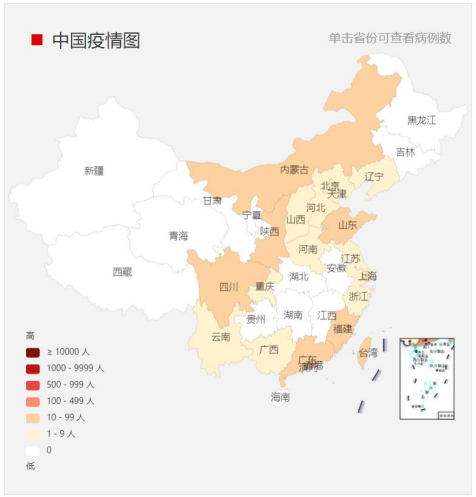
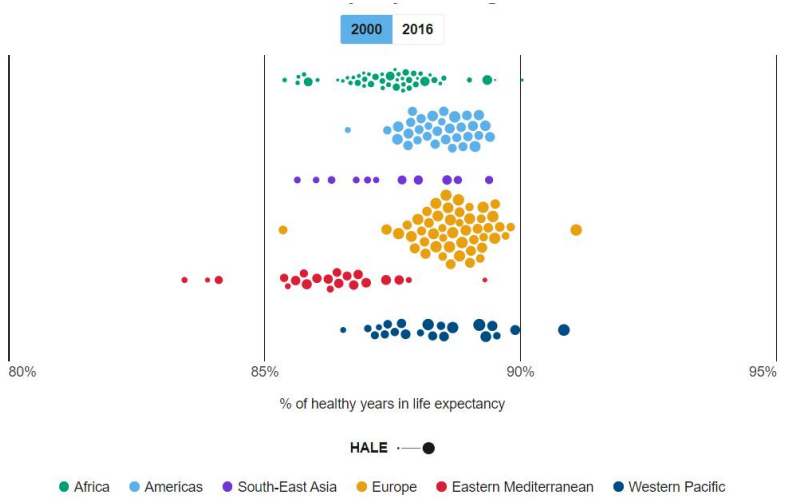
Knowledge Discovery and Data Mining

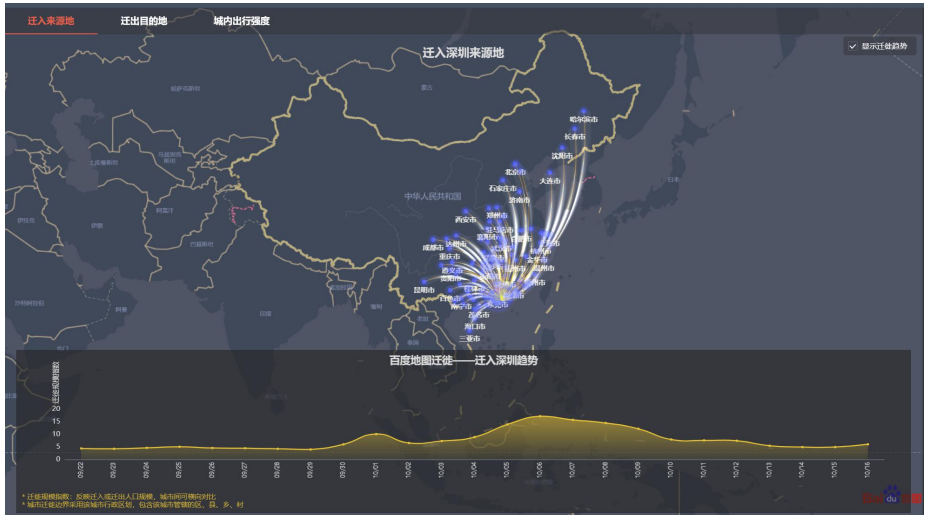
Lab 10 Plotting Data on a Graph and choose the best Graph

