Data Visualization

Steve Iannaccone • 04.21.2017

Overview

Objectives

- Learn fundamentals of graphic design
- Use Seaborn and Bokeh to make great static plots

- Use plotly and folium to make great interactive plots

Graphic Design

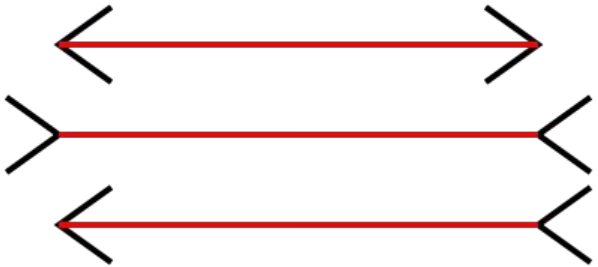
The Thinking Eye

- How do we display information so that it is easily interpretable to the viewer?
- Our eyes are connected to our brains
- This processing comes at a cost

Graphic Design

The Thinking Eye

- We are easily fooled!
- Which line is the longest?



Graphic Design

The Grammar of Graphics

Graphic Design

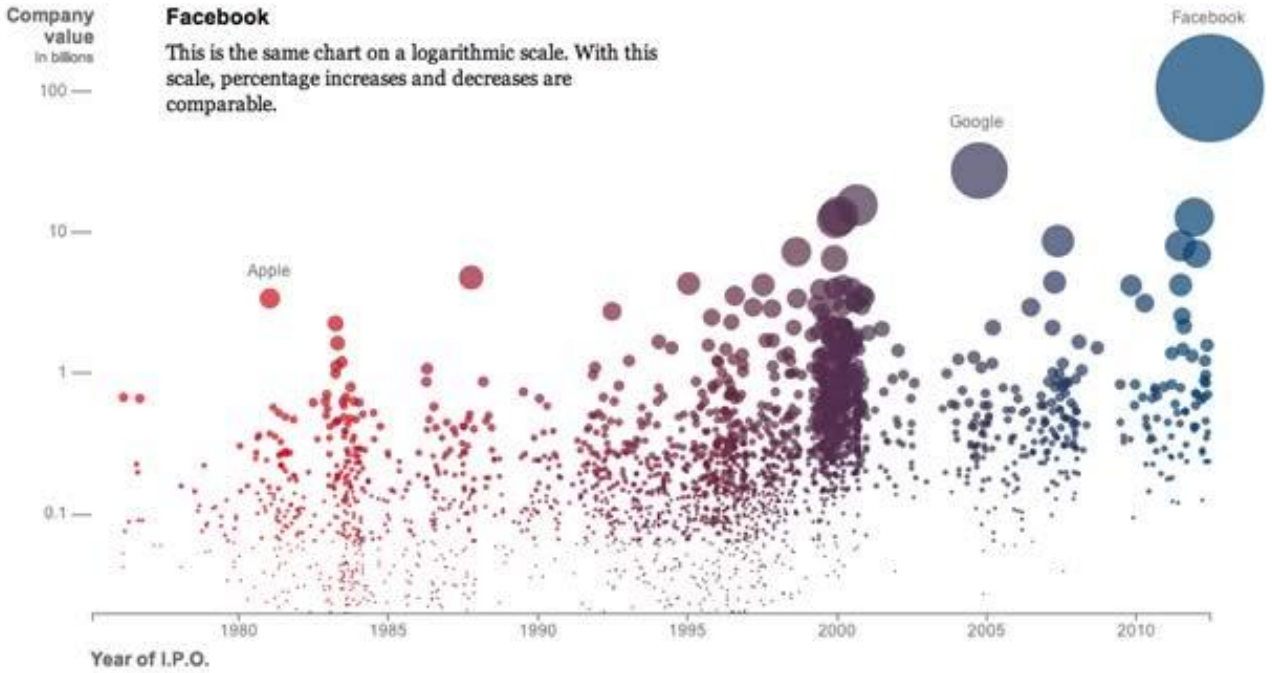
We can Control How We Show Data

- Size
- Color
- Weight
- Position
- Shape
- Scale

Graphic Design

Changing Size

Bubblecharts - Change point sizes to reflect the input data


Company value
In billions
100
10
1
0.1
Facebook
This is the same chart on a logarithmic scale. With this scale, percentage increases and decreases are comparable.
Facebook
Google
Apple
1980
1985
1990
1995
2000
2005
2010
Year of I.P.O.

