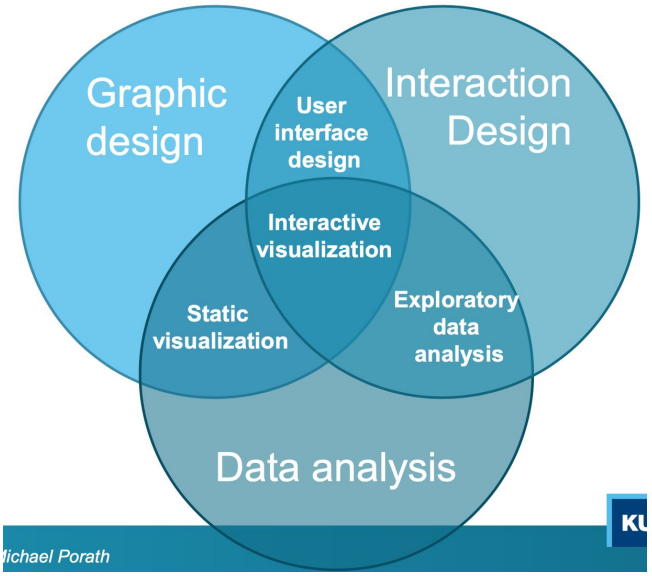
CHECKLIST OF DATA VIZ PRINCIPLES



Lecture 1: Introduction

Tufte's principles:

1. Tell the truth with relevance and integrity of the content
2. The purpose of DV is effective communication.
3. The key of effectiveness is clarity and precision.

Design aesthetics - 5 principles:

1. Show the data
2. Maximize the data-ink ratio, within reason
3. Erase non-data ink, within reason
4. Erase redundant data-ink (less is more)
5. Revise and edit

Summary:

1. Avoid (intentional) misleading
2. Don't use too much decorative ink to:
   - Maximize the data clarity
   - Make data standout from the background
3. Additional data can appear on demand

data-ink ratio = data-ink / total ink used to print graphic
lie factor = size of effect shown in graphic / size of effect in data

Workflow (we did them all):

1. Acquire

2. Parse
3. Filter
4. Mine
5. Represent
6. Refine
7. Interact

Lecture 2: Perception and Principles

ATTENTION TO PRE-ATTENTIVE UNDERSTANDING

Speed of perception can change due to different factors:

- Different shapes pop out
  (example if you have a square shape inside all round squares, you will immediately see the square, because is different from the others)
- Different colour can be pre-attentively identified
- Magnitude estimation (between two object with same shape)
- Perceptual or apparent scaling can compensate magnitude to match perception

QUANTITATIVE, ORDINAL AND CATEGORICAL DATA

Quantitative data
Position / Length / Angle / Slope / Area / Volume / Density / Shape

Ordinal data
Position / Density / Colour saturation / Texture / Connection / Containment / Length / Angle / Slope / Area / Volume

Categorical data
Position / Colour hue / Texture / Connection / Containment / Density / Colour saturation/ Shape / Length / Angle / Slope / Area / Volume

EXPRESSIVENESS TYPES AND EFFECTIVENESS RANKINGS

Effectiveness principle:

- encode most important attributes with highest ranked channels

Expressiveness principle:

- match channel and data characteristics

GESTALT PRINCIPLES
Guideline: Use a combination of closure, common region and layout to ensure that data entities are represented by graphical patterns that will be perceived as figure, not ground.

- figure and ground

- proximity
- similarity
- connectedness
- continuity
- closure
- smallness
- surroundedness

COLOR
Take into consideration color-blindness
red and green >>> absolutely NO for color-blindness
blue and orange >>> always safe

How to use colors:
- hue: categorical data
- saturation: ordinal and quantitative data
- luminance: ordinal and quantitative data

Avoid using gray scales as a method for representing more than a few (two to four) numerical values;

Haloing effect (can be used when you have a colored element in another-color background)
- Enhancing the edges
- Luminance contrast as a highlighting method

Highlighting: make small subset clearly distinct from the rest

Guidelines
- Use more saturated colors for small symbols, thin lines, or small areas.
- Use less saturated colors for large areas.

