

Building “Geographic Data Science...”

Dani Arribas-Bel [[@darribas](https://twitter.com/darribas)]



UNIVERSITY OF
LIVERPOOL

The
Alan Turing
Institute



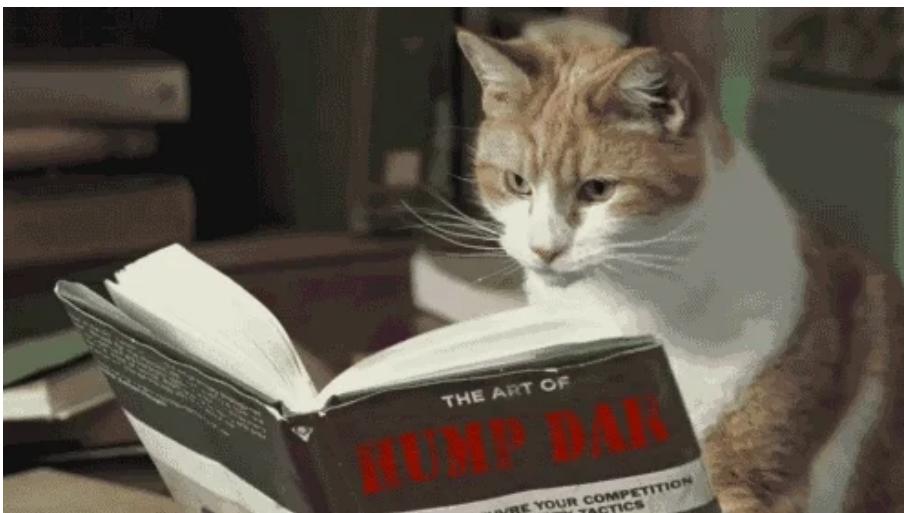
Geographic
Data Science
Lab

This talk



darribas.org/gdsbook_overview/202110

We have a book!



via GIPHY

Almost...



I am this close

via GIPHY

Coming “soon” but...

Coming “soon” but...

... you can already:

- <https://geographicdata.science>
- <https://github.com/gdsbook/book>



The Authors



@sreyog

Serge Rey



@darribas

Dani
Arribas-Bel



@levijohnwolf

Levi Wolf

Today

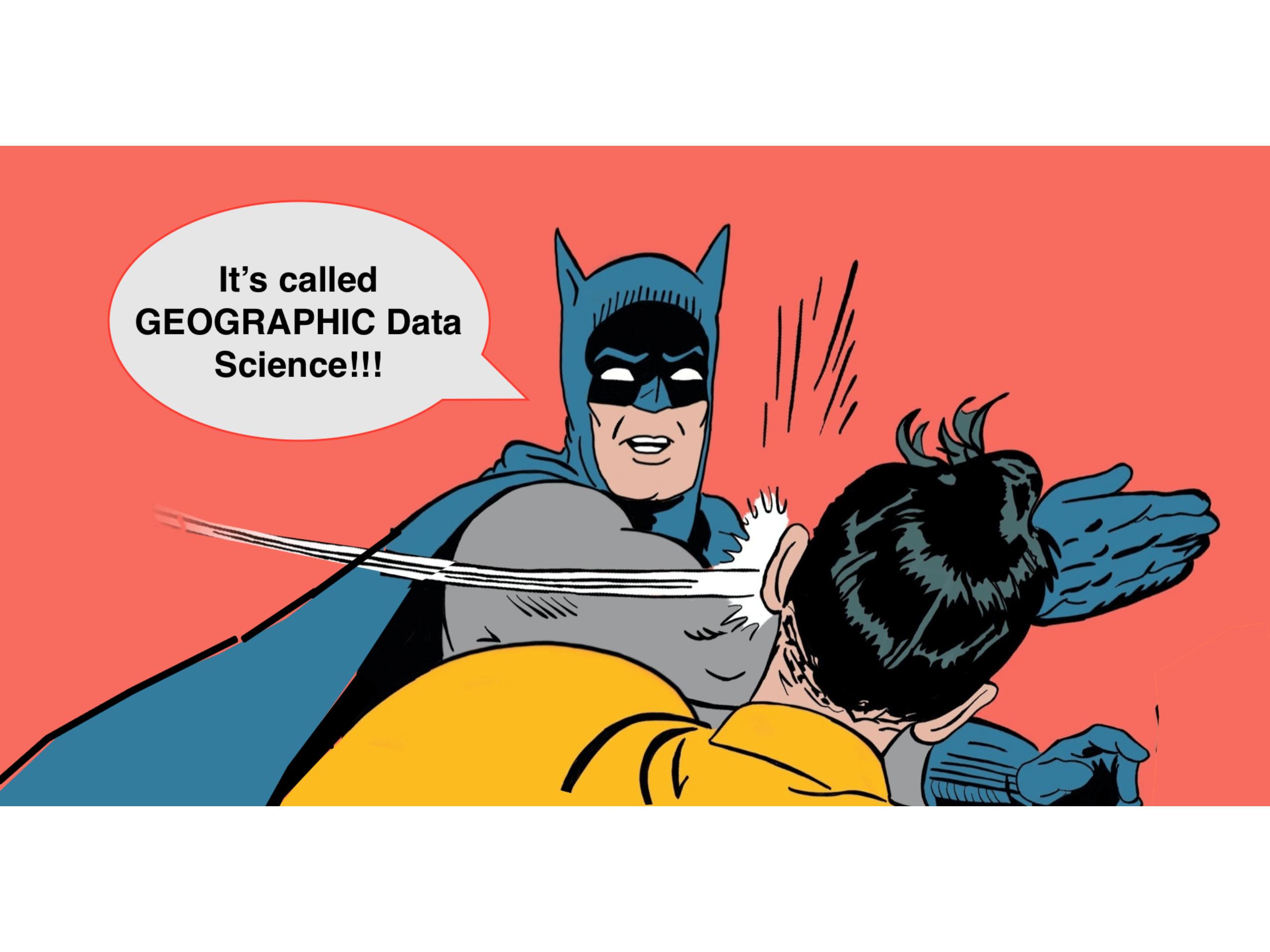
- *Why?* - Vision
- *What?* - Book
- *How?* - Community

