ggplot2: easy graphics with R

Ilmari Ahonen

Department of Mathematics and Statistics
University of Turku
Versatile and rich graphics are widely considered a major strength of the R language. However, achieving pleasing results with the base installation of R often still requires:

- a lot of time
- extensive knowledge of the various plotting functions
- hacking

ggplot2 package by prof. Hadley Wickham aims to simplify the process and bring everything under one common language
The philosophy of ggplot2, based on the Grammar of Graphics\textsuperscript{1} by Leland Wilkinson, separates plot “layers” into distinct categories:

- data and aesthetic mapping ($x$, $y$, color, size, ...)
- statistical transformation (log, mean, histogram, smoother, ...)
- a geometric object (dot, bar, line, box, ...)
- position adjustment (jitter, stack, ...)

ggplot2 installs and loads like any other R package:

```r
# to install: install.packages("ggplot2")
library(ggplot2)
```
We’ll use the `mtcars` data as an example:

<table>
<thead>
<tr>
<th></th>
<th>mpg</th>
<th>cyl</th>
<th>disp</th>
<th>hp</th>
<th>drat</th>
<th>wt</th>
<th>qsec</th>
<th>vs</th>
<th>am</th>
<th>gear</th>
<th>carb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazda RX4</td>
<td>21.0</td>
<td>6</td>
<td>160</td>
<td>110</td>
<td>3.90</td>
<td>2.62</td>
<td>16.46</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Mazda RX4 Wag</td>
<td>21.0</td>
<td>6</td>
<td>160</td>
<td>110</td>
<td>3.90</td>
<td>2.88</td>
<td>17.02</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Datsun 710</td>
<td>22.8</td>
<td>4</td>
<td>108</td>
<td>93</td>
<td>3.85</td>
<td>2.32</td>
<td>18.61</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td>21.4</td>
<td>6</td>
<td>258</td>
<td>110</td>
<td>3.08</td>
<td>3.21</td>
<td>19.44</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td>18.7</td>
<td>8</td>
<td>360</td>
<td>175</td>
<td>3.15</td>
<td>3.44</td>
<td>17.02</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Valiant</td>
<td>18.1</td>
<td>6</td>
<td>225</td>
<td>105</td>
<td>2.76</td>
<td>3.46</td>
<td>20.22</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Duster 360</td>
<td>14.3</td>
<td>8</td>
<td>360</td>
<td>245</td>
<td>3.21</td>
<td>3.57</td>
<td>15.84</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Merc 240D</td>
<td>24.4</td>
<td>4</td>
<td>147</td>
<td>62</td>
<td>3.69</td>
<td>3.19</td>
<td>20.00</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Merc 230</td>
<td>22.8</td>
<td>4</td>
<td>141</td>
<td>95</td>
<td>3.92</td>
<td>3.15</td>
<td>22.90</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Merc 280</td>
<td>19.2</td>
<td>6</td>
<td>168</td>
<td>123</td>
<td>3.92</td>
<td>3.44</td>
<td>18.30</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Merc 280C</td>
<td>17.8</td>
<td>6</td>
<td>168</td>
<td>123</td>
<td>3.92</td>
<td>3.44</td>
<td>18.90</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Merc 450SE</td>
<td>16.4</td>
<td>8</td>
<td>276</td>
<td>180</td>
<td>3.07</td>
<td>4.07</td>
<td>17.40</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Merc 450SL</td>
<td>17.3</td>
<td>8</td>
<td>276</td>
<td>180</td>
<td>3.07</td>
<td>3.73</td>
<td>17.60</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Merc 450SLC</td>
<td>15.2</td>
<td>8</td>
<td>276</td>
<td>180</td>
<td>3.07</td>
<td>3.78</td>
<td>18.00</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cadillac Fleetwood</td>
<td>10.4</td>
<td>8</td>
<td>472</td>
<td>205</td>
<td>2.93</td>
<td>5.25</td>
<td>17.98</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Lincoln Continental</td>
<td>10.4</td>
<td>8</td>
<td>460</td>
<td>215</td>
<td>3.00</td>
<td>5.42</td>
<td>17.82</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Chrysler Imperial</td>
<td>14.7</td>
<td>8</td>
<td>440</td>
<td>230</td>
<td>3.23</td>
<td>5.34</td>
<td>17.42</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Fiat 128</td>
<td>32.4</td>
<td>4</td>
<td>79</td>
<td>66</td>
<td>4.08</td>
<td>2.20</td>
<td>19.47</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Honda Civic</td>
<td>30.4</td>
<td>4</td>
<td>76</td>
<td>52</td>
<td>4.93</td>
<td>1.61</td>
<td>18.52</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Toyota Corolla</td>
<td>33.9</td>
<td>4</td>
<td>71</td>
<td>65</td>
<td>4.22</td>
<td>1.83</td>
<td>19.90</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Let’s investigate the relationship between the number horse powers and fuel consumption. We’ll start by setting these variables as aesthetics:

```r
my.plot <- ggplot(data=mtcars, aes(x=mpg, y=hp))
```

Now we have just an empty plot frame where we still need to add the geometric attributes.
my.plot + geom_point()
Many other geometrics are also grammatically correct

```r
my.plot + geom_point() + geom_line()
```
my.plot + geom_bar(stat="identity")
my.plot + geom_text(label=rownames(mtcars), size=4)
Let’s go with the dot plot for the next examples

\[
p \leftarrow \text{my.plot} + \text{geom_point}()\]
It is easy to add additional aesthetics to existing plots

```r
p + aes(size=cyl)
```
p + aes(size=cyl, col=factor(gear))
Every type of plot can be faceted

\[ p + \text{aes(size=cyl, col=\text{factor(gear)})} + \text{facet_grid(. \sim am)} \]
Adding some typical analysis results, such as errorbars or smoothers, is straightforward

```r
p + geom_smooth(method=lm)
```
p + geom_smooth(method=loess)
Again, additional aesthetics can be added

```r
p + geom_smooth(method=lm) + aes(col=factor(gear))
```
The appearance of the plots can be modified using themes. Let’s continue with the previous plot

```
p2 <- p + geom_smooth(method=lm) + aes(col=factor(gear))
p2
```
p2 + theme_classic()
p2 + theme_minimal()
Now let’s create our own theme

```r
our.awesome.theme <- theme(
  axis.text = element_text(size = 14),
  legend.key = element_rect(fill = "navy"),
  legend.background = element_rect(fill = "lightgreen"),
  legend.position = c(0.1, 0.9),
  panel.grid.major = element_line(colour = "lightgray"),
  panel.background = element_rect(fill = "beige"
)
```
\begin{verbatim}
p2 + theme_minimal() + scale_colour_grey() + 
aes(linetype=factor(gear), shape=factor(gear))
\end{verbatim}
The package documentation and additional resources can be found on the project web page: http://ggplot2.org/
Some additional points:

- `qplot()` behaves similarly to the base R’s `plot()` and can be used for creating simple graphics quickly without using the full potential of the grammar.

- `ggplot2` is not suited for interactive graphics and is often too slow for smooth animations.

- Some commonly used plots, such as scatter plot matrices are difficult to express using the grammar and are thus not straightforward to produce with `ggplot2`. See `GGally` package for help.
Some additional resources that were used for this presentation:

- Rstudio: IDE for R and more
- knitr: embed R analyses and results in tex documents