An Empirical Study of Data Visualisation

Katie Murphy - 1632254

Contents

Chapter 1

Survey

chapter{Data collection}

1.1 Background on survey design

TODO 1 cite: https://books.google.co.uk/books?id=mSRTDwAAQBAJ&lpg=PP1&ots=nynKGurXVG&dq=how%20to%20write%20a%20survey&lr&pg=PP1#v=onepage&q&f=false TODO 2 cite: https://www.idsurvey.com/en/advantages-and-disadvantages-of-phone-survey/ TODO 3 cite: https://books.google.co.uk/books?hl=en&lr=&id=ctow8zWdyFgC&oi=fnd&pg=PR15&dq=survey+methodology&ots=fgcIbBXkS9&sig=LvdOfWOBGrtHFrcVbdfi1ahV970#v=onepage&q=survey%20methodology&f=true TODO 4 cite: https://en.wikipedia.org/wiki/Census TODO 5 cite: https://www.understandingsociety.ac.uk/ TODO 6 cite: https://ora.ox.ac.uk/objects/uuid:526114c2-8266-4dee-b663-351119249fd5 TODO 7 cite: https://www.nature.com/articles/bdj.2008.149 TODO 8 cite: https://www.collinsdictionary.com/dictionary/english/systematic

As explained in [CITE 3 HERE], a survey is a means of obtaining quantitative information regarding opinions and experiences of the respondents in order to explore the views of the target population as a whole. In this book, a survey is noted as a "systematic" method of collecting data, where the author states that the word "systematic" is deliberately used in order to separate surveys from other methods of information collection. The word

"systematic" is defined by the Collins English Dictionary as something that *"is done according to a fixed plan, in a thorough and efficient way"* [CITE 8 HERE], and this reflects that a survey is highly structured, with each participant being asked to answer standardised set of questions with no deviation from the 'fixed plan' of the survey. However, while having a 'fixed plan' is a key defining factor of a survey, whether or not the survey is 'thorough' and 'efficient' depends heavily on the survey structure and design. Designing an effective, systematic survey involves the task of balancing being thorough alongside being efficient, and this will be discussed in more depth later. A survey of a systematic nature as described here has the ability to yield a large volume of easily comparable results as every participant has been asked the same exact questions.

This is in contrast to the more free-form structure of focus groups and unstructured or semi-structured interviews. These are less systematic as they involve back-and-forth discussion between the participants and interviewer, which is discussed in

can be more difficult to analyse statistically, and are better suited for gaining general insights rather than to undergo rigorous statistical testing.

As mentioned above, survey methodology involves asking a sample of the target population a series of standardised questions, which may be presented in mediums such as written questionaires or structured interviews. Depending on the aims of the study, there will be many benefits and downfalls of each method. There may also be times when a combined approach is required to gather the necessary information.

Phone calls and other forms of interview-based survey allow the interviewer to form a personal connection with the survey participant, which can be especially helpful for a company's image if the interviewer is particularly professional or charismatic. Additionally, while the interviewer will still be limited to asking the pre-set questions, the format of such a survey can be considered semi-structured and with much more room for interpretation. This can lend itself to gaining additional insights that may not have otherwise been gathered from a more closed-form paper or online survey. Additionally, the more open format can negate any error as a result of participants misinterpreting questions due to the interviewer's ability to immediately clarify on any misunderstandings. This type of survey also provides an instant response, which is beneficial if there is only a short time frame available in which to gather

information.

However, there are also shortfalls to an interview-based survey method. For instance, although a charismatic interviewer can positively impact the image of whoever is conducting the survey, this could also lead to biases, such as the respondent answering in a way they feel will please the interviwer. Additionally, the image of the organisation could potentially be tainted if the interviwer appears rude or unprofessional, alongside potentially providing bias in the opposite direction. As well as this, telephone surveys are likely to be interpreted as a telemarketing scheme, and thus potentially have a negative impact on the number of willing respondents. The reduced anonymity of this type of survey may also create bias in the way of participants avoiding making statements that could be deemed socially unacceptable, or that they feel they may be judged for, and therefore may not provide answers accurate to their true line of thought.

A second way of presenting survey questions, and the way that will be used in this study, is a self-completed written questionaire. A questionaire may either consist of physical paper forms that are mailed or handed out to people within the target population, or may exist in an online format. As discussed in [CITE 1 HERE], this form of surveying

A questionaire is composed of two types of questions; closed-ended questions, where participants select answers from a given set, and open-ended questions, in which the participants are able to provide free-form answers.

This method of surveying is also often anonymous, which can mitigate the errors caused by participants fearing being judged on their responses. Closed-ended questions are very good for obtaining quantitative data as the answers may be easily categorised and counted. Open-ended questions are generally used to

The UK Household Longitudinal Study is an ongoing study and an example of implementation of a combined use of the above mentioned surveying methods. Initially, in 'wave 1' of the study, a sample of 40,000 households in the UK were selected to be surveyed on a yearly basis. The survey involves all members of each selected household, overall comprising of around 100,000 individials, and asks them a wide range of questions %TODO Expand on this

1.2 Specific goals of survey tool for this study

While visualisations can be a very useful tool for understanding data, they also have the potential to be highly misleading. This section of the study will explore how modifying certain aesthetic features of visualisations can impact perception and interpretation of data, and how these modifications can be exploited in order to mislead the observer. Misleading visualisations may be created in an effort to deliberately influence the viewers' perceptions, or accidentally as a result of poor practice and knowledge surrounding data visualisation. In either case, visualisations have the ability to communicate different messages and stories depending on how they present the data to the observer. There is a large amount of research and literature surrounding this topic, both in terms of providing frameworks for good visualisation practice as well as looking into how various techniques are used to decieve viewers. Results from some of these papers will be replicated, as well as used to form hypotheses which this survey will investigate.

One feature that will be tested is scaling of axes. This is widely regarded in literature to be a commonly used tactic for misleading the observer.

A large amount of the literature exploring misleading tactics in data visualisation focusses mainly on bar plots and line plots for categorical and time series data, and so this is what the survey will also focus on. There is also a large amount of literature surrounding the use of three dimensional vualisations and pie charts, which will not be explored here however will be discussed in later sections.

In particular, the results formulated in [CITE: https://dl.acm.org/doi/pdf/10.1145/3380851.3416762] will be explored. Similarly to our study, the paper uses a survey to explore how deceptive visualisation techniques can be employed as well as their impact on perception of the data. The survey discussed in this paper presents the participant with four plots; a bar plot, a line plot, a pie chart and a bubble plot,

The first example shows how truncating the y-axis of a bar plot can overexaggerate differences in the heights of the bars, perhaps leading to incorrect observations regarding comparisons of values within the data. A first hypothesis that will be tested in this survey is that truncating the y-axis does indeed impact how the data is interpreted.

For example, if the data consists of values, such as profits, at different time points, then truncating the y-axis can at first make the profit increase each year seem much greater than it actually is. Only when one reads the scale on the axis will it be observed that there may only be a very marginal, and potentially almost negligible, difference. Of course if someone were to read the y-axis scaling they would determine that the change was in fact small, but the human brain thrives on visual and pictorial feedback, and thus it is likely that a person would look at this plot and simply draw conclusions based on the plot itself without reference to the truncated axis. Additionally, the use of a logarithmic scale will be investigated, and it will be hypothesised that this will also impact interpretation.

1.3 Survey Design

Although the content of the surveys for this study is not likely to be controversial or highly personal, anonymity is still important as the participants could otherwise potentially feel pressure to give a 'correct' answer, given the mathematical nature of the questions. Anonymity here means that this pressure is potentially reduced and thus the relevant measurement bias may be mitigated. The questionaire will also contain

The publication [CITE: https://dl.acm.org/doi/pdf/10.1145/3380851.3416762] was used as inspiration for survey design.

