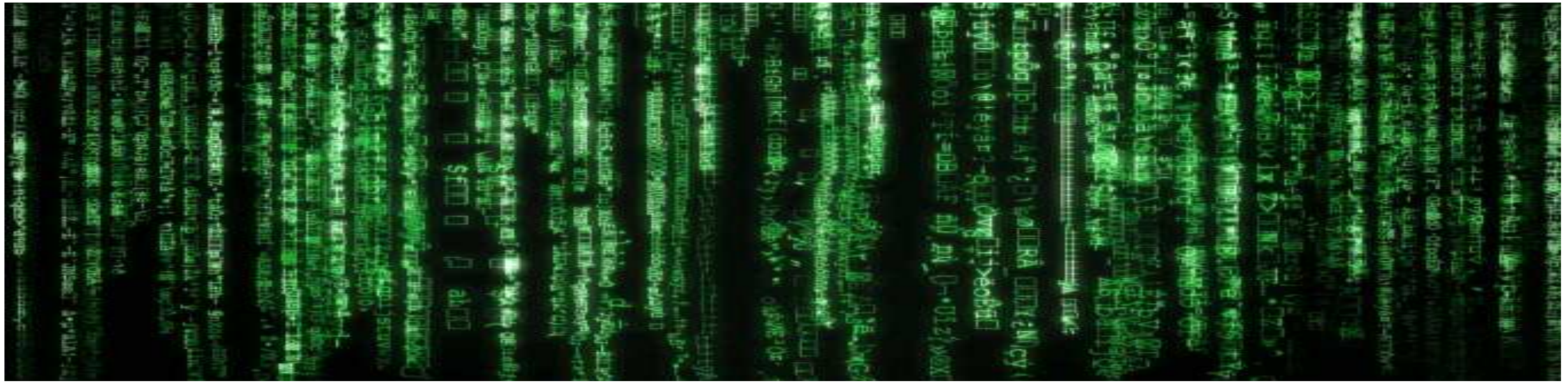


INF 110 Discovering Informatics

More Charts

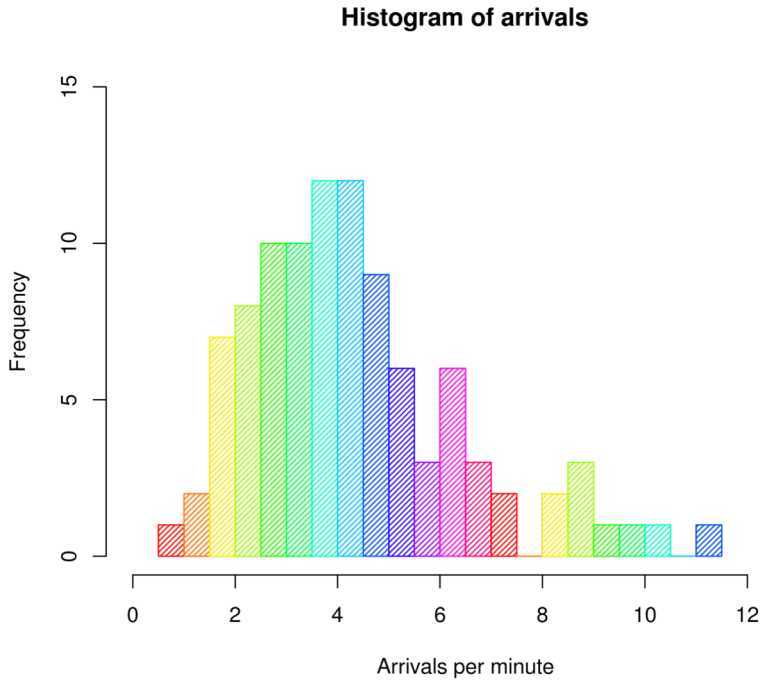
NAU
NORTHERN
ARIZONA
UNIVERSITY

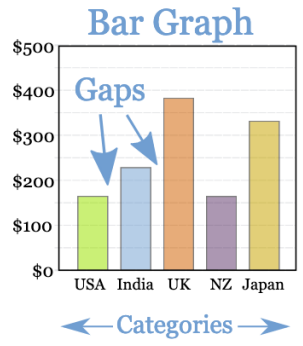
Histograms

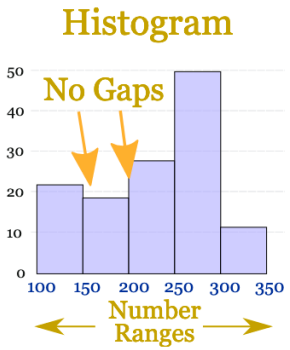
- Represents the distribution of numerical data
- Uses "bins" or ranges of data
- Supports "density estimation"
- No space between bars to distinguish between bar charts

