Data Visualization using Python

Upendra Kumar Tiwari
(Asstt. Professor,ABESIT)

Agenda:

- Why data Visualisation
- What is data visualization
- Introduction to different libraries used for Data Visualisation.
- Some Handson on Data Visualisation using Seaborn and Matplotlib

Why Data Visualization?

Human brain can process information easily when it is in pictorial or graphical form

Why Data Visualization?

Data visualization allows us to quickly interpret the data and adjust different variables to see their effect

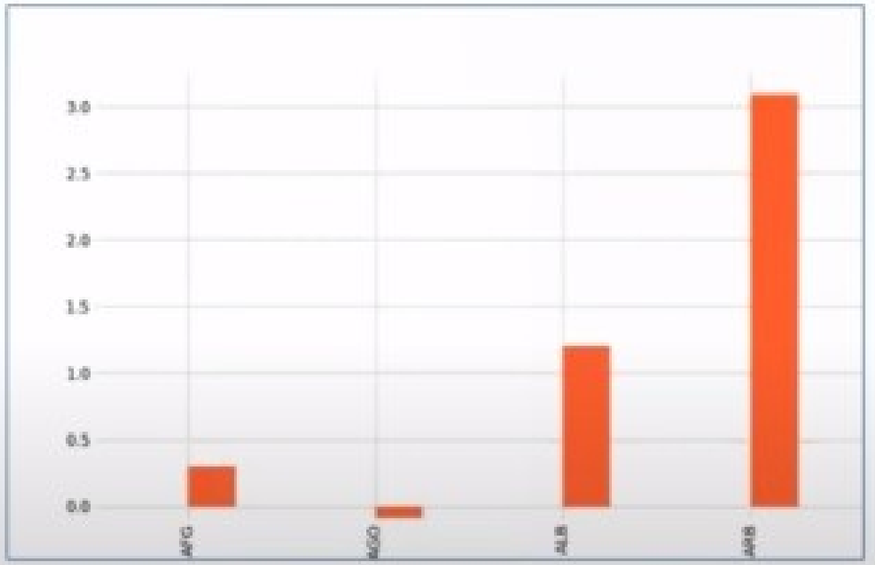
What Is Data Visualization?

Data visualization is the presentation of data in a pictorial or graphical format.

What is Matplolib?

Matplotlib is a Python package used for 2D graphics

Country Name	Country Code	2010	2011	2012	2013	2014
Afghanistan	AFG	20.6	20.9	19.7	21.1	20.8
Angola	AGO	10.8	10.7	10.7	10.6	10.5
Albania	ALB	25.799999	27	28.3	28.7	29.2
Arab World	ARB	25.022214	28.11752	29.11321	29.33531	29.70457
United Arab Emirates	ARE	9.8000002	9.8	9.8	9.9	10
Argentina	ARG	19.5	18.8	18.4	19.7	21.3
Armenia	ARM	38.299999	38.7	35	32.5	35.1
Australia	AUS	11.4	11.4	11.7	12.2	13.1
Austria	AUT	8.8000002	8.2	8.7	9.1	9.2
Azerbaijan	AZE	14.6	14.5	14.3	13.4	13.6
Burundi	BDI	10.8	10.8	10.8	10.8	10.7
Belgium	BEL	22.5	18.6	19.7	23.1	23.6
Benin	BEN	2	2	2	1.8	1.7
Burkina Faso	BFA	5.1999998	5.3	5.2	5.2	5
Bangladesh	BGD	8.1999998	8.2	8.2	8.9	9.1
Bulgaria	BGR	22.9	25.2	28.2	29.7	25.9
Bahrain	BHR	10.2	11.4	10.5	10.6	10.9
Bahamas, The	BHS	36	27.2	30.4	30.8	30.1
Bosnia and Herzegovina	BIH	57.200001	57.1	61.7	57.4	57.5
Belarus	BLR	13.2	12.5	11.8	12	12
Belize	BLZ	20.9	24.3	26	22.4	22