The decision to focus on the two plot types was made in part to follow the existing literature, but also to ensure the survey is not too long. As discussed in [CITE: https://ojs.ub.uni-konstanz.de/srm/article/view/7145], too long a survey can result in higher measurement error due to factors such as waning interest or mental fatigue of respondents, resulting in careless responding and non-response. This is also further explored in [CITE: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=3059&context=etd_all].
While [CITE: https://ojs.ub.uni-konstanz.de/srm/article/view/7145] does conclude that a 'split survey' design, where each respondent is only asked to answer a selection of questions from the whole set, is effective in reducing error while gathering large amount of information, this will not be employed here. The reasoning for this is that there will already be a large set of different surveys being sent, and creating further splits could potentially lead to small sample sizes and thus inconclusive results. Additionally to

this, the paper investigates how placement of questions in the survey can affect responses, concluding that questions asked later in the survey are more susceptible to bias.

%%%%% TODO:

% Perform meta-analysis of different sources to find optimal number of survey questions? % “20 minute rule” -> Don’t go beyond 20 mins % Work out how long each question will take... Guessing 1-2 mins per question => ~10-20 questions? % Long surveys can increase measurement error %

%%%%% Potential sources about survey length %%%%% % https://journals.sagepub.com/doi/pdf/10.2501/IJMR-2017-039 -> Study about ideal length of survey in terms of time % https://ojs.ub.uni-konstanz.de/srm/article/view/7145 -> Study about reducing measurement error due to survey length % https://corescholar.libraries.wright.edu/etd_all/1918/ -> About ‘careless responding’ and ‘insufficient effort responding’ % https://journals.sagepub.com/doi/abs/10.1177/089443930101900202

Two types of questionaire will be written for this study. The first type will present a series of fairly basic visualisations accompanied by questions created to guage whether each presented visualisation has an impact on how the data is interpreted by the survey participant.

The set will consist of three separate surveys, which will be identical up to the visualisation package used. Particularly, one will contain visualisations made with R’s ggplot2, the next with matplotlib from Python, and the last with the JavaScript library D3. These surveys will be distributed to the general public by sharing links on social media platforms such as Facebook. The reasoning behind creating three separate surveys in a variety of languages is to ascertain whether the language used influences the interpretation.

%TODO explain the second type of survey further The second type will be a single survey comparing the implementation of the different languages and will be distributed to a group specialising in visual analytics for pharmaceutical research. This second survey aims to explore opinions on the coding languages themselves, in terms of features such as readability, reproducibility and ease of implementation.

%TODO explain that I’m using questionaires and why

In the set of three surveys we aim to answer questions such as:

- In which form is the given information of interest most accurately interpreted by the viewer?
- What factors of a plot can bias interpretation?
- Does the visualisation tool used have an impact on interpretation?
- Does the tool used have an impact on opinions regarding aesthetic features?

% Idea: Since I have had to extract things from ninja warrior data (lots of data), could show data and then a plot to show how % visualisations are useful to find things you can't immediately see in the data itself. % Idea: Axes in time series plots - Use short and tall axes and ask which varies most? (Maybe use HR data from http://ecg.mit.edu/time-series/) % Idea: Distance between bars in barplot - further apart harder to interpret? % Idea: Does uning a logarithmic scale impact interpretation?

1.4 The survey

Potentially will use google sheets and post all three, and ask people to complete one of them. %TODO Specify further

1.5 Demographic Questions

The questions below are used to

- Please enter your age
- If you are a university student or past university graduate please specify your area of study. (Drop down box: Science, Technology, Engineering, Maths, Arts, Social Sciences, Humanities, Business, N/A, Other (please specify))
- How strongly do you agree with each of the following statements? (Linear scale with 1 - 5, 1=strongly disagree, 5=strongly agree)

-
- I have good spatial awareness skills
-
- I have good observational skills
-
- I have good numerical skills
- Are you colourblind? (Checkbox: Yes, No, Prefer not to answer)
- Do you have any disorders that may affect visual processing? (this could be a general visual processing disorder or dyslexia, dyscalculia etc) ((Checkbox: Yes, No, Prefer not to answer))

1.6 Bar Plot Questions

American Ninja Warrior

The following bar charts present information regarding how many times 4 obstacles were used over the course of 10 seasons of the TV show 'American Ninja Warrior'. Please note that the answers to this section are entirely subjective, and that there are no 'correct' answers.

The first three questions refer to this bar chart, bar chart A.

Figure

%Note: Deliberately design 'bad' as well as good plots?

%%%%% Misleading aspects of plots: %%%%%

% Axis Scales % - Sources: % - https://www.nature.com/articles/s41559-018-0610-7 % - https://web.archive.org/web/20101123050530/http://graphpad.com/faq/file/1487logaxes.pdf % - https://dl.acm.org/doi/fullHtml/10.1145/3231772 % - https://journals.sagepub.com/doi/pdf/10.

1177/1050651920958392 % - https://journals.sagepub.com/doi/pdf/10.1177/1050651920958392 % -

%%%%%

% TODO Write about choosing more 'basic' questions to cut down on number of questions as well as appealing to audience

1.7 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 the language used to make the plots, and the order in which plots have been presented.

2.0.1 2.1.1 Ninja Warrior - Part 1

The first part of the survey consisted of showing the respondents three bar plots representing data regarding how many times four obstacles were used throughout 10 seasons of American Ninja Warrior. The three presented visualisations all showed the same raw data, but used three different y-axis scalings in order to assess whether changing this scale in these ways affects viewer interpretation. The questions asked were designed to test the effect of scale on both reading off exact values and gauging differences in values. Each respondent was asked four questions; two free form answer and two multiple choice. The use of free form answers did result in occasional non-valid answers, such as statements along the lines of "Don't know" when a number was required. Many people also opted to write a number between 0 and 1 when a percentage was required, but these will not be considered invalid, as there were a large number of responses of this type, but rather we will assume that any number between 0 and 1 is considered as the corresponding percentage. Ie. an answer of 0.5 will be considered as 50%.

First we will look at the summary statistics for each of the four questions laid out above for the whole population, before sub-setting for language and survey version number. Each table of summary statistics presents columns for the three plot types; the control plot, the logarithmic plot, and the truncated plot, respectively in that order.

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

This question, the first of the survey, asked participants to type the how many times Salmon Ladder was used, based on the bar plot. The 'correct' answer, or rather the true height of the corresponding bar, was 42. There were three invalid answers in these responses; one for the R versions of the survey and two for the Python versions. The invalid response in the R survey was '41/42', which we will take to be 41.5, and the invalid Python responses were given as 'Don't know' and 'Next to none.'. These two will be considered as 'NA' responses and thus discounted from the analysis of this question. These responses will. however, still be useful in our investigation; both were entered for the logarithmically scaled plot made in Python. The default log scaling in Python uses standard form notation, which perhaps these two participants were less familiar with. Similarly, there were two answers of '10^15' and '10^9', again potentially pointing towards the respondents being less familiar with this notation.

Below we see the summary statistics for the overall population, where we have taken the '41/42' response as 41.5 and omitted the two invalid text answers as well as an additional NA response and thus obtain a sample size of 67.

