COMP 4462
Data Visualization Tutorial

CHEN Chang
PAN Ziqi

Friday 15 November, 2024

Outline

- **Exercise review** (in-class exercise 2, 3, and 4)

- **Tips for midterm 2** (coverage, question types, cheatsheet)

- **Quick review**
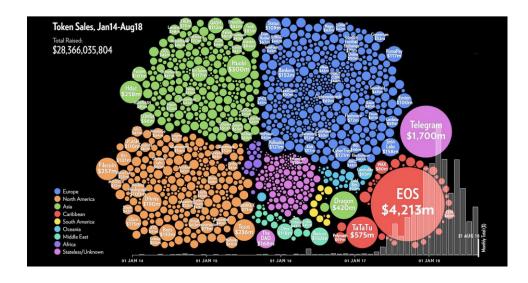
- **Q&A**

In-class exercise 2

Question 1

Figure 1. Cryptocurrency visualization shows the $6.4 billion poured into initial coin offerings (ICOs -- represented by each circle) since 2014.

How would you modify this visualization if you were asked to present the **ICOs** occurring **each day** (Figure 1 shows **Jan 14 - 18**) in **a static graph**?

Note: Feel free to redesign the entire visualization but **keep the types of data shown**.

Present **at least two design alternatives** with drawings and text descriptions explaining your visual encodings (and interactions if any).

Question 1 grading scheme and answers

Grading scheme: 2 alternatives, 2.5 points each

Specifically,

- **ICO**: each cryptocurrency usually has only one initial coin offering
- **Daily**: not monthly; not accumulative
- **Time span**: from Jan 2014 to Aug 2018, more than 4 years
- **Static**: need overview; no tedious interaction (e.g., a view per day)
- **Continent**: encoding for continents (e.g., colors, stacked bars, …)
- **Data-ink rate**

Question 1 grading scheme and answers

Grading scheme: 2 alternatives, 2.5 points each

Candidate answers:

- **Stacked bar chart**: x-axis showing the days and bars (color-continent & length-amount) showing individual ICOs.

- **Calendar view**: inside each cell are the circles of the corresponding day.

Question 2

Figure 2. Napoleon's troops moved to and retreated from Moscow.

1) Modify this map to show the political leaning (**supporting France vs. supporting Russia**) of the **countries and regions**.

2) Design a visualization (**based on Figure 2**) to **encode** **the length of travel time** and **number of soldiers** to show the progress of the war and the turning point(s).

Question 2 grading scheme and answers

1) **(2.5 points)** **Choropleth map** on the base map with color showing the political leaning (color hue corresponding to the representative hue of France and Russia; may use saturation to show the strength of leaning).

2) **(2.5 points)** Use reasonable visual channels to encode the length of travel time and number of soldiers respectively, directly on each line segment.

   *Candidate designs:

   Color saturation/the elevation on the z-axis for duration

   Line thickness for the number of soldiers

In-class exercise 3

Open-ended question

Task: During the COVID-19 pandemic, many European Union (EU) **countries** launched vaccination visualizations to help **monitor the vaccination progress** and update policies accordingly.

Data: Given **two datasets** from Kaggle (COVID-19 world vaccination progress + country profile), illustrate all or a subset of the data (at least four attributes of your choice). The data was **high-dimensional**.
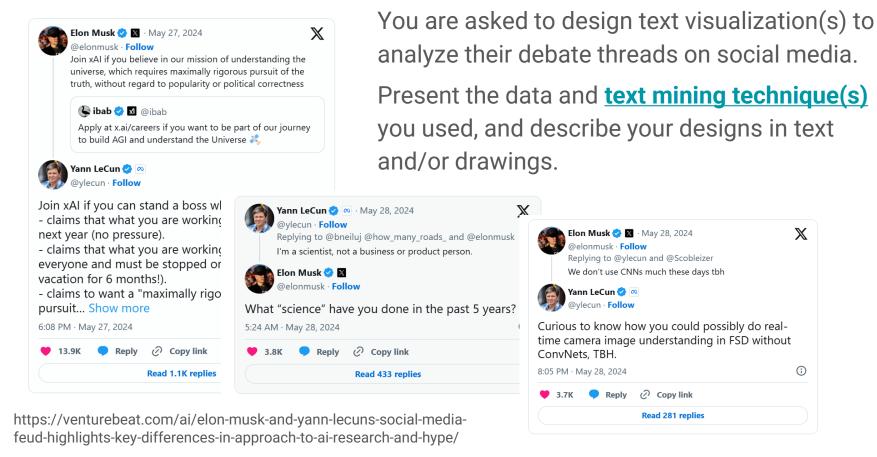
Please design such **a visualization** for EU.

