Karina Sirabian 9/6/20

**Activity 2: Intro to Statistics and Weather**

1. There are 157849 rows and 11 columns in the dataset.
2. A character can be a letter, number, or any other symbol surrounded by quotes. A string is like an array (or list) of single characters. A numeric is a number that can have a decimal or can be negative while an integer is a number that cannot have decimal places or be negative. An integer must be a positive whole number. A factor is a data type used for repeating objects (i.e. integers, numerics, character strings, logicals) in a vector. If you have a vector with repeating character strings, you can turn it into a factor and create labels for each character string. Each label is a level in the factor. An example would be if you have a vector with the days of the week c(Sun, Sat, Weds, Tues, Tues, Mon, Sat, Thurs, Weds, Fri) then you can turn this into a factor, which will have a level for each day of the week. You can rename these levels (example turn Sun into Sunday) and you can reorder the levels from Sun, Sat, Weds, Tues, Mon, Weds, Fri to the actual order of the days (the default order is the order in which they are entered). (See example vector of each data type in activity2.R)
3. The hist function has many parameters (arguments) for users to enter. Here is a list of each argument and what it is doing:

* **X (a vector)**: a vector of values for which the histogram is desired. In our example, datW$TAVE[datW$siteN == 1] is a vector of the average daily temperatures in Aberdeen, WA.
* **Freq (a logical):** If freq is set to be true then the y-axis of the histogram represents the frequencies of each group on the x-axis. If freq is false then the probability density is plotted (so the total area of the histogram equals 1). Freq defaults to TRUE if the breaks are equidistant.
* **Main (string character):** Specifies the main title on the histogram
* **xlab (string character):** Specifies the label on the x-axis
* **ylab (string character):** Specifies the label on the y-axis
* **Col (string character):** A string of either a color name or a hexadecimal value (ex: “#FF0000” entered would be red) to fill the bars. A NULL value would have the bars be unfilled.
* **Border (string character):** A string of either a color name or a hexadecimal value (ex: “#FF0000” entered would be red) to specify the color of the border around the bars. The default is the standard foreground color

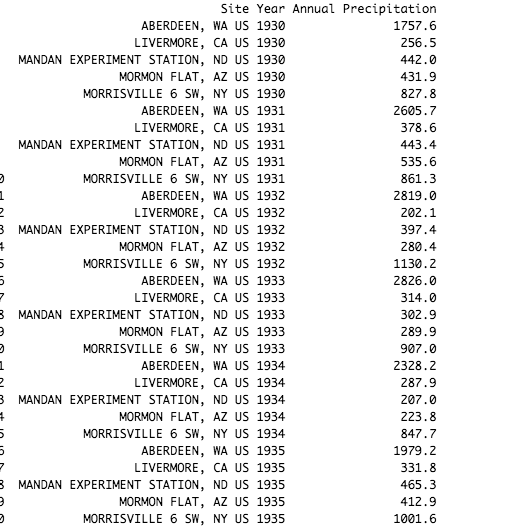
1. The histogram below shows the daily average air temperature in Mormon Flat, Arizona. The distribution does not look like it is normally distributed because it does not follow the standard bell curve pattern. The distribution most resembles the shape of a bimodal distribution because there are two peaks.



1. Let’s assume that the threshold for extreme high temperatures in an area is the 95th percentile of the average daily temperature for said area. In Aberdeen, the current temperature for the 95th percentile is 18.51026 degrees Celsius (see code for calculation). If climate change causes the mean temperature to increase by 4 degrees Celsius, we can calculate the probability that the temperature is above 18.51026 degrees Celsius to be 0.2031656 (see code for calculations).
2. Below is a histogram of the daily precipitation for Aberdeen. To describe the shape of this daily precipitation data, both the exponential and gamma distributions could be used because they can both describe data sets that are skewed to the right and peak around zero.



1. The chart below shows the annual precipitation for each site from the years 1930-1935. To see the annual precipitation from each site for the rest of the years, the code from activity2.R will produce the full version of this chart.



1. The histogram in purple below shows the annual precipitation for Aberdeen. The plot appears to be normally distributed, but some quantities on the far left are lower than those on the far right causing the distribution to be slightly skewed to the left (the median of the data is larger than the mean).

The histogram in green below shows the annual precipitation for Mandan Experiment Station. The plot looks normally distributed and the mean and median for the data are very close together so it is not skewed.





1. P(Annual precipitation in Aberdeen ≤ 700mm) = 5.690109x10^-5 P(Annual precipitation in Mandan Experiment Station ≤ 700mm) = 0.9968318 (see git hub link for code)
2. <https://github.com/Karina-Sirabian/ENVST_206.git>

Activity 2.R