# Machine Learning, Statistics, and Data Mining for Heliophysics

Monica Bobra[1] and James Mason[2]

[1]Stanford University
[2]University of Colorado at Boulder

November 24, 2022

**Abstract**

We present a book entitled "Machine Learning, Statistics, and Data Mining for Heliophysics," an online and open source book available at helioml.org. This book includes a collection of interactive Jupyter notebooks, written in the programming language Python, that walks the reader through the process of applying machine learning, statistics, and data mining techniques on various kinds of solar and space physics data sets to reproduce published results. We consider this book to be a living document with frequent updates. Please contact us if you'd like to submit a chapter!

# We present an online book about machine learning, statistics, and data mining for heliophysics.

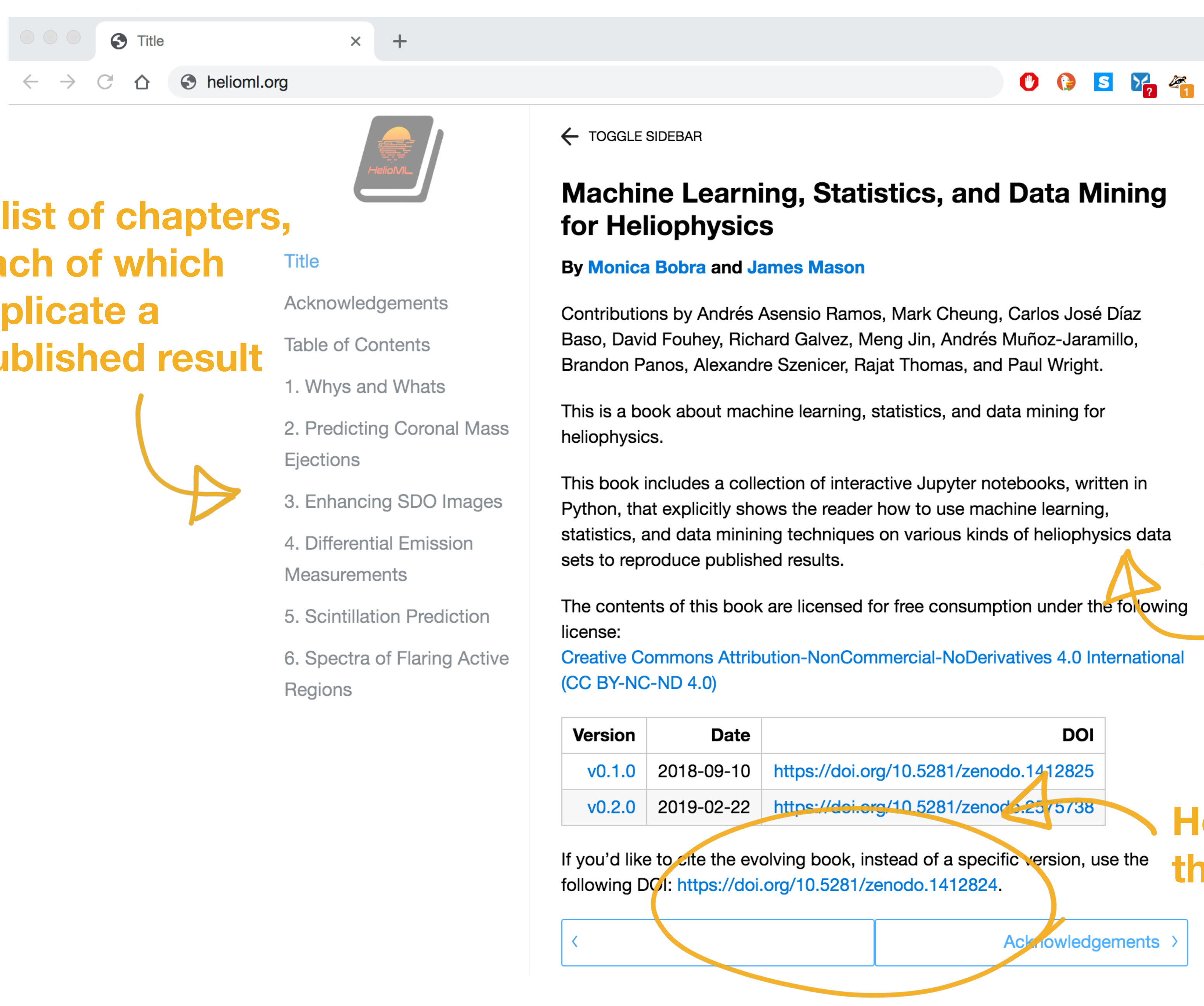
## What's in this book?

Title Page

A closer look at Chapter 6

Click the interact button to run the code – no setup necessary!

