Jongho Jung
Nandita Jha

1	2	3	Σ

Assignment 2

(Due: 14.11.2022)

Task 1 Scales and Visual Mapping

(a)
- Distance (in meters) - Quantitative (Ratio scale)
- Longitudes - Quantitative (Ratio scale)
- Measurements of duration - Quantitative (Ratio scale)
- Point scores in an exam - Quantitative (Interval scale)
- Postal codes -Ordinal Scale
- Capital cities -Nominal Scale

(b)
- **Scenario 1**
  Answer: Bar chart. Bar Charts are suitable for comparison of values as it is easier to understand the difference. So it will be better to use bar chart to compare total number of unemployed persons between are three occupations.
- **Scenario 2**
  Answer: Line Chart. Evolution of Data through time is basically a relational data. So, line charts will be suitable to represent values changing over time.

(c)
- List two visual variables used to represent the quantitative data in each chart.
  - color in both bar and pie chart
  - width in bar chart and position of each attribute value in pie chart. For example the red section is of higher percentage and while it is placed below cyan blue which is of similar area on the pie chart- the difference of which is more in volume is harder to gauge.

- Pick the chart that works best for sorting the values in increasing order. Explain your choice based on the respective visual variables.

Horizontal bar chart is best to represent values of increasing order. Pie chart is a good choice to represent part to whole values. However in this case, the red part and cyan blue part of the pie chart looks very similar so it will be best to represent the election results in horizontal bar chart.

Task 2 Lie Factor

(a) **Explain the term "lie factor" in your own words.** ANSWER: Lie factor is the ratio of perceived size of a value in a chart to the size of in actual value size. The size of effect is the difference between second value and first value divided by the first value.

(b) **Name two reasons why the examples of Figure 2 and Figure 3 are problematic from an InfoVis perspective?** ANSWER: In Figure 2, the width of each bar is quite odd and does not appear proportionate. Secondly there is a very big difference between 'people on welfare' and 'people with a full time job' and this difference will be more emphasized when represented in pie chart. In figure 2, labelling on x-axis is very crowded and the value of 'RAT' and 'PCR' on Y-Axis is difficult to follow when both colors are represented in same bar.. So, two separate line graphs in red and blue color will be better to show the daily covid numbers in the given month.

(c) **Calculate the lie factor of the bar chart in Figure 2, using the formula from the lecture. The left bar has a height of 172 px, the right bar has a height of 36 px. Hint: Use the right bar value as the first value in the lie factor equation.** ANSWER: Lie factor can be calculated as [(second value- first value)/ first value / size of actual effect in data]. According to this: [(172px- 36px)/ 36px / size of actual effect in data] = 3.77 / (108.6 M -101.7M)/101.7 M = 3.77 / 0.067 = 56.2

Task 3 Kernel Density Estimation

(a) Impl

(b)

1. Decreasing bandwidth value (h) to 0.1 will undersmooth and increasing h to 3.0 will oversmooth.