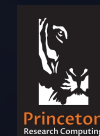


DATA VISUALIZATION IN PYTHON

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



Our team



**Ellen
DiPippo**



**Peter
Ramadge**



**Sarah
McGovern**



**Amy
Winecoff**



**Andrzej
Żurański**



**Brian
Arnold**



**Jose
Garrido**



**Pranay
Anchuri**



Data Visualization in Python



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



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



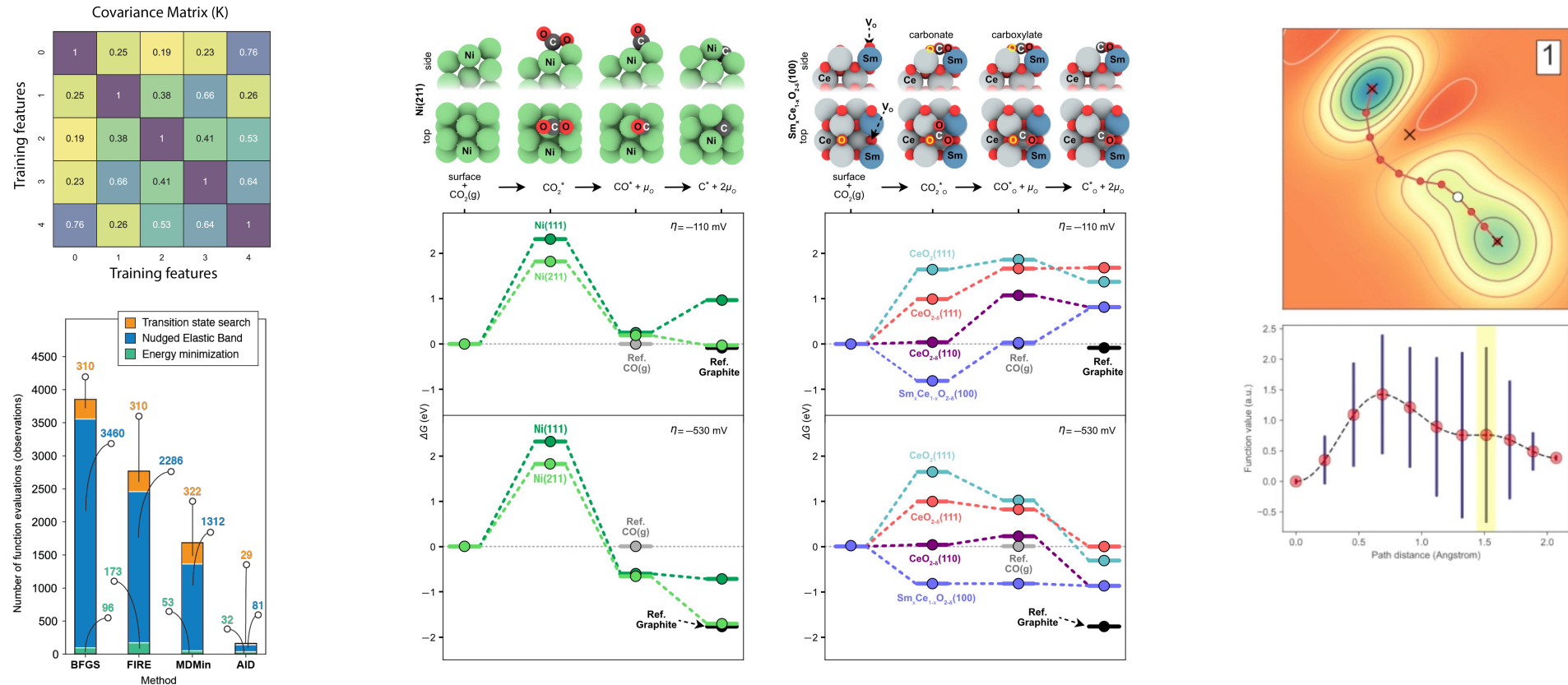
- Plotting library based on matplotlib for **statistical graphics**.



