

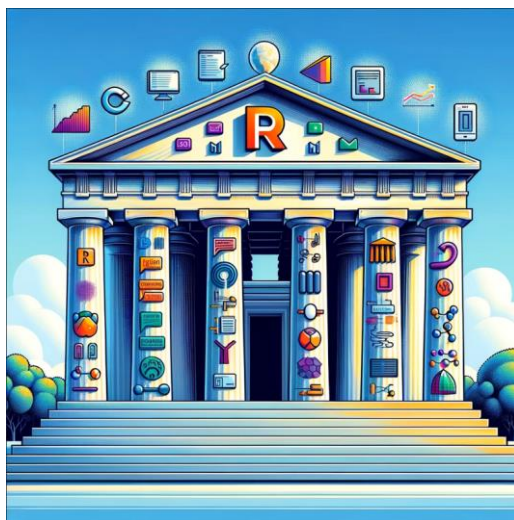
**c(co)Cursos Masters**  
**Disciplina: Programação em R**  
**Professor/a(s): Adriana Bezerra Bessa**  
**Monitor/a(s):**

**1º SEMESTRE DE 2024 – 1º TRIMESTRE**

**WORKBOOK DE ATIVIDADES – VERSÃO ALUNO/A**

**AULA 01: A Linguagem R: uma introdução e seus fundamentos**

## **MOTIVAÇÃO**



Qualquer aprendizado tem como ponto de partida a construção de alicerces sólidos, baseado na apresentação de conceitos fundamentais para que seja possível na sequência progredir na evolução do aprendizado. Tal fato não é diferente para o estudo da linguagem R

A linguagem R é uma ferramenta poderosa para análise de dados, que permite manipular, visualizar e modelar dados de forma eficiente. A primeira aula da linguagem R fornece uma

introdução aos principais conceitos da linguagem, incluindo operadores, tipos de dados, estruturas de dados e operações básicas.

### ATIVIDADE EM GRUPO

As atividades serão feitas em grupos, os quais serão formados aleatoriamente por aula. A definição da quantidade grupos será estabelecido no início do curso, pois depende do total de inscritos na disciplina. Cada grupo poderá ter entre 3 ou 4 alunos no máximo.

#### 1. Tipo de variáveis

Crie um data frame com 3 colunas aonde:

- coluna1: números inteiros de 1 a 3.
- coluna2: números (float) de 1 a 3.
- coluna3: caracteres 1 a 3.

#### 2. Operações

Agora faça as seguintes operações:

- Parte I: Faça a soma da coluna1 com a coluna1. Qual o tipo de resultado? (Inteiro, Float ou caractere).
- Parte II: Faça a soma da coluna1 com a coluna2. Qual o tipo de resultado? (Inteiro, Float ou caractere).
- Parte III: Faça a soma da coluna2 com a coluna3. Qual o tipo de resultado? (Inteiro, Float ou caractere).

#### 3. Navegando no Dataframe

Primeiramente vamos converter o banco de dados “AirPassengers” para um data.frame.

- O nome do banco convertido será “data”.
- Vamos nomear a coluna do banco “data” de “Passengers”.
- Vamos criar uma nova coluna chamado “date” com as datas associadas.
- Vamos converter os dados da coluna “Passengers” para o tipo “numeric”.

### 4. Funções básicas

Utilizando o banco de dados iris, crie uma cópia do banco chamada de data. No banco de dados “data” crie 2 novas variáveis:

- variável 1 ( $\ln \text{Sepal.Length}$ ): deve ser o logaritmo neperiano da variável “Sepal.Length”.
- variável 2 ( $\exp \text{Sepal.Length}$ ): deve ser o exponencial neperiano da variável “Sepal.Length”.

#### Entrega:

- Um script R contendo a resposta para cada um dos itens da atividade em grupo acima.
- O arquivo deve ser postado na pasta “Entrega de Atividade” no Teams, ao final da aula.

## AULA 02: Manipulação de dados no R: da carga ao processo de transformação

### MOTIVAÇÃO

A qualidade dos dados é a pedra fundamental de qualquer projeto de análise de dados; dados ruins resultam em análise incorretas, modelos limitados e ineficazes.



Neste sentido, a aula de manipulação de dados em R ensina aos alunos como preparar dados para análise e modelagem. O processo de data wrangling engloba as etapas de carga, limpeza e transformação de dados.

O pacote tidyverse é uma ferramenta poderosa para data wrangling, fornecendo funções para importar dados, identificar e corrigir erros, e transformar dados em um formato adequado.

Ao final da aula, os alunos estarão preparados para realizar data wrangling de forma eficiente e eficaz.

