

**Universidade Federal do Rio Grande do Sul**  
**Faculdade de Medicina**  
**Programa de Pós-Graduação em Cardiologia e Ciências Cardiovasculares**

**Plano de Ensino e Ementa**

**CAR00110**

*Bioestatística Aplicada em Software R*

Prof. Dr. Lucas Helal

Epidemiologia e Bioestatística Aplicadas em Ambiente R

<b>CAR00110 - Bioestatística Aplicada em Software R (2 créditos)</b>	<b>3</b>
<b>Detalhamento</b>	<b>4</b>
<b>Aviso Prévio</b>	<b>4</b>
<b>No GitHub e Moodle disponibilizados pelo professor, e em dinâmica prévia ao início da disciplina, tutoriais estarão disponibilizados para instalação dos softwares e dependências necessárias, assim como instruções dadas pelo professor em tempo real.</b>	<b>4</b>
Monitora da Disciplina	4
Elegibilidade e Vagas	4
Filosofia da Disciplina	4
Justificativa da Disciplina	4
Objetivo Primário	5
Objetivos Específicos	6
Estrutura do Curso	9
O que é o R?	9
Por que devo usar o R	10
<b>Bibliografia</b>	<b>12</b>
Bibliografia de Referência/Obrigatória	12
Bibliografia Secundária/de Apoio	13
Lista inicial de pacotes básicos em linguagem R para necessários à disciplina	15
Lista de links úteis	16
Para os mais curiosos	17

## CAR00110 - Bioestatística Aplicada em Software R (2 créditos)

*Epidemiologia e Bioestatística Aplicadas em Ambiente R*

2023/2 - segundo termo

Nível: pós-graduação (mestrado acadêmico e profissional, doutorado acadêmico)

---

**Instituição de Ensino Superior:** Universidade Federal do Rio Grande do Sul

**Escola:** Faculdade de Medicina

**Programa de Pós-Graduação:** Programa de Pós-Graduação em Cardiologia e Ciências Cardiovasculares

**Identificação da Disciplina:** CAR00110

**Links Fixos:**

*Sala de Aula Virtual:* <https://meet.google.com/gkd-uzfg-efd>

*Moodle:* [www.moodle.ufrgs.br](http://www.moodle.ufrgs.br)

*Fórum da Disciplina em Slack:* [https://is.gd/CAR00110\\_slack](https://is.gd/CAR00110_slack)

► **Tutoriais de instalação, materiais de apoio e FAQ:** <https://github.com/lhelal1/CAR00110>

**Carga horária e créditos:** 30 horas/2 créditos

**Regente:** Prof. Dr. Lucas Helal

Contatos: [lucas.helal@ufrgs.br](mailto:lucas.helal@ufrgs.br) | [lhelal@hcpa.edu.br](mailto:lhelal@hcpa.edu.br)

GitHub: <https://github.com/lhelal1>

Calendário público para disponibilidade de atendimentos: [https://is.gd/CAR00110\\_cal](https://is.gd/CAR00110_cal)

**Agenda da Disciplina:** segundas-feiras (19:00-21:00) e sextas-feiras (18:00-21:00)

**Início:** 11-08-2023

**Término:** 29-09-2023

**Pré-requisitos obrigatórios:** NA.

**Pré-requisitos desejáveis:** habilidade de leitura e escrita em língua inglesa.

**Método de Avaliação:** por critérios de participação em disciplina.

**Método de Recuperação de Nota:** a decidir com o professor.

**Aprovação:** este Plano de Disciplina e Ementa foi aprovado pela coordenadora do PPG Cardiologia no dia XX/YY/ZZZZ - **Prof<sup>a</sup>. Dr<sup>a</sup>. Carisi A. Polanczyk** (Coordenadora do PPG Cardiologia, UFRGS).

