

# Exercício Programa (EP) 02

*Ana C V de Melo - Alexandre Locci*

*MAC113 - FEA diurno – 2020*

## 1 EP: Informações sobre a Covid-19 no mundo

Este EP utiliza um conjunto de informações disponibilizadas pelo “European Union Open Data Portal” (<https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data>).

## 2 Preparação

1. Abra o RStudio.
2. Na janela de *Files* (canto direito inferior) crie um diretório com seu nome e/ou identificação do EP.
3. Copie do *moodle* todos os arquivos disponíveis para o EP atual.
4. Atualize o diretório de execução do RStudio para a sessão atual: clique em *Session/Set Working Directory/Choose Directory...* e escolha o diretório onde colocou os seus arquivos.
5. Abra o script disponível e pode iniciar a solução do EP.

## 3 Sobre este EP

Apesar de existirem atualmente vários dados sobre a Covid-19, muitas vezes queremos informações que não são exibidas diretamente nos sites que armazenam os dados.

### 3.1 Os dados utilizados:

1. Planilha “*mundoCovid19.csv*”, na qual temos as seguintes colunas de informações (ela está escrita em inglês):
  - **dateRep**: data completa (dia/mês/ano)
  - **day**: dia
  - **month**: mês
  - **year**: ano
  - **cases**: número de casos confirmados
  - **deaths**: número de óbitos confirmados
  - **countriesAndTerritories**: país
  - **geoId**: identificação geográfica
  - **countryterritoryCode**: código do país
  - **popData2018**: população em 2018
  - **continentExp**: continente do país

Antes de iniciar a solução, veja a planilha para entender as informações disponíveis.

## 3.2 Objetivo do EP

Extrair novos dados juntando informações dessa planilha. Muitas vezes nos fazemos perguntas sobre a situação atual da pandemia. Em geral, muitas das nossas perguntas podem ser respondidas a partir de informações que estão disponíveis nas planilhas existentes, mas, para isso, precisamos processar essa informação.

Neste EP, vamos construir novas funções que trazem respostas para algumas perguntas relativas à pandemia.

## 4 Sua Tarefa será:

O seu script deverá implementar as funções para cada um dos exercícios definidos abaixo. **Não mude o nome das funções fornecidas no script** porque precisaremos desses nomes para conseguir corrigir os EPs. V. precisa apenas complementar cada uma das funções pedidas.

Inicialmente o script lê a planilha (já implementado no script fornecido):

- “`mundoCovid19.csv`” (descrita acima)

**obs:** Usaremos em especial a configuração ( `stringAsFactors = FALSE`) para que os nomes dos países e continentes sejam tratados como caracteres puros (não retire isso do script, é para facilitar as operações que vocês irão realizar).

### 4.1 Exercício 1

Ler a planilha dada e criar o respectivo data-frame: código disponível no script fornecido. Você só precisa executar essa parte do script.

```
df_covidMundo <- read.csv(file = "mundoCovid19.csv",
                           header = TRUE, sep = ",",
                           fill = TRUE,
                           stringsAsFactors = FALSE) # lê a planilha)
```

### 4.2 Exercício 2

Quando trabalhamos com dados, principalmente em grande quantidade, precisamos fazer alguns ajuste nestes dados para que não atrapalhem o processamento. Sendo assim, faça uma função (`subs_NA_por_zero`) que dada uma coluna de um data-frame, substitui todos os objetos “NA” pelo valor 0. Iremos substituir todos os objetos “NA” da coluna “deaths” do data-frame “`df_covidMundo`” pelo valor 0 (alguns valores iniciais estavam vazios).

```
head(df_covidMundo)
```

```
##      dateRep day month year cases deaths countriesAndTerritories geoId
## 1 18/05/2020  18    5 2020   262      1      Afghanistan      AF
## 2 17/05/2020  17    5 2020     0     NA      Afghanistan      AF
## 3 16/05/2020  16    5 2020  1063    32      Afghanistan      AF
## 4 15/05/2020  15    5 2020   113     6      Afghanistan      AF
## 5 14/05/2020  14    5 2020   259     3      Afghanistan      AF
## 6 13/05/2020  13    5 2020   280     5      Afghanistan      AF
## countryterritoryCode popData2018 continentExp
## 1                AFG   37172386          Asia
## 2                AFG   37172386          Asia
## 3                AFG   37172386          Asia
## 4                AFG   37172386          Asia
```

```
## 5          AFG      37172386      Asia
## 6          AFG      37172386      Asia
```

```
df_covidMundo["deaths"] <- subs_NA_por_zero(df_covidMundo$deaths)
head(df_covidMundo)
```

```
##      dateRep day month year cases deaths countriesAndTerritories geoId
## 1 18/05/2020 18     5 2020    262      1      Afghanistan      AF
## 2 17/05/2020 17     5 2020     0      0      Afghanistan      AF
## 3 16/05/2020 16     5 2020   1063    32      Afghanistan      AF
## 4 15/05/2020 15     5 2020    113     6      Afghanistan      AF
## 5 14/05/2020 14     5 2020    259     3      Afghanistan      AF
## 6 13/05/2020 13     5 2020    280     5      Afghanistan      AF
##  countryterritoryCode popData2018 continentExp
## 1                    AFG      37172386      Asia
## 2                    AFG      37172386      Asia
## 3                    AFG      37172386      Asia
## 4                    AFG      37172386      Asia
## 5                    AFG      37172386      Asia
## 6                    AFG      37172386      Asia
```

A função `head()` imprime na tela as primeiras linhas de um data-frame...

Observe que na linha 2 o NA foi substituído por zero!

### 4.3 Exercício 3

Qual o número total de óbitos nos países africanos (África) durante o mês de maio (mês 5) de 2020?

Faça uma função, que dados como parâmetros de entrada um data-frame com as informações existentes no `df_covidMundo`, o continente, o mês e o ano, retorne o total de óbitos no continente naquele dado mês.

```
resposta <- obitos_continente_mes(df_covidMundo, "Africa", 5, 2020)
cat("Número de óbitos na África durante o mês de Maio/2020:", resposta)
```

```
## Número de óbitos na África durante o mês de Maio/2020: 1168
```

Veja que podemos fazer perguntas semelhantes para outros continentes e datas. E o melhor é que poderíamos usar a mesma função para responder a essas novas perguntas.

### 4.4 Exercício 4

Dado o avanço da epidemia no mundo no mês de abril, gostaríamos de saber quantos óbitos tivemos no dia 26/4/2020.

Para responder a essa pergunta, você deverá fazer uma função (`obitos_data`) que tem como parâmetros de entrada um data-frame com as informações existentes no `df_covidMundo`, o dia, o mês e o ano, e dá como resultado o total de óbitos naquele dia.

```
resposta <- obitos_dia(df_covidMundo, 26, 4, 2020)
cat("Número de óbitos no mundo no dia 26/04/2020: ", resposta)
```

```
## Número de óbitos no mundo no dia 26/04/2020: 6246
```

Usando essa mesma função, você poderá verificar o número de óbitos em outras datas...

## 4.5 Exercício 5

Apesar de saber que os dados apresentados na planilha *mundoCovid19.csv* depende de como os países notificaram os casos e óbitos, seria interessante saber qual o país notificou o maior número de óbitos em um único dia.

Para responder a essa pergunta, você deverá fazer uma função (`maior_obitos_pais_data`) que tem como parâmetros de entrada um data-frame com as informações existentes no `df_covidMundo`, e dá como resultado o número de óbitos, a data e o país, onde foi notificado o maior número de óbitos até o momento (considerando até a data existente na planilha).

```
resposta <- maior_obitos_pais_data(df_covidMundo)
cat("O maior número de óbitos (", resposta[1],") \n \t ocorreu em ", resposta[2],
    ", \n \t no país: ", resposta[3])
```

```
## O maior número de óbitos ( 4928 )
##      ocorreu em 16/04/2020 ,
##      no país:  United_States_of_America
```

## 4.6 Exercício 6

Considerando as notificações registradas na planilha, em qual dia houve o maior número de óbitos na europa (“Europe”)?

Você deverá fazer uma função (`maior_obitos_continente`) que tem como parâmetros de entrada um data-frame com as informações existentes no `df_covidMundo` e um continente, e dá como resultado a data que em houve o maior número de óbitos naquele continente e o número de óbitos (considerando até a data existente na planilha).

Dica: veja que neste problema você precisará somar todos os óbitos na Europa em cada uma das datas disponíveis... A partir disso, poderá verificar qual dia teve o maior número de óbitos.

```
resposta <- maior_obitos_continente(df_covidMundo, "Europe")
cat("Dia com o maior número de óbitos na Europa: ", resposta[1],
    "\n \t número de óbitos:", resposta[2])
```

```
## Dia com o maior número de óbitos na Europa: 04/04/2020
##      número de óbitos: 5140
```

## 4.7 Exercício 7

Considerando que já sabemos como calcular em qual dia tivemos o maior número de óbitos em um continente ( *Exercício 6* ), gostaríamos de saber qual país daquele continente registrou o maior número de óbitos naquele dia. Gostaríamos de saber, por exemplo, qual país europeu teve o maior número de óbitos no dia em que houve o maior número de óbitos em toda a europa?

Você deverá fazer uma função (`pais_maior_obitos_continente_data`) que tem como parâmetros de entrada um data-frame com as informações existentes no `df_covidMundo`, o continente e uma data, e dá como resultado o país pertencente ao continente com o maior número de óbitos naquela data, junto o número de óbitos do país.

```
data_obitos <- maior_obitos_continente(df_covidMundo, "Europe")
data <- data_obitos[1]
obitos <- data_obitos[2]
resposta <- pais_maior_obitos_continente_data(df_covidMundo, "Europe", data)
cat("\n Número total de óbitos na Europa no dia", data, "--" , obitos)
```

```
##
## Número total de óbitos na Europa no dia 04/04/2020 -- 5140
cat("País europeu com o maior número de óbitos nesta data: ", resposta[1],
    "\n \t com um total de óbitos: ", resposta[2])

## País europeu com o maior número de óbitos nesta data:  France
##          com um total de óbitos:  2004
```

#### 4.8 Dicas importantes:

- Não modifique o que está escrito no script fornecido. Ele foi feito para ajudá-los a resolver o problema.
- Resolva o problema na ordem em que é sugerida. Alguns precisam de conhecimentos dos exercícios anteriores.

#### 4.9 Exemplo de Execução e como o seu EP será avaliado:

- Exemplos de execução mostrados em cada exercício.
- a correção será feita mediante comparação dos resultados gerados pelo seu EP com os resultados gerados pelos nossos scripts para os dados atualizados. Vamos basicamente fazer novas perguntas usando as funções que foram pedidas em cada exercício e comparar com os nossos resultados.

#### 4.10 O que está sendo fornecido no *moodle*

- a planilha:
  - a. *covidMundo19.csv*
- o script da solução parcial do EP
  - a. *EP02.script.R*

#### 4.11 O que V. deve entregar:

- V. deve gravar o seu script solução no arquivo *EP02.script.R* (parte da solução já está no script fornecido). **ESTE É O ÚNICO ARQUIVO A SER ENVIADO NO MOODLE - NÃO ENVIE ARQUIVOS .ZIP.**
- Envie o seu arquivo solução para o *moodle* (**precisa ser esse arquivo mesmo e só ele**). Tenha certeza de que você gravou todas as modificações que fez no arquivo antes de entregar.
- Guarde uma cópia para você de todas as listas e EPs que fizer durante a disciplina.
- Observe a **data de entrega**. Só serão recebidos os EPs (pelo próprio sistema) até aquela data.