

Notes(1): Vision, Color, Jupyter, R

Here are some notes regarding what I learned while in lecture, reading online and implementing it myself.

Vision:

- Vision here represents how we visualize data we have collected and what we can infer from it.
- Data visualization helps us with
 - better communication with the viewer/user about the relationships of data with images,
 - better understanding of how the data is distributed which allows the user to see its trends and patterns more clearly,
 - Better exploration of data ,
 - And better theory construction.
- Data Visualization uses:
 - How data trends over time,
 - How often relevant events happen over time
 - How two or more variables are related to each other
 - What type of audience to target while doing market research
 - How to analyse value and risk within data
 - Encourages viewer to compare different pieces of data

Color:

- Colors are an effective medium for communicating meaning
- Colors are known to have psychological implications. For example: red is perceived as a warm, powerful color and blue is perceived as a cold and calm color.
- Colors help to set the tone of the visual display
- Colors can also be used to evoke different emotions
- Colors are useful to create associations among data
- Colors help to break camouflage and distinctly show data that stands out
- Colors can complement each other and help the data to stand out
- How you represent the data with different colors has a meaning. There are three types:
 - Categorical
 - Sequential
 - Diverging
- Contrasting colors help in visual search
- Colors help to tell a more effective story, engages the user and captures their attention quickly.

Jupyter :

- Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.
- The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.
- Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala.
- Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer.

R:

- I went through the tutoring by getting an understanding of the R environment, how to install it, and an overview of RStudio.
- You are able to remove individual objects from memory, view them, modify them, and reload them from the command-line or by just executing single lines of code in your R script file (they have the obvious extension of .r) in an IDE like RStudio.
- R gives developers access to a REPL (stands for Read–eval–print loop [2]) environment and that's how you are able to do the above actions seamlessly!

Key Features of R:

- Consists of packages for almost any statistical application one can think of. CRAN currently hosts more than 10k packages.
- Comes equipped with excellent visualization libraries like ggplot2.
- Capable of standalone analyses

R vs Python:

- We know that R and Python both are open source programming languages.
- The major purpose of using R is for statistical analysis, on the other hand Python provides the more general approach to data science.
- R is having the most powerful communication libraries that are quite helpful in data science.
- In addition, R is equipped with many packages that are used to perform the data mining and time series analysis.
- If you use R and you want to perform some object-oriented function than you can't use it on R.
- On the other hand, Python is not suitable for statistical distributions.

- R is more functional, Python is more object-oriented.
- R has more data analysis built-in, Python relies on packages.
- R has more statistical support in general
- It's usually more straightforward to do non-statistical tasks in Python.
- While trying to learn more about the differences between R and Python I came across this comparison online:

R	Python
R codes need more maintenance.	Python codes are more robust and easier to maintain.
R is more of a statistical language and, also used for graphical techniques.	Python is used as a general-purpose language for development and deployment.
R is better used for data visualization.	Python is better for deep learning.
R has hundreds of packages or ways to accomplish the same task. It has multiple packages for one task.	Python is designed on the philosophy that "there should be one and preferably only one obvious way to do it". Hence it has few main packages to accomplish the task.
R is easy to start with. It has simpler libraries and plots.	Learning python libraries can be a bit complex.
R supports only procedural programming for some functions and object-oriented programming for other functions.	Python is a multi-paradigm language. It means python supports multiple paradigms like object-oriented, structured, functional, aspect-oriented programming.
R is a command line interpreted language.	Python strives for simple syntax. It has a similarity to the English language.
R is developed for data analysis, hence it has more powerful statistical packages.	Python's statistical packages are less powerful.
R is slower than python but not much.	Python is faster.
R makes it easy to use complicated mathematical calculations and statistical tests.	Python is good for building something new from scratch. It is used for application developments as well.
R is less popular but still, it has many users.	Python is more popular than R

Source: <https://www.educba.com/r-vs-python/>