

ggplot2: a philosophy and a package

Peter D Smits

Committee on Evolutionary Biology
University of Chicago

January 6, 2013

Introduction

there are three main ways of making graphics in R:
base graphics, lattice, grid, ggplot2.

ggplot2 is an R package written (primarily) by Hadley Wickham.
implementation of the Grammar of Graphics (by Leland Wilkinson).
extremely popular, huge community, extremely powerful.

London Cycle Hire Journeys

Thicker, yellower lines mean more journeys



Data: 3.2 Million Journeys (from TfL)
Routing: Ollie O'Brien (@oobr) + OpenStreetMap cc-by-sa
Buildings: OS OpenData Crown Copyright 2011
Map: James Cheshire (@spatialanalysis)

Hadley Wickham

- ▶ professor of statistics at Rice University
- ▶ from New Zealand (oddly common in statistics)
- ▶ author of many R packages (ggplot2, reshape2, plyr, devtools, and more)
- ▶ ggplot and reshape made up most of his PhD thesis

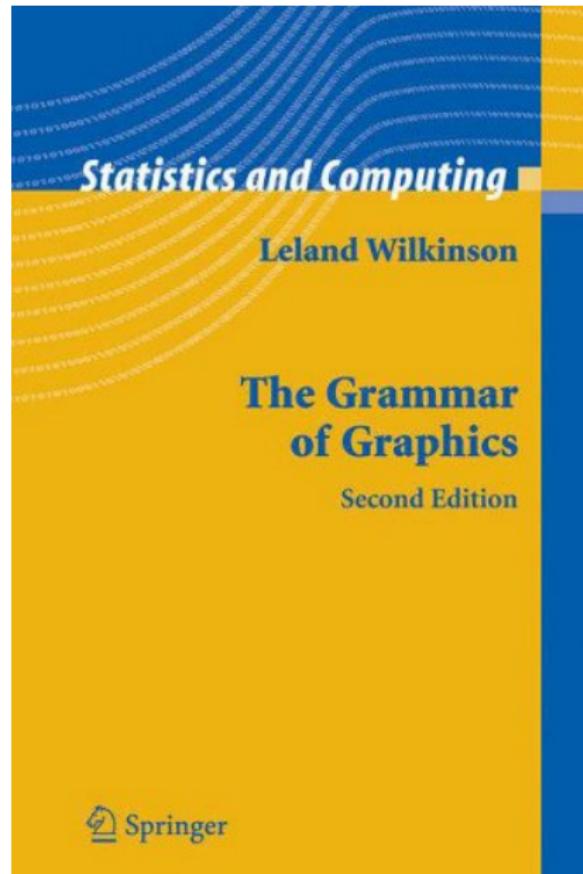


Grammar of Graphics

most texts on graphics were written a while ago (Tukey '77, Tufte '83, Chambers et al. '83, Cleveland '85). GoG newer and more “modern”.

few basic ideas

- ▶ variables
- ▶ aesthetics
- ▶ geometry
- ▶ statistics
- ▶ facets
- ▶ scales, etc.



Basic implementation

- ▶ object (ggplot)
- ▶ aesthetics (aes)
 - ▶ x, y position
 - ▶ size, shape
 - ▶ group, colour
- ▶ geometrics (geom) and statistics (stat)
 - ▶ points, lines, line segments
 - ▶ bars, histograms, boxes
 - ▶ maps (!!)
 - ▶ and more
- ▶ themes

Today's first data

The muscle car data set!

mtcars (part of base R install, along with a lot of other datasets)

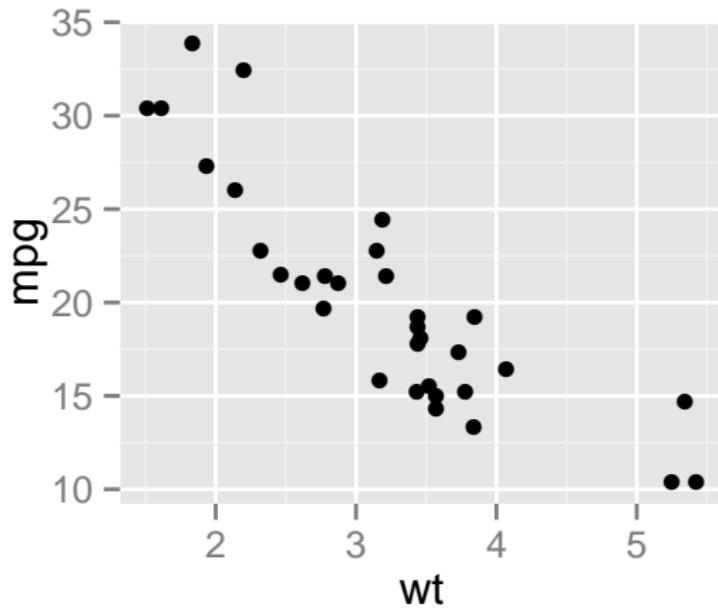
11 cars for 32 variables. Below is tiny subset.

```
## Loading required package: xtable
```

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.00	6.00	160.00	110.00	3.90	2.62
Mazda RX4 Wag	21.00	6.00	160.00	110.00	3.90	2.88
Datsun 710	22.80	4.00	108.00	93.00	3.85	2.32
Hornet 4 Drive	21.40	6.00	258.00	110.00	3.08	3.21
Hornet Sportabout	18.70	8.00	360.00	175.00	3.15	3.44
Valiant	18.10	6.00	225.00	105.00	2.76	3.46

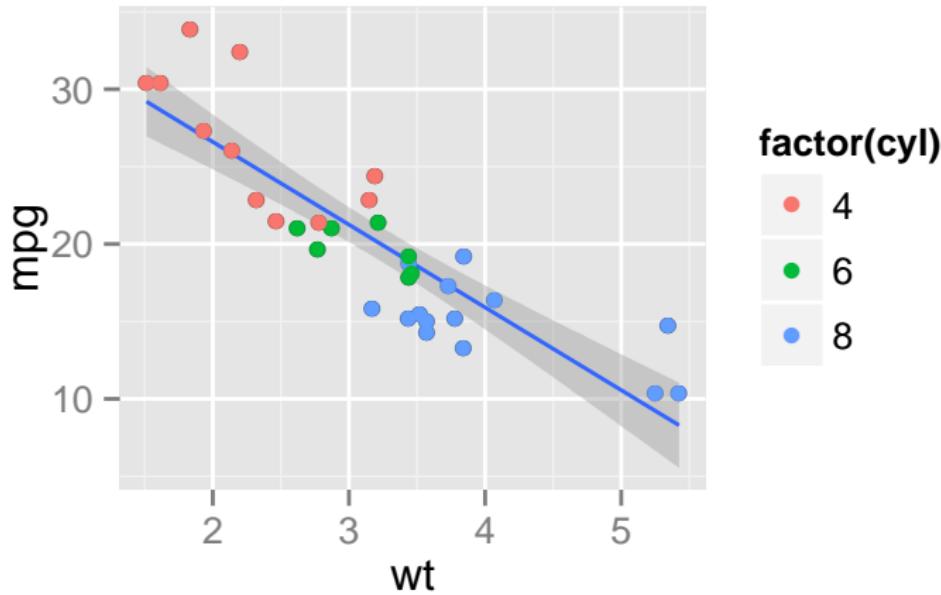
Our first graph

```
library(ggplot2)
g1 <- ggplot(mtcars, aes(wt, mpg)) + geom_point()
g1
```



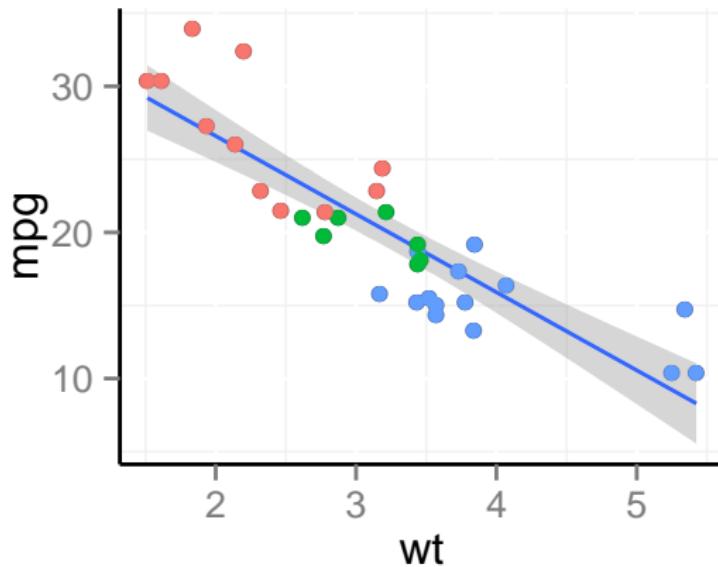
Adding a stat and modifying our first graph

```
g1 <- g1 + stat_smooth(method = "lm")
g1 <- g1 + geom_point(aes(colour = factor(cyl)))
g1
```



Try and make it look prettier

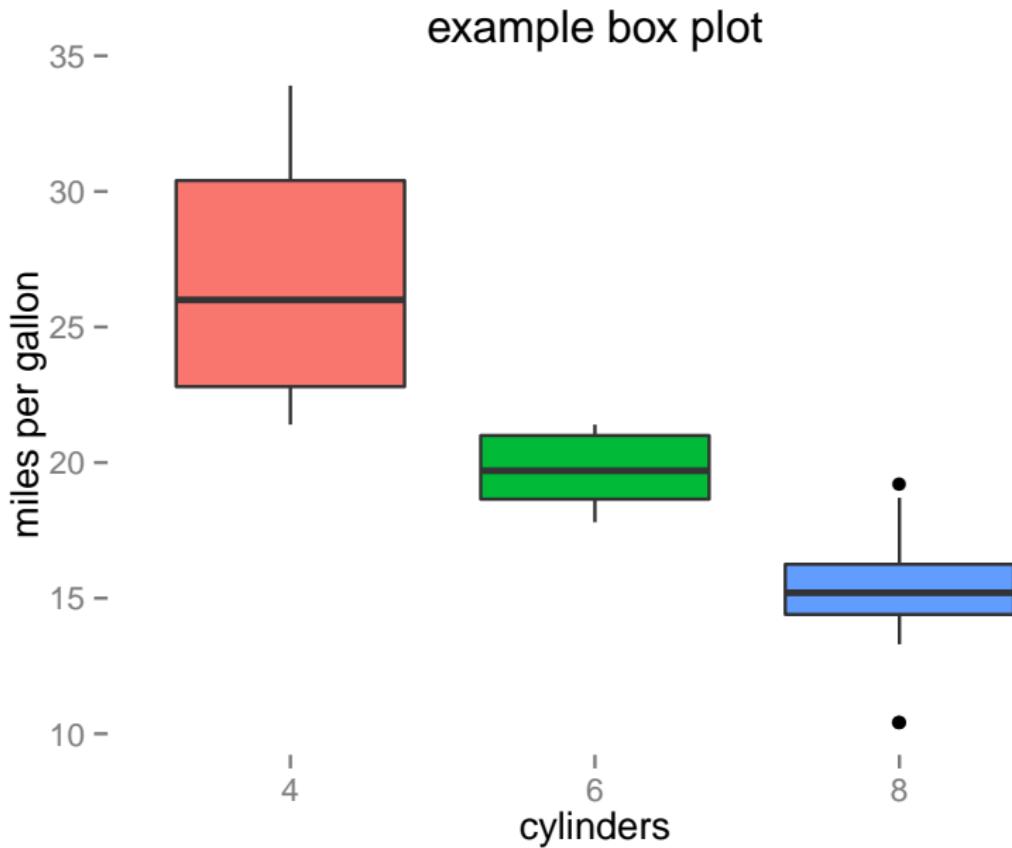
```
g1 <- g1 + theme(legend.position = 'none',
                  axis.line = element_line(colour = 'black'),
                  panel.background = element_blank())
g1
```



Our second graph

```
g2 <- ggplot(mtcars,
              aes(x = factor(cyl),
                  y = mpg,
                  fill = factor(cyl)))
g2 <- g2 + geom_boxplot()
g2 <- g2 + theme(legend.position = 'none',
                  panel.background = element_blank(),
                  panel.grid = element_blank())
g2 <- g2 + labs(x = 'cylinders',
                 y = 'miles per gallon',
                 title = 'example box plot')
```

Our second graph



Today's second dataset

iamonds!

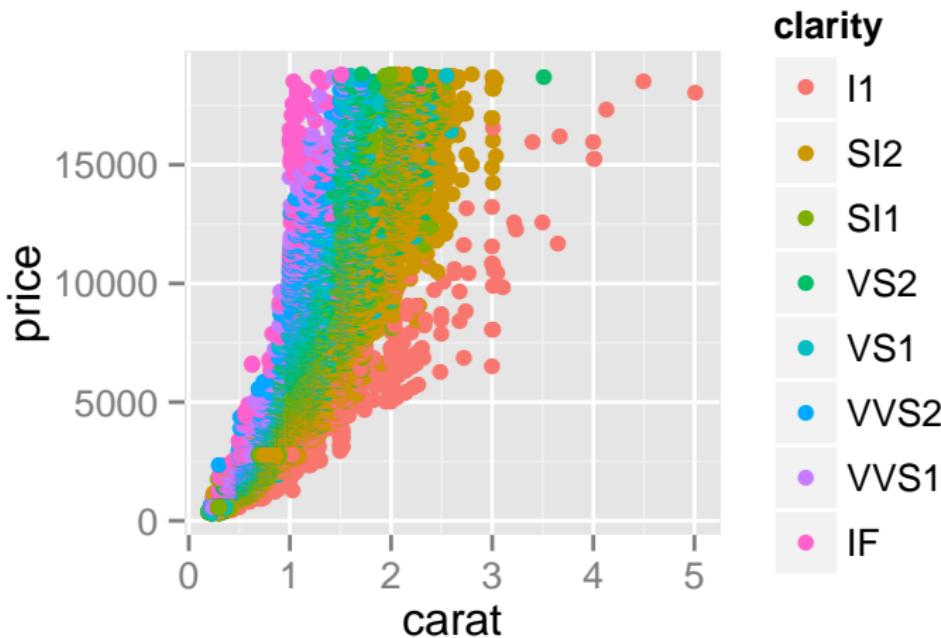
53940 samples for 10 variables.

