



Clinical Graphs using Python Packages

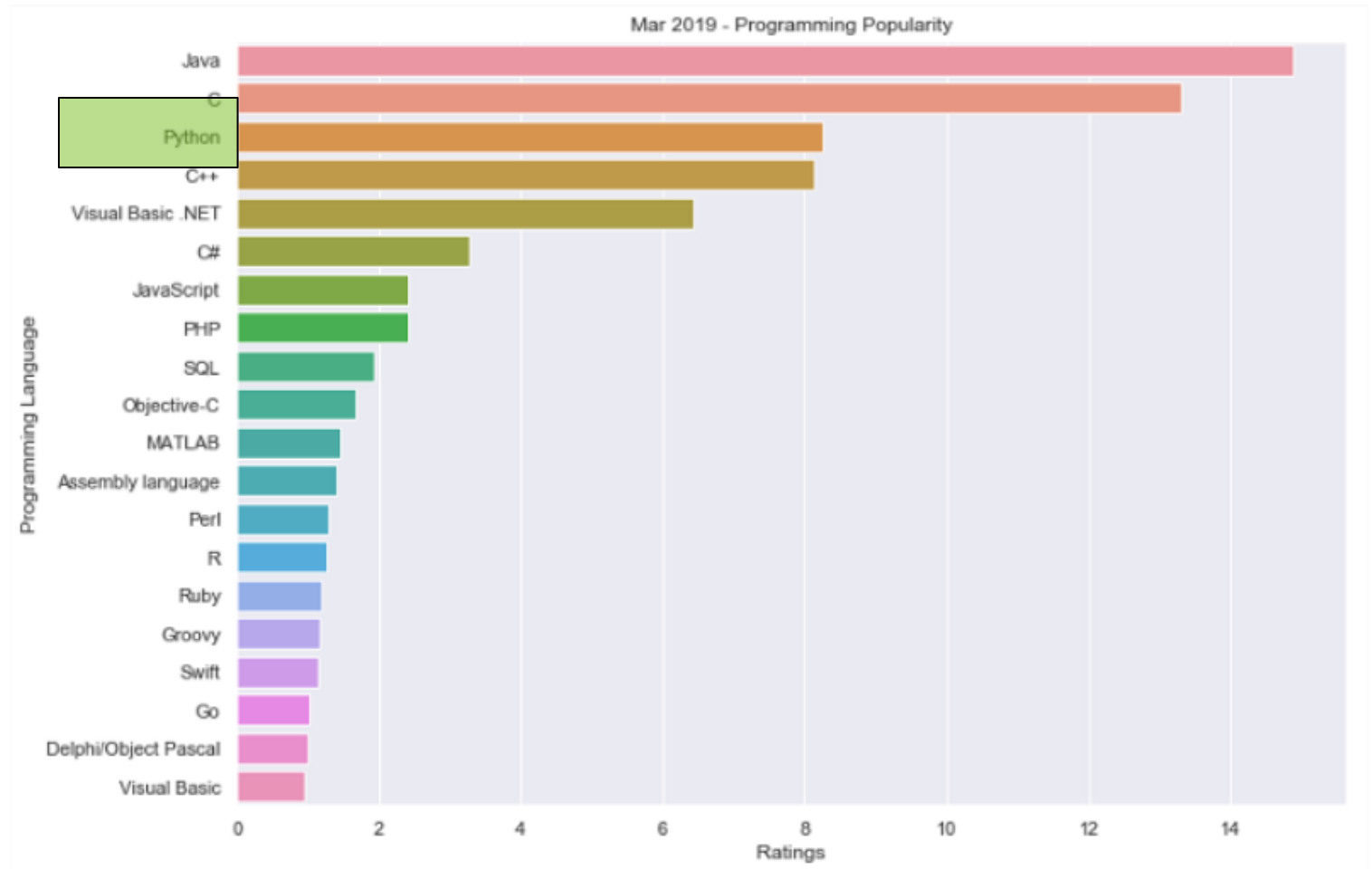
PHUSE EU Connect 2020

Siddharth Kumar, Associate Vice President, Navitas Data Sciences

Disclaimer

 This Presentation does not represent views of Navitas Data Sciences and belong solely to the author. I take full responsibility for the work and errors if any.

