DUXBURY
THOMSON LEARNING

Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Chapter 3 - Day 2

## 3.2 Stem Plots and 3.3 Histograms

**SOCKS** – SHAPE OUTLIERS CENTER SPREAD





DUXBURY
THOMSON LEARNING

Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## 3.2 Introduction to Stem and Leaf

**EXAMPLE: Weight of Female Students**

- For our first example, we use the weights of the 25 female students.

| | | | | |
|---|---|---|---|---|
| 150 | 140 | 155 | 195 | 139 |
| 200 | 157 | 130 | 113 | 130 |
| 121 | 140 | 140 | 150 | 125 |
| 135 | 124 | 130 | 150 | 125 |
| 120 | 103 | 170 | 124 | 160 |

| stem | leaf |
|---|---|
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |

Now add the leaf

*Next Slide*

- **Use the 1st two digits as the stem**
- **Use the 3rd digit as the leaf**

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## Stem and Leaf Example (continued)
### Weight of Female Students

1. Typically we sort the order the stems in increasing order.
2. We also note on the diagram the units for stems and leaves
3. Visually we can see possible outliers.
   - In Chapter 4, we give a precise rule for deciding when an observation is an outlier.

| | |
|---|---|
| 10 | 3 |
| 11 | 3 |
| 12 | 014455 |
| 13 | 00059 |
| 14 | 000 |
| 15 | 00057 |
| 16 | 0 |
| 17 | 0 |
| 18 | |
| 19 | 5 |
| 20 | 0 |

Probable outliers

Stem: Tens and hundreds digits
Leaf: Ones digit




DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.2 Stem and Leaf
## Key Points

***When to use:***

- *Numerical data with small to moderate number of observations*
- *number of possible categories.*

**How to construct:**

1. Select one or more leading digits for **the stem** value.
2. The next digits (or digit) after the stem becomes **the leaf.**
3. List the possible stem values in a vertical column (typically low to high values).
4. Record the leaf for every observation besides the corresponding stem value.
5. **Indicate the units for stem and leaf someplace in the display.**

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.2 Stem and Leaf

## Key Points

***What to Look For*** *The display conveys information about:*

- **A representative or typical value in the data set**
- **The extent of spread about a typical value**
- **The presence of any gaps in the data**
- **The extent of symmetry in the distribution of values**
- **The number and locations of peaks**

**SOCKS – SHAPE OUTLIERS CENTER SPREAD**

5



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Comparative Stem and Leaf Diagram

## Student Weight (Comparing two groups)

When it is desirable to **compare two groups**,

- **back-to-back stem** and leaf diagrams are useful.
- Here is the result from the student weights.
- **What does the comparative stem and leaf show?**
  - **it is clear that the males weigh more (as a group not necessarily as individuals) than the females.**

| *Females* | | *Males* |
|---|---|---|
| 3 | 10 | |
| 3 | 11 | 7 |
| 554410 | 12 | 145 |
| 95000 | 13 | 0004558 |
| 000 | 14 | 000000555 |
| 75000 | 15 | 0005556 |
| 0 | 16 | 00005558 |
| 0 | 17 | 000005555 |
| | 18 | 0358 |
| 5 | 19 | |
| 0 | 20 | 0 |
| | 21 | 0 |
| | 22 | 55 |
| | 23 | 79 |

6

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## Repeated Stem-and-leafs

**GPA example**: The following are the GPAs for the 20 advisees of a faculty member.

- GPA…

| | | | | | | |
|---|---|---|---|---|---|---|
| 3.09 | 2.04 | 2.27 | 3.94 | 3.70 | 2.69 | 2.22 |
| 3.72 | 3.23 | 3.13 | 3.50 | 2.26 | 3.15 | 2.66 |
| 2.80 | 1.75 | 3.89 | 3.38 | 2.74 | 1.65 | |

- **If the ones digit is used as the stem, you only get three groups (1, 2, and 3).**
- **You can expand this a little by breaking up the stems by using each stem twice**
  - letting the 2nd digits 0-4 be a category (L), and
  - the 2nd digits 5-9 be the other category (H).
- ***The next slide gives two versions of the stem-and-leaf diagram.*** →

7



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## Two Repeated Stem-and-leafs

