UNIVERSITY OF COLORADO BOULDER

Extra Credit Assignment

INFO 4602

# Project 4: Experimentation

Yonghao Zhang

05/04/2018

## 1 Part I: Designing an Experiment

1. **Research Question**

   I want to find out which kind of distribution chart is the best visualization for uncertainty.

2. **Hypotheses**

   From the information visualization class, I've got touch with scatter plot, box plot, and violin plot for the showing of uncertainty data. I just found there is new method to visualize uncertainty called Beeswarm Plot. I'm really curious which visualization would be the best for people. My hypothesis would be Beeswarm or violin plot because both of them could show not only the range but also the distribution of data.

3. **Independent Variables**

   I will use same data to compare four visualization methods, including scatter plot, box plot, violin plot and Beeswarm plot.

4. **Dependent Variables/Measures**

   I will ask participants which one is the best visualization from the aspects of distribution, appearance, outlier, Five-number summary and in general.

5. **Control Variables**

   As I mentioned, I will compare four visualizations for the same data. Besides, I want to avoid participants from INFO 4602 class because they may already have exposed to the standard solution for this question.

6. **Description of the Stimuli**

   The data I will use is from the *Andrew Sielen's Block*[1].

7. **Experimental Procedure**

   At the first place, I will use *Google Form*[2] to take a survey. I will show them four visualizations and ask them questions about five dependent variables. All participants need to do is five multiple-choice questions. Then I can use the result to do some analyzes and try to figure out the research question.

8. **Planned Analysis**

   I will analyze the data and characterized the best visualization from all the aspects. There could be more than one best visualizations because they all have meaning to be existed.

## 2 Part II: Building the Apparatus

There are five questions in my *Google Form*[3] based on four graphs:

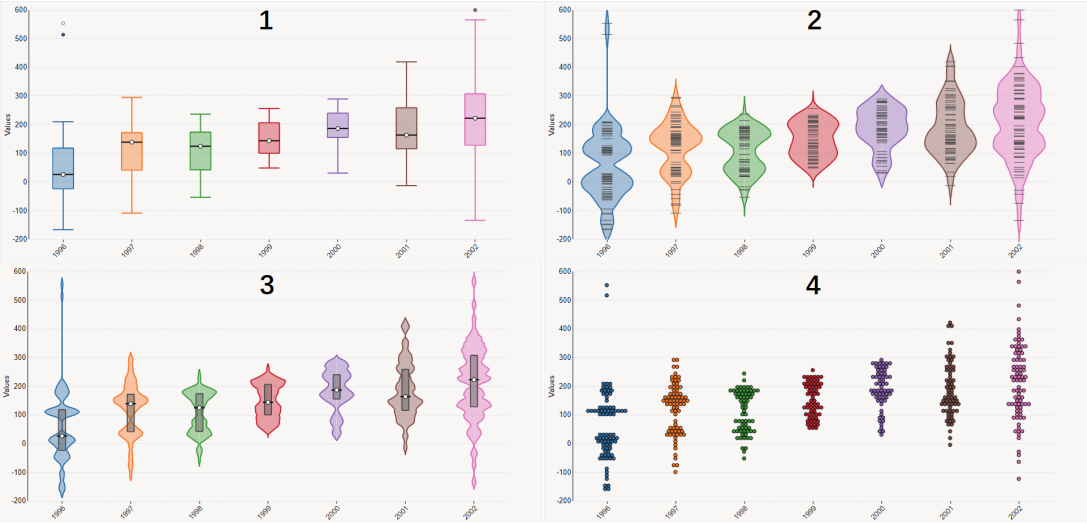


Figure 1: Four visualizations for the same data

1.Which graph below is clearest to see Five-number summary? (min, Q1, median, Q3, max)

2.Which graph below is clearest to see outliers? (ex. data not in majority)

3.Which graph below is clearest to see data distribution?

[1] http://bl.ocks.org/asielen/92929960988a8935d907e39e60ea8417
[2] https://docs.google.com/forms/d/e/1FAIpQLSc9SMk90wl8CQrGmU3frIKZMdLAIOepJK9bPbCZj2Xk97cSxg/viewform?usp=sf_link
[3] https://docs.google.com/forms/d/e/1FAIpQLSc9SMk90wl8CQrGmU3frIKZMdLAIOepJK9bPbCZj2Xk97cSxg/viewform?usp=sf_link

4.Which graph below has the most good-looking?

5.Overall, which graph do think is the best? (Information presenting, Appearance, etc)

## 3 Part III: Conducting the Study

At the last time, there were only seven participants took the survey. I'm regretting to avoid classmates in INFO class now. But it make sense because everyone is busy in final week.

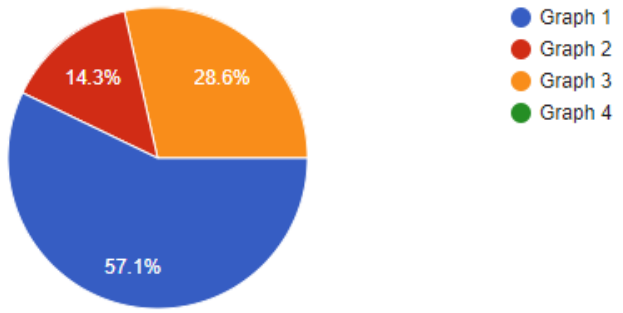


Figure 2: Which graph below is clearest to see Five-number summary?

