Aula 2 - Introdução ao tidyverse

Table of contents

1 Aula 2

Dr. Gabriel Bertolini

2º semestre 2023 - PPGGEO UFRGS

Oficina de programação

Aula introdutória sobre tidyverse.

2 Bibliotecas

Função install.packages pode ser usada para baixar arquivos do CRAN - pacotes da comunidade "oficiais" que estão alinhados as praticas e estilo de programação padronizadas do R. Pacotes do github pode ser instalados, utilizando uma função do pacote devtools: devtools::install.github("link_do_github_exemplo:gabertol/ztR")

os pacotes só precisam ser instalados uma vez, e carregados sempre que necessário usando a função library().

```
#install.packages("tidyverse") # Instalar se necessário
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.1     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

3 Funções de R base úteis

```
vetor<-c(1.1,4.2,2.9,9,1,4,5,0.0001,NA)

# média
mean(vetor,na.rm=TRUE) #média
```

```
[1] 3.400012
```

```
# desvio padrão
sd(vetor)
```

```
[1] NA
```

```r
# minimo
min(vetor)
```

```
[1] NA
```

```r
# maximo
max(vetor)
```

```
[1] NA
```

```r
# extensão
length(vetor)
```

```
[1] 9
```

```r
# cria sequencia de valores
A<-seq(from=pi,
    to=pi^2,
    by=pi/4)

length(A)
```

```
[1] 9
```

```r
# gerador de números com distribuição normal
PER<-rnorm(n=1000000,mean=350,sd=5)

mean(PER)
```

```
[1] 350.0049
```

```r
sd(PER)
```

```
[1] 4.999833
```

```
# gerador de numeros randomicos
rep(c("A","B","C"), 2)
```

```
[1] "A" "B" "C" "A" "B" "C"
```

4 Introdução ao tidyverse

A biblioteca guarda-chuva tidyverse inclui uma serie de pacotes, cada uma para seu fim. Por hora, vamos explorar o pacote dplyr que trata de manipulação básica de bancos de dados. Para isso, vamos primeiramente importar um banco de dados. Para importar, vamos usar a função read.csv

```
BD<-read.csv("./data/geoquimica_granito.csv")

# Roda o objeto no console
#BD


#Abre tabela estilo excel/ selecionar o objeto com mouse e apertar F2 é a mesma coisa
#view(BD)

BD<-BD %>% as_tibble()  # tibble é um tipo particular de dataframe do tidyverse, que facil
```

Trata-se de um banco de dados de XXXX et al. Aqui vamos usar uma série de funções do dplyr e tidyr para operações nesse banco de dados.

4.0.1 Selecionar colunas com select()

```
# Selecionar algumas colunas
# Usar função names() para examinar quais colunas o banco tem
names(BD) # o banco tem 117 colunas, vamos selecionar algumas colunas
```

```
  [1] "X.1"               "X"                 "Tectonic_setting"
  [4] "Location_notes"    "Pluton"            "Colour"
  [7] "Symbol"            "Size_rel"          "Size"
 [10] "SubGroup"          "Group"             "DebonPQ"
 [13] "TASMiddlemostPlut" "Villaseca"         "QANOR"
 [16] "LaRoche"           "Geol_unit_notes"   "Rock_type_notes"
```

```
 [19] "Ref"               "Age"               "Latitude"
 [22] "Longitude"         "SiO2"              "TiO2"
 [25] "Al2O3"             "FeOt"              "MnO"
 [28] "MgO"               "CaO"               "K2O"
 [31] "Na2O"              "P2O5"              "H2O"
 [34] "H2Ot"              "H2O.MINUS"         "LOI"
 [37] "Li"                "Be"                "B"
 [40] "Sc"                "V"                 "Cr"
 [43] "Ni"                "Cu"                "Zn"
 [46] "Rb"                "Sr"                "Y"
 [49] "Zr"                "Nb"                "Cs"
 [52] "Ba"                "La"                "Ce"
 [55] "Pr"                "Nd"                "Sm"
 [58] "Eu"                "Gd"                "Tb"
 [61] "Dy"                "Ho"                "Er"
 [64] "Tm"                "Yb"                "Lu"
 [67] "Hf"                "Ta"                "Pb"
 [70] "Th"                "U"                 "Co"
 [73] "Mo"                "W"                 "Ga"
 [76] "Ge"                "As"                "In"
 [79] "Sn"                "Sb"                "Cd"
 [82] "Q"                 "C"                 "Or"
 [85] "Ab"                "An"                "Ne"
 [88] "Ns"                "Ks"                "Di"
 [91] "Wo"                "Hy"                "Ol"
 [94] "Il"                "Tn"                "Pf"
 [97] "Ru"                "Ap"                "Sum_CIPW"
[100] "Orthoclase"        "Albite"            "Anorthite"
[103] "Quartz"            "Apatite"           "Ilmenite"
[106] "Biotite"           "Amphibole"         "Corundum"
[109] "Rest"              "sum_meso"          "P"
[112] "K"                 "Ti"                "Cr2O3"
[115] "A.NK"              "A.CNK"             "K2O.Na2O"
```