## Using GPA Data

```
1L |
1H | 67
2L | 0222
2H | 6678
3L | 01123
3H | 57789
```

**Diagram 1**

- **Stem: Ones digit**
- **Leaf: Tenths digits**

```
1L |
1H | 65,75
2L | 04,22,26,27
2H | 66,69,74,80
3L | 09,13,15,23,38
3H | 50,70,72,89,94
```

**Diagram 2**

- **Stem: Ones digit**
- **Leaf: Tenths and hundredths digits**

Note: The characters in a stem-and-leaf diagram must all have the same width, so if typing use a fixed character width font such as courier.

8

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Frequency Distributions & Histograms

- **Discrete numeric data -** the frequency tables are **similar to** those produced for **qualitative data**.
- **For example, a survey of local law firms in towns in Iowa gave**

| Number of Lawyers | Frequency | Relative Frequency |
|---|---|---|
| 1 | 11 | 0.44 |
| 2 | 7 | 0.28 |
| 3 | 4 | 0.16 |
| 4 | 2 | 0.08 |
| 5 | 1 | 0.04 |

*Graph it* →

9

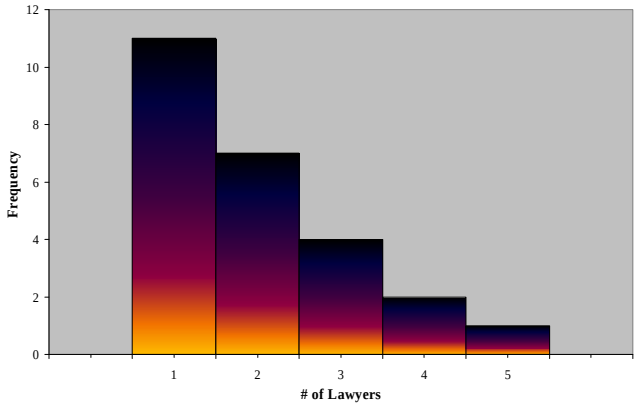


DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Discrete Data Histogram

Lawyer Example

**The number of lawyers in the firm will have the following histogram.**




Clearly, the largest group are single member law firms and the frequency decreases as the number of lawyers in the firm increases.

10

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Histogram for Discrete Data
## Key Points

***When to use:***

- *Discrete numeric data*
- *Works well even for large data sets*

**How to construct:**

1. Draw a horizontal scale, and mark the possible values of the variable.
2. Draw a vertical scale, and mark it with either frequency or relative frequency.
3. Above each possible value draw a rectangle with the value centered (so the rectangle for 1 is centered at 1, the rectangle for 2 is centered at 2 and so on).
4. The height of each rectangle corresponds to either the frequency or relative frequency.
5. Often possible values are consecutive whole numbers, in which case the base width for each rectangle is 1 (that is widths are all the same).





DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Histogram for Discrete Data
## Key Points

***What to Look For:***

- **Central or typical value**
- **Extent of spread or variation**
- **General shape**
- **Location and number of peaks**
- **Presence of gaps and outliers**

**SOCKS – SHAPE OUTLIERS CENTER SPREAD**

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

*3.3 Histogram for Discrete Data*

## Example: "How many textbooks did you purchase last term?"

**50 students were asked this question.**

- **The result is summarized below and**
- **The histogram is on the next slide.**

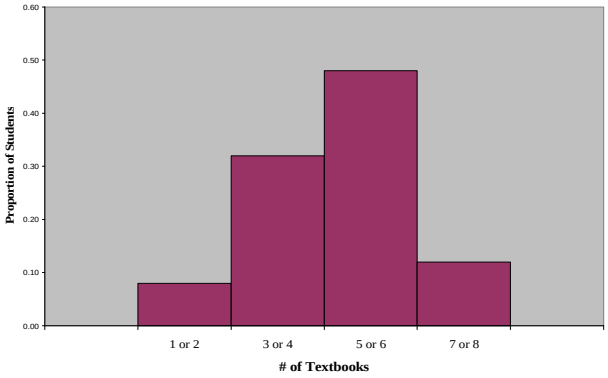
| Number of Textbooks | Frequency | Relative Frequency |
|---|---|---|
| 1 or 2 | 4 | 0.08 |
| 3 or 4 | 16 | 0.32 |
| 5 or 6 | 24 | 0.48 |
| 7 or 8 | 6 | 0.12 |

13



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## "How many textbooks did you purchase last term?"




