



HKUST
VISLAB

HKUST
HCI Initiative

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

[illegible]

Token Sales, Jan14-Aug18

Total Raised:
\$28,366,035,804

Legend:

- Europe
- North America
- Asia
- Caribbean
- South America
- Oceania
- Middle East
- Africa
- Stateless/Unknown

Key Tokens and Amounts:

- EOS \$4,213m
- Telegram \$1,700m
- Huobi \$500m
- Filecoin \$257m
- Dragon \$420m
- TaTuTu \$575m
- Tezos \$236m
- EOS \$300m
- EOS \$150m
- EOS \$100m
- EOS \$50m
- EOS \$25m
- EOS \$12.5m
- EOS \$6.25m
- EOS \$3.125m
- EOS \$1.5625m
- EOS \$0.78125m
- EOS \$0.390625m
- EOS \$0.1953125m
- EOS \$0.09765625m
- EOS \$0.048828125m
- EOS \$0.0244140625m
- EOS \$0.01220703125m
- EOS \$0.006103515625m
- EOS \$0.0030517578125m
- EOS \$0.00152587890625m
- EOS \$0.000762939453125m
- EOS \$0.0003814697265625m
- EOS \$0.00019073486328125m
- EOS \$0.000095367431640625m
- EOS \$0.0000476837158203125m
- EOS \$0.00002384185791015625m
- EOS \$0.000011920928955078125m
- EOS \$0.0000059604644775390625m
- EOS \$0.00000298023223876953125m
- EOS \$0.000001490116119384765625m
- EOS \$0.0000007450580596923828125m
- EOS \$0.00000037252902984619140625m
- EOS \$0.000000186264514923095703125m
- EOS \$0.0000000931322574615478515625m
- EOS \$0.00000004656612873077392578125m
- EOS \$0.000000023283064365386962890625m
- EOS \$0.0000000116415321826934814453125m
- EOS \$0.00000000582076609134674072265625m
- EOS \$0.000000002910383045673370361328125m
- EOS \$0.0000000014551915228366851806640625m
- EOS \$0.00000000072759576141834259033203125m
- EOS \$0.000000000363797880709171295166015625m
- EOS \$0.0000000001818989403545856475830078125m
- EOS \$0.00000000009094947017729282379150390625m
- EOS \$0.000000000045474735088646141895751953125m
- EOS \$0.0000000000227373675443230709478759765625m
- EOS \$0.00000000001136868377216153547393798828125m
- EOS \$0.000000000005684341886080767736968994140625m
- EOS \$0.000000000002842170943040383868484497072265625m
- EOS \$0.0000000000014210854715201919342422485361328125m
- EOS \$0.00000000000071054273576009596712112426806640625m
- EOS \$0.000000000000355271367880047983560562134033203125m
- EOS \$0.0000000000001776356839400239917802810670166015625m
- EOS \$0.00000000000008881784197001199589014053350830078125m
- EOS \$0.000000000000044408920985005997945070266754150390625m
- EOS \$0.0000000000000222044604925029989725351333770703125m
- EOS \$0.00000000000001110223024625149948626766688853515625m
- EOS \$0.000000000000005551115123125749743133833444267798828125m
- EOS \$0.0000000000000027755575615628748715669167221338994140625m
- EOS \$0.0000000000000013877787807814374357834583610669497072265625m
- EOS \$0.00000000000000069388939039071871789172918053347485361328125m
- EOS \$0.000000000000000346944695195359358945864590266737426806640625m
- EOS \$0.0000000000000001734723475976796794729322951333687134033203125m
- EOS \$0.00000000000000008673617379883983973646614756668435670166015625m
- EOS \$0.000000000000000043368086899419919868233073783342178350830078125m
- EOS \$0.000000000000000021684043449709959934116536891667108925390390625m
- EOS \$0.0000000000000000108420217248549799670582684458335544626970166015625m
- EOS \$0.00000000000000000542101086242748998352913422244167772338850830078125m
- EOS \$0.00000000000000000271050543121374499676456711122083886169425390390625m
- EOS \$0.00000000000000000135525271560687249838228355561004193097126970166015625m
- EOS \$0.0000000000000000006776263578034362496911417777800209654856134033203125m
- EOS \$0.00000000000000000033881317890171812484557088889001048274280670166015625m
- EOS \$0.0000000000000000001694065894508590624227854444450052413714033203125m
- EOS \$0.00000000000000000008470329472542953121114272222250026218570166015625m
- EOS \$0.000000000000000000042351647362714765605571361111250013109350830078125m
- EOS \$0.0000000000000000000211758236813573828027856555562500065546754150390625m
- EOS \$0.000000000000000000010587911840678691440139277778125000327733703125m
- EOS \$0.0000000000000000000052939559203393457200696388890625001638668515625m
- EOS \$0.0000000000000000000026469779601696728600348194445312500081933426806640625m
- EOS \$0.0000000000000000000013234889800848364300170959722226

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.

**Elon Musk**   · May 27, 2024
[@elonmusk](#) · [Follow](#)

Join xAI if you believe in our mission of understanding the universe, which requires maximally rigorous pursuit of the truth, without regard to popularity or political correctness

**ibab**   @ibab
Apply at [x.ai/careers](#) if you want to be part of our journey to build AGI and understand the Universe 🚀

**Yann LeCun**   · May 27, 2024
[@ylecun](#) · [Follow](#)

Join xAI if you can stand a boss w/

- claims that what you are working next year (no pressure).
- claims that what you are working everyone and must be stopped or vacation for 6 months!).
- claims to want a "maximally rigorous pursuit..." [Show more](#)

6:08 PM · May 27, 2024

 13.9K  Reply  Copy link

[Read 1.1K replies](#)

**Yann LeCun**   · May 28, 2024
[@ylecun](#) · [Follow](#)

Replying to [@bneiluj](#) [@how_many_roads_](#) and [@elonmusk](#)
I'm a scientist, not a business or product person.




