

Exercício Programa (EP) 01

Ana C V de Melo - Alexandre Locci

MAC113 - FEA diurno – 2020

1 EP: Extratos de Informação da Covid-19 no Estado de São Paulo

Este EP utiliza um conjunto de informações disponibilizadas pela Fundação SEADE (Fundação Sistema Estadual de Análise de Dados), os quais foram acessadas no dia 27/abril/2020.

Na Fundação SEADE (<https://www.seade.gov.br/>), vários dos dados utilizados aqui para o exercício são disponibilizados de forma gráfica e com outros processamentos. Contudo, podemos ainda extrair novas informações a partir dos dados disponibilizados.

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, em particular para o estado de São Paulo, muitas vezes queremos informações que não são exibidas diretamente nos sites que armazenam os dados.

3.1 Os dados utilizados:

Como já mencionado, utilizaremos dados fornecidos na Fundação SEADE. As informações constam de 2 planilhas extraídas do site da SEADE:

1. Planilha “pop.SP.2020.cvs”, na qual temos as seguintes colunas de informações:
 - **Localidades:** nomes dos municípios de SP
 - **Pop:** população total do município
 - **Pop Masc:** população do sexo masculino no município
 - **Pop Fem:** população do sexo feminino no município
 - **Pop Urbana:** população que vive na zona urbana
 - **Pop Rural:** população que vive na zona rural
 - **Grau Urban(porc):** taxa de urbanização do município (em porcentagem)
2. Planilha “covid.municipios.SP.cvs”, na qual temos as seguintes colunas de informações:
 - **Município:** nomes dos municípios onde se tem registrado pelo menos 1 caso da Covid-19
 - **Grande regioao:** tipo de região (interior, RMSP - Região Metropolitana, Capital)
 - **Num Casos:** número de casos confirmados da Covid-19 no município
 - **Num obitos:** número de óbitos pela Covid-19 registrados no município

Os nome de municípios na coluna **Município** da planilha “covid.municipios.SP.csv” é um subconjunto (apenas os que têm casos confirmados) dos nomes dos municípios da coluna **Localidades** da planilha “pop.SP.2020.csv”, a qual possui informação sobre a população dos municípios. Antes de iniciar a solução, veja cada uma das planilhas para entender as informações disponíveis.

3.2 Objetivo do EP

Extrair novos dados juntando informações dessas duas planilhas.

Não estão explícitas, nas planilhas, as informações relativas aos infectados e óbitos em municípios que são predominantemente rurais ou urbanos. Queremos, com este EP, fornecer novas planilhas relativas a extratos de informações para os municípios rurais e urbanos. As informações necessárias e as tarefas a serem realizadas estão descritas no próximo item.

3.3 Sua Tarefa será:

Implementar um programa que:

a. lê inicialmente as 2 planilhas:

- “covid.municipios.SP.csv” (descrita acima) e
- “pop.SP.2020.csv” (descrita acima)

obs: no script fornecido existem dicas para essa leitura. Usaremos em especial a configuração (*as.is = TRUE*) para que os nomes dos municípios sejam tratados como caracteres puros (não retire isso do script, é para facilitar as operações que vocês irão realizar).

b. gera um novo data-frame com as seguintes informações extraídas das planilhas lidas (precisa ser exatamente essa ordem de colunas):

- **Município:** nomes dos municípios onde se tem pelo menos 1 caso da Covid-19
- **Grande regioao:** tipo de região (interior, RMSP - Região Metropolitana, Capital)
- **Num Casos:** número de casos confirmados da Covid-19
- **Num obitos:** número de óbitos registrados pela Covid-19
- **Pop:** população total do município
- **Grau.Urb:** taxa de urbanização do município (em porcentagem)
- **infec.100mil:** número de infectados por Covid-19 no município para cada 100.000 habitantes
- **obitos.100mil:** número de óbitos por Covid-19 no município para cada 100.000 habitantes

c. gera um extrato do data-frame do item b. apenas com os municípios que são predominantemente rurais (taxa de urbanização do município menor que 50%). Este data-frame contém todas as colunas do data-frame do item b.

d. grava em uma planilha (“plan.covid.rural.csv”) o data-frame criado no item c. A planilha gravada deve conter os nomes das colunas, não deve conter os números das linhas e nem aspas (“”) quando escrevemos os dados. Um formato parecido com as planilhas que foram fornecidas para este EP (o script tem algumas dicas).

e. gera um extrato do data-frame do item b. apenas com os municípios que são predominantemente urbanos (taxa de urbanização do município maior ou igual a 50%). Este data-frame contém todas as colunas do data-frame do item b.

f. grava em uma planilha (“plan.covid.urbana.csv”) o data-frame criado no item e. A planilha gravada deve conter os nomes das colunas, não deve conter os números das linhas e nem aspas (“”) quando escrevemos os dados. Um formato parecido com as planilhas que foram fornecidas para este EP (o script tem algumas dicas).

3.4 Dicas importantes:

- Não modifique o que está escrito no script fornecido. Ele foi feito para ajudá-los a resolver o problema.
- Para resolver este EP, vocês precisam apenas do conhecimento até as aulas sobre data-frames. Dá para resolver tudo só com as manipulações que vocês já conhecem sobre data-frames. Isso não impede que usem conhecimentos adquiridos nas próximas aulas.
- Muito Importante: **quando as planilhas são fornecidas, nem sempre elas estão ordenadas pela mesma coluna.** Por exemplo, temos **Localidades** e **Municípios** nas duas planilhas distintas que tratam da mesma informação (nomes de municípios). As planilhas não necessariamente estão ordenadas da mesma forma para essas colunas. Então, se quiser extrair informação dos data-frames da população e da Covid-19 dos municípios, eles precisam estar ordenados da mesma forma. No script há uma sugestão de ordenar os data-frames pelas colunas correspondentes aos municípios.

3.5 Exemplo de Execução e como o seu EP será avaliado:

- considerando as planilhas fornecidas como entrada, seu EP deverá gerar 2 novas planilhas com os conteúdos iguais aos que estão nas planilhas *exemplo.plan.covid.rural.csv* e *exemplo.plan.covid.urbana.csv*.
- as planilhas aqui fornecidas foram extraídas do site no dia 27/abril/2020. Para a correção, utilizaremos novas planilhas com os dados atualizados.
- a correção será feita mediante comparação das 2 planilhas geradas pelo seu EP com as planilhas que deveriam gerar para os dados atualizados.

3.6 O que está sendo fornecido no *moodle*

- as planilhas:
 - a. *pop.SP.2020.csv*
 - b. *covid.municipios.SP.csv*
 - c. *exemplo.plan.covid.rural.csv* e *exemplo.plan.covid.urbana.csv* apenas para V. conferir se a resposta do seu EP está correto. Com as planilhas dadas nos itens a. e b., ele deveria gerar essas duas planilhas
- o script da solução parcial do EP
 - a. *EP01.script.R*

3.7 O que V. deve entregar:

- V. deve gravar o seu script solução no arquivo *EP01.script.R* (parte da solução já está no script fornecido).
- 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.