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.

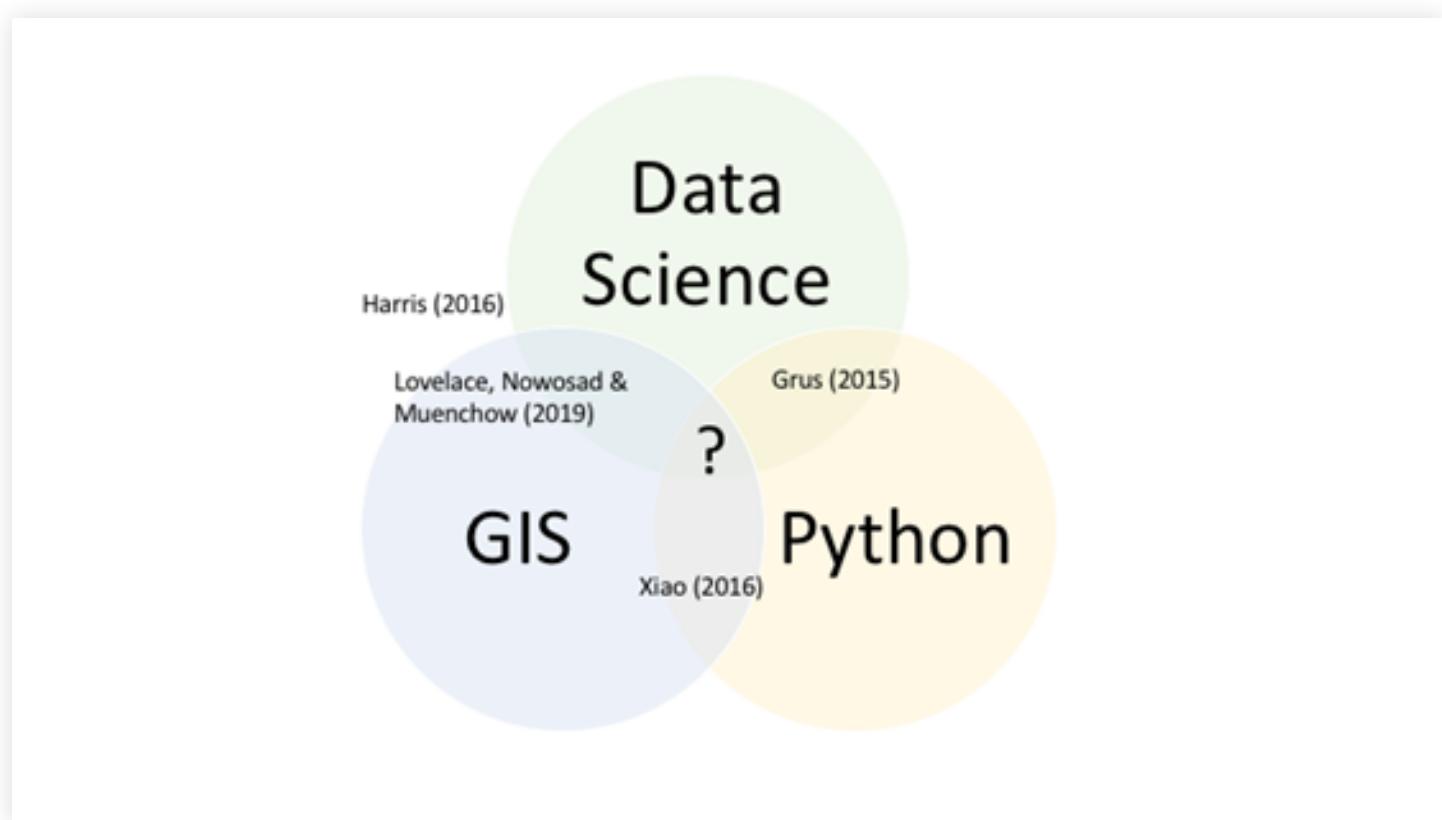


Geographic Data Science

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

What?

The Book is...

