

# Data Science Boot Camp

Databrew

2021-03-25



# Contents

<b>1</b>	<b>Welcome</b>	<b>7</b>
<b>2</b>	<b>Introduction</b>	<b>9</b>
<b>3</b>	<b>Core theory</b>	<b>13</b>
3.1	Principles of data science . . . . .	13
3.2	Data, ethics, and citizenship . . . . .	13
3.3	Principles of data visualization . . . . .	13
3.4	Principles of data writing . . . . .	14
<b>4</b>	<b>Getting started in R</b>	<b>15</b>
4.1	Downloading R and RStudio . . . . .	15
4.2	Running R code: basic math & operators . . . . .	15
4.3	RStudio: Tour, workflow, & scripts . . . . .	15
4.4	Objects: variables and vectors . . . . .	15
4.5	Calling functions . . . . .	15
4.6	Base plots . . . . .	15
4.7	Packages . . . . .	15
4.8	ggplot . . . . .	15
<b>5</b>	<b>Working with data in R</b>	<b>17</b>
5.1	Project workflows . . . . .	17
5.2	Importing data . . . . .	17
5.3	Dataframes: exploration & summarization . . . . .	17

5.4	Data wrangling . . . . .	17
5.5	Exploratory Data Analysis . . . . .	17
5.6	Significance statistics . . . . .	18
5.7	Customizing plots . . . . .	18
5.8	Managing project files . . . . .	18
5.9	Formatting your own data . . . . .	18
<b>6</b>	<b>Your core R toolkit</b>	<b>19</b>
6.1	Uniting and joining data . . . . .	19
6.2	For loops . . . . .	19
6.3	Writing functions . . . . .	19
6.4	Working with text . . . . .	19
6.5	Cleaning messy data . . . . .	19
6.6	Working with factors . . . . .	19
6.7	Matrices & lists . . . . .	19
6.8	Pipes . . . . .	19
6.9	Exporting data & plots . . . . .	19
6.10	Web-based data . . . . .	19
<b>7</b>	<b>Interactive dashboards in R</b>	<b>21</b>
7.1	Shiny dashboards . . . . .	21
7.2	Data entry apps . . . . .	21
<b>8</b>	<b>Databases</b>	<b>23</b>
8.1	Intro: What, why, when (and when not) . . . . .	23
8.2	Platforms: PostgreSQL, MySQL, SQLite, etc. . . . .	23
8.3	Alternatives: NoSQL . . . . .	23
8.4	Practical: spinning up a local DB . . . . .	23
<b>9</b>	<b>Reproducible research &amp; documentation in R</b>	<b>25</b>
9.1	Introduction to Rmarkdown . . . . .	25
9.2	Standards for tables . . . . .	25
9.3	Standards for figures . . . . .	25

<i>CONTENTS</i>	5
<b>10 Version control and teamwork</b>	<b>27</b>
10.1 What is version control? . . . . .	27
10.2 Github . . . . .	27
10.3 Standard git operations . . . . .	27
10.4 Other git platforms . . . . .	27
<b>11 Scientific writing techniques</b>	<b>29</b>
<b>12 Advanced skills</b>	<b>31</b>
12.1 Geographic computing & GIS . . . . .	31
12.2 Mapping . . . . .	31
12.3 Statistical modeling . . . . .	31
12.4 Apply family . . . . .	31
12.5 Iterative analyses . . . . .	31



# Chapter 1

## Welcome

Welcome to the Data Science Boot Camp.

Number	Name
01	intro
02	core theory
03	getting started in R
04	data in r
05	r toolkit
06	dashboards
07	databases
08	documentation
09	version control
10	writing
11	advanced skills
12	references





