Rebeca de Jesus Carvalho

FGV CEPESP rebeca.jesus.carvalho@gmail.com

# Tópicos da aula

- Introdução
- Bases relacionais com a gramática do dplyr
- R e Power Bl
- Laboratório

#### Conteúdo da aula de hoje

- Realizaremos uma breve recapitulação dos assuntos abordados até aqui.
- Veremos como combinar *data frames* de diferentes origens que se relacionam por meio de uma ou mais variáveis chave.
- Para isso, utilizaremos novos verbos do dplyr, de sufixo \_join para trabalhar com bases de dados relacionais.
- E, como atividade bônus, aprenderemos como integrar R e Power Bl.

# O que aprendemos até agora

- Conhecemos a gramática do *dplyr* e suas facilidades na manipulação de dados, especialmente com o emprego do operador **pipe** (%>%).
- Aprendemos a filtrar, selecionar e renomear colunas: filter(), select()
  e rename().
- Também vimos como ordenar, contar, recodificar, criar novas colunas, agrupar e sumarizar nossos dados: arrange(), count(), recode(), mutate(), group\_by() e summarise().

#### O que aprendemos até agora

- Descobrimos algumas das funções básicas disponíveis para explorar um data frame: head(), glimpse(), names(), table() etc.
- Exploramos os operadores aritméticos, relacionais e lógicos e seus diferentes usos.
- Discutimos as diferentes cláusulas condicionais if, else e else if e suas aplicações no R.
- E, finalmente, conhecemos os benefícios das estruturas de repetição: while e for.

#### O que aprendemos até agora

- No último encontro, aprendemos a importar e exportar bases de dados no R, principalmente com o pacote readr: read\_csv e write\_csv.
- Vimos as vantagens de se utilizar o pacote janitor para alterar a estética de nossas tabelas: clean\_names().
- E, por fim, exploramos a diversidade de visualizações possíveis através da utilização da gramática de gráficos do *ggplot2*.

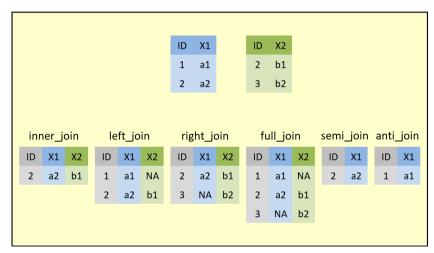
# Combinando data frames

- Combinar data frames é necessário quando as informações que serão utilizadas na análise estão presentes em mais de uma fonte de dados.
- As variáveis usadas para conectar um par de data frames são chamadas de chaves. Uma chave é uma variável (ou conjunto de variáveis) que identifica exclusivamente uma observação. Em casos simples, uma única variável é suficiente para identificar uma observação.

#### Tipos de join

- Há 6 tipos de "joins" em R:
  - Interseção dos dois conjuntos, x e y: inner\_join();
  - Mantém todas as observações em x, ou seja, à esquerda: left\_join();
  - Mantém todas as observações em y, ou seja, à direita: right\_join();
  - Mantém todas as observações em x e y: full\_join();
  - Mantém todas as observações em x que correspondem a y: semi\_join(); e
  - Elimina todas as observações em x que correspondem a y: anti\_join().

#### Visão geral das funções de join do dplyr



R e Power BI

- É possível importar e modelar dados dentro do Power BI com scripts de R, tais quais os que trabalhamos no curso, além de produzir gráficos ou rodar algoritmos implementados na linguagem.
- Entre as vantagens dessa integração, destacam-se, sobretudo:
  - A possibilidade de criar visualizações personalizadas, que não estão disponíveis nativamente no Power BI;
  - A vasta biblioteca de pacotes estatísticos do R que podem ser incorporados ao Power BI; e
  - Importação e análise de dados de diferentes fontes diretamente no Power BI.

#### Laboratório

Agora é o momento de partir para a ação! Temos dois tutoriais agendados para hoje, prontos para serem explorados. Eles estão disponibilizados no GitHub, e você pode acessá-los clicando neste **link**.

Se deixou algum tutorial inacabado dos encontros anteriores, comece por eles. Caso contrário, prossiga.

# Dúvidas?