Data Visualization

Content:

Introduction:

> "Visuals are a powerful way to convey message present information, and persuade audiences (1)."

Do you agree or disagree? Before you answer my question, there are some facts and examples.

- Human's eyes see images in less than **1/10** of a second.
- In the brain, the process of visuals is **60,000** faster than the process of text.
- The **most** information transmitted to the brain is visual.
- Human's eyes review **36,000** visual information per hour.

The above examples show the importance and the power of the visuals. The next question is how should we apply visuals in our project? The answer is **data visualization**. At the begin of this class, we have already learned that visualization is an important part of data analysis. In this post, I will explain data visualization in details. **"what is data visualization?"** and **"why is data visualization important?"** This post will answer two questions. It will also contain the information of data visualization in R, especially the tools. As example, I will show how to make pie-chat in R.

What is Data Visualization?

> "This birth of statistical thinking was also accompanied by a rise in visual thinking: diagrams were used to illustrate mathematical proofs and functions; nomograms were developed to aid calculations; various graphic forms were invented to make the properties of empirical numbers– their trends, tendencies, and distributions—more easily communicated, or accessible to visual inspection (4)."
>
> "The main goal of data visualization is its ability to visualizedata, communicating information clearly and effectively. (5)"

Basically, data visualization is a method/ presentation of data in a pictorial format, such as statistical graphics, plots and information graphics. The main goal of data visualization is to present information *clearly* and *efficiently*. Moreover, *"It involves the creation and study of the visual representation of data (2)."* It is both an art and a science. Data visualization helps people to see analytics in visual, which is easier to grasp difficult identify and confusing concepts in big data. Most of data visualization are created for human consumption. It is interesting that human can distinguish those differences in line length, shape, color and direction without any sign post. Data visualization gives senses in intuitive way.

Why Data Visualization matters?

> "In other words, we see – and use and use – more of our data more of our data and process it with visual pattern recognition as the basis of o and process it with visual pattern recognition as the basis of our interpretation (8)."

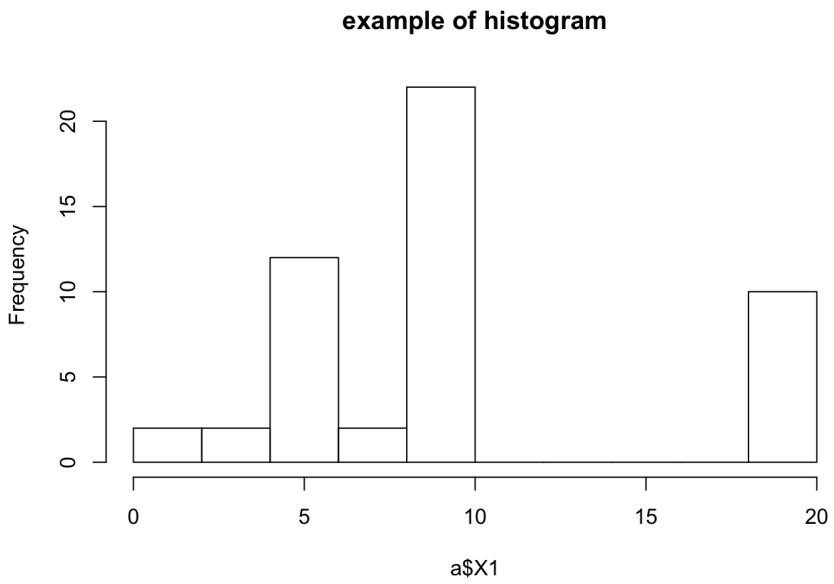
Visual is a powerful way to present information. Especially, human's brains process visual information in the fastest speed. Using plots and graphs to visualize large data is intuitive and efficient. "Data visualization is going to change the way our analysts work with data. They're going to be expected to respond to issues more rapidly. And they'll need to be able to dig for more insights – look at data differently, more

imaginatively. Data visualization will promote that creative data exploration. (Simon Samuel)”