1) Demonstrate your design (sketches/diagrams + text descriptions)

2) Clarify the "What", "Why", and "How" (Munzner's Nested Model of Vis Design).

Open-ended answer

1) - **Correlation and similarity** analysis on high-dimensional data;

   - The data items are the **EU countries**;

   - *Candidate designs/methods: Scatter plot matrix, Parallel coordinates, Dimensionality reduction (PCA, t-SNE, MDS, …), etc.

2) - What: types and attributes of shown data;

   - Why: the **tasks** that can be performed with your visualization(s) to help understand vaccination situations and inform vaccination policies in EU countries;

   - How: your proposed **idioms** and interactions.

In-class exercise 4

Question 1

You are asked to design text visualization(s) to analyze their debate threads on social media.

Present the data and **text mining technique(s)** you used, and describe your designs in text and/or drawings.

Question 1 answers

1) Identify the **topics** involved in Elon Musk and Yann LeCun's social media feud and the topical evolution **over time**.

Text mining technique: text clustering and topic modeling.

*Candidate design: Theme river (+ word cloud) to show topic evolution.

2) Compare the public's **sentiments** towards Elon Musk's and Yann LeCun's posts and the sentiment changes, if any, **over time**.

Text mining technique: sentiment analysis.

*Candidate design: Line/bar/heatmap to show the temporal change of sentiment.

Question 1 answers

3) Analyze who, Elon Musk or Yann LeCun, initiated **each topical thread**, how they **referred to** each other's posts, and how they engaged (@) external sources during their debates.

Text mining technique: entity and relation extraction.

*Candidate design: Node-link diagram to show the connectivity.

Tips for midterm 2

Coverage

The content will cover all lectures and labs up to November 18, with an **emphasis** on materials taught **after Midterm 1**.

- Lecture notes and recordings
- Tutorial notebooks
- In-class exercises (may facilitate your preparation)

Question types

The question types include multiple choice, analysis, problem-solving, and design, **the same as before**.

- Multiple choice
- Analysis
- Problem-solving
- Design

Cheatsheet

Note that each student can bring **one** additional A4-sized self-prepared **cheatsheet** (new in Midterm 2).

Quick review

Review - Data types (NOIR)

- **Categorical**:
  - Nominal *categorized*
  - Ordinal *categorized + ranked*

- **Quantitative**:
  - Interval *categorized + ranked + evenly spaced*
  - Ratio *categorized + ranked + evenly spaced + has a natural zero*

Review - Idiom (Visual Encoding) Design

- Visual encoding composition: mark + visual channel

- Mark: point/line/area/…   *[attribute of] item*

- Visual channel: *attribute value*

Channel Expressiveness Types and Effectiveness Ranks

Review - Idiom (Visual Encoding) Design

- Visual encoding composition: mark + visual channel
- Mark: point/line/area/…   *[attribute of] item*
- Visual channel: *attribute value*
  - Expressiveness types
  - Effectiveness ranks: separability, popout, scalability
  - Encoding with multiple visual channel: Integral/Separable
- Visual encoding grouping

Review - Rules/laws/principles

Please check their definitions and/or examples:

- Gestalt Principles

- Tufte's Principles of How to Achieve Graphical Excellence

- Tamara Munzner's Rules of Thumb

Review - Rules/laws/principles

Please check their definitions and/or examples:

- **Gestalt Principles**:
  - Proximity, Similarity, Connectedness, Enclosure, Continuity, Common fate, Symmetry, Closure, and Figure and ground