“the art of uncovering the insights and trends in data”



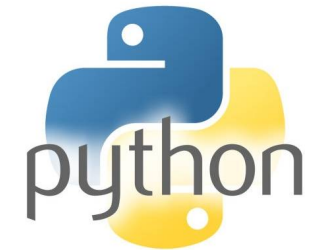
Reference: <https://towardsdatascience.com/visualize-programming-language-popularity-using-tiobeindexpy-f82c5a96400d>

Agenda

- Introduction to Python
- Important Python Packages
- Python Packages for Data Visualization
- Clinical Graph Examples
 - Bar Chart
 - Line Chart
 - Scatter Plot
 - Box Plot
 - 95% Confidence Interval
- 3D & Interactive Figures
- Summary

Introduction to Python

- Python is a popular, powerful open source interpreted programming language by Guido van Rossum and released in 1991.
- Its a high-level language and platform that could be used for both research and development and developing production systems.
- Simple, readable syntax which is similar to the English language.
- Python has various Libraries or Packages to choose from, which makes it a complete programming language
- There are various user created modules which are shared with community through repositories such as GitHub.
- Python is open source and could be downloaded through distribution channel such as Anaconda. **Anaconda** is popular because it brings many of the tools used in data science and machine learning with just one install
- Install IDE (Inegrated Development environment) – provides tools to facilitate Programming
- Jupyter Notebook, PyCharm and Spyder are widely used IDEs



Important Python Packages

scikit-learn

is a Machine Learning library. Built on NumPy, SciPy and matplotlib

NumPy, datetime

is used for working with multi dimensional arrays. The array object in NumPy is called **ndarray**

pandas

is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool

Matplotlib, Seaborn, ggplot, plotly

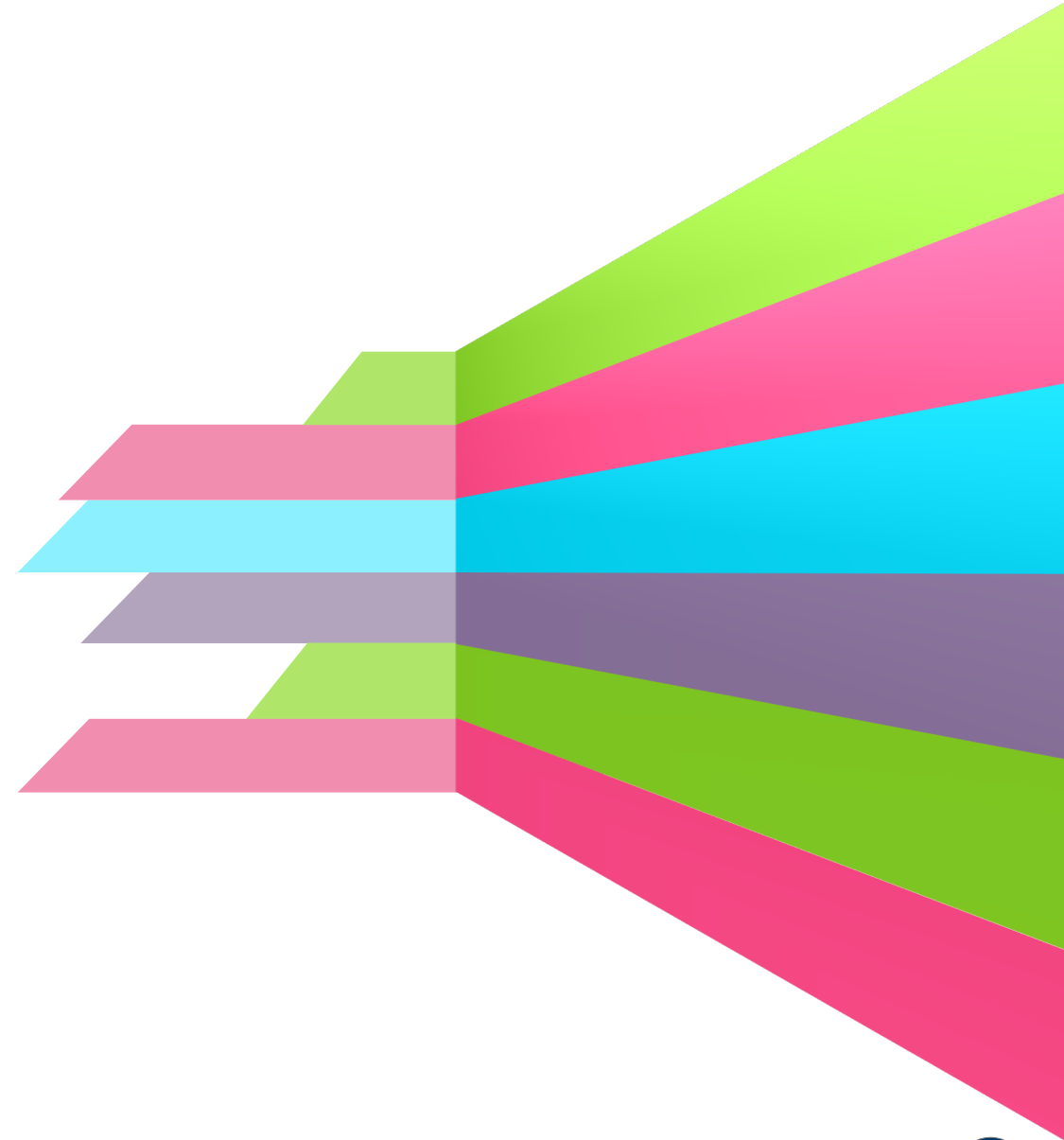
is a comprehensive library for Data Visualization.