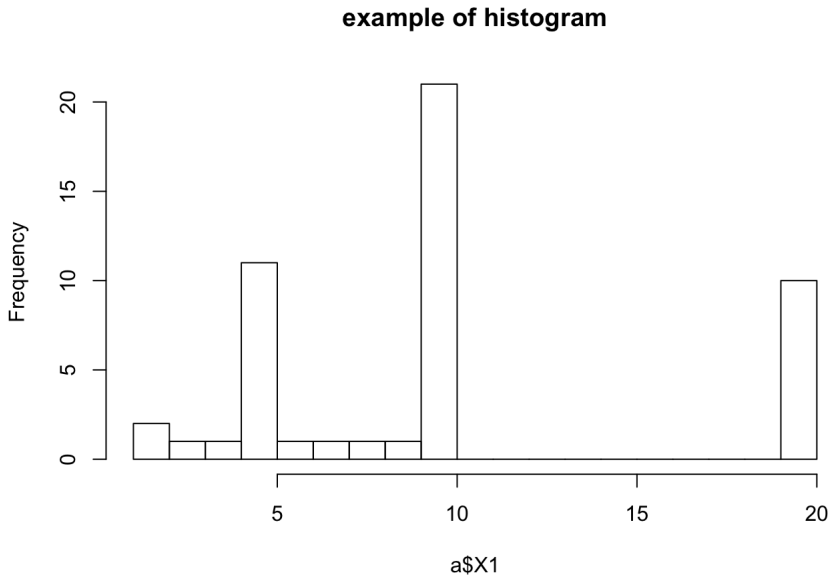
Information of Data Visualization in R:

- The very basic visualization in R:
    - Histogram: it is a plot that breaks the data into bins, which shows the frequency distribution of these bins. Example:

```
# example of histogram
a <- data.frame("1"= c(1:10,rep(10,20), rep(20,10), rep(5,10)), "2" = 1:50)
hist(a$X1, breaks = 10, main = "example of histogram")
```

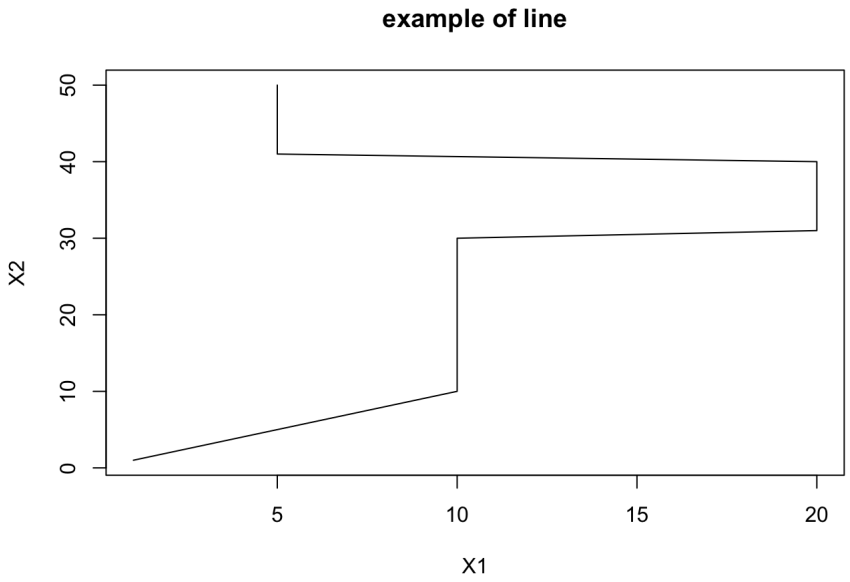



```
# different breaks
hist(a$X1, breaks = 20, main = "example of histogram")
```



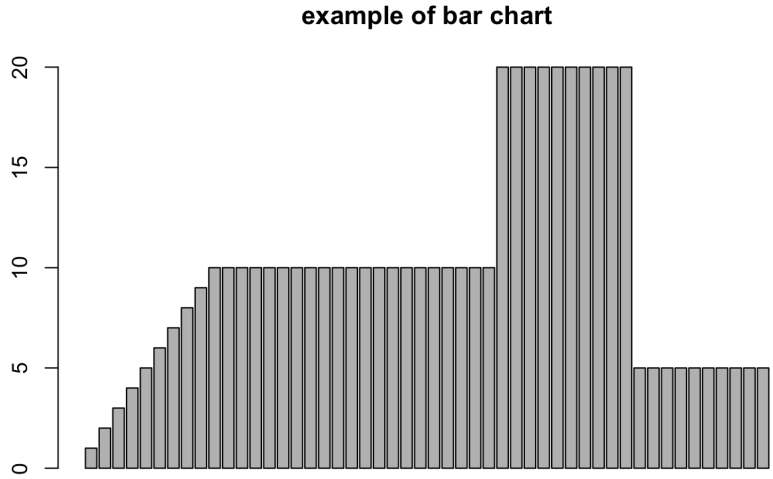
- Line Chart: it is a plot that shows the trend spread over a time period. It is also a good presentation in comparing relative changes. Example:

```
# example of line
plot(a, type = "l", main = "example of line")
```



