

CHAPTER OVERVIEW

5: The Distribution of Data

When we measure something, such as the percentage of yellow M&Ms in a bag of M&Ms, we expect two things:

- that there is an underlying “true” value that our measurements should approximate, and
- that the results of individual measurements will show some variation about that “true” value

Visualizations of data—such as dot plots, stripcharts, boxplot-and-whisker plots, bar plots, histograms, and scatterplots—often suggest there is an underlying structure to our data. For example, we saw in Chapter 3 that the distribution of yellow M&Ms in bags of M&Ms is more or less symmetrical around its median, while the distribution of orange M&Ms was skewed toward higher values. This underlying structure, or distribution, of our data as it effects how we choose to analyze our data. In this chapter we will take a closer look at several ways in which data are distributed.

[5.1: Terminology](#)

[5.2: Theoretical Models for the Distribution of Data](#)

[5.3: The Central Limit Theorem](#)

[5.4: Modeling Distributions Using R](#)

[5.5: Exercises](#)

This page titled [5: The Distribution of Data](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [David Harvey](#).