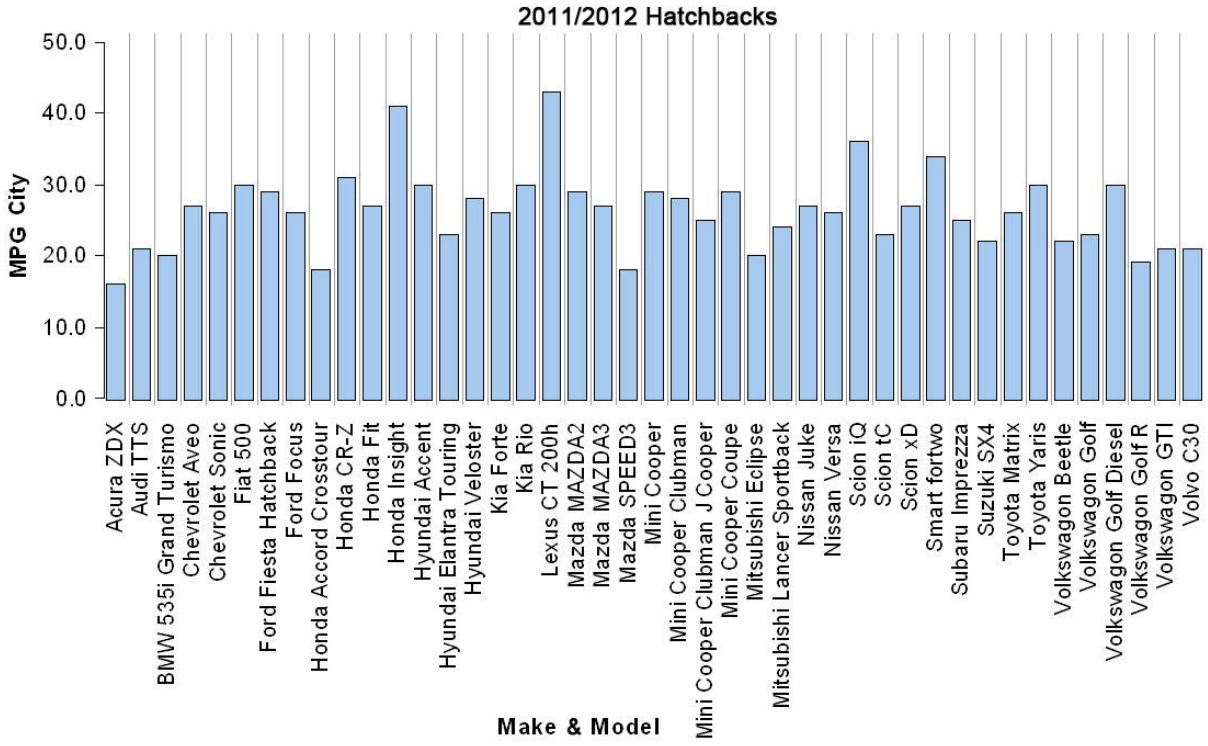
MOD 7 PART 1: DISTRIBUTIONS FOR QUANTITATIVE DATA (DOTPLOTS)

Learning Goal

- For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

1) Use the case-value graph below to answer each of the following questions.



a) How many hatchbacks get 23 mpg in the city?

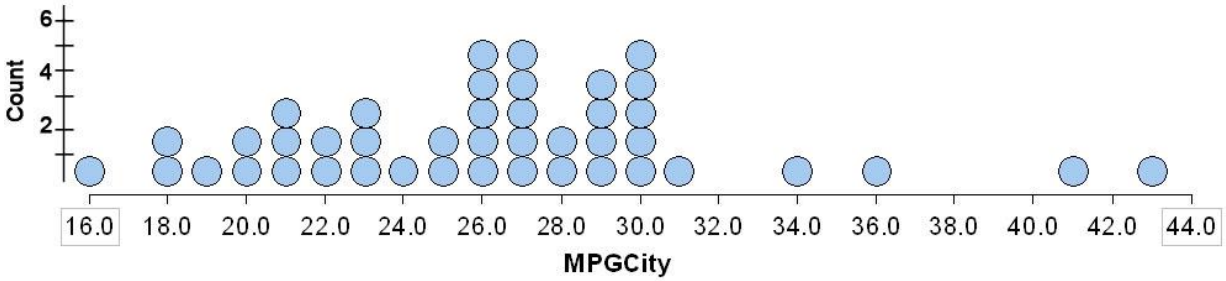
b) What is/are the mpg rate(s) that occur most frequently (this measure of center is the mode)?

c) Which mpg rate is the middle rate (this measure of center is the median)?