Graphic Design

A Quick Side Note About Color

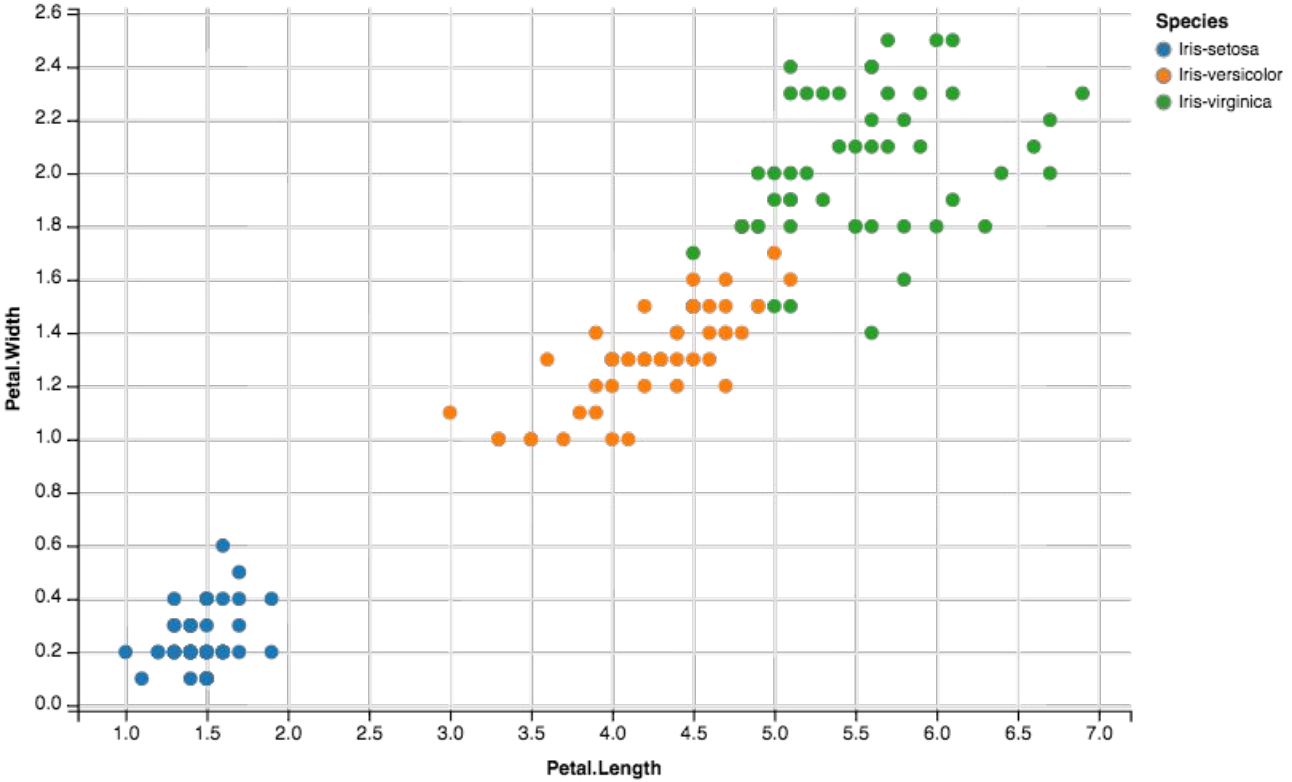
Colorspaces:

- RGBa - (red, green, blue, alpha) usually 8bits per channel
- Hexadecimal - 2 hex values per R, G and B (#00FF00)
- CMYK - Cyan, Magenta, Yellow, Black
- HSV - Hue, Saturation, Value

Graphic Design

Changing Color

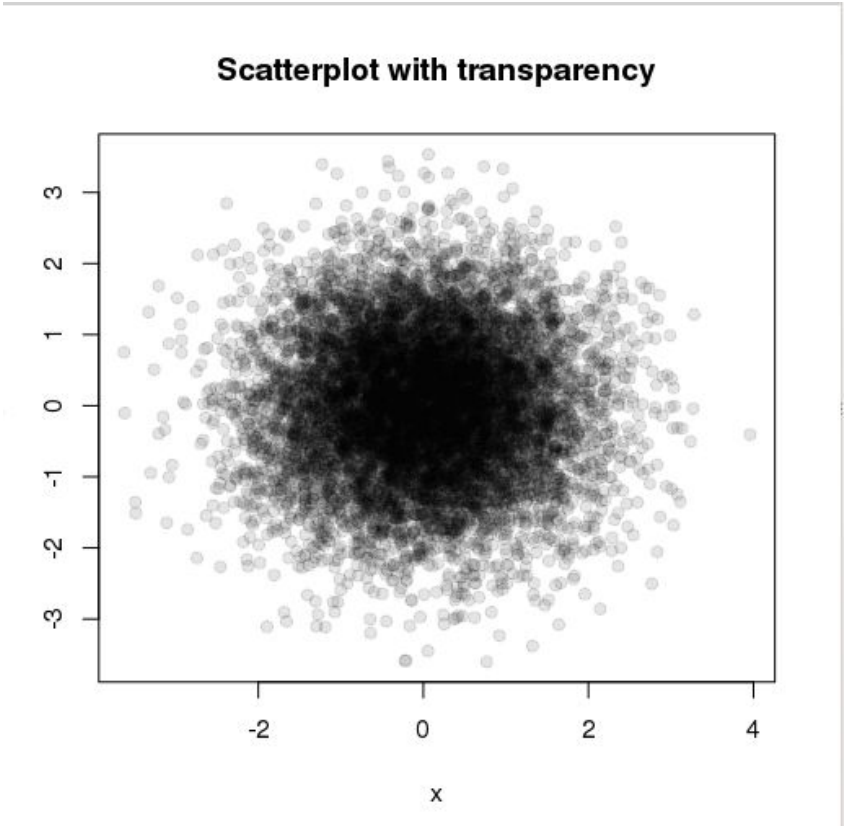
Scatterplots - Change color to reflect classification


Species
Iris-setosa
Iris-versicolor
Iris-virginica
Petal.Width
Petal.Length
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0

