Sections 2.3 and 2.4: Frequency Distributions and their Graphs

We've seen how to use bar charts to represent qualitative data. With a few modifications we do can the same for quantitative data.

Example 1 *Consider the following test scores.*

55	87	54	61
85	88	94	98
109	83	104	105
98	106	102	92
101	95	105	
95	58	78	
71	93	85	
98	88	88	
106	94	74	
105	54	88	

Definition 2 *A **stem-and-leaf** plot organizes quantitative data by splitting each value into a stem and a leaf. Stems are used as rows where leaves represent observations.*

**Example 3** *For the above data, I used the 10s as my stems and single digits as my leaves.*

stem	leaf
5	4 4 5 8
6	1
7	1 4 8
8	3 5 5 7 8 8 8 8
9	2 3 4 4 5 5 8 8 8
10	1 2 4 5 5 5 6 6 9

Example 4 *It would be silly to give each possible integer grade possibility its own bar. Instead we create intervals or **classes** of equal width. Once we set the classes, we determine how many data points fit into each class. The number of data points in a class is referred to as the **class frequency**.*

Class	Frequency
50-59	4
60-69	1
70-79	3
80-89	8
90-99	9
100-109	9

*Next we create a bar chart using classes as our categories and frequency to represent the height of each bar. This type of bar chart is called a **histogram**. Note that the general shape is the same as our previous stem-and-leaf display.*

Test 1 Grades

50-59 60-69 70-79 80-89 90-99 100-109
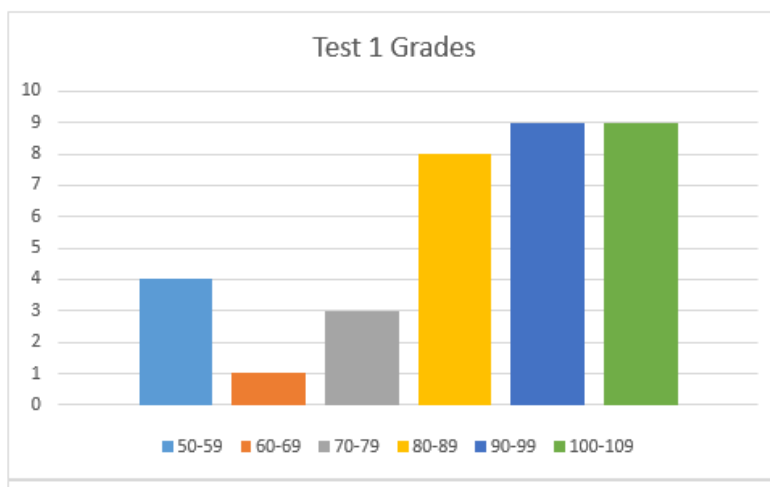
Remark 5 *The advantage of a stem-and-leaf display is the ability to retrieve values from the data set. However, with a large data set, the stem-and-leaf display becomes dense and difficult to read.*

Definition 6 *The smallest value that can appear in a class interval is called the* **lower class limit.**

Definition 7 *The largest value that can appear in a class interval is called the* **upper class limit.**

Definition 8 *The* **class width** *is the difference between the consecutive lower class limits.*

Example 9 *For the* $(80 - 89)$ *class, the lower class limit is 80, the next lower class limit is 90. Thus, the class width is* $90 - 80 = 10$

How does one determine class width? Examine your data. Look to see where the data falls and pick a reasonable number of desired classes. In looking at the test scores above it seemed natural to break the data down into groups of ten and we ended up with six classes. Could we have broken the data into groups of five? Sure. That would have resulted in twelve classes. There is no single correct number of classes to utilize. Be careful not to use too few classes or too many. In doing so you create a chart with too much or too little detail. I chose to begin the lower class limit at 50 rather than at the smallest data point, 54. I though the overall structure of the histogram classes would look better that way.

The provides a clinical way to determine class width. First determine the number of desired classes. Then determine class width as

$$\text{class width} = \frac{\text{largest data value - smallest data value}}{\text{number of classes}}$$

and then round up to a convenient number. Determining esthetically pleasing and statistically informative class intervals is both an art and a science.

Definition 10 *A **mode** is the most frequently occurring observation or number. A mode occurs at the tallest bar in the histogram.*

Problem 11 *What is the mode time that people visit 2nd and Charles in Kennesaw?*

Problem 12 *What is the class width for times that people visit 2nd and Charles in Kennesaw?*
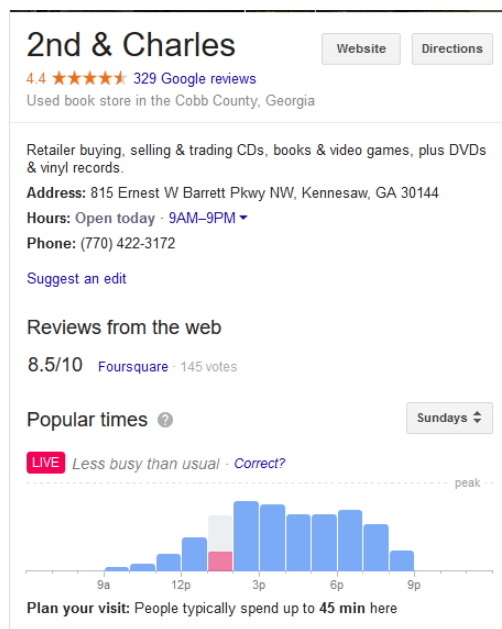
1   Exercises

1. Kokoska 3rd edition Section 2.3: 2.3, 2.57

2. Kokoska 3rd edition Section 2.4: 2.73-2.77, 2.83, 2.85