```
BD %>%  # reparar o uso do comando pype esse codigo significa select(BD,SiO2:Cd)- as funçõ
  select(SiO2:Cd) # selecionar elementos quimicos - SiO2 até Cd
```

```
# A tibble: 3,050 x 59
    SiO2  TiO2 Al2O3  FeOt    MnO   MgO   CaO   K2O  Na2O P2O5        H2O  H2Ot
   <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <chr>     <dbl> <dbl>
 1  65.4 0.827  17.1  4.91 0.0771 1.80   1.81  5.13  2.66 0.2880228    NA    NA
```

```
 2  69.6 0.709  14.9  4.19 0.0540 1.63   1.44  5.02  2.22 0.22098      NA    NA
 3  70.6 0.571  15.3  2.82 0.0532 1.33   1.20  5.60  2.33 0.197212     NA    NA
 4  70.6 0.478  15.7  2.51 0.0407 0.978  1.31  5.28  2.88 0.2001758    NA    NA
 5  70.7 0.532  15.3  2.54 0.0609 1.21   1.24  5.57  2.64 0.2268257    NA    NA
 6  71.0 0.484  15.5  2.47 0.0396 1.01   1.48  5.33  2.56 0.1863152    NA    NA
 7  72.4 0.470  14.3  2.68 0.0422 1.03   1.09  5.05  2.75 0.2026448    NA    NA
 8  73.2 0.428  14.4  2.40 0.0605 0.769  1.06  5.04  2.46 0.2184707    NA    NA
 9  73.2 0.363  14.4  2.25 0.0211 0.678  1.10  5.05  2.86 0.1591542    NA    NA
10  74.8 0.530  13.5  2.73 0.0472 1.18   1.43  3.28  2.30 0.1812674    NA    NA
# i 3,040 more rows
# i 47 more variables: H2O.MINUS <dbl>, LOI <chr>, Li <dbl>, Be <dbl>, B <dbl>,
#   Sc <dbl>, V <chr>, Cr <dbl>, Ni <dbl>, Cu <dbl>, Zn <dbl>, Rb <dbl>,
#   Sr <dbl>, Y <dbl>, Zr <dbl>, Nb <dbl>, Cs <dbl>, Ba <dbl>, La <chr>,
#   Ce <dbl>, Pr <chr>, Nd <chr>, Sm <chr>, Eu <chr>, Gd <chr>, Tb <dbl>,
#   Dy <dbl>, Ho <dbl>, Er <dbl>, Tm <dbl>, Yb <dbl>, Lu <dbl>, Hf <dbl>,
#   Ta <dbl>, Pb <dbl>, Th <chr>, U <dbl>, Co <dbl>, Mo <chr>, W <dbl>, ...
```

```r
BD %>%
  select(X,Tectonic_setting,Latitude,Longitude,Albite:Corundum) %>% # selecionar colunas X
  select(-X) # usar - pra retirar a variavel
```

```
# A tibble: 3,050 x 11
   Tectonic_setting Latitude Longitude Albite Anorthite Quartz Apatite Ilmenite
   <chr>               <dbl>     <dbl>  <dbl>     <dbl>  <dbl>   <dbl>    <dbl>
 1 OROGENIC BELT          NA        NA   22.5      7.11   27.2   0.680    0.786
 2 OROGENIC BELT          NA        NA   18.8      5.68   35.1   0.522    0.673
 3 OROGENIC BELT          NA        NA   19.7      4.68   33.6   0.465    0.542
 4 OROGENIC BELT          NA        NA   24.4      5.19   31.4   0.472    0.454
 5 OROGENIC BELT          NA        NA   22.4      4.65   32.0   0.535    0.506
 6 OROGENIC BELT          NA        NA   21.7      6.11   33.0   0.440    0.460
 7 OROGENIC BELT          NA        NA   23.3      4.07   35.3   0.478    0.446
 8 OROGENIC BELT          NA        NA   20.8      3.84   37.9   0.516    0.407
 9 OROGENIC BELT          NA        NA   24.2      4.39   35.3   0.376    0.344
10 OROGENIC BELT          NA        NA   19.5      5.91   46.2   0.428    0.504
# i 3,040 more rows
# i 3 more variables: Biotite <dbl>, Amphibole <dbl>, Corundum <dbl>
```

4.0.2 Filtrar colunas com filter()