Xuan Song
Songx@sustech.edu.cn

A picture is worth a thousand words

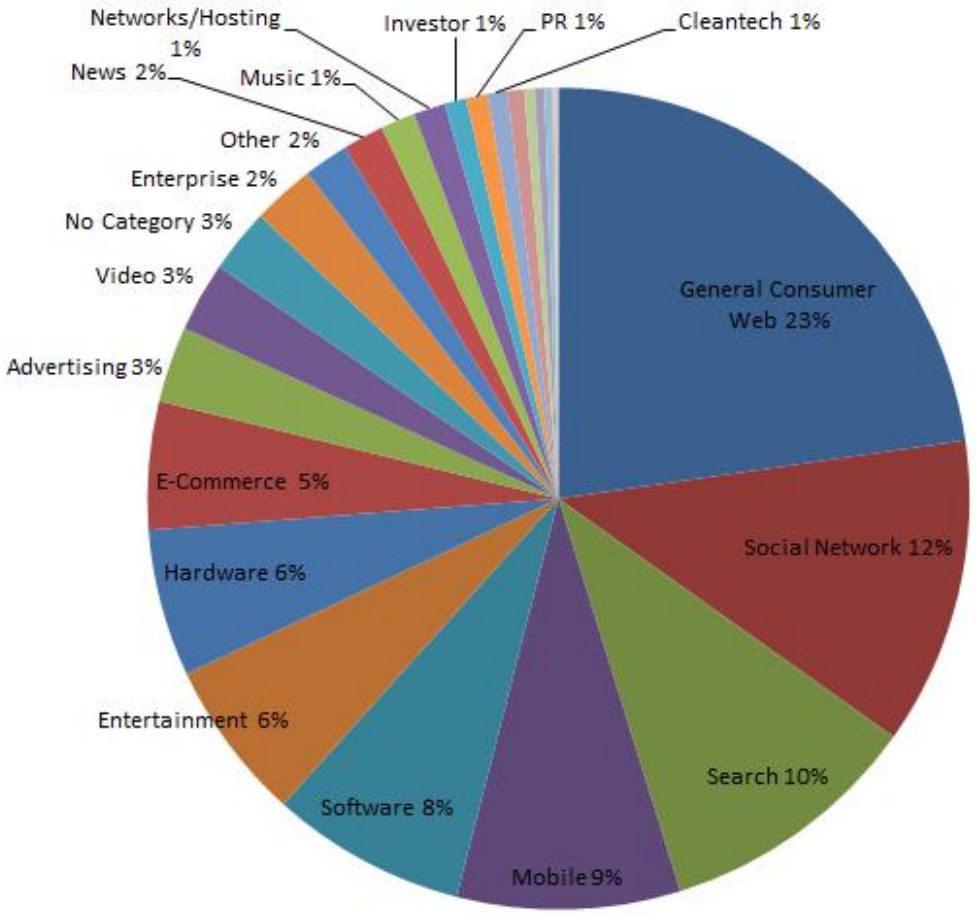






https://wp.m.163.com/163/page/news/virus_report/index.html?_nw_=1&_anw_=1
https://www.who.int/data
https://qianxi.baidu.com/

What to Show?



A Graph like this doesn't really show you a lot!

