

Master of Science in Sustainability Management

SUMA PS5255 Data Analysis & Visualization in Sustainability

Tuesdays, 6:10-8:00 PM

3 credits

Instructors:

Greg Yetman, Senior Staff Associate in the Center for International Earth Science Information Network, email: gyetman@ciesin.columbia.edu, (845) 365- 8980

Valentina Mara, Senior Staff Associate in the Center for International Earth Science Information Network, email: vmara@ciesin.columbia.edu, (845) 365-8904

Office Hours:

TBD

Response Policy:

Questions regarding the assignments will be posted and responded to on the discussion board. Students can expect a response within 24 hours, 7 days a week.

For all other questions, email is our preferred mode of communication.

Course Overview

Data science is an exciting new field of applied research that takes advantage of the ever-growing volume of data being collected to support of decision-making in both the public and private sectors. Similar to traditional statistical analysis, these new approaches have limits and issues that are important to understand before application to problem solving. This course aims to introduce the common methods used in data science, best practices in data management, and the basic scripting skills required to start analyzing data in R and Python. After introducing foundational scripting and data analysis methods, a case study approach will be used to highlight both what can be accomplished with data analysis and the limits of the data and methods used. Lab exercises will teach basic skills in scripting in Python and R and then move to a common approach for data analysis: adapting existing scripts and software libraries to solve applied data problems.

The requirement to understand the interaction of social and natural systems requires data-driven policy decisions, and the ongoing assessment of policies requires rigorous, reproducible assessments of effectiveness for promoting sustainability. Both requirements can be met in part by data science approaches that are applicable to the natural and social sciences and reproducible in academic and applied settings. Data science techniques have been developed to derive insight from large volumes of available data that are often collected for purposes other than the interests of the data scientist. This flexibility in approach means that the techniques used in data science are well adapted to support gaining insights from data relevant for sustainability science. This course has been designed to introduce these techniques in anticipation of increased use in promoting sustainability.

The course has no prerequisites; however, an understanding of statistics and probability will be very useful background, and any previous programming or scripting skills will be applicable to the lab assignments.

This course satisfies the M.S. in Sustainability Management program's Area 2: Quantitative Analysis requirement.

Learning Objectives

This course will provide an introduction to the main tools and methods of data analysis and visualization used in sustainability management. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that are used in data analysis: Python programming language and R, along with a selection of the numerous analysis libraries available for both tools.

By the end of this course, students will:

L1: Be able to interpret data and write data analysis reports using the best practices of data analysis in the fields of sustainability science and sustainability management

L2: Create new and adapt existing R and Python scripts for data acquisition and analysis

L3: Access primary data via Application Programming Interfaces (APIs) and via popular libraries (e.g., numerical Python, Random Forest) from the extensive collection available for Python and R

L4: Identify and apply appropriate predictive models for sustainability data sets

Master of Science in Sustainability Management

L5: Apply industry standard big data processing techniques, including neural networks and the Random Forest model, to address complex sustainability problems