## Detalhamento

### Aviso Prévio

**No GitHub e Moodle disponibilizados pelo professor, e em dinâmica prévia ao início da disciplina, tutoriais estarão disponibilizados para instalação dos softwares e dependências necessárias, assim como instruções dadas pelo professor em tempo real.**

### Monitora da Disciplina

Rafaela Ribeiro-Lucas  
Candidata de Mestrado, PPG Cardiologia  
Universidade Federal do Rio Grande do Sul  
[rglucas@hcpa.edu.br](mailto:rglucas@hcpa.edu.br)

### Elegibilidade e Vagas

São elegíveis para esta disciplina alunos de pós-graduação *stricto sensu* da Universidade Federal do Rio Grande do Sul ou de outras Instituições de Ensino Superior que estejam regularmente matriculados, não restritos ao PPG Cardiologia.

Serão ofertas 20 (vinte) vagas para alunos formais e, alunos ouvintes, deverão contactar o regente da disciplina para verificar a adequabilidade da solicitação.

### Filosofia da Disciplina

Esta disciplina foi idealizada para instrumentalização e exposição de discentes de pós-graduação da área da saúde ao ambiente computacional em R. Com o uso de fundamentos de Epidemiologia e Bioestatística, pretende-se tornar o processo pedagógico aplicado e reduzir o nível de abstração.

Também, a disciplina correrá com o uso conjunto de tecnologias para comunicação em ambientes acadêmicos; versionamento de materiais; e com leve exposição ao pensamento lógico-computacional, dado que o R é uma linguagem de programação. Por fim, os alunos serão expostos à temas diretamente ligados à evolução de suas dissertações, teses, artigos científicos ou quaisquer manuscritos científicos que exijam a apresentação de resultados de forma didática e o mais compreensível possível para o ouvinte/leitor. Isto será feito com a exposição à criação de gráficos, figuras, tabelas, boas práticas computacionais e em pesquisa, e uso de documentos dinâmicos (R Markdown).

### Justificativa da Disciplina

Há alguns anos, a linguagem R ganhou espaço maciço entre pesquisadores e cientistas, frente a outros pacotes estatísticos tradicionais, especialmente os que lidam com alto volume de dados (e.g., bioinformática, finanças, ecologia). Os principais motivos que levaram à tal transição

da comunidade científica à linguagem R para manuseio de dados e procedimentos estatísticos poderiam ser: a) altíssima capacidade de adaptação para qualquer necessidade relacionada à pesquisa, por meio da criação de pacotes/bibliotecas pela comunidade aberta; b) alto poder computacional mesmo em computadores domésticos; c) interligação com a cultura *open-source* - ou seja, o aplicativo é de código aberto (qualquer pessoa pode propor melhorias) e **é gratuito por filosofia de existir** (*se fosse cobrado, provavelmente custaria mais que os pacotes estatísticos disponíveis no mercado*); d) por fim, robusta comunidade de usuários ativos na rede, participantes em fóruns e similares, assim como extensa documentação dos pacotes estatísticos disponíveis (com exemplos), tornando a resolução de problemas e implementação de novos pacotes algo relativamente fácil de lidar.

Eu poderia listar muitos outros, *técnicos e filosóficos*, mas esses por si só já são mais que suficientes para **justificar o porquê de um curso em R para alunos de pós-graduação na área da saúde - área que mais modernamente vem fazendo o movimento e tornando o conhecimento da linguagem cada vez mais imperativa.**

## Objetivo Primário

Esta disciplina tem como objetivo primário introduzir a estatística em linguagem computacional para alunos de pós-graduação na área da saúde, com nenhum conhecimento prévio de linguagem R ou qualquer outro tipo de linguagem de programação, utilizando como base problemas reais da pesquisa clínica, da bioestatística e da epidemiologia.

