Theme in ggplot2 package

Content:

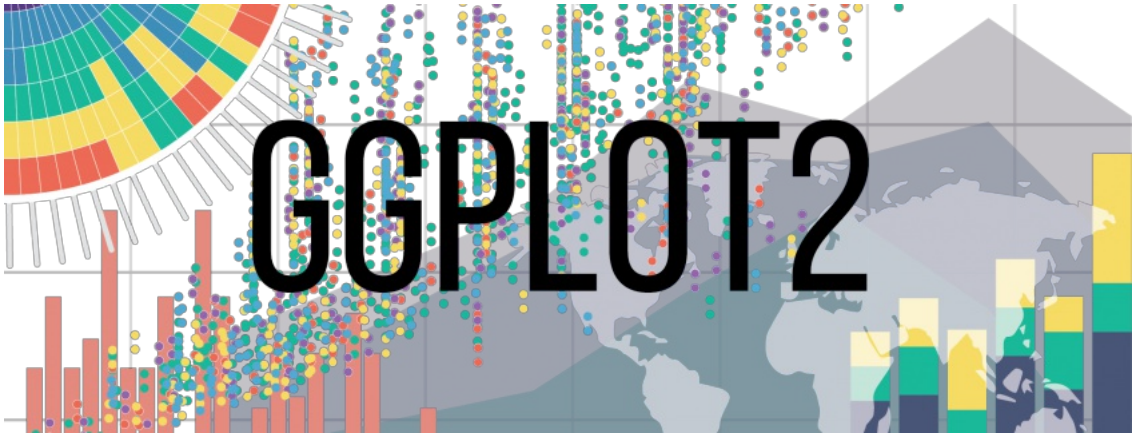
- The introduction of ggplot package

- Why ggplot2?

- What we can do with ggplot2 ?

- Modify components of a theme in ggplot

- Example



(source: https://datasciencedojo.com/wp-content/uploads/12_R-and-GGPLOT2-01-845x321.png)

Introduction:

> "*ggplot2* is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts (1)."
>
> "ggplot2 package, created by Hadley Wickham, offers a powerful graphics language for creating elegant and complex plots (3). Its popularity in the R community has exploded in recent years. Origianlly based on Leland Wilkinson's The Grammar of Graphics, ggplot2 allows you to create graphs that represent both univariate and multivariate numerical and categorical data in a straightforward manner(6). Grouping can be represented by color, symbol, size, and transparency. The creation of trellis plots (i.e., conditioning) is relatively simple(5)"

I'd like to introduce ggplot2. Based on the lecture so far, ggplot2 is a useful package in R, which helps plotting.

Until January 27,2017, There are more than 10,000 R packages available for download (2).

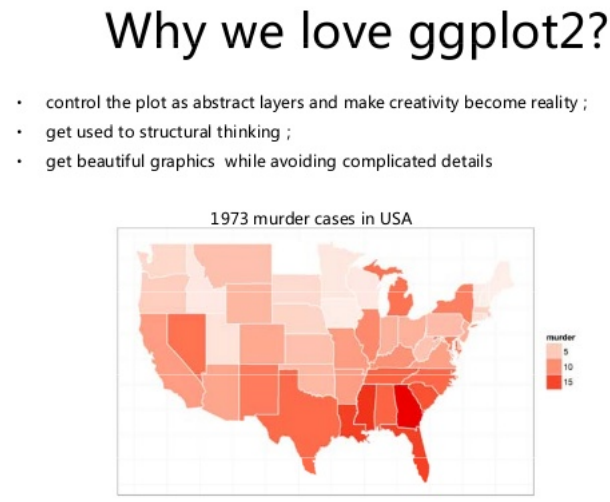
Some of them are common packages (4),

- *Rcpp* "Seamless R and C++ Integration (693288 downloads, 3.2/5 by 10 users)"
- *ggplot2* "An Implementation of the Grammar of Graphics (598484 downloads, 4.0/5 by 82 users)"
- *stringr* "Simple, Consistent Wrappers for Common String Operations.(543434 downloads, 4.1/5 by 18 users)"
- *plyr* "Tools for Splitting, Applying and Combining Data(523220 downloads, 3.8/5 by 65 users)"
- *colorspace* "Color Space Manipulation (476304 downloads, 4.0/5 by 2 users)"

and so on……

What makes ggplot2 special?