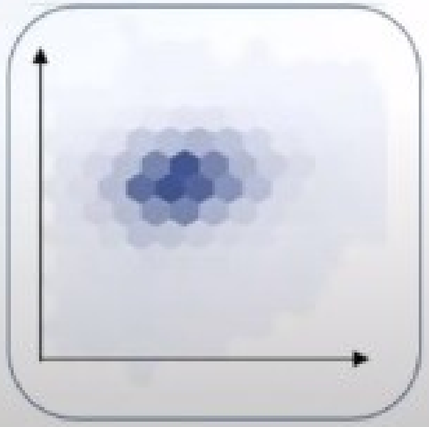
Types of Plots

Bar graph

Histograms

Scatter Plot

Pie Plot



Hexagonal Bin Plot

Area Plot

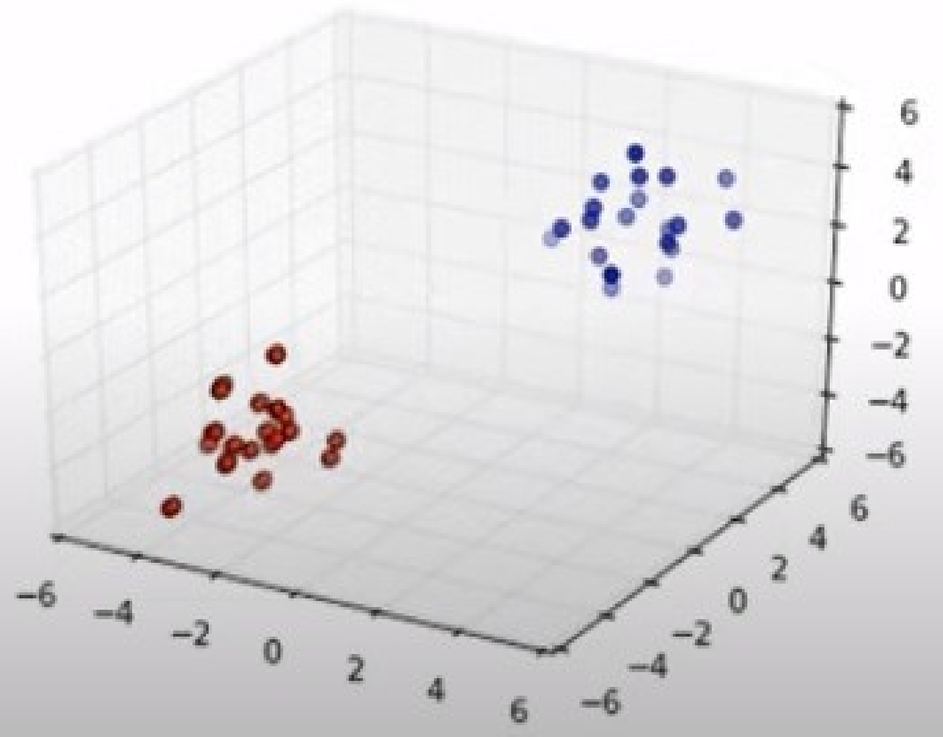
Introduction to *Seaborn*

Seaborn library:

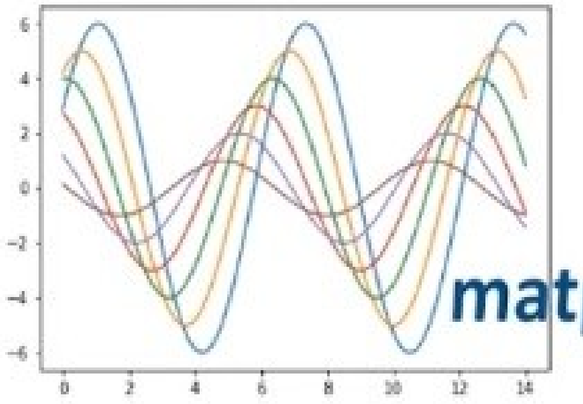
- Used for data visualization & based on Matplotlib
- Seaborn allows the creation of statistical graphics

Functionalities:

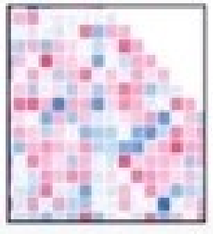
- Allows comparison between multiple variables
- Supports multi-plot grids
- Available univariate and bivariate visualizations
- Availability of different colour palettes
- Estimates and plots linear regression automatically

