Introduction to ggplot2

Andy Grogan-Kaylor

March 20, 2020

Table of Contents
Background ........................................................................................................................................... 2
ggplot in 3 easy steps (maybe 2 easy steps) .................................................................................... 2
  aesthetic: what you want to graph (e.g. x, y, z). ........................................................................... 2
  geom: how you want to graph it. ........................................................................................................ 2
  options: optional titles, themes, etc............................................................................................... 2
A Simple Quick Example ..................................................................................................................... 2
Call The Relevant Libraries .............................................................................................................. 4
Simulated Data .................................................................................................................................. 4
The Essential Idea Of ggplot2 Is Simple ............................................................................................ 5
  ggplot2 Starts By Calling The aesthetic ....................................................................................... 5
  We Then Call The geometry ............................................................................................................ 6
Examples ............................................................................................................................................. 6
  One Continuous Variable At A Time ............................................................................................... 6
    Dotplot .......................................................................................................................................... 6
    Add Some Options ......................................................................................................................... 7
  Different Geoms ............................................................................................................................... 8
  One Categorical Variable at a Time ................................................................................................. 10
  A Categorical Variable and A Continuous Variable .................................................................... 12
    Barchart ....................................................................................................................................... 12
Two Continuous Variables At A Time ............................................................................................... 12
  Basic Scatterplot ............................................................................................................................. 12
  Add Some Options .......................................................................................................................... 13
  Try A Smoother .............................................................................................................................. 13
  Try A Density Plot .......................................................................................................................... 14
  Try a Hexagon Geom .................................................................................................................... 17
  Combine Points and Smoother And Add Some Themes .............................................................. 17
Two Continuous Variables And A Third Categorical Variable ...................................................... 25
Background

R has a number of graphing libraries, including base graphics that are installed whenever you install R.

ggplot2, is a graphing library in R that makes beautiful graphs. ggplot2 graph syntax can be formidably complex, with a somewhat steep learning curve.

That being said, learning ggplot2 is worth the effort for a couple of reasons. First, the graphs are beautiful. Second, ggplot2’s syntax, though seemingly arcane at times, forces you to think about the nature of your data, and the ideas that you are graphing. Lastly, a little bit of knowledge about ggplot2 can go a long way, and can build a powerful foundation for future learning.

ggplot in 3 easy steps (maybe 2 easy steps)

aesthetic: what you want to graph (e.g. x, y, z).

gem: how you want to graph it.

options: optional titles, themes, etc.

A Simple Quick Example

The intent of this tutorial is to build the foundation of this idea that:
A little bit of ggplot can go a long way

and to give you a simple introduction to the idea that any ggplot graph is composed of:

an aesthetic + a geom or two + other optional elements like titles and themes.

So, as a quick and simple example...

```r
library(ggplot2)

ggplot(my_demo_data, # the data that I am using
       aes(x = my_outcome)) + # aesthetic: what I am graphing
       geom_histogram(fill = "red", # geom: how I am graphing it
color = "black") # options: fill = "red"; color = "black"
```

And now, with labels...

```r
ggplot(my_demo_data, # the data that I am using
       aes(x = my_outcome)) + # aesthetic: what I am graphing
       geom_histogram(fill = "red", # geom: how I am graphing it
color = "black") +
       labs(title = "Your Title Here",
            subtitle = "Your Subtitle Here",
            caption = "A Caption, If You Want One",
            x = "my outcome",
            y = "count")
```
This document is a very brief introduction to the basic ideas of ggplot2. More information about ggplot can be found here. More ggplot2 examples can be found here.

**Call The Relevant Libraries**

You will need a few R libraries to work in ggplot. You may only need `library(ggplot2)`, but some of these other libraries may also be helpful.

```r
library(ggplot2) # beautiful graphs
library(ggthemes) # nice themes for ggplot2
library(ggbeeswarm) # "beeswarm" plots
library(cowplot) # arrange graphs
library(pander) # nice tables
library(psych) # nice table of descriptive statistics
```

**Simulated Data**

In this example, we simulate some data. But your own learning of ggplot will progress more quickly if you use data that you have access to, on an issue that you care about.

Here are the first few rows of simulated data:
The Essential Idea Of `ggplot2` Is Simple.

There are 3 essential elements to any ggplot call:

1. **An aesthetic** that tells `ggplot` which variables are being mapped to the `x axis`, `y axis`, (and often other attributes of the graph, such as the `color fill`). Intuitively, the aesthetic can be thought of as **what you are graphing**.
2. A **geom or geometry** that tells `ggplot` about the basic structure of the graph. Intuitively, the geom can be thought of as **how you are graphing it**.
3. Other options, such as a **graph title**, **axis labels** and **overall theme** for the graph.

**ggplot2 Starts By Calling The aesthetic**

For one variable:

```
p <- ggplot(mydata, aes(x = ...))
```

This says there is only one variable running along the horizontal `x` axis in the aesthetic.