Graphic Design

Changing Color

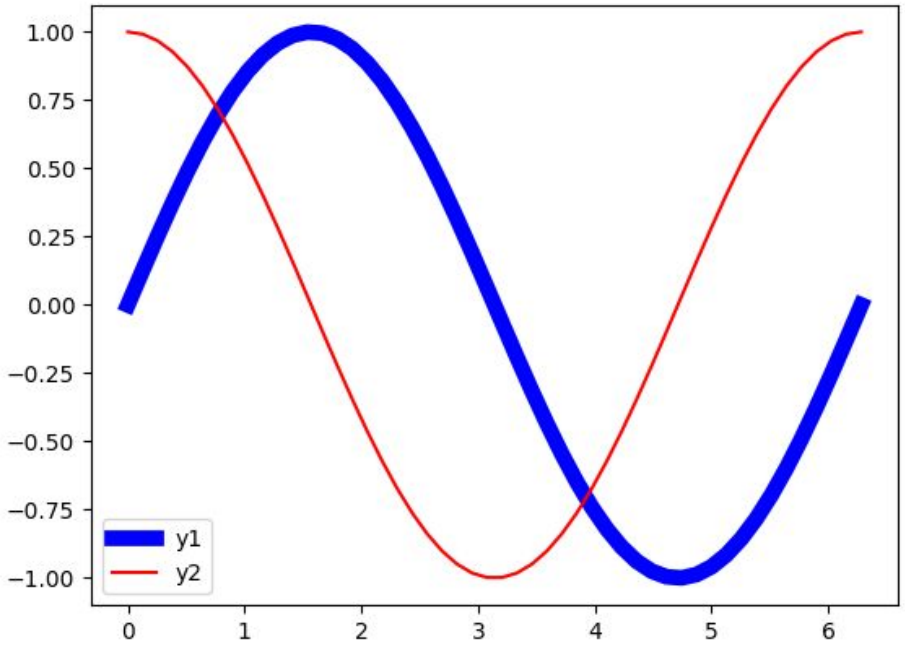
Scatterplots - Change transparency to show density


Scatterplot with transparency
3
2
1
0
-1
-2
-3
-2
0
2
4
x

Graphic Design

Changing Weight

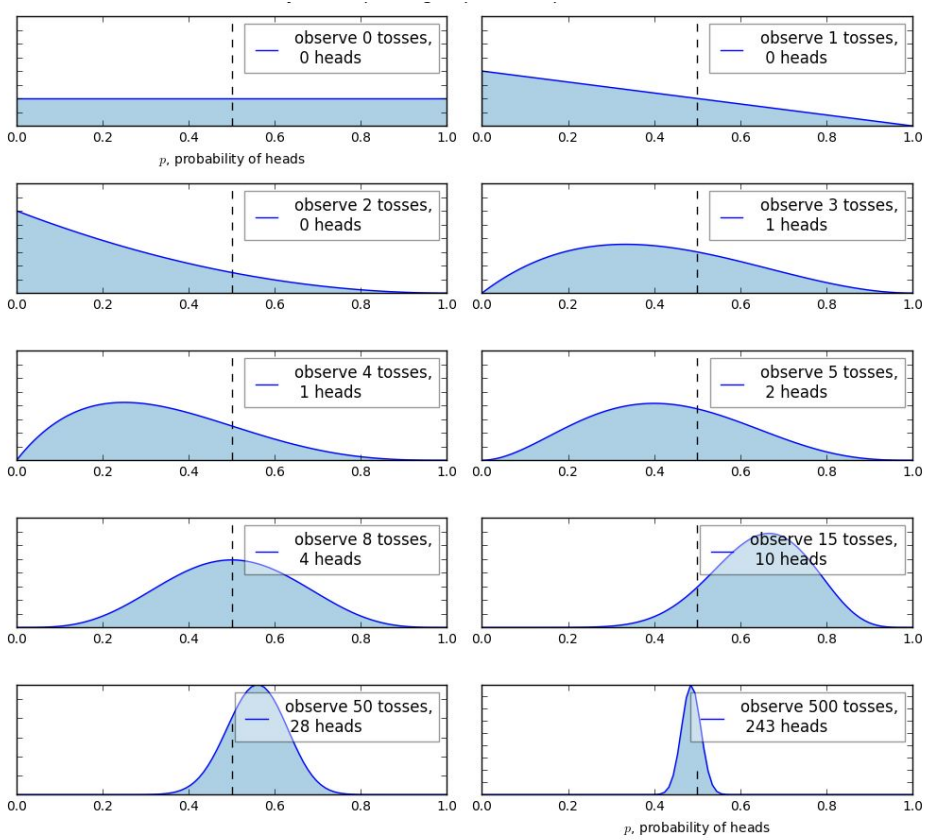
Line plots - Change line width to show importance


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
1
2
3
4
5
6
y1
y2

Graphic Design

Changing Position

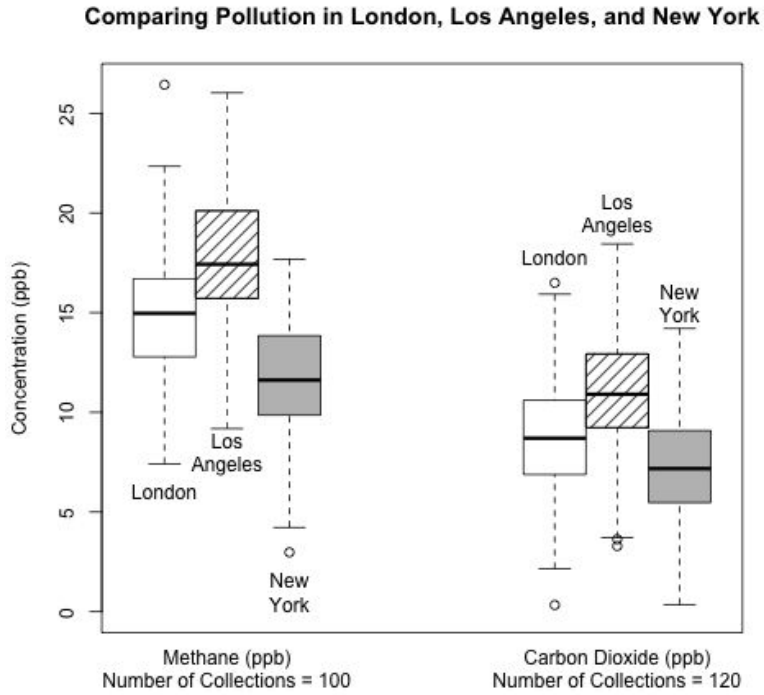
Subplots - We can add multiple plots to the same figure to show that these graphics are related to one another


observe 0 tosses,
0 heads
observe 1 tosses,
0 heads
observe 2 tosses,
0 heads
observe 3 tosses,
1 heads
observe 4 tosses,
1 heads
observe 5 tosses,
2 heads
observe 8 tosses,
4 heads
observe 15 tosses,
10 heads
observe 50 tosses,
28 heads
observe 500 tosses,
243 heads
0.0
0.2
0.4
0.6
0.8
1.0
p, probability of heads
p, probability of heads

