Visualizing Data
Lecture 11

Wan Fang

Southern University of Science and Technology

Agenda

- DO and DON'T
  - Learn from good and bad practices
- Interaction
  - DIKW and the user side gap
  - The Cycle of Encoding and Decoding
  - Information-seeking mantra by Ben Shneiderman

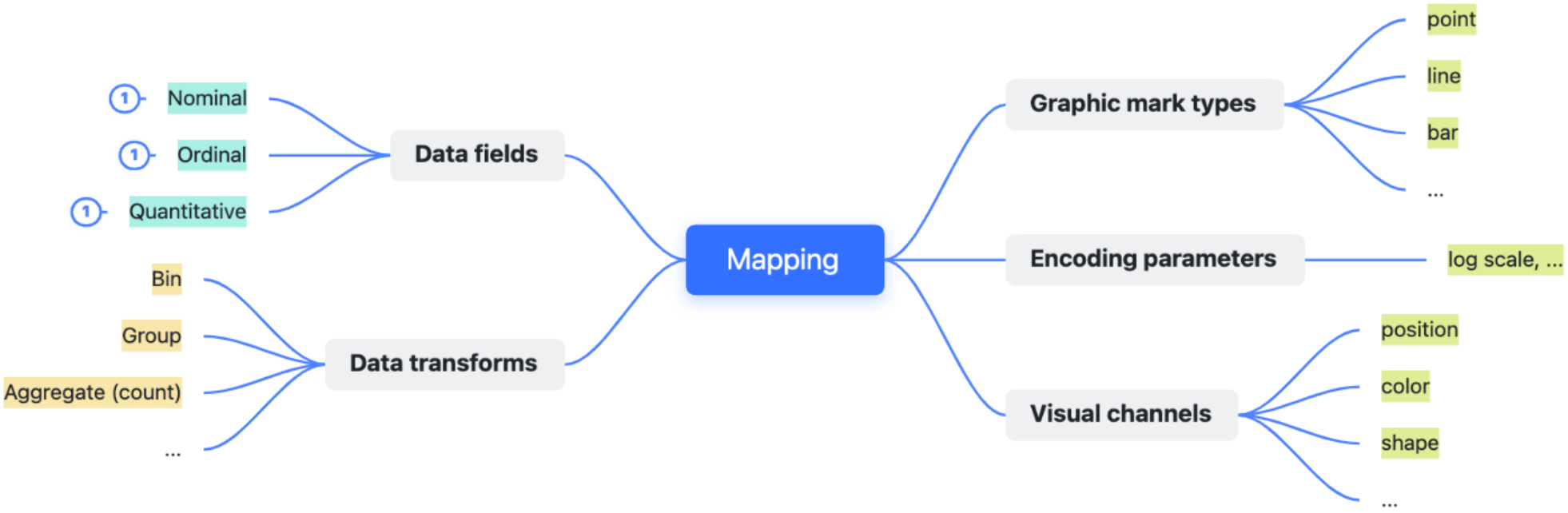
DO and DON'T

Wan Fang

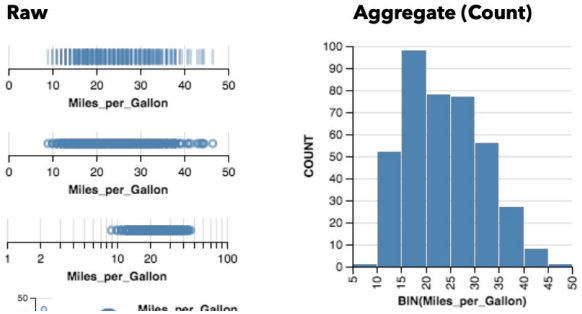
Southern University of Science and Technology

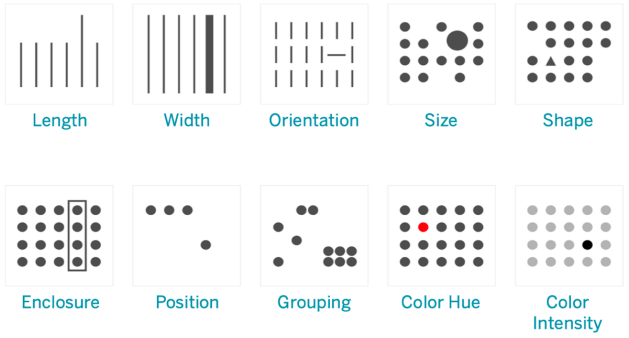
[Adapted from ID430B Data Analytics for Designers, KAIST]

