Programming for Data Analytics

DANL100 (04)

Spring 2022 (CRN: 60148 , January 26, 2022 - May 25, 2022)

Instructor:	Paul Yan
Contact Information:	Email: pyan@geneseo.edu Phone: (585) 245-5260 Office: South Hall 223C
Lecture Hours/Location:	MW: 2:30pm - 3:45pm @ Newton Lecture Hall Room: 209
Teaching method	Face-to-face
Office Hours	MW: 10:00am-12:00pm via Zoom or in my office
Prerequisites:	None
Textbook:	My lecture notes
Websites:	http://canvas.geneseo.edu (for syllabus, HW, list of term projects, etc.) http://datayyy.com/rpy/ (another location for syllabus, slides, videos, etc.)
QR codes	After a few weeks, I will show how to generate the above images (just 3 lines).
Course Description:	We have entered a big-data era. Thus, it is quite important for a business school student to learn how to process, relatively speaking, big data. Among many open-source statistical software, R and Python are the top two. In this course, students learn both R and Python. For the first half semester, students learn how to install R, define a variable, write simple functions, run a loop to process hundred or thousand data sets, and more. After understanding basic concepts and functionalities, R packages will be discussed. Then students learn Python. In terms of data, students learn how to download, and process public data associated with economics, finance and accounting, such as FRED (Federal Reserve Economic Data), UCI Machine Learning Data Depository, SEC quarterly index files, SEC Financial Statement Data Sets, and French's Data Library.
School of Business Mission	Students acquire strong quantitative, analytical, and communication skills while preparing for professional success as socially responsible individuals in today's complex business environment. We strive for teaching excellence and recognize that high-quality faculty scholarship and professional activities advance the impact we have on practice, pedagogy, and our knowledge of business.
Bachelor of Science Program Learning Goals	Competency Goal 1: Our learners will have strong analytical skills Competency Goal 2: Our learners will have strong quantitative skills Competency Goal 3: Our learners will have effective communications skills Competency Goal 4: Our learners will have a thorough understanding of various functional areas of business Competency Goal 5: Our learners will have a multidimensional understanding of social responsibility.

Objectives Course Learning Objectives	After successfully completing this course, students are expected to demonstrate their ability to: • Understand the principles of probability. (LG1) • Understand the properties of distributions. (LG1, LG2) • Apply statistical concepts to many business applications. (LG4) • Collect, organize, describe data and make statistical inference. (LG3). • Understand the concept of confidence interval and use it to make inference about the data. (LG1, LG2)
Computational Tools	R and Python are two languages used for this course. To learn how to install R, watch the following videos. Windows users: https://www.youtube.com/watch?v=ZoPJGmpYJzw Mac users: https://www.youtube.com/watch?v=WJDrYUqNrHg
Data	For this course, students will learn how to retrieve data from many public/open sources, such as UCI Machine Learning Data Depositary, Yahoo!Finance, SEC corporate filing, Federal Research Economics Data Library (FRED) , and Census Data.
Teaching method	Hands-on is critical. I will not go through lecture notes mechanically. Instead, we will do a few in-class exercises during each lecture, kind of a flipped classroom environment. https://study.com/teach/flipped-classroom.html

One-line R command for this course	`source("http://datayyy.com/rpy.txt")  # or`
	`source("http://geneseo.edu/~pyan/rpy.txt")`
	Note that I will explain the above two lines during the first lecture.

Extra credit (optional)	Before the final exam, I will post something about this.
Academic Integrity:	https://www.geneseo.edu/dean_office/dishonesty
Attendance Policy	If you are feeling unwell physically and/or mentally, please do not attend in-person classes. You will not be penalized for non-attendance. The reason is that in-person courses present the highest risk of contagion and members of the college community who may be impacted if ill students attend these classes.
Grade Evaluation:	Homework 30% Midterm (for R software) 30% Final exam 30% Class participation 5% Office hours (I will explain this) 5% ---------- ---------- Total 100%
Course Schedule:	For the detailed schedule, see below. I reserve the right to change the course schedule throughout the semester. Changes to the schedule will be announced in class or via email.
Academic calendar	https://www.geneseo.edu/registrar/academic-calendar

