Data analysis + interactive visualization = Decision making.

Tashfeen Abbasi 22i-2041 DS-D

Lecture 2:-

- Data Visualization (helps users understand complex dataset)

-- Humans produced or will produce 147 zettabytes (ZB) of data in 2024.

-- Anscombe's Quartet → • Set of four datasets that have same descriptive statistics. (mean, Variance, correlation, linear regression etc)
• but appear very different when graphed.

Q:- Type of data & its purpose?

1 -- Scientific Visualization → e.g., visualize molecules, weather patterns.
(visual representation of scientific data)
• employs statistical graphics
• Type → 1D, 2D or 3D space/ scalar or vector fields
• Purpose → • convey scientific data accurately
• reveal structure in data
• encourage exploration of data

2 -- Information Visualization → e.g., tree maps.
(representation of abstract data often non-numerical (Texts))
Type → abstract data (numerical, categorical, text)
Purpose → • reveal basic structure
• encourage exploration

3 -- Statistical graph → e.g., scatter plot, box plot, histogram.
(represent quantitative data)
Type → Astract data (statistical, quantitative)
Purpose → Convey underlying structure. numerical
• happening in data & analyze at same time. (deeper understanding)

4 -- Visual analytics → -- includes elements of information visualization, statistical graphics, information dashboards.
(Combining algos or computer tools with visuals)
Type → Abstract data
Purpose → • Answer specific question
• Not constraint by single display

5 -- Information dashboards → • shows important data in on place. (extremely dense)
Type → Temporal or time series data
Purpose → • Convey large data quickly
• Convey outlier & trends

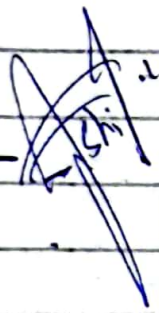
6 -- Infographics → • Combination (visual & text)
(Combines visual elements with textual)
tells story & explain data in a simple way.
Type → Abstract data.
-- Can utilize element of all above to convey information.
Purpose → • Eye catching
• May not be accurate
• not encourage exploration
• not use space efficiently.

Lie factor (how accurately visualization represent the data)

-- Statistical graphics are subset of information visualization.

Lecture 3:- Tufte → • Show data clearly
• Maximize data-ink ratio (use more ink for data, less for decoration)
• Erase non-data ink
• Erase redundant data ink (Avoid repeating data)

-- Graphical excellence → • gives viewer greatest number of ideas in short time with least ink in smaller space.
(more ideas, short time, least ink, smaller space)
• Edward Tufte's principle (Simplicity, Clarity)
• Example, (Napoleon's march to & from Moscow)



-- Graphical integrity → • Visual representation & numerical quantities
(show data variation, not design variation)
• Graphics must not quote out of context
• Clear, detailed, thorough labeling

-- Lie factor = (Size of effect in graphic) / (Size of effect in data)

Lie factor = >1, overstating (dramatic)
Lf = 1, accurate
Lf = <1, understating

Change in effect (data) = (Final value − Initial value) / Initial value × 100
(how much actual data has changed)

Change in graphics = (Final graphic element − Initial graphic element) / Initial graphic element × 1
(how much visual representation has changed)

Data ink = ink used to show data.

- Data ink ratio = data ink / total ink in graphics

- Chart junk → unnecessary elements in charts that do not enhance understanding of data.

- William Cleveland → (1985)
  - clear vision
  - clear understanding
  - scales
  - general strategy.