- *ggplot2* is a data visualization package for the statistical programming language R (3).
- This package was created by Hadley Wickham in 2005. It is a young package. *ggplot* is an implementation of a general scheme for data visualization, which called Leland Wilkinson's Grammar of Graphics.
- The most special thing is that this grammar breaks up graphs into semantic components such as scales and layers.



(source : https://image.slidesharecdn.com/rworkshopiii-3hourstolearnggplot2series-140707095607-phpapp01/95/r-workshop-iii-3-hours-to-learn-ggplot2-series-4-638.jpg?cb=1404727723)

Why choose *ggplot2*?

- special grammar of graphics
- flexible
- high level plot specification
- theme system

How to built the grammar of graphics?

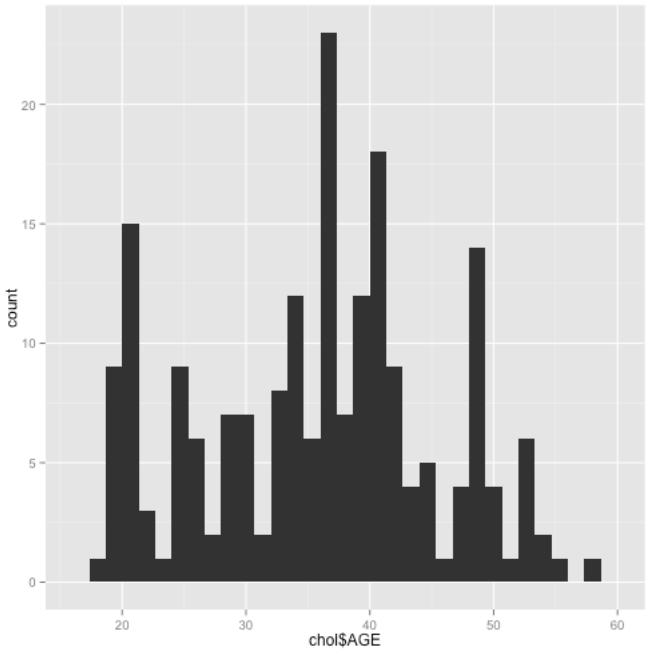
- data
- aes mapping
- scale
- coordinate system
- statistical transformations

Modify components of a theme in ggplot

theme() is a most interesting part for me in *ggplot*, which can help to control the appearance of all non-data components of the plot.

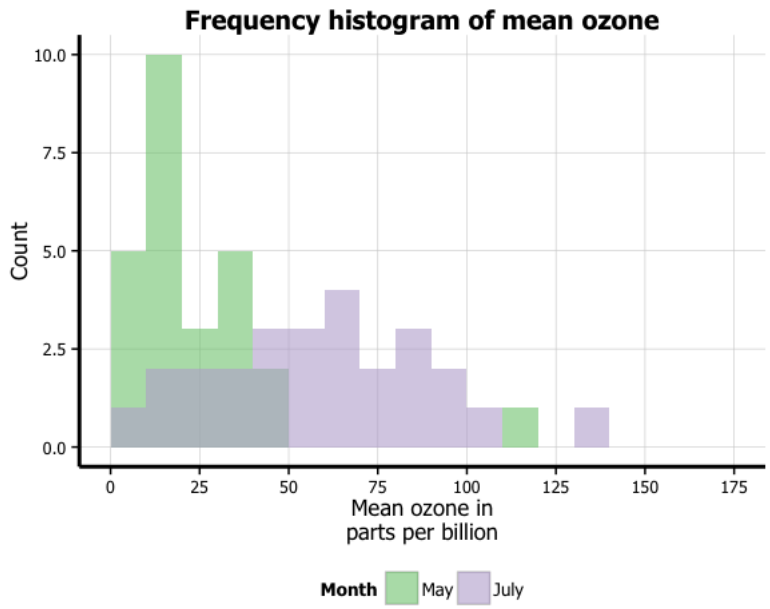
- *line*
- *rect*
- *text*
- *axis.title*
- *legend.background*
- *plot.title*
- *panel.border*
- …

> from simple plot



(source: https://s3.amazonaws.com/assets.datacamp.com/blog_assets/Histograms+R/ggplot2/9392mV3.png)