```
# Filtrar tectonic_setting
unique(BD$Tectonic_setting) # ver valores unicos de Tectonic_setting para saber qual varia
```

```
[1] "OROGENIC BELT"           "INTRAPLATE VOLCANICS"
[3] "ARCHAEAN CRATONS"        "CONVERGENT MARGIN"
[5] "OCEANIC PLATEAU"         "OPHIOLITE"
[7] "RIFT VOLCANICS"          "CONTINENTAL FLOOD BASALT"
[9] "OCEAN ISLAND"
```

```
BD %>%
  select(X,Tectonic_setting,Latitude,Longitude,Albite:Corundum) %>% # reparar funções enca
  filter(Tectonic_setting=="OCEANIC PLATEAU") # filtrando somente o Oceanic Plateau
```

```
# A tibble: 13 x 12
   X         Tectonic_setting Latitude Longitude Albite Anorthite Quartz Apatite
   <chr>     <chr>               <dbl>     <dbl>  <dbl>     <dbl>  <dbl>   <dbl>
 1 46143     OCEANIC PLATEAU      12.5       -70   37.3      8.51  33.3    0.135
 2 46135     OCEANIC PLATEAU      12.5       -70   32.7      9.14   6.31   0.439
 3 46142     OCEANIC PLATEAU      12.5       -70   45.7     15.0   22.7    0.304
 4 46139     OCEANIC PLATEAU      12.5       -70   31.2     24.6   19.9    0.208
 5 46140     OCEANIC PLATEAU      12.5       -70   37.4     17.8   22.7    0.283
 6 46134     OCEANIC PLATEAU      12.5       -70   36.3     23.2    8.53   0.408
 7 9157-ARU~ OCEANIC PLATEAU      12.5       -70   39.1     16.1   21.2    0.337
 8 10446-A10 OCEANIC PLATEAU      -7         156   31.9     13.7   14.9    0.378
 9 10446-A11 OCEANIC PLATEAU      -7         156   50.6      2.22  37.5    0.260
10 10446-A6  OCEANIC PLATEAU      -7         156   32.5     17.1   22.6    0.378
11 10446-A7  OCEANIC PLATEAU      -7         156   31.6     17.7   15.9    0.378
12 10446-A8  OCEANIC PLATEAU      -7         156   34.6     20.2   18.0    0.378
13 10446-A9  OCEANIC PLATEAU      -7         156   29.1     21.1   20.0    0.401
# i 4 more variables: Ilmenite <dbl>, Biotite <dbl>, Amphibole <dbl>,
#   Corundum <dbl>
```

```
BD %>%
  filter(Tectonic_setting=="CONTINENTAL FLOOD BASALT",
         SiO2<60,
         SiO2>50,
         TiO2>1) # filtrar varias colunas - separadas por ,
```

```
# A tibble: 13 x 117
     X.1 X         Tectonic_setting Location_notes Pluton Colour Symbol Size_rel
   <int> <chr>     <chr>            <chr>          <chr>  <chr>   <int>    <dbl>
 1  1952 201163    CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 2  1953 201168    CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 3  1955 85457     CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 4  1957 85452     CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 5  1958 85453     CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 6  1959 85455     CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 7  1960 85456     CONTINENTAL FLO~ NORTH ATLANTI~ Skye   honey~      2      1.6
 8  2055 119793    CONTINENTAL FLO~ ETENDEKA PROV~ Etane~ honey~      2      1.6
 9  2056 119795    CONTINENTAL FLO~ ETENDEKA PROV~ Etane~ honey~      2      1.6
10  2057 99730     CONTINENTAL FLO~ ETENDEKA PROV~ Etane~ honey~      2      1.6
11  2111 8564-MC4  CONTINENTAL FLO~ ETENDEKA PROV~ Messum honey~      2      1.6
12  2113 43064     CONTINENTAL FLO~ ETENDEKA PROV~ Messum honey~      2      1.6
13  2124 8572-MC44 CONTINENTAL FLO~ ETENDEKA PROV~ Messum honey~      2      1.6
# i 109 more variables: Size <dbl>, SubGroup <chr>, Group <chr>, DebonPQ <chr>,
#   TASMiddlemostPlut <chr>, Villaseca <chr>, QANOR <chr>, LaRoche <chr>,
#   Geol_unit_notes <chr>, Rock_type_notes <chr>, Ref <chr>, Age <chr>,
#   Latitude <dbl>, Longitude <dbl>, SiO2 <dbl>, TiO2 <dbl>, Al2O3 <dbl>,
#   FeOt <dbl>, MnO <dbl>, MgO <dbl>, CaO <dbl>, K2O <dbl>, Na2O <dbl>,
#   P2O5 <chr>, H2O <dbl>, H2Ot <dbl>, H2O.MINUS <dbl>, LOI <chr>, Li <dbl>,
#   Be <dbl>, B <dbl>, Sc <dbl>, V <chr>, Cr <dbl>, Ni <dbl>, Cu <dbl>, ...
```

