CS49000-VIZ - Fall 2020

Introduction to Data Visualization

# Visualization Libraries

## Lecture 2

August 27, 2020

# Today

- Announcements
- Pandas for Data Handling
- Matplotlib.pyplot vs. Bokeh
- First Assignment


# Data structures & Data Analysis

- works harmoniously with numpy
- imports csv files into DataFrame objects
- Slice, filter, count, select, replace, dump, group…


- Excellent tutorials and user guide
- Read the doc if unfamiliar
- Very helpful for projects that involve tabular data


# Online doc

# Matplotlib vs Bokeh

- Bar Chart

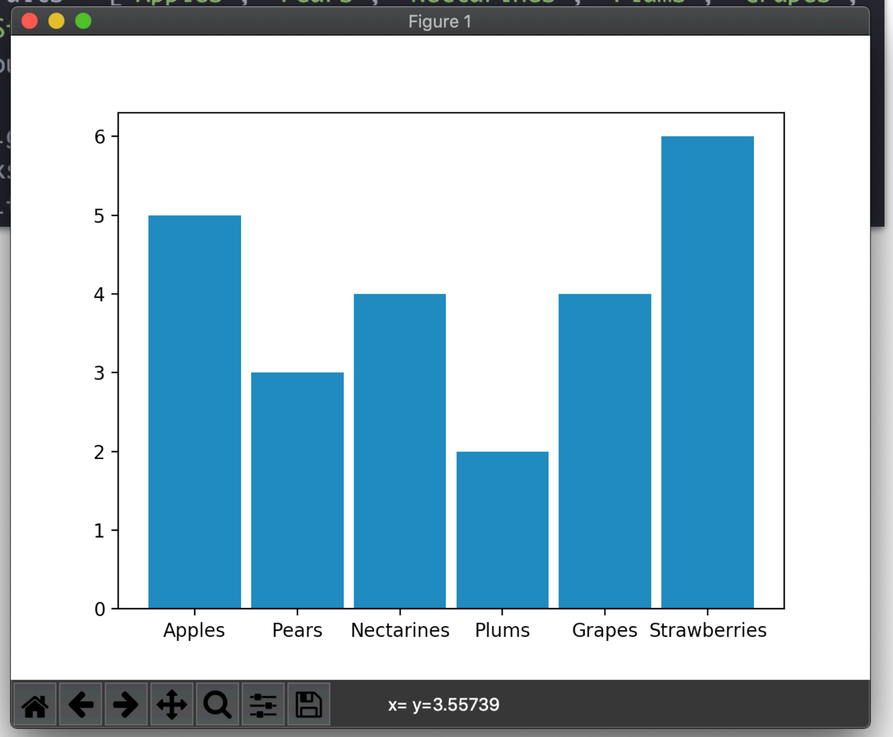
```python
import numpy as np
import matplotlib.pyplot as plt

fruits = ['Apples', 'Pears', 'Nectarines', 'Plums', 'Grapes',
'Strawberries']
counts = [5, 3, 4, 2, 4, 6]

fig, axs = plt.subplots()
axs.bar(fruits, counts, width=0.9)
plt.show()
```

```python
from bokeh.io import output_file, show
from bokeh.plotting import figure

output_file("bars.html")

fruits = ['Apples', 'Pears', 'Nectarines', 'Plums', 'Grapes',
'Strawberries']
counts = [5, 3, 4, 2, 4, 6]

p = figure(x_range=fruits, plot_height=250, title="Fruit Counts",
           toolbar_location=None, tools="")

p.vbar(x=fruits, top=counts, width=0.9)

p.xgrid.grid_line_color = None
p.y_range.start = 0

show(p)
```