The `p <- ...` means that we are **assigning** this graph aesthetic to plot `p`. We can then add other features to plot `p` as we continue our work. This **iterative** nature of `ggplot2` is one of the things that makes it so powerful. As your workflow and your documents become more complex, you can build a simple consistent foundation\(^1\) for your graphs, then add something simple to make a first graph, and a different something simple to make a second graph.

For two variables:

---

\(^1\) By way of illustration, this foundation could be just an aesthetic (e.g. `aes(...)`) alone, or possibly an aesthetic plus a theme (e.g. `theme_tufte()`), plus axis labels to create a consistent look and feel for your graphs across a report.
p <- ggplot(mydata, aes(x = ..., y = ...)) This says there are two variables: one for the horizontal x axis; and another for the vertical y axis, in the aesthetic.

**We Then Call The geometry**

We can then add different geometries to our plot:

For one variable:

+ geom_density() This says add a density geometry to the graph.
+ geom_histogram() This says add a histogram geometry to the graph.
+ geom_violin() This says add a **violin plot** geometry to the graph.
+ geom_beeswarm() This says add a **beeswarm** geometry to the graph.

A **beeswarm** is a creative layout of points that intuitively lets you understand the distribution of a quantity. The **beeswarm** geometry requires separate installation of the **ggbeeswarm** package. You also need to call `library(ggbeeswarm)` to use this geometry.

For two variables:

+ geom_point() This says add a point (scatterplot) geometry to the graph.
+ geom_smooth() This says add a smoother to the graph.

**Examples**

**One Continuous Variable At A Time**

**Dotplot**

```
# call ggplot2 where aesthetic is: x uses our predictor variable

p1 <- ggplot(mydata,  
               aes(x = predictor))

p1 +  
      geom_dotplot(dotsize = .15) # add dotplot geom
```
Add Some Options

```r
p1 +
  geom_dotplot(dotsize = .15,
               fill = "red")  # add dotplot geom in red
  labs(title = "Dotplot of predictor")  # Add title
```
Different Geoms

Histogram

```r
p1 + geom_histogram(fill = "blue", 
                     color="black") + # add histogram geom in blue 
labs(title ="Histogram of predictor") # Add title
```

Density

```r
p1 + geom_density(fill = "gold") + # add density geom in gold 
labs(title ="Density of predictor") # Add title
```
Violin Plot

```r
p2 <- ggplot(mydata,
             aes(x = 1,  # we need an aesthetic with _x_
                 y = predictor))  # & _y_

p2 + geom_violin(fill = "purple") +
     labs(title = "Violin Plot of predictor") # Add title
```
Beeswarm

```r
p3 <- ggplot(mydata,
             aes(x = predictor, # we need an aesthetic with _x_
                 y = 1)) # & _y_

p3 + geom_beeswarm(color = "red",
                      groupOnX = FALSE) +
    labs(title = "Beeswarm Plot of predictor") + # Add title
    theme(axis.title.y = element_blank(),
          axis.text.y = element_blank()) # tweak y axis
```

Beeswarm Plot of predictor

One Categorical Variable at a Time

The easiest way to represent a single categorical variable is likely a bar graph.

Here bars represent the count of observations in each group.

```r
p_barchart <- ggplot(mydata,
                      aes(x = group)) +
    geom_bar(fill = "red")

p_barchart
```
Changing the *aesthetic* slightly results in a *stacked* bar chart. Since all groups are stacked in 1 bar, we have to add information about the colors that we want to use to distinguish the groups.

```r
p_stacked_barchart <- ggplot(mydata,
  aes(x = 1,
       fill = group)) +
  geom_bar() +
  scale_fill_manual(values = c("red", "blue"))
```

`p_stacked_barchart`
A Categorical Variable and A Continuous Variable

Barchart
Here bars represent the average value of our outcome variable for members of each group.

\[
p_{\text{barchart of mean}} \leftarrow \text{ggplot}(\text{mydata},
\text{aes}(x = \text{group}, \# \text{slightly different aesthetic}
\text{y = outcome})) +
\text{stat_summary}(\text{fun.y = mean, \# take the mean of the data}
\text{fill = "blue", \# fill color}
\text{geom = "bar"}) \# \text{we want to summarize data with bars}
\]

\[p_{\text{barchart of mean}}\]

Two Continuous Variables At A Time

Basic Scatterplot
\# call ggplot2 where aesthetic uses both predictor and outcome

\[
p_{4} \leftarrow \text{ggplot}(\text{mydata},
\text{aes}(x = \text{predictor},
\text{y = outcome})) \# \text{set up aesthetic}
\]

\[p_{4} + \text{geom_point()} \# \text{add point geom (scatterplot)}\]
Add Some Options

```r
p4 + # start with basic plot that has only an aesthetic
group_point(color = "blue") + # add point geom in blue
labs(title = "Scatterplot of Outcome by Predictor") # add title
```

Try A Smoother

```r
p4 +
group_smooth() + # add smooth geom
labs(title = "Smother of Outcome by Predictor") # add title
```
Try A Density Plot

