

2.9: Graphing Quantitative Data- Histograms

Another common type of graph for frequency distributions for quantitative variables (interval or ratio scales of measurement) is a histogram.

? Exercise 2.9.1

What other type of chart shows a frequency distribution for quantitative variables? Hint: We've already talked about it.

Answer

Line graph (sometimes called a frequency polygon)

? Exercise 2.9.1

Histograms look more like bar graphs, but are actually more like line graphs. Why?

Answer

Your answer probably talks about how the bars in bar charts don't touch, but the bars (or bins) in histograms do touch. Your answer should say something about the type of variable: quantitative variables (ratio or interval scale of measurements) should be graphed with histograms, and qualitative variables (nominal scale of measurement) should be graphed with bar charts.

Figure 2.9.1 shows a histogram (created by SPSS) of the frequency distribution of the final exam score data from Table 2.2.2 (and also shown in the line graph in Figure 2.8.1).

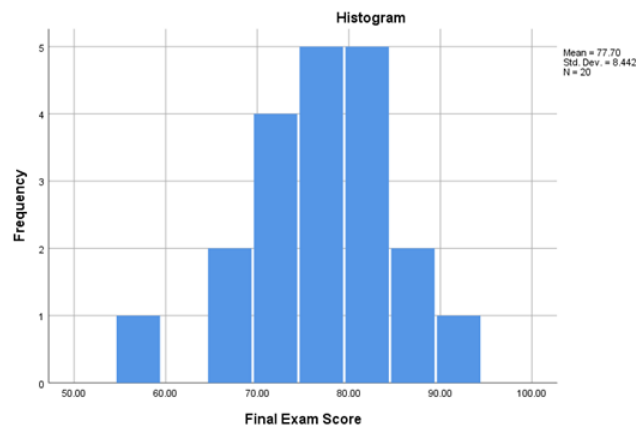


Figure 2.9.1: Histogram of Frequency of Final Exam Scores. (Copyright CC-SA, chart by Dr. MO via data provided by [OpenIntro.org](https://openintro.org))

As with all frequency distributions, the y-axis of a histogram shows the frequency of each score (how many people had each score) and the x-axis has each score. Figure 2.9.1 shows the number of students who in each score category on the y-axis (Frequency), and labels the score categories on the x-axis (Final Exam Score, with each 10 points labeled). To make the histogram, we just count up the number of data points falling inside each bin, then plot those frequency counts as a function of the bins. Voila, a histogram. The 10-point categories are called bins. Bin #1 goes from about 55 to about 59 points, bin #2 goes about 65 to 69 points, and so on until the last bin. The difference in the x-axis between a line graph and a histogram is that the scores are combined into ranges, or categories, in a histogram. This makes it easier to quickly understand what a graph is showing if there are a lot of different scores (like everyone's GPA in your class), but it does lose some accuracy.

Each bar (bin), and can be used to answer questions about the frequency of scores within these bin categories. For example, how many people scored between 90 to 100 points on the Final Exam? The seventh bar (or first on the right), the one between 90 and 100 on the x-axis, tells you how many. Look how tall that bar is. How tall is it? The height is shown on the y-axis, which provides the frequency. One person scored between 90 points and 100 points on the Final Exam.

Just like a line graph, a histogram can show the shape (kurtosis) or skew, and the range of our data set. The shape of the histogram refers to how it goes up and down. The shape tells us where the data is. For example, when the bars are low we know there isn't much data there. When the bars are high, we know there is more data there. So, where is most of the data? It looks like it's mostly in the middle two bins, between 75-85 points. The range of the data tells us the minimum (lowest score) and the maximum (highest score) of the data. In Figure 2.9.1, most of the scores are between 65 points and 95 points, but the minimum scores is between 55-59 points.

When you make a histogram you get to choose how wide each bar will be. For example, below are four different histograms of the same data (a happiness survey). What changes is the width of the bins.