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

Motivation

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



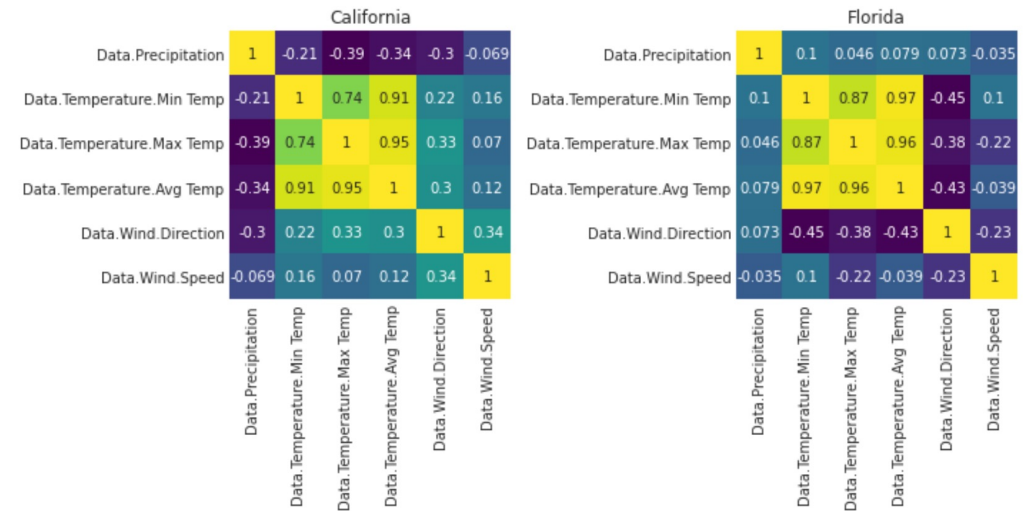
J. A. Garrido Torres, P. C. Jennings, M. H. Hansen, J. R. Boes, and T. Bligaard, *Phys. Rev. Lett.*, **122**, 156001, 2019.

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.

Motivation

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

```
1 states = ['California', 'Washington', 'Florida', 'New York']
2
3 fig, ax = plt.subplots(nrows=1, ncols=len(states))
4 fig.set_size_inches(5 * len(states), 5)
5
6 for i in range(0, len(states)):
7     subset_i = df3[df['Station.State'] == states[i]]
8     corr_i = subset_i.corr()
9     sns.heatmap(corr_i, cmap='viridis', annot=True, ax=ax[i], cbar=False)
10    ax[i].set_title(states[i])
11
12 plt.tight_layout()
```



Motivation

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



A screenshot of the GitHub profile page for PrincetonUniversity. The profile header shows the repository count (196), packages, people (43), and projects. Below the header is a search bar and filters for repository type and language. The main content area lists several repositories: ILAng (Modeling and Verification Platform for SoCs using ILAs), VST (Verified Software Toolchain), STELLOPT (state-of-the-art stellarator optimization code), and PsyNeuLink (block modeling system for cognitive neuroscience). Each repository entry includes its description, tags, statistics (stars, forks, issues), and a commit activity graph. On the right side, there are sections for "Top languages" (Python, C++, MATLAB, Jupyter Notebook, C), "Most used topics" (python, ila, formal, gerrymandering, verification), and "People" (43 contributors).

Motivation

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

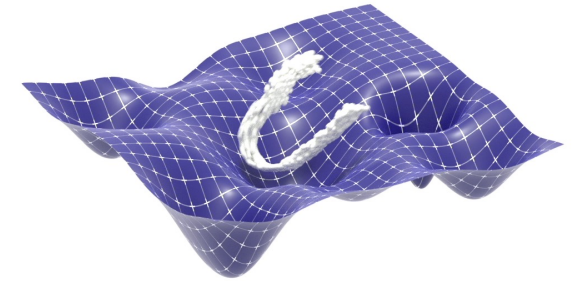


Motivation

- 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.

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?!**

Framework of this session

Before you start please make sure you create a copy of this notebook in your Drive.

Otherwise the modifications that you make to the notebook will be reflected in the notebook of the other participants.

Go to **File** → **Save a copy in Drive** so you can work with your own copy of this notebook.

Data Visualization in Python – Bootcamp

Welcome to the Winter Bootcamp Data Visualization in Python session at the Princeton Institute for Computational Science & Engineering (PICSciE), OIT Research Computing and Center for Statistics and Machine Learning (CSML) at Princeton University.

Through this course we will learn how to interact with some of the most common Python packages for data visualization.

- 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.