Data 210G Spring 2018 Syllabus

Course:	DATA 210G: Elements of Data Science
Instructor:	Mary Rudis
Office:	**Adjunct Faculty office 2nd floor**
Office hours:	Tuesday and Thursday by appointment Before and after class
GBCC Course Number:	**CRN 23480**
Textbooks:	https://leanpub.com/rprogramming **"R Programming for Data Science**" by Roger D. Peng published by Leanpub Please get the book + Datasets + R Code Files **"R For Data Science"** by Hadley Wickham & Garrett Grolemund, published by O'Reilly Media
Other Tools: Resources	**(see last page – Resources and Course Links)**
Required Software:	R and RStudio – in THAT order! https://www.r-project.org/ https://www.rstudio.com/products/rstudio/download/ - download OR https://gbcc-rstudio.org/ ← login credentials needed
Semester:	Spring 2018 January 16 – May 4
Days/Time/Place:	online – via Blackboard LMS
Contact Email:	mrudis@ccsnh.edu

Prerequisite: MATH145G/147G or MATH150G/152G or higher. Some prior programming experience is helpful but not required.

Catalog Description: This course is the foundation for introducing students to key topics in data science, including data acquisition/preparation and exploratory data analysis. Major topics include an introduction to the R programming language and RStudio integrated development environment, working with modern data formats (e.g. XML, CSV, JSON, XLS, XHTML), data import/export (e.g. files, APIs – application programming interfaces – , web sites, databases), finding data to augment analyses, and exploratory data analysis & visualization.

Desired Student Competencies: Upon successful completion of the course, the student will be able to:

1. Use the core data structures of R including vectors, lists and data frames

2. Write and organize analysis scripts that utilize the functional programming nature of R and vectorization model unique to R
3. Work with all modern data formats, including XML, CSV, JSON, XLS (Excel), XHTML (web pages), and understand how to appropriately transform this data for use in structured analysis projects
4. Use web APIs, including Census & Bureau of Labor Statistics data sets and database APIs to import data for structured analysis projects
5. Explain the difference between discrete & continuous data and the basic operations that can be applied to each type
6. Understand the concept of "tidy" data and demonstrate how to transform data for use in structured analysis projects
7. Apply the "split-apply-combine" analysis pipeline paradigm to any data analysis problem
8. Search for authoritative data sets to help augment their analysis projects
9. Visualize data for use in exploratory data analysis as a pre-cursor to statistical analysis of data sets
10. Effectively communicate preliminary results of data analysis projects

Required Materials:

- **R and RStudio (or RStudio Server):**
  There will be Lab activities for this course that cannot be completed without using R. It is assumed all are novice users of R and all code will be given to the student.

Assessment:

Grading:	Participation:	10%
	Quizzes:	25%
	Course Projects (5):	25%
	Final Project:	40%

Participation (10%): This is an online course that has an interactive component to it. Participation is done via weekly virtual office hours – LIVE sessions accessed through Blackboard Collaborate, AND participation in the discussion board/forum also in Blackboard. It is VITAL that students work together to solve programming problems, issues especially when trying to determine how to fix errors in code.

Quizzes (25%): Quizzes will be on Blackboard. It is vital that you have a reliable internet connection. Quizzes not submitted by the deadline date will be recorded as zero. Students are expected to work quiz problems entirely on their own without help from other students. Book/notes and other reference material may be consulted as needed.

Projects (25%): Programming in R takes practice and more practice. Typically these will be assigned early on during a new topic and students will have approximately 2 weeks to complete each one. Students may collaborate on these, as long as proper citation/credit is given.

Final Project (40%): The course final project will be made available in late February. It will be designed so that students will need to use all knowledge about the structures and forms of data, how to create a tidy data set when the originating data is not, performing

some exploratory data analysis and munging with the data to get it ready for the next phase (model building and validation).

Course Grades: please check your grades in your MML Gradebook often!!!!

College Grade	MML Gradebook Numeric
A	93.33 to 100
A-	90.00 to 93.32
B+	86.67 to 89.99
B	83.33 to 86.66
B-	80.00 to 83.32
C+	76.67 to 79.99
***C**	**73.33 to 76.66 *minimum grade to advance**
C-	70.00 to 73.32
D+	66.67 to 69.99
D	63.33 to 66.66
D-	60.00 to 63.32
F	below 60

Technology:

This course is part of the Certificate in Practical Data Science. We will be using a statistical programming tool called "**R**" and **RStudio** which are free and available online OR you can download and install it on your own personal computer or laptop. Instructions for accessing the online tool will be emailed to each student separately.

