# Building “Geographic Data Science…”

Dani Arribas-Bel [@darribas]

UNIVERSITY OF LIVERPOOL

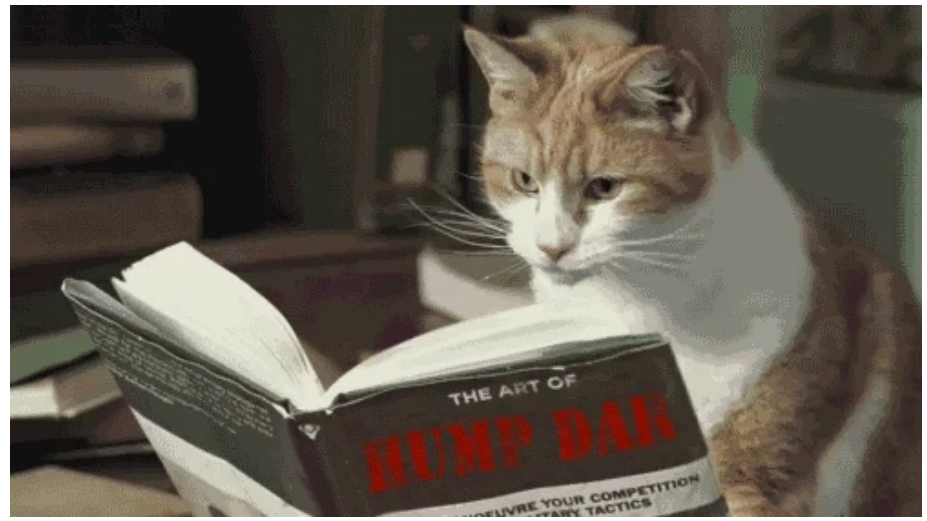
The Alan Turing Institute

Geographic Data Science Lab

# This talk

darribas.org/gdsbook_overview/202110

# We have a book!




via GIPHY

# Almost…

via GIPHY

## Coming “soon” but…

# Coming “soon” but…

… you can already:

- `https://geographicdata.science`
- launch binder
- `https://github.com/gdsbook/book`

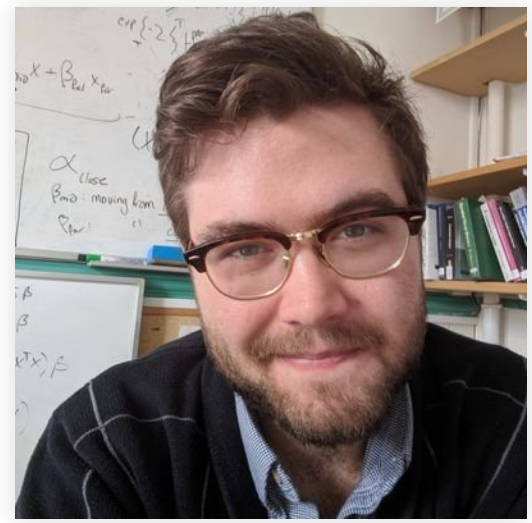
# The Authors

@sreyog

Serge Rey

@darribas

Dani Arribas-Bel



@levijohnwolf

Levi Wolf

# Today

- *Why?* – Vision
- *What?* – Book
- *How?* – Community

*Why?*

Data, data, data

Data Science

. . .

It’s called GEOGRAPHIC Data Science!!!

# Geographic Data Science

geographical analysis

*Geographical Analysis* (2021) **53**, 61–75

Special Issue

## Geographic Data Science

Alex Singleton, Daniel Arribas-Bel

Department of Geography and Planning, University of Liverpool, Liverpool, L69 7ZT, U.K.

