Graphical Displays of Data

Adam G. Fera

February 9, 2018

Frequency Distributions: A Chart of a Different Color

- This is the first step in analyzing data.
- Summarizes data visually to give us a sense of how the data are distributed.
- Check data for coding errors
- Most common types of frequency tables are:
  - Frequency
  - Percentage
  - Combination

Frequency Distributions

A Frequency Table should have 3 BASIC ELEMENTS

1. The complete set of levels or values that define the variable
2. A record of the frequency (number of individuals) in each level or value associated with the variable.
3. Percentage of individuals that fall into each level or value of the variable.

Conventions for Building Distributions

- Title or distribution explains contents
- Variable(s) in table
- Shows data distribution
- Categories in stub of distribution
- Use column headings
- Relevant columns are totaled
- Footnotes added if needed

Frequency Distributions

Created by:

- Listing category
- Tallying number of observations that fit

Scores:

1,1,2,2,2,2,2,3,3,3,3,4,4,5

Score	Frequency
1	2
2	5
3	4
4	2
5	1

Percentage Distributions (PD)

- Created by:
  - Dividing cell frequency by total number of cases and multiplying by 100
- Allows comparisons between similar data sets of different sizes.

Percentage Distributions (PD)

In Addition to the Frequency of each value, researchers want percentages:

- Percent is calculated by using the frequency divided by the total INCLUDING THE MISSING.
- Valid Percent is calculated by using the frequency divided by the total EXCLUDING THE MISSING.
  - If there are missing data, the percent and valid percent columns will be different. If there are no missing data, both columns will match perfectly.
  - When there are missing data, the frequency table will have to values for n. One with the missing data and one without the missing data.

Percentage Distributions Continued

- Cumulative Percent is calculated by summing the valid percent contribution to "n" for the category and above.

Frequency Distribution

A Typical frequency distribution table for the variable "residential area"

1. All the levels of residential area (Urban, Suburban, and Rural) that define the variable are in the first column.
2. The freqeucny associated with each level are in the second column.
    - The number of times each value occurred in the data
3. The percentage assciated with each level is in the third column.
4. The Valid percent is in the fourth column.
5. The Cumulative percent is in the fifth column.

Frequency Table Example

Residential Area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Urban	123	24.6	24.6	24.6
	Suburban	215	43.0	43.0	67.6
	Rural	162	32.4	32.4	100.0
	Total	500	100.0	100.0	

R Example

- Read in the Rikers 1989 data set.
- Run frequencies for the following variables using the `table()` function:
  - SEX [*sex*]
  - DEF FIRST ARREST? [*firar*]

Graphical Representation of Freqeuncies

- Pie charts
- Bar charts
- Histograms

Pie Charts

- Most basic graph type
- Best for nominal data with few categories
- Depicts N
  - Relative size of categories is displayed in terms of "slices".
- **Only** use with:
  - Nominal data with few categories
  - Ordinal data with few categories

