W

EVANS SCHOOL
OF PUBLIC POLICY & GOVERNANCE

UNIVERSITY *of* WASHINGTON

PUBPOL 542-A: COMPUTATIONAL THINKING FOR GOVERNANCE ANALYTICS

Prof. José Manuel Magallanes, PhD.

Winter 2022

E-mail: **magajm@uw.edu**

Office Hours: By appointment

Website: **https://canvas.uw.edu**

Class Hours: 8:30-11:20 (Thursdays)

Class Room: Parrington Hall 220

Course Description

This course is an innovative combination of Data Science and Computational Social Science particularly structured for policy studies. It introduces algorithmic thinking to policy students to solve problems related to data pre-processing, optimization via linear programming, and the modeling of governance complexity via agent-based simulation.

The participants will also develop good practices in reproducibility for professional work in policy research. Every session of the course follows a hands-on approach (laptops needed). Tools to be taught and used include:

- Programming languages (Python/R/Netlogo).

- Version control apps (GitHub).

- Document preparation tools (markdown)

- Advanced spreadsheet functionality.

- Computational representation of governance complexity.

There is no pre-requisite for the course.

Course Objectives

1. Carry-out analytic work following a reproducibility/replicability approach.

2. Master basic data management and analysis principles.

3. Become an effective user of programming languages (without becoming a programmer!).

4. Be a reflective user of data science in complex governance settings.

5. Use computational tools to aid complex decision-making.

Course Book

Magallanes Reyes, J. M. (2017). *Introduction to Data Science for Social and Policy Research: Collecting and Organizing Data with R and Python.* Cambridge University Press, 1 edition edition.

Recommended Readings

- Python Language Documentation: **https://www.python.org/doc/**

- R Manuals: **https://cran.r-project.org/manuals.html**

Software installations required

Students have to install the following software in their computers:

- ANACONDA by Continuum Analytics (choose according to your Operating System):
  https://www.anaconda.com/products/individual

- R (choose according to your Operating System):
  https://cran.r-project.org/

- RStudio Desktop Personal License (choose according to your Operating System)
  https://www.rstudio.com/products/rstudio/download/

- NetLogo (choose according to your Operating System)
  https://ccl.northwestern.edu/netlogo/download.shtml

- GITHUB:
  Get an account at **https://github.com/**,
  and then download the desktop app from **https://github.com/** .

- OpenSolver: If you have a gmail account, we will use it from there (no need to install it now).

Evans School Community Conversation Norms

This course has adopted the Evans School Community Conversation Norms. Please be aware of these norms in interactions with the instructor and other students. At the Evans School, we value the richness of our differences and how they can greatly enhance our conversations and learning. As a professional school, we also have a responsibility to communicate with each other–inside and outside of the classroom–in a manner consistent with conduct in today's increasingly diverse places of work. We hold ourselves individually and collectively responsible for our communication by:

- Listening carefully and respectfully

- Sharing and teaching each other generously

- Clarifying the intent and impact of our comments

- Giving and receiving feedback in a "relationship –building" manner

- Working together to expand our knowledge by using high standards for evidence and analysis

Changes to the Syllabus

The professor reserves the right to make changes to the syllabus during the quarter. The professor will notify students immediately by email and in class if any changes are made.

Grading Policy

Grades are based on three Deliverables:

- **Deliverable I**: Prepare data for analytical work using Python. Make sure you can find a way to integrate the tables:

  1. Collect some data tables. At least three data tables from different files or web pages. One has to be from a website. The other ones can be a file downloaded from a website, or use an API. (0.5 pt)
  2. Clean the data. (0.5 pt)
  3. Format and integrate the data tables. (0.5 pt)
  4. Upload the integrated file to Github, make sure the file can be read into R. (0.5 pt)

- **Deliverable II**: Using the previous file, use R to:

  1. Inform if the variables in the data file represent some latent concept. (0.5 pt)
  2. Inform how the units of analysis could be clustered. (0.5 pt)
  3. Test a hypothesis using regression. (0.5 pt)

- **Deliverable III**: Choose one these (0.5 pt):

1. Solving a optimization problem.

2. Making social network analysis.

3. Making a social simulation.

The remaining 0.5 points for the final grade will be based on how well the code is explained, using markdown (following a tutorial-like style).

Working in groups

The deliverables can be done individually or in pairs. You should have decided who will be in your group by the end of the **first** week of classes.

Course Schedule

Week 01, 01/03 - 01/07:   Data Pre Processing using Python: Data gathering

- Uploading files.

- Scraping tables from the web.

- Using APIs.

- Data from social apps.

- Preparing the Github repo.

Week 02, 01/10 - 01/14:   Data Pre Processing using Python: Data cleaning and Formatting

- Removing or cleaning cell values using regular expressions.

- Setting values in the right measurement scale.

Week 03, 01/17 - 01/21:   Data Pre Processing using Python: Integration

- Reshaping.

- Merging and appending different data sets.

- Saving files.

Week 04, 01/24 - 01/28:   Data Analytics in R: Dimensionality reduction

- Cluster analysis.

- Factor analysis.

Week 05, 01/31 - 02/04:   Data Modeling in R I

- Linear regression.

- Logistic regression.

Week 06, 02/07 - 02/11:   Optimization for Decision Making - I

- Introducing operations research.

- Modeling problems in linear programming.

Week 07, 02/14 - 02/18:   Optimization for Decision Making - II

- Using spreadsheet to solve linear programming problems.

Week 08, 02/21 - 02/25:   Working on your own: preparing deliverables I / II

Week 09, 02/28 - 03/04:   Modeling Governance Complexity I

- Network Analysis.

Week 10, 03/07 - 03/11:   Modeling Governance Complexity II

- Complexity and policy making.

- Agent based-modeling.

Week 11, 03/14 - 03/18:   Final week: submitting last deliverable (III).

Remote teaching protocol

- Remote teaching is always a possibility due to COVID or other situations. If that happened, further directions will be shared by the instructor, so no class is postponed.