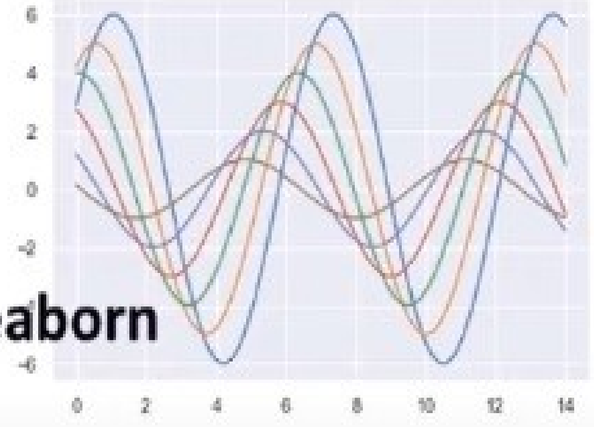


Seaborn fixes two shortcomings of Matplotlib:



- Matplotlib settings are difficult to figure out, Seaborn comes with numerous customized themes and high-level interfaces.
- Matplotlib doesn't serve well when it comes to dealing with DataFrames, while Seaborn functions actually work on DataFrames.



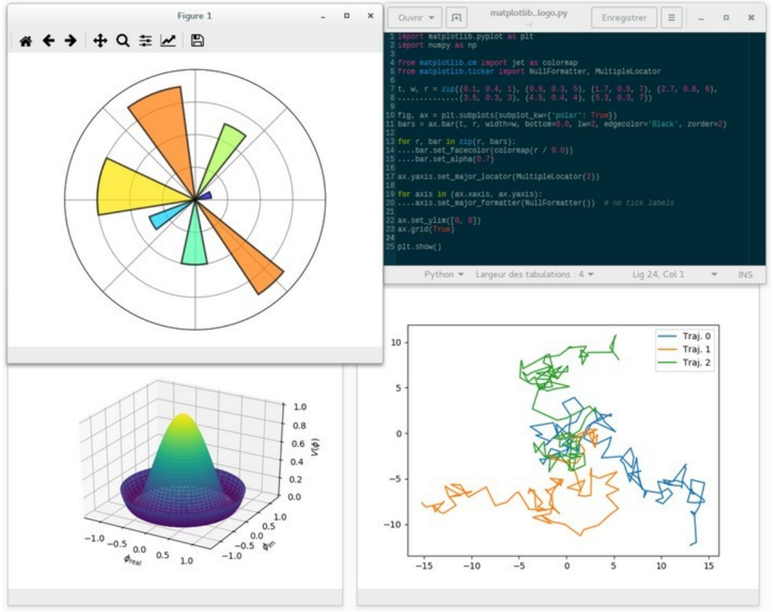


Seaborn Vs Matplotlib

LIBRARIES/TOOLS USED FOR DATA VISUALIZATION IN PYTHON

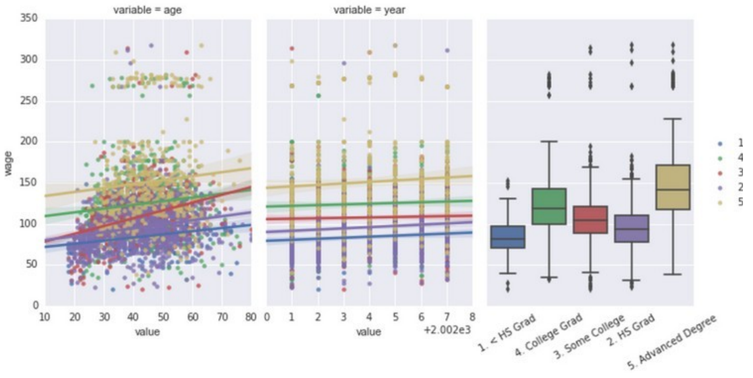
• Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. Used to draft plot such as Histogram, scatter plot, Line chart etc.



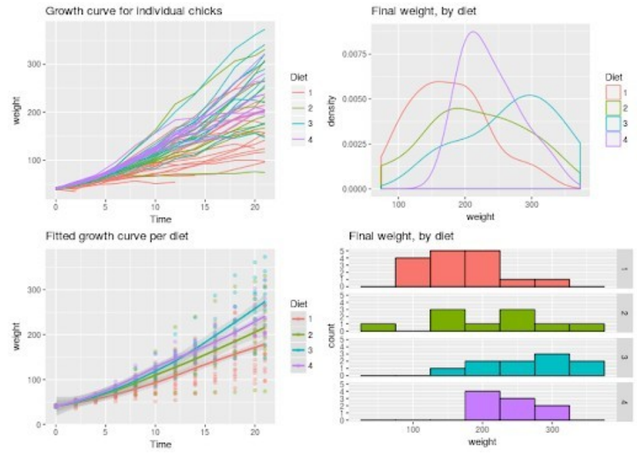
• Seaborn

Seaborn's default styles and color palettes are much more sophisticated than Matplotlib. Seaborn puts visualization at the core of understanding any data. Seaborn is a higher-level library- it's easier to generate certain kinds of plots, including heat maps, time series, and violin plots. Used to plot chart suchh as scatter plot, column chart etc.