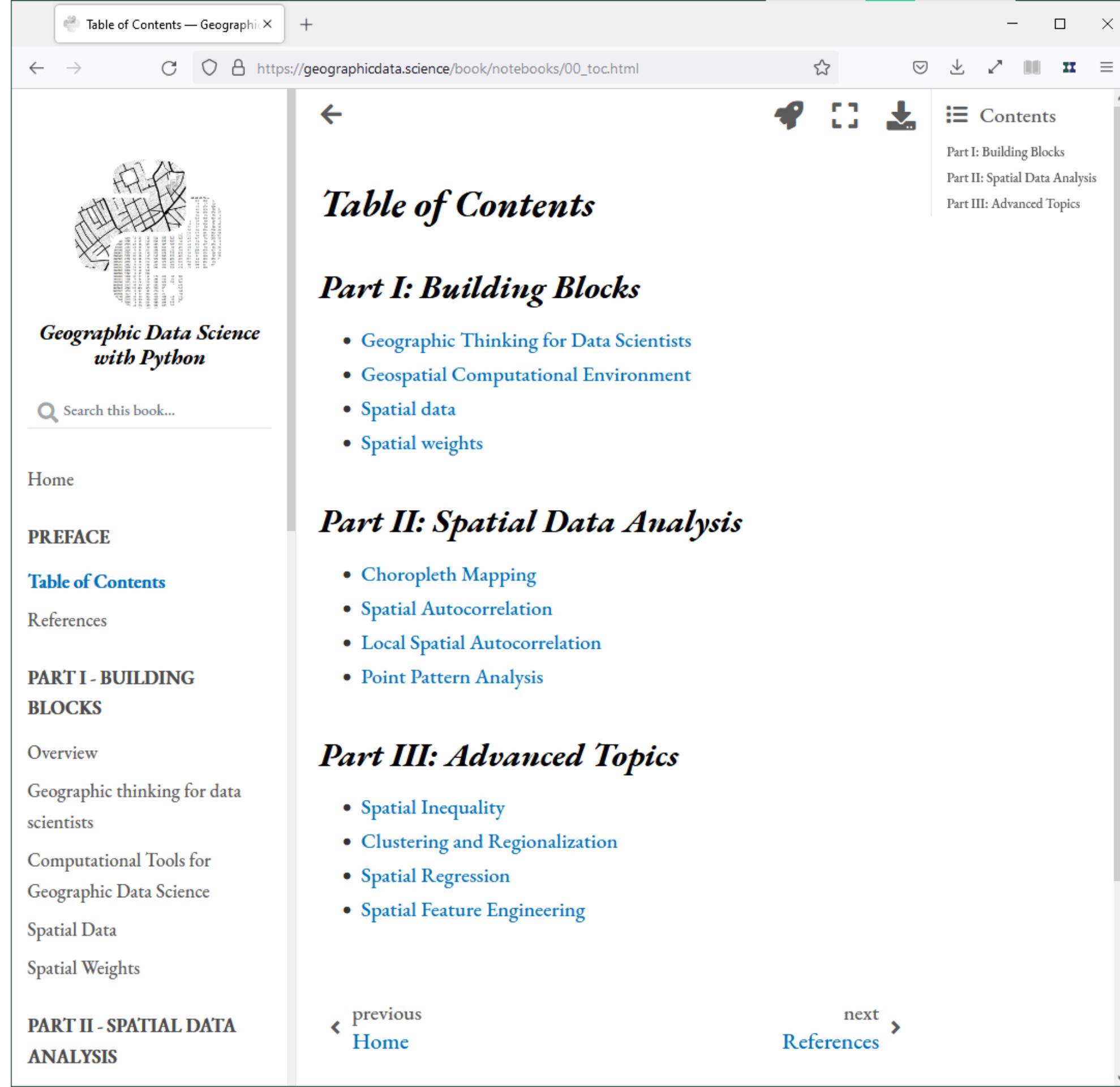


The Book is *not*...