2) Take a look at the questions you answered on the previous page. Were you asked to examine the individual cases and their values (i.e. the individual hatchbacks), or were you asked to describe patterns in the data and/or to create summaries about the group (in this case the group would be 2011/2012 Hatchbacks)?

3) Was it awkward to work with the case-value graph on the previous page? Why or why not?

4) Let's answer the same questions using a dotplot to represent the same data for the variable MPGCity.



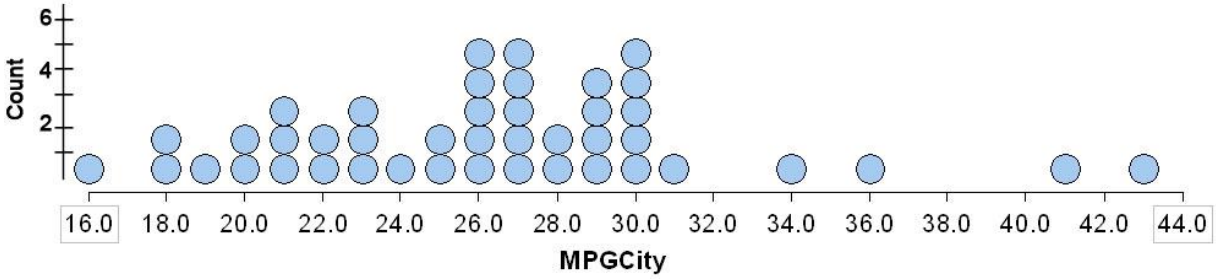
a) How many hatchbacks get 23 mpg in the city?

b) What is/are the mpg rate(s) that occur most frequently?

c) Which mpg rate is the middle rate (this measure of center is typically called the median)? What did you do to find this "middle"?

5) Which graph gives a more immediate summary of the distribution for the MPGCity variable (the case-value graph on the first page of this activity or the dotplot on the second page)? Explain.

6) Use the dotplot to summarize the data even further.



a) What is the overall range for the mpg values?

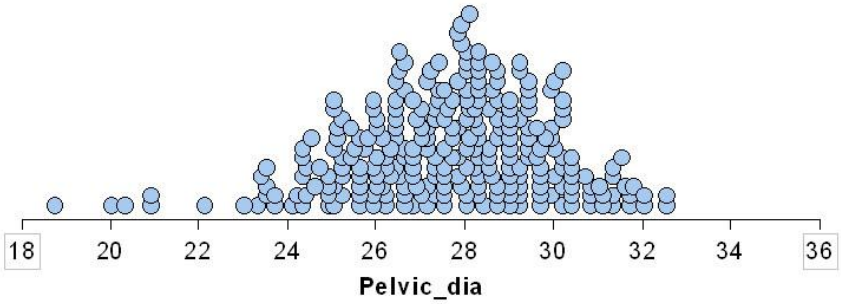
b) Describe the typical interval (the main clump of data) for the mpg values.

c) What is the average mpg (this measure of center is typically called the mean)?

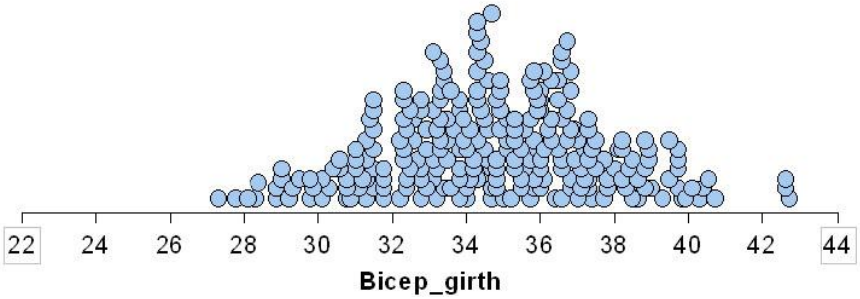
d) What do you think the unusual mpg rates are (these are typically called outliers)?

7) For each of the following dotplots, draw a smooth curve outlining the distribution, and then describe the center and the shape of the distribution (i.e. symmetrical vs skewed). Also note any outliers.