Readings

- Abu Sharkh, M. and I. Gough (2010). Global Welfare Regimes. *Global Social Policy* 10(1): 27-58.
CADDIS – Analyzing data @ http://cfpub.epa.gov/caddis/analytical_tools.cfm
- Barnthouse, L. 2008. The strengths of the ecological risk assessment process: linking science to decision making. *Integrated Environmental Assessment and Management*. 4(3):299-305
- Cutter, S. L., Boruff, B. J. and Shirley, W. L. (2003), Social Vulnerability to Environmental Hazards*. *Social Science Quarterly*, 84: 242–261. doi:10.1111/1540-6237.8402002
- Dilley, M., R. Chen, U. Deichmann, A. Lerner-Lam and M. Arnold. 2005. Natural disaster hotspots: a global risk analysis. Synthesis report. Center for Hazards and Risk Research, Columbia University.
<http://sedac.ciesin.columbia.edu/hazards/hotspots/synthesisreport.pdf>
- Esty, D., M. Levy, T. Srebotnjak, A. de Sherbinin. 2005. 2005 Environmental Sustainability Index. New Haven: Yale Center for Environmental Law & Policy, and Palisades NY: Center for International Earth Science Information Network (CIESIN), Columbia University. (408 pages) <http://www.yale.edu/esi/>
- Gelman, A. and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. New York, Cambridge University Press.
- Grus, Joel. 2015. *Data Science from Scratch: First Principles with Python 1st Edition*, O'Reilly Media. ISBN 978-1491901-42-7
- Helsel, D.R. and R. M. Hirsch, 2002. *Statistical Methods in Water Resources. Techniques of Water Resources Investigations*, Book 4, chapter A3. U.S. Geological Survey. 522 pages.
- Hosmer, D. and S. Lemeshow. 2000. *Applied logistic regression*. 2nd. ed. Wiley.
- Hsu, A., D. Esty, M. Levy, and A. de Sherbinin. 2016. *The 2016 Environmental Performance Index*. New Haven, CT: Yale Center for Environmental Law and Policy.
- Rygel L, O'Sullivan D. and Yarnal, Brent 2006 A Method for Constructing a Social Vulnerability Index: An Application to Hurricane Storm Surges in a Developed Country Mitigation and Adaptation Strategies for Global Change. 11: 741–764
C Springer DOI: 10.1007/s11027-006-0265-6
- Venables, W. N., D. M. Smith and the R Core Team .2012. *An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics, Version 2.15.2 (2012-10-26)*”, ISBN 3-900051-12-7

Resources

Students will be expected to work on their own laptops (Mac, Windows, or Linux) using open source software components from the Conda package management system for Python (<https://conda.io/docs/>) and the open source RStudio development environment for R (<https://www.rstudio.com/products/rstudio/>).

Columbia University Library

Columbia's extensive library system ranks in the top five academic libraries in the nation, with many of its services and resources available online: <http://library.columbia.edu/>.

SPS Academic Resources

The Office of Student Affairs provides students with academic counseling and support services such as online tutoring and career coaching: <http://sps.columbia.edu/student-life-and-alumni-relations/academic-resources>.

Course Requirements (Assignments)

Master of Science in Sustainability Management

Students will be expected to work on their own laptops (Mac, Windows, or Linux) using open source software components from the Conda package management system for Python (<https://conda.io/docs/>) and the open source RStudio development environment for R (<https://www.rstudio.com/products/rstudio/>).

Getting Started with Python and R: (L1, L2)

This is a demonstration and ungraded exercise. This set of demonstrations and in-class student exercises will expose students to the Python and R development environments. The requirements and procedures for installing the software will also be covered.

In this assignment, students will:

1. Start the interactive command prompt and issue basic commands
2. Start an Integrated Development Environment, author, save and run a simple script
3. Generate basic output techniques (print statements, print formatting)
4. Use software libraries and import and use them to perform simple calculations

Individual Lab Assignments (L1, L2, L3, L4)

Lab 1: Introduction to Python. Data types, syntax, functions. This lab will require the students to learn Python syntax so that they may create and run scripts, and perform simple debugging. The first part of the lab assignment will include step-by-step instructions that illustrate the use of variables of different types and Python Syntax. The second part of the assignment will require students to adapt an existing set of scripts that exhibit functional programming to implement new functions that meet the requirements of a specification for data processing. Sample input and output for testing the functions will be provided.

In this assignment, students will:

1. Differentiate between dynamically typed languages such as Python and static languages (e.g., Java).
2. Identify the basic and common data types in Python and their uses, including:
 - a. Strings, for display and qualitative data
 - b. Numeric data (integer, float), for statistical and mathematical operations
 - c. Tuples, for storing and manipulating data pairs
 - d. Lists, for storing strings or numeric data in a convenient structure for manipulation (combining lists and other data types, splitting)
 - e. Dictionaries, for lookup purposes
 - f. Sets, for fast access to unique values
3. Use basic Python syntax for constructing scripts, including:
 - a. Whitespace, case and how Python interprets them
 - b. The `__main__` method and order of flow
 - c. Conventions for variables, functions, classes
 - d. List comprehension
 - e. Sorting data types
4. Use and construct functions in Python by adapting simple examples—a power function and a Fibonacci sequence function—to create new functions that meet a requirement

Lab 2: R/ Exploratory data analysis

In this assignment, students will:

Master of Science in Sustainability Management

1. Set up the work: Introduction to R and R Studio windows.
2. Create short script. Write script preliminaries: clean the workspace, set up working directory
3. Read, write and save data in R
4. Create descriptive statistics
5. Generate graphs for continuous and categorical data

Lab 3: Data Input/Output, Data Parsing, Accessing data via Web APIs. This lab will provide instructions for writing a script to read in and parse sample text data that matches a search pattern and write a subset of the data out to a new file. A template script for accessing the Twitter API will be provided and students will need to adapt the script to access real-time data from the Twitter stream and analysis the data using a simple sentiment analysis, outputting the results to a new data file.