- 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



GIScience + Data Science

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

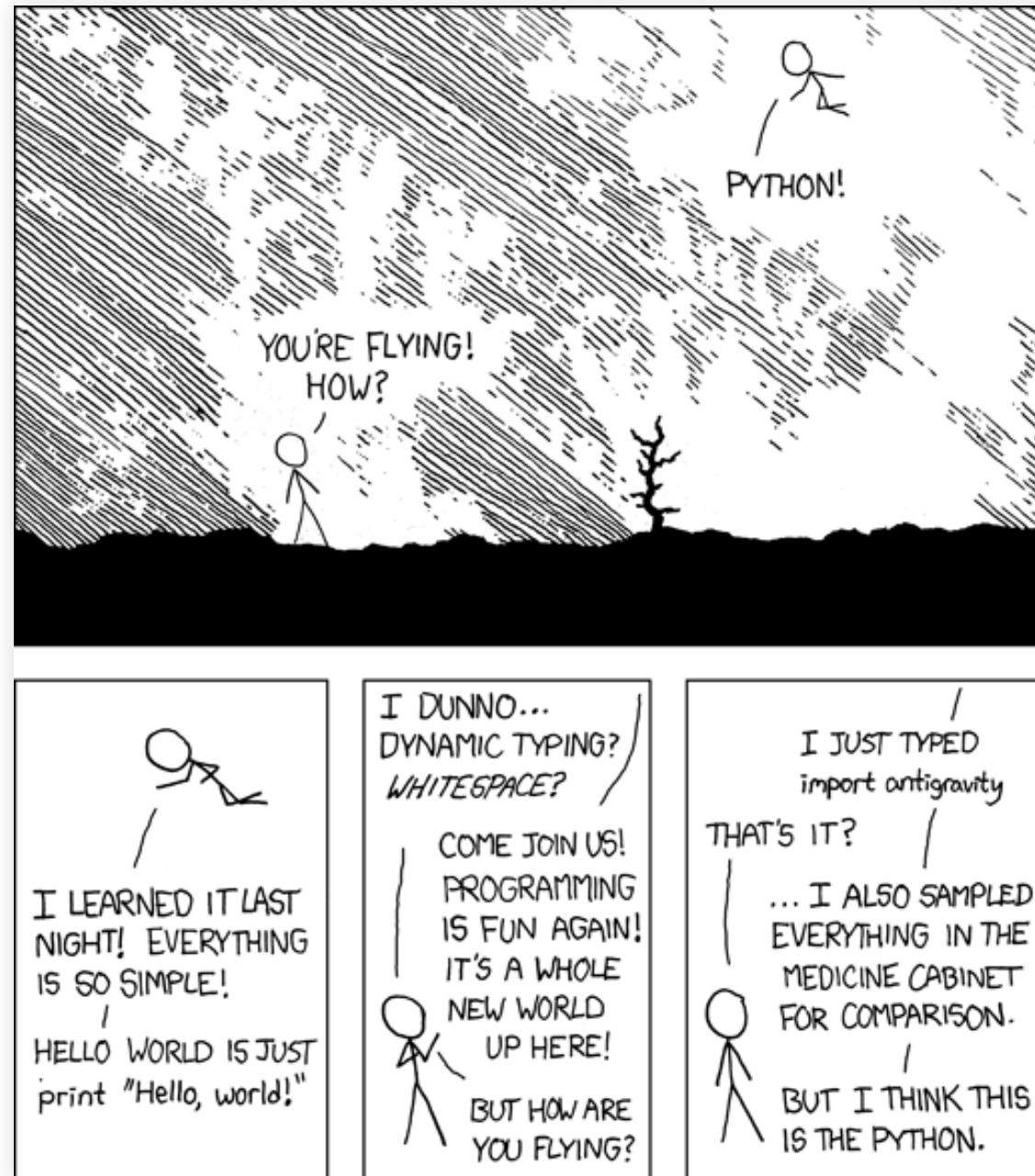
Bonus: Datasets

The screenshot displays three separate browser windows, each representing a different dataset on the DataHub platform:

- AirBnb**: Shows a list of download options for Airbnb data, including "Import requests", "Import parquet as zip", "Import parquet as csv", "Import parquet as sql", and "Import parquet as json".
- Airports**: Shows a preview of the "airports" dataset. The code block shows how to import the data from various URLs. The preview table includes columns: source_id, featurecla, type, name, abbrev, location, gis_code, lat, lon.
- Brexit dataset**: Shows a preview of the "brexit_vote" dataset. The code block shows how to import the data from a URL. The preview table includes columns: source_id, featurecla, type, name, abbrev, location, gis_code, lat, lon.

How?

Python



Source

Radically Open

Welcome!

Geographic Data Science with PySAL and the PyData Stack

This is the site for the book "Geographic Data Science with PySAL and the PyData Stack", by Sergio J. Rey, Dani Arribas-Bel and Levi J. Wolf. Here you can find out about the latest news regarding the book, read more about the [authors](#), or jump straight to the [book](#).

Latest news

Oct 30, 2020
Geographic data science on the road!

Apr 21, 2020
New Data Section available online

Aug 29, 2019
Work in progress...

Aug 24, 2019
Hello world!!!

Subscribe

Geographic Data Science with PySAL and the PyData Stack This book serves as an introduction to a whole new way of thinking systematically about geographic data, using geographical analysis and computation to unlock new insights hidden within data.

Sergio J. Rey, Dani Arribas-Bel & Levi J. Wolf
geographicdatascience@gmail.com

[Home](#) [Authors](#) [Book](#)

gdsbook / book

Code Issues Pull requests Actions Projects Security Insights

master

actions-user GA build of book HTML 12 days ago 341

github Change GA commit message to clarify further 2 months ago

