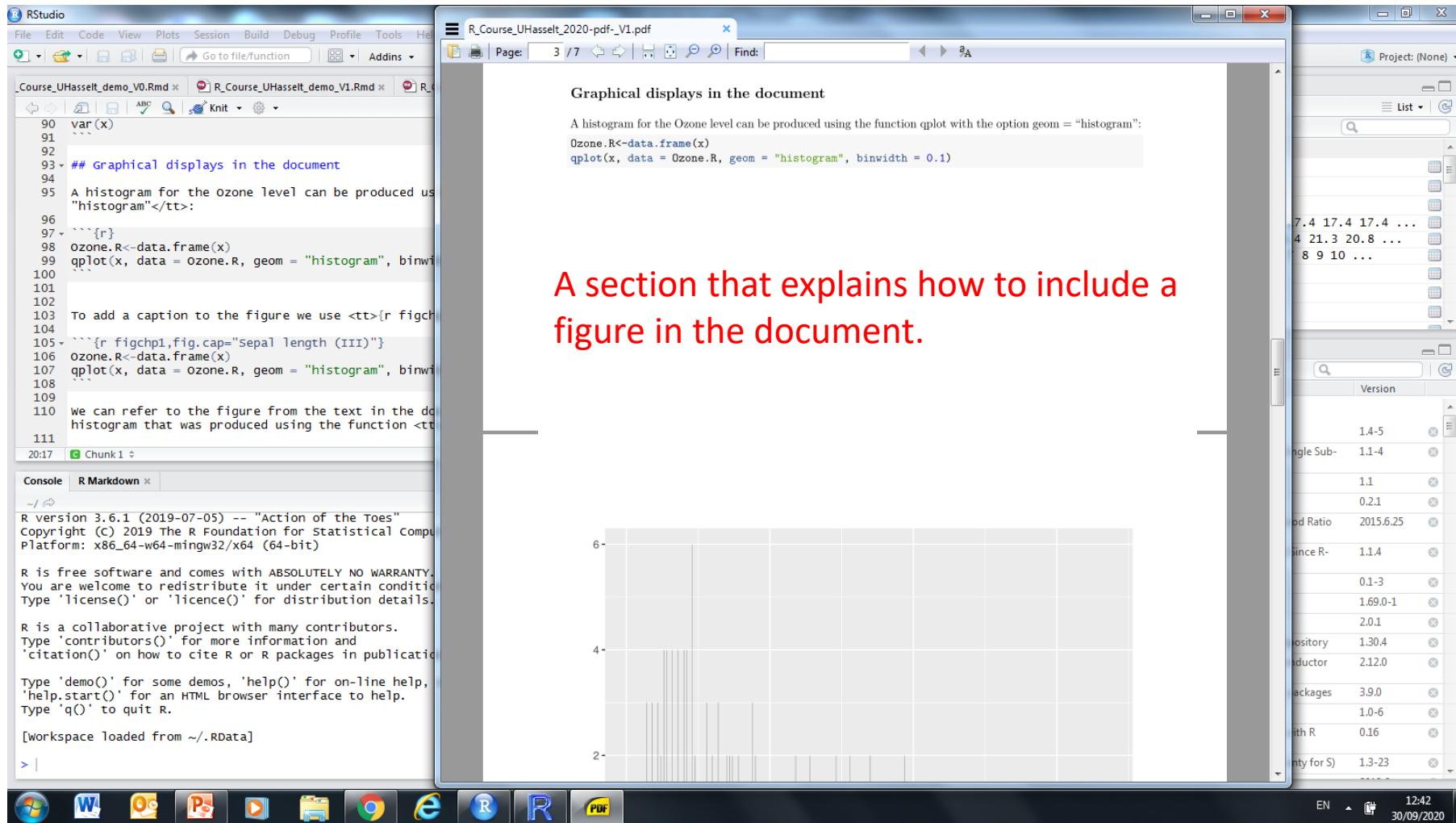
Name Description Version

System Library

| Name            | Description                                                         | Version   |
|-----------------|---------------------------------------------------------------------|-----------|
| abind           | Combine Multidimensional Arrays                                     | 1.4-5     |
| additivityTests | Additivity Tests in the Two Way Anova with Single Sub-class Numbers | 1.1-4     |
| askpass         | Safe Password Entry for R, Git, and SSH                             | 1.1       |
| assertthat      | Easy Pre and Post Assertions                                        | 0.2.1     |
| asypow          | Calculate Power Utilizing Asymptotic Likelihood Ratio Methods       | 2015.6.25 |
| backports       | Reimplementations of Functions Introduced Since R-3.0.0             | 1.1.4     |
| base64enc       | Tools for base64 encoding                                           | 0.1-3     |
| BH              | Boost C++ Header Files                                              | 1.69.0-1  |
| biclust         | BiCluster Algorithms                                                | 2.0.1     |
| BiocManager     | Access the Bioconductor Project Package Repository                  | 1.30.4    |
| BiocStyle       | Standard styles for vignettes and other Bioconductor documents      | 2.12.0    |
| BiocVersion     | Set the appropriate version of Bioconductor packages                | 3.9.0     |
| bitops          | Bitwise Operations                                                  | 1.0-6     |
| bookdown        | Authoring Books and Technical Documents with R Markdown             | 0.16      |
| boot            | Bootstrap Functions (Originally by Angelo Canty for S)              | 1.3-23    |

EN 12:43 30/09/2020

# The PDF output



## Example 3

### Very simple analysis

To run the analysis in the example, use the program in R Studio:

[R\\_course\\_UHasselt\\_2021\(pdf\)\\_V2.Rmd](#)

# Produce a report for an analysis

- How to use Rmd to produce a report about an analysis ?
- Why this is important ?
- Example: the old faithful dataset.
  - Part 1: the analysis.
  - Part 2: the analysis + report using Rmd (you need to run the program
    - [R\\_course\\_UHasselt\\_2020\(pdf\)\\_V2.Rmd](#)

# Exploratory analysis of the Old Faithful dataset

- Old Faithful is a geyser that is found in Yellowstone National Park, in Wyoming.
- It is one of the most famous attractions at Yellowstone National Park.
- Old Faithful is unique because of how long and how often it erupts.



# Exploratory analysis of the Old Faithful dataset

- Our aim is to explore the association between the time between eruptions (waiting time) and the duration of the eruption.
- For the analysis we use **basic graphical R functions** and **functions to calculate descriptive statistics** for the data.
- Output: PDF format.

# The data in R

```
>head(faithful)
```

The name of the data in R

```
eruptions waiting
1 3.600 79
2 1.800 54
3 3.333 74
4 2.283 62
5 4.533 85
6 2.883 55
```

A data frame with two variables:  
eruption and waiting time.

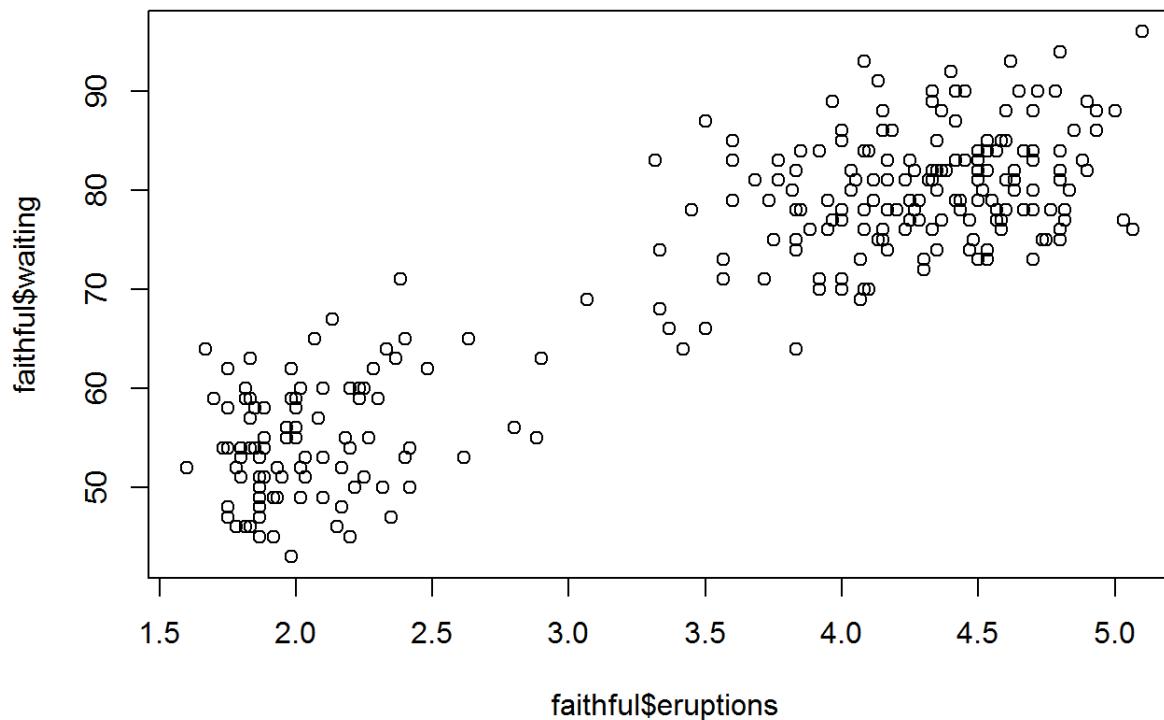
# Eruption time and waiting time

```
plot(faithful$eruptions, faithful$waiting)
```