**Tabela 1.** Cronograma da Disciplina.

Data	Tópico
11/ago/2023	Aula 1 - Utilizando a linguagem R e seus documentos
14/ago/2023	Aula 2 - Sintaxe Básica e Fundamentos de Bioestatística
18/ago/2023	Aula 3 - Manipulação de Variáveis I
21/ago/2023	Aula 4 - Manipulação de Variáveis II
25/ago/2023	Aula 5 - Estatística Descritiva em Linguagem R
01/set/2023	Aula 6 - Estatística Descritiva em Linguagem R II
04/set/2023	Aula 7 - Estatística Descritiva e Inferencial em Linguagem R III
15/set/2023	Aula 8 - Gráficos, Tabelas e Figuras em Linguagem R
22/set/2023	Aula 9 - Gráficos, Tabelas e Figuras em Linguagem R II
29/set/2023	Aula 10 - Utilizando a Linguagem R para resolver problemas comuns que exigem soluções difíceis: randomização, anonimização de dados, simulações de cenários, cálculo amostral

**Nota:** detalhes de cada tópico abaixo.

## Objetivos Específicos

Nesta disciplina, pretendemos que você a termine com as seguintes habilidades (e mais algumas não listadas):

R Básico	Conteúdo
RB1	Eu sou capaz de definir o que a linguagem R é.
RB2	Eu sou capaz de definir o que o R Studio é.
RB3	Eu sei instalar um pacote em R, criar um script em R, criar um documento R Markdown, criar pastas no meu computador pelo próprio ambiente de desenvolvimento etc.
RB4	Eu consigo achar a função <code>help()</code> referente aos pacotes do R.
RB5	Eu saberei comentar corretamente meu código em R.
RB6	Eu saberei usar o padrão de linguagem mais adequado para boas práticas em manejo de dados.
RB7	Eu posso definir o que é um objeto em linguagem R.
RB8	Eu sei atribuir argumentos, funções, etc. aos objetos.
RB9	Eu sei atribuir valor a um objeto em linguagem R.
RB10	Eu sei dar o comando ao ambiente de desenvolvimento para que meus resultados em linguagem em R apareçam somente no console.
RB11	Eu sei definir o que é um ambiente de desenvolvimento e o que é um console.
RB12	Eu sei identificar os tipos de dados que operam em R.
RB13	Eu sei definir um elemento em linguagem R.
RB14	Eu sei definir e criar um vetor em R.
R Dados	Conteúdo
RD1	Eu consigo ler um banco de dados no R Studio.
RD2	Eu consigo criar bancos de dados a partir de um banco de dados previamente importado.
RD3	Eu sei definir e utilizar operadores Booleanos (YES, NO, TRUE, FALSE, OR, AND, IF...).
RD4	Eu sei definir o que é um data frame.
RD5	Eu sei criar um data frame com dados próprios.
RD6	Eu sei extrair uma coluna/variável de qualquer tipo de data frame.
RD7	Eu sei atribuir nomes às variáveis em data frames em R.
RD8	Eu sei lidar com dados faltantes ( <i>missing data</i> ) em R.

