A quick review of R and rmarkdown

jsn-jin @ UCLA

Do not forget to the change the working directory to some folder on your PC.

To set up the working directory, maybe you also want to enter the same folder name, e.g., `setwd("Desktop")`*, in the* **Console** *window.*

Advantages of 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. In particular, you can check this R Markdown cheatsheet https://rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf or the Markdown cheatsheet https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet.

The following are some nice features.

R code chunks

Click on "Insert" and then "R" buttons to embed a R code chunk within the document. Then, click the **Knit** button to create a document that *includes both content as well as the output.*

```r
# basic is the name of this code chunk
x <- 2 - 3 * 4 / 5 ** 2
x
```

```
## [1] 1.52
```

As we can see, the result is 2.3104 (known as R inline code).

Math inside R Markdown

It is convenient to type mathematical expressions. **(Note: Do not leave a space between the $ and your mathematical notation.)**

Example 1 (inline math): `$\sum_{n=1}^{100} n^2$` is rendered as $\sum_{n=1}^{100} n^2$.

Example 2 (equation block)):

$$\sum_{n=1}^{10} n^2$$

or

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$
$$= \frac{\$18 - \$15}{\$15}$$
$$= 0.2 \text{ or } 20\%.$$

See this page for more details on LaTex code.

Block-level elements

Section headers can be written after a number of pound signs, e.g.,

First-level header

Second-level header

Third-level header

Order and unordered list items

Ordered list items start with numbers, and Unordered list items start with *, -, or +, and you can nest one list within another list by indenting the sub-list, e.g.,

1. the first item
2. the second item
3. the third item
   - one unordered item
   - one unordered item

See this page for more details.

Italics and bold

- * or _ can be used to note *emphasis*
- ** or __ can be used to **bold text**
- They can be used ***together*** (text in bold italics)

In the future, you are required to write homework answers using R Markdown!

Short introduction to programming with R

We aim to go over some of the commands in this cheatsheet.

Download dataset

load dataset

To import the excel file with .csv extension, we use the function read.csv. It reads from an Excel spreadsheet and returns a data frame.

```r
getwd()
```

```
## [1] "/Users/haojin/Desktop/Economics/applied-econometrics/r_review"
```

```r
# load the dataset in the working directory
df <- read.csv("MSFT.csv")

# set output options to show only 5 significant digits (optional)
options(digits = 5)

class(df) # type of object
```

```
## [1] "data.frame"
```

```r
head(df) # first 6 rows
```

```
##         Date  Open  High   Low Close Adj.Close   Volume
## 1 2018-01-09 88.65 88.73 87.86 88.22    85.455 19484300
## 2 2018-01-10 87.86 88.19 87.41 87.82    85.068 18652200
## 3 2018-01-11 88.13 88.13 87.24 88.08    85.320 17808900
## 4 2018-01-12 88.67 89.78 88.45 89.60    86.792 24271500
## 5 2018-01-16 90.10 90.79 88.01 88.35    85.581 36599700
## 6 2018-01-17 89.08 90.28 88.75 90.14    87.315 25621200
```

```r
summary(df) # summary statistics
```

```
##         Date          Open            High            Low            Close
##   2018-01-09:  1   Min.   : 86.3   Min.   : 88.1   Min.   : 83.8   Min.   : 85
##   2018-01-10:  1   1st Qu.:101.6   1st Qu.:102.3   1st Qu.:100.5   1st Qu.:101
##   2018-01-11:  1   Median :110.7   Median :111.5   Median :109.4   Median :111
##   2018-01-12:  1   Mean   :116.5   Mean   :117.4   Mean   :115.3   Mean   :116
##   2018-01-16:  1   3rd Qu.:135.9   3rd Qu.:136.7   3rd Qu.:134.2   3rd Qu.:136
##   2018-01-17:  1   Max.   :159.4   Max.   :160.8   Max.   :158.3   Max.   :161
##   (Other)   :497
##    Adj.Close        Volume
##   Min.   : 82.3   Min.   :8.99e+06
##   1st Qu.: 99.5   1st Qu.:2.03e+07
##   Median :108.9   Median :2.54e+07
##   Mean   :114.9   Mean   :2.81e+07
##   3rd Qu.:135.0   3rd Qu.:3.22e+07
##   Max.   :160.6   Max.   :1.11e+08
##
```

For excel file with .xls or .xlsx extension, we can use the function read_excel from the readxl package.

```r
# install.packages('readxl')
# load the 'readxl' package
library(readxl)

df <- read_excel('MSFT.xlsx')
head(df)
```

```
## # A tibble: 6 x 7
##   Date                 Open  High   Low Close `Adj Close`   Volume
##   <dttm>              <dbl> <dbl> <dbl> <dbl>       <dbl>    <dbl>
## 1 2018-01-09 00:00:00  88.7  88.7  87.9  88.2        85.5 19484300
## 2 2018-01-10 00:00:00  87.9  88.2  87.4  87.8        85.1 18652200
## 3 2018-01-11 00:00:00  88.1  88.1  87.2  88.1        85.3 17808900
## 4 2018-01-12 00:00:00  88.7  89.8  88.4  89.6        86.8 24271500
## 5 2018-01-16 00:00:00  90.1  90.8  88.0  88.3        85.6 36599700
## 6 2018-01-17 00:00:00  89.1  90.3  88.8  90.1        87.3 25621200
```

