

2.2: Organizing Qualitative Data

Overview: Types of Data

Nominal (categorical): Values are names or categories. Values do not measure the amount of any characteristic. Values do not have a meaningful order. (It does not make sense to rank the values in an order.)

Ordinal: Values can be arranged in a meaningful order, from smallest to largest. A small value indicates the object has less of the characteristic being measured; a large value indicates the object has more of the characteristic being measured.

Interval: Values are ordinal, having a meaningful order. In addition, the difference between values has meaning. In other words, an equal difference between values represents an equal difference in the characteristic being measured. Thus, it is meaningful to add and subtract values. However, there is not a meaningful zero (a value of zero does not indicate the absence of the characteristic being measured).

Ratio: Values are ordinal (have a meaningful order) and interval (have a meaningful difference between values). In addition, the value of zero has meaning (a value of zero indicates absence of the characteristic being measured). Thus, ratios between values have meaning (if one value is three times as large as another value, then it has three times as much of the characteristic being measured). For this reason, it is meaningful to multiply and divide values. (And, because ratio data is also interval data, it is meaningful to add and subtract values.)

Examples of nominal variables: *nominal = categorical*

Names, colors, gender, religion

Examples of ordinal variables: *(have order/rank)*

Movie ratings (G, PG, PG13, R)

Likert Scales (strongly disagree, disagree, neither, agree, strongly agree)

Examples of interval variables:

Temperature

Examples of ratio variables:

weight, height, age, salary, checking account balance, profit

Data is easier to interpret if it is organized into a visual display. Useful displays for qualitative data include tables, bar graphs, and pie charts.

Tables:

A *frequency distribution* lists each category and the number of data points (occurrences) corresponding to that category. When arranged in the form of table, this is called a *frequency table*.

Tables often include the *relative frequency*.

The *relative frequency* of a category is the proportion (or percentage) of the total observations that fall within that category.

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}} = \frac{\text{Frequency}}{\text{Total \# of data points}}$$

A *relative frequency distribution* lists the categories along with their relative frequencies.

Example 1:

2.26 Primetime Broadcast Shows. From the **TVbytheNumbers** website, we obtained the networks for the top 20 primetime broadcast TV shows by total viewership for the week ending August 18, 2013.

CBS	NBC	NBC	ABC	CBS
CBS	CBS	CBS	CBS	FOX
CBS	CBS	CBS	CBS	NBC
NBC	FOX	CBS	NBC	ABC



Variable = network name
(qualitative)

Values of the variable: CBS, NBC, etc

Network	Frequency	Relative Frequency
CBS	11	$11/20 = 0.55$
NBC	5	$5/20 = 0.25$
ABC	2	$2/20 = 0.10$
FOX	2	$2/20 = 0.10$
	<u>$n = 20$</u>	<u>Sum = 1.00</u>

Bar charts (bar graphs):

A *bar graph* is a visual display in which the category names are along the vertical or horizontal axis, and the frequencies (or relative frequencies) are on the other axis.

Note: Category name = value of the nominal variable

In a *side-by-side bar graph*, the relative frequencies for values of the nominal variable are simultaneously displayed for two or more subgroups (or for two or more years, or for two or more values of some other variable).

Example 2:

2.32 Robbery Locations. The **Department of Justice** and the **Federal Bureau of Investigation** publish a compilation on crime statistics for the United States in *Crime in the United States*. The following table provides a frequency distribution for robbery type during a one-year period.

Robbery type	Frequency
Street/highway	127,403
Commercial house	37,885
Gas or service station	7,009
Convenience store	14,863
Residence	49,361
Bank	5,777
Miscellaneous	48,878

$$n = 291,176$$

Relative Frequency

Percentage
43.7%

$$\frac{127403}{291176} \approx 0.438$$

$$0.130$$

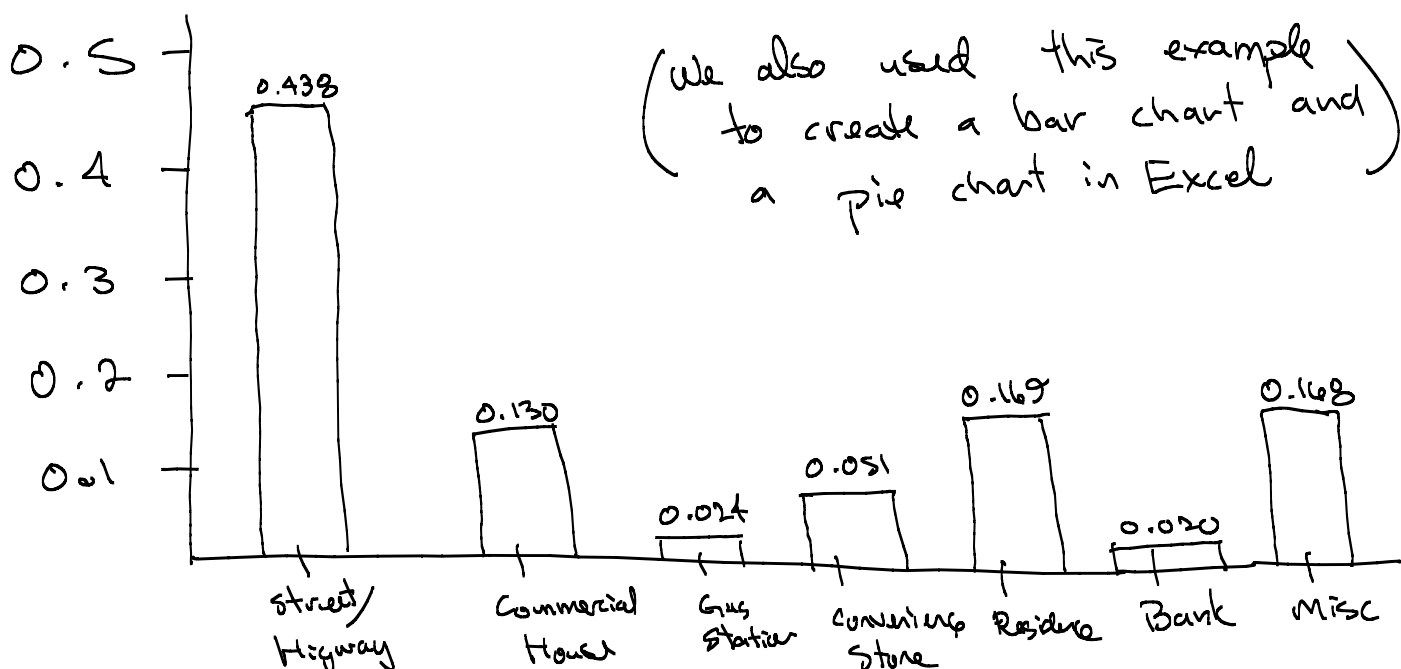
$$0.024$$

$$0.051$$

$$0.169$$

$$0.020$$

$$0.168$$



Pie charts:

A pie chart is a circle divided into sectors, in which each sector represents a category. For each category, the relative frequency is equal to the ratio of the sector area to the total circle area. (In other words, if one-third of the observations are in a given category, then one-third of the circle area will be in the sector corresponding to that category.)

Example 3:

2.34 Freshmen Politics. The Higher Education Research Institute of the University of California, Los Angeles, publishes information on characteristics of incoming college freshmen in *The American Freshman*. In 2000, 27.7% of incoming freshmen characterized their political views as liberal, 51.9% as moderate, and 20.4% as conservative. For this year, a random sample of 500 incoming college freshmen yielded the following frequency distribution for political views.

Political view	Frequency
Liberal	147
Moderate	237
Conservative	116