

PROGRAMA DE DISCIPLINA

IDENTIFICAÇÃO:

Disciplina: Ciência de Dados Aplicada
Créditos: 01 / Semestre/ano: 02/2017
Professor: Ricardo Barros Sampaio e Tainá Raiol
E-mail: ricardo.sampaio@fiocruz.br

A inscrição da disciplina só será confirmada após o preenchimento do formulário online no link:

<https://goo.gl/forms/xZyfmhztQpsN7Ss1>

EMENTA:

- A temática do curso está voltada para a área da ciência de dados na plataforma R de programação no estudo quantitativo de bases de dados da saúde.
- Atividades em ambiente de ensino online com atividades práticas.
- O curso busca desenvolver competências para o estudo e análise quantitativa com aplicação de métodos como mineração de texto, tratamento e visualização de dados e aprendizado de máquina.
- Serão abordados os conceitos introdutórios sobre a Ciência de Dados e a aplicação prática com bases de dados reais.
- As atividades serão desenvolvidas dentro do ambiente online da plataforma DataCamp (<https://www.datacamp.com/home>) e da linguagem R, RStudio e Shiny e para isso serão trabalhadas competências para o seu uso de forma introdutória.

OBJETIVOS:

1. Entender os conceitos principais sobre a Ciência de Dados
2. Estudar a plataforma R e as suas aplicações básicas no tratamento estatístico de dados
3. Estudar o ambiente de programação R, RStudio e Shiny para o uso de forma adequada dos pacotes disponíveis
4. Realizar atividades de importação e limpeza de dados
5. Construir funções em R para o tratamento de dados científicos
6. Aplicar as funcionalidades de tratamento de dados nas bases disponibilizadas com o pacote dplyr
7. Desenvolver soluções para a visualização de dados científicos utilizando o pacote ggplot2
8. Estudar os conceitos de Aprendizado de Máquina incluindo classificação e regressão estatística
9. Desenvolver aplicativos para análise e disponibilização visual de dados estruturados
10. Descrever o processo de análise de dados, desde o levantamento até a análise e apresentação dos resultados

CONTEÚDO PROGRAMÁTICO:



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz
Brasília

PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICAS PÚBLICAS EM SAÚDE

ESCOLA FIOCRUZ DE GOVERNO – EFG/Gereb/Fiocruz

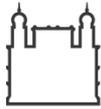
MESTRADO PROFISSIONAL

PROGRAMA DE DISCIPLINA

- a) Plataforma R;
- b) Tratamento de Dados;
- c) Visualização de dados;
- d) Estatística básica;
- e) Aprendizado de Máquinas;
- f) Análise de Redes;
- g) Plataforma Shiny;
- h) Geração de Relatórios.

METODOLOGIA DE ENSINO

- O curso envolve a realização de atividades práticas online (até 40% da pontuação) e um trabalho em grupo (até 60% da pontuação).
- As atividades online são auto-instrutivas (estudo dirigido), e ocorrem no ambiente [datacamp.com](https://aprender.ead.unb.br/course/view.php?id=60) e com suporte de fóruns no ambiente de EAD da UnB (<https://aprender.ead.unb.br/course/view.php?id=60>). Serão disponibilizados 16 (dezesesseis) módulos online mas apenas 8 (oito) serão obrigatórios, cuja conclusão plena até a data e horário limites confere ao aluno 5 pontos cada, totalizando um máximo de 40 pontos. O prazo para finalização das disciplinas obrigatórias é até o dia 08 de outubro.
- O trabalho em grupo (até 60 pontos) consiste na produção de uma análise de uma base de dados real, produzido em grupo de até três alunos, versando sobre problemas brasileiros em Ciência, Tecnologia e Sociedade (CTS).
- O trabalho deve utilizar os recursos de ciência de dados da plataforma R, aprendidos no ambiente online, além da plataforma Shiny (<http://shiny.rstudio.com/>). Os dados e as questões usadas para o trabalho final serão fornecidos pelos professores no encontro de 1º de setembro.
- Espaços para trabalho em grupo e encontro com os professores orientadores poderão ser realizados nas dependências da Escola Fiocruz de Governo as sextas feiras pela manhã no horário acordado pelo grupo.
- Para orientação inicial da disciplina e sobre o trabalho em grupo ocorrerão 2 encontros presenciais de participação obrigatória, no Auditório Interno da Fiocruz-Brasília, no campus Darcy Ribeiro. No dia 11/08 das 14:00 as 18:00 e no dia 01/set das 08:30 as 10:00.
- Visando orientação para realização das atividades online e esclarecimento de outras dúvidas será ofertado apoio presencial, com frequência opcional, em laboratório de informática da



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz
Brasília

PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICAS PÚBLICAS EM SAÚDE

ESCOLA FIOCRUZ DE GOVERNO – EFG/Gereb/Fiocruz

MESTRADO PROFISSIONAL

PROGRAMA DE DISCIPLINA

Fiocruz, nas sextas-feiras, das 16h às 18h, nos seguintes dias: (Lab 1 – excepcionalmente no LINF, subsolo do minhocão) 18/ago; (Lab 2) 25/ago; (Lab 3) 01/set; (Lab 4) 15/set; (Lab 5) 22/set; (Lab 6) 29/set; (Lab 7) 06/out; (Lab 8) 20/out; (Lab 9) 27/out; (Lab 10) 10/nov; (Lab 11) 17/nov; (Lab 12) 24/nov e (Lab 13) 01/dez.

