

Introdução a Métodos Computacionais para Ciência Sociais: Comunicação Política e Mídias Sociais

Tiago Ventura

January 17, 2021

E-mail: venturat@umd.edu

Web: <http://tiagoventura.rbind.io/>

Instructor

- Tiago Ventura
- Ph.D. em Ciência Política (IESP-UERJ), e Ph.D. Candidate em Governo e Política na Universidade de Maryland, College Park.
- Email: venturat@umd.edu
- Pronome: Ele/Dele

Introdução

Até alguns anos atrás, coletar dados em larga escala sobre comportamento social - por exemplo, quem você votou, quem são seus amigos mais próximos, ou quais tipos de notícias você consome, lê e compartilha - era uma atividade cara, custosa, aberta a poucos pesquisadores e em geral raro. Porém, nos tempos atuais de *Big Data*, a situação alterou-se por completo. O mundo atualmente possui uma capacidade maior de armazenar e coletar dados do que há pesquisadores capazes e disponíveis de analisar este volume imenso e desorganizado de dados. E mais, a redução nos custos de adquirir, guardar e acessar dados impõe padrões mais altos para os conhecimentos metodológicos e técnicos dos pesquisadores, sobretudo nas ciências sociais, e tal mudança tem introduzido desafios diversos para pesquisa em ciências sociais na era de *Big Data*

Este curso busca preparar você para conduzir pesquisa em ciências sociais nesta era de *Big data*. O curso possui três objetivos principais: em primeiro lugar, iremos proporcionar aos alunos uma introdução compreensiva sobre métodos computacionais em ciências sociais. A ideia é garantir aos alunos após oito semanas de aula alguma fluência em programação, que permitirá aos mesmos usar tais técnicas em suas próprias pesquisas. A linguagem R será utilizada no curso. Em segundo lugar, iremos discutir artigos aplicados na área de mídias digitais e comunicação política. O objetivo é expor alunos aos que considero exemplos das melhores pesquisas neste campo, e que utilizam os mais atualizados recursos metodológicos. Em terceiro lugar, os alunos terão a oportunidade de aplicar as técnicas e conteúdos adquiridos ao longo do curso em uma ideia original de pesquisa.

Este curso é pensado para alunos de graduação e pós graduação. Enquanto esta ementa pode parecer um tanto longa e intimidadora, não há como adquirir o conjunto de habilidades discutidas neste curso sem um esforço considerável. Dito isto, por favor, avisem-me qualquer preocupação

com o curso no decorrer do semestre. Estou aberto a ajustes e adequações a esta ementa, e o objetivo principal é preparar um curso capaz de ajudar os alunos no longo prazo.

A maior parte das leituras do curso serão em Inglês. Porém, todas as aulas, códigos, e slides estarão em Português. Infelizmente, as melhores pesquisas em Métodos Computacionais para Ciências Sociais, tanto teóricas quanto aplicadas, ainda encontram-se em inglês. E eu acredito firmemente que conhecer tais pesquisa pode gerar insights para suas próprias idéias.

Aulas

As aulas serão todas **virtuais** as segundas e quartas-feiras, das 11:10AM até 12:50 PM.

Objetivos

Este curso não possui nenhum pré-requisito, e é desenhado para alunos com experiências variadas em métodos computacionais. Ao final do curso, eu espero que os alunos sintam-se confortáveis em programar em R e em usar este software em suas próprias pesquisas.

Mais importante, ao final deste curso, minha expectativa é que os alunos conheçam a literature e métodos usados na fronteira do conhecimento em ciências sociais computacionais, e particularment comunicação política. Estes papers certamente guiarão alunos a densenvolver suas próprias idéias.

Comunicação

Site: Uma página na internet para o curso (http://tiagoventura.rbind.io/fgv_css) será disponibilizada para os alunos. Você deve checar o site regularmente e verificar os materiais que utilizaremos ao longo do semestre.