Basic Graph

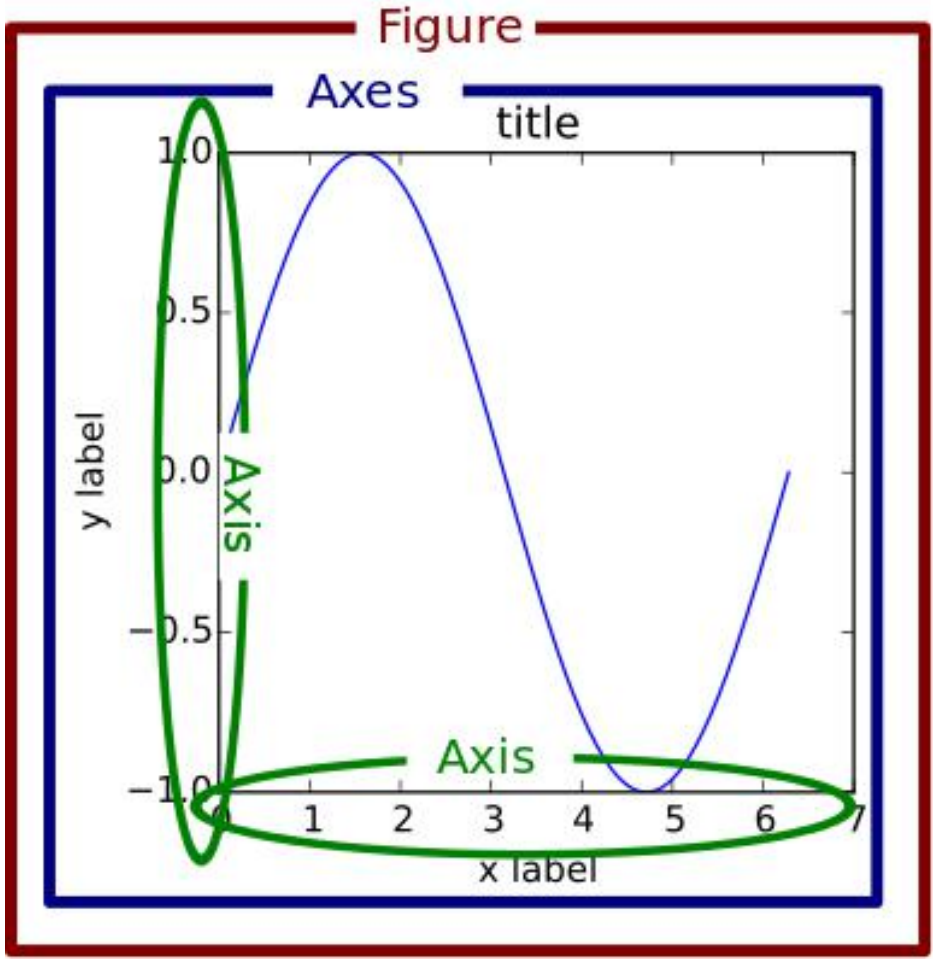


Figure
Axes
title
1.0
0.5
0.0
−0.5
−1.0
y label
Axis
Axis
0
1
2
3
4
5
6
7
x label

Basic Graph

- Plot creation
- Plotting routines
- Basic plot customizations
- Showing and saving your plots

Sample Code

```
import matplotlib.pyplot as plt
# Set the size of figure
plt.figure(figsize=(8,4))

# Prepare the data
x = [1,2,3,4,5,6,7,8]
y = [1,2,3,4,5,6,7,8]
# Plot the data
plt.plot(x, y, 'go',label='dots')

plt.legend()        # Show legend
plt.title('Title') # Add title
plt.xlabel('X')    # Add x label
plt.ylabel('Y')    # Add y label

#save figure
plt.savefig('dots.png', dpi=300)
# Show the plot
plt.show()
```

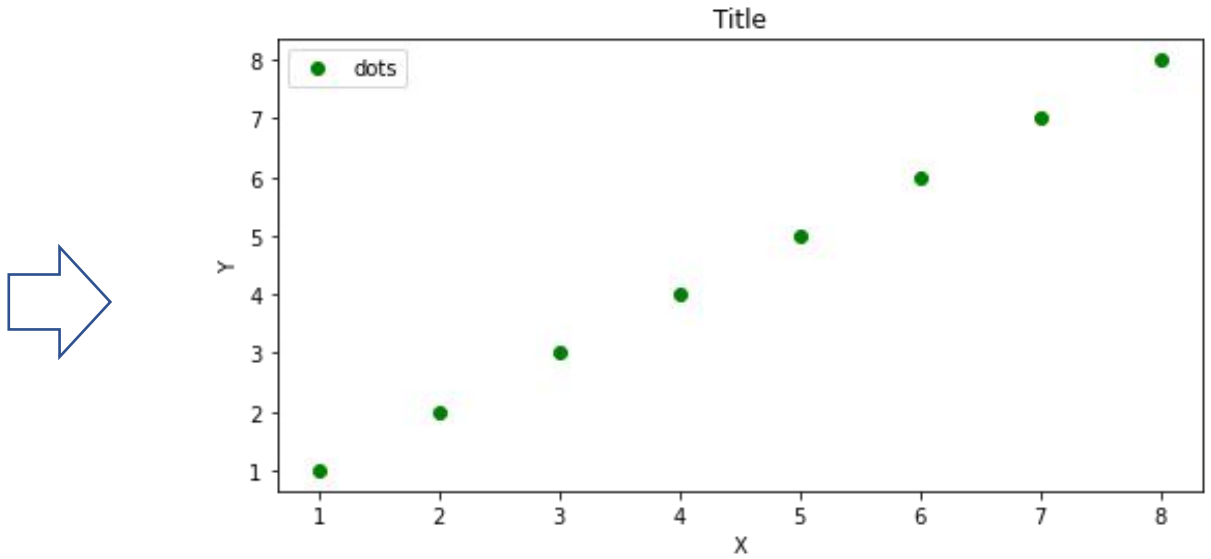
Basic Graph

```
import matplotlib.pyplot as plt
# Set the size of figure
plt.figure(figsize=(8,4))

# Prepare the data
x = [1,2,3,4,5,6,7,8]
y = [1,2,3,4,5,6,7,8]
# Plot the data
plt.plot(x, y, 'go',label='dots')

plt.legend()        # Show legend
plt.title('Title') # Add title
plt.xlabel('X')    # Add x label
plt.ylabel('Y')    # Add y label

#save figure
plt.savefig('dots.png', dpi=300)
# Show the plot
plt.show()
```