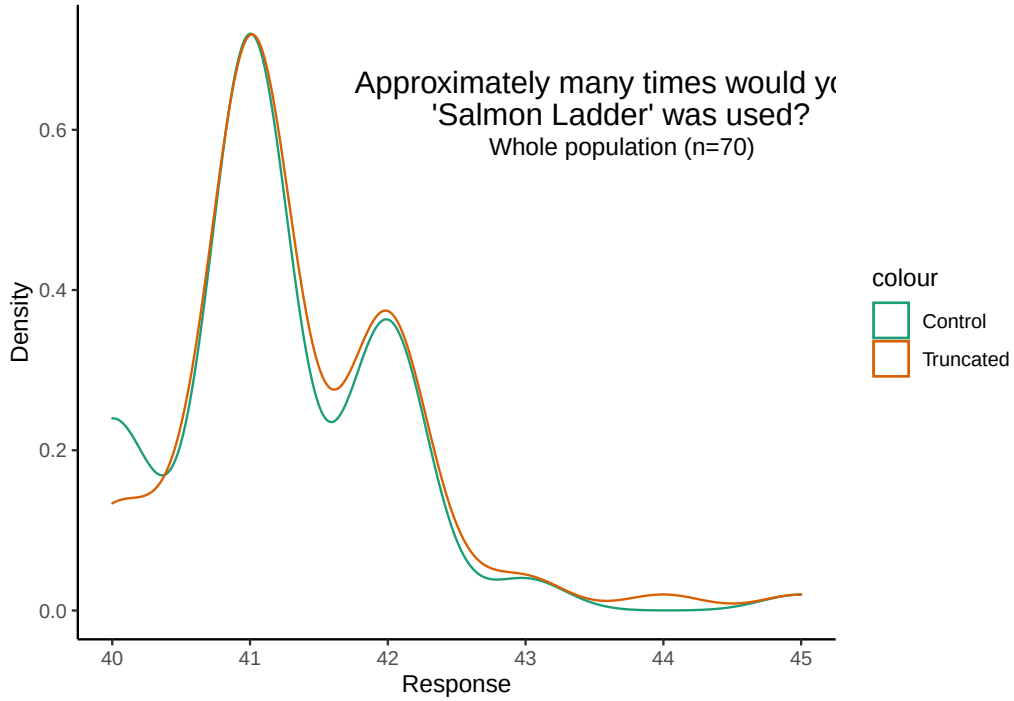
The table below presents the summary statistics for the total population for the first question.

```
## [1] n =  67
```

```
##    con_1_all      log_1_all           trn_1_all
##  Min.   :40.0   Min.   :9.000e+00   Min.   :40.00
##  1st Qu.:41.0   1st Qu.:3.000e+01   1st Qu.:41.00
##  Median :41.0   Median :3.500e+01   Median :41.00
##  Mean   :41.2   Mean   :1.493e+13   Mean   :41.37
##  3rd Qu.:42.0   3rd Qu.:4.000e+01   3rd Qu.:42.00
```

```
##  Max.   :45.0   Max.   :1.000e+15   Max.   :45.00
```

We can see that for the control and truncated plots we have means 41.21 and 41.35 respectively and both have median 41, which at first glance do not appear significantly different from the true value of 42. To investigate this further we will run some statistical tests. To decide if z-tests are applicable here we will first look at the distribution of these variables before running Shapiro-Wilk tests of normality. Below we can see a density plot depicting the distributions of the values for the control and truncated plots.



The two distributions are very similarly shaped, and neither appears similar to a Gaussian curve, hence it is likely that they will violate the normality condition of a z-test. To confirm this hypothesis, Shapiro-Wilk tests for normality are performed. **Control Plot**

```
##
##  Shapiro-Wilk normality test
##
## data:  con_1_all
## W = 0.81359, p-value = 5.596e-08
```

Truncated Plot

```
##
##  Shapiro-Wilk normality test
##
## data:  trn_1_all
## W = 0.82679, p-value = 1.327e-07
```

For both the control plot and truncated plot responses, the Shapiro-Wilk tests give p « 0.05, and thus we reject the hypothesis that these data are normal, and so they do, in fact, violate the normality condition required for a one-sample z-test. However, we can also calculate whether the sample size of 70 dictates that we may still use the z-test as an approximation.

One alternative to using a t-test is to use a Wilcoxon-Mann-Whitney (WMW) test. Note however that this test requires a symmetric distribution with even spread of values about the median. We see below, and from the density plot, that this isn't the case.

Control Plot

```
## [1] Median: 41
```

```
## [1] Number of observations below median: 12
```

```
## [1] Number of observations above median: 22
```

```
## [1] Number of observations at median: 36
```

Control Plot

```
## [1] Median: 41
```

```
## [1] Number of observations below median: 9
```

```
## [1] Number of observations above median: 28
```

```
## [1] Number of observations at median: 33
```

We now move on to consider the one-sample sign test. This is less powerful than the t-test and WMW test, but is required as the data violates the conditions for these two test.

However using a two-sided Wilcoxon test with $\mu = 42$ we achieve $p << 0.05$ for both of the sets of responses, signifying that we do in fact see a statistically significant difference in the location of these responses from the true value. Furthermore, applying one sided one sided tests, we see that the sample value is likely to be less than the true value, meaning respondents tended to underestimate. It does, however, appear that there is not much difference between the average values of responses for each of the two plots, but this comparison will be discussed in a later chapter.

Sign tests for the control Plot

```
##
##  Exact binomial test
##
## data:  42 and length(con_1_all)
## number of successes = 42, number of trials = 70, p-value = 0.1196
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4759319 0.7153289
## sample estimates:
## probability of success
##                    0.6
```

```
##
##  Exact binomial test
##
## data:  greater and length(con_1_all)
## number of successes = 22, number of trials = 70, p-value = 0.002548
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2085298 0.4363330
## sample estimates:
```

```
## probability of success
##              0.3142857

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  con_1_all
## V = 93, p-value = 5.673e-09
## alternative hypothesis: true location is less than 42

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  con_1_all
## V = 93, p-value = 1
## alternative hypothesis: true location is greater than 42
```

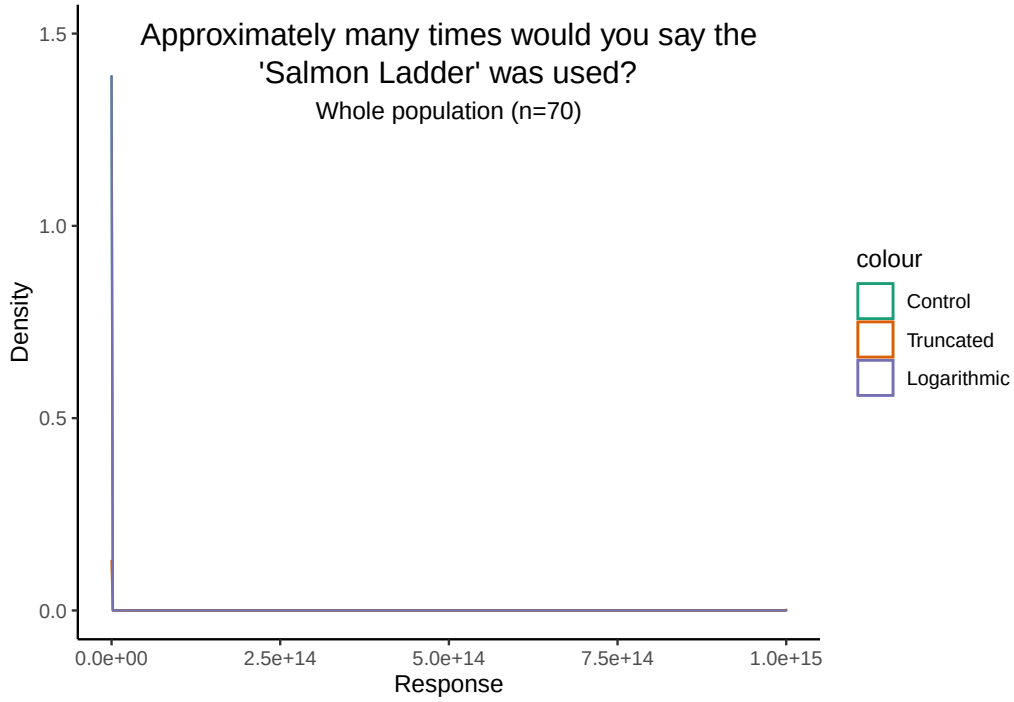
Wilcoxon tests for the truncated Plot

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  trn_1_all
## V = 150, p-value = 3.428e-07
## alternative hypothesis: true location is not equal to 42

