Choosing the suitable chart to present comparison about multiple-variable data

Student Name(s)   
Student Number(s)

# Proposed Hypothesis (What Question Are You Asking?)

(650 words approx.)

Explain the question you are asking in your experiment and why it is important. Also provide an overview of some interesting, important, and relevant existing work that addresses similar questions. For existing work provide appropriate citations and screenshots.

Our question: What types of charts are more suitable for showing comparison on multiple-variables data visualization

Our proposed hypothesis: when changing different type of charts in multiple- variables data visualizations, there will have the different accuracy and speed of answering the questions.

Generally, we need to compare different factors influencing on one

One of the studies on visualization multidimensional data

Furthermore, the literature does not include some studies of how the viewers understand or interpret charts presenting comparison information of multiple variables.

# Experimental Method

## Overview

(400 words approx.)

How will you run your experiment? What are the dependent and independent variables? Are there any confounding variables? How will your design take account of confounding variables? What will the experimental conditions be? What is your overall experimental design (within-group or between-group)? Why?

We made a decision on controlled experiment within group approach to test what kind of charts are the most suitable for comparising information in multiple variables data visualization. In order to make our experiment successfully, the following variables are included in our experiment:

Independent variables:

Chart types: multiple Pie plots, Box chart, multiple histograms, small multiples

Different data comparison: values, proportions and distributions

Different questions: The highest or lowest values, the range and tendency. The statements were of the following three types: (i) “The highest values in X from time1 to time2”, (ii) “The ”. For each chart type and variable combination, two versions of statement type were shown to participants: one that was true and one that was false. Meanwhile, there was an option for impossible to tell. For example, for the data shown in the first graph in Figure 1 the true statement “ the increase trend in Dublin from 2010 to 2017 ” and the false statement “increase trend in Dublin from 2010 to 2017”

Dependent variable: measuring the accuracy of answering statements and speed of answering statements

Controlled variables: lights of room; noise; capability of using computer; gender; age; the same colour for all data

Confounding variables: To be get the most accurate output as much as possible, we take into account confounding variables, which can influence measurement of independent variable. In our case the confounding variables considered are: 1) previous experience; 2) eye sights of the subject; 3) colour blindness; 4) learning effect

In order to begin to answer our questions, there were few basic questions about studying background and identified what’s number under the colour pictures. Firstly, participants need to give their background so that we are able to avoid some participants who study statistics or any related with data subjects. It would help us to avoid previous experience. Secondly, we also had some colourful pictures to test the participants colour blindness. Meanwhile, Task was presented in random order to control for learning effect.

The combination of chart types (4), different comparison data (3) and the questions (3)

## Data collection

(250 words approx.)

What data will you collect during the experiment about participants’ performance and why? Will there be a mix between subjective and objective measurements? How does the data you will collect relate directly to the questions you are asking?

We would collect data about the participants’ performance from the experiments in the following ways:

1. Accuracy to evaluate how well participants answer questions matches the ground truth.
2. Speed of answering the questions from the first visualization to the last data visualization by measuring time taken for their responses in each chart type.
3. A confusion matrix to analyze the types errors made by participants.

To compute the accuracy, we initially depend on majority voting that has been chosen by most participants. The final accuracy values we reported was that the total correct answers by match ground truth divided the total ratings that calculated by qualified participants multiple the number of chart types and the number of questions.

Objective measurements are based on how well the participants performance, such as task completion time, percentage of task completed and errors percentage and so on. On the other hand, subjective measurements are to measure how participants feel or say they actually experience.

In our experiment, using questions in the section 3 data visualization could identify how well the participants performed and using the scale information in section 2.5 practical setup about chart types to see how well they feel experience.  So, we used the mixed objective measurements and subjective measurements.

From data we collected, we could get the information about how people perceive and understand the information through use the different chart types. We could notice that changing different chart types increased or decreased their speed of answering questions and accuracy. Meanwhile, we can see from the confusion matrix that which chart types had the lowest error made by participants.

According to analyze the data collected, this experiment helps in proving our proposed hypothesis by recording speed of answering questions and the accuracy that the answered correctly.

## Selected subjects

(200 words approx.)

Who will you use as subjects in your experiment. Why are these a representative sample? How will you source these subjects?

## Data analysis

(200 words approx.)

How will you analyse the data you collect during the experiment (only very basic details required)? How will this analysis answer the question you originally proposed?

Firstly, data analysis gives us support on establishing the relationship between charts and their  effectiveness by offering us how changing charts type would generate different answers on the same questions as well as the speed of providing the response as we recorded the time. Secondly, the comparison of answers with the ground truth in different visualization can

## Practical setup

(250 words approx.)

In practical terms how will you run your experiment? Will it be online or offline? What instructions will participants be given? What type of room will the experiment be in? Will visualisations be displayed on-screen or on printed paper?

# Data Visualisations

Provide examples of the images you will use in your experiment. Provide visualisations for as many experimental conditions as possible.

# References

Provide complete bibliographic detail for any cited work.