Variations

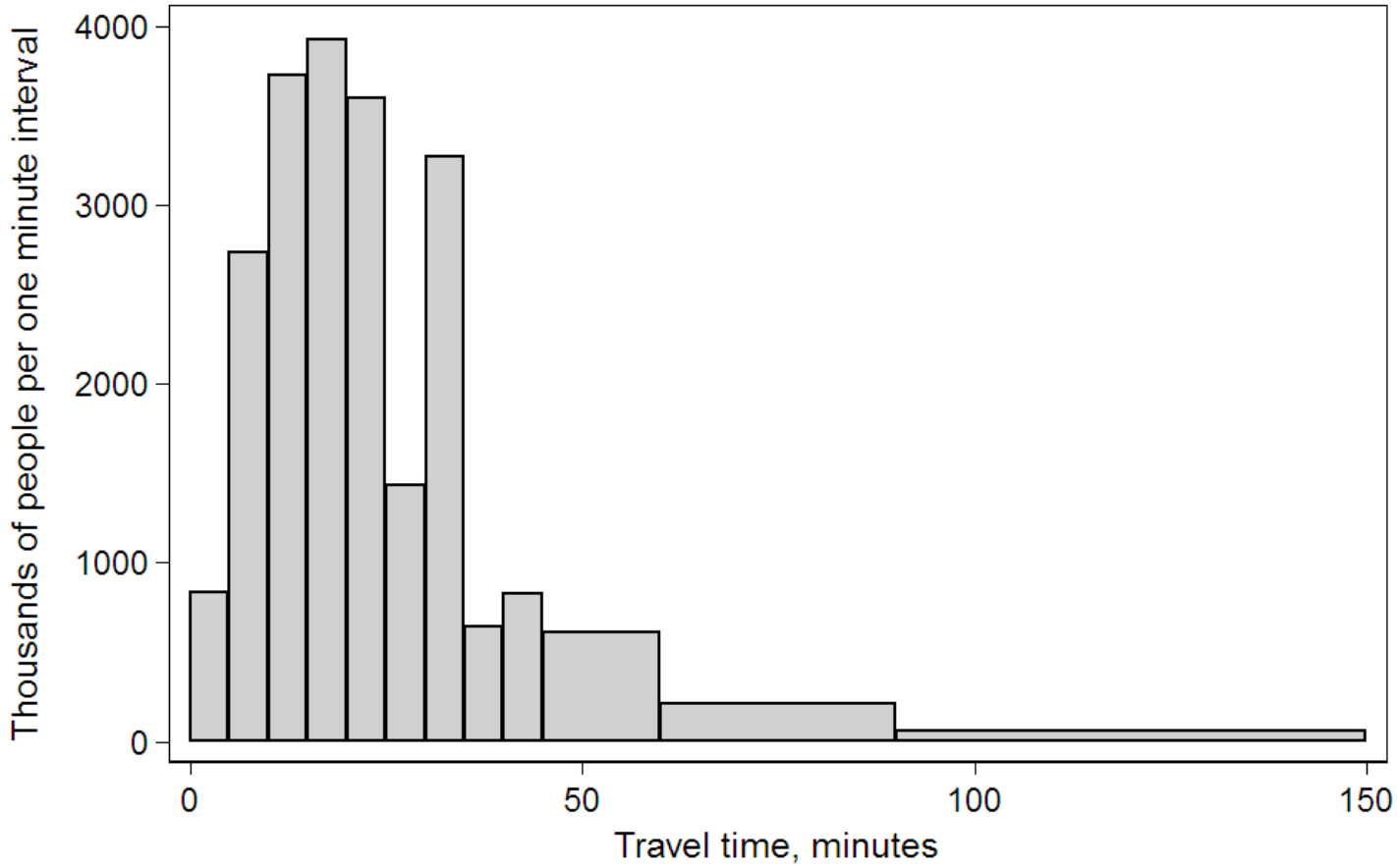
- Non-uniform bin size







U.S. Census Example



Histograms in Python

```
t.hist('column')
```

Optional arguments:

- bins
- units
- normed

Live Code Histogram of Movie Earnings

Tasks: Modify the top movies table to:

- Include gross rounded / truncated to millions
- Only the movie title
- display the results with a histogram
- Set the bins from 300 to 2000 stepped by 100
- Try normed=False

Learning Outcomes

- Creating new columns programmatically
- Create a histogram

Overlaid Charts

- It's often useful to look at multiple variables in a single chart.
- The code to do this looks like:

```
t.chart_method(column_label_of_common_axis,
               array_of_labels_of_variables_to_plot)
```

- But the datascience module can usually figure out the variables from the table automatically (omit the last argument).

Live Code Multi-variable Scatter Charts

Tasks: Use the Galton data set to plot heights with the mother and father's height

Learning Outcomes

- Using multiple variables in a scatter chart

Live Code Multi-variable Line Chart

Tasks: Plot the populations in 2010 and 2014.