```r
BD %>%
  filter(Tectonic_setting %in% c("OROGENIC BELT",
                               "INTRAPLATE VOLCANICS",
                               "ARCHAEAN CRATONS",
                               "CONVERGENT MARGIN")) # filtrar pelos strings usando %in%
```

```
# A tibble: 2,642 x 117
     X.1 X     Tectonic_setting Location_notes     Pluton Colour Symbol Size_rel
   <int> <chr> <chr>            <chr>              <chr>  <chr>   <int>    <dbl>
 1     1 CR03  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 2     2 BB202 OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 3     3 CB02  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 4     4 HB02  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 5     5 CR01  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 6     6 LL01  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 7     7 MP21  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
 8     8 OK07  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16      1.6
```

```
 9      9 LG05  OROGENIC BELT    ////////Langebaan~ Cape ~ brown4     16       1.6
10     10 BB201 OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
# i 2,632 more rows
# i 109 more variables: Size <dbl>, SubGroup <chr>, Group <chr>, DebonPQ <chr>,
#   TASMiddlemostPlut <chr>, Villaseca <chr>, QANOR <chr>, LaRoche <chr>,
#   Geol_unit_notes <chr>, Rock_type_notes <chr>, Ref <chr>, Age <chr>,
#   Latitude <dbl>, Longitude <dbl>, SiO2 <dbl>, TiO2 <dbl>, Al2O3 <dbl>,
#   FeOt <dbl>, MnO <dbl>, MgO <dbl>, CaO <dbl>, K2O <dbl>, Na2O <dbl>,
#   P2O5 <chr>, H2O <dbl>, H2Ot <dbl>, H2O.MINUS <dbl>, LOI <chr>, ...
```

```r
BD %>%
  filter(Tectonic_setting== "OROGENIC BELT" & SiO2>60) # filtrar condicionalmente
```

```
# A tibble: 1,370 x 117
     X.1 X     Tectonic_setting Location_notes     Pluton Colour Symbol Size_rel
   <int> <chr> <chr>            <chr>              <chr>  <chr>  <int>    <dbl>
 1     1 CR03  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 2     2 BB202 OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 3     3 CB02  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 4     4 HB02  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 5     5 CR01  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 6     6 LL01  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 7     7 MP21  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 8     8 OK07  OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
 9     9 LG05  OROGENIC BELT    ////////Langebaan~ Cape ~ brown4     16       1.6
10    10 BB201 OROGENIC BELT    ////////Peninsula~ Cape ~ brown4     16       1.6
# i 1,360 more rows
# i 109 more variables: Size <dbl>, SubGroup <chr>, Group <chr>, DebonPQ <chr>,
#   TASMiddlemostPlut <chr>, Villaseca <chr>, QANOR <chr>, LaRoche <chr>,
#   Geol_unit_notes <chr>, Rock_type_notes <chr>, Ref <chr>, Age <chr>,
#   Latitude <dbl>, Longitude <dbl>, SiO2 <dbl>, TiO2 <dbl>, Al2O3 <dbl>,
#   FeOt <dbl>, MnO <dbl>, MgO <dbl>, CaO <dbl>, K2O <dbl>, Na2O <dbl>,
#   P2O5 <chr>, H2O <dbl>, H2Ot <dbl>, H2O.MINUS <dbl>, LOI <chr>, ...
```

4.0.3 Criar colunas com mutate(), agrupamento com group(), arredondamento com round()

```
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  mutate(l_Si=log(SiO2)) #cria nova coluna com o log da coluna SiO2
```

```
# A tibble: 3,050 x 9
   X     Tectonic_setting  SiO2  TiO2 Al2O3  FeOt    MnO   MgO  l_Si
   <chr> <chr>            <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl>
 1 CR03  OROGENIC BELT     65.4 0.827  17.1  4.91 0.0771 1.80   4.18
 2 BB202 OROGENIC BELT     69.6 0.709  14.9  4.19 0.0540 1.63   4.24
 3 CB02  OROGENIC BELT     70.6 0.571  15.3  2.82 0.0532 1.33   4.26
 4 HB02  OROGENIC BELT     70.6 0.478  15.7  2.51 0.0407 0.978  4.26
 5 CR01  OROGENIC BELT     70.7 0.532  15.3  2.54 0.0609 1.21   4.26
 6 LL01  OROGENIC BELT     71.0 0.484  15.5  2.47 0.0396 1.01   4.26
 7 MP21  OROGENIC BELT     72.4 0.470  14.3  2.68 0.0422 1.03   4.28
 8 OK07  OROGENIC BELT     73.2 0.428  14.4  2.40 0.0605 0.769  4.29
 9 LG05  OROGENIC BELT     73.2 0.363  14.4  2.25 0.0211 0.678  4.29
10 BB201 OROGENIC BELT     74.8 0.530  13.5  2.73 0.0472 1.18   4.31
# i 3,040 more rows
```

