DATA VISUALIZATION IN PYTHON

Wintersession Workshop, 2022
PICSciE and OIT Research Computing
The Center for Statistics and Machine Learning
Princeton University
January 12, 2022

José A. Garrido Torres
Data Scientist (Schmidt DataX)
jagt@princeton.edu

PRINCETON UNIVERSITY

Princeton Research Computing

CENTER FOR STATISTICS AND MACHINE LEARNING

DataX

Our team

PRINCETON UNIVERSITY

CENTER FOR STATISTICS AND MACHINE LEARNING

DataX

Ellen DiPippo

Peter Ramadge

Sarah McGovern

Amy Winecoff

Andrzej Żurański

Brian Arnold

Jose Garrido

Pranay Anchuri

DataX

PRINCETON UNIVERSITY

Data Visualization in Python

- **Python** 2-D/3-D **plotting library** (virtually fully-customizable figures).

- **Create, analyze** and **manipulate** two-dimensional **data**.

- Plotting library based on matplotlib for **statistical graphics**.

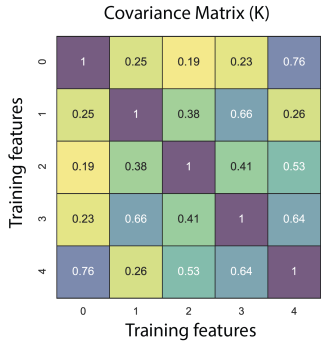
- Serves to create and deploy **interactive** data **visualizations**.

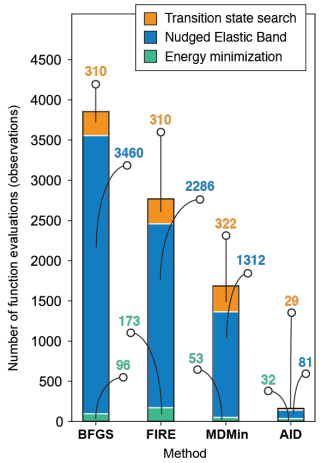
DataX

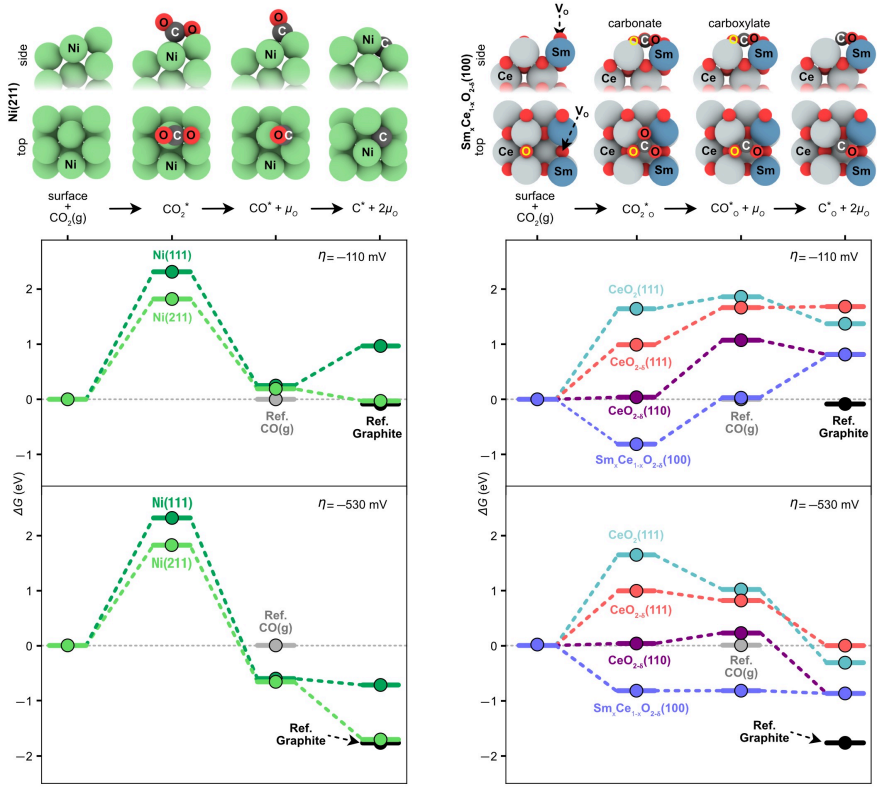
PRINCETON UNIVERSITY

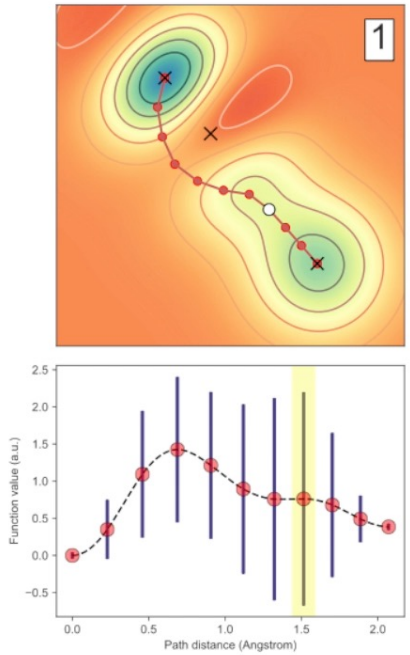
Motivation

- Engage the audience when presenting results through **attractive** and **effective** data **visualization**.









