Introduction to plotting

Presenter: Steve Baskauf
steve.baskauf@vanderbilt.edu

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

CodeGraf landing page

- vanderbi.lt/codegraf

Plotting with pandas

Jean & Alexander Heard
LIBRARIES

Pandas plotting basics

- The row index will be used for the x axis by default (or y axis for horizontal bar graphs).
- Every column is added to the plot automatically as another line (or other feature).
- The plot type can be specified using the **`kind=`** argument.
- Other details such as title, axis limits and ticks, and figure size can be specified.

Subplots

- Subplots can be used instead of plotting all lines on the same axes
- Y axis scales on subplots are independent
- X axis scales are shared by default
- **`figsize=`** argument can be used to control size

Other plot types

scatterplot

- requires specifying X and Y columns as arguments

pie

- must specify a single Y or use subplots

bar

- has vertical (**`bar`**) or horizontal (**`barh`**) kinds
- horizontal bar uses row index as default for Y labels
- index from bottom, so reverse alphabetical by default

Plotting with pyplot

Jean & Alexander Heard
LIBRARIES

What is pyplot?

- **`pyplot`** is a collection of styles from **`matplotlib`**, the workhorse library for Python plotting.
- Built on top of numpy
- pandas plots are matplotlib plots under the hood
- pyplot allows a high degree of control of appearance and a huge number of plot type options

- Conventional import statement:

```
import matplotlib.pyplot as plt
```

`plt.plot()` function

- The **`plt.plot()`** function is a versatile plotting function
- It defaults to X, Y line plot, but can be switched to other kinds like scatterplot
- Can control some features, but is limited

- Default input to pyplot plots is numpy arrays (which include pandas series)

Creating figures and subplots

- **`figure()`** is an object type for an entire graphic
- Figures can have multiple subplots.
- Subplots are confusingly referred to as "axes" (not to be confused with an axis on a plot). So we see **`ax`** used as a subplot object name.
- Subplots are organized in rows and columns
- Subplots can be generated *en masse* as a numpy array of subplots and handled programatically

Pros and cons of using figures and subplots

- Con: the coding is more complicated
- Pro: the degree of control over labeling is greater
- Pro: you can automate generation of multiple subplots

Creating plots programmatically

Jean & Alexander Heard
LIBRARIES

Creating "non-standard" plots

- pyplot allows you to control many aspects of plots
- There are usually multiple ways to create a particular plot.
- No clean overarching theoretical framework that is equivalent to R's "grammar of graphics" (ggplot)
- The challenge is extracting the attributes from your data and assigning them to the plot
- Requires good familiarity with pandas dataframes and series
- Input is generally via NumPy arrays, so basic familiarity with help with many code examples.
- Example: stacked bar graph

Remote Support for Teaching and Research Needs

- Access to digital collections 24/7
- Skype consultations with your subject librarian
- Ask a Librarian: an easy way to submit a question via email
- Live chat available from the Library home page

Jean & Alexander Heard
LIBRARIES

NEED HELP? ASK A LIBRARIAN!

https://www.library.vanderbilt.edu/ask-librarian.php

Jean & Alexander Heard
LIBRARIES