### ATIVIDADE EM GRUPO

As atividades serão feitas em grupos, os quais serão formados aleatoriamente por aula. A definição da quantidade grupos será estabelecida no início do curso, pois depende do total de inscritos na disciplina. Cada grupo poderá ter entre 3 ou 4 alunos no máximo.

#### 1. Download de arquivos para base de dados

Acesse o site do Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org/>). Faça download das series:

- Production of Total Industry in Germany (DEUPROINDMISMEI) em formato XLS.

- Inflation, consumer prices for Germany (FPCPITOTLZGDEU) em formato CSV.

## 2. Carga de dados

- Utilize a biblioteca readxl para fazer a importação do da base de dados da Produção Industrial da Alemanha. Informe o range de dados (ex: "A11:B741"), bem como o tipo de dado de cada coluna observation date: date e DEUPROINDMISMEI: numeric). Chame a base de GER PI.
- Troque o nome das colunas da base GER PI para data e prodInd.
- Utilize a biblioteca readr para fazer a importação do da base de dados da Inflação e preços do consumidor da Alemanha. Informe o tipo de dado de cada coluna DATE: date e FPCPI- TOTLZGDEU: double). Chame a base de GER Price.
- Troque o nome das colunas da base GER Price para data e price.

## 3. Agrupamento de base

Note que a base GER PI, carregada no exercício anterior, está em periodicidade mensal, enquanto a base GER Price, carregada no exercício anterior, está em periodicidade anual. Vamos transformar a base de periodicidade mensal para periodicidade anual.

- Crie uma coluna com a informação do "ano" de cada registro (dica: para isso utilize a biblioteca lubridate).
- Crie um data frame com a informação de agrupamento.
- Agrupe os dados da base GER PI por ano, tirando a média da variável "prodInd" (dica: para isso utilize o summarise do pacote dplyr).
- Salve os resultados do agrupamento em um data frame com nome "GER PI.Anual".

## 4. Transformação de dados

Uma transformação comum é criar uma variável nova que é o logaritmo de uma outra variável.

- Utilizando a base GER PI.Anual, crie uma nova variável que é o logaritmo neperiano da variável "ProdIndAnual" (chame a variável de "ln prod").
- Utilizando a base GER Price, tente criar uma nova variável que é o logaritmo neperiano da variável "price". É mostrado alguma mensagem? Caso afirmativo, por que a mensagem aparece?
- Utilizando a base GER Price, mostre as "Estatísticas descritivas" das colunas, utilizando a função summary. Existe valores negativos?
- Utilizando a base GER Price, filtre as linhas que tem valores de preço abaixo de zero.

### Entrega:

- 
- Um script R contendo a resposta para cada um dos itens da atividade em grupo acima.
  - O arquivo deve ser postado na pasta “Entrega de Atividade” no Teams, ao final da aula.

## AULA 03: A construção de gráficos com a linguagem R

### MOTIVAÇÃO



A visualização de dados desempenha um papel crucial na análise e na comunicação de informações complexas. Gráficos bem projetados têm o poder de simplificar conceitos complexos, revelar tendências e insights ocultos e transmitir informações de forma mais eficaz do que uma descrição textual.

No mundo da análise de dados, as visualizações são uma das principais ferramentas utilizadas. As visualizações dão vida aos dados, permitindo que os padrões se tornem óbvios, tendências e valores discrepantes que, de outra forma, poderiam permanecer ocultos.

### ATIVIDADE EM GRUPO

As atividades serão feitas em grupos, os quais serão formados aleatoriamente por aula. A definição da quantidade grupos será estabelecida no início do curso, pois depende do total de inscritos na disciplina. Cada grupo poderá ter entre 3 ou 4 alunos no máximo.

#### 1. Carga de dados e construção de gráfico

A “Crise das Tulipas”, foi um período de especulação desenfreada e colapso econômico que ocorreu nos Países Baixos no século XVII, durante o auge da popularidade das tulipas como plantas ornamentais.

Neste exercício vamos gerar uma visualização com o objetivo de mostrar como as flores se distribuem em função das suas características básicas como comprimento e largura da sépala.

Para isso vamos utilizar a base de dados Iris. Esta base é um conjunto clássico de informações botânicas que contém medidas de diferentes espécies de flores.

- Carregue o pacote do ggplot.
- Carregue a base de dados Iris.
- Crie um gráfico de dispersão com as seguintes propriedades:
- A variável “Sepal.Length” deve ser colocada no eixo x.
- A variável “Sepal.Width” deve ser colocada no eixo y.
- Cada ponto deve ser colorido baseado na variável “Species”.
- O título do gráfico deve ser “Sepal Length vs Sepal Width”.
- O subtítulo do “Iris Database”.
- O caption do gráfico deve ser “Fonte: Iris database (R-Studio)”.
- O nome do eixo x deve ser “Sepal Length”.
- O nome do eixo y deve ser “Sepal Width”.