data Fix data downloads and update gds_env for ... 2 months ago

docs GA build of book HTML 12 days ago

figures Complete draft of ch01 11 months ago

infrastructure Fix titles so they look OK on side TOC 2 months ago

notebooks GA build of book HTML 12 days ago

.gitignore Ignore NASADEM.hgt files 8 months ago

.nojekyll Create .nojekyll 2 years ago

Dockerfile Removed paired markdowns on Binder setup 2 months ago

LICENSE restructure of infrastructure to move to new JB 2 months ago

Makefile Remove CNAME (taken care of in parent rep...) 2 months ago

README.md Update README.md 15 months ago

appveyor.yml add substance and infrastructure 2 years ago

About

This book serves as an introduction to a whole new way of thinking systematically about geographic data, using geographical analysis and computation to unlock new insights hidden within data.

geographicdata.science

data-science data-analysis-python geographical-information-system geographic-data spatial-analysis spatial-statistics statistics spatial-data-analysis

Readme View license

Commits · gdsbook / book · GitHub

Code Issues Pull requests Actions Projects Security Insights

master

Commits on Nov 6, 2020

- GA build of book HTML actions-user committed 12 days ago ✓ 7375dca
- Merge pull request #102 from ljjwolf/ch2 darribas committed 12 days ago ✓ 4762249
- update spatial data chapter with additional structure ljjwolf committed 12 days ago 63adeb9

Commits on Oct 21, 2020

- GA build of book HTML actions-user committed 29 days ago ✓ d8631b1
- Merge pull request #100 from darribas/master ljjwolf committed 29 days ago ✓ d24db1a
- GA build of book HTML actions-user committed 29 days ago 0400d49
- Ch.12: Remove DEM code to go to blog: fix header hierarchy for cluste... darribas committed 29 days ago 6bbfacb

Commits on Oct 16, 2020

- add sketch of spatial data chapter ljjwolf committed on Oct 16 e67d876

Commits on Sep 21, 2020

- GA build of book HTML actions-user committed on Sep 21 310cad1
- Add refs page darribas committed on Sep 21 39d1da4
- interim build