Here is a tiny part of that dataset.

	carat	cut	color	clarity	depth	table	price
1	0.23	Ideal	E	SI2	61.50	55.00	326
2	0.21	Premium	E	SI1	59.80	61.00	326
3	0.23	Good	E	VS1	56.90	65.00	327
4	0.29	Premium	I	VS2	62.40	58.00	334
5	0.31	Good	J	SI2	63.30	58.00	335
6	0.24	Very Good	J	VVS2	62.80	57.00	336

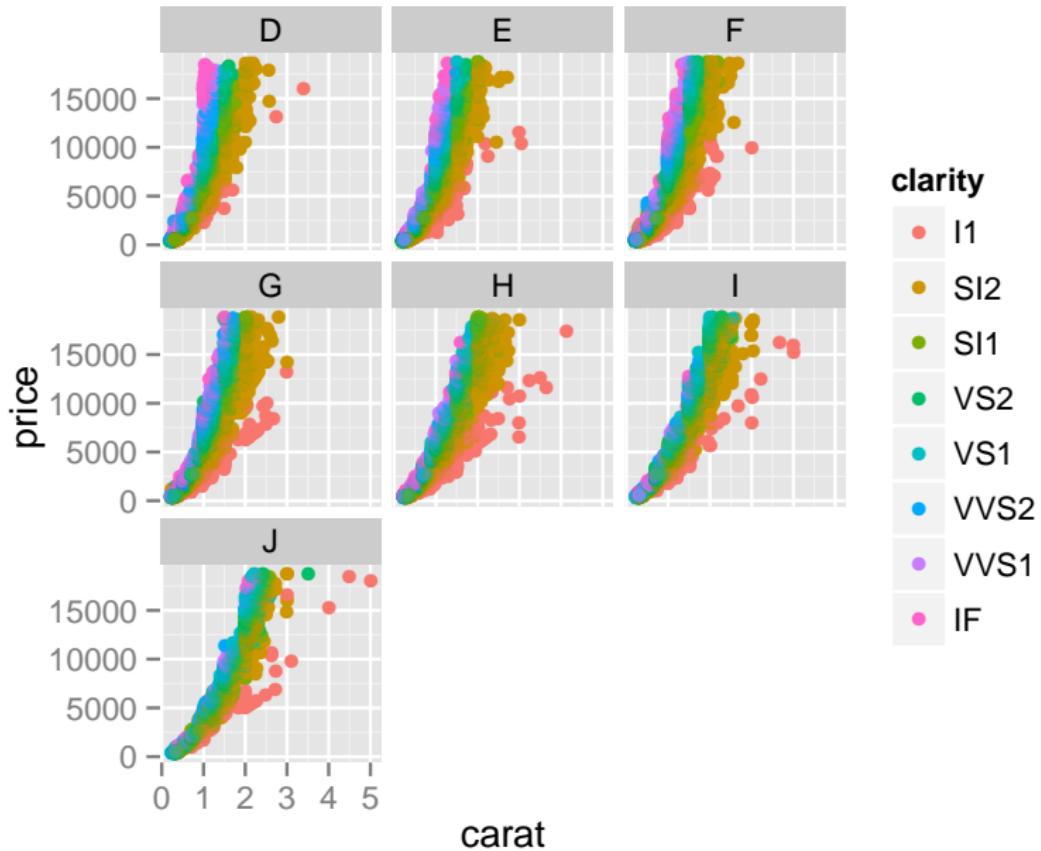
Looking at diamonds

```
d <- ggplot(diamonds,  
             aes(x = carat, y = price,  
                  colour = clarity))  
d <- d + geom_point()  
d
```



Make that look better

```
d <- d + geom_point(alpha = 0.1)
d <- d + facet_wrap(~ color)
```



i++i

Other useful packages

GGally: matrix plots

ggthemes: various canned themes to make your plots prettier (or hilariously ugly)

Useful websites

<http://docs.ggplot2.org/current/> : current ggplot documentation (very good)

<http://groups.google.com/group/ggplot2> : ggplot2 mailing list

<http://wiki.stdout.org/rcookbook/Graphs> : various tips and tricks to get over problems

<http://stackoverflow.com/> : coding question/answer site

<http://stats.stackexchange.com/> : statistics question/answer site

<http://www.r-bloggers.com/> : R blog aggregator