Slack: Nosso seminário também fará uso do Slack para comunicação interna (https://join.slack.com/t/umd-apb9922/shared_invite/zt-kne0ldx2-1J9kE0FJYxjZRA11IUhhNA). A plataforma slack funciona como um fórum aberto de discussão e colaboração. Quaisquer pergunta direcionadas a mim devem ser feitas via Slack. Se você não conhece o Slack, cheque este tutorial (<https://get.slack.help/hc/en-us/articles/218080037-Getting-started-for-new-members>). Na primeira aula, eu enviarei o convite para todos serem incluídos no Slack e discutiremos como usá-lo.

Avaliação

Os alunos serão avaliados de acordo com três components: participação, exercícios de programação, e projeto final.

Participação em Aula (10%)

Ao longo do curso, serão dadas amplas oportunidade de participação aos alunos. Tanto nas discussões sobre artigos aplicados, quanto nos exercícios de programação. Eu avaliarei os alunos com base em seu engajamento nas referidas discussões.

Durante as aulas pedimos que o vídeo dos alunos permaneça sempre ligado. O dos professores também ficará. O uso da câmera é considerado como critério de participação em aulas.

Exercícios. (40%)

Durante a primeira parte do curso (semana 1 até semana 8), os alunos trabalharão em exercícios de programação relacionados a cada tópico da semana. Cada exercício deve ser entregue até a sexta-feira da semana correspondente, até meia noite. Atraso na entrega dos exercícios será penalizado nas notas. Estes exercícios serão frequentes durante as primeiras oito semanas, porém, os farei relativamente curtos. Vocês poderão discutí-los em sala da aula comigo e com seus colegas. Eu encorajo alunos a trabalharem em grupo, porém, cada exercício deve ser submetido individualmente.

Os exercícios devem ser respondidos usando R Markdown, uma linguagem capaz de produzir documentos com bons formatos e combinando resultados, escrita e códigos de R. Para submeter seus exercícios, você precisará compilar seu arquivo de R markdown em formato pdf, e submeter via slack. Nas duas primeiras semanas, iremos praticar diversas vezes este procedimento.

Projeto Final (50%)

Nas últimas semanas do curso, os alunos trabalharão em grupos (3-4) em um projeto final de curso aplicando as técnicas discutidas no decorrer do semestre. O formato para o projeto final é bastante flexível. Os alunos podem optar por iniciar um desenho de pesquisa para um novo artigo, criar um novo banco de dados, ou qualquer outra tipo de aplicação em ciências social computacional, que receba meu aval.

Os alunos terão três semanas para desenvolver o projeto. A primeira semana será dedicada a dividir os alunos em seus grupos e selecionar o tópico e formato do projeto final. A segunda semana será dedicada somente ao desenvolvimento do projetos, em que os alunos poderão conversar em grupos, e receber meu feedback sobre os projetos. A semana final (A2) será dedicada a apresentação dos resultados. Minha expectativa é de que estes projetos auxiliem alunos a escrever novos artigos, iniciar trabalhos relacionados aos seus TCCs e Dissertações, ou criar instrumentos capazes de auxiliar decisões de políticas públicas e transparência de dados.

Materiais

Leituras Principais

Estas são as leituras principais do curso. Todas estão disponíveis gratuitamente on-line. Eu irei prover semanalmente cópias dos artigos para as semanas específicas.

- [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.](#)

Plano do Seminário

Parte I: Programando em R

As primeiras oito semanas do curso focarão em construir uma sólida base em programação em R. Neste período, não focaremos em introdução a estatística, ou outros métodos das ciências sociais

computacionais; o objetivo principal é passar aos alunos as capacidades básicas para entender R e seguir adiante com seus projetos no futuro. Cada semana será dividida da seguinte forma:

Segundas: Aula em que discutirei os conceitos principais da semana, e darei vários exemplos de códigos e aplicações em R.

Quartas Sua vez! Nestes dias, será sua vez de aplicar o que discutimos na aula anterior. Você poderá rever o código, fazer perguntas, e iniciar seus exercícios de programação da semana.

Semana 1: Introdução e Festival da Instalação (Fevereiro 1 - 3)

Tópicos: O que é Ciência Social Computacional? Estrutura do Curso, Instalando R e todas as outras ferramentas que iremos utilizar no curso.