In this assignment, students will:

1. Read and parse text files (CSV data files, HTML documents)
2. Identify different options for handling incomplete or poorly formatted data during parsing
3. Apply technics for basic pattern matching (glob, string comparison)
4. Apply techniques for outputting data to text files such as CSV and Excel spreadsheets
5. Adapt an existing script for accessing a web API such as Twitter to acquire and read live data to accomplish the following:
 - i. Update authentication information to use proper credentials
 - ii. Modify criteria for selecting Tweets to process and store
 - iii. Using API documentation, modify elements of selected Tweets that are stored/processed
 - iv. Summarize the results of a simple sentiment analysis

Lab 4: R/ Cluster analysis with randomly generated data & Cluster analysis literature review

In this assignment, students will:

1. Test relationship between categorical and continuous variables
2. Conduct cluster analysis using two methods: hierarchical and k means
3. Visually represent cluster analysis

Lab 5: Data acquisition and exploration II. This lab will require students to derive statistics and create graphics of the data suitable for inclusion in a report. The students will have to show the ability to interpret the data by answering questions based on the statistics and graphics. Part of the interpretation will be the usefulness of the data for statistical inference. Students will select different sets of U.S. census data to use in the exercise based on guidelines provided in the lab.

In this assignment, students will:

1. Perform exploratory data analysis through the use of Python libraries for statistical analysis using sample data.
Characterize data by deriving descriptive measures, including:
 - a. Basic parametric statistics (min, mean, max, standard deviation)
 - b. Interquartile range, skewness and kurtosis
 - c. Central tendency (spatial measure from a collection of 2D points)
2. Generate charts and plots appropriate for the data and statistical measures, including
 - a. Bar charts
 - b. Scatter plots
 - c. Line graphs
 - d. 2D map plot

Master of Science in Sustainability Management

Lab 6: R/ Principal Component Analysis

In this assignment, students will:

1. Perform Principal Component Analysis
2. Create Social Vulnerability Index using the PCA
 - a. Collect data from Census
 - b. Standardize variables
 - c. Clean data based on descriptive statistics
 - d. Normalize the input variables by calculating z-scores (mean =0, standard deviation =1)
 - e. Calculate Principal Component Analysis (PCA), and retaining the most important components
 - f. Examine the correlations between the input variables and retained components, to determine what characteristic are being represented in each dimension
 - g. Examine the direction of components, so that the higher positive values represent high vulnerability
 - h. Generate scores for each retained dimensions
 - i. Calculate Social Vulnerability Index by summing the most important dimensions' scores
 - j. Calculate Weighted Social Vulnerability Index by multiplying the eigenvalues
 - k. Categorize distribution into deciles, and rank the block groups
 - l. Map the index

Python Lab 7: Data Analysis, Calculating natural hazard risk. This lab will require to students to calculate probabilities of recurrence of natural disasters using sample data on past events, including measures of confidence for the prediction. A sample data set, such as earthquakes greater than a certain magnitude or precipitation anomalies based on a standardized precipitation index, will be provided for use in determining the probability of a threshold being exceeded in a given time period. Using the probability surface derived in the first part of the lab with exposure and vulnerability data provided, the students will develop a relative risk measure for different regions.

