Data Analytics and Visualization

with Applications in Financial Domain

January 14, 2023

HIT, Haldia, India

Sohom Ghosh

About Me

I am a Data Scientist at the Artificial Intelligence, Centre of Excellence of Fidelity Investments. My research interests include Applications of Natural Language Processing (NLP) in FinTech.

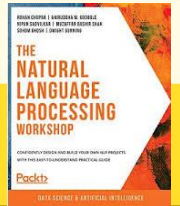
Over the last few years, I have been working on improving digital experience & financial well-being of millions of users across different industries like Financial Services, Internet and so on. In addition to being co-author of the books NLP Fundamentals and The NLP Workshop, I have several publications in proceedings of international conferences and refereed journals.

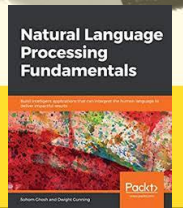
I hold a Master's Degree in Software Systems (with specialization in Data Analytics) from BITS Pilani, India and a Bachelor's Degree in Computer Science & Engineering.

Outside work, I like to play harmonica & cajon. Being an adventure lover and a fitness buff, I believe that "Health is Wealth".

SOHOM GHOSH

sohomghosh.github.io

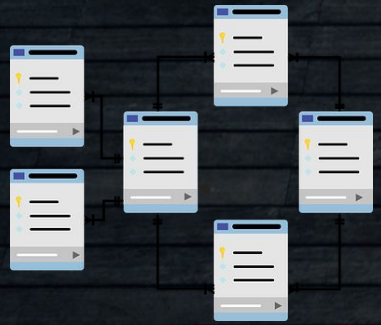




SOHOM
GHOSH

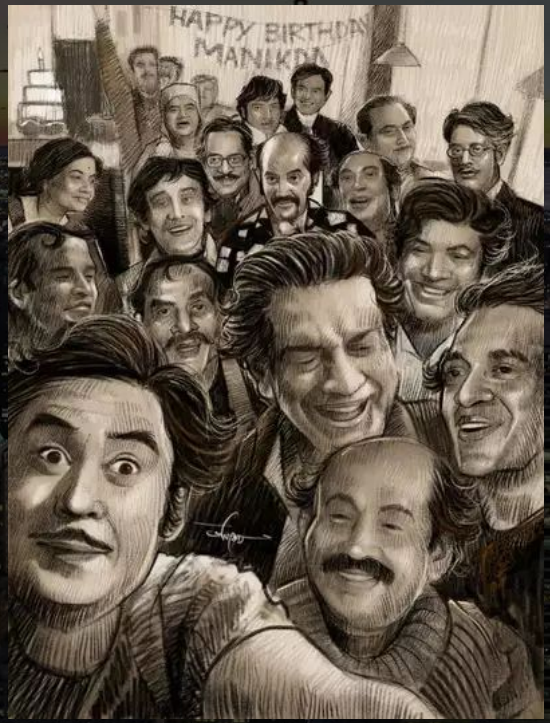
Data and its types

Data→Information→Insights



Structured (e.g. csv, excel), Semi-structured (e.g HTML, XML), Unstructured

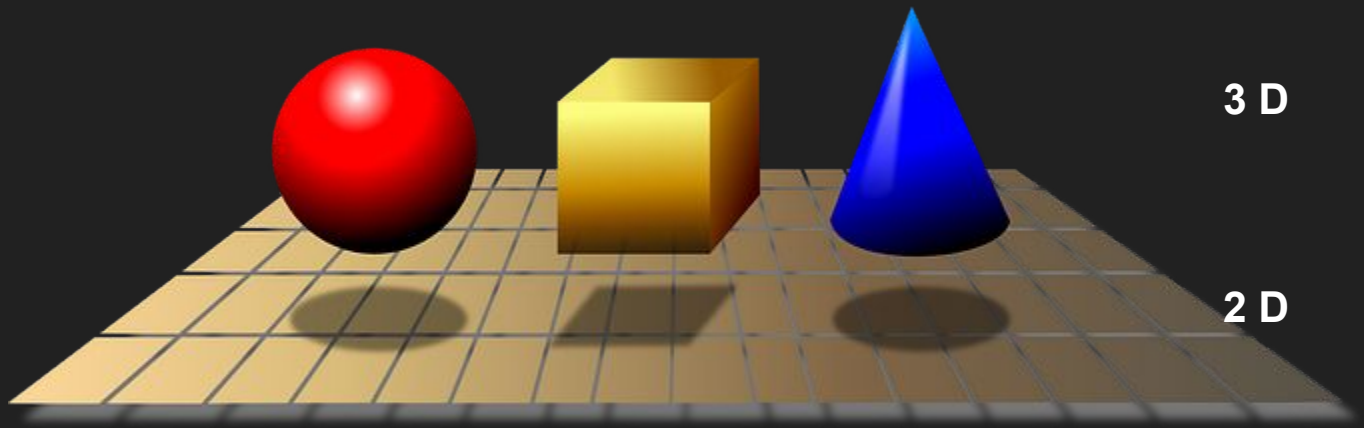
Why should we visualize data?