Code as text; text as codE

The screenshot shows a Jupyter notebook interface. On the left, there's a sidebar with a map icon and the title "Geographic Data Science with Python". Below it are sections for Home, PREFACE, PART I - BUILDING BLOCKS, PART II - SPATIAL DATA ANALYSIS, PART III - ADVANCED TOPICS, and DATASETS. The main area contains a code cell and its output. The code uses the descartes library to plot an alpha shape and a convex hull over a map of Tokyo. The output shows several red circles representing the alpha shape, a blue dashed polygon for the convex hull, and black dots for source points. A legend in the bottom right corner identifies the symbols.

```
from descartes import PolygonPatch # to plot the alpha shape easily
f,ax = plt.subplots(1,1, figsize=(9,9))

# Plot a green alpha shape
ax.add_patch(PolygonPatch(alpha_shape, edgecolor='green',
                           facecolor='green', alpha=.2, label = 'Tighest alpha shape'))

# Include the points for our prolific user in black
ax.scatter(coordinates_t, color='k', marker='.', label='Source Points')
# Add a basemap
ax.imshow(basemap, extent=basemap_extent)

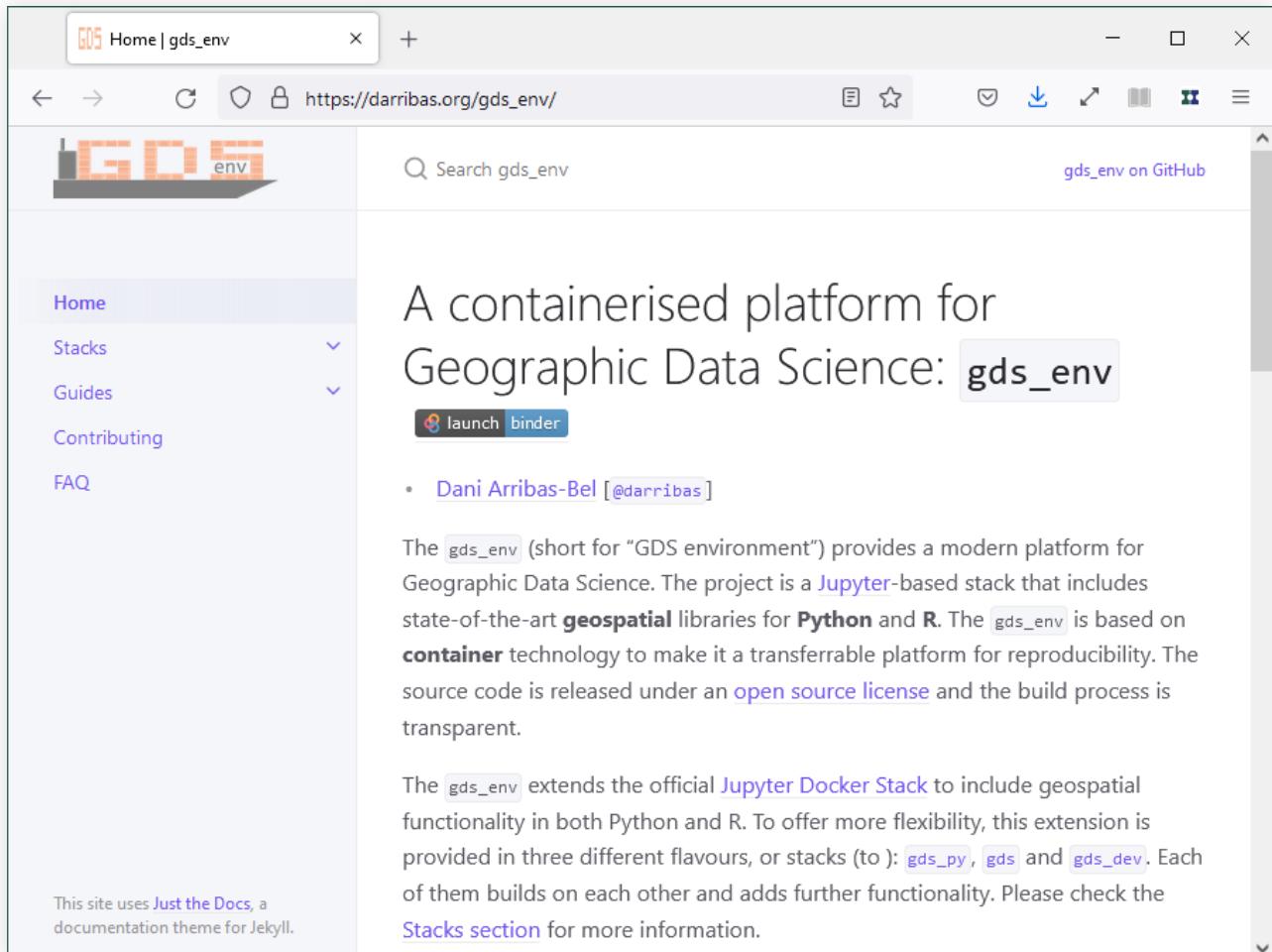
# plot the circles forming the boundary of the alpha shape
for i, circle in enumerate(circles):
    # only label the first circle of its kind
    if i == 0:
        label = 'Bounding Circles'
    else:
        label = None
    ax.add_patch(pit.Circle(circle, radius=alpha, facecolor='none', edgecolor='r', label=label))

# add a blue convex hull
ax.add_patch(pit.Polygon(convex_hull_vertices,
                        closed=True,
                        edgecolor='blue', facecolor='none',
                        linestyle=':', linewidth=2,
                        label='Convex Hull'))

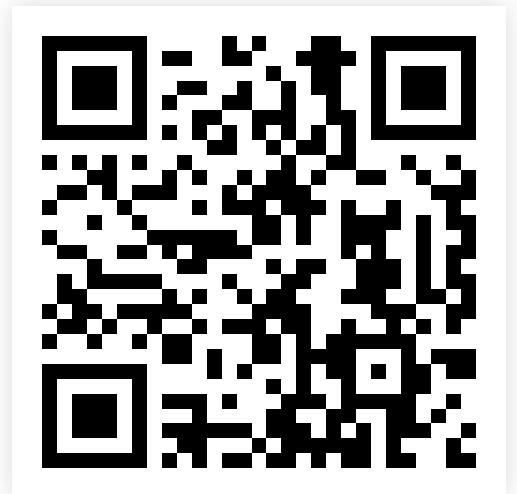
plt.legend()
```