J. A. Garrido Torres, P. C. Jennings, M. H. Hansen, J. R. Boes, and T. Bligaard, Phys. Rev. Lett., **122**, *156001, 2019.*
T. L. Skafte, Z. Guan, M. Machala, C. Gopal, M. Monti, L. Martinez, E. Stamate, S. Sanna, J. A. Garrido Torres, E. Crumlin, M. Garcia-Melchor, M. Bajdich, W. Chueh, C. Graves, Nature Energy, **4**, *846–855, 2019.*

DataX
PRINCETON UNIVERSITY

Motivation

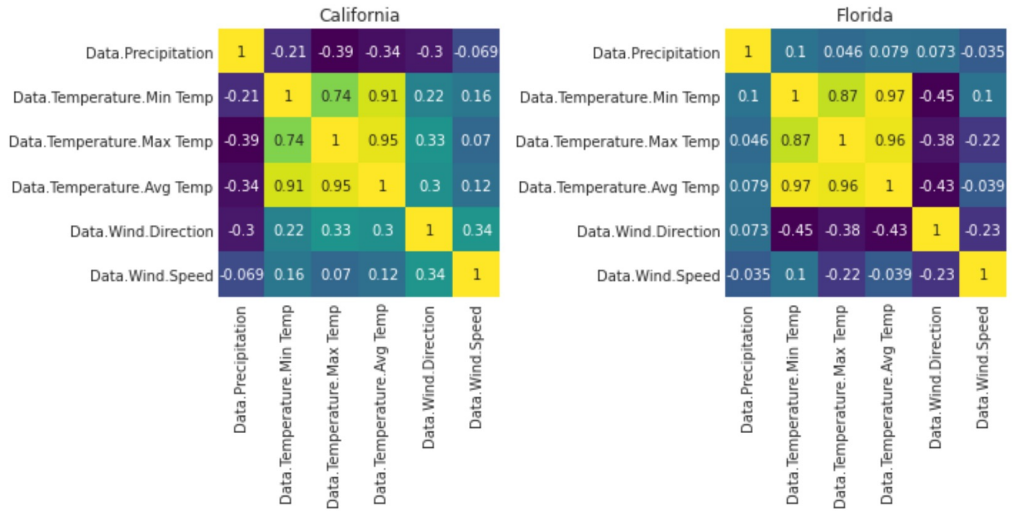
- **Scientific programming** helps overcoming **reproducibility** issues.

```
states = ['California', 'Washington', 'Florida', 'New York']

fig, ax = plt.subplots(nrows=1, ncols=len(states))
fig.set_size_inches(5 * len(states), 5)

for i in range(0 , len(states)):
  subset_i = df3[df['Station.State'] == states[i]]
  corr_i = subset_i.corr()
  sns.heatmap(corr_i, cmap='viridis', annot=True, ax=ax[i], cbar=False)
  ax[i].set_title(states[i])

plt.tight_layout()
```