xlrd, pdfminer

Text mining packages

SASPy

allows Python coders to access SAS data and analytics capabilities.



Python Packages for Data Visualization

Matplotlib

- More than a decade old, widely used
- Used to generate simple and powerful visualizations.

Seaborn

- Built on top of Matplotlib.
- More sophisticated default styles and color palettes

ggplot2

- Ggplot is a Python visualization library based on R's ggplot2 and high level Grammar of Graphics.

Matplotlib

The versatility of Matplotlib can be used to make visualization types such as:

.plot() has several optional parameters.

```
# Example 1 #
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline

#read SAS Datasets
dm = pd.read_sas(filepath_or_buffer = 'demog.sas7bdat',
                 format = 'sas7bdat',
                 encoding="utf-8")

dm = dm.groupby('Age').mean()
dm.plot()

plt.title('PHUSE Europe Connect 2020 - Mean of Height & Weight')
plt.xlabel('Age')
plt.ylabel('Mean')

plt.show()
```

1. "area" is for area plots.
2. "bar" is for vertical bar charts.
3. "barh" is for horizontal bar charts.
4. "box" is for box plots.
5. "hexbin" is for hexbin plots.
6. "hist" is for histograms.
7. "kde" is for kernel density estimate charts.
8. "density" is an alias for "kde".
9. "line" is for line graphs.
10. "pie" is for pie charts.
11. "scatter" is for scatter plots.

Bar Chart

```
# Example 2 #
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

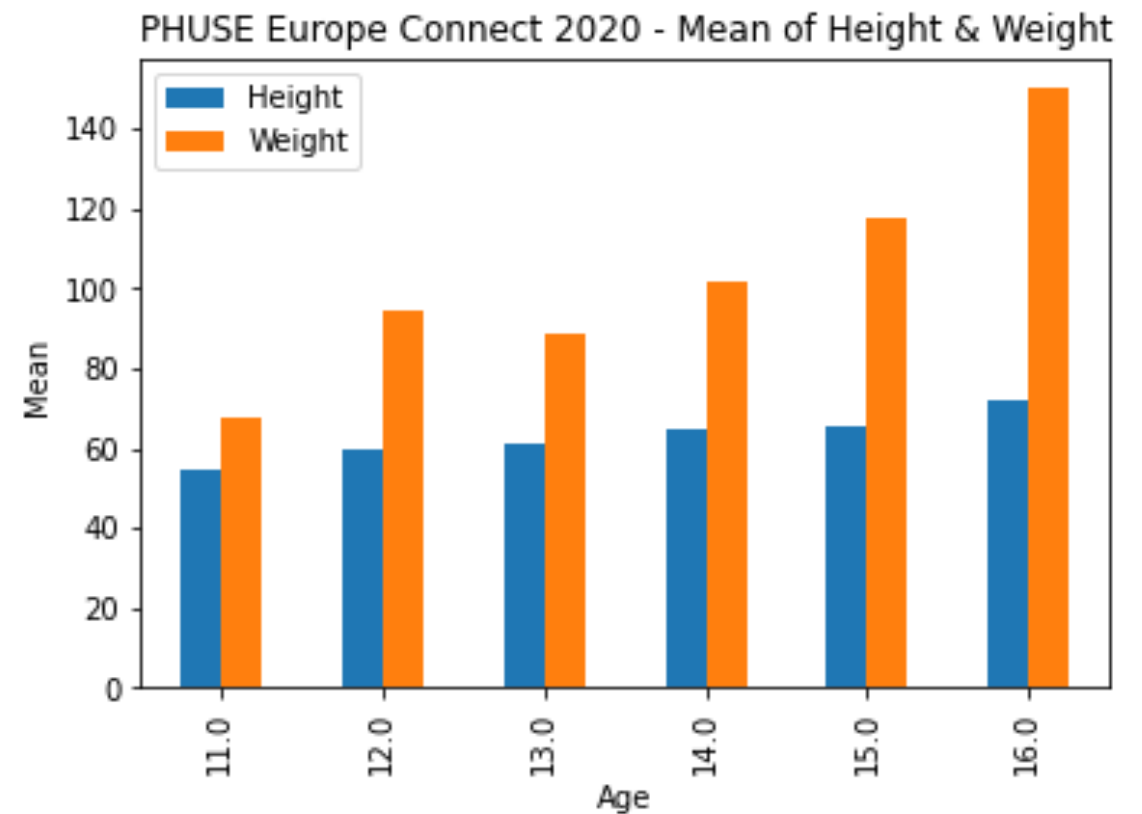
%matplotlib inline

#read SAS Datasets
dm = pd.read_sas(filepath_or_buffer = 'demog.sas7bdat',
                 format = 'sas7bdat',
                 encoding="utf-8")

dm = dm.groupby('Age').mean()
dm.plot.bar()

plt.title('PHUSE Europe Connect 2020 - Mean of Height & Weight')
plt.xlabel('Age')
plt.ylabel('Mean')

plt.show()
```



