Introduction to Data Analytics

ITE 5201

Lecture5-Data Visualization

Instructor: Parisa Pouladzadeh
Email: parisa.pouladzadeh@humber.ca

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 and patterns in data.

Types of Data Visualization

Data Storytelling

- For presentations to organizational decision makers

Data Showcasing

- For presentations to analysts, scientist, mathematicians, and engineers

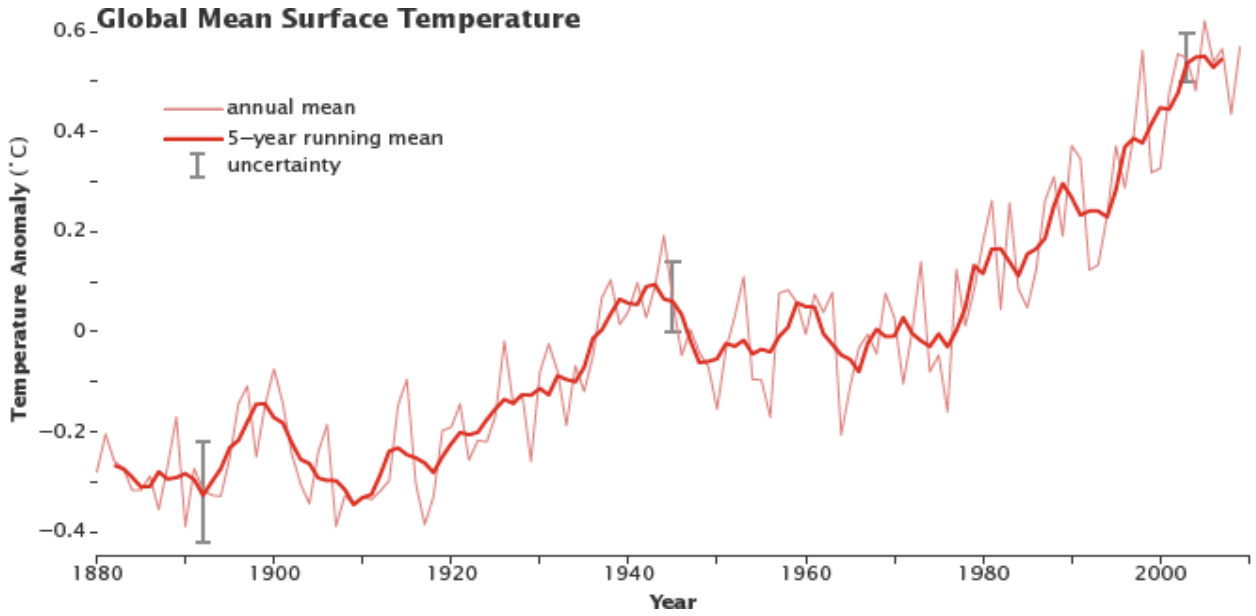
Data Art

- For presentations to activists or to the general public

Data Storytelling

- Make it easy for the audience to get the point. Your data visualization should be:
  - Clutter-free
  - Highly organized
- Audience:
  - Nonanalysts
  - Nontechnical business managers
- Product types:
  - Static images
  - Simple interactive dashboards