Graphic Design

Changing Shape

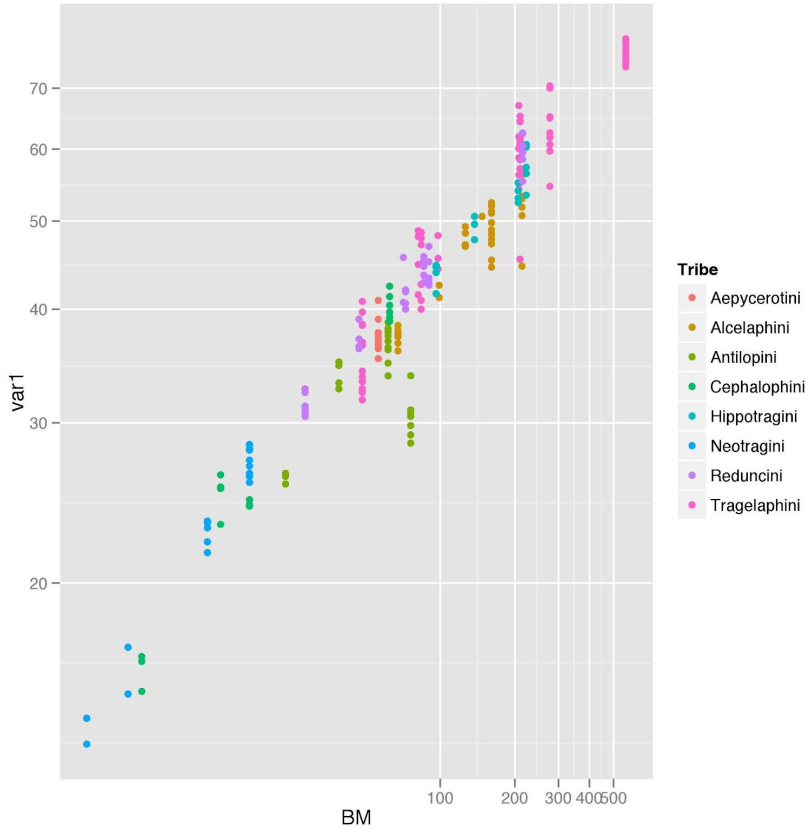
Boxplots - If your plot has to be in B&W we can use shapes and textures to indicate classification


Comparing Pollution in London, Los Angeles, and New York
Concentration (ppb)
0
5
10
15
20
25
London
Los Angeles
New York
London
Los Angeles
New York
Methane (ppb)
Number of Collections = 100
Carbon Dioxide (ppb)
Number of Collections = 120

