

Quantitative Data

Think about gathering information on the question

How many siblings do you have?

1. Similar to the earlier question we examined (*What is your favorite sport?*) explain some of the issues that could arise from collecting data based on this question.

lack of clarity

adopted sibling?
 half? step? foster?
 sibling who died?
 living w/ you or not?

Question may need editing/refinement
 (depends on what you seek)

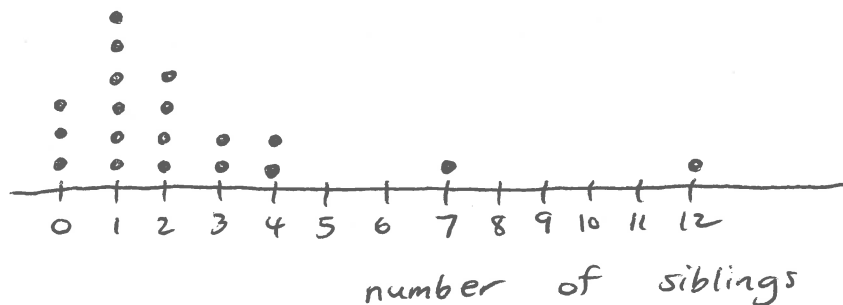
2. Suppose we collect data for the number of siblings; here are the results:

0 1 4 0 2 2 3 1 7 3 1 4 1 2 1 0 12 1 2

Organize the data first (e.g., a frequency table).

# siblings	frequency (how many)
0	3
1	6
2	4
3	2
4	2
7	1
12	1

3. Construct a **dot plot** from the data. A dot plot here would have “number of siblings” running horizontally and a dot for frequency (e.g., you’d have 4 stacked dots for 2 siblings).



4. (We'll do this as a class but you can try them beforehand)

ranges from no siblings to 12 siblings

(a) What is the **range** of the data?

$$\text{range} = \text{max} - \text{min} = 12 - 0 = 12$$

(b) Make some general statements about the layout of the data. What stands out in the dot plot?

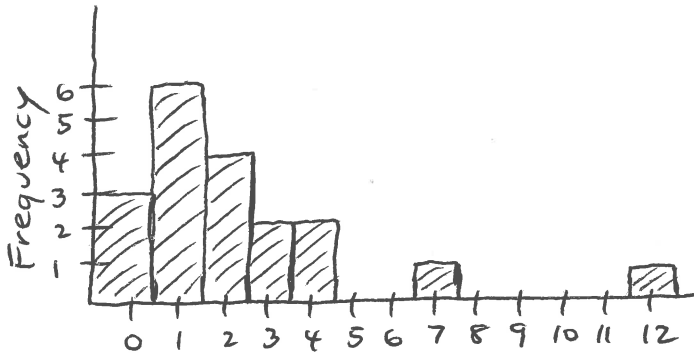
0-2 (cluster) ← lots of data

2 gaps in the data

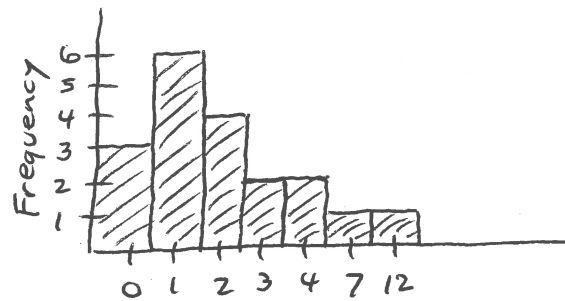
outliers (there is a formal definition for this in statistics)

"most polled have 2 or fewer siblings"

5. Construct a **histogram** to summarize this numerical data. A histogram is like a happy marriage of the dot plot and a bar graph. The bars in your histogram should touch (Why?).



Valid or no?



Measures of Central Tendency (Mean, Median, Mode)

Mean (Average):	Add each data value and divide by the number of data values.
Median:	Arrange the data in numerical order. The median is the middle data value. Note: If there are an even number of data values, then find the mean of the two closest to the middle.
Mode:	The data value that occurs most often.

6. Find the 3Ms for the sibling data (be sure to label one from another). Any surprises? Which one is the center? Why?

$$\bar{x} = \frac{47}{19} \approx 2.47 \text{ (mean)}$$

0 0 0 1 1 1 1 1 2 2 2 2 3 3 4 4 7 12
 ↑
 median = 2 mode = 1

7. Temporarily remove the data point 12 from the list and recalculate the 3Ms. Discuss.

$$\bar{x} = \frac{35}{18} \approx 1.94 \quad \text{median} = 1.5 \quad \text{mode} = 1$$

8. Why do you think we even bother calculating the 3Ms?

"summarizes" the data
 w/ one value