```
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  mutate(across(.cols=SiO2:MgO,~log(.x))) # Operações em massa
```

```
# A tibble: 3,050 x 8
   X     Tectonic_setting  SiO2   TiO2 Al2O3  FeOt   MnO      MgO
   <chr> <chr>            <dbl>  <dbl> <dbl> <dbl> <dbl>    <dbl>
 1 CR03  OROGENIC BELT     4.18 -0.190  2.84 1.59  -2.56  0.586
 2 BB202 OROGENIC BELT     4.24 -0.344  2.70 1.43  -2.92  0.488
 3 CB02  OROGENIC BELT     4.26 -0.561  2.73 1.04  -2.93  0.286
 4 HB02  OROGENIC BELT     4.26 -0.739  2.75 0.920 -3.20 -0.0223
 5 CR01  OROGENIC BELT     4.26 -0.631  2.73 0.930 -2.80  0.188
 6 LL01  OROGENIC BELT     4.26 -0.726  2.74 0.904 -3.23  0.00650
 7 MP21  OROGENIC BELT     4.28 -0.756  2.66 0.986 -3.16  0.0276
 8 OK07  OROGENIC BELT     4.29 -0.848  2.67 0.877 -2.81 -0.263
 9 LG05  OROGENIC BELT     4.29 -1.01   2.66 0.809 -3.86 -0.389
10 BB201 OROGENIC BELT     4.31 -0.635  2.61 1.00  -3.05  0.165
# i 3,040 more rows
```

```r
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%
  mutate(contagem=n())
```

```
# A tibble: 3,050 x 9
# Groups:   Tectonic_setting [9]
   X     Tectonic_setting  SiO2  TiO2 Al2O3  FeOt    MnO   MgO contagem
   <chr> <chr>            <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>    <int>
 1 CR03  OROGENIC BELT     65.4 0.827  17.1  4.91 0.0771 1.80      1471
 2 BB202 OROGENIC BELT     69.6 0.709  14.9  4.19 0.0540 1.63      1471
 3 CB02  OROGENIC BELT     70.6 0.571  15.3  2.82 0.0532 1.33      1471
 4 HB02  OROGENIC BELT     70.6 0.478  15.7  2.51 0.0407 0.978     1471
 5 CR01  OROGENIC BELT     70.7 0.532  15.3  2.54 0.0609 1.21      1471
 6 LL01  OROGENIC BELT     71.0 0.484  15.5  2.47 0.0396 1.01      1471
 7 MP21  OROGENIC BELT     72.4 0.470  14.3  2.68 0.0422 1.03      1471
 8 OK07  OROGENIC BELT     73.2 0.428  14.4  2.40 0.0605 0.769     1471
 9 LG05  OROGENIC BELT     73.2 0.363  14.4  2.25 0.0211 0.678     1471
10 BB201 OROGENIC BELT     74.8 0.530  13.5  2.73 0.0472 1.18      1471
# i 3,040 more rows
```

```r
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%
  count()
```

```
# A tibble: 9 x 2
# Groups:   Tectonic_setting [9]
  Tectonic_setting             n
  <chr>                    <int>
1 ARCHAEAN CRATONS           310
2 CONTINENTAL FLOOD BASALT   191
3 CONVERGENT MARGIN          763
4 INTRAPLATE VOLCANICS        98
5 OCEAN ISLAND                56
6 OCEANIC PLATEAU             13
7 OPHIOLITE                   19
8 OROGENIC BELT             1471
9 RIFT VOLCANICS             129
```

```
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%  # Agrupa os valores por grupos baseados nos tipos tecton
  mutate(across(SiO2:MgO,~ifelse(.x==0,mean(.x),.x)),  # substituir 0 por média por grupo
         SiO2=round(SiO2,digits=2)) #arredondar para dois digitos
```

