

# ggplot2

*Drew Schmidt*

*04/2/2015*

1. Install the ggplot2 package.

The remaining exercises refer to the diamonds dataset. Before exploring it, you need to run:

```
library(ggplot2)
data(diamonds)
```

2. Create a histogram of the “carat” variable.
3. Save the plot you just made as both a pdf and a png.
4. Begin with:

```
g <- ggplot (data=diamonds, aes(x=clarity) )
```

Produce a barplot and a histogram with g (remember, clarity is categorical). Is there a difference?

5. Create scatterplots of price by carat faceted by color. How would you describe the relationship between price and carat across groups?
6. Every plot should tell a story. What story do our scatterplots tell about a diamond’s carat and its price? (Just a short, one sentence explanation; if this seems tricky, you’re probably overthinking it.)
7. Refer to the subset plot above where we restricted the data only to those diamonds with color “J”. Produce a scatterplot with a LOESS fit in the same plot. Do you notice anything striking in this plot (you may have noticed it in another plot above)?
8. Pushing multiple plots to the same “viewport” is not trivial (though there are several packages that help with this, such as `gridExtra`). Work through the following example and try to make sense of what it is doing:

```
g1 <- ggplot(data=diamonds, aes(x=clarity, y=price) )
g2 <- ggplot(data=diamonds, aes(x=price))
```

```
first <- g1 +
  geom_boxplot() +
  ggtitle("Boxplots") +
  scale_y_log10("Log-scale Price") +
  xlab("Clarity")
```

```
second <- g1 +
  geom_jitter(size=1, aes(color=color)) +
  ggtitle("Jitter Plot") +
  xlab("Clarity") +
  ylab("Price")
```

```

third <- g2 +
  geom_histogram(aes(y=..density..), color="black", fill="gray") +
  geom_density(color="black") +
  facet_wrap(~clarity) +
  ggtitle("Densities") +
  xlab("Price") +
  ylab("Density")

# Using grid to control placement of plots
library(grid)
grid.newpage() # ready the plotting device
pushViewport(viewport(layout = grid.layout(2, 2))) # set up plot layout, here a 2x2 grid
print(first, vp=viewport(layout.pos.row=1, layout.pos.col = 1)) # plot 'first' in square (1,1)
print(second, vp=viewport(layout.pos.row=2, layout.pos.col = 1)) # plot 'second' in square (2,1)
print(third, vp=viewport(layout.pos.row=1:2, layout.pos.col = 2)) # plot 'third' in rectangle (1-2, 2)

```

## Suggested Solutions to Selected Exercises

1. `install.packages("ggplot2")`

2. `ggplot(data=diamonds, aes(x=carat) ) + geom_hist()`

3. `ggsave ("plot_pdf.pdf", last_plot()) # pdf`  
`ggsave ("plot_png.png", last_plot()) # png`

4.

```
g + geom_bar() g + geom_histogram() \end{lstlisting}
```

5.

```
ggplot(diamonds, aes(carat, price)) + geom_point() + facet_wrap(~color)
```

The relationship between price and carat looks similar across all colors.

6. There appears to be a fairly strong, positive association between a diamond's carat and its price.

7. `g_subset + geom_point() + geom_smooth()`

The data appears to contain some outliers with extremely high leverage.