

Data Science R.

© . .

2022-03-08

Contents

Chapter 1

“ ” “ ” 123 ,
 ,
 : ,
 ,
 R,
 CRAN.
 ,
 vignetts
 .
 CRISP DM (?),
 :
 Data Science-
 ,
 vnsidorenko@gmail.com.

Chapter 2

1. № 1. Data Science-

: “ ” *Data Science-* (?)
“ ” (?) , *RMarkdown* ,
LaTeX.

2.1 ?

- / , ,
R IDE RStudio.
- (html, pdf, doc) ,
(, / . .).
- LaTeX , : $y(x) = b_0 + b_1x + b_2x^2$.
- BibTeX, ‘Refer-
ences’, .

2.2

2.2.1 Data Science- ?

- Data Science- (- Data Mining-)
CRISP DM (?). CRISP-DM 1.0 Data
Mining (. 1):

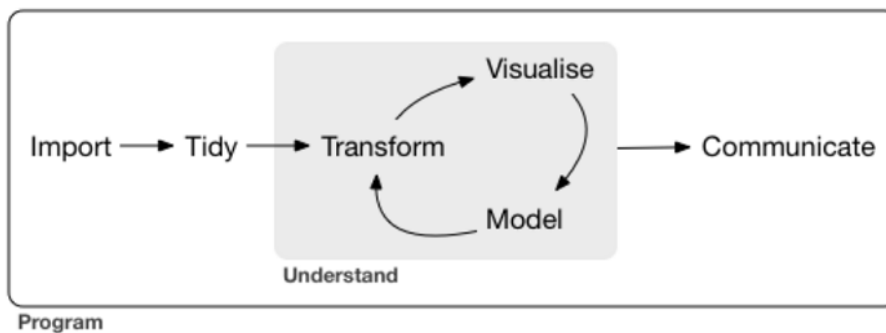


Figure 2.2: . 2. Data Science-

(Import) - (, ,
.)
(Tidy) - “ ” ,
“ - ”, , “ ’ - ”.
(Wrangling, Munging),
(Missing Value Imputation),
(data reduction), (Transforming) . . ,
Tiding+Transforming=Wrangling(Munging).
- - (Transform-Visualise-Model)
Data Mining- , ,
, ,
(Understanding) , ,
- (ommunicate)
, ,
. Data Science-
, , -
, ,
R , IDE RStudio
Data Science- , ,
.

2.2.2

- .doc, pdf, .html .

Data Science,
Data Science.
(Literate Programming) – ,
,
,
(. Markdown, YAML,
HTML, LaTeX),
,
,
(IDE, Integrated Development
Environment,),
(. 3):

- IDE R Studio (?) ;
- R (?) – ;
- RMarkdown (?) – Markdown; (. (?)
RStudio: “*Help\Cheat_Sheets*”);
• R Markdown: ;
- LaTeX .

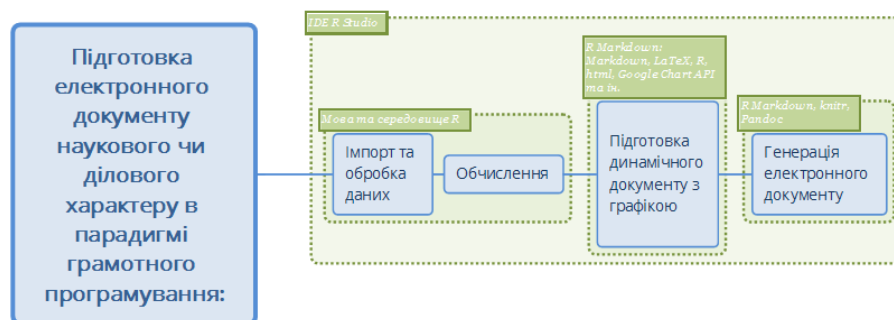


Figure 2.3: . 3.
(*literate programming*)

2.2.3 Markdown RMarkdown

Markdown (: []) – ,
structurally valid XHTML HTML.

, GitHub, Reddit Stack Overflow Markdown

R Markdown (?) – R, Markdown-
IDE RStudio R .

SQL, Stan. Pandoc , R, Python, ++, HTML,
html, doc pdf
- , , , .

2.2.4 R

CRAN, R. R GUI,
, ‘Contributed’, ,

R . ,

2.2.5 RStudio

, RMarkdown ,
IDE RStudio.

2.2.6 RMarkdown-

1. RStudio.
2. RMarkdown- R Notebook, .

2.2.7

Ctrl+Shift+K.

2.3 Markdown-

2.3.1

$$y(x) = b_0x + b_1 + b_2x^2 \quad x \in [x_1; x_2].$$

2.3.2

1. R Markdown, *RMarkdown*.
2. YAML-, (. 4).

```

---
title: "Модуль 1. Базовий. Лабораторна робота №1. Створення основи типового Data Science-проекту"
author: "&copy; [Сидоренко В. М.](https://www.linkedin.com/in/valeriy-sydorenko-6782279a/), 'r format(Sys.time()),
'%"
date: "'r Sys.Date()"
output:
  # pdf_document:
  #   highlight: tango
  #   toc: yes
  # word_document:
  #   highlight: tango
  #   toc: yes
  html_notebook:
    toc: yes # генерація змісту документа
    toc_float: true
    highlight: tango # Колір підсвічування коду
fontsize: 12pt # розмір шрифту
header-includes:
  \usepackage[T2A]{fontenc}
  \usepackage[utf8]{inputenc}
  \usepackage[russian]{babel}
editor_options:
  chunk_output_type: console # вивід результатів обчислень на консоль
bibliography: references_lab.bib # ім'я файлу з БД бібліографічних посилань
---
```

Figure 2.4: . 4. YAML-,

3. LaTeX *RMarkdown*-
\$:

$$y(x)=b_0x+b_1+b_2x^2$$

3. R , @ref(fig:fig_1):

```

#
b0 <- 2
b1 <- 3
b2 <- 1.57

#

x <- seq(-1, 1, .1)
y <- b0 + b1 * x + b2 * x^2

plot(x, y,
      type = "l",
      col = "red",
      main = " ",
      xlab = "x",
      ylab = "y"
)
```