

Project Review: A Visual History of Nobel Prize Winners

Description: Use Python and R to visualize the history of Nobel winners from 1901 to 2016 with Seaborn and Pandas library from Python, Tidyverse and ggplot2 library from R. Some datasets and platform are provided by DataCamp.

Method: Python versus R

As we all know, both Python and R are required for data analysis. R is a functional programming language and people usually use it in Statistics. Python is a Computer Science tool for data analysis, which is an imperative language. However, statistics knowledge is required for analyzing data. Developers can use Python in a number of different areas, whereas they use R mostly in Statistics.

In terms of R, some libraries are helpful for data visualization. It is an open sourced programming language, with full Rdocumentation on Github's CRAN, Bioconductor. R is developed by statisticians, meaning you can learn it without any computer science background. However, the speed of R's processing is slow. It's hard for learners to study R, especially when they are trying to learn GUI from statistical analysis.

In terms of Python, IPython Notebook makes data analysis easier since it can be shared easily with others. Python can be used is such different areas, such as testing, developing, and data analysis. Most developers know Python and can be easily combined with statistics. Although there are some libraries such as Seaborn, Bokeh and Pygal for data visualization,

According to my experience, I prefer R as my primary data analysis tool. It's a functional programming language and I have learned Racket at the University of Waterloo, which is another functional programming language. I self-studied Python myself and is not really familiar with some functions. However, Python is still recommended as it is useful in other areas in the industry.