> to a complecated plot

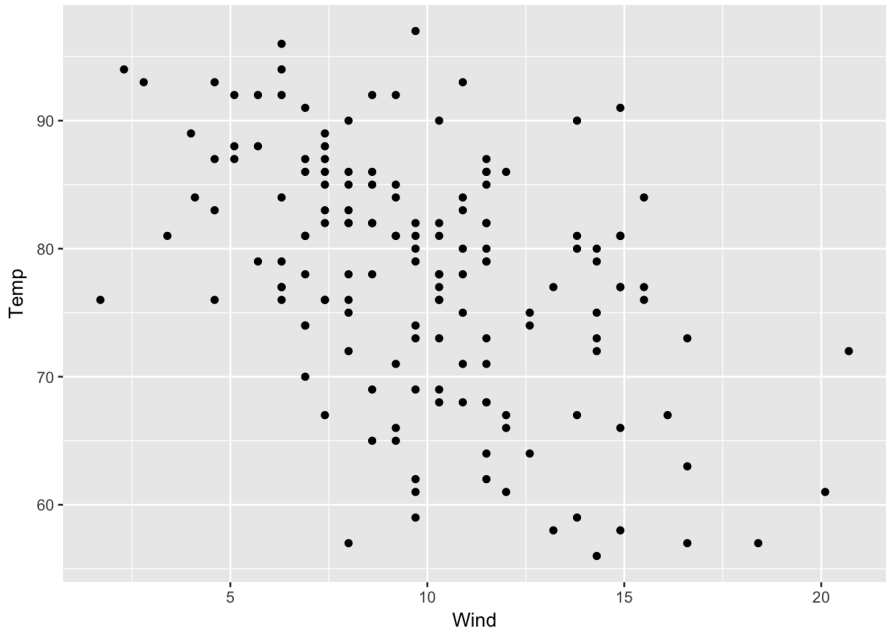


(source: http://t-redactyl.io/figure/histogram_20-1.png)

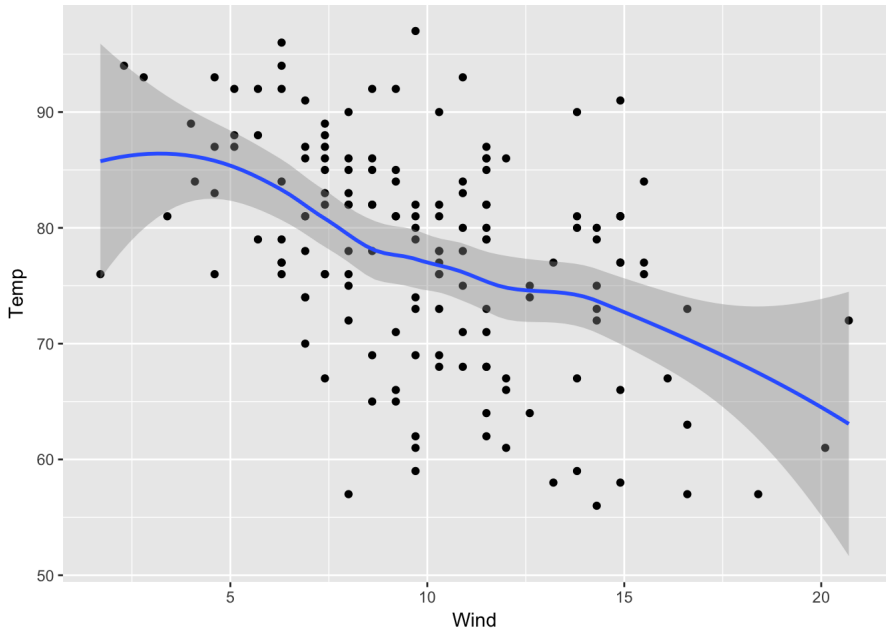
Examples

Example Data: *airquality in datasets*

```
library(datasets)
library(ggplot2)
# creating a now data frame from datasets
Air <- as.data.frame(airquality)
# creating a ggplot graph
p <- ggplot(data = Air, aes(x = Wind, y = Temp))
p1 <- p + geom_point()
p1
```