Basic graphical  
function in R

```
plot(x, y)
```

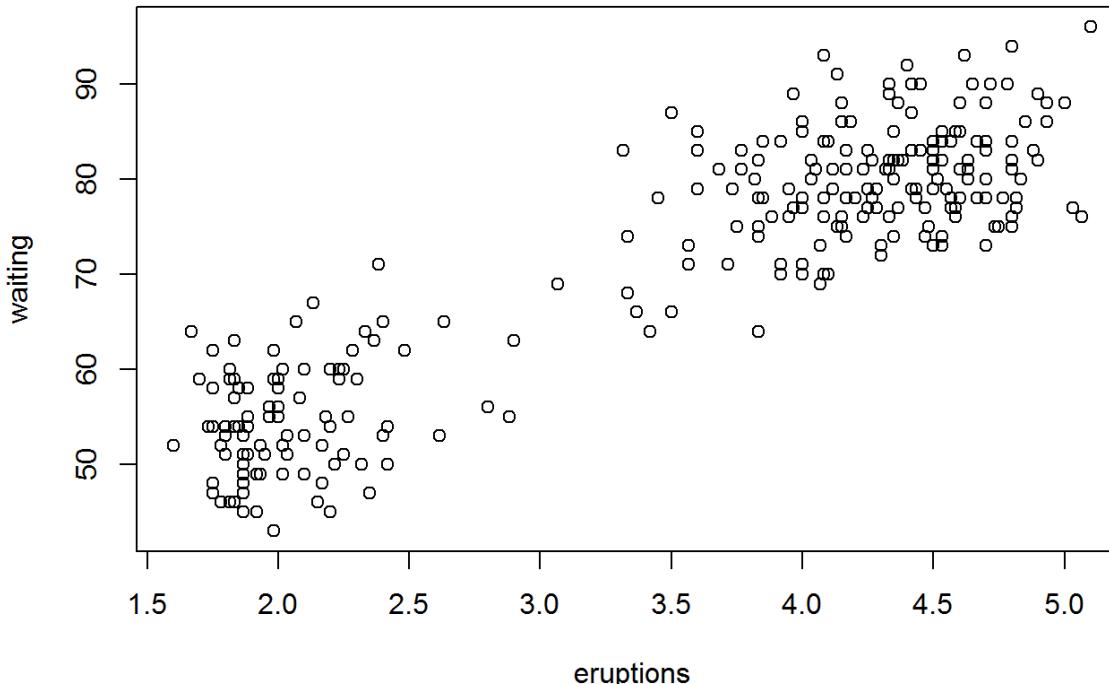


# Eruption time and waiting time

```
plot(faithful$eruptions, faithful$waiting,
 xlab="eruptions", ylab="waiting")

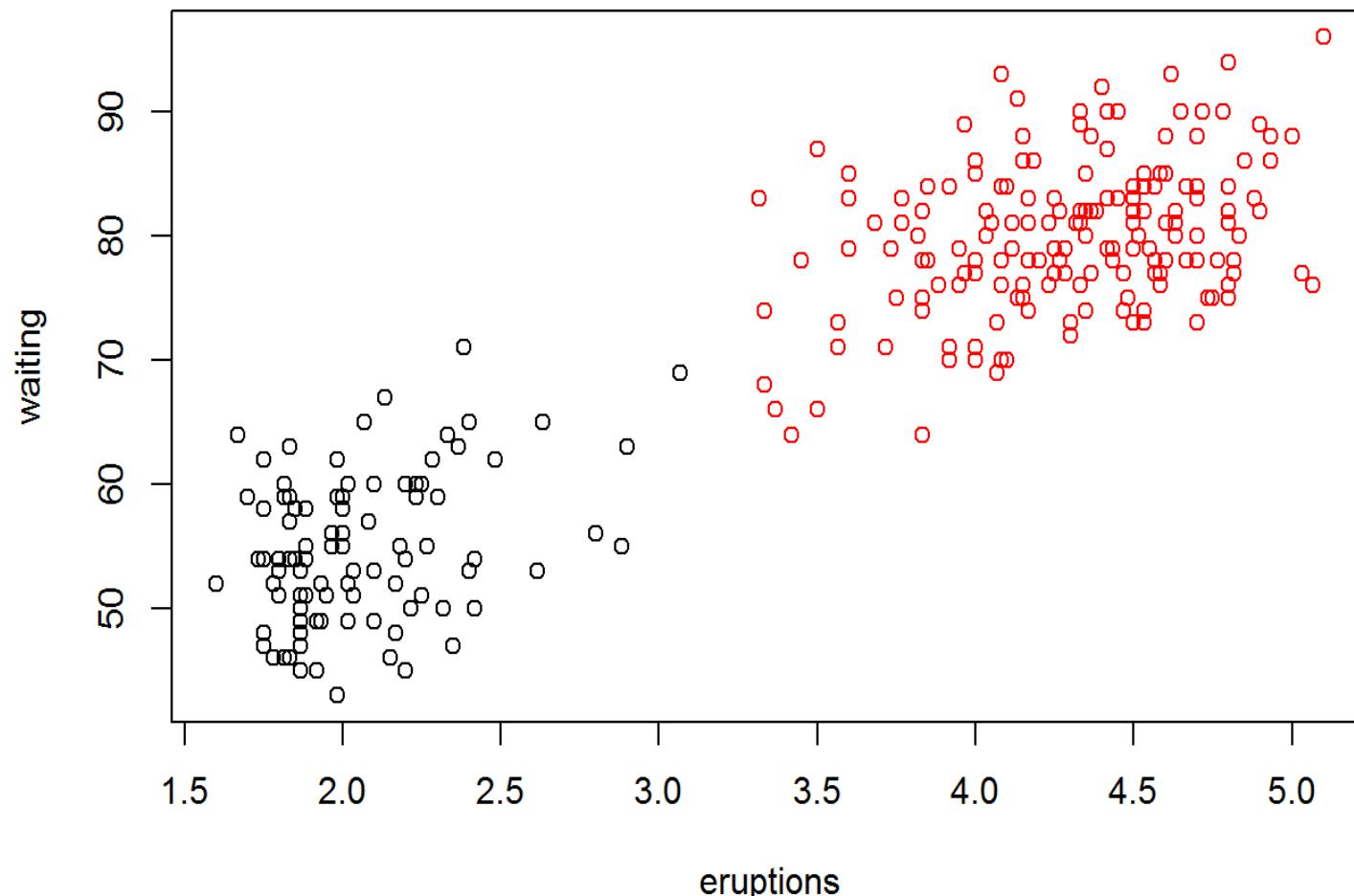
title("The old faithful data")
```

Adding title and text  
for the labels.



# Two clusters ?

The old faithful data



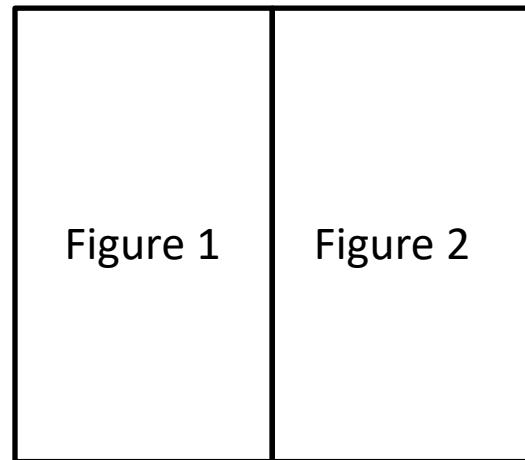
# Distribution of eruption time

```
par(mfrow=c(1, 2))
```

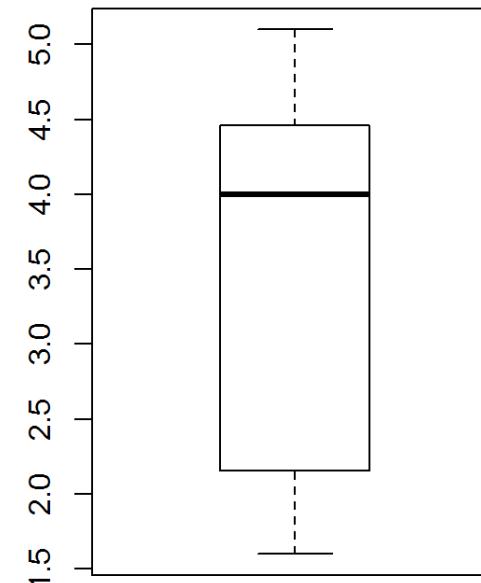
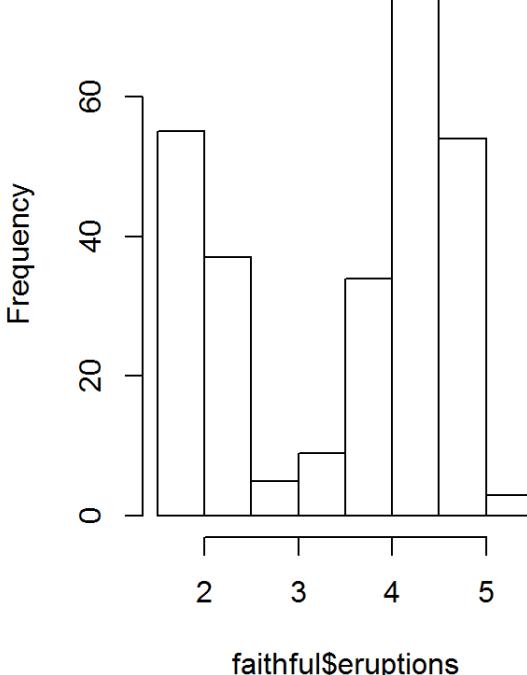
Split the graphical window

```
hist(faithful$eruptions)
boxplot(faithful$eruptions)
```

Split the graphical window



Histogram of faithful\$eruptions

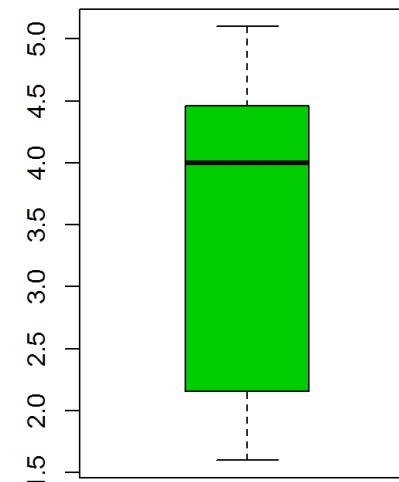
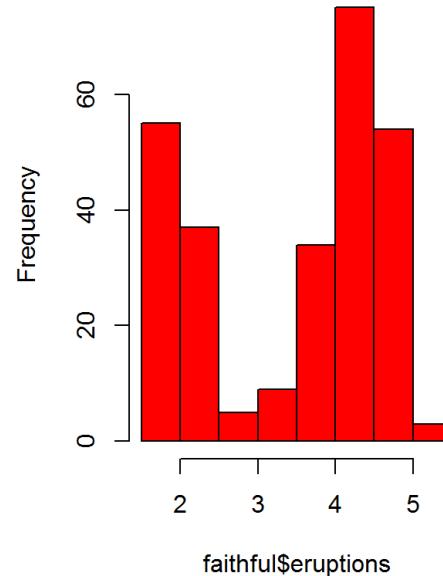


# Distribution of eruption time

```
par(mfrow=c(1, 2))
hist(faithful$eruptions, col=2)
boxplot(faithful$eruptions, col=3)
```

Change plot settings.

Histogram of faithful\$eruptions

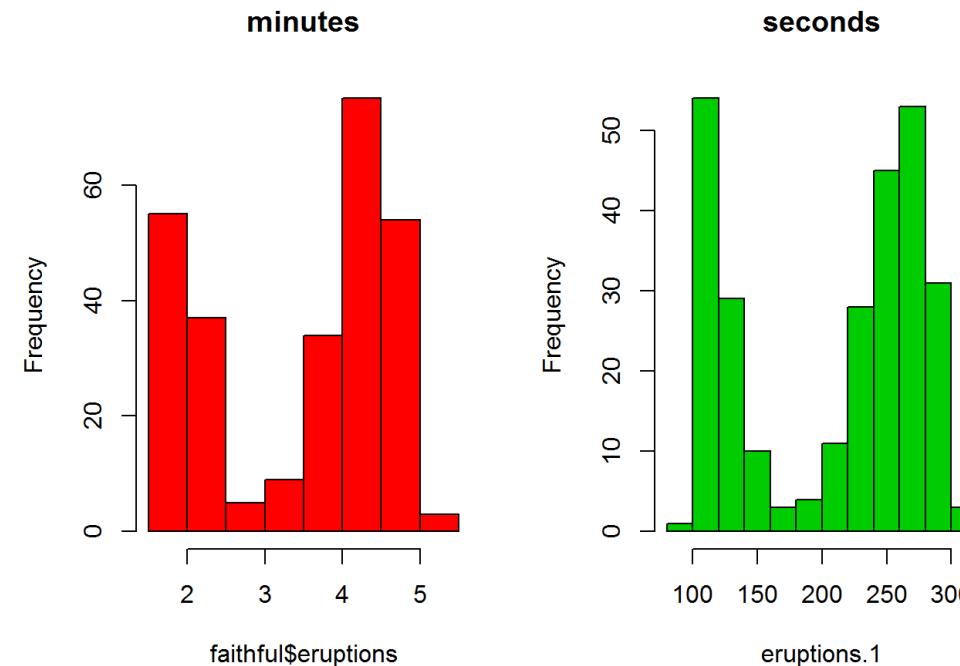


# Distribution of eruption time in seconds

```
eruptions.1<-faithful$eruptions*60
par(mfrow=c(1,2))
hist(faithful$eruptions, col=2,main="minutes")
hist(eruptions.1, col=3,main="seconds")
```

Create a new object:

**eruptions.1**



# The Rmd program

- Repeat the analysis above.
- Produce a report for the analysis.
- Output: PDF format.
- Run the analysis in your own laptop to see the report.
- Rmd program:

R\_course\_UHasselt\_2021(pdf)\_V2.Rmd

# The R markdown program & PDF output

## R\_course\_UHasselt\_2021(pdf)\_V2.Rmd

The screenshot shows the RStudio interface with two main windows. On the left, the R Markdown editor displays the following R code:

```
25 # Introduction
26
27 old Faithful is a geyser that is found in Yellowstone National Park, in wyom
28 attractions at Yellowstone National Park. Old Faithful is unique because of
29 aim is to explore the time between eruptions and the duration of the eruption
30 graphical R functions and functions to calculate descriptive statistics for
31 # The data
32
33 Data recorded from 1990 measured the time between eruptions and the duration
34 minutes. This data set was based on 272 observations of Old Faithful's eruptions.
35 eruption time in mins and waiting time to next eruption (in mins). Both vari
36 ````{r}
37 head(faithful)
38
39
40
41 # Analysis of the old faithful data
42
43 ## Eruption time and waiting time
```

A red arrow points from the line `head(faithful)` to the corresponding output in the PDF window. The PDF window on the right shows the generated document:

**Introduction**

Old Faithful is a geyser that is found in Yellowstone National Park, in Wyoming. It is one of the most famous attractions at Yellowstone National Park. Old Faithful is unique because of how long and how often it erupts. Our aim is to explore the time between eruptions and the duration of the eruption. For the analysis we use basic graphical R functions and functions to calculate descriptive statistics for the data.

**The data**

Data recorded from 1990 measured the time between eruptions and the duration of the eruption, both taken in minutes. This data set was based on 272 observations of Old Faithful's eruptions and consists of 2 variables: eruption time in mins and waiting time to next eruption (in mins). Both variables are numeric.