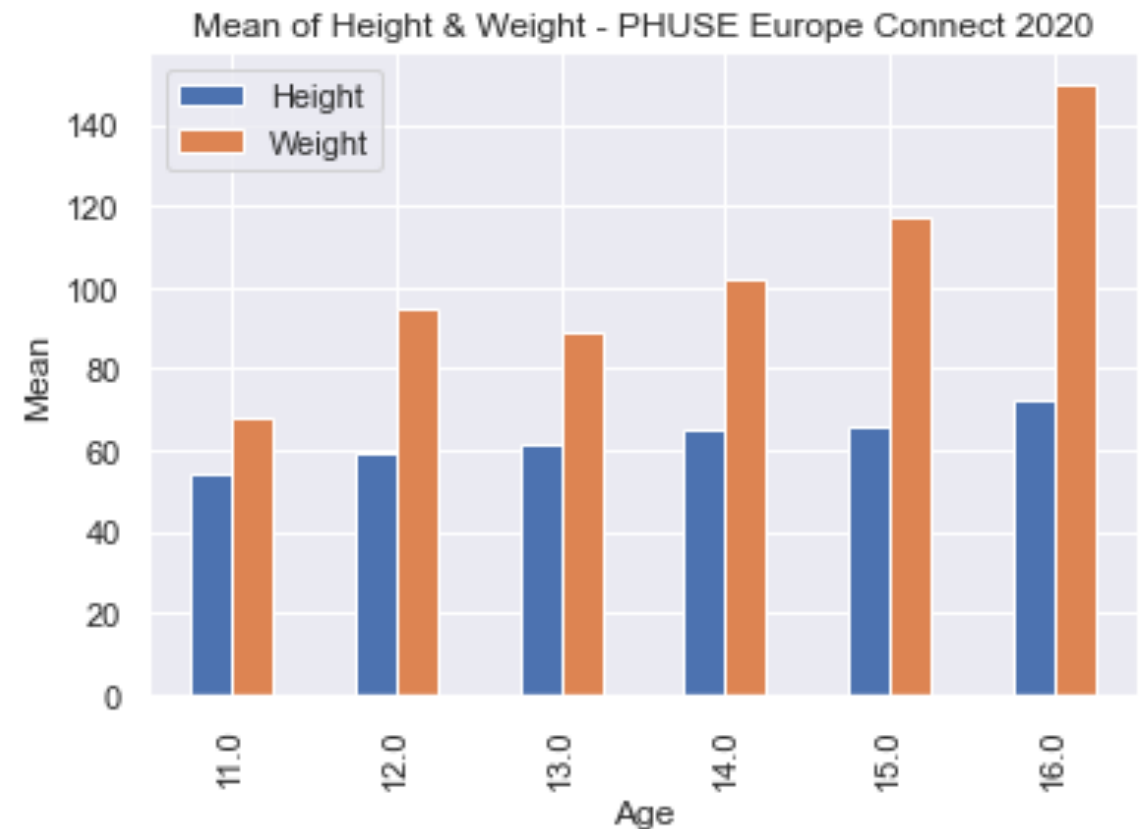
```
# Example 3 - Seaborn
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set()

#read SAS Datasets
dm = pd.read_sas(filepath_or_buffer = 'demog.sas7bdat',
                 format = 'sas7bdat',
                 encoding="utf-8")

dm = dm.groupby('Age').mean()
dm.plot.bar()

plt.title('Mean of Height & Weight - PHUSE Europe Connect 2020')
plt.xlabel('Age')
plt.ylabel('Mean')
plt.show()
```



