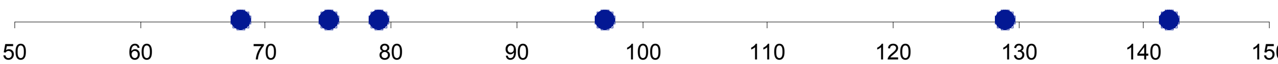
Descriptive Statistics: Central Tendency
Goal: We want one number that will describe the central location of a set of data.

Example Data: Six real estate listings were randomly sampled from the Quad City Times classifieds (08/05/2004).
The prices of the houses (in thousands of dollars): **68 75 79 97 129 142**

1. Let's begin by creating a dotplot of the data.

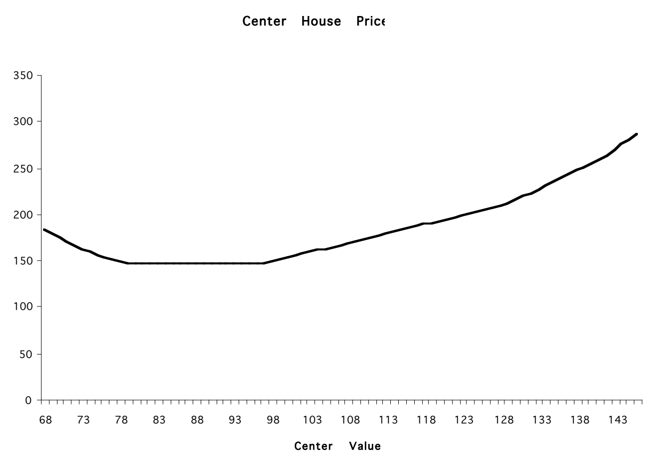


2. Two (potential) graphical methods for finding the center:
   a) Locate a point halfway between the minimum and maximum. Midpoint = $\frac{68+142}{2} = \frac{210}{2} = 105$
   b) Find the balance point (treat each dot as if it has some weight). For this example, the balance point = 98.3333

3. Now let's try to use logic and algebra to find the center of the data. We first need to define what we mean by the *center* of the data.
   Possible definition: Center = the point that minimizes the sum of the distances between each data value and that point.
   We'll assume that distances must be positive values.

   We want to select a center, *c*, that minimizes: $$\sum_{i=1}^{n} |x_i - c| = |68-c| + |75-c| + |79-c| + |97-c| + |129-c| + |142-c|$$

   We could use calculus to find the value of *c* that minimizes this quantity (although the absolute values make it messy).
   Instead, let's use a graphical method to find the center value that minimizes the sum of absolute deviations.
   We can pretty much guess that the center value is going to be between 75 and 110. Let's concentrate on values in this range.

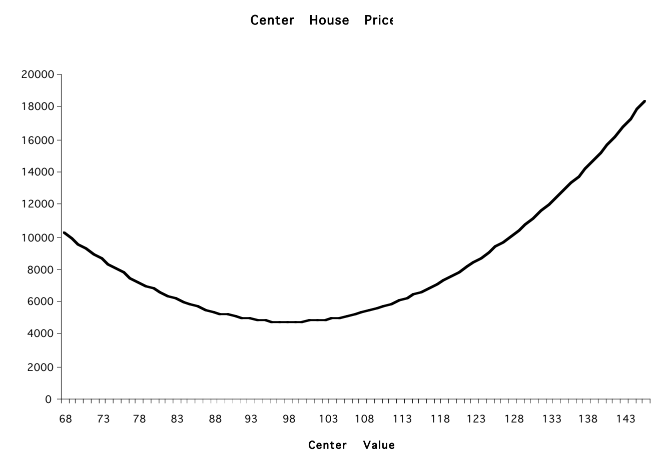


   It looks like this quantity is minimized when we set the center value anywhere between 79 and 97.

We want to select a center, *c*, that minimizes:

$$\sum_{i=1}^{n}(x_i - c)^2 = (68-c)^2 + (75-c)^2 + (79-c)^2 + (97-c)^2 + (129-c)^2 + (142-c)^2$$

Let’s use a graphical method to find the center value that minimizes the sum of squared deviations.



$$\sum_{i=1}^{N}(x_i - a)^2 = (x_1 - a)^2 + (x_2 - a)^2 + \ldots + (x_n - a)^2 = \left(x_1^2 - 2x_1a + a^2\right) + \left(x_2^2 - 2x_2a + a^2\right) + \ldots\left(x_n^2 - 2x_na + a^2\right)$$

$$\frac{d}{dx} = (2x_1 - 2a) + (2x_2 - 2a) + \ldots + (2x_n - 2a) = 2(x_1 + x_2 + \ldots + x_n) - 2(a + a + \ldots + a)$$

$$0 = 2(x_1 + x_2 + \ldots + x_n) - 2(a + a + \ldots + a) \Rightarrow (x_1 + x_2 + \ldots + x_n) = (a + a + \ldots + a) = na$$

$$\frac{(x_1 + x_2 + \ldots + x_n)}{n} = a = \frac{1}{n}\sum x_i$$