Today we are going to investigate whether a distribution is symmetric or skewed, and determine which measure of center best represents a *typical* number, and which measure of variability best represents the distribution.

Reminders:

- **Distributions** are either **skewed** or **symmetrical.**
- We **represent** distributions with **dot plots**, **histograms**, and **box plots.**
- **Mean** is the average, **median** is the middle number, **mode** is the most common number.
- **Standard deviation** and **Interquartile Range (IQR)** both measure the variability.

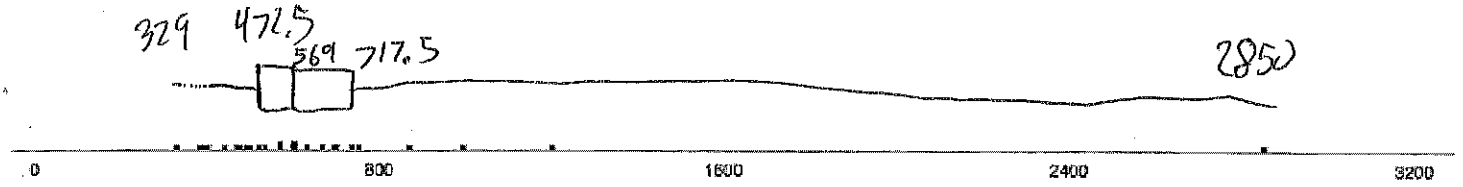
When measuring variability…

- **Standard Deviation** is the best measure if the distribution is **symmetrical**
- **Interquartile Range** is the best measure if the distribution is **skewed**

I went on Zillow.com to find the prices of 25 houses for sale near CHS. Here are the reported prices (in thousands of dollars):

700	534	470	869	695	2850	735	995	440	1200	600	520	490
750	500	629	329	569	400	475	600	665	569	390	384	

1. Below I created a **dot plot** of the distribution. **Draw** a **box plot** using the same number line, above the dot plot (remember you can use the TI-84 calculator to help you find the numbers needed to create the box plot!).



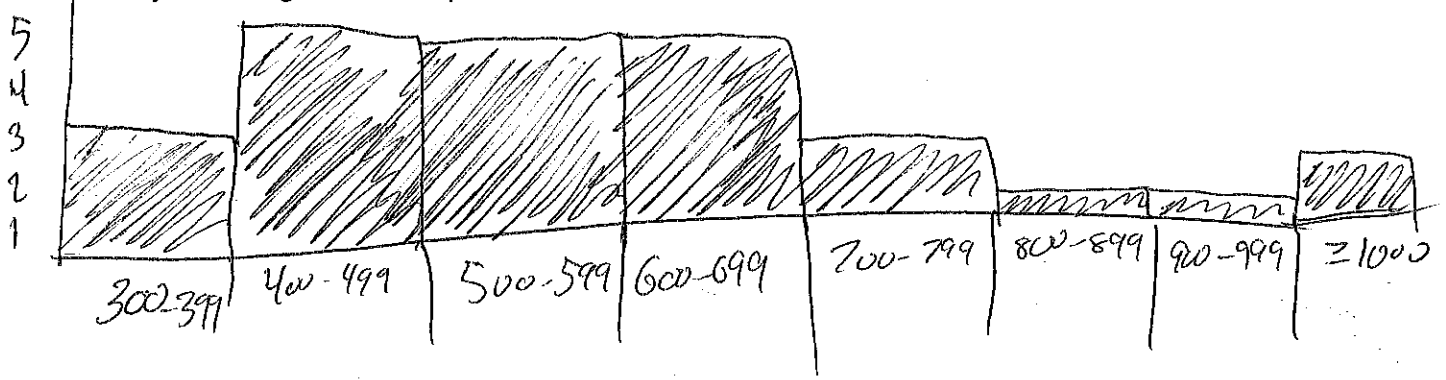
Housing Prices near CHS (in the thousands $)

2. Some of the prices are pretty similar to each other, and the dot plot makes it look like almost every house has a different price. Another way to represent a distribution is with a **histogram.** A histogram is like a dot plot, but instead of each individual number getting a dot, you count how many dots are in an interval. We will use an interval size of 100.

a. First, count how many houses fit into each **interval.**

300-399	400-499	500-599	600-699	700-799	800-899	900-999	≥1000
3	5	5	5	3	1	1	2

b. Draw your histogram in the space below.



3. Calculate the measures of center (median, mode, and the mean) of the distribution.

median = 569, mean = 694, mode = 600

4. Would you describe the distribution as more **symmetric** or **skewed**?

Skewed right

5. What do you think is the best estimate for the price of a *typical* house near CHS (**mean or median**)? Justify your answer.

$569,000 b/c the median is better for skewed distributions.

6. Calculate the standard deviation and interquartile range of the distribution.

$s_x = 491.19$, IQR = 717.5 − 472.5 = 245

7. What do you think is the best estimate for the variability of houses near CHS (interquartile range or standard deviation)? Justify your answer.

IQR. The outliers make the standard deviation too big. The distribution is skewed, so use IQR.

8. Imagine that the outlier housing prices were NOT included in our data.

   a. Which of the three measures of center do you think will change? Which will stay the same? Explain your thinking.

   Mean will change a lot. Median might move a little bit. Mode will not change. The mean is most affected by skew.

   b. Which of the two measures of variability (standard deviation or interquartile range) will change? Which will stay the same? Explain your thinking.

   Standard deviation will change a lot (decrease). IQR will shift, but only slightly.

   c. Now that the outliers aren't included, which measure of variability do you think is a better measure of variability?

   Standard deviation because it is now symmetric.

   d. Fill in the blanks in the sentences:

   i. Because Skewed distributions pull the mean in the direction of the tail, the median is a better measure of center.

   ii. Because Symmetric distributions have the mean in the center of the distribution, the mean is the best measure of a typical data point.

   iii. Because skewed distributions have values that are spread far in one direction, but clustered in the other, the IQR is a better measure of variability.

   iv. Because symmetric distributions have values that are centrally clustered, the Standard deviation is a better measure of variability.