Pie Chart Example

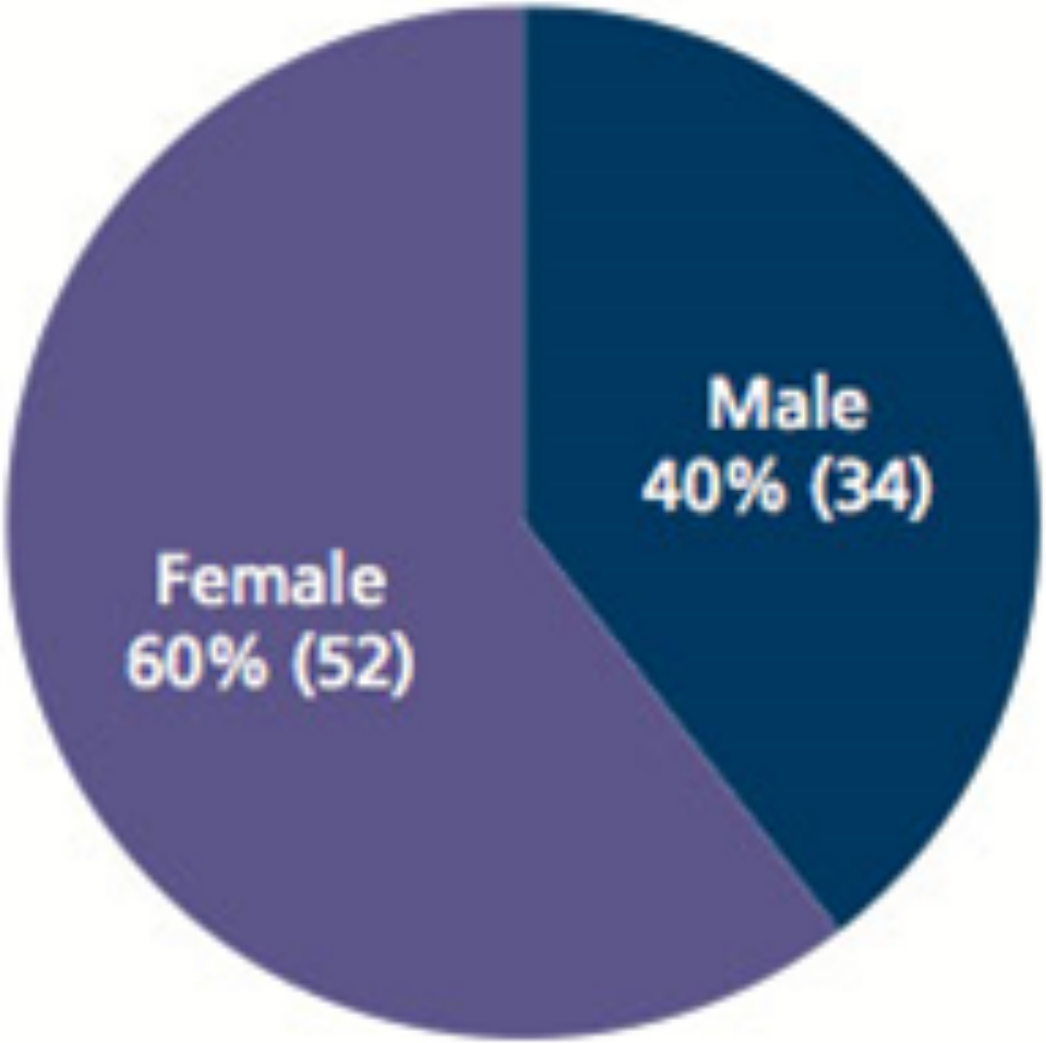


Figure: Pie Chart

R Example

- Read in the Rikers 1989 data set.
- Create a pie chart for the following variables:
  - SEX [*sex*] and
  - DEF FIRST ARREST? [*firar*]

Histograms and Bar Charts

- Histograms and bar charts both use the length of bars to display the relative size of categories.
- Represent frequency distribution plot of:
  - Categories/variable on one axis (usually X)
  - Responses as bars on the other axis (usually Y)
- Bar length can represent the frequency or percentage of that category.

