An Introduction to Data Science for Sensory and Consumer Scientists

John Ennis, Julien Delarue, and Thierry Worch

2020-12-29

Contents

Preface **7**

About the Authors **9**

Introduction 13

1 Introduction **13**

1.1 Core principles in Sensory and Consumer Science 13

1.2 How should sensory and consumer scientists learn data science? . 14

1.3 Caution: Don't that everybody does 14

1.4 Example projects 14

2 What is Data Science? **15**

2.1 History 15

2.2 Workflow 15

2.3 Benefits of data science 15

2.4 How to learn data science 15

2.5 How to use this book 15

2.6 Recommended data science tools 15

3 Getting Started with R **17**

3.1 R 17

3.2 RStudio 17

3.3 Git 17

3.4 GitHub 17

Data Scientific Workflow **21**

4 Example Project **21**

4.1 Background 21

4.2 Other details 21

4.3 Conclusions? 21

5 Data Preparation **23**

5.1 Importation 23

5.2 Organization 23

5.3 Inspection 23

5.4 Manipulation 23

5.5 Cleaning 23

6 Data Analysis **25**

6.1 Transformation 25

6.2 Exploration 25

6.3 Modeling 25

7 Data Visualization **27**

7.1 Principles 27

7.2 Table Mechanics 27

7.3 Chart Mechanics 27

7.4 Examples 27

8 Insight Delivery **29**

8.1 Design principles 29

8.2 Scientific inquiry vs storytelling 29

8.3 Research reformulation 29

8.4 Interactive reporting 29

Reproducible Research **33**

9 Tools for Collaboration **33**

9.1 Principles . . . 33
9.2 Tools . . . 33
9.3 Documentation . . . 33
9.4 Version control . . . 33
9.5 Online repositories for team collaboration . . . 33
9.6 Building a code base . . . 33

10 Automated Reporting **35**

10.1 Excel . . . 35
10.2 Word . . . 35
10.3 PowerPoint . . . 35
10.4 HTML . . . 35

Additional Topics **39**

11 Machine Learning **39**

11.1 Concepts and general workflow (training/test) . . . 40
11.2 Unsupervised learning . . . 40
11.3 Semisupervised learning . . . 40
11.4 Supervised learning . . . 40
11.5 Predictive modeling . . . 40
11.6 Interpretability . . . 40
11.7 Cmputer vision . . . 40
11.8 Other methods and resources . . . 40

12 Text Analysis **41**

12.1 Data import . . . 41
12.2 Analysis . . . 41

13 Graph Databases **43**

Conclusion **47**

14 Conclusion **47**

Preface

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.

About the Authors

John Ennis ...

Julien Delarue ...

Thierry Worch ...

Introduction

Chapter 1

Introduction

Sensory and consumer science (SCS) is consider 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.

1.1 Core principles in Sensory and Consumer Science

1.1.1 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.

1.2 How should sensory and consumer scientists learn data science?

1.3 Caution: Don't that everybody does

1.4 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