The screenshot shows a GitHub pull request merge page. The URL is https://github.com/gdsbook/book/runs/1285952518?cl. The page displays a list of actions for a merge pull request from darribas/master. One action, "build-html-and-deploy", is highlighted in blue and has a status of "succeeded 29 days ago in 2m 50s". Other actions listed include "Set up job", "Checkout", "Setup Miniconda", "Build HTML", "Commit files", "Push changes", "Post Setup Miniconda", "Post Checkout", and "Complete job". The GitHub interface includes a header with navigation links like Why GitHub?, Team, Enterprise, Explore, Marketplace, Pricing, and a sidebar with contents like Introduction, Location, Location, Location, Visualization, and Centrography.

Runs *anywhere...*



A screenshot of a web browser window displaying the [gds_env](https://darribas.org/gds_env/) documentation. The page features a sidebar with links to Home, Stacks, Guides, Contributing, and FAQ. The main content area includes a search bar, a GitHub link, and a "launch binder" button. It describes the project as a containerised platform for Geographic Data Science, mentioning Python and R stacks, and extends the official Jupyter Docker Stack to include geospatial functionality. A QR code is visible on the right side of the page.



darribas.org/gds_env

... by anyone

Software Installation Guide

Installation Guide

Purpose

Mac OS Installation

- Installation
- Running Python

Linux Installation

- Installation
- Running Python

Windows 10 Pro/Education

- Installation
- Running Python

Windows 10 Home/pre-10

- Installation
- Running Python

Windows Specifications

Contributors

References

Published with bookdown

https://gdsl-ul.github.io/soft_install/

Software Installation Guide

Francisco Rowe, Dani Arribas-Bel
2021-06-17

Purpose

This resource provides step-by-step descriptions on how to install and run Python for Geographic Data Science from your own computer.

Select your Operating System and follow the steps.

What is your operating system?

```
graph TD; MacOS --> A1[Choose A1]; Linux --> A2[Choose A2]; Windows --> A2; A2 --> A2p[Which version?  
See below on how to check your version]; A2p --> A3[Choose A3]; A2p --> A4[Choose A4]; A3 --> A3p[Windows Pro]; A4 --> A4p[Windows Home or pre-10]
```



gdsl-ul.github.io/soft_install

Try it out...