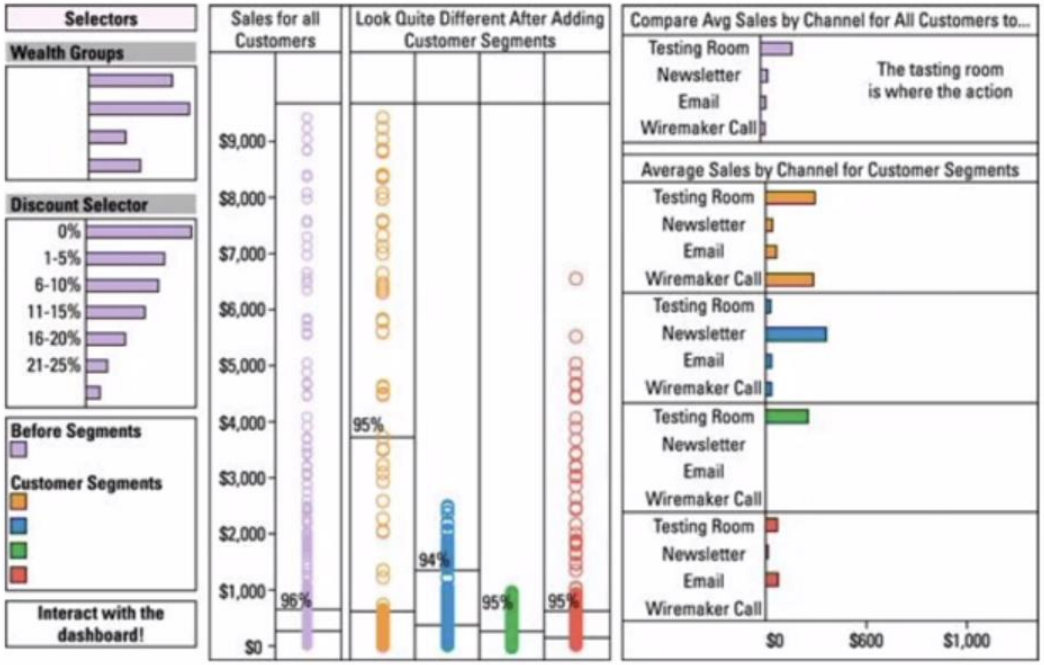
Example


Global Mean Surface Temperature
annual mean
5–year running mean
uncertainty
Temperature Anomaly (°C)
0.6
0.4
0.2
0
−0.2
−0.4
1880
1900
1920
1940
1960
1980
2000
Year

Data Showcasing

- Showcase lots of data so your audience members can think for themselves
- Your data visualization should be:
  - Highly contextual
  - Open ended
- Audience:
  - Analysts
  - Engineers, mathematicians, scientists
- Product types:
  - Static images
  - Interactive dashboards

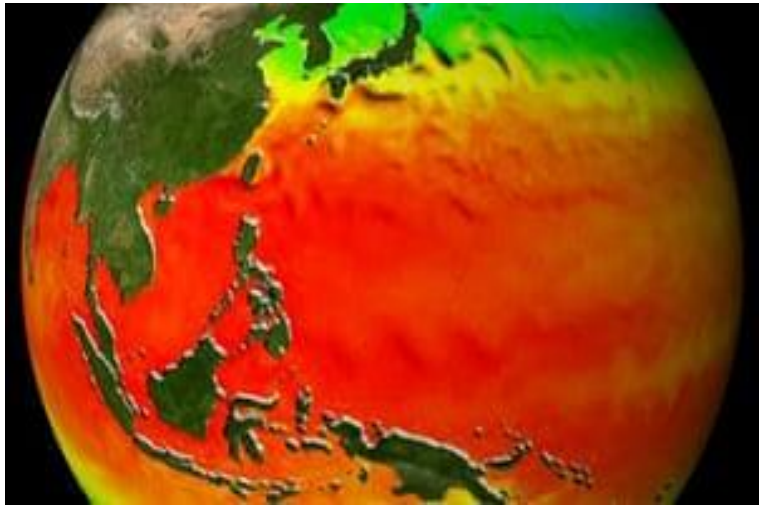
Example


Selectors
Wealth Groups
Discount Selector
0%
1-5%
6-10%
11-15%
16-20%
21-25%
Before Segments
Customer Segments
Interact with the dashboard!
Sales for all Customers
$9,000
$8,000
$7,000
$6,000
$5,000
$4,000
$3,000
$2,000
$1,000
$0
96%
Look Quite Different After Adding Customer Segments
95%
94%
95%
95%
Compare Avg Sales by Channel for All Customers to...
Testing Room
Newsletter
Email
Wiremaker Call
The tasting room is where the action
Average Sales by Channel for Customer Segments
Testing Room
Newsletter
Email
Wiremaker Call
Testing Room
Newsletter
Email
Wiremaker Call
Testing Room
Newsletter
Email
Wiremaker Call
Testing Room
Newsletter
Email
Wiremaker Call
$0
$600
$1,000

