



M1b: Tools

DSE 10200: Introduction to Data Science

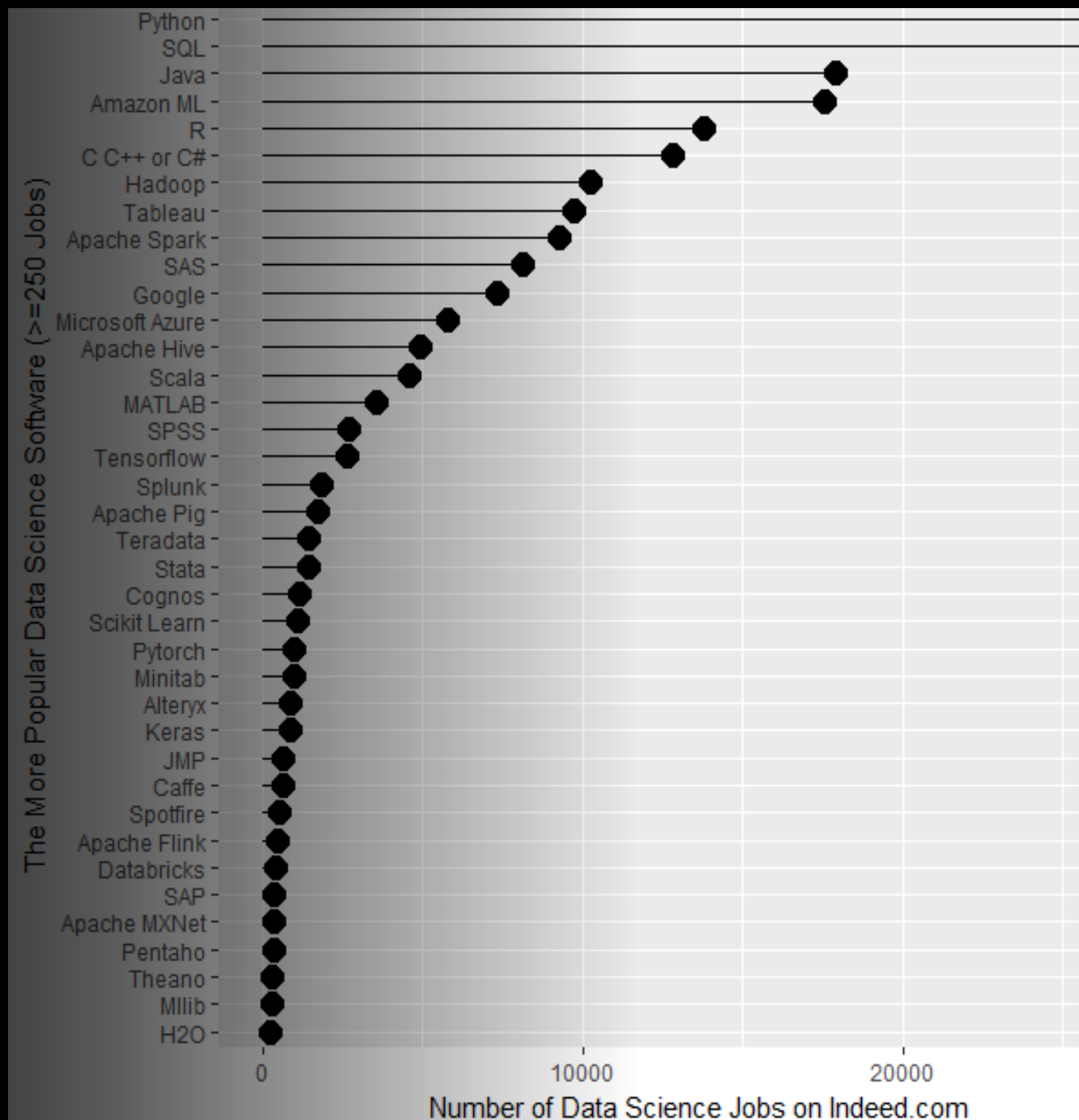