Scatter Plot

```
# Example 4 - Scatter Plot #
# PHUSE Europe Connect 2020 #

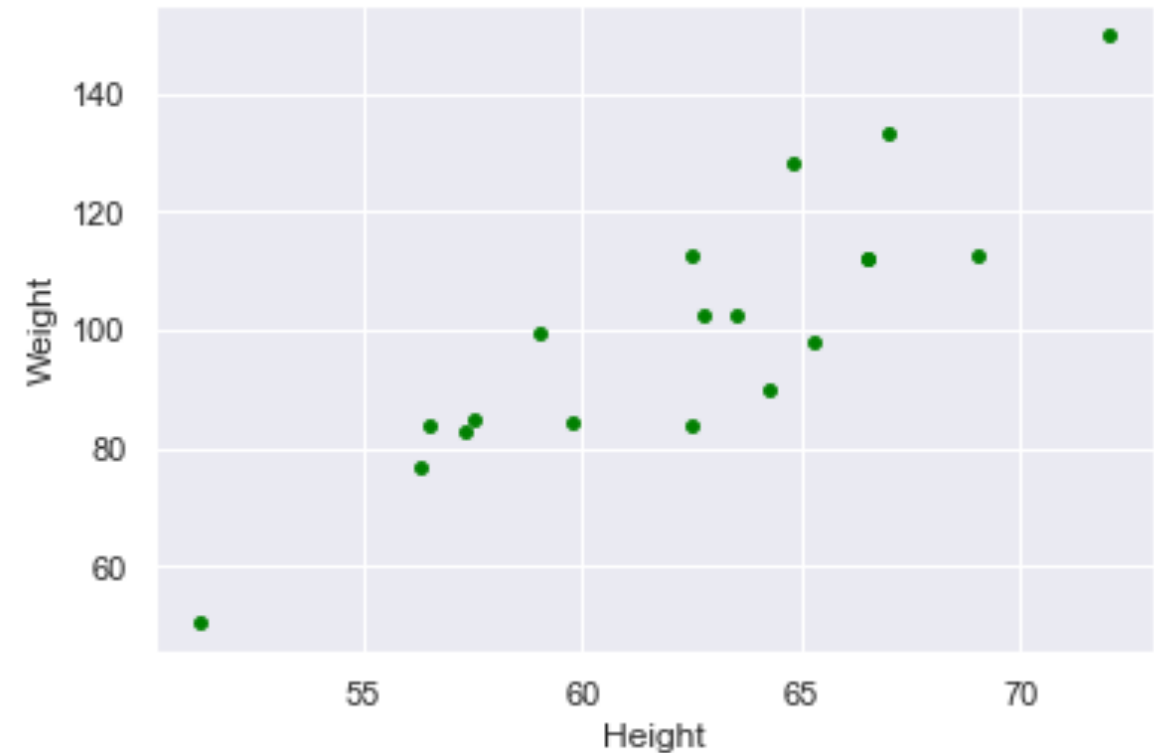
import pandas as pd
import numpy as np
import seaborn as sns

import matplotlib.pyplot as plt

#read SAS Datasets
dm = pd.read_sas(filepath_or_buffer = 'demog.sas7bdat',
                  format = 'sas7bdat',
                  encoding = "utf-8")

###Scatter plot
dm.plot.scatter(x = "Height",
                y = "Weight",
                s = 15,
                marker='o',
                linewidth=1,
                color='green')

# Function to show the Plot
plt.show()
```



