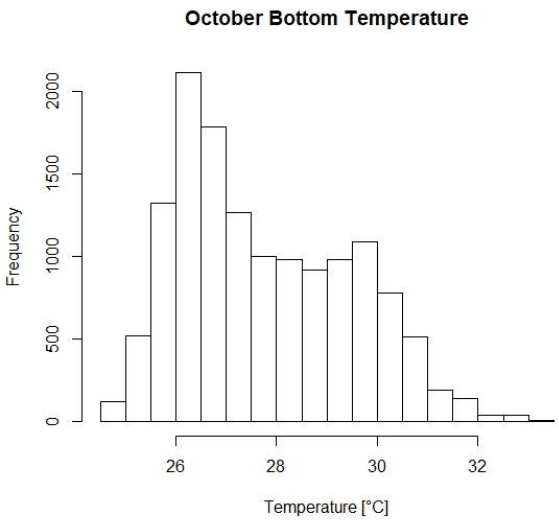
Summarizing Skewed Distributions, Finding Outliers, & Creating Box Plots Worksheet

KEY TERMS

- **Median**: *a number that measures the "center" of the data; the "middle value" when the data has been sorted from smallest to largest*
  - To find the median, arrange all numbers in order from smallest to largest, then find the middle number. If two numbers are in the middle, average them.
- **Quartiles:** *numbers that separate the data into quarters*
- **Interquartile Range (IQR):** *a number that indicates the spread of the middle half or the middle 50% of the data. It is the difference between the third quartile (Q3) and the first quartile (Q1)*
  - IQR = Q3 - Q1
- **Outlier:** *an observation of data that does not fit the rest of the data; sometimes called an "extreme value"*
  - To find where outliers lie, find the lower and upper fences ("cutoffs") for unusual values:
    - Lower fence = Q1 - 1.5(IQR)
    - Upper fence = Q3 + 1.5(1QR)
- **Boxplot:** *a visualization that shows where the bulk of the data lie*
  - To create a boxplot, the box is drawn from Q1 to Q3 with a line for the median inside the box. Whiskers are drawn to the most extreme values within the fences and potential outliers are marked with an asterisk.

Illowsky, B., & Dean, S. (2018). Introductory statistics
Gould, R., & Ryan, C. N. (2015). Introductory statistics: Exploring the world through data. Pearson.



Here is a histogram of bottom temperature data from the Mokauea loko i'a from October 2019. This data has been cleaned where duplicates and outliers have been removed.
Temperature is a critical parameter that needs to be understood to perform restoration efforts to the Mokaeua loko i'a. Help us to understand October 2019's bottom temperatures by completing the exercises below.

What is the shape of this histogram?

Given the shape of this histogram, complete the exercises below.

The above data is real data collected from research at Mokauea loko i'a.
Additional Mokauea loko i`a data can be explored here: http://grogdata.soest.hawaii.edu/

EXERCISE 1: Median

Here is a random sample of 10 data points (rounded to the nearest tenth) from the October 2019 bottom temperature data. Calculate the median of this sample. Round your answer to the nearest tenth.

27.3, 28.2, 29.5, 27.7, 28.8, 26.6, 30.0, 26.3, 31.1, 26.0

EXERCISE 2: IQR

With the same sample above, find Quartile 1, 2, and 3. Round your answer to the nearest tenth.

Q1 =

Q2 =

Q3 =

Now calculate the IQR of the dataset.

EXERCISE 3: Outliers

Find the upper and lower fences to figure out where outliers lie. Continue to the above sample. Round to the nearest tenth.

Based on your calculations, are there any outliers present in the sample?

Would 33.8°C be considered an outlier? Why or why not?

Exercise 4: Boxplot

Using your answers found in the previous exercises, complete the Five Number Summary, and create a boxplot with those numbers.

Five Number Summary

1) Lowest value: _____
2) Q1: _____
3) Q2/Median: _____
4) Q3: _____
5) Highest value: _____