By now, it should be no secret that the toolkit of a data scientist is very diverse. Many tools exist to describe data, from summary statistics (mean, median, mode) to data visualizations (scatterplots, boxplots, histograms). While the latter come in all shapes and sizes, many of the visualizations look similar and can be confusing to the untrained eye. What’s the difference between line graphs and scatter plots? Area charts and multiple bar charts? When would one use a donut chart? While there are many such questions, I will be focusing on the difference between histograms and bar/column charts. Bar and column charts are essentially the same chart, but one has a 90-degree orientation difference.

A bar/column chart is a visual method to present *categorical* data with rectangular bars. Each bar in the chart represents a *discrete* category. The categories are displayed on the x-axis, while the measured values for each category are displayed on the y-axis and indicated by the height of each bar. A bar chart is primarily used for comparison and contrast of values across several categories, such as comparing the statistics of several countries or a company’s sales statistics. At my job, I used bar charts to display the volume of proteins produced by several bacterial species.

Below is an example of a bar chart displaying the per capita daily oil consumption of different countries. The x-axis displays the categories, in this case the country, and the y-axis displays the measured value, in this case the daily oil consumed per person. It should be noted that each category is not limited to one value; many bars can be used to indicate multiple values for each category in what is called a *grouped bar chart.*

While bar charts compare values across categorical variables, histograms allow data scientists to view the probability distribution, or the function that gives the probabilities of each value occurring. Histograms adopt a similar format to vertical bar charts; however, each x-value represents a *non-discrete quantity* or “range” rather than a discrete category and each y-value measures the *frequency* of each x-axis quantity, not a value. A value’s frequency is directly proportional to its probability, and by plotting a distribution using a histogram, a data scientist can easily visualize the probability of certain events or instances occurring. A data scientist can also quickly visualize the mean, standard deviation, mode, and percentiles based on the shape of the distribution.

Histograms are commonly used for three purposes, though these are by no means the only ones: 1) to determine how process occurrences change between two years, 2) to determine the most common probabilities for a process, or 3) to determine the shape of a distribution (normal distributions are easier to work with, as we will learn). Below is an example taken from the book *Think Stats,* where the frequency of each pregnancy length is shown. Immediately, it is clear that 39 weeks is the most common pregnancy length with a very low spread seen between lengths. We can conclude based on the shape alone that a pregnancy length more than a few weeks away from 39 weeks is abnormal. Sure enough, the actual mean and standard deviation were found to be 38.6 and 2.7 weeks, paralleling what we could determine from the histogram.

To summarize, bar charts compare values across individual categories while histograms visualize the probability distribution of a dataset. They are like apples and oranges; while they may have some similarities, they are used for different purposes.

References:

Downey, A. B. (2014). *Think Stats.* O’Reilly Media.

Duvey, U. (2020). *Is there a difference between a bar chart and a histogram?* Quora. Retrieved on September 12, 2020 from https://www.quora.com/Is-there-a-difference-between-a-bar-chart-and-a-histogram.

Surbhi, S. (2016). *Difference Between Histogram and Bar Graph.* Key Differences. Retrieved on September 12, 2020 from https://keydifferences.com/difference-between-histogram-and-bar-graph.html.

*Top Ten Oil Consuming Countries* (2014). Plotly Chart Studio. Retrieved on September 12, 2020 from <https://chart-studio.plotly.com/~11006/36.embed>