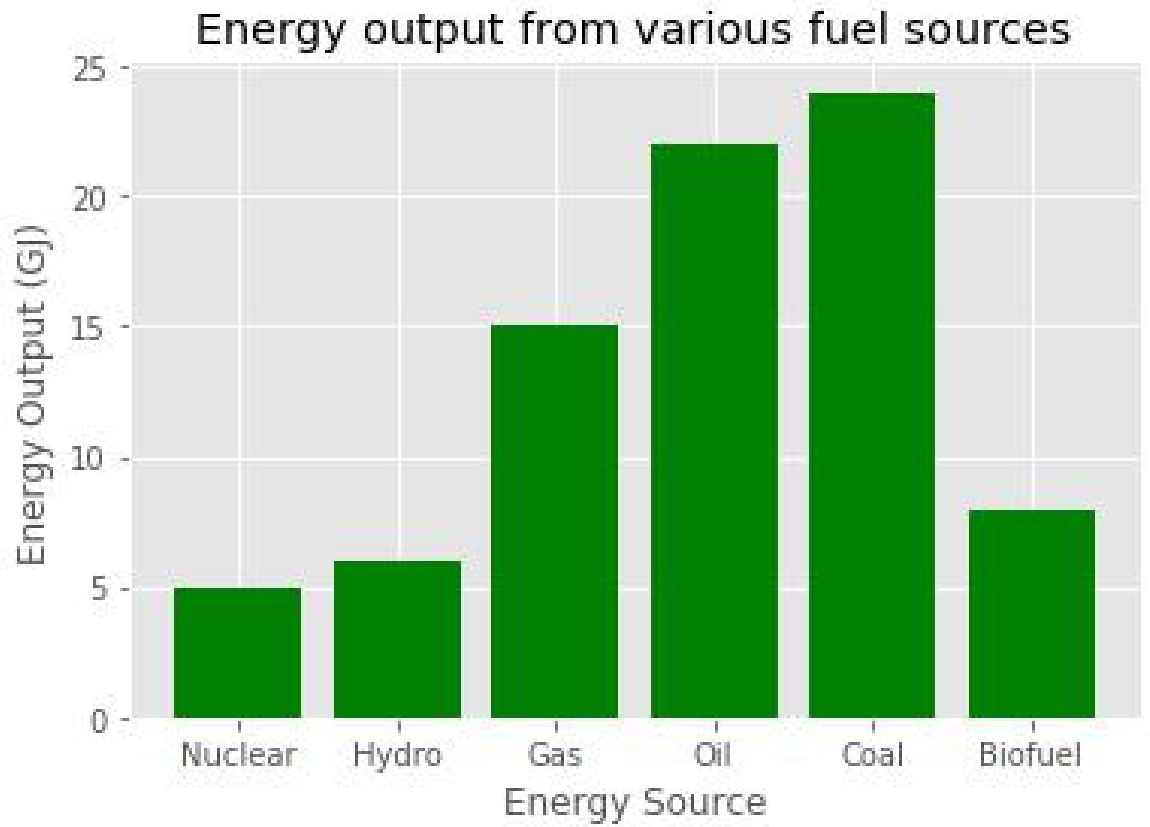
If you have any problems about the functions of matplotlib, you could refer to the following link:
https://matplotlib.org/3.3.2/api/pyplot_summary.html

What can we accomplish by a graph?