**Analysis of the old faithful data**

Eruption time and waiting time

By looking at the scatter plot, we can see a visual representation of the data.

```
plot(faithful$eruptions,faithful$waiting)
```

The PDF window also includes a red box highlighting the data frame output:

|      | eruptions | waiting |
|------|-----------|---------|
| ## 1 | 2.880     | 79      |
| ## 2 | 1.800     | 54      |
| ## 3 | 3.333     | 74      |
| ## 4 | 2.283     | 62      |
| ## 5 | 4.533     | 85      |
| ## 6 | 2.883     | 55      |

# The R markdown program & PDF output

The screenshot shows the RStudio interface with two main windows. On the left, the 'Code' window displays R code for generating a scatterplot of 'faithful' data. A red arrow points from the text 'The output' in the PDF window back to the R code in the 'Code' window. The 'Console' window at the bottom shows the command used to generate the PDF and the output of pandoc.exe indicating a parsing error. On the right, the 'PDF' window shows the resulting PDF document. The PDF contains the R code as part of the text, followed by a red box highlighting the plot command, and a scatterplot titled 'The old faithful data' with axes labeled 'eruptions' and 'waiting'. The word 'The output' is overlaid in red text on the plot area.

R code as a part of your text in the output

```
[1] 96
and the minimum eruption time is
min(faithful$eruptions)
[1] 1.6
Let us look once again in the scatterplot. We can add labels name and title
plot(faithful$eruptions,faithful$waiting,
xlab="eruptions",ylab="waiting")
title("The old faithful data")
```

The old faithful data

The output

Waiting vs Eruptions Scatterplot

In the next figure we use different colors for the two clusters.

```
plot(faithful$eruptions,faithful$waiting,
xlab="eruptions",ylab="waiting")
```

# What did we see today ?

- R Studio.
- R markdown + output.
- Basic graphical functions in R and how to control the figure (title, colors etc).
- Do the analysis in the file:

**R\_course\_UHasselt\_2021 (pdf) \_v2 .Rmd**

# Extra sildes

Source vs Visual Editor

# Source vs Visual Editor

The screenshot shows the RStudio interface with a blue arrow pointing to the 'Knit' button in the toolbar above the code editor.

**Code Editor (Source View):**

```
1 ---
2 title: "Demo 2"
3 author: "Rudradev Sengupta, Bernard Lang'ir and Ziv Shkedy"
4 date: "05 October 2024"
5 output:
6 html_document: default
7 pdf_document: default
8 word_document: default
9 params:
10 snapshot: lubridate::ymd_hms("2015-01-01 12:30:00")
11 start: lubridate::ymd("2015-01-01")
12 ---
13
14 ``{r setup, include=FALSE}
15 knitr::opts_chunk$set(echo = TRUE)
16 ```
17
18 ## R Markdown
19
20 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and
MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
21
22 When you click the Knit button a document will be generated that includes both content as well as
the output of any embedded R code chunks within the document. You can embed an R code chunk like
this:
23
24 # Demo 2
```

**Environment View:**

Environment is empty

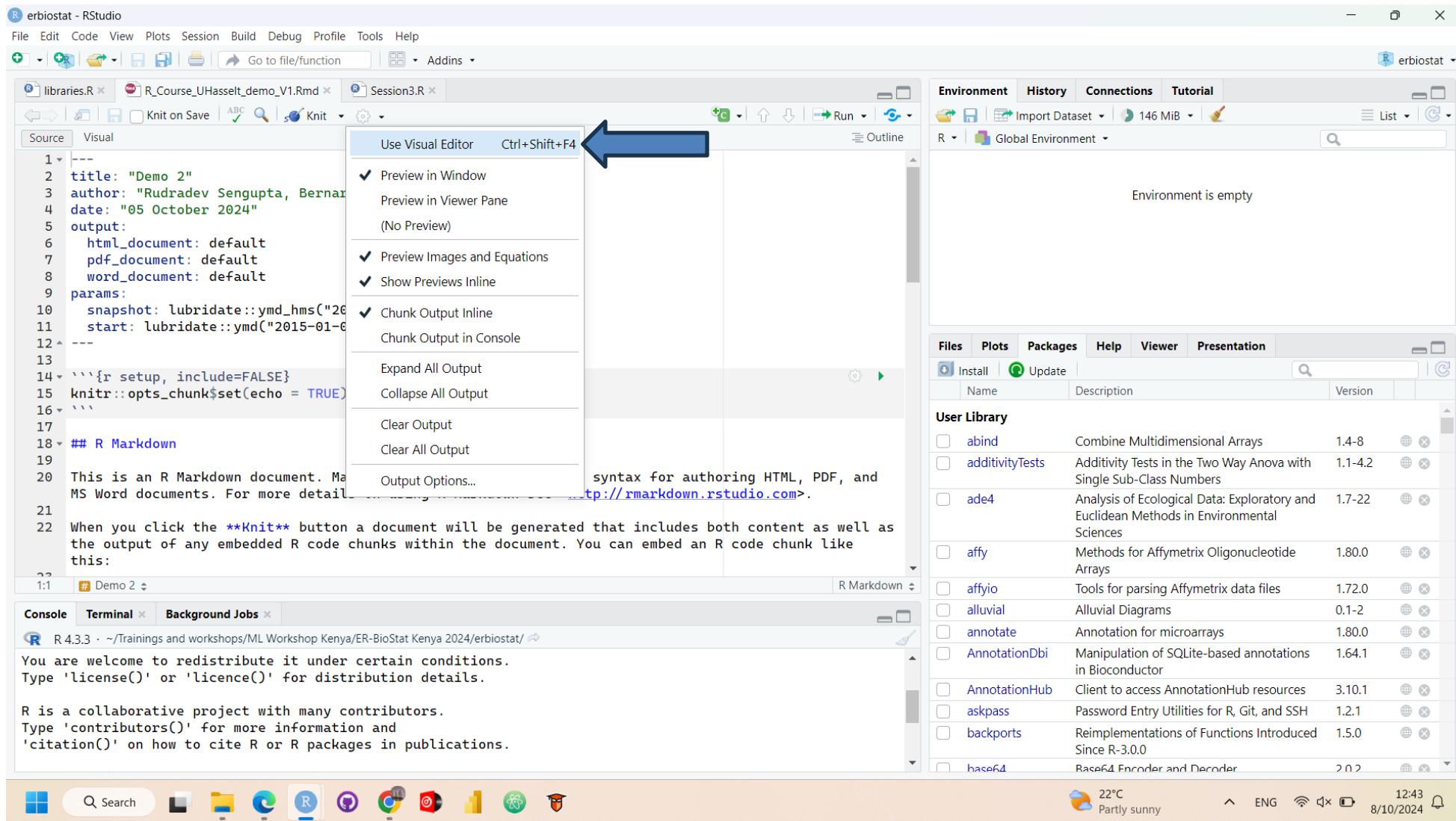
**Packages View:**

| Name            | Description                                                                              | Version | Action  |
|-----------------|------------------------------------------------------------------------------------------|---------|---------|
| abind           | Combine Multidimensional Arrays                                                          | 1.4-8   | Install |
| additivityTests | Additivity Tests in the Two Way Anova with Single Sub-Class Numbers                      | 1.1-4.2 | Install |
| ade4            | Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences | 1.7-22  | Install |
| affy            | Methods for Affymetrix Oligonucleotide Arrays                                            | 1.80.0  | Install |
| affyio          | Tools for parsing Affymetrix data files                                                  | 1.72.0  | Install |
| alluvial        | Alluvial Diagrams                                                                        | 0.1-2   | Install |
| annotate        | Annotation for microarrays                                                               | 1.80.0  | Install |
| AnnotationDbi   | Manipulation of SQLite-based annotations in Bioconductor                                 | 1.64.1  | Install |
| AnnotationHub   | Client to access AnnotationHub resources                                                 | 3.10.1  | Install |
| askpass         | Password Entry Utilities for R, Git, and SSH                                             | 1.2.1   | Install |
| backports       | Reimplementations of Functions Introduced Since R-3.0.0                                  | 1.5.0   | Install |
| base64d         | Base64 Encoder and Decoder                                                               | 202     | Install |

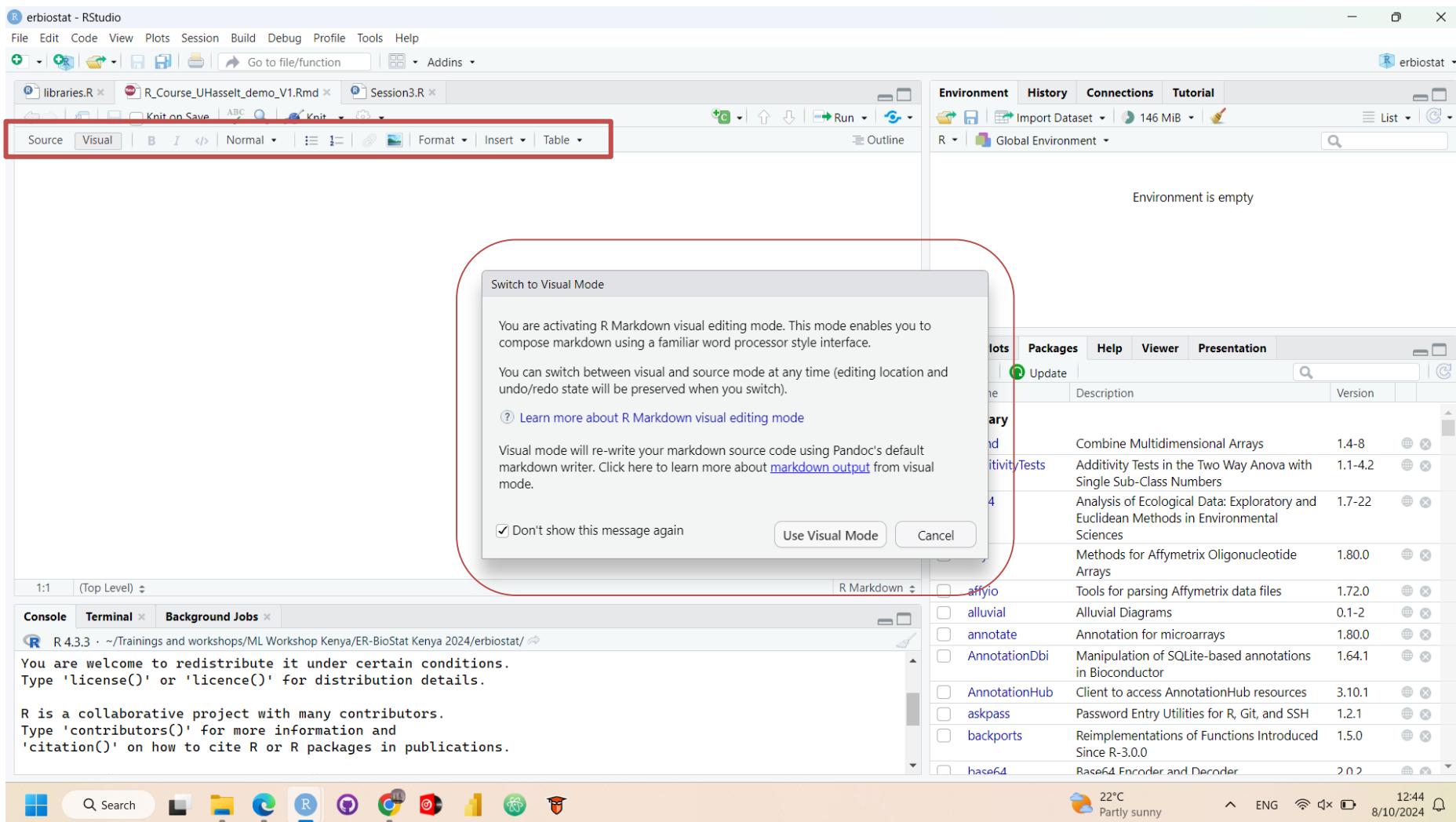
**System Tray:**

22°C Partly sunny 12:43 8/10/2024

# Source vs Visual Editor



# Source vs Visual Editor



# Source vs Visual Editor

The screenshot shows the RStudio interface with the Source tab selected in the top-left panel. The main code editor displays R Markdown code:

```

title: "Demo 2"
author: "Rudradev Sengupta, Bernard Osang'ir and Ziv Shkedy"
date: "05 October 2024"
output:
 html_document: default
 pdf_document: default
 word_document: default
params:
 snapshot: lubridate::ymd_hms("2015-01-01 12:30:00")
 start: lubridate::ymd("2015-01-01")
---|

{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

The R Markdown preview pane below shows the rendered content:

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

The Environment pane on the right indicates that the environment is empty.

The Packages pane lists the installed packages:

| Name            | Description                                                                              | Version |
|-----------------|------------------------------------------------------------------------------------------|---------|
| abind           | Combine Multidimensional Arrays                                                          | 1.4-8   |
| additivityTests | Additivity Tests in the Two Way Anova with Single Sub-Class Numbers                      | 1.1-4.2 |
| ade4            | Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences | 1.7-22  |
| affy            | Methods for Affymetrix Oligonucleotide Arrays                                            | 1.80.0  |
| affyio          | Tools for parsing Affymetrix data files                                                  | 1.72.0  |
| alluvial        | Alluvial Diagrams                                                                        | 0.1-2   |
| annotate        | Annotation for microarrays                                                               | 1.80.0  |
| AnnotationDbi   | Manipulation of SQLite-based annotations in Bioconductor                                 | 1.64.1  |
| AnnotationHub   | Client to access AnnotationHub resources                                                 | 3.10.1  |
| askpass         | Password Entry Utilities for R, Git, and SSH                                             | 1.2.1   |
| backports       | Reimplementations of Functions Introduced Since R-3.0.0                                  | 1.5.0   |
| base64          | Base64 Encoder and Decoder                                                               | 202     |

The bottom status bar shows system information: 22°C, Partly sunny, 12:44, ENG, 8/10/2024.

# Source vs Visual Editor

The screenshot shows the RStudio interface with two main panes. The left pane is the 'Source' editor, which displays R Markdown code. The right pane is the 'Visual' editor, which displays the generated HTML output. A red box highlights the 'Knit' button in the Source editor toolbar. A blue arrow points from the 'Knit' button in the Source editor to the 'Knit' button in the Visual editor toolbar.

**Demo 2**  
Rudradev Sengupta, Bernard Osang'ir and Ziv Shkedy  
05 October 2024

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the Knit button a document will be generated that includes both content as well as the output of any **embedded R code** chunks within the document. You can embed an R code chunk like this:

```
project_start <- params$start
summary(mtcars)
```

| ##         | mpg    | cyl           | disp          | hp             |
|------------|--------|---------------|---------------|----------------|
| ## Min.    | 10.40  | Min. :4.000   | Min. :71.1    | Min. :52.0     |
| ## 1st Qu. | 15.43  | 1st Qu.:4.000 | 1st Qu.:120.8 | 1st Qu.:96.5   |
| ## Median  | 19.20  | Median :6.000 | Median :196.3 | Median :123.0  |
| ## Mean    | 20.09  | Mean :6.188   | Mean :230.7   | Mean :146.7    |
| ## 3rd Qu. | 22.80  | 3rd Qu.:8.000 | 3rd Qu.:326.0 | 3rd Qu.:180.0  |
| ## Max.    | 33.90  | Max. :8.000   | Max. :472.0   | Max. :335.0    |
| ##         | drat   | wt            | qsec          | vs             |
| ## Min.    | 2.760  | Min. :1.513   | Min. :14.50   | Min. :0.0000   |
| ## 1st Qu. | 3.080  | 1st Qu.:2.581 | 1st Qu.:16.89 | 1st Qu.:0.0000 |
| ## Median  | 3.695  | Median :3.325 | Median :17.71 | Median :0.0000 |
| ## Mean    | 3.597  | Mean :3.217   | Mean :17.85   | Mean :0.4375   |
| ## 3rd Qu. | 3.920  | 3rd Qu.:3.610 | 3rd Qu.:18.90 | 3rd Qu.:1.0000 |
| ## Max.    | 4.930  | Max. :5.424   | Max. :22.90   | Max. :1.0000   |
| ##         | am     | gear          | carb          |                |
| ## Min.    | 0.0000 | Min. :2.000   | Min. :1.000   |                |

**Including project metadata:**

```
R Markdown
Project: ML Workshop Kenya/ER-BioStat Kenya 2024
Author: Rudradev Sengupta, Bernard Osang'ir, Ziv Shkedy
Date: 05 October 2024
```

R 4.3.3 · ~/Trainings and workshops/ML Workshop Kenya/ER-BioStat Kenya 2024

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

[Workspace loaded from ~/Trainings and workshops/ML Workshop Kenya/ER-BioStat Kenya 2024]

22°C Partly sunny 12:46 8/10/2024

# Source vs Visual Editor

The screenshot shows the RStudio interface with a large blue arrow pointing from the Source Editor on the left to the Visual Editor on the right.

**Source Editor (Left):**

- File menu: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar icons: New, Open, Save, Run, Knit, Addins.
- Project navigation: libraries.R, R\_Course\_UHasselt\_demo\_V1.Rmd\*, Session3.R.
- File tabs: Go to file/function, Addins.
- Text area:
  - R Markdown**
  - This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
  - When you click the **Knit** button a document will be generated that includes output of any **embedded R code** chunks within the document. You can
  - Code chunk:

```
{r cars}
project_start <- params$start
summary(mtcars)
```
- Console tab: R Markdown, Console, Terminal, Render, Background Jobs.
- Console output:

```
R 4.3.3 · ~/Trainings and workshops/ML Workshop Kenya/ER-BioStat Kenya 2024/erbiostat/
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 ~/Trainings and workshops/ML Workshop Kenya/ER-BioStat Kenya 2024/erbiostat/.RData]
```

**Visual Editor (Right):**

- Environment tab: Environment, History, Connections, Tutorial.
- Environment pane: Import Dataset, 149 MiB, List, Global Environment (empty).
- Packages tab: Packages, Help, Viewer, Presentation.
- Packages pane: User Library (list of installed packages).
- System status bar: 22°C, Partly sunny, ENG, 12:47, 8/10/2024.

# Source vs Visual Editor

The screenshot shows the RStudio interface with two main panes demonstrating different editing modes.

**Left Pane (Source Editor):**

- Header:** R erbiostat - RStudio
- Menu Bar:** File Edit Code View Plots Session Build Debug Profile Tools Help
- Toolbar:** Go to file/function, Session3.R, Outline, Run, Knit, Addins
- Text Area:** R Markdown document content.
- Table:** A table with columns "Column", "Report", and "P-value". A blue arrow points upwards from the table, labeled "Column". The text "Direct editing" is overlaid on the table area.
- Text Below Table:** Demonstration for visual
- Code Editor:** Shows R code: 

```
fr cars
```
- Console:** Displays R version information and the R logo.

**Right Pane (Visual Editor):**

- Header:** erbiostat
- Menu Bar:** Environment History Connections Tutorial
- Toolbar:** Import Dataset, 149 MiB, List, Global Environment
- Text Area:** Environment is empty
- Packages Tab:** Shows the User Library with various R packages listed.
- System Status Bar:** 22°C, Partly sunny, ENG, 12:47, 8/10/2024

# Source vs Visual Editor

The screenshot shows the RStudio interface with two panes demonstrating the difference between Source and Visual editors.

**Left Pane (Source Editor):**

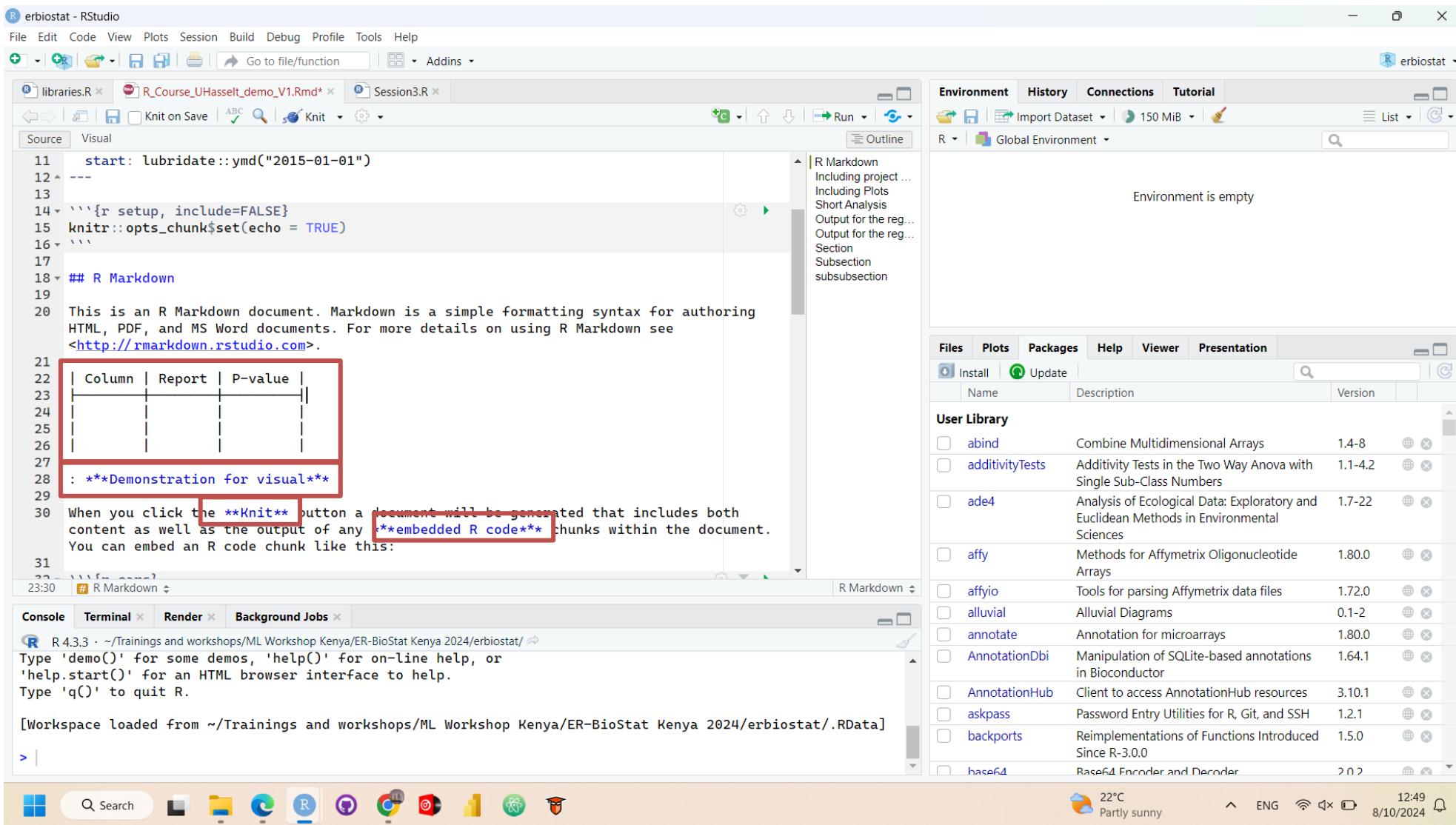
- Header:** erbiostat - RStudio
- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help
- Toolbar:** Go to file/function, Addins
- File List:** libraries.R, R\_Course\_UHasselt\_demo\_V1.Rmd, Session3.R
- Editor:** Displays R Markdown code with a red arrow pointing to the "Report" column header of a table.
- Console:** Shows R code and output related to the mtcars dataset.

**Right Pane (Visual Editor):**

- Title Bar:** ~/Trainings and workshops/ML Workshop Kenya/ER-BioStat Kenya 2024/erbiostat/R\_Course\_UHasselt... R\_Course\_UHasselt\_demo\_V1.html
- Toolbar:** Publish, Find
- Content Area:**
  - ## Demo 2
  - Rudradev Sengupta, Bernard Osang'ir and Ziv Shkedy
  - 05 October 2024
  - ## R Markdown
  - This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
  - Demonstration for visual**
  - A table with columns: Column, Report, P-value.
  - Text explaining that clicking Knit generates a document including both content and R code output.
  - An R code block: 

```
project_start <- params$start
summary(mtcars)
```
  - An R code output block showing the summary statistics for the mtcars dataset.
- Sidebar:** Shows installed packages: s (1.4-8), anova with (1.1-4.2), ratory and (1.7-22), eotide (1.80.0), files (1.72.0), 0.1-2 (1.80.0), notations (1.64.1), resources (3.10.1), and SSH (1.2.1). It also shows a note about Base64 Encoder and Decoder.
- System Tray:** Shows battery level (22°C), network status, and date/time (12:48, 8/10/2024).

# Source vs Visual Editor



# What did we see until now ?

- R Studio.
- R markdown + output.
- Basic graphical functions in R and how to control the figure (title, colors etc).
- Do the analysis in the file:
  - [\*\*R\\_course\\_UHasselt\\_2021\(pdf\)\\_V2.Rmd\*\*](#)
- Explore use of Source and Visual Editor



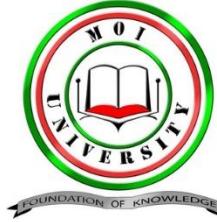
# Case study: The airquality data

R program: [Airquality.R](#)

Rmd program: [Case studies\\_Prog1\\_V1.Rmd](#)

# What do we do in this session ?

- We conduct a simple analysis for the variable **Wind speed** in the airquality data:
  - Summary statistics.
  - Graphical display: histogram.
  - Confidence interval.
  - Test of hypothesis.
- Focus:
  - How to conduct the analysis in R ?
  - How to produce an output ?
  - How to combine text and software output in the same document ?



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

# The airquality data

R Program: Airquality.R

Part 1  
The dataset

# The airquality data in R

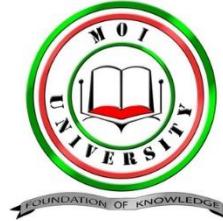
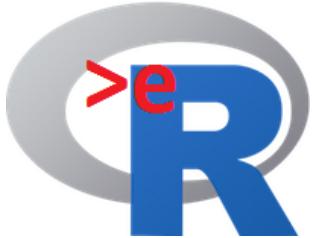
```
> dim(airquality) → The R object for the data: 153
[1] 153 6 observations and 6 variables.
> names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"