**Elon Musk**   · May 28, 2024
[@elonmusk](#) · [Follow](#)

What "science" have you done in the past 5 years?




5:24 AM · May 28, 2024

 3.8K  Reply  Copy link

[Read 433 replies](#)




**Elon Musk**   · May 28, 2024
[@elonmusk](#) · [Follow](#)

Replying to [@ylecun](#) and [@Scobleizer](#)
We don't use CNNs much these days tbh

**Yann LeCun**   · May 28, 2024
[@ylecun](#) · [Follow](#)

Curious to know how you could possibly do real-time camera image understanding in FSD without ConvNets, TBH.

8:05 PM · May 28, 2024

 3.7K  Reply  Copy link

[Read 281 replies](#)

<https://venturebeat.com/ai/elon-musk-and-yann-lecuns-social-media-feud-highlights-key-differences-in-approach-to-ai-research-and-hype/>

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).

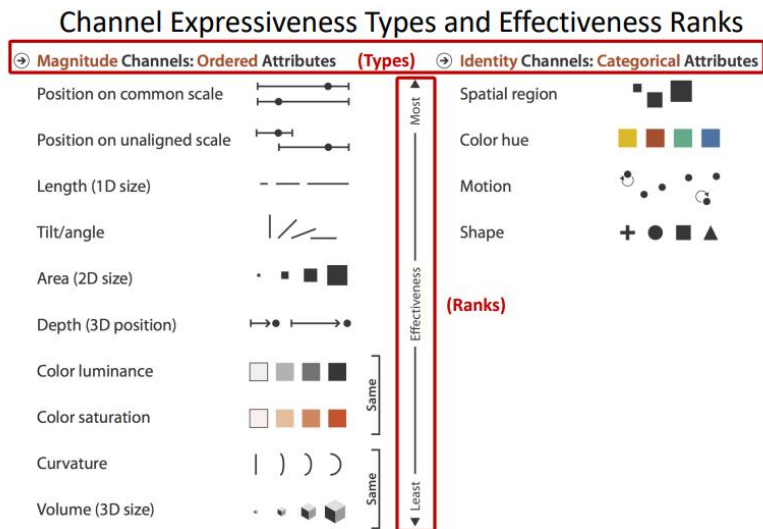
Quick review

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*



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, Squarified layout, Voronoi, Radial
 - **Division (layered) layout:**
 - Layered diagrams, Sunburst diagram
 - Matrix layout
 - 3D layout

Review - Graph visualization

- **Visual Clutter Reduction:**

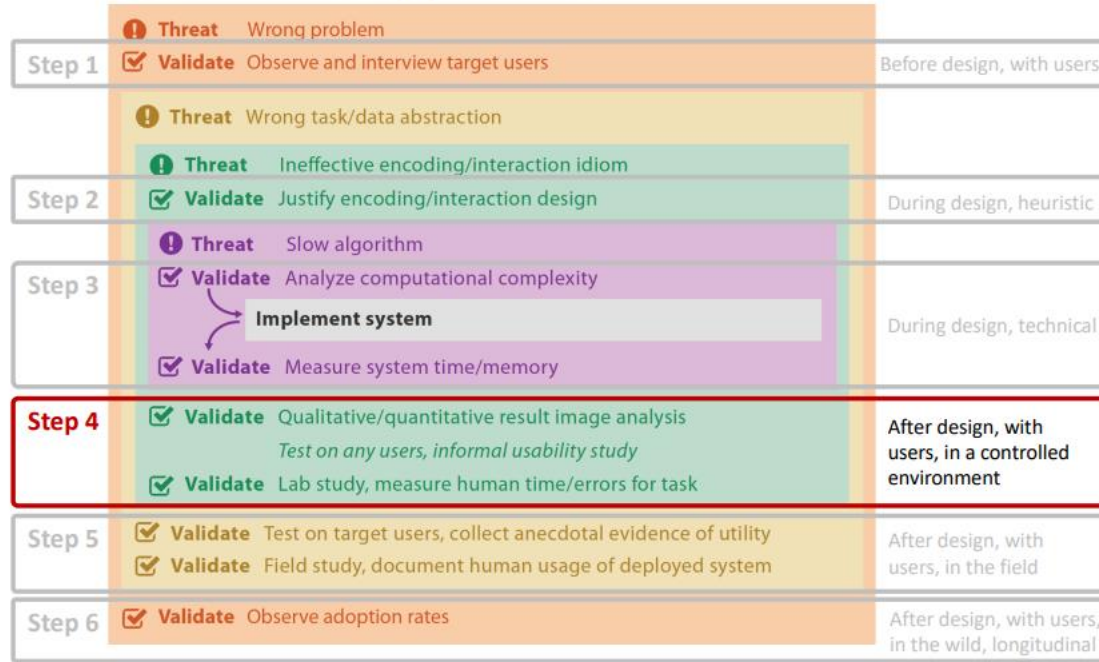
- Visual clutter: **excess items** (or their representation or organization) lead to a degradation of performance at some task.
- Reduction approaches:
 - **Edge-centric:** Edge bundles, Confluent drawings, Hierarchical trees
 - **Node-centric:** Limitations of node clustering
 - **Appearance-centric:** 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



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.