Mapping Data to Visual Variables



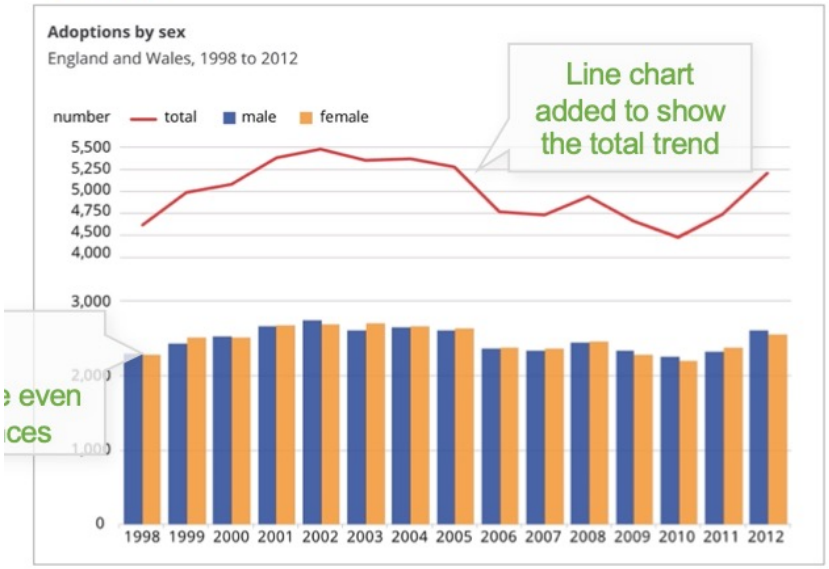
1D: Quantitative





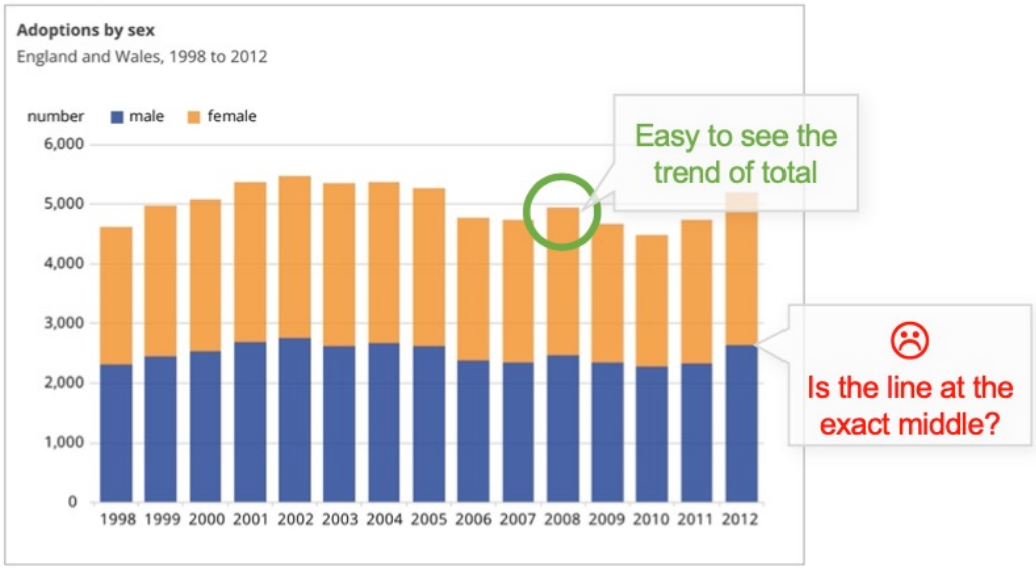
Place things to compare side-by-side

GOOD



Grouped Bar Chart

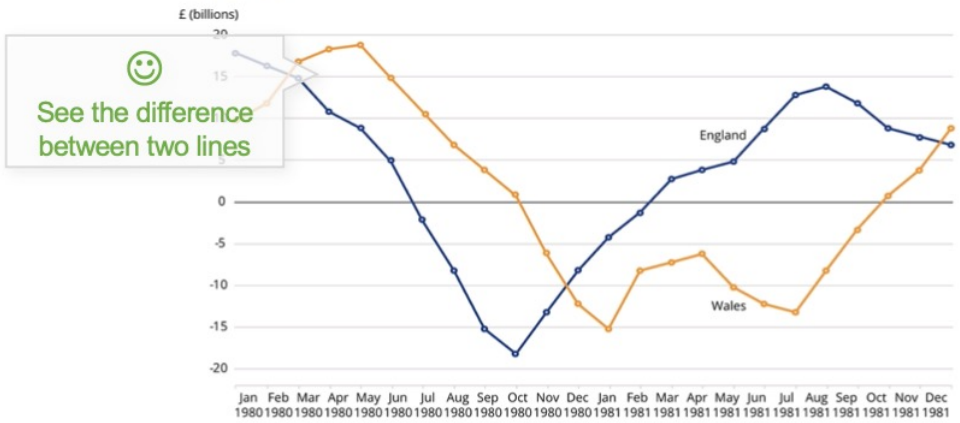
BAD



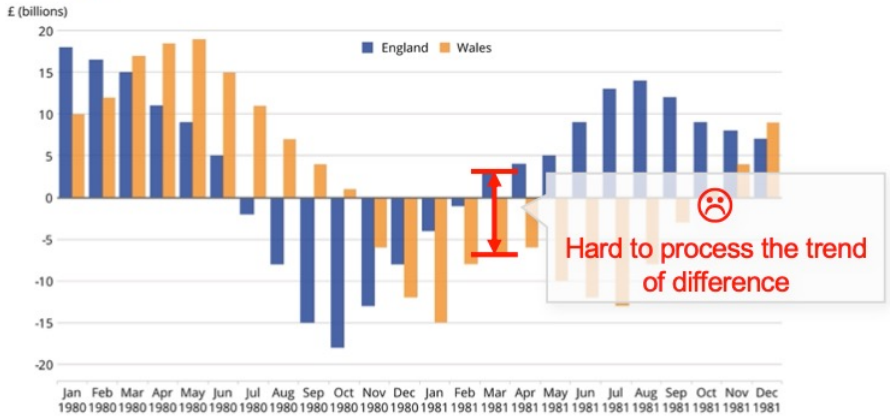
Stacked Bar Chart

Minimize viewer's effort to catch the main signal

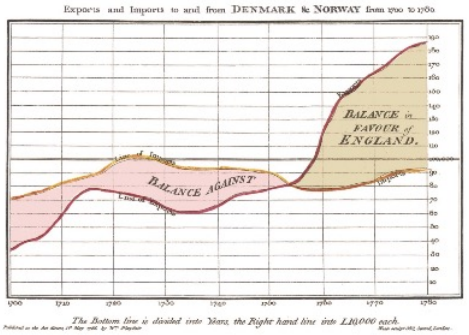
GOOD



BAD



BETTER



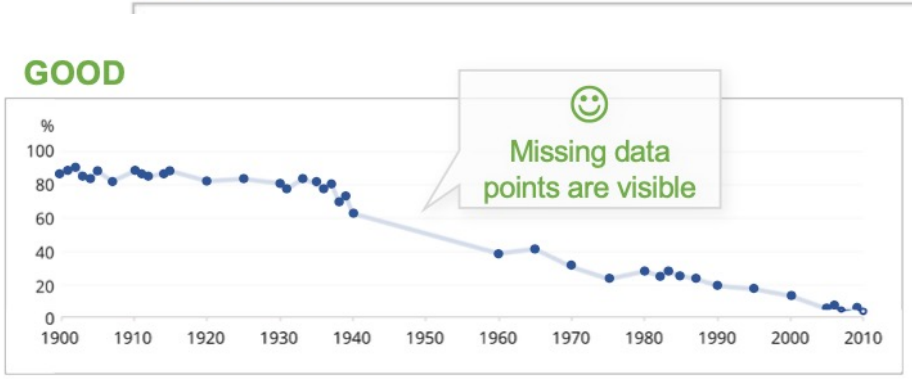
William Playfair's famous line chart

TAKEAWAY.

- Different analytic tasks might require different chart types for the same data
- Minimize the viewer's effort to catch the main signal

Beware of the common assumption of the chart type

Line chart assumes that data points are evenly distributed



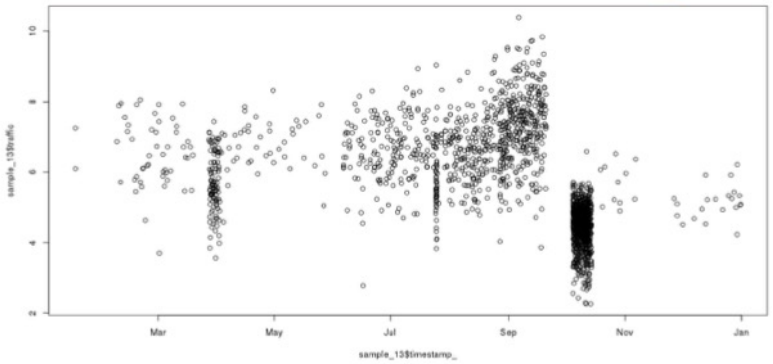
Scatter plot + dimmed line chart

: The line implies the chart focuses on timeline trends. However, missing dots warn "This part of the timeline has no data point."

Simple line chart

: Line charts imply that the axis is regular (i.e. data points are evenly-spaced). Viewers may have a false assumption for unevenly-spaced time series.

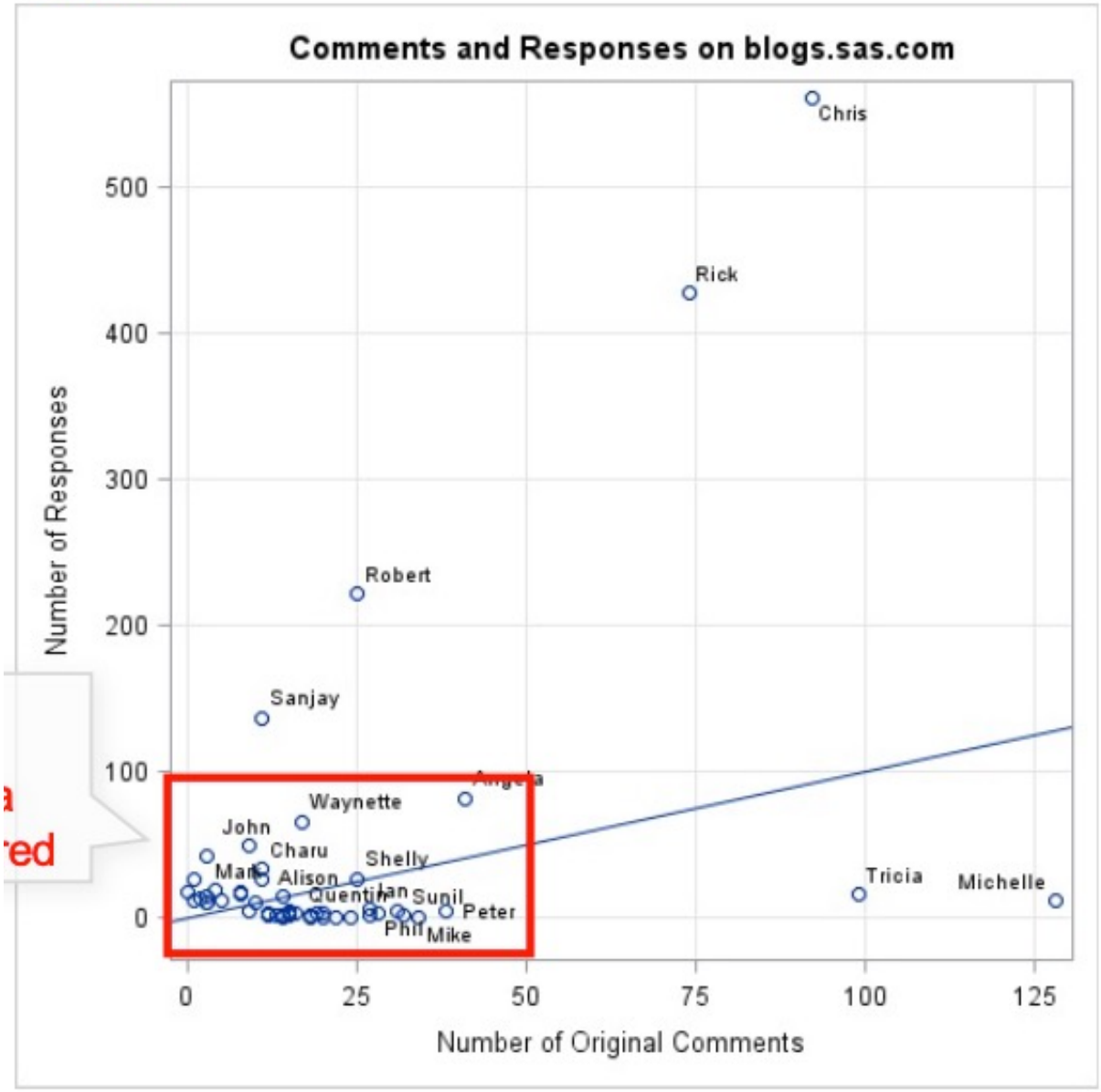
An Extreme Case of Unevenly Spaced Time Series (link)



TAKEAWAY.

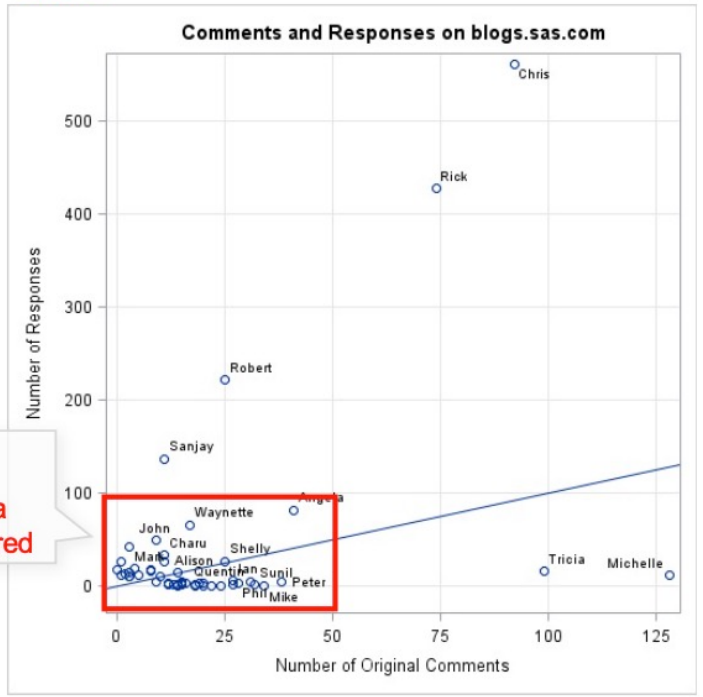
- Make missing data points visible.
- Beware of the common assumption of the chart type (e.g. line charts look like data points are evenly distributed)
- You can infer missing data points but that's not part of visualization