RD9	Eu sei definir uma matriz e uma lista de dados.
RD10	Eu sei extrair elementos de matrizes e listas de dados.
RD11	Eu sei converter um data frame em uma lista.
RD12	Eu sei trocar elementos dentro de um objeto.
RD13	Eu sei definir e criar um fator
<b>R Dados Manipulação</b>	<b>Conteúdo</b>
RDM1	Eu sei identificar duplicatas de observações.
RDM2	Eu sei como checar números frequentes no R.
RDM3	Eu sei utilizar funções para avançar em processos de manejo de dados com pacotes nativos do R: <code>subset()</code> , <code>apply()</code> , <code>sort()</code> , <code>order()</code> , <code>cbind()</code> , <code>rbind()</code> , <code>merge()</code> , <code>str()</code> etc.
RDM4	Eu fui apresentado ao pacote <code>tidyverse()</code> e suas dependências para manejo avançado de banco de dados.
RDM5	Eu sei extrair componentes de números em formato DATA (i.e., data calendário)
RDM6	Eu sei adicionar uma nova variável em um banco de dados em R; eu sei retirar uma variável de um banco de dados em R.
RDM7	Eu sei criar uma variável derivada (variável produto de variáveis pré-existentis) em R.
RDM8	Eu sei manipular classes de dados em R - transformar dados contínuos para categóricos; criar variáveis binárias, etc.
RDM9	Eu sei visualizar um banco de dados.
RDM10	Eu sei filtrar um banco de dados.
RDM11	Eu sei salvar um banco de dados em qualquer extensão utilizando a linguagem R (para ser lido em SPSS, Stata, SAS, etc.)
<b>R Boas Práticas</b>	<b>Conteúdo</b>
RBP1	Eu saberei dizer por que um código limpo é importante.
RBP2	Eu saberei dizer por que um código robusto é importante - e saberei definir o que é um código robusto.
RBP3	Eu saberei utilizar a taxonomia usual em códigos, ainda que esteja fazendo Estatística.
RBP4	Eu saberei a importância de um código reprodutível.
RBP5	Eu saberei a importância de um código que seja interpretável no longo termo, quando eu revisitá-lo depois de anos sem lê-lo.
RBP6	Eu saberei a importância de ter (quase) certeza que conseguirei reproduzir algo que produzi tempos atrás; ou que terceiros possam fazê-lo sem precisar me consultar.
<b>R Estatística</b>	<b>Conteúdo</b>
RE1	Eu saberei diferenciar probabilidade, densidade e aleatoriedade, assim como saberei utilizar funções em linguagem R para gerar distribuições respectivas.
RE2	Eu saberei definir o que é uma lista pseudoaleatorizada.
RE3	Eu saberei amostrar aleatoriamente elementos de um dado objeto.

RE4	Eu sei calcular estatísticas descritivas tanto para dados contínuos quanto para dados categóricos.
RE5	Eu sei apresentar as estatísticas descritivas tanto forma de gráfico quanto tabelas em linguagem R.
RE6	Eu sei operar funções para inferência estatística de dados contínuos - modelos lineares.
RE7	Eu sei operar funções para inferência estatística de dados contínuos/categóricos que necessitam de parametrização - modelos lineares generalizados.
RE8	Eu sei estimar incidências e razões de incidências (riscos relativos) por diferentes tipos de modelos - Poisson, binomial negativa, etc.
RE9	Eu sei estimar razão de chances por meio de regressão logística;
RE10	Eu sei estimar intervalos de confiança em qualquer nível arbitrário.
RE11	Eu sei estimar o tempo até o evento de situações exemplo.
RE12	Eu sei testar pressupostos de normalidade e resíduos para dados contínuos.
RE13	Eu sei utilizar e interpretar testes de hipótese de menor complexidade em R, como o teste t de Student, a análise de variância, etc.
RE14	Eu sei calcular amostra para diferentes tipos de delineamentos de estudo, e também sei como encontrar material para outros tipos de estudos.
RE15	Eu sei gerar listas de randomização de diferentes maneiras em ambiente R.
<b>R Plots e Documentos</b>	<b>Conteúdo</b>
RP1	Eu sei plotar um resultado com funções nativas em R.
RP2	Eu sei plotar um ou mais resultados com funções nativas do R.
RP3	Eu sei plotar um ou mais resultados utilizando o pacote ggplot2( ).
RP4	Eu sei customizar gráficos de acordo com meus critérios.
RP5	Eu sei como gerar gráficos que podem compor minha dissertação, tese ou artigos.
RP6	Eu sei gerar relatórios em R Markdown.
RP7	Eu sei gerar relatórios em R Markdown em formato .png e também sei recorrer a recursos mínimos de HTML se necessário.
RP8	Eu sei gerar relatórios em R Markdown em formato .pdf e também sei recorrer à funções básicas de recursos alfabéticos em LaTeX no mesmo documento.
RP9	Eu sei gerar relatórios em R Markdown em formato .docx e depois editá-lo como quiser no meu editor de texto externo.



## Estrutura do Curso

Esta disciplina será ofertada de forma inteiramente remota, com as aulas acontecendo sempre por meio do mesmo link disponibilizado.

