Desafio 2: Resumindo um Banco de Dados

Entregar até 14h00 de 17/04/2020

Beatriz Milz

HTML compilado em: 10 de abril de 2020

## Carregar pacotes  
  
library(tidyverse) # pacote que inclui diversos pacotes para ciência de dados  
  
library(readr) # pacote usado para abrir a base  
  
library(janitor) # pacote usado para limpar a base  
  
library(gt)  
  
library(flextable)  
library(officer)

## Roteiro

1. Baixe e abra o banco de dados “CENSOESC\_2004\_RO.CSV”. Incomumente este banco usa o separador “|”. O banco nacional é 1GB então já filtramos o banco para o estado de Roraima para deixar ele mais leve e accessível. Os nomes das variáveis são mais ou menos auto-explicativas, e descrevemos alguns em mais detalhe ao longo do desafio.

censoesc\_2004\_RO <- read\_delim("https://raw.githubusercontent.com/JonnyPhillips/Ciencia\_de\_Dados/master/Desafios/CENSOESC\_2004\_RO.csv",   
 "|", escape\_double = FALSE, trim\_ws = TRUE) %>% janitor::clean\_names()  
  
  
# glimpse(censoesc\_2004\_RO)

1. Os dados contém escolas não-ativas (veja a variável CODFUNC); tire elas do banco de dados. O que representa uma observação no banco de dados agora?

escolas\_ativas <- censoesc\_2004\_RO %>% filter(codfunc == "Ativo")  
  
# glimpse(escolas\_ativas)

**R:**  Cada observação no novo banco de dados representa uma escola ativa, no ano de 2004, no estado de Roraima

* 1. Quantas escolas há no estado, quantas por municípo (inclua uma tabela), e qual é o município com o maior número de escolas?

# Quantas escolas ATIVAS há no estado?  
escolas\_no\_estado <- escolas\_ativas %>% distinct(codigo\_escola) %>% nrow()  
  
  
# Quantas escolas por município?  
  
escolas\_por\_munic <- escolas\_ativas %>% group\_by(munic) %>% tally() %>% arrange(-n)  
  
# Qual é o município com o maior número de escolas?  
  
munic\_com\_mais\_escolas <- escolas\_por\_munic %>% slice(1) %>% .[[1]]  
  
munic\_com\_mais\_escolas\_n <- escolas\_por\_munic %>% slice(1) %>% .[[2]]

**R:**  No estado de Roraima, em 2004, existiam 789 escolas ativas, sendo que o município BOA VISTA era o município com o maior número de escolas ativas (149). A seguir, está representada a tabela @ref(tab:escolasmunic), com as relação do número de escolas por município.

escolas\_por\_munic %>% flextable() %>%   
 set\_header\_labels(munic = "Município", n = "Número de escolas ativas") %>%   
 autofit() %>% theme\_booktabs() %>% flextable::set\_caption("Tabela de número de escolas ativas por município, em RR")

Tabela de número de escolas ativas por município, em RR

| Município | Número de escolas ativas |
| --- | --- |
| BOA VISTA | 149 |
| RORAINOPOLIS | 62 |
| NORMANDIA | 61 |
| CARACARAI | 59 |
| ALTO ALEGRE | 55 |
| UIRAMUTA | 55 |
| MUCAJAI | 51 |
| CANTA | 50 |
| IRACEMA | 45 |
| CAROEBE | 44 |
| PACARAIMA | 43 |
| AMAJARI | 42 |
| BONFIM | 36 |
| SAO LUIZ | 20 |
| SAO JOAO DA BALIZA | 17 |

* 1. Adicione uma última linha na tabela que tenha o total de escolas no estado.

escolas\_por\_munic %>%  
 add\_row(munic = "Total", n = sum(.$n))

## # A tibble: 16 x 2  
## munic n  
## <chr> <int>  
## 1 BOA VISTA 149  
## 2 RORAINOPOLIS 62  
## 3 NORMANDIA 61  
## 4 CARACARAI 59  
## 5 ALTO ALEGRE 55  
## 6 UIRAMUTA 55  
## 7 MUCAJAI 51  
## 8 CANTA 50  
## 9 IRACEMA 45  
## 10 CAROEBE 44  
## 11 PACARAIMA 43  
## 12 AMAJARI 42  
## 13 BONFIM 36  
## 14 SAO LUIZ 20  
## 15 SAO JOAO DA BALIZA 17  
## 16 Total 789

* 1. Quantos professores há no estado de Roraima, e quantos por município (inclua uma tabela)?
  2. Qual é o porcentagem de professores entre todos os empregados (professores e funcionários) em cada município (inclua uma tabela)?

1. Calcule a média e o desvio padrão do número de professores por escola em cada município (inclua uma tabela). Qual município há a maior variabilidade no número de professores por escola?
2. Quantos professores trabalham em cada uma das redes federais, estaduais, municipais e particulares (a variável DEP)? Calcule o porcentagem de professores em cada rede (inclua uma tabela).
3. Qual é o porcentagem de escolas com laboratório de informático, por município (inclua uma tabela)? Qual é o nome do município com o maior porcentagem de escolas com laboratório de informático?
4. O banco de dados do INEP contém quatro variáveis (VDG1CA,VDG1C1,VDG1C3,VDG1C4) com o número de professores por nível de ensinamento.
   1. Vire o banco de dados do formato largo para o formato longo, para que as quatro variáveis VDG1CA, VDG1C1, VDG1C3, VDG1C4 se tornam uma variável que descreve o nível de ensinamento do professor, e uma outra variável que descreve o número de professores.
   2. No novo banco de dados longo, recodifique as variáveis: VDG1CA para ‘Creche’, VDG1C1 para ‘Pré-escola’, VDG1C3 para ‘Fundamental’, VDG1C4 para ‘Médio’.
   3. Resuma o número de professores por nível de ensinamento no estado inteiro.
5. Usando o banco de dados ‘longo’ criado na questão 8:
   1. Calcule a média e o total do número de professores em cada nível de ensinamento em cada município (inclua uma tabela)
   2. Calcule o porcentagem dos professores em cada nível de ensinamento, por município, em uma tabela, e o município com o maior porcentagem dos seus Professores alocado à nível Fundamental
   3. Calcule o porcentagem dos professores em cada município, por nível de ensinamento, em uma tabela, e o município com o maior porcentagem de Professores de Ensino Fundamental.
6. Voltando para o seu banco de dados ‘largo’, gere uma amostra aleatória de duas escolas estaduais em cada município. Mostre os resultados numa tabela.
7. Agora, queremos gerar uma tabela com nomes dos municípios nas linhas e rede (Estadual, Municipal, Federal, Particular; variável DEP) nas colunas. Nas células, colocamos o número de escolas de cada rede em cada município. Há vários jeitos de realizar isso, mas vamos seguir os passos abaixo:
   1. Calcule o número de escolas por município e rede. Todas as redes existem em todos os município?
   2. Transforme as variáveis do município e rede em fatores.
   3. Use complete (de Tutorial 3) para criar uma tabela ‘completa’ com todas as combinações de município e rede possível, mesmo que não existe uma rede específica num município específico. Preenche os valores NA com ‘0’, porque já sabemos que não temos escolas nestes casos.
   4. Use uma função de pivot para virar o tibble e gerar o tamanho de tabela desejada (municípios nas linhas, redes nas colunas).
8. Para as tabelas que você produziu acima, volte para gerar tabelas estáticos bem formatados com kable(). Verifique que todas as tabelas têm um título, as colunas apropriadas, formatação de números, etc.