```
# A tibble: 3,050 x 8
# Groups:   Tectonic_setting [9]
   X     Tectonic_setting  SiO2  TiO2 Al2O3  FeOt    MnO   MgO
   <chr> <chr>            <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>
 1 CR03  OROGENIC BELT     65.4 0.827  17.1  4.91 0.0771 1.80
 2 BB202 OROGENIC BELT     69.6 0.709  14.9  4.19 0.0540 1.63
 3 CB02  OROGENIC BELT     70.6 0.571  15.3  2.82 0.0532 1.33
 4 HB02  OROGENIC BELT     70.6 0.478  15.7  2.51 0.0407 0.978
 5 CR01  OROGENIC BELT     70.7 0.532  15.3  2.54 0.0609 1.21
 6 LL01  OROGENIC BELT     71.0 0.484  15.5  2.47 0.0396 1.01
 7 MP21  OROGENIC BELT     72.4 0.470  14.3  2.68 0.0422 1.03
 8 OK07  OROGENIC BELT     73.2 0.428  14.4  2.40 0.0605 0.769
 9 LG05  OROGENIC BELT     73.2 0.363  14.4  2.25 0.0211 0.678
10 BB201 OROGENIC BELT     74.8 0.530  13.5  2.73 0.0472 1.18
# i 3,040 more rows
```

4.0.4

4.0.5 Contar com count() ou remoldar dataframe com reframe()

```
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%  # Agrupa os valores por grupos baseados nos tipos tecton
  mutate(across(SiO2:MgO,~ifelse(.x==0, mean(.x),.x)),  # substituir 0 por média por grupo
         across(SiO2:MgO,~round(.x,digits=2))) %>%
  reframe(n=n(),
          across(.cols=SiO2:MgO,~round(mean(.x,na.rm = TRUE),2)))
```

```
# A tibble: 9 x 8
  Tectonic_setting              n  SiO2  TiO2 Al2O3  FeOt   MnO   MgO
  <chr>                     <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARCHAEAN CRATONS            310  69.6  0.31  15.4  2.43  0.04  1.33
2 CONTINENTAL FLOOD BASALT    191  67.5  0.53  14.6  4.21  0.13  0.5
```

```
3 CONVERGENT MARGIN          763  66.4  0.52  15.7  3.92  0.08  1.7
4 INTRAPLATE VOLCANICS        98  70.3  0.32  14.1  3.35  0.08  0.49
5 OCEAN ISLAND                56  65.2  0.53  16.2  4.02  0.16  0.44
6 OCEANIC PLATEAU             13  63.6  0.52  16.1  5.05  0.1   2.59
7 OPHIOLITE                   19  61.6  1.07  15.0  6.03  0.17  2.41
8 OROGENIC BELT             1471  69.8  0.41  14.4  2.85  0.07  1.31
9 RIFT VOLCANICS             129  69.3  0.43  13.8  3.66  0.1   0.49
```

4.1 Alterando a forma do dataframe com pivot_longer e _wider()

```
# Dado longo
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%  # Agrupa os valores por grupos baseados nos tipos tecton
  mutate(across(SiO2:MgO,~ifelse(.x==0, mean(.x),.x)),  # substituir 0 por média por grupo
         across(SiO2:MgO,~round(.x,digits=2))) %>%
  reframe(n=n(),
          across(.cols=SiO2:MgO,~round(mean(.x,na.rm = TRUE),2))) %>%
  pivot_longer(cols=SiO2:MgO)
```

```
# A tibble: 54 x 4
   Tectonic_setting             n name  value
   <chr>                    <int> <chr> <dbl>
 1 ARCHAEAN CRATONS           310 SiO2  69.6
 2 ARCHAEAN CRATONS           310 TiO2   0.31
 3 ARCHAEAN CRATONS           310 Al2O3 15.4
 4 ARCHAEAN CRATONS           310 FeOt   2.43
 5 ARCHAEAN CRATONS           310 MnO    0.04
 6 ARCHAEAN CRATONS           310 MgO    1.33
 7 CONTINENTAL FLOOD BASALT   191 SiO2  67.5
 8 CONTINENTAL FLOOD BASALT   191 TiO2   0.53
 9 CONTINENTAL FLOOD BASALT   191 Al2O3 14.6
10 CONTINENTAL FLOOD BASALT   191 FeOt   4.21
# i 44 more rows
```

```
#dado largo
BD %>%
  select(X,Tectonic_setting,SiO2:MgO) %>%
  group_by(Tectonic_setting) %>%  # Agrupa os valores por grupos baseados nos tipos tecton
  mutate(across(SiO2:MgO,~ifelse(.x==0, mean(.x),.x)),  # substituir 0 por média por grupo
```

```
         across(SiO2:MgO,~round(.x,digits=2))) %>%
  reframe(n=n(),
          across(.cols=SiO2:MgO,~round(mean(.x,na.rm = TRUE),2))) %>%
  pivot_longer(cols=SiO2:MgO) %>% # Alongar dataframe para forma tidy
  ungroup() %>%
  group_by(name) %>%
  pivot_wider(id_cols=Tectonic_setting,values_from=value) # Expandir lateralmente o datafr
```