## Chapter 2

# Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

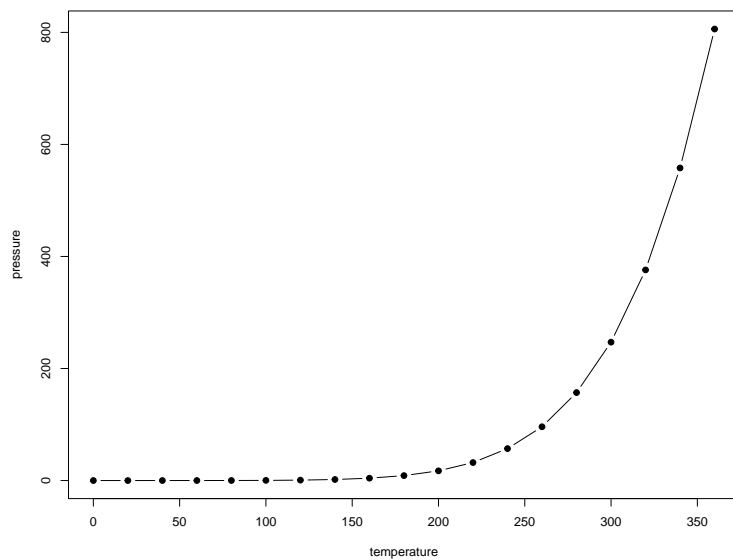


Figure 2.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).



# Chapter 3

## Core theory

### 3.1 Principles of data science

#### 3.1.1 What is data science?

#### 3.1.2 What is the data life cycle?

#### 3.1.3 What is a pipeline?

#### 3.1.4 Data science ‘in the wild’

#### 3.1.5 The reproducibility crisis

### 3.2 Data, ethics, and citizenship

### 3.3 Principles of data visualization

#### 3.3.1 Bad examples

#### 3.3.2 Good examples

#### 3.3.3 Edward Tufte

#### 3.3.4 Grammar of graphics

#### 3.3.5 Design principles

### 3.3.6 Plots & power: the politics of plotting

## 3.4 Principles of data writing

## Chapter 4

# Getting started in R

4.1 Downloading R and RStudio

4.2 Running R code: basic math & operators

4.3 RStudio: Tour, workflow, & scripts

4.4 Objects: variables and vectors

4.5 Calling functions

4.6 Base plots

4.7 Packages

4.8 ggplot





## Chapter 5

# Working with data in R

### 5.1 Project workflows

### 5.2 Importing data

#### 5.2.1 Working directories

#### 5.2.2 Reading in data

### 5.3 Dataframes: exploration & summarization

### 5.4 Data wrangling

#### 5.4.1 Data transformation, filtering, & grouping

#### 5.4.2 The tidyverse and tibbles

#### 5.4.3 Transformation with dplyr

### 5.5 Exploratory Data Analysis

#### 5.5.1 Exploring distributions

#### 5.5.2 Variable types & statistics

### 5.5.3 Descriptive statistics

## 5.6 Significance statistics

### 5.6.1 Comparison tests

### 5.6.2 Correlation tests

## 5.7 Customizing plots

### 5.7.1 Base plots

### 5.7.2 ggplot

## 5.8 Managing project files

## 5.9 Formatting your own data

## Chapter 6

# Your core R toolkit

6.1 Uniting and joining data

6.2 For loops

6.3 Writing functions

6.4 Working with text

6.5 Cleaning messy data

6.6 Working with factors

6.7 Matrices & lists

6.8 Pipes

6.9 Exporting data & plots

6.10 Web-based data



## Chapter 7

# Interactive dashboards in R

### 7.1 Shiny dashboards

### 7.2 Data entry apps



## Chapter 8

# Databases

- 8.1 Intro: What, why, when (and when not)
- 8.2 Platforms: PostgreSQL, MySQL, SQLite, etc.
- 8.3 Alternatives: NoSQL
- 8.4 Practical: spinning up a local DB





## Chapter 9

# Reproducible research & documentation in R

### 9.1 Introduction to Rmarkdown

### 9.2 Standards for tables

### 9.3 Standards for figures



## Chapter 10

# Version control and teamwork

10.1 What is version control?

10.2 Github

10.3 Standard git operations

10.4 Other git platforms



## Chapter 11

# Scientific writing techniques



## Chapter 12

# Advanced skills

12.1 Geographic computing & GIS

12.2 Mapping

12.3 Statistical modeling

12.4 Apply family

12.5 Iterative analyses

12.5.1 Randomization

12.5.2 Simulations





# Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.21.