

Análise dos Dados com R

Projetos em R

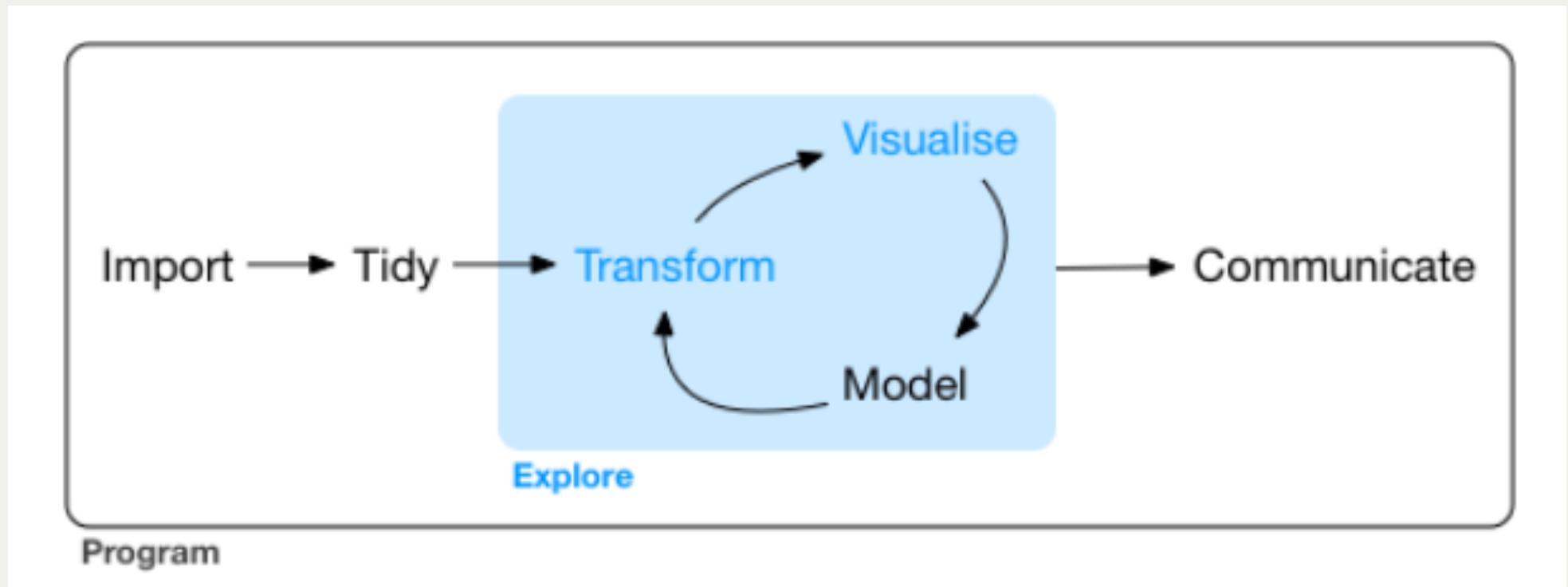
James R. Hunter, PhD
Retrovirologia, EPM, UNIFESP

2023-10-03

Semana Passada

- Falamos das variáveis e estruturas de variáveis
 - `vector`, `data.frame`, `tibble`
- Hoje, falamos de como usar esses objetos
 - Como combinar em *frases* que executam comandos e análises
 - Loops
 - Declarações condicionais (*if ... then ... else*)

Diagrama de Processo dos Projetos



Fluxo de Trabalho do Projeto

- Diagrama *Tidyverse* mostra passos principais para organizar tarefas computacionais
- **Porém**
 - Precisa muito planejamento **antes** que podemos começar com *Import*
 - Diretamente no início de um projeto



Questões Iniciais no Planejamento

– Biológico/Médico

- O que é o tópico de minha pesquisa?
 - A questão na escala grande
- O que desejo calcular com minha análise?
- O que a ciência tinha dito sobre este tópico antes?
 - Procura na literatura científica
 - Quais são bons recursos (PubMed, Google Scholar, etc.)?
 - Quando terei lido o suficiente?

Questões Iniciais - Tipo do Estudo

- O que é o tipo de estudo que irei realizar
 - Experimental
 - Observacional
 - Caso-Controle
 - Outro tipo?
- O que é minha hipótese?
- Quais experimentos/trabalho de campo/recolhimento dos dados apoiam esta hipótese?

