



PLANO DE ENSINO

NOME DO PROFESSOR: Francisco Alves de Oliveira Júnior		NOME DO CURSO: Administração Pública			
DISCIPLINA: Tópicos Especiais em Administração Pública I - Análise e visualização de dados com R		CRÉDITO:	PERÍODO:	CARGA HORÁRIA TOTAL: 30 hs	
EMENTA: Processo de análise de dados. Pesquisa reproduzível. Introdução à programação. Visualização de Dados: gramática dos gráficos; uso do pacote ggplot2. Manipulação de Dados. Introdução à programação funcional.					
OBJETIVO GERAL: O objetivo desta disciplina é apresentar ao aluno a utilização do ambiente computacional R para manipulação e visualização de dados, com foco em dados tabulares, utilizando práticas de pesquisa reproduzível. O enfoque não é a utilização do R para fins de modelagem estatística, no entanto, conceitos básicos como estatísticas descritivas e distribuições de probabilidade serão utilizados para análise e interpretação de bases de dados reais. Essa disciplina não assume conhecimento prévio em programação. Ao final, espera-se que o aluno tenha desenvolvido habilidades computacionais úteis para análise de dados.					
OBJETIVOS ESPECÍFICOS		CONTEÚDO PROGRAMÁTICO	CARGA HORÁ	ESTRATÉGIA METODOLÓGICA	
DATA	TEMA DA AULA			TÉCNICAS DIDÁTICO	RECURSOS

06/03/2020	Processo de análise de dados	Introdução. Visão geral sobre o processo de coleta, importação, tratamento e limpeza, transformação, visualização, modelagem, e comunicação de dados.	2h	Exposição oral	Laboratório
13/03/2020	Pesquisa reproduzível	Introdução ao R e ao RStudio. Controle de versão com git e colaboração com GitHub. Relatórios dinâmicos. Uso do pacote rmarkdown. Organização de projetos de análise de dados.	2h	Aula prática	Laboratório
22/05/2020	Pesquisa reproduzível (cont.)	Introdução ao R e ao RStudio. Controle de versão com git e colaboração com GitHub. Relatórios dinâmicos. Uso do pacote rmarkdown. Organização de projetos de análise de dados.	2h	Videoaula (Disponibilizada em 18/05)	Rstudio Cloud, GitHub, Youtube, Moodle
29/05/2020	Introdução à programação	A linguagem de programação R: comandos básicos, estruturas de dados.	2h	Videoaula (Disponibilizada em 25/05)	Rstudio Cloud, GitHub, Youtube, Moodle

05/06/2020	Introdução à programação (cont.)	A linguagem de programação R: comandos básicos, estruturas de dados.	2h	Videoaula (Disponibilizada em 01/06)	Rstudio Cloud, GitHub, Youtube, Moodle
19/06/2020	Manipulação de dados	Dados estruturados: dados tabulares; dados relacionais; <i>tidy data</i> . Limpeza e transformação de dados. Operador <code>`%>%`</code> (<i>pipe</i>). Uso do pacote <i>tidyr</i> . Uso do pacote <i>dplyr</i> .	2h	Videoaula (Disponibilizada em 08/06)	Rstudio Cloud, GitHub, Youtube, Moodle
			2h	Videoaula (Disponibilizada em 15/06)	Rstudio Cloud, GitHub, Youtube, Moodle
26/06/2020	Manipulação de dados (cont.)	Dados estruturados: dados tabulares; dados relacionais; <i>tidy data</i> . Limpeza e transformação de dados. Operador <code>`%>%`</code> (<i>pipe</i>). Uso do pacote <i>tidyr</i> . Uso do pacote <i>dplyr</i> .	2h	Videoaula (Disponibilizada em 22/06)	Rstudio Cloud, GitHub, Youtube, Moodle
03/07/2020	Manipulação de dados (cont.)	Dados estruturados: dados tabulares; dados relacionais; <i>tidy data</i> . Limpeza e transformação de dados. Operador <code>`%>%`</code> (<i>pipe</i>). Uso do pacote <i>tidyr</i> . Uso do pacote <i>dplyr</i> .	2h	Videoaula (Disponibilizada em 29/06)	Rstudio Cloud, GitHub, Youtube, Moodle
10/07/2020	Visualização de dados	Gramática dos gráficos. Gráficos em camadas. Principais tipos de visualizações. Uso do pacote <i>ggplot2</i> .	2h	Videoaula (Disponibilizada em 06/07)	Rstudio Cloud, GitHub, Youtube, Moodle
17/07/2020	Visualização de dados (cont.)	Gramática dos gráficos. Gráficos em camadas. Principais tipos de visualizações. Uso do pacote <i>ggplot2</i> .	2h	Videoaula (Disponibilizada em 13/07)	Rstudio Cloud, GitHub, Youtube, Moodle
24/07/2020	Visualização de dados (cont.)	Gramática dos gráficos. Gráficos em camadas. Principais tipos de visualizações. Uso do pacote <i>ggplot2</i> .	2h	Videoaula (Disponibilizada em 20/07)	Rstudio Cloud, GitHub, Youtube, Moodle
31/07/2020	Visualização de dados (cont.)	Gramática dos gráficos. Gráficos em camadas. Principais tipos de visualizações. Uso do pacote <i>ggplot2</i> .	2h	Videoaula (Disponibilizada em 27/07)	Rstudio Cloud, GitHub, Youtube, Moodle
07/08/2020	Introdução à programação funcional	Funções e programação funcional. Uso do pacote <i>purrr</i> .	4h	Videoaula (Disponibilizada em 03/08)	Rstudio Cloud, GitHub, Youtube, Moodle