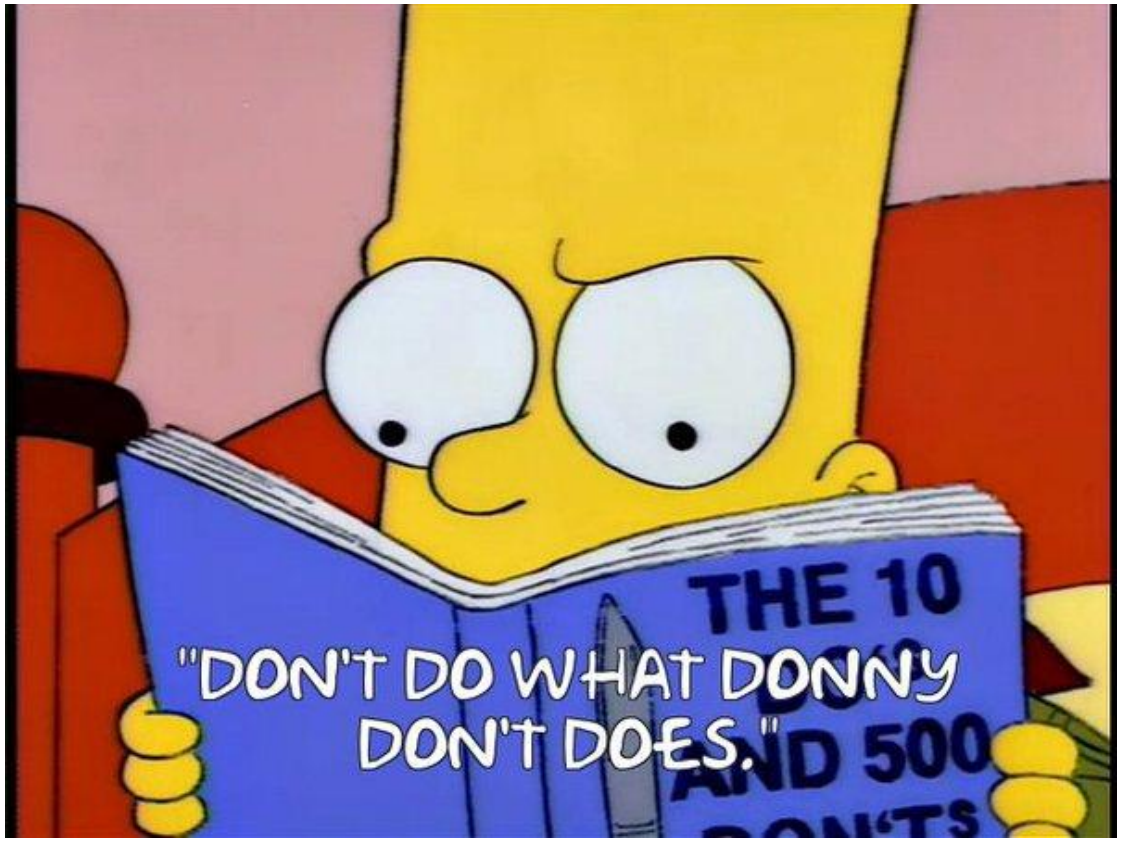
Graphic Design

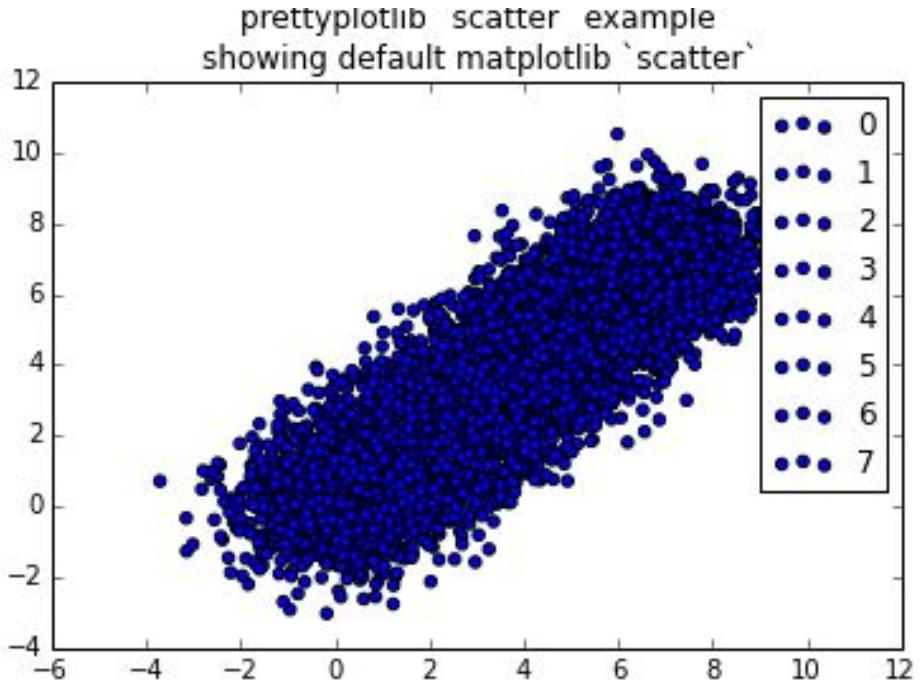
Changing Scale

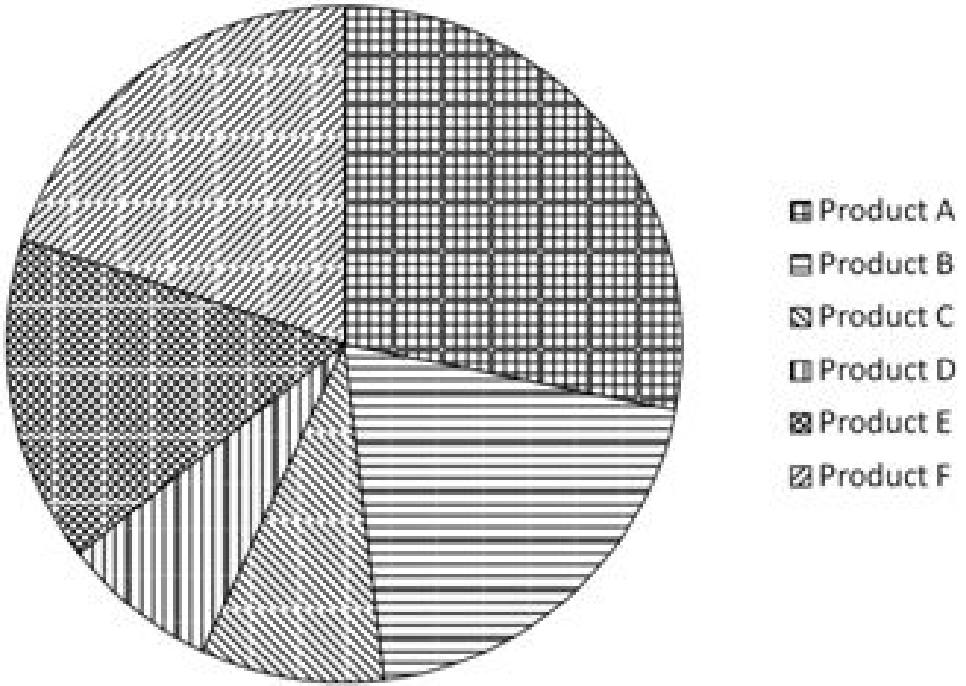
Log plots - If your data is not linear, you can scale your axes to show logarithmic effects

