

Antes de Mais Nada ...

James Hunter

20 de abril de 2020

Nos capítulos seguintes, nós vamos aprender a linguagem de estatística e programação R e sua utilização na análise de dados. O “approach” será bem prática. Vamos começar com a instalação do software necessário, continuar com operações básicas em R, e seguir todos os passos de um projeto de análise de dados até um relatório final. No longo do caminho à conclusão da matéria, você terá muitas oportunidades para praticar suas habilidades.

Texto e Fontes

Deixe-me ressaltar que estou escrevendo e desenvolvendo todos os materiais da matéria em R, no formato *R Markdown*: apostila, slides, etc. R com RStudio tem uma gama de recursos super abrangente. Nós vamos ver neste curso rápido só uma pequena porção deles. Durante o curso, vou apresentar vários documentos que você pode usar como referência ou tutorial para aprender parte do universo R/RStudio. Um deles tratará de *R Markdown*: como preparar documentos, programas, apresentações, blogs, etc. com ele.

Quando inclui programação em R, vou usar o seguinte fonte:

```
sub_dados <- dados %>%
  select(dia20:dia25)

glimpse(sub_dados)
skimr::skim(sub_dados)
sub_dados %>% map_dbl(sd)
```

As vezes, vai ter um ponto importante que não cabe linearmente no texto. Este tipo de ponto eu chamo um **VSS**, um *verbum sapientibus satis*, que é Latím para “um palavra é suficiente para os sábios”, em honra da minha professora de Latím de ensino médio, Profa. Walden.

Antes da Primeira Aula

Para estudar uma linguagem como R, temos que ter ela instalada onde podemos acessar ela e escrever e executar programas (“*scripts*”). Os próximos paragrafos apresentarão duas maneiras em que podemos aproveitar de R. Primeiro, descreverei como instalar R e RStudio no seu laptop. Quando temos aulas, seja presencial ou à distância, deve ter ele perto porque você vai querer seguir os slides, executar alguns dos programas que desenvolveremos e consultar vários referências. Segundo, vou explicar um pouco como usar “*RStudio Cloud*”, onde vou colocar uma cópia dos materiais do curso se você está longe de seu computador.

Depois disso, vou mostrar como você pode carregar um número de “pacotes”, que contêm programação útil para o curso. A maioria das ferramentas para manipular dados, visualizar eles e fazer as análises ficam nesses pacotes.

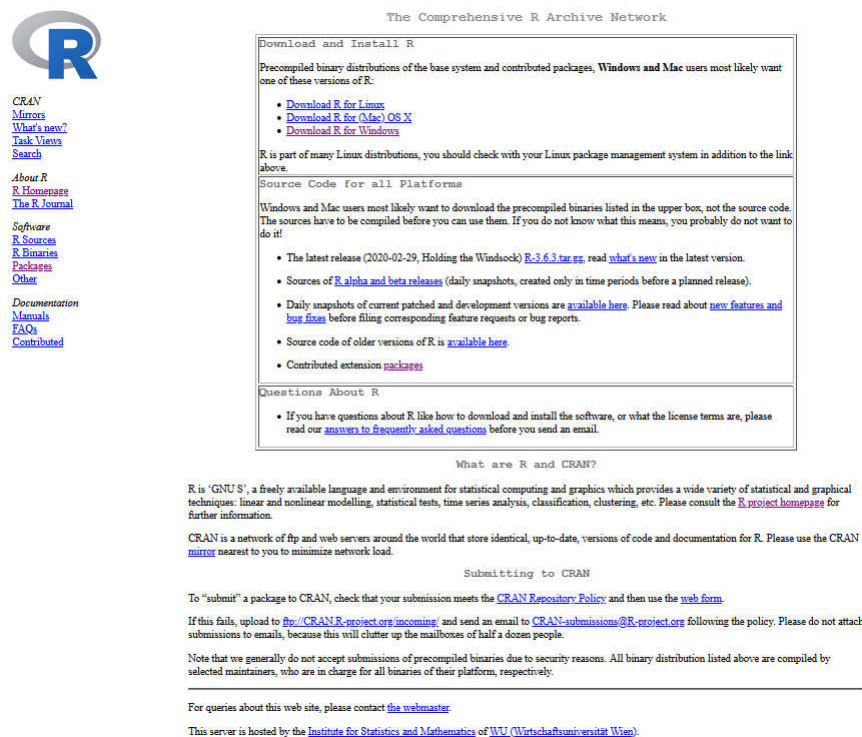
VSS Você precisará usar R e RStudio para completar o Entrance Quiz. Para usar, tem de instalar, né?

Instalar R no Seu Computador

Você pode achar R no site seguinte:

<https://cran.r-project.org/> (<https://cran.r-project.org/>)

CRAN é o grupo oficial que mantem R e garanta que todos os componentes funcionam sozinhos e juntos. Esta é um dos motivos que R está tornando muito popular – a qualidade do produto é muito alta. Figura 1 mostra a tela principal de R.



CRAN

No dia 28 de abril, CRAN vai liberar a nova versão de R, v.4.0.0 (“Arbor Day”). Na linha da tela acima de “*latest release*”, vai aparecer este número invés de 3.6.3.

Existem versões para Windows, MacOS (o novo nome de OSX) e para vários sabores de Linux. Na parte superior da tela, clique no link que é apropriado para seu sistema operacional e seguir as instruções para download do software. Vou detalhar aqui alguns dos passos para Windows e para MacOS.

Processo de Instalação de R (para os Windows)

1. Clique no link “Download R for Windows”
2. Na próxima tela, clique no link “base”
3. Clique no link: “Download R x.x.x for Windows” onde “x.x.x” é o número de versão.

O site vai fazer o download do programa como é normal para os downloads. Você pode instalar o programa seguindo seu procedimento normal e R vai depositar no Desktop um ícone.

Instalação do RStudio

Primeiro, porque precisamos este programa? RStudio dá asas para R. Ao nível mais básico, RStudio é nossa interface para R. É um programa baseado num editor de texto onde você escreve os scripts que vai desenvolver em R. Mas, é realmente o canivete suíço de R. Lá pode gerenciar os arquivos, estabelecer projetos que colocam todos os arquivos de um projeto junto, fazer controle de versões, e ver seus resultados, relatórios, etc. Em todas nossas aulas, vou usar RStudio para controlar todas as demonstrações de programação, e RStudio é o software em que fiz toda a documentação da matéria. RStudio também é **de graça**, sem custo, *free*. Existe uma versão comercial, mas vocês não precisam ela.

Para instalar, navegue diretamente á página <https://www.rstudio.com/products/rstudio/download/> (<https://www.rstudio.com/products/rstudio/download/>), que permite que você evita muita informação desnecessária.

Esta tela tem muita informação. Então precisa rolar para baixo até que você chega nos “RStudio Desktop 1.2.xxxx”. Parece como a figura seguinte:

RStudio Desktop 1.2.5042 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:

DOWNLOAD RSTUDIO FOR WINDOWS
1.2.5042 | 149.84MB

Requires Windows 10/8/7 (64-bit)



All Installers

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

RStudio 1.2 requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

OS	Download	Size	SHA-256
Windows 10/8/7	RStudio-1.2.5042.exe	149.84 MB	5d4ed644
macOS 10.13+	RStudio-1.2.5042.dmg	126.89 MB	74ea88ab
Ubuntu 14/Debian 8	rstudio-1.2.5042-amd64.deb	96.41 MB	655e2757
Ubuntu 16	rstudio-1.2.5042-amd64.deb	104.07 MB	e2f15cc2

RStudio Download

Agora, pode fazer o download para o sistema operacional seu e instalar o programa seguindo as instruções no módulo de instalação.