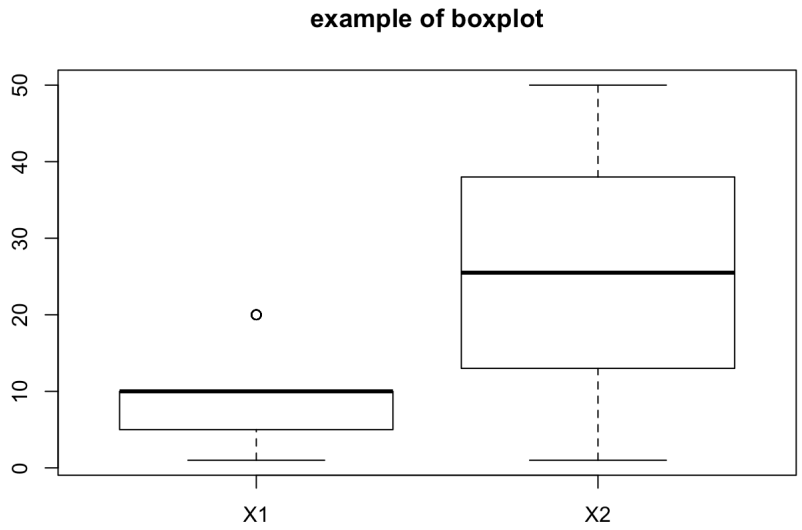
- Bar Chart: Bar plots are also suitable for showing comparisons like line chart. Moreover, it shows the cumulative totals across several groups. Example:

```
#example of bar chart
barplot(a$X1, main = "example of bar chart")
```



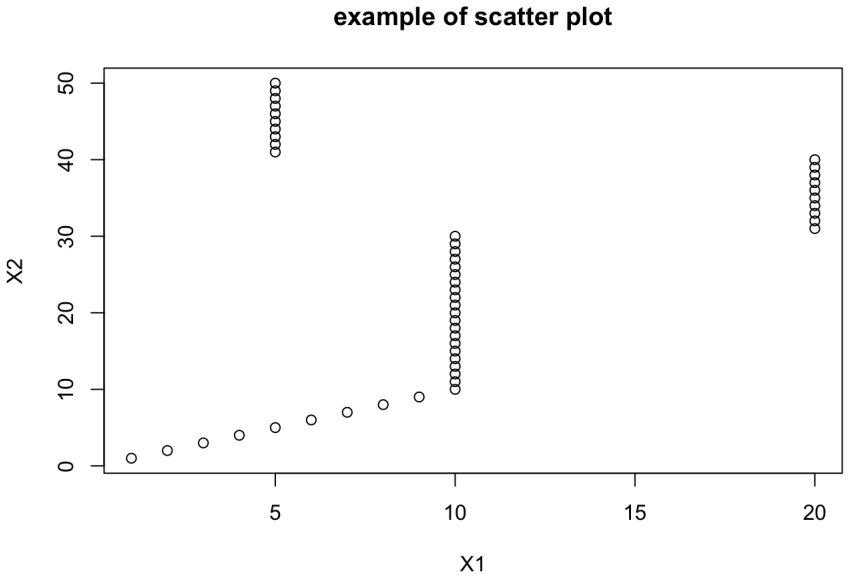
- Box Plot: *It is thus useful for visualizing the spread of the data is and deriving inferences accordingly (3).* There are five statistically significant numbers. They are the minimum, the 25th percentile, the median, the 75th percentile and the maximum. Example:

```
#example of boxplot
boxplot(a, main = "example of boxplot")
```