Variables names:
Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island.
Wind: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia
Airport.
```

# The airquality data in R

```
> head(airquality)
```

|   | zone | Solar.R | Wind | Temp | Month | Day |
|---|------|---------|------|------|-------|-----|
| 1 | 41   | 190     | 7.4  | 67   | 5     | 1   |
| 2 | 36   | 118     | 8.0  | 72   | 5     | 2   |
| 3 | 12   | 149     | 12.6 | 74   | 5     | 3   |
| 4 | 18   | 313     | 11.5 | 62   | 5     | 4   |
| 5 | NA   | NA      | 14.3 | 56   | 5     | 5   |
| 6 | 28   | NA      | 14.9 | 66   | 5     | 6   |



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

# The airquality data

Part 2

Analysis plan + analysis in R

# Analysis plan

- Response: wind speed.
- Analysis:
  - Calculate mean and SD.
  - Construct 95% C.I. for the population mean.
  - Test the hypothesis for the population mean.

## Case 1: using a $N(0,1)$

If

$$X \sim N(\mu, \sigma^2)$$

then:  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

And  $Z_{\bar{X}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

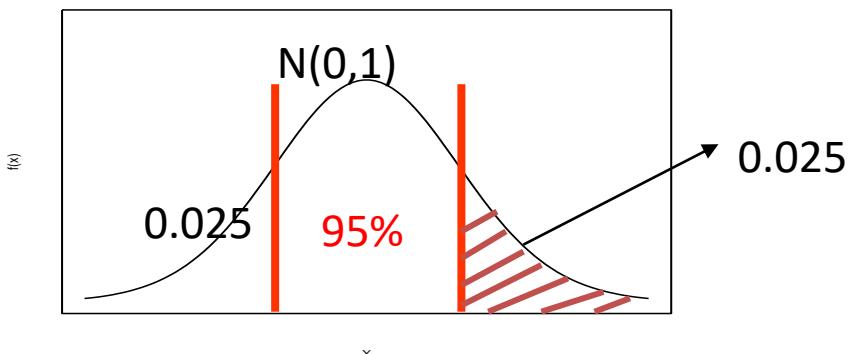
1.  $X$  has a normal distribution with unknown  $\mu$  and known  $\sigma^2$ .

# CI for case 1

Step 1: example, choose  $1-\alpha = 0.95$

Step 2: Case 1, thus:  $Z_{\bar{X}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

Step 3: critical point:



# CI for case 1

Step 4: calculate the point estimator :  $\bar{x}$

Step 5: calculate the CI

For this, we know:

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1.96\right) = 0.95$$

or, after the conversion of the formula:

$$P\left(\bar{X} - 1.96 \times \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + 1.96 \times \sqrt{\frac{\sigma^2}{n}}\right) = 0.95$$

## CI for case 1

$$P\left(\bar{X} - 1.96 \times \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + 1.96 \times \sqrt{\frac{\sigma^2}{n}}\right) = 0.95$$
$$P(L \leq \mu \leq R) = 1 - \alpha$$

So, a  $(1-\alpha)$  CI for  $\mu$  is :

$$\left[ \bar{x} - z \sqrt{\frac{\sigma^2}{n}}, \bar{x} + z \sqrt{\frac{\sigma^2}{n}} \right]$$

# Two sided hypothesis testing

The mean under  $H_0$  is not equal to the mean under  $H_1$ :

$$H_0 : \mu = \mu_{H_0}$$

$$H_1 : \mu \neq \mu_{H_0}$$

null hypothesis

alternative  
hypothesis

two sided test problem

# Test statistic

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

# R functions for the analysis

- `mean()`
- `var()`
- `z.test()`

R program for the analysis: `Airquality.R`

# R program

```
wind<-airquality$Wind
M.wind<-mean(wind)
SD.wind<-sqrt(var(wind))
```

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

# R program

```
wind=na.omit(airquality$Wind)
```

```
library(TeachingDemos)
```

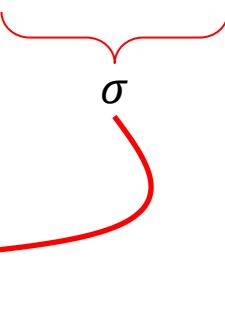
```
z.test(wind, sd=SD.wind)
```

$\sigma$

$$\left[ \bar{x} - z \sqrt{\frac{\sigma^2}{n}}, \bar{x} + z \sqrt{\frac{\sigma^2}{n}} \right]$$

# R program

**z.test**(wind, SD.wind, **mu=9**)

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$


$$H_0: \mu = 9$$

$$H_1: \mu \neq 9$$

# Output

```
> wind<-airquality$Wind
> M.wind<-mean(wind)
> SD.wind<-sqrt(var(wind))
> M.wind
[1] 9.957516
> SD.wind
[1] 3.523001
```

# Output (C.I)

```
> wind=na.omit(airquality$Wind)
> library(TeachingDemos)
> z.test(wind, sd=SD.wind)
```

One Sample z-test

```
data: wind
z = 34.961, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample
mean = 0.28482, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
9.399284 10.515749
sample estimates:
mean of wind
9.957516
```

$$\left[ \bar{x} - z \sqrt{\frac{\sigma^2}{n}}, \bar{x} + z \sqrt{\frac{\sigma^2}{n}} \right] = [9.399, 10.515]$$

# Output (test of hypothesis)

```
> z.test(wind, SD.wind, mu=9)
One Sample z-test

data: wind
z = 3.3619, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample
mean = 0.28482, p-value = 0.0007742
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
 9.399284 10.515749
sample estimates:
mean of wind
 9.957516
```

$H_0: \mu = 9$   
 $H_1: \mu \neq 9$

**A two sided test**

## Case 3: using a $t_{(n-1)}$ distribution

An estimate for the standard error of the sample mean.

$$SE = \frac{s}{\sqrt{n}}$$

The standard deviation.  
The sample size.

If the sample comes from a population whose **distribution is normal** and the **variance is unknown**, and the **sample size is small** , then :

$$T_{\bar{X}} = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$$

A Student t-distribution with  $(n-1)$  degrees of freedom is denoted by  $t(n-1)$ .

# R functions for the analysis

- In our case, n=153 (n>30 !!!) but, for illustration, we will use t distribution for the test.
- `mean()` .
- `var()` .
- `t.test()` .

# R program

$$t_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$$

$t.\text{test}(\text{wind}, \text{mu}=9)$

$H_0: \mu = 9$   
 $H_1: \mu \neq 9$

- Two sided t test with unknown variance.
- $n > 30$ .

# Output (C.I)

```
> t.test(wind, mu=0)
```

One Sample t-test

```
data: wind
t = 34.961, df = 152, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
```

**95 percent confidence interval:**

**9.394804 10.520229**

sample estimates:

mean of x  
9.957516

$$\left[ \bar{x} - t \sqrt{\frac{s^2}{n}}, \bar{x} + t \sqrt{\frac{s^2}{n}} \right] = [9.394, 10.520]$$

Not relevant

# Output (test of hypothesis)

```
t.test(wind, mu=9, var.equal=TRUE)
```

One Sample t-test

```
data: wind
t = 3.3619, df = 152, p-value = 0.0009794
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
 9.394804 10.520229
sample estimates:
mean of x
9.957516
```

$$\begin{aligned} H_0: \mu &= 9 \\ H_1: \mu &\neq 9 \end{aligned}$$

A two sided test



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

# The airquality data

## Part 3 The Rmd program

Rmd program for the analysis: [Case Study\\_Prog1\\_V1.Rmd](#)

# The output

- We run the R markdown file.
- Produce output format:
  - Word document.

# The Rmd program

The screenshot shows the RStudio interface with an Rmd file open. The left pane displays the Rmd code, and the right pane shows the Global Environment and Data panes.

**Document setup.** (Annotation: A red curly brace groups lines 1-9)

```
1 ---
2 title: 'Case study 1: analysis of the daily average wind speed in New York in 1973'
3 output:
4 word_document: default
5 pdf_document: default
6 html_document: default
7 subtitle: Foundations for inference using R
8 layout: page
9 ---
```

**Many R packages, not all needed.** (Annotation: A red curly brace groups lines 19-34)

```
10 |
11 |
12 ````{r setup, include=FALSE}
13 options(htmltools.dir.version = FALSE)
14 knitr::opts_chunk$set(echo = TRUE,
15 message = FALSE,
16 warning = FALSE,
17 eval = TRUE,
18 tidy = FALSE)
19 library(knitr)
20 library(tidyverse)
21 library(deSolve)
22 library(minpack.lm)
23 library(ggpubr)
24 library(readxl)
25 library(gamlss)
26 library(data.table)
27 library(grid)
28 library(png)
29 library(nlme)
30 library(gridExtra)
31 library(mvtnorm)
32 library(e1071)
33 library(lattice)
34 library(ggplot2)
```

**RStudio Status Bar:**

- Windows icon
- Google Chrome icon
- Case studies
- Downloads
- Inbox (3) - t...
- RStudio
- eR\_biosat\_i...
- Meeting 21...
- Cloud icon: 18°C
- ENG
- 10:55
- 21/05/2024
- 2 notifications

# The Rmd program

The screenshot shows the RStudio interface with two tabs open: "Case studies\_Prog1\_V1.Rmd\*" and "Visualization.Rmd". The left pane displays R code for loading the airquality dataset and performing exploratory analysis. The right pane shows the "Global Environment" tab with the "airquality" dataset listed. Red annotations with arrows point from specific lines of code to descriptive text labels:

- Line 48: "# Case study 1: The wind speed in the airquality dataset" → **Section**
- Line 50: "## Exploratory analysis of the daily average of the wind speed" → **Subsection**
- Line 54: "```{r, echo=TRUE, message=FALSE, warning=FALSE}" → **Dimension of the data,**
- Line 55: "dim(airquality)" → **variables names and first**
- Line 60: "head(airquality)" → **6 lines of the data**

Large red text on the right side of the image reads: "R code".

128

# Choose the output

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Print, and Go to file/function.
- Project Bar:** Project: (None).
- Source Editor:** Displays an Rmd file with code and comments. A red box highlights the Knit dropdown menu, and a red arrow points to the "Knit to Word" option.
- Environment Tab:** Shows the Global Environment and Data section with various R objects listed.
- Plots Tab:** Shows the current plots.
- Packages Tab:** Shows the current packages.
- Help Tab:** Shows help documentation.
- Viewer Tab:** Shows the current viewer settings.
- Presentation Tab:** Shows presentation settings.

# The Word document output

The screenshot shows a Microsoft Word document window titled "Case-studies\_Prog1\_V1 [Compatibility Mode] - Word". The ribbon menu is visible at the top, with "Home" selected. The main content area contains the following text:

**Case study 1: analysis of the daily average wind speed in New York in 1973**

**Foundations for inference using R**

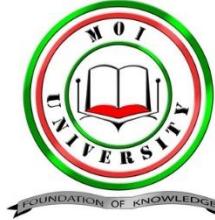
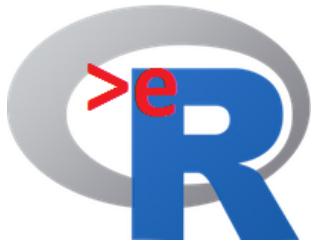
**Case study 1: The wind speed in the airquality dataset**

**Exploratory analysis of the daily average of the wind speed**

The airquality dataset is a R object gives information about 153 daily air quality measurements ( $n = 153$ ) in New York, May to September 1973.