Generally, there are 5 different things you can accomplish with a graph:

1.Comparison
2.Composition
3.Distribution
4.Relationship
5.Trending

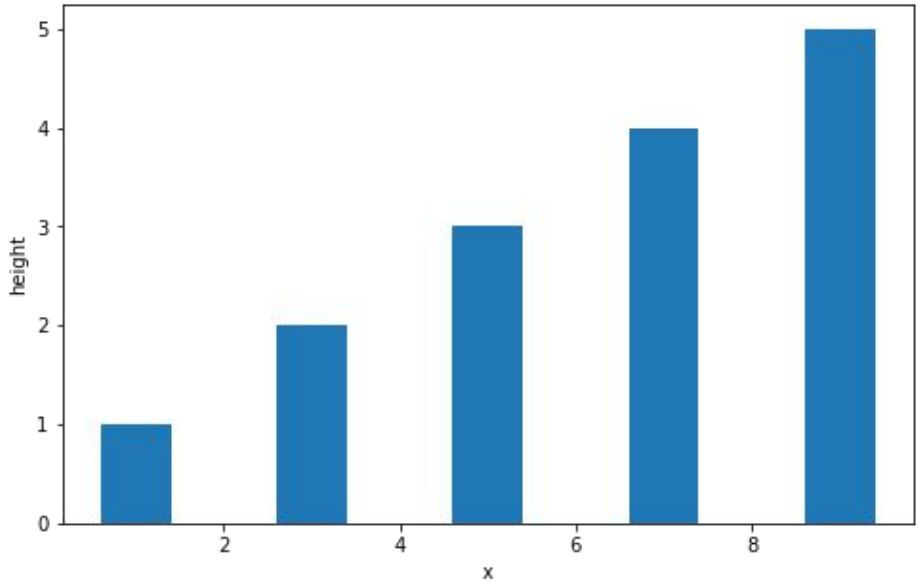
Bar/Column



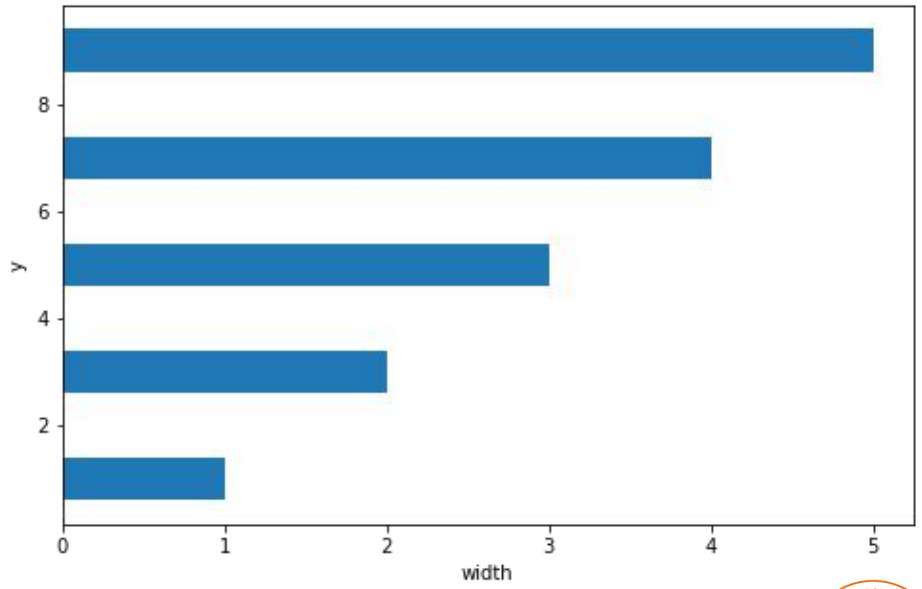
- **Use consistent colors throughout the chart**, selecting accent colors to highlight meaningful data points or change over time.
- **Use horizontal labels** to improve readability.
- **Starts the y-axis at 0** to appropriately reflect the values in your graph.