Variables

```r
open <- df$Open
close <- df$Close
class(open)
```

```
## [1] "numeric"
```

```r
head(open)
```

```
## [1] 88.65 87.86 88.13 88.67 90.10 89.08
```

```r
open[1:5] # Square brackets! 1 and 5 are indices
```

```
## [1] 88.65 87.86 88.13 88.67 90.10
```

for loop

```r
length(close)
```

```
## [1] 503
```

```r
for (t in 1:length(close)){
  df$diff[t] <- close[t] - open[t]
}
```

```
## Warning: Unknown or uninitialised column: `diff`.
```

```r
# or you can do
df$diff_new = close - open
head(df)
```

```
## # A tibble: 6 x 9
##   Date                 Open  High   Low Close `Adj Close` Volume     diff
##   <dttm>              <dbl> <dbl> <dbl> <dbl>       <dbl>  <dbl>    <dbl>
## 1 2018-01-09 00:00:00  88.7  88.7  87.9  88.2        85.5 1.95e7 -0.430
## 2 2018-01-10 00:00:00  87.9  88.2  87.4  87.8        85.1 1.87e7 -0.0400
## 3 2018-01-11 00:00:00  88.1  88.1  87.2  88.1        85.3 1.78e7 -0.0500
## 4 2018-01-12 00:00:00  88.7  89.8  88.4  89.6        86.8 2.43e7  0.930
## 5 2018-01-16 00:00:00  90.1  90.8  88.0  88.3        85.6 3.66e7 -1.75
## 6 2018-01-17 00:00:00  89.1  90.3  88.8  90.1        87.3 2.56e7  1.06
## # ... with 1 more variable: diff_new <dbl>
```
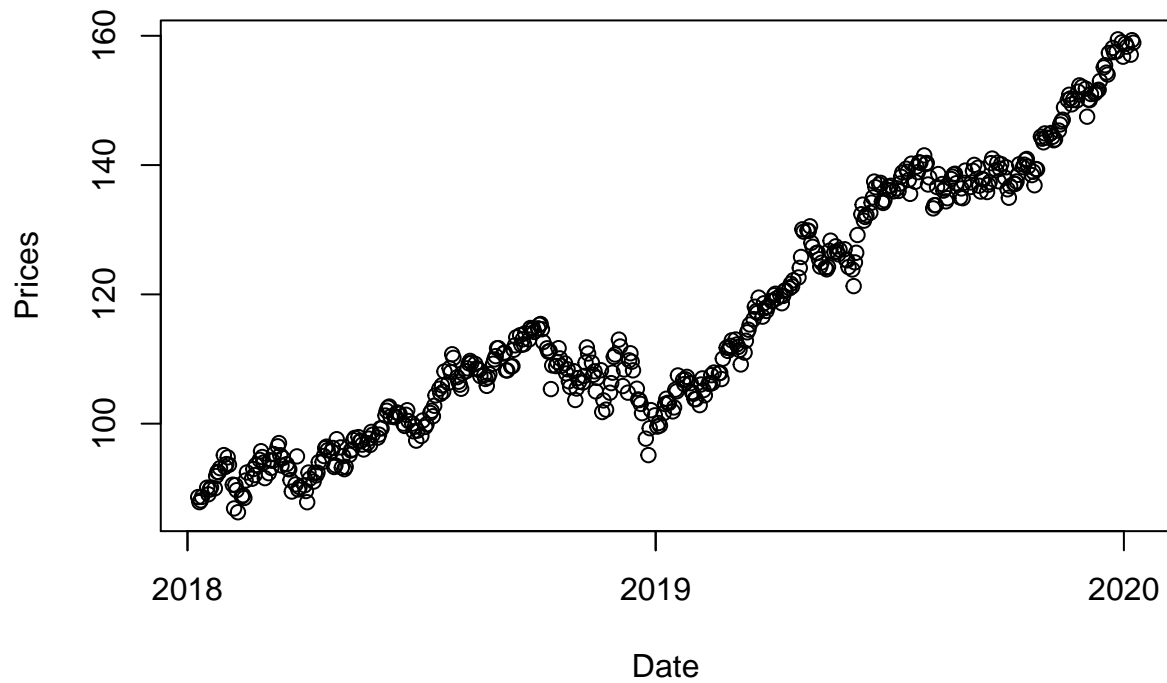
Including Plots

You can also embed plots, for example:

Opening MSFT Prices

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

(We can use ggplot2 to create better figures.)