- Visando ainda oferecer mais flexibilidade e expansão dos horizontes da ciência de dados, serão oferecidas 5 (cinco) palestras sobre temas de ciência de dados, no Auditório interno da Fiocruz, das 08:30h às 12h:

- 18/08 (manhã) Plataforma R e Tidy Data
- 01/09 (manhã) Apresentação das bases e conceitos gerais do DPLYR e GGLOT 2
- 20/10 (manhã) Análise de Redes com R
- 10/11 (manhã) Plataforma Shiny e Visualização de Redes
- 24/11 (manhã) Aprendizado de máquina

- Todos os trabalhos finais realizados pelos alunos e que tiverem recebido pontuação superior a 40 (quarenta) pontos serão colocados em domínio público. A orientação e avaliação dos trabalhos finais serão feitas com o apoio de outros professores da UnB que puderem colaborar com a disciplina.

- A conclusão do curso envolve ainda uma atividade social com participação em congresso, com data a ser definida.

Pré requisitos:

- O nível de exigência da disciplina é avançado e é essencial o domínio de leitura e escuta em língua inglesa
- Recomendável, mas não exigido, conhecimento sobre lógica de programação e ferramentas para tratamento e visualização de dados

Serão demandados exercícios e leituras regulares, e, portanto, o(a) aluno(a) deve dispor de 4 horas de atividade extra-sala por semana

RECURSOS INSTRUCIONAIS:

Laboratório de informática para mentoria
Sala de aula para exposição de conteúdo
Acesso a ambiente online para realização das atividades



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz
Brasília

PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICAS PÚBLICAS EM SAÚDE

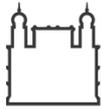
ESCOLA FIOCRUZ DE GOVERNO – EFG/Gereb/Fiocruz

MESTRADO PROFISSIONAL

PROGRAMA DE DISCIPLINA

Proposta de cronograma do curso

Encontros	Descrição da Conteúdo	Carga horário prevista
Dia 11 ago 14:00 as 18:00 FIOCRUZ	Apresentação da disciplina e inscrição no ambiente Data Camp	4 horas
Dia 18 ago 08:30 as 12:00 FIOCRUZ	Plataforma R e Tidy Data	4 horas
Dia 01 set 08:30 as 12:00 FIOCRUZ	Apresentação dos data sets e divisão em grupos Pacote DPLYR Pacote GGLOT 2	4 horas
Dia 20 out 08:30 as 12:00 FIOCRUZ	Análise de Redes com R	4 horas
Dias 10 nov 08:30 as 12:00 UnB/FCI	Plataforma Shiny e Visualização de Redes	4 horas
Dias 24 nov 08:30 as 12:00 UnB/FCI	Aprendizado de máquina	4 horas
Dias 08 dez 08:30 as 12:00 UnB/FCI	Apresentação dos trabalhos finais	4 horas
Laboratórios		
18 ago 16:00 as 18:00	Laboratório de Informática – Subsolo do minhocão (ICC) norte	2 horas
Dias: 25/08; 01/09 15/09; 22/09 29/09; 06/10 20/10; 27/10 10/11; 17/11 24/11; 01/12 Horário: 16:00 as 18:00	Laboratório de Informática – Escola Fiocruz de Governo	2 horas
Reunião Alunos / Professores		
Dias: 25/08; 22/09 06/10; 27/10 17/11; 01/12 Horário: 08:00 as 12:00	Espaço para discussão com equipe e professores	4 horas



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz
Brasília

PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICAS PÚBLICAS EM SAÚDE

ESCOLA FIOCRUZ DE GOVERNO – EFG/Gereb/Fiocruz

MESTRADO PROFISSIONAL

PROGRAMA DE DISCIPLINA

AVALIAÇÃO:

- O rendimento na disciplina será avaliado de duas formas principais:
 - o 5 pontos por realização de cada um dos 8 módulos online obrigatórios, totalizando 40 pontos.
 - o 60 pontos por entrega de relatório e aplicativo com análise de dados, efetuado em grupo com até três alunos, sendo 30 pontos pelo relatório em formato R Markdown e 30 pontos do aplicativo em Shiny com disponibilização dos dados.
- A participação no congresso com a organização e/ou apresentação dos resultados poderá contar pontos extras

REFERÊNCIAS BIBLIOGRÁFICAS:

- Referência Principal
- WICKHAM, Hadley; GROLEMUND, Garrett. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. USA: O'Reilly, 2016.
-
- Referências Auxiliares
- GROLEMUND, Garrett. Hands-On Programming with R: Write your own functions and simulations. USA: O'Reilly, 2014.

BIBLIOGRAFIA COMPLEMENTAR:

- <https://www.datacamp.com/courses/free-introduction-to-r>
- <https://www.datacamp.com/courses/intermediate-r>
- <https://www.datacamp.com/courses/intermediate-r-practice>
- <https://www.datacamp.com/courses/importing-data-in-r-part-1>
- <https://www.datacamp.com/courses/importing-data-in-r-part-2>
- <https://www.datacamp.com/courses/cleaning-data-in-r>
- <https://www.datacamp.com/courses/importing-cleaning-data-in-r-case-studies>
- <https://www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial>
- <https://www.datacamp.com/courses/joining-data-in-r-with-dplyr>
- <https://www.datacamp.com/courses/data-visualization-in-r>
- <https://www.datacamp.com/courses/data-visualization-with-ggplot2-1>
- <https://www.datacamp.com/courses/introduction-to-data>
- <https://www.datacamp.com/courses/exploratory-data-analysis>
- <https://www.datacamp.com/courses/exploratory-data-analysis-in-r-case-study>
- <https://www.datacamp.com/courses/correlation-and-regression>
- <https://www.datacamp.com/courses/reporting-with-r-markdown>
- <http://shiny.rstudio.com>