Data Science

Transforming Data into Knowledge and Vision

Introduction to R

William Yu, PhD

Economist

UCLA Anderson Forecast

Correlation vs causality

- The stock market slump might correlate/predict recession
- The stock market slump might cause recession
  - Only when there is a wealth effect
- It is the best to know the true causality. But it is not necessary in order to predict well.

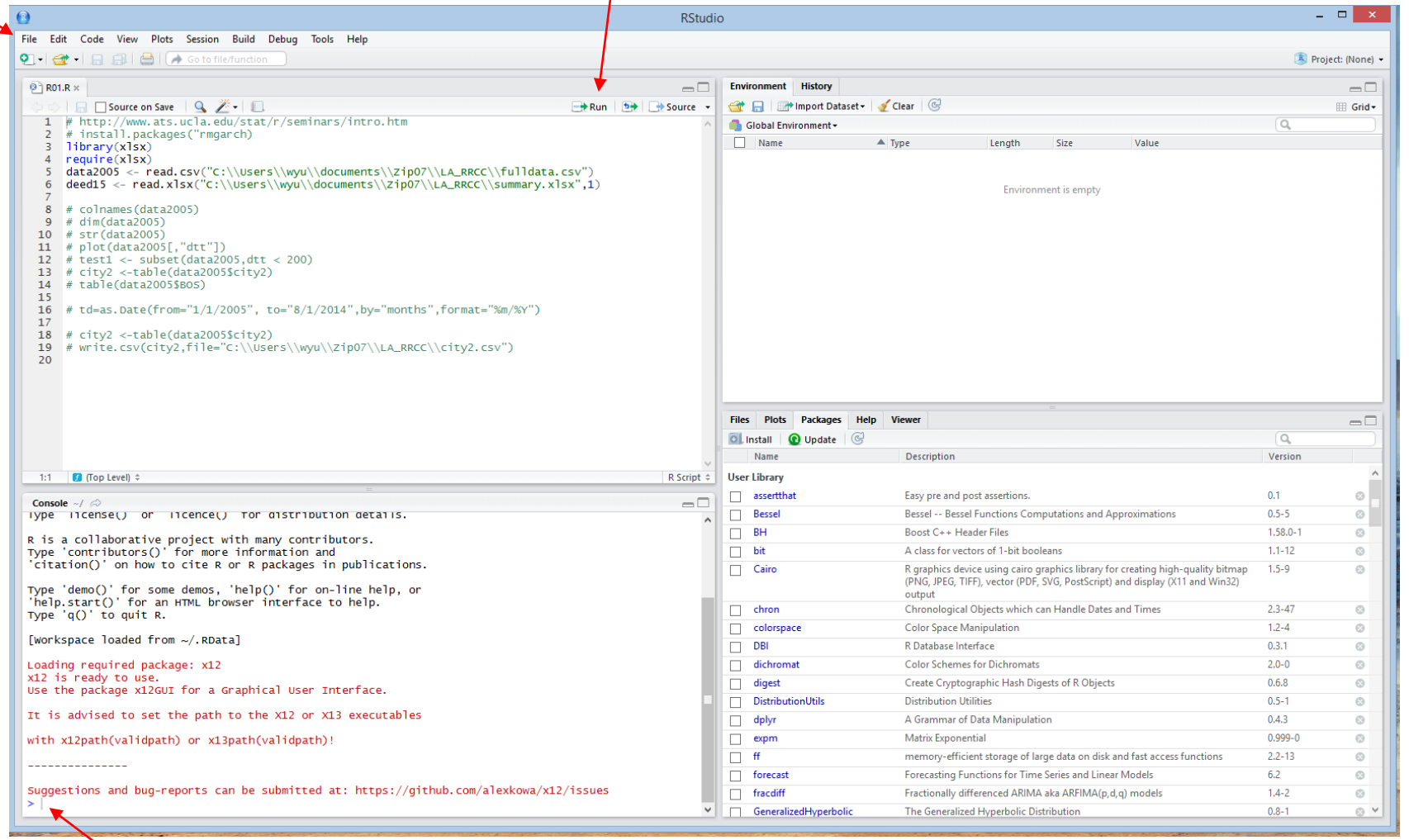
We still can use Excel

- Install Analysis ToolPak into your Excel Spreadsheet.
- In Excel, click FILE and then click Options → Add-ins → In Manage box, select Excel Add-ins and click Go → In the Add-ins available box, select the Analysis ToolPak and then click OK
- Click DATA on the toolbar, you should see the Data Analysis on the toolbar.

RStudio

Open the script:
Open File
Write a new script:
New File→ R Script

Highlight and click and run the program!