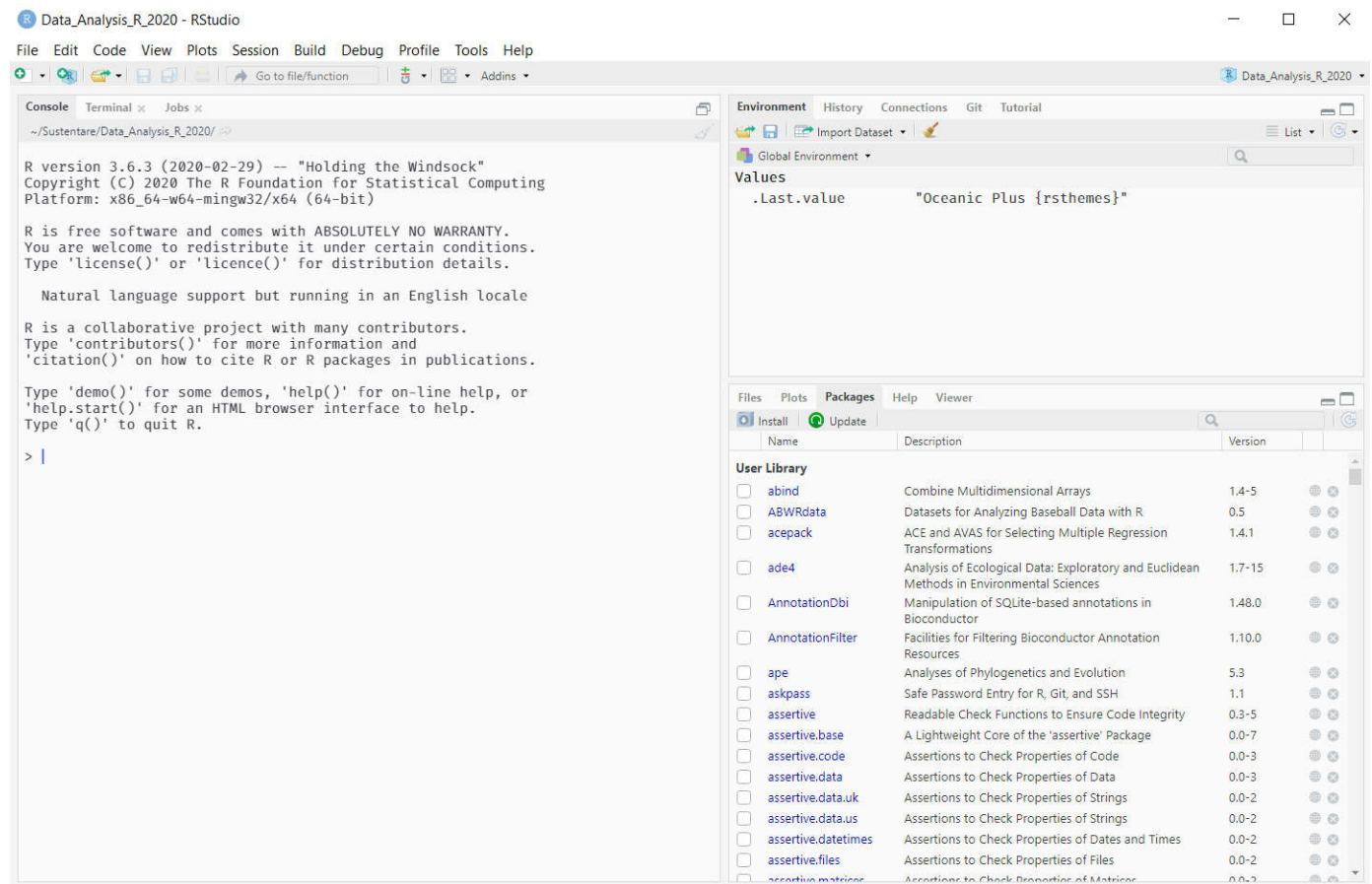
Se vocês precisam ajuda adicional com a instalação, coloquei um vídeo no meu canal de YouTube chamado “MAD CB Instalar R”. Para alcançar o canal, vá para: <https://www.youtube.com/channel/UCbvgZ8RYeTtjgAKE-jub5A> (<https://www.youtube.com/channel/UCbvgZ8RYeTtjgAKE-jub5A>)

Iniciando RStudio

No seu Desktop ou nos menus, clique no ícone de RStudio (não o ícone de R) duas vezes. RStudio vai abrir e R vai abrir automaticamente dentro de RStudio. Num dos painéis (normalmente o painel no lado esquerdo) vai aparecer o *Console*, onde R executa as funções. No lado direito têm dois painéis. Naquele para cima são várias abas relacionadas aos variáveis que R guarda na memória ativa do computador e comandos que você

executa. No painel inferior, têm abas que mostram informação sobre os arquivos na pasta “*working directory*”, o lugar onde R acha os arquivos (data, programas) que você quer executar, os pacotes que você tem no computador e as páginas de assistência (“*Help*”) para R. Vou falar sobre todos esse tópicos mais tarde. Utilizando as Preferências (no item de menu de “RStudio”), você pode controlar a aparência da tela).