- The largest group of students bought 5 or 6
- The graph appears symmetric with no outliers
- *Notice graph has a gap for 0 textbooks.*

14

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Histogram for Continuous Data

**Description of Example "Student Weight":**

- Consider the student weights of 79 students.
- The weights fall between 103 (lowest) and 239 (highest).
  - The range of the dataset is 239-103=136.

*How many class intervals should we make?*





DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## Example: "Student Weight Dataset"

Here is the Frequency Distribution

- Choosing a width of 15 we have the following frequency distribution.
  - **We have 10 groups (min:103, max:239, 136/15=10)**

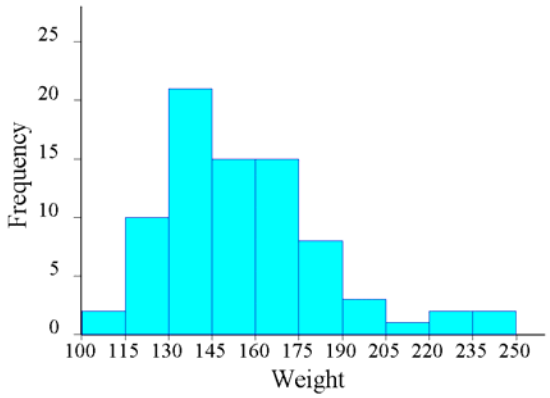
| Class Interval | Frequency | Relative Frequency |
|---|---|---|
| 100 to <115 | 2 | 0.025 |
| 115 to <130 | 10 | 0.127 |
| 130 to <145 | 21 | 0.266 |
| 145 to <160 | 15 | 0.190 |
| 160 to <175 | 15 | 0.190 |
| 175 to <190 | 8 | 0.101 |
| 190 to <205 | 3 | 0.038 |
| 205 to <220 | 1 | 0.013 |
| 220 to <235 | 2 | 0.025 |
| 235 to <250 | 2 | 0.025 |
| | 79 | 1.000 |

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Histogram for Continuous Data

1. Notice how boundaries of the class intervals on a horizontal axis are marked

2. The following histogram is for the **frequency** table of the weight data.




17



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# 3.3 Histogram for Continuous Data

## When the Class Interval Widths are Equal

### Key Points

**When to use**:

- **Continuous numeric data**
- Works well even for large data sets

**How to construct:**

1. Decide how many groups or "classes" you want to break up the data. Typically somewhere between 5 and 20.
2. Determine the "starting point" for the lowest group.
3. Mark the boundaries of the class intervals on a horizontal axis.
4. Use either frequency or relative frequency on the vertical scale.
5. Draw a rectangle for each class so that the edges are at the class boundaries.
6. The height of each rectangle corresponds to frequency selected.

### What to Look For:

**SOCKS – SHAPE OUTLIERS CENTER SPREAD**

18

DUXBURY
THOMSON LEARNING
Professor David Crystal
Roche
Dept

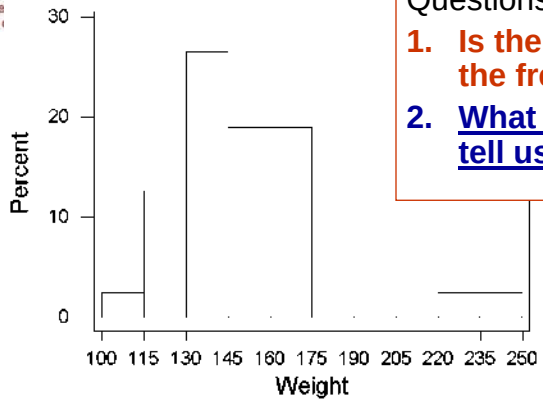
# Relative Frequency Histogram for Continuous Data

**Another continuous histogram**

- This is based on **relative frequencies** labeled percent.

