

# An Introduction to Data Science for Sensory and Consumer Scientists

John Ennis, Julien Delarue, and Thierry Worch

2020-12-28



# Contents

<b>Index</b>	<b>9</b>
<b>1 Index</b>	<b>9</b>
<b>Introduction</b>	<b>13</b>
<b>2 Introduction</b>	<b>13</b>
2.1 Core principles in Sensory and Consumer Science . . . . .	13
2.2 How should sensory and consumer scientists learn data science? .	14
2.3 Caution: Don't that everybody does . . . . .	14
2.4 Example projects . . . . .	14
<b>3 What is Data Science?</b>	<b>15</b>
3.1 History . . . . .	15
3.2 Workflow . . . . .	15
3.3 Benefits of data science . . . . .	15
3.4 How to learn data science . . . . .	15
3.5 How to use this book . . . . .	15
3.6 Recommended data science tools . . . . .	15
<b>4 Getting Started with R</b>	<b>17</b>
4.1 R . . . . .	17
4.2 RStudio . . . . .	17
4.3 Git . . . . .	17
4.4 GitHub . . . . .	17

<b>Data Scientific Workflow</b>	<b>21</b>
<b>5 Example Project</b>	<b>21</b>
5.1 Background . . . . .	21
5.2 Other details . . . . .	21
5.3 Conclusions? . . . . .	21
<b>6 Data Preparation</b>	<b>23</b>
6.1 Importation . . . . .	23
6.2 Organization . . . . .	23
6.3 Inspection . . . . .	23
6.4 Manipulation . . . . .	23
6.5 Cleaning . . . . .	23
<b>7 Data Analysis</b>	<b>25</b>
7.1 Transformation . . . . .	25
7.2 Exploration . . . . .	25
7.3 Modeling . . . . .	25
<b>8 Data Visualization</b>	<b>27</b>
8.1 Principles . . . . .	27
8.2 Table Mechanics . . . . .	27
8.3 Chart Mechanics . . . . .	27
8.4 Examples . . . . .	27
<b>9 Insight Delivery</b>	<b>29</b>
9.1 Design principles . . . . .	29
9.2 Scientific inquiry vs storytelling . . . . .	29
9.3 Research reformulation . . . . .	29
9.4 Interactive reporting . . . . .	29

<i>CONTENTS</i>	5
<b>Reproducible Research</b>	<b>33</b>
<b>10 Tools for Collaboration</b>	<b>33</b>
10.1 Principles . . . . .	33
10.2 Tools . . . . .	33
10.3 Documentation . . . . .	33
10.4 Version control . . . . .	33
10.5 Online repositories for team collaboration . . . . .	33
10.6 Building a code base . . . . .	33
<b>11 Automated Reporting</b>	<b>35</b>
11.1 Excel . . . . .	35
11.2 Word . . . . .	35
11.3 PowerPoint . . . . .	35
11.4 HTML . . . . .	35
<b>Additional Topics</b>	<b>39</b>
<b>12 Machine Learning</b>	<b>39</b>
12.1 Concepts and general workflow (training/test) . . . . .	40
12.2 Unsupervised learning . . . . .	40
12.3 Semisupervised learning . . . . .	40
12.4 Supervised learning . . . . .	40
12.5 Predictive modeling . . . . .	40
12.6 Interpretability . . . . .	40
12.7 Computer vision . . . . .	40
12.8 Other methods and resources . . . . .	40
<b>13 Text Analysis</b>	<b>41</b>
13.1 Data import . . . . .	41
13.2 Analysis . . . . .	41
<b>14 Graph Databases</b>	<b>43</b>

**Conclusion****47****15 Conclusion****47**

# Index





# Chapter 1

## Index

Welcome to the website for *Introduction to Data Science for Sensory and Consumer Scientists*. This book being written in the open and is currently under development.



# Introduction



## Chapter 2

# Introduction

Sensory and consumer science (SCS) is considered as a pillar of food science and technology and is useful to product development, quality control and market research. Most scientific and methodological advances in the field are applied to food. This book makes no exception as we chose a cookie formulation dataset as a main thread. However, SCS widely applies to many other consumer goods so are the content of this book and the principles set out below.

### 2.1 Core principles in Sensory and Consumer Science

#### 2.1.1 Measuring and analyzing human responses

Sensory and consumer science aims at measuring and understanding consumers' sensory perceptions as well as the judgements, emotions and behaviors that may arise from these perceptions. SCS is thus primarily a science of measurement, although a very particular one that uses human beings and their senses as measuring instruments. In other words, sensory and consumer researchers measure and analyze human responses. To this end, SCS relies essentially on sensory evaluation which comprises a set of techniques that mostly derive from psychophysics and behavioral research. It uses psychological models to help separate signal from noise in collected data [ref O'Mahony, D.Ennis, others?]. Besides, sensory evaluation has developed its own methodological framework that includes most refined techniques for the accurate measurement of product sensory properties while minimizing the potentially biasing effects of brand identity and the influence of other external information on consumer perception [Lawless & Heymann, 2010]. A detailed description of sensory methods is beyond the scope of this book and many textbooks on sensory evaluation methods are available to