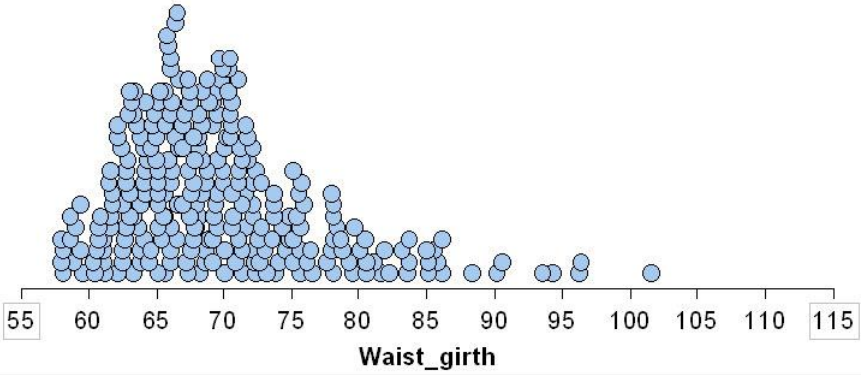
Pelvic diameter for 260 women who were exercising servarl times a week.



Bicep girth for 247 men who were exercising several times a week.



Waist girth for 260 women who were exercising several times a week.

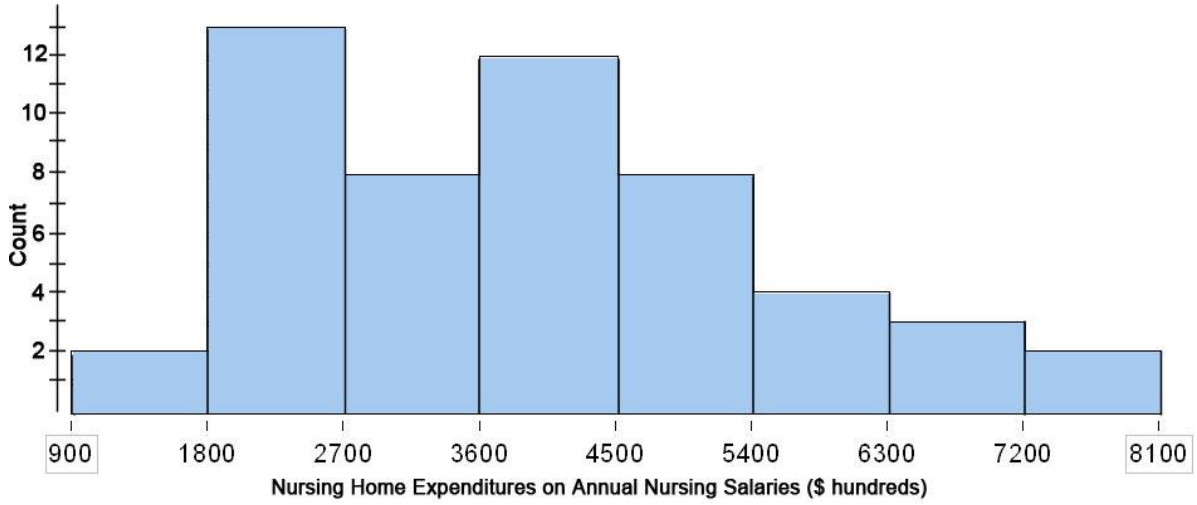


MOD 7 PART 2: DISTRIBUTIONS FOR QUANTITATIVE DATA (HISTOGRAMS)

Learning Goal

- For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

1) The data in the histogram below were collected by the Department of Health and Social Services of the State of New Mexico and cover 52 of the 60 licensed nursing facilities in New Mexico in 1988. The quantitative variable *Expenditures on Nursing Salaries* is in hundreds of dollars.