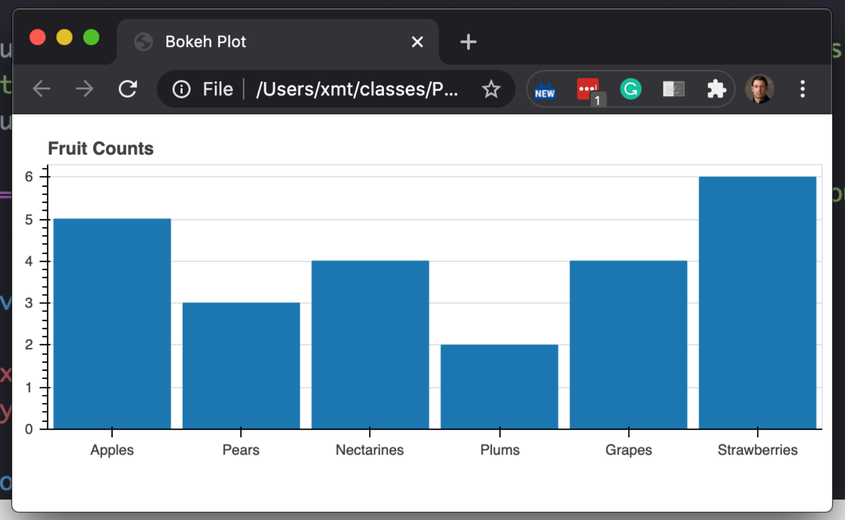
# Matplotlib vs Bokeh

- Bar Chart

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

fruits = ['Apples', 'Pears', 'Nectarines', 'Plums', 'Grapes',
'S
co

fi
ax
pl
```




```
from bokeh.io import output_file, show
from bokeh.plotting import figure

output_file("bars.html")

