

### 3.1: Visualize the Data

The first step in this one-factor modeling process is to determine whether or not it *looks* as though a linear relationship exists between the predictor and the output value. From our understanding of computer system design that is, from our *domain-specific knowledge* we know that the clock frequency strongly influences a computer system's performance. Consequently, we must look for a roughly linear relationship between the processor's performance and its clock frequency. Fortunately, R provides powerful and flexible plotting functions that let us visualize this type relationship quite easily.

This R function call:

```
> plot(int00.dat[, "clock"], int00.dat[, "perf"], main="Int2000", xlab="Clock", ylab="Performance")
```

generates the plot shown in Figure 3.1. The first parameter in this function call is the value we will plot on the x-axis. In this case, we will plot the clock values from the int00.dat data frame as the independent variable

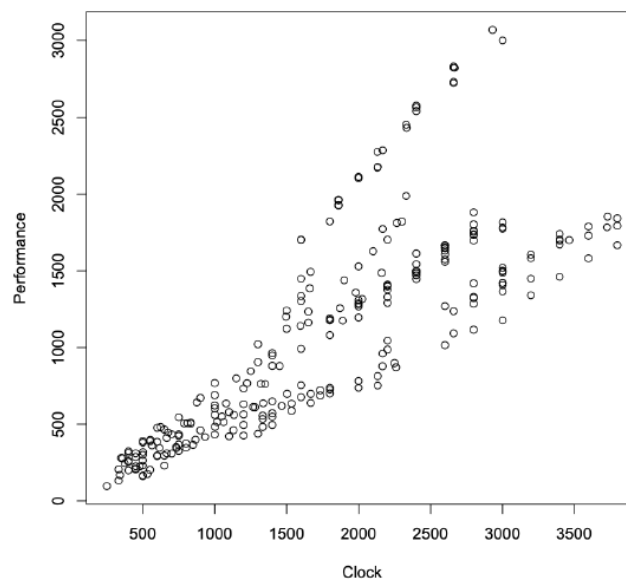


Figure 3.1: A scatter plot of the performance of the processors that were tested using the Int2000 benchmark versus the clock frequency.

on the x-axis. The dependent variable is the perf column from `int00.dat`, which we plot on the y-axis. The function argument `main="Int2000"` provides a title for the plot, while `xlab="Clock"` and `ylab="Performance"` provide labels for the x and y-axes, respectively.

This figure shows that the performance tends to increase as the clock frequency increases, as we expected. If we superimpose a straight line on this scatter plot, we see that the relationship between the predictor (the clock frequency) and the output (the performance) is roughly linear. It is not perfectly linear, however. As the clock frequency increases, we see a larger spread in performance values. Our next step is to develop a regression model that will help us quantify the degree of linearity in the relationship between the output and the predictor.

This page titled [3.1: Visualize the Data](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [David Lilja \(University of Minnesota Libraries Publishing\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.