Metodologias Informacionais com R

Módulo I: Introdução à Linguagem R

Telmo dos Santos Klipp (telmo.klipp@inpe.br)

Telmo dos Santos Klipp

Graduação em Tecnologias da Informação e Comunicação (UFSC, 2009-2013)

Mestrado em Engenharia de Computação (FURG, 2013-2015)

Bolsista CNPq no Laboratório de Computação para Clima Espacial (LCCE/INPE, 2015-2020)

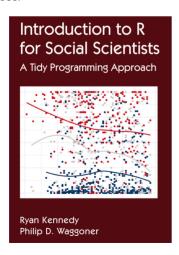
Doutorando em Computação Aplicada (INPE, 2020-atual)

1 Informações Gerais sobre o Curso

- Materiais disponibilizados via Classroom (entre com o código jdq7j5o);
- O aprendizado requer a prática que será constante nas aulas;
- Será cobrado ao menos uma atividade semanal **\(\F**.

Bibliografia Básica:

 Kennedy, R., & Waggoner, P. D. (2021). Introduction to r for social scientists: a tidy programming approach. CRC Press.



Bibliografia Complementar:

- Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. "O'Reilly Media, Inc.". Disponível em: https://r4ds.had.co.nz/index.html. Acesso em: 12 de maio, 2023. (Online)
- Damiani, A. et. al., (2022). Ciência de Dados em R. Curso-R. Disponível em: https://livro.curso-r.com. Acesso em: 12 de maio, 2023. (Online)
- de Aquino, J. A. (2014). R para cientistas sociais. Editora da UESC (editus). Disponível em: http://www.uesc.br/editora/. Acesso em: 12 de maio, 2023.
- de Oliveira, P. F., Guerra, S., McDonnell, R. (2018).
 Ciência de Dados com R: Introdução. Editora IBPAD.
 Disponível em: https://cdr.ibpad.com.br/index.html.
 Acesso em: 12 de maio, 2023. (Online)

Gostaria de Conhecer Vocês

Dinâmica em grupo:

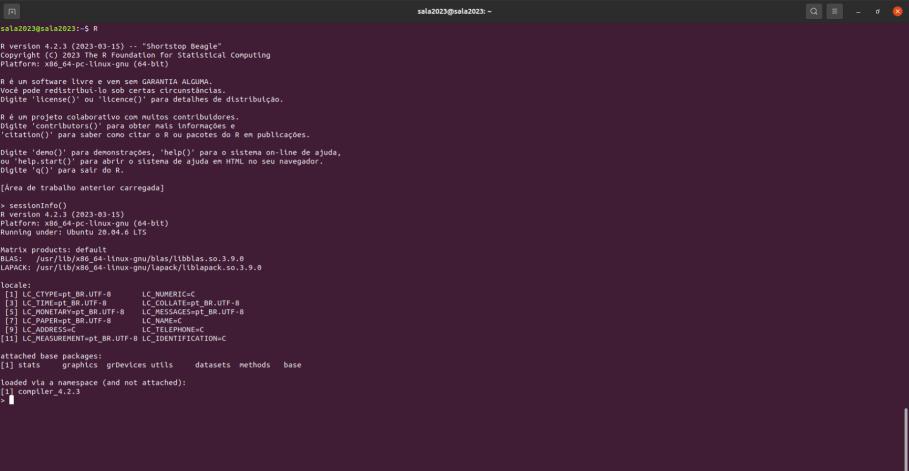
Por favor se apresente.

▼ Tarefa 1 para a semana:

Disponibilize um painel com informações sobre você no mural online da platafoma Padlet.

O que é o R?