"A picture is worth a thousand words"

Source: Tribute to Satyajit on his 99th Birth Anniversary. An illustration by Aniket Mitra.

What we can/(not) visualize?



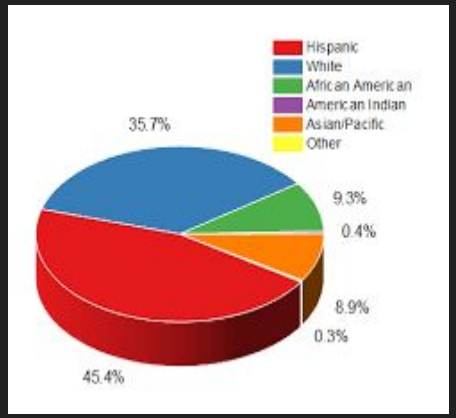
Beyond 3D ?

3 D

2 D

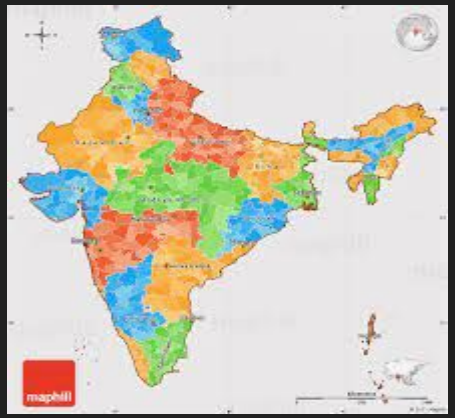
1 D

Common mistakes in Data visualization

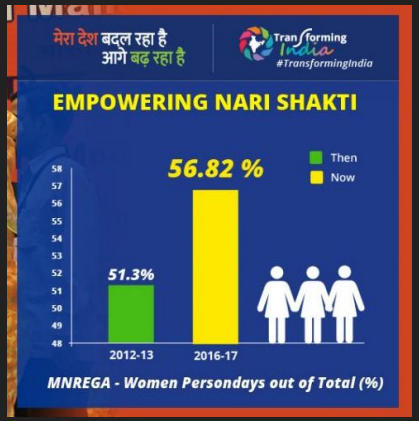


1

2



3



4

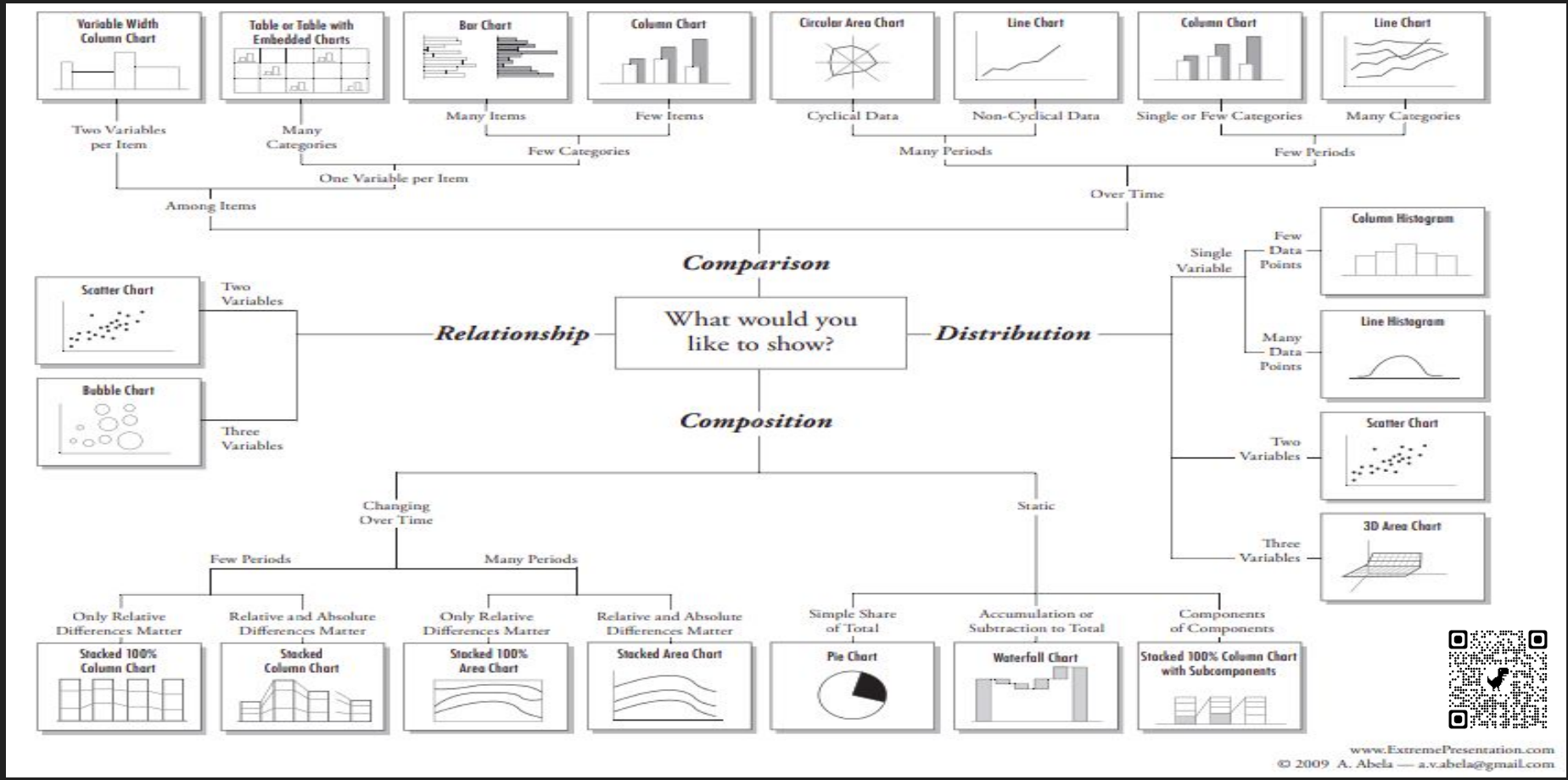
Other points to remember:

- Don’t forget to label the axes
- Avoid putting too much information in one go
- For publishing in academic journals try to use shape(s) instead of colour(s).

Sources:https://www.originlab.com/doc/Origin-Help/3D-Color-PieChart, http://www.maphill.com/india/simple-maps/political-map/cropped-outside/, https://www.kdnuggets.com/2017/10/5-common-mistakes-bad-data-visualization.html

