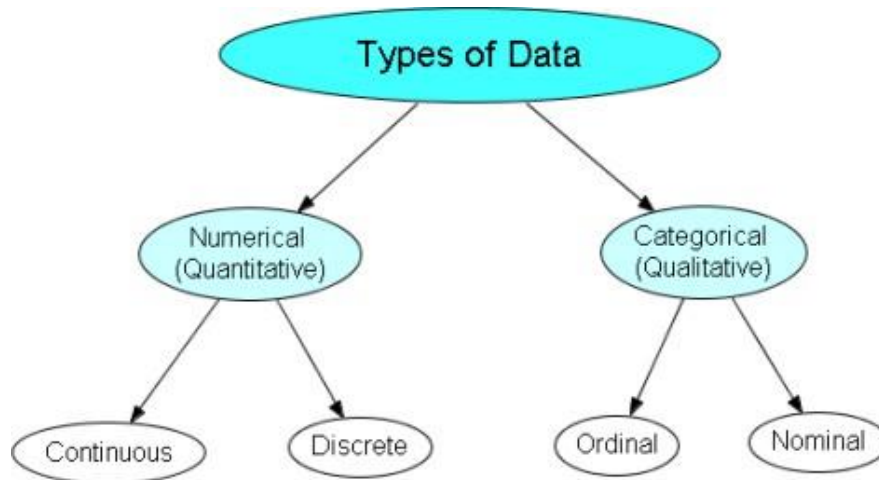


Section 1.2 – Organizing and Displaying Categorical Data

MDM4U

David Chen

Part 1: Types of Variables



Numeric/Quantitative Variable: A variable that takes _____ values for which it makes sense to find an average. These variables can be either _____ or _____.

Continuous Variable: A numeric variable that can have an _____ number of values in a given interval. Measurable with _____.

Examples:

Discrete Variable: A numeric variable that can take on only a _____ number of values within a given range. Usually measured with integer values only.

Examples:

Categorical/Qualitative Variable: A variable that places an individual into one of several _____ or _____. Categorical variables may have categories that are naturally ordered (_____ variables) or have no natural order (_____ variables).

Ordinal Variable: A categorical variable that has a _____ of its possible values, but the distances between the values are undefined.

Example:

Nominal Variable: Type of categorical variable that describes a name, label, or category with **no** _____.

Example:

Part 2: Frequency Tables

To make an accurate picture of data, the first thing we have to do is make 'piles'. For categorical data, 'piling' is easy. We just count the number of cases corresponding to each category. We can organize these counts into a _____, which records the totals and category names.

Frequency tables are used to _____ data.

Example 1:

Grade 12's were asked when their spares were and these were the results:

A, B, C, D, A, D, D, B, A, C, A, C, B, B, B, A, D, C, A, A, B, D, C, A, B
B, A, C, C, D, A, B, A, B, B, B, D, D, A, D, D, C, A, D, C, D, A, B, B, A

The problem with data that is presented like this is that you can't 'see' what is going on. Organize the data in to a frequency table to better see the distribution of data.

Spare	Frequency

Counting the frequency is useful, but sometimes we want to know the _____ of data in each category, so we make a _____.

A relative---frequency table shows the frequency of a data group as a _____ or _____ of the whole data set.

Spare	Frequency	Relative Frequency

Part 3: Bar Graphs

Graphs are used to _____ data. Bar graphs, segmented bar graphs, pie charts, and pictographs are appropriate types of graphs for displaying the data of _____ variables. Bar graphs can also be used for discrete numeric variables.

A bar graph displays the distribution of a categorical variable, showing the counts (frequency) for each category next to each other for easy comparison.

A bar graph consists of parallel bars of equal widths (_____) with lengths proportional to the _____ of the variables they represent.

