

## Categorical Data

Categorical Data: Data are not numbers but *categories*.

1. What is wrong with collecting data based on the question,

*What is your favorite sport?*

Think of a few things that should be clarified so that when people answer this question, we will obtain useful data. This list can be quite varied and looooooong if you really think about it.

- There are many different sports.
- Some people may consider certain things as real sports and some won't.

2. Suppose we eventually adjust the question to read,

*What competitive team sport do you most enjoy watching?*

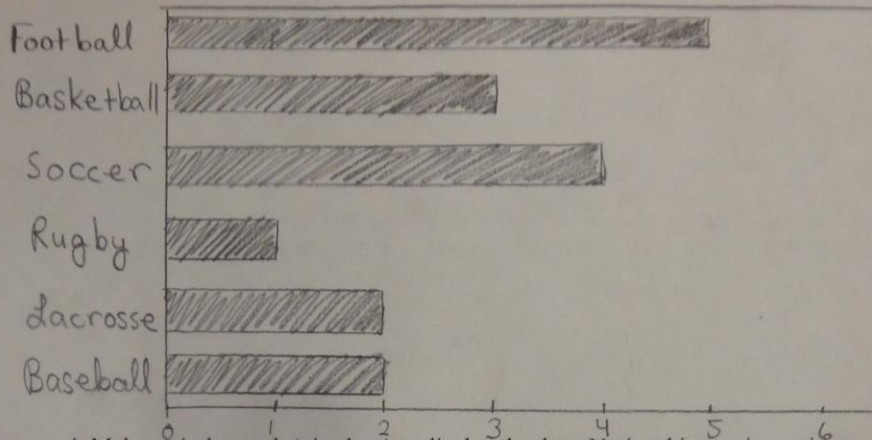
These are the responses you get from different individuals:

~~Football, Basketball, Soccer, Soccer, Rugby, Football, Football, Basketball, Soccer, Lacrosse, Soccer, Basketball, Baseball, Football, Baseball, Football, Lacrosse~~

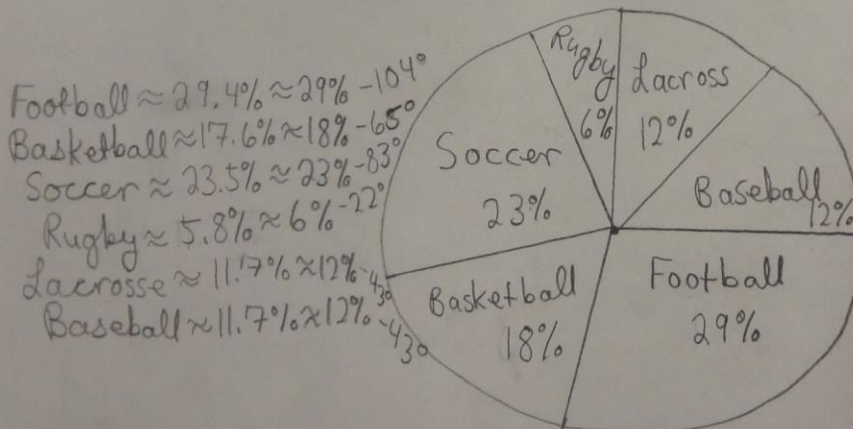
In table form, categorize the data and show the number of times (the frequency) each category was chosen. This is called a **frequency table**.

Sport	Tally	Frequency
Football		5
Basketball		3
Soccer		4
Rugby		1
Lacrosse		2
Baseball		2

3. Make a neat bar graph to display this data. Label the graph to minimize any misinterpretation.



4. Make a circle graph (pie chart) to display the data. Notice this requires using percent and degrees to ensure an accurate graph. Organize this information neatly so you can then construct a pie chart.



5. Look at both the bar graph and the pie chart. They each have their own advantages and disadvantages. List some of these below.

- The bar graph shows a specific amount of people that chose each sport.
- The pie chart shows the percentages of people that chose each sport.

6. **Analysis.** Write down a few statements that can be immediately inferred from one of the graphs.

- Football is the most popular
- Rugby is least popular
- Lacrosse & baseball are equally liked.

I like the responses to #1 here too:

Categorical Data: Data are not numbers but *categories*.

1. What is wrong with collecting data based on the question,

*What is your favorite sport?*

Think of a few things that should be clarified so that when people answer this question, we will obtain useful data. This list can be quite varied and looooooong if you really think about it.

- is this a favorite sport to be played? OR a favorite sport to watch?
- what counts as a sport?
- does it have to be a sport using equipment?
- does it have to be a team sport?

2. Suppose we eventually adjust the question to read,

*What competitive team sport do you most enjoy watching?*

These are the responses you get from different individuals:

Football, Basketball, Soccer, Soccer, Rugby, Football, Football, Basketball, Soccer, Lacrosse, Football, Basketball, Soccer, Lacrosse, Football, Lacrosse

Also note that bar graphs can be vertical or horizontal:

3. Make a neat bar graph to display this data. Label the graph to minimize any misinterpretation.

Sport	# of people
Football	5
Basketball	3
Soccer	4
Rugby	1
Lacrosse	2
Baseball	2

4. Make a circle graph (pie chart) to display the data. Notice this requires using percent and degrees to ensure an accurate graph. Organize this information neatly so you can then construct a pie chart.

Football:  $\frac{5}{17} \rightarrow 30\% \rightarrow 108^\circ$   
Basketball:  $\frac{3}{17} \rightarrow 18\% \rightarrow 65^\circ$   
Soccer:  $\frac{4}{17} \rightarrow 24\% \rightarrow 86^\circ$   
Rugby:  $\frac{1}{17} \rightarrow 6\% \rightarrow 22^\circ$   
Lacrosse:  $\frac{2}{17} \rightarrow 12\% \rightarrow 44^\circ$   
Baseball:  $\frac{2}{17} \rightarrow 12\% \rightarrow 44^\circ$

5. Look at both the bar graph and the pie chart. They each have their own advantages and disadvantages. List some of these below.

The bar graph is nice because you can easily see which is most popular. The pie chart is nice because you can see how much of the whole group is associated with each sport.

6. **Analysis.** Write down a few statements that can be immediately inferred from one of the graphs.

- Football is the most popular
- Rugby is the least popular
- it is more difficult to compare sizes in the pie chart, in my opinion.