Note with R4DS

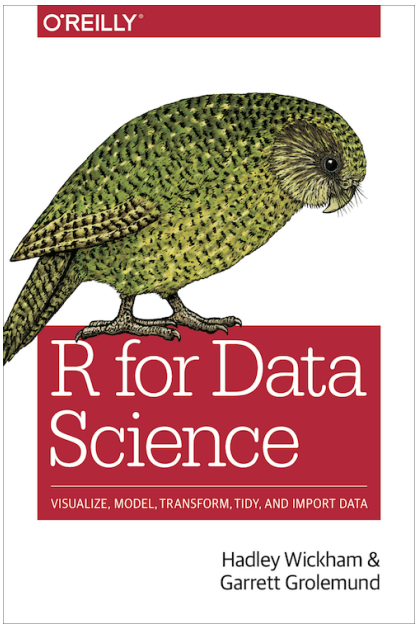
2019-05-17

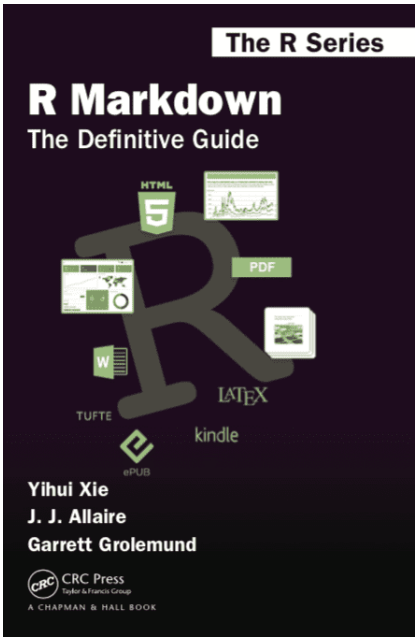
Contents

About this notebook

This notebook is my practice after reading those books:

- **R for Data Science**
- **R Markdown: The Definitive Guide**
- **bookdown: Authoring Books and Technical Documents with R Markdown**





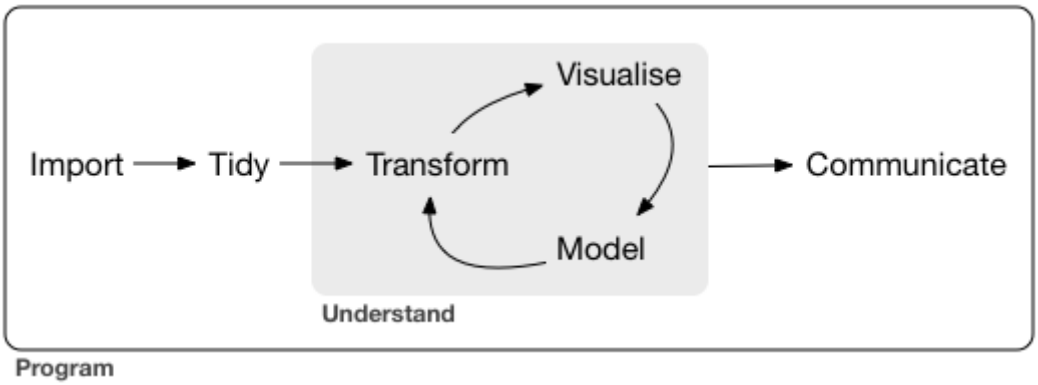
Chapter 1

Introduction

Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge. The goal of "R for Data Science" is to help you learn the most important tools in R that will allow you to do data science. After reading this book, you'll have the tools to tackle a wide variety of data science challenges, using the best parts of R.

1.1 Data Science Project process:

Data science is a huge field, and there's no way you can master it by reading a single book. The goal of this book is to give you a solid foundation in the most important tools. Our model of the tools needed in a typical data science project looks something like this:



1.2 What you won't learn

- Big Data
- Python, Julia, and friends
- Non-rectangular data
- Hypothesis confirmation

1.3 Prerequisites

1.3.1 R

- Download R from CRAN: https://cran.r-project.org
- Cloud mirror: https://cloud.r-project.org (which automatically figures it out for you.)

1.3.2 RStudio

- Download and install it from http://www.rstudio.com/download
- RStudio IDE Cheat Sheet: https://www.rstudio.com/resources/cheatsheets/#ide

1.3.3 The tidyverse packages

Install the tidyverse packages:

```
if (!require("tidyverse")) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.1     v purrr   0.3.2
## v tibble  2.1.1     v dplyr   0.8.1
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Load it with the library() function:

```
library(tidyverse)
```

Update the packages:

```
tidyverse_update()
```

1.3.4 Other packages

In this book we'll use three data packages from outside the tidyverse:

```
install.packages(c("nycflights13", "gapminder", "Lahman"))
```

Chapter 2

Introduction

Chapter 3

Data visualisation