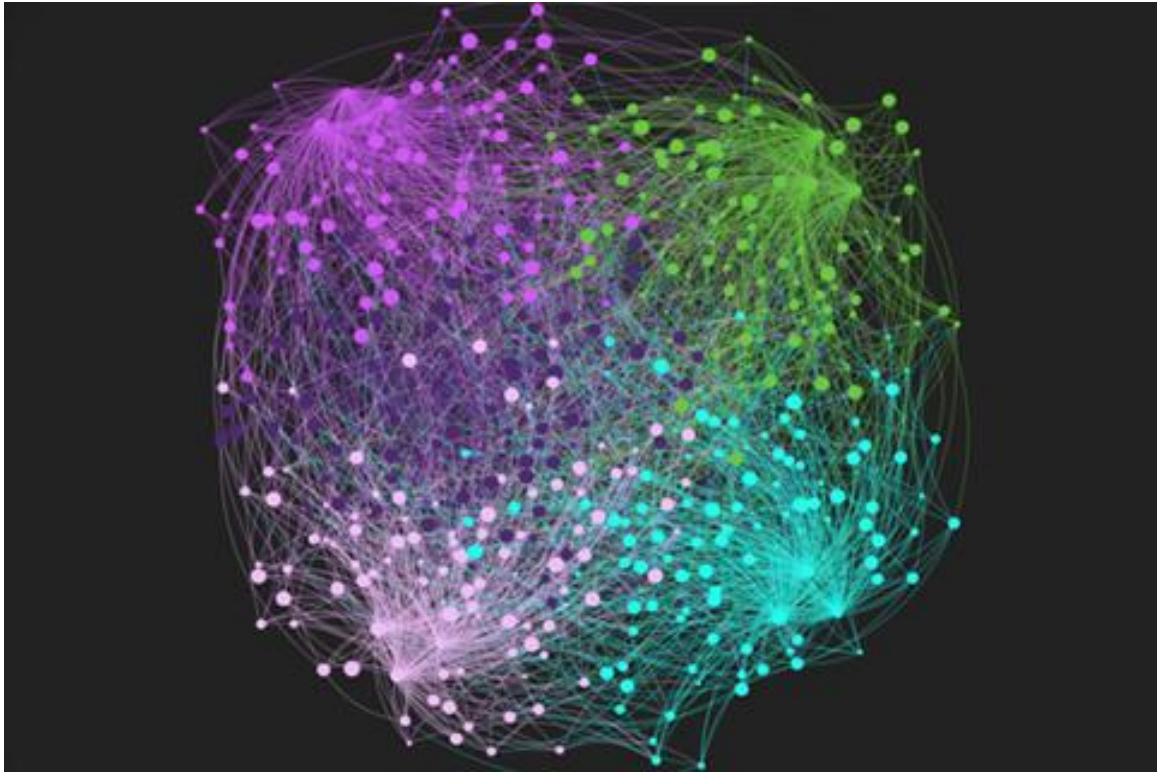

70
60
50
40
30
20
var1
100
200
300
400
500
BM
Tribe
Aepycerotini
Alcelaphini
Antilopini
Cephalophini
Hippotragini
Neotragini
Reduncini
Tragelaphini

Some Pitfalls

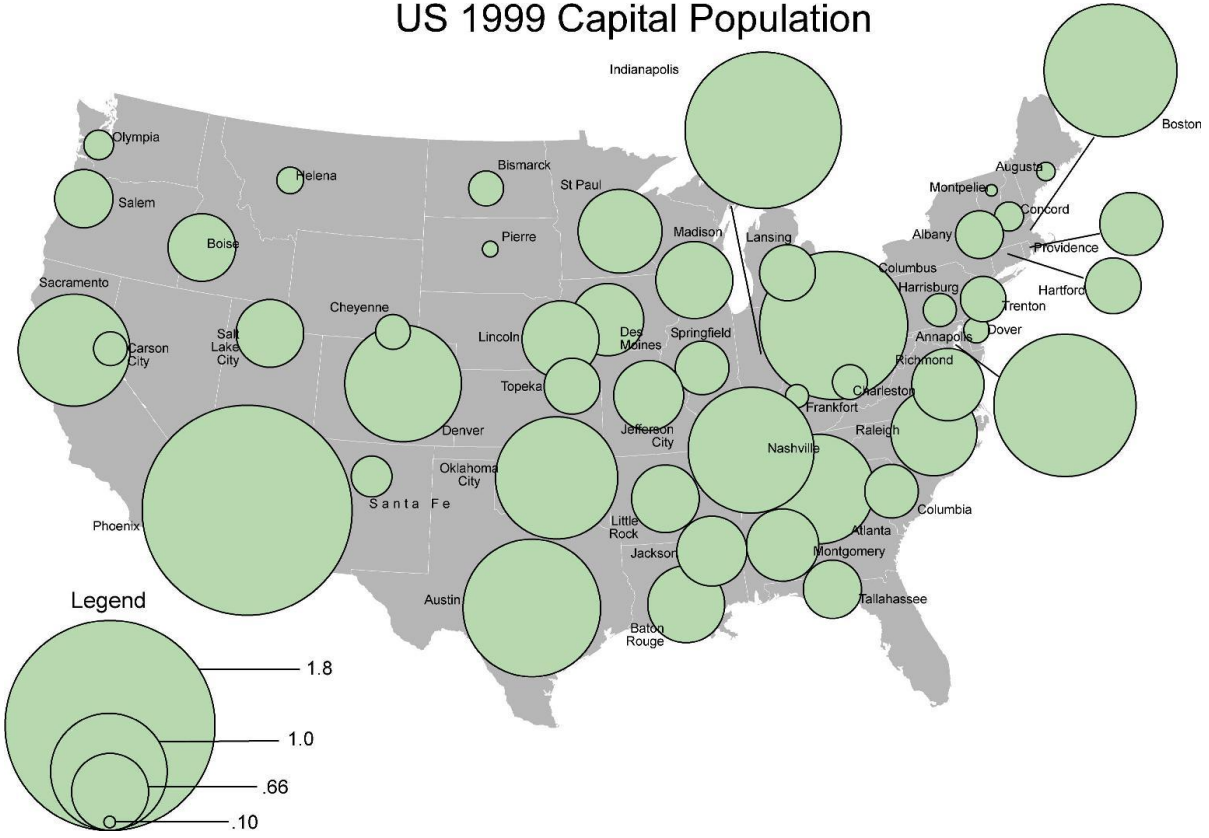

"DON'T DO WHAT DONNY
DON'T DOES."
THE 10
AND 500


prettyplotlib scatter example
showing default matplotlib `scatter`
12
10
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6
8
10
12
0
1
2
3
4
5
6
7


