UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

Development of E-learning materials using R & R markdown

Ziv Shkedy (Hasselt University)

International symposium on current trends in modeling and software development in data science and Statistics

Cape Town, 20/02/24-23/02/24

Visit us on Facebook ER-BioStat

GitHub https://github.com/eR-Biostat

twitter @erbiostat

Overview

- Starting point: Rudradev's session,
  - Linear regression using R markdown.
  - Output development.
- From R markdown to development of E-learning course.
- From an output on my laptop to a course online…

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

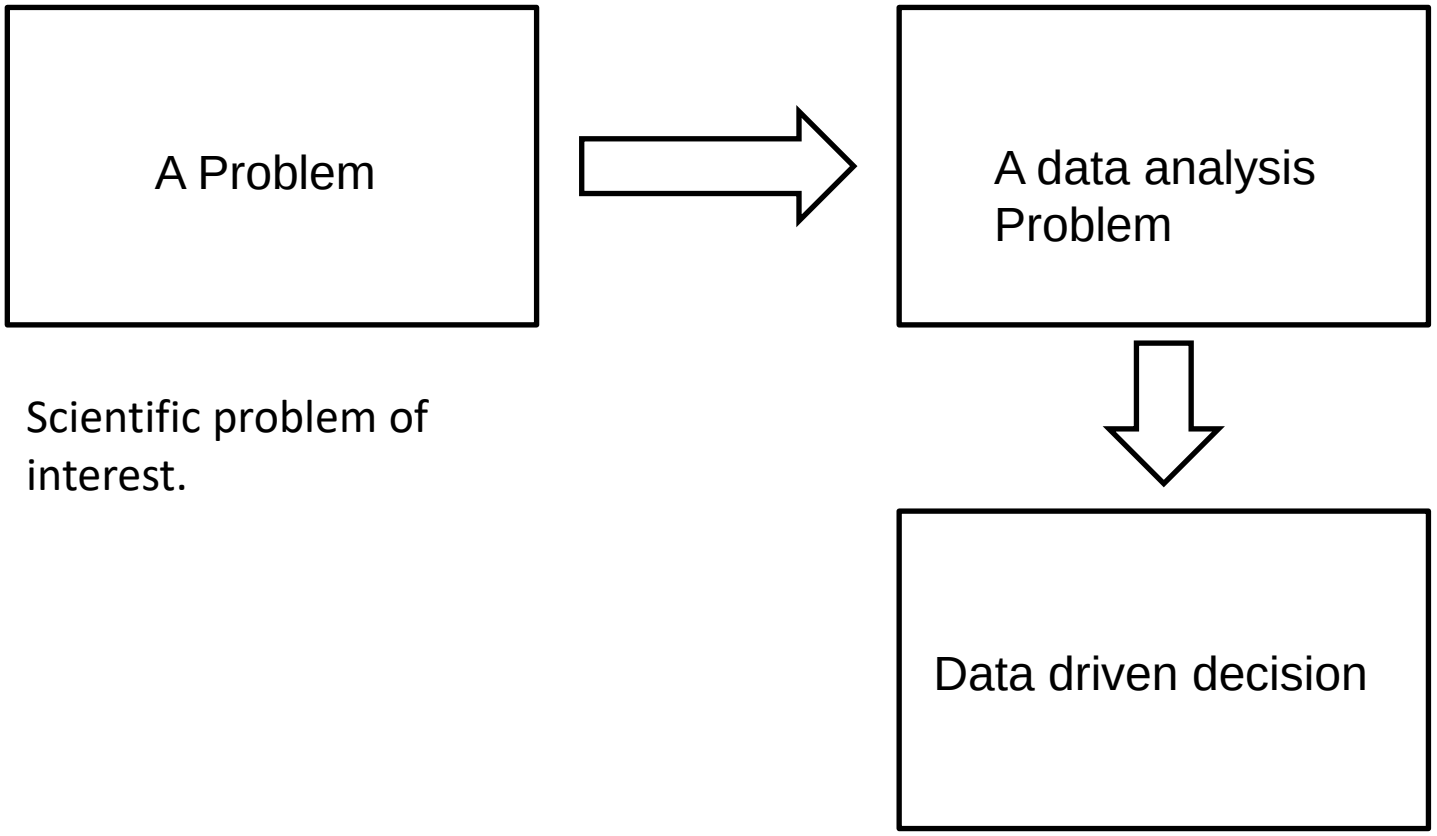
vliruos SHARING MINDS, CHANGING LIVES

Short summary of Rudradev's session

Linear regression using R markdown

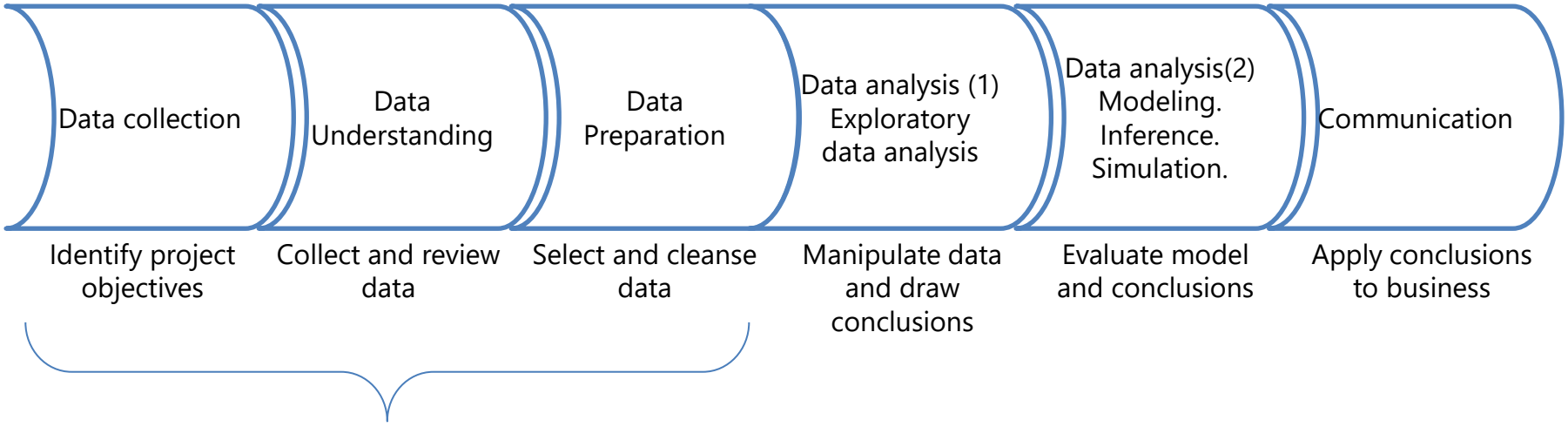
Steps in data analysis

- Data analysis approach in the course:



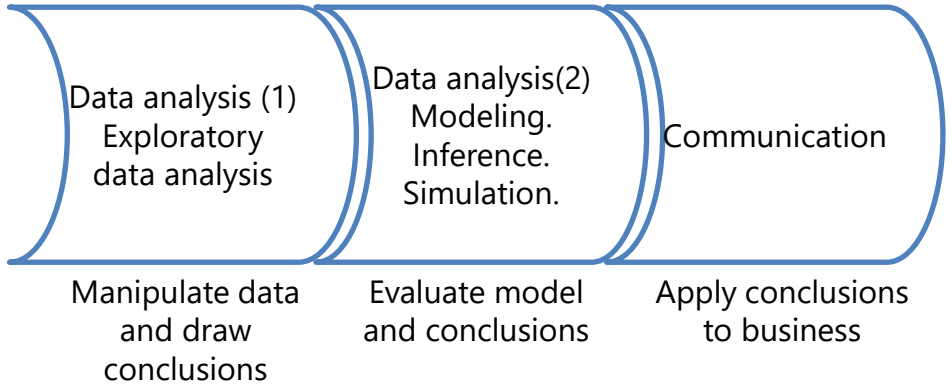
Steps in data analysis

- Steps related to data analysis:



Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.



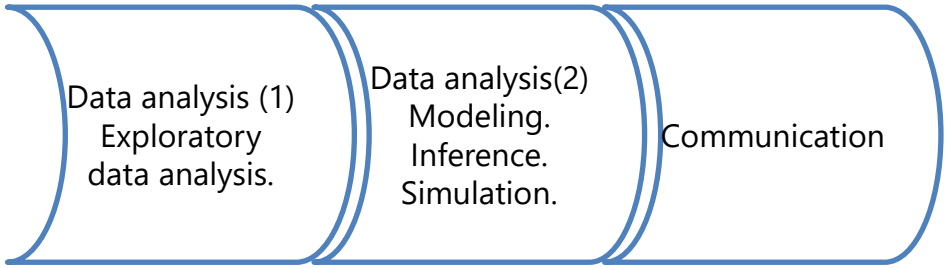
Scientific problem of interest:
how to model the association ?

Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.

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

Methodology: simple linear regression.



Scatterplot

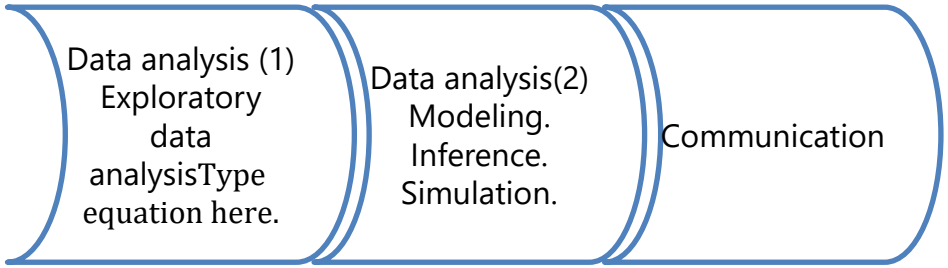
A linear regression model.

A report.

The solution

Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.



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

Scatterplot.

Linear regression model.

A report.

Methodology: simple linear regression.

We "translate" the methodology to software usage

ggplot2()

lm()

R markdown to produce a HTML file.

We develop software to produce the solution and to communicate the solution

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

The `mtcars` data

Part 1

Analysis using basic R programming

R Program: er_prog1_SA_2024.R

The `mtcars` data in R

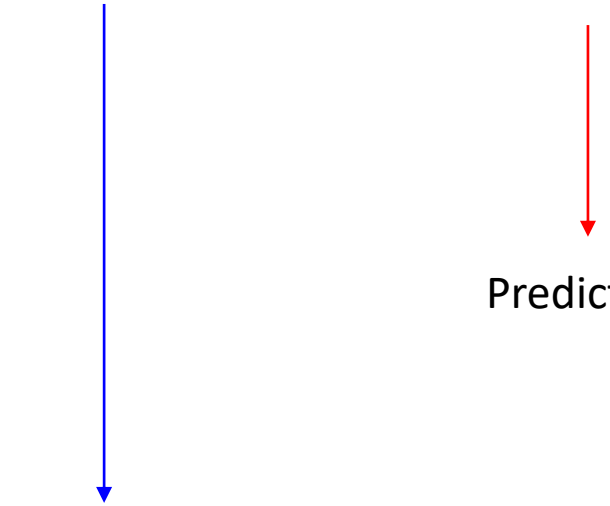
```
> dim(mtcars)
[1] 32 11
> names(mtcars)
 [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
[11] "carb"
```

The R object for the data: 32 observations and 11 variables.

Variables names.

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
>
```



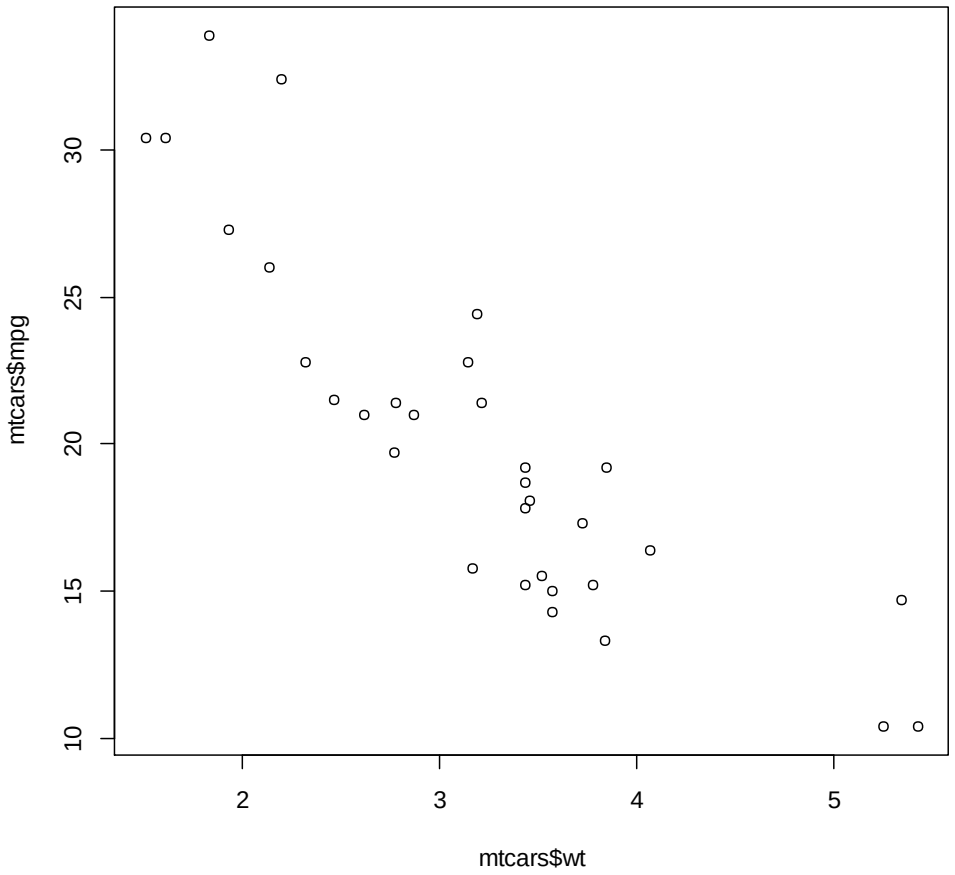
Response

Predictor

The `mtcars` data in R

```
> plot(mtcars$wt,mtcars$mpg)
> cor(mtcars$wt,mtcars$mpg)
[1] -0.8676594
```

- R functions to produce the plot and the correlation.
- Basic functions in R



Simple linear regression in R

```
lm(response~predictor)
```

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

Simple linear regression in R: the `mtcars` data

```
>fit.lm=lm(mtcars$mpg~mtcars$wt)
```

The R object that contains the results of the fitted model.

$$mpg_i = \beta_0 + \beta_1 wt_i + \varepsilon_i$$

```
> 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

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
```

- Output:
  - Parameter estimates etc.

R code for the analysis

```
dim(mtcars)
names(mtcars)
head(mtcars)
plot(mtcars$wt,mtcars$mpg)
cor(mtcars$wt,mtcars$mpg)
fit.lm=lm(mtcars$mpg~mtcars$wt)
summary(fit.lm)
```

- Produce the plot.
- Calculate the correlation.
- Fit the model:

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

- Print the estimated model.

R Program: er_prog1_SA_2024.R

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

The `mtcars` data

Part 2

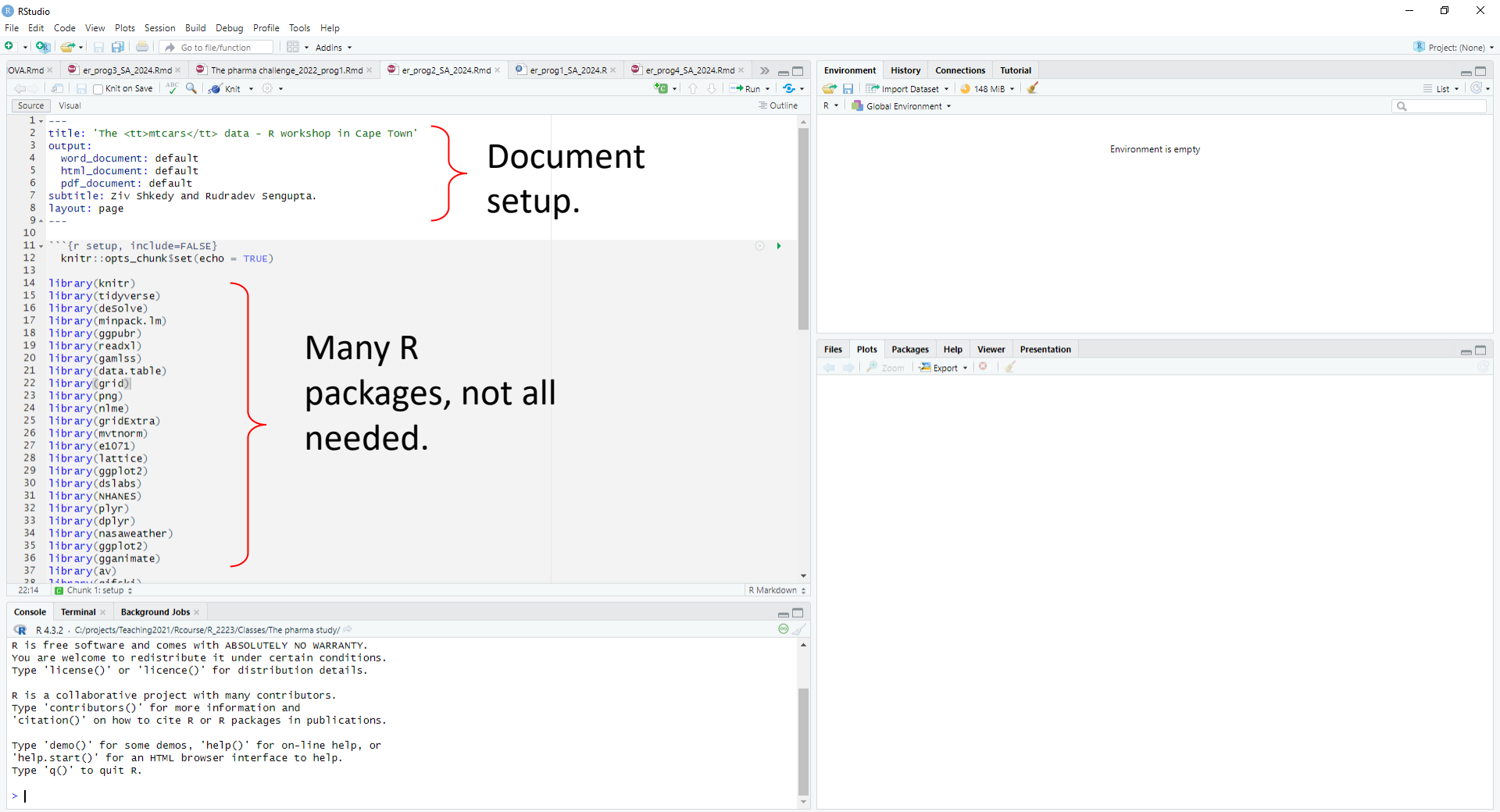
Analysis using basic Rmd programming

R Program: er_prog2_SA_2024.Rmd

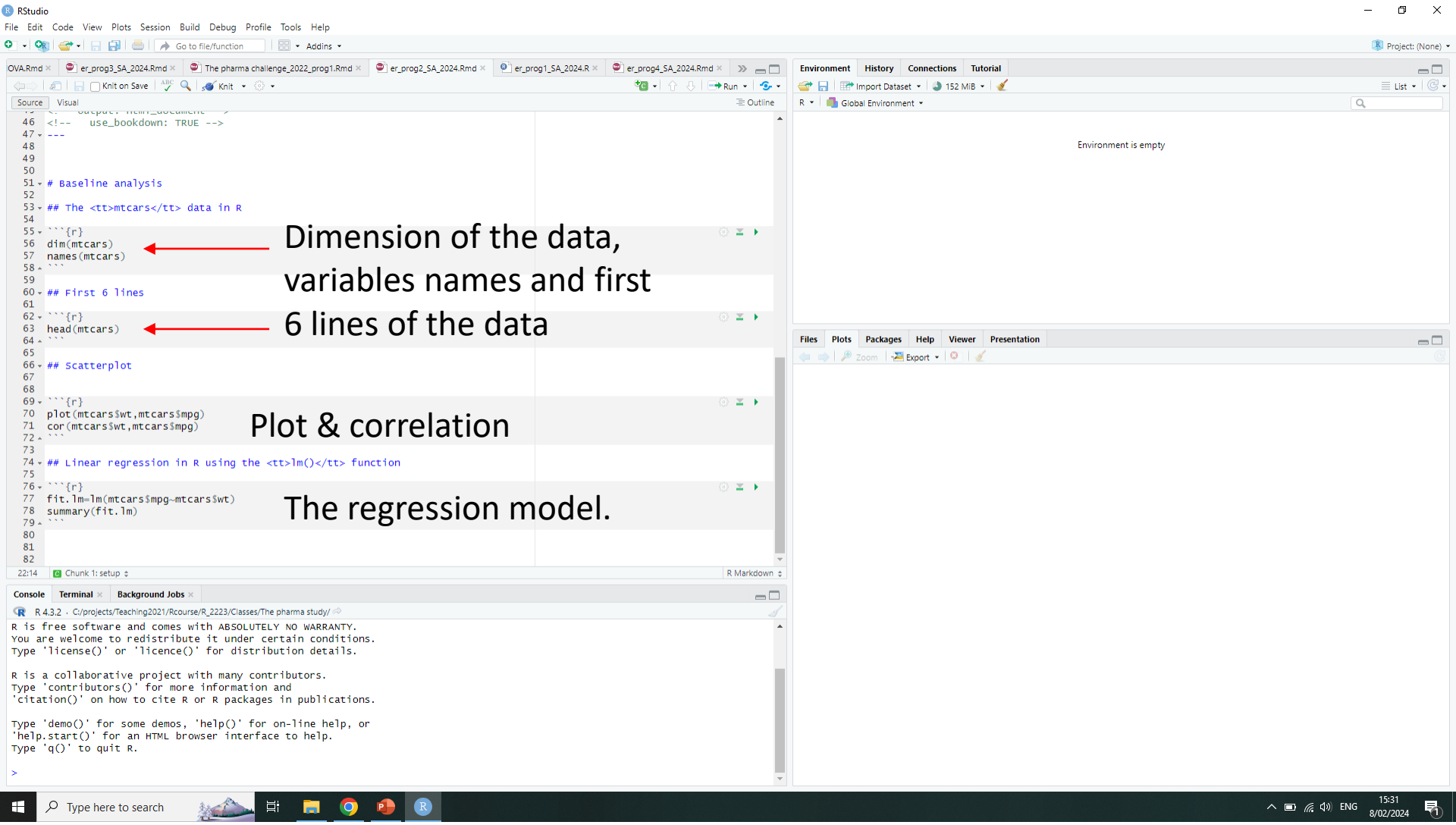
The output

- We run the same analysis as before.
- Use R markdown.
- Produce the possible output formats:
  - HTML.
  - PDF.
  - Word doc.

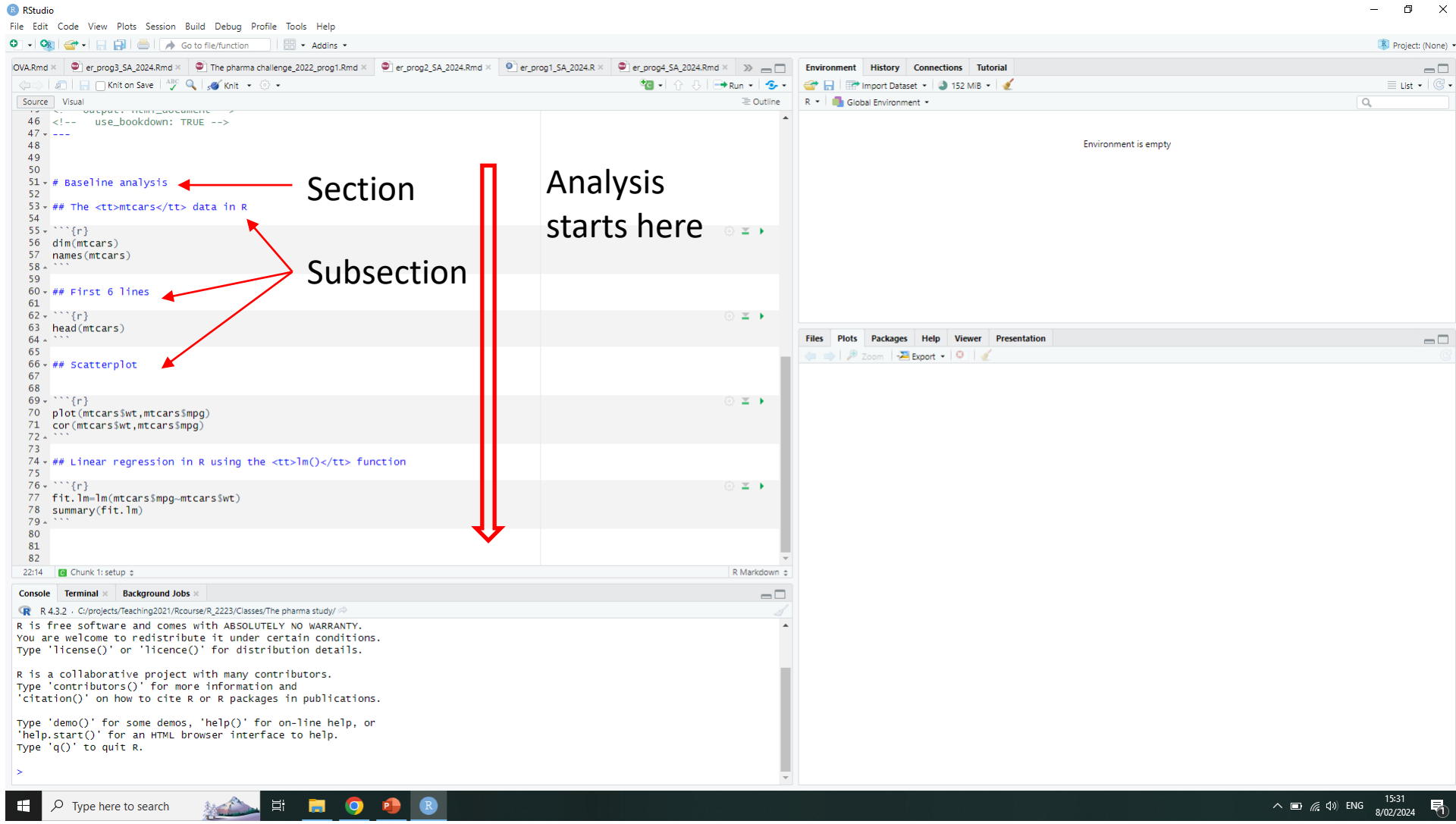
The Rmd program



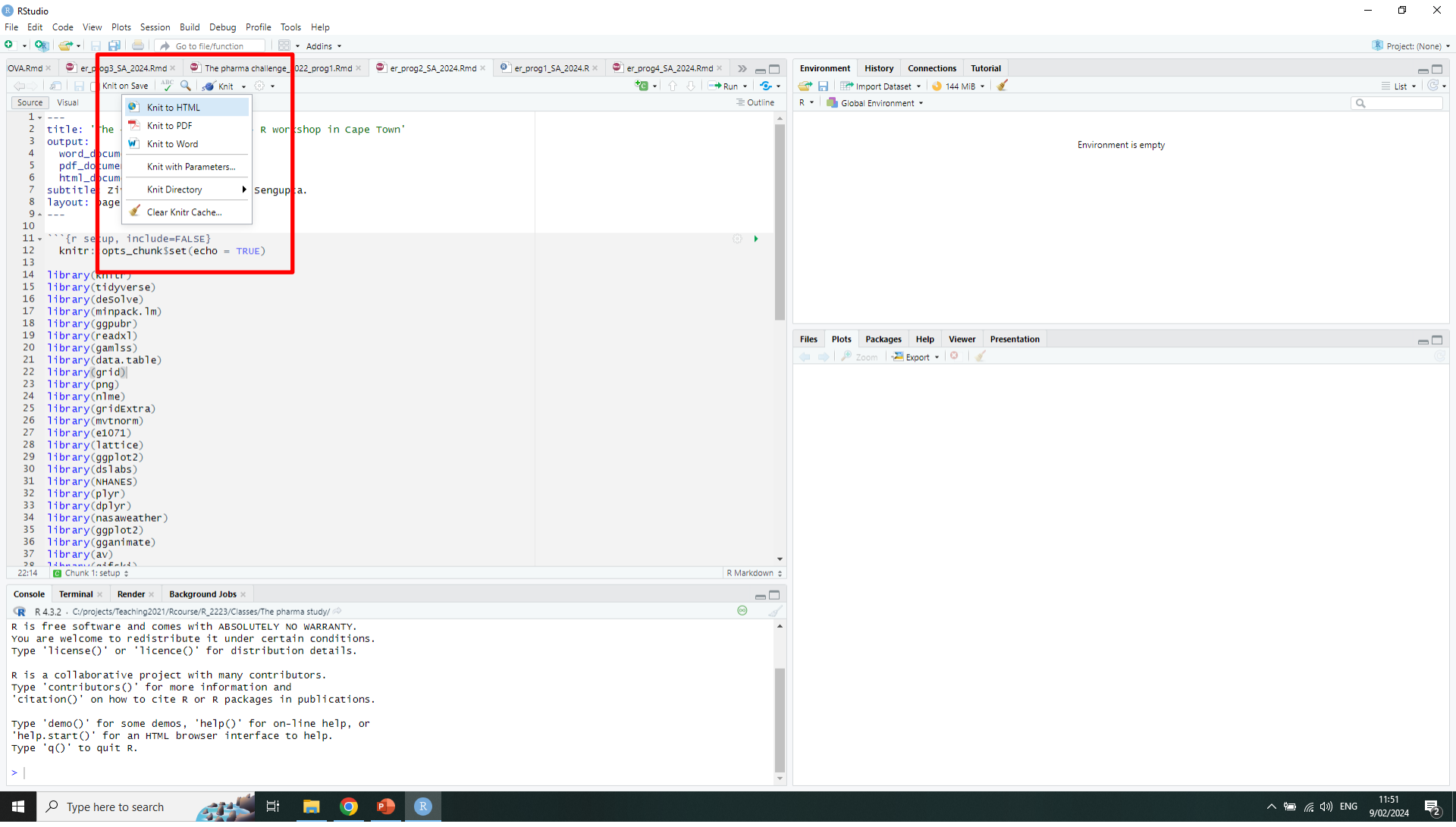
The Rmd program



The Rmd program



Choose the output



The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog2_SA_2024.html

er_prog2_SA_2024.html | Open in Browser | Find | Publish

Section

Subsection

The `mtcars` data - R workshop in Cape Town

Ziv Shkedy and Rudradev Sengupta.

Baseline analysis

The `mtcars` data in R