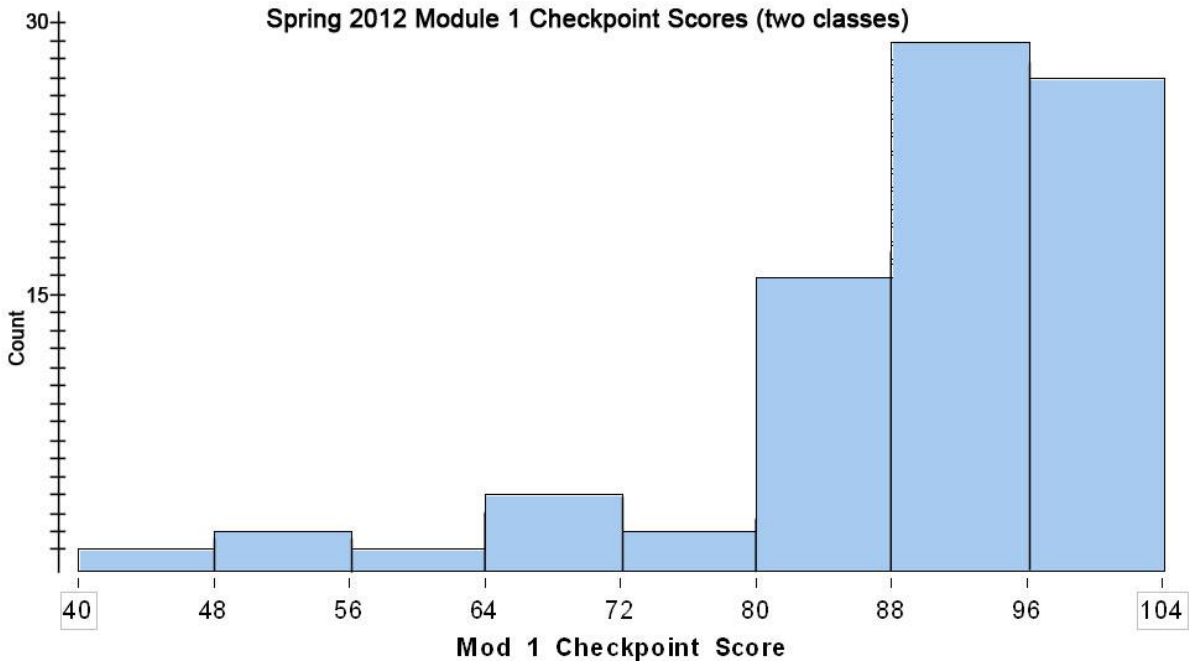
a) How many nursing homes spent between $450,000 and $540,000 annually on nursing salaries?

b) What was the frequency of nursing homes that spent between $180,000 and $270,000 annually on nursing salaries?

c) What percentage of nursing homes spent between $450,000 and $540,000 annually on nursing salaries (this percentage is also called the **relative frequency**)? Explain how you found the relative frequency of nursing homes that spent between $450,000 and $540,000 annually on nursing salaries.

d) What percentage of nursing homes spent between $180,000 and $270,000 annually on nursing salaries (again … this percentage is called a **relative frequency**)?

2) The following is a histogram indicating the distribution of scores on the Spring 2012 Module 1 Checkpoint for an instructor's combined statistics classes.



3) What percentage of students scored below 80% (assume left-hand endpoints are included in each bin)?

4) For each of the following questions, answer the question if the histogram in number 2 above provides enough information to answer it. If not, write "not enough information".

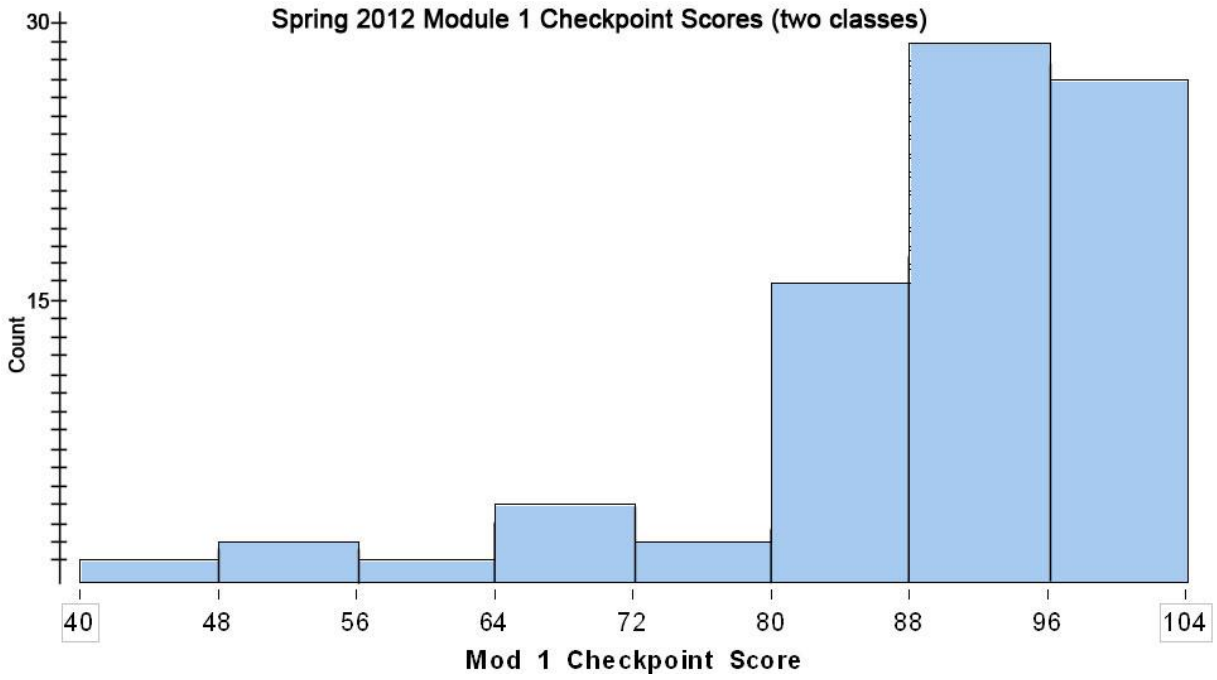
   a) What percentage of the students who took the exam scored at least 88 and less than 96 points?