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  trn_1_all
## V = 150, p-value = 1.714e-07
## alternative hypothesis: true location is less than 42

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  trn_1_all
## V = 150, p-value = 1
## alternative hypothesis: true location is greater than 42
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```

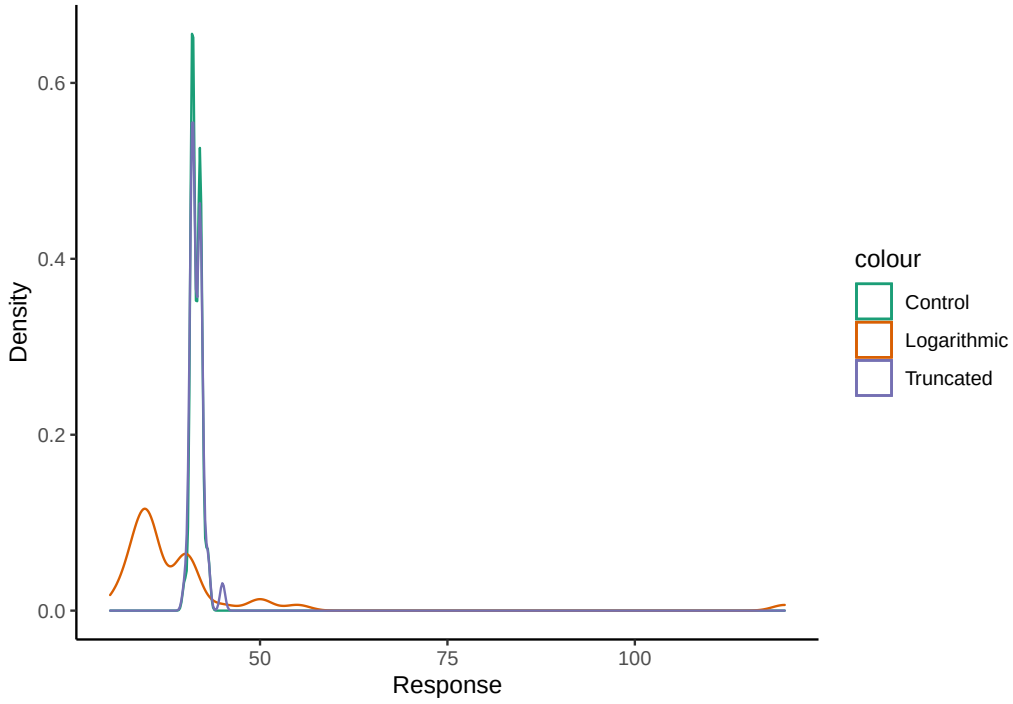


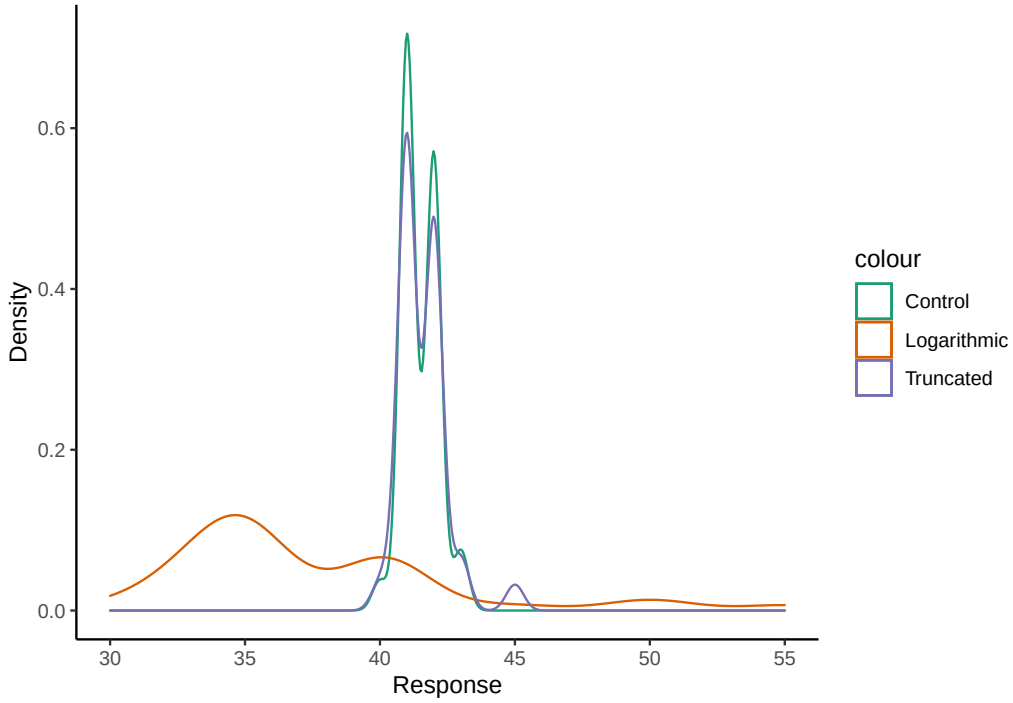
For both of these we also have a minimum of 40 and maximum of 45, giving a range of 5.

Now consider the values for the log plot. We see a much higher mean of 1.493e+13 which, on closer inspection of the languages (below), is attributed fully to the python versions of the plot.

Summary statistics of the R versions

```
##     con_1_r         log_1_r          trn_1_r
##  Min.   :40.00   Min.   : 30.00   Min.   :40.00
##  1st Qu.:41.00   1st Qu.: 35.00   1st Qu.:41.00
##  Median :41.00   Median : 35.00   Median :41.00
##  Mean   :41.49   Mean   : 39.74   Mean   :41.57
##  3rd Qu.:42.00   3rd Qu.: 40.00   3rd Qu.:42.00
##  Max.   :43.00   Max.   :120.00   Max.   :45.00
```