Avoid overcrowded markers



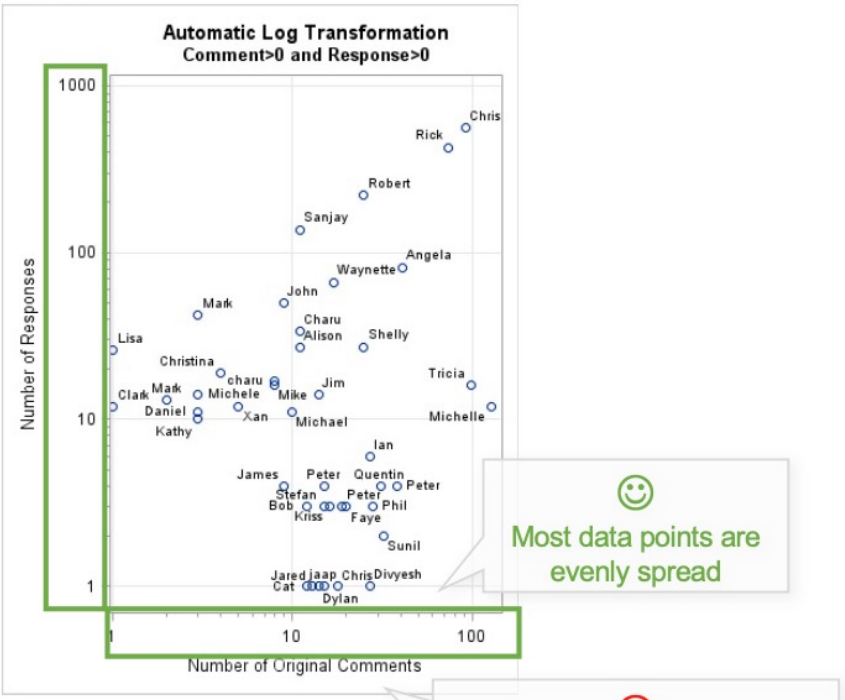
Avoid overcrowded markers

BAD



☹ Too many data points are cluttered

GOOD



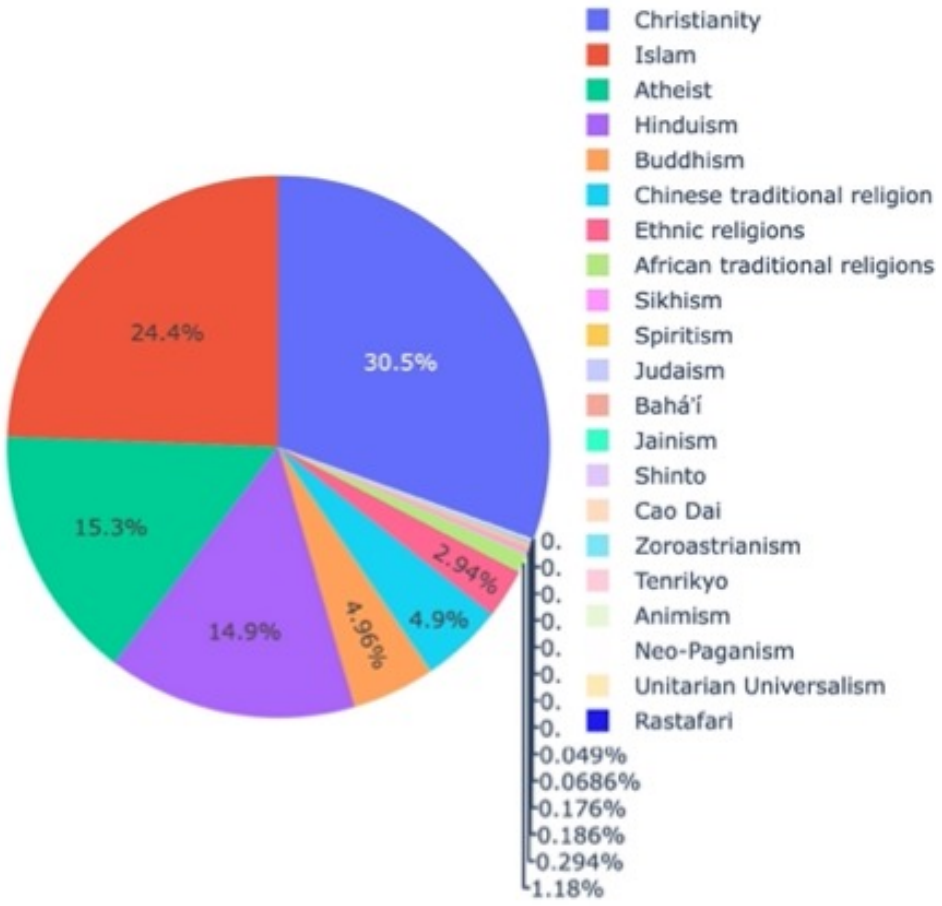
☺ Most data points are evenly spread

☹ Log axis might be hard to interpret for some audiences

How to visualize Part-to-Whole

For large # of distinct items

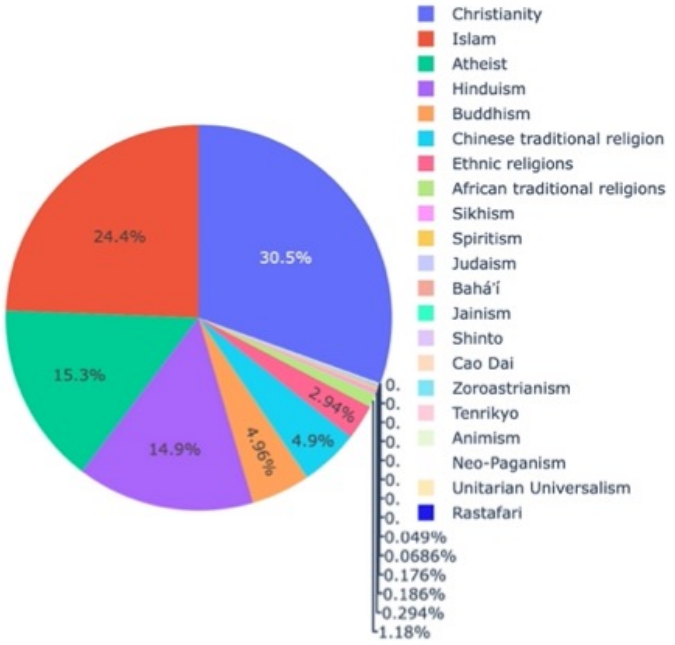
Pie charts are sensitive to the cardinality (i.e. # of distinct items should not be much larger than six); Also it's not easy to visually compare ratios of items



How to visualize Part-to-Whole

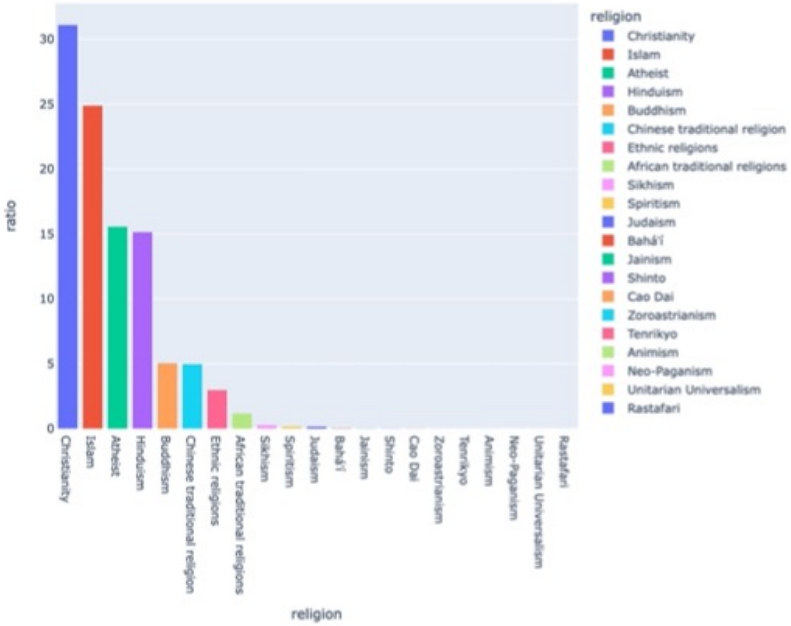
For large # of distinct items

BAD



Pie charts are sensitive to the cardinality (i.e. # of distinct items should not be much larger than six); Also it's not easy to visually compare ratios of items

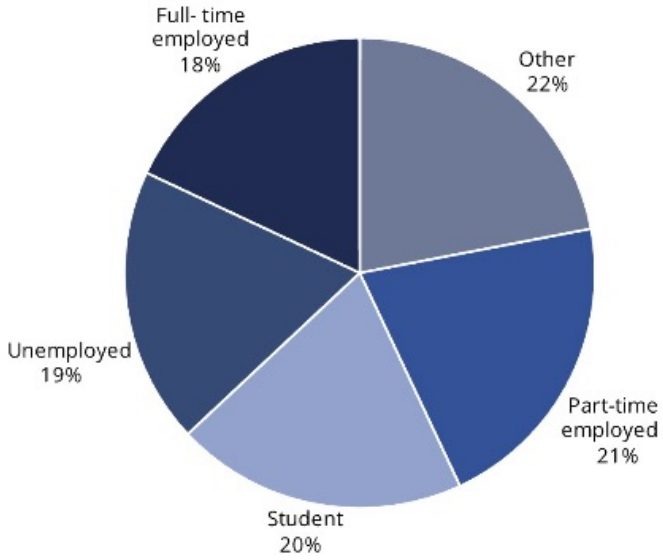
GOOD



Bar charts are better with large # of distinct items.

How to visualize Part-to-Whole

When there's no dominant item,

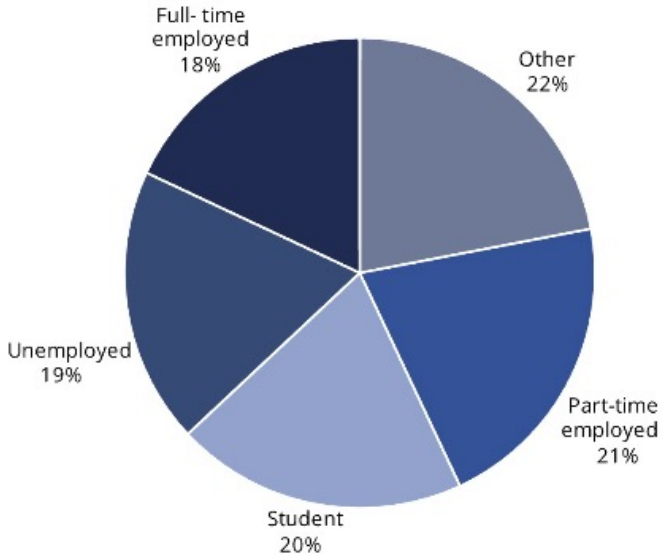


Pie charts are bad at comparing similar ratios. Labels are required then.