fru
'St
cou

p =

p.v

p.x
p.y

sho
```



# Matplotlib vs Bokeh

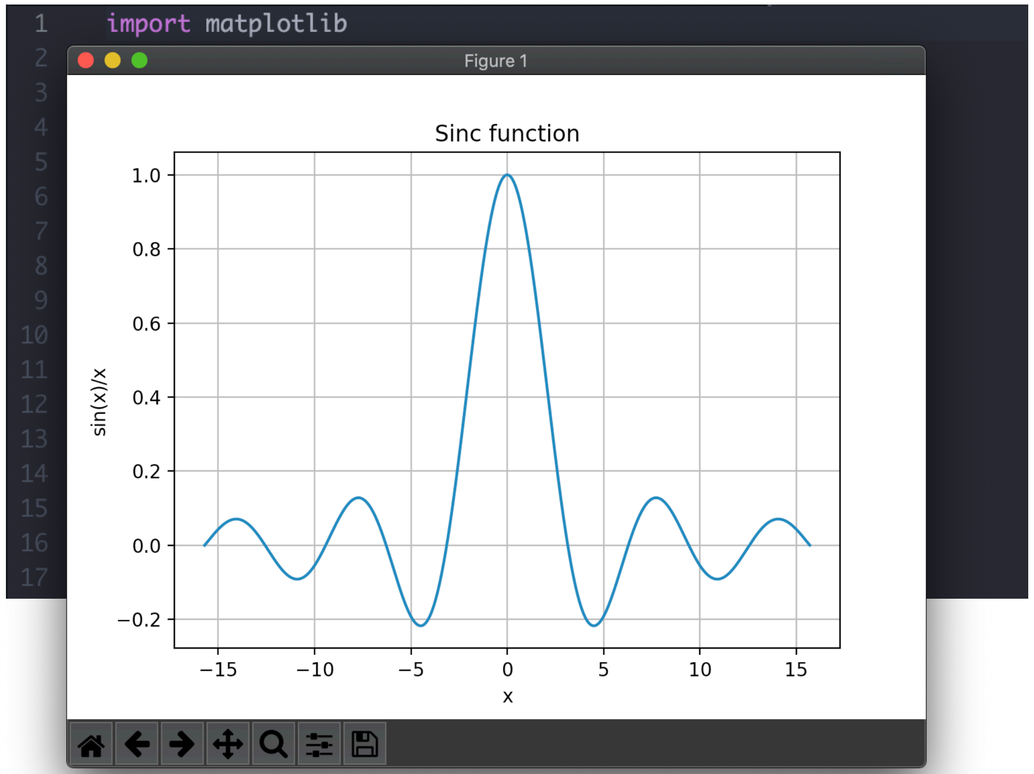
- Line Chart

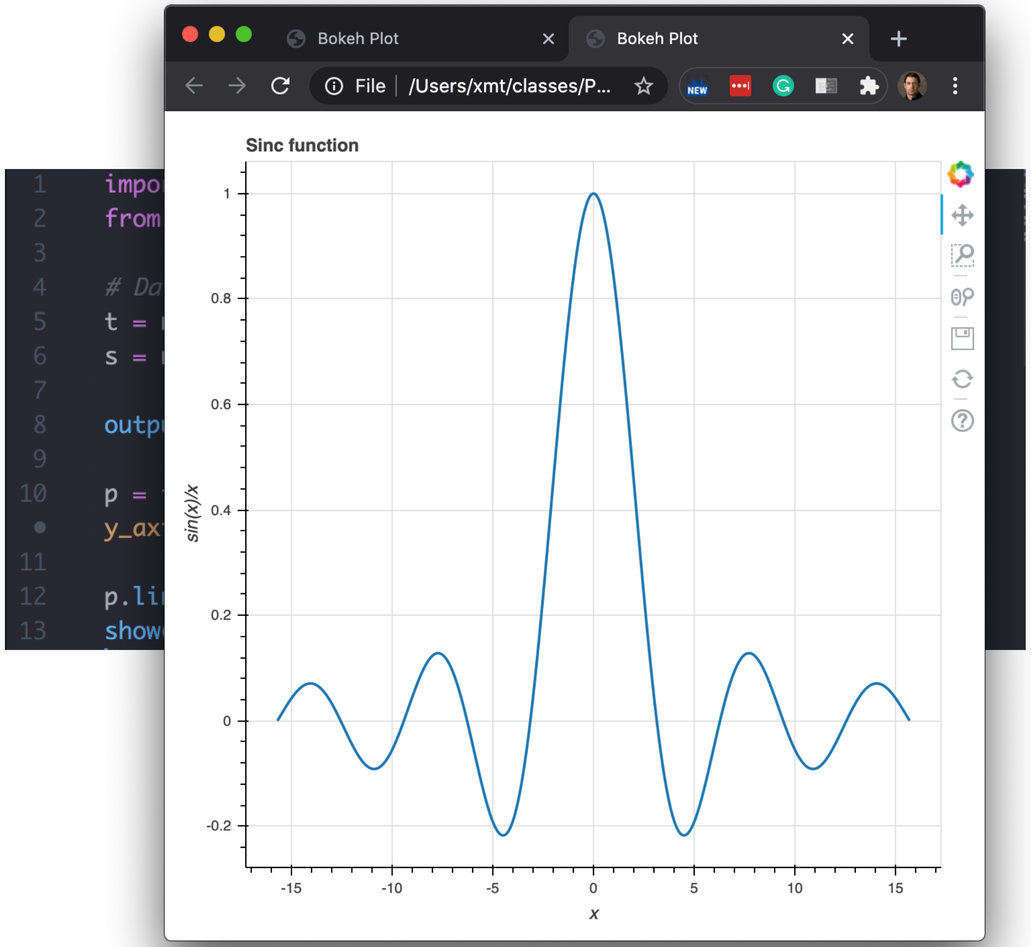
```python
import matplotlib
import matplotlib.pyplot as plt
import numpy as np

# Data for plotting
t = np.arange(-5.0*np.pi, 5.0*np.pi, 0.01)
s = np.sin(t)/t

fig, ax = plt.subplots()
ax.plot(t, s)

ax.set(xlabel='x', ylabel='sin(x)/x',
       title='Sinc function')
ax.grid()

fig.savefig("sinc.png")
plt.show()
```

```python
import numpy as np
from bokeh.plotting import figure, output_file, show