Use easy-to-remember and consistent color codes in color pallets
Red, green, blue and yellow are hard-wired into the brain as primaries. If it is necessary to remember a color coding, these colors are the first that should be considered.

Chromo Stereopsis
If we use in the same image two far pure colors the eye is notable to focus both of them
>>>>> DO NOT use red and blue together in a black background, and do not use blue in a red background, nor vice versa.

Guidelines
- Beware of interactions between some colors (e.g. red/blue)
- Use can be good: for highlighting, creating 3D effect, etc.
- Resolve if unintended by: 1) using colors that are less saturated; 2) surrounding the contrasting colors with a background that moderates the effect of their different wavelengths; 3) separating the contrasting colors.

To summarize:
Color is excellent for labeling and categorization.
(However, only small number of colors can be used effectively)

- To show detail in visualization, always have considerable luminance contrast between background and foreground.
- Simultaneous contrast with background colors can dramatically alter color appearance, making color look like another.
- Beware of interaction between colors (e.g. red/blue).
- Small color coded objects should be given high saturation.
- Red, green, blue and yellow are hard-wired into the brain as primaries. If it is necessary to remember a color coding, these colors are the first that should be considered.
- Remember that colors have meanings: use appropriate color palettes for qualitative, quantitative and ordinal data.
- Respect the color blind.

Lecture 3: Representation

Encoding of values:

1. Univariate data:
   box plot
   histogram/bargram (vertical/horizontal)
   wordcloud (limited information shown; not recommended)
   Anscombe's quartet
2. Bivariate data:
   scatter plot
   time series
   linked histogram
3. Trivariate data:
   choropleth map
   scatterplot matrix (smart solution example: the highlighting of one element in one plane can be brushed into the remaining planes);
   * Hugo’s genre/year/ego-position chart is a non-static representation of trivariate data
4. Hypervariate data:
   parallel coordinates
   star/radar lot
   mosaic plot
   infocanvas
   icons/faces (not recommended)

Encoding of relations:

- Relation: a logical or natural association between two or more things; relevance of one to another; connection.
- Relationship: The association/relevance/connection itself.

1. Use lines and colours to indicate relations and relationships between groups:
   network
   chord diagram
2. Use maps and diagrams to explore the intersection:
   Venn diagram
   cluster map.
   More intersections of entities brings more effectiveness
   i) Effectiveness = time of success / time of exploration
   ii) Items which are relevant in more than one aspect are usually more valuable. Displaying multiple aspects of relevance visually helps the user to explore the items.
3. Use trees to show hierarchical data:
   trees
   cone trees
   cam trees
   hyperbolic trees
   tree map
   sunburst

Lecture 4: Interaction and Presentation

INTERACTIVE REPRESENTATIONS
Interactive representations can be set in the intersection between Graphic design, Interactive design and Data analysis, and also in the intersection between User Design Interface, Static Visualization and Exploratory Data Analysis

WHY INTERACTION IS USEFUL? INTERACTION HAS AN EXPLORATORY ASPECT
In these cases a static representation is not enough:
Scale (too many data points, too many different dimensions)
Storytelling
Exploration
Learning

WHAT IS THE USER INTENT?
/ Select and focus (pick a detail to keep track of it)
/ Reconfigure (show a different arrangement)
/ Encode (show a different representation - change visual variables)
/ Abstract-elaborate (show more or less detail)
/ Filter (show something conditionally)
/ Connect (show related items - brushing and linking)

Guidelines:

- Avoid ambiguities in interaction. Enhance the user navigation. Use stepped interaction in discrete information spaces.
- Support the user to form an action plan >> he must understand what to do and how.