```
dim(airquality)
[1] 153 6
names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
head(airquality)
Ozone Solar.R Wind Temp Month Day
```

At the bottom left, there is a green circular icon with a white letter "G". The status bar at the bottom shows "Page 3 486 words English (United States)". On the right side of the status bar, there are zoom controls and a page number "130".



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

# The airquality data

## Part 4

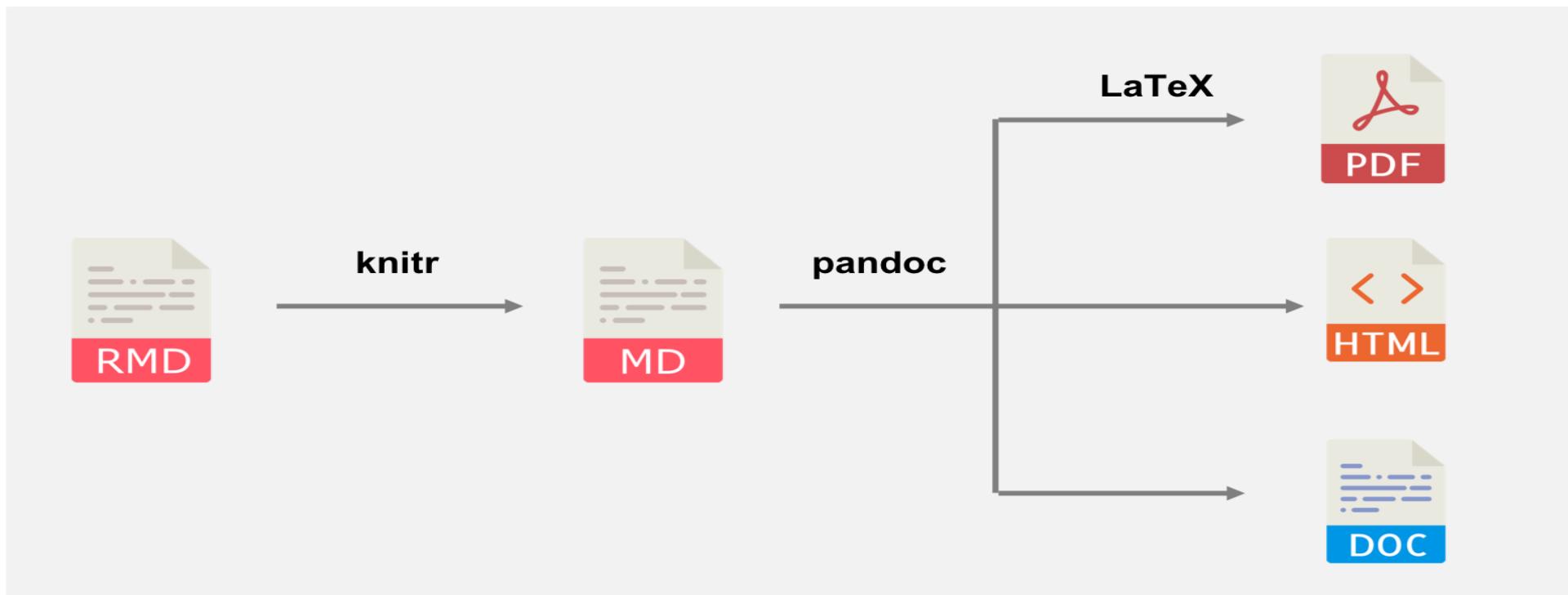
### How to produce the Word output ?

# Reproducible Research

- Aim: create an output in a Word document.
- Can be used to communicate the analysis' results with other people in the organization.
- Not all potential readers are interested on “how to do the analysis”.
- We DO NOT aim to develop a report for the analysis but to provide a document from which the results can be seen and discuss by different people in the organization.

# The Rmd file

- Analyses → high quality report.
- Rmarkdown – Different dynamic and static formats (html, pdf, word, books, dashboard, e.t.c).



# The Word output

The screenshot shows a Microsoft Word document titled "Case-studies\_Prog1\_V1 [Compatibility Mode] - Word". The document content is as follows:

- Section Title:** Case study 1: analysis of the daily average wind speed in New York in 1973
- Section Subtitle:** Foundations for inference using R
- Section Description:** Case study 1: The wind speed in the airquality dataset
- Section Subdescription:** Exploratory analysis of the daily average of the wind speed
- Text:** The airquality dataset is a R object gives information about 153 daily air quality measurements ( $n = 153$ ) in New York, May to September 1973.
- R Code:**

```
dim(airquality)
[1] 153 6
names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
head(airquality)
Ozone Solar.R Wind Temp Month Day
```

Annotations with red curly braces highlight the title and the analysis output section.



## Part 4.1: How to set up the Word file?

# The Rmd file

- We use Rmd file to
  - Conduct the analysis.
  - Set up the document.
- We use a Word file in order to
  - Present & communicate the result.

# Set up the document

The screenshot shows the RStudio interface with two R Markdown files open: 'Case studies\_Prog1\_V1.Rmd' and 'Visualization.Rmd'. The 'Case studies\_Prog1\_V1.Rmd' file contains R code for setting up the document, including the line 'output: word\_document: default'. A red arrow points from this line to a red bracket on the right side of the screen. Below this bracket, the text 'Set up the Word document: Word\_document: default' is displayed in blue. To the right of the bracket, a Microsoft Word document titled 'Case study 1: analysis of the daily average wind speed in New York in 1973' is shown. The document includes a subtitle 'Foundations for inference using R' and some text about the airquality dataset. A red arrow points from the right side of the RStudio interface towards the Microsoft Word document.

```
1 ---
2 title: 'Case study 1: analysis of the daily average wind speed in New York in 1973'
3 output:
4 word_document: default ←
5 pdf_document: default
6 html_document: default
7 subtitle: Foundations for inference using R
8 layout: page

9
10
11
12 ````{r setup, include=FALSE}
13 options(htmltools.dir.version = FALSE)
14 knitr::opts_chunk$set(echo = TRUE,
15 message = FALSE,
16 warning = FALSE,
17 eval = TRUE,
18 tidy = FALSE)
19 library(knitr)
20 library(tidyverse)
21 library(deSolve)
22 library(minpack.lm)
23 library(ggpubr)
24 library(readxl)
25 library(gamlss)
26 library(data.table)
27 library(grid)
28 library(png)
29 library(nlme)
30 library(gridExtra)
31 library(mvtnorm)
32 library(e1071)
33 library(lattice)
34 library(ggplot2)
35
10:1 (Top Level) ⇲
```

Console

Project: (None)

Environment History Connections Tutorial

Data

- hist\_bmi List of 11
- NHANES 10000 obs. of 76 variables
- NHANES summarv 1 obs. of 2 variables
- List of 6
- 1 obs. of 2 variables
- 1 obs. of 2 variables
- List of 11

values

File Plots Packages Help Viewer Presentation

Case study 1: analysis of the daily average wind speed in New York in 1973

Foundations for inference using R

Case study 1: The wind speed in the airquality dataset

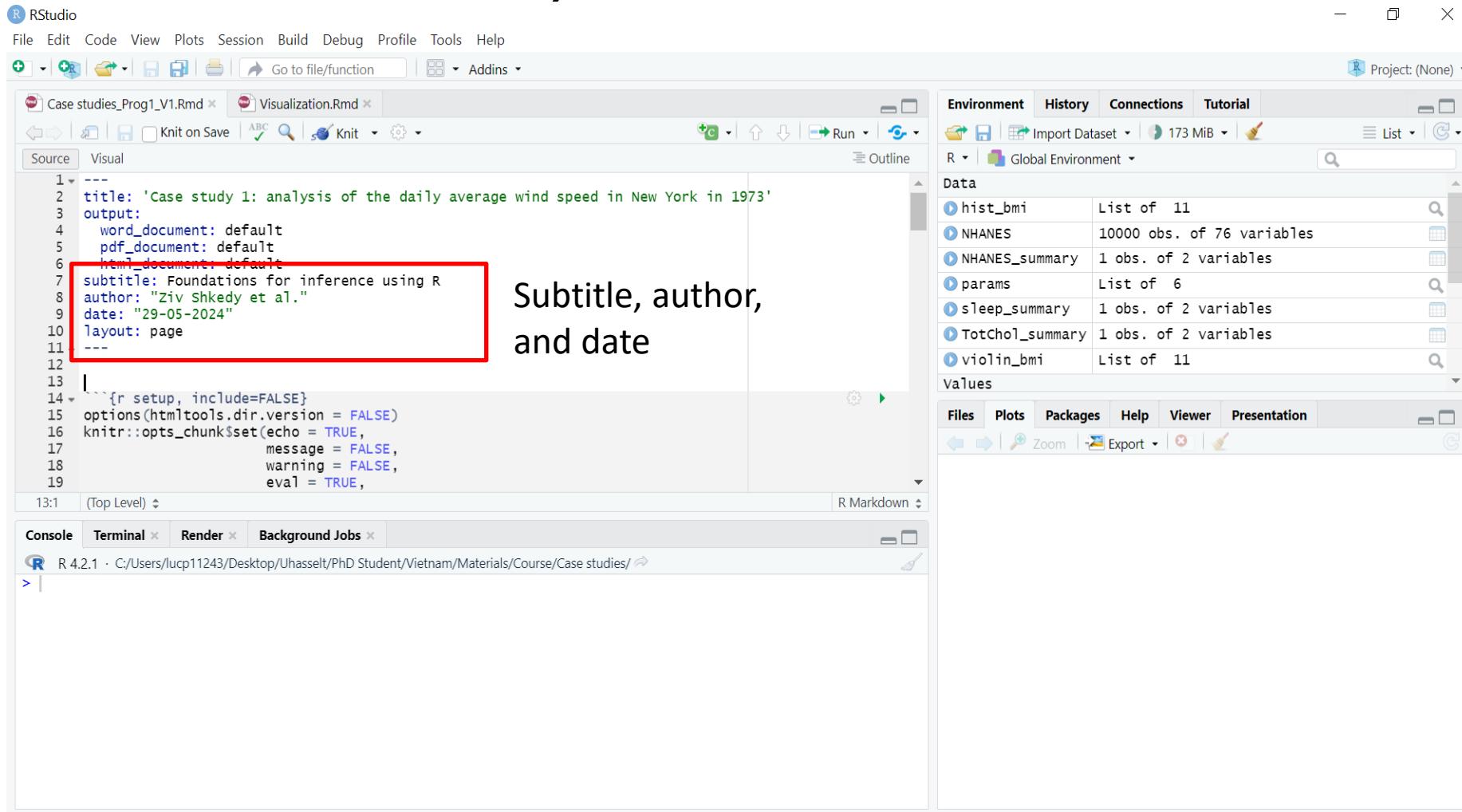
Exploratory analysis of the daily average of the wind speed

The airquality dataset is a R object gives information about 153 daily air quality measurements ( $n = 153$ ) in New York, May to September 1973.

```
dim(airquality)
[1] 153 6
names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
head(airquality)
Ozone Solar.R Wind Temp Month Day
```

18°C 10:55 21/05/2024 2

# Titles, authors and dates



## Subtitle, author, and date

# Titles, authors and dates

The screenshot shows a Microsoft Word document window titled "Case-studies\_Prog1\_V1 [Compatibility Mode] - Word". The ribbon menu is visible at the top, with the "Home" tab selected. The main content area contains the following text:

**Case study 1: analysis of the daily average wind speed in New York in 1973**

**Foundations for inference using R**

Ziv Shkedy et al.  
29-05-2024

**Case study 1: The wind speed in the airquality dataset**

**Exploratory analysis of the daily average of the wind speed**

The airquality dataset is a R object gives information about 153 daily air quality measurements ( $n = 153$ ) in New York, May to September 1973.

```
dim(airquality)
[1] 153 6
names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
head(airquality)
Ozone Solar.R Wind Temp Month Day
```

At the bottom left, there is a green circular icon with a white letter "G". The status bar at the bottom shows "Page 3 491 words English (United States)".

A red box highlights the main title text. A blue arrow points from the subtitle text to the word "Sub title" on the right.



## Part 4.2: The Word document and the Rmd program in details.

# Section, subsection, subsubsection

The screenshot shows the RStudio interface with the following elements:

- File menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source tab:** Displays R code in a script named "Case studies\_Prog1\_V1.Rmd".

