General Education Statistics
Class Notes
Organizing Quantitative Data (Section 2.2)

We will see many different kinds of displays here meant to summarize the data. We will use a worksheet to cover many types of displays. We will learn how to read them as well.

Discrete or Continuous Data:

The first step in summarizing quantitative data is to determine whether the data are discrete or continuous.

If the data are discrete and there are relatively few different values of the variable, the categories of data (**classes**) will be the observations (as in qualitative data).

If the data are discrete, but there are many different values of the variables, or if the data are continuous, the categories of data (the **classes**) must be created using intervals of numbers.

Definition: A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same and *the rectangles touch each other*.

Definition: A **stem-and-leaf plot** uses digits to the left of the rightmost digit to form the **stem.** Each rightmost digit forms a **leaf**. We see this on this section's worksheet.

For example, a data value of 147 would have 14 as the stem and 7 as the leaf. Another example using the number 15.6 would use 15 as the stem and 6 as the leaf.

Definition: A **dot plot** is drawn by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed. It will look similar to a bar graph.

Definition: Time series data: If the value of a variable is measured at different points in time, the data are referred to as **time series data**.

A **time-series graph** is obtained by plotting the time in which a variable is measured on the horizontal axis and the corresponding value of the variable on the vertical axis. Line segments are then drawn connecting the points. The term **line graph** is also used to describe this graph.

1

EXAMPLE Drawing a Histogram for Discrete Data

Suppose we asked 50 households how many cars they had. Draw a frequency histogram for this "number of cars per household" data.

# of Cars	Frequency	Relative Frequency
0	4	$\frac{4}{50} = 0.08$
1	13	$\frac{13}{50} = 0.26$
2	22	0.44
3	7	0.14
4	3	0.06
5	1	0.02

expl 1: Histograms for discrete data are like bar graphs but the bars *will* touch each other. Also, we will draw each bar so that it is centered above the value on the axis. Because of this, we place 0 cars to the right of the vertical axis, not where you would expect it on the Cartesian plane.

Let's draw this histogram. Start with two labeled axes and a title. Your horizontal axis should be "Number of cars" with "Frequency" on the vertical axis.
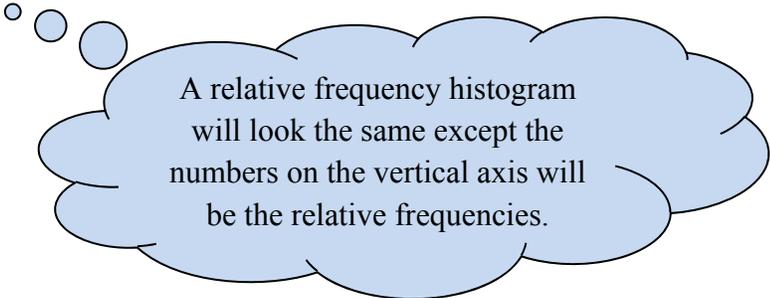
A relative frequency histogram will look the same except the numbers on the vertical axis will be the relative frequencies.

Displaying Continuous Data:

Definitions: The **lower class limit** of a class is the smallest value within the class while the **upper class limit** of a class is the largest value within the class. For the example below, the lower class limit of the first class is 25. The lower class limit of the second class is 35. The upper class limit of the first class is 34. Notice the classes do not overlap but encompass all possible values.

The **class width** is the difference between consecutive lower class limits. The class width of the data is $35 - 25 = 10$.

Number of Persons Aged 25-64 Who Are Currently Work-disabled	
Age	**Number (in thousands)**
25-34	2,132
35-44	3,928
45-54	4,532
55-64	5,108
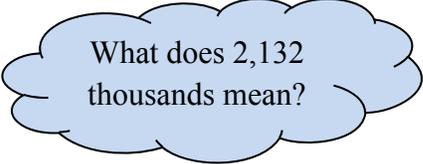
What does 2,132 thousands mean?

You may have to determine the classes for your data. Here are some guidelines.

1. Determine how many classes you want. This amounts to how many bars your histogram will have. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer classes you should have.

2. Determine the class width by computing the following and then rounding up.

$$\text{Class width} = \frac{\text{largest data value } - \text{ smallest data value}}{\text{number of classes}}$$
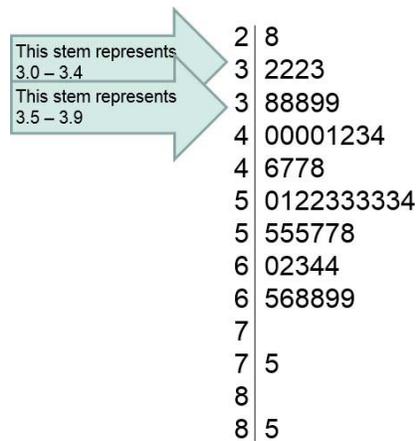
3. Choose the Lower Class Limit of the First Class. You should choose the smallest observation in the data set or a convenient number slightly lower than the smallest observation.

4. Define the Lower Class Limits of the Subsequent Classes starting with the Lower Class Limit of the First Class and adding the Class Width. Do this repeatedly until you have the Lower Class Limits for all Classes.