07/08/2020	ATIVIDADE SUBSTITUTIVA	Matéria toda semestre		Avaliação	
14/08/2020	EXAME ESPECIAL	Matéria toda semestre		Avaliação	

AVALIAÇÃO:

A avaliação será composta de:

- 1) Reprodução, utilizando o ggplot2, de uma visualização de dados constante de um artigo científico ou de uma notícia de jornalismo de dados (30pts)
- 2) Revisão por pares de 1 (5pts)
- 3) Criação de um relatório dinâmico contendo análise exploratória de uma base de dados (30pts)
- 4) Revisão por pares de 3 (5pts)
- 5) Exercícios práticos em sala de aula (30pts)

ATIVIDADES PRÁTICAS SUPERVISIONADAS:

As atividades práticas supervisionadas consistem em:

- Trabalhos individuais
 - Revisão por pares
- Trabalhos em grupo
 - Reprodução de uma visualização de dados
 - Criação de análise exploratória de uma base de dados

BIBLIOGRAFIA BÁSICA:

Wickham, Hadley, and Garrett Golemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 1 edition. Sebastopol, CA: O'Reilly Media. Disponível em: <<https://r4ds.had.co.nz/>>.

Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. 1 edition. Princeton, NJ: Princeton University Press. Versão preliminar disponível em: <<http://socviz.co/>>

Wilson, Greg, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven H. D. Haddock, et al. 2014. “Best Practices for Scientific Computing.” *PLoS Biol* 12 (1): e1001745. <https://doi.org/10.1371/journal.pbio.1001745>. Disponível em < <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745> >

BIBLIOGRAFIA COMPLEMENTAR:

Gentzkow, Matthew, and Jesse Shapiro. “Code and Data for the Social Sciences: A Practitioner’s Guide,” 2014. Disponível em:

<<https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf>>

Bryan, Jennifer. 2018. “Excuse Me, Do You Have a Moment to Talk About Version Control?” *The American Statistician* 72 (1): 20–27.

<https://doi.org/10.1080/00031305.2017.1399928>. Versão preliminar disponível em: <<https://peerj.com/preprints/3159/>>

Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. “Packaging Data Analytical Work Reproducibly Using R (and Friends).” *The American Statistician* 72 (1): 80–88.

<https://doi.org/10.1080/00031305.2017.1375986>. Versão preliminar disponível em <<https://peerj.com/preprints/3192/>>

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016 edition. New York, NY: Springer. Disponível em: <<https://ggplot2-book.org/>>

Broman, Karl W., and Kara H. Woo. 2018. “Data Organization in Spreadsheets.” *The American Statistician* 72 (1): 2–10. <https://doi.org/10.1080/00031305.2017.1375989>.

Versão preliminar disponível em <<https://peerj.com/preprints/3183/>>

ASSINATURA DO PROFESSOR:	DATA:	ASSINATURA DO COORDENADOR DO CURSO:	DATA
--------------------------	-------	-------------------------------------	------