### **Entrega:**

- Arquivo no formato ‘png’ com as dimensões 800 x 600 pixels.
- O arquivo deve ser postado na pasta “Entrega de Atividade” no Teams, ao final da aula.



### AULA 04: Criando um projeto completo no R

#### MOTIVAÇÃO



Essa última aula de programação em R tem como objetivo consolidar o aprendizado dos alunos, evidenciando estruturas de controle, funções e permitindo que eles apliquem seus conhecimentos em um projeto completo. Será possível, de forma sucinta, iniciarmos e finalizarmos um pequeno projeto de análise de dados.

As estruturas de controle são essenciais para controlar o fluxo de execução do código, permitindo a tomada de decisões e a iteração sobre conjuntos de dados. As funções desempenham um papel vital na simplificação e organização do código, tornando-o mais claro e menos suscetível a erros.

O mini-projeto final permitirá aos alunos aplicar todo o conhecimento adquirido ao longo do curso, abordando a maioria dos aspectos e pacotes explorados durante o curso e contendo todas as camadas abordadas: a camada de dados, de lógica e do usuário. O projeto consiste em fazer a análise dos dados da COVID-19.

Ao final da aula, os alunos terão uma compreensão abrangente da programação em R e estarão preparados para enfrentar desafios reais no desenvolvimento de projetos e análises de dados.

#### ATIVIDADE EM GRUPO

As atividades serão feitas em grupos, os quais serão formados aleatoriamente por aula. A definição da quantidade grupos será estabelecida no início do curso, pois depende do total de inscritos na disciplina. Cada grupo poderá ter entre 3 ou 4 alunos no máximo.

##### 1. Criando um projeto no R

Crie um projeto no RStudio com a seguinte estrutura:

- Um diretório chamado: “Database”.
- Um diretório chamado: “Graficos”.

- Baixe a base de dados do site <https://github.com/owid/covid-19-data/tree/master/public/data>.
- Salve a base no diretório "Database" com o nome "owid-covid-data.xlsx".

Caso você não esteja utilizando o R-Studio apenas organize seu diretório de maneira a seguir a estrutura sugerida.

### 2. Camada de dados(Data Loader)

Crie um script chamado "DataLoader.r", o script deve realizar as seguintes tarefas:

- Abrir o banco de dados de nome "owid-covid-data.xlsx" que está no diretório "Database".
- A coluna data deve ser convertida para o tipo data. (garantindo que a coluna é do tipo "Date").
- O script deve filtrar a base e mostrar apenas os seguintes países: "Brazil", "United States", "Mexico", "Germany", "France", "United Kingdom".
- O script deve selecionar apenas as variáveis: location, date, total cases, new cases.

### 3. Função

- Crie uma função que recebe um vetor com datas e determina o dia da semana bem como faz uma contagem das semanas e salve em um script separado.
- A função deve retornar um data.frame com as seguintes colunas: date, weekday, week. Exemplo:

date	weekday	week
2020-01-01	4	1
2020-01-02	5	1
2020-01-03	6	1
2020-01-04	7	1
2020-01-05	1	2
2020-01-06	2	2
2020-01-07	3	2
2020-01-08	4	2
2020-01-09	5	2
2020-01-10	6	2
2020-01-11	7	2
2020-01-12	1	3

### 4. Camada de lógica

Crie um script chamado "main.r", o script deve fazer as seguintes tarefas:

- Carregar a função criada anteriormente.
- Carregar o script DataLoader.r.
- Utilizar a função criada para determinar qual a semana associada a cada registro na base "covid data". (dica utilizar join).
- Excluir a informação da ultima semana pois essa pode estar incompleta.

### 5. Camada de usuário

Crie uma rotina que cria 2 gráficos e salva eles no diretório "Graficos".

- Gráfico 1: utilizando os dados agrupados, crie um gráfico de linha do total de mortes por covid- 19, com uma série para cada país selecionado. (eixo x: Semana, eixo y: Total de mortes).
- Gráfico 2: utilizando os dados agrupados, crie um gráfico de linha das novas de mortes por covid-19, com uma série para cada país selecionado. (eixo x: Semana, eixo y: Novas mortes)
- Salve os dois gráficos no diretório "Graficos". (Configurações recomendadas: scale=1, units = "in", dpi = 300, width = 10.4, height = 5.85).

### Entrega:

- O script main.r com todas as atividades requisitadas e os dois gráficos solicitados.
- O arquivo deve ser postado na pasta " Entrega de Atividade" no Teams, ao final da aula.