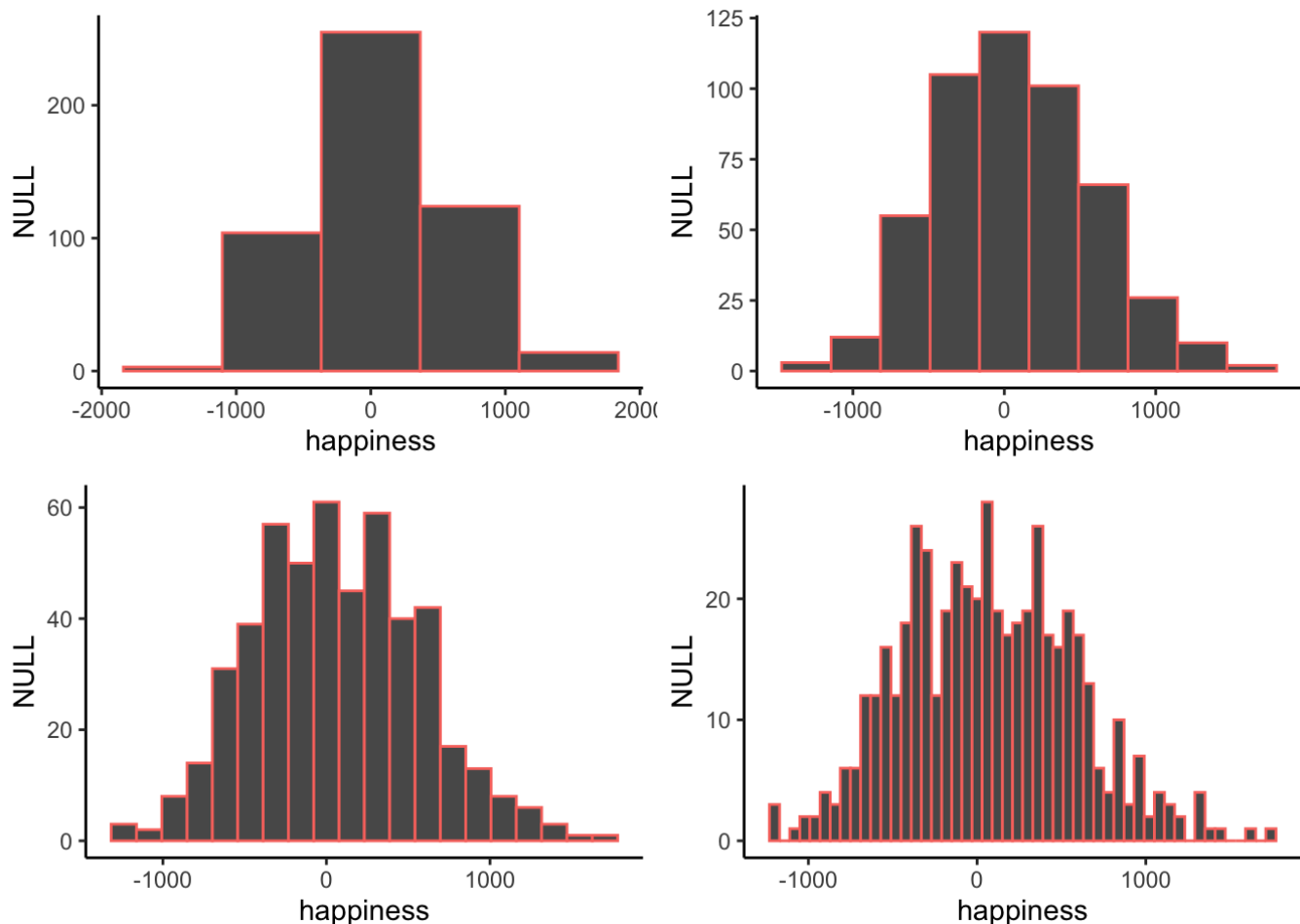


Figure 2.9.2: Four histograms of the same data using different bin widths. (CC-BY-SA [Matthew J. C. Crump](#) from [Answering Questions with Data- Introductory Statistics for Psychology Students](#))