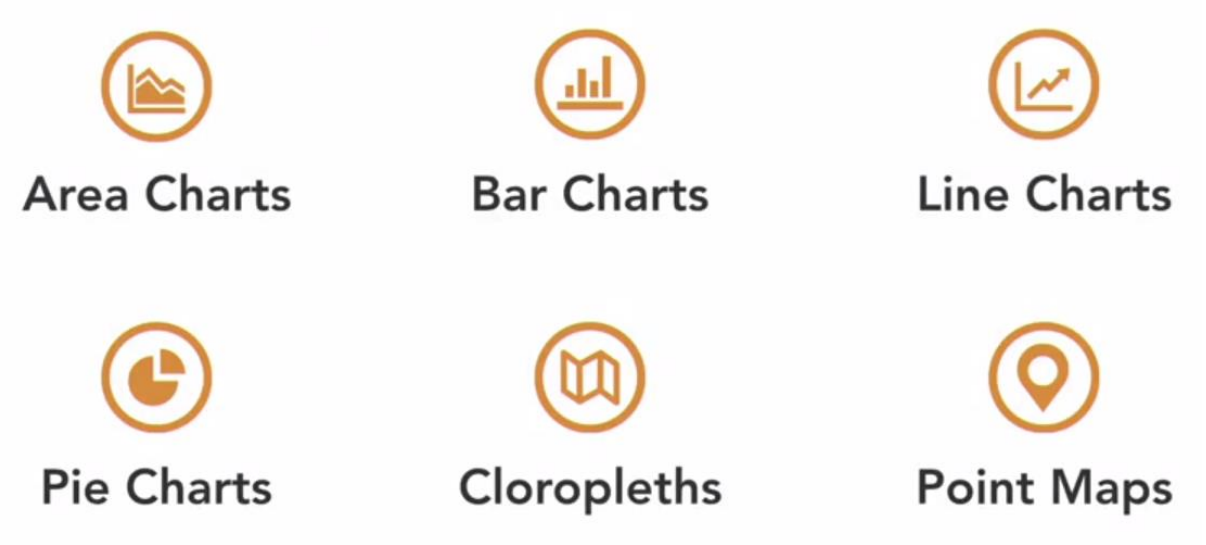
Data Art

- Use your data visualization to make a statement
- Your data visualization should be:
  - Attention getting
  - Creative, controversial
- Intended Audience:
  - Idealists, artists, Social activists
- Product types:
  - Static images

Example



Graphics of data storytelling



Choropleth map

- A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.
- First read the instructions and colour legend/key to understand what the shading means.

Point map


WHITEHORSE
IQALUIT
YELLOWKNIFE
ST. JOHN'S
JASPER
EDMONTON
WHISTLER
SASKATOON
BANFF
CAPE BRETON
VANCOUVER
CALGARY
CHARLOTTETOWN
VICTORIA
REGINA
FREDERICTON
WINNIPEG
QUEBEC CITY
HALIFAX
MONTREAL
OTTAWA
TORONTO
NIAGARA FALLS

Steps to choosing data graphics

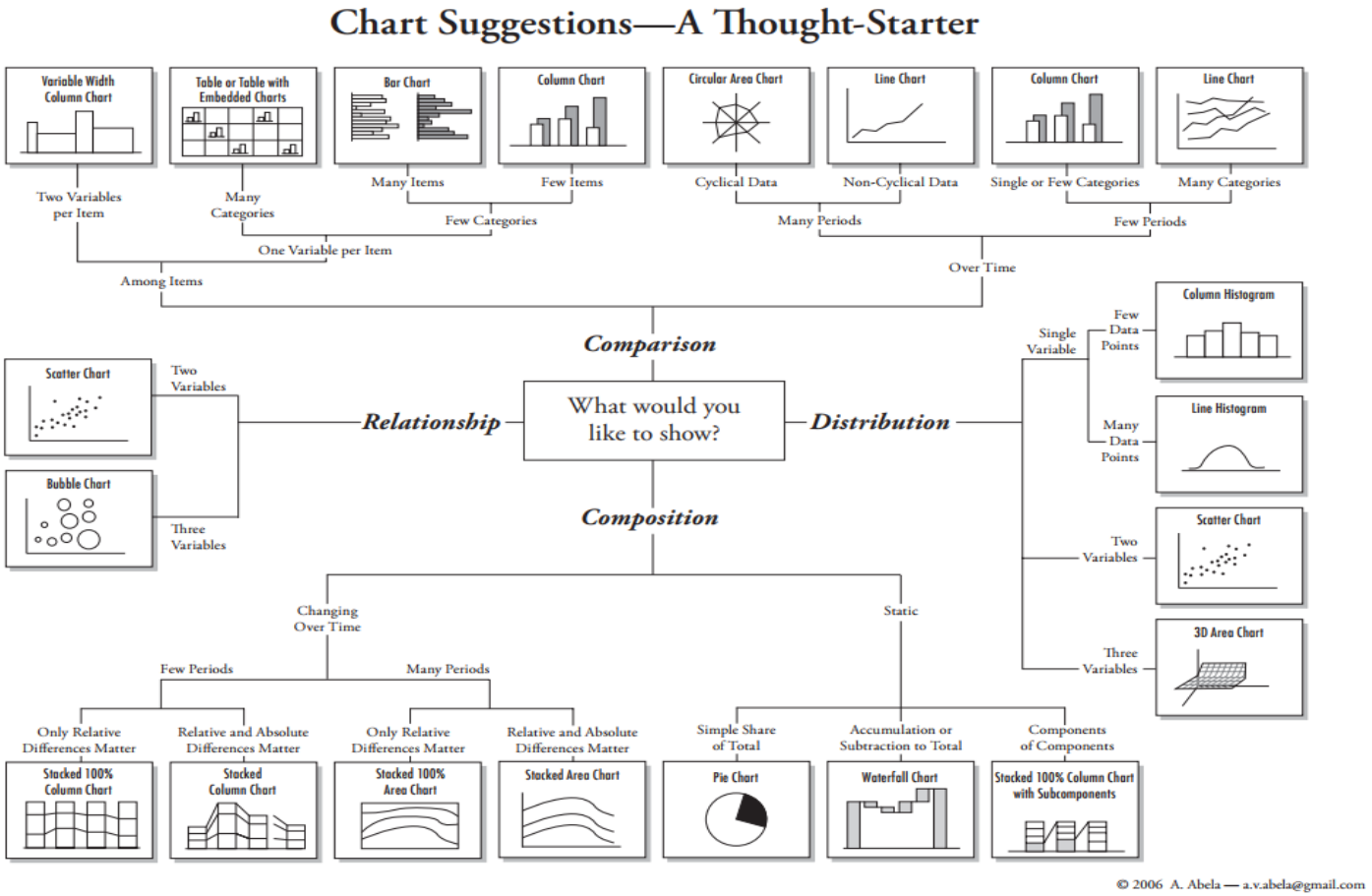
Make a list of question that your data visualization is meant to answer.