Técnica Analítica Principal

- Qual técnica analítica usarei?
 - Quais usararam outros pesquisadores?
 - Funcionaram??
 - Produziram resultados interessantes?

Dados para Meu Projeto

- Quais dados eu preciso para meu projeto?
 - Variáveis categóricas
 - Variáveis de resultado
 - Numéricas
 - Outros tipos de dados
- Quantos dados vou precisar?
 - Número dos casos (**n**)
 - Poder estatístico para conseguir alcançar significância

No Caminho para os Projetos em RStudio

Pastas Ativas

- Quando você inicia RStudio
- Programa usa a pasta de seu ajuste de “Tools/Global Options”

setwd() & getwd()

- Para criar uma nova pasta ativa (*working directory*), precisa usar o comando `setwd()`
 - Com o **nome completo do caminho** em aspas como argumento
- Exemplo - minha pesquisa da tese de doutorado sobre HIVAIDS
 - `setwd(C:/Users/james/OneDrive/Documents/HIVAIDS/Drug_Holiday")`
- Função complementária `getwd()`
 - Sem argumento
 - Mostra o que é a pasta ativa

```
1 getwd()
```

```
[1] "/Users/jameshunter/Documents/MAD/Bioinformatica_APLICADA_2023/rmodule/r_aulas23"
```


Porque Precisamos Projetos?

- Vai trabalhar em projetos diferentes
 - Bases de dados diferentes
 - *Scripts* (programas) diferentes
 - Documentos diferentes
- Jogando todos seus arquivos de todos os projetos em uma pasta R vai enloquecer você
 - Muitos projetos podem ter acima de 500 arquivos
- Planejamento do fluxo do projeto implica organização
- Construir um projeto de análise de dados em R baseado num espaço de trabalho distinto e reservado

Projetos em RStudio -- Trabalho Feito Melhor

- Tudo mundo tem uma estrutura diferente das pastas nos seus computadores
 - Tenho projetos gerais de computação na pasta R
 - Todos meus projetos sobre HIV ficam em minha pasta HIVAIDS
 - Todas as versões desta matéria ficam na minha pasta MAD
- O que é a estrutura das suas pastas?
- Projetos de R criam sua pasta ativa própria
 - Se for na rede de trabalho seria a mesma estrutura para tudo mundo no grupo
 - Não tem mais os caminhos de pasta quilométricos
 - Que implica digitando eles errados

The Tidyverse Warning/Threat!

A Equipe Tidyverse ameaça:

Se a primeira linha de um script de R é:

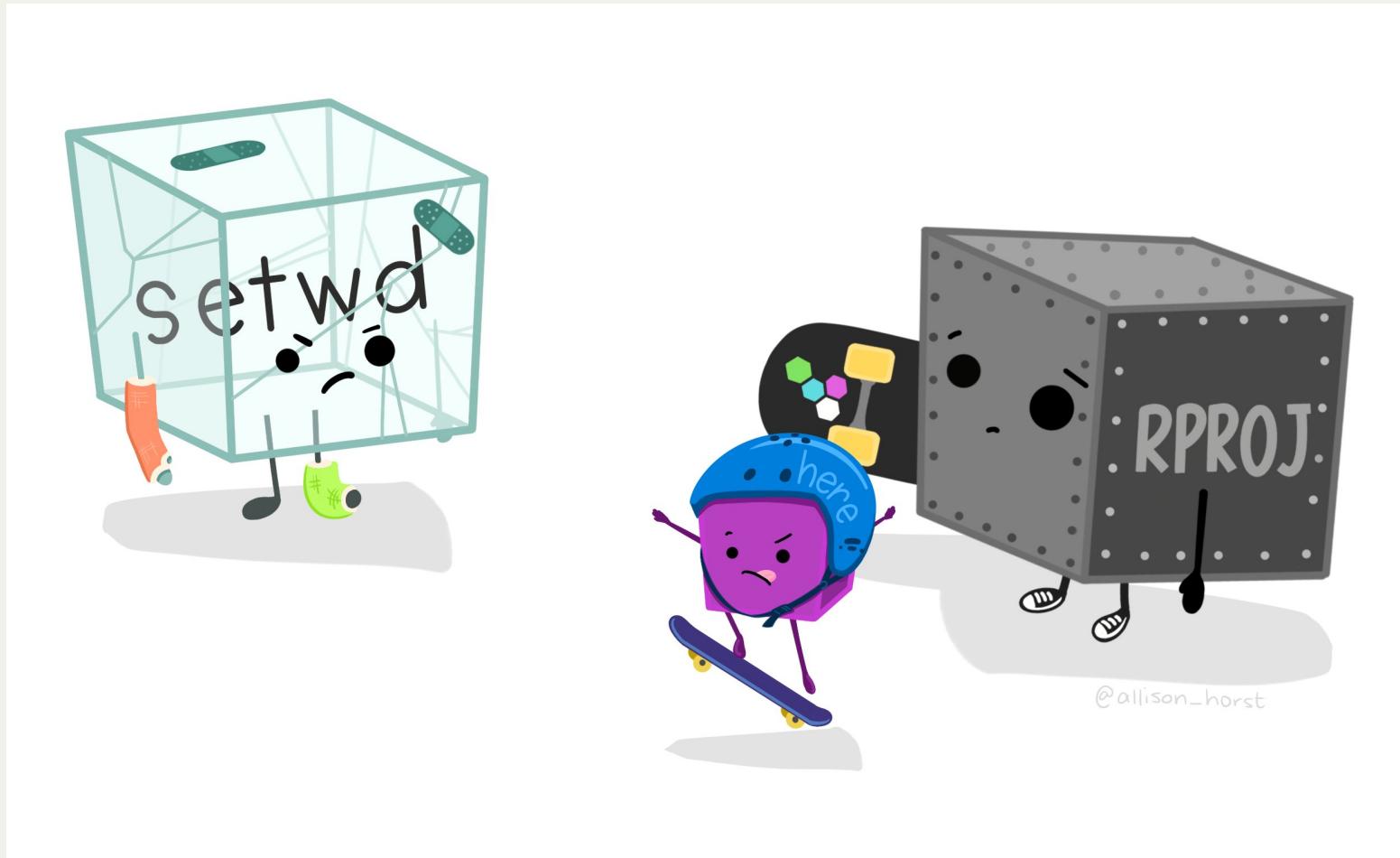
```
setwd("C:\Users\jenny\path\that\only\I\have")
```

Alguém da equipe virá para seu escritório e **incendiar o seu computador**

Você foi avisado!!!

A melhora maneira a evitar tal consequência é criar um novo projeto de RStudio para cada nova atividade.

Projetos



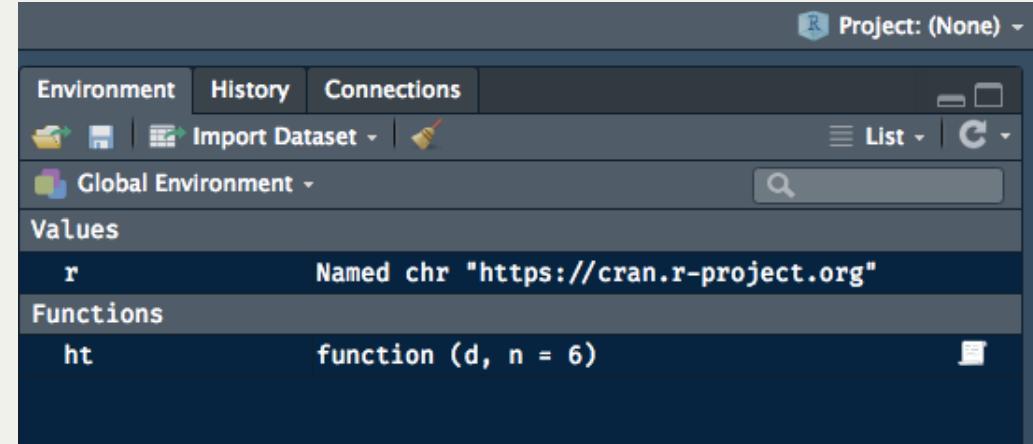
Source: Alison Horst, <https://github.com/allisonhorst/stats-illustrations/blob/master/rstats-artwork>

Como Criar um Projeto de RStudio

- Procure o texto
Project: (None)

- No topo da tela
acima de aba de

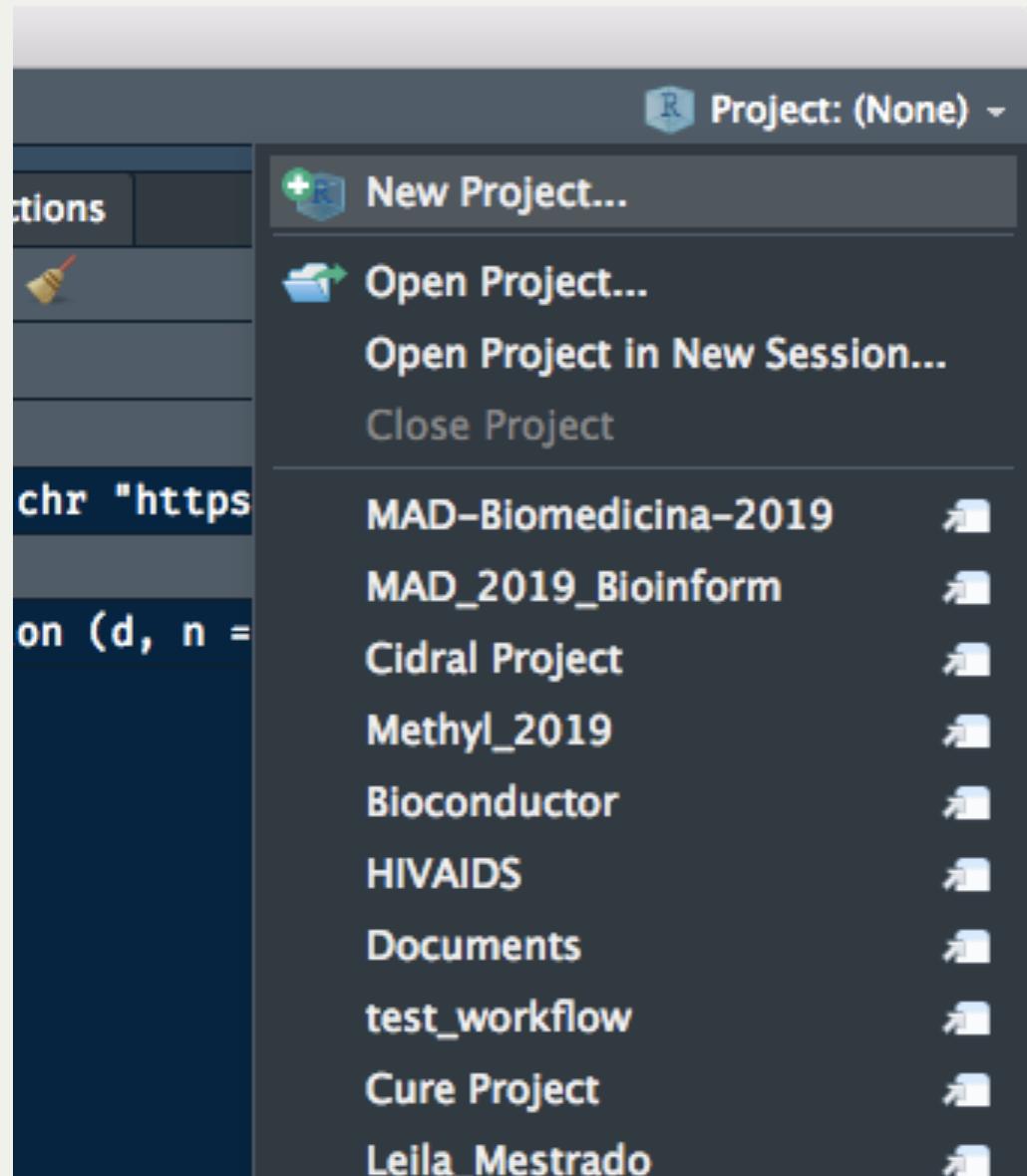
Environment

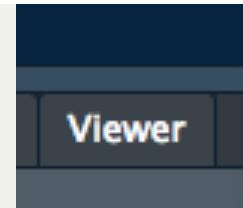


- Até agora, não tem um projeto **ativo**

Criar Novo Projeto

- Clique na seta a direta do “(None)”
- Menu suspenso abrirá
- Clique opção New Project ...





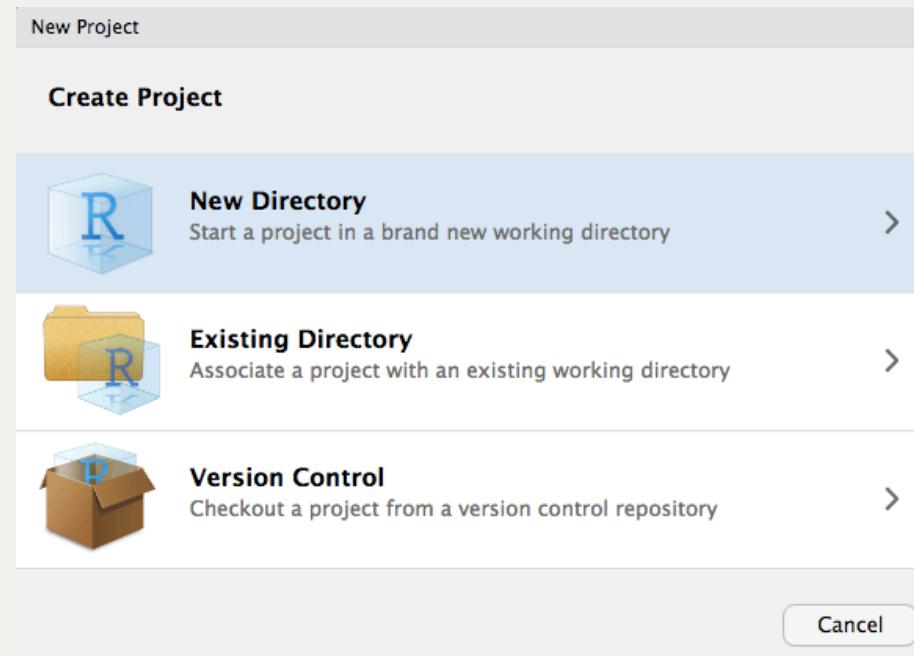
[Clear Project List](#)

[Project Options...](#)

⇧⌘,

Criar Janela do Projeto

- Opções para escolher uma pasta nova ou existente
- Opção **Version Control** mais avançada
- Neste caso, escolhe **New Directory**