Figure 2 shows box plot is the best to read five-number summary,with a support rate of 57.1 percent. A small number of people think violin plot and bean plot is good with percentage of 28.6 and 14.3 separately.

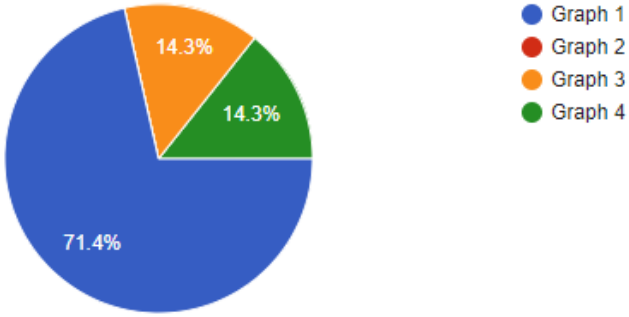


Figure 3: Which graph below is clearest to see outliers?

The evidence from Figure 3 shows that the clearest graph to see outliers is still box plot, the second clearest were violin plot and beeswarm plot, with the same favorability rating of 14.3 percent.

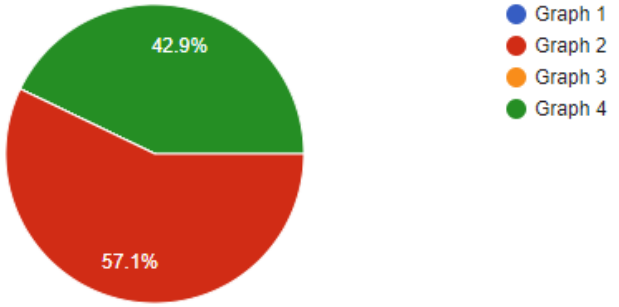


Figure 4: Which graph below is clearest to see data distribution?

It's clear from the information given in Figure 4 that the best graph to see data distribution is the bean plot, and violin plot is only one vote behind it.

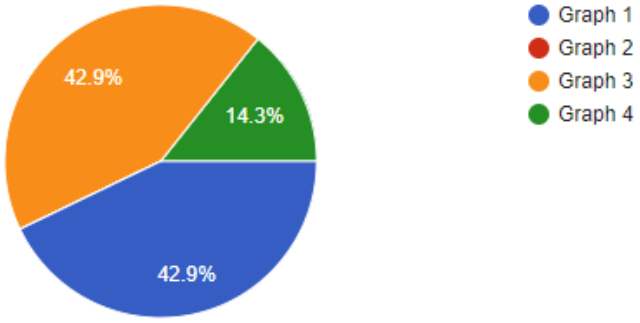


Figure 5: Which graph below has the most good-looking?

As shown in Figure 5, the majority of people answering the survey evenly split to box plot and violin plot with same approval rate of 42.9 percent. Unfortunately, there is only one person think beeswarm plot has a good looking.

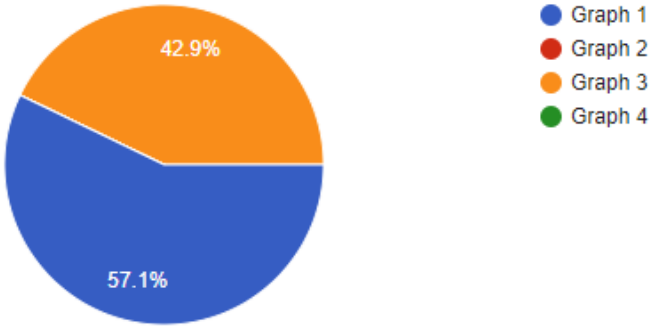


Figure 6: Overall, which graph do think is the best?

From the information shown in Figure 6[1], people only believe box plot and violin plot is good overall. My hypothesis, beeswarm plot, is defeated by this pie chart once and for all.

# 4 Part IV: Inferential Analysis

As the what I mentioned in Planned Analysis part, all these graphs have a reason to be existed, so I will categorized these four visualizations in different aspect by their performance. From the perspective of inferential statistics, most people in the population think box plot is the best graph to demonstrate minimum, Q1, medium, Q3, maximum, and outliers. Besides, with the shared first place with violin plot in the aspect of appearance, box plot is the best plot to represent uncertainty overall. However, there is no space for box plot in the field of distribution. The violin plot, learned from class, its width of the colored region corresponds to the probability density function of t-distribution. In addition, the bean plot also shows the smoothed density of points and it is more complete way to represent a distribution than more summarized lots such as box plot. Therefore, box plot is the best for representing the statistic information and bean chart is the best for represent the data distribution.

# References

[1] A. Sielen, "Violin polt + box plot v3," 2018.