The difference between ggplot2 and R base graphics

Arleen Wang

10/29/2017

Introduction

Until today, we have learnt several different ways to draw a graph.
At the beginning of the class, we have learnt how to use the `plot` to create a graph. As we known, such as `plot`, `boxplot`, `barplot`, `hist` are belongs to the R garphics package.
And later on, we have learnt to use the `ggplot2` package to make a amazing beautiful graph.
When I using code to do the graph, most of the times, I'm not struggle how to draw a graph, I'm struggle how to draw a goodlooking graph.
Yes, I have to admit I'm a little bit OCD on it.
However, every time when I can decorate a beautiful graph, I feel awesome and proud of myself.

Why `R base graphics`

1, Easy to get me where I want to go.
2, Fast to graph what I need.
3, Many different type of graphs.

Why `ggplot2`

1, Follows a grammar,just like any language.
2, It defines basic components that make up a sentence.
3, Supports a countinuum of experise.

The Advantages and disadvantages of `Base R graphics` and `ggplot2`

`base R graphics`

1) Advantages:

-Documentation and examples. Because base graphics have been around as long as R itself, there is a huge accumulation of information for these functions.
-Ease of customization, especially through the sequential layering of graphical elements. A typical thought process – reflected faithfully in the associated R code – for a fancy plot is to set up a coordinate system and then gradually add elements like axes, data points, fitted curves, legends.

2) Disadvantages:

-Awkward workflow for juxtaposition of many related plots. In base graphics, the user must deal with laying out the individual plots, working to keep axes consistent.
-No built-in support for encoding additional information, especially categorical, via color or line type etc.
-Good thing it's easy to customize because you'll be doing it all the time in base graphics. Seriously. Tedious stuff that truly is better handled by a computer is constantly left for you to grind out on your own.

`ggplot2`

1) Advantages:

-Multi-panel conditioning in `ggplot2`, it is easy to break data into subsets and display them as an array of related plots. -Mapping variables into aesthetics. The encoding of a factor, such as male vs. female, via shape or color, is just the tip of the iceberg here.
-Combination of different layers, such as raw data depicted via a scatterplot overlaid with a fitted smooth regression surrounded by a shaded prediction region.
-The over-arching grammar is a valuable system within which to make the graphics of your dreams.

2) Disadvantages:

-Need to understand the Big Idea – the Grammar of Graphics – before individual functions and examples start to make much sense.
-Monolithic/unexpected calling style.
-Difficulty of extreme customization.
-Speed.

Compare the `R base graphics` with `ggplot2`

first we need install the package.

```r
#loada package
library("ggplot2")
```

```r
dat <- data.frame(
  time = factor(c("Lunch","Dinner"), levels=c("Lunch","Dinner")),
  total_bill = c(14.89, 17.23)
)
```