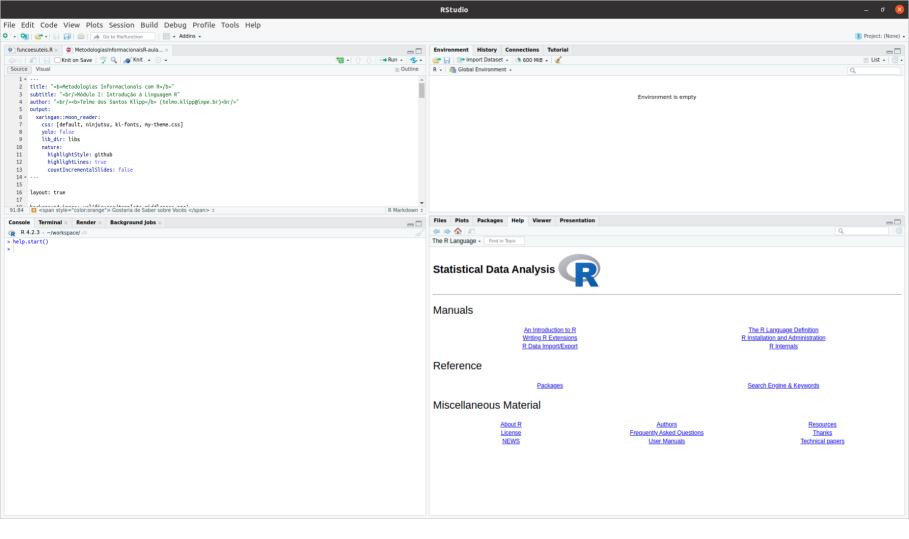
- Ambiente integrado de recursos de software para manipulação, cálculo e exibição gráfica (Venables et al., 2023);
- Linguagem de programação multiparadigma;
- Desenvolvido com foco em manipulação, estatística e análise de dados;
- É gratuito e *open source* (código aberto);
- Possui inúmeros pacotes distribuitos via The Comprehensive R Archive Network.



O que é o R?

RStudio?

- É um ambiente de desenvolvimento integrado (IDE) para R e Python;
- Ferramenta que possui diversas funcionalidades como edição e depuração de código, console e histórico de comandos, gerenciamento e visualização de dados;
- Tem as funcionalidades distribuídas em painéis por meio de abas;
- É disponibilizado em versão grátis e comercial pela Posit.



RStudio?

Por onde começar?

- Fazer o dowload do R em https://cran.r-project.org e seguir os passos para instalação;
- Fazer o dowload do RStudio em https://posit.co/download/rstudio-desktop e seguir os passos para instalação.

O R e o RStudio já estão instalados?

Ter as versões atualizadas do R e RStudio pode ser importante. Digite os respectivos comandos no console: version e rstudioapi::getVersion() ou rstudioapi::versionInfo().

version

```
##
## platform
                  x86_64-pc-linux-gnu
## arch
                  x86 64
## os
                  linux-gnu
## system
                  x86_64, linux-gnu
## status
## major
## minor
                  3.1
                  2023
## year
## month
                  06
## day
                  16
                  84548
## svn rev
## language
## version.string R version 4.3.1 (2023-06-16)
## nickname
                  Beagle Scouts
```

```
rstudioapi::versionInfo()
## $citation
## To cite RStudio in publications use:
##
     Posit team (2023). RStudio: Integrated Development Environment
##
##
     Posit Software, PBC, Boston, MA. URL http://www.posit.co/.
##
## A BibTeX entry for LaTeX users is
##
     @Manual{,
##
##
       title = {RStudio: Integrated Development Environment for R},
       author = {{Posit team}},
##
   organization = {Posit Software, PBC},
##
      address = {Boston, MA},
##
      year = {2023},
##
       url = {http://www.posit.co/},
##
     }
##
##
## $mode
## [1] "desktop"
##
## $version
## [1] '2023.6.1.524'
##
```

Por que usar R?

- É gratuito e open source (regularmente atualizado e mantido);
- Aprender a usar o R não precisa ser exaustivo (amplo e fácil acesso à documentação, fóruns e comunidade para obter informações);
- Linguagem de programação e ambiente de trabalho ao mesmo tempo;
- Executar tarefas de ciência de dados e estatística dispondo de várias ferramentas (distribuídas na forma de pacotes), possivelmente, direcionadas a sua área e domínio de interesse;
- Pelas metodologias que fundamentam alguns pacotes do R como *literate* programming (Knuth, 1992), gramar of graphichs (Wilkinson, 2012), entre outras abordagens.

Contrapontos:

• Aprender a escrever código.

Porém, existem ganhos:

- Flexibilidade e controle;
- Reprodutibilidade.





