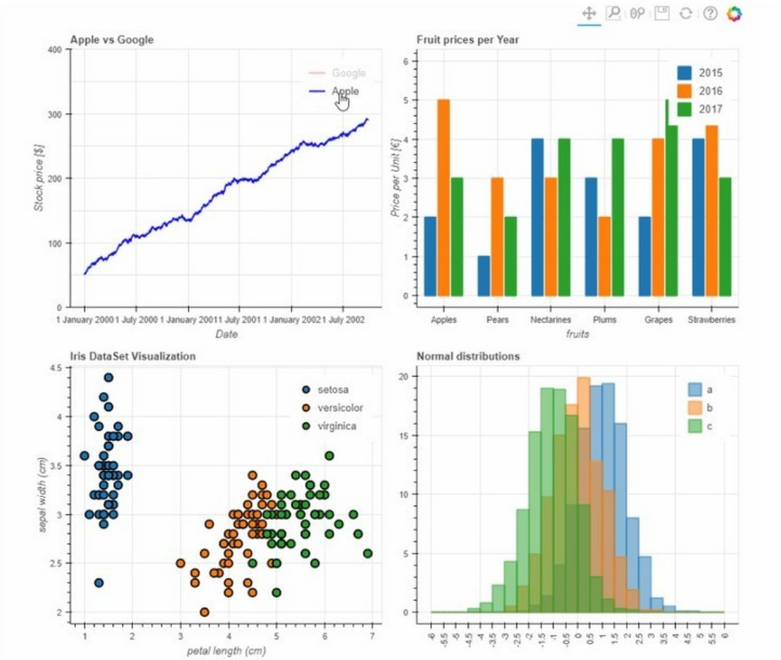
• ggplot Chart

Ggplot operates differently compared to Matplotlib: it lets users layer components to create a full plot. For example, the user can start with axes, and then add points, then a line, a trend line, etc. The Grammar of Graphics has been hailed as an "intuitive" method for plotting, however seasoned Matplotlib users might need time to adjust to this new mindset.Used to draft scatter plot, histogram etc.



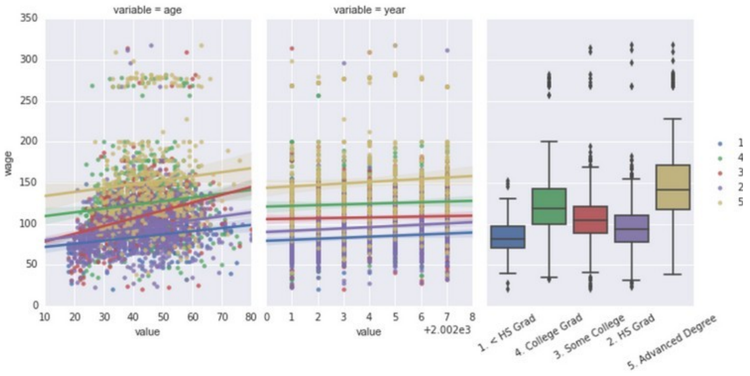
• Bokeh

Bokeh is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high- performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications. Used to draft line graph, column chart etc.



• Plotly

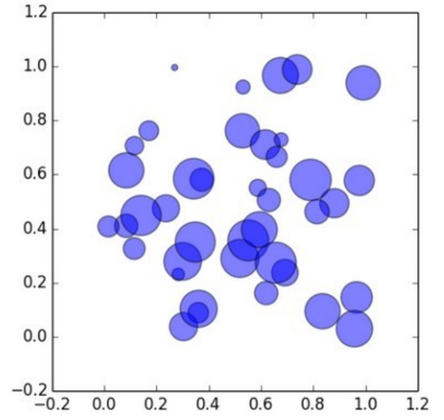
While Plotly is widely known as an online platform for data visualization, very few people know that it can be accessed from a Python notebook. Like Bokeh, Ploty's strength lies in making interactive plots, and it offers contour plots, which cannot be found in most libraries. Used to draft graph such as Box Plot Chart, Scatter plot etc.