Spatial Feature Engineering — [X](#)

https://geographicdata.science

90%

...

Binder

Colab

Spatial Feature Engineering

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 engineering* helps to construct useful data for modelling.

Geographic Data Science with Python

Search this book

Home

PREFACE

Table of Contents

References

What is spatial feature engineering?

Feature Engineering Using Map Matching

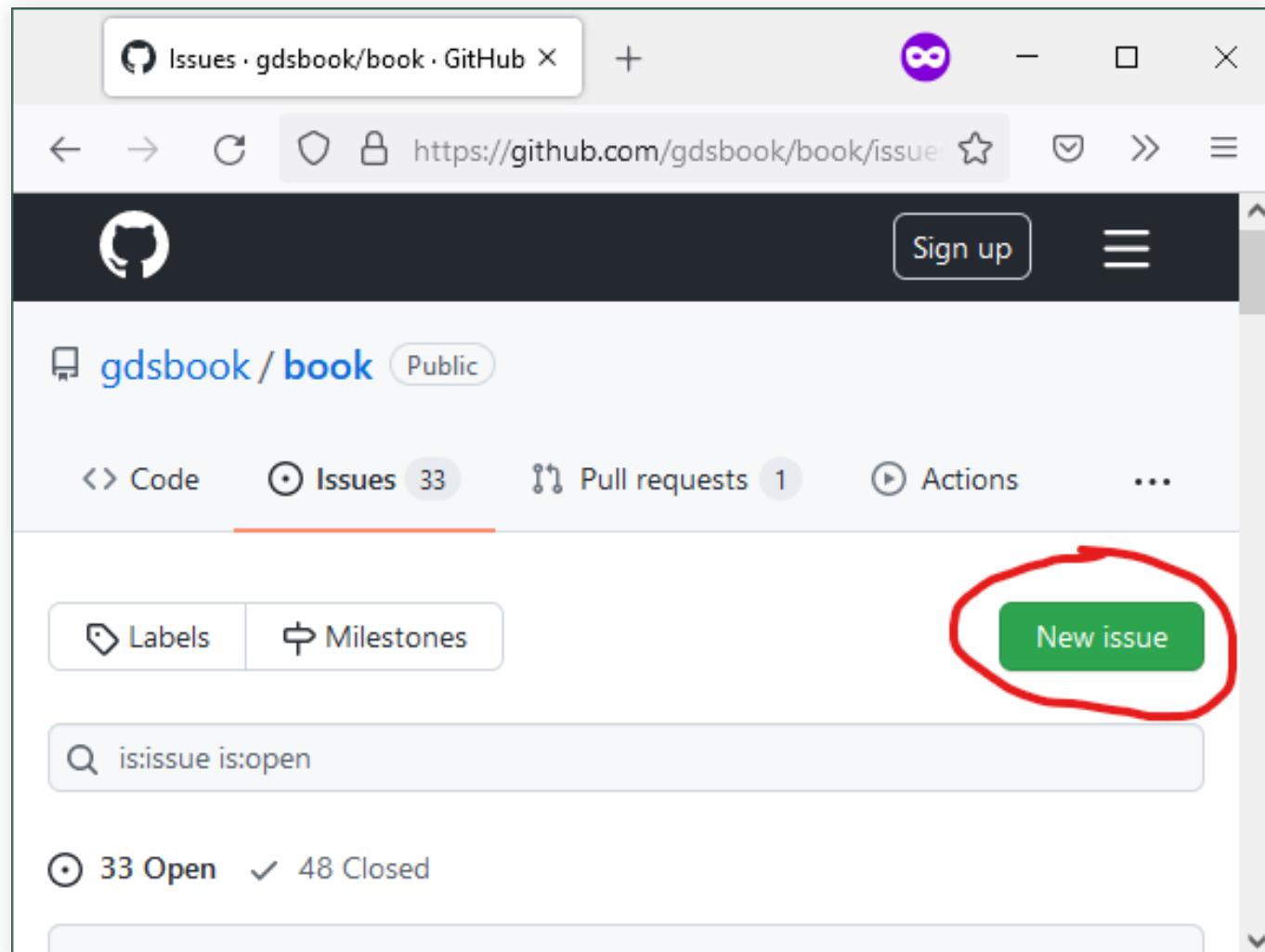
Feature Engineering using Map Synthesis

Conclusion

Questions

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.