```
dim(mtcars)
```

```
## [1] 32 11
```

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

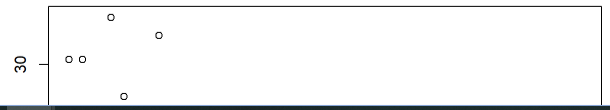
First 6 lines

```
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
```

Scatterplot

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



Type here to search

ENG 15:31 8/02/2024

The PDF output

er_prog2_SA_2024.pdf - Adobe Acrobat Reader (32-bit)

File Edit View Sign Window Help

Home Tools er_prog2_SA_2024.... Sign In

1 / 3 125%

The mtcars data - R workshop in Cape Town

Ziv Shkedy and Rudradev Sengupta.

Baseline analysis

The mtcars data in R

```
dim(mtcars)
```

```
## [1] 32 11
```

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

First 6 lines

```
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
```

Scatterplot

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

Search tools

Export PDF

Edit PDF

Create PDF

Comment

Combine Files

Organize Pages

Request E-signatures

Fill & Sign

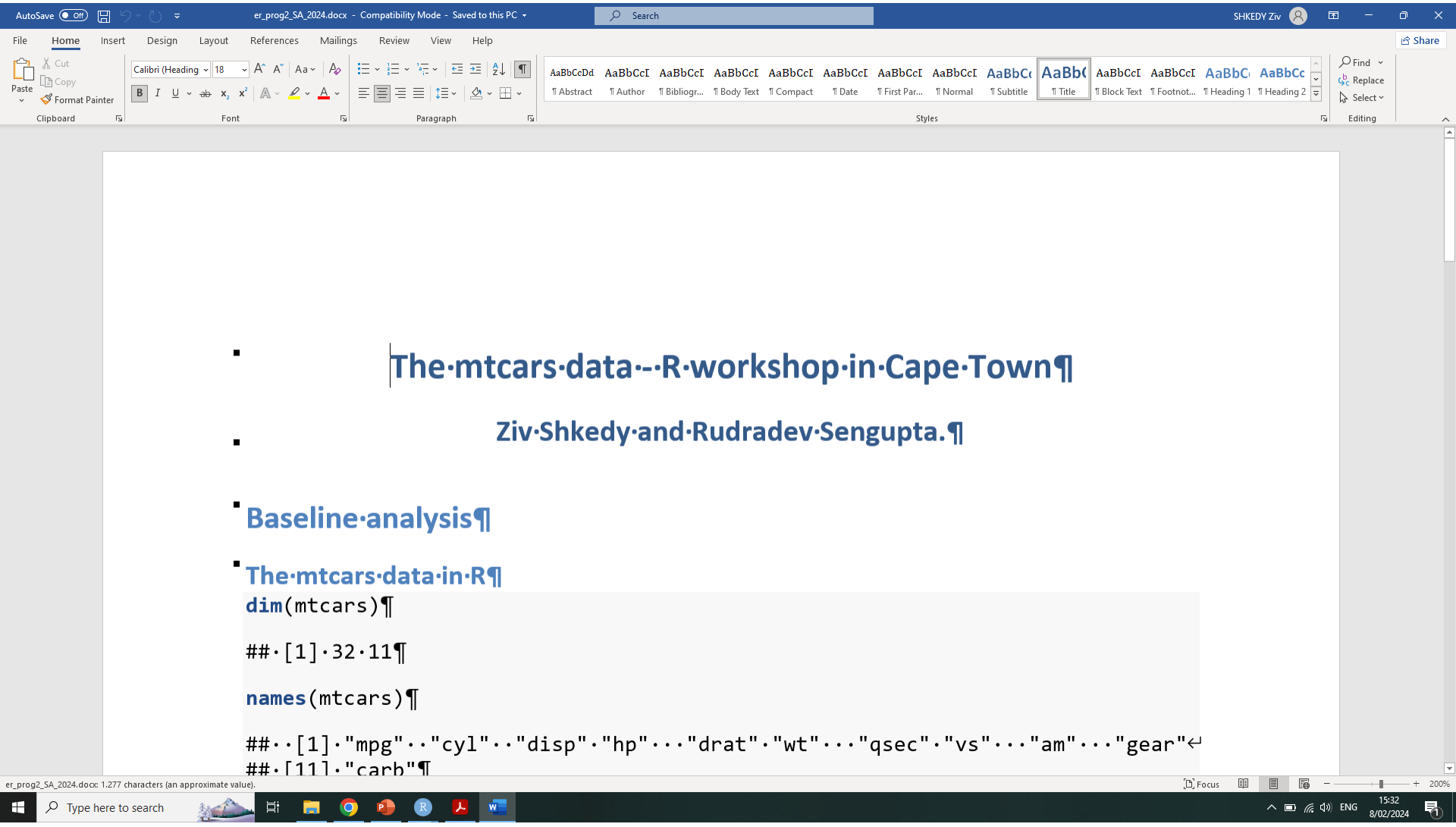
More Tools

Convert, edit and e-sign PDF forms & agreements

Free 7-Day Trial

Type here to search ENG 15:32 8/02/2024

The Word doc output



UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

The `mtcars` data

Part 3

Advance HTML output

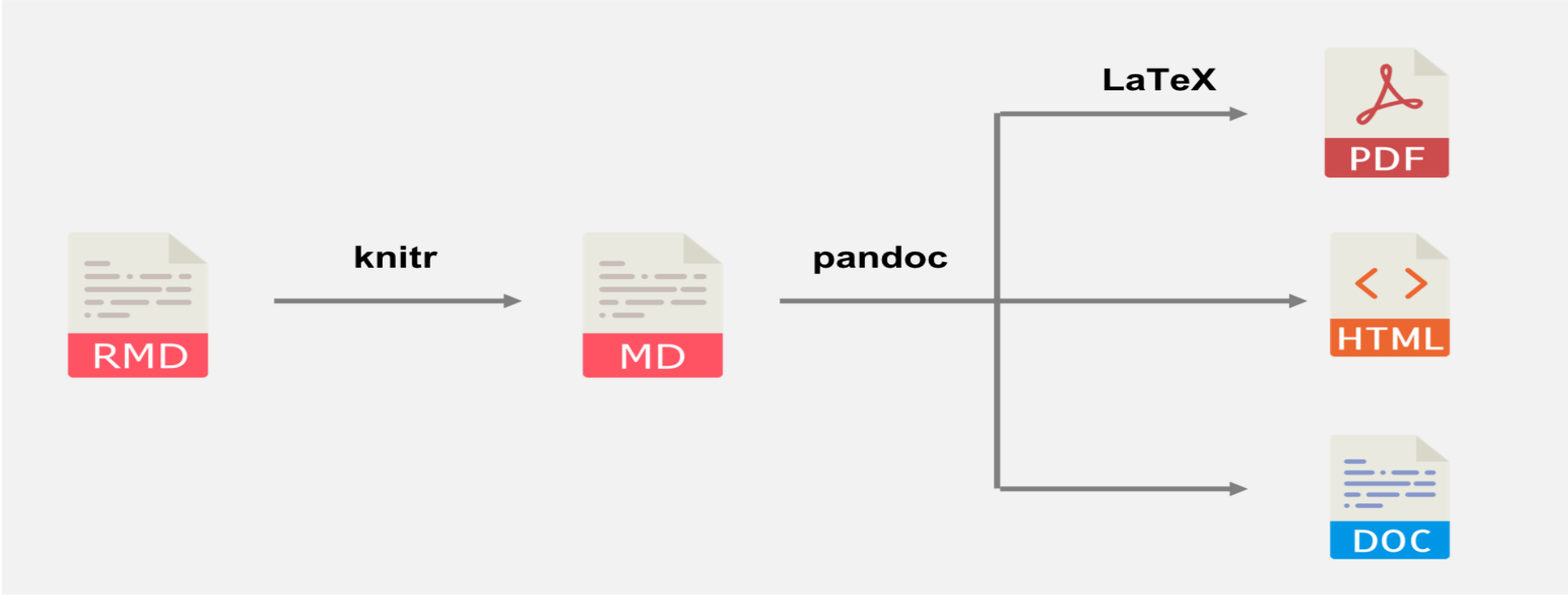
R Program: er_prog3_SA_2024.Rmd

Reproducible Research

- Aim: create an output in a different (highest) quality.
- 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 statistic formats (html, pdf, word, books, dashboard, e.t.c).



Knitr: A General-Purpose Package for Dynamic Report Generation in r. https://yihui.org/knitr/.

The HTML output

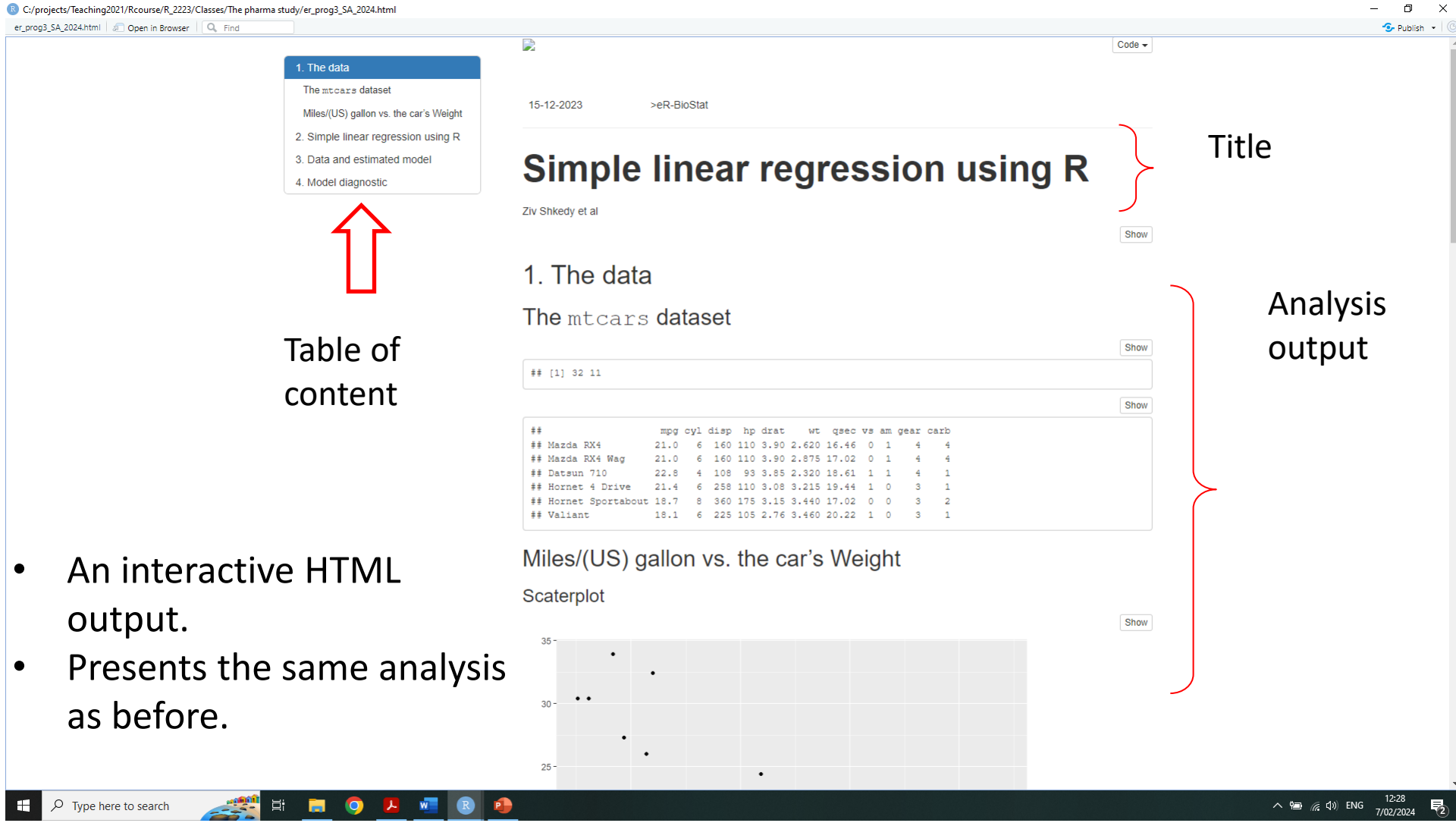


Table of content

Title

Analysis output

- An interactive HTML output.
- Presents the same analysis as before.

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

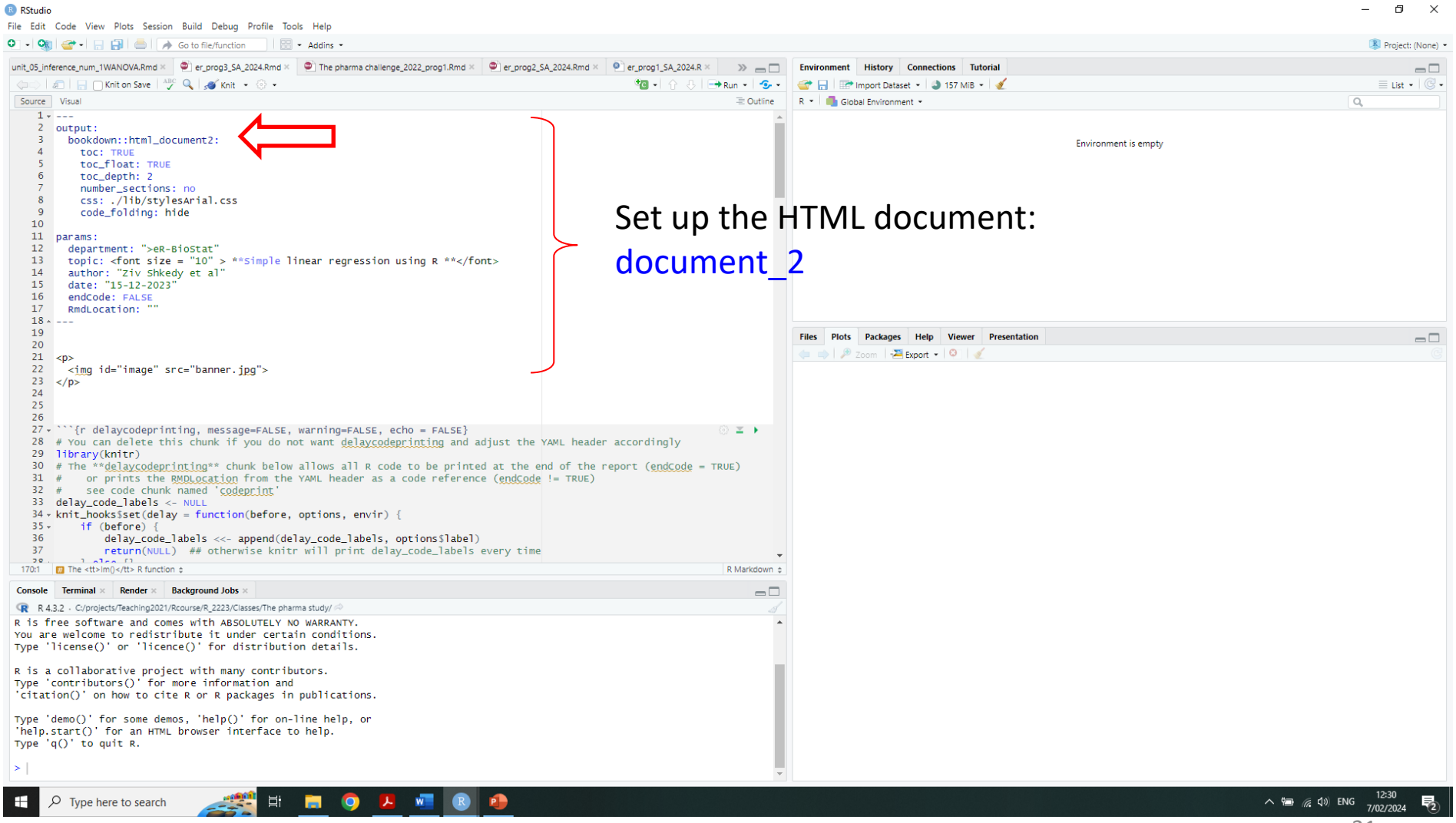
vliruos SHARING MINDS, CHANGING LIVES

Part 3.1:
How to set up the HTML file ?

The Rmd file

- We use Rmd file to
  - Conduct the analysis.
  - Set up the document.
- We use html file to
  - Present & communicate the result.

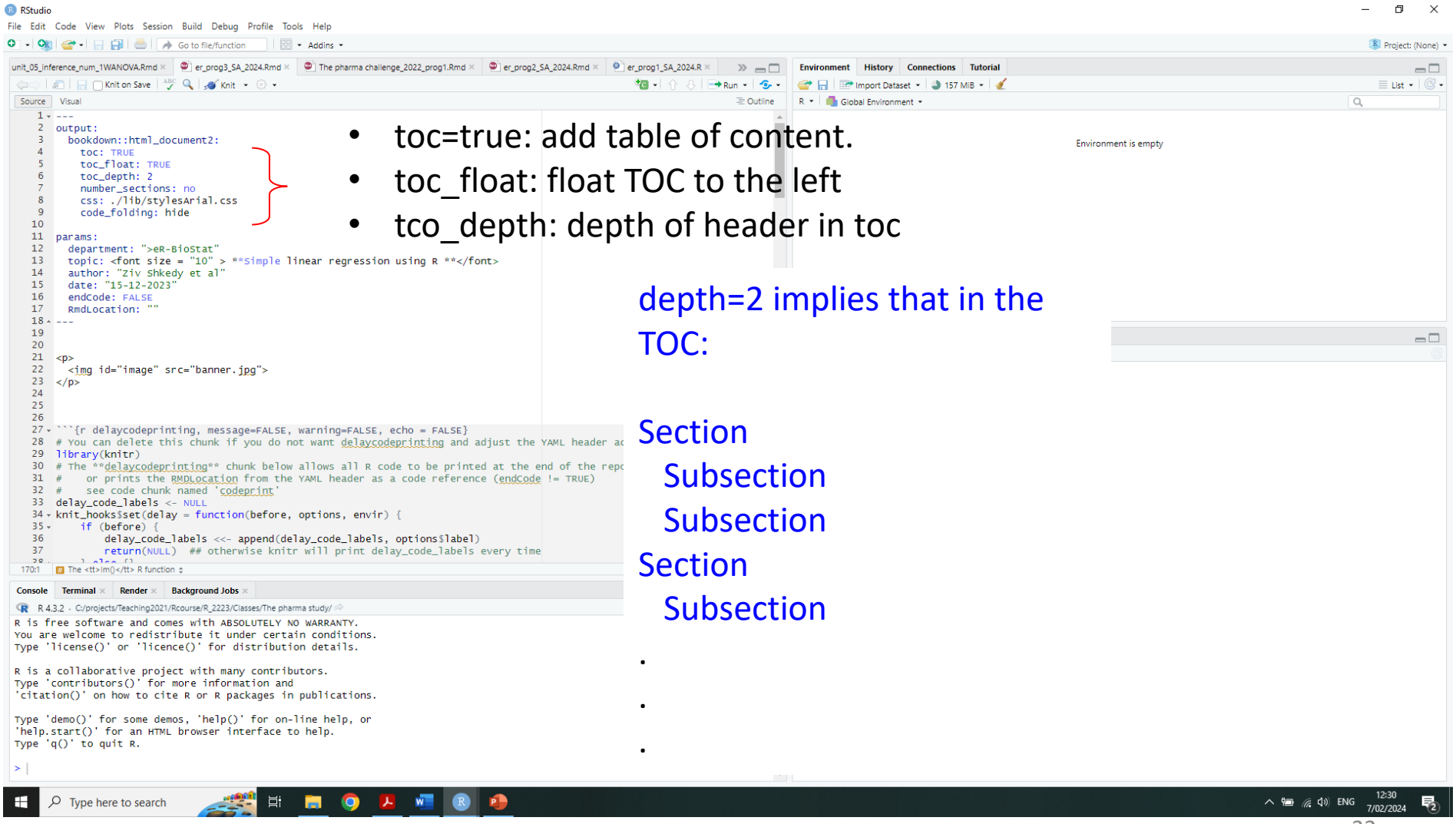
Set up the document



Set up the HTML document:

document_2

Set up the document



- toc=true: add table of content.
- toc_float: float TOC to the left
- tco_depth: depth of header in toc

depth=2 implies that in the TOC:

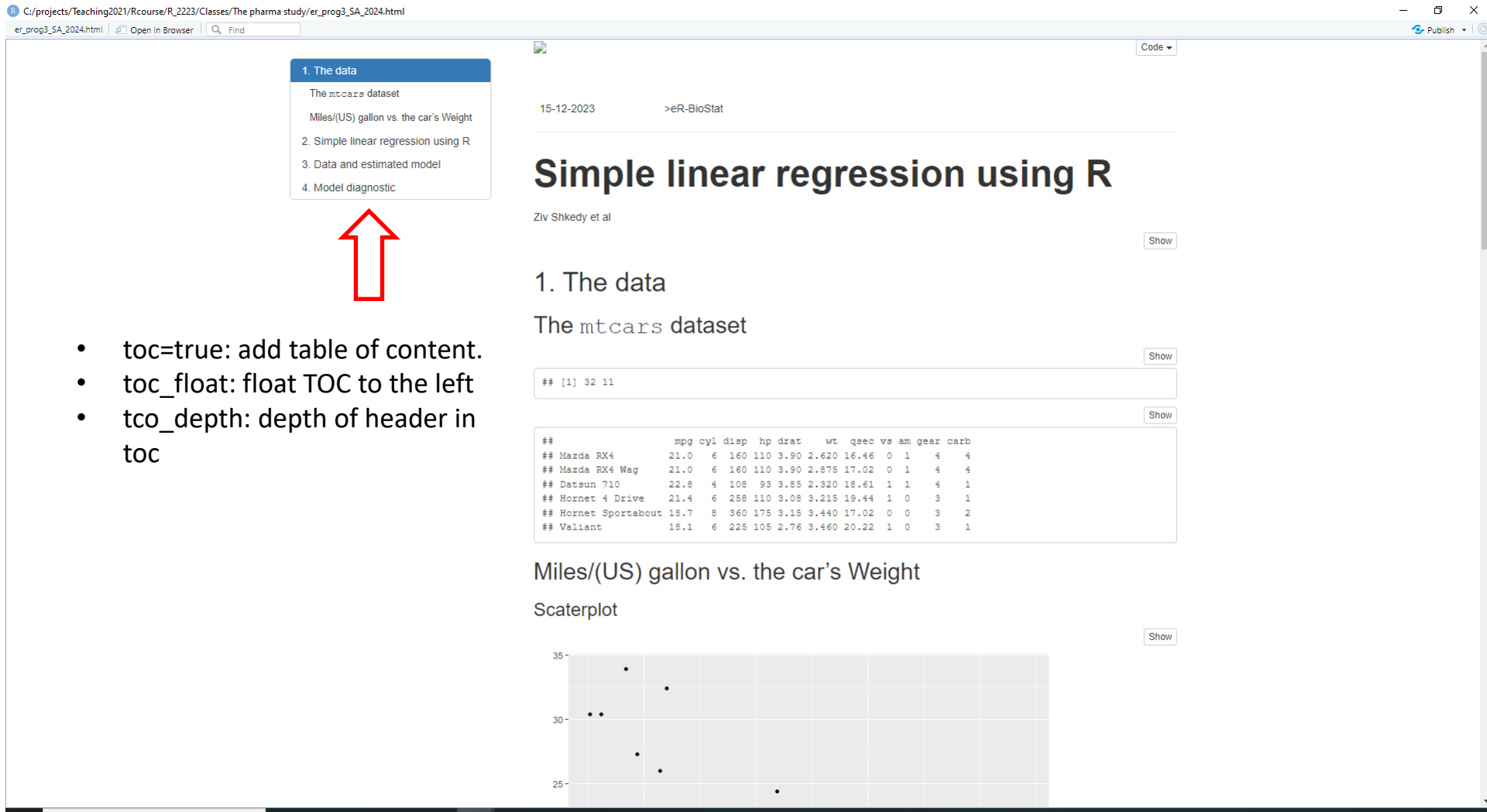
Section
  Subsection
  Subsection
Section
  Subsection

.

.

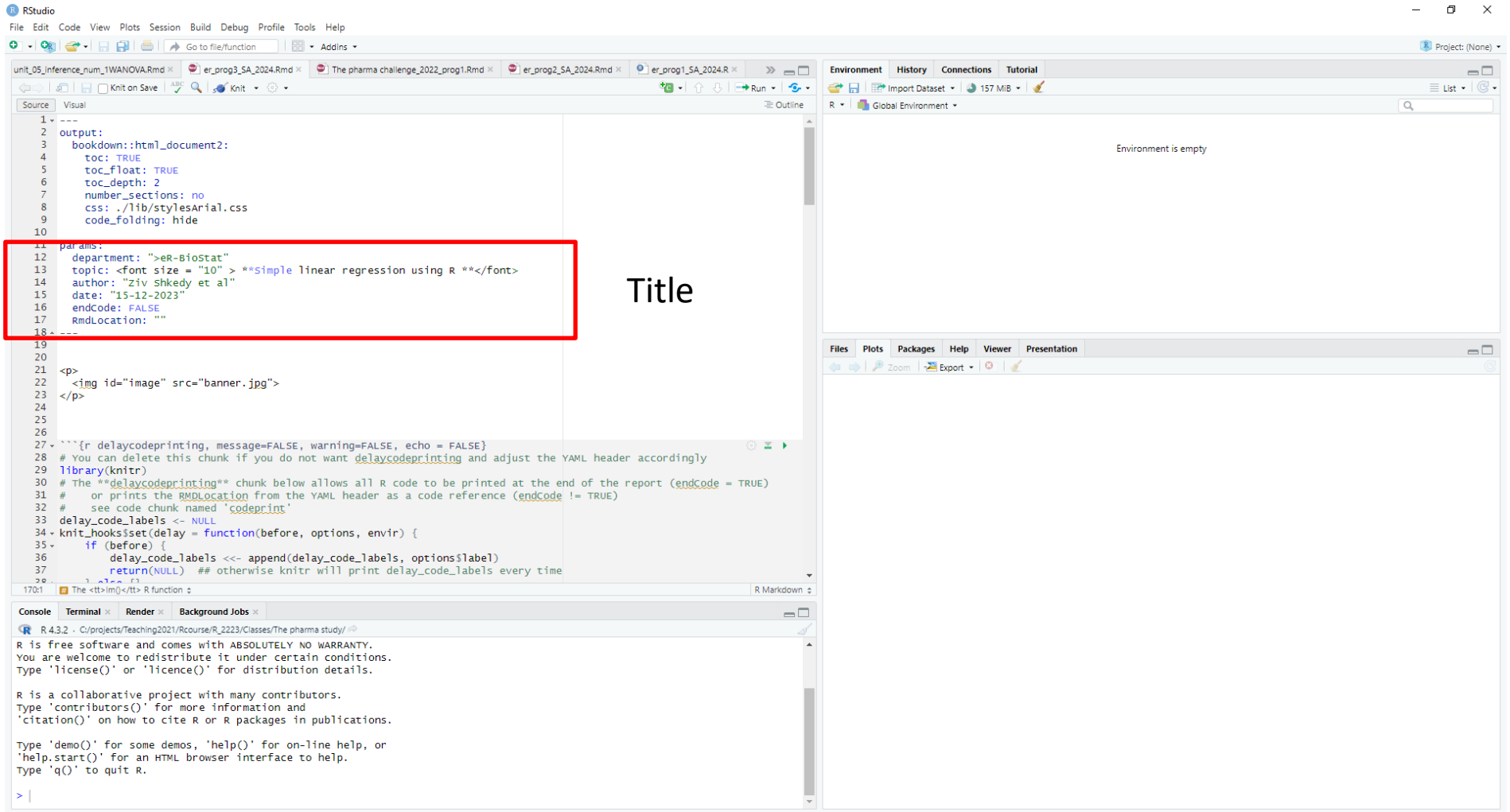
.

The HTML file



- toc=true: add table of content.
- toc_float: float TOC to the left
- tco_depth: depth of header in toc

Titles, authors and dates



Titles, authors and dates

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

1. The data
   - The mtcars dataset
   - Miles/(US) gallon vs. the car's Weight
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

Code

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

Show

Title

1. The data

The mtcars dataset

Show