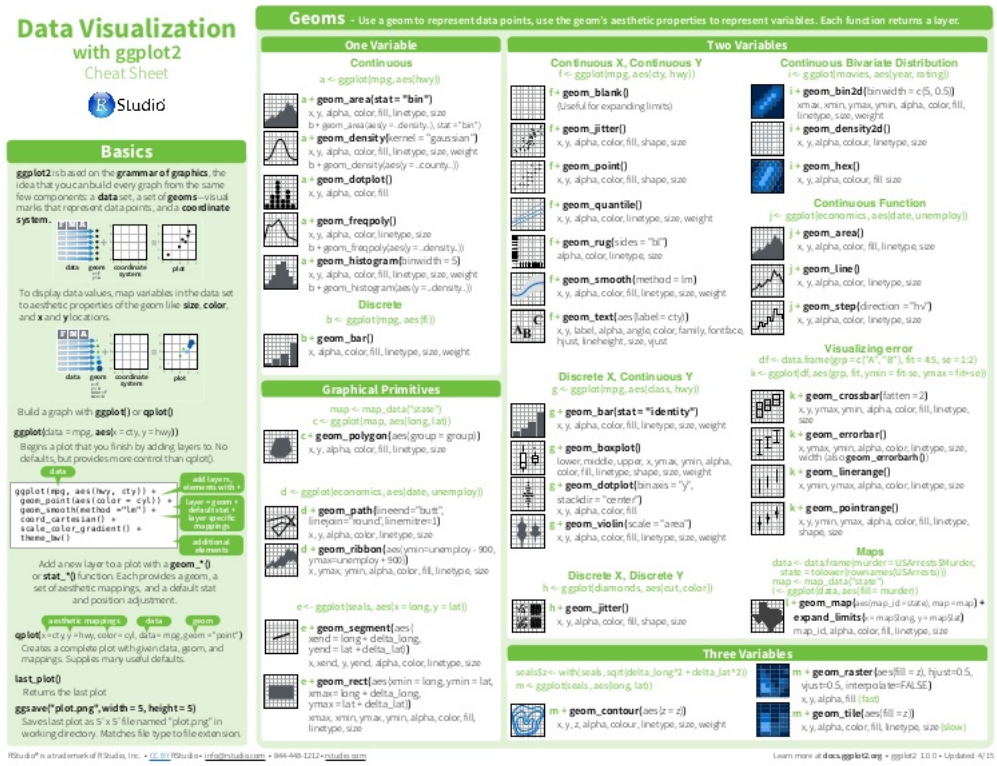
- Scatter plot: it helps in visualizing data intuitively and for simple data inspection. It shows the distribution of data. Example:

```
# example of plot
plot(a, main = "example of scatter plot")
```



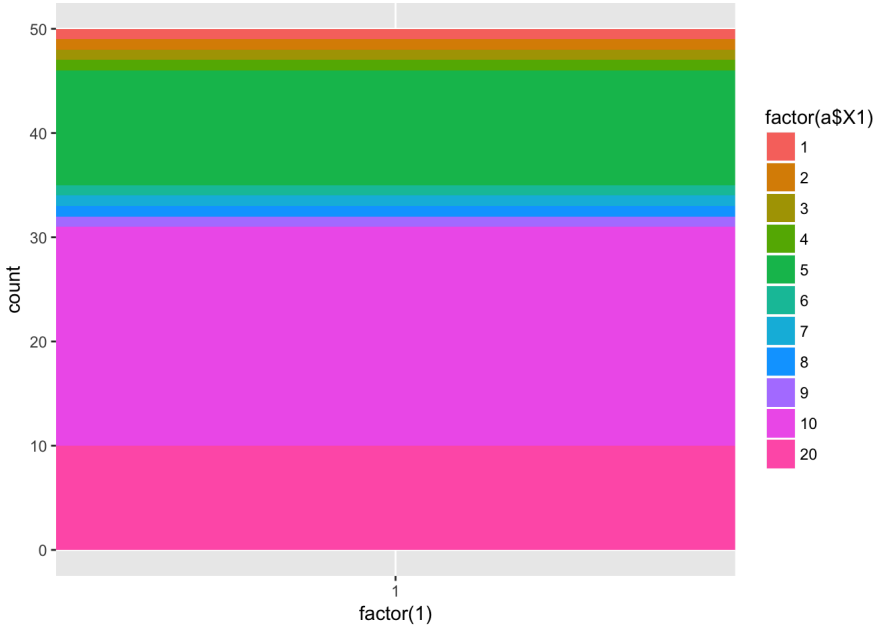
- The adventure visualization in R:
- **ggplot2**: 

we learn this in class, simply, it is a plotting system for R, which based on the grammar of graphics. It cares about many details and tries to provide a powerful model of graphics.