Instructor: Michael Grossberg



python

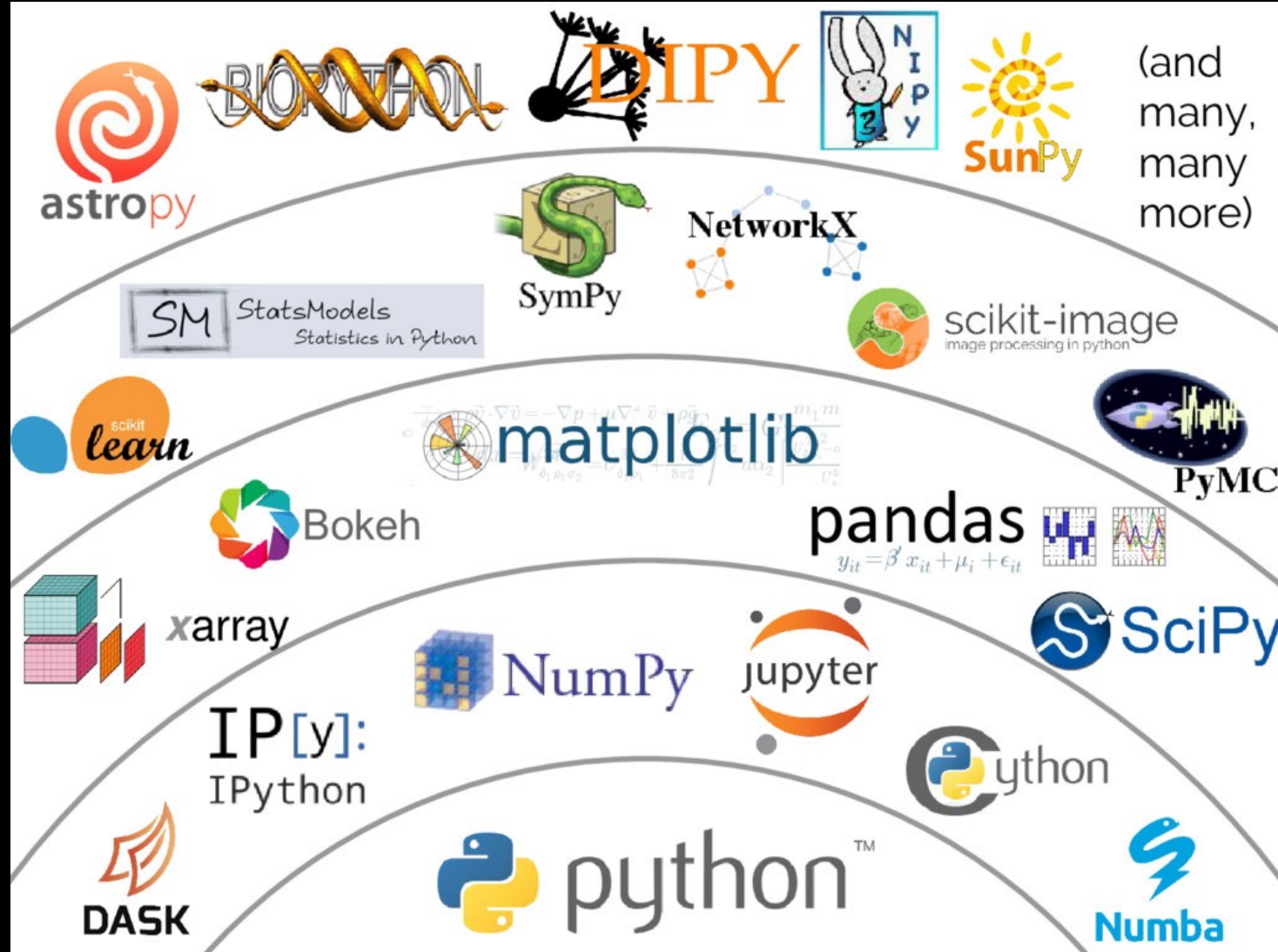
Why Python?