How to visualize Part-to-Whole

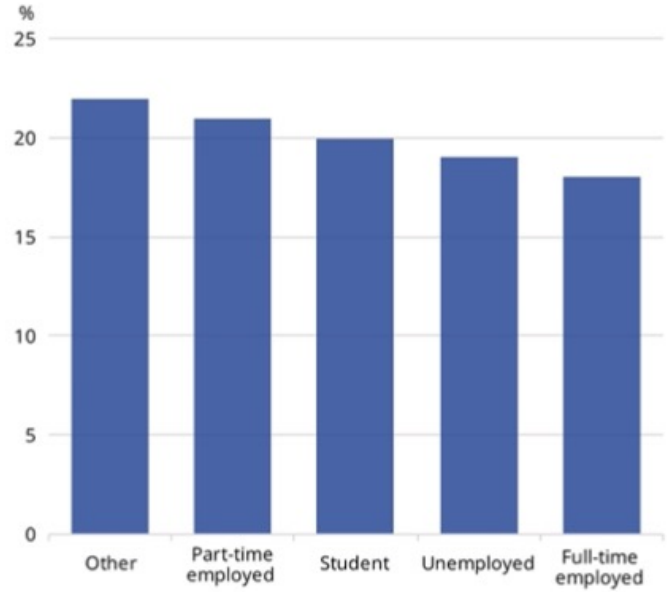
When there's no dominant item,

BAD



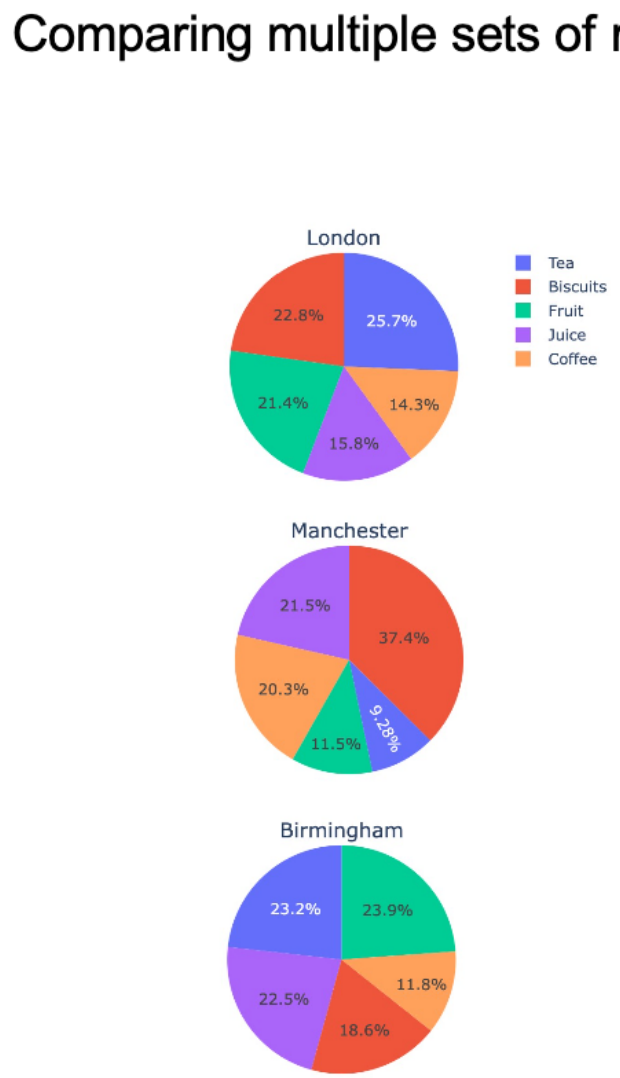
Pie charts are bad at comparing similar ratios. Labels are required then.

GOOD



Bar charts are better at comparing similar ratios (i.e. side-by-side comparison is the best)

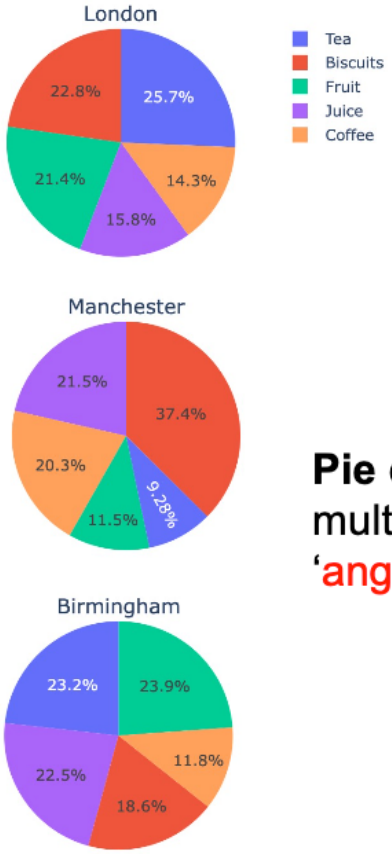
How to visualize Part-to-Whole



How to visualize Part-to-Whole

Comparing multiple sets of ratios

BAD

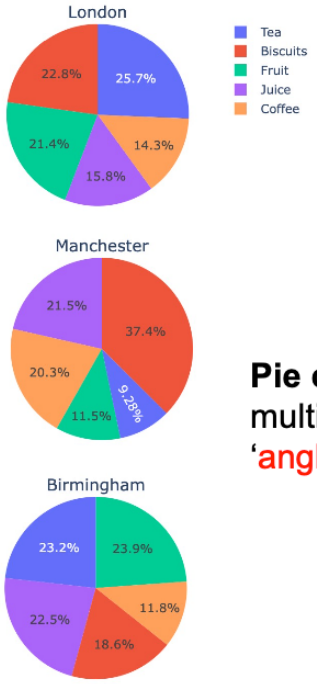


Pie charts are bad at comparing multiple sets of ratios because 'angle' is hard to compare.

How to visualize Part-to-Whole

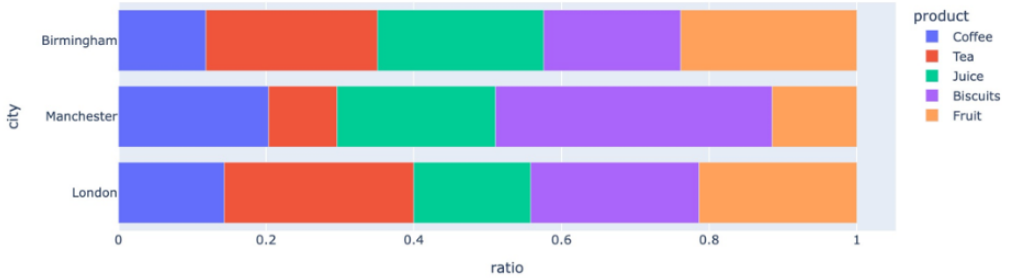
Comparing multiple sets of ratios

BAD



Pie charts are bad at comparing multiple sets of ratios because 'angle' is hard to compare.

GOOD



Stacked bar charts are good at comparing multiple sets of ratios. 'Length' is a good encoding to compare values.

If Pie charts have only bad things, why do people use it a lot?

There could be historical, aesthetic, and other reasons. My favorite explanation is that "*Pie charts are applicable to part-to-whole relationship only*". That means, even before reading text, readers would correctly guess the meaning of pie charts. In contrast, bar charts are effective for a wide range of data and questions – which means viewers have to read title, legends, and other textual information to interpret.

(optional reading) Why human love pie charts

How to visualize survey results on Likert scale

What is Likert scale?

- Commonly used in survey or user study; Captures the feeling of intensity for a given item
- Is Likert scale an ordinal or interval measure? Still lots of debates!

To what extent do you agree with the following statement?

"I will vote for the policy A"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

1. Amieson, Susan (2004). "Likert Scales: How to (Ab)use Them". Medical Education. 38 (12): 1217–1218.
2. Norman, Geoff (2010). "Likert scales, levels of measurement and the "laws" of statistics". Advances in Health Science Education. 15 (5): 625–632.

How to visualize survey results on Likert scale

What is Likert scale?

- Commonly used in survey or user study; Captures the feeling of intensity for a given item
- Is Likert scale an ordinal or interval measure? Still lots of debates!

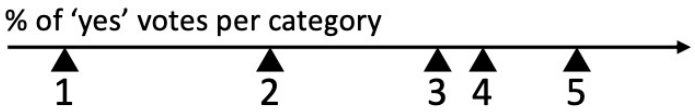
Likert is Ordinal ☹

"Participants tend to avoid extreme responses. Thus the categories are not equally distributed."

"In many situations, participants have motivations to lean toward positive / negative responses."

To what extent do you agree with the following statement?
"I will vote for the policy A"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5



Likert is Interval ☺

*"A well-designed question (i.e. having a clear neutral position, symmetrical and equidistant categories) can **approximate an interval measurement**, with large N"*

1. Amieson, Susan (2004). "Likert Scales: How to (Ab)use Them". Medical Education. 38 (12): 1217–1218.
2. Norman, Geoff (2010). "Likert scales, levels of measurement and the "laws" of statistics". Advances in Health Science Education. 15 (5): 625–632.

How to visualize survey results on Likert scale

What is Likert scale?

- Commonly used in survey or user study; Captures the feeling of intensity for a given item
- Is Likert scale an ordinal or interval measure? Still lots of debates!

Likert is Ordinal ☹

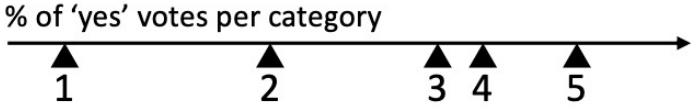
"Participants tend to avoid extreme responses. Thus the categories are not equally distributed."

"In many situations, participants have motivations to lean toward positive / negative responses."

To what extent do you agree with the following statement?

"I will vote for the policy A"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5



Likert is Interval ☺

*"A well-designed question (i.e. having a clear neutral position, symmetrical and equidistant categories) can **approximate an interval measurement**, with large N"*

"The legitimacy of assuming an interval scale for Likert-type categories is an important issue, because the appropriate descriptive and inferential statistics differ for ordinal and interval variables."

1. Amieson, Susan (2004). "Likert Scales: How to (Ab)use Them". Medical Education. 38 (12): 1217–1218.
2. Norman, Geoff (2010). "Likert scales, levels of measurement and the "laws" of statistics". Advances in Health Science Education. 15 (5): 625–632.

