

# Analytics Jumpstart

**pandas methods for exploratory analysis**

---

Nashville Software School



# For today

- More pandas

- `df.value_counts()`

- `df.describe()`

- `df.info()`

- `df.reset_index()`

- Intro to Exploratory Data Analysis

- methods for learning more about the data
  - plots for learning more about the data



# Get Data / Process + Clean Data / Exploratory Data Analysis

## Statistics and other info

*series.value\_counts()* – returns the frequency of each unique value in a pandas series (or DataFrame column)

*series.reset\_index()* – moves the index value to a column and converts the series to a DataFrame

*df.describe()* – to get summary statistics about quantitative data

*df.info()* – to get information about the DataFrame

*df.isnull().sum()* – to get counts of missing values



# Get Data / Process + Clean Data / Exploratory Data Analysis

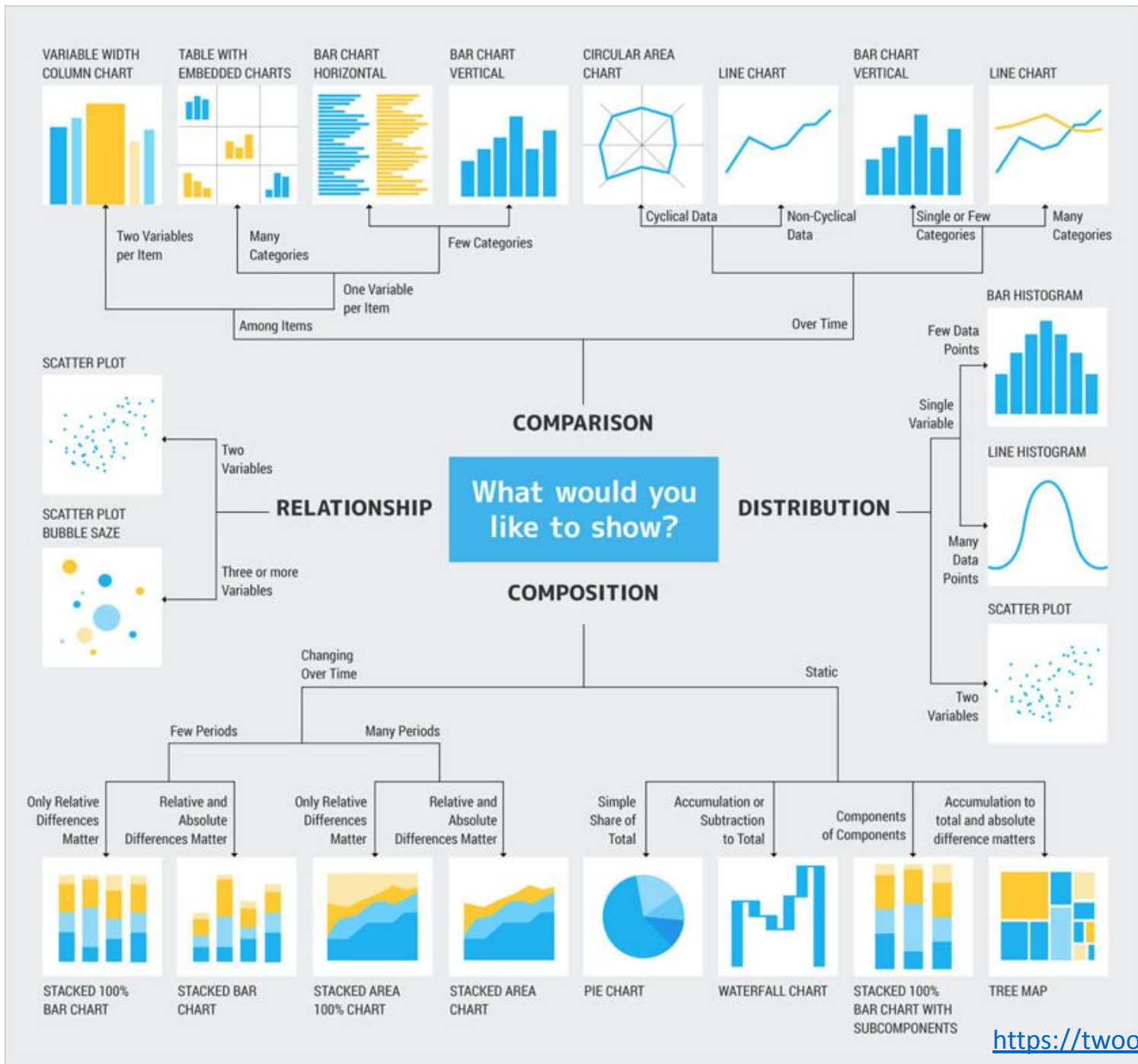
## Exploratory Plots

```
import matplotlib.pyplot as plt  
import seaborn as sns
```

[https://matplotlib.org/api/\\_as\\_gen/matplotlib.pyplot.plot.html#examples-using-matplotlib-pyplot-plot](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html#examples-using-matplotlib-pyplot-plot)

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





# Reminders

- Build upon your work in the same notebook each week. Just open it and add to it.
- If the code in a cell did not run as expected, modify the code in that cell (not a new one)
- Remove any unused/un-useful cells
- Beware of the changing state of objects in your notebook
  - Example – if you create a df and drop 3 columns and then go back to add code to look at the head() in the same cell, you are re-running the command to drop the 3 columns which are no longer there!



# Questions?