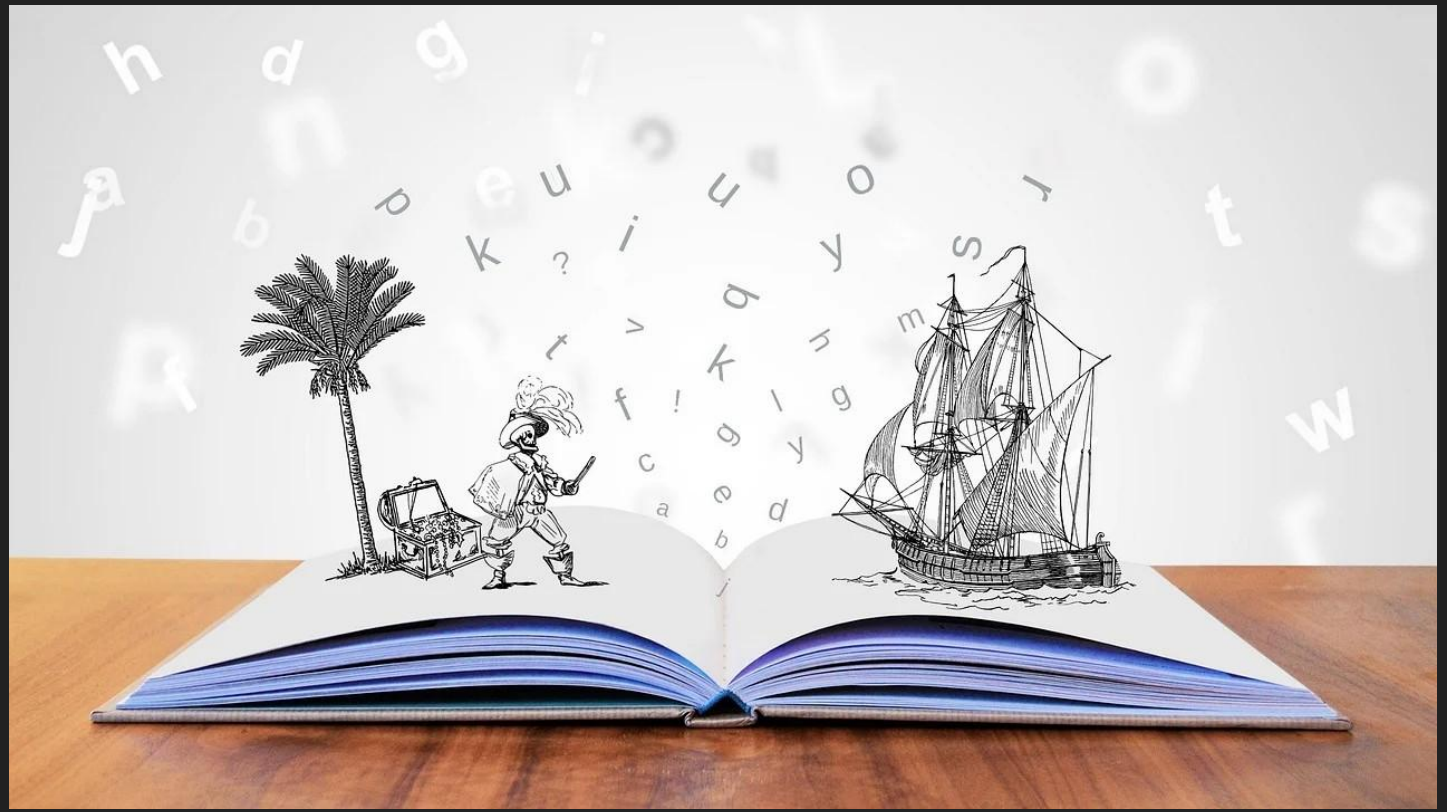
Different kinds of visualization

How to decide which visualization would be appropriate?



Source: https://extremepresentation.com/wp-content/uploads/choosing-a-good-chart-09-1.pdf

Remember: “You use data visualization to tell stories”



Case study 1: Visualizing Loan Application Data

Link: https://colab.research.google.com/ Dataset: https://www.kaggle.com/burak3ergun/loan-data-set/version/1

Case study 2: Visualizing Financial News

- Word Cloud
- Named Entities
- Dependency trees
- Scatter Plots using TSNE of Embeddings

Link: https://colab.research.google.com/

Dataset: https://www.kaggle.com/ankurzing/sentiment-analysis-for-financial-news

Bonus

Static versus Dynamic Data visualization

Dynamic Data visualization

- Interactive Dashboards
- Live Updates

Commercial Tools: Tableau, Qlikview

Demonstration: Gradio (https://gradio.app/demos/)

Time Series with Gradio Live Demonstration:

- Outbreak Forecast

https://colab.research.google.com/github/gradio-app/gradio/blob/main/demo/outbreak_forecast/run.ipynb

Thank you

Any questions?

All the training materials are available at:

https://github.com/sohomghosh/Data_Visualization_FDP

Reach me at:

Email: sohom1ghosh@gmail.com

Website: https://sohomghosh.github.io/

LinkedIn: https://www.linkedin.com/in/sohomghosh