DataX

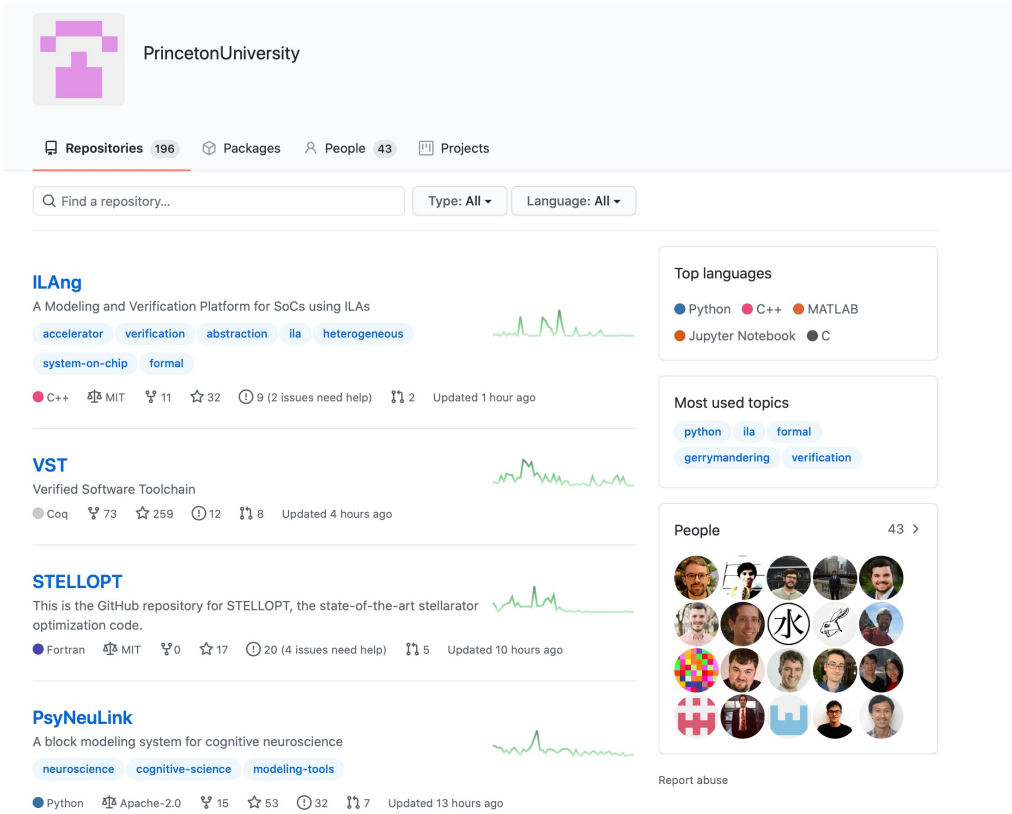
PRINCETON UNIVERSITY

Motivation

- Keeping **track** of the **changes** (version control) you make to your code.

GitHub

GitLab



DataX

PRINCETON UNIVERSITY

Motivation

- **Coding** is **fun**! It offers a large variety of tools to solve problems in an efficient manner.

DataX

PRINCETON UNIVERSITY

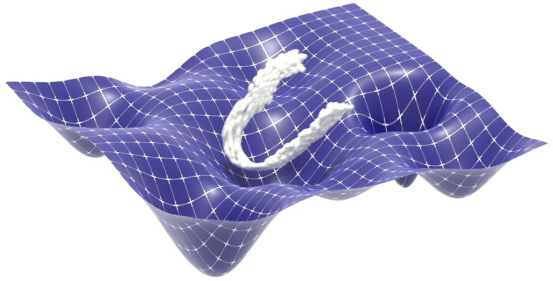
Motivation

- Engage the audience when presenting results through **attractive** and **effective** data **visualization**.
- **Scientific programming** helps overcoming **reproducibility** issues.
- Keeping **track** of the **changes** (version control) you make to your code.
- **Coding** is **fun**! It offers a large variety of tools to solve problems in an efficient manner.

DataX
PRINCETON UNIVERSITY

Motivation

- Engage the audience when presenting results through **attractive** and **effective** data **visualization**.
- **Scientific programming** helps overcoming **reproducibility** issues.
- Keeping **track** of the **changes** (version control) you make to your code.
- **Coding** is **fun**! It offers a large variety of tools to solve problems in an efficient manner.

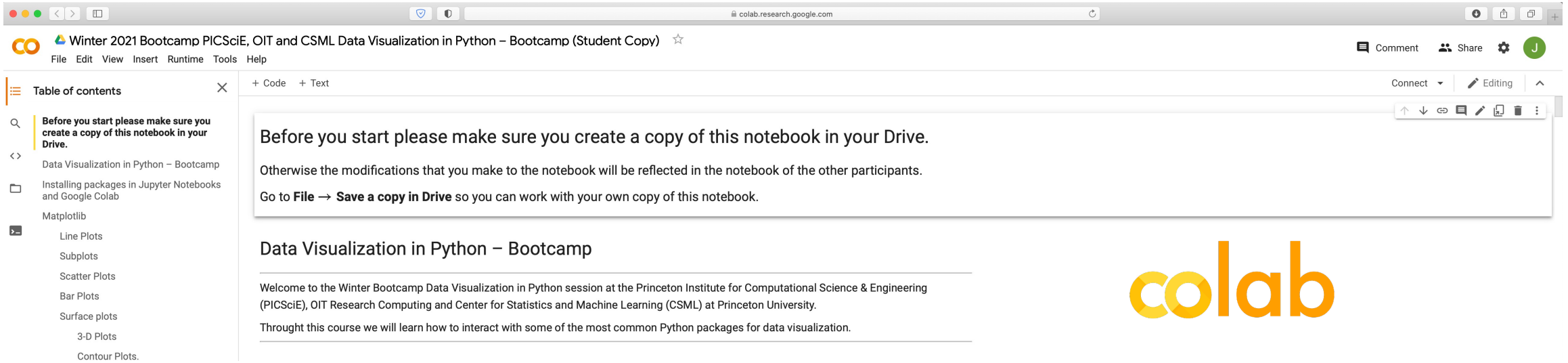


This is a review of my motivation for data visualization in Python but… how about you?!

DataX

PRINCETON UNIVERSITY

Framework of this session



- Notebook for the **Hands-on session** (including short-exercises).
- Notebook with **Exercises** (for practicing after the session).
- Notebook with the **Answers** for the previous Exercises.

DataX
PRINCETON UNIVERSITY