UBC CPSC 436V Midterm

12 Mar 2020

Closed book, no electronic devices. Cell phones must be turned off and put away. Place your photo ID face up on your desk.

One single-sided sheet of notes (8.5"x11") is allowed, keep it so that you can reuse it for the final.

Do not open the exam until told to do so. Answer the questions in the space provided. Do **NOT** write on the back side of the pages.

There are 75 points, you have 75 minutes.

Name _____

Signature _____

Student Number_____

5-digit CS userid _____

Q1: Attribute types *[8 pts]*
What type of attribute are the following? Circle only one of the three choices.

1.	Type of cheese (eg Swiss, Brie)	categorical	ordinal	quantitative
2.	Eye color (eg brown, blue)	categorical	ordinal	quantitative
3.	Class mark (eg C, A, F)	categorical	ordinal	quantitative
4.	Tire pressure (eg 60 psi, 70 psi)	categorical	ordinal	quantitative
5.	First name (eg Alice, Bob)	categorical	ordinal	quantitative
6.	ISBN numbers (eg 978-1-4665-0891-0)	categorical	ordinal	quantitative
7.	Unemployment rate (eg 6%, 3.72%)	categorical	ordinal	quantitative
8.	Starbucks drink sizes (eg venti, grande)	categorical	ordinal	quantitative

Q2: True/False *[12 pts]* Circle only one of the two choices.
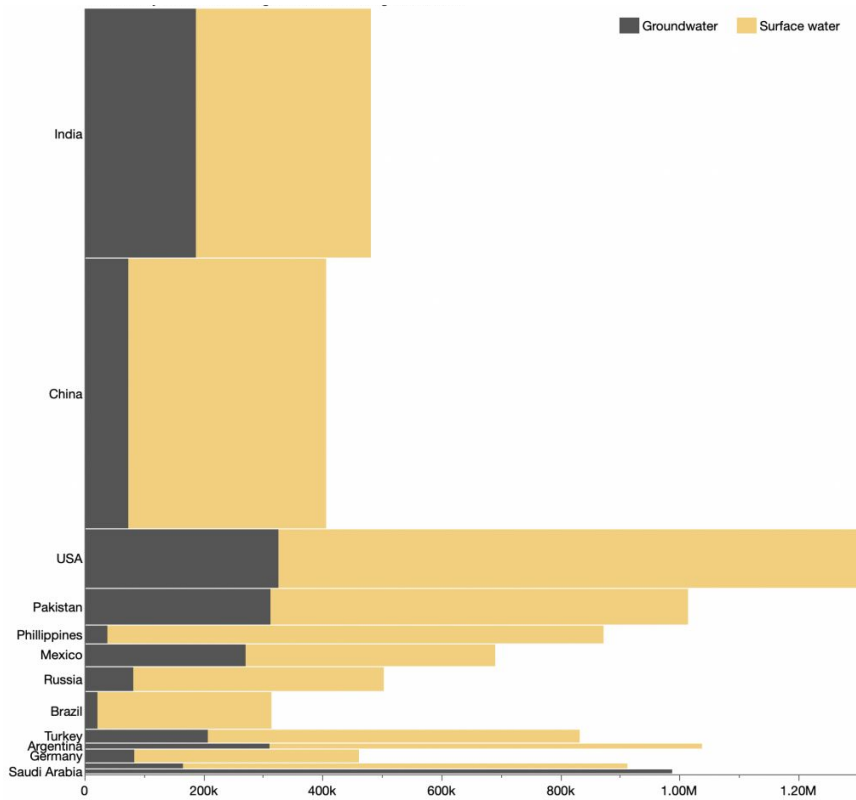
1. T F The color channels of luminance and hue have similar characteristics because they both convey magnitudes
2. T F The combination of {FirstName, LastName} is a suitable unique key in a table with 10K items representing people
3. T F Showing a distribution requires multiple attributes
4. T F A continuous colormap to show sequential quantitative attributes should have fully saturated hues on each end and a zero point in the middle with a desaturated color such as white or grey
5. T F Temporal data may have both cyclic and hierarchical structure
6. T F The {action, target} pair of compare shapes describes an appropriate task for tabular data
7. T F Scatterplots are an appropriate visual encoding for the task of showing correlation between two quantitative attributes
8. T F The strategy of deriving new data can be combined with the strategy of a single view that changes over time
9. T F Aligned area is the most accurately perceived visual channel
10. T F Roughly 3-4 bins of categorical color are discriminable when the colored regions are small and scattered
11. T F Position and color hue are fully separable channels, but size and color are not
12. T F The human perceptual system is well suited for delivering relative judgements for the hue channel and absolute judgements for the saturation and luminance channels
13. T F Line charts are well suited for the task of assessing trends
14. T F Heatmaps are compact and highly scalable but need to have their rows and columns reordered.
15. T F Radial layouts are particularly suitable for cyclic data
16. T F Choropleth maps can be misleading when there is a lot of variability in region size
17. T F Segmented rainbow colormaps are suitable for categorical data because they are highly saturated
18. T F Rainbow colormaps are suitable for spatial data and quantitative attributes in tables because they are continuous
19. T F Bivariate colormaps are often misinterpreted because 8% of men are red-green colorblind
20. T F Animated transitions can be used for constrained navigation
21. T F Partitioning data into juxtaposed small multiple views is suitable for tabular data but not spatial data
22. T F Linked highlighting provides a way to show the connections between items across multiform multiple views
23. T F Blue-orange colormaps or redundantly encoding both hue and luminance are both suitable alternatives to red-green colormaps for colorblind users.
24. T F Interactive navigation uses a camera metaphor with a changing viewpoint.