```
50 # Case study 1: The wind speed in the airquality dataset
51 ## Exploratory analysis of the daily average of the wind speed
52 The <tt>airquality</tt> dataset is a R object gives information about 153 daily air quality
53 measurements ($n=153$) in New York, May to September 1973.
54
55 ##{r, echo=TRUE, message=FALSE, warning=FALSE}
56 dim(airquality)
57 names(airquality)
58
59 ##{r, echo=TRUE, message=FALSE, warning=FALSE}
60 head(airquality)
61
62 The variable of primary interest, <tt>Wind</tt>, is the average wind speed in miles per hour at
63 0700 and 1000 hours at LaGuardia Airport. We use the R package <tt>ggplot2</tt> to explore the
64
```
- Section header:** A red arrow points to the word "The" in line 52, which is highlighted in blue and preceded by two hash symbols, indicating it is a section header.
- Subsection header:** A red arrow points to the word "Wind" in line 62, which is highlighted in blue and preceded by three hash symbols, indicating it is a subsection header.
- Free text:** A red box highlights the explanatory text in line 63, and a red arrow points from this box to the text "free text".
- Environment tab:** Shows the global environment with various objects listed.
- Console tab:** Shows the R session path: R 4.2.1 - C:/Users/lucp11243/Desktop/Lhassell/PhD Student/Vietnam/Materials/Course/Case studies/
- Taskbar:** Shows various application icons and the date/time: 19°C, 12:31, 21/05/2024.

- In addition to the code, we can add free text in the Rmd file.

# Section, subsection, subsubsection

Case-studies\_Prog1\_V1 [Compatibility Mode] - Word NGUYEN Thi Huyen

File Home Insert Draw Design Layout References Mailings Review View MathType Add-ins Help Foxit PDF Tell me what you want to do Share

Cut Copy Format Painter Paste Clipboard

Font Paragraph Styles Editing

Case study 1: analysis of the daily average wind speed in New York in 1973

Foundations for inference using R

Ziv Shkedy et al.

29-05-2024

Case study 1: The wind speed in the airquality dataset

Exploratory analysis of the daily average of the wind speed

The airquality dataset is a R object gives information about 153 daily air quality measurements ( $n = 153$ ) in New York, May to September 1973.

free text

```
dim(airquality)
[1] 153 6
names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
head(airquality)
Ozone Solar.R Wind Temp Month Day
```

section

subsection

The code is shown as a part of the output

Page 3 491 words English (United States)

# Code in the Rmd file

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New Project, Open, Save, Knit, Run, and Addins.
- Source Editor:** Displays an Rmd file with the following code:

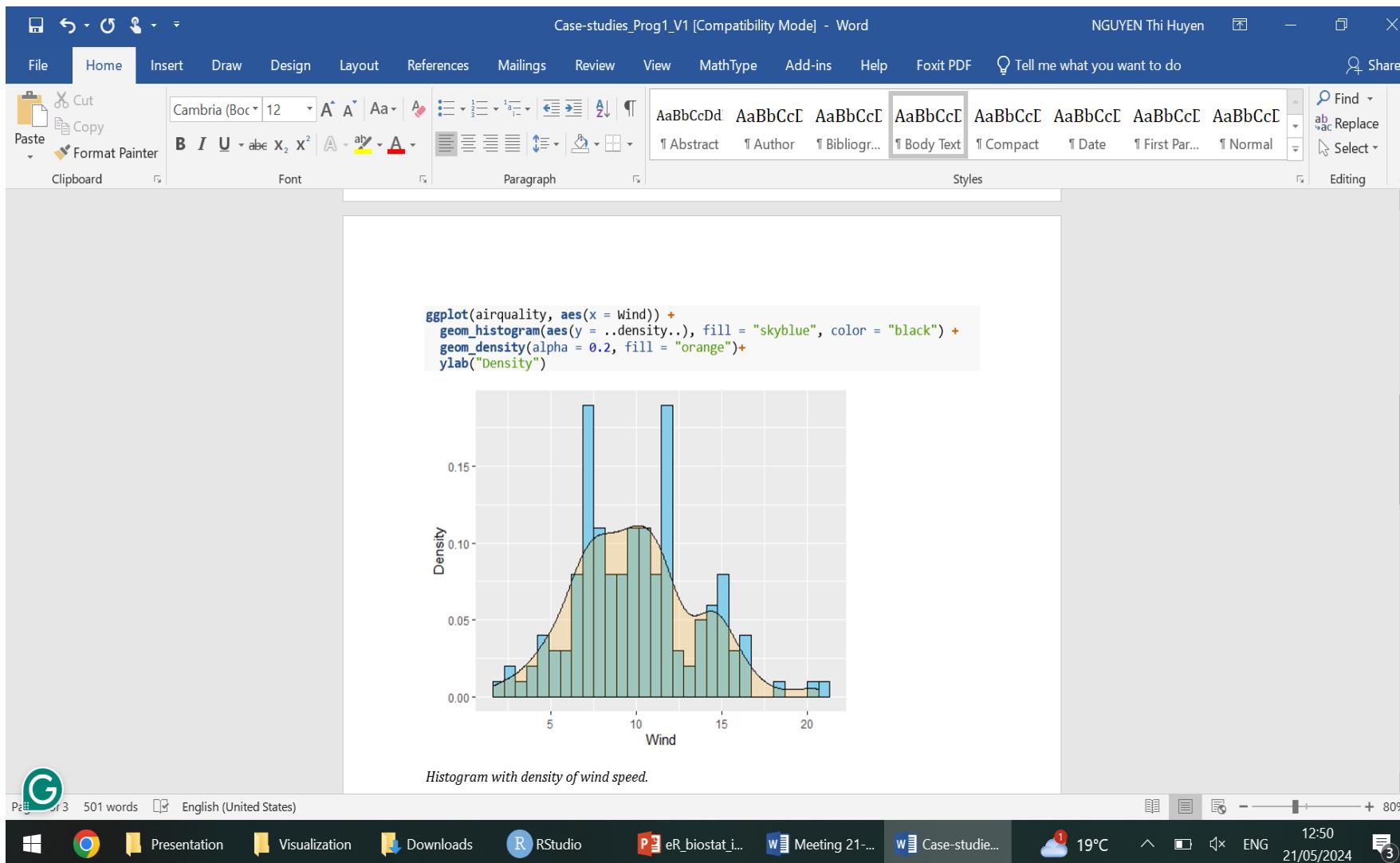
```
58 names(airquality)
59
60
61 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
62 head(airquality)
63
64
65 The variable of primary interest, <tt>Wind</tt>, is the average wind speed in miles per hour at
66 0700 and 1000 hours at LaGuardia Airport. We use the R package <tt>ggplot2</tt> to explore the
67 data. Figure 1 shows histogram of wind speed.
68
69 ```{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Histogram with density of wind speed"}
70 ggplot(airquality, aes(x = Wind)) +
71 geom_histogram(aes(y = ..density..), fill = "skyblue", color = "black") +
72 geom_density(alpha = 0.2, fill = "orange")+
73 ylab("Density")
74
75 The sample mean (\bar{x}) and standard deviation (s) of the wind speed are 9.957516 and
76 # Exploratory analysis of the daily average of the wind speed
```

- Environment View:** Shows the global environment with objects like hist\_bmi, NHANES, params, etc.
- Console View:** Shows the R version (R 4.2.1) and the current working directory (C:/Users/lucp11243/Desktop/Uhasselt/PhD Student/Vietnam/Materials/Course/Case studies).
- Bottom Taskbar:** Includes icons for Windows, Google Chrome, Present..., Visualiz..., Download..., RStudio, eR bios..., Meetin..., Case-st..., Rmd pr..., Section..., 19°C, ENG, 12:46, 21/05/2024, and a notification icon.

A red box highlights the ggplot2 code for creating a histogram, and a red arrow points from this box to the text "The code for histogram".

The code for histogram

# The output in the Word file



# Code in the Rmd file

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Knit, Run, and Addins.
- Project:** Project: (None)
- Source Editor:** Displays an Rmd file with the following code:

```
83
84 ## A 95% confidence interval for the population mean
85
86 We construct a 95% confidence interval for the population mean using the R function
<tt>z.test</tt>. We assume that σ = 1 and the function uses the $N(0,1)$ to select the
critical value $Z_{\alpha/2}$. For a 95% confidence interval, $Z_{\alpha/2}=1.96$.
87 A 95% confidence interval for the wind speed is [9.399284; 10.515749]. This means that we are
95% confident that the true average wind speed lies within this range.
88
89 ````{r, echo=TRUE, message=FALSE, warning=FALSE}
90 library(TeachingDemos)
91 wind=na.omit(airquality$Wind)
92 z.test(wind, sd=SD.wind)
93
94
95
96
97
98
98:1 # A 95% confidence interval for the population mean
```
- Text Overlay:** "The code for confidence interval." is displayed in the center of the Source Editor area.
- Environment Tab:** Shows the global environment with objects like hist\_bmi, NHANES, and violin\_bmi.
- Console Tab:** Shows the R version (R 4.2.1) and the current working directory (C:/Users/lucp11243/Desktop/Uhasselt/PhD Student/Vietnam/Materials/Course/Case studies).
- Taskbar:** Shows various application icons including Windows, Google Chrome, Microsoft Word, Microsoft Excel, RStudio, eR\_biosta..., Meeting..., Case-stu..., Testing..., and system status icons for battery, temperature (19°C), and date/time (21/05/2024, 12:57).

# The output in the Word file

Case-studies\_Prog1\_V1 [Compatibility Mode] - Word

NGUYEN Thi Huyen

AaBbCcDd AaBbCcE AaBbCcE AaBbCcE AaBbCcE AaBbCcE AaBbCcE AaBbCcE

¶ Abstract ¶ Author ¶ Bibliogr... ¶ Body Text ¶ Compact ¶ Date ¶ First Par... ¶ Normal

Find Replace Select Editing

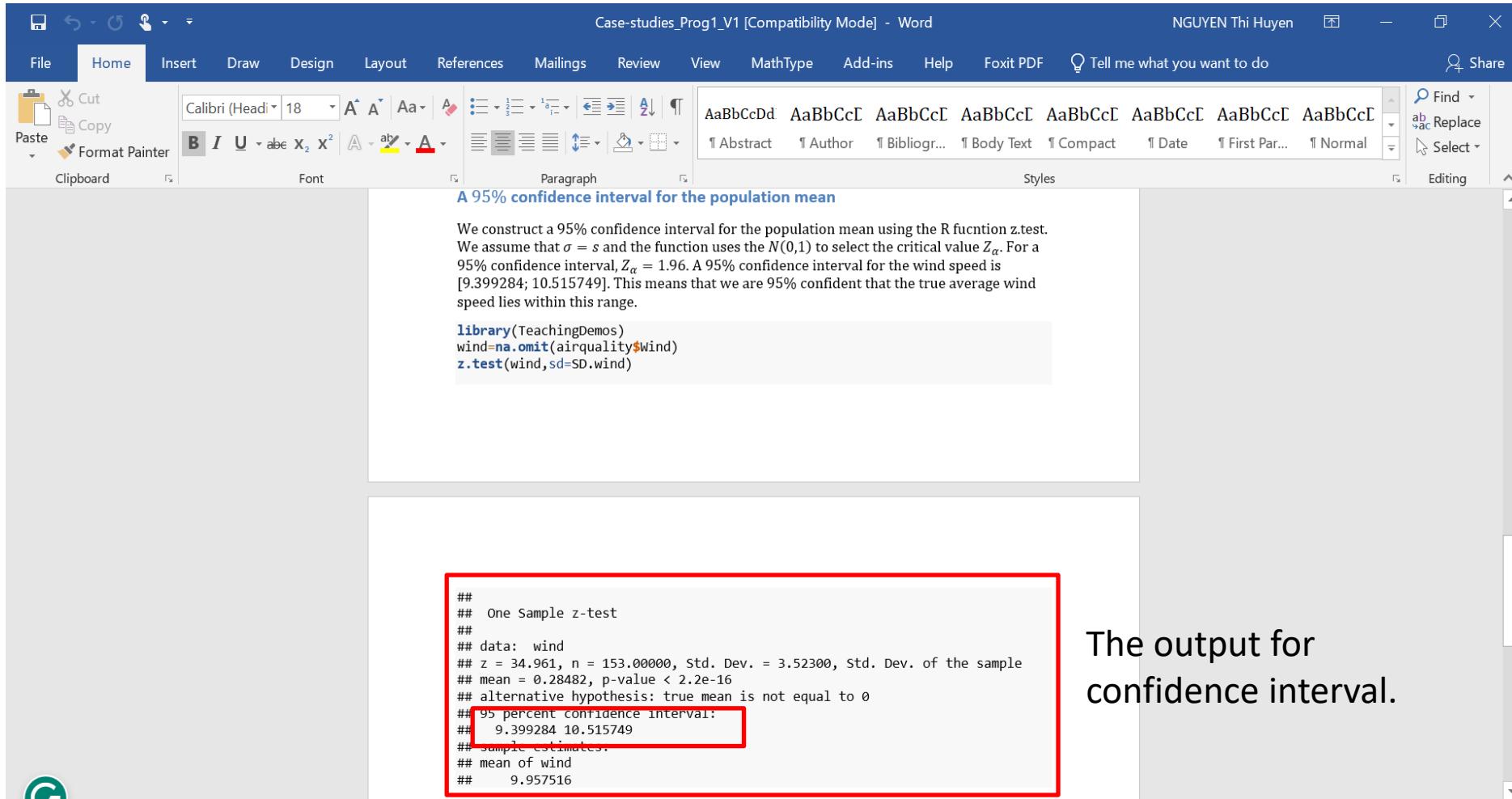
**A 95% confidence interval for the population mean**

We construct a 95% confidence interval for the population mean using the R function `z.test`. We assume that  $\sigma = s$  and the function uses the  $N(0,1)$  to select the critical value  $Z_\alpha$ . For a 95% confidence interval,  $Z_\alpha = 1.96$ . A 95% confidence interval for the wind speed is  $[9.399284; 10.515749]$ . This means that we are 95% confident that the true average wind speed lies within this range.