Learning Outcomes

- Creating a multi-variable line chart

Live Code Multi-variable Bar Chart

Tasks: Bar chart the ethnicities in the CA 2014 data set.

Learning Outcomes

- Creating a multi-variable bar chart

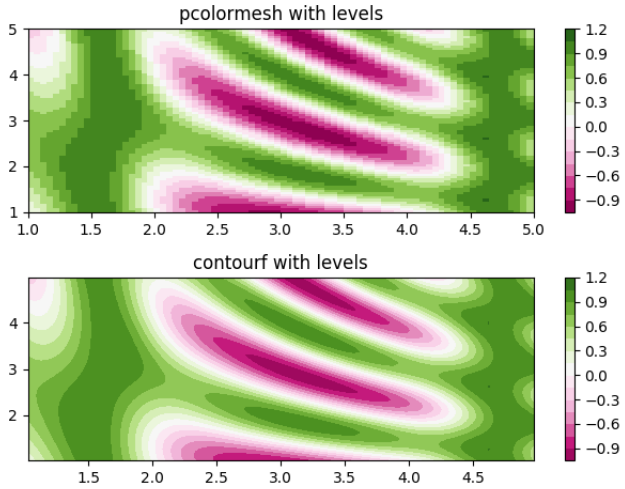
Matplotlib

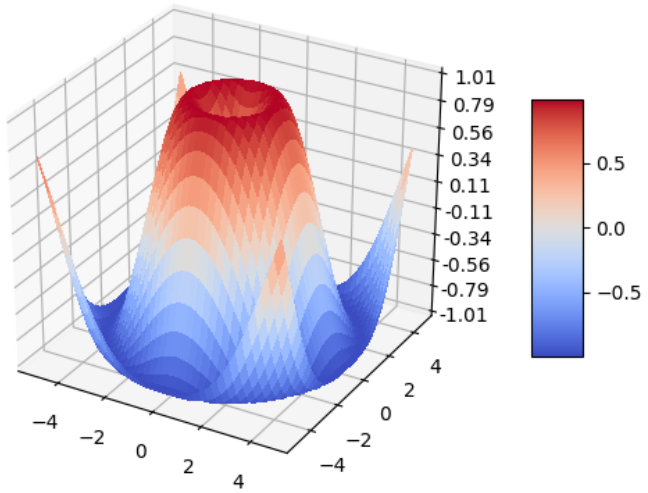
- The datascience module provides a wrapper around matplotlib designed to make it easier to use:

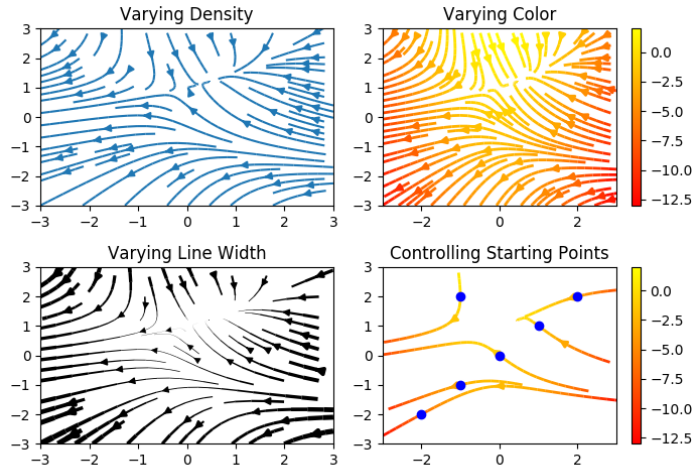
https://matplotlib.org

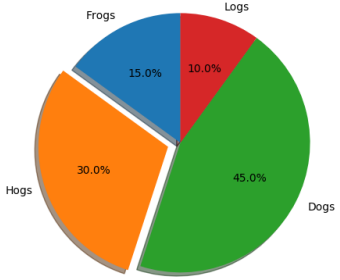
- We won't cover matplotlib or other plotting libraries in depth but our new scientific visualization class will..