O painel de Console mostrará que a versão atual (3.5.0) de R está carregado e pronta para operação.



RStudio Tela Inicial

Instalar os Pacotes Necessários

Durante o curso, nós vamos usar vários pacotes com funções adicionais que vão além as de R básico (“*Base R*”). Vocês podem instalar os mais importantes através de um script que já preparei para você. Trabalhando com o script vai permitir que vocês comecem de aprender como usar o R. O script tem o nome “*instalar_pacotes.R*”.

O seguinte é o texto do programa. É muito simples. Só 2 comandos.

```
pacotes <- c("tidyverse", "broom", "car", "caret", "corr", "data.table",
            "descr", "devtools", "gapminder", "ggpubr", "ggvis", "glue",
            "gmodels", "here", "Hmisc", "hms", "janitor", "jsonlite",
            "kableExtra", "knitr", "lattice", "librarian", "lubridate",
            "magrittr", "mice", "nortest", "nycflights13", "outliers",
            "pROC", "psych", "RColorBrewer", "Rcpp", "readxl", "ROCR",
            "shiny", "styler", "usethis", "titanic")

install.packages(pacotes)
```

O primeiro comando, especificando os pacotes para carregar, usei 2 comandos importantes para R. Eu quero usar o nome `pacotes` para referir à lista dos 38 pacotes. Não quero teclar o nome deles cada vez preciso usar. Chato. Então, estou *atribuindo* a palavra `pacotes` para ficar no lugar da lista inteira. Faça isso com o simbolo de uma seta pontado para a esquerda “<-”, composto do caráter “<” e o hífen “-”. Este é chamada o operador *assignment* e não é equivalente ao sinal de “==”, que é uma condição lógica (uma coisa sendo igual a uma outra). Entretanto, R quer ser tão amigável que permite que você usa o sinal “=” para atribuição.

Uma dica sobre o operador de atribuição. Só use o “<-” e não o “=” . Em algum momento, você vai confundir os 2 usos de “=” . Todos nós fazemos. Desenvolve o hábito agora de usar o “<-” e pode evitar esse erro.

O segundo comando nesta expressão é o `c()` . Este função quer dizer “concatenar” ou colocar os itens que seguem num vetor. Os itens podem ser qualquer tipo de dado que R reconhece. No capítulo X, exploraremos os tipos de dados que pode usar em R. Outra novidade é que agora você viu um **vetor**. Um vetor é um matriz, mas de uma dimensão só. Depois que você executa esta função, vai ver na janela de *Environment* a variável `pacotes` e que ela é do tipo “caráter” (`chr`) e tem elementos 1 até 38.



Vetor Pacotes

Na função `install.packages(pacotes)` , R vai inserir o conteúdo do vetor `pacotes` no lugar da palavra e assim executar.

Cadê `instalar_pacotes.R`?

Temos uma cópia de `instalar_pacotes.R` no SER. Também tem outra no RStudio Cloud, o recurso que explicarei na próxima seção.

Onde Devo Colocar `instalar_pacotes.R`?

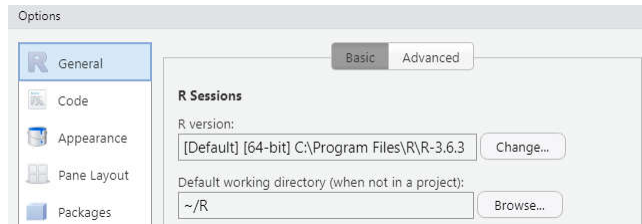
Obviamente, não dá para executar `instalar_pacotes.R` do SER. Precisa pôr ele num lugar onde R pode achar ele. (Com RStudio Cloud, já está localizado no lugar certo.) Chamamos isso um *working directory* (“pasta de trabalho”).

O Que É Um *Working Directory*?

