

M7 – Data Visualization and Analytics



‘A picture is worth a thousand words’

What is Data Visualization

- Data visualization is the graphical representation of information and data. By using [visual elements like charts, graphs, and maps](#), data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.
- In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

Advantages of data visualization

- Quickly see trends and outliers
- Storytelling with purpose
- Internalize quickly
- Easily sharing information
- Interactively explore opportunities.
- Visualize patterns and relationships.
- Non-technical representation of data

Disadvantages of data visualization

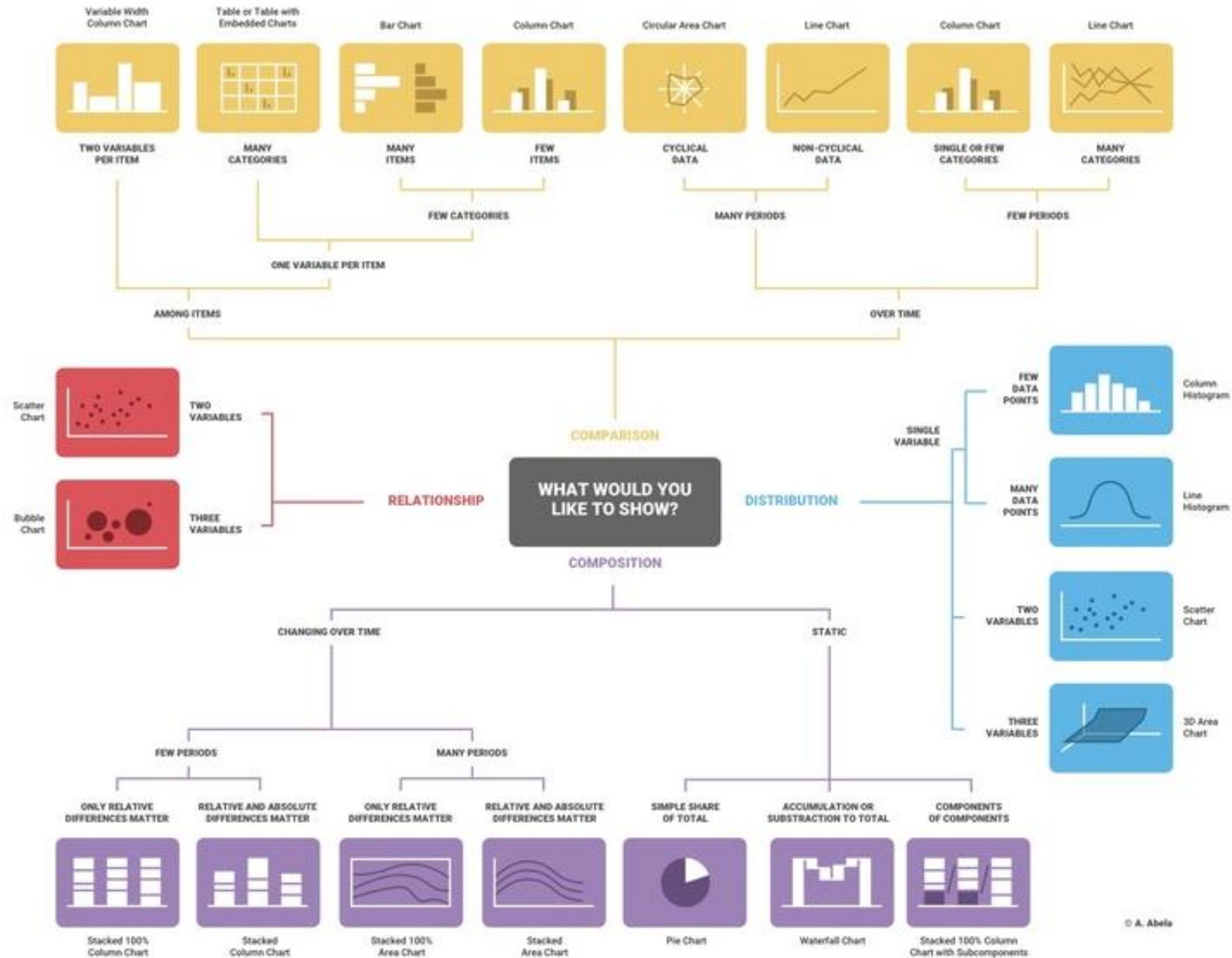
- Biased or inaccurate information.
- Correlation doesn't always mean causation.
- Core messages can get lost in translation.

General Types of Visualizations

- **Chart:** Information presented in a tabular, graphical form with data displayed along two axes. Can be in the form of a graph, diagram, or map. [Learn more.](#)
- **Table:** A set of figures displayed in rows and columns. [Learn more.](#)
- **Graph:** A diagram of points, lines, segments, curves, or areas that represents certain variables in comparison to each other, usually along two axes at a right angle.
- **Geospatial:** A visualization that shows data in map form using different shapes and colors to show the relationship between pieces of data and specific locations. [Learn more.](#)
- **Infographic:** A combination of visuals and words that represent data. Usually uses charts or diagrams.
- **Dashboards:** A collection of visualizations and data displayed in one place to help with analyzing and presenting data. [Learn more.](#)