   b) What is the lowest score on the Module 1 Checkpoint?

   c) What percentage of students scored less than 90?

   d) How many students did not pass the exam, if "not passing" is a score of 70 or less?

5) The following is a histogram indicating the distribution of scores on the Spring 2012 Module 1 Checkpoint for two statistics classes.



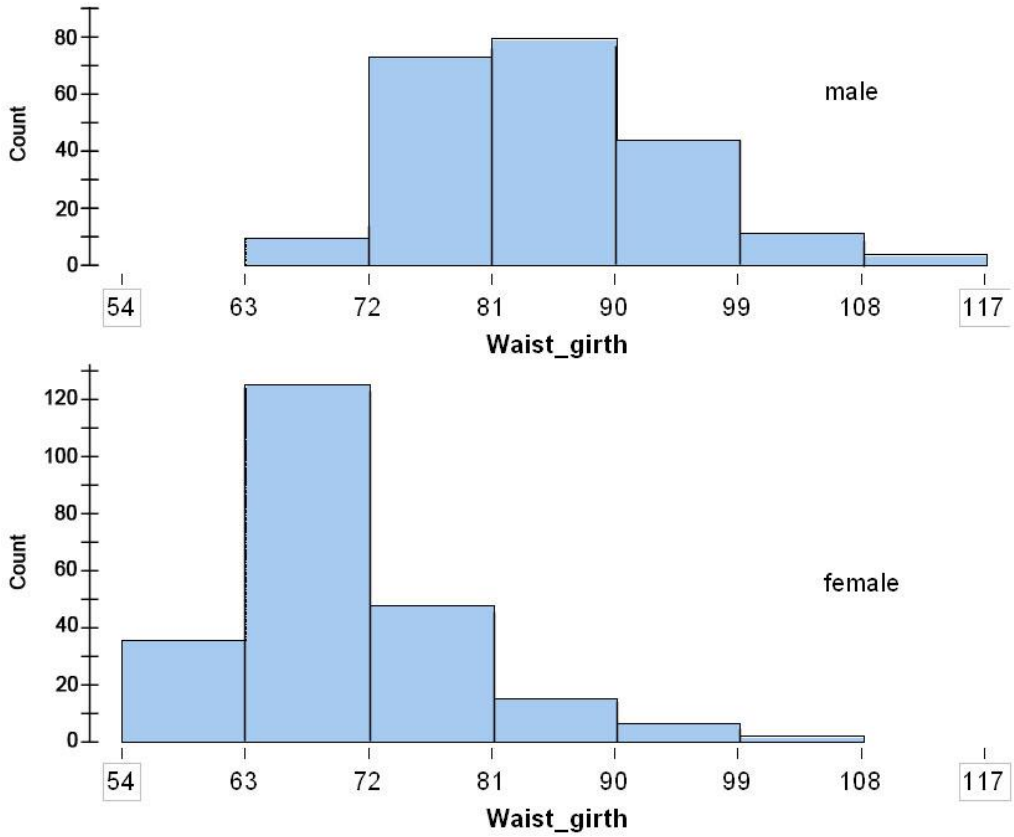
a) Describe where the center appears to be in this distribution (again … try to write like a statistician).

b) What is the most frequently occurring score (i.e. the mode)?

c) Assume this data is representative of all Math 160 students using the same learning materials. In general (no numbers needed), what could you say about the probability of randomly finding a student who scored very low on the checkpoint? In general (no numbers needed), what could you say about the probability of finding a student who scored fairly high on the checkpoint?

d) Describe the shape of this histogram, and try to write like a statistician.

6) Here are data from adults (247 men and 260 women) who were exercising several hours a week. Indicate whether you think the following statements are **valid** or **invalid** (and try to explain why).



a) Typical females have a smaller waist girth than males.

b) The spread in waist girth is smaller for females than for males.

c) Assume that approximately 20% women fall into the third bin from the left. A medium size pair of pants will fit a woman with a waist girth between 72 and 76 centimeters, so a medium size pair of pants will fit about 20% of the women in this sample.