




Programming for Data Analytics

DANL100 (01)

Spring 2021 (CRN:59996, Feb 1, 2021 – May 20, 2021)

Instructor:	Paul Yan
Contact Information:	Email: pyan@geneseo.edu Phone: (585) 245-5260
Lecture Hours/Location:	TR: 1:00pm - 2:45pm @ South Hall 338
Online learning	http://datayyy.com/rpy/ Note: students could take this course online.
Office Hours	TR: 3:00pm-4:30pm via Zoom
Prerequisites:	Math 213 or 221, or equivalent courses
Textbook:	My lecture notes
Websites:	http://canvas.geneseo.edu (for syllabus, HW, list of term projects, etc.) http://datayyy.com/rp/ (another location for syllabus, slides, videos, etc.)
QR codes	 <p>After a few weeks, I will show how to generate the above images (just 3 lines).</p>
Course Description:	<p>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.</p>
School of Business Mission	<p>Students acquire strong quantitative, analytical, and communication skills while preparing for professional success 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.</p>
Bachelor of Science Program Learning Goals	<p>LG1: Our students will have strong analytical skills LG2: Our students will have strong quantitative skills LG3: Our students will have effective communications skills LG4: Our students will have a thorough understanding of various functional areas of business</p>
Objectives Course Learning Objectives	<p>After successfully completing this course, students are expected to demonstrate their ability to:</p> <ul style="list-style-type: none"> Understand the principles of probability. (LG1)

	<ul style="list-style-type: none"> • 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 Tool	R is our tool. To learn how to install, 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 Depository, 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 slides one after another. Instead, we will do a few in-class exercises, kind of a flipped classroom environment. https://study.com/teach/flipped-classroom.html														
One-line R command for this course	<code>source("http://datayyy.com/abc.txt") # or</code> <code>source("http://geneseo.edu/~pyan/abc.txt")</code> Note that I will explain the above two lines during the first lecture.														
Extra oral exam (optional)	For extra credit. This will be conducted via zoom and it is optional.														
Academic Integrity:	https://www.geneseo.edu/dean_office/dishonesty														
Attendance Policy (I)	Attending classes regularly is required. Before-class preparation and in-class participation is an integral part of this course. Students are strongly encouraged to participate in class discussions and ask questions. Students are encouraged to discuss current events relevant to this course or their own experiences. Homework problems are regularly assigned.														
Attendance Policy (II)	For some reasons if you cannot attend classes regularly, make sure that you read the related chapters, lecture notes (slides), videos and finish assignments.														
Grade Evaluation:	<table> <tr> <td>HW (for R software)</td> <td>20%</td> </tr> <tr> <td>Midterm (for R software)</td> <td>25%</td> </tr> <tr> <td>HW (for Python software)</td> <td>20%</td> </tr> <tr> <td>Final exam (80% Python 20%R)</td> <td>25%</td> </tr> <tr> <td>Class participation (including zoom office-hour)</td> <td>10%</td> </tr> <tr> <td>-----</td> <td>-----</td> </tr> <tr> <td>Total</td> <td>100%</td> </tr> </table>	HW (for R software)	20%	Midterm (for R software)	25%	HW (for Python software)	20%	Final exam (80% Python 20%R)	25%	Class participation (including zoom office-hour)	10%	-----	-----	Total	100%
HW (for R software)	20%														
Midterm (for R software)	25%														
HW (for Python software)	20%														
Final exam (80% Python 20%R)	25%														
Class participation (including zoom office-hour)	10%														
-----	-----														
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 2021: <https://www.geneseo.edu/registrar/academic-calendar>

Weekly Course Schedule (tentative and subject to change)

#	Date	Topics	Description (F for Finance)	Data case
1	dd/mm dd/mm	Introduction writing functions	A short survey, self-intro, syllabus discussion Chapter 1: R Installation, basics and value assignments assignment, basic math functions: mean(),min(),max(), median(), sd(), as a scientific calculator Chapter 2: Writing simple functions, one-line functions, multi- lines, add help, double_f(), pv_f(), call functions	Watching a video
2	dd/mm dd/mm	Data sources Data input	Chapter 3: Open data: UCI, Yahoo!Finance, French's Data Lib etc Chapter 4: data input, input data from external files, read.csv(), read.table(), readLines(), etc	HW #1
3	dd/mm dd/mm	Simple Data manipulation Data output	Chapter 5: Simple data manipulation, choose certain row, columns, seq(), merge(), sort() functions, sort data Chapter 6: data output: csv (comma separated value), txt (text), R data sets (.RData, .rds), write.csv(), write.table(), save(), save.image(), sink() functions	HW #2
4	dd/mm dd/mm	R loops, if else if, Date var	Chapter 7: R loops, if else, if else if, for() and while() loop, logic or () and logic and (&) Chapter 8: Date variable, simple plots and graphs	HW #3
5	dd/mm dd/mm	Subset, etc. Matrix	Chapter 9: subset, combine data sets, and merge subset() function, more on merge Chapter 10: Matrix manipulation, scalar, vector and matrix , is.vector(), is.matrix(), as.matrix(), matrix()	HW #4
6	dd/mm dd/mm	Data frame List String manipulation	Chapter 11: Data frame and list data.frame(), colnames(), rowname() Chapter 12: Simple string manipulation grep(), sub(), gsub(), .*, () vs. []	HW #5
7	dd/mm dd/mm	Introduction to R packages	Chapter 13: Introduction to R packages library(), require(), help(package=lmtest) Chapter 14: Monte Carlo Simulations Chapter 15: Reading SAS, pickle, binary and zip files	HW #6
8	dd/mm dd/mm		Before mid-term review Midterm	