• PygalChart

Pygal offers interactive plots that can be embedded in a web browser. The ability to output charts as SVGs is its prime differentiator. For work involving smaller datasets, SVGs will do just fine.

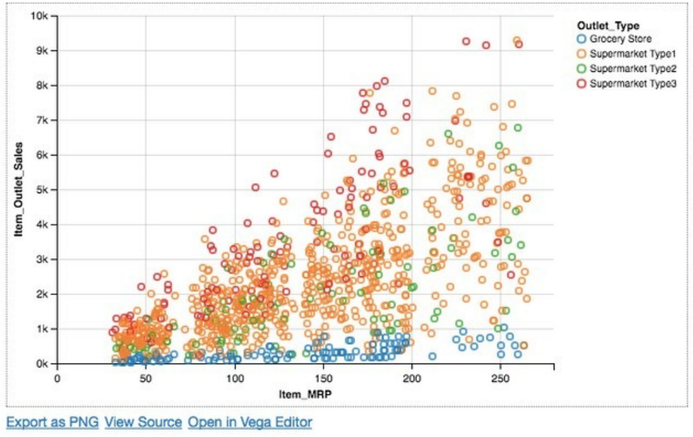
However, for charts with hundreds of thousands of data points, they become sluggish and have trouble rendering.It’s easy to create a nice-looking chart with just a few lines of code since each chart type is packaged into a method and the built-in styles are great. Used to draft scatter chart etc.



• Atlair

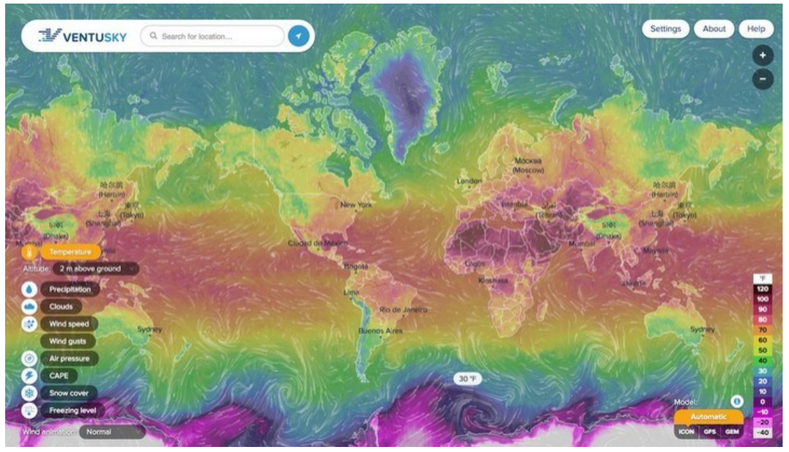
Altair is a declarative statistical visualization python library based on Vega-Lite. You only need to mention the links between data columns to the encoding channels, such as x-axis, y-axis, color, etc. and the rest of the plotting details are handled automatically. This makes Altair simple, friendly and consistent. It is easy to design effective and beautiful visualizations with a minimal amount of code using Altair. Used to plot scatter chart etc.

```
Chart(df).mark_point().encode(
    x='Item_MRP',
    y='Item_Outlet_Sales',
    color='Outlet_Type'
)
```



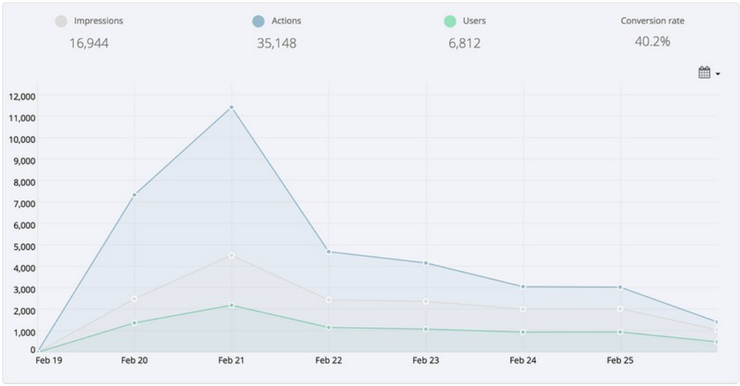
• Folium

Folium makes it easy to visualize data that's been manipulated in **Python** on an interactive Leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing Vincent/Vega visualizations as markers on the map.