Questions:

1. **Is the shape different from the frequency histogram?**
2. **What does the histograms tell us?**




*19*



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Histogram for Continuous Data

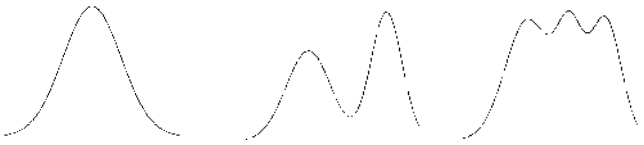
- **What does the histograms tell us?**

- The weights appear to be centered around 150 lbs
- The distribution of the weights is **unimodal** with a peak around 140 lbs
- There a few substantially large values
- The large Outliers skew the data to the right - a **positively skewed distribution**.

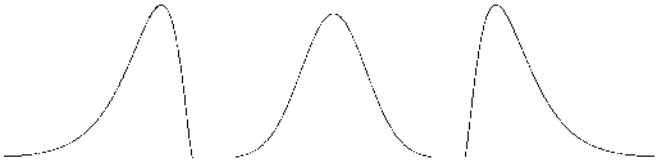
*20*

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Illustrated Distribution Shapes



Unimodal  Bimodal  Multimodal



Skew negatively  Symmetric  Skew positively

21



DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

## 3.3 Cumulative Relative Frequency Table

**Example:** We will continue to use… **"Student Weight Data"**

- If we keep track of the proportion of that data that falls below the upper boundaries of the classes, we have a **cumulative relative frequency table**.

| Class Interval | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|
| 100 to < 115 | 0.025 | 0.025 |
| 115 to < 130 | 0.127 | 0.152 |
| 130 to < 145 | 0.266 | 0.418 |
| 145 to < 160 | 0.190 | 0.608 |
| 160 to < 175 | 0.190 | 0.797 |
| 175 to < 190 | 0.101 | 0.899 |
| 190 to < 205 | 0.038 | 0.937 |
| 205 to < 220 | 0.013 | 0.949 |
| 220 to < 235 | 0.025 | 0.975 |
| 235 to < 250 | 0.025 | 1.000 |

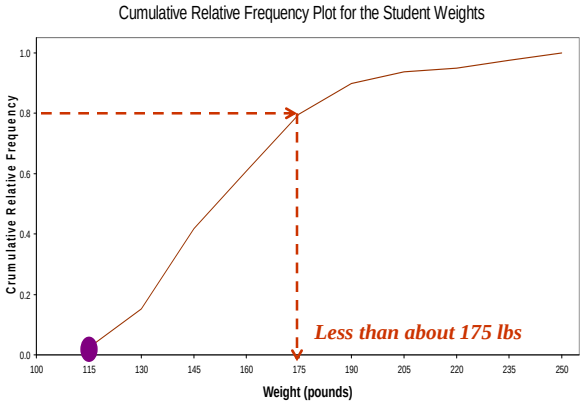
22

DUXBURY
THOMSON LEARNING

Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

# Cumulative Relative Frequency Plot

- Graph the cumulative relative frequencies
- Use the upper endpoint of the corresponding interval
- For example, start by plotting the first cum. relative frequency of .025 at 115 lbs, and so on.
- Using the graph, what is the weight for 80% of the students?







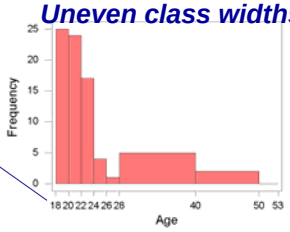

**It is unclear if these Histograms will be on the AP Exam.**

**POSTPONED TO A LATER DATE based on determination if there are any applications for them.**

# *Histograms with uneven class widths*

*(NOT COVERED AT THIS TIME)*

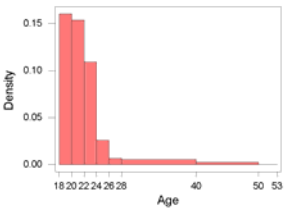
***Uneven class widths***




**To correct the distortion**,

•we create a density histogram. The vertical scale is called the density and the density of a class is calculated by

$$\text{density} = \text{rectangle height} = \frac{\text{relative frequency of class}}{\text{class width}}$$




**CORRECTED HISTOGRAM:**
This choice for the density makes the area of the rectangle equal to the relative frequency.