What is your data visualization types?(storytelling, showcasing or art?

What data graphic type are preferable for the type of data visualization?

Tests out different types of data graphics to see which one is more meaningful?

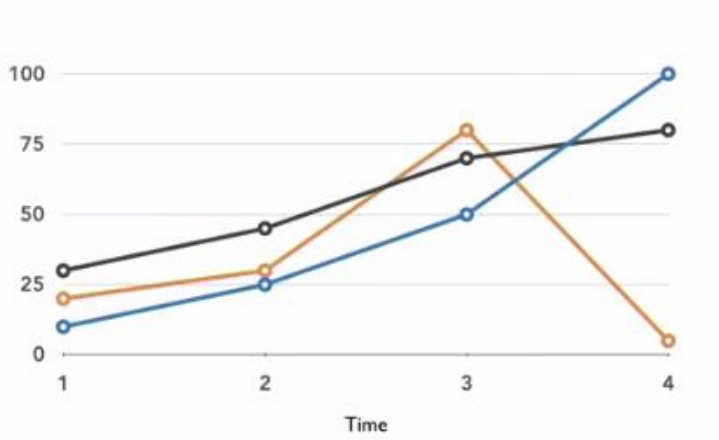
Creating context with color



Charts

Line Charts

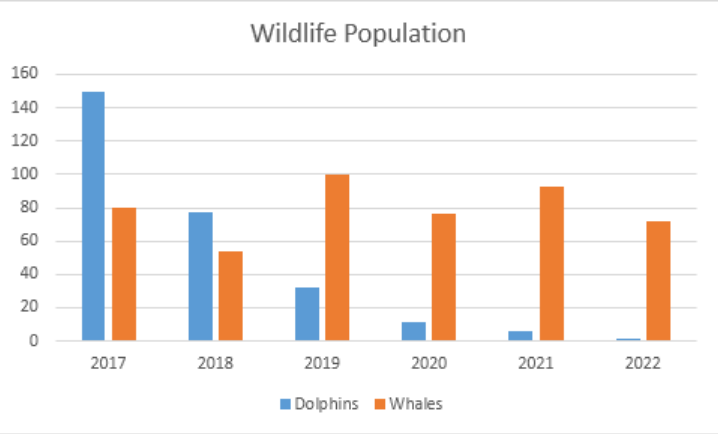
- Show the change in the value of an attribute with respect to an x-variable with is often Time.
- Can be use to visually compare the values of several attributes.



