

# FLS 6397 - Introdução à Programação e Ferramentas Computacionais para as Ciências Sociais

*Leonardo S. Barone e Fernando Limongi*

*01-12-2016*

versão online do programa disponível em <https://github.com/leobarone/FLS6397/>

## 1. Informações básicas

### **Semestre:**

1o de 2016

### **Responsáveis:**

Leonardo S. Barone e Fernando Limongi

### **Horário:**

Segundas-feiras, no período da manhã (horário de início a definir)

### **Sala:**

a definir

### **Número de Créditos:**

8

## 2. Apresentação

O curso oferece aos estudantes de pós-graduação em ciências sociais uma iniciação à lógica de programação e ao uso ferramentas computacionais para (1) coleta (2) organização e preparação e (3) apresentação de dados para pesquisa social. 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, análise de dados ou de ciências sociais computacionais.

Este curso pretende preencher uma lacuna de formação comum em alun@s de pós-graduação em ciências sociais referente à habilidade de lidar com dados e ferramentas computacionais para a condução de suas pesquisas.

O curso está dividido em duas partes. A primeira parte do curso foca na preparação do ambiente computacional, apresentação de ferramentas e “alfabetização” e nas linguagens de programação utilizadas no curso – R, Git, SQL, Markdown, etc. A segunda parte foca em aplicações das habilidades desenvolvidas a problemas de manipulação de grandes bases de dados, coleta de dados na internet e via uso de APIs de terceiros, organização de documentos para análise textual, confecção de gráficos e mapas ou demais tópicos de interesse da turma. Os tópicos serão apresentadas definitivamente no syllabus no começo do semestre e abaixo segue uma lista provisória.

### 3. Estrutura do curso

#### Aulas

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.

#### Leituras e materiais

Na versão final do programa serão indicadas leituras para cada tópico. O material de estudo básico, entretanto serão os tutoriais preparados para cada aula.

As 3 atividades abaixo (originalmente de uma oficina oferecida no DCP) são exemplos de tutoriais para auto-aprendizado que comporiam uma aula (clique no link da atividade): - [Atividade 1] ([https://github.com/leobarone/Webscraping\\_R\\_XML\\_Legislativo\\_2016\\_1/blob/master/Atividade\\_1.Rmd](https://github.com/leobarone/Webscraping_R_XML_Legislativo_2016_1/blob/master/Atividade_1.Rmd)) - [Atividade 2] ([https://github.com/leobarone/Webscraping\\_R\\_XML\\_Legislativo\\_2016\\_1/blob/master/Atividade\\_2.Rmd](https://github.com/leobarone/Webscraping_R_XML_Legislativo_2016_1/blob/master/Atividade_2.Rmd)) - [Atividade 3] ([https://github.com/leobarone/Webscraping\\_R\\_XML\\_Legislativo\\_2016\\_1/blob/master/Atividade\\_3.Rmd](https://github.com/leobarone/Webscraping_R_XML_Legislativo_2016_1/blob/master/Atividade_3.Rmd))

#### Atividades e tempo de dedicação

A cada semana haverá um desafio correspondente ao tópico. 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; (3) criar um corpus de notícias da internet; (4) abrir e organizar os dados do ENEM ou Censo Populacional. Os desafios exigirão dedicação extra-classe e são parte fundamental do curso.

Entre aulas, tutoriais e desafios, espera-se que cada alun@ dedique de 8h a 12h por semana à disciplina.

#### Avaliação

A avaliação é composta pela entrega dos desafios e de um projeto final de organização de dados. 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 encojorar estudantes iniciantes.

### 4. Requisitos

Não é necessário nenhum conhecimento prévio de programação, pacotes estatísticos ou manejo de conjuntos de dados. O curso é recomendado para tanto alun@s 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 d@s inscric@s, e seguindo todos os passos desde a preparação do ambiente de computação até a apresentação de resultados.

É recomendado que @s alun@s inscric@s 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 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 aceit@s como ouvintes.

## 5. Agenda

A ser definida em breve.

## 6. Tópicos (PROVISÓRIO)

1. Básico da Programação em R
2. Estruturas de dados e manipulação de bases em R
3. Tabelas e Gráficos em R
4. SQL e dados volumosos
5. Git básico
6. Markdown básico
7. LaTeX básico
8. Captura de dados na internet
9. Textos, corpus e processamento de linguagem natural
10. Mapas e GIS
11. Redes e grafos

## 7. Bibliografia (PRELIMINAR)

### Livros

- Jones, O., Maillardet, R., & Robinson, A. (2009). Introduction to Scientific Programming and Simulation Using R (1 edition). Boca Raton, FL: Chapman and Hall/CRC.
- Makhabel, B. (2015). Learning Data Mining with R. Packt Publishing Ltd.
- Munzert, S., Rubba, C., MeiBner, P., & Nyhuis, D. (2014). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.
- Nolan, D., & Lang, D. T. (2013). XML and Web Technologies for Data Sciences with R. Springer Science & Business Media.
- Teetor, P. (2011). R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics. O'Reilly Media, Inc.
- Wickham, Hadley. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
- Wickham, Hadley and Francois, Romain (2016). dplyr: A Grammar of Data Manipulation. R package version 0.5.0. <https://CRAN.R-project.org/package=dplyr>
- Zuur, A., Ieno, E. N., & Meesters, E. (2009). A Beginner's Guide to R (2009 edition). Dordrecht; New York: Springer.

### Ferramentas

- Barone, Leonardo Sangali, Aslan, Alexia and McDonnell, Robert Myles (2016). bRasilLegis: R Tools for Brazilian Chamber of Deputies (Camara Dos Deputados) Data. R package version 1.0. <https://CRAN.R-project.org/package=bRasilLegis>
- Ingo Feinerer and Kurt Hornik (2015). tm: Text Mining Package. R package version 0.6-2. <https://CRAN.R-project.org/package=tm>
- Duncan Temple Lang and the CRAN Team (2016). XML: Tools for Parsing and Generating XML Within R and S-Plus. R package version 3.98-1.4. <https://CRAN.R-project.org/package=XML>