Procedures:

Log in to Blackboard: the platform that will be used for quizzes for this course. You are responsible for accessing all course assignments via our LMS Blackboard. You are expected to log into Blackboard, your college email, and use RStudio several times each week.

If you are having difficulty meeting the demands of this course for any reason, please contact the instructor to schedule time to discuss the matter. Often alternative arrangements can be made.
If you are having any problems with Blackboard or have other computer related questions or problems please contact the student help desk at 603-427-7621 or http://www.greatbay.edu/helpdesk

College Email:

I send all class announcements to you at your college address via BB. You are responsible to check your college email regularly for any and all communication sent to you. Email addresses can be found at the Student Information (SIS) site. **Any correspondence not using college email is subject to quarantine and/or removal from the CCSNH email server and will not be answered.** If there are technical questions that need to be answered, please access the helpdesk. The address is http://greatbay.edu/helpdesk. Personal email addresses cannot be used because SIS and Blackboard are not set up to accommodate them. Please note: when you register in MML please use your **college email address** so we can communicate!! I check email regularly and will do my best to reply to your email within 24 hours. I only reply to email from your college address.

Attendance/Participation Policy:
Participation is extremely important! It is your responsibility to keep up with online discussion board topics, be ACTIVE in this area each week, and to check in with the instructor during weekly LIVE sessions scheduled via Collaborate. Students do not need to be present for the entire LIVE session. It can be limited to 15 minutes if no further help is needed. But attendance during this time is mandatory!

Registrar Procedures: Add/Drop Deadlines and Procedures:
Please see the Planning Calendar (at end of this document) for deadlines with the registrar: deadline to drop with full refund, deadline to drop with a "W", and deadline to drop with "WP/WF".

In order to drop this class you must fill out an add/drop form and return this form to the One Stop office for processing. ***Simply ceasing to attend class or contacting the instructor does not constitute officially dropping the course***.

You may withdraw from the course and receive the grade of W, which does not impact your GPA (see Planning Calendar on the last page for this deadline date). No signature from the instructor is required. If past this deadline, the next option is to drop with a "WP/WF". This option does require an instructor signature and can affect GPA. **If you are receiving Financial Aid, you should speak with that office before deciding to withdraw**.

Classroom/Discussion Board/Online Decorum:
In order to create a healthy and productive learning environment there are expectations of all of us. These include:

- Civility in behavior and respect for the rights and opinions of others seeking to learn is highly valued. Please be courteous and respectful at all times. When disagreeing with another's point of view, no belittling, condemnation or put-downs will be tolerated; keep all comments constructive and helpful in nature.
- It is expected that students will be respectful of GBCC staff and GBCC students. GBCC has a no bullying policy. The **Student Code of Conduct** can be found in the **Student Handbook**. The handbook can be found on the College website. Students are responsible for being familiar with the Student Code of Conduct.
- It is expected that we will behave in a manner that is ethically consistent with college policy and in a manner that could not be interpreted as cheating or plagiarism.
- The college defines Plagiarism as, "the intentional or unintentional failure to immediately, accurately, and completely cite and document the source of any language, ideas, summaries, hypotheses, conclusions, interpretations, speculations, graphs, charts, pictures, etc., or other material not entirely your own. This includes failure to cite work of your own that you have used previously."
- Cheating is defined as using or attempting to use unauthorized materials, information, or study aids in any academic exercise or activity. Failure to meet this expectation could result in a grade of F for the course.
- Consequences for acts of plagiarism and/or cheating result in a minimum of 0 for that assignment/quiz/test and maximum F for the course, depending on the severity of the infringement of the policy.

- I welcome and encourage your input intended to improve the quality of your educational experience.

Cancellations:
If an emergency arises, I will post a cancellation announcement of a LIVE session to BB and reschedule.

Assistance:
Sometimes additional assistance is needed. I encourage you to help each other when possible. I will be available for questions via email. There are also "R Programming" forums where help with coding is provided.
https://stackoverflow.com/questions/tagged/r can be a programmer's best friend. It is your responsibility to choose from and utilize available assistance if you need it.