Charts

Bar Charts

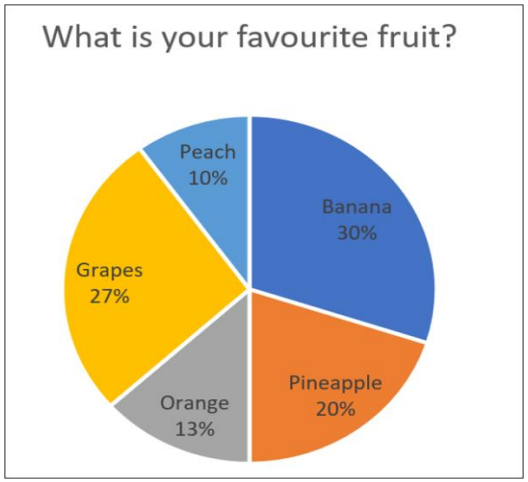
- Respect data attribute values within a particular data category by using bars of different heights.
- Bar Charts represent observation counts within categories.



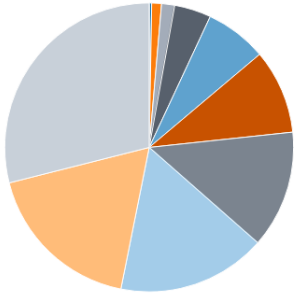
Charts

Pie Chart

- A whole and entire set of categorical data is represented by the complete circle and the proportions of observation are presented by slices.



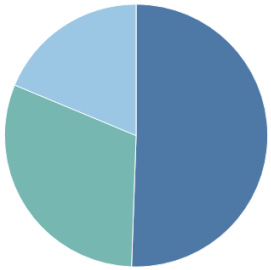
What about pie charts?



Commonly used to show parts of a whole

However…

- Hard to judge relative size of pie slices – **better at differentiating length**
- Take up a lot of space to **present little information**
- **Require labels and good color contrast** to even be usable (often difficult)



Best use is when one overwhelmingly larger value than the rest – no need to focus on actual values

Graphs are useful for?

Line graphs can also be used to compare changes over the same period of time for more than one group.

Pie charts are best to use when you are trying to compare parts of a whole. They do not show changes over time.

Bar graphs are used to compare things between different groups or to track changes over time.

Data visualization libraries in Python

Matplotlib

Seaborn

Time series

A set of observations, results, or other data obtained over a period of time, usually at regular intervals:

- Monthly sales figures
- quarterly inventory data
- and daily bank balances are all time series.

- A time series plot is a graph that you can use to evaluate patterns and behavior in data over time.

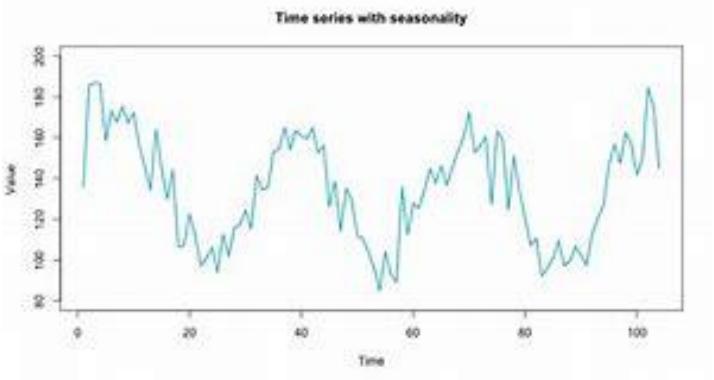
Times series components

Trends: This refers to the movement of a series to relatively higher or lower values over a long period of time.

- For example, when the Time Series Analysis shows a pattern that is upward, we call it an Uptrend,
- when the pattern is downward, we call it a Down trend
- if there was no trend at all, we call it a horizontal or stationary trend

Times series components

- Seasonality: This refers to a repeating pattern within a fixed time period.
  - Trend happens for a period of time and then disappears. However Seasonality keeps happening within a fixed time period.
  - For example, when it's Christmas, you discover more candies and chocolates are sold and this keeps happening



Times series components

Irregularity: This is also called noise.

Irregularity happens for a short duration and it’s non depleting.

- A very good example is the case of Ebola. During that period, there was a massive demand for hand sanitizers

Times series components

Cyclic: This is when a series is repeating upward and downward movement.

It usually does not have a fixed pattern. It could happen in 6months, then two years later, then 4 years, then 1 year later. These kinds of patterns are much harder to predict.