Histograms and Bar Charts

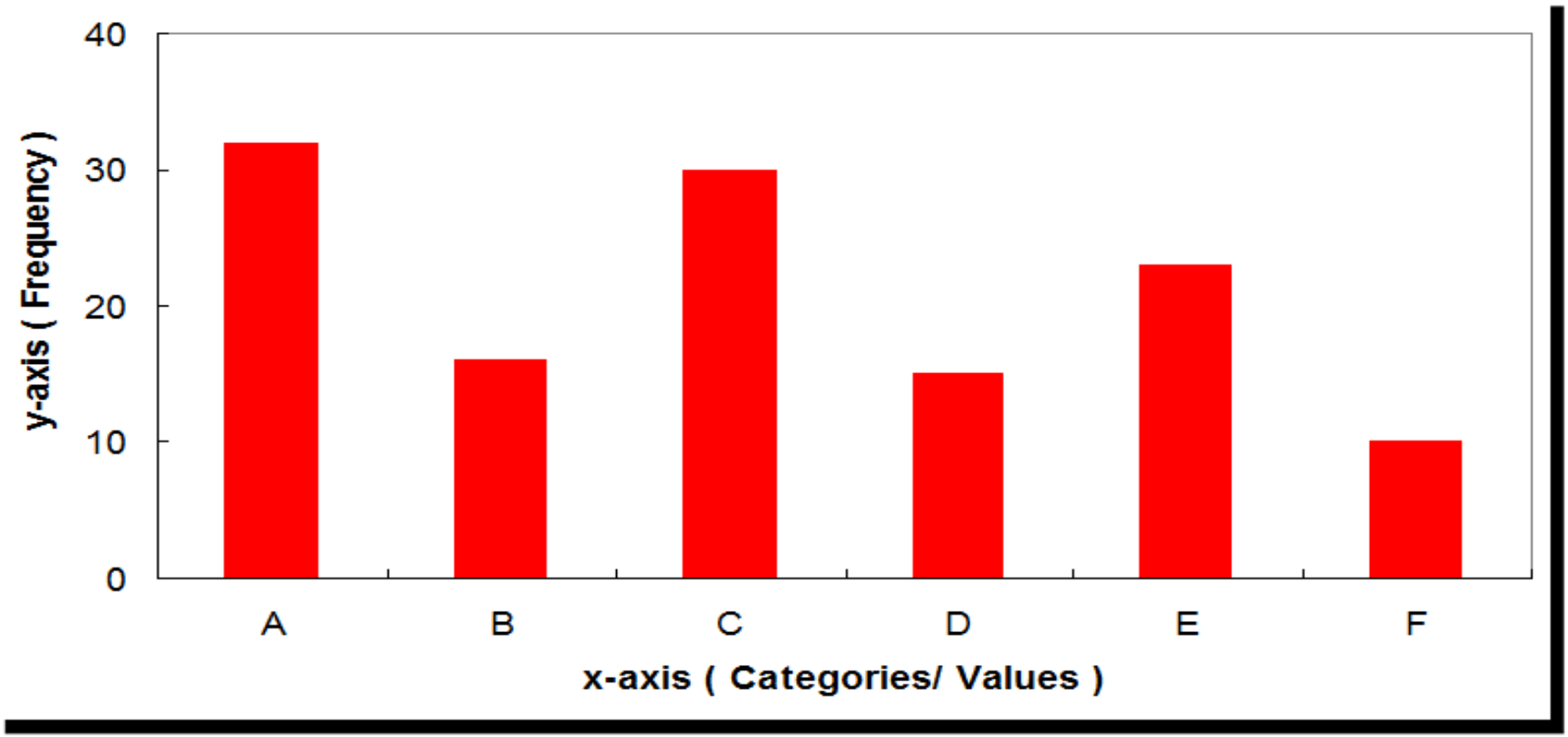
When constructing:

- Y axis starts at Zero and uses equal intervals to the maximum.
- X axis generally starts at Zero and is
- Use proper labeling, legend, and footnotes
- Build on equal class intervals (if grouped)
- Total frequency should equal "N"

Bar Charts

- Bar charts are used with nominal or ordinal level data
- Bars of a Bar Chart are separated
  - Shows category/variable division
- Separated bars show lack of continuity
- Can be oriented horizontally or vertically
- Nominal variables often placed in horizontal bar chart for emphasis