- Tufte's Principles of How to Achieve Graphical Excellence
- Tamara Munzner's Rules of Thumb

Review - Rules/laws/principles

Please check their definitions and/or examples:

- Gestalt Principles

- Tufte's Principles of How to Achieve Graphical Excellence:
  - Graphical integrity (*Lie factor, Consistent scale, and Present data in context*), Data-ink ratio, and Chart junk avoidance

- Tamara Munzner's Rules of Thumb

Review - Rules/laws/principles

Please check their definitions and/or examples:

- Gestalt Principles

- Tufte's Principles of How to Achieve Graphical Excellence

- Tamara Munzner's Rules of Thumb:

  - No unjustified 3D, No unjustified 2D, Eyes beat memory, Resolution over immersion, Visualization mantra (*Overview first, zoom and filter, details on demand*), Function first form next, Responsiveness is required, …

Review - Multivariate data visualization

- Multivariate Data: high dimensionality, interested in correlations between various attributes

- Visualization Techniques:

  - **Geometric Projection**:

    - Plots: Scatter plot, Scatter plot matrix, Parallel coordinates, …

    - Concept: Visual clutter reduction

  - Layout Density

  - Hierarchical Display

  - Iconography (Glyph)

- **Dimensionality Reduction**:

  - PCA, t-SNE, MDS

Review - Text visualization

- Text Mining Techniques:
  - Information extraction, Text summarization, Opinion mining and sentiment analysis, Text clustering, and Text classification
- Text visualization designs:
  - Word level: Word clouds (of all kinds), …
  - Document level: Typography, Treemap, Fingerprinting, …
  - Corpus level:
    - **Topic**, **Sentiment**: Heatmap, Scatter plot
    - **Trend**: Theme river, Area graph, Line graph
    - Discourse: Node-link diagram
    - Relation: Bubble chart, Pixel graph

Review - Graph visualization

- Graph Encoding Idioms:
  - **Node-link layout**:
    - General graph (of different variations)
    - Parallel layout: Sankey diagram
    - Tree
  - **Enclosure (nested) layout**:
    - Treemap: Slice-and-dice layout, Squarifed layout, Voronoi, Radial
  - **Division (layered) layout**:
    - Layered diagrams, Sunburst diagram
  - Matrix layout
  - 3D layout

Review - Graph visualization

- **Visual Clutter Reduction**:
  - Visual clutter: excess items (or their representation or organization) lead to a degradation of performance at some task.
  - Reduction approaches:
    - Edge-centric: Edge bundles, Confluent drawings, Hierarchical trees
    - Node-centric: Limitations of node clustering
    - Appearance-centric: Sampling, Filtering, Clustering, Changing point size/opacity
    - Space-centric
- **Interactions** and Navigation:
  - Select, Zoom and Pan, Filter (Dynamic query controls, Magic lens), Animation

Review - Validation and evaluation

Validation Approaches

		Threat — Wrong problem	
Step 1		Validate — Observe and interview target users	Before design, with users
		Threat — Wrong task/data abstraction	
		Threat — Ineffective encoding/interaction idiom	
Step 2		Validate — Justify encoding/interaction design	During design, heuristic
		Threat — Slow algorithm	
Step 3		Validate — Analyze computational complexity	During design, technical
		Implement system	
		Validate — Measure system time/memory	
Step 4		Validate — Qualitative/quantitative result image analysis	After design, with users, in a controlled environment
		Test on any users, informal usability study	
		Validate — Lab study, measure human time/errors for task	
Step 5		Validate — Test on target users, collect anecdotal evidence of utility	After design, with users, in the field
		Validate — Field study, document human usage of deployed system	
Step 6		Validate — Observe adoption rates	After design, with users, in the wild, longitudinal

Q&A now

Midterm 2
Next Friday
22 Nov (80 min)

Reminder:

- During the lecture in the regular classroom
- In-person, paper-based, close-book, close-note
- One additional A4-sized self-prepared **cheatsheet**
- Bring your student **ID card**, (color) pens, pencils, erasers, rulers, etc.