Python Most Popular Data Science Language



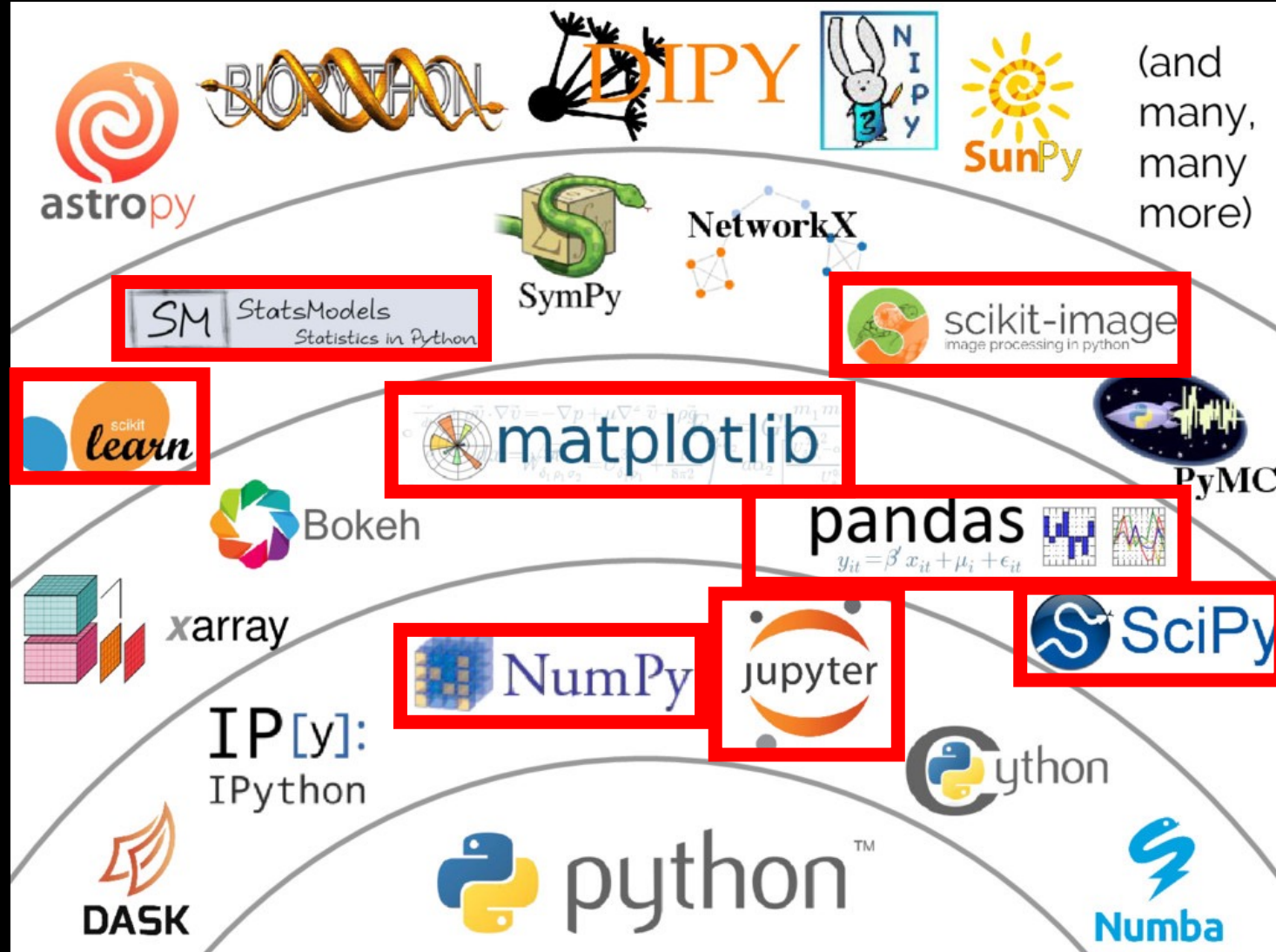
Python most
popular
programming
language