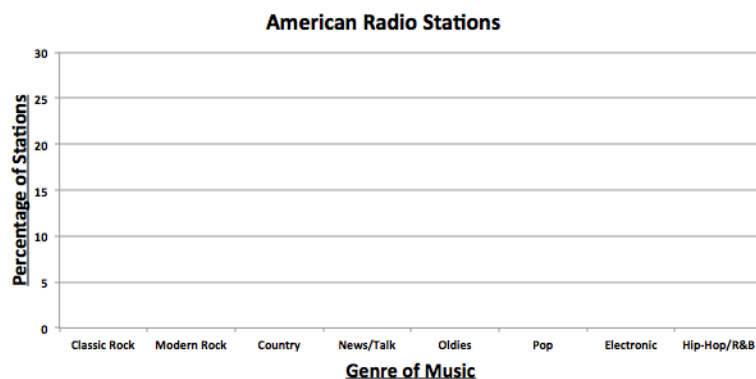
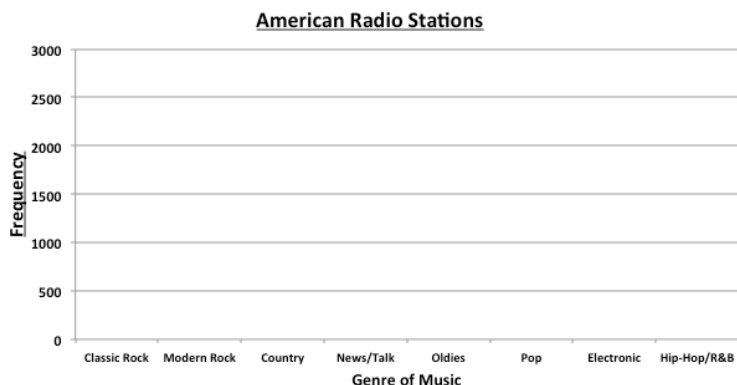
Example 2:

The following frequency table shows the number of different U.S radio stations broken up by category based on the kind of music they broadcast.

I. Complete the relative frequency column

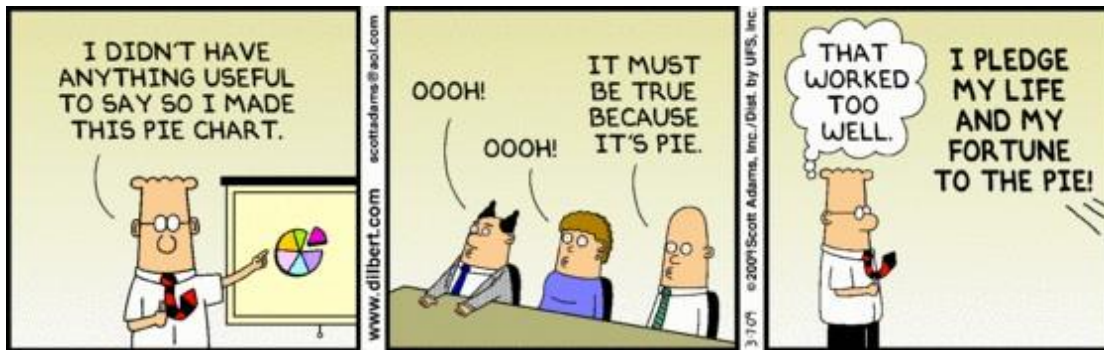
Genre	Frequency	Relative Frequency
Classic Rock	569	
Modern Rock	869	
Country	2066	
News/Talk	2179	
Oldies	906	
Pop	2575	
Electronic	626	
Hip-Hop/R&B	450	
Total	10240	

II. Use the table to create two bar graphs. The first showing frequencies and the second showing relative frequencies of each category.



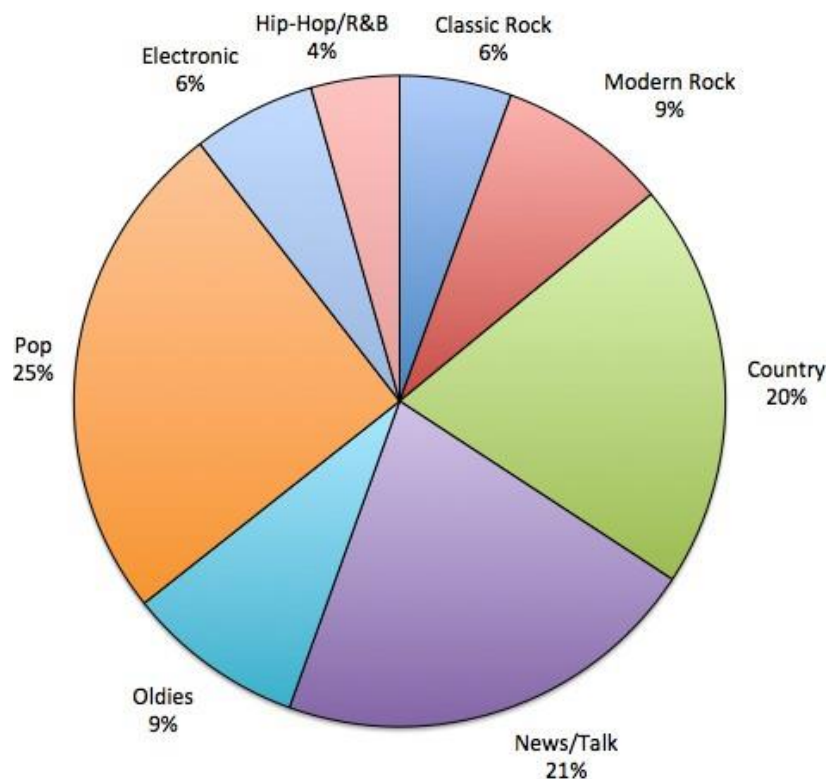
What do you notice about the shapes of the distributions?

