

(/)



Curriculum

Professional Foundations ^

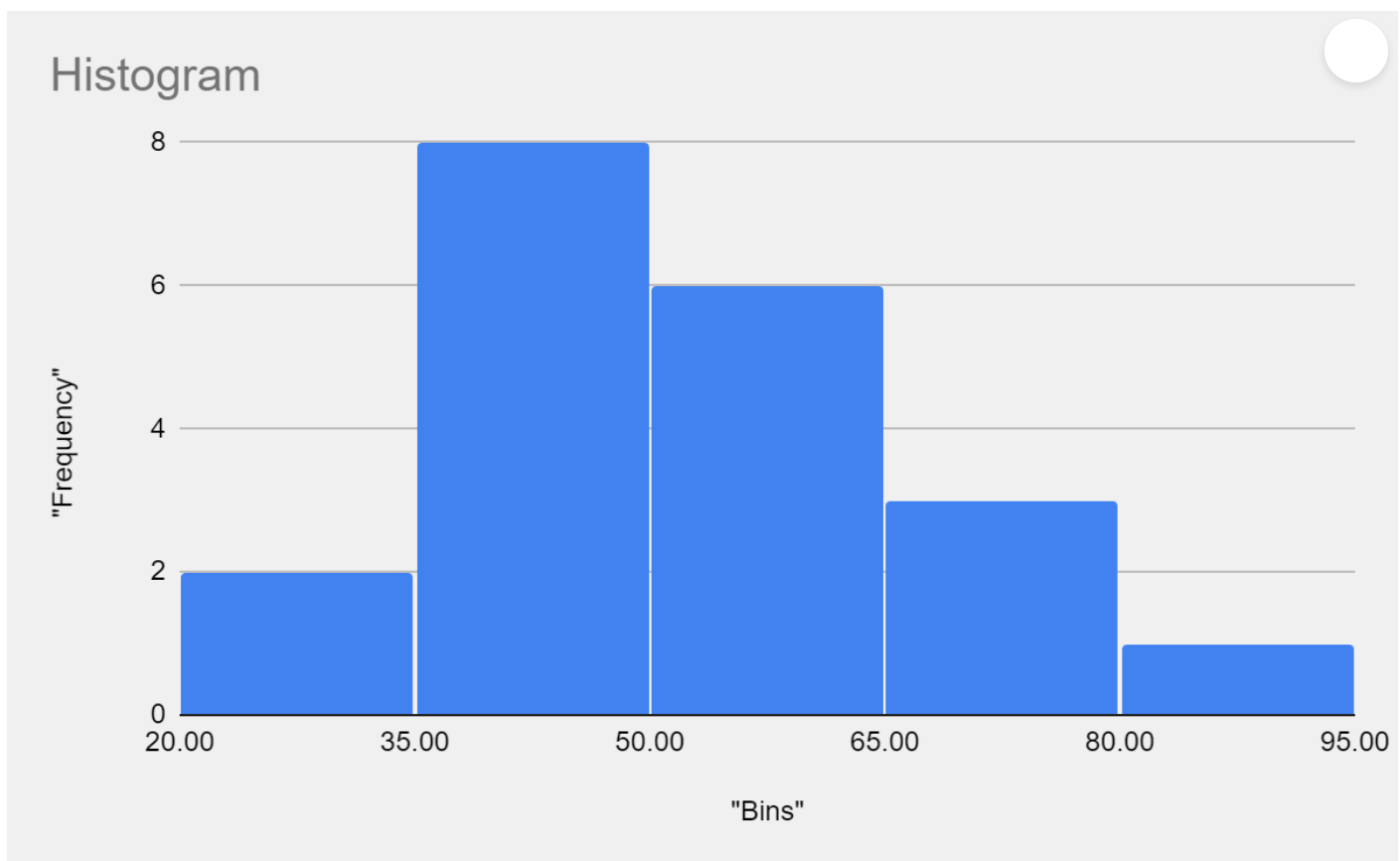
Average: 97.49% v

Week 5 ≡

More Charts & Graphs: Histogram

Histogram

A histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of **continuous data**. This allows the inspection of the data for its underlying distribution (e.g., normal distribution), outliers, skewness, etc. An example of a histogram, and the raw data it was constructed from, is shown below.



Continuous data refers to data that can take any value within a given range. It can be measured with great precision and includes decimal or fractional values. Continuous data is essentially infinite and allows for a smooth and unbroken spectrum of possibilities. Examples of continuous data include height, weight, temperature, and distance.

How do you construct a histogram from a continuous variable?



Let's use the below data to understand how to create a histogram.