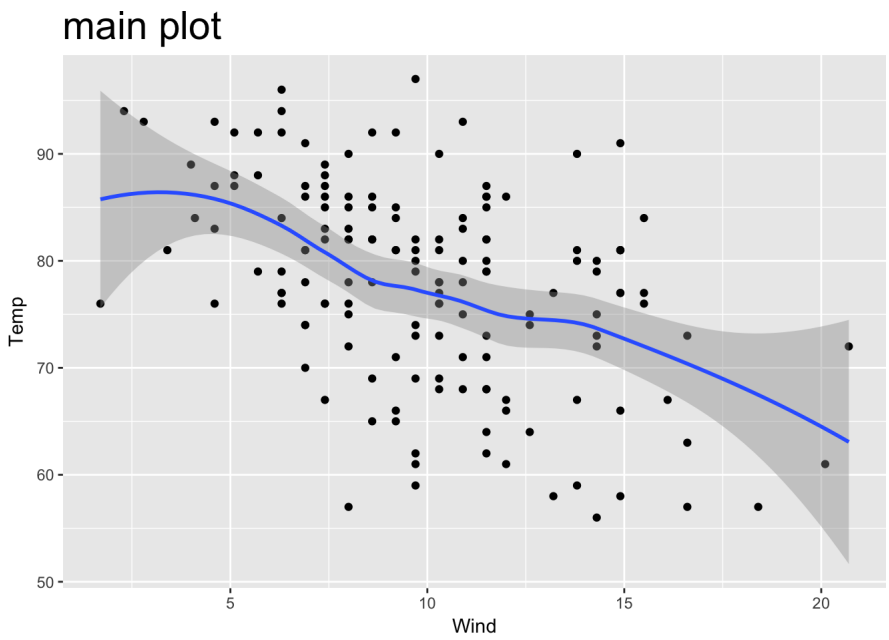

```
# add a smooth line to plot
p2 <- p1+geom_smooth(method = "loess")
p2
```



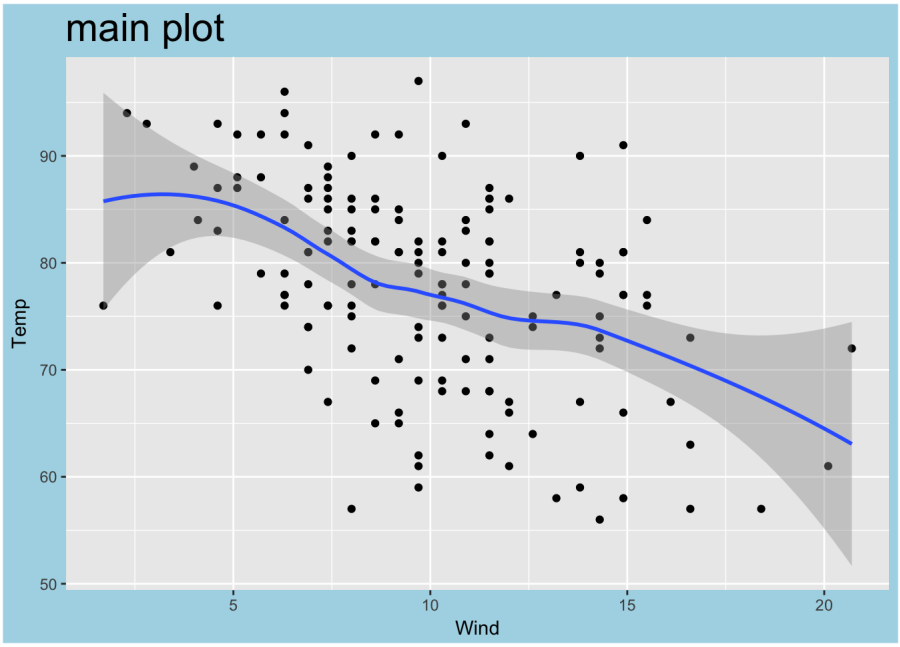
It shows that the grammar of ggplot2 is piece by piece

Then, let us use theme add something.

