

Graphics the ggplot2 Way

Drew Schmidt

April 2, 2015

Contents

- Plotting Data in R
- Introduction to ggplot2
- Deconstructing a ggplot2 Plot
- Using ggplot2
- Advanced Features and Tweaks
- Wrapup

Plotting Data in R

Some Options

- Base graphics
- Grid
- Lattice
- ggplot2
- And [about 40 other packages](#)

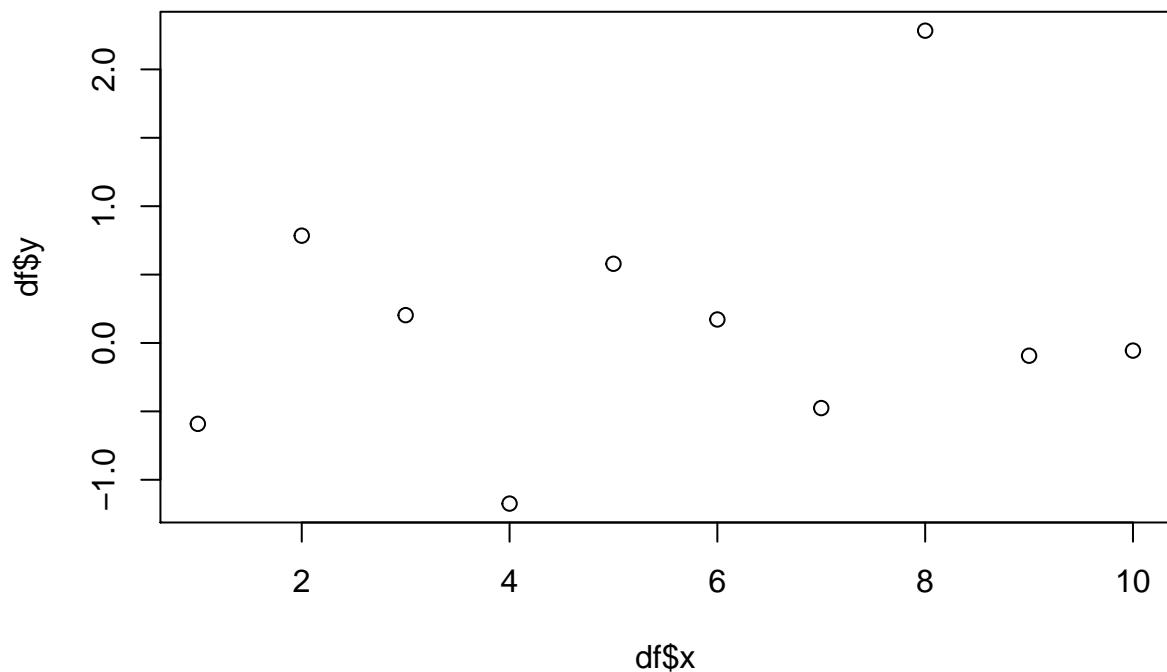
```
df <- data.frame(x=1:10, y=rnorm(10))
head(df)
```

```
##   x         y
## 1 1 -0.5912573
## 2 2  0.7845191
## 3 3  0.2034944
## 4 4 -1.1740205
## 5 5  0.5790318
## 6 6  0.1718351
```

Base Graphics

Makes easy things easy, and hard things impossible.

```
plot(df$x, df$y)
```



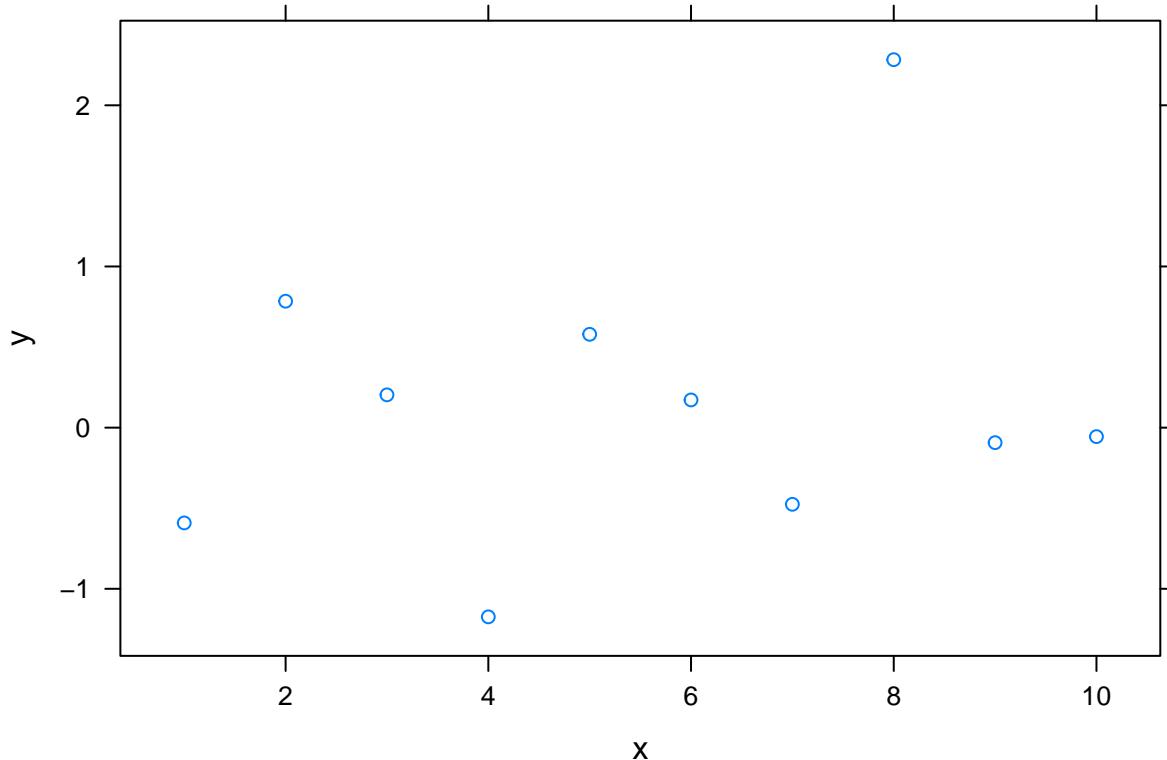
Grid

“Assembly language” for graphics.

