Advanced Research Methods in Leadership and Public Affairs (mpa.6020)

Course Schedule	Thursday, 09:00 AM to 11:50 PM
Location	Computer Lab (Room #120), Building #22 @ The Ridges
Office Hours	By appointment or luck
Contact Information	Room #115, Building #21, @ The Ridges
	(740)-597-1949 ruhil@ohio.edu

Special Accommodations: Ohio University students with disabilities are assured equal access and full participation in the university's programs and services. As such if you have special needs please let me know on the first day of class so that we can make suitable arrangements for you to participate fully in this course.

Course Policy: Attendance is expected for all scheduled class meetings except in cases of "legitimate absences" as defined by the University (illness, death in the family, religious observances, jury duty, or University-sponsored activity). Absences that fall outside this purview do not qualify for make-up exams or homework assignments. The penalty for academic misconduct is a grade of F on any evaluated assignment affected by the misconduct.

Course Description

This course assumes a working knowledge of basic applied statistics and a strong desire to acquire valuable data skills. This course is designed to provide you with the technical skills needed to manage all aspects of real-world data acquisition, data munging, visualization, and analysis. We do all of this in an open-source environment, relying centrally on (R & RStudio). Every week is a workshop that introduces you to specific approaches for meeting specific goals — working with Census data, creating interactive visualizations, managing what seem to be very large and unwieldy data-sets, calculating and plotting quantities of interest, and much more.

Your final product will be a project you submit for a grade. As we work through the first few weeks, you will be gathering the data you intend to use for your project. You will be encouraged to practice everything we do in class with your own data-set, if relevant, so that your R skills get better every week. If you do not practice enough you will impede your grasp of the material and end up with few useful practical data skills. To ensure that you do practice I will give you very specific graded assignments that will require you to carry out specific operations.

Note: I will not cover any of the theory underlying the statistical approaches covered in MPA 6010. If you need to revisit that material review the textbook used for MPA 6010. If we have time I may walk you through how to conduct the basic statistical analysis with R.

Caution: This course is not for the faint of heart. You will have an opportunity to learn a lot about handling data but to master this skill be prepared to practice, to get frustrated as often as you feel elated. If you are looking for an easy course, then this is not the course for you. Look elsewhere. If you take pride in building something through toil and trouble, why, then, you are welcome to stay!

Learning Objectives

- Literacy in data acquisition, data cleaning, and general data management

- Literacy in building static and dynamic (including interactive) visualizations

- Demonstrated ability to utilize a statistical software package for building and managing a database and carrying out exploratory data analysis and/or basic data analytics

Required Texts & Other Materials

Required:

1. Course website. Here you will find the slides and other instructional material I will use in our weekly meetings.

2. Data Management and Visualization with R. This live text will give you a framework for using R and RStudio when working with data – gathering it, cleaning it, and manipulating it prior to carrying out some analyses, building visualizations and so on.

3. Data Analysis for Leadership and Public Affairs with R. This is a live statistics text-book with R code that shows you how to run statistical tests, fit a linear regression model, etc. I may touch upon some material from here just so you understand how to do most of what we did in MPA 6010 with R. There may be some chapters on advanced statistical techniques; feel free to work through these if you wish to.

Handouts & other materials: I will provide other material (either in a printed format or as links to pdf/html files) as needed. Readings may be added/deleted depending upon how the course progresses.

Software & Hardware: We will use R (with RStudio) for all analysis and graphics. You will be running everything in the Cloud unless you want to install the software on your own machine. In the latter event you will ideally have a laptop (windows/os x/linux) with 8GB of RAM minimum. It will be your responsibility to keep both software suites updated, including any packages being used in class. For your personal machine, make sure you download

1. the latest version of R from one of the US mirrors here and

2. the latest version of RStudio for your operating system from here.

Should we need to, I will show you how to install another free utility, MikTex for Windows and MacTex for Macs. Do not install this until I say so.

Grade Requirements

Your grade is based on ONE project that you have to work on throughout the semester and submit in a prescribed format no later than during Exam week.

