CS196

Data Science

Data Visualization (hardcore)

Week 5 of 196 (week 6 of school) :)

Week 4 Challenge Answers

1) Calculate the survival rates of passengers by class (First, Second, Third)

In [94]:
```
titanic_df.groupby('pclass').mean().survived
```

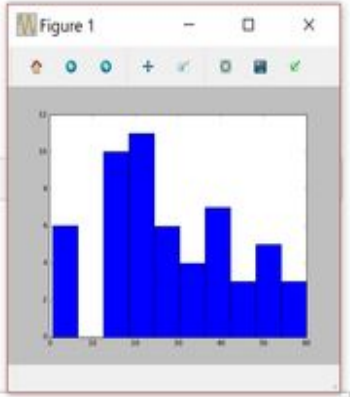
Out[94]:
```
pclass
1    0.629630
2    0.472826
3    0.242363
Name: survived, dtype: float64
```

2) Calculate the average fare paid by those who survived compared to the fare paid by those who didn't

In [65]:
```
titanic_df.groupby('survived').mean().fare
```

Out[65]:
```
survived
0    22.117887
1    48.395408
Name: fare, dtype: float64
```

3) Plot the ages of the female survivors that embarked at Cherbourg



In [*]:
```
import numpy as np
import matplotlib.pyplot as plt
data = titanic_df[titanic_df.embarked=='C'][titanic_df.sex=="female"][titanic_df.age==titanic_df.age][titanic_df.survived==1].age
plt.hist(data.values)
plt.show()
```

```
C:\Users\mattc\Anaconda2\lib\site-packages\ipykernel\__main__.py:3: UserWarning: Boolean Series key will be reindexed to match D
ataFrame index.
  app.launch_new_instance()
```

Data Visualization

Why Bother right?

- Easily Digestible / Sharable
  - Reduces the ‘wall-of-text’ effect
- Visually exposes patterns
- A medium for **explaining** and **exploring** data

Overview

Static

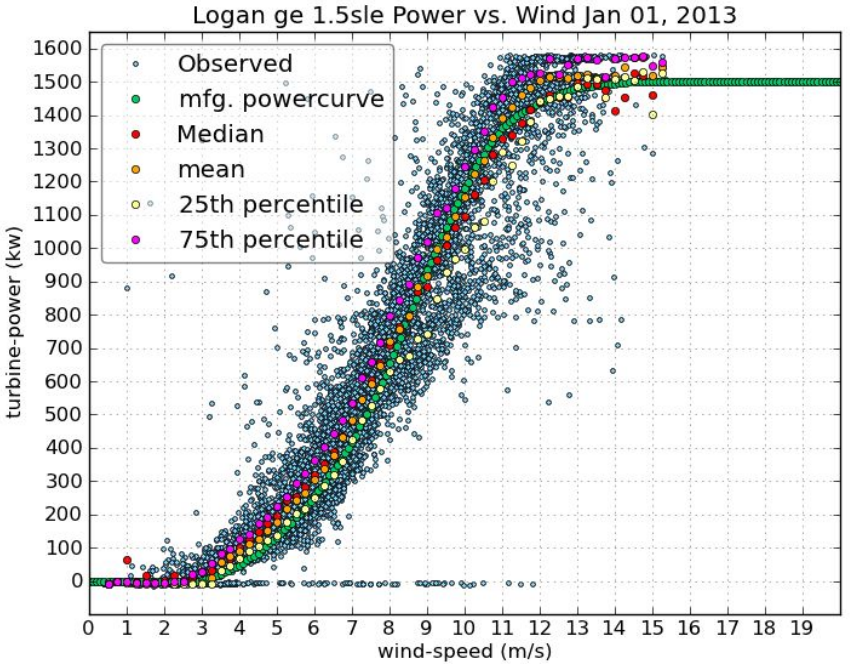
- Matplotlib (Python)
- Seaborn (Python)
- Ggplot2 (R)
- Chart.js (Python)

Dynamic

- D3.js (Javascript)
- Plotly (Proprietary w/ various APIs)
- Tableau (Proprietary)
- Highcharts (Javascript)