```
# A tibble: 9 x 7
  Tectonic_setting          SiO2  TiO2 Al2O3  FeOt   MnO   MgO
  <chr>                    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARCHAEAN CRATONS          69.6  0.31  15.4  2.43  0.04  1.33
2 CONTINENTAL FLOOD BASALT  67.5  0.53  14.6  4.21  0.13  0.5
3 CONVERGENT MARGIN         66.4  0.52  15.7  3.92  0.08  1.7
4 INTRAPLATE VOLCANICS      70.3  0.32  14.1  3.35  0.08  0.49
5 OCEAN ISLAND              65.2  0.53  16.2  4.02  0.16  0.44
6 OCEANIC PLATEAU           63.6  0.52  16.1  5.05  0.1   2.59
7 OPHIOLITE                 61.6  1.07  15.0  6.03  0.17  2.41
8 OROGENIC BELT             69.8  0.41  14.4  2.85  0.07  1.31
9 RIFT VOLCANICS            69.3  0.43  13.8  3.66  0.1   0.49
```

5 Estudo de caso - Vulcões do Tidytuesday

Objetivo: Revisar tidyverse e fazer exercícios. Os dados são disponibilizados no link, e contem o video para analise desse banco no youtube (https://github.com/rfordatascience/tidytuesday/tree/master).

Esse link apresenta mais infos sobre o banco de dado e fontes:

https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-05-12/readme.md

Importação de dados de um banco em excel com readxl:

```
# Ilustrativo
#volc<-read_excel("volcano.xlsx") %>%
#  as_tibble()
```

Opção B, baixar diretamente do github do tidytuesday:

```
volc <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mas
```

```
Rows: 958 Columns: 26
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (18): volcano_name, primary_volcano_type, last_eruption_year, country, r...
dbl  (8): volcano_number, latitude, longitude, elevation, population_within_...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

5.1 select()

```r
volc %>% select(volcano_name,-country,latitude:elevation)
```

```
# A tibble: 958 x 4
   volcano_name    latitude longitude elevation
   <chr>              <dbl>     <dbl>     <dbl>
 1 Abu                 34.5     132.        641
 2 Acamarachi         -23.3     -67.6      6023
 3 Acatenango          14.5     -90.9      3976
 4 Acigol-Nevsehir     38.5      34.6      1683
 5 Adams               46.2    -121.       3742
 6 Adatarayama         37.6     140.       1728
 7 Adwa                10.1      40.8      1733
 8 Afdera              13.1      40.9      1250
 9 Agrigan             18.8     146.        965
10 Agua                14.5     -90.7      3760
# i 948 more rows
```

```r
volc %>% select(!latitude:elevation)
```

```
# A tibble: 958 x 23
   volcano_number volcano_name   primary_volcano_type last_eruption_year country
            <dbl> <chr>          <chr>                <chr>              <chr>
 1         283001 Abu            Shield(s)            -6850              Japan
 2         355096 Acamarachi     Stratovolcano        Unknown            Chile
 3         342080 Acatenango     Stratovolcano(es)    1972               Guatem~
 4         213004 Acigol-Nevseh~ Caldera              -2080              Turkey
 5         321040 Adams          Stratovolcano        950                United~
 6         283170 Adatarayama    Stratovolcano(es)    1996               Japan
```

```
 7          221170 Adwa           Stratovolcano        Unknown          Ethiop~
 8          221110 Afdera         Stratovolcano        Unknown          Ethiop~
 9          284160 Agrigan        Stratovolcano        1917             United~
10          342100 Agua           Stratovolcano        Unknown          Guatem~
# i 948 more rows
# i 18 more variables: region <chr>, subregion <chr>, tectonic_settings <chr>,
#   evidence_category <chr>, major_rock_1 <chr>, major_rock_2 <chr>,
#   major_rock_3 <chr>, major_rock_4 <chr>, major_rock_5 <chr>,
#   minor_rock_1 <chr>, minor_rock_2 <chr>, minor_rock_3 <chr>,
#   minor_rock_4 <chr>, minor_rock_5 <chr>, population_within_5_km <dbl>,
#   population_within_10_km <dbl>, population_within_30_km <dbl>, ...
```

```r
volc %>% select(starts_with("population"))
```

```
# A tibble: 958 x 4
   population_within_5_km population_within_10_km population_within_30_km
                    <dbl>                   <dbl>                   <dbl>
 1                   3597                    9594                  117805
 2                      0                       7                     294
 3                   4329                   60730                 1042836
 4                 127863                  127863                  218469
 5                      0                      70                    4019
 6                    428                    3936                  717078
 7                    101                     485                   18645
 8                     51                    6042                    8611
 9                      0                       0                       0
10                   9890                  114404                 2530449
# i 948 more rows
# i 1 more variable: population_within_100_km <dbl>
```

```r
volc %>% select(contains("within"))
```

```
# A tibble: 958 x 4
   population_within_5_km population_within_10_km population_within_30_km
                    <dbl>                   <dbl>                   <dbl>
 1                   3597                    9594                  117805
 2                      0                       7                     294
 3                   4329                   60730                 1042836
 4                 127863                  127863                  218469
 5                      0                      70                    4019
```

```
 6                    428                    3936                    717078
 7                    101                     485                     18645
 8                     51                    6042                      8611
 9                      0                       0                         0
10                   9890                  114404                   2530449
# i 948 more rows
# i 1 more variable: population_within_100_km <dbl>
```

```
#- **stars_with()**: colunas que começam com um prefixo

#- **ends_with()**: colunas que terminam com um sufixo

#- **contains()**: colunas que contêm uma string

#- **last_col()**: última coluna
```

