# Introduction to Computational Tools for Social Science: Digital Media

Tiago Ventura

January 14, 2021

E-mail: venturat@umd.edu Web: http://tiagoventura.rbind.io/

## Instructor


- Tiago Ventura
- Ph.D. in Political Science (IESP-UERJ), and Ph.D. Candidate in Government and Politics at the University of Maryland, College Park.
- Email: venturat@umd.edu
- Pronouns: He/Him


## Introduction

Until a few years ago, collecting data at a large scale about social behavior - whom you voted for, your closest friends, or the news you consume - was expansive, relatively niche, and ultimately rare. But in the our current *Big Data era*, this is no longer the case. Today, social behavior is efficiently recorded, stored, and observed at an unimaginable scale; As a matter of fact, the world has nowadays a marginally higher capacity to store and collect data than available researchers to analyze this large and messy amount of information. More important, the reduction in the costs of acquiring, storing, and accessing data imposes higher standards in terms of researchers' technical skills, and has subsequently introduced new methodological challenges to conduct academic research in the digital era.

This course will prepare you to conduct social science research in the digital era. We have three primary goals: First, this course will provide students with a general overview of research methods in computational social science. The course will establish enough fluency in computational programming so that students can then move forward with more advanced tools. R programming will be used throughout the course. Second, we will discuss recently published papers on digital media and political communication to expose students to top-notch computational methods, and cultivate an understanding of how these tools can be applied in social science research. Third, students will have an opportunity to apply their programming skills on original research, under a supervised project, discussed in detail below. This application will cement the skills discussed throughout the course.

This course is designed for both undergraduate and graduate students. And I designed a syllabus in which the final product be beneficial beyond this course. I understand this syllabus might be daunting, but I am flexible with making adjustments throughout the semester. That being

said, please communicate any concerns with me throughout the semester. I am open to adjusting learning methods and assignments to your needs.

Although this syllabus and some of the required readings are in English, all of the lectures and coding materials will be in **Portuguese**. Unfortunately, most of the best computational social science research has been published in English, and I firmly believe that knowing what has been published in the top-journals helps students develop their own new research ideas.

## Time and location

Classes will be held **virtually** on Mondays and Wednesdays from 11:10 AM to 12:50PM.

## Course Objectives

The course has no pre-requisite, and is designed for students at all levels of computational methods experience. At the end of the course, I expect students to be comfortable using R for research in social science, to be able to write R functions, execute them in parallel, and use Purrr for functional programming, as well as gain experience to confidently build scripts to collect data online.

Most importantly, I expect that by the end of the course students will be knowledgeable about the recent scholarship and cutting-edge methods employed by scholars in the field of computational social science, political communication, and digital media. I expect these papers to guide students toward new and future exciting research ideas.

## Communication

**Class Website**: A class website (http://tiagoventura.rbind.io/fgv_css) will be used throughout the course and should be checked on a regular basis for lecture materials and required readings.

**Class Slack Channel:** The class also has a dedicated slack channel (https://join.slack.com/t/umd-apb9922/shared_invite/zt-kne0ldx2-1J9kE0FJYxjZRA11lUhhNA). The channel serves as an open forum to discuss, collaborate, pose problems/questions, and offer solutions. Students are encouraged to pose any questions they have there, as this will provide the professor and TA the means of answering the question so that all can see the response. If you're unfamiliar with Slack, please consult the following start-up tutorial (https://get.slack.help/hc/en-us/articles/218080037-Getting-started-for-new-members). I will forward the invite link to our Slack Channel to all the students registered in the course.

## Evaluation

Students will be evaluated according to three components: participation in class, programming problem sets, and a final project.

### Participation in Class (10%)

Students will have multiple opportunities to participate and engage in discussions throughout the course. I will assess the students engagement and evaluate them on this criteria. Keeping your camera open for the most part of the lecture will be used as a criteria for participation. If you have any issue with keeping your camera on during the classes, please talk to me in private.