Product A
Product B
Product C
Product D
Product E
Product F



US 1999 Capital Population




Secaucus
North Bergen
West New York
Union City
Manhattan
Roosevelt Island
Sutton Place
Hoboken
Hunters Point
Woodside
Jackson Heights
Jersey City
Hudson Exchange
Paulus Hook
New York
Soho
Cooperative Village
Wall Street
Middle Village
Glendale
Bushwick
Queens Midtown Expy
Brooklyn Queens Expy
Grand Central Pkwy
Northern Blvd
Queens Blvd
Jersey Turnpike Ext
A
B
C
D
E
F
G
H
I
J

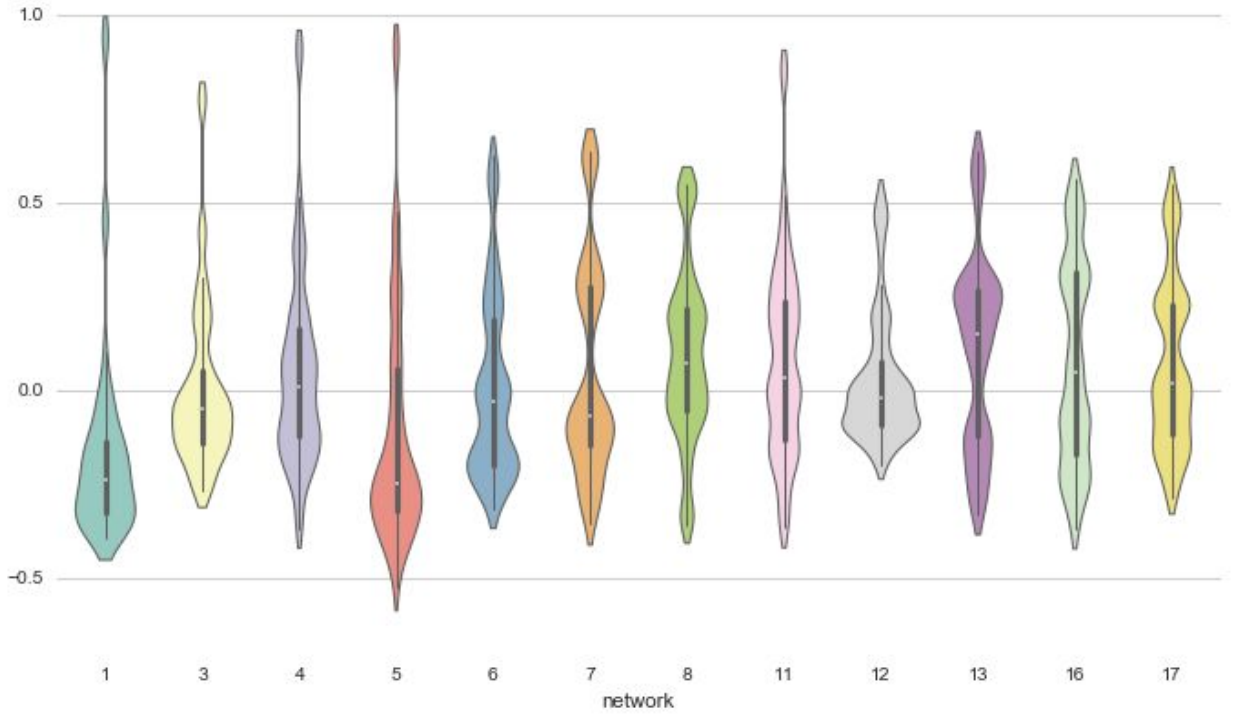
Using Seaborn & Bokeh

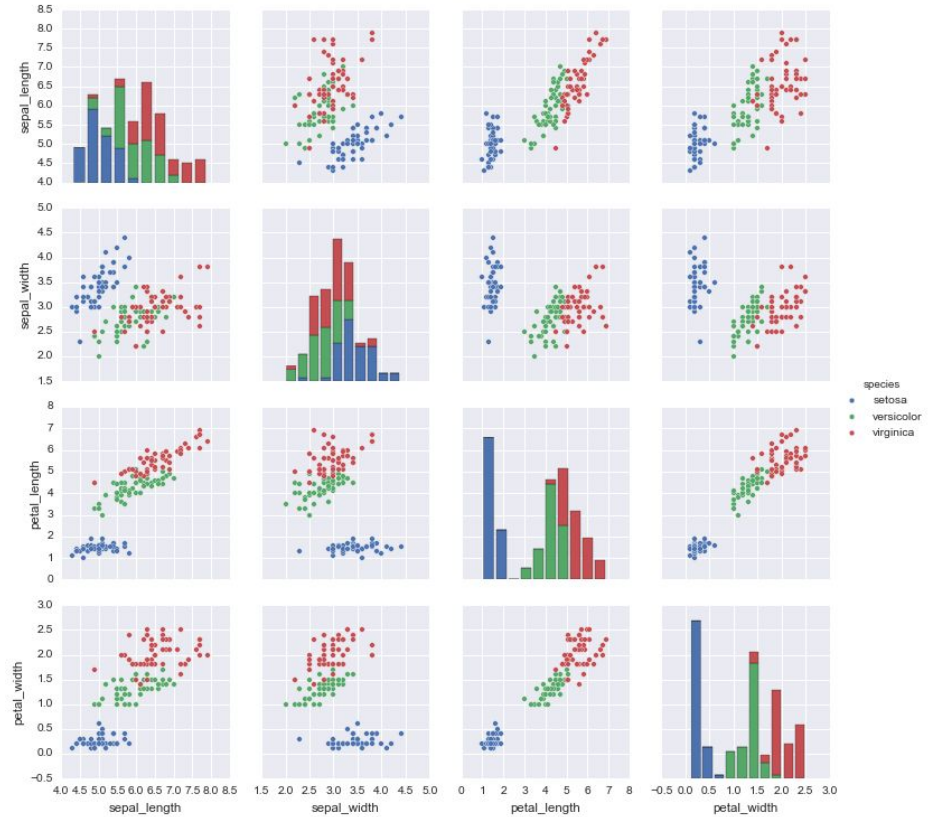
Advanced Plotting in Python

Seaborn

A python package that uses MatPlotLib under the hood

- Simple to use
- Great plots out of the box
- Some added features


1.0
0.5
0.0
−0.5
1
3
4
5
6
7
8
11
12
13
16
17
network


sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica

Advanced Plotting in Python

Seaborn

Check out their Gallery for code examples:

https://seaborn.pydata.org/examples/index.html

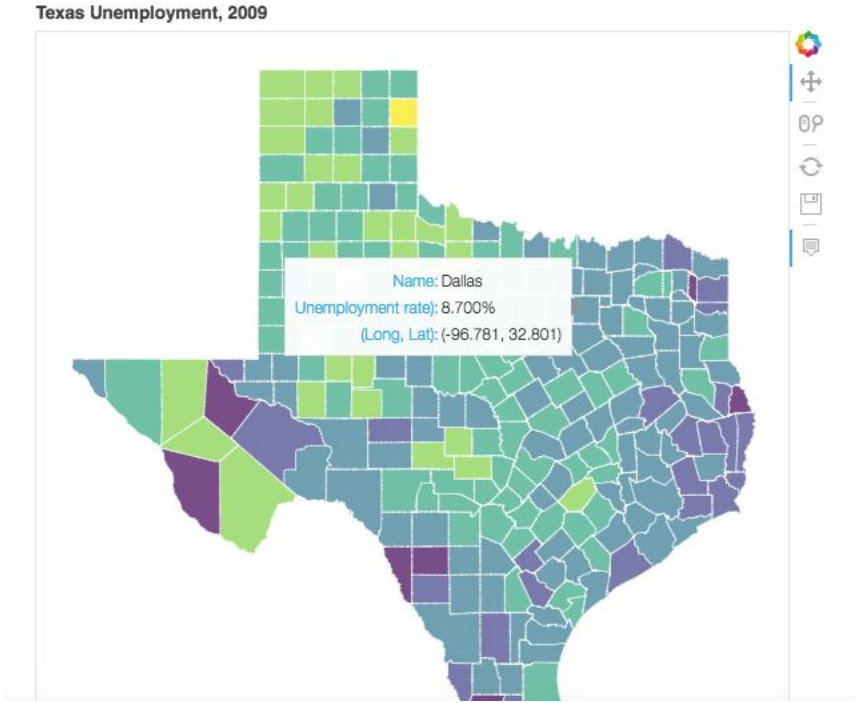
Advanced Plotting in Python

Bokeh

A python package that uses JavaScript under the hood and your web-browser to display

- Great looking plot, perfect if you need to host your figure online
- Can add in interaction *(like hover states)*
- But, less simple to use *(interactions are awkward at best)*

texas.py ¶


Texas Unemployment, 2009
Name: Dallas
Unemployment rate): 8.700%
(Long, Lat): (-96.781, 32.801)

Using Plotly & Folium

Interactive Plotting

Plotly

A python package written by Hadley Wickham that uses D3 under the hood

- Making interactive plots less cumbersome
- Input arguments look like JSON objects
- Documentation is hard to read and navigate

Interactive Plotting

Plotly

