

Data Visualization Techniques Using Python

13 - 16 July, 2020

© ViSER LLC

Layout of Bootcamp



- Day 1
 - Python recap
 - Data Visualization basics
 - Intro to Matplotlib
- Day 2
 - Bokeh/Seaborn installation
 - Intro to Seaborn
 - Plotting with Seaborn
- Day 3
 - Seaborn continued
 - Intro to Bokeh
 - Plotting with Bokeh
- Day 4
 - Plotting with Bokeh, continued
- Important information
 - email
viserllc@gmail.com
 - course material link

PYTHON RECAP



- Packages/Library (pandas, numpy, matplotlib, seaborn, scikit)
- dataframe
- accessing dataframe columns/rows
- python functions
- csv (comma separated values) files
- Anaconda distribution
- Jupyter notebook

Data Visualization

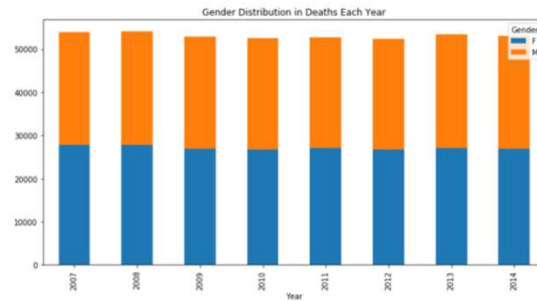


- Presentation of data and analysis with **pictures** or **graphs/charts**
- Enables the visual presentation of analytics
 - easy to grasp difficult concepts
 - can identify new patterns
- Caters to the way, human brain processes information
 - using charts or graphs to visualize large amounts of complex data is **easier** than scanning over spreadsheets or reports
- **Very important part of data analysis**

Analysis Result - Table vs Chart



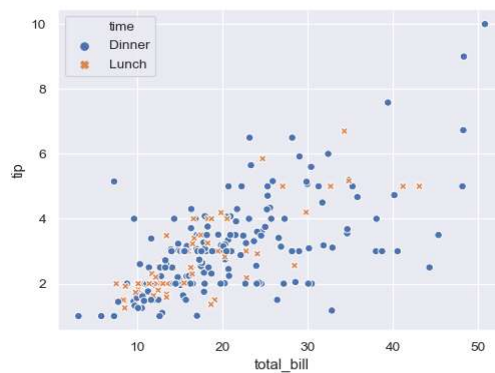
Year	Gender	
2007	F	27749.0
	M	26247.0
2008	F	27816.0
	M	26322.0
2009	F	26941.0
	M	25879.0
2010	F	26675.0
	M	25830.0
2011	F	27075.0
	M	25651.0
2012	F	26766.0
	M	25654.0
2013	F	27133.0
	M	26254.0
2014	F	26916.0
	M	26090.0



Types of Charts(Graphs)

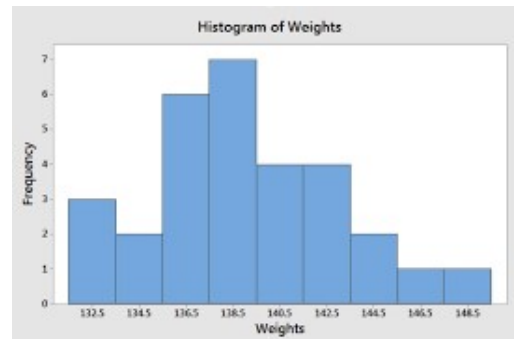
Scatter Plots

- Shows relationship between 2 variables
- Suitable for numerical data



Histogram

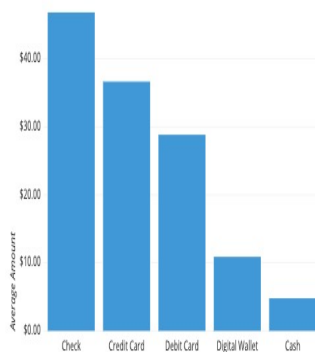
- A Series of bins, showing the frequency of observations of a given variable
- Used for numerical data



Types of Charts(Graphs)

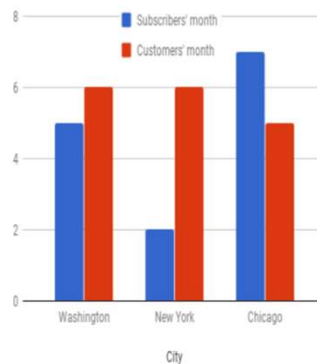


Bar Graphs (Simple)

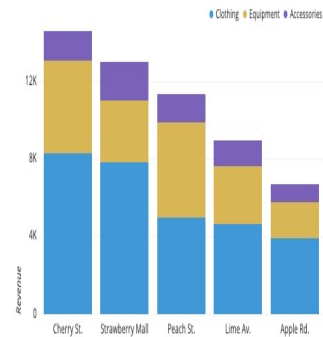


Bar Graphs (Grouped)

Subscribers' month and Customers' month



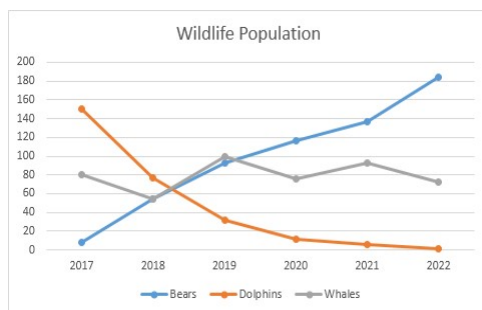
Bar Graphs (Stacked)



Types of Charts(Graphs)