# Data for plotting
t = np.arange(-5.0*np.pi, 5.0*np.pi, 0.01)
s = np.sin(t)/t

output_file('sinc.html')

p = figure(title = 'Sinc function', x_axis_label = 'x',
y_axis_label = 'sin(x)/x')

p.line(t, s, line_width=2)
show(p)
```

# Matplotlib vs Bokeh

- Line Chart



# Matplotlib vs Bokeh

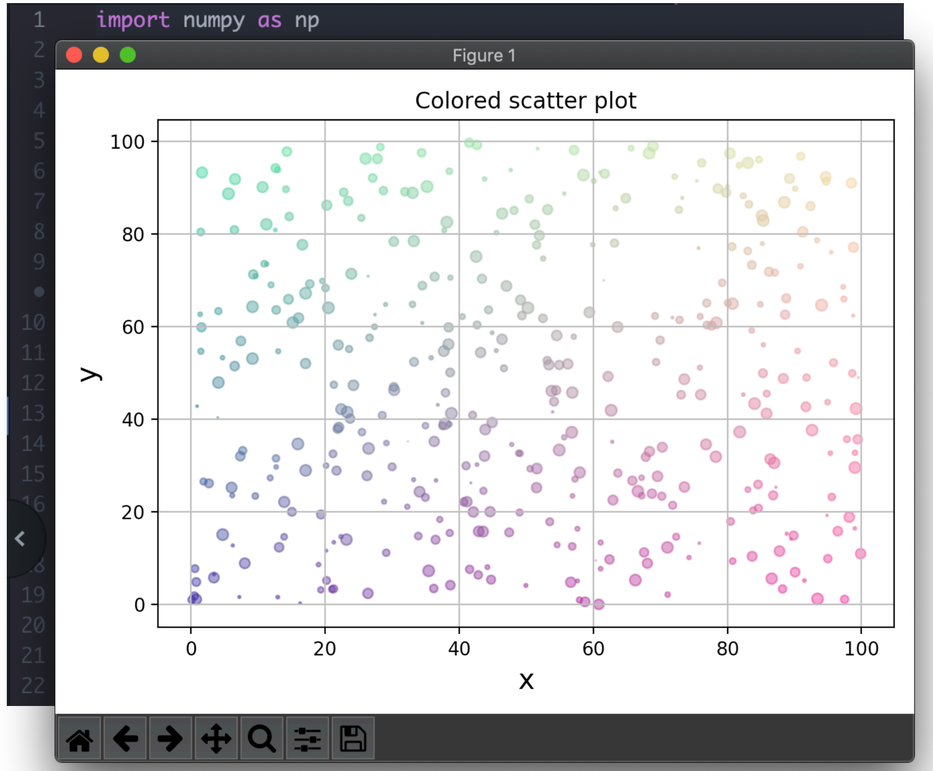
- Scatter plot

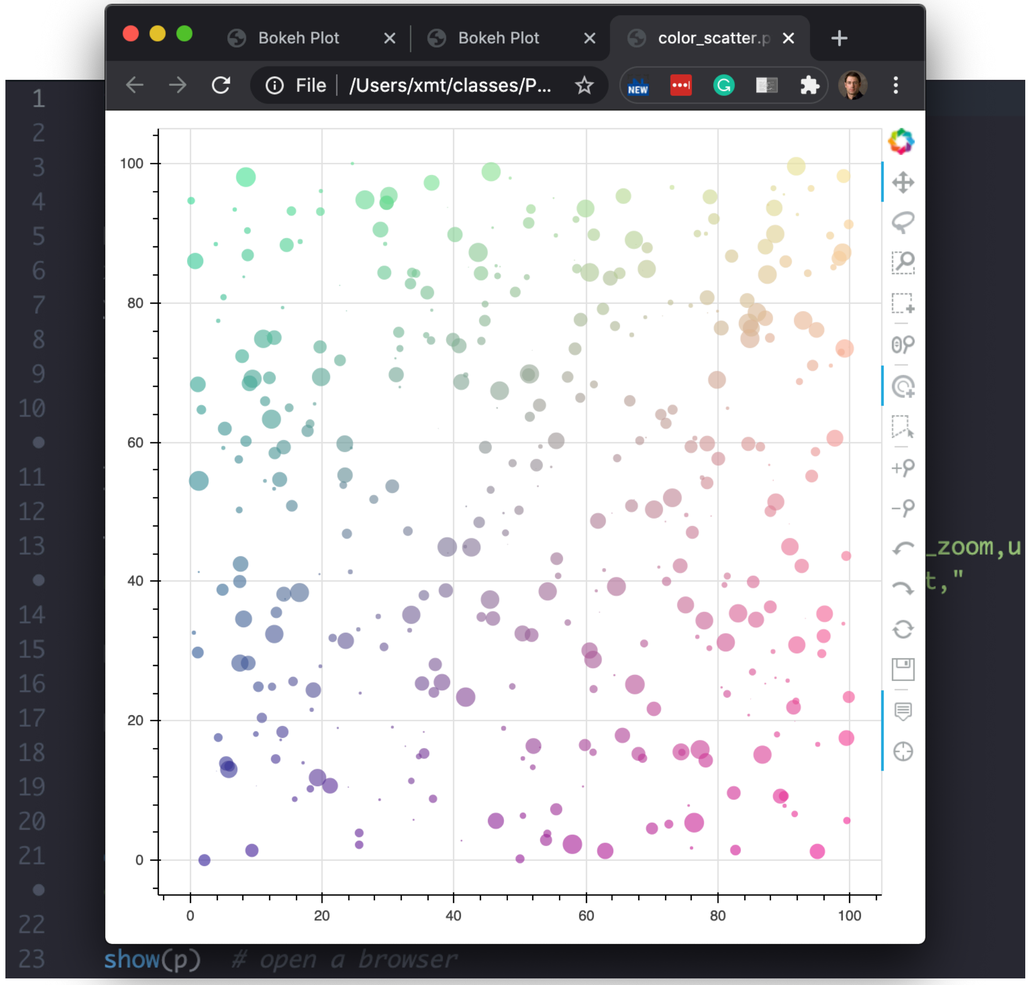
```python
import numpy as np
import matplotlib.pyplot as plt