From a percentage grade to a letter grade

Percentage grade	Letter grade
$grade \geq 90\ \%$	A
$85\% \leq grade < 90\%$	A-
$80\% \leq grade < 85\%$	B+
$75\% \leq grade < 80\%$	B
$70\% \leq grade < 75\%$	B-
$65\% \leq grade < 70\%$	C+
$60\% \leq grade < 65\%$	C
$55\% \leq grade < 60\%$	C-
$grade < 55\%$	F

Course Schedule: Academic calendar Spring 2022: https://www.geneseo.edu/registrar/academic-calendar

Weekly Course Schedule (tentative and subject to change)

#	Date	Topics	Description (F for Finance)	Data case
1	1/26	Introduction	A short survey, self-intro, syllabus discussion Chapter 1: What is data? What are data analytics? R Installation, basics and value assignments assignment, basic math functions: mean(),min(),max(), median(), sd(), as a scientific calculator	
2	1/31 2/2	writing functions	Chapter 2: Writing simple functions, one-line functions, multi-lines, add help, double_f(), pv_f(), call functions Chapter 3: Open data: UCI, Yahoo!Finance, French's Data Lib etc	Watching a video
3	2/7 2/9	Data sources Data input	Chapter 4: data input, input data from external files, read.csv(), read.table(), readLines(), etc Chapter 5: Simple data manipulation, choose certain row, columns, seq(), merge(), sort() functions, sort data	HW #1
4	2/14 **2/16**	Simple Data manipulation Data output	Chapter 6: data output: csv (comma separated value), txt (text), R data sets (.RData, .rds), write.csv(), write.table(), save(), save.image(), sink() functions **Diversity Summit - No classes**	HW #2
5	2/21 2/23	R loops, if else if Plots/Graphs	Chapter 7: R loops, if else, if else if, for() and while() loop, logic or (|) and logic and (&) Chapter 8: Simple plots and graphs	HW #3
6	2/28 3/2	Subset, etc. Matrix	Chapter 9: Date variable, subset, combine data sets, and merge subset() function, more on merge [optional] Chapter 10: Matrix manipulation, scalar, vector and matrix , is.vector(), is.matrix(), as.matrix(), matrix() Chapter 11: Simple string manipulation, grep(), sub(), gsub()	HW #4
7	3/7 3/9	Data manipulation	Chapter 12: Introduction to R packages library(), require(), help(package=lmtest) Before mid-term review Optional: Chapter 13: Zip,SAS,Pickle,Google drive Pre-midTerm review	HW #5
8	**3/14** **3/16**		**March 14-18/Mon-Fri/Spring Break - No classes**	

#	Date	Topics	Description	Topic
9	**3/21** 3/23	Second part: Python	**midterm** Chapter 14: Python basics	HW #6 (Python)
10	3/28 3/30	5 modules	Chapter 14: (continued) Chapter 15: Introduction to Python modules	HW #7 (Python)
11	4/4 4/6	Data input	Chapter 15: (continued) Chapter 16: Data input	HW #8 (Python)
12	4/11 4/13	Data output, Loops, conditions	Chapter 17: Data output Chapter 18: Python loops, conditions	HW #9 (Python)
13	4/18 4/20	Data manipulation Plot, graphs	Chapter 19: Data manipulation Chapter 20: Graphs, plots and visualization	HW #10 (Python)
14	4/25 4/27	String with Python	Chapter 21: String manipulation Chapter 22: Projects type I and II (optional)	HW #11?
15	5/2 5/4	Reviews	Review I: comparison between R and Python Review II: comparison between R and Python	
16	5/9 5/11	Review	Review II: comparison between R and Python (backup day)	
	5/18	**Final**	`> .exam("mw2:30")` DATE TIME COURSE 1 Monday.May.16 12:00-2:30pm(3)12:00-3:20pm(4) MWF12:30-1:20 2 Monday.May.16 12:00-2:30pm(3)12:00-3:20pm(4) MW/MF/WF12:30-2:10 (4) 3 Wednesday.May.18 3:30-6:00pm(3)3:30-6:50pm(4) MW2:30-3:45 4 Wednesday.May.18 3:30-6:00pm(3)3:30-6:50pm(4) MW2:30-4:10(4) (see the schedule shown below) http://datayyy.com/webs/final2022spring.html	

