

Chapter 2: Summarizing and Graphing Data

More Basic Terms

Raw data --- numbers and category labels that are collected, but not yet processed

Variable --- a characteristic that differs from one individual to the next

Observational unit (observation) --- single individual who participates in a study

1

Explanatory and Response Variables

Many questions are about the **relationship** between *two variables*.

It is useful to identify one variable as the **independent variable (explanatory variable, predictor, covariate)** and the other variable as the **dependent variable (response variable)**.

Generally, the *value of the independent variable* for an individual is thought to **partially explain** the *value of the dependent variable* for that individual.

2

Describing Qualitative Data

Class---a category into which qualitative data can be classified

Class frequency---number of observations in the data set falling in a given class

Class relative frequency---class frequency divided by the total number of observations in the data set

$$\text{class relative frequency} = \frac{\text{class frequency}}{n}$$

3

Describing Qualitative Data

Numerical summaries for one or two categorical variables

Count how many fall into each category.

Calculate the percent in each category.

4

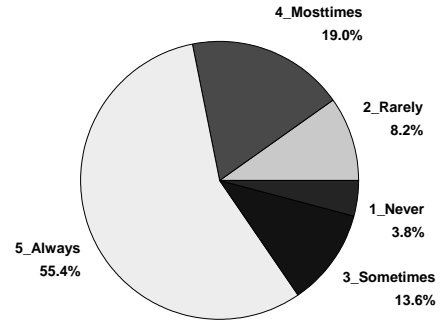
Table 2.1 Seatbelt Use by Twelfth-Graders When Driving

Response	Count	Percent
Always	1686	55.4%
Most times	578	19.0%
Sometimes	414	13.6%
Rarely	249	8.2%
Never	115	3.8%
Total	3042	100%

©2008 Thomson Higher Education

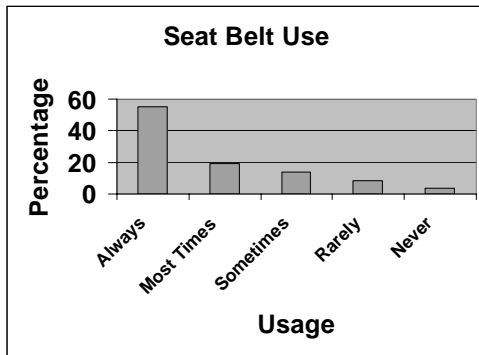
5

Describing qualitative data



6

Describing qualitative data—Bar Graph



7

If working with two variables, have the categories of the explanatory variable define the rows and compute row percentages.

Table 2.2 Gender and Seatbelt Use by Twelfth-Graders When Driving

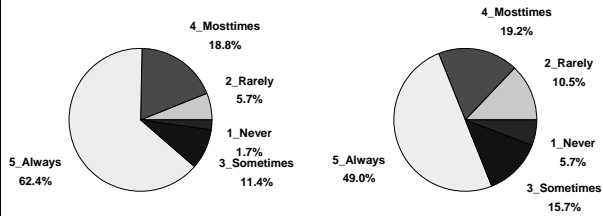
	Always	Most Times	Sometimes	Rarely	Never	Total
Female	915 (62.4%)	276 (18.8%)	167 (11.4%)	84 (5.7%)	25 (1.7%)	1467 (100%)
Male	771 (49.0%)	302 (19.2%)	247 (15.7%)	165 (10.5%)	90 (5.7%)	1575 (100%)

©2008 Thomson Higher Education

8

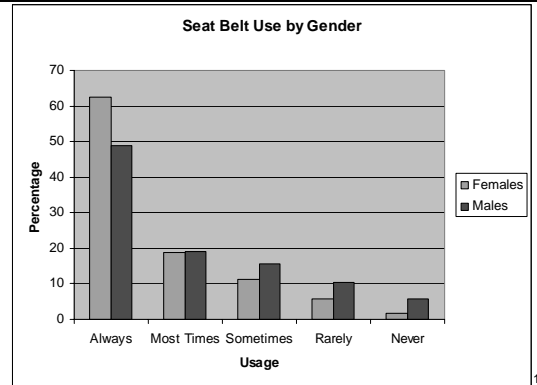
Describing qualitative data

Teen Seatbelt Usage: Females (left), Males (right)



