The idea is: we sacrifice some of the flexibility and composability inherent to ggplot2 in R for a menu of prescribed presentation solutions. This is a package to produce plots while you are in the middle of another task.

For example the plot below showing both an observed discrete empirical distribution (as stems) and a matching theoretical distribution (as bars) is a built in "one liner."

set.seed(52523)

d <- data.frame(wt=100\*rnorm(100))

WVPlots::PlotDistCountNormal(d,'wt','example')

The graph above is actually the product of a number of presentation decisions:

* Using a discrete histogram approach to summarize data (instead of a kernel density approach) to create a presentation more familiar to business partners.
* Using a Cleveland style dot with stem plot instead of wide bars to emphasize the stem heights represent total counts (and not the usual accidental misapprehension that bar *areas* represent totals).
* Automatically fitting and rendering the matching (properly count-scaled) normal distribution as thin translucent bars for easy comparison (again to try and de-emphasize area).

All of these decisions are triggered by choosing which plot to use from the WVPlots library. In this case we chose WVPlots::PlotDistCountNormal. For an audience of analysts we might choose an area/density based representation (by instead specifying WVPlots::PlotDistDensityNormal) which is shown below:

WVPlots::PlotDistDensityNormal(d,'wt','example')

Switching the chosen plot simultaneously changes many of the details of the presentation. WVPlots is designed to make this change simple by insisting an a very simple unified calling convention. The plot calls all insist on roughly the following arguments:

* frame: data frame containing the data to be presented.
* xvar: name of the x variable column in the data frame.
* yvar: name of the y variable column in the data frame (not part of the shown density plots!).
* title: text title for the plot.

This rigid calling interface is easy to remember and makes switching between plot types very easy. We have also make title a required argument, as we feel all plots should be labeled.

What we are trying to do is separate the specification of exactly what plot we want from the details of how to produce it. We find this separation of concerns and encapsulation of implementation allows us to routinely use rich annotated graphics. Below are a few more examples:

set.seed(34903490)

x = rnorm(50)

y = 0.5\*x^2 + 2\*x + rnorm(length(x))

frm = data.frame(x=x,y=y,yC=y>=as.numeric(quantile(y,probs=0.8)))

frm$absY <- abs(frm$y)

frm$posY = frm$y > 0

WVPlots::ScatterHist(frm, "x", "y", smoothmethod="lm",

title="Example Linear Fit")

set.seed(34903490)

y = abs(rnorm(20)) + 0.1

x = abs(y + 0.5\*rnorm(20))

frm = data.frame(model=x, value=y)

frm$costs=1

frm$costs[1]=5

frm$rate = with(frm, value/costs)

frm$isValuable = (frm$value >= as.numeric(quantile(frm$value, probs=0.8)))

gainx = 0.10 # get the top 10% most valuable points as sorted by the model

# make a function to calculate the label for the annotated point

labelfun = function(gx, gy) {

pctx = gx\*100

pcty = gy\*100

paste("The top ", pctx, "% most valuable points by the model\n",

"are ", pcty, "% of total actual value", sep='')

}

WVPlots::GainCurvePlotWithNotation(frm, "model", "value",

title="Example Gain Curve with annotation",

gainx=gainx,labelfun=labelfun)

set.seed(52523)

d = data.frame(meas=rnorm(100))

threshold = 1.5

WVPlots::ShadedDensity(d, "meas", threshold, tail="right",

title="Example shaded density plot, right tail")

set.seed(34903490)

frm = data.frame(x=rnorm(50),y=rnorm(50))

frm$z <- frm$x+frm$y

WVPlots::ScatterHistN(frm, "x", "y", "z", title="Example Joint Distribution")

set.seed(34903490)

x = rnorm(50)

y = 0.5\*x^2 + 2\*x + rnorm(length(x))

frm = data.frame(x=x,yC=y>=as.numeric(quantile(y,probs=0.8)))

WVPlots::ROCPlot(frm, "x", "yC", TRUE, title="Example ROC plot")

We know this collection doesn’t rise to the standard of a complete "grammar of graphics." But it can become (through accumulation) a re-usable repository of a number of specific graphing tasks done well. It is also a chance to eventually document presentation design decisions (though we haven’t gotten far on that yet). The complete set of graphs is shown in the WVPlots\_example vignette.