#	Date	Topics	Description	Topic
9	dd/mm dd/mm	Second part: Python	Introduction to Python Chapter 16: Python basics Chapter 17: Python programs, using Python as an calculator	
10	dd/mm dd/mm		Chapter 18: Introduction to Python modules Chapter 19: Introduction to NumPy and SciPy modules np.npv(),np.pv(),np.irr()	HW #7 (Python)
11	dd/mm dd/mm		Chapter 20: Data input/output, financial calculator Chapter 21: Data manipulation, pandas module, pd.merge(), pd.pivot_table()	HW #8 (Python)
12	dd/mm dd/mm		Chapter 22: Python loops, Chapter 23: Statistical Analysis of Time Series	HW #9 (Python)
13	dd/mm dd/mm		Chapter 24: Monte Carlo Simulations Chapter 25: Matrix manipulation	HW #10 (Python)
14	dd/mm dd/mm		Chapter 26: String manipulation Chapter 27: Linear regressions	HW #11 (Python)
15	dd/mm dd/mm		Chapter 28: Various distributions and hypothesis tests Chapter 29: Optimization	HW #12 (Python)
	dd/mm	Final	TBA	

References (most of them are for term projects)

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

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

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127-1139.

UCI machine learning data depository. <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/uploads/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>

Yan, Yuxing, 2015, Red vs. Blue Stocks: Politics and Profitability of Firms, *Journal of Business and Policy Research* 10 (1), 117-138.

Yan, Yuxing, An internet connected financial calculator, 2012, *Journal of Accounting and Finance* 12(5), 59-70.

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

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

The following menu will pop up.

```
*-----*
* Programming for Data Analytics      2021 by Yan *
*-----*
* .c1  R installation,basics, value assignment *
* .c2  Writing simple R functons          *
*-----*
* >.c1      # go to chapter 1 (a dot in front of c1) *
* >.uu      # go to the utility menu          *
* >.ba      # back to this 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`), we see the following instructions.

```
> .c1
function(i){
" i Chapter 1: R basics
- -----
 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() functions
18 alarm() and finding out all available demos
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 in front of uu), you will see the following menu.

```
*-----*
* Utilities          -- short-cut --    *
*-----*
* .allChapters      # .all              *
* .calendar          # .cal              *
* .videos            # .v                *
* .inClassEx        # .ice              *
*-----*
* >.ice              # see a list of ice  *
* >.uu               # back to utilites  *
* >.ba               # back to main menu *
*-----*
```

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

```
*-----*
* Programming for Data Analytics          2021 by Yan    *
*-----*
*      R (1st half semester)      |      Python (2nd half semester)      *
*-----*
* .c1 R basics                     .c16 Python basics                    *
* .c2 Simple functions             .c17 Python as an calculator        *
* .c3 Open data                    .c18 Functions: financial calculator *
* .c4 Data input                   .c19 Intro 2 Python modules         *
* .c5 Simple data manipulation     .c20 Intro 2 NumPy and SciPy modules *
* .c6 Data output                  .c21 Data input/output              *
* .c7 Loops,if-else                .c22 Python loops, conditions       *
* .c8 Date var, simple plot        .c23 Data manipulation: pandas      *
* .c9 subset,merge data sets       .c24 Visualization: matplotlib module *
* .c10 Matrix operation            .c25 Matrix operation               *
* .c11 data frame and list         .c26 String manipulation            *
* .c12 String manipulation         .c27 Monte Carlo simuations         *
* .c13 Intro to R packages         .c28 Distributions/hypothesis tests *
* .c14 R package: 5 examples       .c29 Regresson and optimization     *
* .c15 SAS,binary,zippped data     .c30 Term projects                  *
*-----*
```