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

The purpose of this set of questions was to decipher whether altering the type of y-scaling used would notably affect perception of the data, in terms of both gauging exact values as well as giving subjective opinions on differences between values. As described previously, the set of questions in the first section is as follows:

- **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 each of the four questions laid out above for the whole population, before subsetting for language, degree subject, and survey version number. Each table of summary statistics presents columns for the three plot types; the control plot, the truncated plot, and the log-scaled plot, respectively in that order.

2.1.1 Whole Population (N=70)

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

The table below presents the summary statistics for the total population for the first question. Here the 'correct' answer, or rather the true value of the corresponding bar, was 42.

```
##    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
```

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. 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.

Wilcoxon tests for the control Plot

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  con_1_all
## V = 93, p-value = 1.135e-08
## alternative hypothesis: true location is not equal to 42
```

```
##
##  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
```

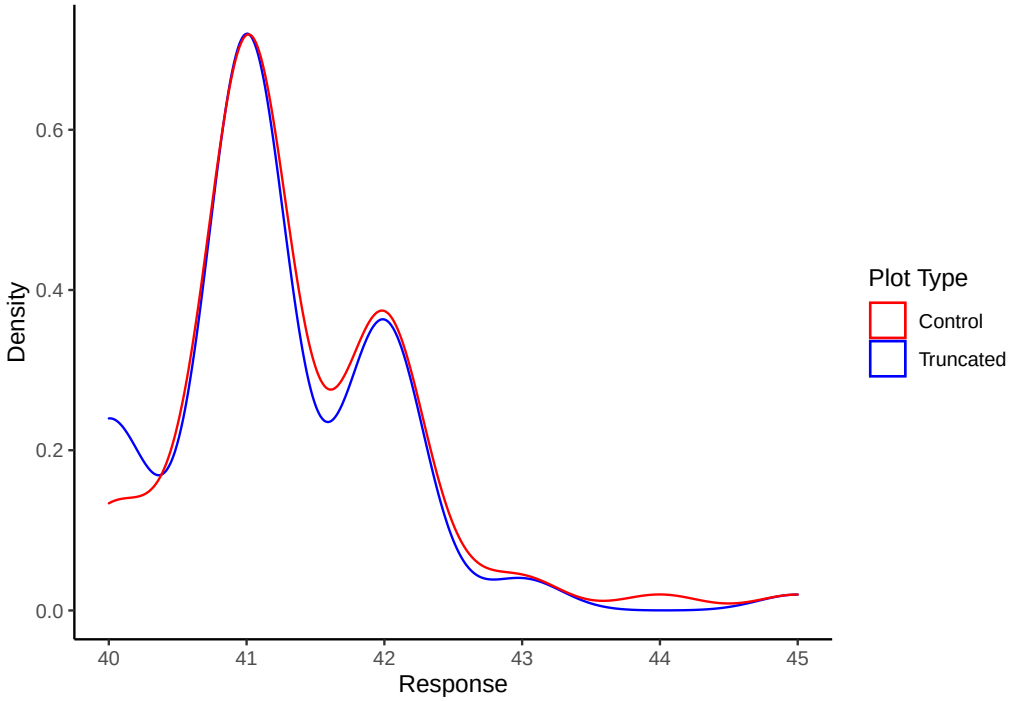
Wilcoxon tests for the truncated Plot

```
##
##  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
```

A Wilcoxon test was chosen as the data is non-normal, which we can see from both the density plots below in addition to the results of the Shapiro-Wilk normality tests, whereby we once again have p « 0.05, and thus there is significant evidence to reject the hypothesis that these data are normal, and so they violate the normality condition required for a one-sample t-test.



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 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
```

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.