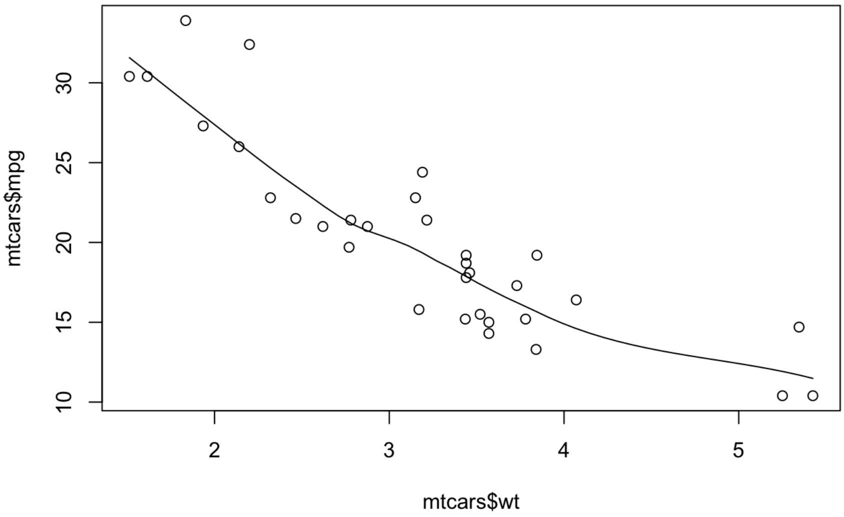
Basic plots

```r
head(mtcars)
```

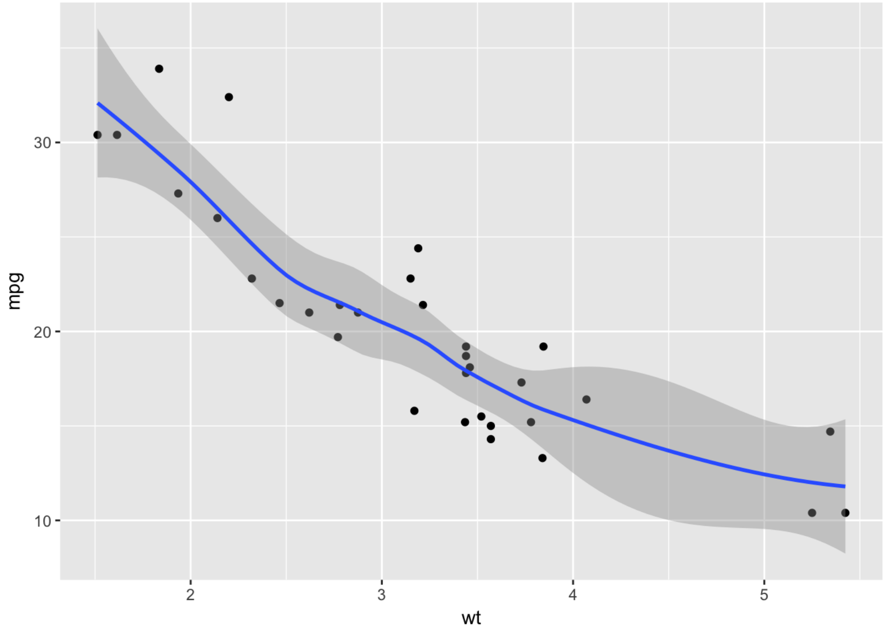
```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

scatter.smooth V.S. ggplot

```
scatter.smooth(mtcars$wt, mtcars$mpg)
```



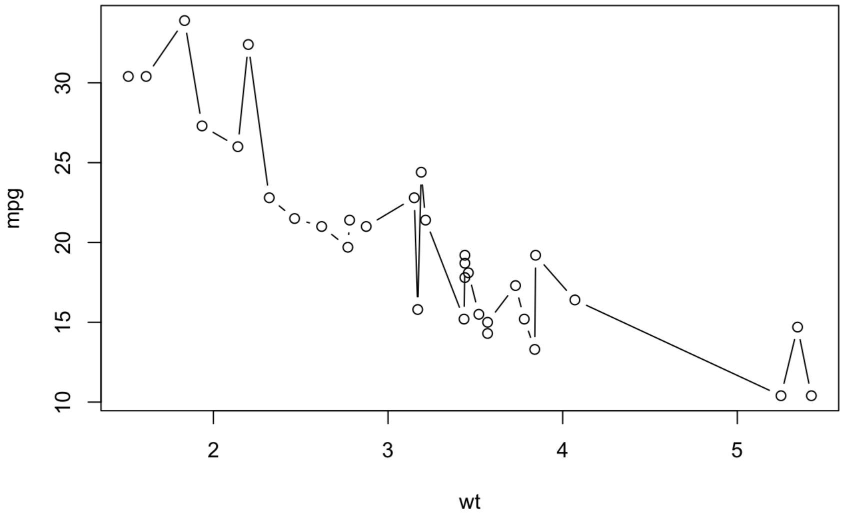
```
ggplot(mtcars, aes(x=wt, y=mpg)) + geom_point() + geom_smooth(method=loess)
```



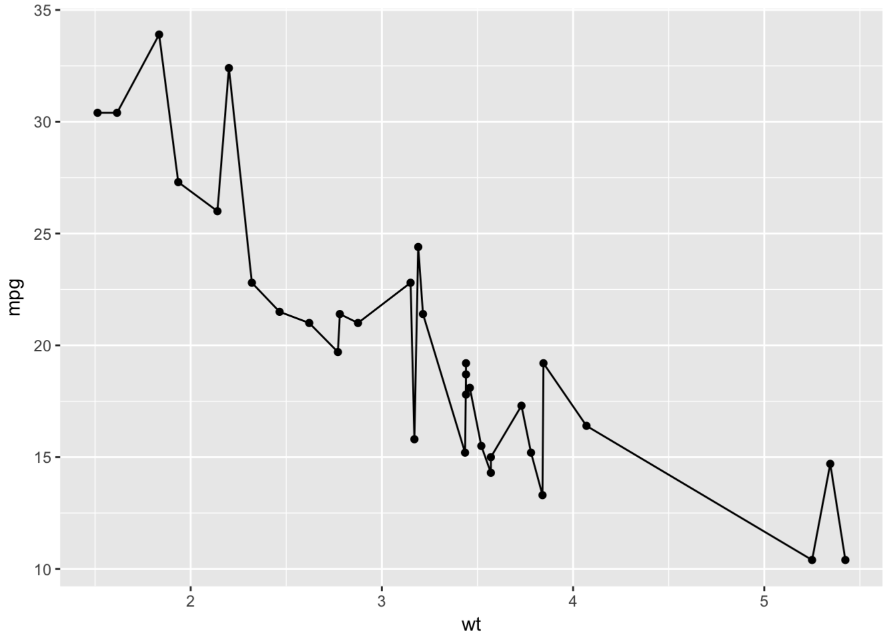
Compare these two graphs, we can see ggplot2 could just split up data by a specific grouping and then show each separately on a graph without the user having to individually write code. The dash area make the graph looks more visualization. Contrary, the `scatter.smmoth` only contains the plots and line. It's more basic than ggplot graph.

plot V.S ggplot

