

A Minimal Book Example

Yihui Xie

2019-07-01

Contents

1	Prerequisites	5
2	Introduction	7
2.1	The Policy Analyst's Challenge	7
2.2	The Data Scientist's Challenge	8
2.3	Information Asymmetry	9
2.4	Goals	10
3	Literature	11
4	Methods	13
5	Applications	15
5.1	Example one	15
5.2	Example two	15
6	Final Words	17

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")  
# or the development version  
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading #.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.name/tinytex/>.

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 4.

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

2.1 The Policy Analyst's Challenge

Data Scientists have an interesting position in policy circles these days. They have been coming into meeting rooms as evangelists for an innovative new way of approaching age-old problems that have vexed policy makers for decades. They are able to enter the room with their heads held high, MacBooks tucked gently under their arms as they assume the head of the table. They wield a new vocabulary and confident promises that they can revolutionize whatever policy problem you're having with just a few simple scripts.

These data scientists have command over a powerful language of technical terms that have been propelled to hypnotic power thanks to sensational articles from business magazines. Companies like the Economist and Harvard Business Review have formed a symbiotic relationship with data scientists – The latter shows a proof of concept of an interesting new idea, and the former will propel the idea to the stratosphere with their articles asking if a new technological world order is upon us.

Data scientists use this new language, along with a screen with several command lines open on one half of the screen, and long files of code on the other before they open the sleek new interface to a product that will solve your problem. Like a wizard uttering a spell, they use technical terms like *artificial intelligence*, *gradient descent* or *algorithm* to assert their authority of the room. All the

“non-technical” folks in the room are immediately subject to a sinister form of information asymmetry. The articles that appear in their inboxes tell them that these new technologies are the key to staying ahead in the new world, and the data scientist in front of them are providing an example of this right in front of them. The example they see is compelling, and if the data scientist’s claims are true, how could they refuse them?

The data scientist has a clear advantage in these situations. Policy analysts are often desperate to learn about this new environment that is being thrust upon them. The articles being written in every business magazine are telling them that they will be left behind if they don’t adopt these technologies. Unfortunately for these analysts, they don’t have the time or the resources to even learn about the techniques that data scientists champion in their meetings. Their arms are twisted in these meetings. They can’t question the techniques and assumptions that data scientists are employing before their eyes, and they can’t refuse to give these methods a shot for fear of ridicule from management above and below. If they reject the data scientist, they are often mocked as a dinosaur, and are warned that they are holding the organization back from modernization.

The data scientist is now ready to go in for the kill. The policy analysts are powerless, and are all but forced to accept the data scientist’s proposal. The data scientist asks for a contract to continue working on this project. They ask for more data, more resources, and time. The policy analyst is in no position to refuse, and accepts. Perhaps the manager and analyst walking out of the meeting are excited for what’s coming next. Or maybe they’re confused, skeptical and disoriented. To the data scientist, it almost doesn’t matter. They walked into that meeting room with a game plan to wrestle the analysts into submission, and it worked without a hitch, as it almost often does.

I’ve seen this story unfold in front of me countless times. I’ve seen managers in the policy space rubbing their head trying to understand why these methods are starting to work. When the analysts poke holes in the demonstration, data scientists are quick to remind them that the method they’re proposing is a *proof of concept*, and not nearly completed. This neutralizes the analysts, leading them right into the palm of the data scientist’s hands.

2.2 The Data Scientist’s Challenge

The story is a little different from the data scientist’s perspective. They have seen an institutional problem, and they see a pool of data that is not being utilized by the policy analysts. They’ve done the research in the techniques and methods that could address this problem. They know that there is an *open source* package that can implement the statistical techniques necessary to help this problem, and they know how to get a hold of the data. They get to work quickly, excited about the impact that their technology can have. After a week

or two of *sprinting* on this problem, they finish their proof of concept and quickly arrange a meeting to show the exciting new technology they have.

The data scientist prepares their presentation. They drum up a couple of quick slides to explain what they see the problem is, and what the technology they've developed can do to solve the problem. They purposely omit the technical mathematical concepts, replacing it with intuitive explanations of the technology—not because they want to hide behind these techniques, but because they want to focus on the results. The mechanics behind technology can be intimidating and would distract from what's actually going on. While they would love to talk about the methods and intuition behind the mathematics that makes their new technology possible, but they have so much to cover and so little time to do so.

The data scientist believes in their methods, and knows that they can do great things with just a little bit more time, and a little more data. Of course the policy analyst is going to find errors in the technology, the data scientist only had two weeks to build this proof of concept! The algorithms behind the technology could be so much more effective with a little tweaking, and with more descriptive data. There are going to be errors in the assumptions, because the data scientist didn't access to the expertise that the policy analysts have spent years accumulating. They are trying to communicate what the technology can do now, and what it could do in the future—if only they could get the managers to sign on and collaborate in the development.

2.3 Information Asymmetry

The policy analysts and managers in our meeting are forced to deal with a troubling disadvantage. They have to decode the all the information presented in a different language to them. They have to think critically about the implications of the technique they likely have never seen, and make a call on whether to pursue this project.

This is an unfortunate case of asymmetric information on both sides. The data scientists want to communicate the work they're doing, but they can only guess as to the exact needs of the client. The analysts themselves are caught unaware of the mechanics behind the technology they're being pressured to endorse. The asymmetry could lead to the two parties talking past each other, not really addressing the other's questions. At worst, loads of time and money can be spent developing a product that doesn't quite fit the needs of the client. It's a tricky situation to navigate, often requiring one of the party's to understand the exact situation of the other and translate their methods to the other's language. This communication is a challenging art that comes with experience.

For policy analysts, an understanding of the techniques used in a data scientist's toolkit can help bridge the communication gap between the two. In more sinister cases, a knowledge of the intuition behind data science's most sophisticated techniques could prove most useful as a defense from clueless colleagues

or nefarious consultants.

2.4 Goals

The goal of this book is to introduce some of the most important concepts in data science to a non-technical policy analyst. Each chapter will present a different set of concepts to the reader, explaining the intuition and ideas behind some of the most popular technology in the world. The hope is that the reader will understand the concepts in data science well enough to identify the main ideas behind the technology they see every day. They will have better judgement in implementing data science in their own projects, and will be able to communicate better with data scientists.

I find the ideas behind data science to be beautiful. There is elegant simplicity in some of data science's most popular workflows. I believe these ideas can stand on their own, and can be communicated without technical background.

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2019) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Chapter 3

Literature

Here is a review of existing methods.

Chapter 4

Methods

We describe our methods in this chapter.

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.

Bibliography

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

Xie, Y. (2019). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.11.