

Section 1.1

Displaying Distributions with Graphs

- **Issue:** How do we examine and describe the main features of a data set?
- We do *exploratory data analysis*. It involves
 - examining each variable
 - examining relationships between variables
- This is accomplished using
 - graphs (today)
 - numerical summaries (Friday)

Types of Graphs:

Categorical Variables:

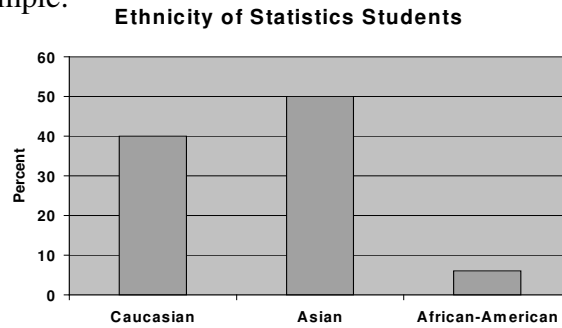
- 1) Bar Graphs
- 2) Pie Charts

Quantitative Variables:

- 1) Histograms
- 2) Stemplots
- 3) Time Plots
- 4) Box Plots (Monday)
- 5) Dot Plots (Monday)

Bar Charts

- Give percents or counts in different categories
- Example:

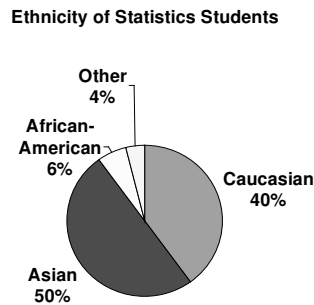


Statistics 528 - Lecture 2

3

Pie Charts

- Give percentages in different categories
- Must include all categories that make up a whole
- Example:



Statistics 528 - Lecture 2

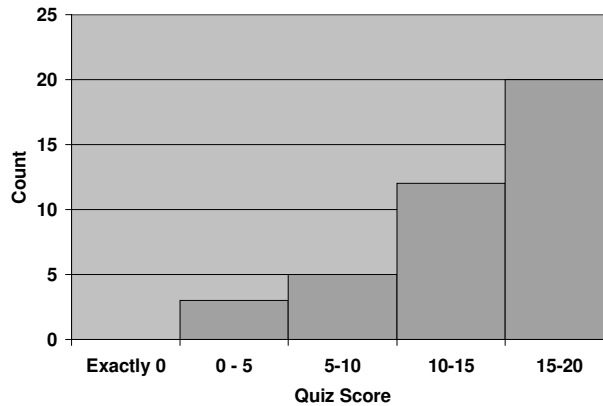
4

Histograms

- A histogram breaks up *quantitative* variables into intervals and displays the count or percentage of the observations that fall into each interval.
- Useful for large datasets
- Example:

Score	Count	Percent
Exactly 0	0	0%
0-5	3	$3/40 = 7.5\%$
5-10	5	$5/40 = 12.5\%$
10-15	12	30%
15-20	20	50%
Total	40	100%

Note: Each class interval includes the right endpoint but not the left endpoint of the interval.
 Statistics 528 - Lecture 2 5



Statistics 528 - Lecture 2

6

-
1. Histograms look like bar charts without the gaps between bars.
 2. If frequencies (same as counts) are used and the intervals are all of the same width, then the sum of the heights of the bars must equal the total number of observations. This is called a **frequency histogram**.
 3. If percents are used, then the sum of the percents of all the bars must equal 100%. This is called a **relative frequency histogram**.

Determining the appropriate number of bins: describing the distribution of Poplar Clone 252 heights - 8 bins (default) vs. 3 bins.

Stemplots

- Good for small - medium size data sets
- Advantage (over histograms): can recreate the original data set using them

To make a stemplot:

- 1) ****Sort the data (smallest to largest)**** Separate each observation into a **leaf** which is the final digit, and the **stem** which consists of the remaining digits.
- 2) Write the stems in a column in consecutive order with the smallest at the top, and draw a vertical line to the right of this column.
- 3) Write each leaf in the row to the right of its stem with the leaves in consecutive order from smallest to smallest to largest.

Example: Initial blood pressure readings are taken for a group of subjects taking part in an experiment to see if calcium in the diet lowers blood pressure.

Data:

Calcium Group

112 111 107 136 107 112 102 123 129 110

Placebo Group

117 123 112 110 112 109 114 98 119 102 130

Sorted Calcium Data: 102, 107, 107, 110, 111, 112, 112, 123, 129, 136

Stemplot:

10	 	2 7 7
11	 	0 1 2 2
12	 	3 9
13	 	6

Back-to-back stemplots are useful to compare two sets of data.

Distribution of Calcium and Placebo Groups

Placebo		Calcium
8	9	
9 7	10	2 7 7
9 7 4 2 2 0	11	0 1 2 2
3	12	3 9
0	13	6

Time Plots

- Plot of a variable against the time it was measured.
- Useful for highlighting:
 - 1) Seasonal Variation - pattern that repeat at regular intervals
 - 2) Trend - persistent, long-term rise or fall

Example:

