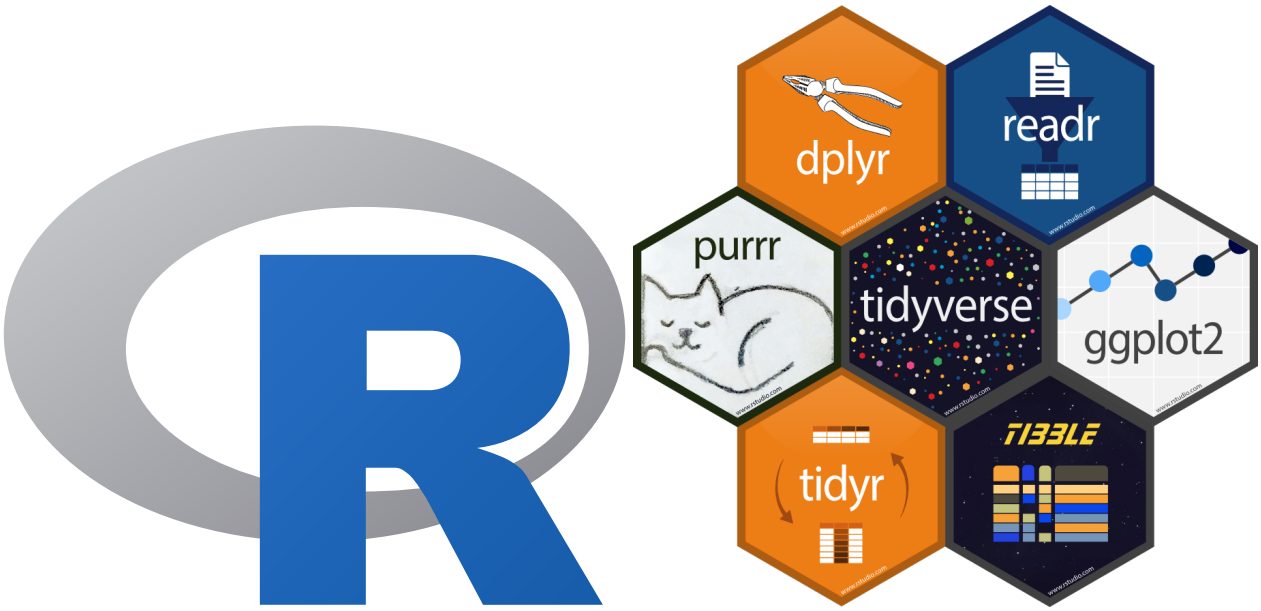
Ciência de dados com R e o Tidyverse

Laboratório de Análise de Dados em Demografia - LADDEM

24 a 28 de Maio de 2021

E-mail: viniciusmaia108@gmail.com
Horários para dúvidas: 10:00-15:00
Endereço: LADDEM Virtual

Web: zlkrvsm.github.io/Oficina-de-R/
Horário de aulas: 19:00-21:30
Sala: *online*



Introdução

Este curso introduzirá uma outra maneira de organizar seu trabalho no R, através da suite de pacotes "tidyverse". Não se trata de reaprender a utilizar o R, ou obter resultados diferentes ou melhores, mas de tornar o código mais legível para outros usuários, mais prático para o analista e facilitar a integração entre pacotes e ferramentas existentes no R. O tidyverse é uma obra frequentemente descrita como "opinionated", que significa dizer que foram feitas escolhas durante seu desenvolvimento que podem desagradar usuários de longa data, mas nossa expectativa é que todos possam encontrar ferramentas úteis nesses pacotes, independente de se eles consideram vantajoso mudar substancialmente sua forma de trabalhar no software. O núcleo do tidyverse são 8 pacotes utilizados para tarefas cotidianas de análise de dados, são eles `ggplot2`, `dplyr`, `tidyr`, `readr`, `purrr`, `tibble`, `stringr` e `forcats`.

O curso é **gratuito** e organizado com apoio do **Núcleo de Estudos de População "Elza Berquó"**. Os participantes receberão um certificado referente ao curso.

Público-alvo

Alunos, pesquisadores, docentes, gestores e sociedade em geral. O participante deve ter conhecimentos básicos de R para poder aproveitar as atividades do curso. Conceitos e utilização de funções comuns, data.frames, vetores, variáveis e seus tipos são necessários, embora ressalto que faremos atividades simples e didáticas e ninguém precisa sentir-se intimidado pelo desafio. Se você está em dúvida se está capacitado para participar, a resposta provavelmente é "sim".

Objetivos

1. Introduzir usuários do R a uma forma distinta de trabalhar com bancos de dados, visualizações e iterações.
2. Capacitar antigos usuários para compreender e utilizar novos paradigmas da linguagem.
3. Oferecer uma nova perspectiva para os programas/rotinas mais utilizados.
4. Fomentar a aplicação do R na análise de resultados e como ferramenta de gestão.

Organização

Professor: Vinícius de Souza Maia (viniciusmaia108@gmail.com)
Instrutora: Natália Martins Arruda (natimarruda@gmail.com)

Vinícius e Natália são pesquisadores no LADDEM/NEPO e no PROMALARIA - "Ciência de dados aplicada em informações Epidemiológicas e Demográficas como estratégia para simulação e vigilância de casos de malária na Amazônia brasileira", coordenado pela Profa. Luciana Correia Alves. O projeto é uma colaboração entre pesquisadores da Unicamp, da Unesp e do IFSP, com apoio do CNPq e da Fundação Bill e Melinda Gates.

Horários e datas

De Segunda, 24 de Maio de 2021 à Sexta, 28 de Maio de 2021, das 19h às 21h30, com um intervalo de 10 minutos.

Aulas

Segunda, 24 de Maio de 2021:

Readr, Tibbles e Tidy data, um outro olhar sobre o data.frame.

Terça, 25 de Maio de 2021:

Stringr, Forcats e Dplyr, variáveis e operações com banco de dados.

Quarta, 26 de Maio de 2021:

Ggplot2: Visualizando dados com a gramática dos gráficos.

Quinta, 27 de Maio de 2021:

Purrr: Uma breve introdução a programação funcional.

Sexta, 28 de Maio de 2021:

Mini-projeto: colocando a mão na massa com o tidyverse.

Material para leitura

Damiani, Athos, Beatriz Milz, Caio Lente, Daniel Falbel, Fernando Correa, Julio Trecenti, Nicole Luduvice, and William Amorim (2021). *Ciência de dados em R*. Em português. Curso-R. https://livro.curso-r.com/index.html/.

RStudio Inc. *Apply functions Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/purrr.pdf.

RStudio Inc. *Data Import Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/data-import.pdf.

RStudio Inc. *Data Transformation Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/data-transformation.pdf.

RStudio Inc. *Data Visualization Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/data-visualization-2.1.pdf.

RStudio Inc. *Factors Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/factors.pdf.

RStudio Inc. *String manipulation Cheatsheet*. Em inglês. https://github.com/rstudio/cheatsheets/blob/master/strings.pdf.

Wickham, Hadley and Garrett Grolemund (2017). *R for Data Science*. Em inglês. O'Reilly. https://r4ds.had.co.nz/.