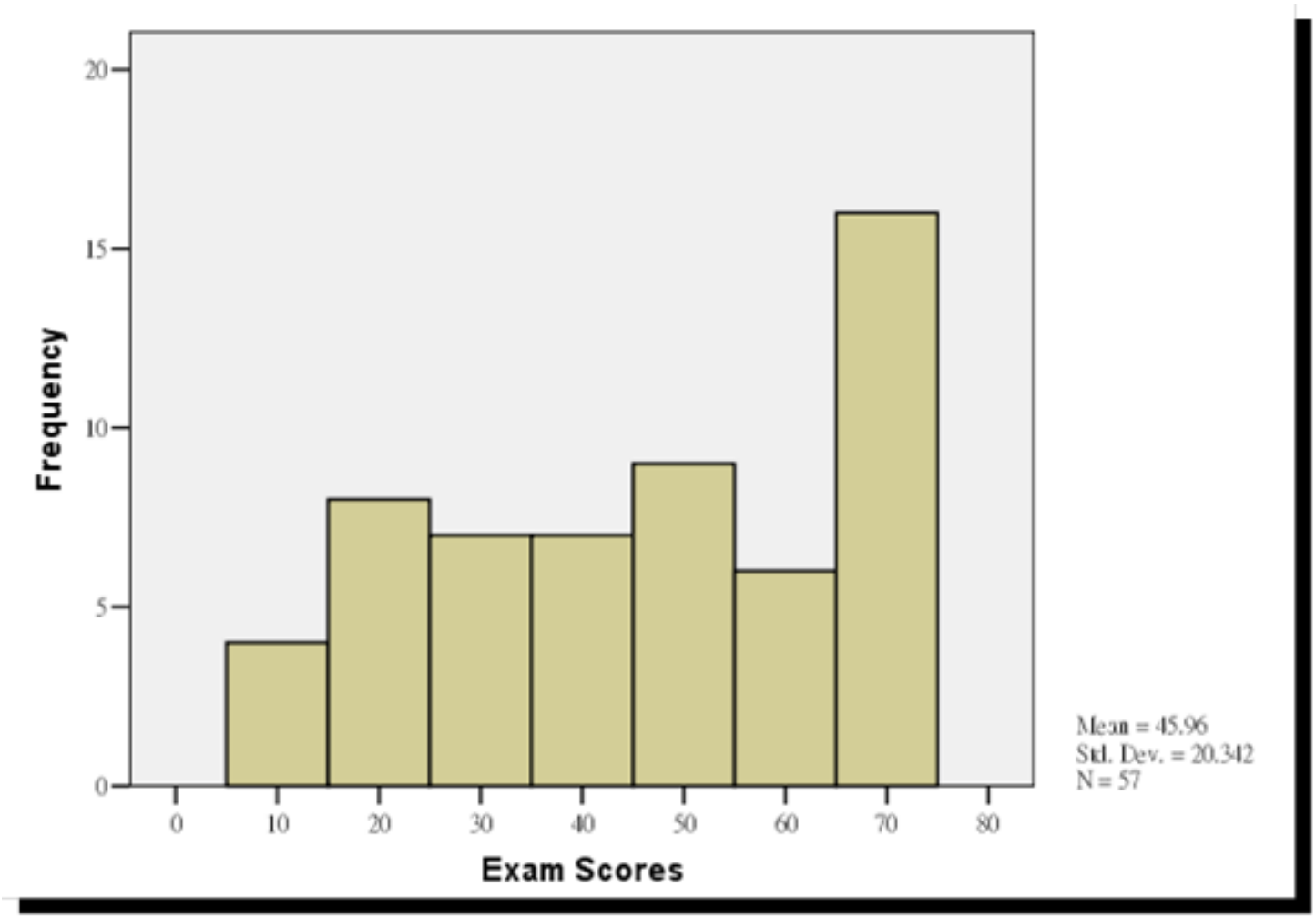
Bar Chart Example



Histograms

- Histograms are used with: Interval and Ratio level data
- Bars of the histogram are connected showing the continuity of the data

Histogram Example



R Example

- Read in the Rikers 1989 data set.
- Create a bar chart for sex and defendant's first arrest:
  - SEX [*sex*] and
  - DEF FIRST ARREST [*firar*]
- Create a histogram for defendant's age:
  - Age [*age*]

Proportions, Rates, and Percentages

To address the problem of unequal group sizes, we can convert the raw frequencies to:

1. Proportions
2. Rates
3. Percentages

Proportions

Proportions represent the fraction of all observations that fit in a category and have a value between 0 and 1.

- To calculate a proportion, divide the number of the obeservations in the category by the total number of observations.
- For example, if there are ten people and four of them are men, to find the proportion of men in the group, divide four by ten.
- $\frac{N_{category}}{N_{total}} = \frac{Men}{Total\ People} = \frac{4}{10} = 0.40$

Rates

Rates represent the number of observations of a category per unit population.

- To calculate a rate, multiply the proportion by the desired unit population.
- For example, to calculate the rate of men per 100,000 population, multiply the proportion (0.40) by the desired unit population (100,000).
- $\textit{Proportion} * \textit{Unit population} = 0.40 * 100,000 = 40,000 \textit{ per } 100,000$

Percentages

Percentages are a special type of rate and represent the number of observations of a category per unit population of 100.

- To calculate a percentage, multiply the proportion by 100 as your unit population.
- For example, to calculate the percentage of men, multiply the proportion (0.40) by the desired unit population (100).
- $Proportion * Unit\ population = 0.40 * 100 =$ 40 *per* 100 *or* 40%

Conclusion

- Graphical representation can be an effective way to examine, interpret, and communicate data.
- Frequency distributions show how data are arranged
- Graphical representation should not replace interpretation and analysis
- Charts need supporting discussion