```
df <- mtcars[order(mtcars$wt),]
with(df, plot(wt, mpg, type="b"))
```



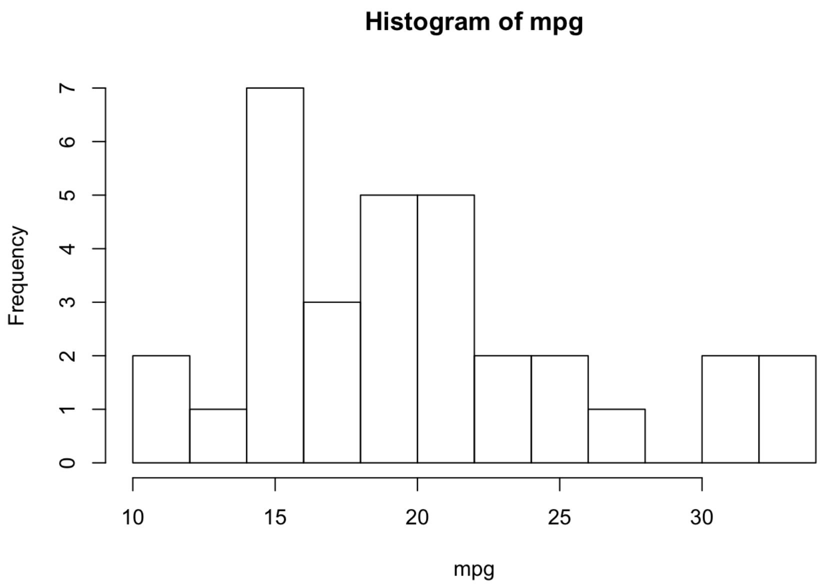
```
ggplot(mtcars, aes(x=wt, y=mpg)) + geom_point() + geom_line()
```



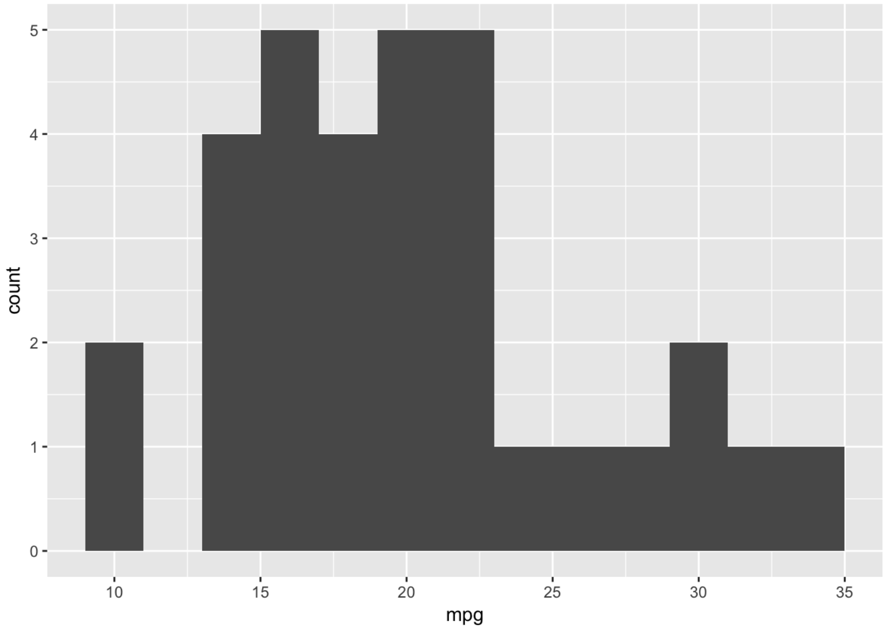
Since the ggplot have some aesthetics set to make the graph connect the dots, such as in this case, I use `geom-point` and `geom_line` to make the graph looks much better than `plot` graph.
And also we can see `plot` graph make a small circle instead of a dot. And between line and line, it's fractured.

histogram V.S. ggplot