O *Working Directory* é o lugar onde R espera achar todos os arquivos que vai precisar (dados e scripts). Você pode achar ele em RStudio clicando o item do *toolbar* da aba *Files* chamado *More/Go To Working Directory* e RStudio vai levar você lá. Se o *Working Directory* não é a pasta que você esperou, pode navegar na lista dos arquivos para a pasta que quer usar como o *Working Directory* e designar esta pasta como o seu *Working Directory*.

Você vai querer estabelecer um *Working Directory* padrão para as ocasiões quando você não está trabalhando num projeto (outro conceito de RStudio que vai esperar um pouco). Pode fazer isso através do item do menu `Tools/Global Options/General/Basic` . Lá a segunda item é o “*Default working directory (when not in a project)*.”. Crie uma pasta ou selecione uma que já criou. A figura abaixo mostra meu, `~/R` , ou seja, a pasta `R` no meu disco rígido. O “~” antes do R quer dizer toda a estrutura das pastas acima do `R` .

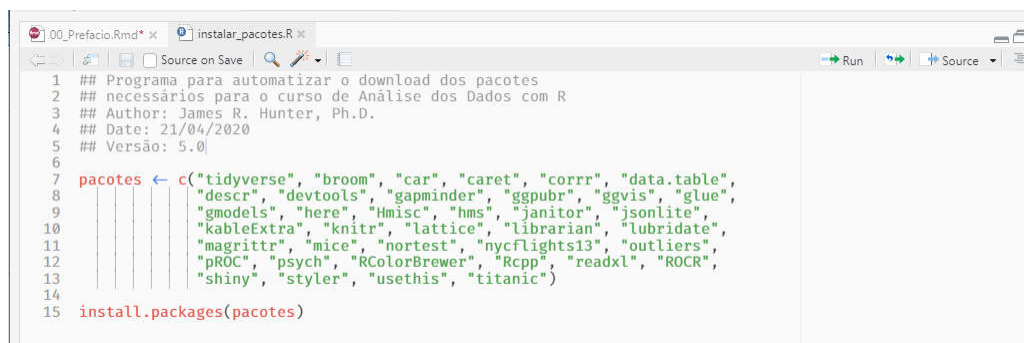
Quando você criou o *working directory*, só precisa usar seu sistema operacional para copiar `instalar_pacotes.R` para ela e estará pronto a executar o programa.



Vetor Pacotes

Instalar os Pacotes

Finalmente, é a hora de carregar o programa com os comandos e executá-lo. Para carregar, só precisa clicar no nome do arquivo `instalar_pacotes.R` e ele vai abrir numa janela acima do Console.



RStudio Janela de Programas

Para executar o programa, você pode clicar na palavra *Source* na barra acima do programa. Vai precisar um pouco de tempo, mas vai carregar no seu computador esses pacotes e outros em que eles dependem para funcionar.

Instalação do R e RStudio em Outros Sistemas Operacionais

Os passos são o mesmo que o procedimento para Windows ou até um pouco mais simples (para MacOS). Siga os passos detalhados aqui com os equivalentes de seu sistema.

Uso do RStudio Cloud

Cloud é um novo recurso de RStudio. Criei um projeto em *Cloud* para cada um dos quatro sessões de aula. No seu browser, vá para projeto que conforme a sessão que você deseja e lá achará todos os arquivos necessários para aquela sessão, dados e scripts.