9

Describing qualitative data—Bar Graph



10

Describing Quantitative Data

Example: 111 air temperature readings

67	72	74	62	65	59	61	69	66	68	58
64	66	57	68	62	59	73	61	61	67	81
79	76	82	90	87	82	77	72	65	73	76
84	85	81	83	83	88	92	92	89	73	81
80	81	82	84	87	85	74	86	85	82	86
88	86	83	81	81	81	82	89	90	90	68
86	82	80	77	79	76	78	78	77	72	79
81	86	97	94	96	94	91	92	93	93	87
84	80	78	75	73	81	76	77	71	71	78
67	76	68	82	64	71	81	69	63	70	75
76										

11

Describing Quantitative Data

Stem-and-leaf plots

5 : 7899

6 : 11122344

6 : 5566777888899

7 : 0111222333344

7 : 5566666677778888999

8 : 00011111111112222222333444

8 : 555666667778899

9 : 00012223344

9 : 67

In this example the stems correspond to the values of the 10's digits; the leafs are the values of the 1's digits

12

Describing Quantitative Data

5 : 7	7 : 6666667777	
5 : 899	7 : 8888999	
6 : 111	8 : 0001111111111	
6 : 223	8 : 222222333	
6 : 4455	8 : 444555	
6 : 66777	8 : 66666777	
6 : 888899	9 : 0001	• Data arranged in ascending order
7 : 0111	9 : 22233	
7 : 2223333	9 : 44	• Easy to identify individual measurements
7 : 4455	9 : 67	

13

Describing Quantitative Data

Histograms

- x -axis divided into intervals (best to use equal class/interval sizes); between 6 and 15 intervals is a good number
- y -axis gives the frequency (count) or relative frequency of the measurements that fall into each interval
 - Draw a bar with corresponding height
 - Decide rule to use for values that fall on the border between two intervals

14

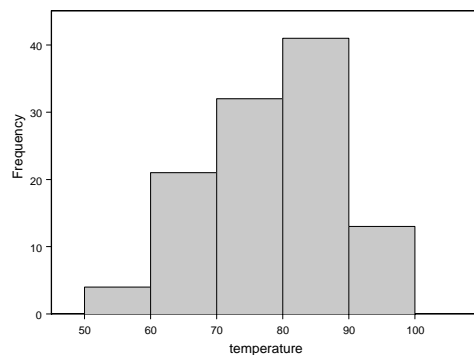
Describing Quantitative Data

Histograms (continued)

- The proportion of total area under the histogram that falls above a particular interval on the x -axis equals the relative frequency of measurements contained in the interval
- Cannot identify individual measurements