```
with(mtcars, hist(mpg, breaks=10))
```



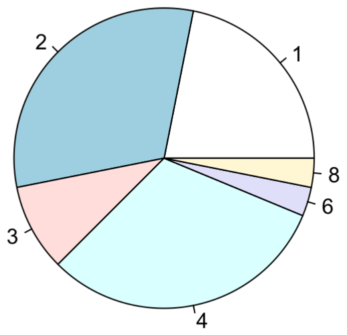
```
ggplot(mtcars, aes(x=mpg)) + geom_histogram(binwidth=2)
```



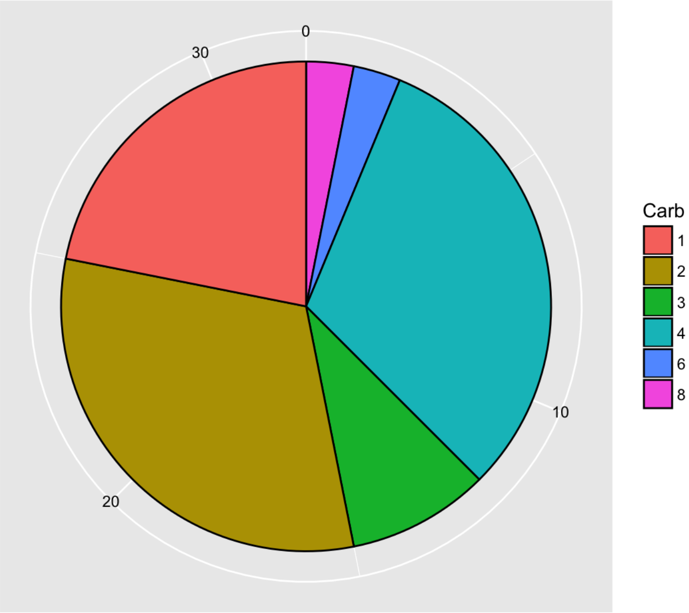
When we draw a histogram, ggplot helps us to read the x-axis more clearly and equably.
The tick labels are smaller than the axis labels and a light gray. Comtrary, the R base graphics make the histogram axis not equably，y-axis indicate the frequency, not like ggplot it's counting number.
Both of them are great graph, it depends on which one you are more familiar.

pie V.S. ggplot

