

Exame de R/

Vitor Pereira

15/09/2022

Exame de R

- Bom dia! Não se assustem com o exame! Não consultem ostensivamente os colegas e, se precisar falar com alguém, fale baixo. O exame é individual.
- Vocês podem consultar todos os copitos disponíveis.
 - O copito que fizemos em Markdown está em `R_STP/markdown/Aula_R`. Podem acessar a versão HTML no navegador ou em pdf.
 - Podem acessar também o livro: www.livro.curso-r.com
 - Podem também acessar todos os copitos disponíveis no R em `Tools/Cheatsheet`
- Consultem os códigos das aulas anteriores e os códigos que estão na nossa apresentação em markdown. Podem copiar a vontade os códigos.
- Respondam atentamente a todas as questões.
- Nem todas as questões precisam ser respondidas em sequência. Se estás com dificuldade em uma questão, passe para a próxima e depois volte.
- O exame deverá ser respondido em um arquivo RMarkdown. Preferencialmente em PDF, mas para aqueles que não instalaram o MikTex, podem fazer em HTML.
- Certifique-se de que você entrou no R pelo projeto `R_STP`. Faça um Git Pull.
- Para criar um arquivo RMarkdown, vá em `File/New File/RMarkdown`. Escolha a opção documento. Salve imediatamente o documento na pasta Exame, com seu nome. Não faça o Git Push até ser autorizado.
- Durante o exame, não irei verificar os erros. Copie e cole as mensagens de erro no google, use o google translate e tente interpretar o erro.
- Coloque seu nome tanto no cabeçalho do RMarkdown como na primeira linha de texto que irá escrever no exame. Se eu não conseguir identificá-lo, não terei como corrigir.

Questão 1

- Comece um chunk de R no Markdown. Ao longo do exame, todos os chunks terão apenas {r} precedido das 3 aspas. Não se esqueça de colocar as 3 aspas ao final do chunk.
- Comece colocando o seguinte chunk de R no seu markdown, depois de escrever seu nome.

```
knitr::opts_chunk$set(echo = TRUE)
```

- Abra outro chunk e inclua um cabeçalho com os caminhos e o pacman. No pacman, você irá carregar os seguintes pacotes: dplyr, ggplot2, sf, readxl, readr, janitor, epiDisplay e tidyr. Não se esqueça de configurar a hora de São Tomé e Príncipe.
- Abra outro chunk de R. Nesse chunk, você deverá abrir a base “pre_escolar_2021-2022.xlsx”, que está na pasta R_STP/input. “Receba” esta base com o nome de base_pre.
- Mostre as 5 primeiras linhas da base. Dica: Olhe para nossa primeira aula (R_STP/aula1_1.R).
- Mostre um resumo, um sumário da base

Questão 2- Limpeza

- Limpe os nomes das colunas da base através da função `clean_names()`
- Retire as colunas vazias

Dica: Olhe o código “`R_STP/code/2_abre_bases.R`” (linha 30)

Questão 3- Limpeza

- Através do comando `tab1`, do pacote `epiDisplay`, mostre a distribuição do período da aula. (Lembre-se do cifrão (`$...`)). Olhe a `aula1_1` como usar o `tab1`.
- Quantas grafias diferentes você encontrou para o período da manhã? E para o período da tarde?
- Crie uma coluna adicional que será igual a 1 se for do período da manhã, e zero caso contrário. Essa variável terá o nome de `manha`, sem acento. Pode ignorar os NAs, ou seja, eles podem ser todos zero.
- Dica: Utilizem o `mutate(case_when)`. Olhem na apresentação como fazer.
- Dica: Para facilitar, podem usar `TRUE ~ 0`

Questão 4- Manipulação

- Utilize a base que está em “R_STP/tmp/base_pre4.RData”. Veja como abrir uma base em RData na apresentação. “Receba” esta base como base_pre5.
- Retire da base todos as crianças que estão a estudar pela 3a vez (no_de_frequencia). Não retire da base as que estão a estudar por mais de 3 vezes.
- Dica. Filtre a base.
- Salve a base como base_pre6.

Questão 5- Manipulação

- Utilize novamente a base que está em “R_STP/tmp/base_pre4.RData”. Veja como abrir uma base em RData na apresentação. “Receba” novamente esta base como base_pre5.
- Gere uma base com:
 1. O total de alunos por distrito/regiao, escola e por gênero
 2. A média de idade dos alunos por escola e por gênero
 3. O percentual de alunos que frequenta a educação infantil pela manhã, por escola e por gênero.

Dica: Utilize a variável periodo_num, para facilitar sua vida.

Dica: Não se esqueça de atentar para os missing values (... , na.rm=TRUE)

- Receba esta base como base_pre7.

Questão 6- Manipulação

- Utilize a base 7 da questão anterior.
- Queremos que a base esteja organizada da seguinte maneira:
 1. Cada linha deve ser uma combinação de distrito e escola
 2. Teremos 6 colunas, além da de escola: (Total de alunos, masculino; Total de alunos, feminino; média de idade, masculino; média de idade, feminino; percentual da manhã, masculino; percentual da manhã, feminino)
- Pense antes. Queremos pivotar essa base para long ou wide? Lembre-se, long [é mais vertical, e wide é mais larga. De onde virão os nomes? De onde virão os valores?
- Olhe o código da apresentação para ver a sintaxe.
- Receba a nova base com o nome de base_pre_8

Questão 7- Gráfico

- Abra a base que está em tmp, como o nome de “base_pre8.RData”. Receba com o nome de base_pre9
- Faça um gráfico de pontos, com o total de alunos do sexo masculino no eixo x, e a média de idade das alunas (feminino), no eixo y.
- O tamanho de cada ponto deverá ser proporcional ao total de alunas do sexo feminino
- As cores dos pontos deverao indicar o distrito.
- Não é necessário mexer nos eixos nem no título

Questão 7- Mapa

- Abra o shapefile de STP através da linha de código abaixo

```
maps_STP <- st_read(paste0(input,"shapes/stp_admbnda_adm2_gadm_2020.shp" ))
```

- Crie a base_pre10, recebendo a base_pre9 e incluindo uma variável com os códigos dos distritos na base do shapefile.

Dica: Olhe na apresentação. Podes utilizar esse pedaço de código para economizar tempo:

```
mutate(ADM2_PCODE=
  case_when(distrito == "R.A.Príncipe" ~ "ST0101" ,
            distrito == "Água Grande" ~ "ST0201" ,
            distrito == "Cantagalo" ~ "ST0202" ,
            distrito == "Caué" ~ "ST0203" ,
            distrito == "Lembá" ~ "ST0204" ,
            distrito == "Lobata" ~ "ST0205" ,
            distrito == "Mé-Zochi" ~ "ST0206"))
```

Questão 8- Mapa

- Faça a união entre o shapefile (maps_STP) e a base_pre10
- Preste atenção. Qual é a chave de ligação?
- Dica: A base_pre10 pode vir pela esquerda. Olhe a apresentação.
- Receba essa base unida com o nome de base_pre11
- Aplique as definições de mapa a base_pre11, recebendo com o nome de base_pre12

Dica:

```
base_pre12 <-st_as_sf(base_pre11)
```

Questão 9- Mapa

- Faça um mapa da média de idade das meninas por distrito/região em STP utilizando o pacote ggplot2

Questão 10- Escreva o que quiser sobre o curso