```
## [1] 32 11
```

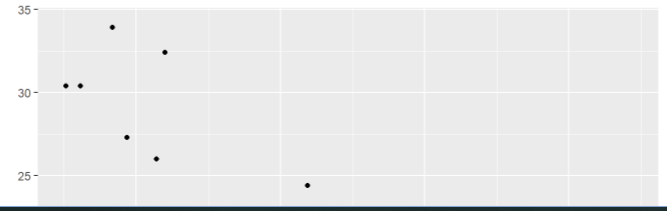
Show

```
##                   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
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show



Type here to search ENG 12:28 7/02/2024

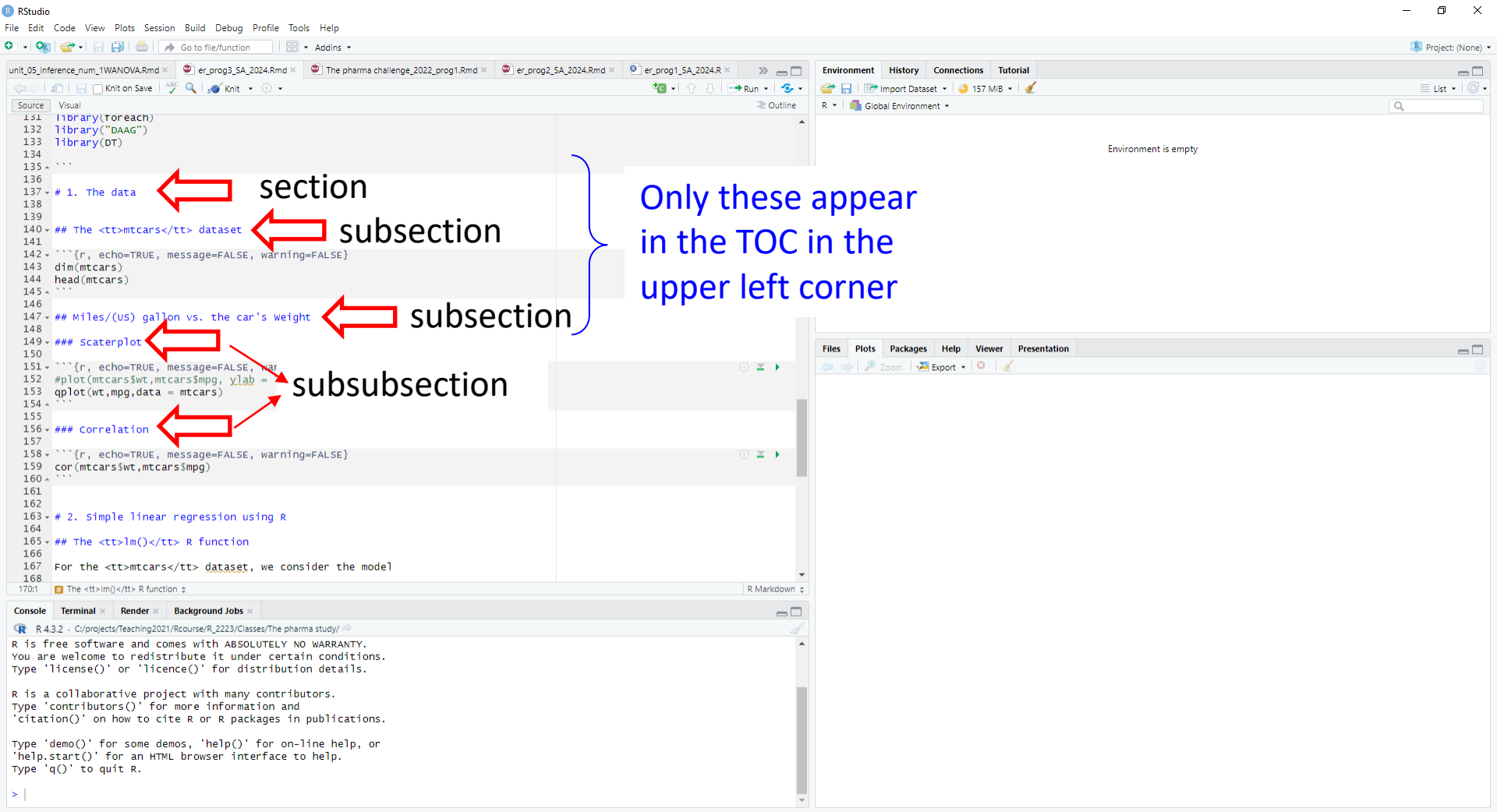
UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

Part 3.2:
The HTML file and the Rmd program in details.

Section, subsection, subsubsection



Section, subsection, subsubsection

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

Depth=2
Only sections
and subsections

Ziv Shkedy et al

Show

1. The data section

The mtcars dataset subsection

Show

```
## [1] 32 11
```

Show

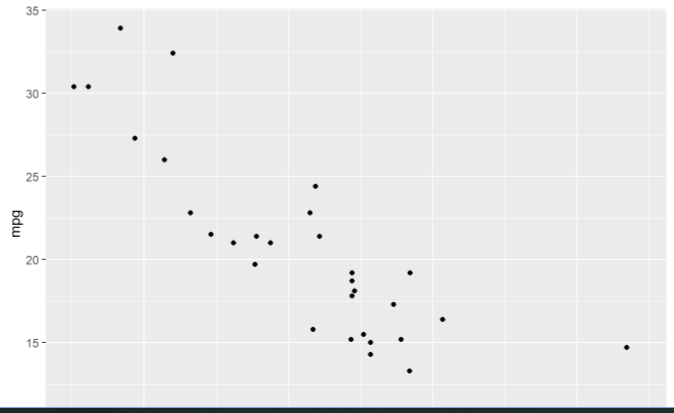
```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 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
```

Miles/(US) gallon vs. the car's Weight subsection

Since depth=2, the subsubsection will not appears in the TOC

Scaterplot subsubsection

Show



Type here to search

ENG 12:37 7/02/2024

Analysis code

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

The code for the analysis

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

Code

Show All Code

Hide All Code

1. The data
   - The mtcars dataset
   - Miles/(US) gallon vs. the car's Weight
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

1. The data

The mtcars dataset

Show

```
## [1] 32 11
```

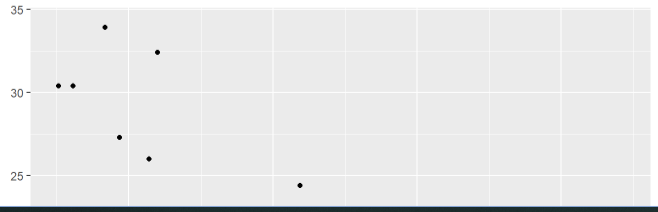
Show

```
##                   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
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show



We can choose if we want to show the code or to hide the code.

Type here to search ENG 12:40 7/02/2024

Reading the external file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

Code

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

Show

1. The data

The mtcars dataset

Show

```
## [1] 32 11
```

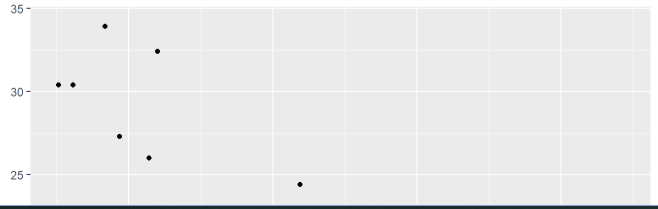
Show

```
##                   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
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show

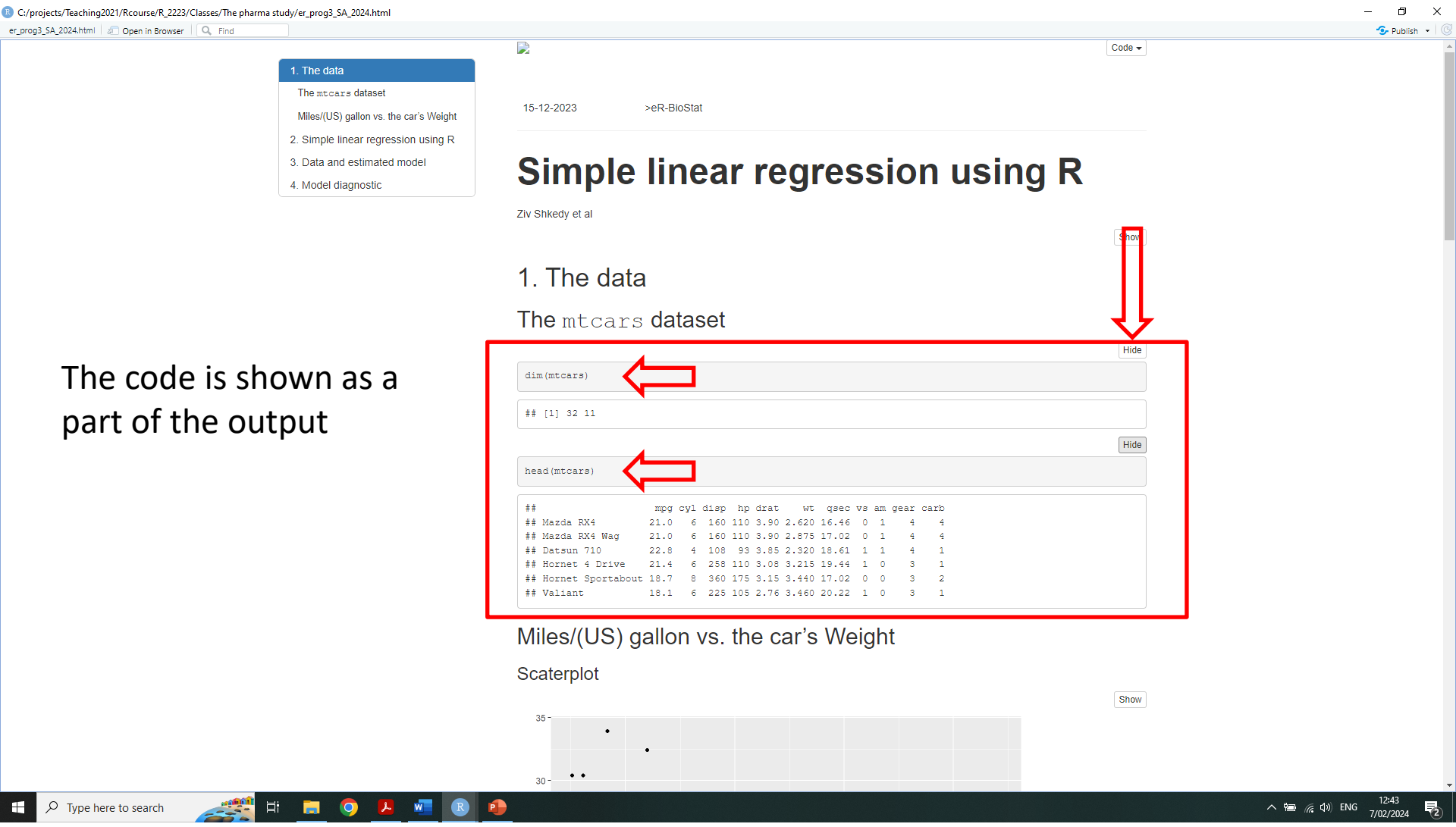


The code is not shown as a part of the output.