User should asks himself Where can I go? How do I get there? What lies beyond?
Where can I usefully go from here?
Overall, User must assess benefit of each available movement.

SENSITIVITY is a movement in information space and the interaction required to achieve it.
Can use dynamic query interface
SCENT is the perceived benefit associated with a movement in the information space, evaluated following the interpretation of one or more cues.

BREADCRUMBS can potentially help the user to understand where he is (in a chronology, or in a location)

HOW TO PRESENT A VISUALISATION:
Some applications can be useful in a presentation, given that every presentation has always 2 conditions: space limitations and time limitations:

- OVERVIEW and DETAIL
- DISTORTION TECHNIQUES
- SUPPRESSION (Basic idea: more relevant information presented in great detail; the less relevant information presented as an abstraction)
- DEGREE OF INTEREST (DOI)
- ZOOM and PAN (especially for maps)
- RAPID SERIAL VISUAL PRESENTATION
- EYE GAZE

Lecture 5: Informational Dashboards

Characteristics of dashboards:

1. Dashboards are visual displays.
2. Dashboards display information needed to achieve specific objectives.
3. A dashboard fits on a single computer screen.
4. Dashboards are used to monitor information at a glance.
5. Dashboards have small, concise, clear, intuitive display mechanisms.
6. Dashboards are customized.

Avoid 13 common mistakes:

1. Exceed the boundaries of a single screen (information on dashboards is often fragmented in 1 or 2 ways: Separated into discrete screens wo which one must navigate or separated into different instances of a single screen that are accessed through some form of interaction / require scrolling)

2. Supplying inadequate context for the data
3. Displaying excessive detail or precision
4. Choosing a deficient measure (use of measures that fail to directly express the intended message)
5. Choosing inappropriate display media (pie charts are difficult to interpret accurately; horizontal bar graphs are better. Other types of graphs can be equally ineffective)

6. Introducing meaningless variety
7. Using poorly designed display media (don't put a legend next to a pie chart because this way your eyes have to go back and forth between the graph and the legend; order of the slices shouldn't be random; bright colors produce sensory overkill.
8. Encoding quantitative data inaccurately
9. Arranging the data poorly (most important data should be prominent; data that should be compared ought to be arranged and visually designed to encourage comparisons)
10. Highlighting important data ineffectively or not at all
11. Cluttering the display with useless decoration
12. Misusing or overusing color (too much color undermines its power)
13. Designing an unattractive visual display

Strategies to create more effective dashboards

1. Reducing the non-data pixels
2. Enhancing the data pixels
3. Designing dashboards for usability
4. Exceptionally well organized
5. Condensed (primarily in the form of summaries and exceptions)
    - Summarize the data as a single number
    - Show critical values exceptions (show the one or two data points that require attention)
6. Specific to and customized for the audience and objectives
7. Eliminate and enhance data pixels
8. Displayed using concise and often small media that communicate the data and its message in the clearest and most direct way possible
9. Support meaningful comparisons
    - Combine items into a single table or graph
    - Place items close to one another
    - Linking items in different groups using a common color
    - Include comparative values (ratios, percentages, etc)

Lecture 6: Communication

<u>Choose a data strategy</u>

1. visualize insights, not data itself
2. consider data proxies, not data itself
3. visualize data interpretations, not data itself

Choose a visualization strategy
4. follow a design approach
5. avoid visualization pitfalls
6.follow good practice
Choose a user engagement strategy
7. visualize a purpose
8. integrate narrative and storytelling principles
Be critical
9. visualization is subjective

Where is the insight value of your visualization situated???
(describe-interpret-predict-prescribe)

Visualization purpose:

- Functional purpose: ANALYZE? EXPLAIN? INSPIRE? - The purpose is significant in how to represent the data
- Engagement purpose: SHOW - TELL - ENGAGE

Narrative: structure a Story:
How - When - Where - Who - What - Why