### Problems Sets (40%)

During the first part of the course (week 1 to week 8), students will work in problem sets related to each programming skill discussed for the specific week. Each problem set is due every Friday, by midnight, for the corresponding week. The assignments will be frequent but each of them should be fairly short, and you will have the opportunity to discuss them in class. You are encouraged to work in groups, but the work you turn in must be your own. Group submission of homework, or turning in copies of the same code or output, is not acceptable. Remember, the only way you actually learn how to write code is to write code.

Problem sets should be completed using R Markdown, a markup language used to produce well-formatted documents with embedded R code and outputs. To submit your homework, knit the R Markdown file to PDF and then submit the PDF file through Slack (unless otherwise noted). This process will be detailed in our first week of the course.

### Final Project (50%)

At the end of the course, students will work in smaller groups towards a final project, using the skills covered throughout the course. I am extremely flexible to what students can use as an final project. I am open to projects such as a new research paper, creating a new dataset, or another type of computational social science application.

Students will have three weeks to work together in this project. The first week will be dedicated to organize the smaller groups, and selecting the topic. The second week will be solely devoted for students to formally receive my support on their project. Students will meet with me and discuss their goals, objectives, and plan for the project. In the last week of class (A2), students will present the initial results of their projects. I expect these projects to be a starting point for future academic publications, end-of-course dissertations, or public-facing policy projects.

## Required Materials

### Main Readings

Below are the primary required readings for this course. I will provide the remaining article readings weekly. As noted above, most of the literature is in English, but I will work to look for equivalent readings in Portuguese. All these books are freely available on-line.

- Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press.
- Wickham, H., Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc.
- Silge, J., Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc.

## Course Plan

### Part I: Programming Skills

This first eight weeks of the course will focus on building a solid foundation on R programming. This is not going to be an introduction to statistics, computer science, or specialized social science methods. The focus will be on practical skills necessary to be successful in further work, including your final project. The weeks will be divided as follows:

**Mondays**: Lecture. I will discuss the main concepts for the week and walkthrough relevant coding.

**Wednesdays** Your turn. You will apply Monday's lesson and work on hands-on-programmings tasks. These meetings will be a opportunity for you to ask me questions and work together on your problem sets.

### Week 1: Introduction and Installfest (Februrary 1 - 3)

*Main Topics: What is computational social science? Structure of the course. Installing R and all the tools we will be using across the course*

*Readings*

- Bit by Bit, Chapter 1 and 2

### Week 2: Introduction to Programming in R (February 8-10)

*Main Topics: Crash course in R. Overview of basic R Programming*

### Week 3: Reproducibility, and Version Control (February 22 - 24)

*Main Topics: Making R Files Reproducible: Best Practices + Rmarkdown, Version Control with Github*

### Week 4: Data Wrangling (March 1 - 3)

*Main Topics: Introduction to Tidyverse*

### Week 4: Data Visualization (March 8 - 10)

*Main Topics: Introduction to GGPLOT*

### Week 5: Functional Programming in R (March 15 - 17)

*Main Topics: Functions. What is a function and how do you write one in R? Functional programming with Purrr. Parallel Processing in R.*

### Week 6: Webscraping I: Screen Scraping (March 22- 24)

*Main Topics: Using Rvest to scrap static websites, Cleaning Htmls and Xmls files*

### Week 7: Webscraping II: APIs (March 29 - 31)

*Main Topics: Acessing APIs, and Cleaning Json Files*

## Part II: Computational Social Science and Applications

From here, we will move our work towards reading and understanding applied papers in Computational Social Science. I have selected topics and articles that apply some of the best applications of new computational methods to important social science questions.

During the first meeting of each week (Mondays), we will discuss together the applied papers. I will not lecture on the paper, this will be a open discussion in which your engagement is key. In the second meeting for each week, I will show you how to implement some of the methods used in each of the papers (hands-on programming).

Readings with an asterisk (*) are mandatory. You will read at most two papers per week. The other readings are only suggested. Some suggested readings can be added to the syllabus throughout the course.