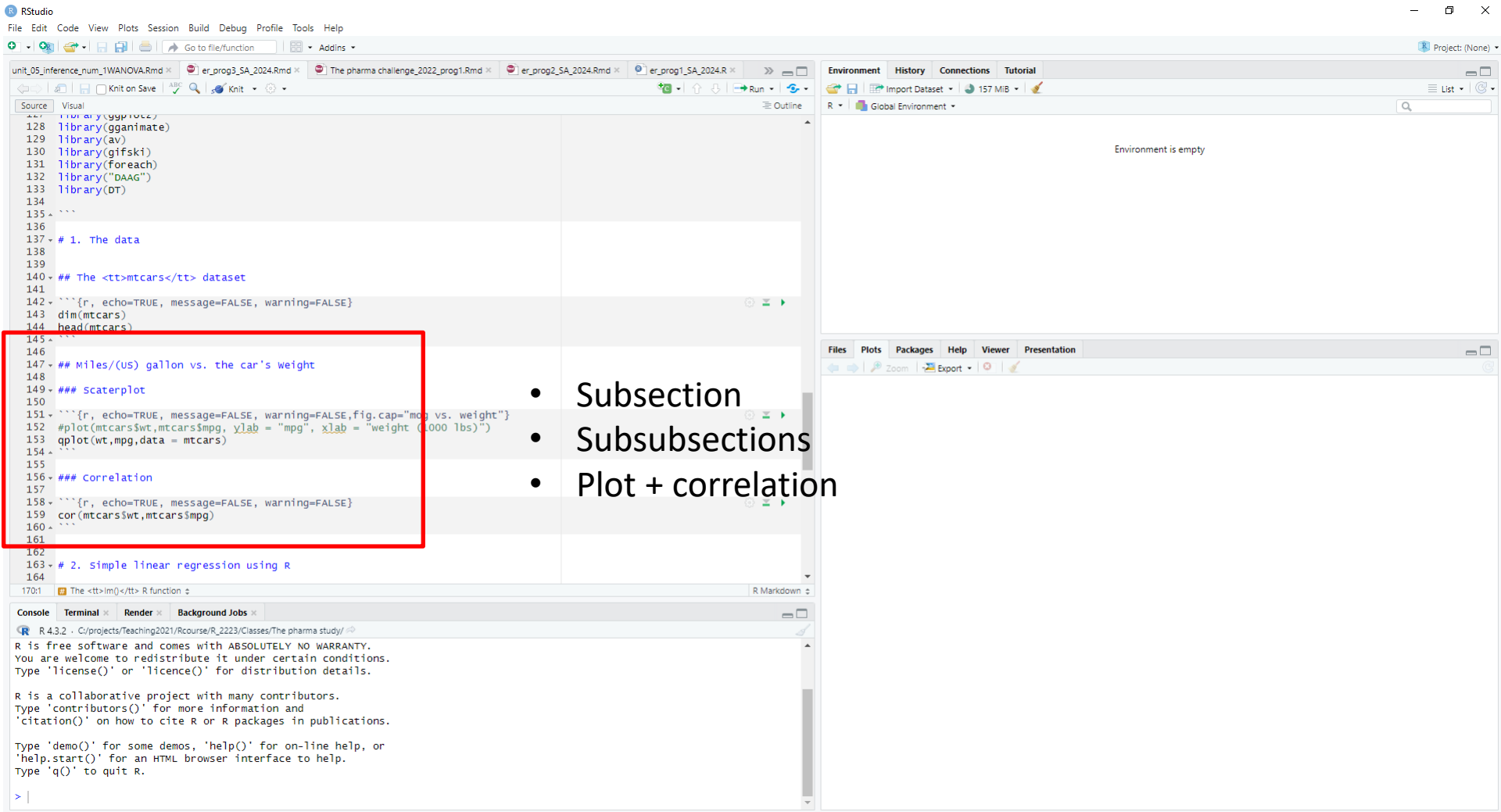
Type here to search ENG 12:42 7/02/2024

The HTML file

The code is shown as a part of the output



Code in the Rmd file



- Subsection
- Subsubsections
- Plot + correlation

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

1. The data
   - The `mtcars` dataset
   - Miles/(US) gallon vs. the car's Weight
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

```
## 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
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show

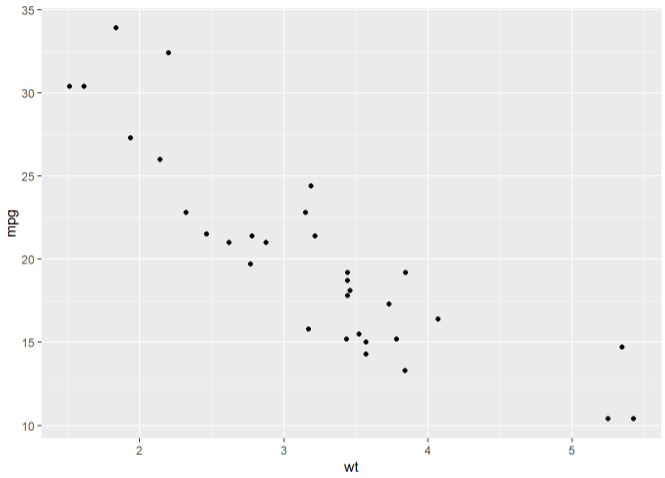


Figure 1: mog vs. weight

Correlation

Show

```
## [1] -0.8676594
```

2. Simple linear regression using R

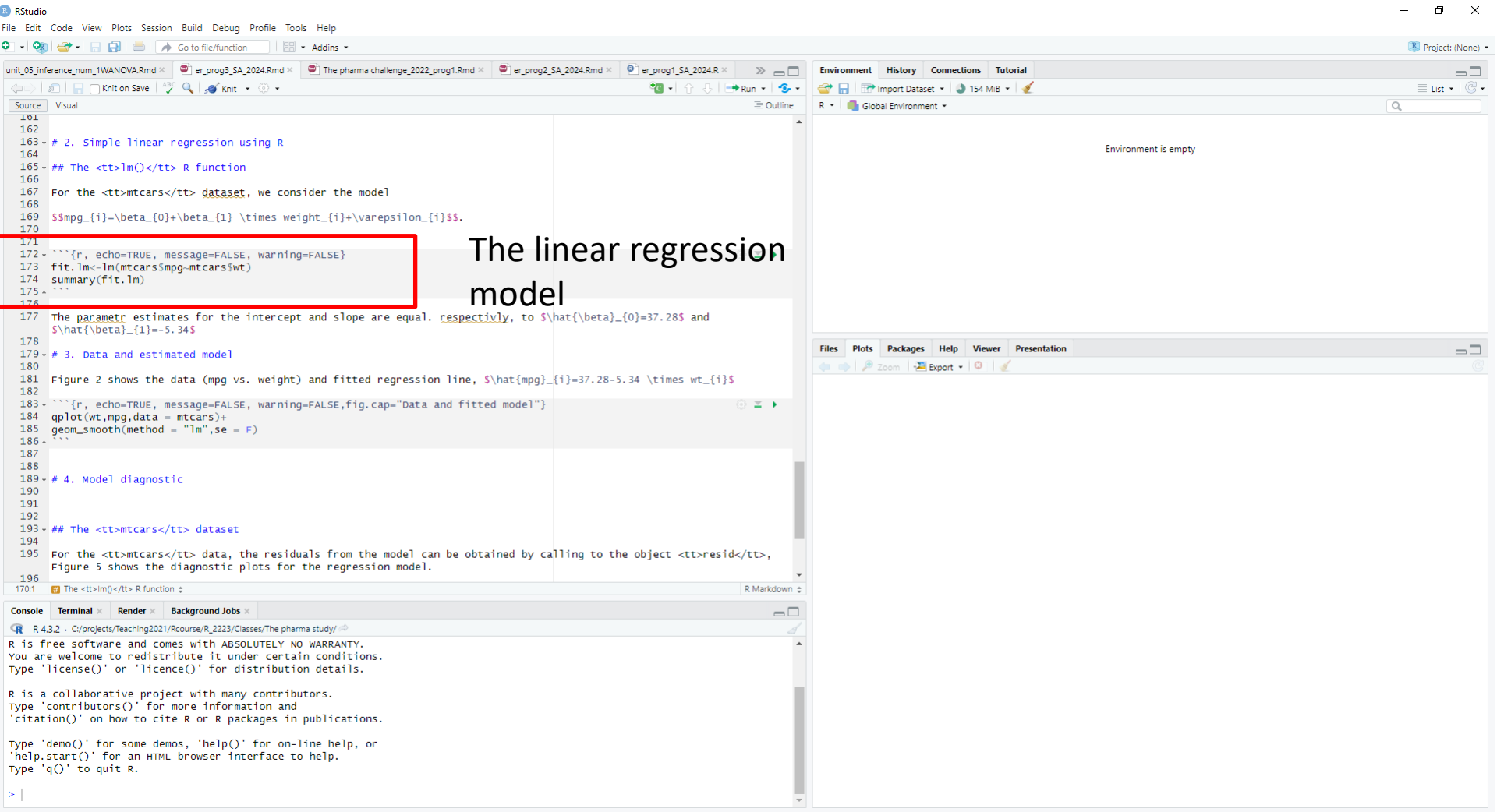
The `lm()` R function

For the `mtcars` dataset, we consider the model

$$mpg_i = \beta_0 + \beta_1 \times weight_i + \varepsilon_i$$

- Subsection
- Subsubsections
- Plot + correlation

Code in the Rmd file



The linear regression model

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

Correlation

Show

```
## [1] -0.8676594
```

2. Simple linear regression using R

The `lm()` R function

For the `mtcars` dataset, we consider the model

$$mpg_i = \beta_0 + \beta_1 \times weight_i + \varepsilon_i$$

Hide

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

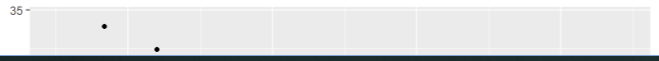
```
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$wt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## mtcars$wt    -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 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
```

The parametr estimates for the intercept and slope are equal. respectivly, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

3. Data and estimated model

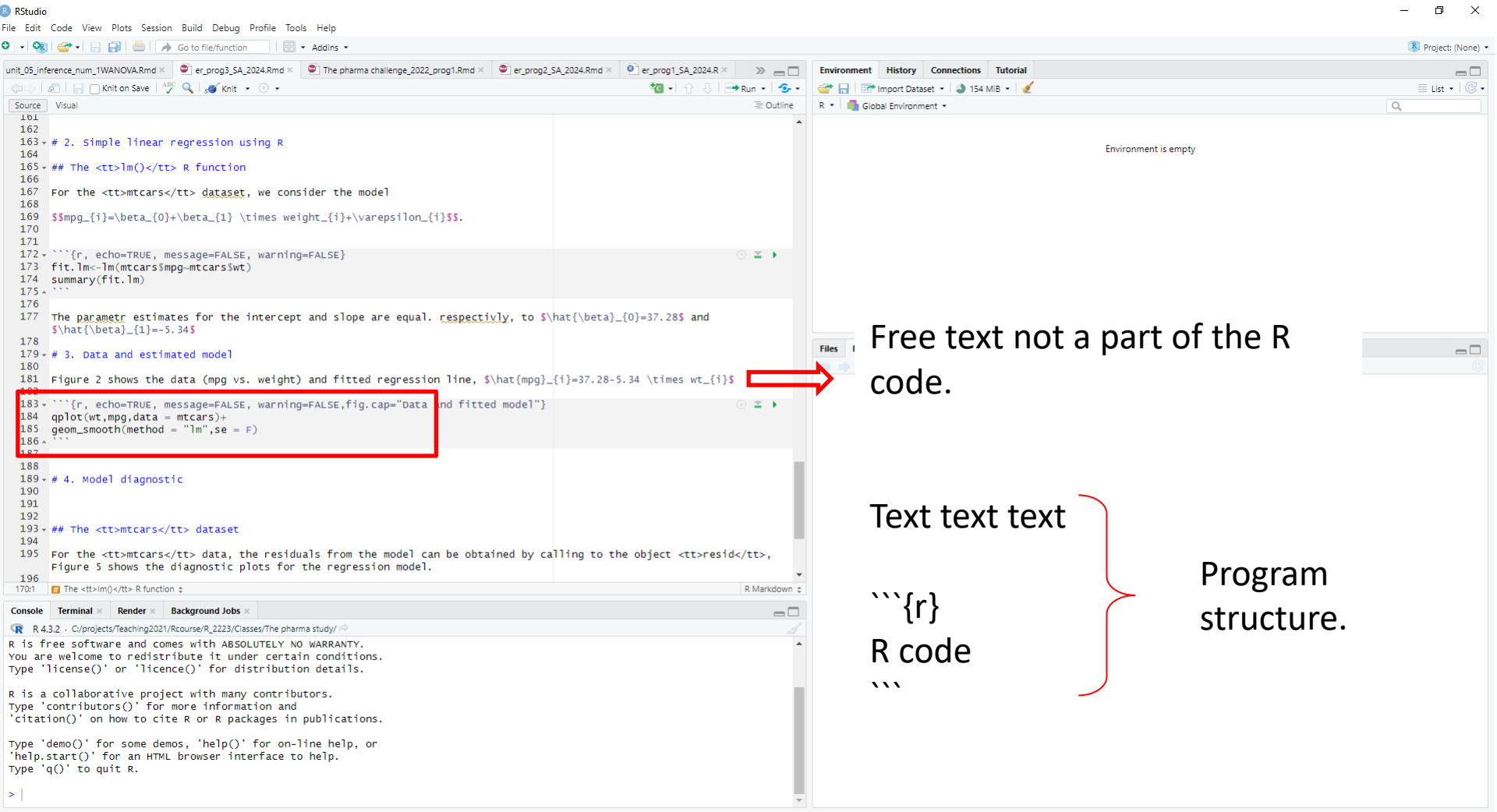
Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{mpg}_i = 37.28 - 5.34 \times wt_i$

Show



Code for the model is shown.

Code in the Rmd file



Free text not a part of the R code.

Text text text

```{r}
R code
```

Program structure.

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html

er_prog3_SA_2024.html | Open in Browser | Find | Publish

1. The data
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

```
##
## 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
```

The parametr estimates for the intercept and slope are equal. respectivly, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

3. Data and estimated model

Free text

Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{mpg}_i = 37.28 - 5.34 \times wt_i$

Show

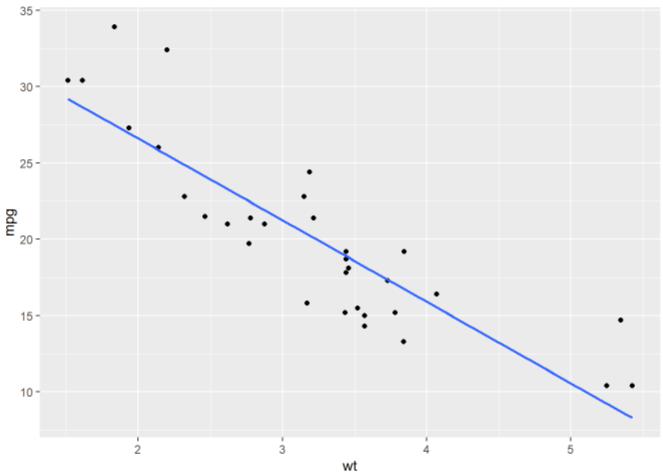


Figure 2: Data and fitted model

4. Model diagnostic

The `mtcars` dataset

For the `mtcars` data, the residuals from the model can be obtained by calling to the object `resid`, Figure 5 shows the diagnostic plots for the regression model.

Show