How to visualize survey results on Likert scale

What is Likert scale?

- Commonly used in survey or user study; Captures the feeling of intensity for a given item
- Is Likert scale an ordinal or interval measure? Still lots of debates!

Likert is Ordinal ☹

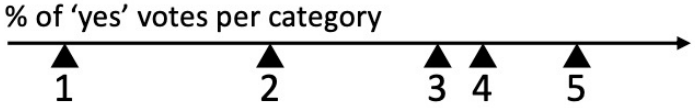
"Participants tend to avoid extreme responses. Thus the categories are not equally distributed."

"In many situations, participants have motivations to lean toward positive / negative responses."

To what extent do you agree with the following statement?

"I will vote for the policy A"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5



Likert is Interval ☺

*"A well-designed question (i.e. having a clear neutral position, symmetrical and equidistant categories) can **approximate an interval measurement**, with large N"*

Take-away messages:

- ✓ Due to many potential biases, it's safe to treat them as ordinals.
- ✓ If you need interval values, carefully frame the question and categories so that they are symmetrical and equidistant.

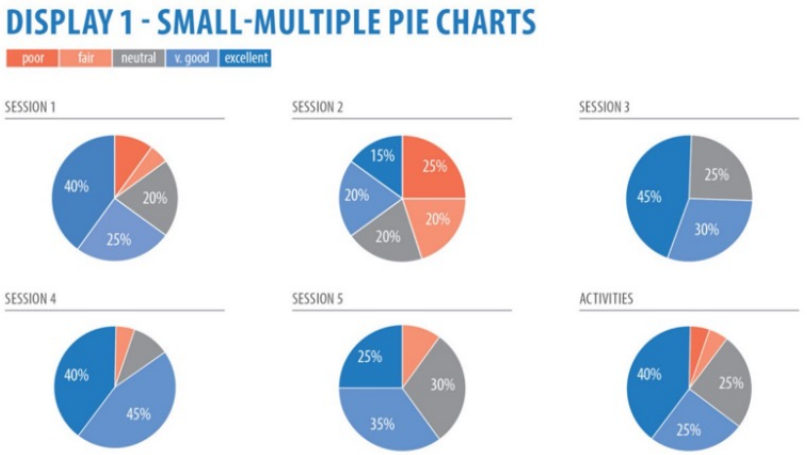
1. Amieson, Susan (2004). "Likert Scales: How to (Ab)use Them". Medical Education. 38 (12): 1217–1218.
2. Norman, Geoff (2010). "Likert scales, levels of measurement and the "laws" of statistics". Advances in Health Science Education. 15 (5): 625–632.

How to visualize survey results on Likert scale

Q. How would you rate the quality of the sessions and activities?

Responses

	Poor	Fair	Neutral	Very Good	Excellent
Session 1	2	1	4	5	8
Session 2	5	4	4	4	3
Session 3	0	0	5	6	9
Session 4	0	1	2	9	8
Session 5	0	2	6	7	5
Activities	1	1	5	5	8



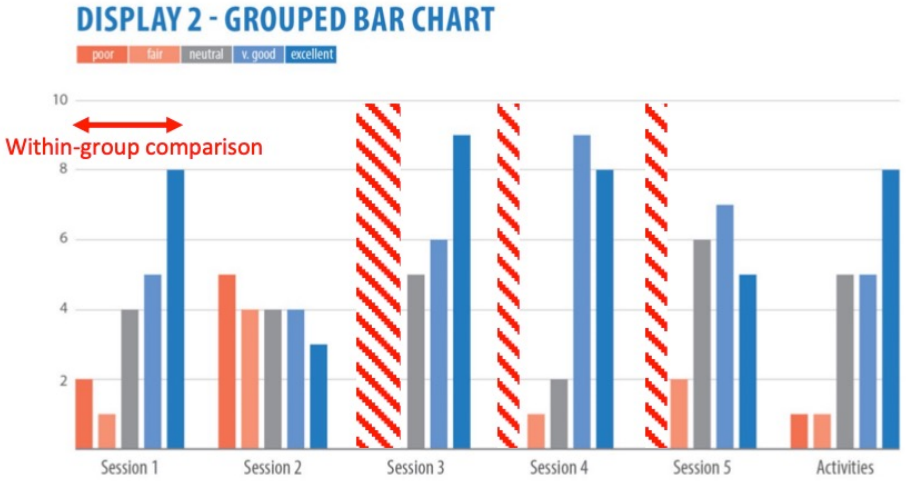
☹ **Pie charts** are bad at comparing multiple sets of ratios because 'angle' is hard to compare. Making it even worse, pies are cyclic ('Poor' and 'Excellent' are placed right next to each other)

☺ No significant benefits of using pie charts

https://medium.com/nightingale/seven-different-ways-to-display-likert-scale-data-d0c1c9a9ad59

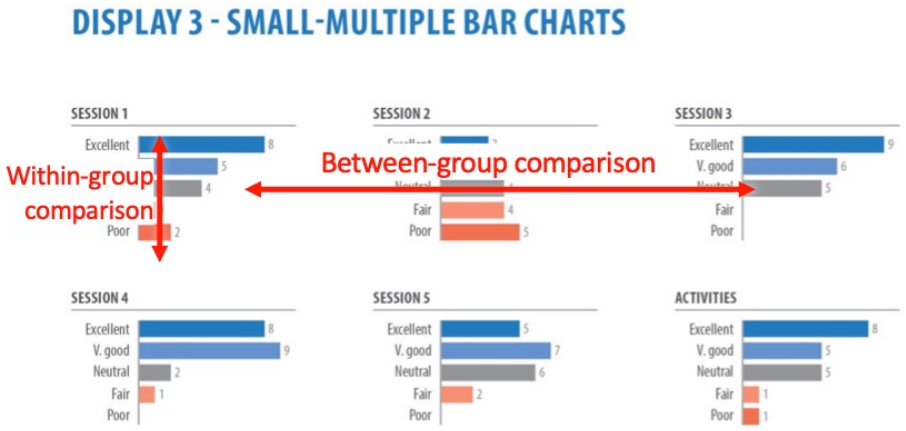
How to visualize survey results on Likert scale

How to visualize Likert values as **Ordinal**



☹ **Grouped bar charts** cannot represent categories with zero responses. Hard to compare between groups

☺ Easy to interpret distribution within a group

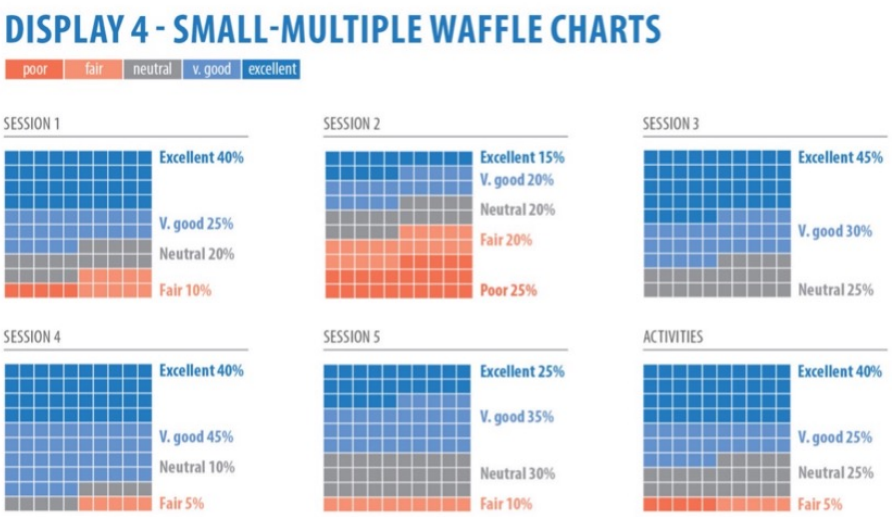


☹ Bar charts are very generic. Viewers have to interpret the meaning (“this is Part-to-Whole information of a Likert scale”) from text.

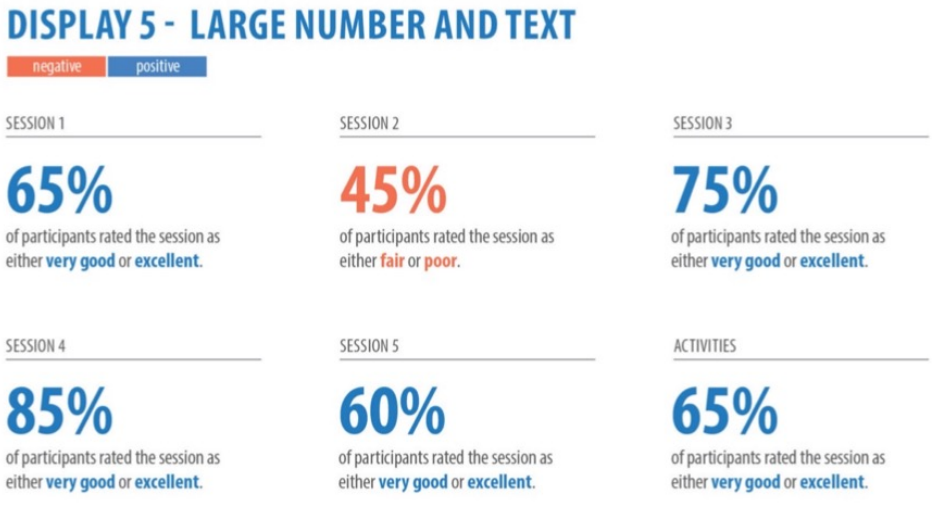
☺ Both between-group and within-group comparisons are well supported