References

Academic calendar: https://www.geneseo.edu/registrar/academic-calendar

Quandl, financial/accounting/economics data platform, https://www.quandl.com/

UCI machine learning data depositary. https://archive.ics.uci.edu/ml/index.php

Yan, Yuxing, 2018, Financial Modeling using R, *Lagaia Books*, Amazon link at http://datayyy.com/webs/amazonR2018.html

Yan, Yuxing, 2018, A Trend in Business Education, *International Journal of Education and Social Science*, 5(75), http://www.ijessnet.com/uploades/volumes/1575190910.pdf

Yan, Yuxing, 2017, Python for Finance (2nd edition), *Packt Publishing*, Amazon link at *http://datayyy.com/webs/amazonP4F2.html*

Yan, Yuxing, 2017, Teaching programming skills to finance students: how to design and teach a great course, *Financial Innovation*, https://link.springer.com/article/10.1186/s40854-017-0081-x

Appendix A: After launching R, type the following command.

```
source("http://datayyy.com/rpy.txt")
```

The following menu will pop up.

```
*-------------------------------------------------------*
* Programming for Data Analytics        2022 by Yan    *
*-------------------------------------------------------*
* .c1  R installation,basics, value assignment          *
* .c2  Writing simple R functions                       *
*-------------------------------------------------------*
* >.c1       # go to chapter 1 (a dot in front of c1)  *
* >.uu       # go to the utility menu                   *
* >.rpy      # back to this main menu                   *
*-------------------------------------------------------*
```

Every function or menu is self-explanary. Typing its name would give you an introduction plus a few examples.

Appendix B: After typing `.c1`, (note that there is a dot in front of c1), will we see the following instructions.

```
> .c1
function(i){
" i  Chapter 1: R basics
  -  -------------------------------------
  0  Survey results
  1  Download and install R
  2  Launch/quit R, one line for this course
  3  Clear console, 2 types of comments
  4  3 ways to assign a value/values & how to show its value
  5  Use up and down arrow keys to recall the previous commands
  6  R is case sensitive
  7  Normal operations: +, -, /, ^ (power)
  8  listing function ls()
  9  remove a variable or several variables
 10  remove all variables
 11  use meaningful variable names
 12  current working directory
 13  head() and tail() functions
 14  mean() for calculating mean and sd() for standard deviation
 15  put several commands on one line
 16  Using .nLetterFunctions() to show all n-letter functions
 17  help(), and example() functionss
 18  File transfer betweem your computer and the virtual lab
 19  YouTube
 20  Links

Example #1:>.c1    # see the above list
Example #2:>.c1(1) # see the first explanation
```

Appendix C: Type .uu (a dot infront of uu), you will see the following menu.

```
> .uu
function(){
"
 *-----------------------------------------*
 * Utilities          -- short-cut --      *
 *-----------------------------------------*
 * .allChapters       # .all               *
 * .calendar          # .cal               *
 * .macUsers          # .mac               *
 * .virtualLab        # .vl # letter l or 1 *
 * .inClassEx         # .ice               *
 *-----------------------------------------*
 * >.ice              # see a list of ice   *
 * >.uu               # back to utilites    *
 * >.rpy              # back to main menu   *
 *-----------------------------------------*
"
```

Appendix D: To see all chapters covered, type `.all` or `.all()`, shown below (subject to chagne)

```
> .all()

 *-------------------------------------------------------------------------*
 * Programming for Data Analytics                    2022 by Yan           *
 *-------------------------------------------------------------------------*
 *    R (1st half semester)              Python (2nd half semester)        *
 *-------------------------------------------------------------------------*
 * .c1  R basics                       .c14 Python basics                  *
 * .c2  Simple R functions             .c15 Python modules                 *
 * .c3  Open data                      .c16 Python data input              *
 * .c4  Data input                     .c17 Python data output             *
 * .c5  Simple data manipulation       .c18 Python loops, cond             *
 * .c6  Data output                    .c19 Data manipulation              *
 * .c7  Loops,if else, logic or, and   .c20 Graphs and plots               *
 * .c8  Plot, graph and moving GIF     .c21 Python string manipulations    *
 * .c9  Date var, data.frame and list   --- below is optional ---------    *
 * .c10 Matrix manipulation            .c22 Project types I and II         *
 * .c11 String manipulation            .c23 Project type III               *
 * .c12 Introduction to R packages     .c24 Term Projects                  *
 * .c13 Zip,SAS,Pickle,Google Drive                                        *
 *-------------------------------------------------------------------------*
>
```