Simple Density

```r
p4 +
  geom_density2d(color = "blue") + # add density geom
  labs(title = "Density Plot of Outcome by Predictor") # add title
```
Filled Density

While not strictly necessary, the use of `scale_fill_gradient` seems to improve the presentation. You can choose your own colors.

```r
p4 +
  stat_density_2d(aes(fill = ..level..),
                  geom = "polygon") + # add filled density geom
  scale_fill_gradient(low = "blue",
                      high = "red") +
  labs(title = "Density Plot of Outcome by Predictor") # add title
```

Add Points

```r
p4 +
  stat_density_2d(aes(fill = ..level..),
                  geom = "polygon") + # add filled density geom
  geom_point(color = "orange") +
  scale_fill_gradient(low = "blue",
                      high = "red") +
  labs(title = "Density Plot of Outcome by Predictor") # add title
```
Use a Raster Geom Instead

```r
p4 +
  stat_density_2d(geom = "raster",
                  aes(fill = ..density..),
                  contour = FALSE) +
  scale_fill_gradient(low = "blue",
                     high = "red") +
  labs(title = "Density Plot (Raster) of Outcome by Predictor")  # add title
```
Try a Hexagon Geom

`geom_hex` may be a useful visualization, especially when there is the possibility of overplotting due to many many points.

```r
p4 + geom_hex() + scale_fill_gradient(low = "blue", high = "red") + labs(title = "Hexagon Plot of Outcome by Predictor") # add title
```

Combine Points and Smoother And Add Some Themes

Themes Included With `ggplot2`

*Default ggplot2 Theme*

```r
p4 + geom_point() + # point geom
gem_smooth() + # add smooth geom
labs(title = "Scatterplot And Smoother of Outcome", subtitle = "nby Predictor") + # add title
theme_grey() # default theme
```
The “minimal” theme

```r
p4 +
  geom_point() + # point geom
  geom_smooth() + # add smooth geom
  labs(title = "Scatterplot And Smoother of Outcome \nby Predictor") + # add title
  theme_minimal() # default theme
```
Themes requiring `ggthemes()`

The themes below make use of `library(ggthemes)` which you will need to install.

"538" Theme

```r
p4 +
  geom_point() + # point geom
  geom_smooth() + # add smooth geom
  labs(title = "Scatterplot And Smoother of Outcome \nby Predictor") + # add title
  theme_fivethirtyeight() + # "538"-like theme
  scale_color_fivethirtyeight() # "538"-like colors
```
Scatterplot And Smoother of Outcome by Predictor

```
$\text{p4 \ +}$
\begin{itemize}
  \item `geom_point()` \# point geom
  \item `geom_smooth()` \# add smooth geom
  \item `labs(title = "Scatterplot And Smoother of Outcome by Predictor")` \# add title
  \item `theme_solarized()` \# Google Docs theme
  \item `scale_colour_solarized()` \# Google Docs colors
\end{itemize}
```

"Solarized Theme"
Scatterplot And Smoother of Outcome by Predictor

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```

```
```
```
Scatterplot And Smoother of Outcome by Predictor

outcome

predictor

"Economist" Theme

p4 +

  geom_point() + \# point geom
  geom_smooth() + \# add smooth geom
  labs(title = "Scatterplot And Smoother of Outcome by Predictor") + \# add title
  theme_economist() + \# Economist magazine theme
  scale_colour_economist() \# Economist magazine colors
```
"Tufte" Theme

# same plot with theme and geom based on the work of Edward Tufte

```r
p4 +
  geom_point() +
  geom_smooth(color = "red") +
  theme_tufte() +
  labs(title = "Scatterplot And Smoother of Outcome by Predictor")
```
```r
p4 +
  geom_point() + # point geom
  geom_smooth() + # add smooth geom
  labs(title = "Scatterplot And Smoother of Outcome by Predictor") + # add title
  theme_gdocs() + # Google Docs theme
  scale_colour_gdocs() # Google Docs colors
```

"Google Docs Theme"
Two Continuous Variables And A Third Categorical Variable

Modify the aesthetic to include `group`.

```r
p5 <- ggplot(mydata,
              aes(x = predictor,
                  y = outcome,
                  color = group))  # aesthetic includes color by group
            
p5 + geom_point() +
        geom_smooth() +
        theme_economist() +
        scale_color_economist() +
        labs(title = "Scatterplot And Smoother of Outcome by Predictor")
```
Add facets or “small multiples” by group

```r
p5 +
  geom_point() +
  geom_smooth() +
  facet_wrap(~group) + # facets or "small multiples" by group
  theme_economist() +
  scale_color_economist() +
  labs(title = "Scatterplot And Smoother of Outcome \nby Predictor")
```
There Is A Lot More That Can Be Done With ggplot2

More information can be found at ggplot2.

More ggplot2 examples can be found here.

Graphics made with the ggplot2 graphing library created by Hadley Wickham.

Available online at https://www.umich.edu/~agrogan

Quick Introduction to ggplot2 by Andrew Grogan-Kaylor is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Last updated: March 20 2020 at 19:38