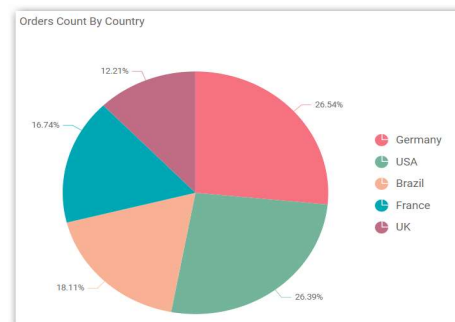
Line Charts

- Good for time-series data
- One line represent one variable
- Can be used to see the movement of several variables



Piechart

- Circular graph divided into slices
- Each slice represents the proportion of one variable in whole

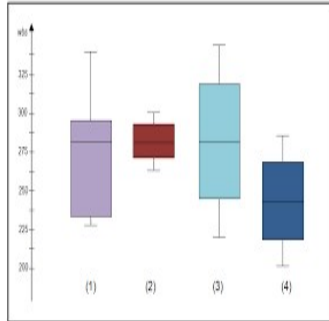


Types of Charts(Graphs)



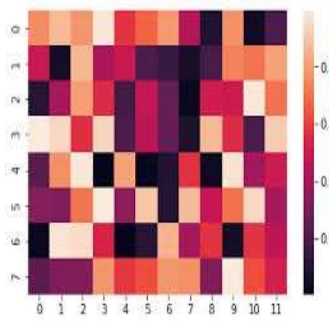
Box Plots

Standard way of displaying the dataset based on 5 properties: the minimum, the maximum, the sample median, and the first and third quartiles



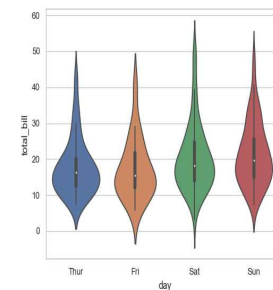
Heatmap

Two-dimensional representation of data in which values are represented by colors



Violin Plot

Similar to box plots, except that they also show the probability density of the data at different values



Data Visualization Python Libraries



- Python has several data visualization libraries to create very simple to very complex visualizations
- Some of them are:
 - Matplotlib
 - Seaborn
 - Bokeh
 - Plotly
 - GGplot

Matplotlib

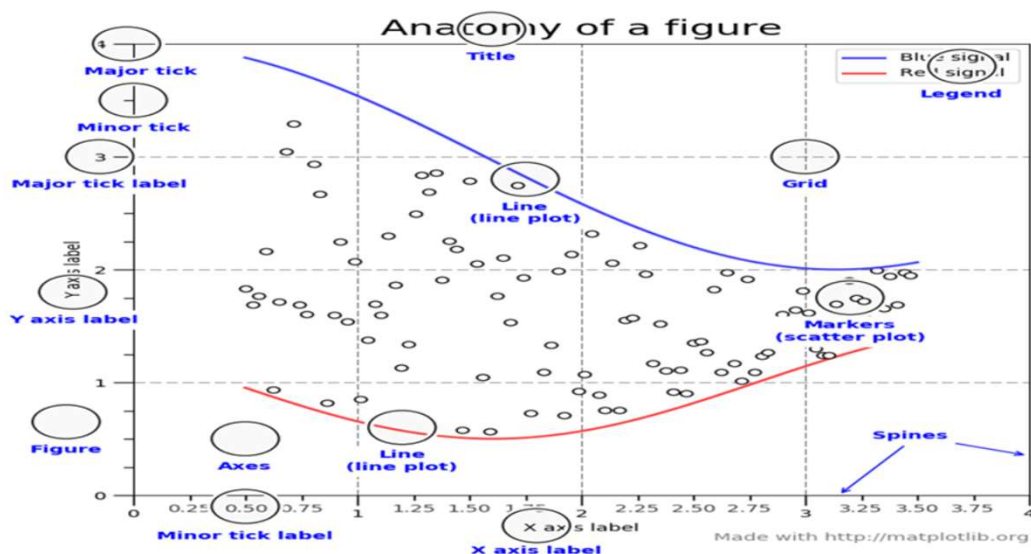


- Very flexible 2-D plotting library in Python
- Supports all the popular charts (lots, histograms, power spectra, bar charts, error charts, scatterplots, etc)
- There are extensions that can be used to create advanced visualizations like 3-D plots
- Can be customized at every level

Important links:

- <https://matplotlib.org/3.1.0/gallery/index.html>
- [matplotlib official documentation \(https://matplotlib.org/users/index.html\)](https://matplotlib.org/users/index.html)

Components of a Matplotlib Figure



Matplotlib Basics



- Matplotlib graphs data on

- **Figures**

- like windows, Jupyter widgets, etc.
 - keeps track of titles, figure legends, etc

- **Axes**

- area where points can be specified in terms of x-y coordinates (or theta-r in a polar plot, or x-y-z in a 3D plot, etc.)

- **Axis**

- takes care of setting the graph limits and generating the ticks (the marks on the axis) and ticklabels

A given figure can contain many Axes, but a given Axes object can only be in one Figure

fig, ax = plt.subplots() # Create a figure containing a single axes.

ax.plot([1, 2, 3, 4], [1, 4, 2, 3]) # Plot some data on the axes.

Bokeh/Seaborn Installation



Windows

- Step 1:

- Go to Windows app and find "Anaconda Navigator" OR
 - type "Anaconda Navigator" in windows search bar (bottom left corner of screen)

- Step 2

- Click "Powershell Prompt" ((base) PS C:\Users\bipva> --- (yours will be different depending on where you have installed Anaconda)
 - Type `pip install seaborn / pip install bokeh` (for latest version) OR
 - Type `conda install seaborn / conda install bokeh` (for any version)
 - Follow the instructions

- Step 3

- Open Jupyter Notebook from Anaconda Navigator (Mac) OR typing "jupyter notebook" in windows search bar