(/)

25 22 36 38 46 45 48 46 55 55 52 58 55 68 67 61 72 80 90 91

To construct a histogram from a continuous variable you first need to split the data into intervals, called bins. In the example above, age has been split into bins, with each bin representing 15 years starting at 20 years. Each bin contains the number of occurrences of scores in the data set that are contained within that bin. For the above data set, the frequencies in each bin have been tabulated along with the scores that contributed to the frequency in each bin (see below):

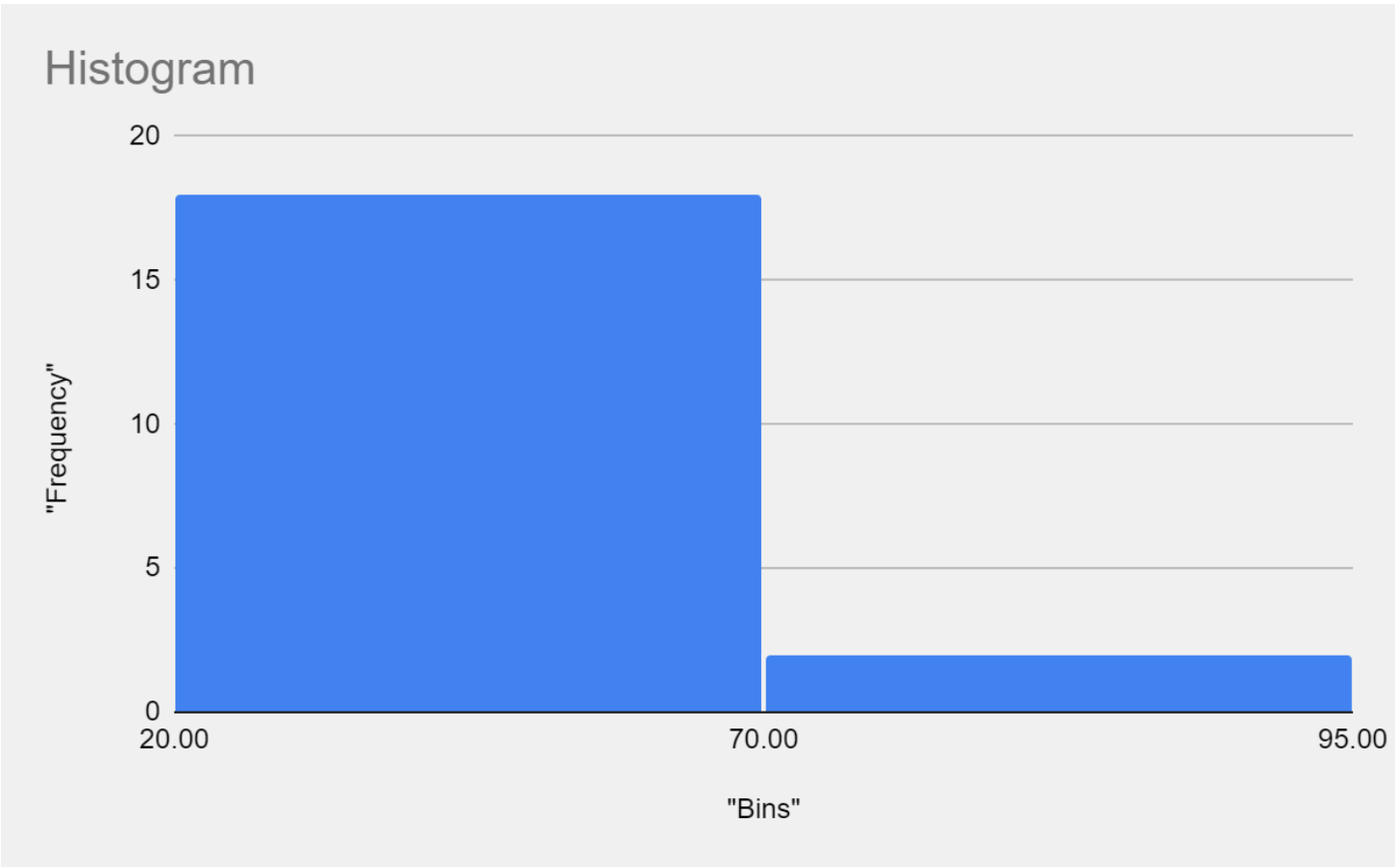
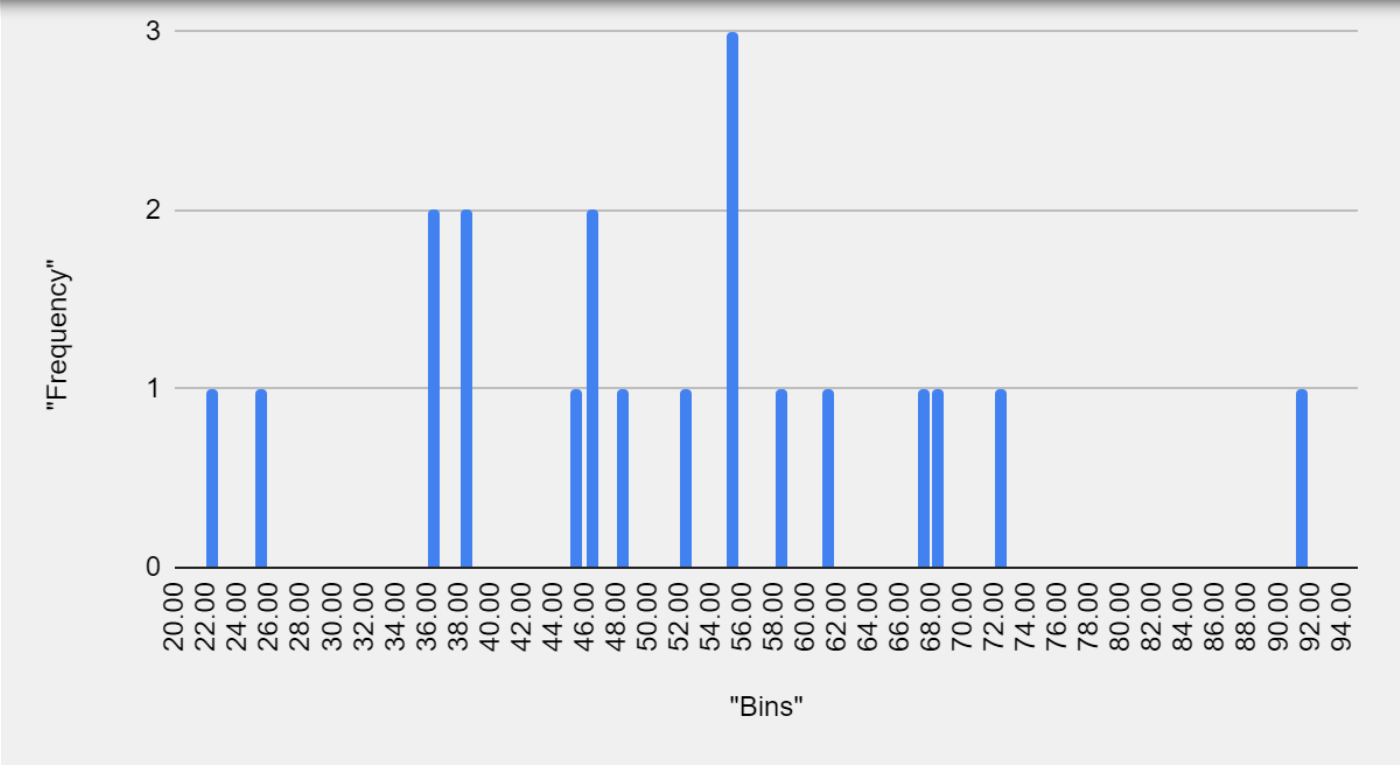
Bin	Frequency	Scores Included in Bin
20-30	2	25,22
30-40	4	36,38,36,38
40-50	4	46,45,48,46
50-60	5	55,55,52,58,55
60-70	3	68,67,61
70-80	1	72
80-90	0	-
90-100	1	91