```
##          1          2          3          4          5          6          7
## -2.2826106 -0.9197704 -2.0859521  1.2973499 -0.2001440 -0.6932545 -3.9053627
```

Type here to search

ENG 12:50 7/02/2024

Short discussion

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

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

Development of E-learning materials using R markdown

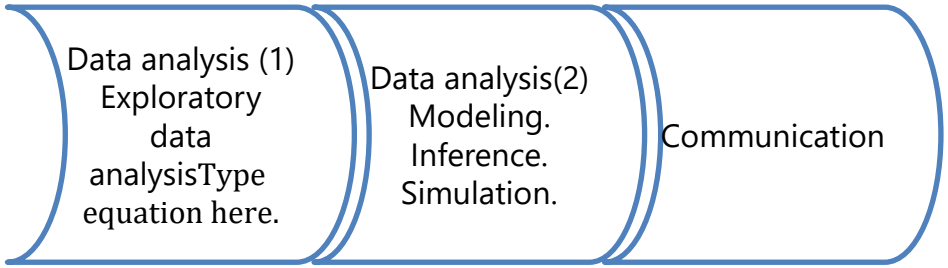
Part 4

The course online (1)

Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.

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



Boxplot by treatment group.

A simple linear regression.

A report.

Methodology: simple linear regression.

We "translate" the methodology to software usage

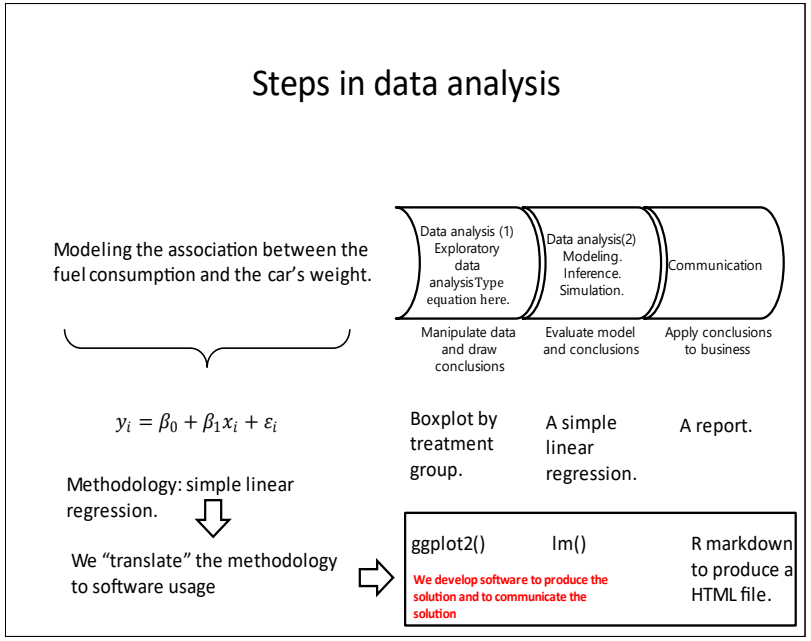
ggplot2()

lm()

R markdown to produce a HTML file.

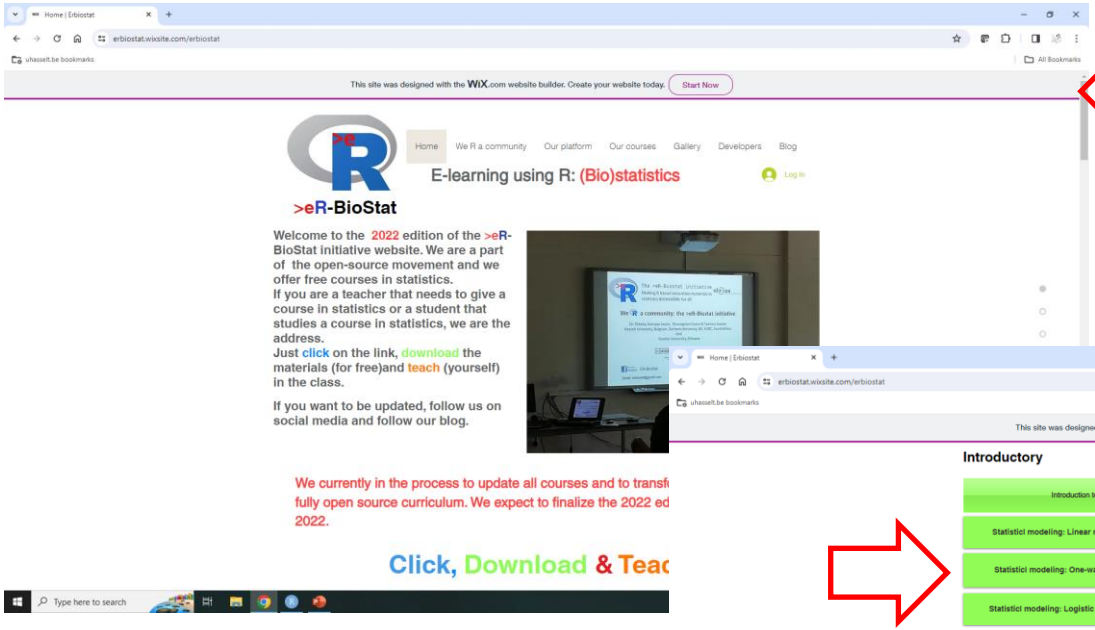
We develop software to produce the solution and to communicate the solution

Developing a course about linear regression



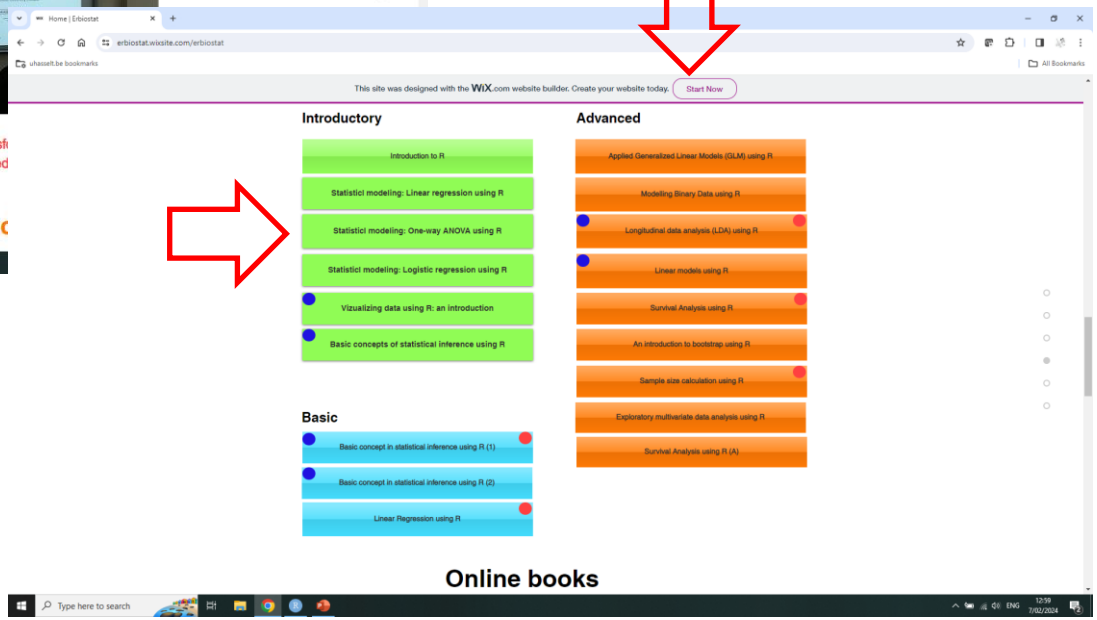
- Suppose that we do not need to produce a report for an analysis but…
- Our aim: development of education materials for a course about linear regression.

Introduction to statistical modeling using R



>eR-Biostat website.

List of courses.



https://erbiostat.wixsite.com/erbiostat

Introduction to statistical modeling using R

Home | Erbiostat | Home | IntroStatMod

erbiostat.wixsite.com/introstatmod

uhasselt.be bookmarks | All Bookmarks

Introduction to Statistical modeling using R

Home About Topics Online tutorials Contact

>eR-BioStat

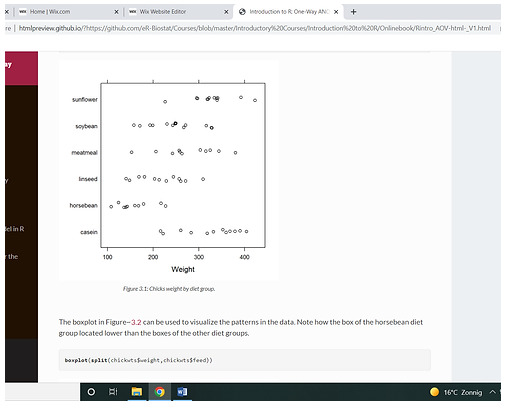
This group of courses is focused on statistical modelling and covers the following topics:

- Simple linear regression using R.
- One-way ANOVA using R.
- Logistic Regression using R.

The courses can be given together, as three parts of a course about statistical modelling or separately as a part of a specific course in statistical modelling.

The courses were developed at an undergraduate level courses (for both statistician and non statisticians) .

This is an open source course and all source files used to produce the slides are available online (in PP, Tex or Rmd formats).



About

General information about the course and course materials and the study methods used in the course.

Topics

In this page, the course is presented in a typical slides format. The course

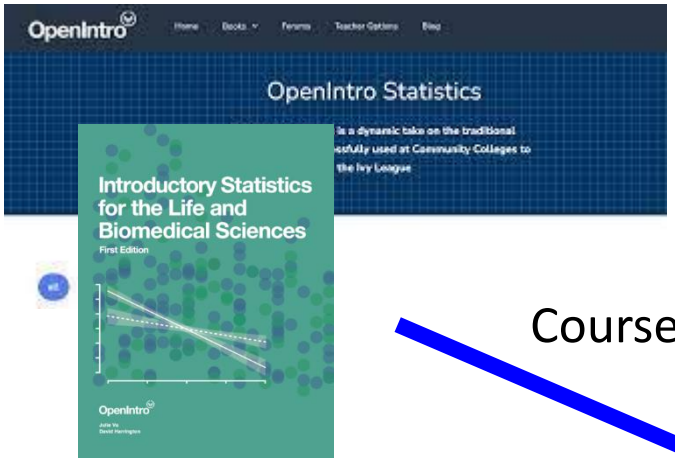
Online tutorials

In this page, suppurating online tutorials are given in different formats. The online

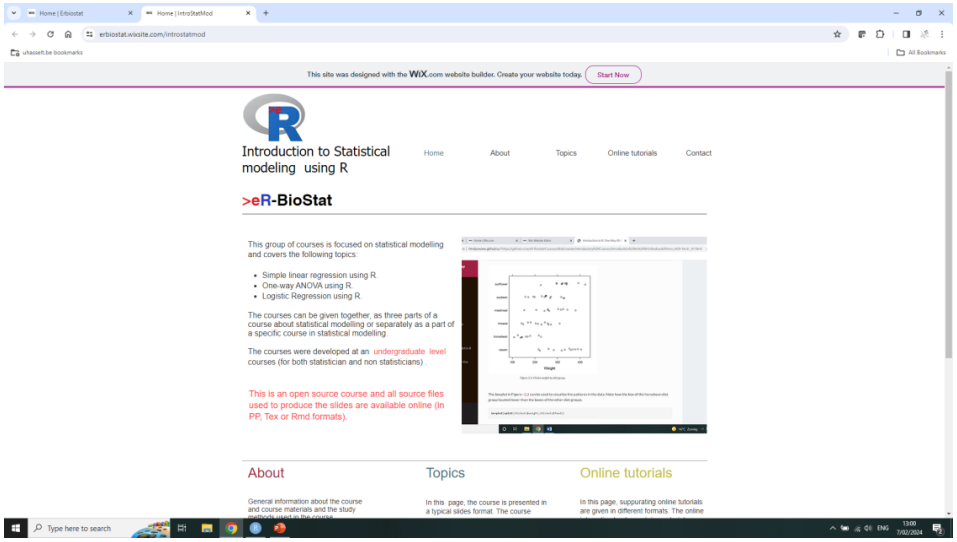
Type here to search ENG 13:00 7/02/2024

https://erbiostat.wixsite.com/introstatmod

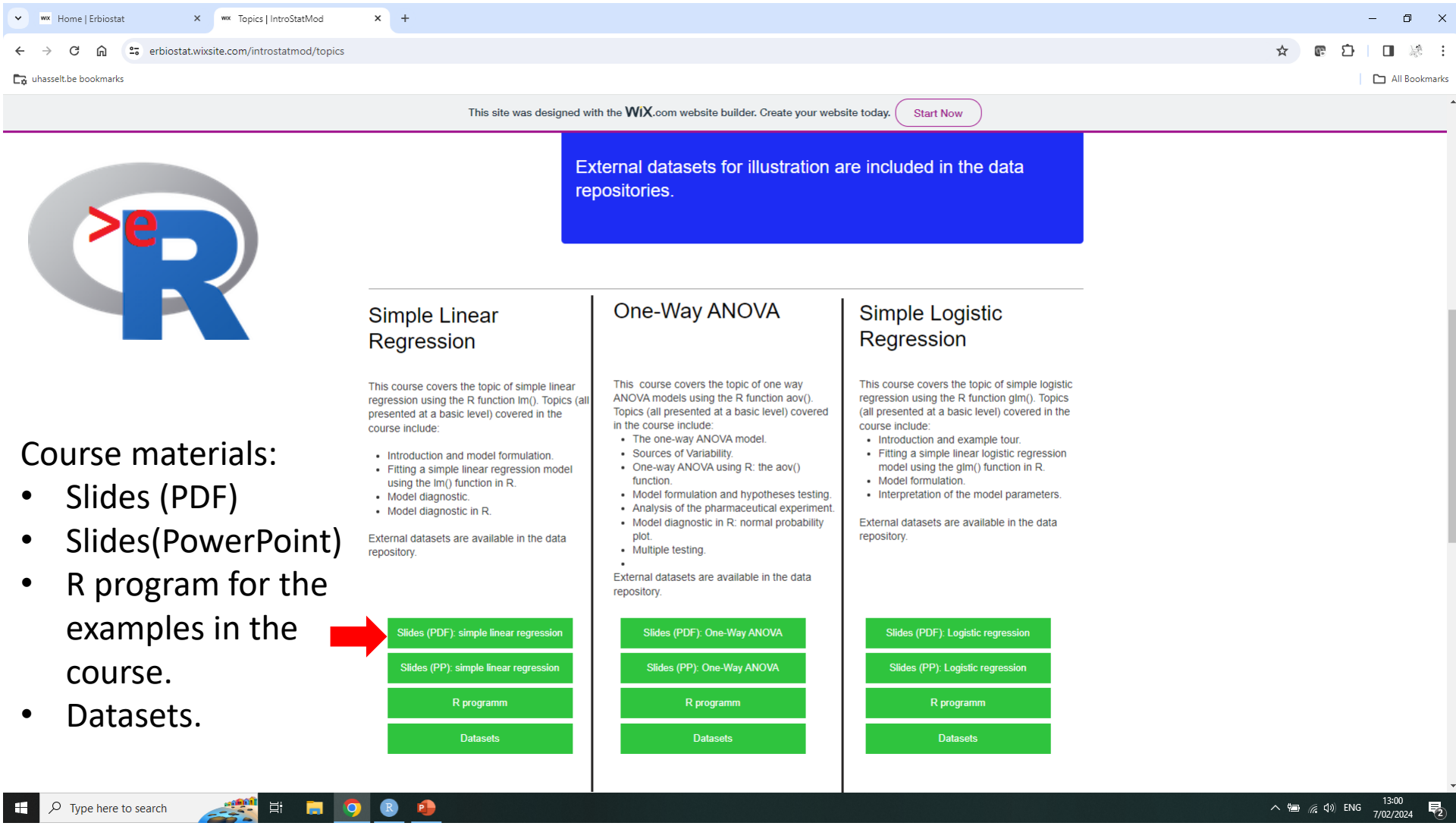
Introduction to statistical modeling using R



Course materials from two sources



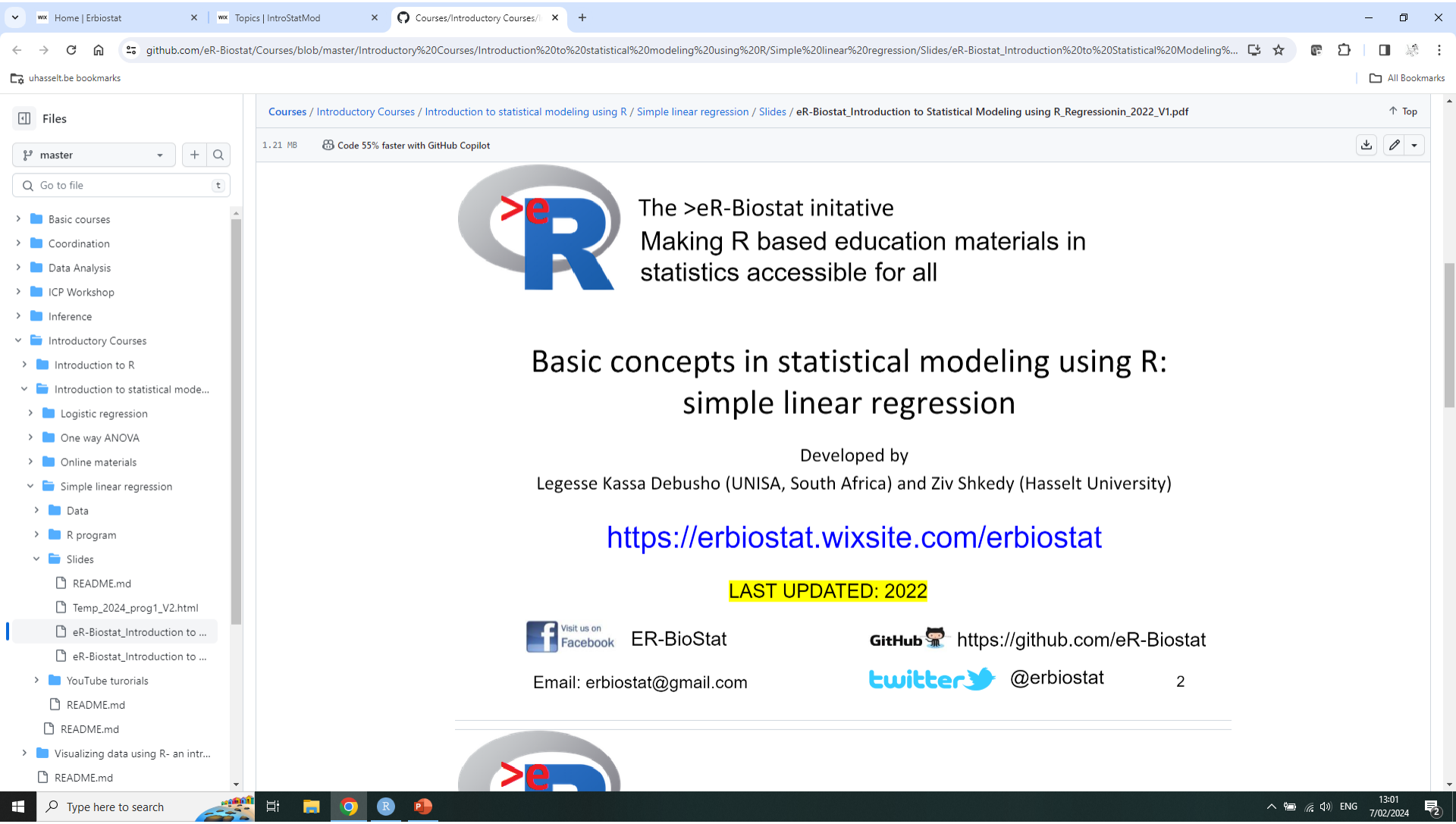
Course I



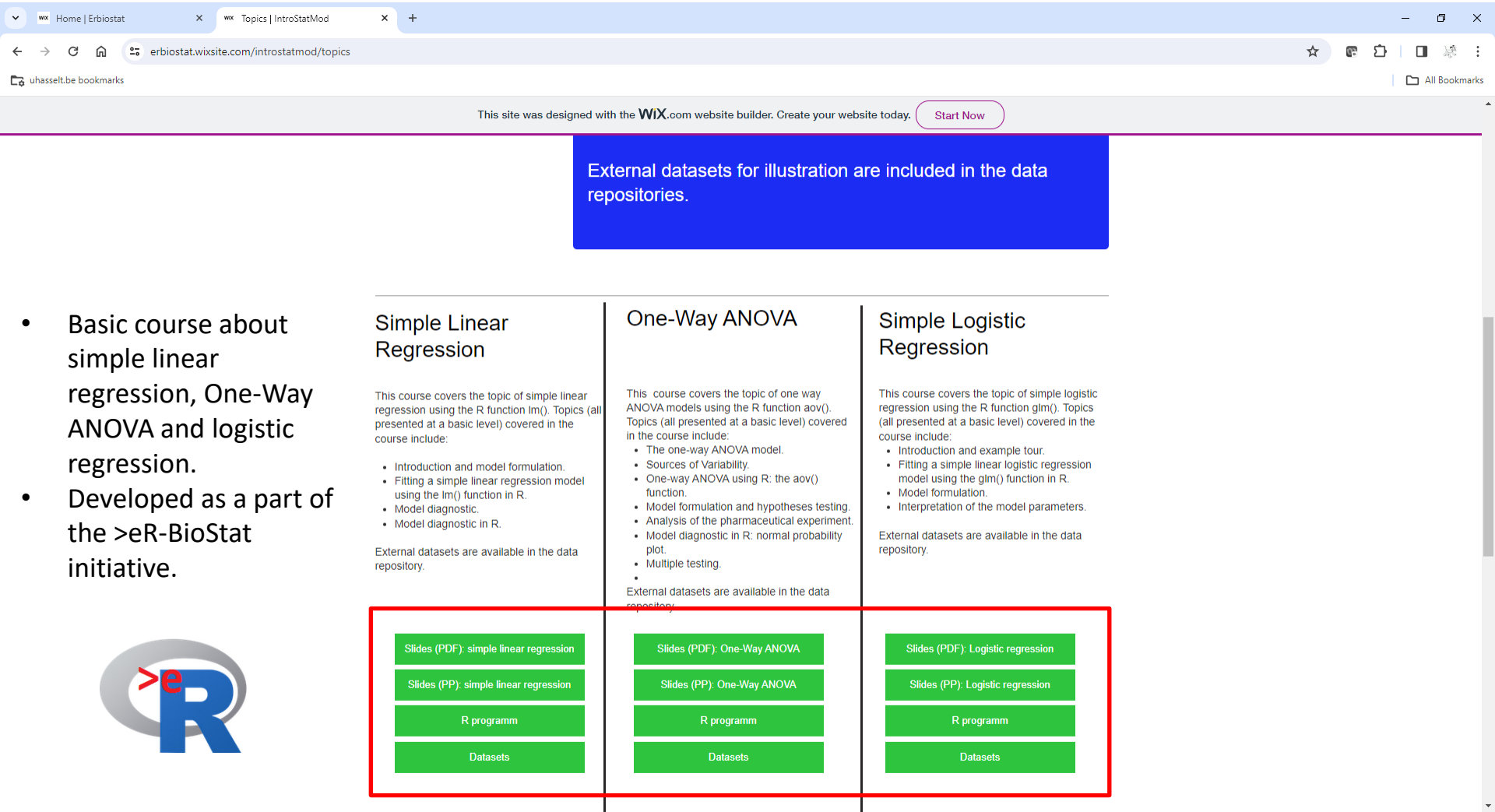
Course materials:

- Slides (PDF)
- Slides(PowerPoint)
- R program for the examples in the course.
- Datasets.

Examples of the slide

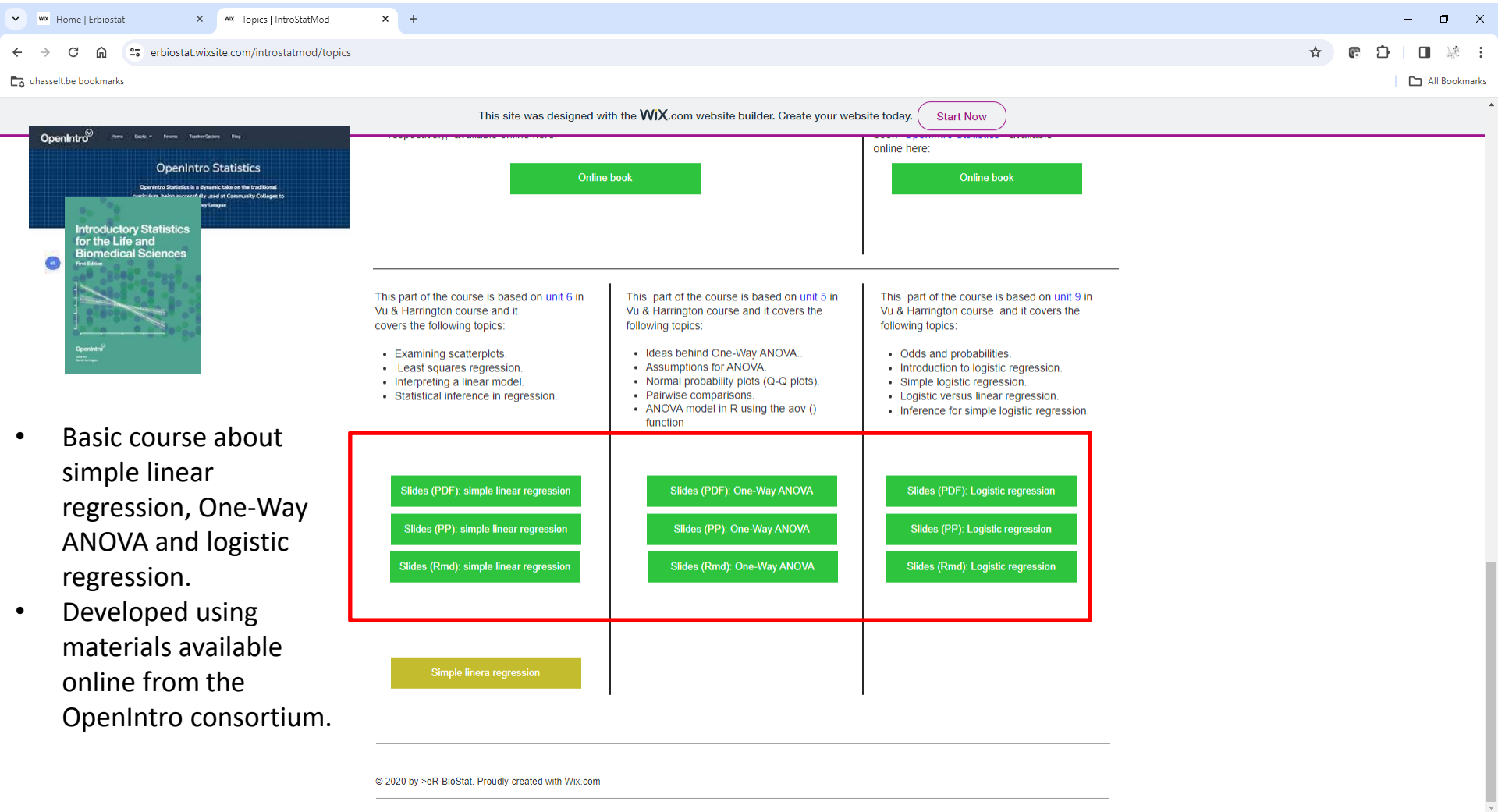


Course materials



- Basic course about simple linear regression, One-Way ANOVA and logistic regression.
- Developed as a part of the >eR-BioStat initiative.

Course II



- Basic course about simple linear regression, One-Way ANOVA and logistic regression.
- Developed using materials available online from the OpenIntro consortium.

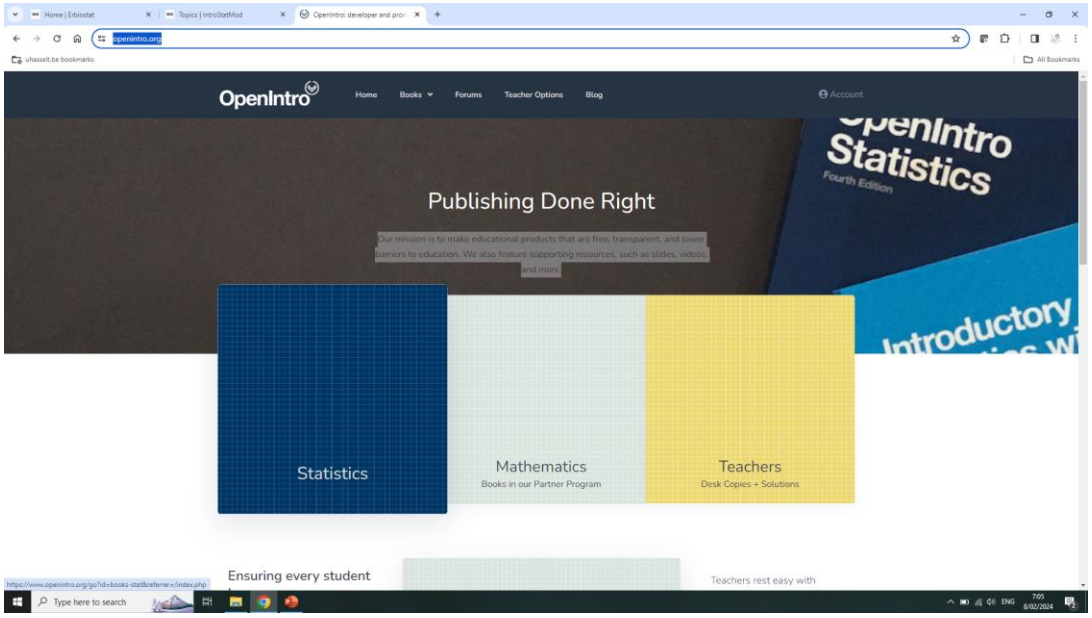
openintro.org

- The OpenIntro project was founded in 2009 to improve the quality and availability of education by producing exceptional books and teaching tools that are free to use and easy to modify.
- Our inaugural effort is *OpenIntro Statistics*. Probability is optional, inference is key, and we feature real data whenever possible.
- Files for the entire book are freely available at openintro.org.

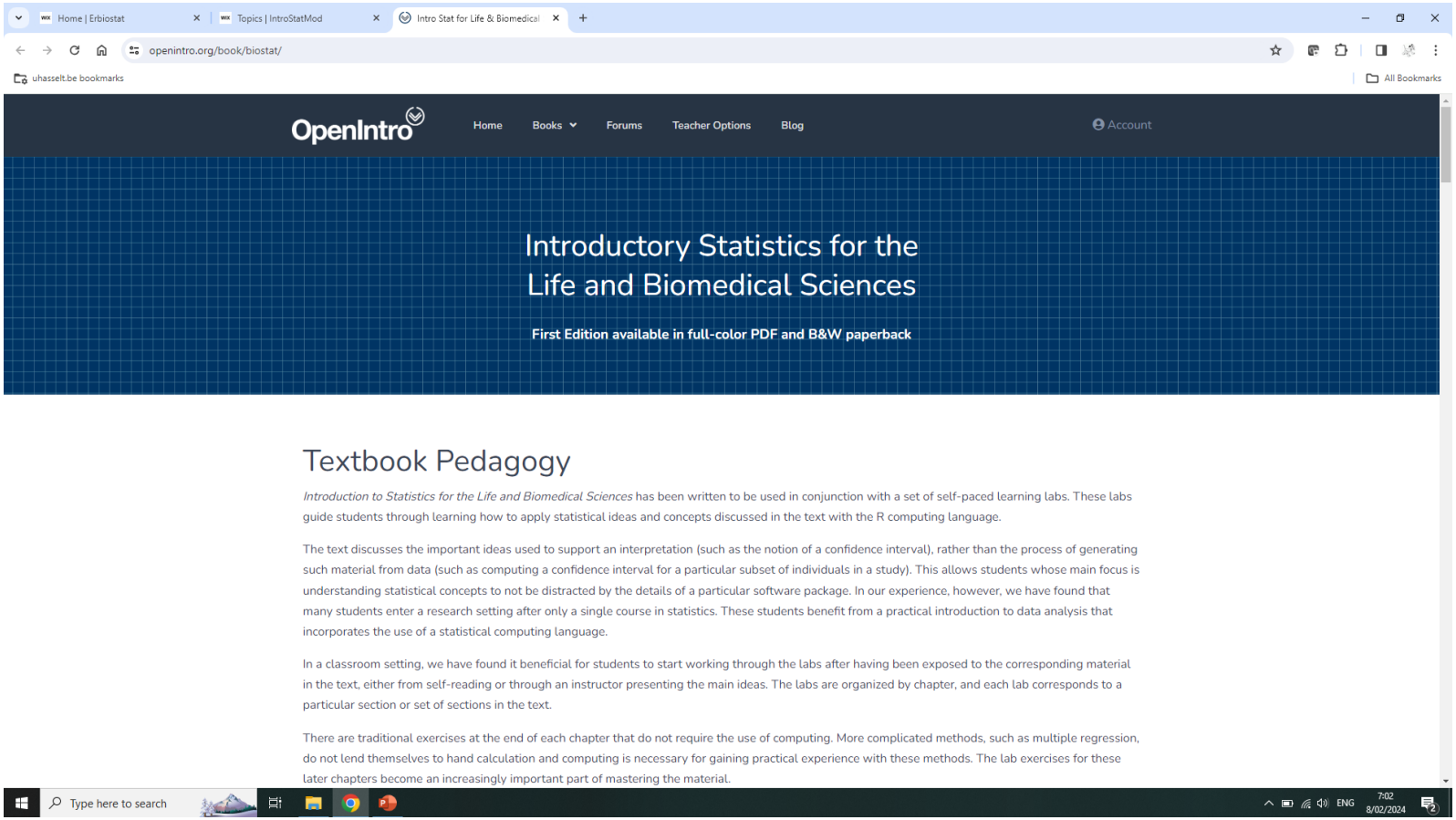
openintro.org

https://www.openintro.org/

“Our mission is to make educational products that are free, transparent, and lower barriers to education. We also feature supporting resources, such as slides, videos, and more.”

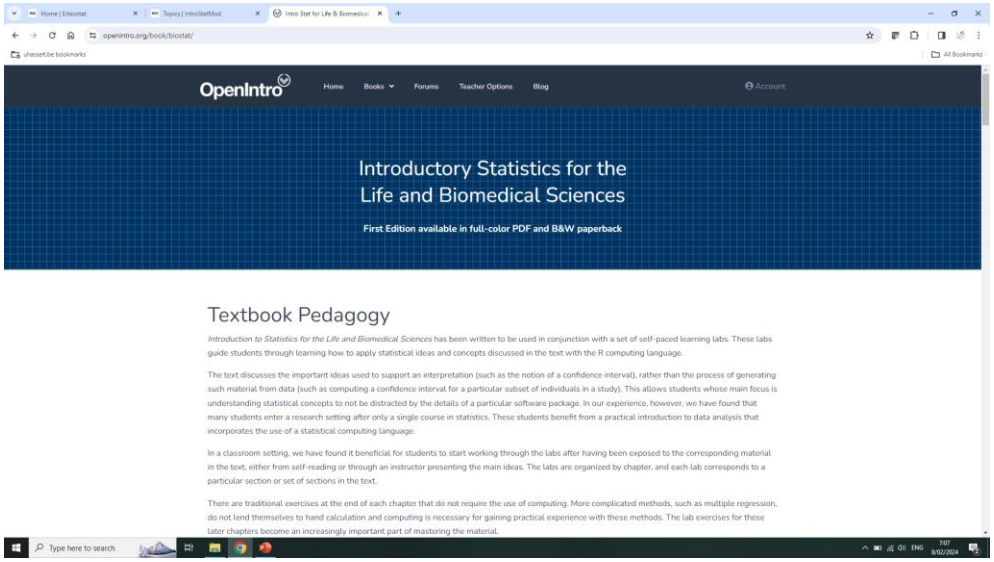


Introductory Statistics for the Life and Biomedical Sciences



https://www.openintro.org/book/biostat/

Introductory Statistics for the Life and Biomedical Sciences



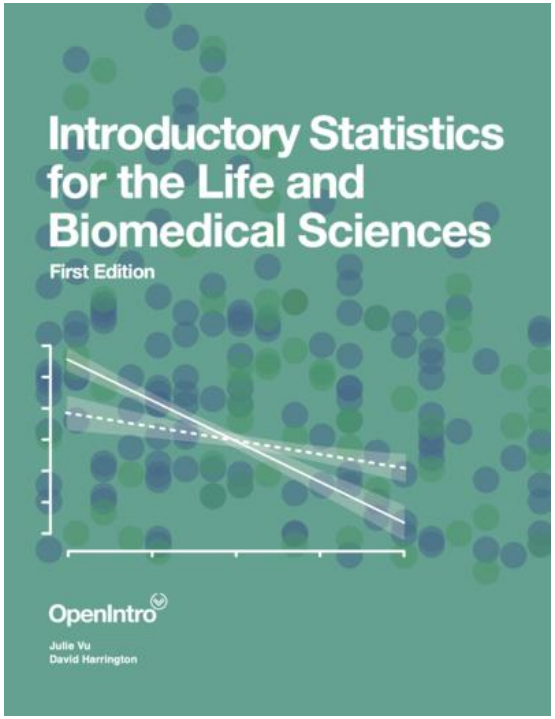
OpenIntro Biostatistics: developing open source materials for teaching and learning applied statistics with R.

Developed by Dave Harrington and Julie Vu.

Planning to work in health or life sciences:

- Motivated more by application than theory.
- No previous statistics courses.
- Familiarity with basic algebra.
- No or limited experience with computing.

Introductory Statistics for the Life and Biomedical Sciences



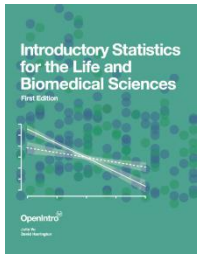
- Written by Dave Harrington and Julie Vu.
- Available on line (for free) on PDF format.
- Focus on theory and practice.
- Undergraduate/graduate levels.
- Available (for free) online.
- Source files for the book in Tex are available.
- Presentations and practical sessions (labs) are available online for free.

Course II

Home | Erbiostat | Topics | IntroStatMod

erbiostat.wixsite.com/introstatmod/topics

uhasselt.be bookmarks | All Bookmarks



Online book

online here:

Online book

This part of the course is based on unit 6 in Vu & Harrington course and it covers the following topics:

- Examining scatterplots.
- Least squares regression.
- Interpreting a linear model.
- Statistical inference in regression.

Slides (PDF): simple linear regression

Slides (PP): simple linear regression

Slides (Rmd): simple linear regression

Simple linera regression

This part of the course is based on unit 5 in Vu & Harrington course and it covers the following topics:

- Ideas behind One-Way ANOVA..
- Assumptions for ANOVA.
- Normal probability plots (Q-Q plots).
- Pairwise comparisons.
- ANOVA model in R using the aov () function

Slides (PDF): One-Way ANOVA

Slides (PP): One-Way ANOVA

Slides (Rmd): One-Way ANOVA

This part of the course is based on unit 9 in Vu & Harrington course and it covers the following topics:

- Odds and probabilities.
- Introduction to logistic regression.