In this assignment, students will:

1. Develop probabilities based on past events
2. Assess exposure and vulnerability from publicly available data
3. Normalize different data measures (exposure and vulnerability) to dimensionless index values for comparison
4. Calculate relative risk for different regions (comparing risk across populations)

Lab 8: R/ Disasters and Developments

In this assignment, students will:

1. Perform Ordinary Least Squares (OLS) Linear Regression
2. Review of linear regression with R, and interpret the regression output
3. Interpret coefficient of categorical independent variables
4. Interpret coefficients as function of variable transformation
5. Perform Regression diagnostic and interpret the results

Python Lab 9: Data Analysis II, predicting the distribution of human population. This lab will require the students to use existing data on population and its spatial covariates (elevation, land cover, water bodies, etc.) to calculate a prediction surface of human population distribution using the Random Forests model. The lab submission will include a plot of the surface and its standard deviation, along with answers to questions that will guide the student towards interpreting the variations in standard deviation. The second part of the lab will use a convolutional neural network and supplied sample data to extract features

Master of Science in Sustainability Management

(buildings) from remotely sensed imagery. A sample script implementing a neural network will be provided with suggestions on how to adapt it to the imagery and training data provided with the lab. The students will be submit the modified code, buildings identified from the imagery, and accuracy measures of the classification.

In this assignment, students will:

1. Implement the Random Forests statistical mode
2. Interpret the output of the Random Forests model, including measures of variability in the prediction
3. Gain a basic understanding of convolutional neural networks for extracting features from imagery
4. Run a convolution neural network using the Python Multiprocessing module (parallel data processing)
5. Perform an accuracy assessment of classified data based on training sets (precision/recall and overall accuracy)

Lab 10: R/ 10 Overview of logistic regression

In this assignment, students will:

1. Review logistic regression with R, and interpret the regression output
2. Transform odds ratios to probabilities
3. Perform Regression
4. Perform regression diagnostic and interpretation of the results

Midterm Examination (L1, L2, L3, L4)

The midterm exam will consist of an in-class written lab assignment with additional short answer essay questions. Prompts will require students to demonstrate an understanding of analysis concepts, and use R and Python software. Students will be required to:

Python:

1. Identify basic and common data types in Python and their uses
2. Use basic Python syntax for constructing scripts,
3. Parse text files (CSV data files, HTML documents)
4. Perform data analysis through the use of Python libraries for statistical analysis using sample data
5. Generate charts and plots appropriate for the data and statistical measures

R:

6. Create short R script
7. Perform cluster analysis
8. Construct visual analysis of cluster analysis

Final Examination (L4, L5)

The final exam will consist of practical applied exercises that will assess the ability to use the theory and software taught throughout the course to solve environmental data analysis problems. Students will be given a The final exam will be a take-home test distributed at the conclusion of the final class session on April 24, 2018. The final exam is due in Canvas by 11:59 PM on May 1, 2018.

Examples of the typed of applied exercises students will be expected to complete in the final exam include:

Master of Science in Sustainability Management

- Calculate probabilities of recurrence of natural disasters using sample data on past events, including measures of confidence for the prediction.
- Use existing data on population and its spatial covariates (elevation, land cover, water bodies, etc.) to calculate a prediction surface of human population distribution using the Random Forests model

Evaluation/Grading

Final course grades will be assigned by the lecturer. The grade break-down is as follows:

- Attendance and Participation – 10% of final grade
- Practical Lab Assignments – 40% of final grade
- Midterm Exam – 20%
- Final Exam – 30%

The details for evaluation and grading of each of these assignments:

Participation

Attendance alone does not count towards your participation grade. Class participation is important and you may also gain participation credit by participating on the discussion forum where you can reflect on the assigned readings each week. Participation will be graded on a scale of 0-100.

Practical Lab Assignments

There will be an accompanying hands-on based laboratory section for each of our analysis topics. The labs are submitted electronically via Canvas. If multiple files are to be submitted, create a separate folder with the number of the lab you are submitted the work for. Each lab will be weighted equally for the final course grade. Each lab assignment will be graded on a scale of 0-100.

Midterm Exam

The midterm exam will consist an in-class written exam with short answers to demonstrate your understanding of analysis concepts, and basic exercises to demonstrate your understanding of R and Python software. The midterm exam will be graded on a scale of 0-100.

Final Exam

The final exam will consist of practical applied exercises that will assess your ability to use the theory and software taught throughout the course to solve environmental data analysis problems. The final exam will be a take-home test distributed at the conclusion of the final class session on April 24, 2018. The final exam is due in Canvas by 11:59 PM on May 1, 2018. The final exam will be graded on a scale of 0-100.