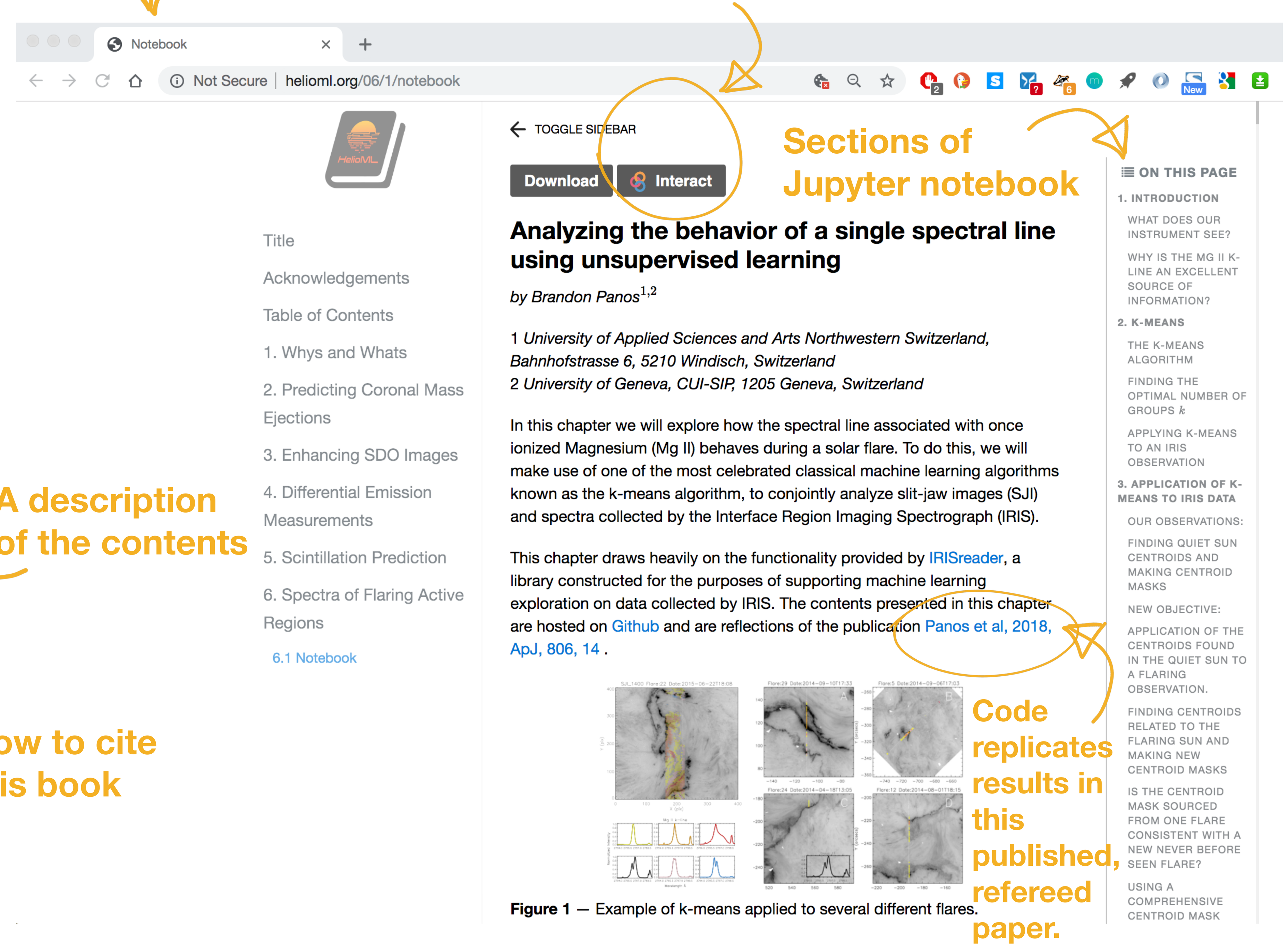
A list of chapters, each of which replicate a published result




A description of the contents

How to cite this book

Sections of Jupyter notebook




Code replicates results in this published, refereed paper.

## Why did we write this book?

### To teach readers how to create replicable results.

"To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis, unless such information is restricted by non-public data policies. That information should include the data, study methods, and computational environment."

*– National Academies of Sciences, Engineering, and Medicine report on Reproducibility and Replicability in Science (2019)*

### To show readers how to use modern analysis techniques on heliophysics data.

"…scientific research in many disciplines is becoming more and more dependent on the careful analysis of large datasets. This analysis requires a skill-set as broad as it is deep: scientists must be experts not only in their own domain, but in statistics, computing, algorithm building, and software design as well."

*– Jake VanderPlas, author of The Python Data Science Handbook (2016)*

We consider this book a living document. Please contact us if you'd like to submit a chapter!


**By** Monica Bobra (Stanford University) and James Mason (University of Colorado at Boulder)

**With contributions by** Andrés Asensio Ramos, Mark Cheung, Carlos José Díaz Baso, David Fouhey, Richard Galvez, Meng Jin, Andrés Muñoz-Jaramillo, Brandon Panos, Alexandre Szenicer, Rajat Thomas, and Paul Wright


Take a picture to load the book