nvironmen tal



























Business Intelligence













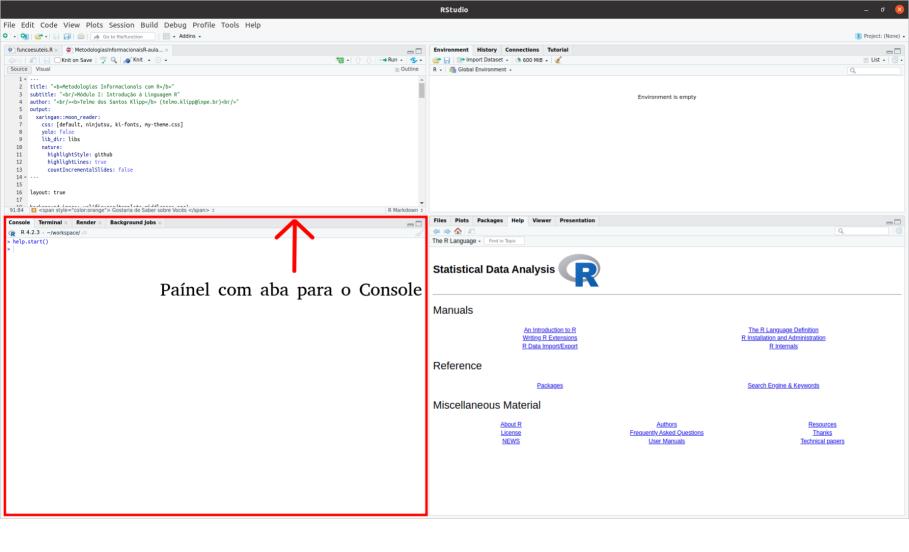




Algumas observações...

- O R é relativamente novo, desenvolvido no início dos anos 1990 (Ihaka e Gentleman, 1996) com base na também linguagem/ambiente S para estatística;
- Atualmente: muitos colaboradores, materiais online, blogs e conferências (UseR, EARL, posit::conf);
- Vamos focar no uso de pacotes baseados em metodologias intuitivas como programação letrada e gramática de gráficos.

Então vamos realmente começar!



RStudio: nossa ferramenta de trabalho - 🔀 Console

RStudio - Console

O console é um ambiente interativo onde podemos digitar comandos (ordens) - logo após o simbolo > (conhecido como *prompt*) - e enviar para o R ao teclar Enter. Os comandos são interpretados (avaliados) no mesmo instante para receber resultados. Teste alguns comandos:

```
1 + 1
## [1] 2
1+6
## [1] 7

"O RStudio é um ambiente de desenvolvimento integrado."
## [1] "O RStudio é um ambiente de desenvolvimento integrado."
```

RStudio - Tonsole

Teste alguns comandos:

```
1 + '1'

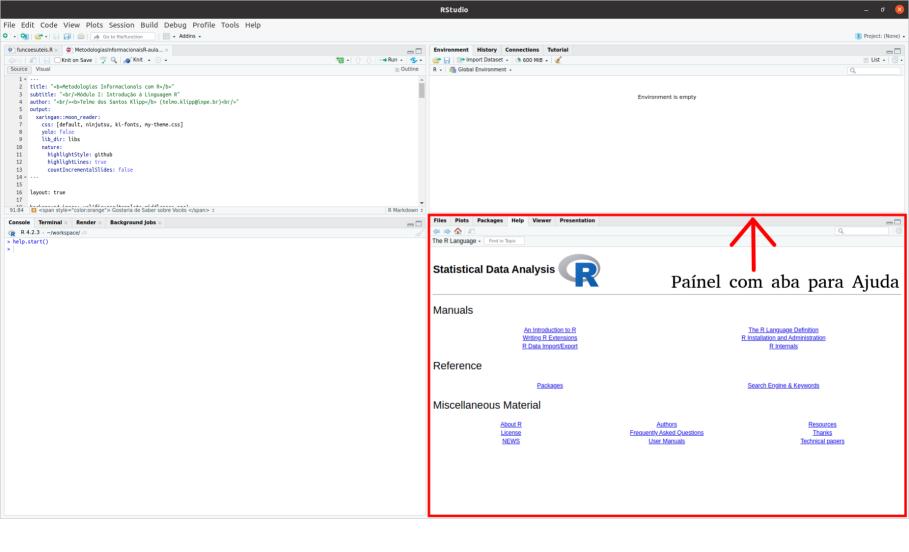
(1 + 1

Hoje é quarta feira!
```

RStudio - Console

Teste alguns comandos:

```
2**2; 3^3; 21 %% 2; 14 %% 2;
 ## [1] 4
 ## [1] 27
 ## [1] 1
 ## [1] 0
 sqrt(25); round(2.55, digits = 1); ceiling(2.01); floor(2.99);
 ## [1] 5
 ## [1] 2.5
 ## [1] 3
 ## [1] 2
round(sqrt(122.563^2, 2))
floor(122.563, digits = 2)
```



