

FLS 6397 - Introdução à Análise de Dados, Programação e Visualização para as Ciências Sociais

1. Informações básicas

Primeiro Semestre, 2022

DCP - FFLCH - USP

Sexta-feiras, 9h - 13h

Glauco Peres da Silva, Rafael Magalhães, Jonathan Phillips

2. Apresentação

Como passamos de um banco de dados bruto para um relatório ou artigo publicado com estatísticas precisas, tabelas claras e gráficos convincentes? Existem diversas ferramentas poderosas para facilitar o processo de análise de dados, a pesquisa e a publicação, mas elas podem parecer intimidadores e complexos. Este curso oferece aos estudantes de pós-graduação em ciências sociais uma iniciação acessível ao uso de software para a coleta, limpeza, análise, comparação e visualização de dados.

O foco do curso é o desenvolvimento da habilidade de programação para solução de problemas diversos relacionados ao manejo de dados com fins de pesquisa. Note-se que não é um curso de metodologia de pesquisa ou estatística. No final do curso alunos serão capazes de desenvolver um script que baixar os dados, organizar-los, calcular medidas relevantes, construir tabelas, gráficos e mapas, e gerar um relatório final para uso na sua dissertação/tese, tudo em uma forma verificável, reproduzível, transparente e documentado.

3. Software

Adotamos o uso de língua de R no curso como uma língua aberta, com muito suporte disponível online e uma ferramenta bem usado nas ciências sociais. Nota que o treinamento em R permite uma transição muito mais fácil para outras línguas no futuro, e as princípios de programação e tratamento de dados são bem transferíveis.

4. Estrutura do curso

Aulas, leituras e materiais

As aulas serão compostas por breves apresentações dos tópicos e por longos laboratórios, com tutoriais para auto-aprendizado e acompanhamento dos instrutores e assistentes. Espera-se que a turma pratique exaustivamente, dentro e fora de sala de aula, as técnicas aprendidas.

Atividades e tempo de dedicação

Ao longo do curso as participantes deverão solucionar desafios correspondentes aos tópicos, usando bancos de dados comuns nas ciências sociais. Exemplos de desafios: (1) organizar automaticamente dados eleitorais a partir do repositório de dados do TSE; (2) elaborar um mapa com dados municipais a partir do DATASUS. Os desafios exigirão dedicação extra-classe e são parte fundamental do curso.

No final do curso as estudantes deverão elaborar um projeto individual relacionado a sua própria pesquisa/interesses.

Entre aulas, tutoriais, leituras e desafios, espera-se que cada aluna ou aluno dedique de 8h a 12h por semana à disciplina.

Avaliação

A avaliação é composta pela entrega dos desafios, do projeto final individual e da participação. A atribuição de nota para os desafios e projetos entregues priorizará o esforço e engenhosidade apresentados (leia-se “código com erros, mas bem elaborado”) em detrimento da finalização do desafio (leia-se “código funcionando plenamente”) como forma de encorajar estudantes iniciantes.

5. Requisitos

Não é necessário nenhum conhecimento prévio de programação, pacotes estatísticos ou manejo de conjuntos de dados. O curso é recomendado para tanto alunas e alunos que já têm alguma noção quanto para estudantes que morrem de medo de computadores. O objetivo é criar um ambiente confortável para o aprendizado de técnicas programação, independentemente da habilidade das inscritas, e seguindo todos os passos desde a preparação do ambiente de computação até a apresentação de resultados.

É recomendado que as participantes já tenham concluído ou esteja cursando algum curso de métodos de pesquisa (de qualquer abordagem) ou de análise de dados, seja do programa ou da IPSA-USP Summer School. É um curso adequado para estudantes em qualquer etapa do mestrado ou doutorado, desde que tenham disponibilidade para realizar as atividades extra-classe.

Estudantes do DCP-USP que podem se matricular regularmente não serão aceitas como ouvintes.

6. Tópicos

1. Introdução ao curso, R e Rmarkdown
2. Abrindo e manipulando bancos de dados
3. Organização e limpeza de dados
4. Calculando estatísticas resumidas
5. Construindo tabelas
6. Visualização de dados e a gramática de gráficos
7. Juntando bancos de dados
8. Mapas e análises espaciais
9. Testes estatísticos e modelos de regressão
10. Produzindo relatórios reproduzíveis com Git e Latex
11. Funções e repetição
12. Mineração de Textos