Como recursos pedagógicos, faremos uso de: a) sala de aula virtual; b) ambiente em linguagem R; c) ambiente em linguagem R em nuvem, de forma colaborativa; d) uso de diapositivos; e) leituras obrigatórias; f) leituras optativas; g) tarefas obrigatórias; h) lista de exercício optativas; i) códigos e scripts compartilhados; j) canal para interação entre os integrantes da disciplina; k) repositório online administrado pelo professor, de livre acesso, com materiais adicionais.

O nosso portal de acesso, requisição de tarefas, entrega de tarefas dentre outros é oficialmente o **Moodle UFRGS**. De forma concomitante, mas não oficial, você poderá encontrar materiais no repositório <https://github.com/lhelal1/CAR00110>, não sendo obrigatória a visita ao GitHub caso você não queira/não se sinta à vontade. Haverá um canal de comunicação da disciplina no **Slack**, que atuará como fórum/portal de conversas/tira-dúvidas (*muito comum no ecossistema de quem trabalha com tecnologia*),

As aulas serão ministradas de **forma síncrona** e os critérios de avaliação constam neste documento.

## O que é o R?

Nós teremos uma aula só disso, mas antes de você se matricular/engajar na disciplina, eu tenho certeza que iria gostar de saber algumas curiosidades sobre o que está por vir.

O *R Foundation* é uma organização sem fins lucrativos que trabalha para o interesse público. Originalmente, foi criada por dois cientistas de dados da University of Auckland, NZ - Ross Ihaka e Robert Gentleman. Existe uma linguagem de programação chamada S, voltada à estatística. O R foi inspirado na linguagem S e sua interface S-PLUS, assim como o seu nome (jogo de palavras) combinado com as iniciais dos criadores. Para quem acha que o R é um pacote estatístico, ledo engano.

Trata-se de uma linguagem de programação, como Java, JavaScript, C, C++, Python e tantas outras (citei algumas que podem ser de conhecimento de vocês já). Isso traz algumas consequências pedagógicas importantes: necessidade de raciocínio lógico na implementação dos comandos estatísticos; sua estrutura de organização - tópico mais avançado, que não será abordado neste curso), não dispensando o uso de funções, loops, laços e outros algoritmos comuns na programação.

Porém o R tem uma particularidade - é uma linguagem desenvolvida para fins estatísticos (descritivos e inferências), que junta o melhor que a liberdade computacional pode nos dar com a sua finalidade estatística. É claro que, das consequências pedagógicas, por vezes acaba se tratando de barreira para alguém adotar a linguagem como a sua “de referência”; ou vem imbuído

de preconceitos em relação ao suposto nível de complexidade, que é rapidamente *down regulated* depois de algumas poucas aulas.

Tecnicamente, é provável que a principal característica da linguagem R seja o fato de ser o que chamamos de *linguagem orientada a objetos* - ou seja, os códigos e linhas de comando estatísticas que criamos são guardadas em objetos com nomes arbitrados por nós e então assim os manuseamos, o que encurta radicalmente o trabalho. Também, é uma linguagem que é chamada, no campo da programação, de *alto nível* - isto é, se assemelha bastante à linguagem humana, reduzindo substancialmente o nível de abstração frente à outras linguagens; e, por fim, é também o que chamamos de uma *linguagem multi-paradigma* - paradigma de programação é como se discrimina uma linguagem de programação em relação ao seu modo de funcionamento e estruturação:

a) é procedural (computa procedimentos em cadeia); b) é orientado a objetos; c) é funcional - isto é, entende os comandos por meio de funções explícitas declaradas pela pessoa; d) reflexiva (tem a capacidade de se auto corrigir, re fatorar, etc.); e) é imperativa (como na língua humana, quando declara-se algo em tom imperativo, há a noção de ordem - ou seja, em R, frequentemente ordena-se ao à linguagem que se execute o comando; f) interpreta vetores, matrizes e sequências - em vez de interpretar um objeto por vez, pode interpretar vários objetos de uma vez só na forma de conjunto.