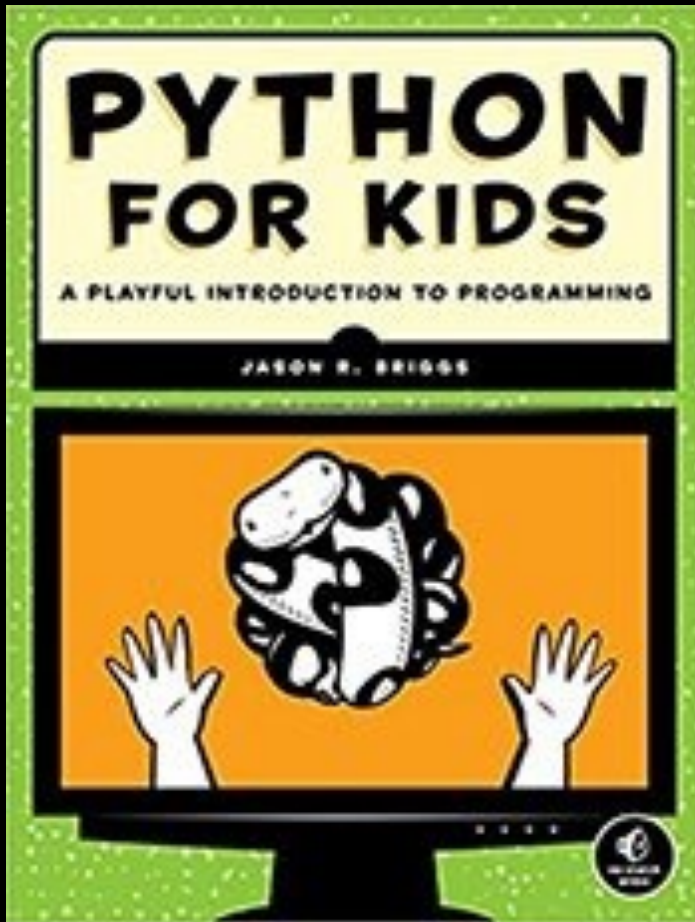
Python as MATLAB Replacement

Libraries
like MATLAB toolboxes



Python as MATLAB Replacement

Libraries
like MATLAB toolboxes

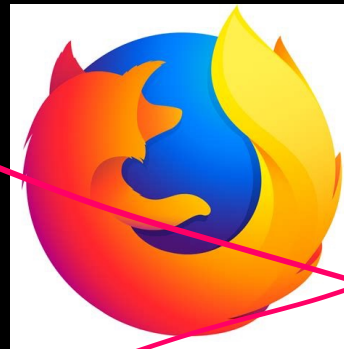


Taught in elementary, middle schools and high schools

Python Easy (easier)
to learn

Is python the best language for everything?

Is python the best language for everything?



Mobile and Web => JavaScript Better


Python Pretty
Good at Most
Things



Math Library

[Install](#) [Documentation](#) [Learn](#) [Community](#) [About Us](#) [Contribute](#)

NumPy




The fundamental package for scientific computing with Python

[GET STARTED](#)

NumPy 1.23.0 released

<https://numpy.org/>

Pandas

 pandas

About us ▾ Getting started Documentation Community ▾ Contribute

pandas



pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool,
built on top of the Python programming language.

[Install pandas now!](#)

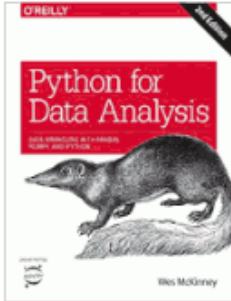
Latest version: 1.4.4

- What's new in 1.4.4
- Release date:
Aug 31, 2022
- Documentation (web)
- Download source code

Follow us



Get the book



Getting started

- Install pandas
- Getting started

Documentation

- User guide
- API reference
- Contributing to pandas
- Release notes





Community

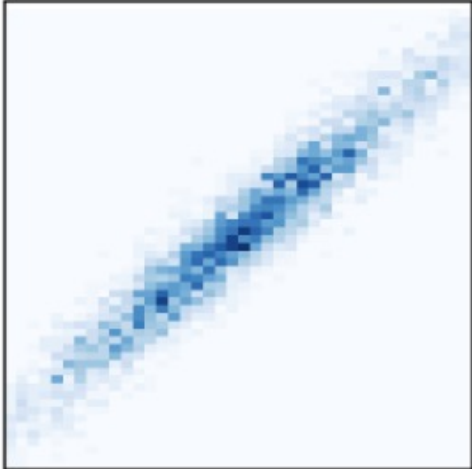
- About pandas
- Ask a question
- Ecosystem

With the support of:

<https://pandas.pydata.org/>

Matplotlib

 [Plot types](#) [Examples](#) [Tutorials](#) [Reference](#) [Usage guide](#) [Develop](#) [Release notes](#)   




hist2d(x, y)

Matplotlib: Visualization with Python


Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

- Create publication quality plots.
- Make interactive figures that can zoom, pan, update.
- Customize visual style and layout.
- Export to many file formats.
- Embed in JupyterLab and Graphical User Interfaces.
- Use a rich array of third-party packages built on Matplotlib.


Try Matplotlib (on Binder) →




Getting Started




Examples



Reference



Cheat Sheets

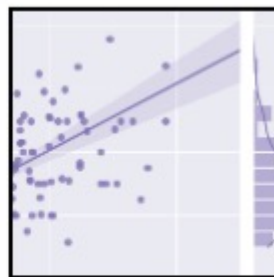
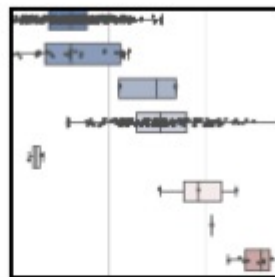
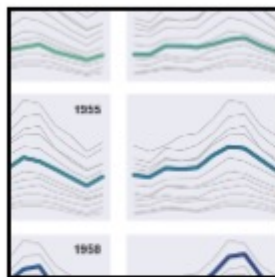
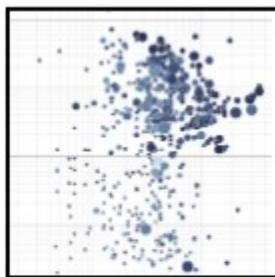
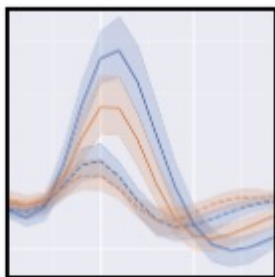
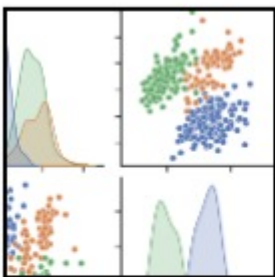


