Introduction to plotting

Presenter: Steve Baskauf
vanderbi.lt/codegraf

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard
LIBRARIES

CodeGraf landing page

- vanderbi.lt/codegraf

Plotting with pandas

Pandas plotting basics

- The row label 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.

Other plot types

scatterplot
- requires specifying X and Y columns as arguments

pie
- must specify a single Y (which manifests as a Series)

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

Matplotlib and pyplot

What is pyplot?

- **pyplot** is a submodule of **matplotlib** that contains a collection of plot styles.

- Built on top of numpy

- pyplot allows a high degree of control of appearance and a huge number of plot type options

- Accessible via two application interfaces: object-oriented and "pyplot"

- Conventional import statement:

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

Figures with subplots

Subplots

- Subplots can be used instead of plotting all data series on the same axes
- Y axis scales on subplots are independent
- X axis scales are shared by default

Creating figures and subplots

- **`figure()`** is an object type for an entire graphic
- **`figsize=`** argument can be used to control size of overall figure
- 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

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

For more videos like this, visit the CodeGraf landing page

vanderbi.lt/codegraf