

# An Introduction to Data Science for Sensory and Consumer Scientists

John Ennis, Julien Delarue, and Thierry Worch

2020-11-11



# Contents

<b>Introduction</b>	<b>9</b>
<b>1 Introduction</b>	<b>9</b>
1.1 How should sensory and consumer scientists learn data science? .	9
1.2 Caution: Don't that everybody does . . . . .	9
1.3 Example projects . . . . .	9
<b>2 What is Data Science?</b>	<b>11</b>
2.1 History . . . . .	11
2.2 Workflow . . . . .	11
2.3 Benefits of data science . . . . .	11
2.4 How to learn data science . . . . .	11
2.5 How to use this book . . . . .	11
2.6 Recommended data science tools . . . . .	11
<b>3 Getting Started with R</b>	<b>13</b>
3.1 R . . . . .	13
3.2 RStudio . . . . .	13
3.3 Git . . . . .	13
3.4 GitHub . . . . .	13
<b>Data Scientific Workflow</b>	<b>17</b>
<b>4 Example Project</b>	<b>17</b>
4.1 Background . . . . .	17

4.2	Other details . . . . .	17
4.3	Conclusions? . . . . .	17
<b>5</b>	<b>Data Preparation</b>	<b>19</b>
5.1	Importation . . . . .	19
5.2	Organization . . . . .	19
5.3	Inspection . . . . .	19
5.4	Manipulation . . . . .	19
5.5	Cleaning . . . . .	19
<b>6</b>	<b>Data Analysis</b>	<b>21</b>
6.1	Transformation . . . . .	21
6.2	Exploration . . . . .	21
6.3	Modeling . . . . .	21
<b>7</b>	<b>Data Visualization</b>	<b>23</b>
7.1	Principles . . . . .	23
7.2	Table Mechanics . . . . .	23
7.3	Chart Mechanics . . . . .	23
7.4	Examples . . . . .	23
<b>8</b>	<b>Insight Delivery</b>	<b>25</b>
8.1	Design principles . . . . .	25
8.2	Scientific inquiry vs storytelling . . . . .	25
8.3	Research reformulation . . . . .	25
8.4	Interactive reporting . . . . .	25
	<b>Reproducible Research</b>	<b>29</b>
<b>9</b>	<b>Tools for Collaboration</b>	<b>29</b>
9.1	Principles . . . . .	29
9.2	Tools . . . . .	29
9.3	Documentation . . . . .	29

<i>CONTENTS</i>	5
9.4 Version control . . . . .	29
9.5 Online repositories for team collaboration . . . . .	29
9.6 Building a code base . . . . .	29
<b>10 Automated Reporting</b>	<b>31</b>
10.1 Excel . . . . .	31
10.2 Word . . . . .	31
10.3 PowerPoint . . . . .	31
10.4 HTML . . . . .	31
<b>Additional Topics</b>	<b>35</b>
<b>11 Machine Learning</b>	<b>35</b>
11.1 Concepts and general workflow (training/test) . . . . .	36
11.2 Unsupervised learning . . . . .	36
11.3 Semisupervised learning . . . . .	36
11.4 Supervised learning . . . . .	36
11.5 Predictive modeling . . . . .	36
11.6 Interpretability . . . . .	36
11.7 Computer vision . . . . .	36
11.8 Other methods and resources . . . . .	36
<b>12 Text Analysis</b>	<b>37</b>
12.1 Data import . . . . .	37
12.2 Analysis . . . . .	37
<b>13 Graph Databases</b>	<b>39</b>
<b>Conclusion</b>	<b>43</b>
<b>14 Conclusion</b>	<b>43</b>
Welcome to the website for <i>Introduction to Data Science for Sensory and Consumer Scientists</i> . This book being written in the open and is currently under development.	



# Introduction





# Chapter 1

## Introduction

- 1.1 How should sensory and consumer scientists learn data science?
- 1.2 Caution: Don't that everybody does
- 1.3 Example projects



## Chapter 2

# What is Data Science?

### 2.1 History

### 2.2 Workflow

#### 2.2.1 Data preparation

#### 2.2.2 Data analysis

#### 2.2.3 Insight delivery

### 2.3 Benefits of data science

#### 2.3.1 Reproducible research

#### 2.3.2 Other benefits (machine learning?)

### 2.4 How to learn data science

### 2.5 How to use this book

### 2.6 Recommended data science tools



## Chapter 3

# Getting Started with R

### 3.1 R

### 3.2 RStudio

### 3.3 Git

### 3.4 GitHub



# Data Scientific Workflow





## Chapter 4

# Example Project

4.1 Background

4.2 Other details

4.3 Conclusions?



## Chapter 5

# Data Preparation

5.1 Importation

5.2 Organization

5.3 Inspection

5.4 Manipulation

5.5 Cleaning



## Chapter 6

# Data Analysis

### 6.1 Transformation

### 6.2 Exploration

### 6.3 Modeling



## Chapter 7

# Data Visualization

### 7.1 Principles

### 7.2 Table Mechanics

### 7.3 Chart Mechanics

### 7.4 Examples





## Chapter 8

# Insight Delivery

8.1 Design principles

8.2 Scientific inquiry vs storytelling

8.3 Research reformulation

8.4 Interactive reporting



# Reproducible Research



## Chapter 9

# Tools for Collaboration

### 9.1 Principles

### 9.2 Tools

#### 9.2.1 GitHub

#### 9.2.2 R scripts

#### 9.2.3 RMarkdown

#### 9.2.4 Shiny

### 9.3 Documentation

### 9.4 Version control

### 9.5 Online repositories for team collaboration

### 9.6 Building a code base

#### 9.6.1 Internal functions

#### 9.6.2 Packages



## Chapter 10

# Automated Reporting

### 10.1 Excel

### 10.2 Word

### 10.3 PowerPoint

#### 10.3.1 Charts

#### 10.3.2 Tables

#### 10.3.3 Bullet Points

#### 10.3.4 Images

### 10.4 HTML





# Additional Topics





## Chapter 11

# Machine Learning

### 11.1 Concepts and general workflow (training/test)

### 11.2 Unsupervised learning

#### 11.2.1 Cluster analysis

#### 11.2.2 Factor analysis

#### 11.2.3 Principle components analysis

#### 11.2.4 t-SNE

### 11.3 Semisupervised learning

#### 11.3.1 PLS regression

### 11.4 Supervised learning

#### 11.4.1 Regression

#### 11.4.2 K-nearest neighbors

#### 11.4.3 Decision trees

#### 11.4.4 Black boxes

##### 11.4.4.1 Random forests

##### 11.4.4.2 SVMs

##### 11.4.4.3 Neural networks

### 11.5 Predictive modeling

## Chapter 12

# Text Analysis

### 12.1 Data import

#### 12.1.1 Data sources

#### 12.1.2 Tokenizing

#### 12.1.3 Lemmatization, stemming, and stop word removal

### 12.2 Analysis

#### 12.2.1 Frequency counts and summary statistics

#### 12.2.2 Word clouds

#### 12.2.3 Contrast plots

#### 12.2.4 Sentiment analysis

#### 12.2.5 Bigrams and word graphs



## Chapter 13

# Graph Databases





# Conclusion



## Chapter 14

## Conclusion



# Appendices