Notice that, unlike a bar chart, there are no “gaps” between the bars (although some bars might be “absent” reflecting no frequencies). This is because a histogram represents a continuous data set, and as such, there are no gaps in the data (although you will have to decide whether you round up or round down scores on the boundaries of bins).

Choosing the correct bin width

There is no right or wrong answer as to how wide a bin should be, but there are rules of thumb. You need to make sure that the bins are not too small or too large. Consider the histogram we produced earlier (see above), then look at the following histograms using the same data, but have either much smaller or larger bins, as shown below:

(/)



We can see from the histogram on the left that the bin width is too small because it shows too much individual data and does not allow the underlying pattern (frequency distribution) of the data to be easily seen. At the other end of the scale is the diagram on the right, where the bins are too large, and again, we are unable to find the underlying trend in the data.

(/)

alone doesn't tell the full story—it's the combined height and width of the bar that gives the complete picture. Sometimes people think the height alone shows how often things occurred, but that's only true when all the bars are the same width. If the bars have different widths, you need to look at both the height and the width to understand the frequency.

What is the difference between a bar chart and a histogram?

The major difference is that a histogram is only used to plot the frequency of score occurrences in a continuous data set that has been divided into classes, called bins. Bar charts, on the other hand, can be used for a great deal of other types of variables including ordinal and nominal data sets.

This is how to make and use a histogram chart on Excel or Google Sheets.

How to Make Histograms

How to Make a Histogram in Excel 2016

**References:**

1. <https://youtu.be/is14ehdy7jo>

(/)

(/concepts/104336?project_id=101000)

(/concepts/104338?project_id=101000)

Copyright © 2024 ALX, All rights reserved.