Sequential Palette

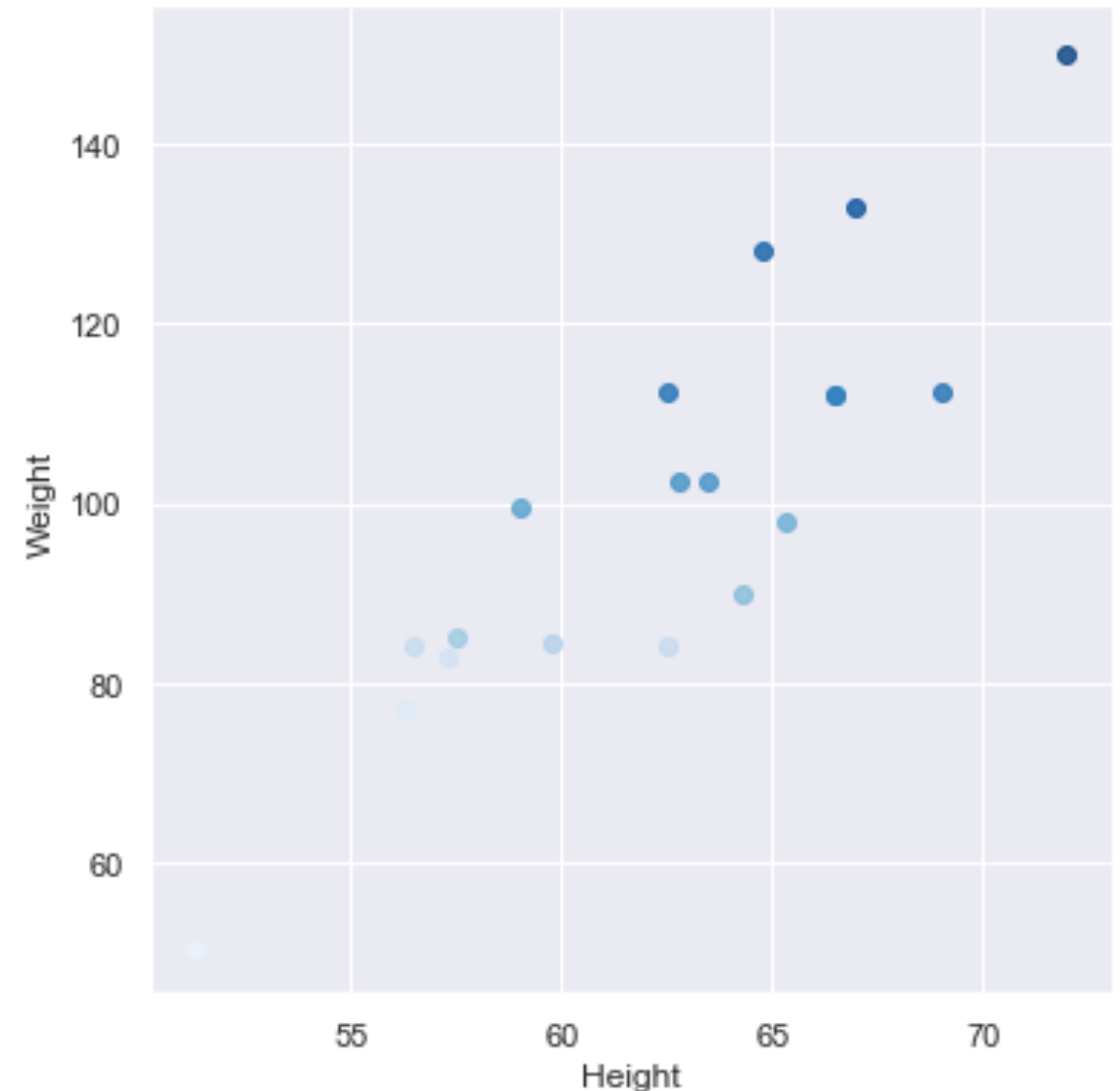
```
# Example 5 - Sequential Palette #
# PHUSE Europe Connect 2020    #

import pandas as pd
import numpy as np
import seaborn as sns

import matplotlib.pyplot as plt

#Seaborn
sns.lmplot(x = "Height",
           y = "Weight",
           data = dm,
           fit_reg=False,
           hue="Weight",
           height=5.5,
           legend=False, |
           palette="Blues")

# Function to show the Plot
plt.show()
```