FINAL GRADING SCALE

Grade	Percentage
A+	98–100 % (for Extraordinary work)
A	93–97.9 %
A-	90–92.9 %
B+	87–89.9 %
B	83–86.9 %
B-	80–82.9 %
C+	77–79.9 %

Master of Science in Sustainability Management

C	73–76.9 %
C-	70–72.9 %
D	60–69.9 %
F	59.9% and below

ASSIGNMENT	% Weight
Participation	10%
Practical Lab Assignments	40%
Midterm Exam	20%
Final Exam	30%

Course Policies

Participation and Attendance

You are expected to complete all assigned readings, attend all class sessions, and engage with others in online discussions. Please arrive for class on time, attend all classes, and to stay until the end of class. If you need to miss a class or leave early for any reason, please discuss the absence with the instructors.

Late work

Work that is not submitted on the due date noted in the course syllabus without advance notice and permission from the instructor will be graded down 10% of its point value for each day it is late (e.g., from a maximum of 10 points to a maximum of 9 for work that is one day late).

Citation & Submission

All labs should be submitted via Canvas (not via email). Any data sources used beyond those provided should be cited by method of access (URL/API name) and date/time the data were retrieved.

School Policies

Copyright Policy

Please note—Due to copyright restrictions, online access to this material is limited to instructors and students currently registered for this course. Please be advised that by clicking the link to the electronic materials in this course, you have read and accept the following:

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

Academic Integrity

Columbia University expects its students to act with honesty and propriety at all times and to respect the rights of others. It is fundamental University policy that academic dishonesty in any guise or personal conduct of any sort that disrupts the life of the University or denigrates or endangers members of the University community is unacceptable and will be dealt with severely. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources, and assistance upon which that work is based. Academic honesty in class assignments and exams is expected of all students at all times.

Master of Science in Sustainability Management

SPS holds each member of its community responsible for understanding and abiding by the SPS Academic Integrity and Community Standards posted at <http://sps.columbia.edu/student-life-and-alumni-relations/academic-integrity-and-community-standards>. You are required to read these standards within the first few days of class. Ignorance of the School's policy concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

Accessibility

Columbia is committed to providing equal access to qualified students with documented disabilities. A student's disability status and reasonable accommodations are individually determined based upon disability documentation and related information gathered through the intake process. For more information regarding this service, please visit the University's Health Services website: <http://health.columbia.edu/services/ods/support>.

Course Schedule/Course Calendar

Class # and Date	Lecture Topics	Readings (Completed by)	Assignment (Due on this Day)
1 – 01/16/2018	Introduction and overview to data analysis and visualization in sustainability - Class Outline - Case studies	<ul style="list-style-type: none"> Grus, Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7. Preface, pp i-xvi Chapter 1, Introduction, pp 1-13 	
2 – 01/23/2018	Introduction to Python - Data types - Syntax - Methods/functional programming	<ul style="list-style-type: none"> Grus, Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7. Chapter 2, a Crash Course in Python, pp 15-36 	Getting started with Python and R
3 – 01/30/2018	Introduction and Exploratory data analysis (EDA) using R - Data summaries - Crosstabulation - Graphical display	<ul style="list-style-type: none"> Venables, W. N., D. M. Smith and the R Core Team .2012. An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics, Version 2.15.2 (2012-10-26)", ISBN 3-900051-12-7 Chapters 8, 9, 12, pp 49-60, 89-106 Helsel, D.R. and R. M. Hirsch, 2002. Statistical Methods in Water Resources. Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey.. Chapter 1 & 2, pp 1-60. CADDIS – Analyzing data @ http://cfpub.epa.gov/caddis/analytical_tools.cfm Barnthouse, L. 2008. The strengths of the ecological risk assessment process: linking science to decision making. Integrated Environmental Assessment and Management. 	Lab 1: Introduction to Python