Program/
Script

Environment,
History

Console
/Command

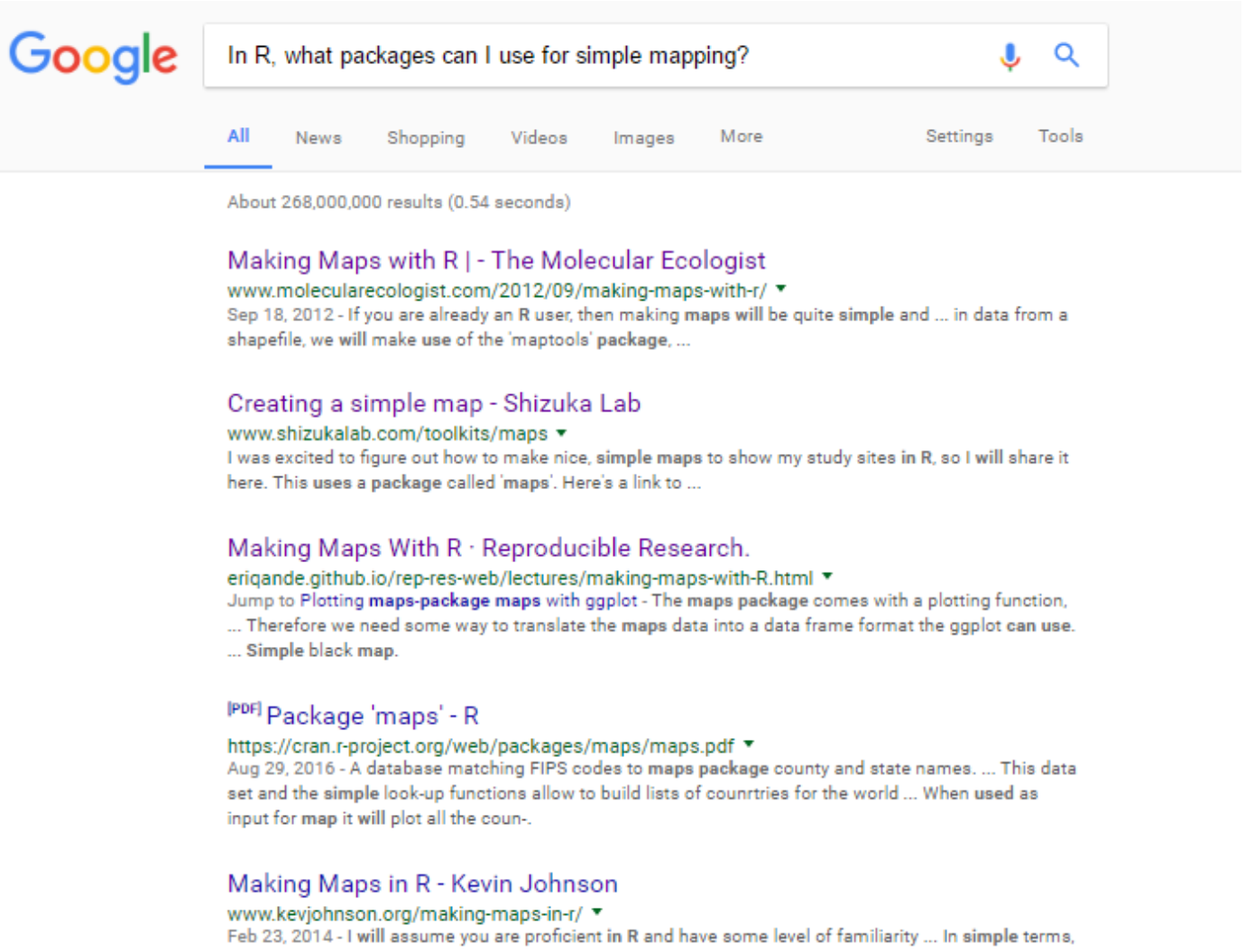
Files,
Plots,
Packages,
Helps,
Viewer

Type your command here

Introduction and data control in R

- Online help
  - help(mean), ?sd, ?rnorm
- Basic operation
  - +, -, *, /,^, exp(), log(), log10(), sqrt()
  - Assignment and create new variables, =, <-
  - Comment operator:
    - everything after # until the end of the line is a comment
  - Legal names of objects
    - Mostly used: "."
    - No number at the first letter: "3angeles" → Wrong!
    - Names are **Case Sensitive**
    - **Data cannot contain comma:** 1600 is Good while 1,600 is Wrong!
  - Construction of vector and matrix
    - > x=c(1,3,7,16)
    - > x
    - [1] 1 3 7 16
    - > x^2
    - [1] 1 9 49 256
    - > y = c(3:12)
    - > y
    - [1] 3 4 5 6 7 8 9 10 11 12
- Matrix
  - > y1 = matrix(0,4,5) # a 4 (rows) by 5 (columns) matrix of zeros
  - > y2 =matrix(1:10, nrow=2) # a 2 by 5 matrix
  - > y3 =matrix(1:10, ncol=5) # the same as y2
  - > y4 = y2+y3
  - > y5 = t(y2) # the transpose of y2, which is 5 by 2 matrix
  - > dim(y1) # the size /dimensions of y1
  - > nrow(y1) # the number of rows of y1
  - > ncol(y2) # the number of columns of y2
  - > length(c(1,4, 0, 7,9)) # the length of a vector
- Matrix element
  - > y2[2,3] # the second row and the third column of y2
  - > y2[2,] # the second row of y2
  - > y2[,5] # the fifth column of y2
  - > y2[,3:5]
- Logical value: TRUE(T) and FALSE (F)
  - Logical operators: >, <. >=, <=, !=
  - > x1 = 1:5
  - > x1 <3
- Previous command: ↑
- Generate random number with standardized normal distribution
  - > x2=rnorm(1000)
  - > plot(x2)
  - > hist(x2) # a simple histogram plot (distribution)
- **Data frame**: a matrix with names above the columns
- **Time Series**: a data frame with time date embedded
- Packages: you need to install packages into your computer only for the first time
- Library: you need to run library every time you open the script

Google Search: In R, how can I do @#$^&*?



How to reduce health care spending?

-- A case on Head Start program 2016

Q09: When your child is sick, where do you First go to get help?				
	Pre N	(%)	Post N	(%)
call family or friends	85	3.85	32	1.71
doctor or phone-line	880	39.91	126	6.72
look in book	51	2.31	1054	56.18
no response	27	1.22	4	0.21
take to ER	822	37.28	4	0.21
treat at home	340	15.42	656	34.97
Total	2205	100.00	1876	100.00