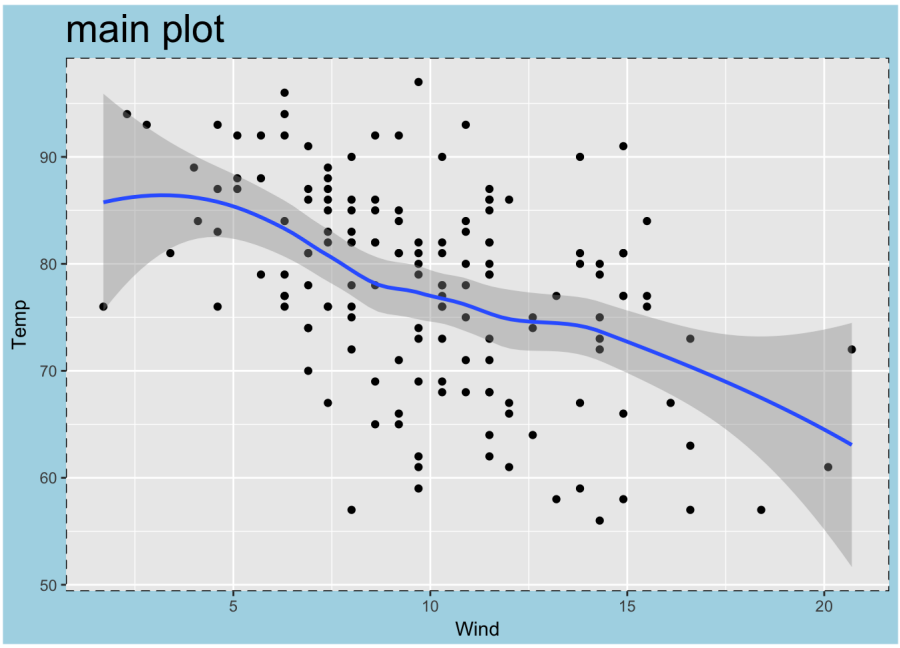
```
# add the main name of this plot with front size = 2
p2 <- p2 + labs(title = "main plot")+ theme(plot.title = element_text(size = rel(2)))
p2
```



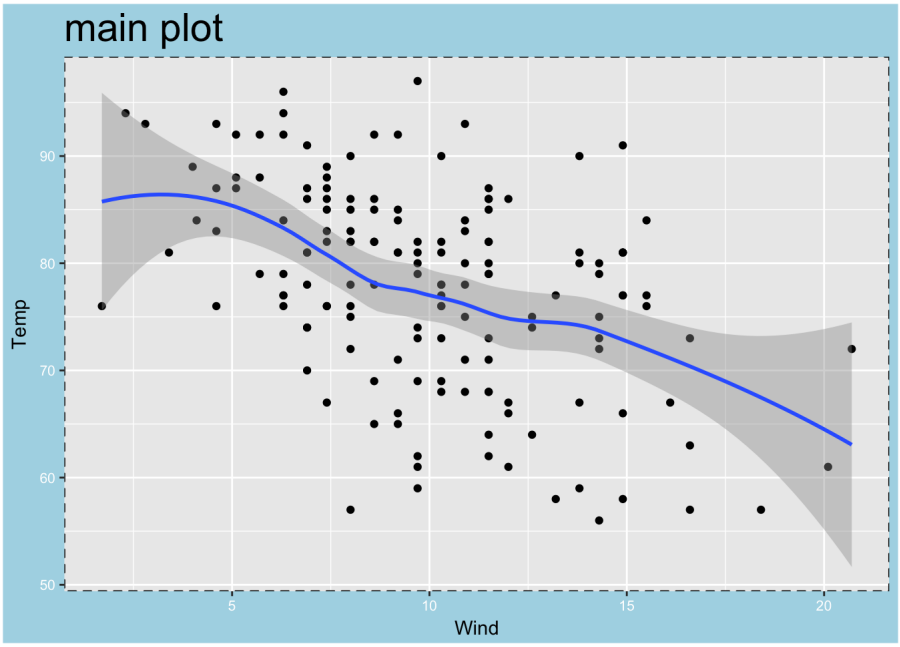

```
# change the color of background
p2 <- p2 +theme(plot.background = element_rect(fill = " light bLue"))
p2
```




```
# remove the border of plot
p2 <- p2 + theme(panel.border = element_rect(linetype = "dashed", fill= NA))
p2
```

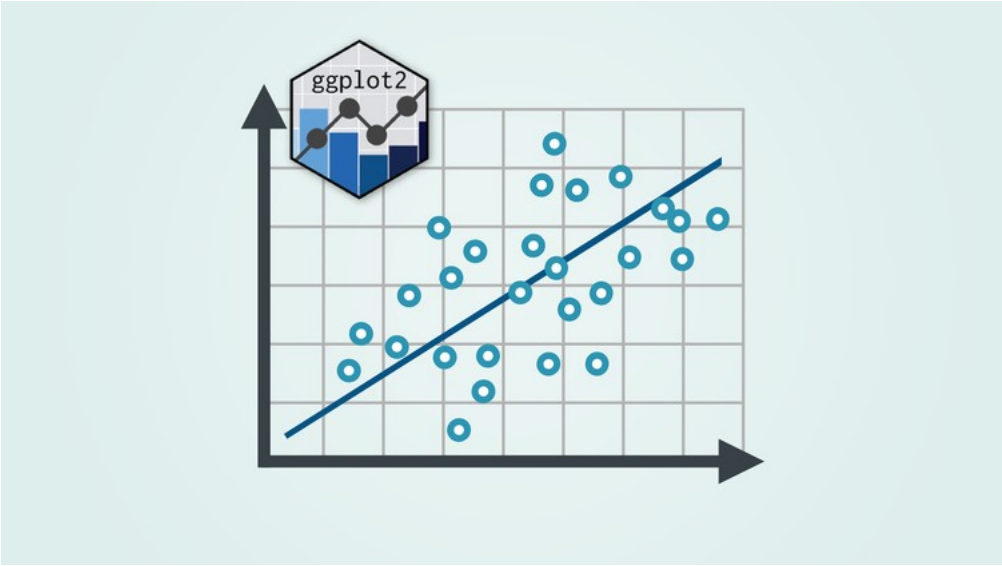



```
# change color of axis to blue
p2 <- p2 + theme(axis.text = element_text(size = 8,colour = "white"))
p2
```




```
##
```

The future development of ggplot2:



(source :https://udemy-images.udemy.com/course/750x422/365790_185f_3.jpg)

So far, ggplot2 is a good method for people to output plot, however, there are three main things that you can't do with ggplot.

- 3-dimensional graphics
- Graph-theory type graphs
- Interactive graphics

For example, the shiny-app part in HW04, I use ggplot to output plot first. It much harder than ggvis.

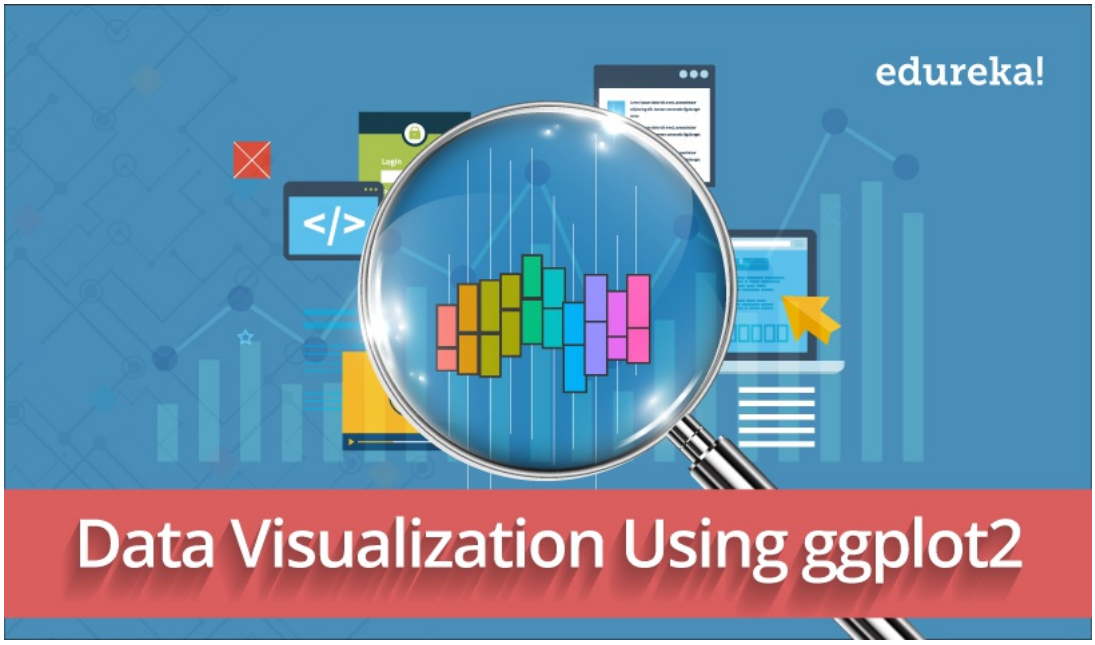
> "Being able to visualize the data is important today, and it's going to be even more important going forward (7)."