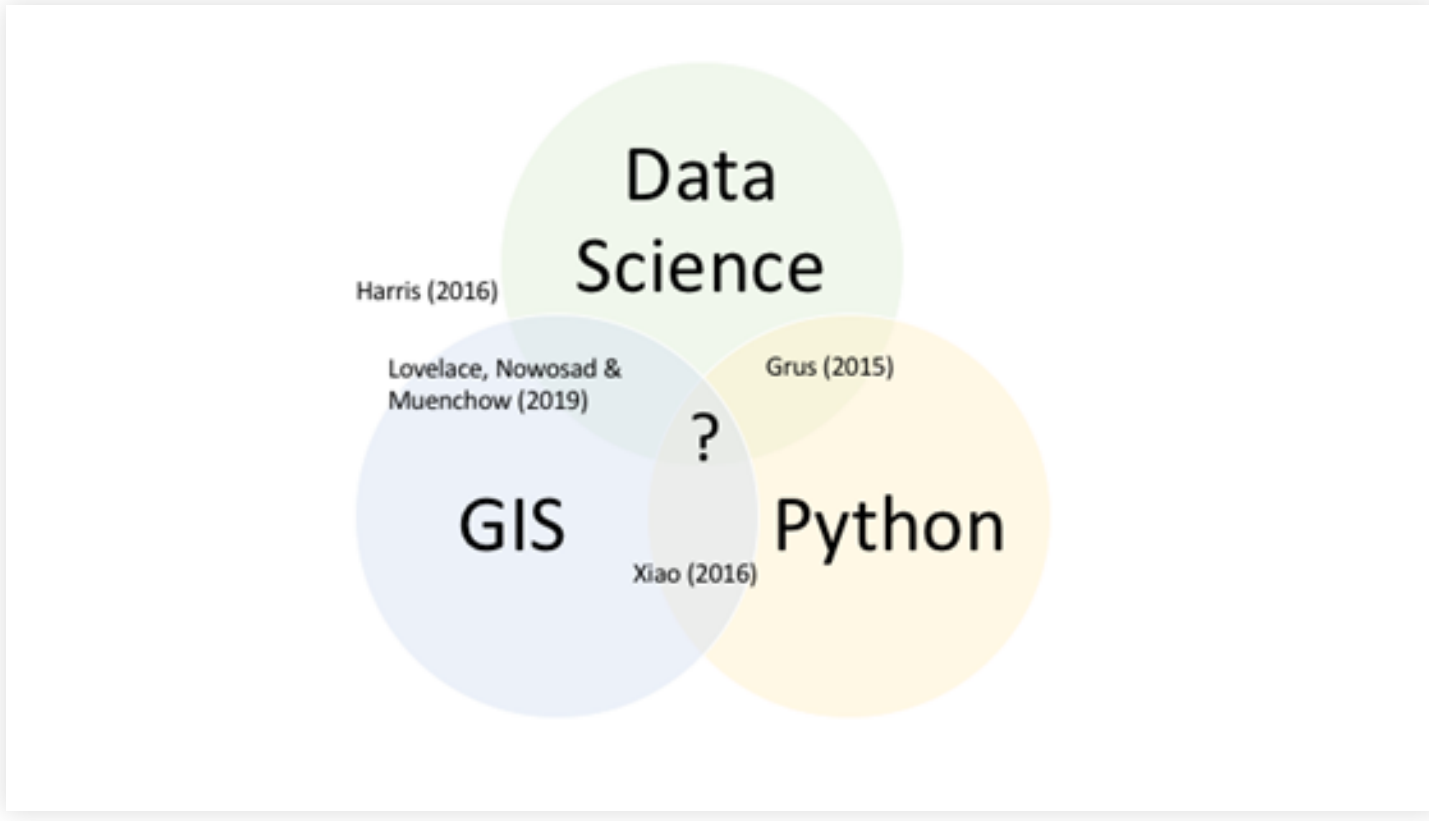
*It is widely acknowledged that the emergence of "Big Data" is having a profound and often controversial impact on the production of knowledge. In this context, Data Science has developed as an interdisciplinary approach that turns such "Big Data" into information. This article argues for the positive role that Geography can have on Data Science when being applied to spatially explicit problems; and inversely, makes the case that there is much that Geography and Geographical Analysis could learn from Data Science. We propose a deeper integration through an ambitious research agenda, including systems engineering, new methodological development, and work toward addressing some acute challenges around epistemology. We argue that such issues must be resolved in order to realize a Geographic Data Science, and that such goal would be a desirable one.*

# Geographic Data Science

- GIScience ↔ Data Science
- Foster innovation (avoid reinventing the wheel)
- Grow a community around collaboration (1+1>2)
  - **Diverse** (`!= white male engineer`)
  - **Inclusive** (*embrace different perspectives*)
  - **Welcoming** (*make it easy to join… and stay*)

# *What?*

# The Book is…


Data
Science
Harris (2016)
Lovelace, Nowosad &
Muenchow (2019)
Grus (2015)
?
GIS
Python
Xiao (2016)

# The Book is *not*…

- A GIS starter
- An introduction to programming
- An in-depth volume (rather *in-breath*)

# The Book is for…

- Data Scientists who work on spatial problems
- GIScientists who want to "update"
- (Social) Scientists getting started in geospatial

# Table of Contents

Contents

Part I: Building Blocks

Part II: Spatial Data Analysis

Part III: Advanced Topics

## Part I: Building Blocks

- Geographic Thinking for Data Scientists
- Geospatial Computational Environment
- Spatial data
- Spatial weights

## Part II: Spatial Data Analysis

- Choropleth Mapping
- Spatial Autocorrelation
- Local Spatial Autocorrelation
- Point Pattern Analysis

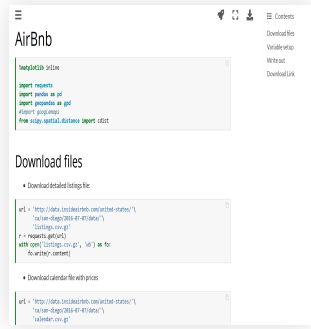
## Part III: Advanced Topics

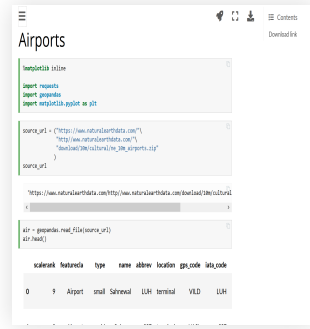
- Spatial Inequality
- Clustering and Regionalization
- Spatial Regression
- Spatial Feature Engineering

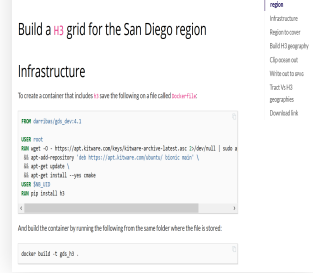
# GIScience + Data Science

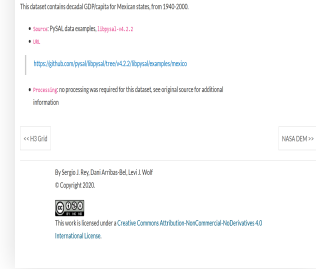
- `geopandas`
- `rasterio`
- `pysal`
- `osmnx`
- `contextily`
- `pandas`
- `xarray`
- `scikit-learn`
- `matplotlib`
- `seaborn`