## Por que devo usar o R

Jamais ousaria dizer que você deve usar algo em detrimento de outro. Os pacotes estatísticos de alto nível disponíveis no mercado são todos igualmente excelentes. O que eu poderia dizer é que, caso você queira usar o R, seria interessante você saber de mais algumas coisas: a) é um programa extremamente leve para a maioria dos procedimentos estatísticos corriqueiros em saúde; b) a comunidade científica da área da saúde, básica e aplicada, já fez o mesmo movimento que outras áreas preceptoras, reconhecendo as vantagens de conduzir procedimentos estatísticos em linguagem R; c) é possível fazer pesquisa clínica e patrocinada utilizando linguagem R - recentemente, o *Food and Drug Administration* (US FDA), agência regulatória dos Estados Unidos (tal qual a ANVISA no Brasil) publicou um extenso memorando com a declaração de anuência para aceite nos pedidos de aprovação de medicamentos e devices com utilização de linguagem R, em todo o ciclo de vida do projeto - <https://www.r-project.org/doc/R-FDA.pdf>.

Uma fama atribuída também ao R na pesquisa clínica é a sua capacidade gráfica e interlocução com outras linguagens de programação. Não é incomum nos depararmos com belíssimas apresentações de resultados realizadas por meio do pacote *ggplot2* (que será abordado na disciplina); ou a geração de documentos dinâmicos, que integram texto e resultados

- os famosos *R Markdowns*, facilitando a vida do pesquisador na elaboração de artigos, relatórios ou mesmo material para aulas (que também será conteúdo deste curso).

Porém, em minha concepção, a maior vantagem que se pode obter ao usar a linguagem R é a evolução em raciocínio lógico e capacidade analítica que se acarreta enquanto humanos de carne e osso, fora da rotina de apertar botões e interpretar *outputs*. Comigo, ocorreu a experiência de mudar minha forma de enxergar os problemas; a minha forma de encarar o método científico - no sentido *stricto sensu*, em sua forma mais conceitual possível. Caso você venha a se engajar com a linguagem, que exige bastante de você no dia a dia, eu fortemente espero que o mesmo também aconteça. Como professor e como ser humano.

## Bibliografia

### Bibliografia de Referência/Obrigatória

1. R CRAN. **R Package Manual**. 2023. <https://cran.r-project.org/>
2. Git —distributed-even-if-your-workflow-isnt. **Git Core Team**. 2023: <https://git-scm.com/>
3. Golemund G; Wickham H. **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**. 2023. 2<sup>nd</sup> Edition, 1<sup>st</sup> release, Ed. O'Reilly, USA: <https://r4ds.hadley.nz/>  
Copyright: CC BY Non-Commercial and Non-Derivs 3.0 (CC BY-NC-ND 3.0 - [open access book](#)). Garret Golemund, Hadley Wickham. All rights reserved. ISBN: 978-1-491-91039-9
4. Brandvain Y. **Applied Biostats**. 2023. Open Source.  
<https://bookdown.org/ybrandvain/Applied-Biostats/#>
5. Neal P, Cavey D. **Foundations of Statistics**. R Bookdown. 2022.  
[https://bookdown.org/peter\\_neal/math4081\\_notes/](https://bookdown.org/peter_neal/math4081_notes/)
6. Damiani A, Milz B, Lente C et al. **Ciência de Dados em R**. Instituto de Matemática e Estatística da Universidade de São (IME-USP) e R-Ladies. 2022: <https://livro.curso-r.com/>.  
Copyright: CC BY Universal. De construção e revisão aberta à comunidade.
7. Wickham H, Chang W, Henry L et al. **ggplot2: The Grammar of Graphics**. 2023.  
<https://ggplot2.tidyverse.org/>
8. Wickham H. **tidyverse R Package**. 2023. <https://tidyverse.tidyverse.org/>