The project will be finalized with my approval and can assume many forms, including an interactive visualization that you build using the data you have gathered. However, the project cannot be a slapdash arrangement of data and analyses but instead it must reflect coherent, substantive questions. These questions could be any issue in (or outside) the public affairs domain. For example, past students have used data to track the socioeconomic and demographic transition of Ohio counties over several decades, compared the performance of traditional public schools versus charter schools in Ohio over a decade, looked at gun violence and attempts at gun control over the last three decades, looked at issues of gender equality using country-level data that go back to the 1960s, and more, and so on. Some of these projects culminated in a research paper, others led to storyboards, posters, and even interactive data visualizations that were presented at Ohio University's Student Research and Creative Activity Expo. You choose what will be the final form your project will assume.

You can work on a project that is related to your thesis (should you be working on one). However, you cannot use this course to do data analysis for your thesis and expect to get credit. Instead, you have to produce work that goes sufficiently beyond your thesis. In the past, for example, students have analyzed data related to their thesis topic and then branched out to look at another sub-field within their primary area of interest.

You cannot work on a marginal project such as mapping all the birds in Ohio, or looking at opioid deaths, etc. The project has to have a narrative, questions you want to ask of the data, and a well-developed write-up of your findings. This is why you have a project clearance period and process (more on that below).

1. **Jan 24** – before 9:00 AM please send me a 1-2 page memo outlining your project ideas. This document will be an RMarkdown-generated html file that you send to me via Teams. I will send you feedback in one to two days.

2. **Jan 31** – before 9:00 AM please send me the revised memo outlining your project ideas. This document will be an RMarkdown-generated html file that you send to me via Teams. I will send you feedback in one to two days.

3. **Feb 07** – If I have suggested tasks or next steps for you, respond no later than 9:00 AM with a progress report

4. **Feb 14** – Submit a final project plan no later than 9:00 AM

5. **Apr 29** – Final project due no later than 9:00 AM

There is no set schedule for the assignments but these will be contained and due within one or two weeks of being assigned to you.

There will be no class during Spring Break – **Mar 06-12, 2022**.

Topics covered in class

1. Introduction to R

   - *Recommended for self-learning*

     (a) Chapter 1
     (b) Garrett Grolemund. 2014. "Introduction to R (Part I)."
     (c) Garrett Grolemund and Hadley Wickham. "R for Data Science."
     (d) W. N. Venables, D. M. Smith and the R Core Team. "An Introduction to R
     (e) R Tutorial: An Introduction to R
     (f) Emmanuel Paradis. 2005. "R for Beginners."
     (g) Trevor Martin. "The Undergraduate Guide to R'
     (h) The OpenIntro Team. "Introduction to R and RStudio."

2. Data Visualization with R

   - *Recommended for self-learning*

     (a) Chapter 2
     (b) Data Visualization for Social Science – Kieran Healy – Chapter 1
     (c) Alberto Cairo: The Functional Art, Chs 1-3
     (d) Alberto Cairo (2015): Reflections on the Challenges and Pitfalls of Evidence-Driven Visual Communication
     (e) Alberto Cairo: Of Conjectures and Uncertainty
     (f) Data Visualization for Social Science – Kieran Healy
     (g) Cookbook for R
     (h) ggplot2 reference guide
     (i) STHDA with ggplot2

3. Basic Data Operations in R

   - *Recommended for self-learning*

     (a) Chapter 3
     (b) Cookbook for R
     (c) r tutor
     (d) Quick-R

4. Data Acquisition with R

   - *Recommended for self-learning*

     (a) Chapter 4

5. Aggregating Data and other Operations in R

   - *Recommended for self-learning*

     (a) Chapter 5
     (b) Wickham: Tidy Data

    (c) dplyr
    (d) dplyr and tidyr
    (e) tidyr
    (f) zevross: tidyr dplyr and more
    (g) data.table

6. Working with dates and times in R

    • *Recommended for self-learning*

    (a) Chapter 6
    (b) Eric Zivot
    (c) Dates and Times Made Easy with lubridate
    (d) R for Data Science
    (e) Kan Nishida
    (f) Paul HIemstra
    (g) dates and times

7. Census Data in R

    • *Recommended for self-learning*

    (a) Chapter 9

8. Advanced Data Visualizations in R

    • *Recommended for self-learning*

    (a) Chapters 7 and 8
    (b) plotly book
    (c) Shiny user showcase
    (d) mapping choropleths and other wonders
    (e) more mapping