N = 400
x = np.random.random(size=N) * 100
y = np.random.random(size=N) * 100
radii = np.random.random(size=N) * 1.5
colors = [
    "#%02x%02x%02x" % (int(r), int(g), 150) for r, g in
    zip(50+2*x, 30+2*y)
]

fig, ax = plt.subplots()
ax.scatter(x, y, c=colors, s=25*radii, alpha=0.5)

ax.set_xlabel('x', fontsize=15)
ax.set_ylabel('y', fontsize=15)
ax.set_title('Colored scatter plot')

ax.grid(True)
fig.tight_layout()

plt.show()
```

```python
import numpy as np

from bokeh.plotting import figure, output_file, show

N = 400
x = np.random.random(size=N) * 100
y = np.random.random(size=N) * 100
radii = np.random.random(size=N) * 1.5
colors = [
    "#%02x%02x%02x" % (int(r), int(g), 150) for r, g in
    zip(50+2*x, 30+2*y)
]

TOOLS="hover,crosshair,pan,wheel_zoom,zoom_in,zoom_out,box_zoom,u
ndo,redo,reset,tap,save,box_select,poly_select,lasso_select,"

p = figure(tools=TOOLS)

p.scatter(x, y, radius=radii,
          fill_color=colors, fill_alpha=0.6,
          line_color=None)

output_file("color_scatter.html", title="color_scatter.py
example")

show(p)  # open a browser
```

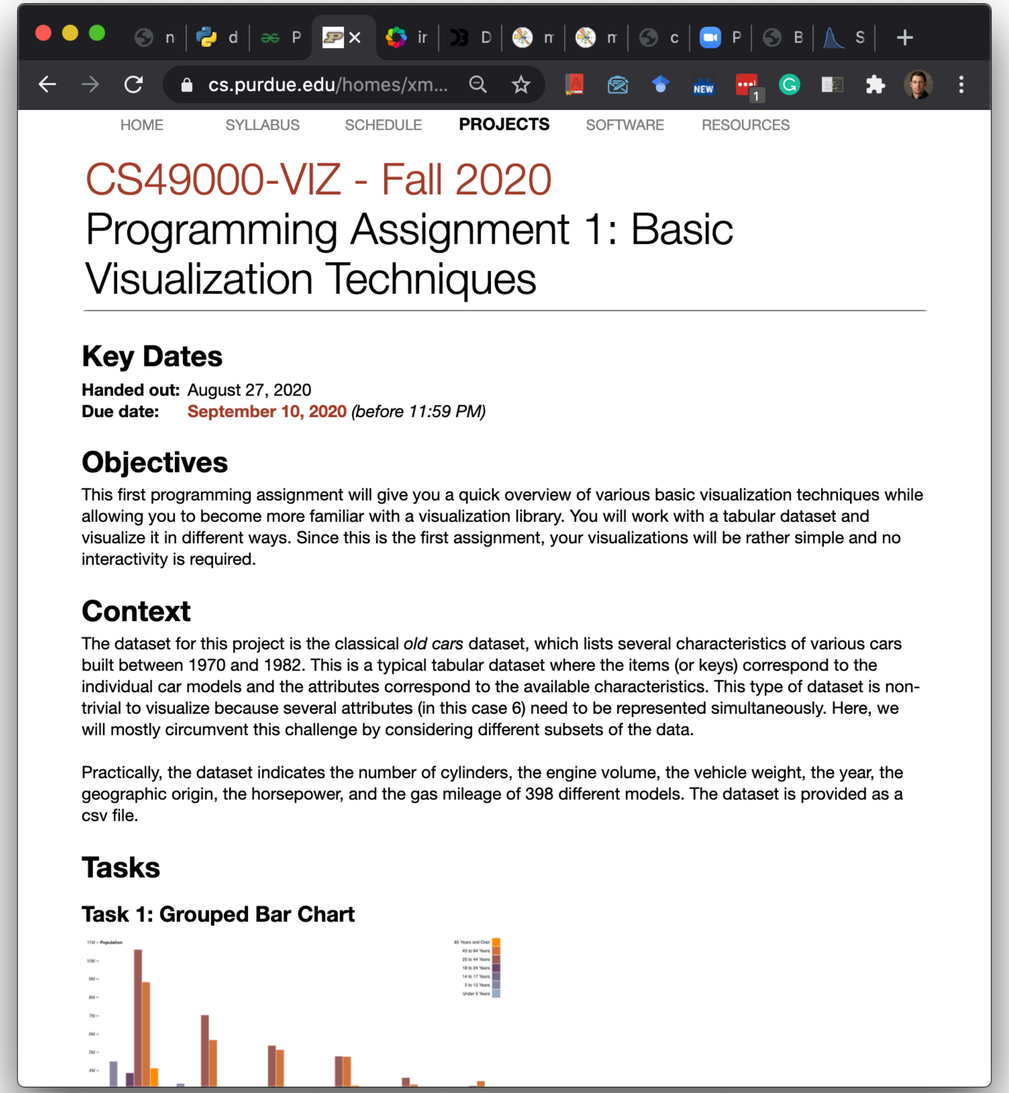
# Matplotlib vs Bokeh

- Scatter plot



# First Project




due 09/10