## Bibliografia Secundária/de Apoio

A bibliografia secundária/de apoio se encontra no GitHub do professor: <https://github.com/lhelal1/CAR00110/>

### Identificadores:

CAR00110-REF1

1. Evans MJ, Rosenthal JS. Probability and Statistics: The Science of Uncertainty. 2009.

CAR00110-REF2

2. Horton NJ, Pruijm R, Kaplan DT. A Student's Guide to R. 2015.

CAR00110-REF3

3. Vandemeulebroecke M. The R language - a short companion. Version 1.2, 2023.

CAR00110-REF4

4. De Jonge E, van der Loo M. An introduction to data cleaning with R. Statistics Netherlands. 2014.

CAR00110-REF5

5. Greg W, Bryan J, Cranston K et al. Good enough practices in scientific computing. **PLoS Comput Biol**, 2017;13(6), e1005510.

CAR00110-REF6

6. Alter G, Gonzalez R. Responsible practices for data sharing. **Am Psych**, 2018;73(2), 146.

CAR00110-REF7

7. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. **Gastr Hep Bed to Bench**, 2013. 6(1), 14.

CAR00110-REF8

8. Uschner D, Schindler D, Hilgers RD et al. randomizeR: an R package for the assessment and implementation of randomization in clinical trials. **J Stat Software**, 2018;85:1-22.

CAR00110-REF9

9. Duthie B. **Creating simulated data sets in R**. R Bookdown, 2022.

CAR00110-REF10

10. Wendt CJ, Anderson GB. Ten simple rules for finding and selecting R packages. **PLoS Compute Biol**, 2022;18(3): e1009883.

CAR00110-REF11

11. Abd-El-Barr M, Rewini-El H. **Fundamentals of Computer Organization and Architecture**. Springer, 2005.

CAR0110-RE12

12. Ramachandram P. Bayesian estimation and inference. **Mathematical Statistics with Application in R**. Elsevier, 2021.

## Lista inicial de pacotes básicos em linguagem R para necessários à disciplina

1. `readr()` - lê pacotes em R (pode vir instalado nativamente)
2. `pacman()` - instala e instala pacotes de forma conjunta, e podendo executar de forma simultânea a mais pacotes;
3. `haven()` - lê e salva bancos de dados em formatos extensíveis, como `.dta` (Stata), `.sas` (SAS), `.sav` (SPSS) etc.
4. `foreign()` - idem.
5. `rmarkdown()` - transforma scripts em documentos dinâmicos;
6. `knitr()` - dá suporte à execução do `rmarkdown()`;
7. `stringr()` - para manipulação de strings (dados em formato de palavras);
8. `tidyverse()` - pacote robusto com múltiplas funções, desde manipulação de bancos de dados à criação de gráficos
9. `dplyr()` - embutido no `tidyverse()`, para a manipulação de banco de dados
10. `ggplot2()` - criação de gráficos e figuras
11. `kableExtra()` - criação de tabelas automáticas
12. `lubridate()` - manipulação de variáveis em formato data/calendário
13. `gapminder()` - possui bancos de dados para manuseio
14. `tableone()` - criação de tabelas descritivas
15. `gt()` - idem

Os pacotes são baixados e instalados no próprio ambiente de desenvolvimento R, por meio das funções `install.packages("nome do pacote")` e, após, `library(nome do pacote - uso opcional de aspas)`. É importantíssimo ressaltar que a linguagem R faz distinção entre letras maiúsculas e minúsculas, espaços faltantes, parênteses faltantes, etc (case-sensitive). Por exemplo, para instalar o pacote pela função `install.packages()`, **é necessário** que o nome do pacote esteja **entre aspas**.

Pacotes **específicos** serão apresentados, baixados e carregados em aula, de acordo com cada função que queiramos cumprir.