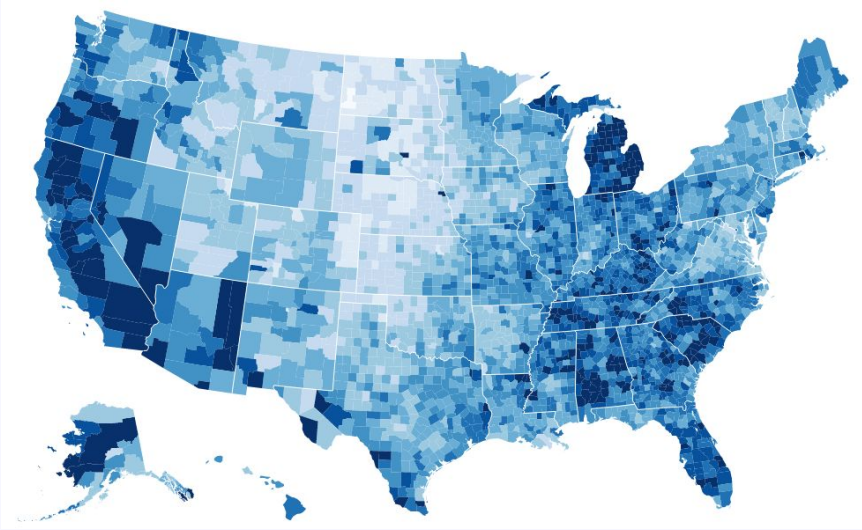
Matplotlib

- Many built-in plot types
- Used mostly for static visualizations
- Great for making ‘production-ready’ plots
- Also used heavily in academia



D3.js

- Best for in-browser visualizations
- Not as 'batteries included' as matplotlib
    - D3 is a DOM-data binding library at its root, forces you to do a lot of boilerplate



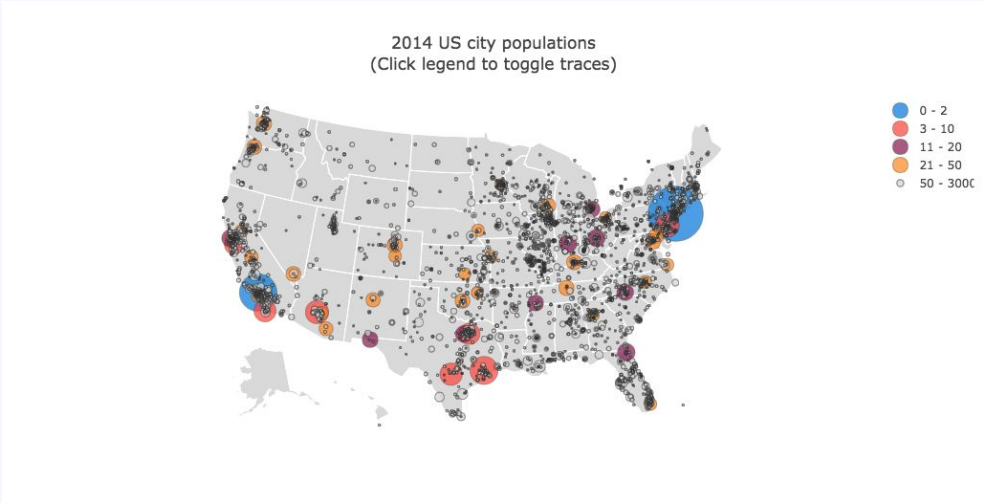
C3.js and Friends

- Lots of alternative D3-derived
  - Similar to how many Python libraries are derived from matplotlib
- Examples:
  - C3.js - D3-based
  - Recharts - D3 charts for React
  - NVD3 - Reusable D3-based charts
  - Chart.js - Simple JS Charting Library
  - Chartist - “Simple Responsive Charts”

CHARTIST.JS

Plotly

- Dynamic Plots
  - Gives you dynamic plots without having to ‘reinvent the wheel’ each time
- APIs for various languages (Python, JS, R, etc.)
- Not everything is free... :(



Jupyter Time

http://github.com/CS196Illinois/Data_Hackerspace

CS196

Week 5 Challenge

Surprise us with visualizations