Part 4: Pie Charts



A pie chart shows the distribution of a categorical variable as a 'pie'. They slice the circle into pieces whose sizes are _____ to the fraction of the whole in each category. A pie chart is best used when trying to show a category's relation to the whole. Pie charts are awkward to make by hand, but technology will do the job for you.

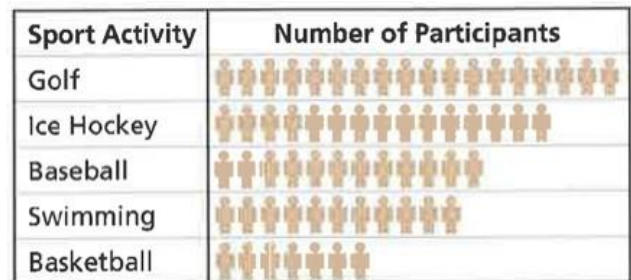
Here is a pie chart showing the data for the U.S radio stations from the previous example:



Part 5: Pictographs

A _____ is a symbolic representation of data. The following pictograph displays the number of participants, aged 15 and older, in the five most popular sports activities in Canada.

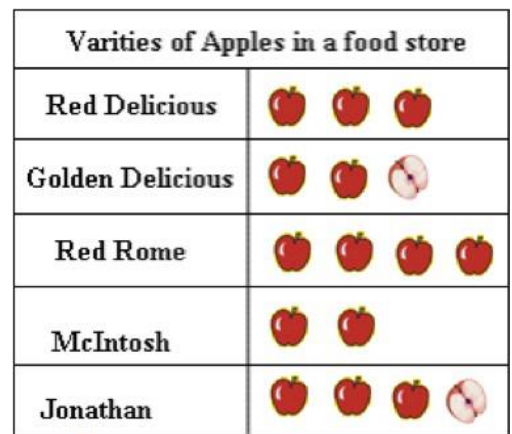
How many people aged 15 and older play hockey?



Legend:  represents 100 000 people

Example 3:

a) How many red delicious apples are in the store?



 = 10 apples  = 5 apples

b) How would you represent 11 apples?

Problems with Pictographs:

- Pictographs can make a graph more interesting but...

Part 6: Contingency Tables and Segmented Bar Graphs

We have learned some techniques for analyzing the distribution of a single categorical variable. If a data set involves two categorical variables, we use a _____. A two---way table of counts organizes data about two categorical variables measured from the same set of individuals.

Example 4: Only 32% of those aboard the Titanic survived. Was that survival rate the same for all ticket classes? To answer that question, we can arrange the counts for the two categorical variables, survival and ticket class, in a two---way table.

Survival	Class				
		First	Second	Third	Crew
	Total	711	1490	2201	
Alive	203	118	178	212	711
Dead	122	167	528	673	1490
Total	325	285	706	885	2201

In this case, survival is our _____ and class is our _____. The margins of the table give totals. When analyzing a contingency table, the goal is to see if the variables depend on each other. This can be done by looking at the two possible _____.

If we think that class might depend survival, then we should look at the distribution of the _____ percentages. This is the conditional distribution for class based on survival.

