An Empirical Study of Visualisation of Data

Katie Murphy

1632254

March 8, 2021

0.1 Introduction

- What is visualisation?
- Understanding theory of visualisation and implementation in Python, R and D3.
- Types of visualisation (ie. static, dynamic and interactive)

Principles of data vis:

1. The audience (who?)
2. The message or questions (what?)
3. The circumstances under which the audience interacts with the data (smartphone/laptop etc) (how?)

0.2 Literature reviews

Look at advances in visualisation practices based on when papers were written.
(ie. in an older paper they preferred this, but in a more recent paper they preferred this)
Is this down to the author's personal preference or do multiple papers back this up?

The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations

0.3 Implementation of Visualisation

- In Plotly, ggplot and D3
- Ease of use
- Consider data analysis tools available
- Interactivity of visualisations (pros/cons) (using Shiny, plotly and D3)
- Reproducibility of visualisations (ie. use of out of software editing such as word or powerpoint to change labels etc)
- Publication ready output?
- Concept of storytelling

0.4 Respondent Study

- How the human mind interprets visuals
- Focus groups and/or survey
- Replicate research
- Intuitive visualisations

Chapter 1

Data collection

1.1 Background on survey design

1.2 Specific goals of survey for this chapter

1.3 The survey

Cohorts

- Students
- People who work in data visualisation and/or statistical programming
- General population

Information about participants

- Please select your age category. (Under 18, 18-25, 26-35, 36-45, Above 45)
- If you are a student or past university graduate please specify your area of study. (maybe drop down box: Science, Technology, Engineering, Mathematics, Arts, Social Sciences, Humanities, Business, Other, N/A)
- How would you rate your spatial awareness skills? (Scale/slider: Poor, Okay, Good, Very Good)
- How would you rate your observation skills? (Scale/slider: Poor, Okay, Good, Very Good)
- How would you rate your mathematics skills? (Scale/slider: Poor, Okay, Good, Very Good)
- Do you consider yourself to be colourblind? (Binary checkbox: Y/N)

Type of questions

- Take categorical data and show different types of plot with no numbers and as questions such as "which would you say is the largest" etc - Pie chart, bar chart, pictogram.

1.4 Conclusion

Chapter 2

Univariate Analysis

This chapter will discuss basic univariate analysis and summary statistics from the survey results, alongside what could be inferred from these. We will look at each section individually and perform multiple initial comparisons whereby we subset for various factors, such as degree subject, the language used to make the plots, and the order in which plots have been presented.

2.1 American Ninja Warrior, Part 1 - Y-Scaling

This section was testing to see whether altering the type of y-scaling used would affect perception of the data, in terms of both gauging exact values as well as giving subjecting opinions on differences between values. For each of the three plots we ask the following four questions:

- *Q1 - "Approximately many times would you say the 'Salmon Ladder' was used?"*
- *Q2 - "Approximately how much more than 'Log Grip' would you say 'Salmon Ladder' was used?"*
- *Q3 - "Approximately how much more than 'Quintuple Steps' would you say 'Salmon Ladder was used?"*
- *Q4 - "In your opinion, approximately how many times would you say 'Log Grip' was used, as a percentage of the number of times 'Salmon Ladder' was used?"*

First we will look at the summary statistics for of the four questions laid out above for the whole population, before looking into whether perceptions differ depending on the language used as well as how the order in which the three plots were observed can cause any discrepancies.

2.1.1 Whole Population (N=70)

Below we see a table of summary statistics for the three plot types; the control plot, the

Q1 - Approximately many times would you say the 'Salmon Ladder' was used?

```
## [1] n = 70
```

```
##    con_1_all       log_1_all           trn_1_all
##  Min.   :40.00   Min.   :9.000e+00   Min.   :40.00
##  1st Qu.:41.00   1st Qu.:3.000e+01   1st Qu.:41.00
##  Median :41.00   Median :3.500e+01   Median :41.00
##  Mean   :41.21   Mean   :1.493e+13   Mean   :41.35
##  3rd Qu.:42.00   3rd Qu.:4.000e+01   3rd Qu.:42.00
##  Max.   :45.00   Max.   :1.000e+15   Max.   :45.00
##                  NA's   :3
```

NAs

```
##      index value
## [1,] "23"  "Don't know"
## [2,] "25"  "Next to none."
## [3,] "68"  NA
```

```
##                                    uni sp_aware obs_skl num_skl cblind vis_p
## 101                         Technology        4       4       3     No
## 121                               <NA>        4       3       3     No
## 105 Sustainability/geological science        3       4       3     No    AD
```

Q2 - Approximately how much more than 'Log Grip' would you say 'Salmon Ladder' was was used?

```
## [1] n = 70
```

```
##    con_2_all       log_2_all       trn_2_all
##  Min.   :3.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.250   1st Qu.:2.250   1st Qu.:5.000
##  Median :5.000   Median :3.500   Median :6.000
##  Mean   :5.357   Mean   :3.671   Mean   :5.871
##  3rd Qu.:6.000   3rd Qu.:5.000   3rd Qu.:7.000
##  Max.   :7.000   Max.   :7.000   Max.   :7.000
```

Q3 - Approximately how much more than 'Quintuple Steps' would you say 'Salmon Ladder' was used?

```
## [1] n = 70
```

```
##    con_3_all       log_3_all       trn_3_all
##  Min.   :3.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.250   1st Qu.:2.250   1st Qu.:5.000
##  Median :5.000   Median :3.500   Median :6.000
##  Mean   :5.357   Mean   :3.671   Mean   :5.871
##  3rd Qu.:6.000   3rd Qu.:5.000   3rd Qu.:7.000
##  Max.   :7.000   Max.   :7.000   Max.   :7.000
```

Q4 - In your opinion, approximately how many times would you say 'Log Grip' was used, as a percentage of the number of times 'Salmon Ladder' was used?

```
## [1] n = 70
```

```
##    con_4_all       log_4_all       trn_4_all
##  Min.   : 5.00   Min.   : 0.10   Min.   :  1.00
##  1st Qu.:50.00   1st Qu.: 0.50   1st Qu.: 14.38
##  Median :50.00   Median : 0.75   Median : 50.00
##  Mean   :47.66   Mean   :12.86   Mean   : 39.81
##  3rd Qu.:50.00   3rd Qu.: 0.90   3rd Qu.: 50.00
##  Max.   :75.00   Max.   :90.00   Max.   :100.00
##  NA's   :3       NA's   :4       NA's   :2
```

NAs

```
##      index con_4_all log_4_all trn_4_all
## [1,]    11        NA        NA        NA
## [2,]    48        NA        NA        48
## [3,]    60        NA        NA        NA
## [4,]    68        50        NA        50
```

```
##                                    uni sp_aware obs_skl num_skl cblind vis_p
## 12                             Science        2       3       2     No
## 17                         Engineering        4       4       4     No     Y
## 25                           Geography        4       4       4   <NA>
## 105 Sustainability/geological science        3       4       3     No    AD
```