**RoyStock 1.0**

First using web scraping tools in R to Scrap the 100 most popular stocks of the day. Then determine the top 10 good performance stocks among them, using %change as standard. After that using financial data extracting tool to pull off the 10 stock’s history performance in the recent year/month, finally using ggplot2 to visualize the data, compared curves.

Section 1 Web Scraping

Material: Mohan’s thesis. (4 hours)

Package: rvest

After import in url and load html code, we have two options: 1) study html code to fully understand it; 2) find a way to use selector gadget and find the coordinate. Once we have the sheet of 100 most popular stocks of the day, we can sort it.

Section 2 Stock Code List

Material: list [http://www.r-tutor.com/r-introduction/list (1](http://www.r-tutor.com/r-introduction/list%20%20%20%20%20(1) hour)

Package: none

Make a list of 100 element, each of the element is a stock code, named with the corresponding company’s name.

Section 3 Data pull off

Material: tutorial

<https://cran.r-project.org/web/packages/tidyquant/vignettes/TQ01-core-functions-in-tidyquant.html>

<https://www.datacamp.com/community/tutorials/r-web-scraping-rvest>

Package: tidyquant, Lubridate

Using the stock code to pull off data from yahoo finance, using Lubridate to determine the time period of the history data we want.

Section 4 visulization

Material: ggplot2 tutorial <http://r-statistics.co/ggplot2-Tutorial-With-R.html>

Package: ggplot2

Using ggplot 2 to draw comparison curves to realize intuitive visualization.

Additional Step: may need to use dplyr to manipulate data. <http://genomicsclass.github.io/book/pages/dplyr_tutorial.html>

**RoyStock 1.1**

The 1.1 version will focus on optimizing the running process

1. Generate the code list of top performed stocks before pulling off history data, so every time we run it, it will only pull of the needed stocks
2. About install package function, write it in a more professional way: check if installed, only install if didn’t
3. Some stocks didn’t have whole year’s history, right now this would limit the length of history price of all stocks. Try to find a way to show it on final plots while keeping the stocks that have full year’s history data. (fixing needed in the merge part)

**dplyr** in R: <http://genomicsclass.github.io/book/pages/dplyr_tutorial.html>

**RoyStock 1.2**

The 1.2 version will focus on monitor the process and decorating the result.

1. Apply version control (**Github**)
2. Explore more about **ggplot2**

**RoyStock 2.0**

Main thing about 2.0 version is: adding changeable parameters!

Include The length of history as adjustable parameter (dig deeper in **lubridate**).

Adding different ways of determining “top performed stocks”, e.g. %change of not only today but also in recent month, recent year, or other performance index. (Need new web scraping source). The point is about giving include different “options” in user defined functions in R.

**RoyStock 3.0 (TBD)**

Now it’s time to apply R Shiny!

Create user-friendly UI to provide no-learning-curve experience!