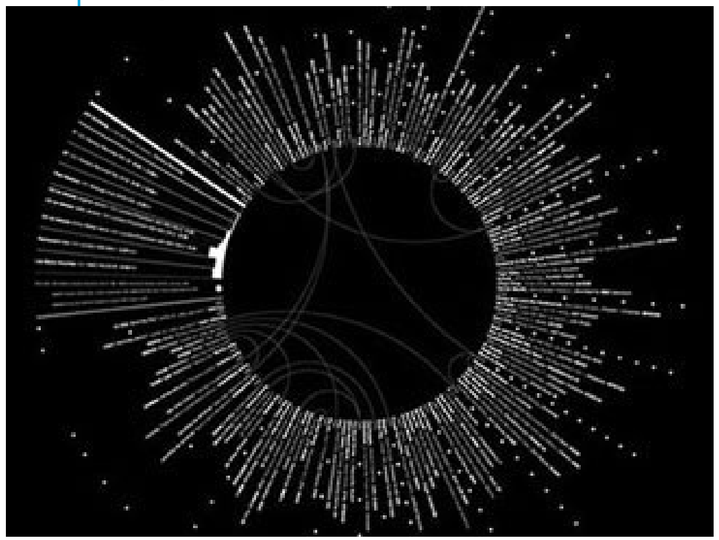
• Gleam

Gleam is inspired by R's Shiny package. It allows the user to turn any analysis into interactive web apps using only Python scripts. Gleam users don't need to know HTML, CSS, or JavaScript to do this. Gleam works with any Python data visualization library. Once users have created a plot, they can build fields on top of it to filter and sort data.

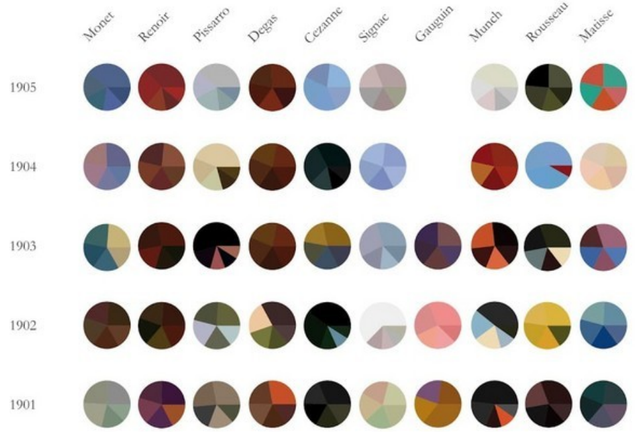


Application of Data Visualization

Cinema: explaining a movie plot through data visualization

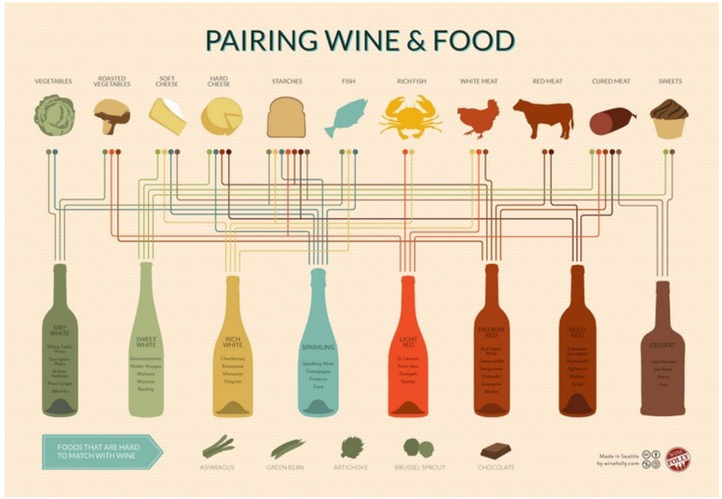


Art: analyzing the color palettes of great artworks



National geographic: cartography

Gastronomy in pictures



Literature and astronomy

Different Link

1. for Matplotlib https://matplotlib.org/
2. for Seaborn https://seaborn.pydata.org/
3. for ggplot chart https://pypi.org/project/ggplot/
4. for Bokeh https://bokeh.org/
5. for plotly https://plotly.com/
6. for pygal http://www.pygal.org/
7. for atlair https://altair-viz.github.io/
8. for geoplot https://residentmario.github.io/geoplot/user_guide/Working_with_Geospatial_Data.html

“Thank You”