Benford's law investigation

The following table lists the heights of buildings which are the tallest in the world for different categories (e.g. Tallest tower, tallest bridge etc.) Heights are listed in both metres and feet.

Height (m)	Height (ft)		Height (m)	Height (ft)
829.8	2,722		189	620
640	2,100		187	614
634	2,080		182	597
629	2,063		178.3	585
604	1,982		175	574
601	1,972		173.7	570
529	1,736		172	564
472	1,549		171	561
458	1,503		167.6	550
452	1,482		167.5	550
442	1,450		162	530
419.7	1,377		160	525
400	1,312		152.4	500
385	1,263		141.3	463.6
380	1,247		138.98	456
366.8	1,203		138.8	455.2
354	1,161		139	456
342	1,122		138	453
314	1,031		122	400
305	1,001		118	387
291	954		118	387
273.8	898		117.5	386
270	886		110	361
265	870		100	328
262	860		85.4	280
248	814		85	279
214	702		82.5	271
202	663		76	249
192	630		23.2	76

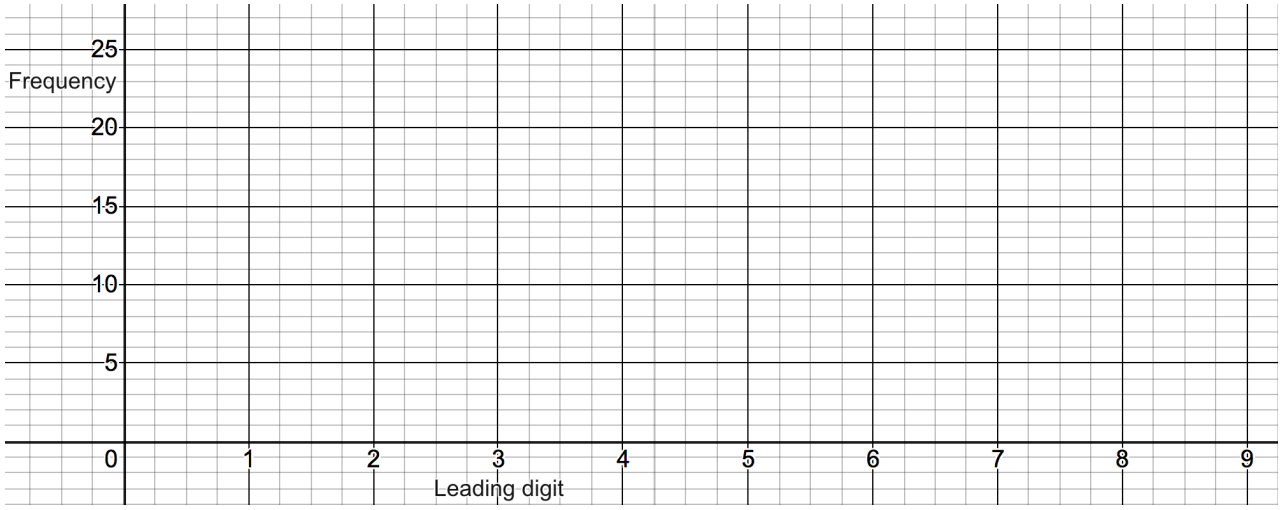
Data from List of Tallest Buildings and Structures

1. What would predict to see when you count the frequency of the leading digits (the first non-zero)?

2. (i) For the building heights in metres fill in the following frequency table.

Leading digit	1	2	3	4	5	6	7	8	9
Frequency									

(ii) Plot your points on the graph below.



3. (i) For the building heights in feet fill in the following frequency table.

Leading digit	1	2	3	4	5	6	7	8	9
Frequency									

(ii) Plot your points on the graph above.

4. Comment on your results.

5. Benford's Law states that for this kind of data the leading digit d occurs with probability $P(d)$

$$P(d) = \log(d+1) - \log(d)$$

(i) Show that this law can be stated as:

$$P(d) = \log\left(1 + \frac{1}{d}\right)$$

(ii) Use Benford's law to calculate the expected frequency for the leading digits of building heights in this sample. Plot these points on your graph and compare the graphs.

(iii) We can use Desmos to fit a regression curve for the building height expected frequency f for each leading digit d. Use Desmos to fit an exponential regression of the form:

$$f = ae^{bd} + c$$

6. Another example of data which fits Benford's Law is the distribution of the leading digits of 2^n. The first 96 numbers in this sequence were calculated and listed below:

Leading digit	Frequency	Percentage	Expected
1	29	30.2 %	
2	17	17.7 %	
3	12	12.5 %	
4	10	10.4 %	
5	7	7.3 %	
6	6	6.3 %	
7	5	5.2 %	
8	5	5.2 %	
9	5	5.2 %	

Calculate the percentage expected by Benford's Law and compare this to the actual data.

7. Benford's Law is used by some investigators. Can you think why this may be useful? Investigate online some of the real life use Benford's Law and the reason behind it.