Lattice

A powerful system for producing graphics,

```
library(lattice)
xyplot(y~x, df)
```



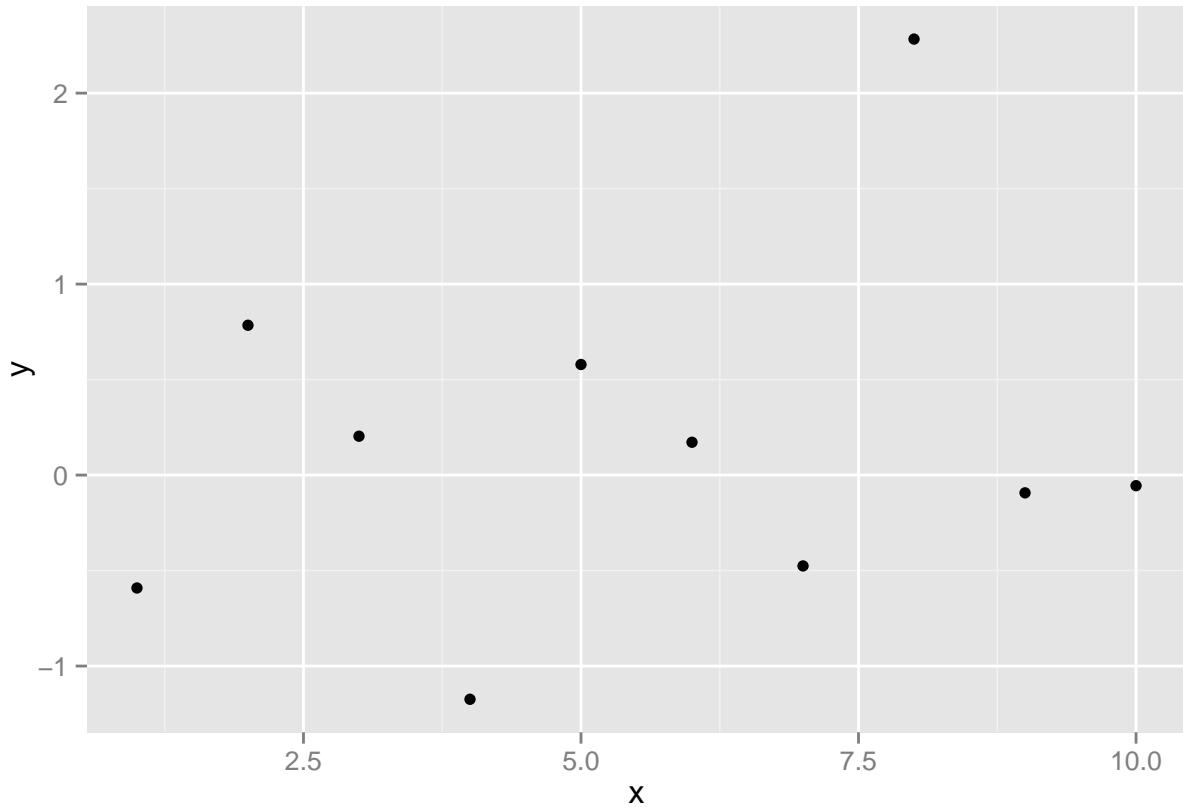
ggplot2

Powerful and flexible; a darling of the R community.

```
library(ggplot2)

## Loading required package: methods

ggplot(df, aes(x, y)) + geom_point()
```



Introduction to ggplot2

Ways of Plotting

- `qplot()` — Shorthand
- Layer syntax — Extremely rare “in the wild”.
- geom/stat syntax — Most popular method.
- autoplot — Methods (as in OO) for plotting objects.

General Procedure

- Put your data into a dataframe.
- Declare which dataframe ggplot should use and set aesthetics.
- Add layers via geom/stat functions.
- Set options, faceting, labels, titles, ...

?????

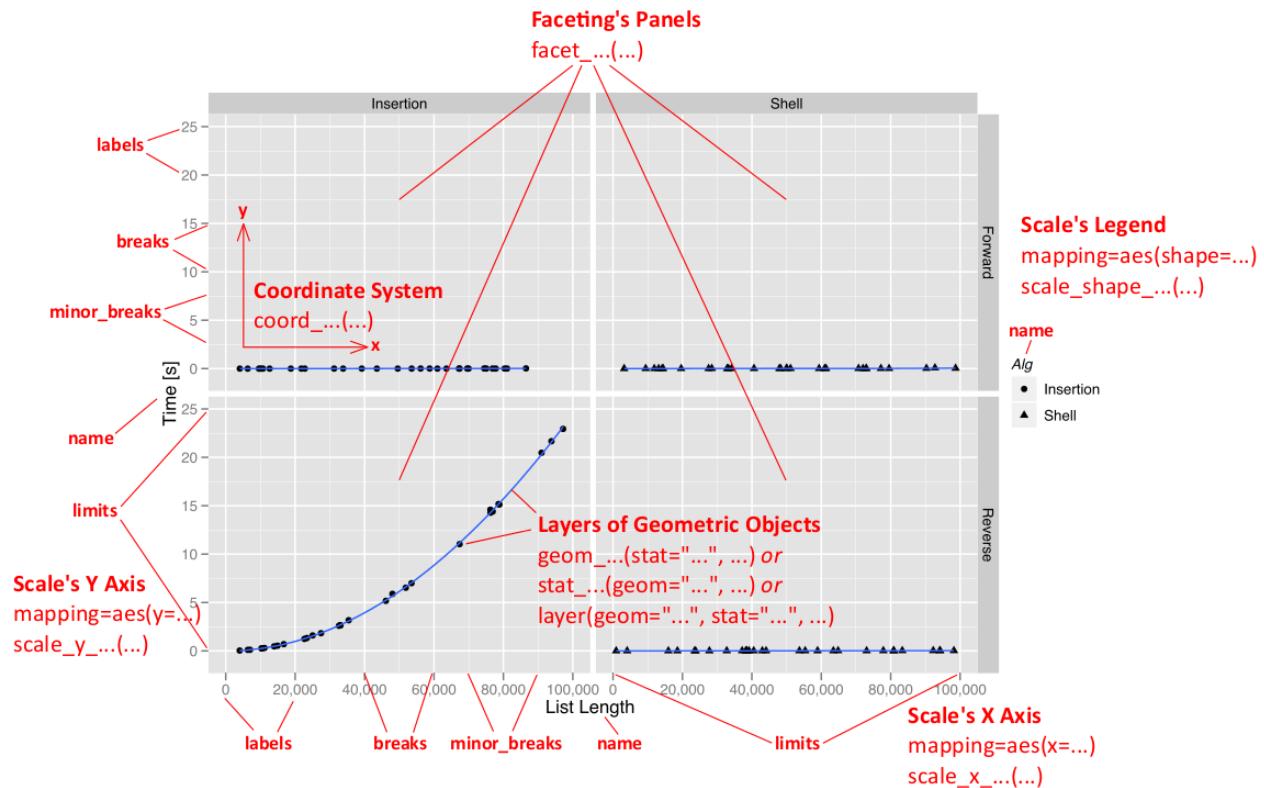
Geoms? Aesthetics? Facets?