Bar Chart

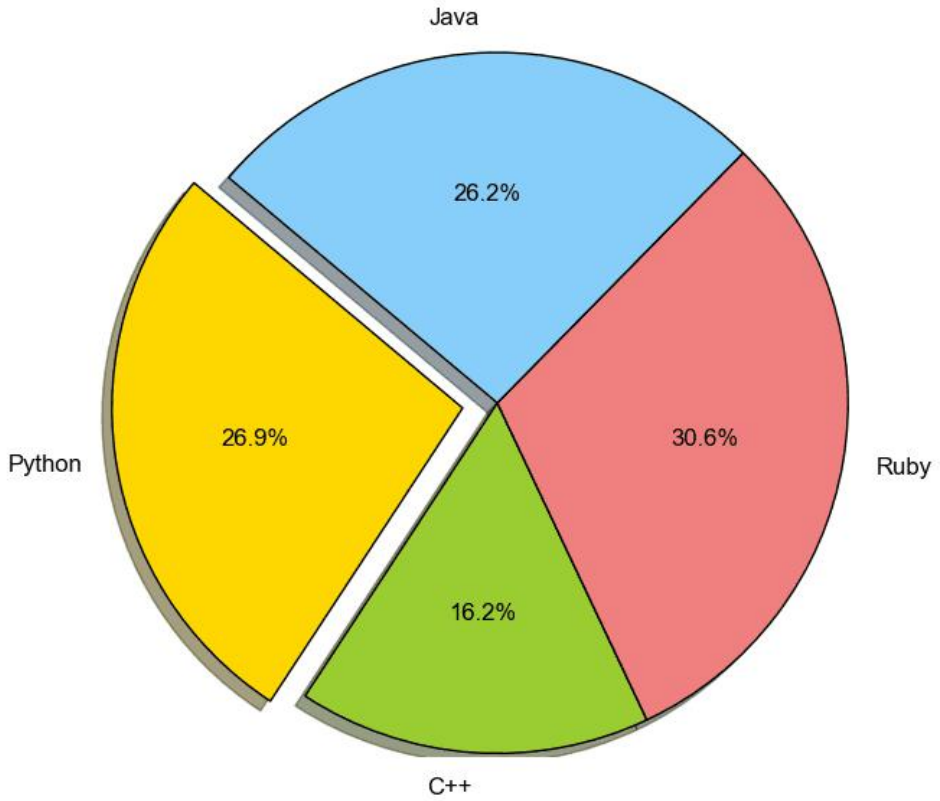
- Make a bar plot:
  - `matplotlib.pyplot.bar(x, height, width=0.8)`



- Make a horizontal bar plot:
  - `matplotlib.pyplot.barh(y, width, height=0.8)`

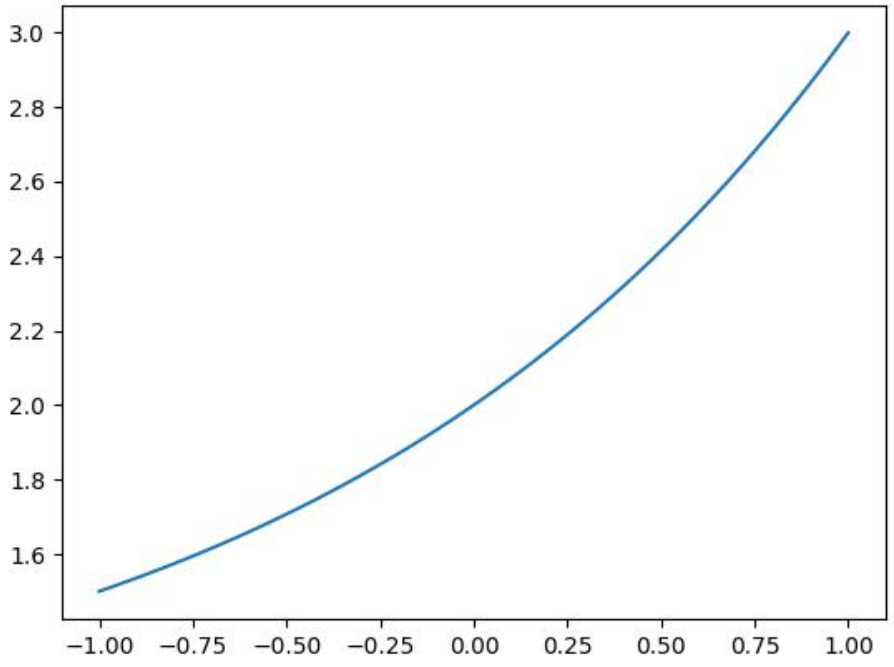


Pie



- **Don't illustrate too many categories** to ensure differentiation between slices.
- **Ensure that the slice values add up to 100%.**
- **Order slices according to their size.**

