

# Lição de Casa 1

James R. Hunter, PhD.

18 de setembro de 2020

Nesta lição da casa, vamos trabalhar com algumas problemas verdadeiras e alguns simplificados. As respostas precisam ser submetidas antes de **2 de outubro** por email: jameshunterbr@gmail.com.

## Problema 1

O arquivo `trplasma.csv` (no GitHub) mostra todas as mutações que uma amostra dos pacientes com HIV em falha virológica tiveram no gene transcriptase reversa do vírus. Um “1” na célula da planilha indica que a mutação estava presente e “0” demonstra ausência da mutação naquele momento.

A primeira coluna é o número de código do paciente e todas as outras colunas representam as mutações. Os nomes das colunas podem ser interpretados como:

- “tr” para *transcriptase reversa*,
- o número seguinte como a posição do aminoácido (codon) com a mutação
- o código para semana de exame de sangue (“bl” = *baseline* ou “\_12” = 12 semanas).

Primeira tarefa: importar o arquivo usando `readr::read_csv()` para um tibble chamada `trplas`.

Segunda tarefa: Transforme os dados para um conjunto *tidy*, usando as funções de `tidyr`.

Dica: planeje a transformação. Quais são as variáveis? Como vai dividir o código da mutação? Desenhe no papel como vai aparecer o tibble.

## Problema 2

Carregue na memória o arquivo “`pac_demo.xlsx`” (no GitHub) como `pac_data`, e responda às perguntas seguintes.

- Existe uma diferença entre as médias de `copias_cv` entre pacientes de São Paulo e de Rio Grande de Sul? Quanto?
- Quando falamos de carga viral de HIV, normalmente falamos de uma transformação das cópias do vírus em log10. Usando a função `log10()`, crie uma nova variável da valor logaritmico de carga viral, `log_cv` e salvar ele de volta para `pac_data`.
- O que é a idade mediana das pacientes com baixa contagem das células CD4+ (`contagem_cd4`), ou seja, um valor em baixo de 200?
- Este valor é maior ou menor da idade mediana das pessoas com contagem de CD4+ maior de 200? Porque você acha que esta diferença existe?

## Problema 3

No pacote `nycflights13`, o pior linha aérea em demoras de chegada nos destinos foi Frontier Airlines. Agora vamos considerar o destino que teve o pior desempenho em termos de demora. O diagrama da estrutura de

`nycflights13` fica nos slides de Aula 3.

Qual é o tempo médio de demora na chegada para todos os destinos que os vôos de Nova York e o que é o nome do pior?

Qual é o tempo médio de demora na chegada para todos os destinos que os vôos de Nova York e o que é o nome do pior?