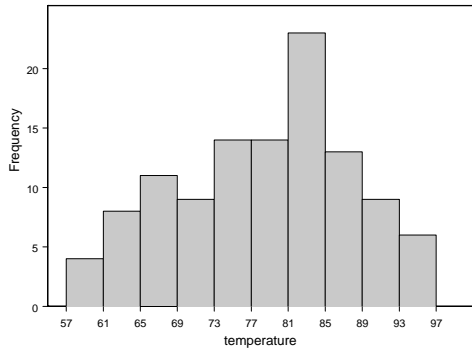
15

Describing Quantitative Data--Histogram

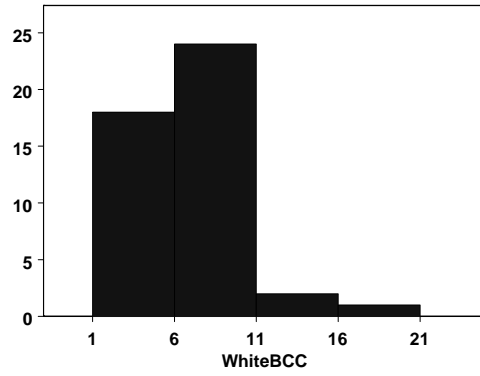


16

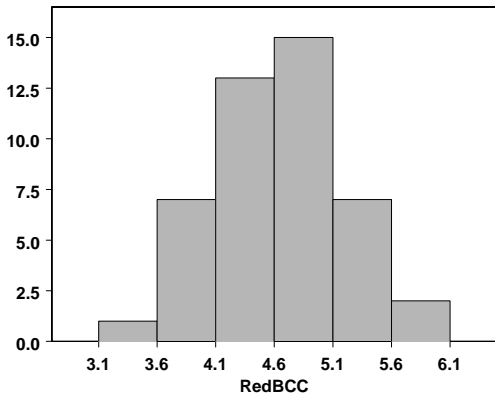
Describing Quantitative Data--Histogram



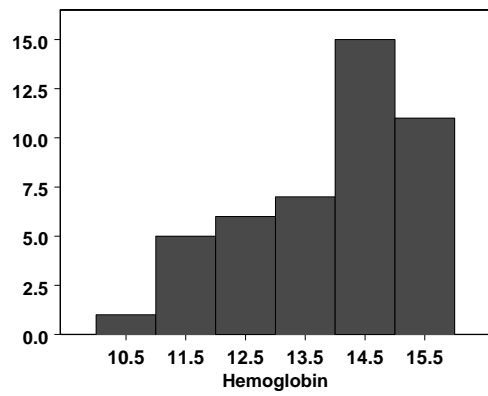
17



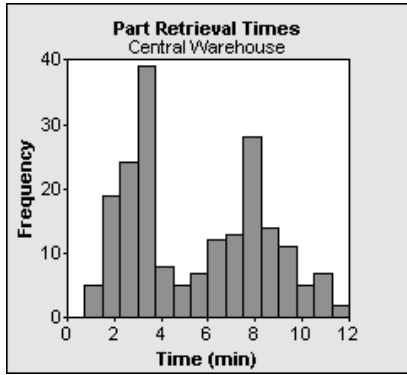
18



19

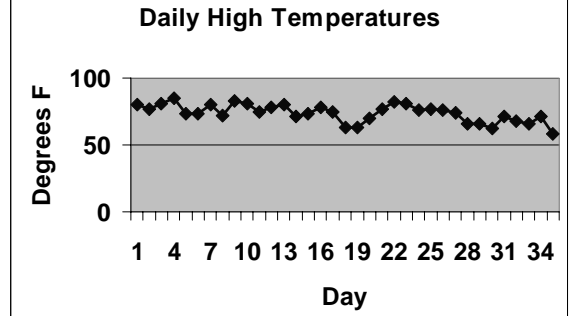


20



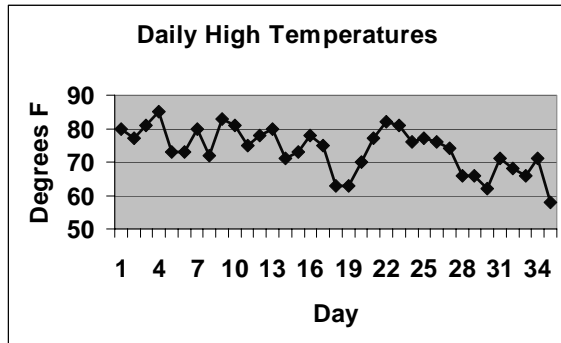
http://www.statit.com/support/quality_practice_tips/histograms_with_two_or_more_peak.shtml 21

Describing Quantitative Data—Time Series



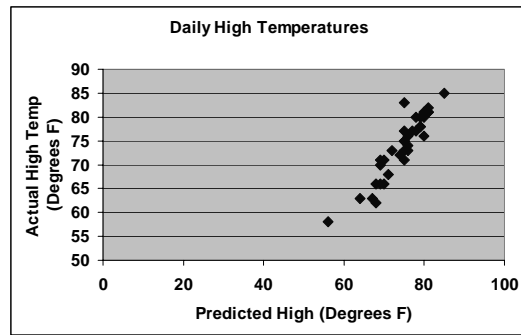
22

Describing Quantitative Data—Time Series



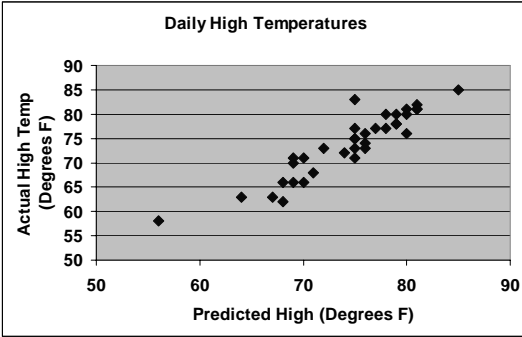
23

Describing Quantitative Data—Scatterplot



24

Describing Quantitative Data—Scatterplot



25