0.6
0.4
0.2
0.0
Density
50
75
100
Response
colour
Control
Logarithmic
Truncated

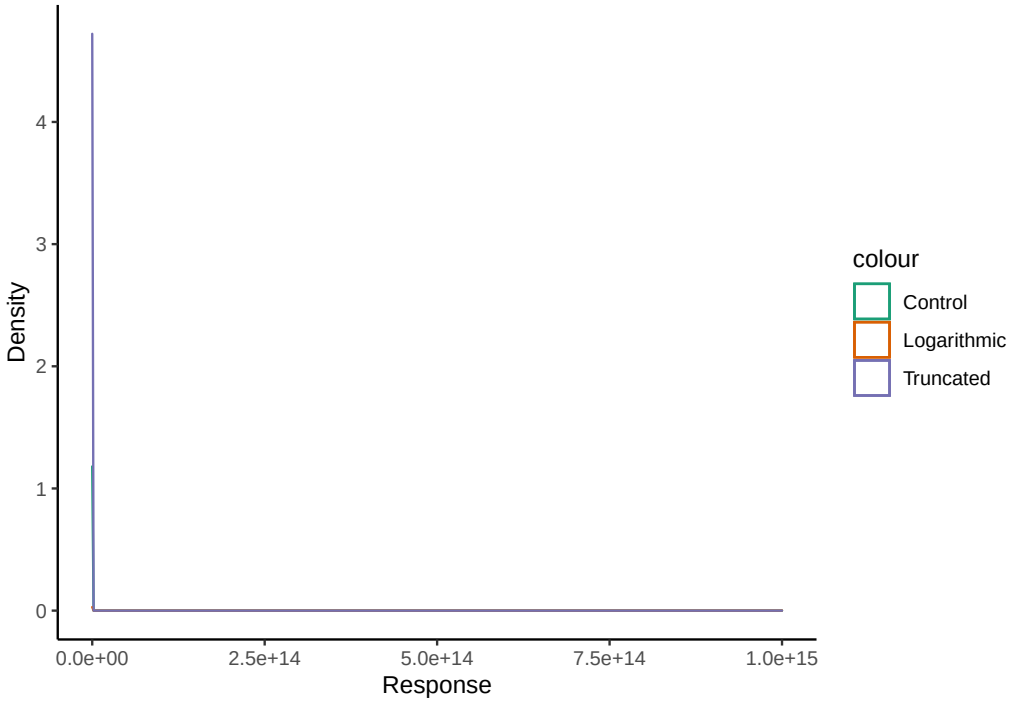

0.6
0.4
0.2
0.0
Density
30
35
40
45
50
55
Response
colour
Control
Logarithmic
Truncated

Summary statistics of the Python versions

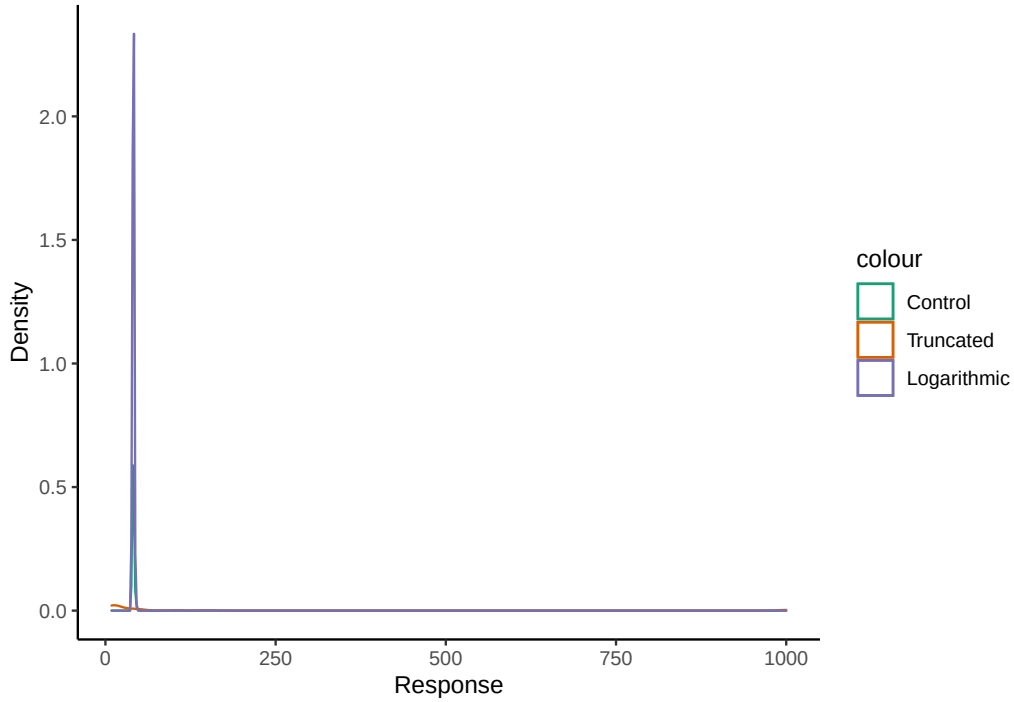
```
##     con_1_py        log_1_py            trn_1_py
##  Min.   :40.00   Min.   :9.000e+00   Min.   :40.00
##  1st Qu.:40.00   1st Qu.:1.200e+01   1st Qu.:41.00
##  Median :41.00   Median :1.500e+01   Median :41.00
##  Mean   :40.87   Mean   :3.571e+13   Mean   :41.10
##  3rd Qu.:41.00   3rd Qu.:5.000e+01   3rd Qu.:41.25
##  Max.   :45.00   Max.   :1.000e+15   Max.   :44.00
##                  NA's   :3
```

The responses contributing very heavily to this very large mean were written as 10^15 and 10^9, alongside two responses of 1000 one of 100. This lends to the idea that using matplotlib's default standard form notation for the log scale may have misled some participants who perhaps are less familiar with standard form. Adding to this conclusion are the NA values, two of which were responses of "Don't know" and "Next to None.". Looking at the R version, we have a mean of 39.74 and median 35 which, after once again applying the Wilcoxon tests, shows statistically significant under-estimation of the value, but are much closer to the true value than that of the Python plot.

```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```




```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```



Wilcoxon tests for the R log Plot

```
## Warning in wilcox.test.default(log_1_r, mu = 42, alt = "t"): cannot compute
## exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  log_1_r
## V = 104, p-value = 0.0001786
## alternative hypothesis: true location is not equal to 42

## Warning in wilcox.test.default(log_1_r, mu = 42, alt = "l"): cannot compute
## exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
```

```
## data:  log_1_r
## V = 104, p-value = 8.931e-05
## alternative hypothesis: true location is less than 42
```

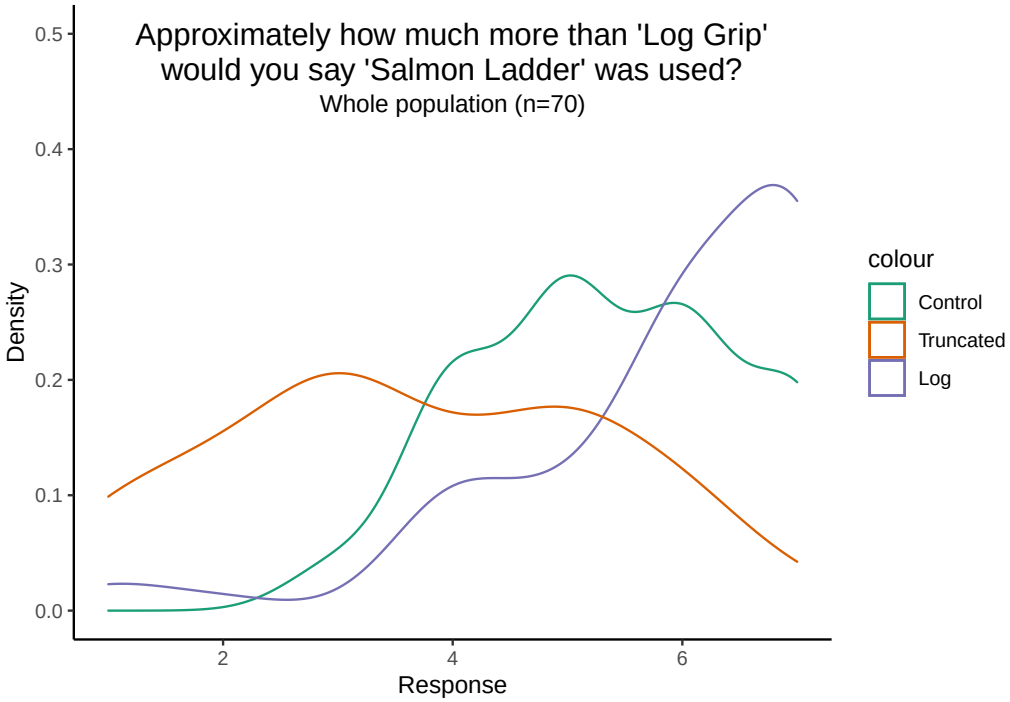
```
## Warning in wilcox.test.default(log_1_r, mu = 42, alt = "g"): cannot compute
## exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  log_1_r
## V = 104, p-value = 0.9999
## alternative hypothesis: true location is greater than 42
```

NA values for log plot data

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

Approximately how much more than 'Log Grip' would you say 'Salmon Ladder' was was used? Now we consider the results from the second question, in which the participants were asked to respond on a scale from 1-5. The density plot below depicts the distribution of results for each of the three plot types. We see that the distributions appear to once again be non-normal.