How to visualize survey results on Likert scale

How to visualize Likert values as **Ordinal**



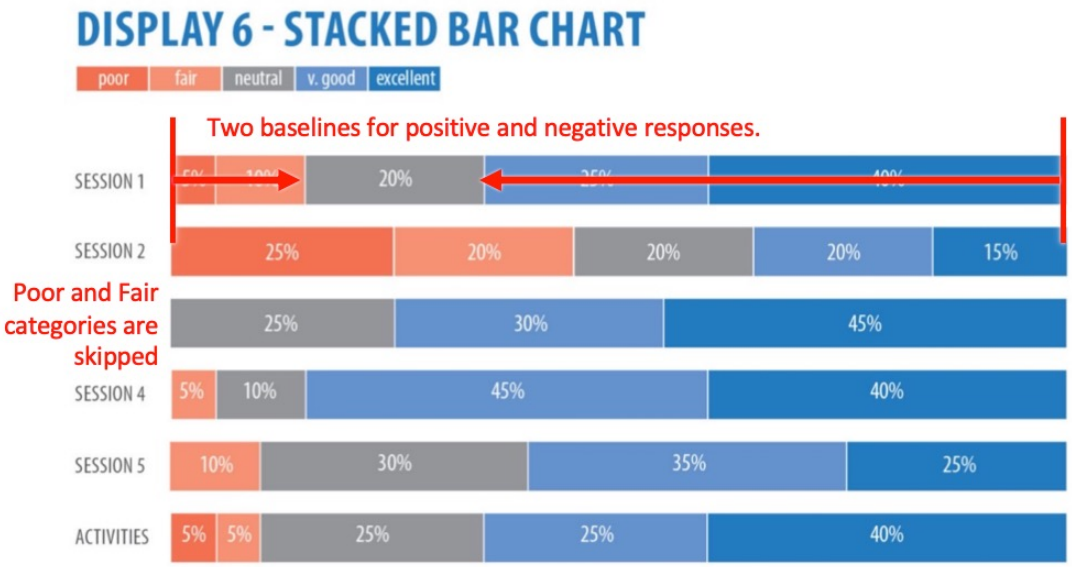
☺ Very Fancy. Each waffle is clearly part-to-whole information. (a perfect alternative of pie charts?!)
☹ Within-group or Between-group comparisons are not as easy as with bar charts; Empty categories are skipped (same with the group bar charts)



☺ Communicate specific insights with minimal cost; Useful as a summary of other charts
☹ Subjective; Huge information loss; Not allowing viewers to find their own insights

How to visualize survey results on Likert scale

How to visualize Likert values as **Ordinal**

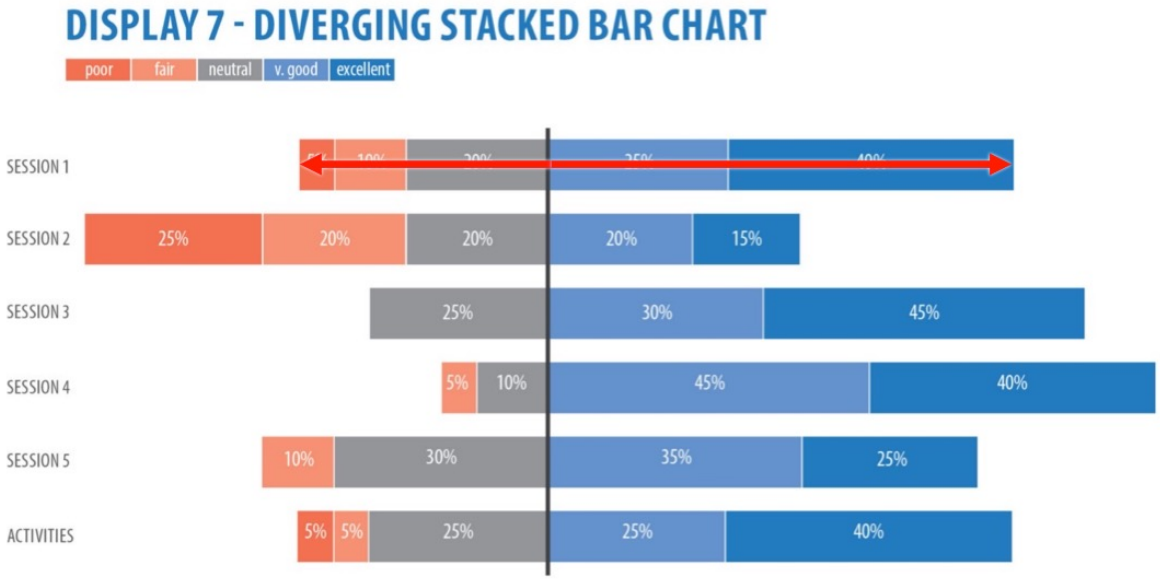


☺ Efficient use of screen real-estate (i.e. can show many sessions in small space); Both between and within-group comparison are well-supported; Since bars have the same total height, it's clear that they show part-to-whole information; Category ordering is maintained.

☹ Empty categories are skipped; There are two baselines for positive and negative categories (i.e. hard to compare).

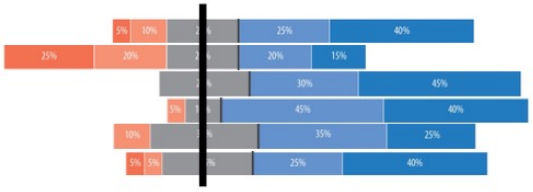
How to visualize survey results on Likert scale

How to visualize Likert values as **Ordinal**



☺ Aligning bars at the common baseline between neutral and v.good; Optimized for comparing positive / negative categories both within and between-groups

☹ Almost perfect. But why not align at the center of neutral?



Summary

- There is no silver-bullet chart. Each chart design has pros and cons.

- Bar chart variations are very versatile
  - Grouped, Stacked, and Aligned bar charts can cover most analytic tasks
  - “Being widely used” makes bar charts hard to guess its meaning (as opposite to pie charts).

- Use multiple charts in combination for products or storytelling
  - E.g. coordinated charts; dashboard; data story; infographics, etc

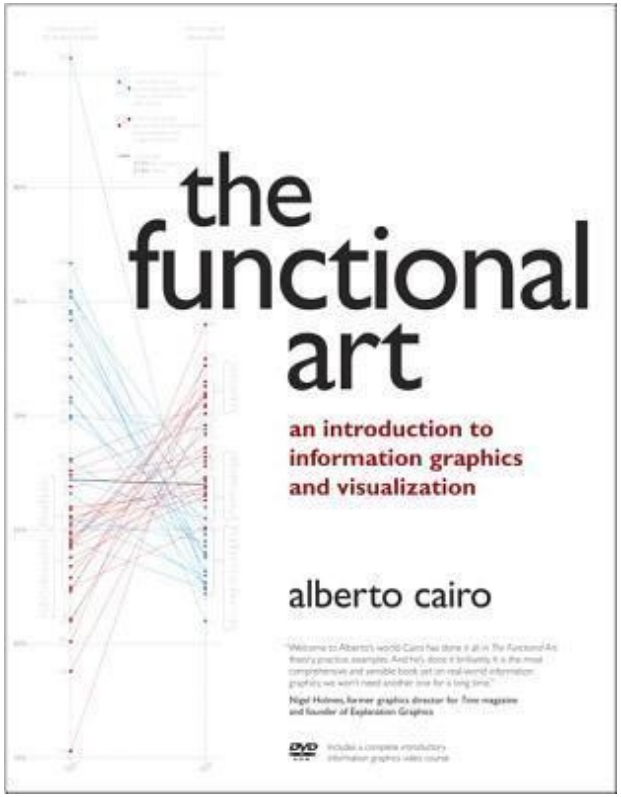
Interaction

Wan Fang

Southern University of Science and Technology

[Adapted from Kai Wong's article on Design for information exploration]

What makes a good visualization?



"A good graphic has two basic goals: It presents information and then allows users to explore that information."

DIKW and the User side gap

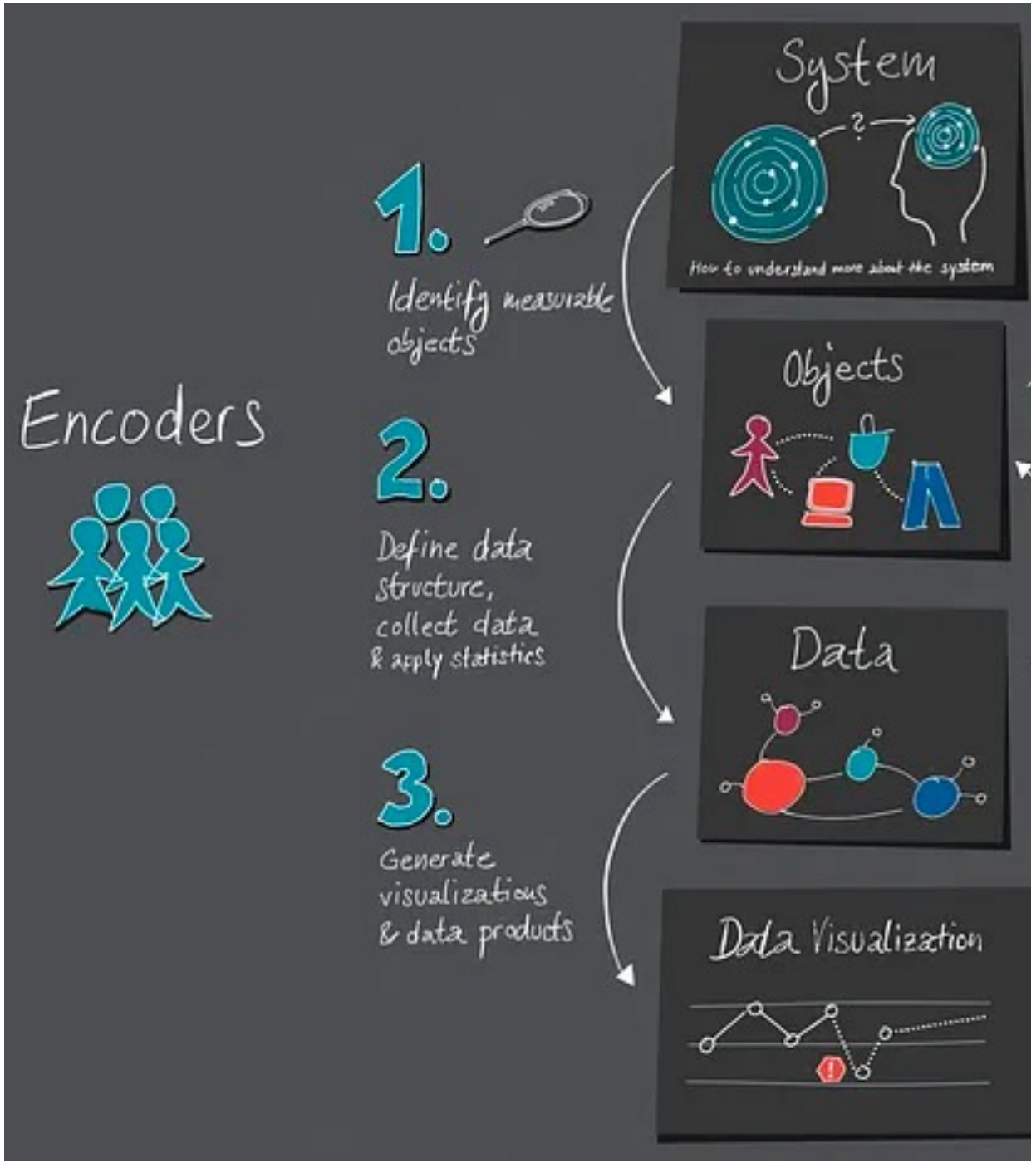
- DIKW information model

- To jump from Information to Knowledge requires an understanding of how your user understands or finds patterns in the data.

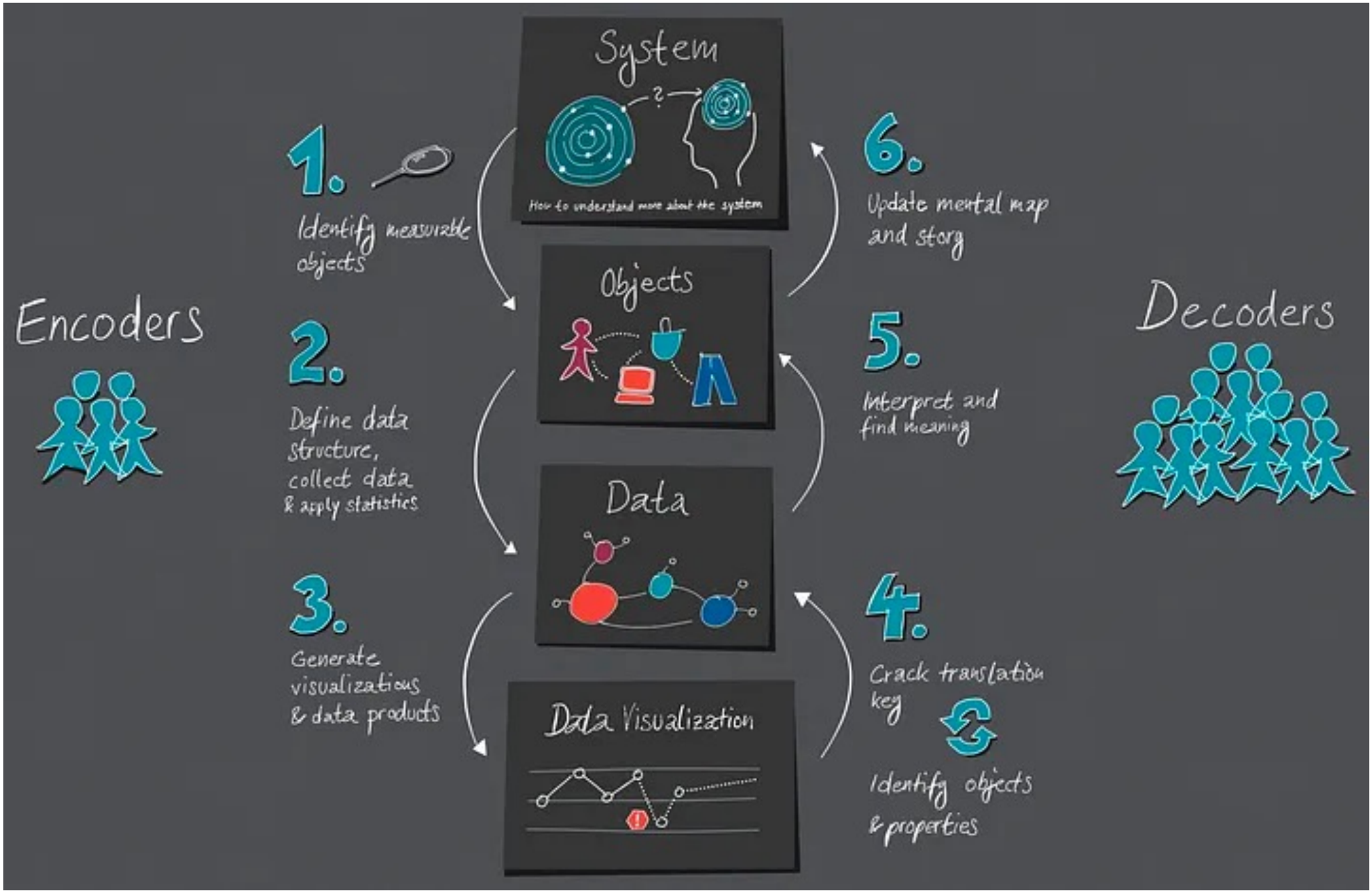
- **How data goes from the designer to the user?**

https://en.wikipedia.org/wiki/DIKW_pyramid

The Cycle of Encoding and Decoding



Interaction



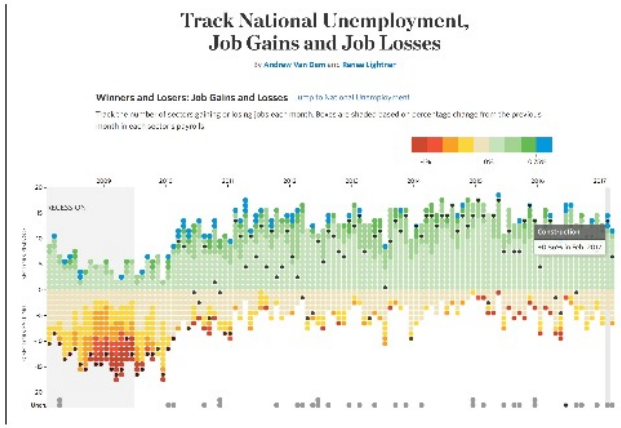
Designers encode, users decode

- Without standardized specifications, we're relying on our diverse audience to decode the data in their minds.
- How do we design visualizations that can be easily decoded?

Ben Shneiderman's information-seeking mantra

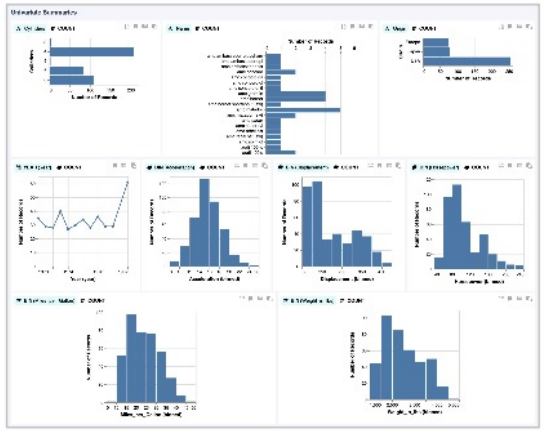
1. Overview first 2. Zoom and Filter 3. Detail-on-demand

The initial view of a **dashboard** provides the data overview



Data journalism article begins with an overview that introduces the topic and the dataset

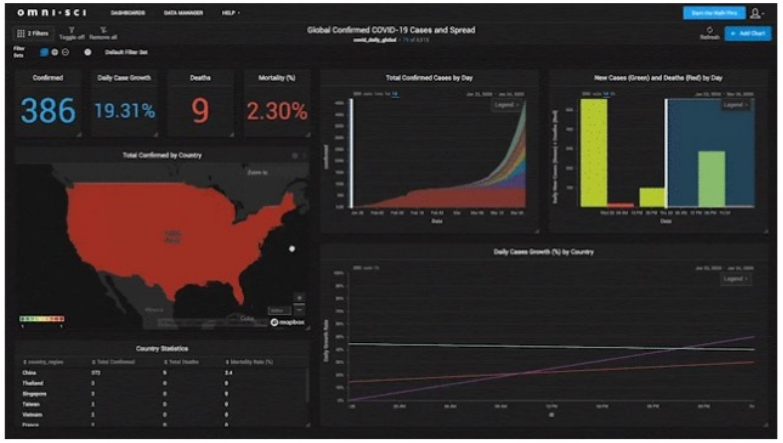
http://graphics.wsj.com/job-market-tracker/



The initial view of **an AI-supported visualization tool** (voyager) begins 1D charts per columns (i.e. **univariate** analysis)

Ben Shneiderman's information-seeking mantra

1. Overview first
2. Zoom and Filter
3. Detail-on-demand



Brushing & Linking
Multiple charts in a dashboard are connected with each other. Brushing (i.e. selecting items) in one chart filter the same items in other charts.

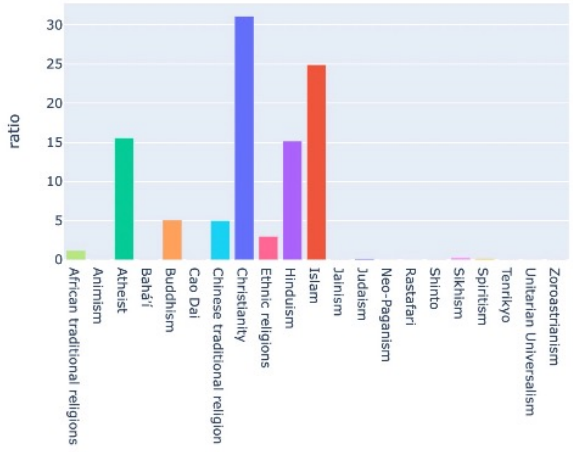
Filtering UI
Data journalism articles often provide filtering UI for readers to find an interesting part of the dataset, and go deeper.

2D, 3D, and higher-dimension charts
After finding interesting columns, analysts would combine them to narrow down the scope and to find more specific insights (i.e. **multi-variate** analysis)

Ben Shneiderman's information-seeking mantra

1. Overview first
2. Zoom and Filter
3. Detail-on-demand

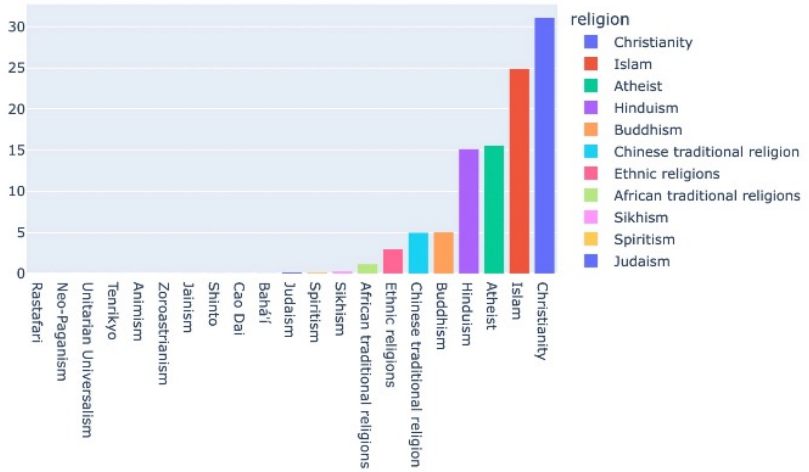
Sorted by Category (A->Z)



Switch Sorting Mechanism

Different sorting methods helps viewers focus on different patterns and items. Thus sorting is a variation of zoom and filtering process

Sorted by Ratio (small->large)

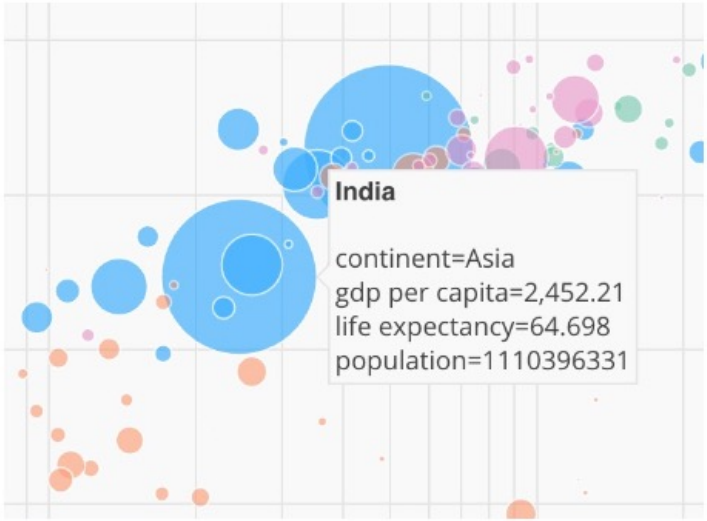


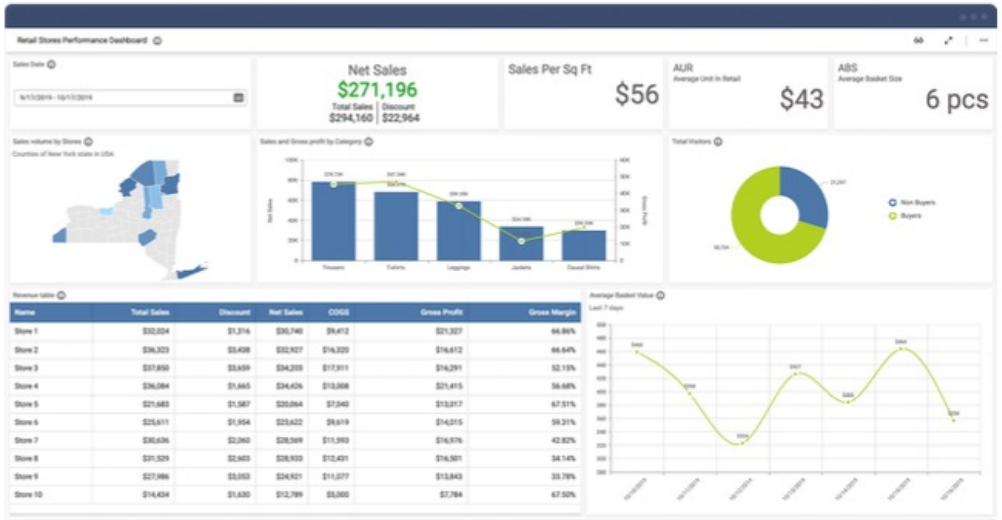
Ben Shneiderman's information-seeking mantra

1. Overview first

2. Zoom and Filter

3. Detail-on-demand





Mouse-over Tooltip

Most EDA tools allow viewers to see raw data of individual items as tool-tip

Data table filtered by user selection

Most EDA tools offer data tables that shows raw data of currently selected items

Basic Selection Methods

- Point Selection
  - Mouse Hover / Click
  - Touch / Tap
  - Select Nearby Element (e.g., Bubble Cursor)
- Region Selection
  - Rubber-band (rectangular) or Lasso (freehand)
  - Area cursors (“brushes”)
- Brushing & Linking
  - Select (“brush”) a subset of data; See selected data in other views

Dynamic Queries

- Textual Queries

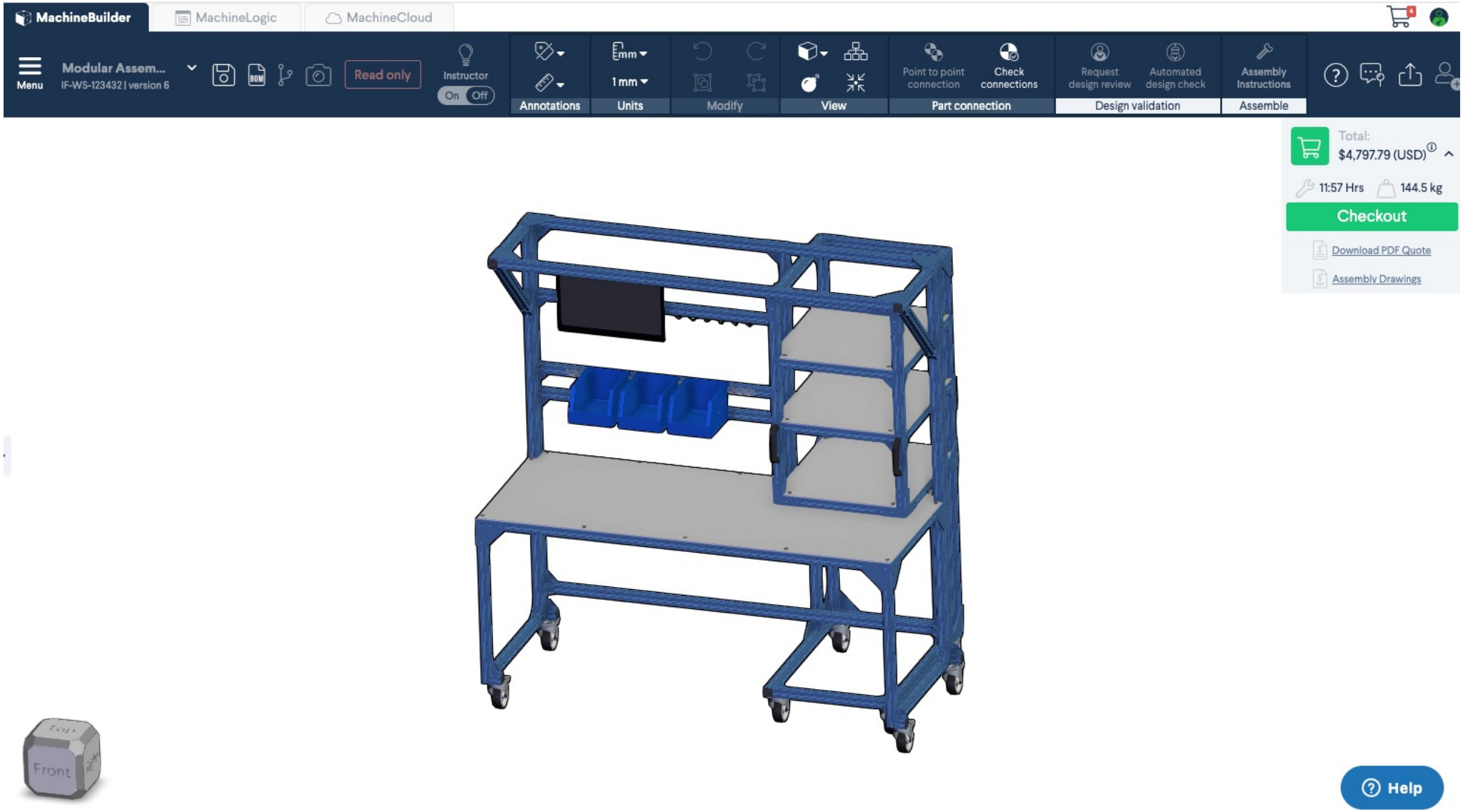
```
SELECT house FROM seattle_homes
WHERE price < 1,000,000 AND bedrooms > 2
ORDER BY price
```

Dynamic Browser : DC Home Finder

IdNumber	Dwelling	Address	City
2	House	5256 S. Capitol St.	Beltsville, MD
4	House	5536 S. Lincoln St.	Beltsville, MD
5	House	5165 Jones Street	Beltsville, MD
8	House	5007 Jones Street	Beltsville, MD
9	House	4872 Jones Street	Beltsville, MD
17	House	5408 S. Capitol St.	Beltsville, MD

- Direct Manipulation

3D Dynamic Queries



Summary of Interactivity

- Most visualizations are interactive
  - Even passive media elicit interactions
- Good visualizations are task dependent
  - Pick the right interaction technique
  - Consider the semantics of the data domain

- Taxonomy of Interactions
  - **Data and View Specification**: Filtering, Sorting
  - **View Manipulation**: Select, Navigate, Brushing & Linking, Dynamic Queries
  - **Process and Provenance**: Record, Annotate, Share, Guide

Assignment 2 on **Interactive Visualization**

- Tasks
  - Find at least three interactive visualization of the same or similar data sample, analyze, compare, and explain how they visualized the data story
  - Build a visualization that enables interactive exploration of a dataset of your own choosing
  - Present and explain the tool/graph you've used and how you visualized it

- Notes
  - Goals: gain hands-on implementation experience, & think about the effectiveness of the specific techniques for your chosen data
  - Yes, you can choose the same data sample as Assignment 1, or another one, but please note that this time we focus on the interactive visualization techniques

https://ds363.ancorasir.com/

Thank you~

Wan Fang
Southern University of Science and Technology