```
with(mtcars, pie(table(carb)))
```



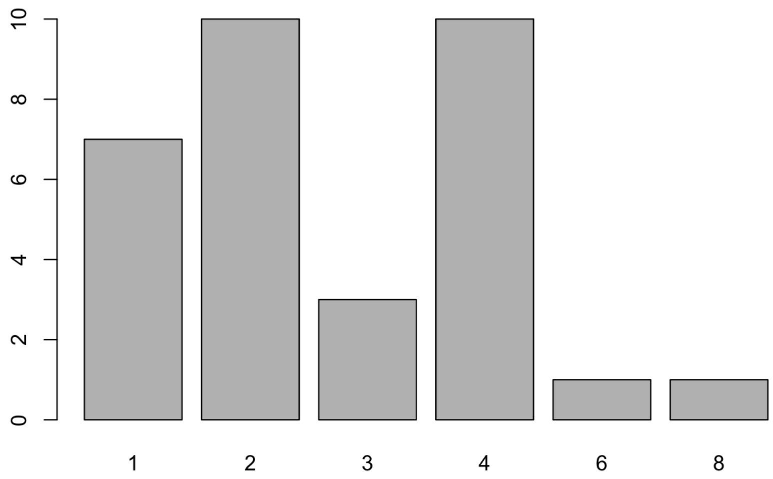
```
df <- data.frame(table(mtcars$carb))
colnames(df) <- c("Carb", "Freq")
ggplot(df, aes(x=1, y=Freq, fill=Carb)) +
geom_bar(stat="identity", color="black") +
coord_polar(theta="y") +
theme(axis.ticks=element_blank(),
      axis.text.y=element_blank(),
      axis.text.x=element_text(colour='black'),
      axis.title=element_blank())
```



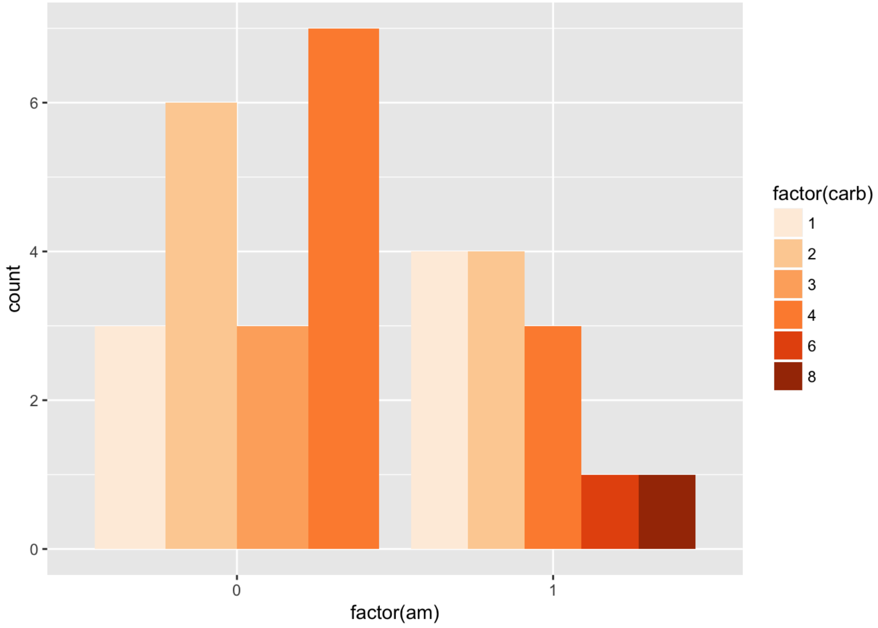
When we use the R base graphics to draw, we can basically use the `with()` function, imports the data, and then use `table()`, we can get a basic pie graph. On the other hand.
if we use `ggplot`, we can use bright color to separate the different carbs in this case. And different number shows different colors.

barplot V.S. ggplot

```
with(mtcars, barplot(table(carb)))
```



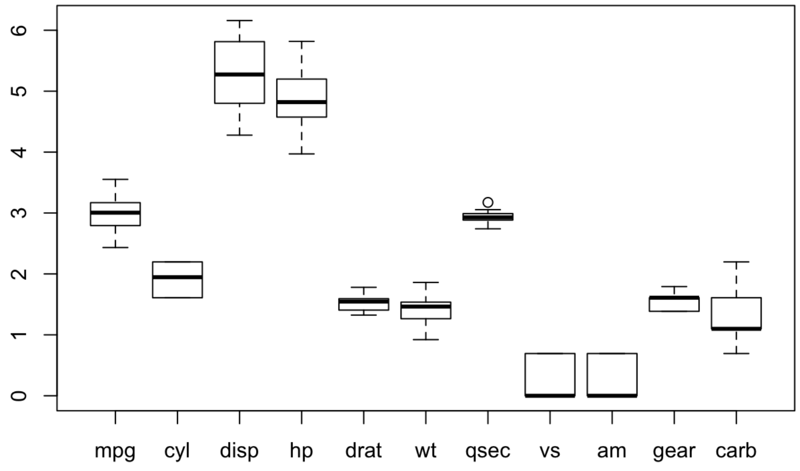
```
ggplot(mtcars, aes(x=factor(am), fill=factor(carb))) +
geom_bar(position="dodge") +
scale_fill_brewer(palette="Oranges")
```



The `barplot` just give us a rough graph, we even can't read the x-axis clearly.
But the `ggplot` use two aestheics which is `geom_bar` and `scale_fill_brewer` make the each matrix from left to the right gradually deepen the orange color。

boxplot V.S. ggplot

