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

Specific Learning Objectives:

- Distinguish between categorical and quantitative variables;
- Identify graphs that represent the distribution of a quantitative variable;
- Analyze the distribution of a quantitative variable using a dot plot. Describe the shape, give a general estimate of center, and determine the overall range.

Overview:
In this activity you will practice analyzing the distributions of quantitative variables using descriptions of shape, center and spread. We will focus on dot plots. If you encounter vocabulary that you do not know (or don't remember), check out the vocabulary list on the last page of the previous activity.

1) Here is a partial spreadsheet of 2011-2012 data for a set of hatchback cars.

Car make and model	City miles per gallon	Drive	EPA size class	Engine
Acura ZDX	16 mpg	All wheel drive	Sport utility	6 cylinder
Audi TTS	14 mpg	All wheel drive	Midsize car	10 cylinder
Chevrolet Aveo	25 mpg	Front wheel drive	Compact car	4 cylinder

   a) Who are the individuals described by this data?

   b) How many variables are shown in the spreadsheet?

   c) Which variables are categorical variables?

   d) Which variables are quantitative variables?

   e) Describe what the values in the spreadsheet tell us about the Chevrolet Aveo.

2) Below you will see a case-value graph and a dot plot graph of the 2011-2012 data for a set of hatchback cars. Use one or both graphs to answer the questions below. Jot down notes or draw on the graphs to show how you determined your answers.


2011/2012 Hatchbacks
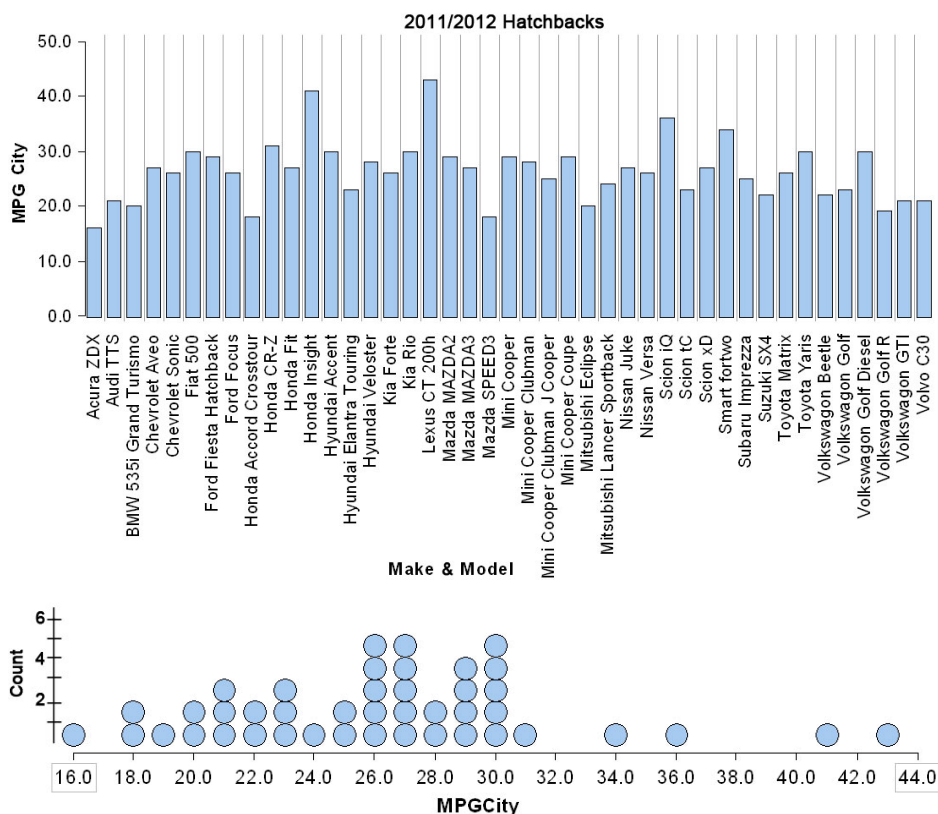
a) What is the best city mpg for this group of hatchbacks? _____

What model of hatchback gets the highest city miles per gallon?

What is the worst city mpg for this group of hatchbacks? _____

What model of hatchback gets the worst city mpg?