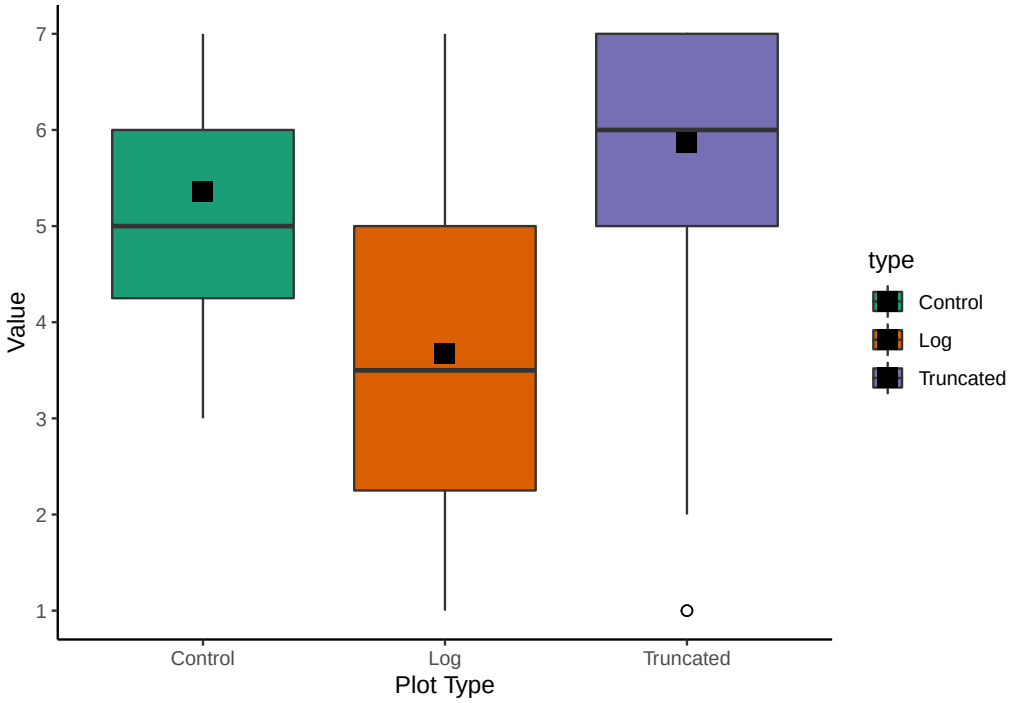



```
##    con_2_all       log_2_all       trn_2_all
##  Min.   :3.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.250   1st Qu.:2.250   1st Qu.:5.000
##  Median :5.000   Median :3.500   Median :6.000
##  Mean   :5.357   Mean   :3.671   Mean   :5.871
##  3rd Qu.:6.000   3rd Qu.:5.000   3rd Qu.:7.000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

An initial look at the table of summary statistics reveal means of 5.375, 3.671 and 5.871 respectively for the control, log and truncated plots, meaning that for the 'baseline' control plot participants, on average, judged the difference to be moderately significant, with the perceived difference being smaller for the log plot and larger for the truncated plot. This is consistent with results from [[[CITE chrome-extension://cbnaodkpfinfiipjblikofhlhlcickei/src/pdfviewer/web/viewer.html?file=file:///C:/Users/Katie/Downloads/YangVargasRestrepoStanleyMarsh%20(2020).pdf]]], in which the researchers, similar to this survey, showed participants a series of control bar plots alongside those with a truncated axis, and concluded that the

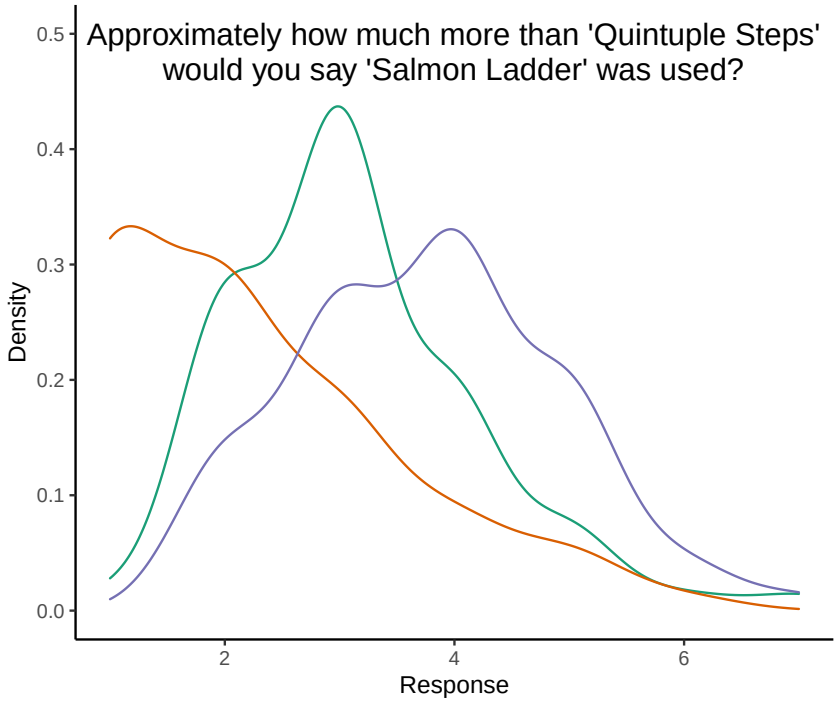
difference in values for the truncated axis were perceived to be larger than those of the control plots.

The box plot shows these results for each plot type.



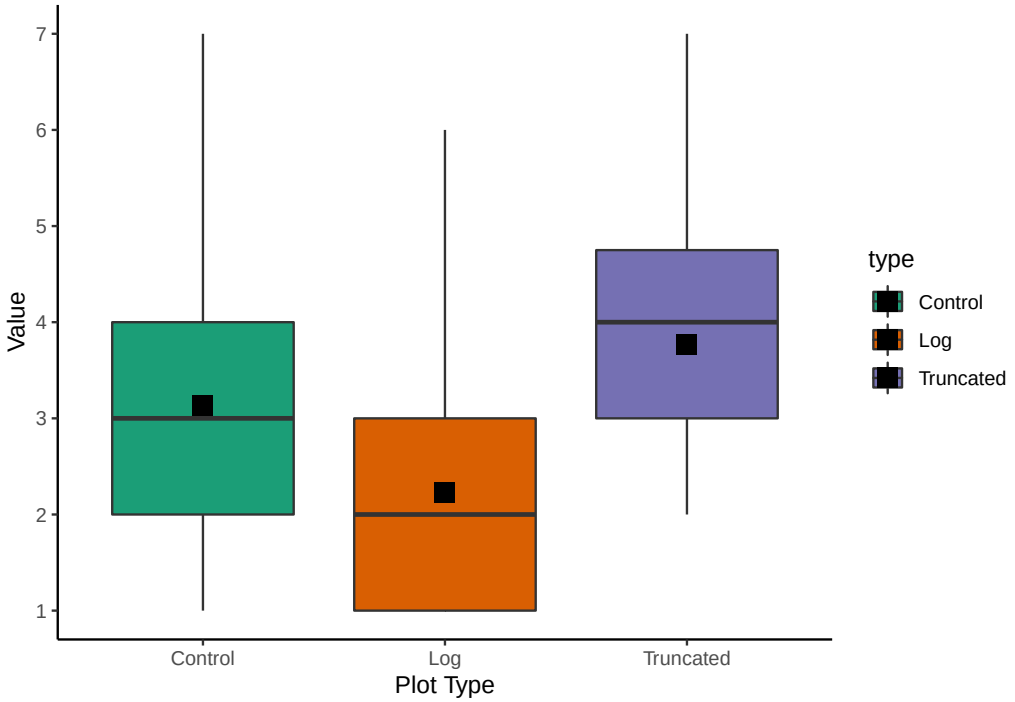
We see that the interquartile range for the control plot is smallest of the three at 1.75, followed by the truncated plot at 2, and then the log plot at 2.75. This depicts that overall, there was more of a consensus in the subjective perception of the difference for the control plot than the other two, and less agreement between participants for the logarithmic scale. This could be once again hypothesised to be due to misunderstanding of the standard form notation as well as the logarithmic scaling itself. The black squares represent the means here, and we can see that for the first two boxes, the mean is higher than the median, perhaps signifying a positive skew, with a slightly negative skew for the truncated plot.

Approximately how much more than 'Quintuple Steps' would you



say 'Salmon Ladder' was used?