## Week 8: Working with Social Media Data (April 19 - 26)

**Applied Papers**

- Calvo, Ernesto and Natalia Aruguete. 2016. Time to Protest: Polarization and Time-to-Retweet in Argentina. Journal of Communication. *
- Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. "Tweeting from left to right: Is online political communication more than an echo chamber?." Psychological science 26, no. 10 (2015): 1531-1542. *
- Shugars, Sarah, and Nicholas Beauchamp. "Why Keep Arguing? Predicting Engagement in Political Conversations Online." SAGE Open 9, no. 1 (January 1, 2019): 2158244019828850. https://doi.org/10.1177/2158244019828850.
- Bakshy, E., Messing, S., Adamic, L. 2015. Exposure to ideologically diverse news and opinion on Facebook. science, 348(6239), 1130-1132.

**Programming Skills:**

- Accessing the Twitter API
- Reading Data From Twitter
- Mapping news, and content activation from Twitter
- Ideological Scaling using Twitter data

## Week 9: Experiments on Social Media Behavior (April 28 - May 3)

**Applied Papers:**

- Calvo, E. and Ventura, T., 2020. Will I Get Covid? Partisanship, Social Media Frames, and Perceptions of Health Risk in Brazil? *
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295-298. *
- Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences, 117(27), 15536-15545.
-

**Programming Skills:**

- Intro to Experiments and Randomization of Online Surveys.

## Week 10: Recruiting Survey Respondents Through Online Tools (May 5)

**Applied Papers:**

- Samuels, D.J. and Zucco, C., 2013. Using Facebook as a subject recruitment tool for survey-experimental research. Available at SSRN 2101458. *

- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." Political analysis 20, no. 3 (2012): 351-368.

### Week 12: Digital Field Experiments (May 10 - 12)

**Applied Papers:**

- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. Political Behavior, 39(3), 629–649.*
- Victoria Asbury, Keng-Chi Chang, Katherine McCabe, Kevin Munger, Tiago Ventura; The Effect of Streaming Chat on Perceptions of Debates. Under Review.*
- Bail et al (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37): 9216-9221.
- King, G., Pan, J., Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. Science, 345(6199), 1251722.

**Programming Skills:**

- Creating a Bot on Twitter.
- Intro to Mturk in R.

### Week 13: Text Analysis: Overview (May 17 - 19)

**Applied Papers**

- Grimmer, J., Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3), 267-297.*
- Wilkerson, J. and Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. Annual Review of Political Science, 20, 529:544.

**Programming Skills:**

- Text Data in R, Tidytext, and Quanteda.
- Silge, J., Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc.

### Week 14: Text Analysis: Supervised Classification and Topic Models (May 24 - 26)

**Applied Papers:**

- Moreira, D., 2020. Com a Palavra os Nobres Deputados: Ênfase Temática dos Discursos dos Parlamentares Brasileiros. Dados, 63(1).*
- Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A. and Parnet, O. (2016). A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting With Party Candidates. Journal of Communication, 66: 1007–1031.*
- Barberá, Pablo, Amber E. Boydstun, Suzanna Linn, Ryan McMahon, and Jonathan Nagler. "Automated Text Classification of News Articles: A Practical Guide." Political Analysis, undefined/ed, 1–24.

- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. American Journal of Political Science, 58(4), 1064-1082.
- Terman, R. (2017). Islamophobia and Media Portrayals of Muslim Women: A Computational Text Analysis of US News Coverage. International Studies Quarterly, 61(2): 489-502.

**Programming Skills:**

- Structural Topic Models in R

## Part III: Final Project

### Week 16: Final Project: Research Speed Dating (May 31 - June 2)

- Picking your groups and topics.

### Week 17: Final Project: Working in Groups (June 7 - 9 )

- Working Together and Presenting initial results

### Week 18: Final Project: Final Presentation (June 14 - 16)

- Final Presentation