They are trying to run a business, so they imply that you need an account to make plots...Thankfully this isn't true. Just run it in `offline` mode.

Interactive Plotting

Plotly

```
$ pip install plotly
```

```
import plotly
import plotly.plotly as py
from plotly.graph_objs import *
from plotly.offline import download_plotlyjs,
    init_notebook_mode, plot, iplot
```

Interactive Plotting

Plotly

Now in plotly's examples wherever you see:

```
plotly.iplot()
```

or:

```
py.iplot()
```

Use this instead:

```
plotly.offline.plot()
```

Interactive Plotting

Plotly

Example:

https://plot.ly/python/line-and-scatter/#style-scatter-plots

Interactive Plotting

Plotly

This writes a local HTML file that has all the necessary JS functions minified in the script tag. This file allows you to interact with your graph as a stand alone file.

Interactive Plotting

Plotly

Parameters of the plot look like JavaScript Object Notation *(JSON)*.

In Python, this means it looks like nested `dict()` calls.

Interactive Plotting

Plotly

```
marker = dict(
         size = 10,
         color = 'rgba(255, 182, 193, .9)',
         line = dict(
             width = 2,
         )...
```

Interactive Plotting

Plotly

Using APIs like Mapbox are awesome at plotting on top of great looking maps.

This will often require setting up an account to get access-tokens to use.

Interactive Plotting

Folium

Allows us to use Python to interact with Leaflet.JS mapping tools much in the same why Mapbox and Plotly do, but without any accounts or access-tokens.

Recap

1. What are the things we can change to make better figures?
2. What is a hexadecimal color, and where do we see them most often?
3. Name at least 3 packages that let us make plots in python.

Assignment

Use folium to make an interactive map.