- Simple logistic regression.
- Logistic versus linear regression.
- Inference for simple logistic regression.

Slides (PDF): Logistic regression

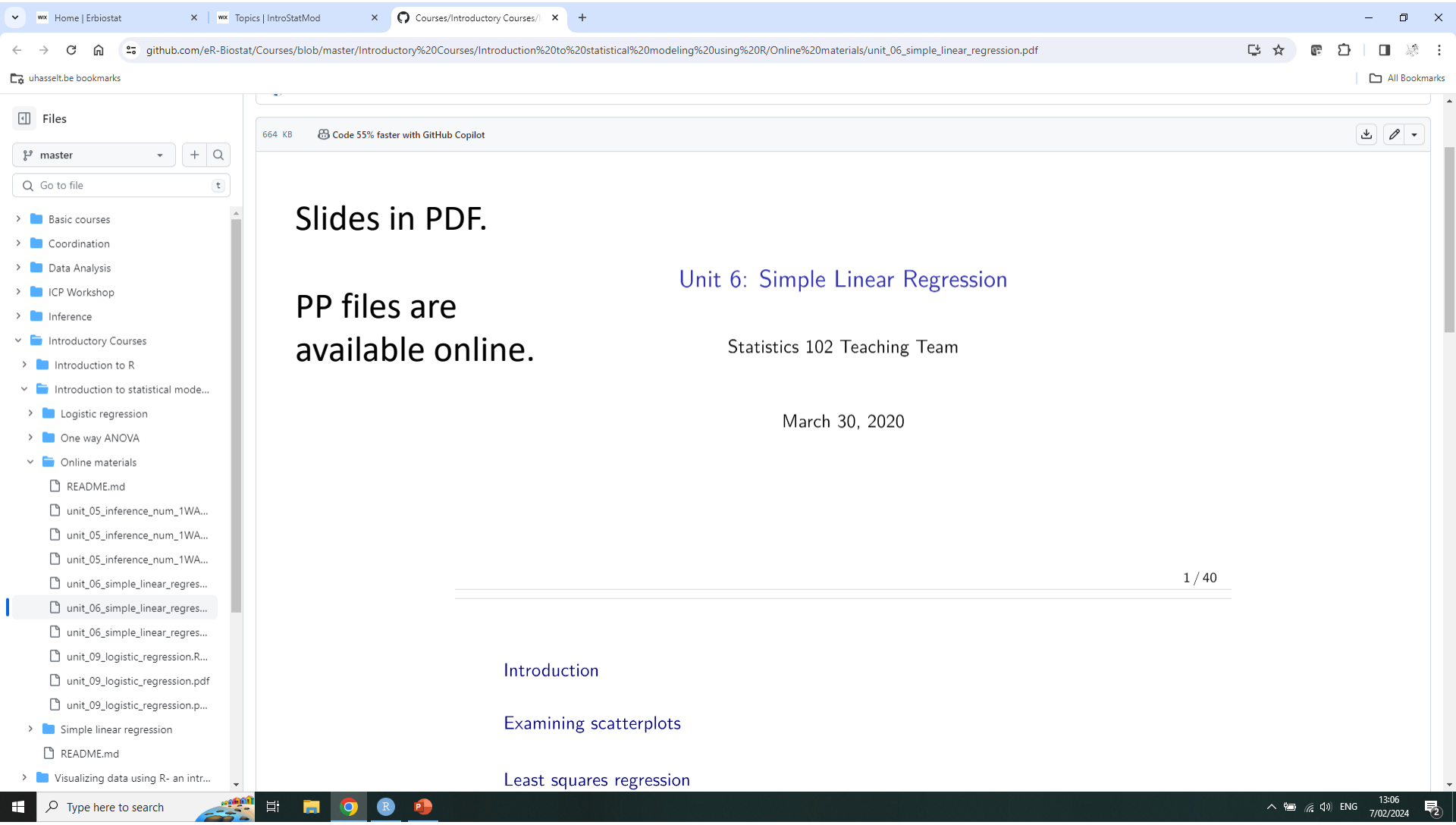
Slides (PP): Logistic regression

Slides (Rmd): Logistic regression

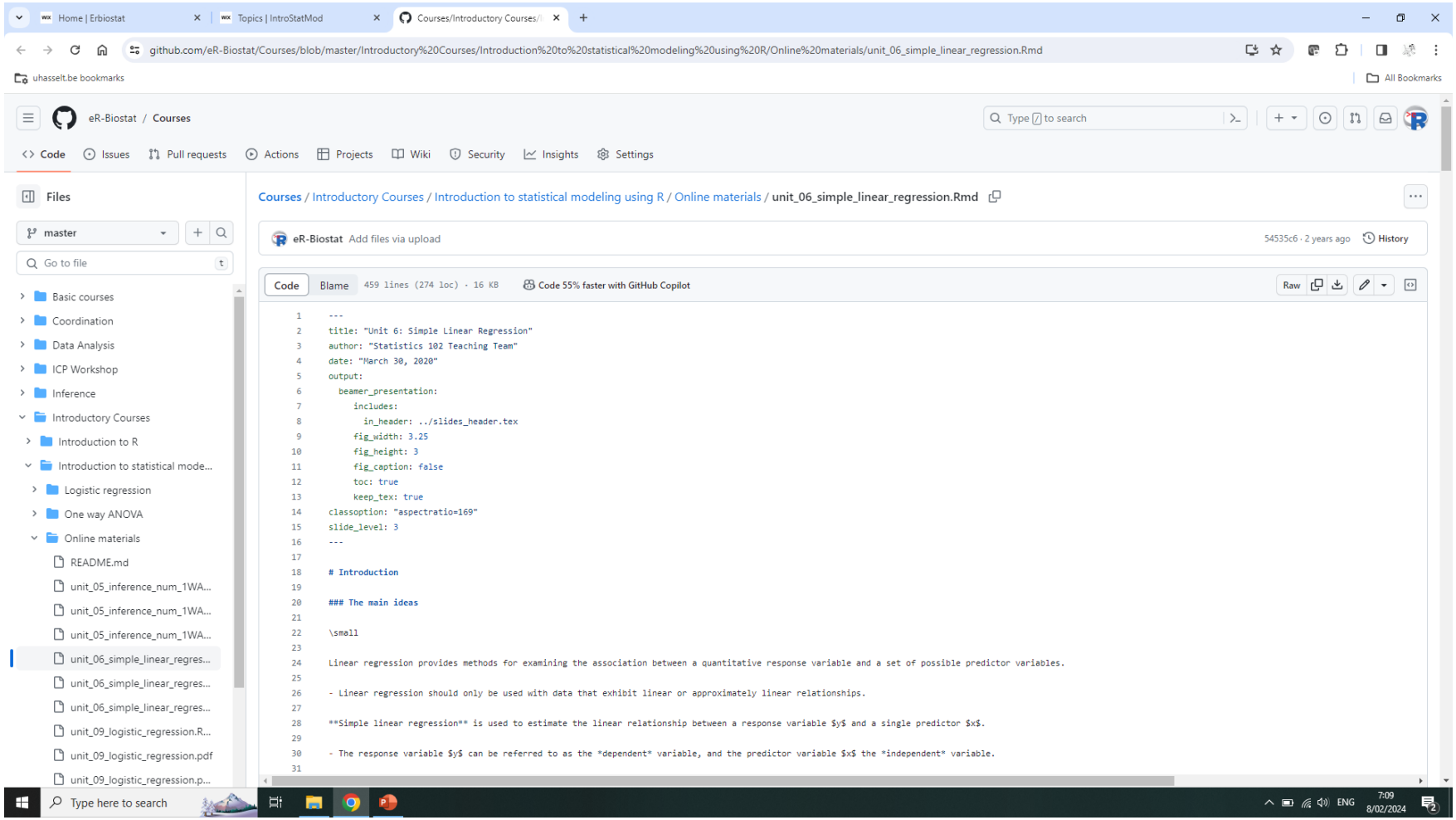
Type here to search

ENG 13:02 7/02/2024

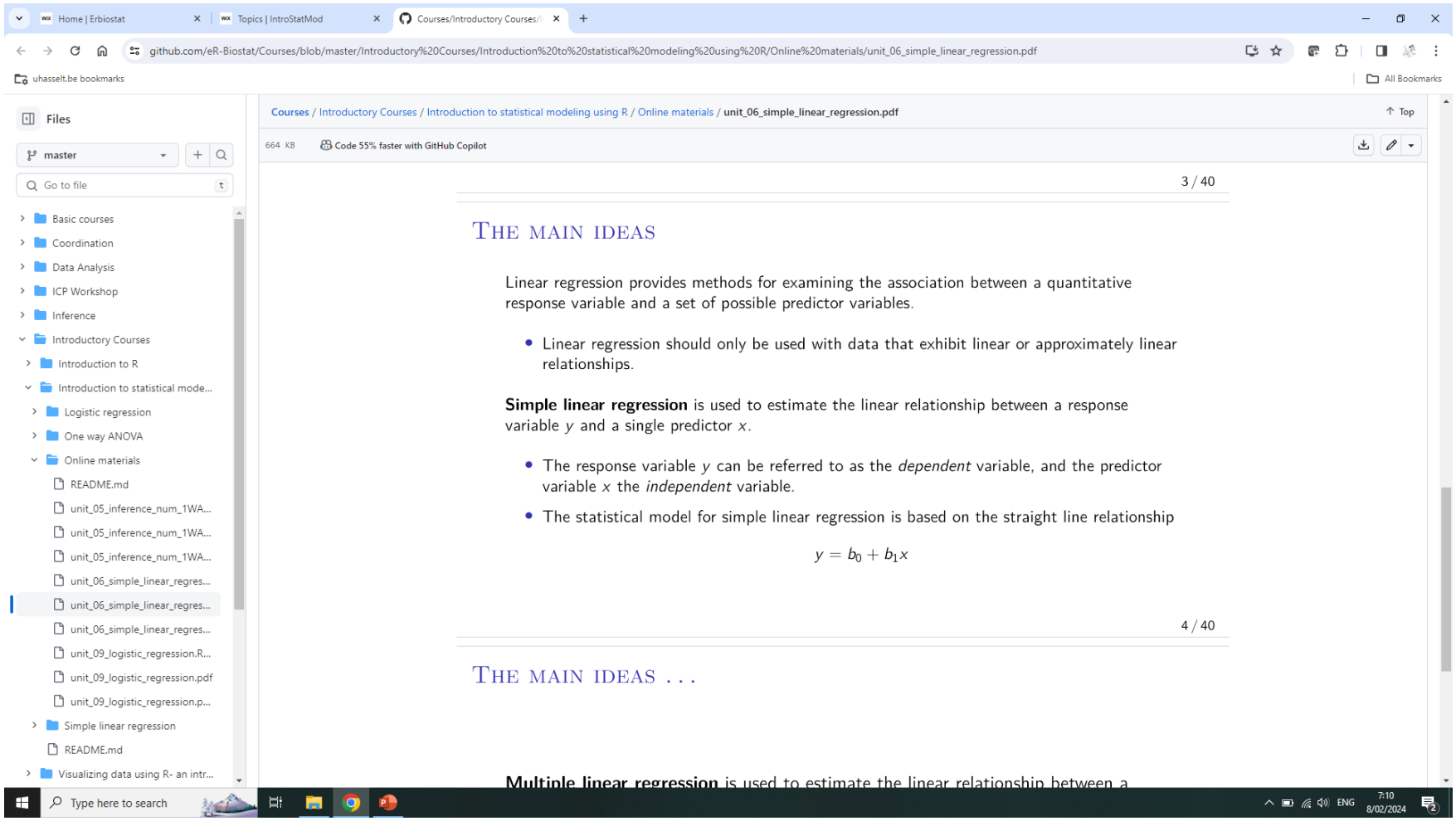
Slides


Home | Erbiostat
Topics | IntroStatMod
Courses/Introductory Courses/
github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Online%20materials/unit_06_simple_linear_regression.pdf
uhasselt.be bookmarks
All Bookmarks
Files
master
Go to file
Basic courses
Coordination
Data Analysis
ICP Workshop
Inference
Introductory Courses
Introduction to R
Introduction to statistical mode...
Logistic regression
One way ANOVA
Online materials
README.md
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_09_logistic_regression.R...
unit_09_logistic_regression.pdf
unit_09_logistic_regression.p...
Simple linear regression
README.md
Visualizing data using R- an intr...
664 KB
Code 55% faster with GitHub Copilot
Slides in PDF.
PP files are available online.
Unit 6: Simple Linear Regression
Statistics 102 Teaching Team
March 30, 2020
1 / 40
Introduction
Examining scatterplots
Least squares regression
Type here to search
ENG
13:06
7/02/2024

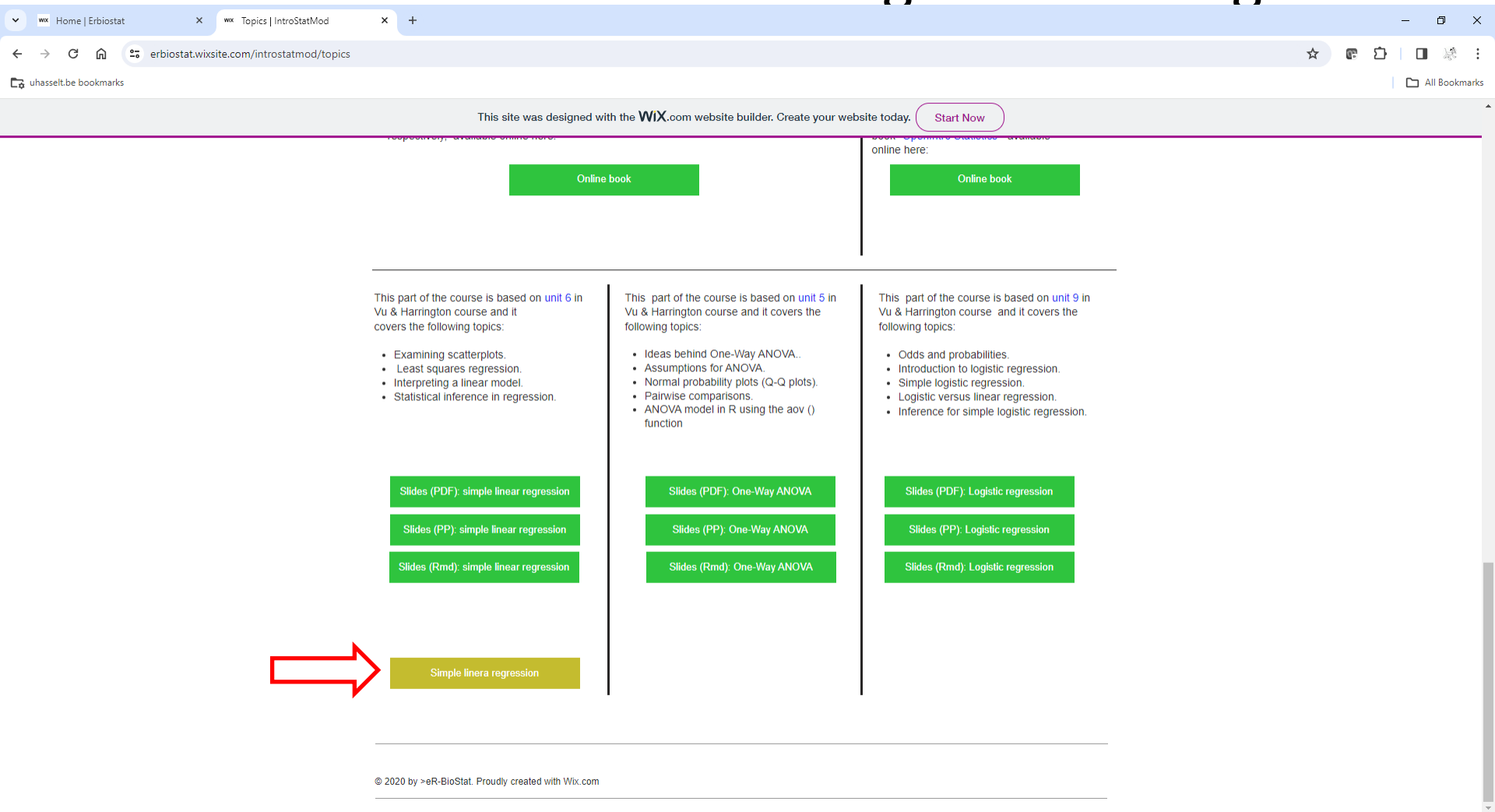
The Rmd file for the slides



A PDF output


Home | Erbiostat
Topics | IntroStatMod
Courses/Introductory Courses/
github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Online%20materials/unit_06_simple_linear_regression.pdf
uhasselt.be bookmarks
All Bookmarks
Files
master
Go to file
Basic courses
Coordination
Data Analysis
ICP Workshop
Inference
Introductory Courses
Introduction to R
Introduction to statistical mode...
Logistic regression
One way ANOVA
Online materials
README.md
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_09_logistic_regression.R...
unit_09_logistic_regression.pdf
unit_09_logistic_regression.p...
Simple linear regression
README.md
Visualizing data using R- an intr...
Courses / Introductory Courses / Introduction to statistical modeling using R / Online materials / unit_06_simple_linear_regression.pdf
Top
664 KB
Code 55% faster with GitHub Copilot
3 / 40
THE MAIN IDEAS
Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables.
Linear regression should only be used with data that exhibit linear or approximately linear relationships.
Simple linear regression is used to estimate the linear relationship between a response variable y and a single predictor x.
The response variable y can be referred to as the dependent variable, and the predictor variable x the independent variable.
The statistical model for simple linear regression is based on the straight line relationship
$y = b_0 + b_1 x$
4 / 40
THE MAIN IDEAS . . .
Multiple linear regression is used to estimate the linear relationship between a
Type here to search
ENG
7:10
8/02/2024

Course I + II:
online text about linear regression using R



A part of the website of the course

Home | Erbiostat | Topics | IntroStatMod | Temp_2024_prog1_V2.knit

htmlpreview.github.io/?https://github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Simple%20linear%20regression/Slides/Temp_2024_prog1_V2.html

uhasselt.be bookmarks | All Bookmarks

Code

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

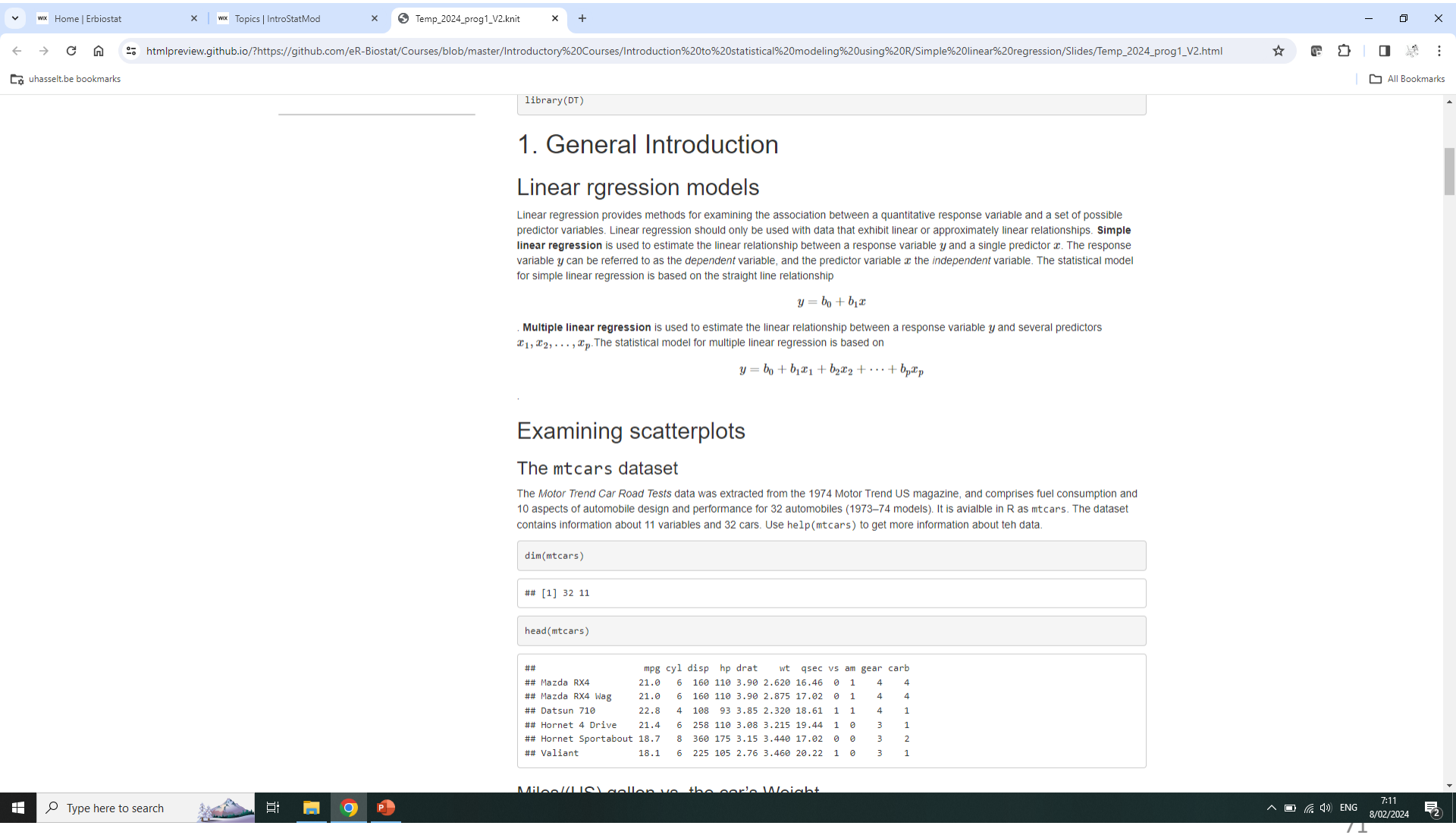
```
## load/install libraries
.libPaths(c("./Rpackages",.libPaths()))
library(knitr)
library(tidyverse)
library(deSolve)
library(minpack.lm)
library(ggpubr)
library(readxl)
library(gamlss)
library(data.table)
library(grid)
library(png)
library(nlme)
library(gridExtra)
library(mvtnorm)
library(e1071)
library(lattice)
library(ggplot2)
library(dslabs)
library(NHANES)
library(plyr)
library(dplyr)
library(nasaweather)
library(ggplot2)
library(gganimate)
library(av)
library(gifski)
library(foreach)
library("DAAG")
library(DT)
```

1. General Introduction

Linear regression models

Type here to search | ENG 7:11 8/02/2024

A part of the website of the course



UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

vliruos SHARING MINDS, CHANGING LIVES

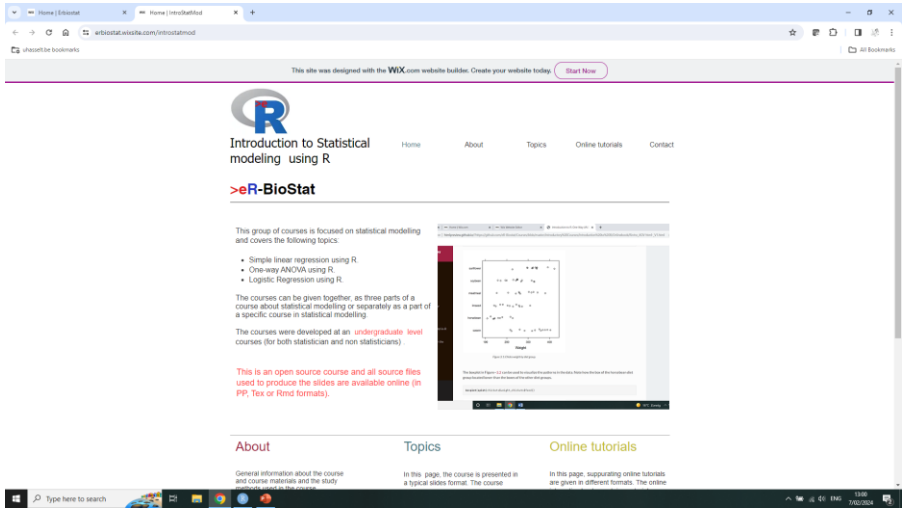
Development of E-learning materials using R markdown

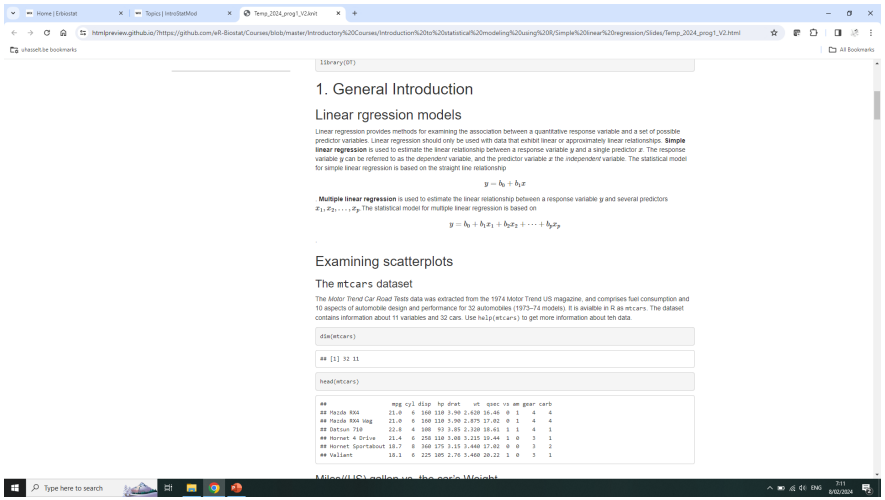
Part 5a: developing the content

R Program: er_prog4_SA_2024.Rmd

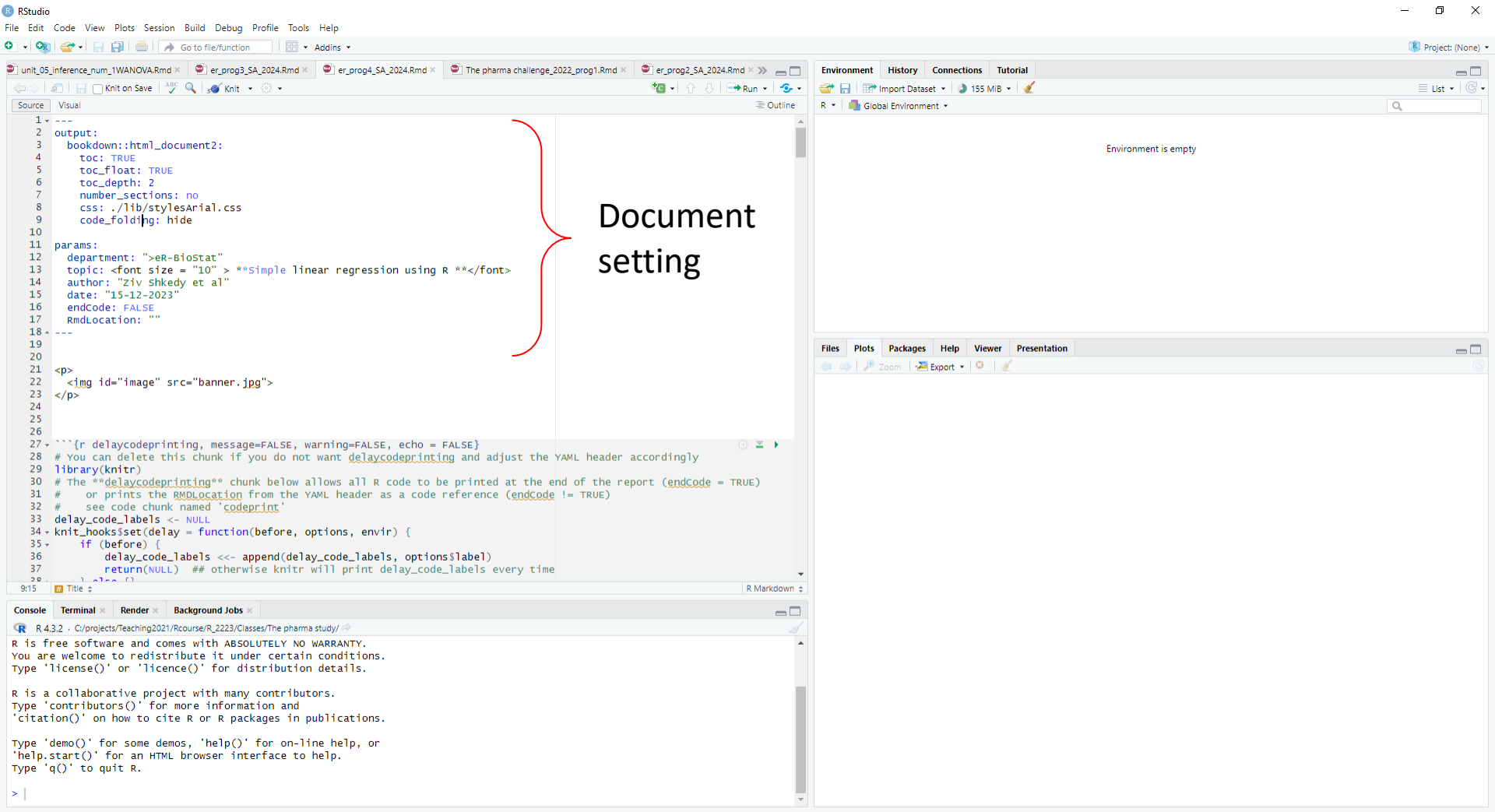
Content on the website

- The Rmd file for the online book: the program and output.





The Rmd program



The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html

er_prog4_SA_2024.html | Open in Browser | Find | Publish

Code

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

Show

1. General Introduction

Linear rgression models

Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables. Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable y and a single predictor x. The response variable y can be referred to as the *dependent* variable, and the predictor variable x the *independent* variable. The statistical model for simple linear regression is based on the straight line relationship

$$y = b_0 + b_1 x$$

. **Multiple linear regression** is used to estimate the linear relationship between a response variable y and several predictors $x_1, x_2, \ldots, x_p$.The statistical model for multiple linear regression is based on

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

.

Examining scatterplots

The `mtcars` dataset

The *Motor Trend Car Road Tests* 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). It is avialble in R as `mtcars`. The dataset contains information about 11 variables and 32 cars. Use `help(mtcars)` to get more information about teh data.

Show

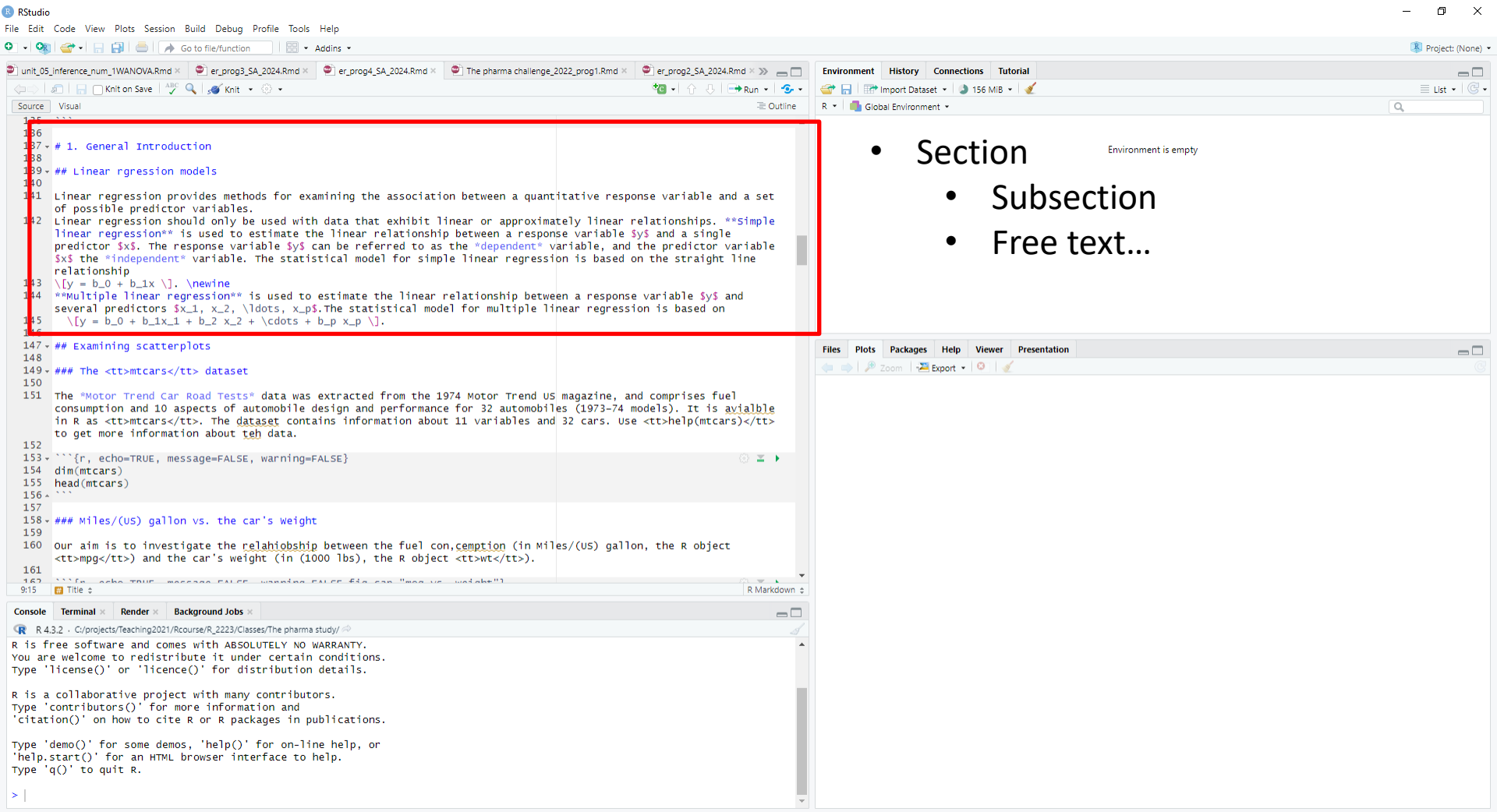
```
## [1] 32 11
```

Show

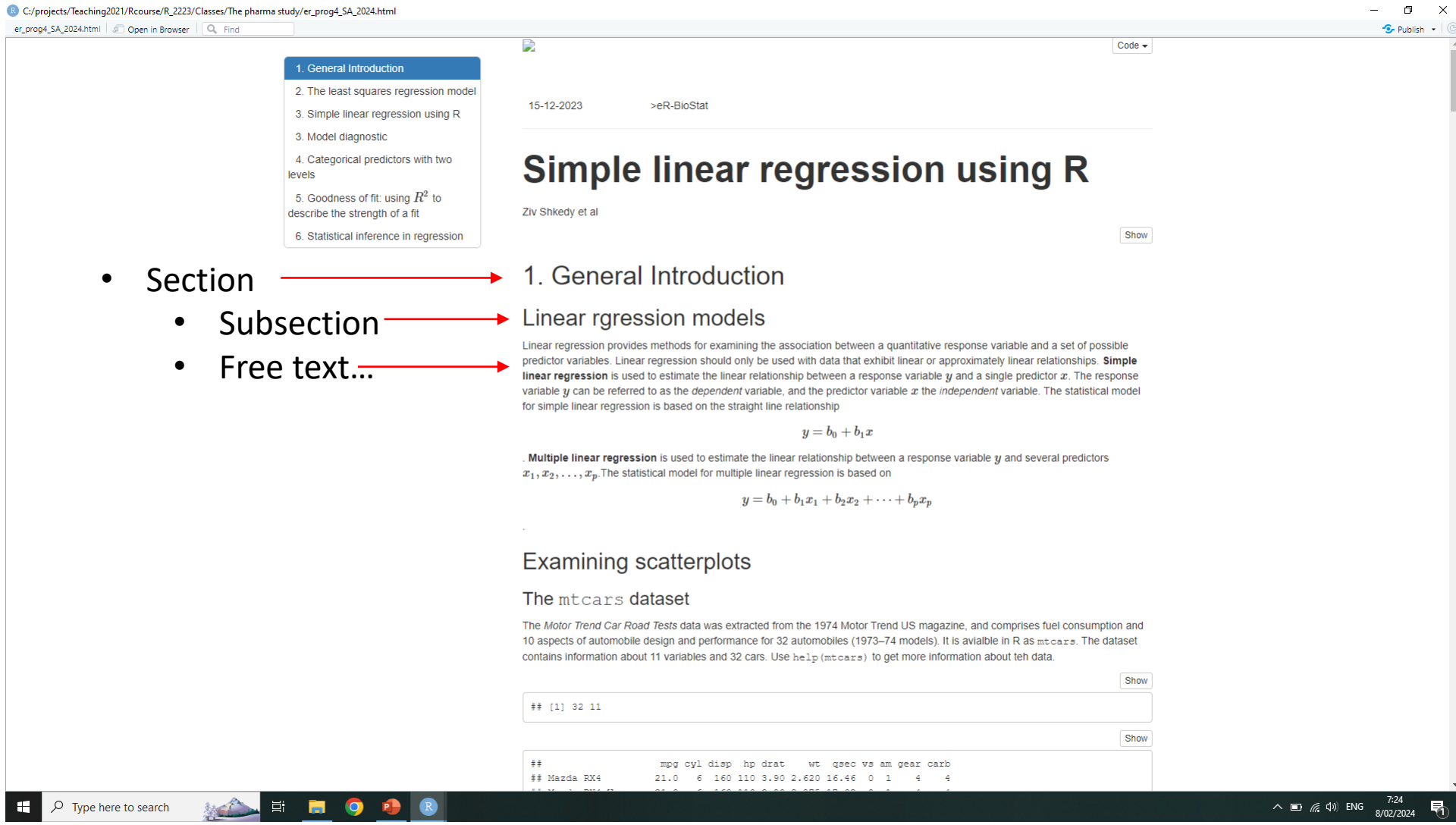
```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
```

Type here to search | ENG 7:24 8/02/2024

The Rmd program

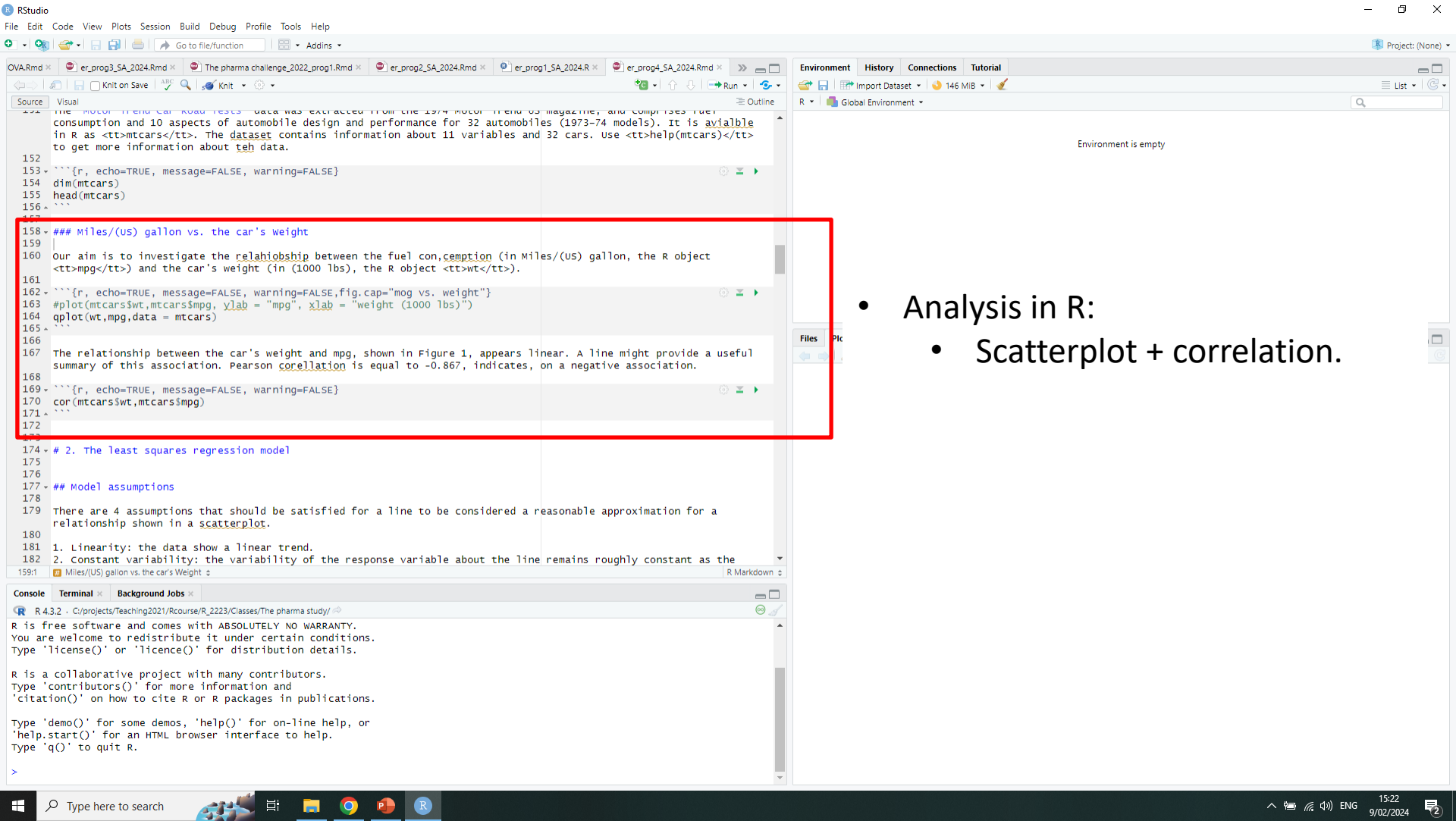


The HTML output



- Section
  - Subsection
  - Free text…

The Rmd program



- Analysis in R:
  - Scatterplot + correlation.

The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html

er_prog4_SA_2024.html | Open in Browser | Find | Publish


