Graphics Contents[1]

In his book, *The Grammar of Graphics*, Wilkinson gives us a language for describing the contents of a graphic. He decomposes a graphic into seven basic elements as follows:

DATA Functions that create variables from data. We often do not need to specify any function (aside from the identity) if the data are in a data frame.

TRANS Transformations, if any, to be applied to the variables, e.g. percentile rank, log, inverse of normal CDF, cut/group, etc.

FRAME Algebraic expression that defines the frame. A one-dimensional frame is typically specified by a single variable, e.g. **x**. A two dimensional frame by **x*y**, where * is interpreted as "cross" or "by" or "Cartesian product." Other symbols include, + which stands for "blend" or "union", and / which denotes stratification or nesting.

SCALE Dimensions on which the graphics orient themselves, such as categorical, interval, log, and power. In addition, the scale contains information about tick mark locations and format (e.g. scientific notation).

COORD Coordinate system to use, such as polar and cartesian, plus, if needed, information about how to reflect, rotate, stretch, dilate, and translate.

GUIDE Guiding notation such as axes, legends, markers, etc.

GRAPH Graphing functions to appear in the frame. Simple examples include, *point* and *line*.

A few concrete examples, help convey what these elements mean. Consider the plot of cross-validated Type I and II errors for the log odds ratio from the Niave Bayes classifier applied to the ling spam data (Figure 1). this graphic would be described as follows:

DATA: **logOdds, type1, type1**

FRAME: **logOdds * (type1 + type2)**

SCALE: interval(dim1 from -30 to 60)

SCALE: interval(dim2 from 0 to 0.05)

GUIDE: line(vertical at -20)

GUIDE: line(horizontal at type I error for logOdds = -20)

GUIDE: line(horizontal at type II error for logOdds = -20)

GUIDE: Other guides include legend, title, ylabel, xlabel

[1] Adapted from Wilkinson's book, The Grammar of Graphics

GRAPH: line(color("Type1" + "Type II"))

FRAME: **logOdds * (type1 + type2)**

SCALE: interval(dim1 from -150 to 250)

SCALE: interval(dim2 from 0 to 0.8)

COORD: stretch(dim1,dim2,0.3,0.5)

COORD: translate(dim1,dim2,0.6,0.3)

GRAPH: line(color("Type1" + "Type II"))

The language used to describe the elements is informal and only meant to be a general guideline for the content. What is important is to note that ideas behind the description of the graphic. Note that the elements TRANS and COORD do not appear in the specification of the first graphic because the variables do not need to be transformed and the coordinate system is the default. The first FRAME statement says that the variables type1 and type2 will blend, but they are differentiated by the color as specified in the GRAPH element. Not all of the GUIDEs are listed here for reasons of space, but note that the markers are noted as they are atypical guides.

The second graphic is superposed on the first. This is denoted via the COORD element, which indicates that this graphic is shrunk in both dimensions (stretch) and positioned in the first frame (translate).