Wilcoxon tests for the R log Plot

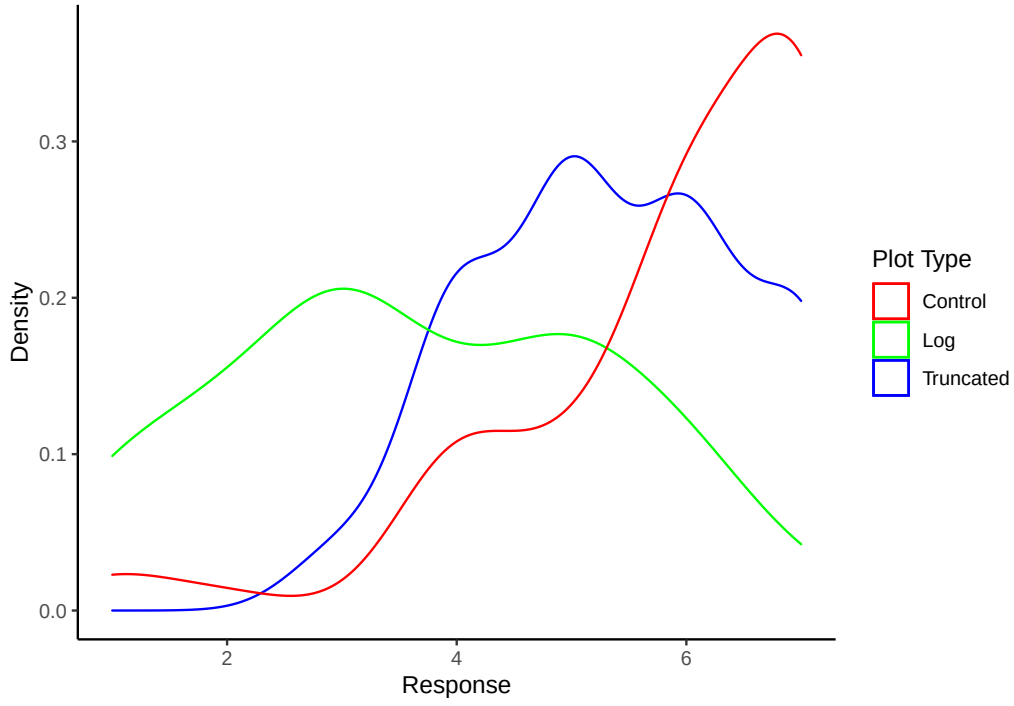
```
## Warning in wilcox.test.default(log_1_r, mu = 42, alt = "t"): cannot compute
## exact p-value with ties


##
##  Wilcoxon signed rank test with continuity correction
##
## data:  log_1_r
## V = 142, p-value = 0.0008878
## 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 = 142, p-value = 0.0004439
## 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 = 142, p-value = 0.9996
## 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
```

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

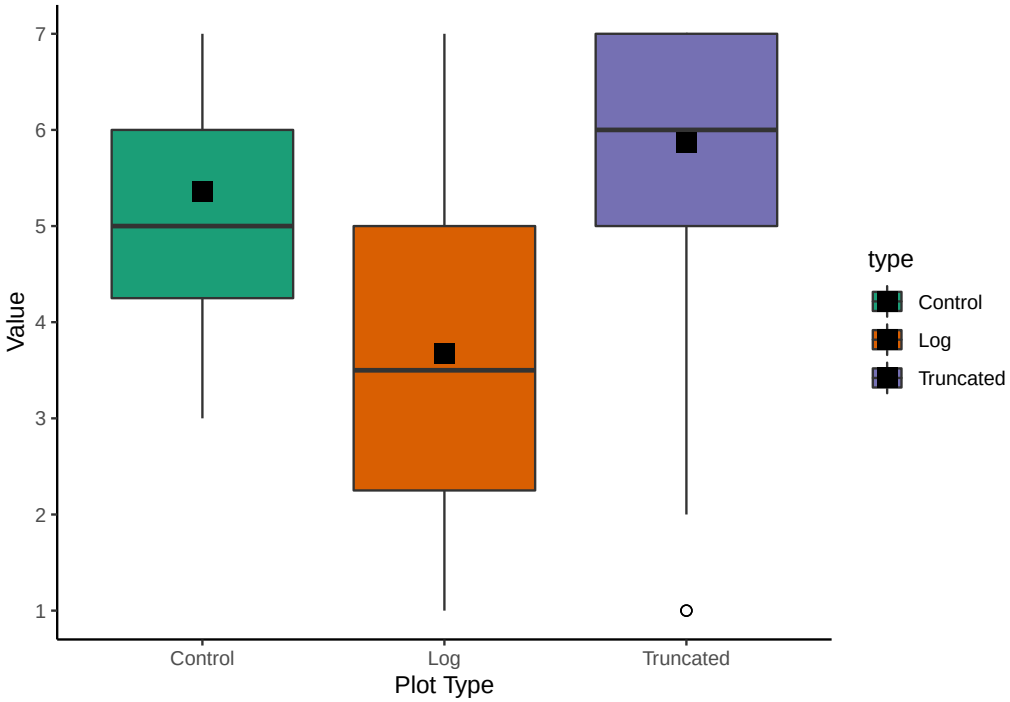
Q2 - Approximately how much more than 'Log Grip' would you say 'Salmon Ladder' was was used? Now we consider the results from the second question, in which the participants were asked to respond on a scale from 1-5. The density plot below depicts the distribution of results for each of the three plot 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.

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
```