```
## 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
```

Miles/(US) gallon vs. the car's Weight

Our aim is to investigate the relahiobship between the fuel con,cemption (in Miles/(US) gallon, the R object `mpg`) and the car's weight (in (1000 lbs), the R object `wt`).

Show

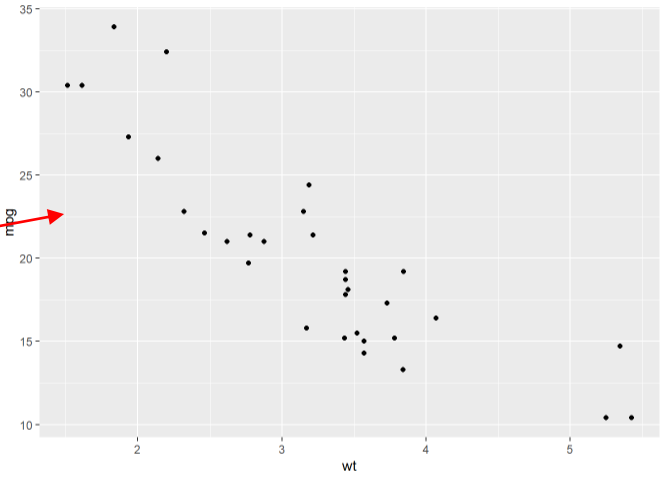


Figure 1: mog vs. weight

The relationship between the car's weight and mpg, shown in Figure 1, appears linear. A line might provide a useful summary of this association. Pearson corellation is equal to -0.867, indicates, on a negative association.

Show