Documentation

<https://matplotlib.org/>

Seaborn

seaborn: statistical data visualization



Seaborn is a Python data visualization library based on [matplotlib](#). It provides a high-level interface for drawing attractive and informative statistical graphics.

For a brief introduction to the ideas behind the library, you can read the [introductory notes](#) or the [paper](#). Visit the [installation page](#) to see how you can download the package and get started with it. You can browse the [example gallery](#) to see some of the things that you can do with seaborn, and then check out the [tutorial](#) or [API reference](#) to find out how.

To see the code or report a bug, please visit the [GitHub repository](#). General


Contents

- [Introduction](#)
- [Release notes](#)
- [Installing](#)
- [Example gallery](#)
- [Tutorial](#)
- [API reference](#)

Features

- [Relational: API | Tutorial](#)
- [Distribution: API | Tutorial](#)
- [Categorical: API | Tutorial](#)
- [Regression: API | Tutorial](#)
- [Multiples: API | Tutorial](#)
- [Style: API | Tutorial](#)
- [Color: API | Tutorial](#)

Sklearn

 [Install](#) [User Guide](#) [API](#) [Examples](#) [Community](#) [More ▾](#)

scikit-learn

Machine Learning in Python

[Getting Started](#) [Release Highlights for 1.1](#) [GitHub](#)

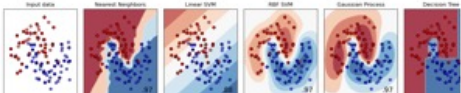
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: SVM, nearest neighbors, random forest, and more...




Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, nearest neighbors, random forest, and more...



Clustering

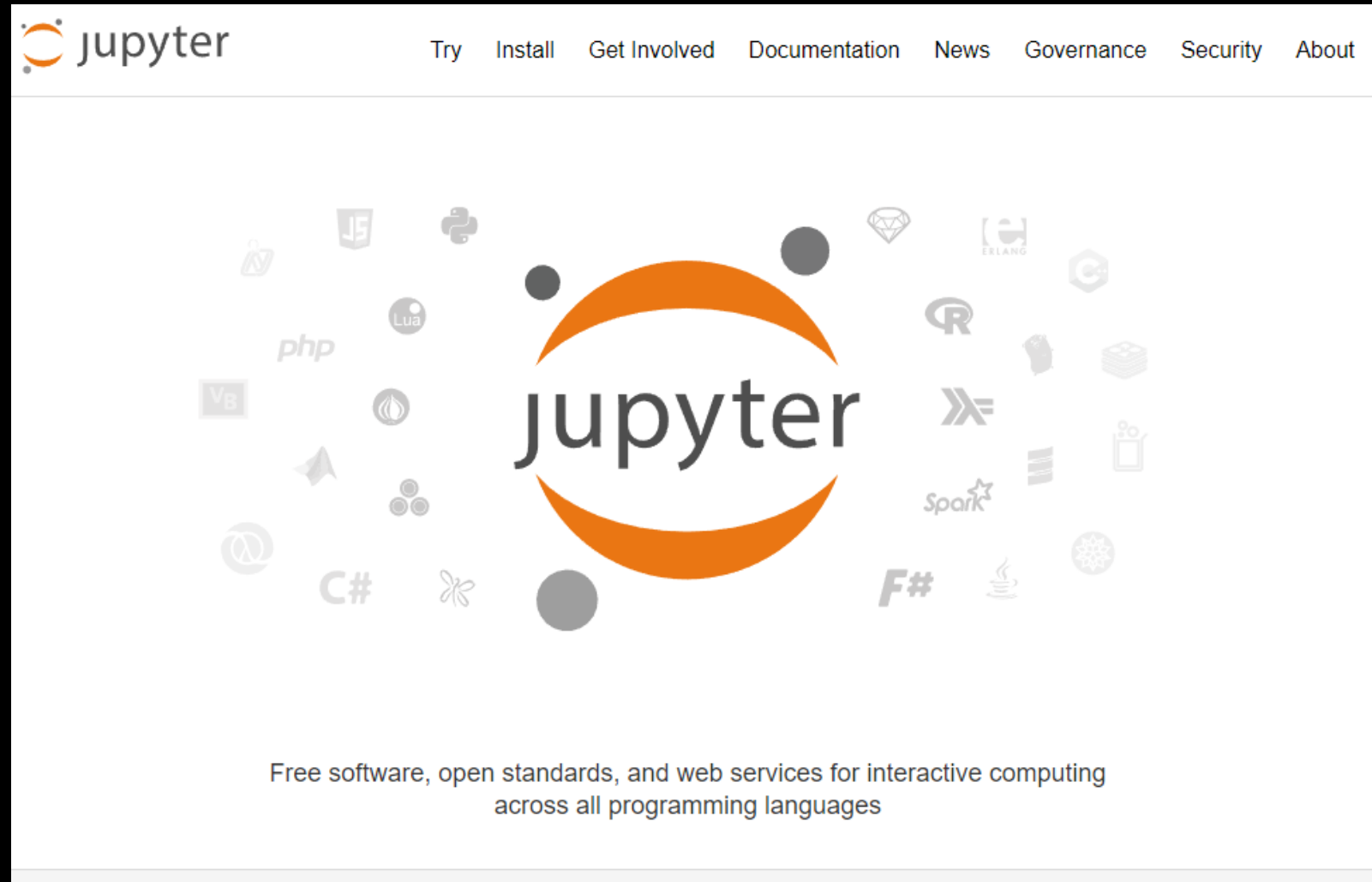
Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering, mean-shift, and more...