```
##    con_3_all       log_3_all       trn_3_all
##  Min.   :1.000   Min.   :1.000   Min.   :2.000
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:3.000
##  Median :3.000   Median :2.000   Median :4.000
##  Mean   :3.129   Mean   :2.229   Mean   :3.771
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:4.750
##  Max.   :7.000   Max.   :6.000   Max.   :7.000
```



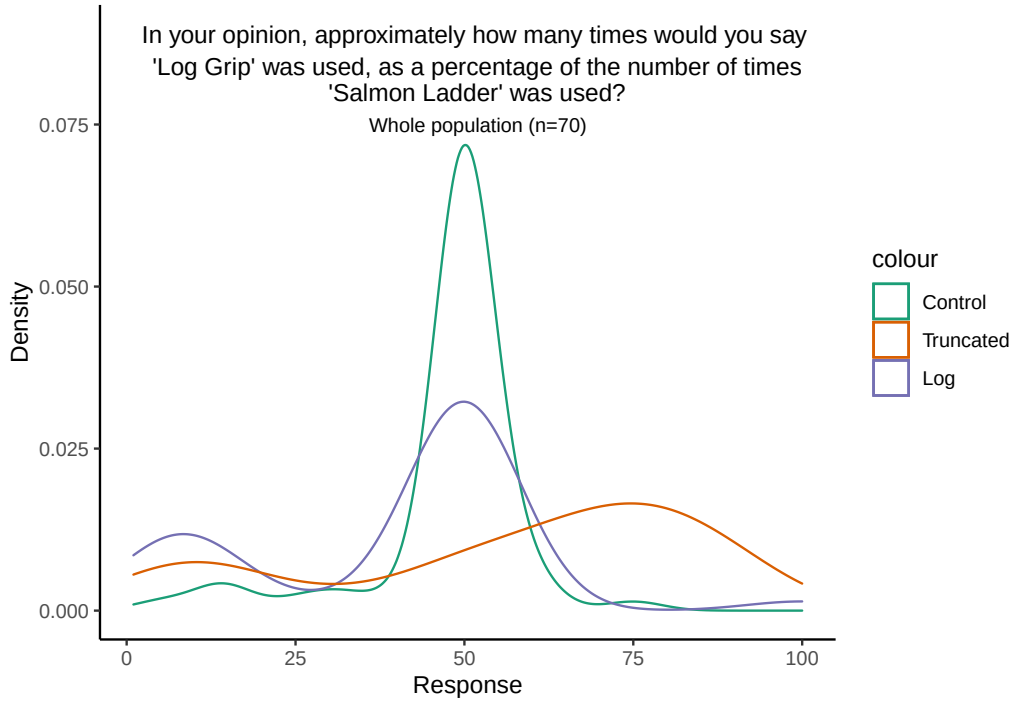
In your opinion, approximately how many times would you say 'Log Grip' was used, as a percentage of the number of times 'Salmon Ladder' was used?

```
##    con_4_all       log_4_all       trn_4_all
##  Min.   : 5.00   Min.   : 3.00   Min.   :  1.00
##  1st Qu.:50.00   1st Qu.:42.25   1st Qu.: 14.38
##  Median :50.00   Median :65.50   Median : 50.00
##  Mean   :47.66   Mean   :56.12   Mean   : 39.81
##  3rd Qu.:50.00   3rd Qu.:80.00   3rd Qu.: 50.00
##  Max.   :75.00   Max.   :90.00   Max.   :100.00
##  NA's   :3       NA's   :4       NA's   :2
```

```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 2 rows containing non-finite values (stat_density).
```

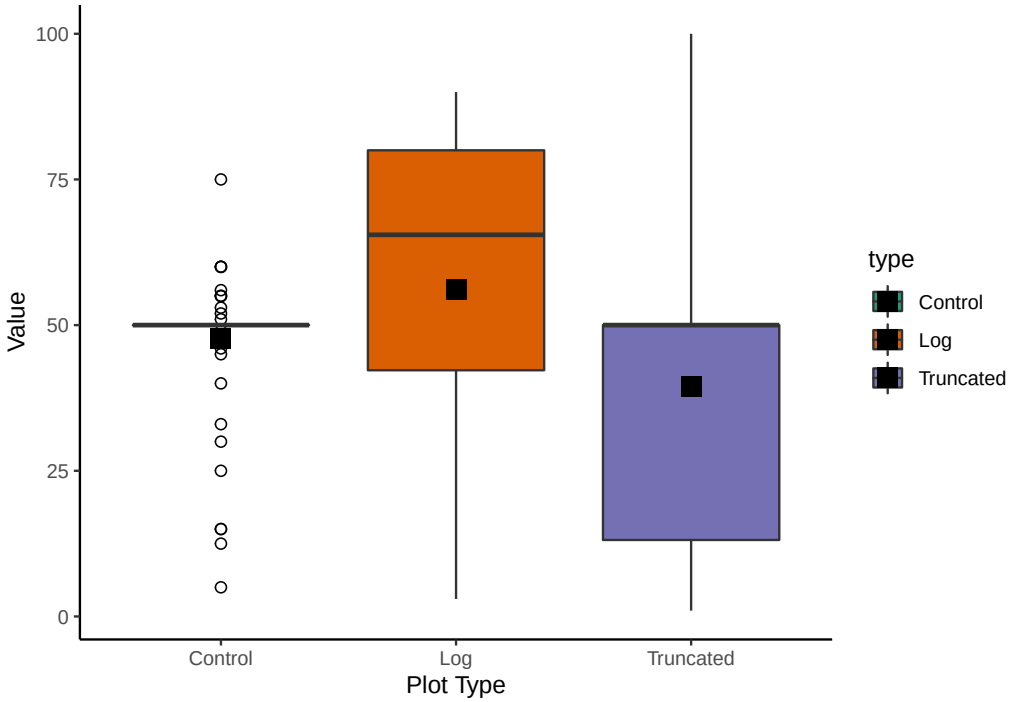
```
## Warning: Removed 4 rows containing non-finite values (stat_density).
```



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
```