CHART SUGGESTIONS - A THOUGHT-STARTER

<https://infogram.com/page/choose-the-right-chart-data-visualization>



Examples of Bad Visualization

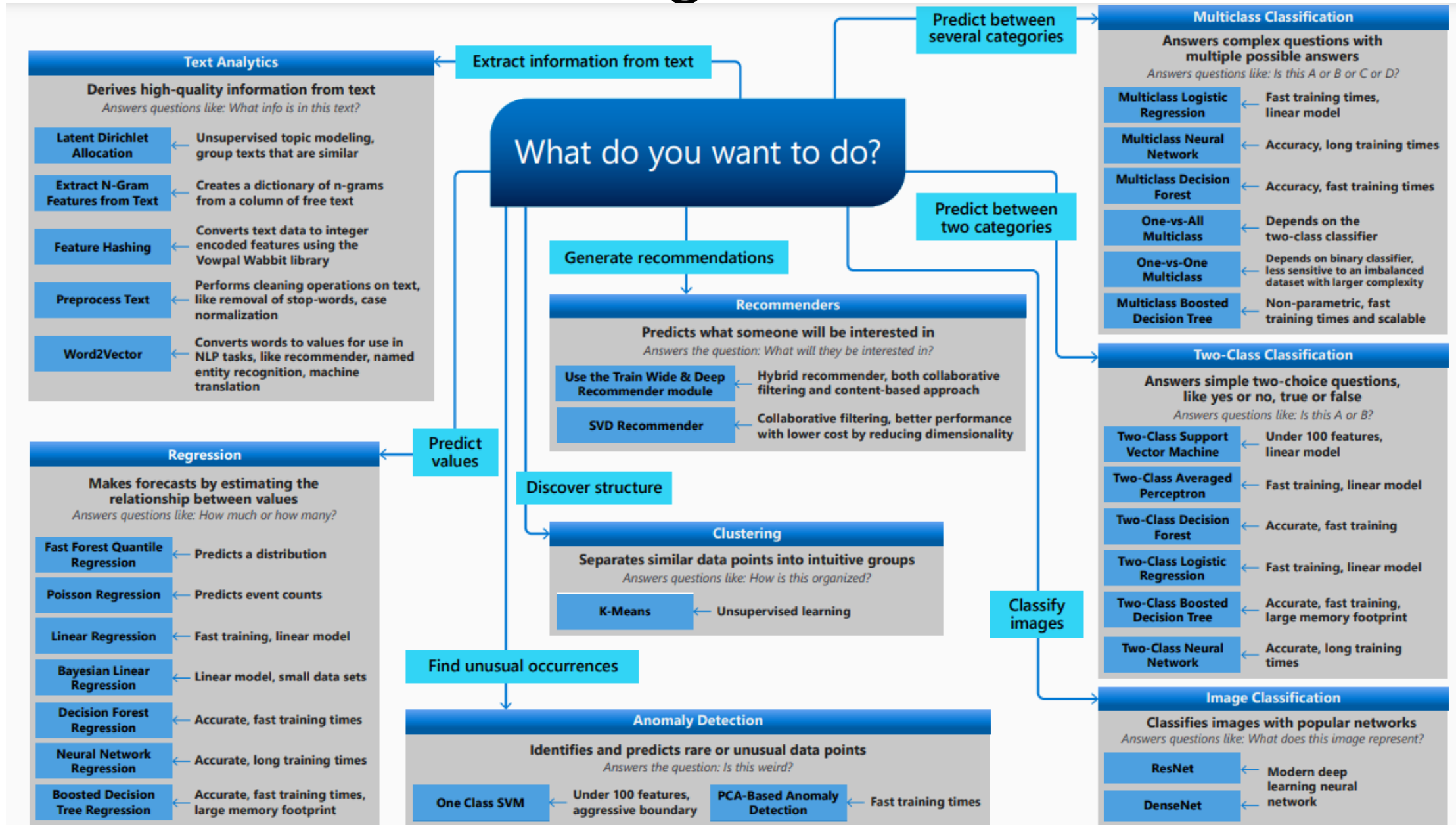
- <https://junkcharts.typepad.com/>

Key Characteristics of Data Visualization

(from New York Times)

1. Clarity of context and purpose
2. Respect for the reader
3. Editorial rigor and integration
4. Clarity of questions
5. Data research and preparation
6. Visual restraint
7. Layout and placement
8. Diversity of techniques
9. Technical Execution
10. Annotation

ML Algorithms



ML Algorithms cheatsheet

- <https://canvas.northwestern.edu/courses/199515/files/folder/unfiled?preview=16813217>
- Please refer to the file uploaded under module 7.

Week 7 Assignment

Use Case Approach: Ease of Data Analysis and Modeling

Evaluate your list of 10 use cases as they relate to an Overall Ease of Data Analysis and Modeling (1 – 5, 5 being easiest and 1 being most complex)

Use Case	What business question are you trying to answer with data visualization and modeling?	What kind of data visualization would provide useful insights?	What kind of machine learning, forecasting or analytics is suggested?	What algorithmic approach do you expect to lead with to solve this?	Overall Ease of Data Analysis and Modeling Score
Childhood Obesity Analysis	Is BMI increasing / decreasing? Is there a change due to these factors - lifestyle (Food type, Kind of Activities)	Use simple bar charts differentiated by the lifestyle	We could model as multi-class classification (Obese, Overweight, Healthy, Underweight) adjusting for the multi factors. Ex: Decision Tree, Logistic Regression (One Vs. Rest), Starting point Vs End point model: Any sample could end up in any of the 16 possible cases.	Stochastic Gradient approach or Generalized linear models (if it is a logistic regression) If it is decision tree – minimizing cross entropy loss function using stochastic gradient	4.5