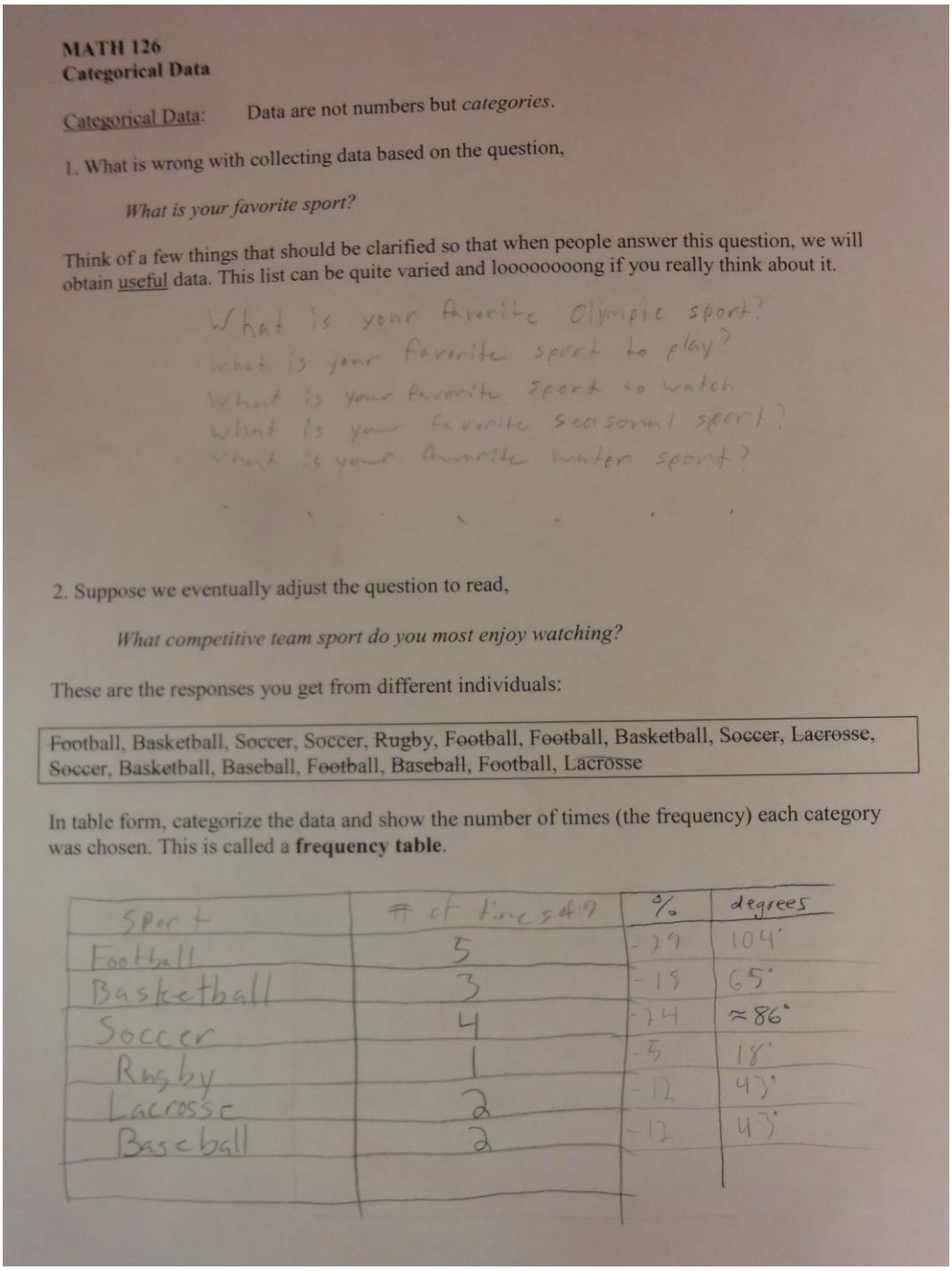


MATH 126
Categorical Data

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 looooooooong if you really think about it.

