Applied Geodata Science

Benjamin Stocker (lead), Koen Hufkens (contributing), Pepa Aran (contributing)

2022-11-22

Contents

About this book

This book accompanies the course(s) *Applied Geodata Science*, taught at the Institute of Geography, University of Bern.

The course introduces the typical data science workflow using various examples of geographical and environmental data. With a strong hands-on component and a series of input lectures, the course introduces the basic concepts of data science and teaches how to conduct each step of the data science workflow. This includes the handling of various data formats, the formulation and fitting of robust statistical models, including basic machine learning algorithms, the effective visualisation and communication of results, and the implementation of reproducible workflows, founded in Open Science principles. The overall course goal is to teach students to tell a story with data.

Course plan

1. Getting started
2. Programming primer
3. Data wrangling
4. Data visualisation
5. Data variety
6. Code management
7. Open Science practice

   MILESTONE 1: Communicating a reproducible workflow (→ LO1)

8. Regression
9. Supervised machine learning fundamentals
10. Random Forest
11. Neural Networks
12. Interpretable machine learning
13. Unsupervised machine learning

    MILESTONE 2: Identify patterns and demonstrate how explained (→ LO2)

Chapter 1

Getting started

Chapter lead author: Pepa Aran

TBC

Contents:

```
- Lecture (Beni): Data revolution, opportunities, challenges; explain relevance and why new metho
- installing environment
- workspace management
- R, RStudio
- R libraries, other libraries and applications
```

Chapter 2

Programming primers

Chapter lead author: Pepa Aran

TBC

Contents:

```
- Lecture (Beni): Models and data
- Base R
- variables, classes
- data frames
- loops
- conditional statements
- functions
- input and output
- intro to visualisation
- Performance assessment: [link](https://stineb.netlify.app/files/ex1.pdf) to my exercise, [link
```

Chapter 3

Data wrangling

Chapter lead author: Benjamin Stocker

Contents:

```
- Lecture (Beni): Tidy data, "bad" data
- Data frame manipulations with tidyverse
- Tidy data
- Dealing with missingness, bad data, outliers
- Imputation (note also imputation as part of the modelling workflow)
- Performance assessment: **CAT 1,** [link](https://stineb.github.io/esds_book/ch-02.html#exercis
```

Chapter 4

Data visualisation

Chapter lead author: Benjamin Stocker

Contents:

- Lecture (Isabelle Bentz?): The art of visualising data, grammar of graphics
- Exercise: Develop decision tree for what type of visualisation to apply
- Performance assessment: Interactive work sequence

Chapter 5

Data variety

Chapter lead author: Koen Hufkens

Contents:

- Lecture (Mirko): Mapping data
- Data formats, standards, metadata
- Geographic data
- Scraping, wget
- APIs

Chapter 6

Code management

Chapter lead author: Koen Hufkens

Contents:

```
- git: repositories, stage, commit, push, fork, pull request, fetch upstream
- Performance assessment: **CAT 2**
```

Chapter 7

Open science practices

Chapter lead author: Koen Hufkens

Contents:

```
- Lecture (Koen): Open science - history, motivation, reproducibility crisis, current initiatives
- Environmental data repositories
- Methods to create visualised reproducible workflow
- RMarkdown files
- Performance assessment: **CAT 3**, [link to Dietze exercise on pair coding](https://github.com/

**MILESTONE 1: Communicating a reproducible workflow (→ LO1)**
```