Line

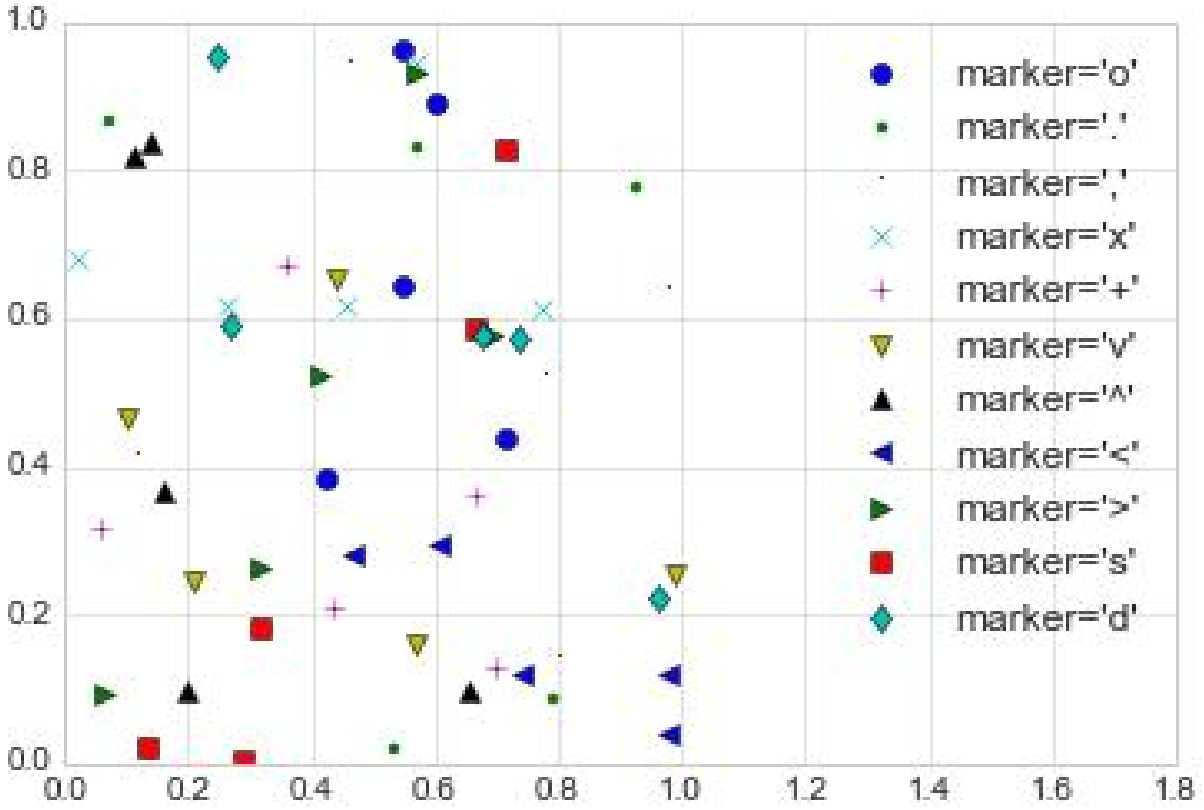


- **Use solid lines only**.
- **Don't plot more than four lines** to avoid visual distractions.
- **Use the right height** so the lines take up roughly 2/3 of the y-axis' height.

Line Chart

- Plot y versus x as lines and/or markers:
  - `plot(x, y, 'go--', linewidth=2, markersize=12)`
  - `plot(x, y, color='green', marker='o’, linestyle='dashed',linewidth=2, markersize=12)`

Scatter



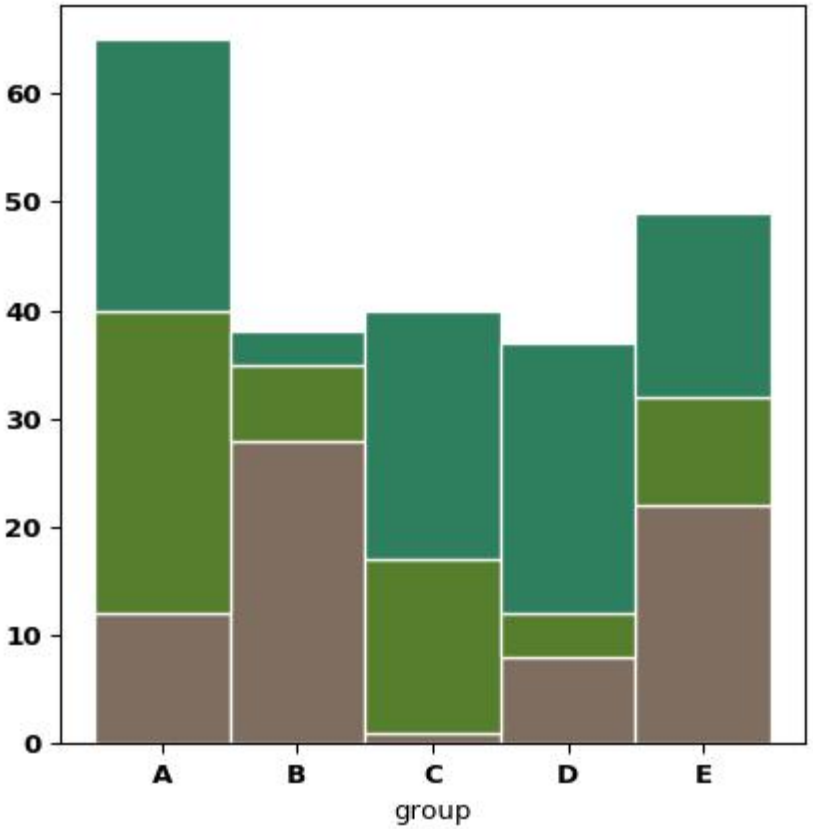
- **Include more variables**, such as different sizes, to incorporate more data.
- **Start y-axis at 0** to represent data accurately.
- **If you use trend lines, onle use a maximum of two** to make your plot easy to understand.

