Python for data visualization

ETHAN YOO, DATA SCIENCE GRADUATE SPECIALIST

MARCH 21, 2023

ETHAN.YOO@RUTGERS.EDU

Outline

- Why Python?
- Terminology
- Installing packages
- Getting started with Matplotlib
- Getting started with Plotly
- Additional resources
- Demonstration

Why Python?

- Free and open source
- General-purpose programming language
- Interpreted language
- Use cases and existing applications
  - **Web development** (e.g., Django and Flask)
  - **GUI development** (e.g., PyQt)
  - **Scientific and numeric** (e.g., Pandas)
  - **Software development** (e.g., Trac)
  - **System administration** (e.g., Ansible)

Terminology

- Variable
- Function
- Return
- Script
- Modules and packages

Installing packages

Using pip

```
python -m pip install numpy
python -m pip install matplotlib plotly seaborn
```

Using conda

```
conda install numpy
conda install matplotlib plotly seaborn
```

Other useful packages

- Pandas (a required dependency for seaborn)
- JupyterLab

Plots in Matplotlib

- Basic: Two variables
  - `plot(x, y)`
  - `scatter(x, y)`
  - `bar(x, height)`
  - Others: Stacked area plots, stem plots, step plots, or fill the area between two horizontal curves
- Statistical plots
  - Others: Box and whisker plots, error bars, event or raster plots, violin plots,
- Other plot types: Arrays and fields, unstructured coordinates, 3D

Getting started with Matplotlib

```
import matplotlib.pyplot as plt
import numpy as np

# Return a specific number of evenly spaced numbers
# For example, assign 11 values between 0 and 20 to x
x = np.linspace(0, 20, 11)

y = x * 2

plt.plot(x,y)
plt.show()
```

Getting started with Seaborn

```
import seaborn as sns
import numpy as np

# Same data as previous slide
x = np.linspace(0, 20, 11)
y = x * 2

sns.relplot(x=x, y=y)


# Depending on where Seaborn is running
import matplotlib
matplotlib.pyplot.show()
```

Charts in Plotly

- Basic charts
- Statistical charts
- Other chart types: Scientific charts, financial charts, maps, 3D charts

Getting started with Plotly

```
import plotly.express as px

fig = px.bar(x=["a", "b", "c"], y=[1, 3, 2])

fig.show()
```

Additional Resources

- **Coding is political** and **Coding for social justice**
- Documentation for **Matplotlib**, **Seaborn**, **and** **Plotly**
- **Humble Bundle** (books and courses)
- **Kaggle** (e.g., Python and Data Visualization)
- **Python Charts**
- The open access version of ***Python for Data Analysis***
- **Python Crash Course, third edition**: Cheat sheets
- **Software Carpentry**: Programming with Python; Plotting and Programming in Python