```
## [1] -0.8676594
```

2. The least squares regression model

Model assumptions

There are 4 assumptions that should be satisfied for a line to be considered a reasonable approximation for a relationship shown in a

- Analysis in R:
  - Scatterplot.
  - Correlation.

Text without the code.

Type here to search

ENG 7:26 8/02/2024

The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html

er_prog4_SA_2024.html | Open in Browser | Find | Publish


```
## 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
```

Miles/(US) gallon vs. the car's Weight

Our aim is to investigate the relahiobship between the fuel con,cemption (in Miles/(US) gallon, the R object `mpg`) and the car's weight (in (1000 lbs), the R object `wt`).

Hide

```
#plot(mtcars$wt,mtcars$mpg, ylab = "mpg", xlab = "weight (1000 lbs)")
qplot(wt,mpg,data = mtcars)
```

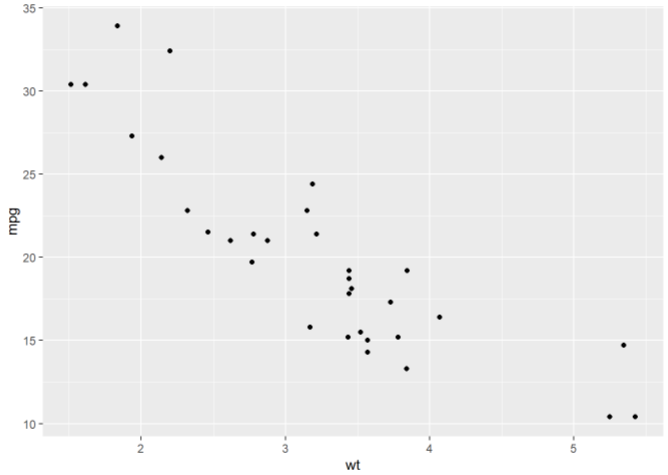


Figure 1: mog vs. weight

The relationship between the car's weight and mpg, shown in Figure 1, appears linear. A line might provide a useful summary of this association. Pearson corellation is equal to -0.867, indicates, on a negative association.

Show

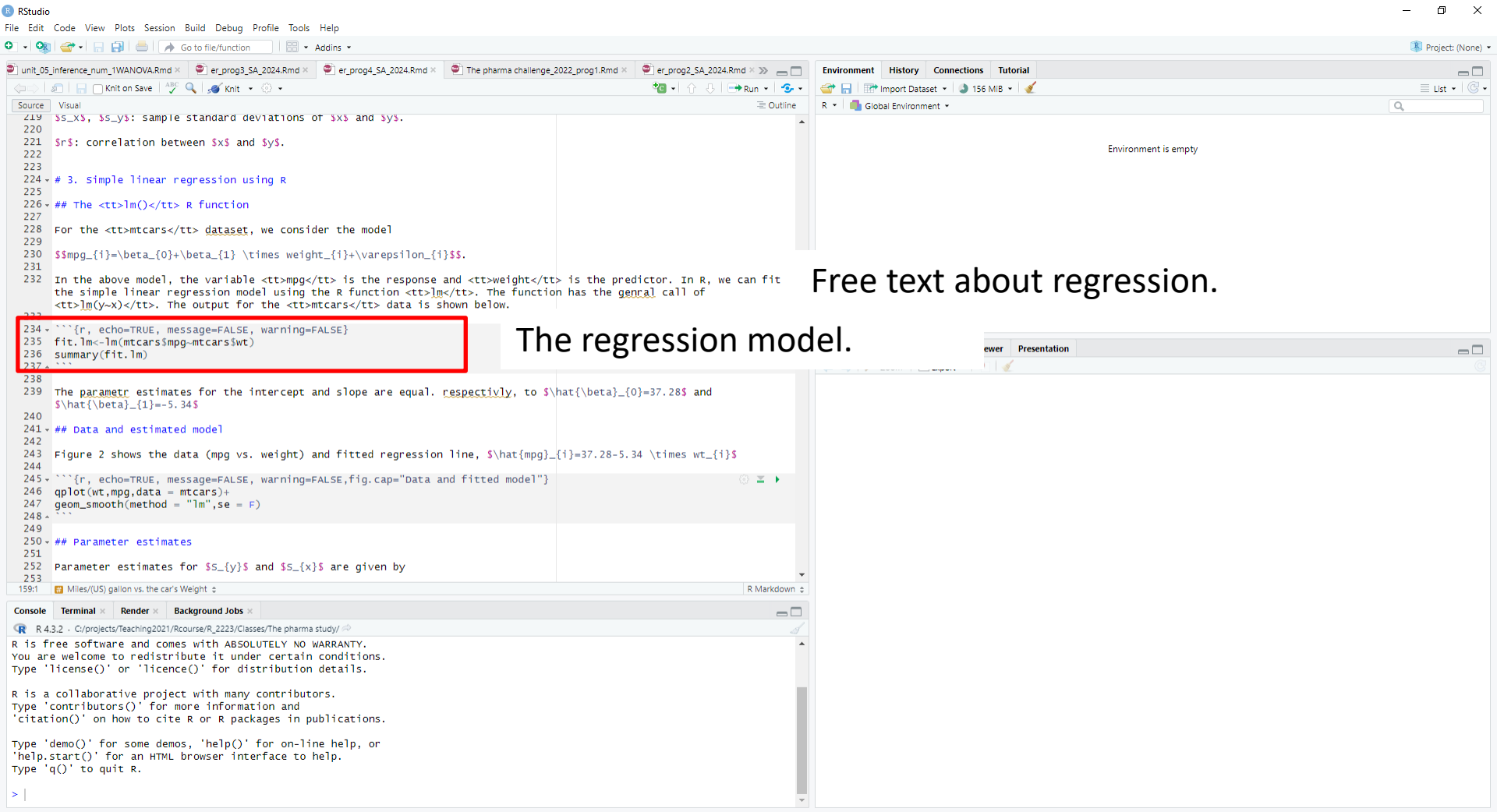
```
## [1] -0.8676594
```

2. The least squares regression model

Text with the code.

Type here to search | ENG 7:26 8/02/2024

The Rmd program



Free text about regression.

The regression model.

The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html

er_prog4_SA_2024.html | Open in Browser | Find | Publish

r: correlation between x and y.

3. Simple linear regression using R

The `lm()` R function

For the `mtcars` dataset, we consider the model

$$mpg_i = \beta_0 + \beta_1 \times weight_i + \varepsilon_i$$

In the above model, the variable `mpg` is the response and `weight` is the predictor. In R, we can fit the simple linear regression model using the R function `lm`. The function has the genral call of `lm(y~x)`. The output for the `mtcars` data is shown below.

Hide

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

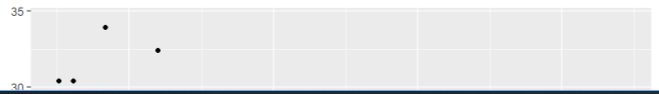
```
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$wt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## mtcars$wt    -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 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
```

The parametr estimates for the intercept and slope are equal. respectivly, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\widehat{mpg}_i = 37.28 - 5.34 \times wt_i$

Show



Free text about regression.

- The regression model:
  - Code.
  - Output.

Type here to search | ENG | 7:27 8/02/2024

Short summary

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

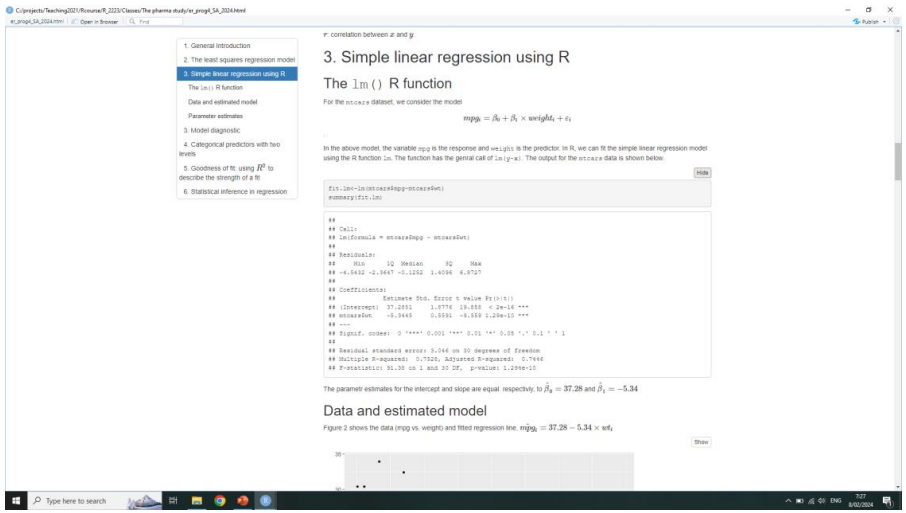
vliruos SHARING MINDS, CHANGING LIVES

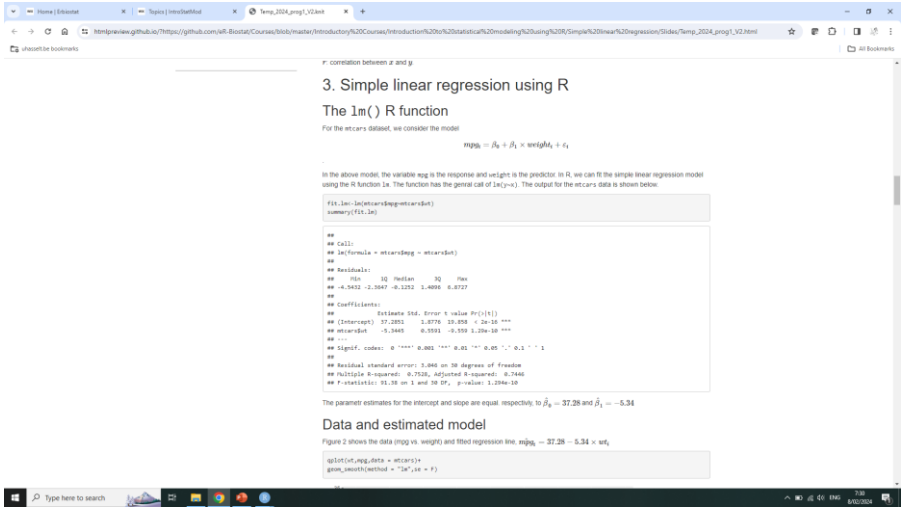
Development of E-learning materials using R markdown

Part 5b: the >eR-BioStat approach

How can we create a course online??

From a laptop to a website..

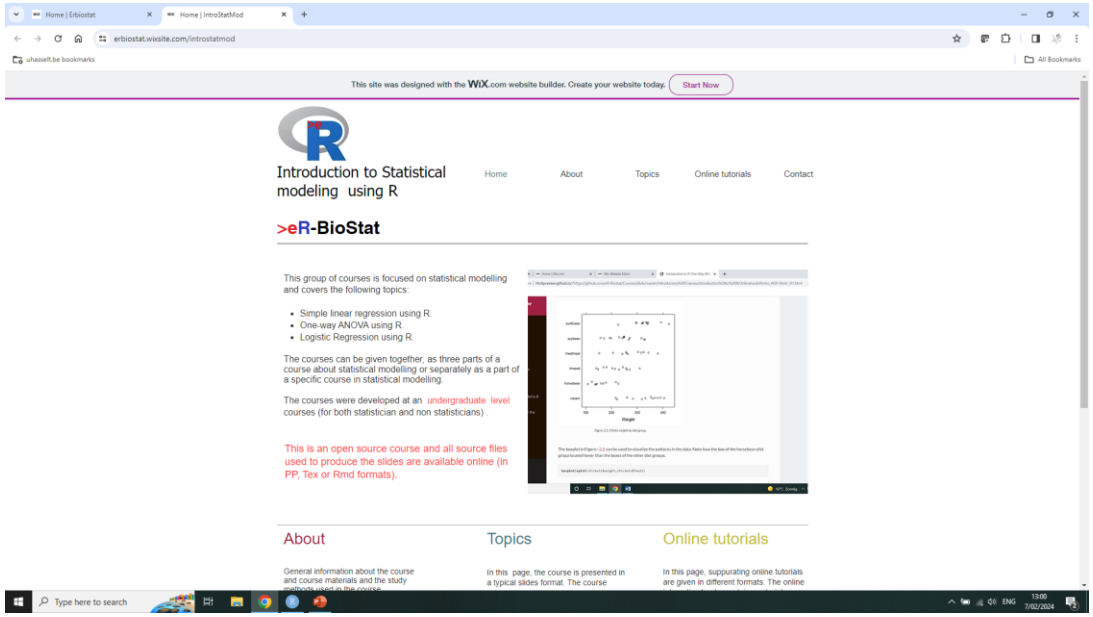




The HTML file on the laptop….

The HTML file as a part of the course's website.

Introduction to statistical modeling using R



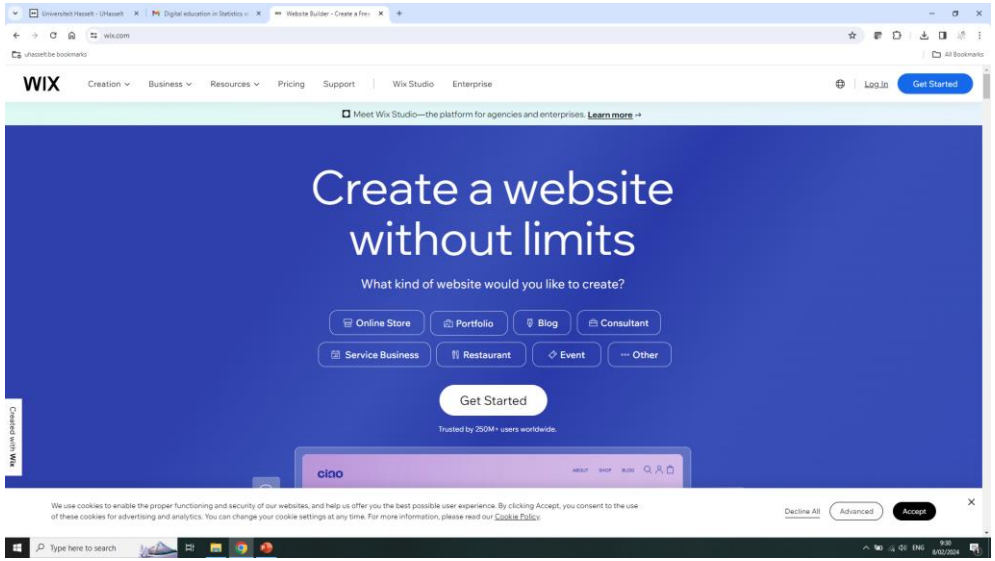
https://erbiostat.wixsite.com/introstatmod

- Website.
- Storage of materials:
  - Slides.
  - Programs.
  - Datasets.
  - Etc.
- Software for the analysis ?
- Storage space & cost ?

- Our approach: bring costs to zero…

How can I develop a website for my course ?

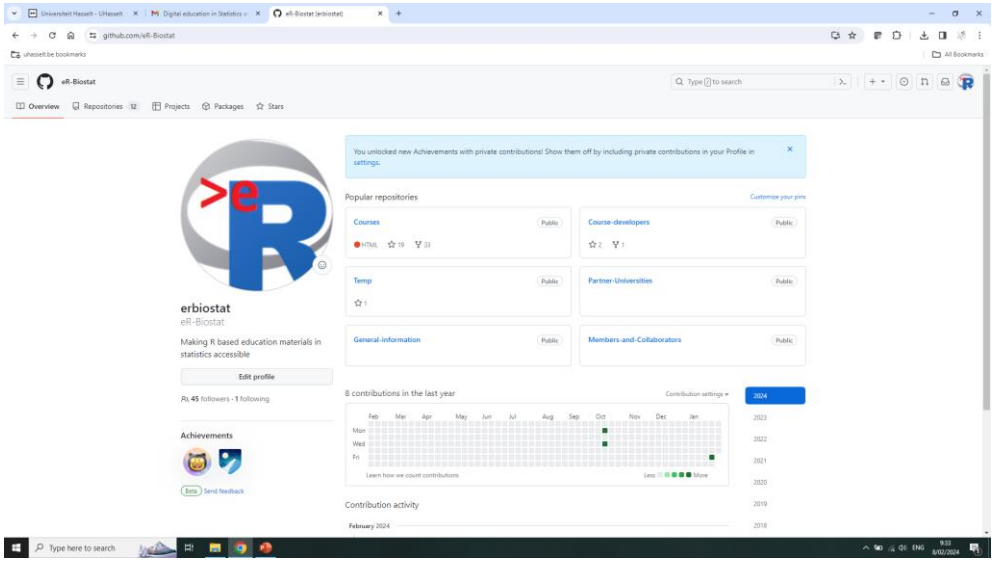
- Websites for the courses were developed using WIX.
- Free.
- Easy to use and learn.



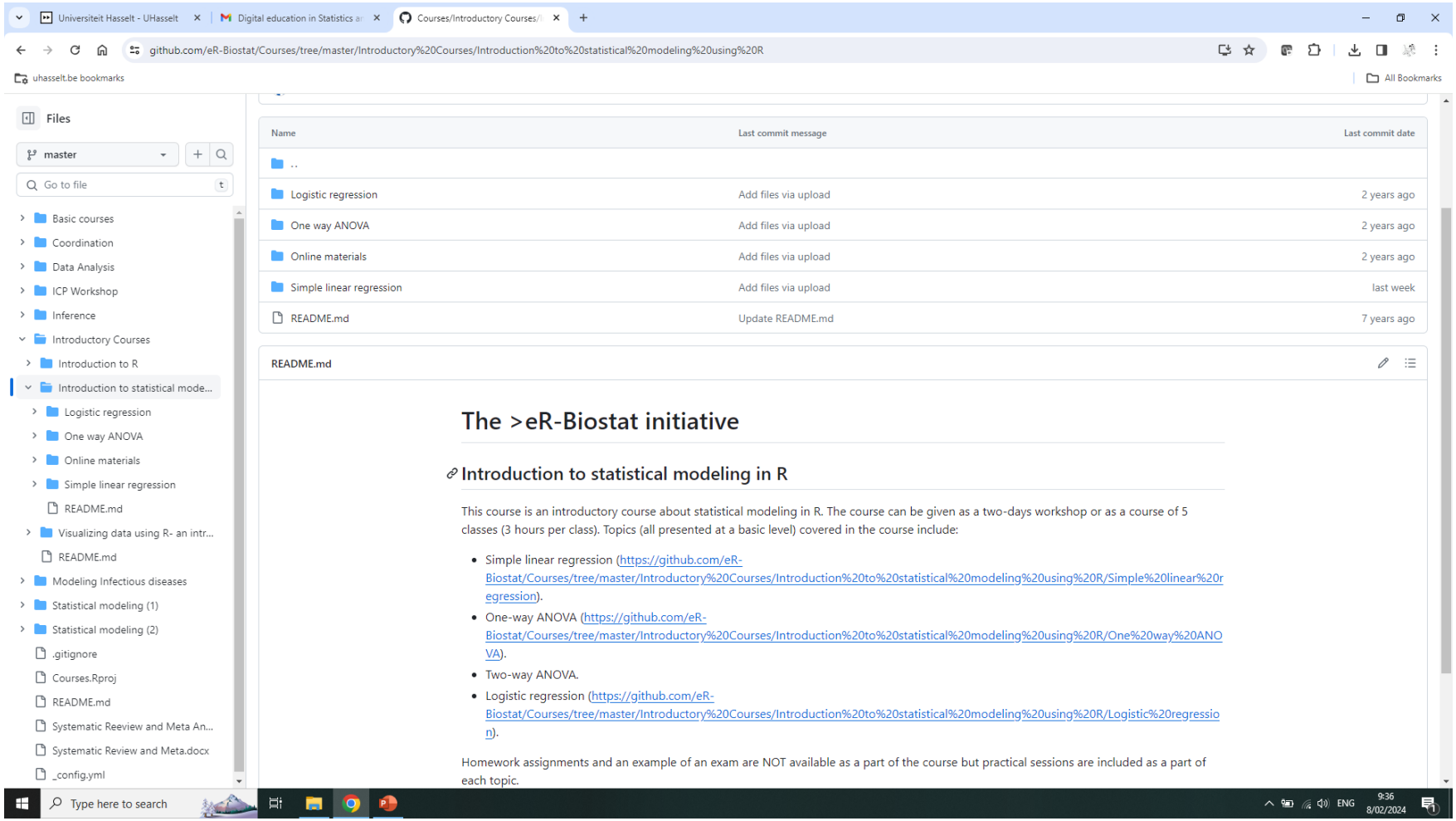
https://www.wix.com/

Where can I store the course materials (slides, programs, notes…..?

- All course materials are stored on Github:
  - Slides.
  - Programs.
  - Datasets.
  - R markdown programs.
- Free and unlimited.



https://github.com/eR-Biostat

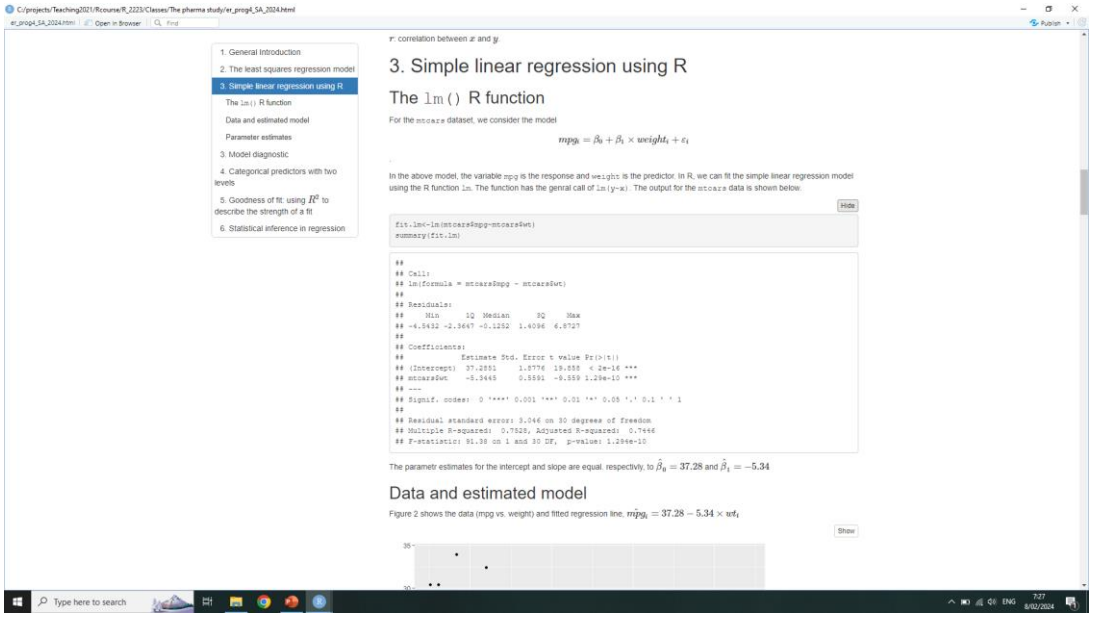


https://github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses

Software for the analysis

- We use R but……

Summary



- Content produced using R markdown.
- Store online on GitHub.
- Presented online in a website developed using WIX.
- Data analysis using R.

- Costs=0 !!!

What is available to the users ?

- Who are the users ? Teachers & students & others.
- What is available ? Everything.
- An open source approach:
  - Slides.
  - R programs for examples.
  - R programs for the slides.
  - PowerPoints files.

Discussion

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