Master of Science in Sustainability Management

		4(3):299-305	
4- 02/06/2018	Python: Data acquisition and exploration - File input/output - Web APIs: tapping the Twitter stream	<ul style="list-style-type: none"> Grus , Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7. Chapter 3, Visualizing Data, pp 37-47, Chapter 9, Getting Data, pp 103-120 	Lab 2: Exploratory data analysis
5 – 02/13/2018	Multivariate analysis using R I - Introduction and basic concepts - k-means cluster analysis - hierarchical cluster analysis	<ul style="list-style-type: none"> Abu Sharkh, M. and I. Gough (2010). Global Welfare Regimes. <i>Global Social Policy</i> 10(1): 27-58. 	Lab 3: Data Input/Output, Data Parsing, Accessing data via Web APIs.
6 – 02/20/2018	Python: Data acquisition and exploration II - Visualizing data - Exploratory data analysis: characterizing populations with census and survey data	<ul style="list-style-type: none"> Grus , Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7, Chapter 5, Statistics, pp 48-58, Chapter 10, Working with data, pp 121-140 	Lab 4: Cluster analysis
7 - 02/27/2018	Multivariate Analysis using R II - Principal Component Analysis - Calculation of Social Vulnerability	<ul style="list-style-type: none"> Cutter, S. L., Boruff, B. J. and Shirley, W. L. (2003), Social Vulnerability to Environmental Hazards*. <i>Social Science Quarterly</i>, 84: 242–261. doi:10.1111/1540-6237.8402002 Rygel L, O’Sullivan D. and Yarnal, Brent 2006 A Method for Constructing a Social Vulnerability Index: An Application to Hurricane Storm Surges in a Developed Country Mitigation and Adaptation Strategies for Global Change. 11: 741–764 C Springer DOI: 10.1007/s11027-006-0265-6 	Lab 5: Data acquisition and exploration II
8 - 03/06/2018	MIDTERM EXAM		
03/13/2018	SPRING BREAK		
9 - 03/20/2018	Python: Data Analysis, Calculating natural hazard risk - Developing probabilities based on past events - Assessing exposure and simple vulnerability factors with public	<ul style="list-style-type: none"> Grus , Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7. Chapter 6, Probability, pp 69-80. Chapter 7, Hypothesis and Inference, pp 81-92 	Lab 6: Introduction to Principal Component Analysis

Master of Science in Sustainability Management

	data - Calculating risk and estimating disaster impacts		
10 - 03/27/2018	Guest lecture – developing environmental sustainability indicators (Marc Levy or Alex De Sherbinin)	<ul style="list-style-type: none"> Hsu, A., D. Esty, M. Levy, and A. de Sherbinin. 2016. The 2016 Environmental Performance Index. New Haven, CT: Yale Center for Environmental Law and Policy. Esty, D., M. Levy, T. Srebotnjak, A. de Sherbinin. 2005. 2005 Environmental Sustainability Index. New Haven: Yale Center for Environmental Law & Policy, and Palisades NY: Center for International Earth Science Information Network (CIESIN), Columbia University. (408 pages) http://www.yale.edu/esi/ 	Lab 7: Data Analysis, Calculating natural hazard risk
11 - 04/03/2018	Regression techniques using R - OLS - Interpretation of regression outputs - Regression diagnostic	<ul style="list-style-type: none"> Gelman, A. and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. New York, Cambridge University Press. Chapter 3 & 4, pp 31-74 	
12 - 04/10/2018	Python: Data Analysis II , Predicting the distribution of human population - Prediction of population based on covariate data with Random Forests - Using computer vision techniques to map settlements	<ul style="list-style-type: none"> Grus , Joel. 2015. <i>Data Science from Scratch: First Principles with Python</i> 1st Edition, O'Reilly Media. ISBN 978-1491901-42-7. Chapter 11, Machine Learning, pp 141-150, Chapter 17, Decision Trees 201-212 	Lab 8: Disasters and Development
13 - 04/17/2017	Logistic Regression with R - Odds ratio and probabilities - Interpretation of regression outputs - Regression diagnostic - Interaction effects	<ul style="list-style-type: none"> Hosmer, D. and S. Lemeshow. 2000. Applied logistic regression. 2nd. ed. Wiley. Chapters 1-3, pp 1-88. Gelman, A. and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. New York, Cambridge University Press. Chapter 5, pp 79-105 	Lab 9: Data Analysis II, predicting the distribution of human population.
14 – 04/24/2017	Course Conclusion -Final Exam Distributed		Lab 10: Logistic regression
05/01/2018	Final Exam		Final Exam due at 11:59 PM

Master of Science in Sustainability Management

--	--	--	--