What is your favorite Olympic sport?
What is your favorite sport to play?
What is your favorite sport to watch
What is your favorite seasonal sport?
What is your favorite water 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:

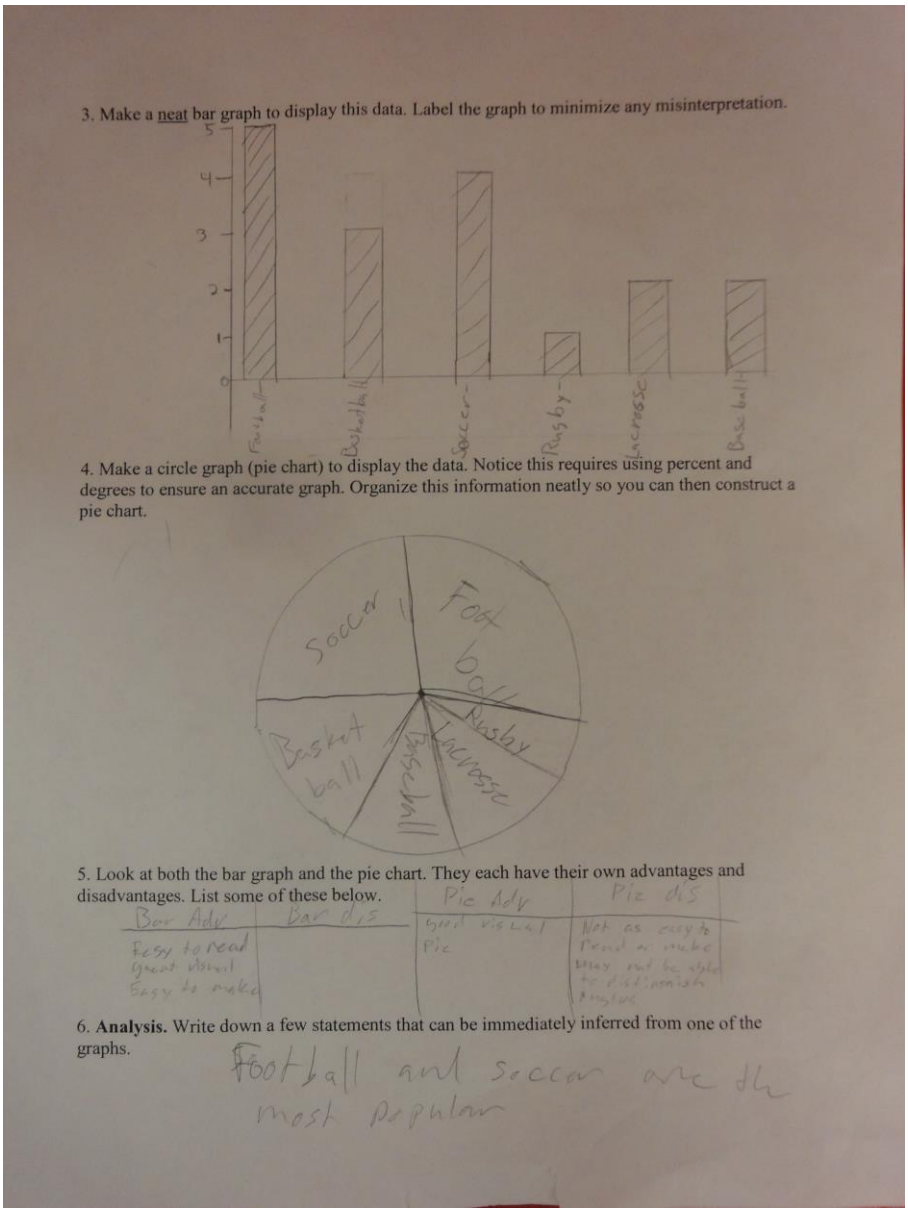
Football, Basketball, Soccer, Soccer, Rugby, Football, Football, Basketball, Soccer, Lacrosse, Soccer, Basketball, Baseball, Football, Baseball, Football, Lacrosse

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	# of times it's said	%	degrees
Football	5	~29	104°
Basketball	3	~18	65°
Soccer	4	~24	≈86°
Rugby	1	~5	18°
Lacrosse	2	~12	43°
Baseball	2	~12	43°

Question 1 addresses some important issues: Is this your favorite sport to play? To watch? If watching, do we mean on television or live in person? Also, what is the definition of a sport? Is dancing a sport? Checkers? Poker? Individual views on these issues may alter the responses you receive. *Not asking a clear question may result in an ambiguous data set.*

Garbage in = Garbage out.



Comments on #5: The bar graph shows the frequency in each category. In other words, the raw data is still visible. In the pie graph, the raw data is lost but something else is gained: you can see the percentage of the whole (difficult to detect in the bar graph). Moreover, you can see what is happening with combinations of categories. For example, a quick glance at the pie graph tells us that more than half of the responses were either football or soccer. This is more difficult to see in the bar graph.

Another nice pie graph: