- **Statistical Analysis**

This is just a very brief overview to broadly direct your data analysis. We are more interested in what you might observe from the data than the sophistication of your statistical analysis. If you group the data using the simple techniques contained in this overview then that should be sufficient to make some reasonable observations.

Lets say, rather than the data provided you had the following data to analyse…

User	Response
1	4
2	5
3	2
4	3
5	3
6	4
7	3
8	3
9	5
10	5
11	4

What can we say about this data set? Well the basis statistical measures are the number of observations (in this case 11), the minimum value (in this case 2) and the maximum value (in this case 5). 11 is not a statistically significant number so we would not have much confidence in any observations, and (from observation) we might note that there is a reasonable spread of values between the minimum and maximum.

Another very common statistical measure is the mean, or arithmetic average of the distribution. The average is often used to represent a distribution of values, but it is actually a very poor representation of the data whenever the distribution is not evenly spread either side of the mean. Take the following illustration for example, where the distribution is not evenly spread and the average is not a useful or accurate representation of the depth distribution of the river…
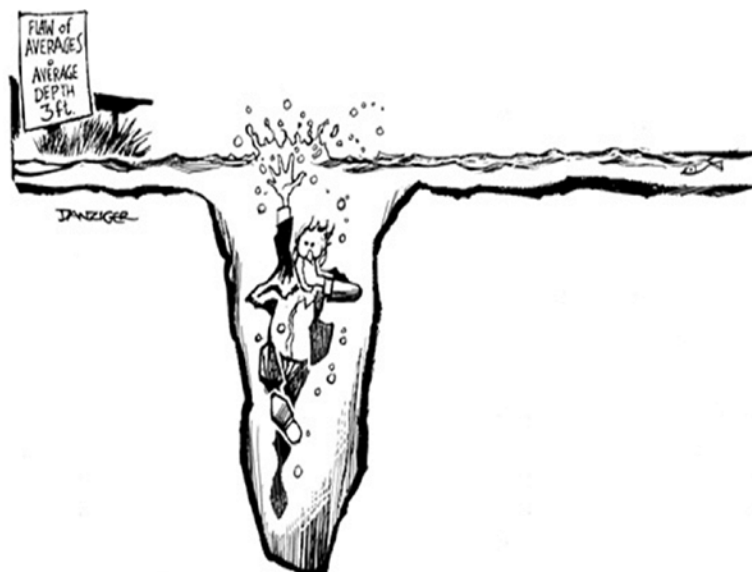
For this reason we do NOT recommend use of mean/average to describe a distribution. What other ways can we represent them?

One way is graphically. To do this we must count up the frequency of each value and display as a bar chart. For our example data set, that amounts to the following:

Value	Frequency
1	0
2	1
3	4
4	3
5	3
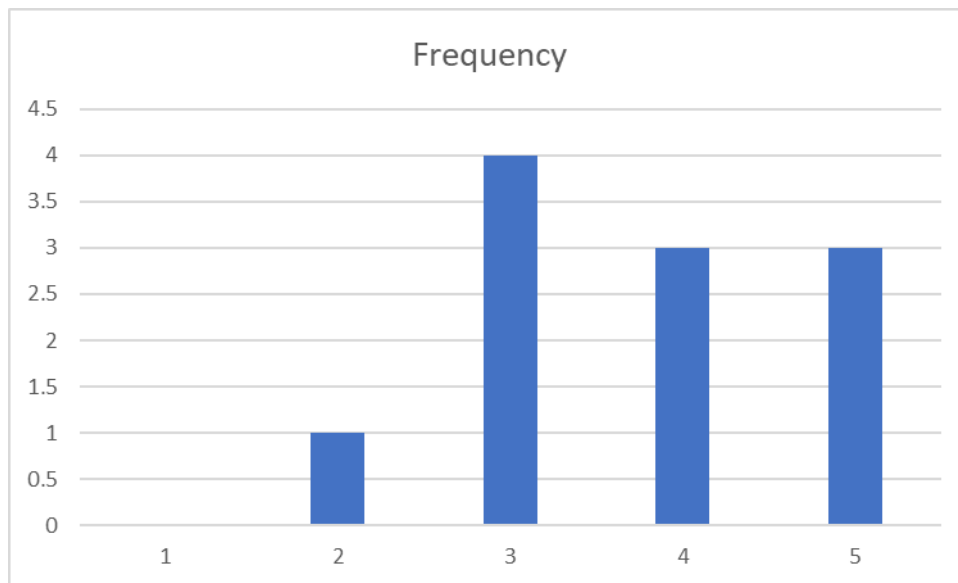
This tells us a great deal about the distribution – which is the most frequent value, how the distribution is spread, etc. We can see that the data values are generally towards the higher values, and there are very few low values present.