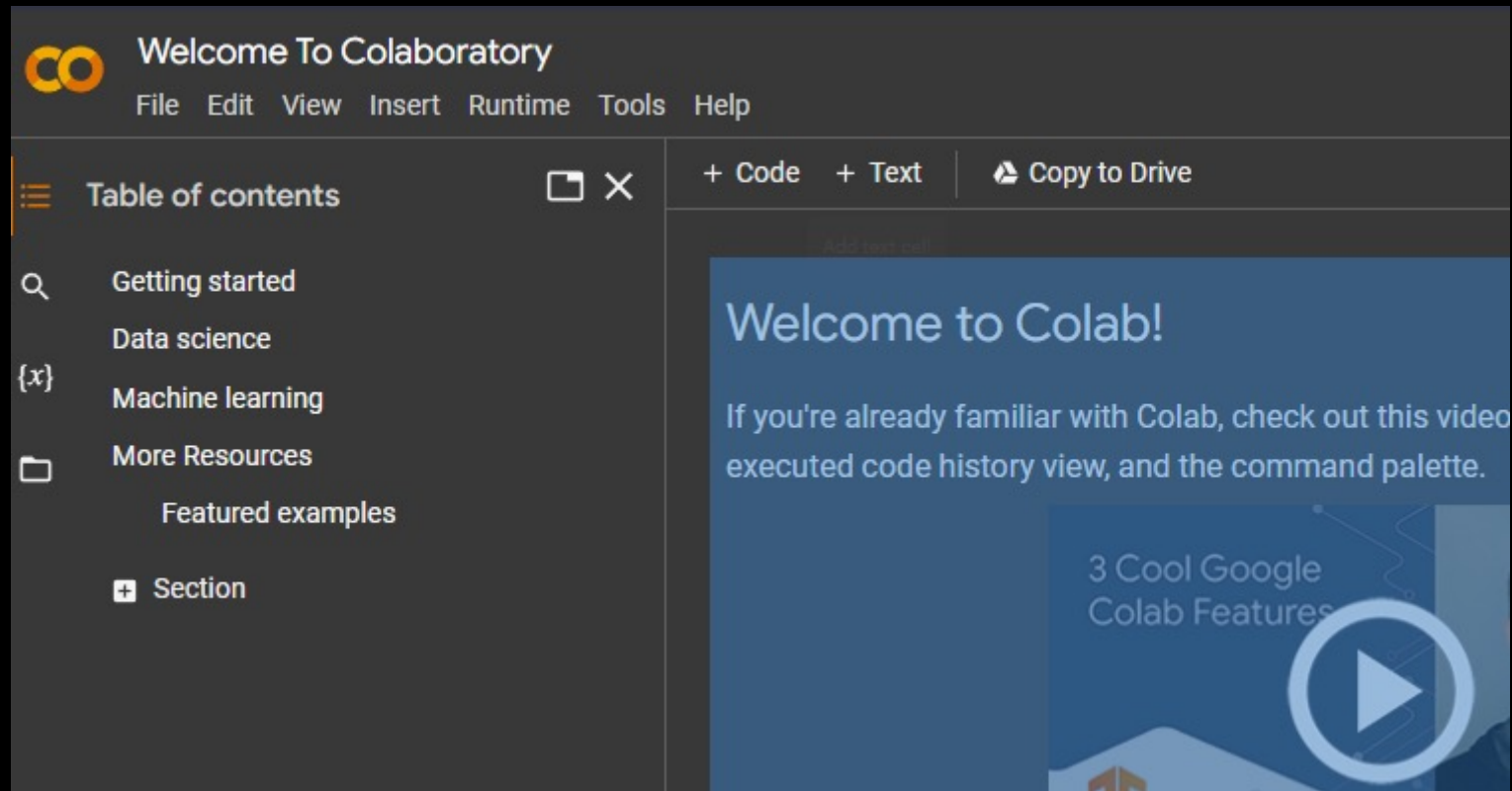


Jupyter



<https://jupyter.org/>

Google Colab



<https://colab.research.google.com/>

Anaconda