```
library(TeachingDemos)
wind=na.omit(airquality$Wind)
z.test(wind, sd=SD.wind)
```

## One Sample z-test
##
## data: wind
## z = 34.961, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample
## mean = 0.28482, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 9.399284 10.515749
## sample estimates:
## mean of wind
## 9.957516

The output for confidence interval.



Page 3 501 words English (United States)

Windows Presentation Visualization Downloads RStudio eR\_biotat\_i... Meeting 21... Case-studies... 19°C 12:56 ENG 21/05/2024 3

# Code in the Rmd file

The screenshot shows the RStudio interface with an Rmd file open. The code in the Source tab is as follows:

```
Test of hypothesis about the population mean
Testing the hypotheses whether the wind speed is equal to 9 versus a two-sided alternative hypothesis at the significant level of 0.05 can be formulated by:
$H_0: \mu = 9 $; $H_1: \mu \neq 9$.

$H_0: \mu = 9$ Vs. $H_1: \mu \neq 9$.

We use the z.test() function and specify mu=9.
```{r, echo=TRUE, message=FALSE, warning=FALSE}
z.test(wind, SD.wind, mu=9)
```
Since p-value = 0.0007742 which is much smaller than alpha = 0.05 there is sufficient evidence
Test of hypothesis about the population mean
```

A red box highlights the code block starting with `z.test(wind, SD.wind, mu=9)`. A callout box labeled "The code for testing hypotheses." points to this highlighted area.

The RStudio environment pane shows the following data objects:

| Object          | Type                       | Description |
|-----------------|----------------------------|-------------|
| hist_bmi        | List of 11                 |             |
| NHANES          | 10000 obs. of 76 variables |             |
| NHANES_summary  | 1 obs. of 2 variables      |             |
| params          | List of 6                  |             |
| sleep_summary   | 1 obs. of 2 variables      |             |
| TotChol_summary | 1 obs. of 2 variables      |             |
| violin_bmi      | List of 11                 |             |

The status bar at the bottom shows the system tray icons and the date/time: 12:57, 21/05/2024.

# The output in the Word file

Case-studies\_Prog1\_V1 [Compatibility Mode] - Word

NGUYEN Thi Huyen

File Home Insert Draw Design References Mailings Review View MathType Add-ins Help Foxit PDF Tell me what you want to do Share

Cut Copy Format Painter

Font Paragraph Styles

Test of hypothesis about the population mean

Testing the hypotheses whether the wind speed is equal to 9 versus a two-sided alternative hypothesis at the significant level of 0.05 can be formulated by:

$H_0: \mu = 9$  Vs.  $H_1: \mu \neq 9$ .

We use the z.test() function and specify mu=9\$.

```
z.test(wind, SD.wind, mu=9)
```

```

One Sample z-test

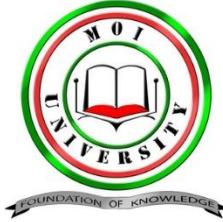
data: wind
z = 3.3619, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample
mean = 0.28482, p-value = 0.0007742
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
9.399284 10.515749
sample estimates:
mean of wind
9.957516
```

Since p-value = 0.0007742 which is much smaller than  $\alpha = 0.05$ , there is sufficient evidence to say that the mean of the wind speed is not equal to 9.

The output for testing hypotheses.

Page 3 of 3 501 words English (United States)

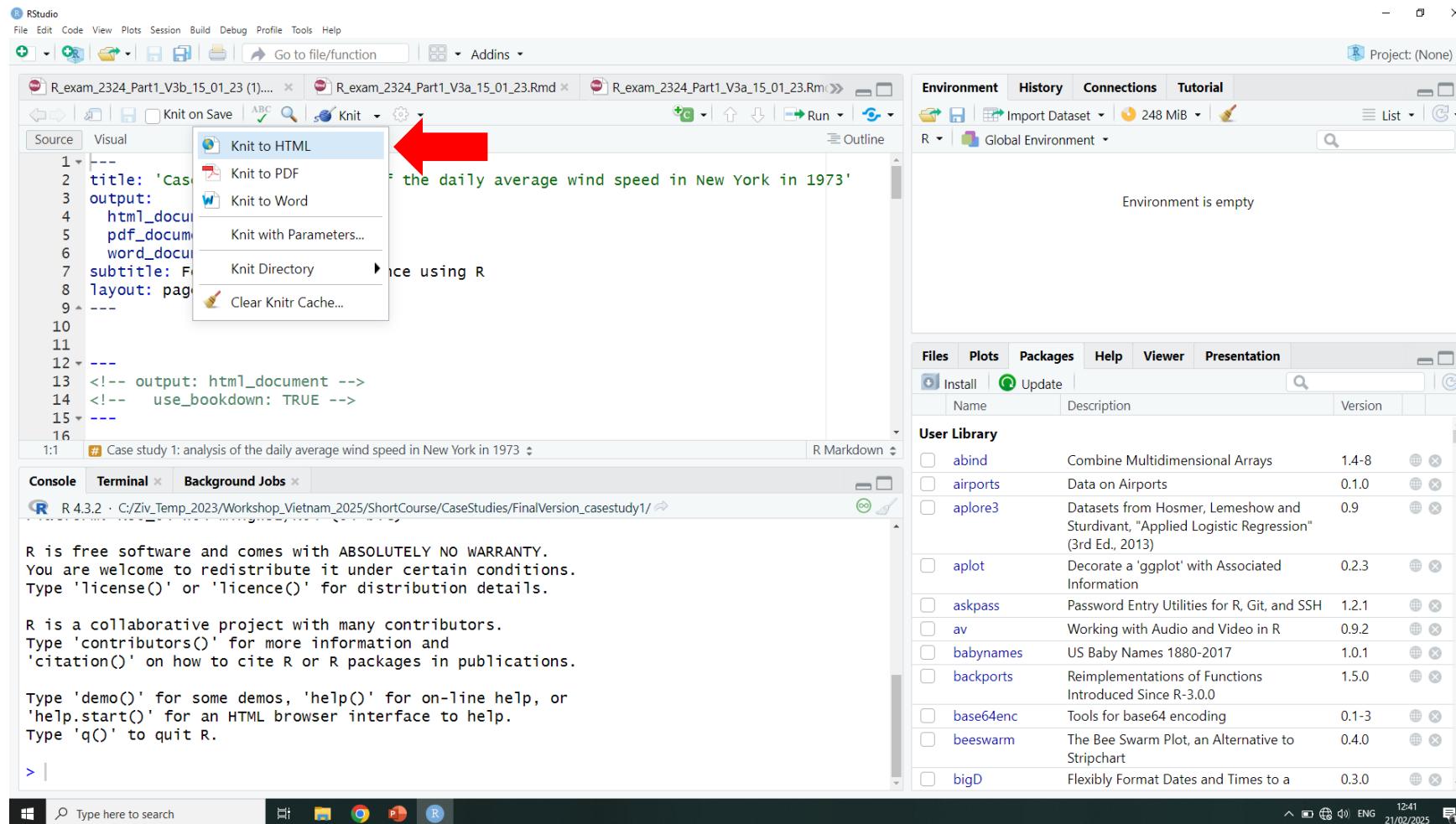
Presentation Visualization Downloads RStudio eR\_biotstat\_i... Meeting 21... Case-studies\_Prog1\_V1 19°C 12:56 21/05/2024 3



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

## Part 5: How to set up the HTML file?

# Setup the HTML document



The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins.
- Source Editor:** Shows R code for knitting a document. A red arrow points to the "Knit to HTML" option in the Knit dropdown menu.
- Console:** Displays the R startup message and a prompt (> |).
- Environment:** Shows an empty environment.
- Global Environment:** Shows an empty global environment.
- Packages:** Shows the User Library with various packages listed.
- System Status:** Shows system status including battery level, network, and date/time (12:41 21/02/2025).

# The output

Case study 1: analysis of the daily average wind speed in New York in 1973

Foundations for inference using R

## Case study 1: The wind speed in the airquality dataset

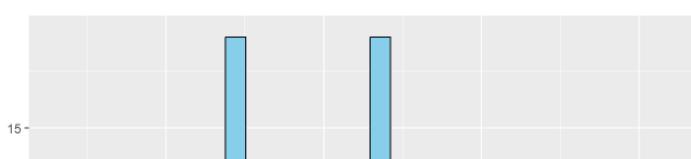
### Exploratory analysis of the daily average of the wind speed

The airquality dataset is a R object gives information about 153 daily air quality measurements ( $(n=153)$ ) in New York, May to September 1973.

```
dim(airquality)
[1] 153 6
head(airquality)
Ozone Solar.R Wind Temp Month Day
1 41 190 7.4 67 5 1
2 36 118 8.0 72 5 2
3 12 149 12.6 74 5 3
4 18 313 11.5 62 5 4
5 NA NA 14.3 56 5 5
6 28 NA 14.9 66 5 6
```

The variable of primary interest, Wind, is the average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport. We use the R package ggplot2 to explore the data. Figure 1 shows histogram of wind speed.

```
ggplot(airquality, aes(x = Wind)) +
 geom_histogram(fill = "skyblue", color = "black")+
 ylab("Frequency")
```



151

# The Rmd file

The screenshot shows the RStudio interface with several red annotations:

- section**: A red box highlights the word "section" in the Rmd file's code.
- subsection**: A red box highlights the word "subsection" in the Rmd file's code.
- free text**: A red box highlights the text "The <tt>airquality</tt> dataset is a R object gives information about 153 daily air quality measurements (\$n=153\$) in New York, May to September 1973."
- R code**: A red box highlights the code block starting at line 59.

The Rmd file content:

```
51 ~~~~
52
53
54 # Case study 1: The wind speed in the airquality dataset section
55
56 ## Exploratory analysis of the daily average of the wind speed subsection
57
58 The <tt>airquality</tt> dataset is a R object gives information about 153 daily air free text
quality measurements ($n=153$) in New York, May to September 1973.
59
60 ~~~{r, echo=TRUE, message=FALSE, warning=FALSE}
61 dim(airquality)
62 head(airquality)
63 ~~~
64
65 The variable of primary interest. <tt>wind</tt>. is the average wind speed in miles
1:1 # Case study 1: analysis of the daily average wind speed in New York in 1973 R Markdown
```

The R console output:

```
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.
```

The Environment tab in the top right shows "Environment is empty".

The Packages tab in the bottom right shows the User Library:

| Name      | Description                                                                                  | Version |
|-----------|----------------------------------------------------------------------------------------------|---------|
| abind     | Combine Multidimensional Arrays                                                              | 1.4-8   |
| airports  | Data on Airports                                                                             | 0.1.0   |
| aplore3   | Datasets from Hosmer, Lemeshow and Sturdivant, "Applied Logistic Regression" (3rd Ed., 2013) | 0.9     |
| aplot     | Decorate a 'ggplot' with Associated Information                                              | 0.2.3   |
| askpass   | Password Entry Utilities for R, Git, and SSH                                                 | 1.2.1   |
| av        | Working with Audio and Video in R                                                            | 0.9.2   |
| babynames | US Baby Names 1880-2017                                                                      | 1.0.1   |
| backports | Reimplementations of Functions Introduced Since R-3.0.0                                      | 1.5.0   |
| base64enc | Tools for base64 encoding                                                                    | 0.1-3   |
| beeswarm  | The Bee Swarm Plot, an Alternative to Stripchart                                             | 0.4.0   |
| bigD      | Flexibly Format Dates and Times to a                                                         | 0.3.0   |

# The output

- R code is a part of the text in the document.

Case study 1: analysis of the daily average wind speed in New York in 1973 **section**

Foundations for inference using R

Case study 1: The wind speed in the airquality dataset **subsection**

Exploratory analysis of the daily average of the wind speed

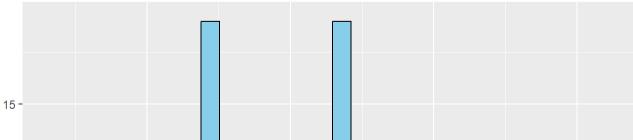
The airquality dataset is a R object gives information about 153 daily air quality measurements ( $(n=153)$ ) in New York, May to September 1973.

```
dim(airquality)
[1] 153 6

head(airquality)
Ozone Solar.R Wind Temp Month Day
1 41 190 7.4 67 5 1
2 36 118 8.0 72 5 2
3 12 149 12.6 74 5 3
4 18 313 11.5 62 5 4
5 NA NA 14.3 56 5 5
6 28 NA 14.9 66 5 6
```

The variable of primary interest, Wind, is the average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport. We use the R package ggplot2 to explore the data. Figure 1 shows histogram of wind speed.

```
ggplot(airquality, aes(x = Wind)) +
 geom_histogram(fill = "skyblue", color = "black") +
 ylab("Frequency")
```



A histogram showing the frequency distribution of wind speed. The x-axis is labeled 'Wind' and ranges from approximately 5 to 50. The y-axis is labeled 'Frequency' and ranges from 0 to 15. There are two prominent blue bars: one between 10 and 15, and another between 30 and 35. The background grid is light gray.

# Discussion

- R Studio + R markdown:
- Easy to use.
- Text + code.
- Output:
  - Standard: HTML, PDF, DOC.
  - Example: Word.doc.