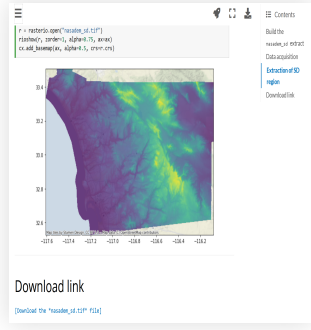
# Bonus: Datasets

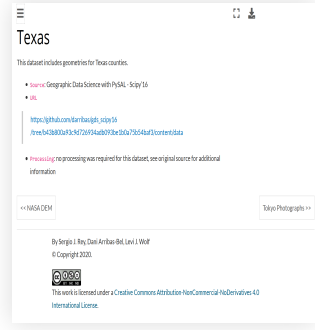




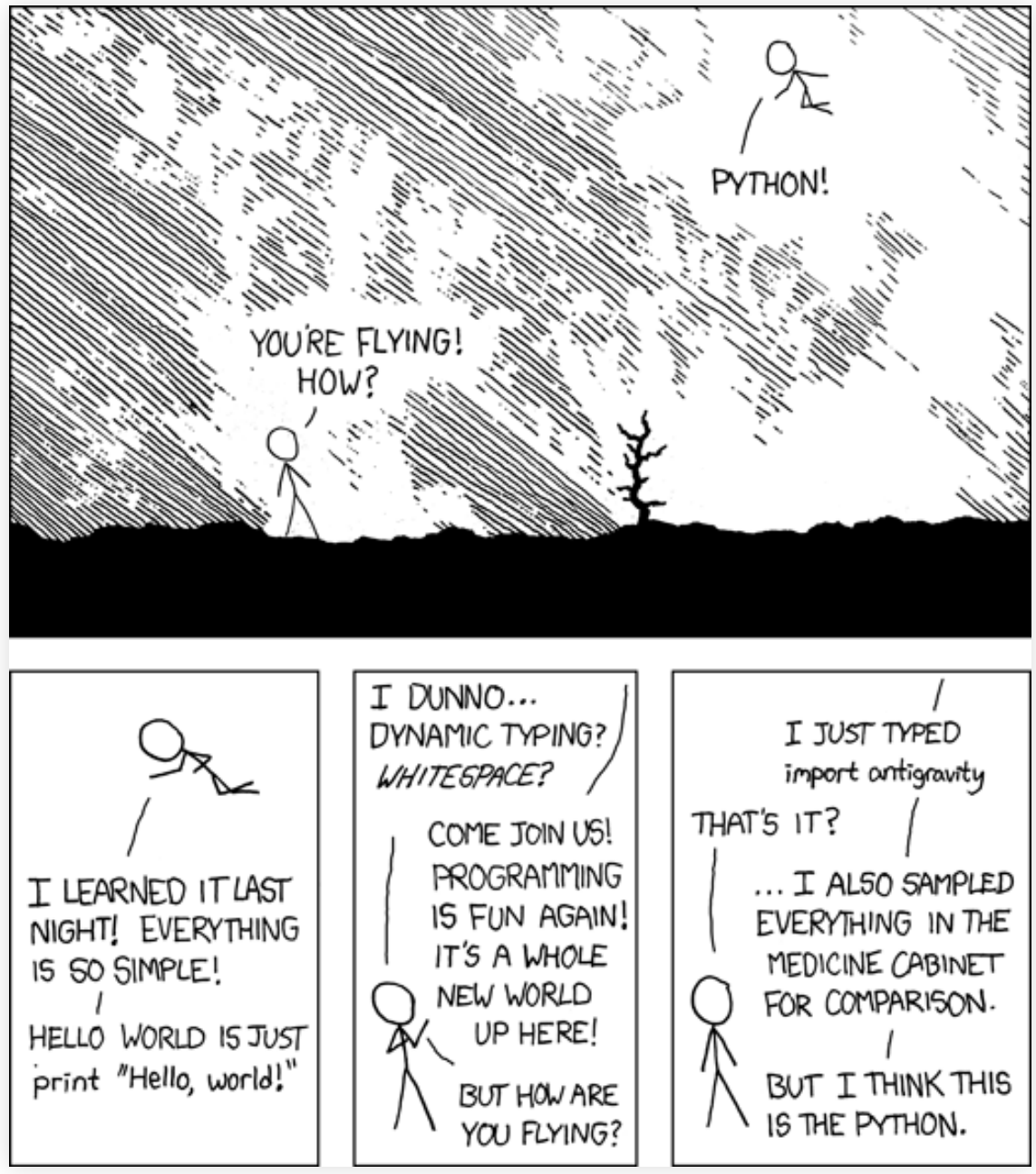



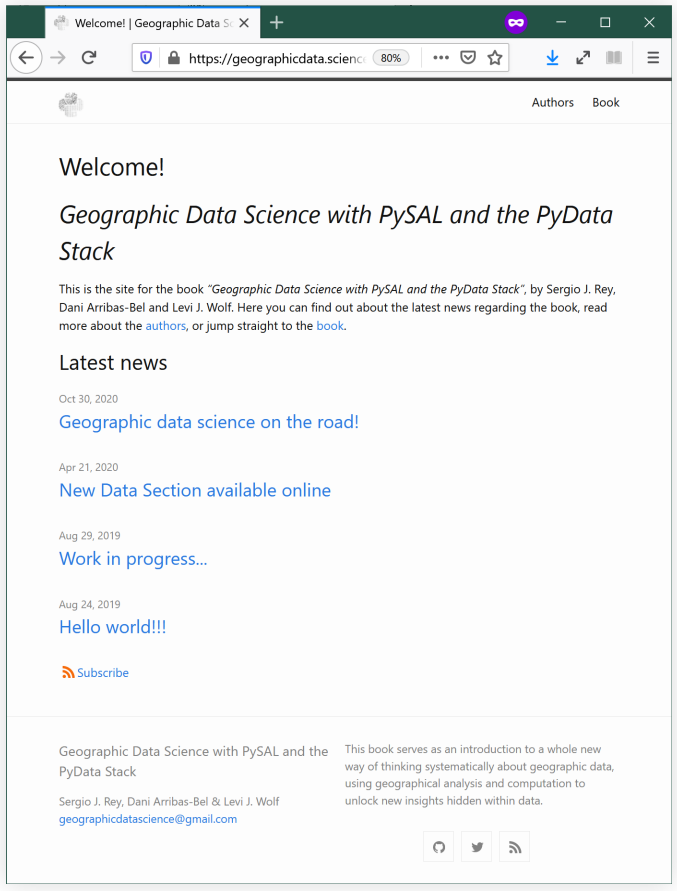
# *How?*

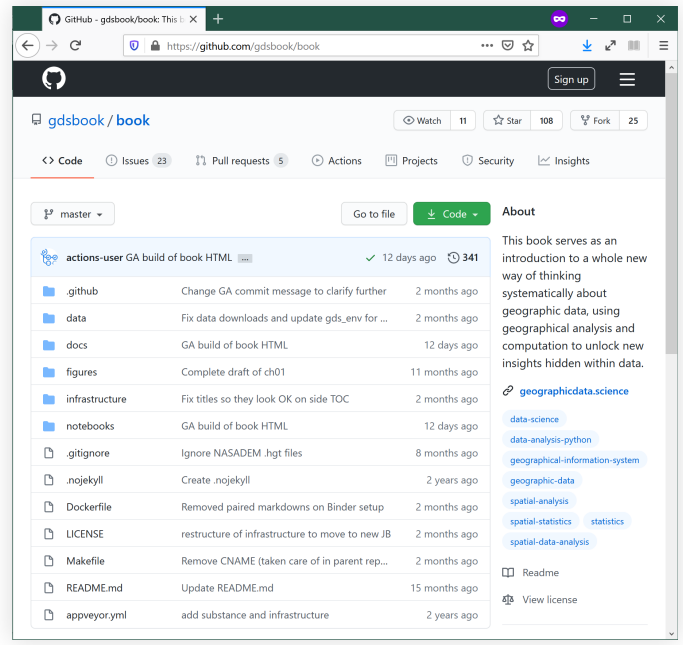
# Python

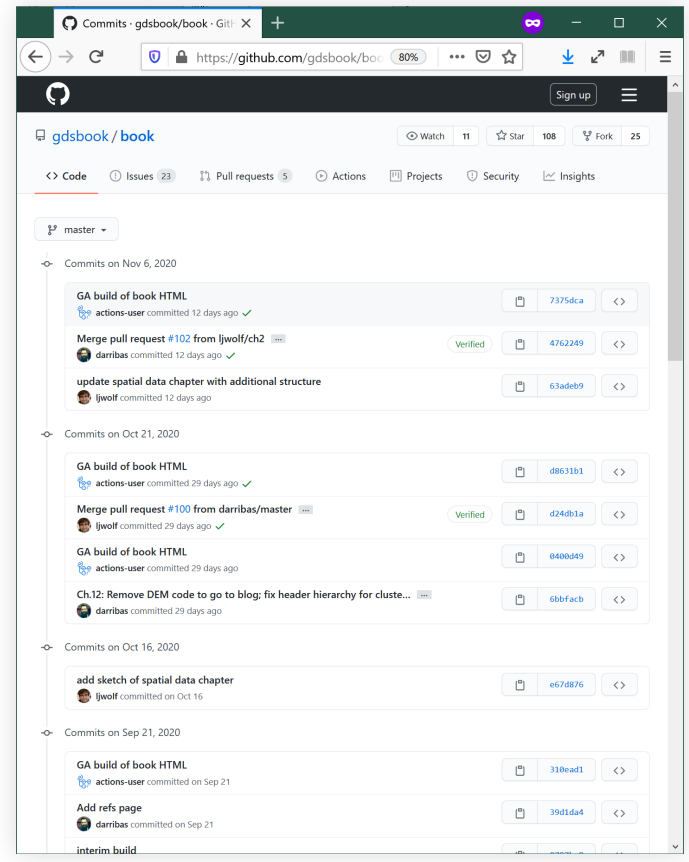



Source

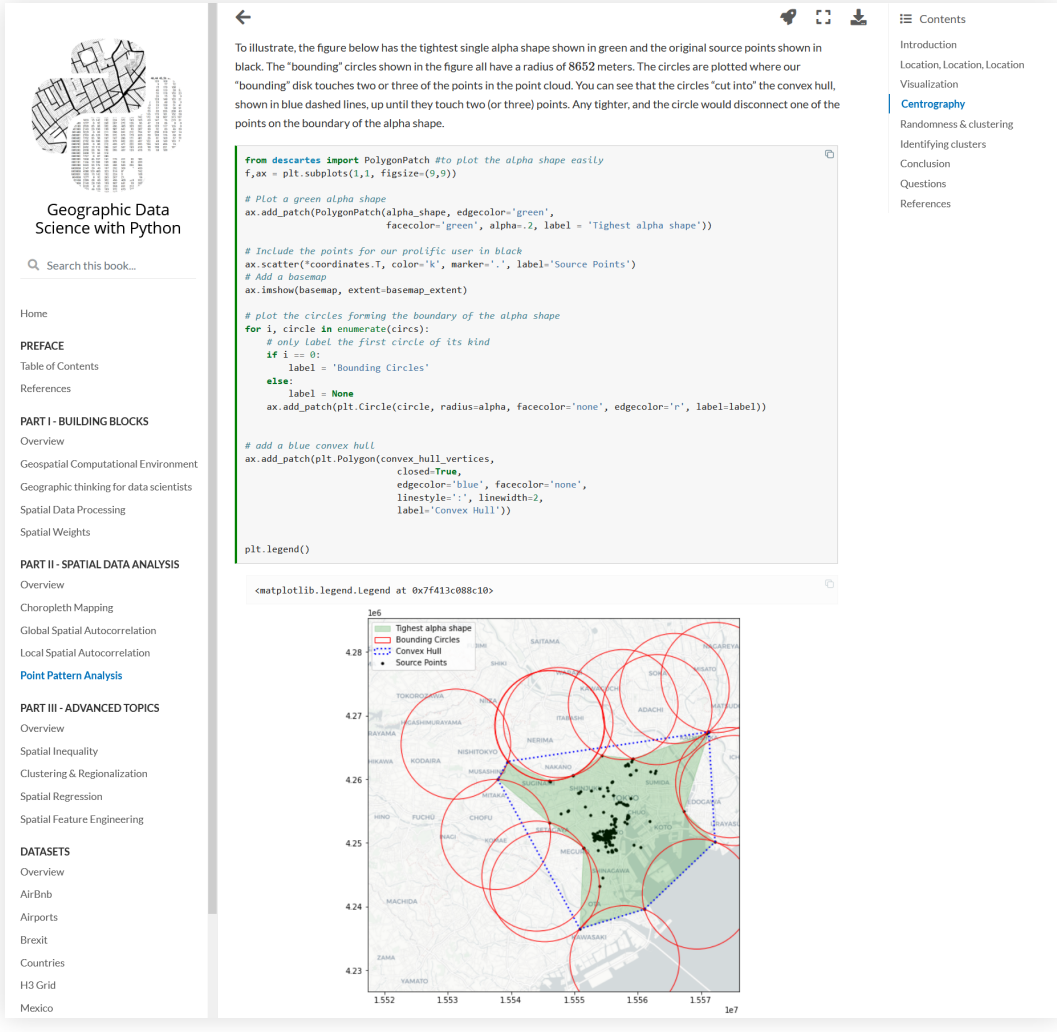