All of the histograms have roughly the same overall shape: From left to right, the bars start off small, then go up, then get small again. In other words, as the numbers get closer to zero, they start to occur more frequently. We see this general trend across all the histograms. But, some aspects of the trend fall apart when the bars get really narrow. For example, although the bars generally get taller when moving from -1000 to 0, there are some exceptions and the bars seem to fluctuate a little bit. When the bars are wider, there are less exceptions to the general trend. How wide or narrow should your histogram be? It's a Goldilocks question. Make it just right for your data.

Students often get confused between a bar graph and a histogram because they both have bars. There are two differences, however. First, in a histogram, the bars must touch; there should **not** be space between the categories that were created from the quantitative data. This leads to the second difference between histograms and bar charts: Histograms show quantitative variables, while bar charts show qualitative variables (nominal scale of measurement).

Histogram Interpretation

The following are the same questions asked about the Figure 2.8.1, the line graph showing the frequency distribution of this same data. You might find it interesting how the answers change a little depending on how the data is presented.

1. What kind of graph is Figure 2.9.1?
 1. This is a histogram. Histograms are used to display quantitative data, but it combines each score into groups. In Figure 2.xxx, the groups are 10-point ranges.
2. What does the x-axis measure in Figure 2.9.1?
 1. The x-axis on histograms show categories based on each score. In Figure 2.9.1, the software created 10-point categories for each Final Exam Score that a student earned.
3. What does the y-axis measure in Figure 2.9.1?
 1. The y-axis is the axis that goes up and down. For most charts discussed in this textbook, the y-axis will be frequencies. In Figure 2.9.1, that means that the y-axis shows how many students earned scores on the Final Exam in each 10-point range.
4. Is Figure 2.9.1 skewed? If so positively or negatively? If not, is the graph tall/narrow, medium/normal, or wide/flat?
 1. Because there is a gap between the lowest score category and the next one, I would say that this histogram is negatively skewed. However, it looks less skewed than the same data in the line graph in Figure 2.8.1. In fact, just by looking at Figure 2.9.1 and not knowing the exact scores, I might have said that it is not skewed and that it has approximately normal kurtosis (mesokurtic).
5. What do you notice from Figure 2.9.1? What pops out to you?
 1. Two things that pop out to me. First, there seemed to be the most scores around the middle of the distribution. The second thing I notice is that one column that's to the left. This shows that one student (frequency = 1) score below 60 points on the Final Exam.
6. What does Figure 2.9.1 make you wonder about?
 1. This histogram still makes me wonder if the professor ended up "curving" the exam so that the highest earning score was now like earning 100% on the Final Exam. If I haven't said this before, THIS IS NOT MY CLASS.
7. What is a catchy headline for Figure 2.9.1?
 1. Final Exam Scores are Normal [This is catchy if you know what a Normal Distribution is; more in Ch. 4!]
8. How could you summarize the info in Figure 2.9.1 into one sentence?
 1. The class did really well on the Final Exam, although one student scored pretty low.
9. Who might want to know the information in Figure 2.9.1?
 1. I still am guessing that students who are going to take this class from this professor might be interested. College administrators might also want to know so that they can see that most students are passing their classes.

Summary

Making a histogram is our first act of officially *summarizing* the data. We are no longer look at the individual bits of data, instead we will see how the numbers group together. Let's look at a histogram of the happiness data, and then explain it.

Contributors and Attributions

- Matthew J. C. Crump (Brooklyn College of CUNY)
-

Dr. MO (Taft College)

This page titled 2.9: Graphing Quantitative Data- Histograms is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michelle Oja.

- 2.2: Look at the data by Matthew J. C. Crump is licensed CC BY-SA 4.0. Original source: <https://www.crumplab.com/statistics/>.