Localização das Aulas no RStudio Cloud

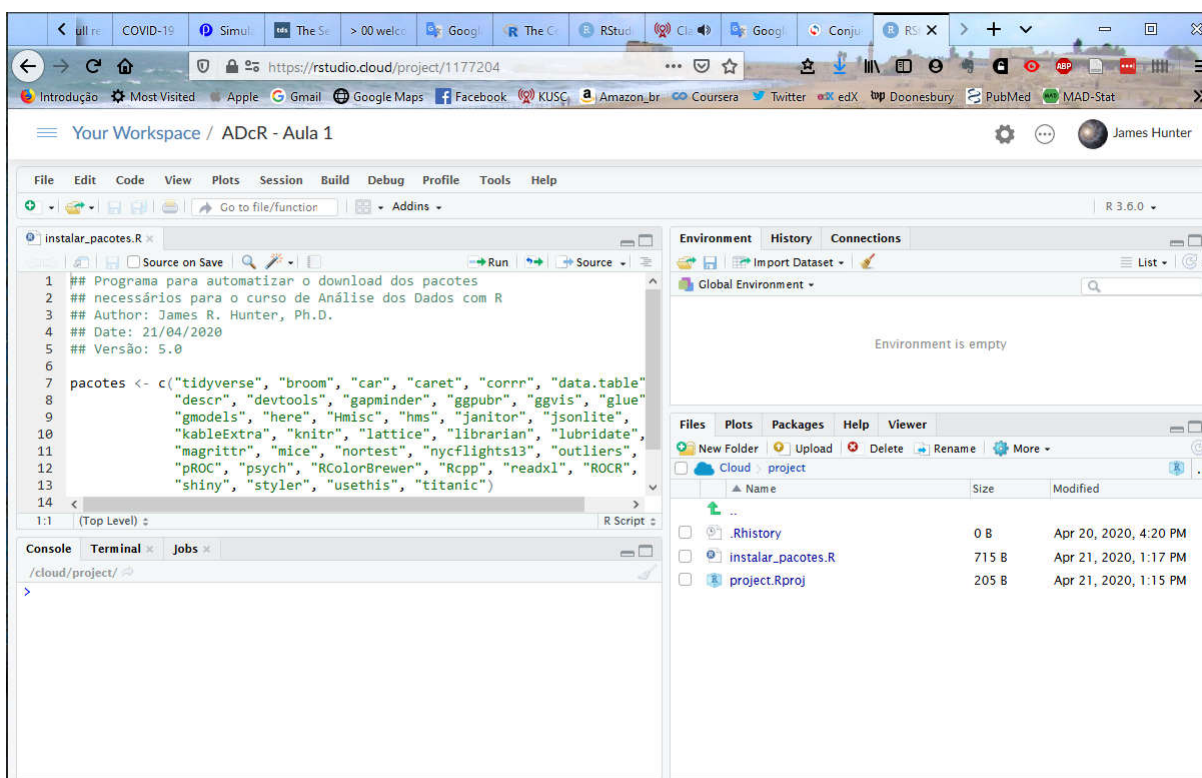
Aula	Project
ADcR Aula 1	https://rstudio.cloud/project/1177204 (https://rstudio.cloud/project/1177204)
ADcR Aula 2	https://rstudio.cloud/project/1181159 (https://rstudio.cloud/project/1181159)

Aula Project

ADcR Aula 3 <https://rstudio.cloud/project/1181172> (<https://rstudio.cloud/project/1181172>)

ADcR Aula 4 <https://rstudio.cloud/project/1181165> (<https://rstudio.cloud/project/1181165>)

Cada projeto parece como a tela de RStudio que você já viu como mostra a figura seguinte. Tem os mesmos componentes e capacidades. Pode fazer o seu trabalho lá e transferir um script que você escreve lá para seu laptop depois (copiar e colar). Você pode acessar o RStudio Cloud de qualquer computador e todos os browsers principais (Chrome, Firefox, Safari, Edge, etc.).



RStudio Janela de Programas

Para Trabalhar com RStudio Cloud

Você precisa abrir uma conta com RStudio Cloud para ter acesso. O processo é super facil e não custa nada.

VSS Importância de Fazer Esta Fase Rápido

Se você demora de instalar os softwares e tenta praticar um pouco com eles até o momento antes que as aulas começam (11 de maio), vai sofrer. Sem duvida. Por causa de nosso inimigo, Sr. SARS-CoV-2 e a doença ele causa, temos um tempo reduzido para nosso trabalho. Este põe mais ônus em você para chegar na primeira aula preparada.

What's Next

No próximo capítulo, vou traçar um pouco da historia de R, como chegou de ser uma linguagem de análise importante e como eu uso. Vou descrever meu trabalho modelando como COVID-19 está espalhando em

Brasil e através do mundo. Assim, pode ver que R aplica bem às tarifas analíticas importantes no mundo.

Vou fornecer também lá uma serie de recursos que pode ajudar você dominar R e fazer dela uma ferramenta útil.