Deconstructing a ggplot2 Plot

4 Components

1. Geoms — physical components (point, line, path, polygon)
2. Aesthetics — visual cues (size, rotation, thickness, gradient, shape, color)
3. Coordinates
4. Faceting — coplotting (more on this later)

Customization Nearly Unlimited



Geom/Stat Functions

```
geom_abline geom_aesthetics geom_area geom_bar geom_bin2d geom_blank geom_boxplot geom_contour
geom_crossbar geom_density geom_density2d geom_dotplot geom_errorbar geom_errorbarh geom_freqpoly
geom_hex geom_histogram geom_hline geom_jitter geom_line geom_linerange geom_map geom_path
geom_point geom_pointrange geom_polygon geom_quantile geom_raster geom_rect geom_ribbon
geom_rug geom_segment geom_smooth geom_step geom_text geom_tile geom_violin geom_vline

stat_abline stat_aesthetics stat_bin stat_bin2d stat_bindot stat_binhex stat_boxplot stat_contour
stat_density stat_density2d stat_ecdf stat_ellipse stat_function stat_hline stat_identity stat_qq
stat_quantile stat_smooth stat_spoke stat_sum stat_summary stat_summary2d stat_summary_hex
stat_unique stat_vline stat_ydensity
```

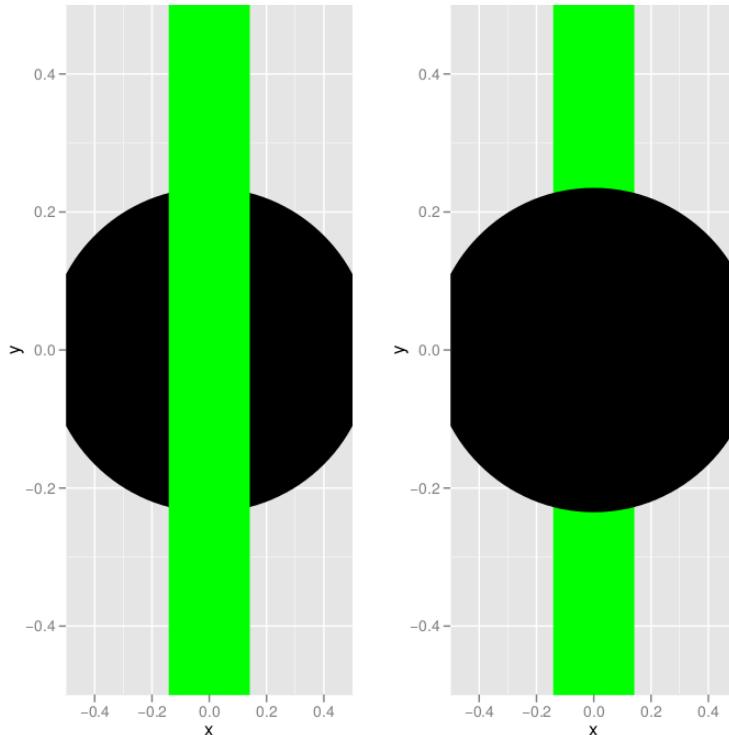
For explanations and examples other than those provided here, see the ggplot2 reference manual <http://had.co.nz/ggplot2/>

Layering

Adding layers is not necessarily commutative!

```
g + geom_point() + geom_line()
```

```
g + geom_line() + geom_point()
```



Using ggplot2

The Diamonds Dataset

```
library(ggplot2)

data(diamonds)
head(diamonds)

##   carat      cut color clarity depth table price     x     y     z
## 1  0.23    Ideal   E    SI2   61.5     55   326 3.95 3.98 2.43
## 2  0.21  Premium   E    SI1   59.8     61   326 3.89 3.84 2.31
## 3  0.23      Good   E    VS1   56.9     65   327 4.05 4.07 2.31
## 4  0.29  Premium   I    VS2   62.4     58   334 4.20 4.23 2.63
## 5  0.31      Good   J    SI2   63.3     58   335 4.34 4.35 2.75
## 6  0.24  Very Good   J   VVS2   62.8     57   336 3.94 3.96 2.48
```

```

nrow(diamonds)

## [1] 53940

summary(diamonds)

##      carat          cut        color       clarity
##  Min.   :0.2000   Fair     : 1610   D: 6775   SI1    :13065
##  1st Qu.:0.4000  Good    : 4906   E: 9797   VS2    :12258
##  Median :0.7000  Very Good:12082  F: 9542   SI2    : 9194
##  Mean   :0.7979  Premium  :13791   G:11292   VS1    : 8171
##  3rd Qu.:1.0400  Ideal    :21551   H: 8304   VVS2   : 5066
##  Max.   :5.0100                    I: 5422   VVS1   : 3655
##                               J: 2808   (Other): 2531
##      depth          table       price        x
##  Min.   :43.00   Min.   :43.00   Min.   : 326   Min.   : 0.000
##  1st Qu.:61.00  1st Qu.:56.00  1st Qu.: 950   1st Qu.: 4.710
##  Median :61.80  Median :57.00  Median : 2401   Median : 5.700
##  Mean   :61.75  Mean   :57.46  Mean   : 3933   Mean   : 5.731
##  3rd Qu.:62.50  3rd Qu.:59.00  3rd Qu.: 5324   3rd Qu.: 6.540
##  Max.   :79.00  Max.   :95.00  Max.   :18823   Max.   :10.740
##
##      y              z
##  Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 4.720   1st Qu.: 2.910
##  Median : 5.710   Median : 3.530
##  Mean   : 5.735   Mean   : 3.539
##  3rd Qu.: 6.540   3rd Qu.: 4.040
##  Max.   :58.900   Max.   :31.800
##

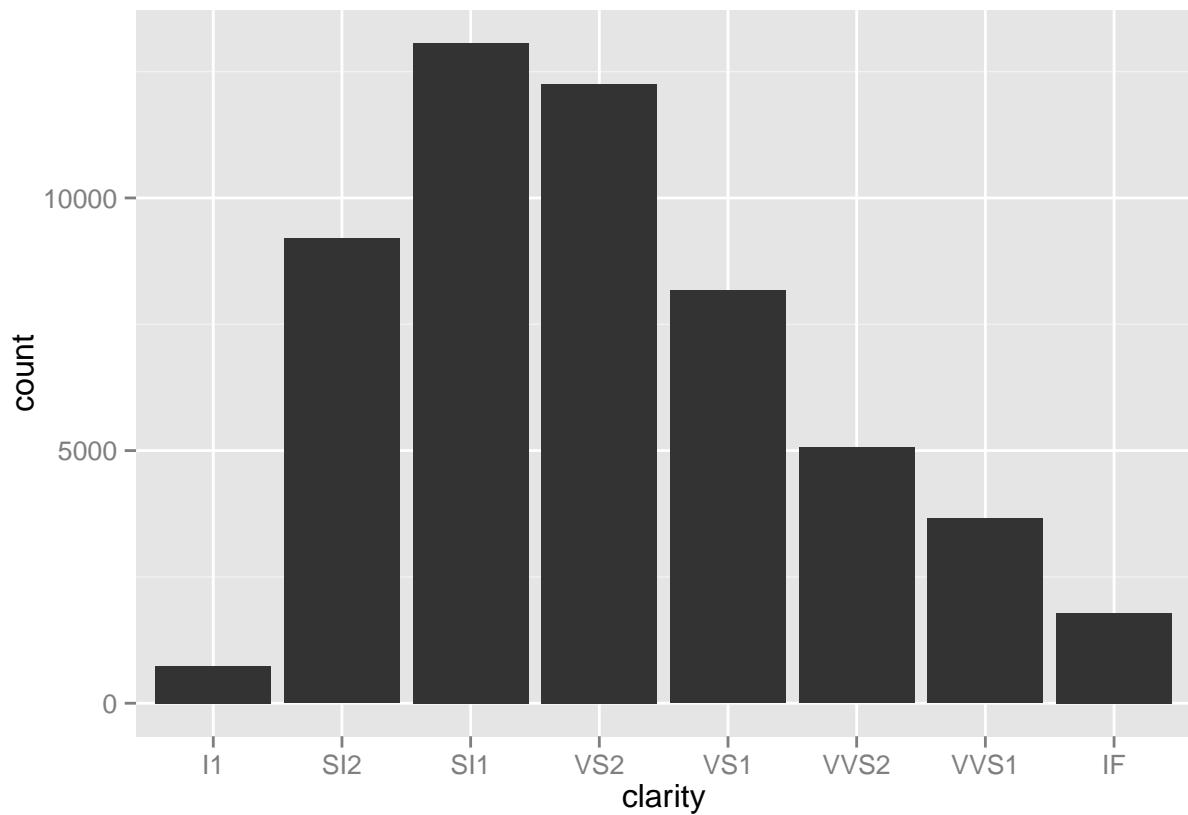
```

Barplot

```

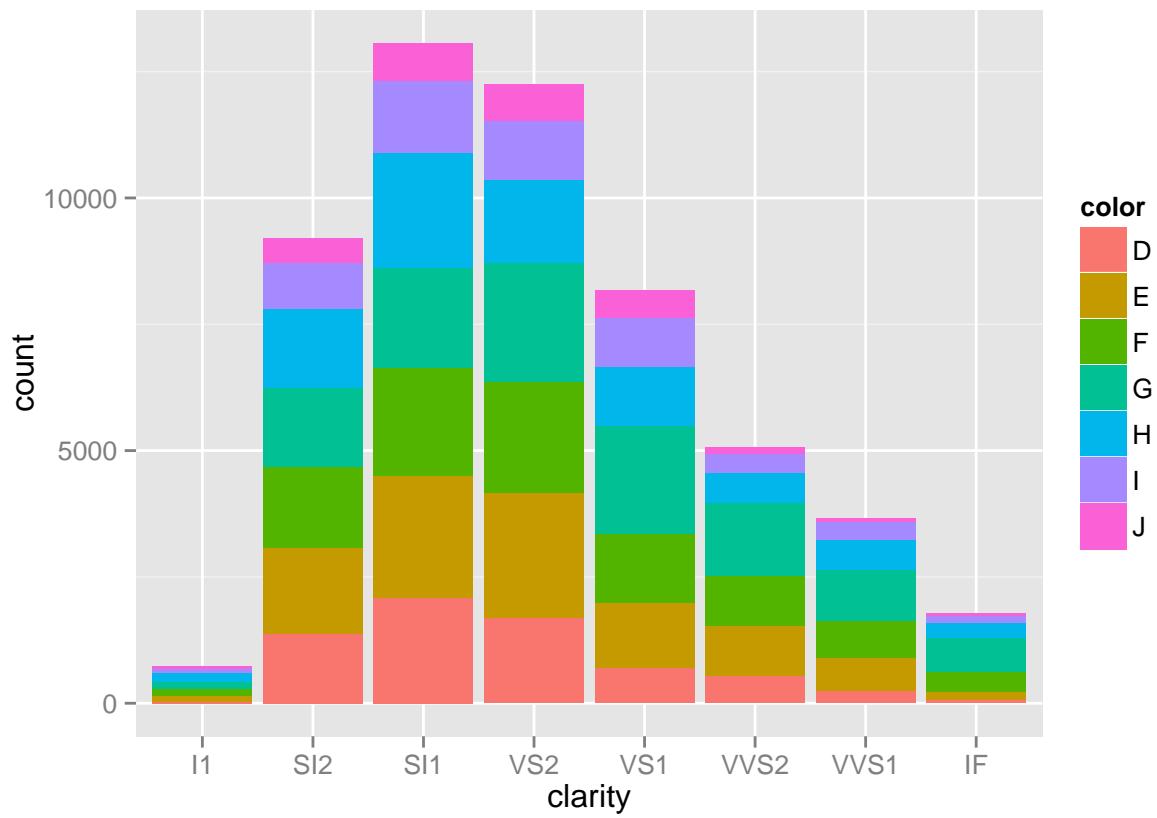
ggplot(data=diamonds, aes(x=clarity) ) + geom_bar()

```



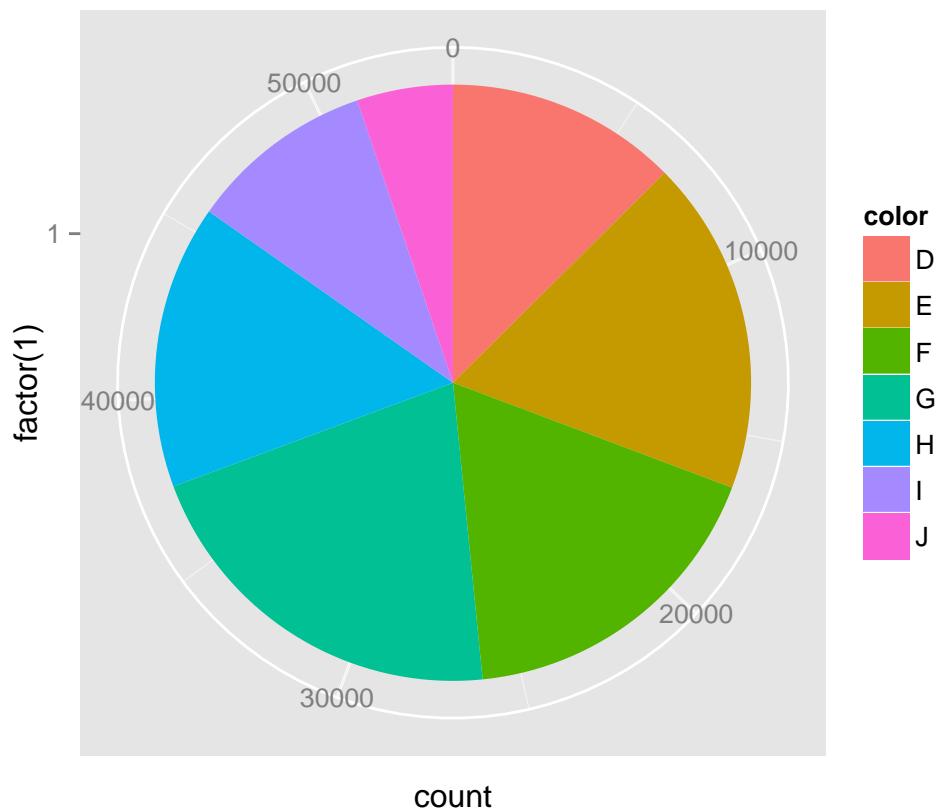
Barplot

```
ggplot(data=diamonds, aes(x=clarity, fill=color) ) + geom_bar()
```



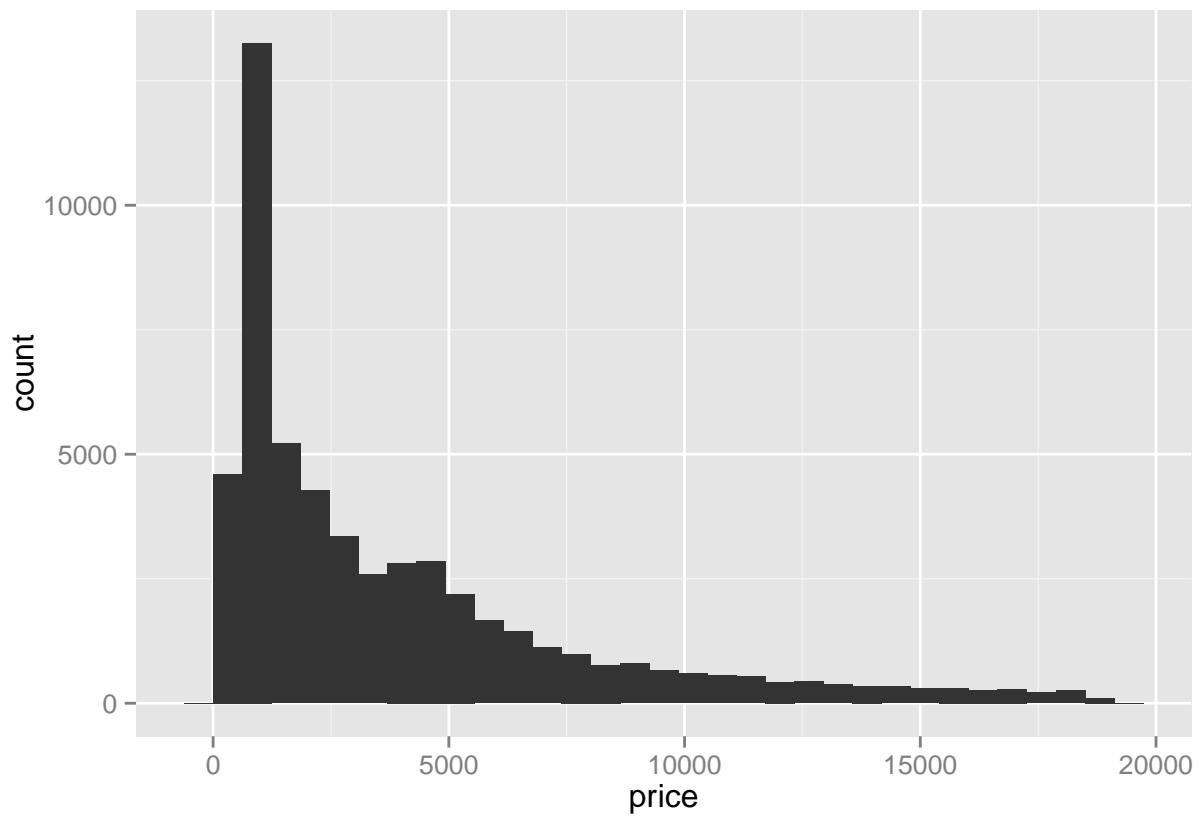
Pie Chart

```
ggplot(data=diamonds, aes(x=factor(1), fill=color) ) +  
  geom_bar(width=1) +  
  coord_polar(theta="y")
```



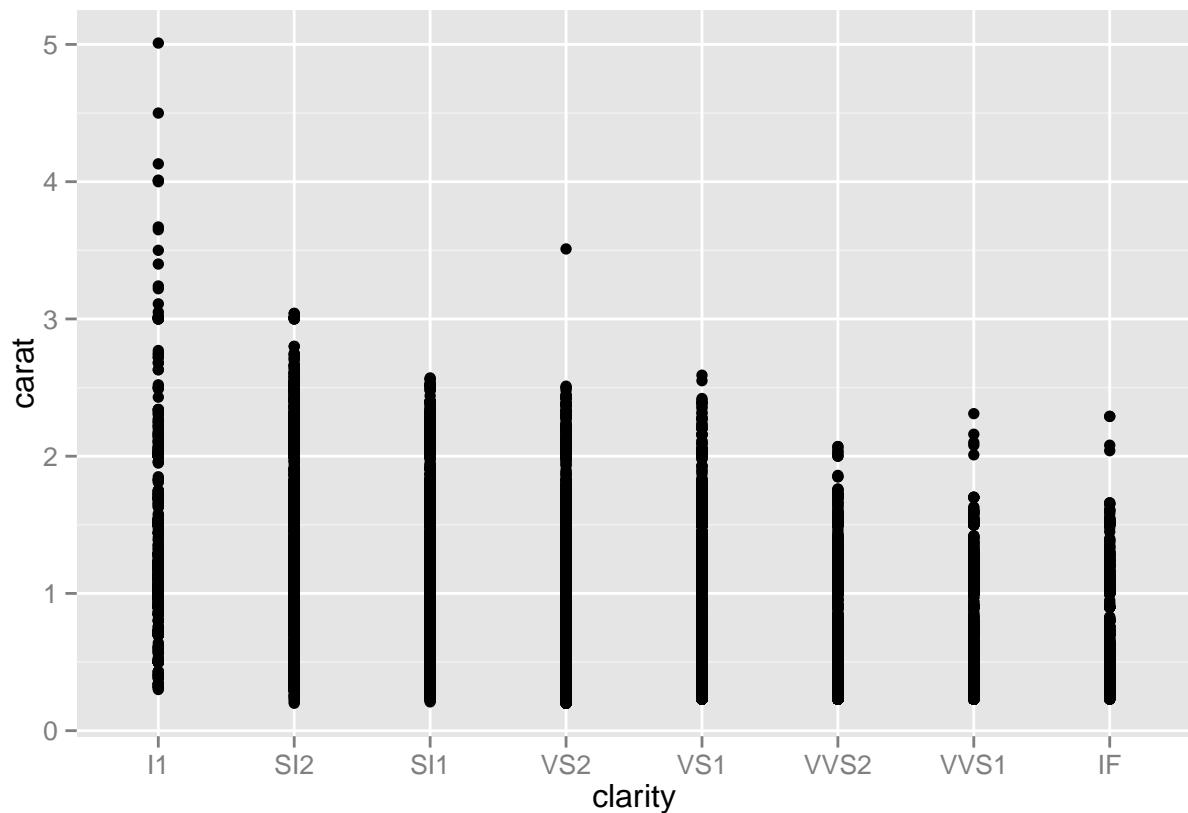
Histogram

```
ggplot(data=diamonds, aes(x=price) ) + geom_histogram()
```



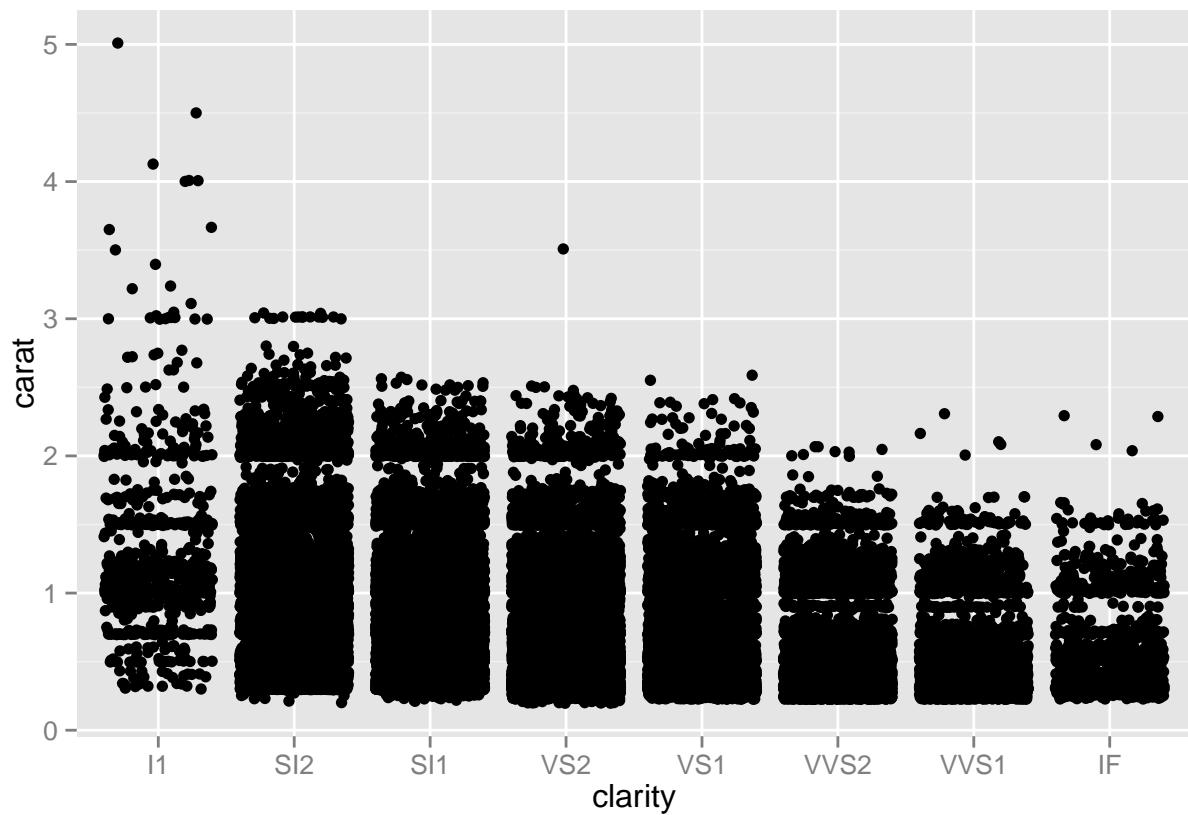
Scatterplots

```
g <- ggplot(data=diamonds, aes(x=clarity, y=carat) )  
g + geom_point()
```



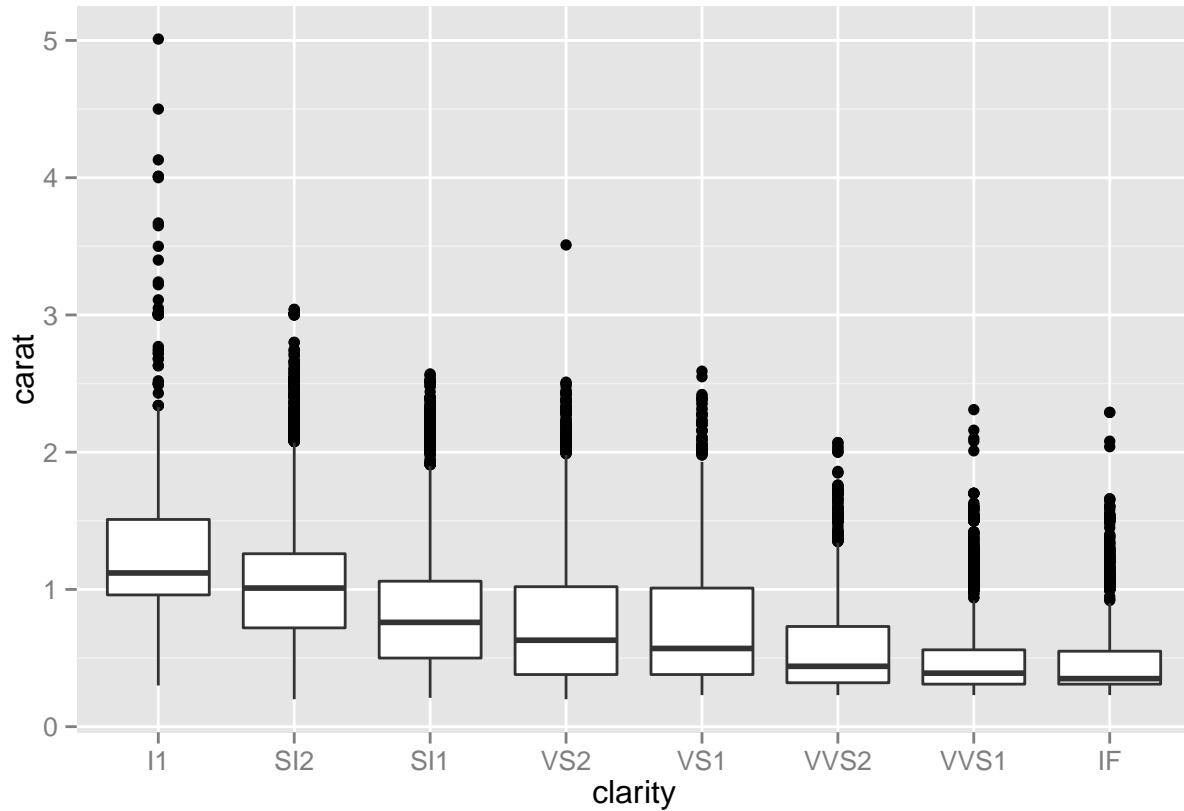
Scatterplot with jitter

```
g + geom_jitter()
```



Boxplot

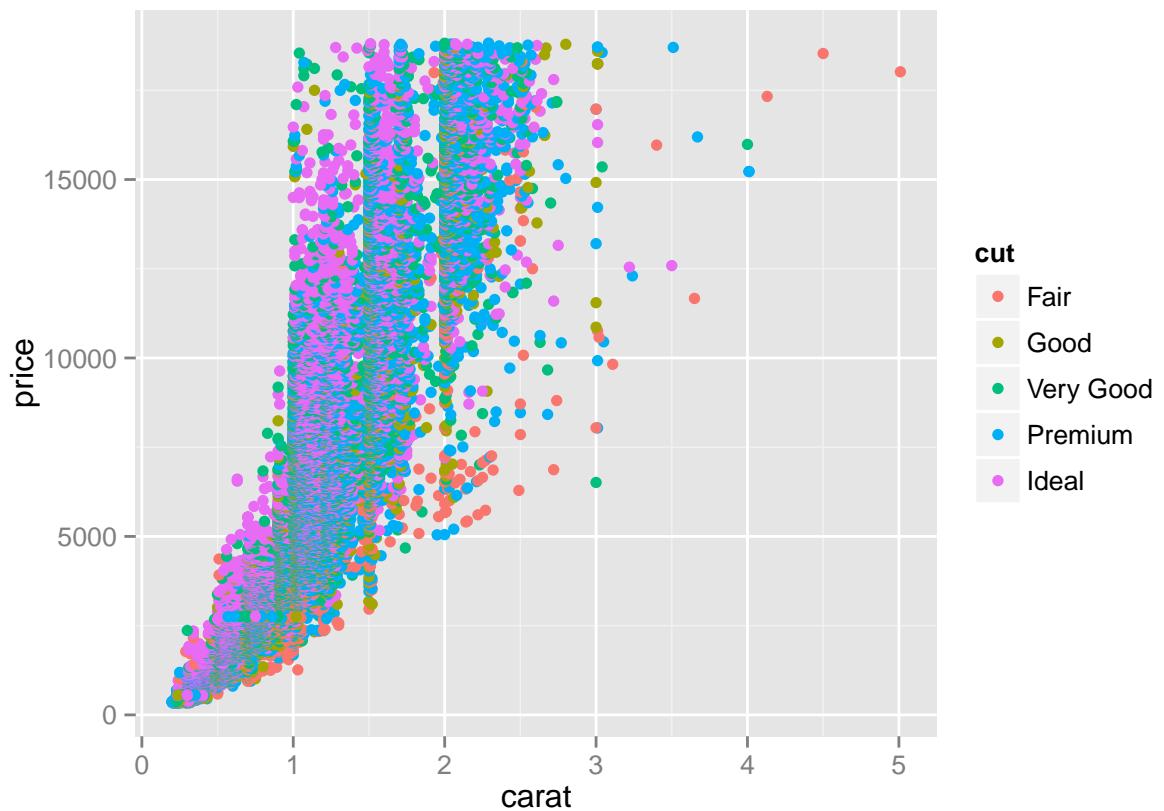
```
g + geom_boxplot()
```



Advanced Features and Tweaks

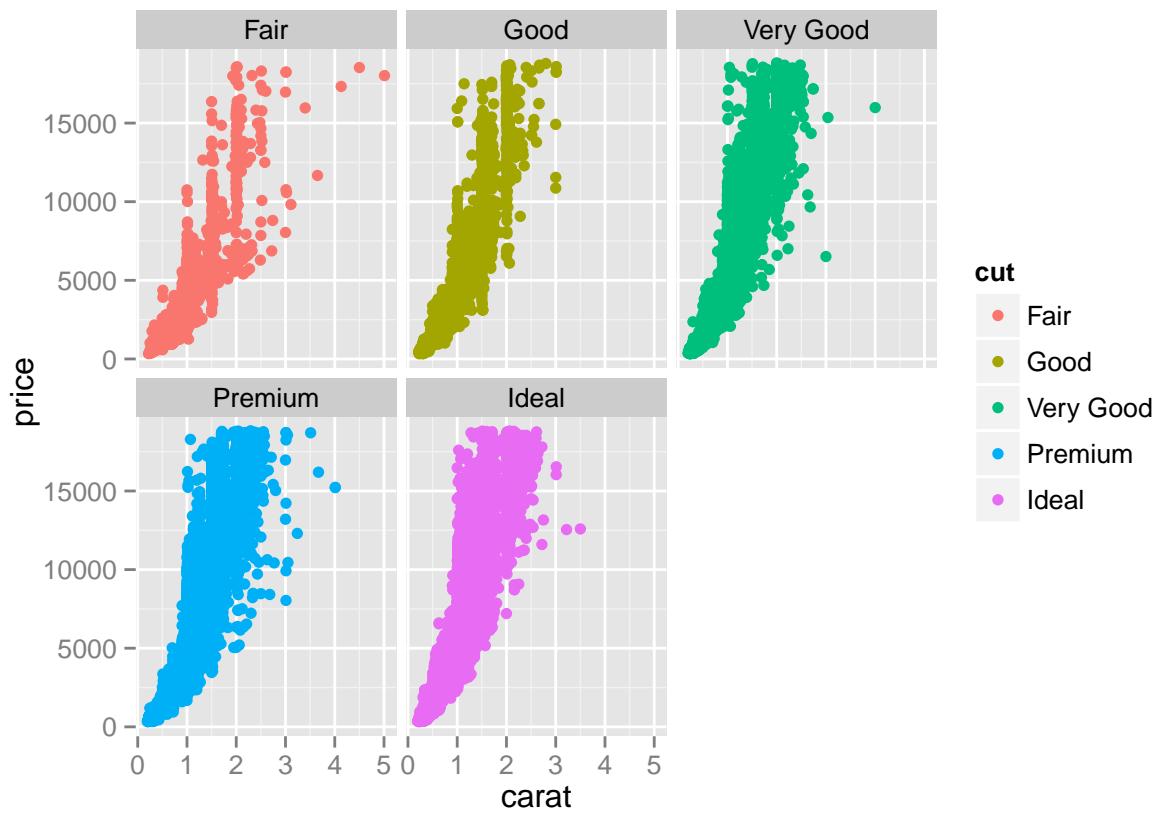
Faceting

```
g <- ggplot(data=diamonds, aes(x=carat, y=price) ) + geom_point(aes(color=cut))
g
```