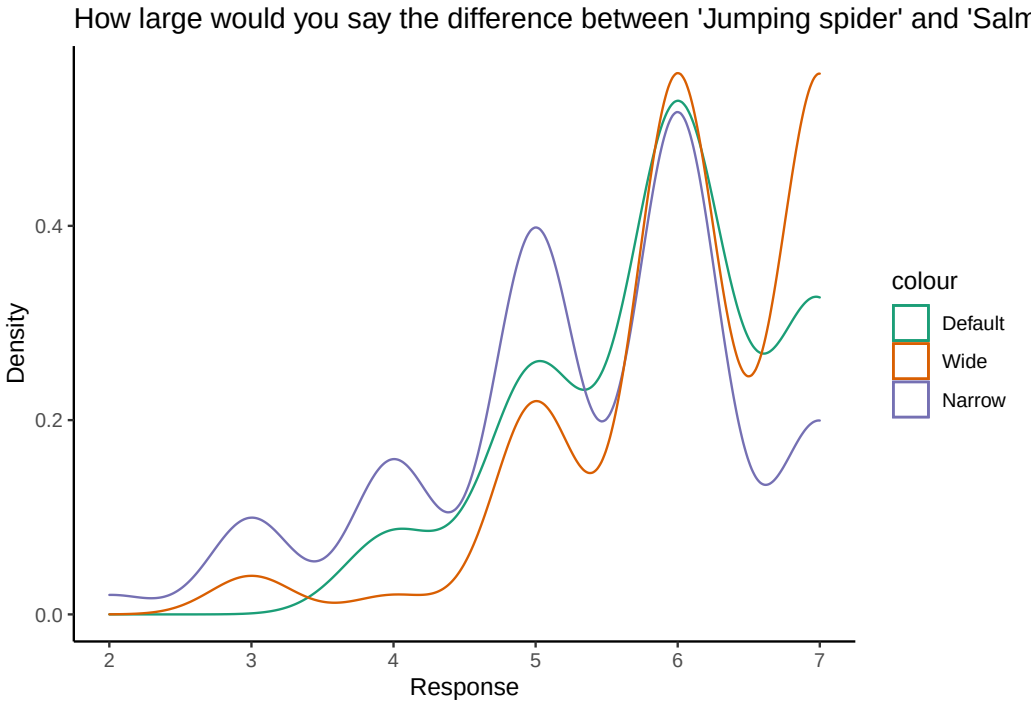
```
##                                      uni sp_aware obs_skl num_skl cblind vis_pro
## 12                               Science        2       3       2      No      No
## 17                           Engineering        4       4       4      No     Yes
## 25                             Geography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



2.0.2 2.1.2 Ninja Warrior - Part 2

How large would you say the difference between 'Jumping spider' and 'Salmon Ladder' is?

```
##    def_1_all       wid_1_all       nar_1_all
##  Min.   :4.000   Min.   :2.000   Min.   :3.000
##  1st Qu.:5.000   1st Qu.:5.000   1st Qu.:6.000
##  Median :6.000   Median :6.000   Median :6.000
##  Mean   :5.914   Mean   :5.357   Mean   :6.129
##  3rd Qu.:7.000   3rd Qu.:6.000   3rd Qu.:7.000
##  Max.   :7.000   Max.   :7.000   Max.   :7.000
```



How large would you say the difference between 'Log Grip' and 'Floating Steps' is?

```
##    def_2_all       wid_2_all       nar_2_all
##  Min.   :2.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3.000   Median :3.000   Median :3.000
##  Mean   :3.057   Mean   :3.057   Mean   :3.214
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :7.000   Max.   :5.000   Max.   :7.000
```

How many times would you say 'Floating Steps' were used?

```
##    def_3_all       wid_3_all       nar_3_all
##  Min.   :26.00   Min.   :24.00   Min.   :23.00
##  1st Qu.:27.12   1st Qu.:27.00   1st Qu.:27.00
##  Median :28.00   Median :28.00   Median :28.00
##  Mean   :27.97   Mean   :28.04   Mean   :27.39
```

```
##  3rd Qu.:28.00   3rd Qu.:29.00   3rd Qu.:28.00
##  Max.   :33.00   Max.   :30.00   Max.   :29.00
```

Comparisons

	A	B	C
Which of the three bar charts do you find most aesthetically pleasing?	26	25	18
Which bar chart do you feel is easiest to read and interpret?	30	17	23
Which bar chart do you find hardest to read and interpret?	14	27	29

2.0.3 2.1.3 Ninja Warrior - Part 3

How many times would you say 'Floating Steps' were used in the Finals (Regional/City) round?

```
##              Min. 1st Qu. Median     Mean 3rd Qu. Max.
## Stacked         9      10     11 14.32857      14   35
## Side by side   10      11     11 11.80000      12   40
```

How many times would you say 'Log Grip' was used in the Finals (Regional/City) round?

```
##              Min. 1st Qu. Median      Mean 3rd Qu. Max.
## Stacked         6       8      9 10.571429      10   25
## Side by side    2       8      9  9.057143      10   15
```

Please select the statement you feel applies to the bar chart above.

```
##              Equal Less More
## Stacked         27   31   11
## Side by side    60    5    2
```

Which obstacle do you think was used MORE in Finals (Regional/City) rounds, 'Log Grip' or 'Floating Steps'?

```
##              Equal Less More
## Stacked         56    2   12
## Side by side    57    4    9
```

Which bar chart do you feel is easiest to read and interpret?

```
##
## Side by side      Stacked
##           59           11
```

```
##       Side by side Stacked
## Set A           10       3
## Set B           11       1
## Set C            9       1
## Set D           11       1
## Set E            8       3
## Set F           10       2
```

Which colour scheme do you find most aesthetically pleasing?

```
##        A  B
## Set A  3 10
## Set B  1 11
## Set C  9  1
## Set D  1 11
## Set E  8  3
## Set F 10  2
```

Do you feel that one of the colour schemes makes it easier to read and interpret? If so, please select which one.

```
##
##   Default Greyscale   Viridis
##        30         5        35
```

```
##
## Default Viridis
##      12      13
```

```
##
##   Default Greyscale
##        18         4
```

```
##
## Greyscale   Viridis
##         1        22
```

2.0.4 2.1.4 Sales - Part 1

How much would you say sales of each company increased between January and December? [Company A]

```
##         Separate Truncated   Zeroed
## Min.    1.000000  1.000000 1.000000
## 1st Qu. 2.000000  2.000000 1.000000
## Median  3.000000  2.000000 1.000000
## Mean    3.043478  2.414286 1.371429
## 3rd Qu. 4.000000  3.000000 1.750000
## Max.    7.000000  7.000000 3.000000
```

How much would you say sales of each company increased between January and December? [Company B]

```
##         Separate Truncated   Zeroed
## Min.    1.000000  1.000000 1.000000
## 1st Qu. 4.000000  4.000000 2.000000
## Median  5.000000  6.000000 2.000000
## Mean    4.826087  5.144928 2.478261
## 3rd Qu. 6.000000  7.000000 3.000000
## Max.    7.000000  7.000000 6.000000
```

How large would you say the drop in sales between April and July of Company A is?

```
##         Separate Truncated   Zeroed
## Min.    1.000000  1.000000 1.000000
## 1st Qu. 3.000000  2.000000 1.000000
## Median  4.000000  3.000000 1.000000
## Mean    4.028571  2.814286 1.571429
## 3rd Qu. 5.000000  3.000000 2.000000
## Max.    7.000000  7.000000 6.000000
```

2.0.5 2.1.5 Sales - Part 2

Based on the above graph, how large would you say the difference is between the number of sales Company C makes and the number of sales Company D makes?

```
##         Truncated Zeroed
## Min.     2.000000    1.0
## 1st Qu.  4.000000    2.0
## Median   4.000000    3.0
## Mean     4.271429    2.7
## 3rd Qu.  5.000000    3.0
## Max.     7.000000    5.0
```