Leituras

- Bit by Bit, Chapter 1 and 2

Semana 2: Introdução ao R (Fevereiro 8-10)

Tópicos: Curso Rápido em R, objetos, operações básicas, R Studio.

Semana 3: Replicabilidade e Controle de Versão. (Fevereiro 22 - 24)

Tópicos: Rmarkdown e Git-Github

Semana 4: Manipulação de Dados (Março 1 - 3)

Tópicos: Introdução ao Tidyverse

Semana 4: Visualização de Dados (Março 8 - 10)

Tópicos: ggplot2

Semana 5: Programação Funcional em R (Março 15 - 17)

Tópicos: Funções, Loops, Purr e Processamento em Paralelo.

Semana 6: Raspagem de Dados I: Raspando Websites (Março 22- 24)

Tópicos: Rvest e Limpando arquivos Htmls and Xmls

Semana 7: Raspagem de Dados II: APIs (Março 29 - 31)

Tópicos: APIs e Arquivos Json

Parte II: Ciência Social Computacional: Artigos Aplicados

A partir daqui, avançaremos para a leitura e compreensão de artigos aplicados em Ciências Sociais Computacionais. Esta será a parte mais interessante do curso!

Selecionei tópicos e artigos que considero os melhores exemplos no uso novos métodos computacionais e respondendo a questões importantes das ciências sociais.

Durante a primeira reunião de cada semana (em geral, segundas-feiras), discutiremos juntos os artigos. Não vou fazer uma palestra sobre os artigos, esta será uma discussão aberta em que o seu envolvimento é a chave.

Cada aluno deve trazer **uma pergunta** e **um comentário** sobre o aspecto mais importante de ao menos um dos textos para cada uma de nossas aulas.

Na segunda reunião de cada semana, mostrarei como implementar alguns dos métodos usados em cada um dos artigos (programação prática).

As leituras com um asterisco (*) são obrigatórias. Você vai ler no máximo dois artigos por semana. As outras leituras são apenas sugeridas. Algumas leituras sugeridas podem ser adicionadas ao programa ao longo do curso.

Semana 8: Dados de Mídias Sociais (Abril 19 - 26)

Artigos Aplicados

- 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.

Técnicas de Programação:

- Acessando dados na API do Twitter.
- Processando dados do Twitter.
- Analizando Notícias e Ativação de Conteúdo com dados do Twitter.
- Escalonamento ideológico com dados do Twitter.

Semana 9: Experimentos sobre comportamento em mídias sociais. (Abril 28 - Maio 3)

Artigos Aplicados:

- 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.

-

Técnicas de Programação:

- Introdução a Experimentos e Randomização em Surveys On-line.

Semana 10: Recrutamento para Surveys On-line (Maio 5)

Artigos Aplicados:

- 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.

Semana 12: Experimento Digitais de Campo (Maio 10 - 12)

Artigos Aplicados:

- 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.

Técnicas de Programação::

- Criando Robôs no Twitter.
- Breve Introdução ao Mturk.

Semana 13: Análise de Texto: Visão Geral (Maio 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.

Técnicas de Programação:

- Tidytext e Quanteda para Análise de Texto.
- [Silge, J., Robinson, D. \(2017\). *Text mining with R: A tidy approach*. O'Reilly Media, Inc.](#)

Semana 14: Análise de Texto: Aprendizado Supervisado e Modelo de Tópicos. (Maio 24 - 26)

Artigos Aplicados:

- 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.

Técnicas de Programação:

- Modelo de Tópicos Estruturais em R

Parte III: Projeto Final

Semana 16: Projeto Final: Research Speed Dating (May 31 - June 2)

- Selecionando seus grupos e tópicos

Semana 17: Projeto Final: Mãos a Obra (June 7 - 9)

- Sessões dedicadas para vocês desenvolverem seus trabalhos finais

Semana 18: Projeto Final: Apresentação (June 14 - 16)

- Apresentação Final