

ADVANCED RESEARCH METHODS IN LEADERSHIP AND PUBLIC AFFAIRS (mpa.6020)

Course Schedule	Thursday, 09:00 AM to 11:50 PM
Location	Room #105, Building #21 @ The Ridges
Office Hours	By appointment or luck
Contact Information	Room #115, Building #21, @ The Ridges (740)-597-1949 ruhil@ohio.edu

© The lectures, classroom activities, and all materials associated with this class and developed by the instructor are copyrighted in the name of (Ruhil) on January 9, 2018.

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 builds upon material covered in MPA 6010 while also providing you with the technical skills needed to manage all aspects of real-world data acquisition, munging, visualization, and analysis. We will do all of this in an open-source environment, relying centrally on (R & RStudio). The first few weeks of class will be tutorials that will introduce you to the software and walk you through "how to do with R what we did in MPA 6010 with SPSS." The majority of the class will be devoted to graphics with **ggplot2** and allied packages, and with **data munging**.

As we work through the first few weeks, you will be gathering the data you intend to use in this course. You will be encouraged to practice everything we do in a given week with your own data-set, if relevant, so that you get better at building simple graphics and analyses in R. If you do not practice enough you will impede your grasp of the material and end up with few useful practical data skills.

Note: I will not cover any of the theory underlying the statistical approaches covered in MPA 6010 in depth. If you need to revisit that material **review the textbook** and ask questions. I will 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, management and statistical analysis
- 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

Required Texts & Other Materials

Required:

1. **Data Management and Visualization with R**. This live text will give you a framework for working with 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.
2. **Data Analysis for Leadership and Public Affairs with R**. This is a live statistics textbook with R code that shows you how to run tests, a regression, etc. I will 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. It is your responsibility to keep both updated, including any packages being used in class. 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.**

You will need a laptop (windows/os x/linux) that you bring to every class meeting, with the charger.

You will need to install the free version of **Slack** on your laptop and the free Slack App for your smartphone as well. Create an account with your OU email but pick a different password, not what you use for your OU account. We will use Slack instead of email to communicate questions and answers and swap code and files while you are working on your projects. This way neither the code nor the data get corrupted in shipping and I can diagnose coding errors faster in slack.

Should we need to, in the first class session 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

There are no exams or assignments in this course. Rather, you will complete a data-intensive project. The project must be finalized after negotiating with me, and **the agreement must be in place on or before Feb 16, 2018.**

The project 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 as a group but if you do, make sure every group-member is pulling his/her weight. In addition, you will have to secure my approval for the project since what a lone individual can accomplish over the course of a semester is far less than what a group working on a single project can. This raises the bar for group projects.

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 even branched out to look at another sub-field within their primary area of interest.

Naturally, your grade is based on the (i) palpable effort you put into your project, and (ii) on the quality of the final product.

Topics covered in class

1. Introduction to R

- Required: [Chapter 1](#)
- *Recommended for self-learning*
 - [Garrett Grolemund. 2014. "Introduction to R \(Part I\)."](#)
 - [Garrett Grolemund and Hadley Wickham. "R for Data Science."](#)
 - [W. N. Venables, D. M. Smith and the R Core Team. "An Introduction to R"](#)
 - [R Tutorial: An Introduction to R](#)
 - [Emmanuel Paradis. 2005. "R for Beginners."](#)
 - [Trevor Martin. "The Undergraduate Guide to R"](#)
 - [The OpenIntro Team. "Introduction to R and RStudio."](#)

2. Data Visualization with R

- Required: [Chapter 2](#)
- Required: [Data Visualization for Social Science – Kieran Healy – Chapter 1](#)
- *Recommended for self-learning*
 - [Alberto Cairo: The Functional Art, Chs 1-3](#)
 - [Alberto Cairo \(2015\): Reflections on the Challenges and Pitfalls of Evidence-Driven Visual Communication](#)

- Alberto Cairo: Of Conjectures and Uncertainty
- Data Visualization for Social Science – Kieran Healy
- Cookbook for R
- ggplot2 reference guide
- STHDA with ggplot2

3. Basic Data Operations in R

- Required: Chapter 3
- *Recommended for self-learning*
 - Cookbook for R
 - r tutor
 - Quick-R

4. Data Acquisition with R

- Required: Chapter 4

5. Aggregating Data and other Operations in R

- Required: Chapter 5
- *Recommended for self-learning*
 - Wickham: Tidy Data
 - dplyr
 - dplyr and tidyr
 - tidyr
 - zevross: tidyr dplyr and more
 - data.table

6. Working with dates and times in R

- Required: Chapter 6
- *Recommended for self-learning*
 - Eric Zivot
 - Dates and Times Made Easy with lubridate
 - R for Data Science
 - Kan Nishida
 - Paul Hlemstra
 - dates and times

7. Advanced Data Visualizations in R

- Required: Chapters 7 and 8
- *Recommended for self-learning*
 - plotly book
 - Shiny user showcase
 - mapping choropleths and other wonders
 - more mapping