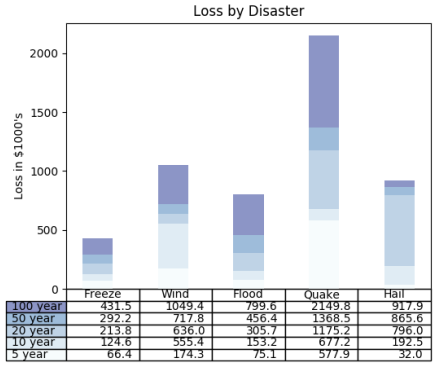
Matplotlib Examples



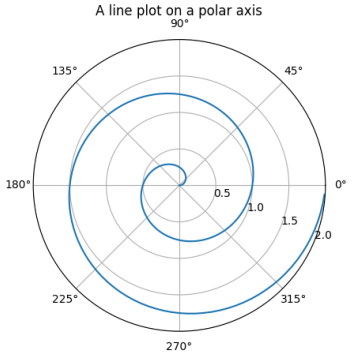


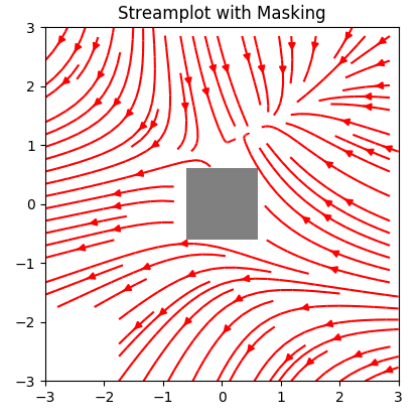






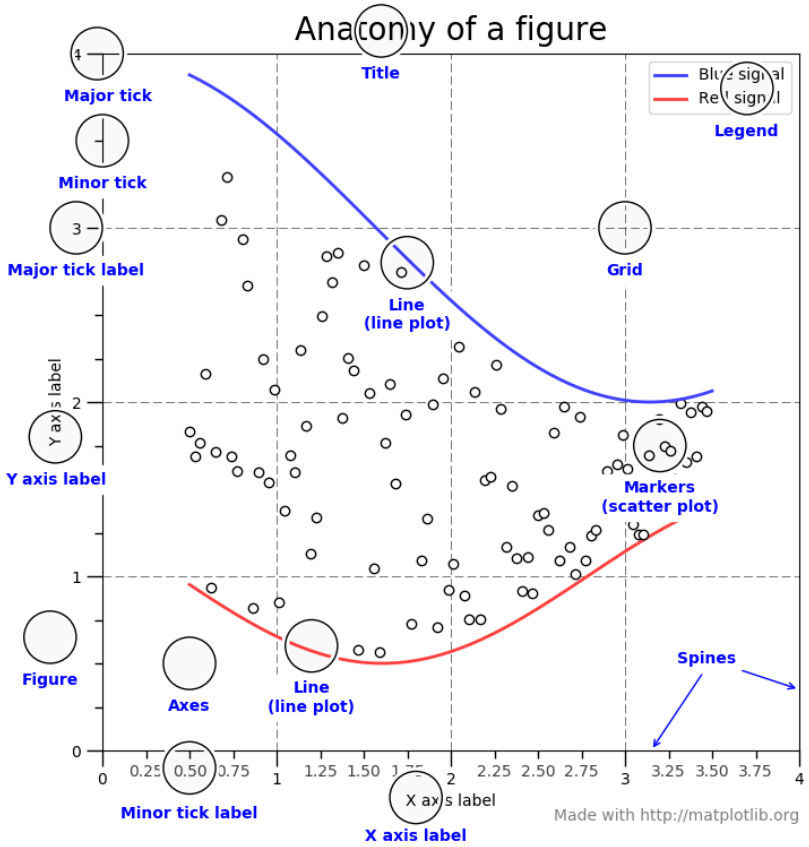
	Freeze	Wind	Flood	Quake	Hail
100 year	431.5	1049.4	799.6	2149.8	917.9
50 year	292.2	717.8	456.4	1368.5	865.6
20 year	213.8	636.0	305.7	1175.2	796.0
10 year	124.6	555.4	153.2	677.2	192.5
5 year	66.4	174.3	75.1	577.9	32.0





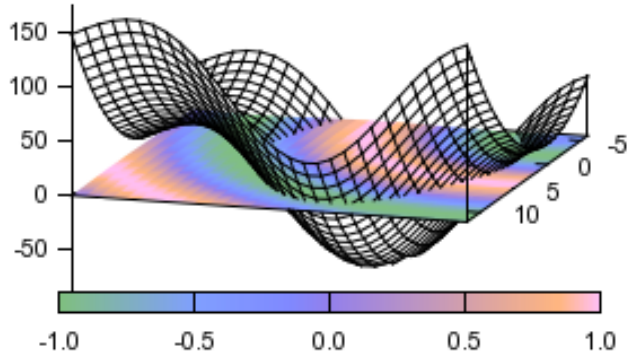
Matplotlib Customization

- Almost everything can be customized!
- Output can be set to PDF, SVG, PNG, etc.
- Great for generating figures for scientific publications



gnuplot

- Older (30+ years)
- Faster
- More stable
- Uses graphing "language"



Easier for quick command line graphs (if you know how)

http://www.gnuplot.info/

BUT: matplotlib is being developed more actively, Python is easier to use.

Live Code Using Google Trends

Tasks: Using Google Trends, which is more popular, gnuplot or matplotlib?

Learning Outcomes

- Using Google Trends to get and plot data


end