Sullivan Section 2.1

Gene Quinn

Organizing Qualitative Data

qualitative data consists of *nonnumerical* measures that characterize of classify individuals.

Organizing Qualitative Data

qualitative data consists of *nonnumerical* measures that characterize of classify individuals.

By *nonnumerical* measures, we mean measures for which arithmetic operations make no sense.

Organizing Qualitative Data

qualitative data consists of *nonnumerical* measures that characterize of classify individuals.

By *nonnumerical* measures, we mean measures for which arithmetic operations make no sense.

So a measure like student ID number, even though it consists of numbers, would be considered nonnumerical because adding two student id numbers would not be meaningful.

Some characteristics like gender are clearly qualitative.

Frequency

A **frequency distribution** lists categories of data along with the *number* of occurrences in each category.

Frequency

A **frequency distribution** lists categories of data along with the *number* of occurrences in each category.

Each category is paired with a number that represents a **count** of individuals who fall into that category.

Frequency

A **frequency distribution** lists categories of data along with the *number* of occurrences in each category.

Each category is paired with a number that represents a **count** of individuals who fall into that category.

Example: Age distribution in a course:

Age	Count
17	8
18	12
19	5

Frequency

Example: Years of Formal Education of Adults in a Town

Years	Count
$0-6$	1476
$7-9$	3413
$10-11$	14282
12 or more	57345

Relative Frequency

A **relative frequency** is the proportion or percent of observations that fall within a given category, and is given by the formula:

$$\text{Relative frequence} = \frac{\text{frequency}}{\text{sum of all frequencies}}$$

Relative Frequency

A **relative frequency** is the proportion or percent of observations that fall within a given category, and is given by the formula:

$$\text{Relative frequence} = \frac{\text{frequency}}{\text{sum of all frequencies}}$$

A **relative frequency distribution** lists the relative frequencies for each category of data.

Relative Frequency

Example: Age distribution in a course:

Age	Relative Frequency
17	0.32
18	0.48
19	0.20

Relative Frequency Example

We calculated the relative frequencies by applying the formula cell by cell, using the fact that the total for all categories is 25.

So, the value *Relative Frequency* column is the value in the *Count* column divided by 25.

Age	Count	Relative Frequency
17	8	0.32
18	12	0.48
19	5	0.20

Bar Graphs

We have seen that frequency data and relative frequency data can be represented in a natural way as **tables**:

Years	Count
$0-6$	1476
$7-9$	3413
$10-11$	14282
12 or more	57345

Bar Graphs

We have seen that frequency data and relative frequency data can be represented in a natural way as **tables**:

Years	Count
$0-6$	1476
$7-9$	3413
$10-11$	14282
12 or more	57345

An alternative way to represent qualitative data is **graphically**

Bar Graphs

A **bar graph** is a graphical representation constructed by drawing rectangles whose height is proportional to the frequency or relative frequency of each category.

Bar Graphs

A **bar graph** is a graphical representation constructed by drawing rectangles whose height is proportional to the frequency or relative frequency of each category.

Bar graphs are one of the most common ways to represent qualitative data graphically.

Pareto Charts

A bar graph whose bars are drawn in order of decreasing frequency or relative frequency is called a **Pareto chart**.

Pareto Charts

A bar graph whose bars are drawn in order of decreasing frequency or relative frequency is called a **Pareto chart**.

You can convert any bar graph to a Pareto chart by arranging the bars in decreasing order.

Side-by-Side Bar Graphs

A **side-by-side** bar graph is a variation on the idea of a bar graph that can be used to compare two sets of data.

Suppose our age table contained data from two classes instead of one:

Age	MA145A	MA145B
17	8	6
18	12	17
19	5	2

Side-by-Side Bar Graphs

A **side-by-side** bar graph is a variation on the idea of a bar graph that can be used to compare two sets of data.

Suppose our age table contained data from two classes instead of one:

Age	MA145A	MA145B
17	8	6
18	12	17
19	5	2

This data could be displayed in a side-by-side bar graph.

Pie Charts

A **pie chart** is a circle divided into sectors, with each sector representing a category of data.

Pie Charts

A **pie chart** is a circle divided into sectors, with each sector representing a category of data.

The area of each sector is proportional to the frequency (or relative frequency) of the category.

Pie Charts

A **pie chart** is a circle divided into sectors, with each sector representing a category of data.

The area of each sector is proportional to the frequency (or relative frequency) of the category.

The *angle* subtended by each sector, in degrees, is 360 times the relative frequency for the category corresponding to that sector.