# Radically Open

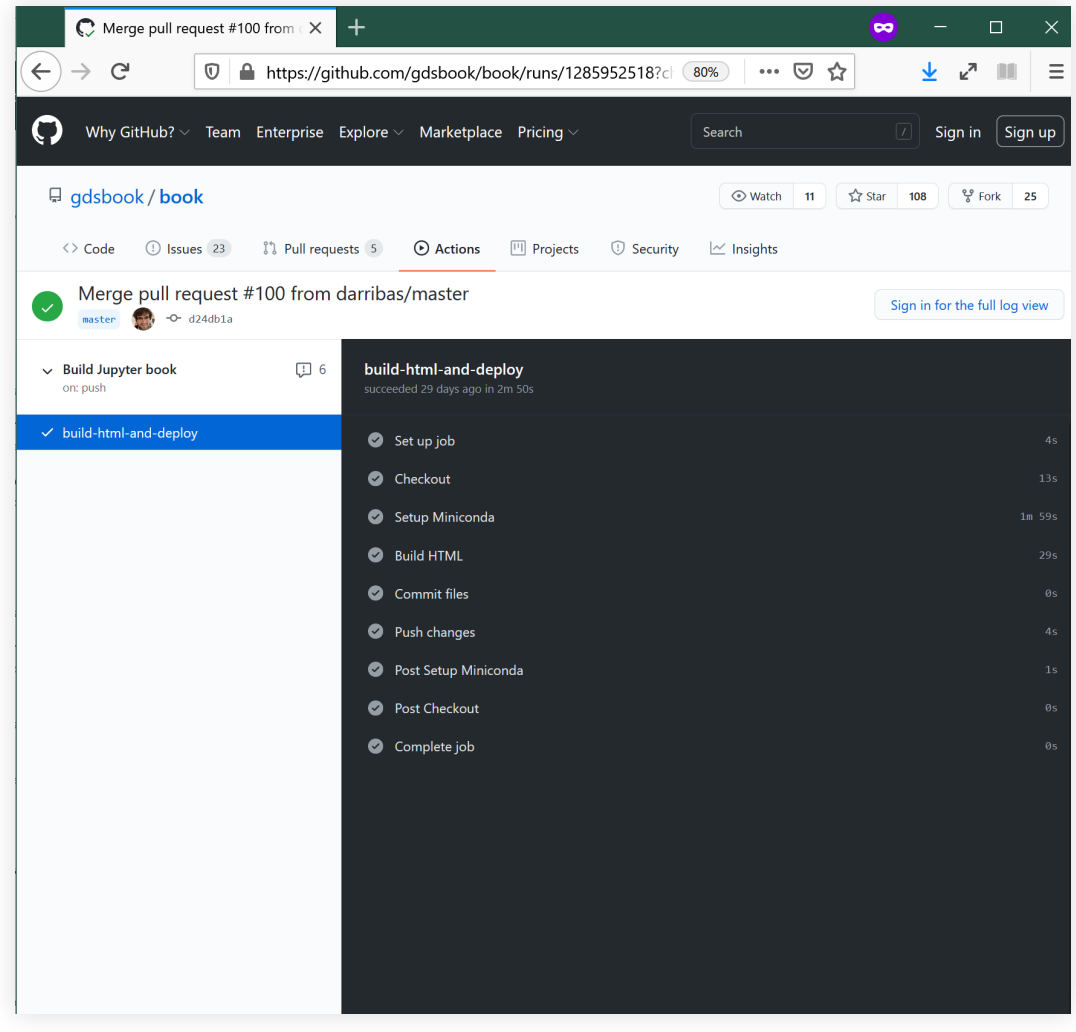


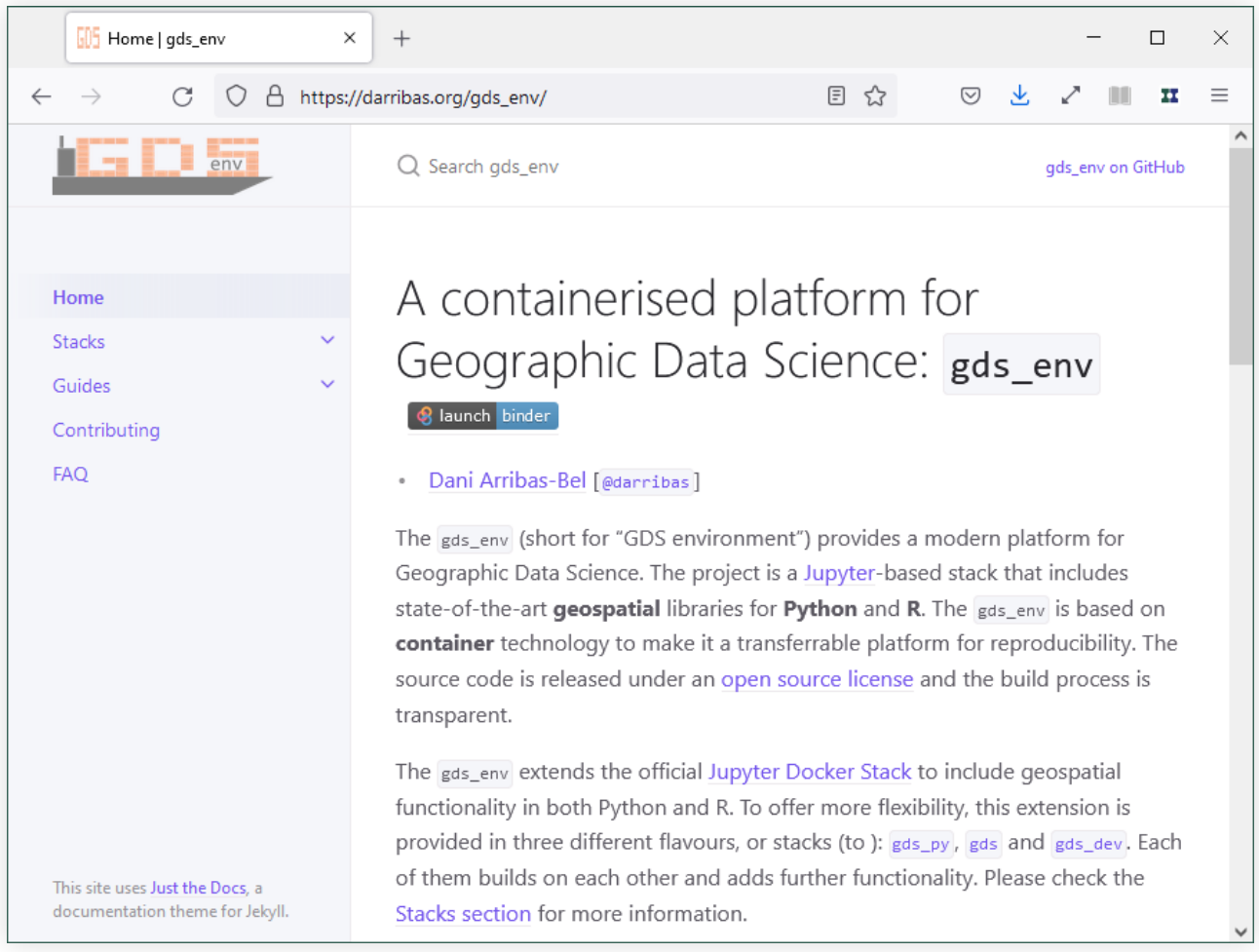



# Code as text; text as codE



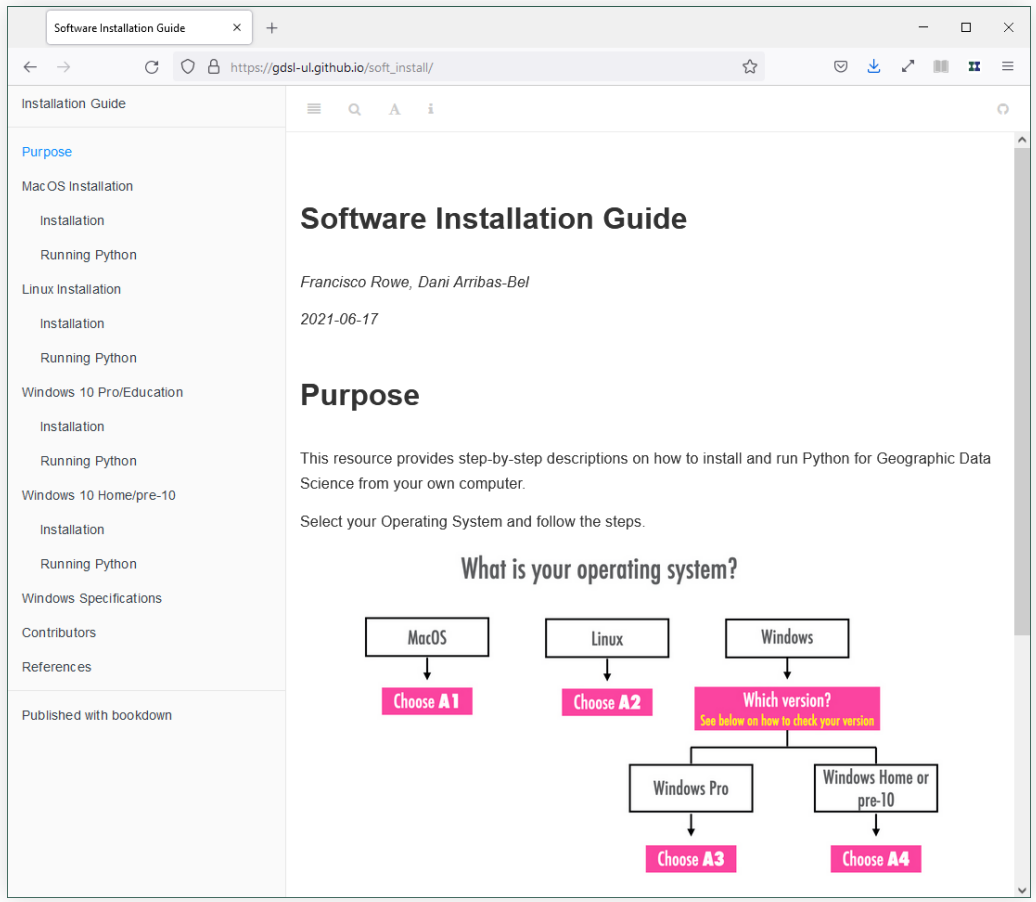



# Runs *anywhere*...



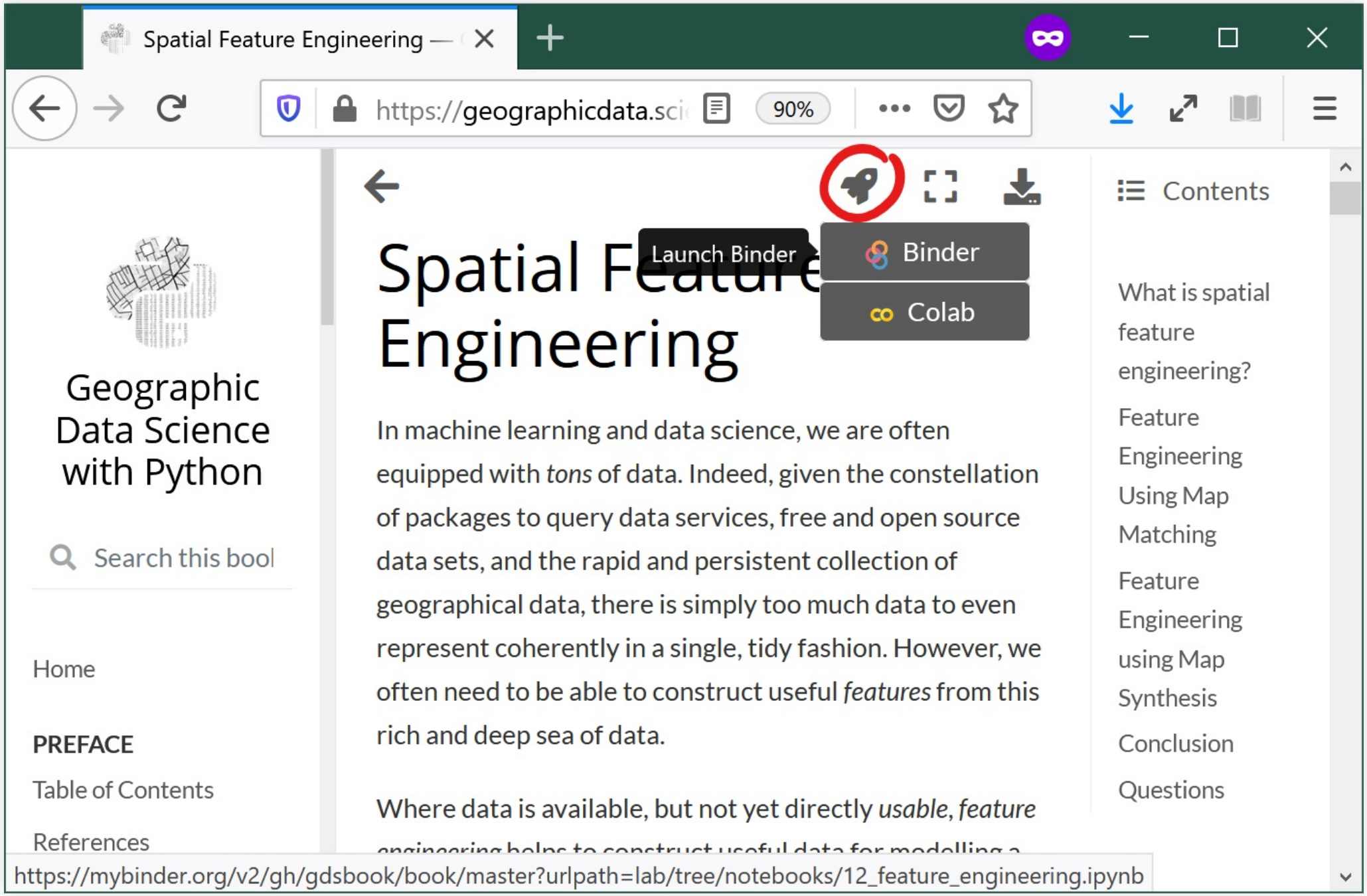

darribas.org/gds_env

# … by *anyone*

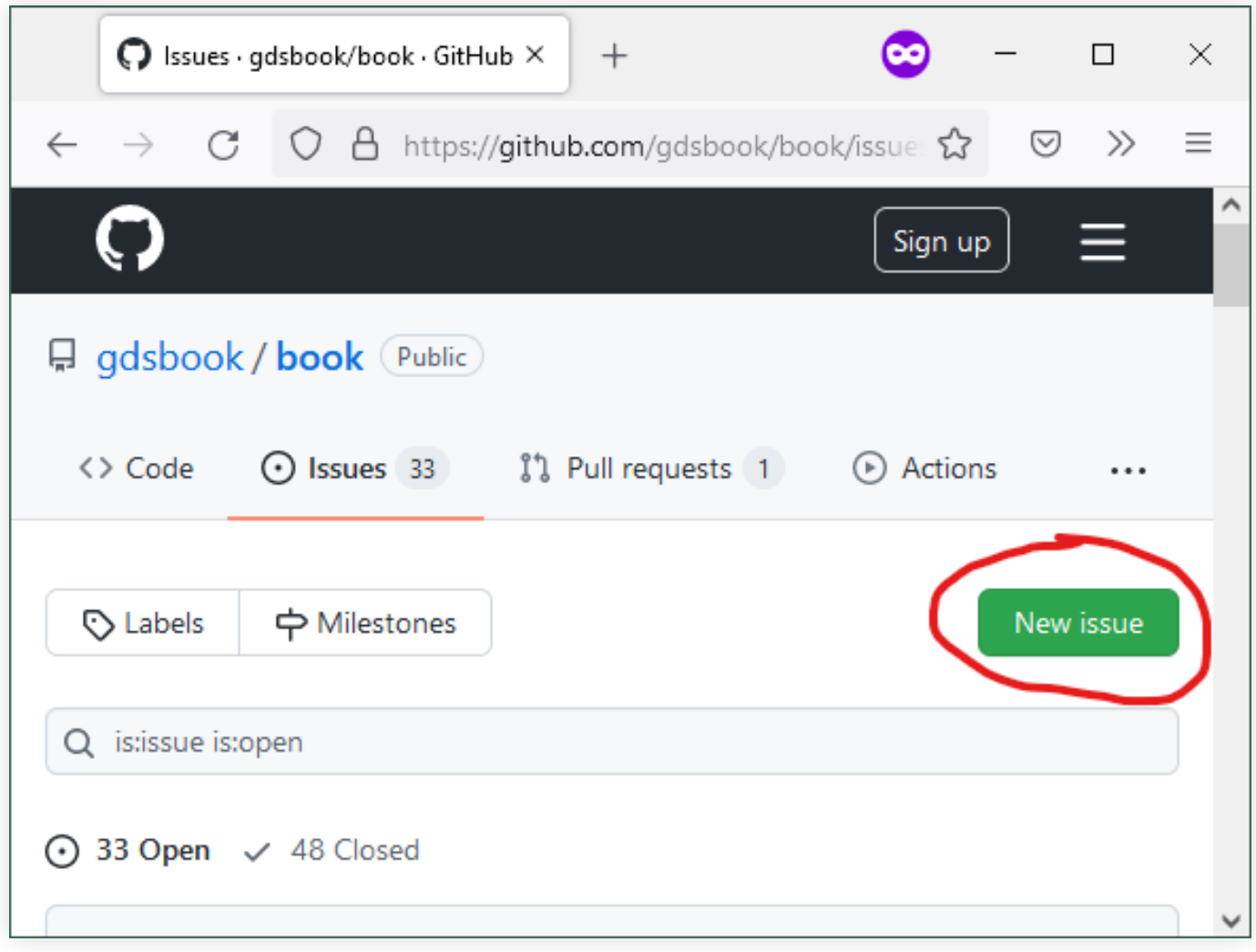



gdsl-ul.github.io/soft_install

# Try it out…


Spatial Feature Engineering —
https://geographicdata.sci
90%
Launch Binder
Binder
Colab
Spatial Feature Engineering
Contents
What is spatial feature engineering?
Feature Engineering Using Map Matching
Feature Engineering using Map Synthesis
Conclusion
Questions
Geographic Data Science with Python
Search this book
Home
PREFACE
Table of Contents
References
In machine learning and data science, we are often equipped with *tons* of data. Indeed, given the constellation of packages to query data services, free and open source data sets, and the rapid and persistent collection of geographical data, there is simply too much data to even represent coherently in a single, tidy fashion. However, we often need to be able to construct useful *features* from this rich and deep sea of data.
Where data is available, but not yet directly *usable*, *feature*
https://mybinder.org/v2/gh/gdsbook/book/master?urlpath=lab/tree/notebooks/12_feature_engineering.ipynb

# . . .and make it better!!!




`github.com/gdsbook/book/issues/new`

# Building "Geographic Data Science…"

Dani Arribas-Bel [@darribas]

UNIVERSITY OF LIVERPOOL

The Alan Turing Institute

Geographic Data Science Lab

[PDF version of these slides]

These slides are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.