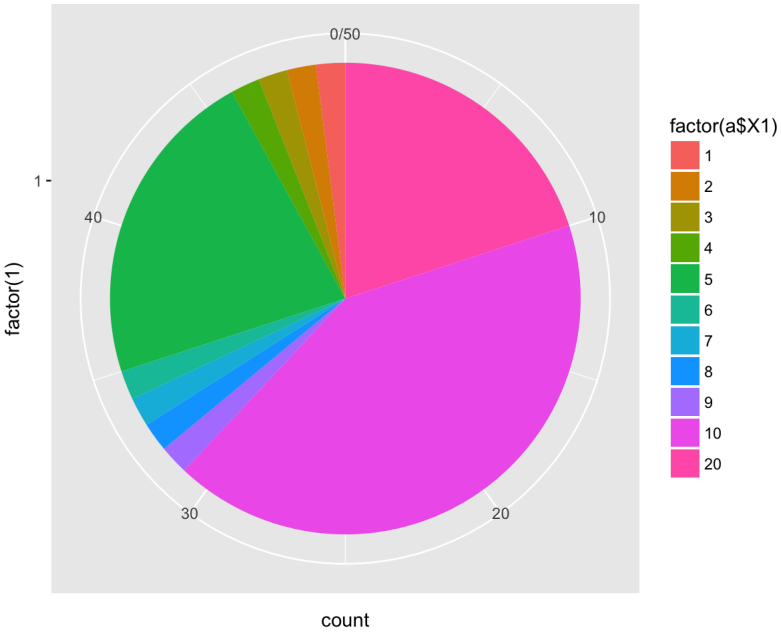


For example, I will show how to use ggplot2 to visualization a pie-chart.

```
# example of pie-chart in ggplot2
library(ggplot2)
# first create a bar plot
bar <- ggplot(a, aes(x = factor(1), fill = factor(a$X1))) + geom_bar(width = 1)
bar
```




```
# second transfer this bar plot into polar chart by adding coord_polar
pie <- bar + coord_polar(theta = "y")
pie
```



The future development of data visualization:

> “Being able to visualize the data is important today, and it’s going to be even more important going forward (3).”

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.

Summary

In this post, I introduce what is data visualization and explain why it is important. I also connect data visualization to lecture by using examples to show how it works. As what I mentioned before, data visualization is powerful and potential method.

Works Cited

1. *“8 Great Examples of Complex Data Visualized.”* Maptive, 12 Feb. 2016, www.maptive.com/8-great-examples-complex-data-visualized-2/.

2. https://cran.r-project.org/web/packages/hexbin/hexbin.pdf
3. *"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.
4. http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf
5. Friedman, Vitaly. *"Data Visualization and Infographics."* Smashing Magazine, 14 Jan. 2008, www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/.
6. techATstate. *"Tech@State: Data Visualization - Keynote by Dr Edward Tufte."* YouTube, YouTube, 7 Aug. 2013, www.youtube.com/watch?v=g9Y4SxgfGCg.
7. https://www.sfu.ca/gis/geog_x55/web355/icons/11_lec_vweb.pdf
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. chttps://www.sas.com/en_us/insights/big-data/data-visualization/_jcr_content/socialShareImage.img.png
10. http://www.dmeforpeace.org/sites/default/files/images/081114_FeatureBlog_0.png

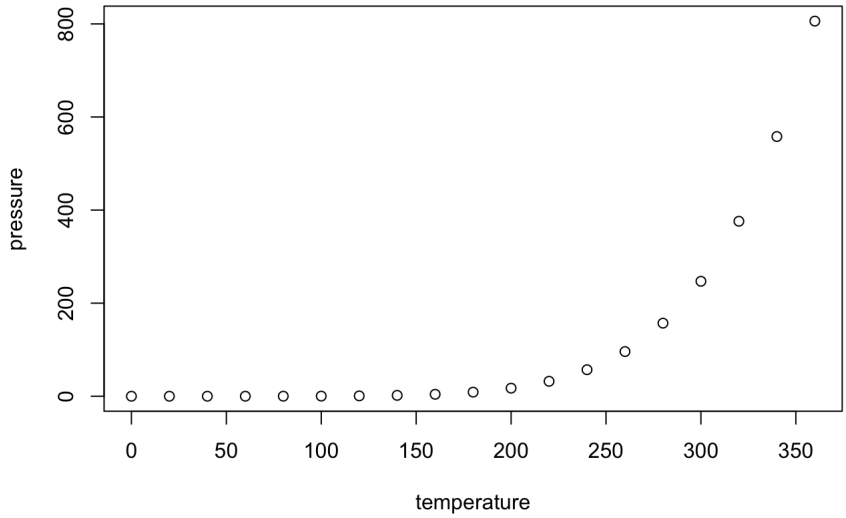
R Markdown

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.