readers seeking more information. However, just to give a brief overview, it is worth remembering that sensory methods can be roughly divided into three categories, each of them bearing many variants: - Discrimination tests that aim at detecting subtle differences between two products. - Descriptive analysis (DA), also referred to as ‘sensory profiling’, aims at providing both qualitative and quantitative information about product sensory properties. - Hedonic tests. This category gathers affective tests that aim at measuring consumers’ liking for the tested products or their preferences among a product set. Each of these test categories generates its own type of data and related statistical questions in relation to the objectives of the study. Typically, data from difference tests consist in series of correct/failed binary answers depending on whether judges successfully picked the odd sample(s) among a set of three or more samples. These are used to determine whether the number of correct choices is above the level expected by chance. Conventional descriptive analysis data consist in intensity scores given by each panelist to evaluated samples on a series of sensory attributes, hence resulting in a product x attribute x panelist dataset (Figure 1). Note that depending on the DA method, quantifying means other than intensity ratings can be used (ranks, frequency, etc.). Most frequently, each panelist evaluates all the samples in the product set. However, the use of balanced incomplete design can also be found when the experimenters aim to limit the number of samples evaluated by each subject. Eventually, hedonic test datasets consist in hedonic scores (ratings for consumers’ degree of liking or preference ranks) given by each interviewed consumer to a series of products. As for DA, each consumer usually evaluates all the samples in the product set, but balanced incomplete designs are sometimes used too. In addition, some companies favor pure monadic evaluation of product (i.e. between-subject design or independent groups design) which obviously result in unrelated sample datasets. Sensory and consumer researchers also borrow methods from other fields, in particular from sociology and experimental psychology. Definitely a multidisciplinary area, SCS develops in many directions and reaches disciplines that range from genetics and physiology to social marketing, behavioral economics and computational neuroscience. So have diversified the types of data sensory and consumer scientists must deal with.

## **2.2 How should sensory and consumer scientists learn data science?**

## **2.3 Caution: Don’t that everybody does**

## **2.4 Example projects**

## Chapter 3

# What is Data Science?

### 3.1 History

### 3.2 Workflow

#### 3.2.1 Data preparation

#### 3.2.2 Data analysis

#### 3.2.3 Insight delivery

### 3.3 Benefits of data science

#### 3.3.1 Reproducible research

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

### 3.4 How to learn data science

### 3.5 How to use this book

### 3.6 Recommended data science tools





## Chapter 4

# Getting Started with R

### 4.1 R

### 4.2 RStudio

### 4.3 Git

### 4.4 GitHub



# Data Scientific Workflow



## Chapter 5

# Example Project

5.1 Background

5.2 Other details

5.3 Conclusions?



## Chapter 6

# Data Preparation

6.1 Importation

6.2 Organization

6.3 Inspection

6.4 Manipulation

6.5 Cleaning





## Chapter 7

# Data Analysis

### 7.1 Transformation

### 7.2 Exploration

### 7.3 Modeling



## Chapter 8

# Data Visualization

### 8.1 Principles

### 8.2 Table Mechanics

### 8.3 Chart Mechanics

### 8.4 Examples



## Chapter 9

# Insight Delivery

9.1 Design principles

9.2 Scientific inquiry vs storytelling

9.3 Research reformulation

9.4 Interactive reporting



# Reproducible Research





## Chapter 10

# Tools for Collaboration

### 10.1 Principles

### 10.2 Tools

#### 10.2.1 GitHub

#### 10.2.2 R scripts

#### 10.2.3 RMarkdown

#### 10.2.4 Shiny

### 10.3 Documentation

### 10.4 Version control

### 10.5 Online repositories for team collaboration

### 10.6 Building a code base

#### 10.6.1 Internal functions

#### 10.6.2 Packages



## Chapter 11

# Automated Reporting

### 11.1 Excel

### 11.2 Word

### 11.3 PowerPoint

#### 11.3.1 Charts

#### 11.3.2 Tables

#### 11.3.3 Bullet Points

#### 11.3.4 Images

### 11.4 HTML



# Additional Topics





## Chapter 12

# Machine Learning

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

### 12.2 Unsupervised learning

#### 12.2.1 Cluster analysis

#### 12.2.2 Factor analysis

#### 12.2.3 Principle components analysis

#### 12.2.4 t-SNE

### 12.3 Semisupervised learning

#### 12.3.1 PLS regression

### 12.4 Supervised learning

#### 12.4.1 Regression

#### 12.4.2 K-nearest neighbors

#### 12.4.3 Decision trees

#### 12.4.4 Black boxes

##### 12.4.4.1 Random forests

##### 12.4.4.2 SVMs

##### 12.4.4.3 Neural networks

### 12.5 Predictive modeling



# Chapter 13

## Text Analysis

### 13.1 Data import

#### 13.1.1 Data sources

#### 13.1.2 Tokenizing

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

### 13.2 Analysis

#### 13.2.1 Frequency counts and summary statistics

#### 13.2.2 Word clouds

#### 13.2.3 Contrast plots

#### 13.2.4 Sentiment analysis

#### 13.2.5 Bigrams and word graphs



## Chapter 14

# Graph Databases



# Conclusion



## Chapter 15

## Conclusion





# Appendices