Differential Palette

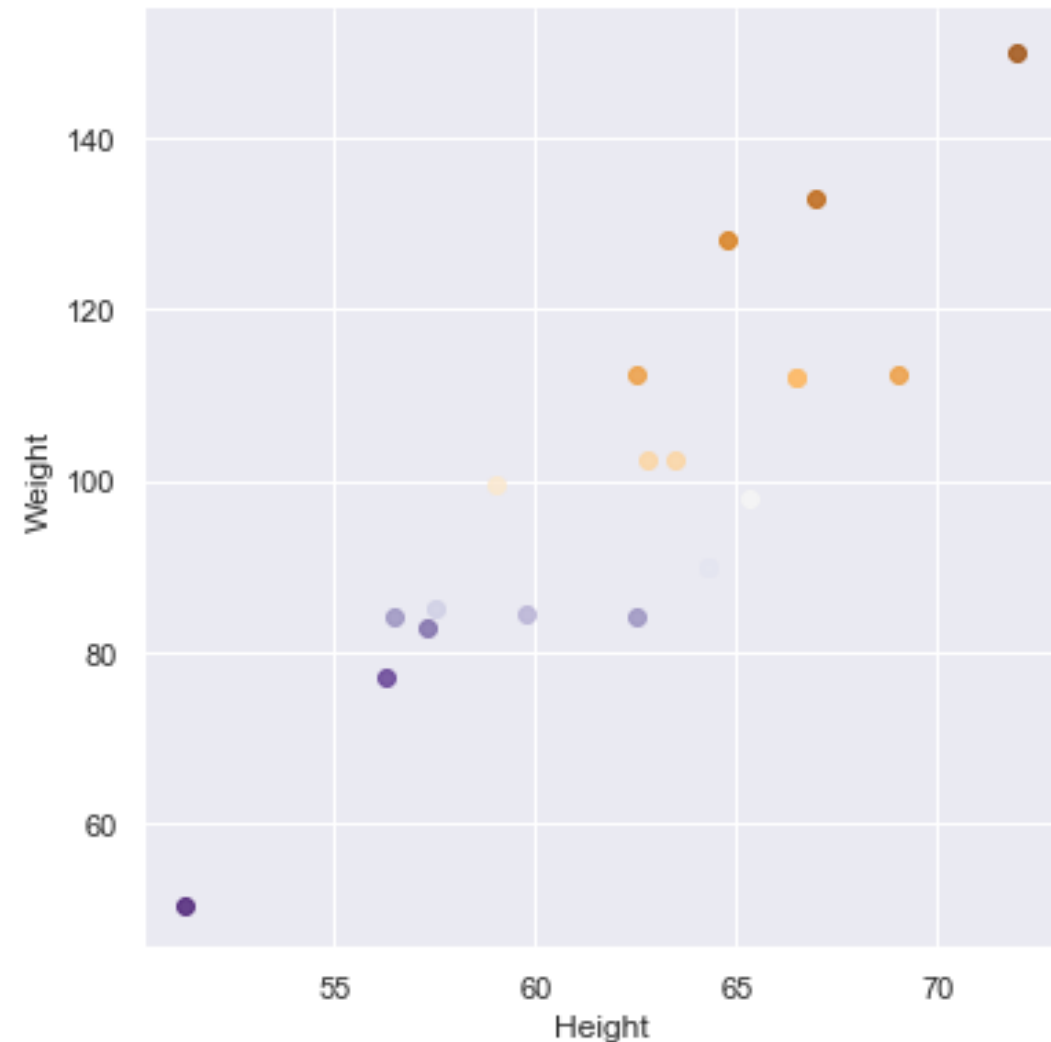
```
# Example 6 - Differential Palette #
# PHUSE Europe Connect 2020      #

import pandas as pd
import numpy as np
import seaborn as sns

import matplotlib.pyplot as plt

#Seaborn
sns.lmplot(x = "Height",
           y = "Weight", data=dm,
           fit_reg=False, hue="Weight",
           height=5.5, legend=False,
           palette="PuOr_r")

# Function to show the Plot
plt.show()
```



