

## 11.1: What Do We Mean By Structure?

The signals we measure include contributions from determinate and indeterminate sources, with the determinate components resulting from the analytes in our sample and with the indeterminate sources resulting from noise. When we describe our data as having structure, or that we are looking for structure in our data, our interest is in the determinate contributions to the signal. Consider, for example, the data in the following figure, which shows the visible spectra for 24 samples at 635 wavelengths.

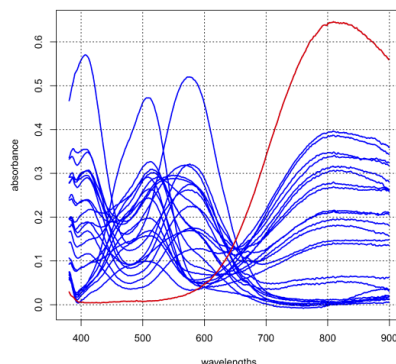


Figure 11.1.1: Visible spectra for 24 samples recorded at 635 wavelengths between 380.5 nm and 889.5 nm. The spectrum in red highlights one of the 24 spectra included in this data set.

Each curve in this figure, such as the one shown in red, is one of the 24 samples that make up this data set and shows the extent to which each of the 635 discrete wavelengths of light are absorbed by that sample: this is the determinate contribution to the data. Looking closely at the spectrum shown in red, we see small variations in the absorbance superimposed on the determinate signal: this is the indeterminate contribution to the data.

Although when first examined, the 24 spectra in Figure 11.1.1 may create a sense of disorder, there is a clear underlying structure to the data. For example, there are four apparent peaks centered at wavelengths around 400 nm, 500 nm, 580 nm, and 800 nm. Each of the individual spectra include one or more of these peaks. Further, at a wavelength of 800 nm, we see that some samples show no absorbance, and presumably lack whatever analyte is responsible for this peak; other samples, however, clearly include contributions from this analyte. This is what we mean by finding structure in data. In this chapter we explore three tools for finding structure in data—cluster analysis, principal component analysis, and multivariate linear regression—that allow us to make sense of that structure.

This page titled [11.1: What Do We Mean By Structure?](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [David Harvey](#).