Simon Samuel, head of customer value modeling for bank in UK, has a view that data visualization will place an important position in future. It is true that human society step into big data era, which means we should find more efficient way to process and present data in an easier way for decision makers. Data visualization is exactly a **potential** method.

In my first post, I explain the inportance of Data Visualization. Now, there is a good mehtod we can use to improve our plot.

Summary

In this post, I introduce more about what is ggplot2 and explain the advantages of ggplot2. I also connect ggplot2 to lecture by using examples, which is based on lecture but more focuse on theme() part. As what I mentioned in last post, data visualization is powerful and potential method. Meanwhile, *ggplot2* the package of R is really helpful. To satisfy different conditios, theme() part can control many non-data elements.The important point is not only about the data itself, but also about how to exhibition the result *efficiently* and *evidently*.



(source:https://cdn.edureka.co/blog/wp-content/uploads/2017/10/data-visualization-using-ggplot2.png)

Works Cited

1. "ggplot2." *ggplot2*, ggplot2.org/.
2. "KDnuggets." KDnuggets Analytics Big Data Data Mining and Data Science, www.kdnuggets.com/2015/06/top-20-r-packages.html.

3. “Ggplot2.” Wikipedia, Wikimedia Foundation, 31 Oct. 2017, en.wikipedia.org/wiki/Ggplot2.
4. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
5. Robert Kabacoff -. “Graphics with ggplot2.” Quick-R: ggplot2 Graphs, www.statmethods.net/advgraphs/ggplot2.html.
6. techATstate. *“Tech@State: Data Visualization - Keynote by Dr Edward Tufte.”* YouTube, YouTube, 7 Aug. 2013, www.youtube.com/watch?v=g9Y4SxgfGCg.
7. *“Data visualization: A wise investment in your big data future.”* SAS, www.sas.com/en_us/insights/articles/analytics/data-visualization-a-wise-investment-in-your-big-data-future.html.
8. Quoted in M. Wood. 1994. *“The Traditional Map.”* in Visualization in Geographic Information Systems, Edited by Hilary Hearnshaw & David Unwin. John Wiley & Sons.
9. https://datasciencedojo.com/wp-content/uploads/12_R-and-GGPLOT2-01-845x321.png
10. https://image.slidesharecdn.com/rworkshopiii-3hourstolearnggplot2series-140707095607-phpapp01/95/r-workshop-iii-3-hours-to-learn-ggplot2-series-4-638.jpg?cb=1404727723
11. http://www.rmanchester.org/presentations/2013/05/ManchesterR_-_Introduction_to_ggplot2_-_Dennis_Prangle_-_20130502.pdf

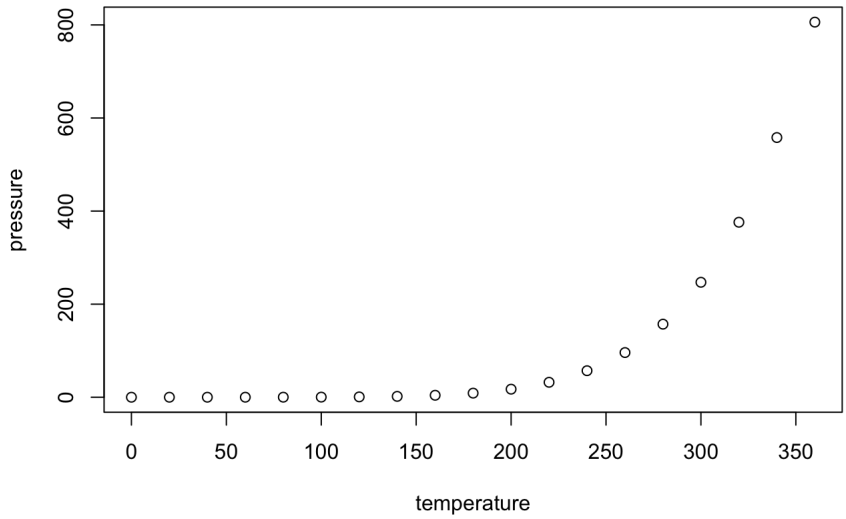
R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.