```
boxplot(log1p(mtcars))
```



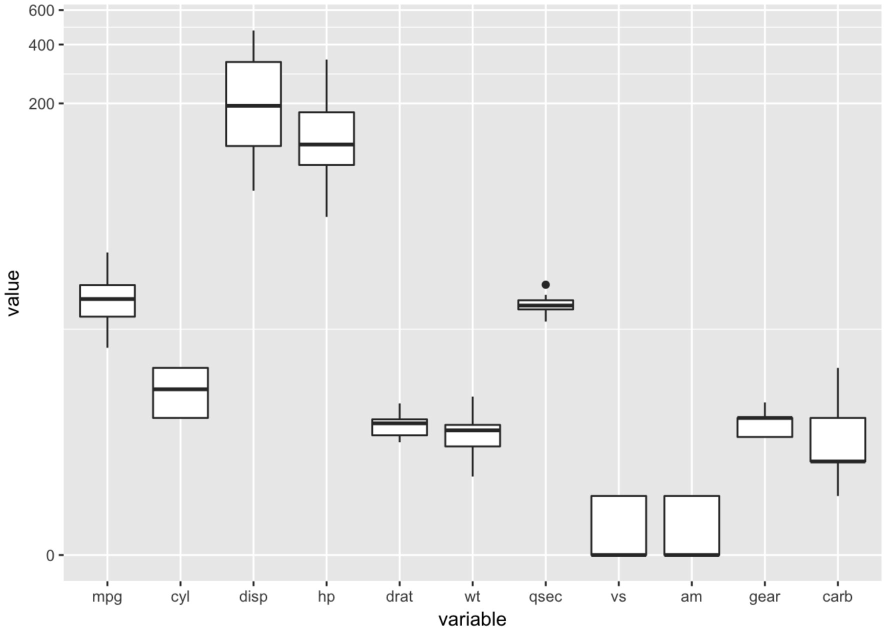
```
df <- head(mtcars)
```

```
library(reshape2)
```

```
df <- melt(mtcars)
```

```
## No id variables; using all as measure variables
```

```
ggplot(df, aes(x=variable, y=value)) +
geom_boxplot() +
scale_y_continuous(trans="log1p")
```



Use the `boxplot`, we can see there is no main title on the axies. Each variables on the axis not that clearly shows like as ggplot. But it does a easy and fast way to graph a boxplot.
Use the `ggplot`, after we import the data, we can show the x-axis and y-axis vaiable names in different range. Then use the `geom_bocplot` and `scale_y_continuous` to set where the limits y axis starts and stops.

Take home message

Now that I looked at ggplot2 more closely, do I want to switch away from base graphics? Yes
Base R plotting is indeed a bag of tricks. They're often great tricks (especially for their time, and relative to what came before!). But why does plot expect one row per observation, while matplot expects one column per group? Because they're two different ways to build two different plots. Why do you call plot(…, type = "l") to add the first line to a graph, then lines to add each additional one? Because that's what those two different tricks are for. So to Jeff, faceting, color scales, legends, and all the things that made the above example shorter are useful tricks that ggplot2 has that that base R doesn't. They're "points scored" in the ggplot2 column. in this post and elsewhere, that ggplot2 is restrictive: that there is some superset of graphs that cannot be expressed in ggplot2 but can in base plotting. I'm skeptical of this, simply because I've been looking for such a graph in several years of professional ggplot2 use and it's pretty rare to run into one.

References

1,http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.htm
2, https://flowingdata.com/2016/03/22/comparing-ggplot2-and-r-base-graphics/
3,https://www.stat.ubc.ca/~jenny/STAT545A/block90_baseLatticeGgplot2.html
4,http://varianceexplained.org/r/teach_ggplot2_to_beginners/
5,https://en.wikipedia.org/wiki/Ggplot2#Comparison_with_base_graphics_and_other_packages
6,https://www.quora.com/How-do-R-programmers-choose-among-plotting-systems-base-lattice-ggplot2-etc
7,http://duke-biostats.github.io/ngs-2015/BaseGraphicsGGPlotComparison.html