Student Support and CAPS:
The College is committed to providing support for students with disabilities. Any student with physical, learning, attention, and/or psychological disabilities is encouraged to visit the Center for Academic Planning and Support (CAPS) and make an appointment with the Coordinator of Disability Support Services. More information is available at http://greatbay.edu/student-services/academic-support/disability-services

The **Center for Academic Planning and Support (CAPS**) has a variety of services to support your academic success at Great Bay Community College. These include:

- Math drop-in assistance from a GBCC math instructor (no appointment needed). Drop-in times at posted in CAPS and on CAPS website. http://greatbay.edu/student-services/tutoring <- click on campus tutoring schedule
- Free tutoring support on an individual 1-1 basis. For more information about tutoring times, access, availability contact Tutor Coordinator, **Chelsea Jackman**, at 603-427-7623 or email: cjackman@ccsnh.edu. Schedules are posted at CAPS and updated periodically during each semester at www.greatbay.edu/CAPS
- Study Groups organization and tutor facilitation
- CAPS open computer lab with various course software.
- Individual academic counseling for setting goals, organization, time management and finding strategies to increase learning and performance
- Assistive technology
- ESOL support, nontraditional/gender equity and single parent support
- Career exploration
- Academic workshops – times posted each semester.
- Locations are: 210 at Portsmouth campus and 115 at Rochester campus

Check with CAPS to see all that is available for support for this class.

Success:
Success in Data Science is achieved through practice. Students learn at different rates and come into this course with different levels of mastery of the course objectives. On average, ***11(eleven) hours per week*** is a good estimate of time spent on coding, viewing and taking notes from lecture videos and study of reading material in order to be

successful. Ingredients for **success** include excellent participation, a positive attitude, full use of available educational materials in class, completing and understanding homework and quiz assignments before deadlines, seeking assistance when needed, using the built-in HELP features of RStudio when needed, and a willingness to devote the time necessary to master the material.

Planning Calendar – Spring 2018

Data 210	Elements of Data Science	Textbook Readings and Topics
T-R		
Jan 16		Peng: Chapters 1-4, pages 1-12
Jan 18	Quiz #1 available	
Jan 22	Data Types and Why it Matters	Peng: Chapter 5, pages 13-23; Wickham: Chapter 2
	Saving Your Work	
Mon Jan 29	**DEADLINE TO DROP w/ 100% REFUND**	
Jan 30	Saving Work and Storing Data	Peng: Chapters 6, 8; Wickham: Chapter 4
Feb 1	Project One Due!	**Peng: Chapter 17** is needed when writing your first actual program
Fri, Feb 2	**DEADLINE TO ADD w/ instructor permission**	
Feb 6	Project Data and Outside World	Peng: Chapter 9; Wickham: Chapter 6
Feb 8		
Feb 13	Reading and Parsing Data	Peng: Chapter 7; Wickham: Chapter 8
Feb 15	Project Two Due!	
Mon Feb 19	**President's Day: CLOSED**	
Feb 20	Subsetting and Vectorized Operations	Peng: Chapters 10, 11
Feb 22		
Feb 27	Dates and Times	Peng: Chapter 12; Wickham p 134-137
Mar 1		
Mar 6	Plotting with ggplot2	Wickham: p. 1-21 of Chapter 1
Mar 8	Project Three Due!	
Mar 12-16	**SPRING BREAK: CLOSED**	
Mar 20		
Mar 22	Managing data frames (dplyr)	Peng: Chapter 13
Mon Mar 26	**DEADLINE TO DROP w/ a "W"**	
Mar 27	Control Structures in R	Peng: Chapter 14
Mar 29	Project Four Due!	
Apr 3	Writing Your First Function	Peng: Chapter 15
Apr 5		
Apr 10	Looping and the "apply" functions	Peng: Chapter 18
Apr 12		
Apr 17	Regular Expressions and the "grep" function	Peng: Chapter 19
Apr 19	Project Five Due!	
Fri Apr 20	**DEADLINE TO DROP w/ a "WP/WF"**	
Apr 24	Spread, Gather, Separate, Pull to create a Tidy Dataset	Wickham: Chapter 9
Apr 26		
May 1	May Day!	
May 3	**FINAL PROJECT**	DUE NO later than Thursday May 3 at 11:59 PM

DATA 210G Elements of Data Science (with R)

Primary Reading Material with Exercises

"R Programming for Data Science" by Roger D. Peng, 12/22/2016, published by leanpub.com, https://leanpub.com/rprogramming

"R for Data Science: Import, Tidy, Transform, Visualize and Model Data", 1st edition, by Wickham and Grolemund, published by O'Reilly media https://www.amazon.com/Data-Science-Transform-Visualize-Model/dp/1491910399

Supporting Videos and Lectures

R Programming via Johns Hopkins University

Some Helpful Examples

RMarkdown – Submitting Projects For Class

RMarkdown – Submitting Projects For Class (Another Example)

Plotting with ggplot2 part one

Plotting with ggplot2 part two

Another Intro to ggplot2 in R

Tidyr Package and Tidy Data

Data For Use With Videos and Tutorials (If not already in RStudio)

MAACS data set

US County Data Set (already tidy)