Box Plot

```
# Example 7 - Box Plot      #
# PHUSE Europe Connect 2020 #

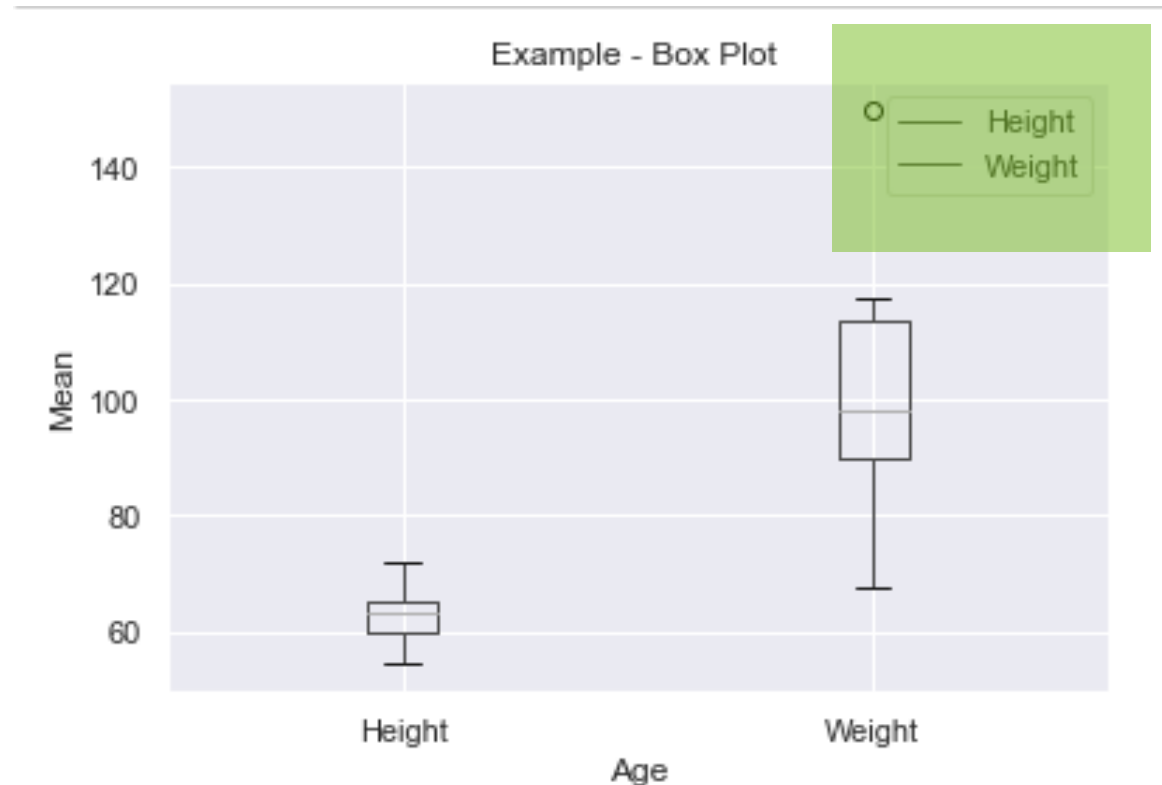
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline

dm = dm.groupby('Age').mean()
dm.boxplot(['Height', 'Weight'])

plt.title('Example - Box Plot ')
plt.xlabel('Age')
plt.ylabel('Mean')
plt.legend(['Height', 'Weight'], loc=1)

plt.show()
```



3 Dimensional

- **Mplot3d** - allows users to create simple 3D graphs with the same "look-and-feel" as matplotlib's 2D plots.
- Mplot3d support for simple 3d graphs including surface, wireframe, scatter, and bar charts.
- Another package **Mayavi** is a very powerful and featureful 3D graphing library.
- For advanced 3D scenes and excellent rendering capabilities, Mayavi is highly recommended.

Interactive

- Plotly - plotly is an interactive visualization library.
- Python library is used to design especially interactive graphs.
- It can plot various graphs and charts like histogram, barplot, boxplot, spreadplot and many more.

Summary

- 
 Selecting right tool for the right task
- 
 Clinical & Statistical Programming Teams are exploring tools and mechanisms to effectively interpret Data in a cost effective and timely manner.
- 
 SAS is preferred language across industry, due to its efficiency, support availability and perceived regulatory acceptance
- 
 Languages like R and **Python** are gaining popularity owing to its open source nature and increasing development of machine learning and artificial intelligence libraries.
- 
 Alternate Programming languages needs to be integrated to automate, and expedite Data Analysis and especially utilize the power of **Data Visualization**

References

- Jupyter Home <http://jupyter.org/>
- pandas <https://pandas.pydata.org/>
- matplotlib <https://matplotlib.org/>
- scikit-learn <https://scikit-learn.org/stable/>
- SASPy Documentation <https://sassoftware.github.io/sas>
- SAS Software Github Page <https://github.com/sassoftware>
- <https://support.sas.com/en/software/saspy.html>
- <https://github.com/sassoftware/saspy-examples/>
- Pandas Plots: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.plot.scatter.html>



Thank You

For any questions, please contact

Name: Siddharth Kumar

Email: Siddharth.kumar@navitaslifesciences.com