For each of the above questions, indicate which graph (case-value or dot plot) was the easiest to use to answer the question.

b) How many hatchbacks get 25 mpg in the city? Which graph was the easiest to use to answer this question?

c) What are the mpg rates that occur most frequently for hatchbacks? Which graph was the easiest to use to answer this question?

d) If you had to pick one mpg rate to represent this data, what would it be? Why did you choose that value?

3) The case-value graph and the dotplot are two different graphs of the same data. What do you see as the advantages and disadvantages of each type of graph?

4) Statisticians make graphs to summarize data, so they prefer to use graphs that show the distribution of the data. A statistical distribution is defined as "an arrangement of the values of a variable showing their observed frequency of occurrence." Here is another definition of a statistical distribution: "a representation that shows the possible values of a variable and how often the variable takes those values." Which graph, the case-value graph or the dotplot, shows the distribution of the variable MPGCity?

5) For each of the following dotplots, draw a smooth curve outlining the distribution, and then describe the shape of the distribution using course vocabulary (See the definitions of symmetric, skewed left, skewed right.)

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

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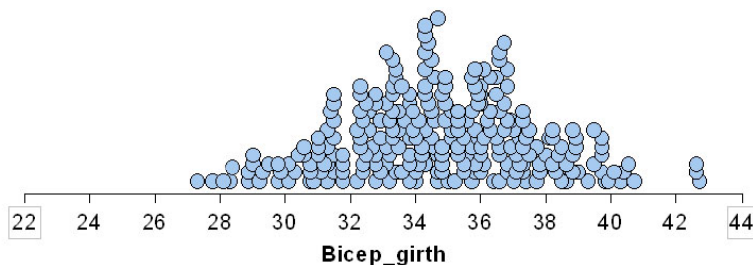
Waist girth for 260 women who were exercising several times a week.

Waist_girth
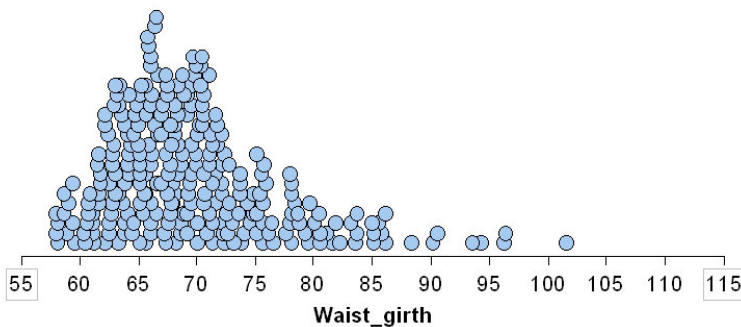
6) Suppose that almost everyone does well on the first exam with a few people (who did not study) performing poorly. What is the shape of the distribution of exam scores?

7) Thinking like a statistician: describing shape, center and spread.
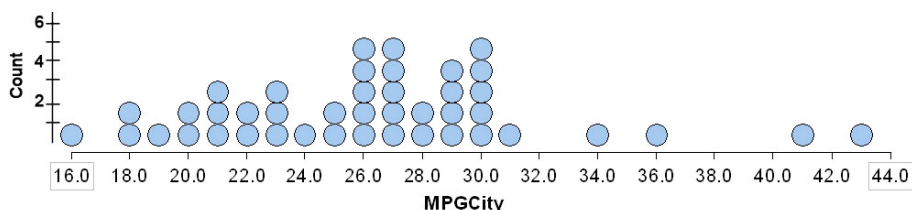
a) *Thinking about shape:* Our shape descriptions (symmetric, skewed left, skewed right) don't fit this distribution. How would you describe this distribution's shape?

b) *Thinking about center:* Previously, you chose one value to represent the distribution of mpg ratings for hatchbacks. What was that value? _____

To represent the distribution, Ann decided to calculate the average mpg rating using the mean. She got 34 mpg. Do you think Ann made a mistake? How can you tell without doing any calculations?

c) *Thinking about spread:* Spread is a description of the variability we see in the data.

Here the city mpg for these hatchbacks varies from a low of _____ mpg to a high of _____ mpg.

What is the overall range for the mpg values (max – min)? _____

Typical hatchbacks have a city mpg that varies from about _____ mpg to about _____mpg.

d) *Thinking about deviations from the pattern*: Are there any unusually larger or unusually small mpg rates? If so, what are the unusual values?

8) Here are dot plots of the sugar content (grams per serving) for some adult cereals and child cereals. Compare the two distributions by comparing shapes, estimates of center, and spreads.