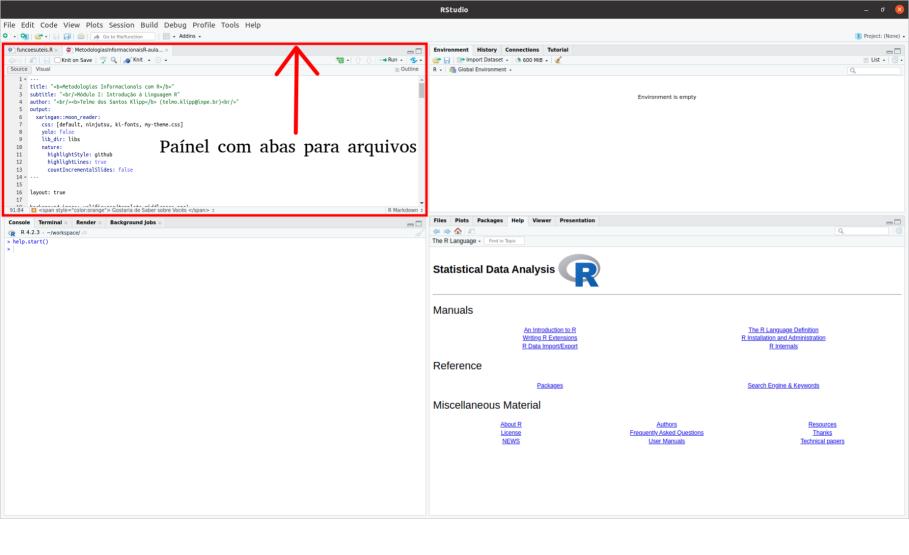
RStudio: nossa ferramenta de trabalho - ② Ajuda

RStudio - ? Ajuda

Nos temos acesso à documentação e manuais com help.start(). Não somente isso, a função help() garante acesso à informação documentada sobre operadores (ex: aritméticos e lógicos) e estruturas de controle da linguagem R. Também sobre funções e outros objetos com exemplos de uso e funcinamento. Em sintese, help() fornece ajuda para uso do sistema R.

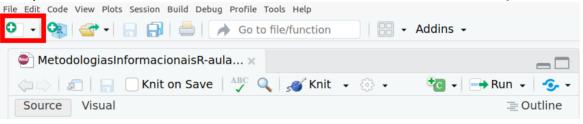
Teste alguns comandos:

```
help("%%") # ou ?"%%"
?sqrt # ou help(sqrt)
```

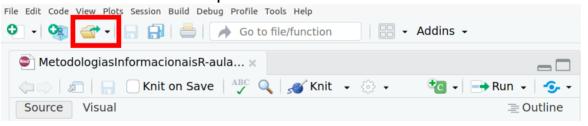


RStudio: nossa ferramenta de trabalho - 🖵 Arquivos

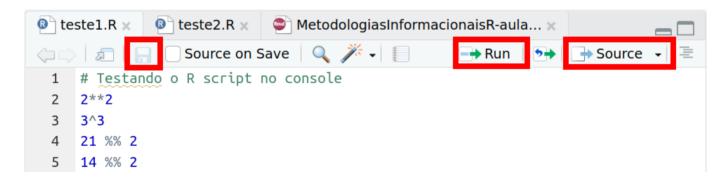
- Nesse painel ficam disponíveis os scripts (arquivos de código) ou outros tipos de arquivo para edição.
- Crie um novo arquivo R script (sua extensão deve ser nome_do_arquivo.R), segundo alguma das opções:
 - 1. No menu principal acesse: File > New File > R script; ou
 - 2. Presione Crtl + Shift + N; ou
 - 3. Clique no ícone do menu secundário e selecione R script:



- Já possui um R script ou outro tipo de arquivo e quer abri-lo. Faça isso segundo alguma das opções:
 - No menu principal acesse: File > Open File e procure/selecione no diretório (pastas) o nome do arquivo; ou
 - 2. Presione Crtl + O; ou
 - 3. Clique no ícone do menu secundário e procure/selecione no diretório o nome do arquivo:



 Digite alguns comandos no seu R script e salve o arquivo usando o ícone para salvar (conforme imagem abaixo ou Ctrl + S). É possível executar os comandos do script um por um usando o ícone Run ou usando o ícone Source, disponíveis no menu da aba do próprio arquivo (canto superior direito), conforme imagem abaixo:



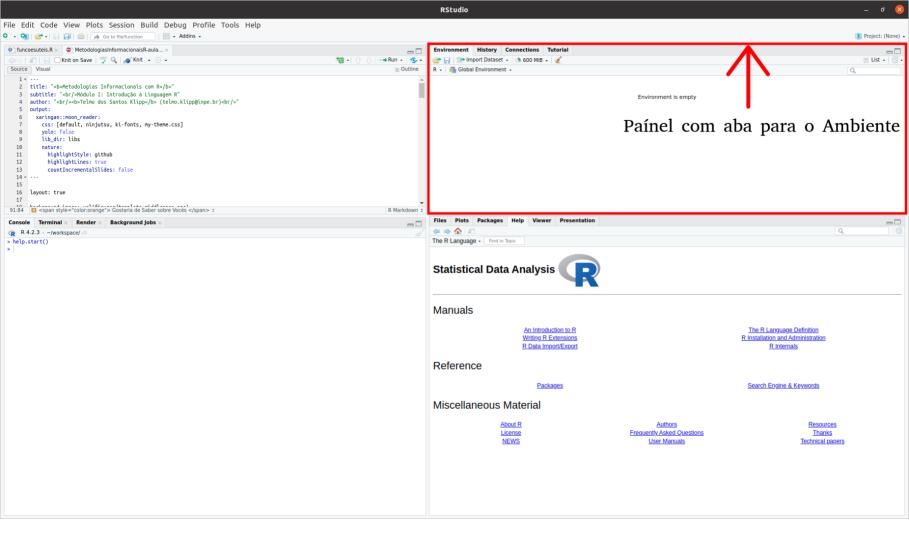
 Também é possível carregar o R script usando a função source("teste1.R"). Teste esse comando no console para executar seu script. Qual resultado será produzido?

 Para este segundo caso, qual resultado será produzido com source("teste2.R")?

 A função source() do pacote base (conhecido como R base) possui parâmetros que modificam seu funcionamento (comportamento) padrão. Teste os comandos abaixo usando seu R script:

```
source("teste1.R", echo = TRUE)
source("teste1.R", print = TRUE)
```

Não se preocupe! Entenderemos sobre funções e pacotes, assim como, representação de dados no R (objetos) durante o curso.



RStudio: nossa ferramenta de trabalho - 🖴 Ambiente

RStudio - Ambiente

- Nesta aba são listadas informações dos objetos e funções carregados no ambiente do R (sessão atual) por meio de campos. É possível interagir, bem como, importar/exportar dados e obter informações sobre alocação de memória (do computador) ou liberar memória que esta sem uso no R.
- Algumas dessas tarefas estão disponíveis por meio dos comandos:

```
ls() # listar objetos/funções
objects() # similar ao ls
gc() # liberar a memória sem uso
```

- Ao longo do curso também veremos funções para importar/exportar dados.
- A aba Histórico ao lado da aba Ambiente também é bastante útil. Essa aba possui o histórico de comandos executados no console. Talvez você já tenha percebido, também é possível acessar o histórico de comandos interativamente no console usando as setas ↑ e ↓ do teclado.

Tarefa 2 para a semana.

 As funções paste() e paste0() do pacote base concatenam (unem) blocos de texto (ou caracteres) conforme o exemplo abaixo:

```
paste("Boa", "tarde!")
## [1] "Boa tarde!"
```

- Qual a diferença entre as funções paste() e paste0()?
- Ao concatenarmos elementos para formar uma expressão na forma

```
paste("O dobro de", 2, "é", 2*2, "!")
## [1] "O dobro de 2 é 4 !"
```

no resultado, é possível notar que existe um espaço entre os caracteres "4" e "!". Por que você acha que isso ocorreu? Que resultado obtemos usando paste@()? Desenvolva uma solução para a mesma concatenação, cujo resultado seja "O dobro de 2 é 4!". Crie um R script que mostre o mesmo resultado ao ser carregado com source().

Referências

- Venables, W. N., D. M. Smith, and the R Core Team (2023). An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Ed. by R. D. C. Team. Version Version 4.3.O. Available at: https://cran.r-project.org/manuals.html. The R Foundation for Statistical Computing. Vienna, Austria. 105 pp.
- Knuth, D. E. (1992). Literate Programming. USA: Center for the Study of Language and Information. ISBN: 0937073806.
- Wilkinson, L. (2012). The grammar of graphics. Vol. 2. Berlin, Heidelberg: Springer. ISBN: 0387245448.
- Ihaka, R. and R. Gentleman (1996). "R: A Language for Data Analysis and Graphics". In: Journal of Computational and Graphical Statistics 5.3, pp. 299–314.

Metotologias Informacionais com

Muito Obrigado pela Atenção!