Homework 1

Chapter 1 HW:  
Consider Figure 1.1 (page 2) with respect to some particular scientific research study, such as a research project that the reader might be undertaking.

What is involved in the data analysis of the study at all its stages, from processing raw data to eventual communication of results?

* The researchers first had to aggregate the data by group and time-period across the range of patients. As it appears from the figure, the raw data was compiled in a tabular fashion to produce a table with columns Patient, group, 0-30 days, 0-365 days. The interaction being studied appears to be the effect of treatment on the incidence of stroke over the course of 0-30 days and 0-365 days. This figure communicates enough of the important context of the study without displaying enough data to allow the user to immediately come to their own conclusion about the results.

How much time is likely to be needed at each stage?

* Initial storage and aggregation of the data should be brief, about fifteen minutes assuming it already exists in a csv friendly format. Analyzing the data may require some research regarding the important outcomes of medical studies related to this paper. This will effect the tools used to visualize the data and what conclusion to draw. Ideally, given the number of samples in the study, this would take about 3 hours.

What specifically needs to be done at each stage?

* Stage 1: Store the data.
* Stage 2: Load the data into the tool of choice (excel, RStudio, etc.).
* Stage 3: Analyze the data for trends, outliers, or other insights.
* Stage 4: Create a visual to effectively communicate the conclusion.

Will R be sufficient for all the stages of analysis of the study?

* R would be sufficient for the study, as it has all of the required tools to analyze this relatively simple dataset.
  + This statement assumes that the dataset already exists in a tabular state and it can be imported into R.

What other tools are required?

* Excel or python may be required depending on the initial format of the data. These tools would be used to tidy up the raw data for analysis.