Worksheet: Data Displays:
This worksheet will review pie charts and bar graphs. It will then lead us through line graphs (also called time-series graphs), histograms, stem-and-leaf plots, and data tables. We will also explore adjusting the scale of a line graph to change the meaning of it. We will briefly look at the concepts of outlier and the distribution of the data.

Stem-and-leaf Plots with Split Stems:
When data appear rather bunched, we can use split stems. The stem-and-leaf plot shown below is an example. As with the determination of class intervals in the creation of frequency histograms, judgment plays a major role. There is no such thing as the correct stem-and-leaf plot. However, some plots are better than others.

This stem represents 3.0 – 3.4
This stem represents 3.5 – 3.9

```
2 | 8
3 | 2223
3 | 88899
4 | 00001234
4 | 6778
5 | 0122333334
5 | 555778
6 | 02344
6 | 568899
7 |
7 | 5
8 |
8 | 5
```

This data represent the unemployment rate in each of the fifty United States plus the District of Columbia in June, 2008.

What is missing? Add a title and legend.
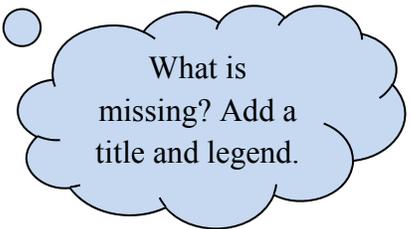
Dot Plots:
Below is a dot plot for the data about the number of cars in 50 sampled households.

Number of Cars in Households
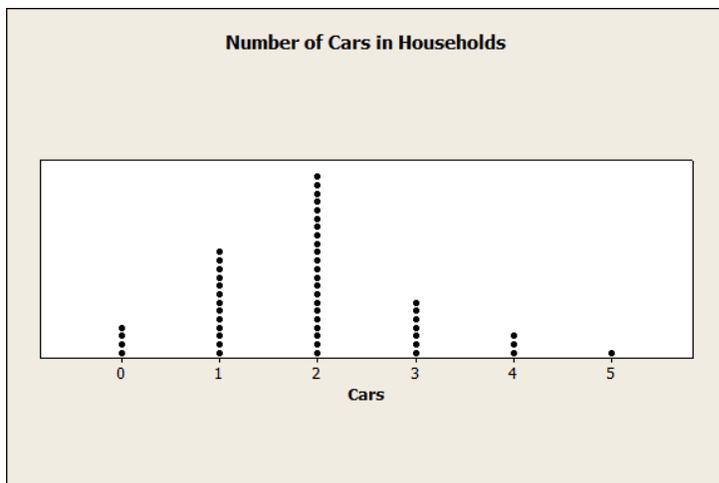
Cars

Here, we simply put a dot above each possible number of cars for each observation.

This seems like a pain in the neck (to read) but it does give us an idea of the shape of the distribution of data…

4

Shape of the Distribution:

There are four main types of shapes we will discuss. They are

1. **Uniform distribution:** the frequency of each value of the variable is evenly spread out across the values of the variable,

2. **Bell-shaped (symmetric) distribution:** the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle,

3. **Skewed right:** the tail to the right of the peak is longer than the tail to the left of the peak,

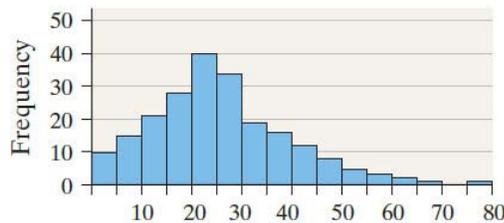4. **Skewed left:** the tail to the left of the peak is longer than the tail to the right of the peak.

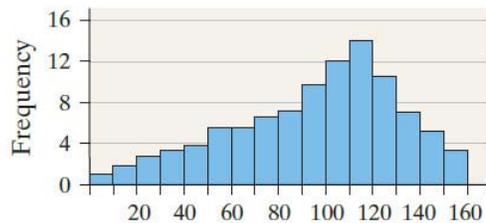To understand this further, consider the histograms of four different distributions.
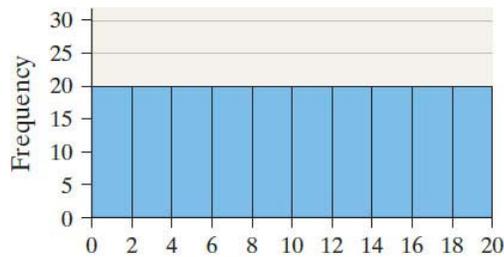
(a) Uniform (symmetric)
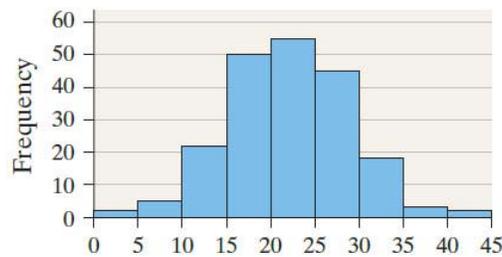
(b) Bell-shaped (symmetric)

(c) Skewed Right

(d) Skewed Left

expl 2: Consider the dot plot shown on the previous page. Which shape do you think it has?

Number of Cars in Households

Cars