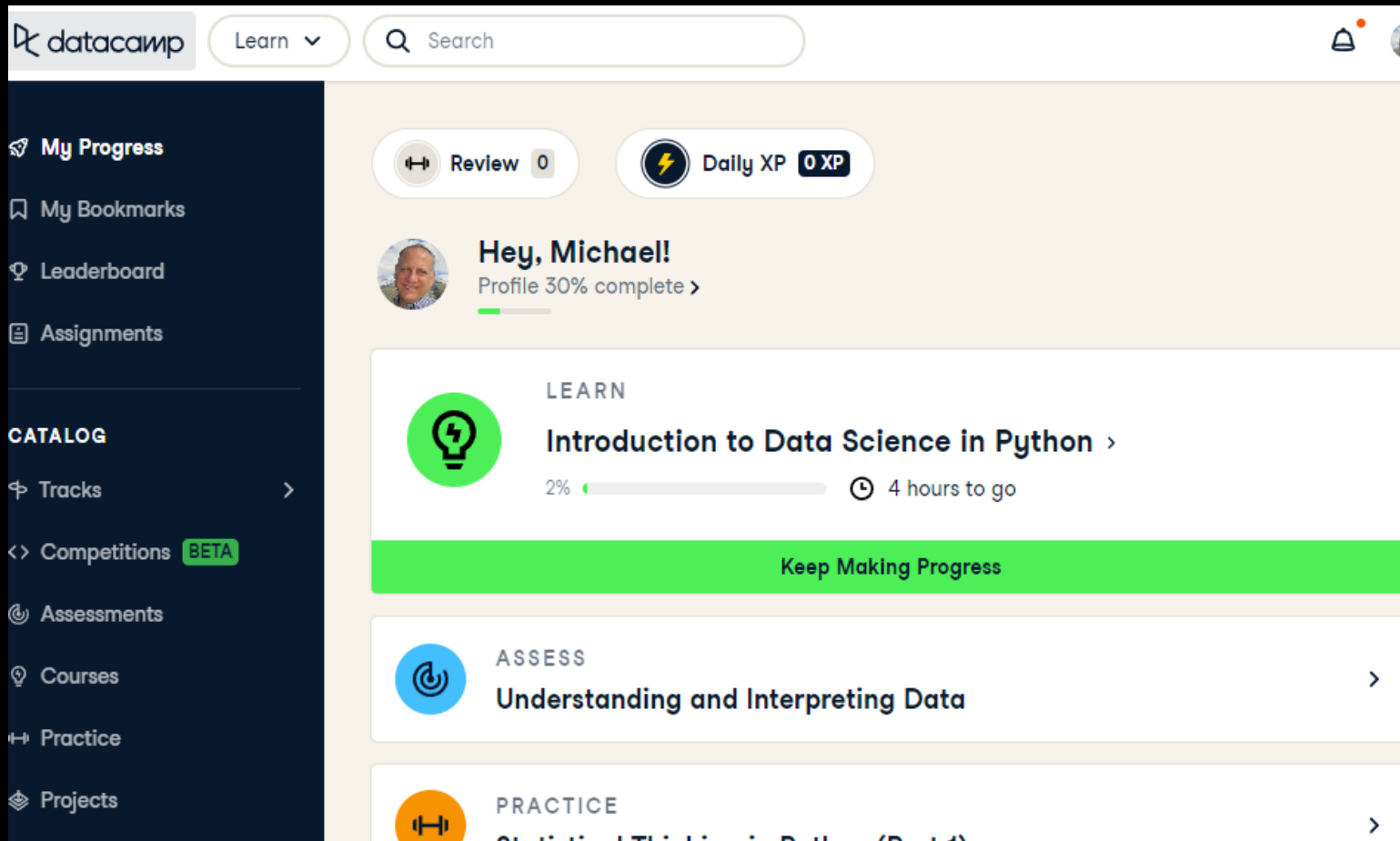


Data science technology for

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of

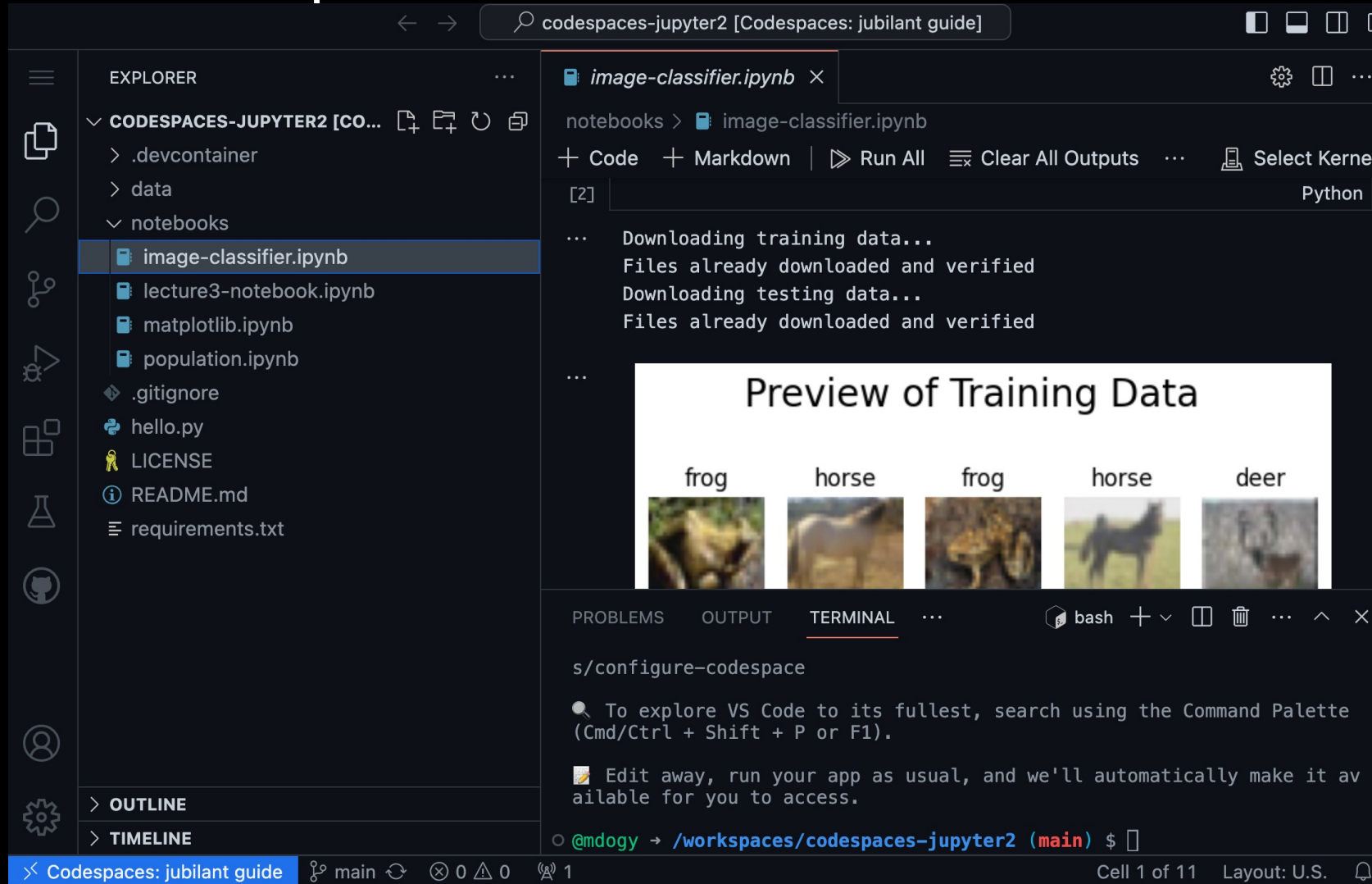
<https://www.anaconda.com/>

Data Camp



<https://app.datacamp.com/learn>

Github Codespaces



<https://docs.github.com/en/codespaces/getting-started/quickstart>