## Lista de links úteis

- O que é o R - <https://rstudio.github.io/r-manuals/r-lang/Introduction.html>
- Quais universidades ao redor do mundo adotam a linguagem R como parte integrante do currículo - <https://www.ufrgs.br/wiki-r/index.php?title=Universidades>
- Como instalar o R - [https://www.ufrgs.br/wiki-r/index.php?title=Instalação do R](https://www.ufrgs.br/wiki-r/index.php?title=Instalação_do_R)
- Como instalar o R Studio - [https://www.ufrgs.br/wiki-r/index.php?title=Instalação do RStudio](https://www.ufrgs.br/wiki-r/index.php?title=Instalação_do_RStudio)
- Introdução às Boas Práticas em Programação - [https://www.ufrgs.br/wiki-r/index.php?title=Style code](https://www.ufrgs.br/wiki-r/index.php?title=Style_code)
- Estrutura Funcional da IDE R Studio - [https://www.ufrgs.br/wiki-r/index.php?title=Modelo de programa](https://www.ufrgs.br/wiki-r/index.php?title=Modelo_de_programa)
- Utilizando o R Markdown - <https://www.ufrgs.br/wiki-r/index.php?title=Markdown>
- Salvar e carregar objetos - [https://www.ufrgs.br/wiki-r/index.php?title=Salvar/Carregar imagens dos objetos](https://www.ufrgs.br/wiki-r/index.php?title=Salvar/Carregar_imagens_dos_objetos)
- Operadores em modo **pipe** - [https://www.ufrgs.br/wiki-r/index.php?title=Operadores Pipe](https://www.ufrgs.br/wiki-r/index.php?title=Operadores_Pipe)
- Importar e Exportar Bancos de Dados - [https://www.ufrgs.br/wiki-r/index.php?title=Importar e Exportar dados](https://www.ufrgs.br/wiki-r/index.php?title=Importar_e_Exportar_dados)
- Manipulação de dados - [https://www.ufrgs.br/wiki-r/index.php?title=Manipulação de dados](https://www.ufrgs.br/wiki-r/index.php?title=Manipulação_de_dados)
- Limpeza de dados - [https://www.ufrgs.br/wiki-r/index.php?title=Limpeza de dados](https://www.ufrgs.br/wiki-r/index.php?title=Limpeza_de_dados)
- Missing data - [https://www.ufrgs.br/wiki-r/index.php?title=Tratamento de Missing Values](https://www.ufrgs.br/wiki-r/index.php?title=Tratamento_de_Missing_Values)



## Para os mais curiosos

- Noções de lógica geral: <https://www.inf.ufrgs.br/~mrpritt/lib/exe/fetch.php?media=inf05508:lpc-notas-3014.pdf>
- Sobre lógica de programação (o documento é de 1986 mas praticamente nada mudou; notar autoria do mesmo :-)): <https://cds.cern.ch/record/319152/files/AT00000453.pdf>
- Usando  $L^A T_E X$  em R Markdown: <https://everyday.codes/tutorials/how-to-use-latex-in-rmarkdown/>
- Dicionário de caracteres em  $L^A T_E X$  (+ de 18 mil símbolos): <https://tug.ctan.org/info/symbols/comprehensive/symbols-letter.pdf>
- Utilizando linguagem de marcação HTML (que poderá ser integrada com o R): [https://www.w3schools.com/html/html\\_intro.asp](https://www.w3schools.com/html/html_intro.asp)
- Integrando seu ambiente desenvolvedor em R (neste caso, R Studio) ao seu Git/GitHub: <https://happygitwithr.com/rstudio-git-github.html>
- Utilizando o seu prompt/terminal (aquele que os hackers usam): <https://gist.github.com/bradtraversy/cc180de0edee05075a6139e42d5f28ce>
- Utilizando plataformas *on-cloud* para proteção de bancos de dados: <https://zenodo.org/>
- Sobre a encriptação de pastas, documentos e outras coisas off-cloud (somente em seu computador - nota: **MUITO CUIDADO! EU NÃO ME RESPONSABILIZO CASO VOCÊ QUEIRA!**): <https://www.wired.com/story/how-to-encrypt-any-file-folder-drive/>