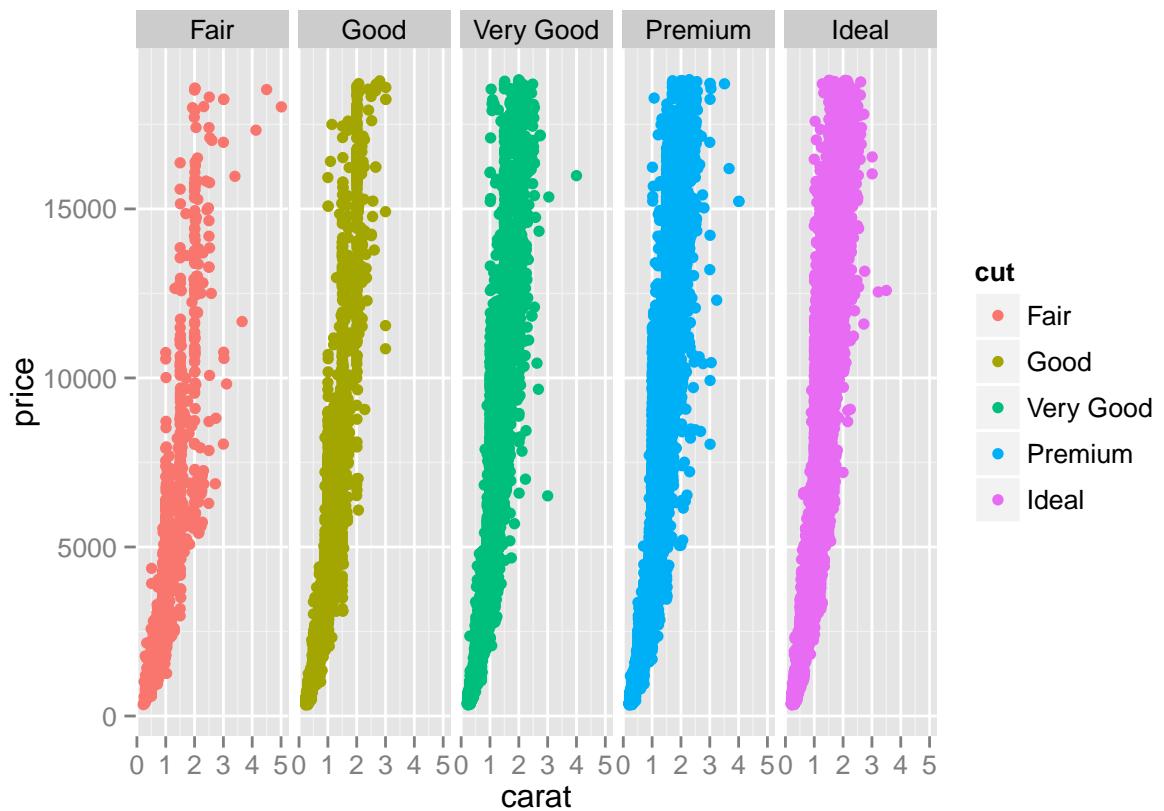
Faceting

```
g + facet_wrap(~ cut)
```



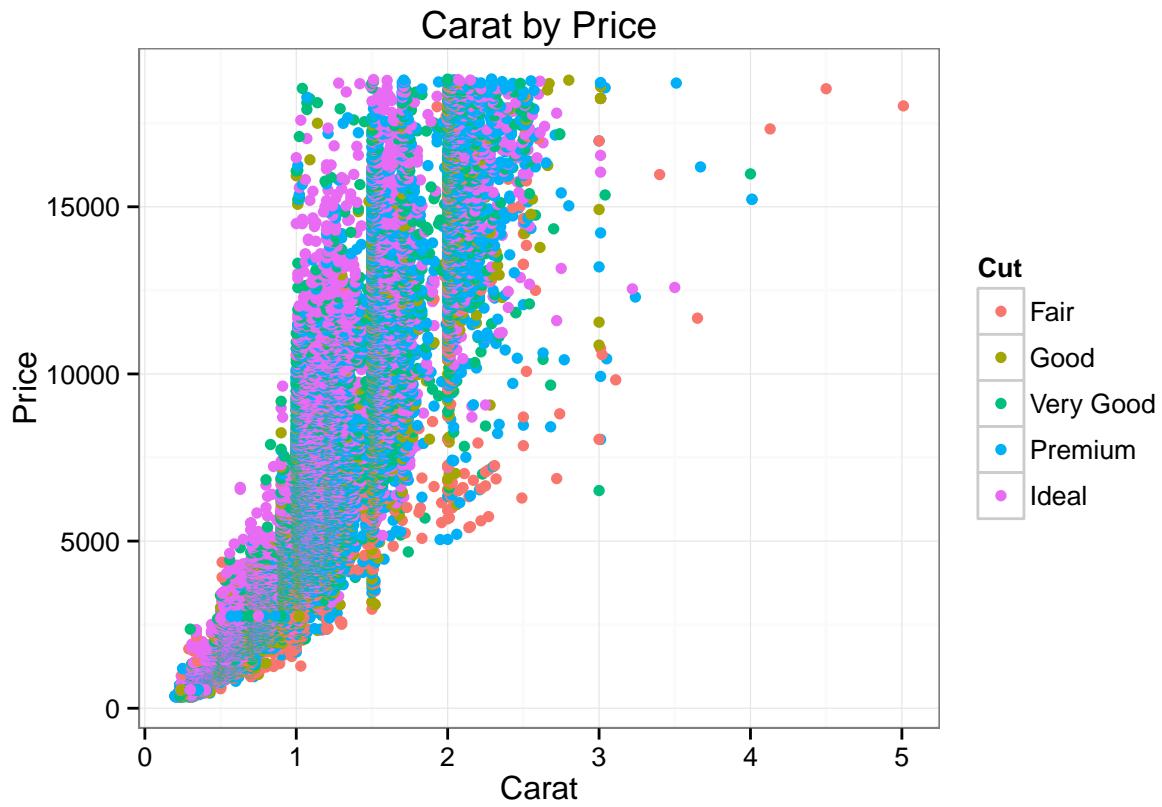
Faceting

```
g + facet_grid(~ cut)
```



Options

```
g + xlab("Carat") + ylab("Price") + ggtitle("Carat by Price") +  
  labs(color="Cut") + theme_bw()
```



Saving Plots

```
pdf("path/to/outfile.pdf")
myplot # or last_plot()
dev.off()

ggsave("path/to/outfile.pdf", myplot)
```

See also:

```
?pdf
?ggsave
```

Faceting

Wrapup

Where to Learn More

- Reference Manual <http://had.co.nz/ggplot2/>
- CRAN page <http://cran.r-project.org/web/packages/ggplot2/index.html>
- Google Group <https://groups.google.com/group/ggplot2>
- [ggplot2] tag on stackoverflow
- Official Book <http://ggplot2.org/book/> (available in print)

Questions?

This presentation is available at github.com/wrathematics/2015SFSURworkshop

Exercises are also available there.