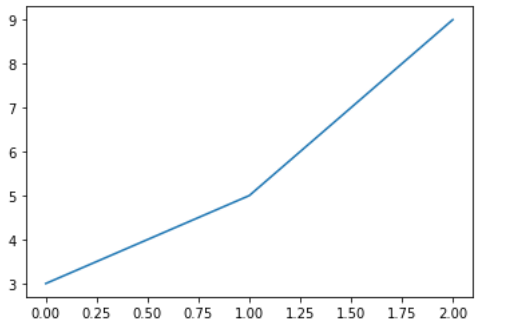
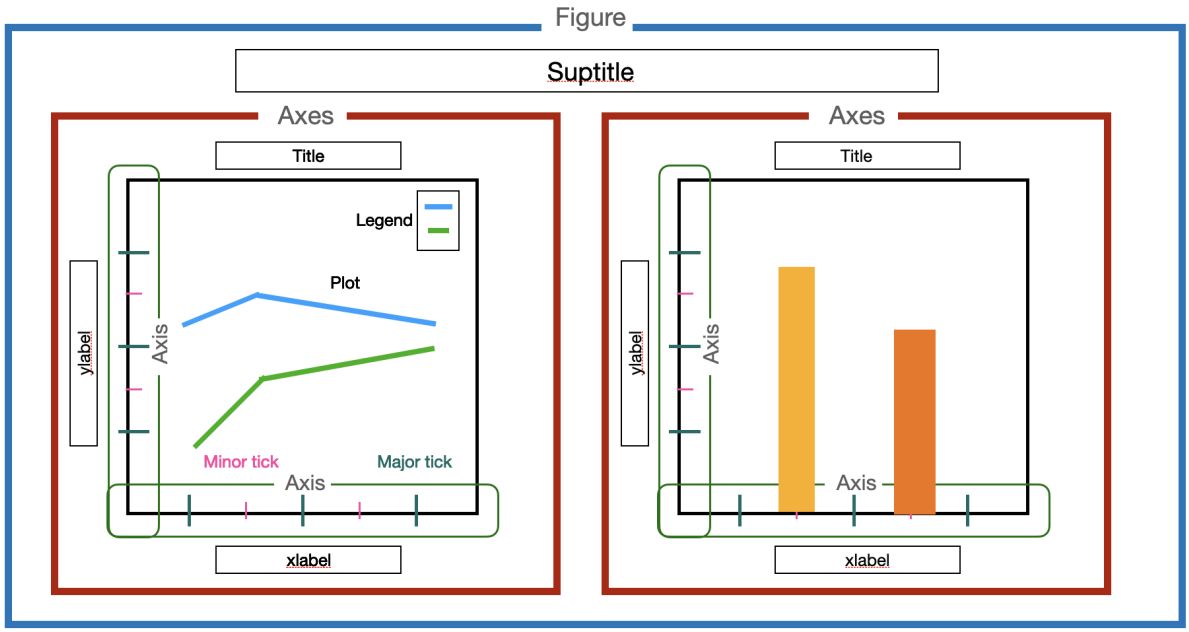
Link: https://colab.research.google.com/drive/1BOyZUnmNjbDYclLb4h5hxXNilCfWAWgy?usp=sharing

```
import pandas as pd
import numpy as np


# Where all do you think Data Voz. is helpful or needed?
# Exploratory - see some difficult patterns, EDA
# Explanatory - I want to create a stort out of it and I want to stake holders


# Art and Science Data Visualisation

# Science
# Understanding the anatomy of a plot
# How to choose which plot is right to answer my question?'

# Art
# Choosing thje right scale, axis, ticks and labels
# Identify and removing clutters
# Ways to highlight the information


# Intro to Matplotlib


import matplotlib.pyplot as plt
import seaborn as sns


# (0,3) (1,5) (2,9)


x_val = [0,1,2]
y_val = [3,5,9]


plt.plot(x_val, y_val)
```

```
[<matplotlib.lines.Line2D at 0x7f1bdf7c7e50>]
```






```
# how to choose the plot


# Data
# Rows - Records, Samples, Volume, Data-points
# Columns - Features, Attributes, Characteristics


# Columns - Continous, Categorical
# Categorical
# Oridinal - inherant ordering
# Nominal - no ordering


# Thumb for deciding the right plot?

# Question
# 1. How many variables/features are involved in answering my question?
  # Univatiate
  # Bivariate
  # Multi-variate
# 2. What are data-types of different variables involved?
  # Numerical
  # Categorical


#·Univariate·Data·Visualisation
··#·N·-·histogram,·....
··#·C·-·pie-chart,·....
#·Bivariate·Data·Visualisation
··#·N,·N·-·scatter,·...
··#·N,·C·-
··#·C,·C·-
#·Multi-variate·Data·Visualisatiob·(3·variables)
··#·N,·N,·N·-·
··#·N,·N,·C·-·
··#·C,·C,·N·-
··#·C,·C,·C·-
```

```
!wget https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/021/299/original/final_vg1_-_final_vg_%281%29.csv?1670840
```

```
--2023-02-07 17:19:23--  https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/021/299/original/final_vg1_-_fina
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 108.157.172.183, 108.157.172.173, 108.157.172.
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|108.157.172.183|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2041483 (1.9M) [text/plain]
Saving to: 'final_vg.csv'

final_vg.csv        100%[===================>]   1.95M  --.-KB/s    in 0.1s
```

```
2023-02-07 17:19:24 (16.5 MB/s) - 'final_vg.csv' saved [2041483/2041483]
```

```
import pandas as pd
import numpy as np
data = pd.read_csv('final_vg.csv')
data.head()
```

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales
0	2061	1942	NES	1985.0	Shooter	Capcom	4.569217	3.033887
1	9137	¡Shin Chan Flipa en colores!	DS	2007.0	Platform	505 Games	2.076955	1.493442
2	14279	.hack: Sekai no Mukou ni + Versus	PS3	2012.0	Action	Namco Bandai Games	1.145709	1.762339
3	8359	.hack//G.U. Vol.1//Rebirth	PS2	2006.0	Role-Playing	Namco Bandai Games	2.031986	1.389856
4	7109	.hack//G.U. Vol.2//Reminisce	PS2	2006.0	Role-Playing	Namco Bandai Games	2.792725	2.592054

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16652 entries, 0 to 16651
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Rank          16652 non-null  int64  
 1   Name          16652 non-null  object 
 2   Platform      16652 non-null  object 
 3   Year          16381 non-null  float64
 4   Genre         16652 non-null  object 
 5   Publisher     16594 non-null  object 
 6   NA_Sales      16652 non-null  float64
 7   EU_Sales      16652 non-null  float64
 8   JP_Sales      16652 non-null  float64
 9   Other_Sales   16652 non-null  float64
```

✓ 0s completed at 22:49