Spreadsheets are very good for providing this kind of representation. They also have functions available to automatically calculate other key values as follows:

Number	11	
Minimum	2	
Maximum	5	
Mean	3.73	
Median	4	
Mode	3	
SD	1.01	
Spread	2.72	4.74

The Median value is the value of the value with as many values above as below when all values are listed in order, or value number 6 (as the middle of 11 values) in the following re-ordered list.

The Mode value is the most frequent value.

2

SD, or standard deviation, is a measure of how spread out the values are around the Mean. The smaller the SD relative to the Mean, the tighter the distribution. The Spread is calculated by first subtracting the SD from the Mean and then adding the SD to the Mean.

User	Response	Ordered
1	4	2
2	5	3
3	2	3
4	3	3
5	3	3
6	4	4
7	3	4
8	3	4
9	5	5
10	5	5
11	4	5

Any standard spreadsheet will calculate these values for you given any list of data values.

We are simply asking you to analyse the data provided and to make some observations about what the data tells you as a comparison between conventional and VR training methods in terms of how the students perform in there exit tests. Is one approach consistently better than the other? Does one improve performance for all students equally over the other, or does one change the distribution of results in any way? Are there any exceptions to the general observation? Is every key learning aspect affected in the same way?

- ## **Performance Data**

The data provided in the spreadsheet has been collected to compare learning performance of workers undergoing conventional (powerpoint) *versus* video game based teaching.

At the end of each training session, workers are required to complete a test comprising 10 questions. These questions cover the following key aspects:

- Working at Height
- PPE
- Lifting Operations
- Electricity
- Vehicle Hazards
- Working Near Water

There are 3 different test papers, each comprising an individual collection of questions.

In total, there were 1,294 respondents who undertook conventional training and 1,268 respondents who undertook training using video games.

Correct	Conventional Paper A		Paper B		Paper C		Virtual Paper A		Paper B		Paper C	
0	0	0.00%	1	0.23%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
1	0	0.00%	1	0.23%	1	0.24%	0	0.00%	0	0.00%	0	0.00%
2	2	0.45%	1	0.23%	1	0.24%	0	0.00%	0	0.00%	0	0.00%
3	0	0.00%	1	0.23%	6	1.43%	0	0.00%	3	0.67%	0	0.00%
4	5	1.12%	3	0.70%	4	0.95%	1	0.23%	5	1.12%	1	0.25%
5	6	1.35%	13	3.03%	8	1.91%	5	1.17%	9	2.02%	11	2.77%
6	10	2.24%	14	3.26%	4	0.95%	20	4.69%	10	2.25%	13	3.27%
7	73	16.37%	68	15.85%	43	10.26%	41	9.62%	66	14.83%	45	11.34%
8	133	29.82%	83	19.35%	90	21.48%	90	21.13%	96	21.57%	77	19.40%
9	134	30.04%	144	33.57%	106	25.30%	163	38.26%	147	33.03%	135	34.01%
10	83	18.61%	100	23.31%	156	37.23%	106	24.88%	109	24.49%	115	28.97%
total	446		429		419		426		445		397	
overall					1294						1268	

The data records the frequency with which each group scored a given number of questions correctly, separated into the three test papers.

It also shows how each group scored across each of the key learning aspects. Given there are different numbers of questions for each learning aspect in each of the three test papers, for consistency the distributions have been converted to 6 point scales. This accounts for some of the strange distributions. For example, if there is only one question for a given aspect on a particular test paper, there will only be values for 0 (1 question answered incorrectly) or 6 (1 question answered correctly).

Correct	Conventional Height		PPE		Lifting		Electricity		Vehicle		Water	
0	73	5.64%	52	6.01%	50	3.86%	19	1.47%	142	10.97%	17	1.31%
1	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
2	20	1.55%	0	0.00%	15	1.16%	49	3.79%	0	0.00%	0	0.00%
3	354	27.36%	0	0.00%	363	28.05%	100	7.73%	0	0.00%	0	0.00%
4	133	10.28%	0	0.00%	86	6.65%	254	19.63%	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	714	55.18%	813	93.99%	780	60.28%	872	67.39%	1152	89.03%	1277	98.69%
total	1294		865		1294		1294		1294		1294	
Correct	Virtual Height		PPE		Lifting		Electricity		Vehicle		Water	
0	54	4.26%	44	5.35%	39	3.08%	20	1.58%	108	8.52%	13	1.03%
1	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
2	33	2.60%	0	0.00%	9	0.71%	56	4.42%	0	0.00%	0	0.00%
3	324	25.55%	0	0.00%	317	25.00%	92	7.26%	0	0.00%	0	0.00%
4	116	9.15%	0	0.00%	94	7.41%	247	19.48%	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	741	58.44%	779	94.65%	809	63.80%	853	67.27%	1160	91.48%	1255	98.97%
total	1268		823		1268		1268		1268		1268	