Scatter plot

- A pandas scatter plot
  - `pandas.DataFrame.plot.scatter(x = col_x_name, y = col_y_name,s=None)`

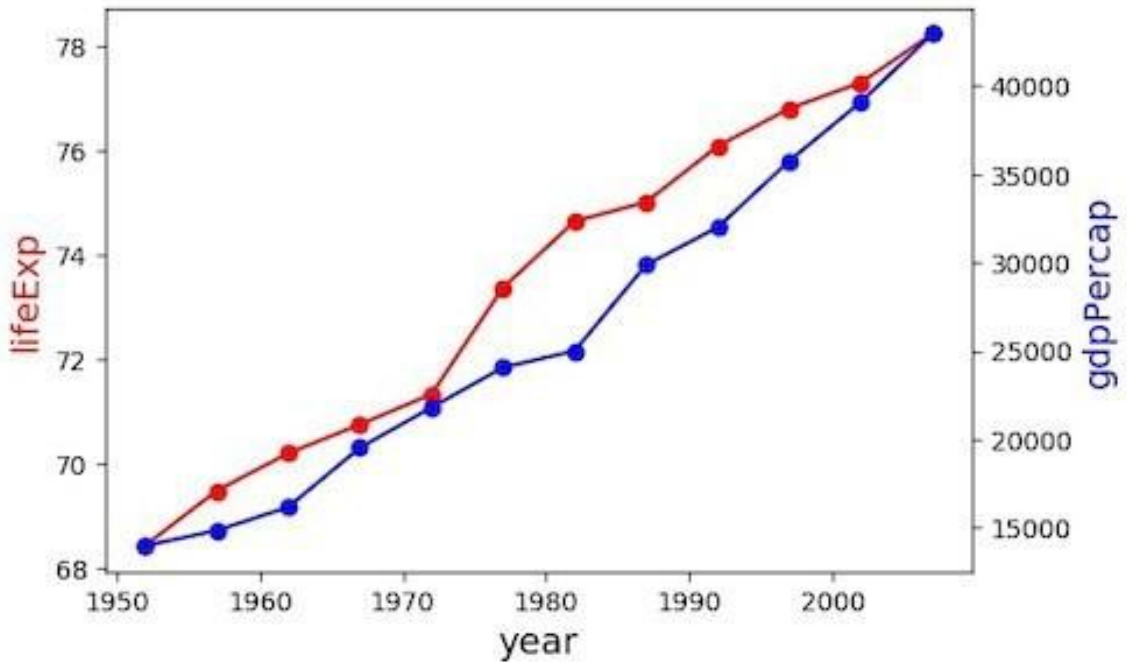
- A matplotlib scatter plot
  - `matplotlib.pyplot.scatter(x,y,s=None)`

Stacked Bar/Column



- **Best used to illustrate part-to-whole relationships**.
- **Use contrasting colors for greater clarity**.
- **Make chart scale large enough** to view group sizes in relation to one another.

Dual-Axis



- **Use the y-axis on the left side for the primary variable** because brains are naturally inclined to look left first.
- **Use different graphing styles** to illustrate the two data sets.
- **Choose contrasting colors for the two data sets**.

Exercise 1

Freely explore the following datasets and present meaningful outcomes.

Exercise 2

Compare countries by happiness and other human metrics.

EXTRA

Find out “Dragons Effect on Winning”, and predict game outcome.

End of Lab 10