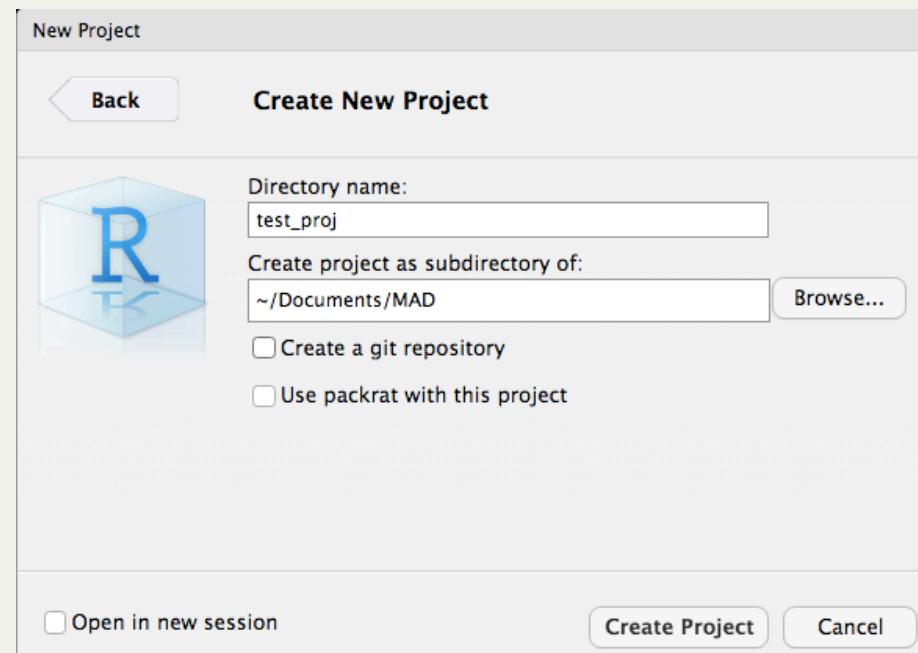


Tipos de Projetos Novos

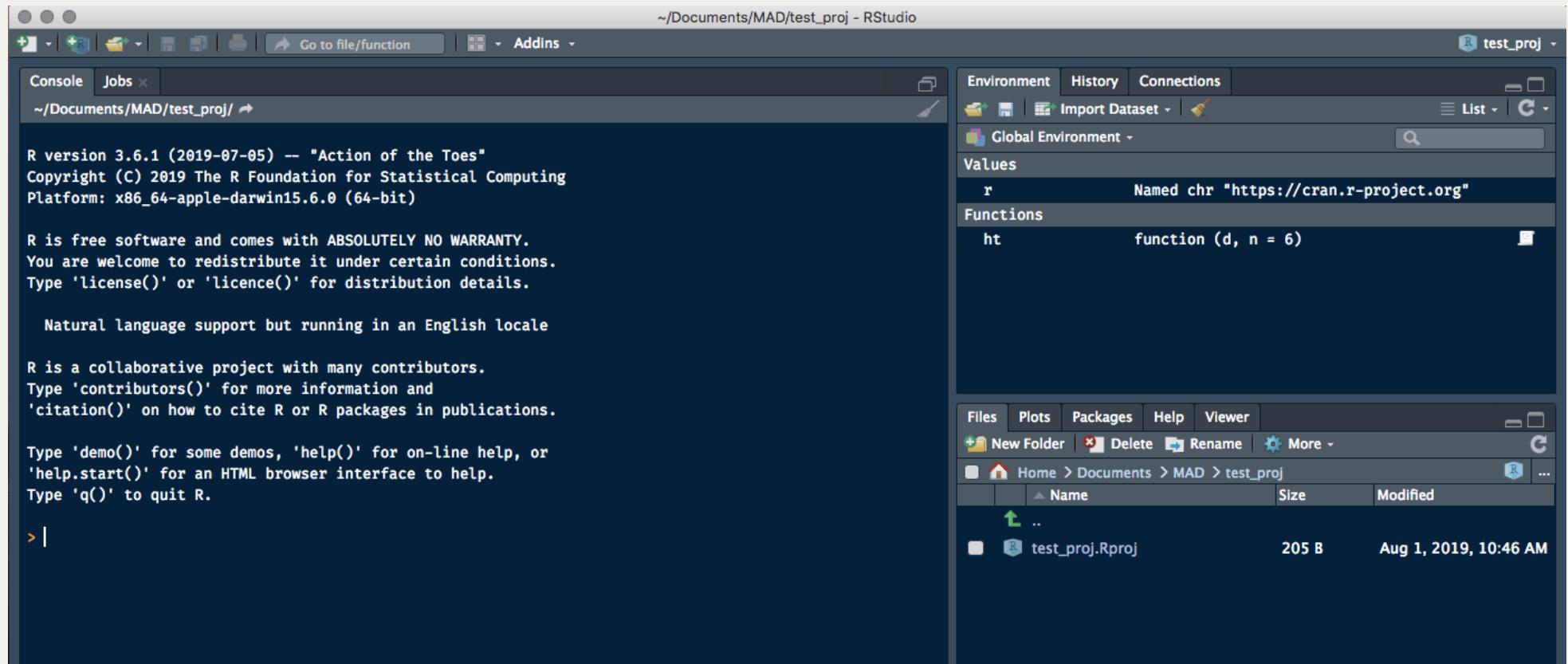
- RStudio tem muitos tipos de projetos
- Por enquanto, queremos um *New Project* simples: clique lá

Tela de Criação do Projeto Final

- Dê um nome e um locale no disco
- Clique o botão **Create Project**
- Por enquanto, esqueça as caixas de **git** e **renv**



Eis Seu Projeto



Sub-Pastas

Minha Estrutura das Pastas Preferida

- Sub-pastas
 - Dados
 - Dados brutos
 - Documentos
 - Gráficos
 - Programas/Scripts
 - Slides

Here



Source: Alison Horst, <https://github.com/allisonhorst/stats-illustrations/blob/master/rstats-artwork/here.png>

Problema

Solução: `here::here()`

- Pacote `here` faz uma coisa só
 - Mostra para você o caminho completo para sua pasta ativa
 - Numa forma simples: não tem argumentos

```
1 here::here()  
[1] "/Users/jameshunter/Documents/MAD/Bioinformatica_APLICADA_2023/rmodule/r_aulas2"
```


Ir à Sub-Pasta **graphics** do Projeto

- Give the function the argument "**graphics**"
- Thereafter, can use variable **gr** to refer to full path of location

```
1 gr <- here::here("graphics")
2 gr
[1] "/Users/jameshunter/Documents/MAD/Bioinformatica_APLICADA_2023/rmodule/r_aulas23/graphics"
```


Fine Parte 1