Survival	Class				
		First	Second	Third	Crew
	Total	711	1490	2201	
Alive	203	118	178	212	711
Dead	122	167	528	673	1490
Total	325	285	706	885	2201

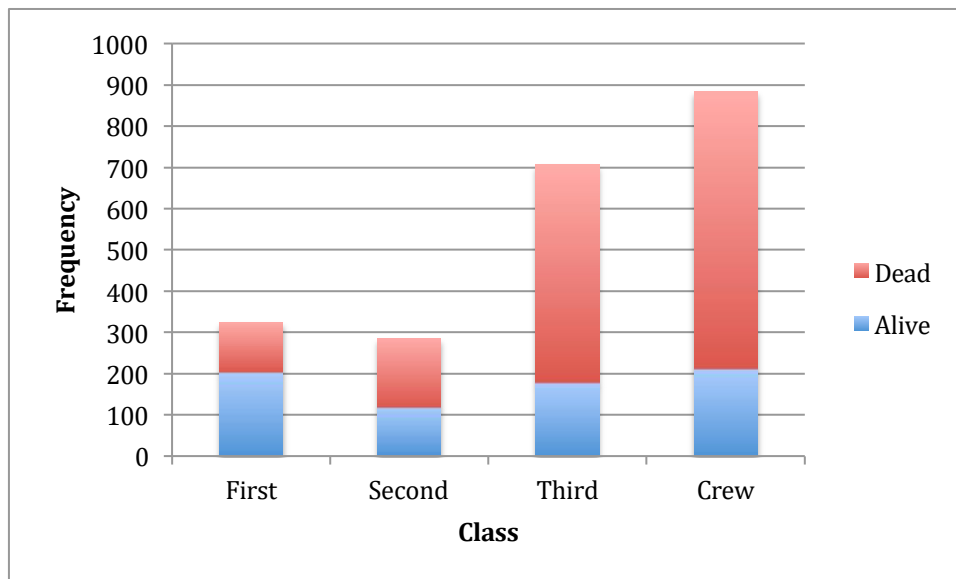
However, in this scenario it would make more sense to determine if survival depends on class. To do this, we should look at the _____ percentages. This is the conditional distribution for survival based on class.

Survival	Class				
		First	Second	Third	Crew
	Total	711	1490	2201	
Alive	Count	203	118	178	212
Dead	Count	122	167	528	673
Total	Count	325	285	706	885
Alive	% of Column	62.5%	41.4%	25.2%	24.0%
Dead	% of Column	37.5%	58.6%	74.8%	76.0%
Total	% of Column	100%	100%	100%	100%

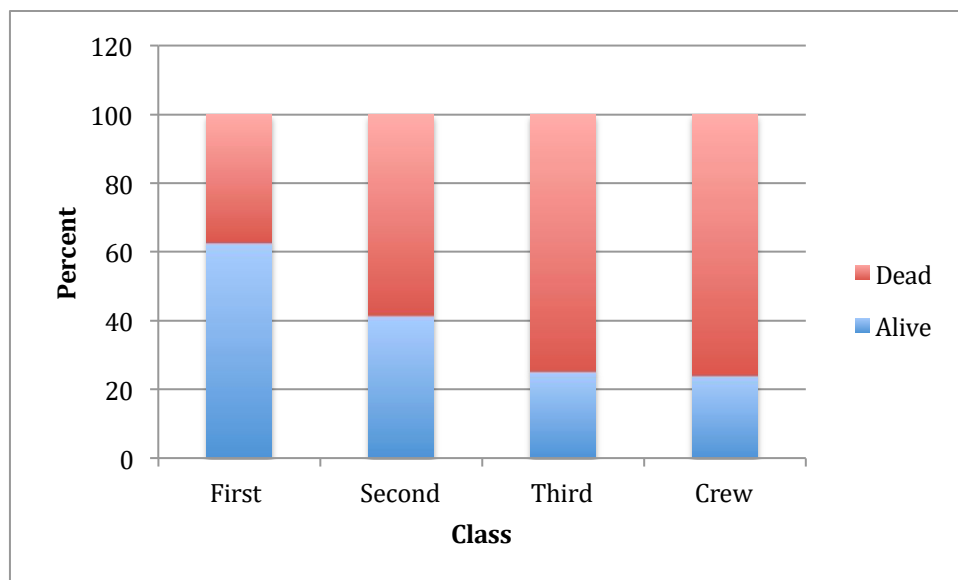
Looking at how the percentages change across the row, it sure seems that class influenced whether a persons survived or not. _____ of first class passengers survived while only _____ of third class passengers survived.

Two---way tables are often displayed using _____.

Example: Segmented bar graph of survival based on class using frequencies



Example: Segmented bar graph of survival based on class using conditional percentages



Note: The bars of each graph have the same proportions but it is easier to see in the second graph that first class passengers had the highest proportion of survivors.