Q3: Water usage three ways *[15 pts]*

Credit: https://www.forthgo.com/blog/2019/12/18/bar-mekko-chart-study/

**A. Water usage by source.** water use (area) = population (height) x per capita use (width)

Ordered by total water usage.   Source: fao.org AQUASTAT. Dark grey - groundwater. Light grey - surface water. X-axis is liters per person per year.

B. x-axis is percent of country's usage

C.

Water usage by source and population
Ordered by total water usage. Source: fao.org AQUASTAT

| | Groundwater | Surface water | Population |
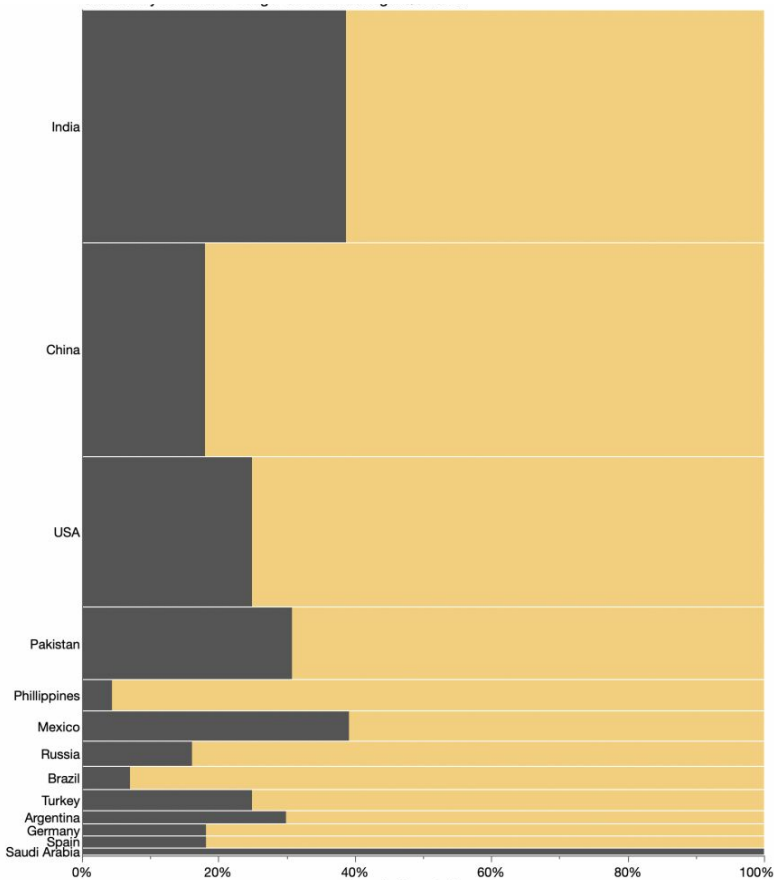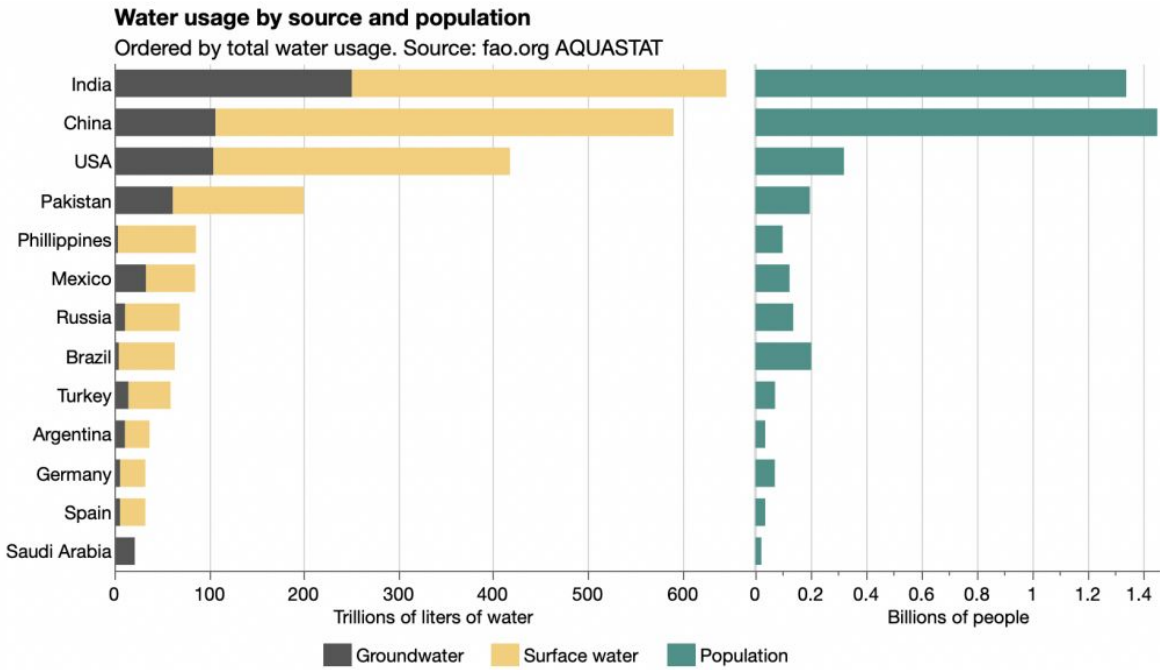
Above are three different visual encodings of the same dataset (A, B, and C). State one disadvantage for each approach, analyzing in terms of how the channels of length (horizontal), height (vertical), area (length x height of segments/bars), and color are used to encode the attributes of water source (ground vs surface water), water usage, and population.

A disadvantage:

B disadvantage:

C disadvantage:

Q4 Maps and Multiples [14 pts]

Credit: https://github.com/jwoLondon/30dayMapChallenge

For each of the two visualizations below (A, B), answer the following questions:

- What type of view coordination is used (single view, multiform, small multiples, overview/detail multiform, overview/detail same form)? Justify your choice very briefly in terms of whether the visual encoding is same or different, and whether data is shared between views or subsetted.
- For each kind of mark used, state the mark type and which channels are used to encode what attribute on that mark.

A. US State to State Migration, 2018. Each small map represents a single state's out-migrants, showing where in the US they moved. Darker grey indicates larger numbers of people moving. Black squares represent those who did not move.

US State to State Migration, 2018
Each small map represents a single state's out-migrants, showing where in the US they moved.
Darker red indicates larger numbers of people moving. Black squares represent those who did not move.
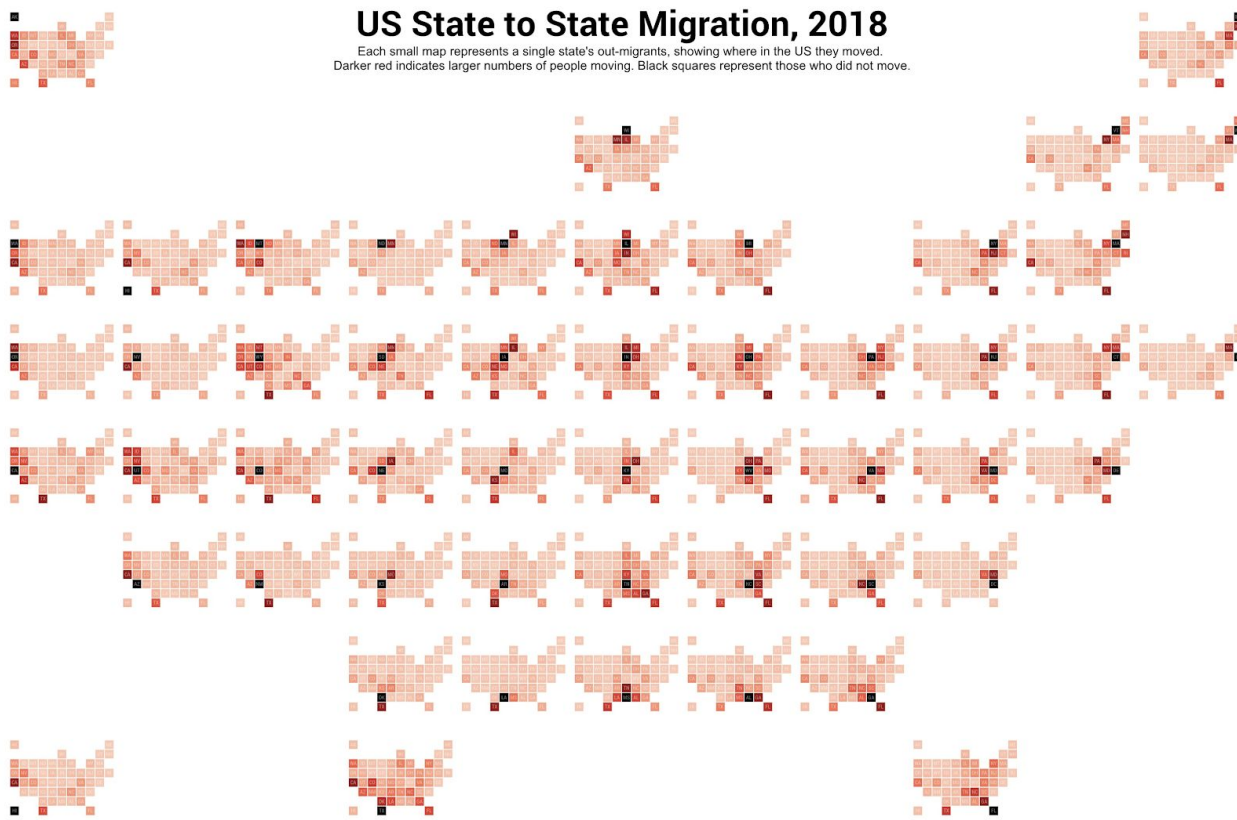
B. I'm Made Up For You. 27,346 tweets containing the phrase "x% of statistics are made up". Sized by the value of x. Content from tweets posted between 2007-2017. Geolocations are made up.
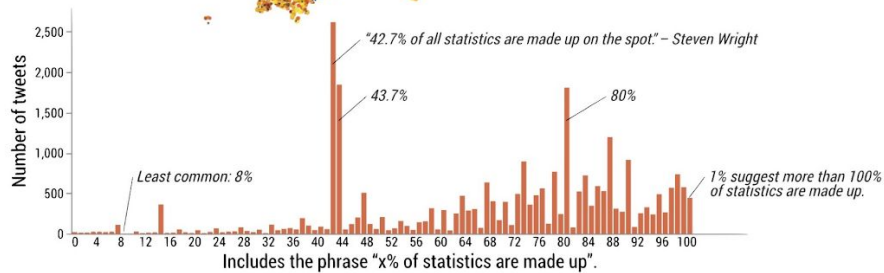
I'm Made Up For You

27,346 tweets containing the phrase "x% of statistics are made up".

Coloured and sized by the value of x.
Content from tweets posted between 2007-2017.
Geolocations are made up.

% of statistics
claimed to be
made up
- 0
- 20
- 40
- 60
- 80
- 100

"42.7% of all statistics are made up on the spot." – Steven Wright

43.7%

80%

Least common: 8%

1% suggest more than 100% of statistics are made up.

Number of tweets

2,500
2,000
1,500
1,000
500
0

0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96 100

Includes the phrase "x% of statistics are made up".

Q5: Sketching Electronic Health Record Interface *[26 pts]*

You are given an electronic health record dataset consisting of events that may have a single timestamp, or they may have both start and end times. Events also have geographic locations. Each patient may have up to 100 events associated with them. There are up to 100 patients in the dataset. The event type attribute has 5 levels: 4 main types of events (emergency, clinic, home, medication), and one catch-all 'other' category for the rest. The task is to assess whether patients who use clinics for routine health care are less likely to use emergency services, and whether this tendency is different for patients with complex medication needs vs. simple or no medication usage. Sketch two views: an overview showing all patients, and a detail view showing the data for only a single selected patient. Explain interactive functionality with words and arrows. Briefly justify your design choices with respect to alternatives.