5.2 filter()

```
volc %>% select(volcano_name,country,latitude:elevation) %>%
         filter(country== "Antarctica")
```

```
# A tibble: 17 x 5
   volcano_name        country    latitude longitude elevation
   <chr>               <chr>         <dbl>     <dbl>     <dbl>
 1 Andrus              Antarctica    -75.8    -132.       2978
 2 Berlin              Antarctica    -76.0    -136        3478
 3 Buckle Island       Antarctica    -66.8     163.       1239
 4 Deception Island    Antarctica    -63.0     -60.7       602
 5 Erebus              Antarctica    -77.5     167.       3794
 6 Hudson Mountains    Antarctica    -74.3     -99.4       749
 7 James Ross Island   Antarctica    -64.2     -57.8      1630
 8 Melbourne           Antarctica    -74.4     165.       2732
 9 Morning             Antarctica    -78.5     164.       2723
10 Penguin Island      Antarctica    -62.1     -57.9       180
11 Pleiades, The       Antarctica    -72.7     166.       3040
12 Royal Society Range Antarctica    -78.2     163.       3000
13 Seal Nunataks Group Antarctica    -65.0     -60.0       368
14 Siple               Antarctica    -73.4    -127.       3110
15 Takahe              Antarctica    -76.3    -112.       3460
```

```
16 Toney Mountain      Antarctica    -75.8    -116.        3595
17 Waesche             Antarctica    -77.2    -127.        3292
```

```
volc %>% select(volcano_name,country,latitude:elevation) %>%
         filter(!country== "Chile") %>%
         filter(elevation>1000,
                latitude>0,
                longitude>0)
```

```
# A tibble: 252 x 5
   volcano_name     country  latitude longitude elevation
   <chr>            <chr>       <dbl>     <dbl>     <dbl>
 1 Acigol-Nevsehir  Turkey       38.5      34.6      1683
 2 Adatarayama      Japan        37.6     140.       1728
 3 Adwa             Ethiopia     10.1      40.8      1733
 4 Afdera           Ethiopia     13.1      40.9      1250
 5 Aira             Japan        31.6     131.       1117
 6 Akademia Nauk    Russia       54.0     159.       1180
 7 Akagisan         Japan        36.6     139.       1828
 8 Akan             Japan        43.4     144.       1499
 9 Akita-Komagatake Japan        39.8     141.       1637
10 Akita-Yakeyama   Japan        40.0     141.       1366
# i 242 more rows
```

```
# %in% comando logico para declarar vetores para filter

volc %>% select(volcano_name,country,latitude:elevation) %>%
         filter(!country %in%  c("Chile","Argentina","Japan"))
```

```
# A tibble: 809 x 5
   volcano_name    country       latitude longitude elevation
   <chr>           <chr>            <dbl>     <dbl>     <dbl>
 1 Acatenango      Guatemala         14.5     -90.9      3976
 2 Acigol-Nevsehir Turkey            38.5      34.6      1683
 3 Adams           United States     46.2    -121.       3742
 4 Adwa            Ethiopia          10.1      40.8      1733
 5 Afdera          Ethiopia          13.1      40.9      1250
 6 Agrigan         United States     18.8     146.        965
 7 Agua            Guatemala         14.5     -90.7      3760
 8 Agua de Pau     Portugal          37.8     -25.5       947
```

```
 9 Agung            Indonesia        -8.34      116.        2997
10 Ahyi             United States    20.4       145.         -75
# i 799 more rows
```

```
# | ou

# & e

volc %>%
  select(volcano_name, country, latitude:elevation) %>%
  filter(country == "Antarctica" | latitude < 10)
```

```
# A tibble: 382 x 5
   volcano_name country          latitude longitude elevation
   <chr>        <chr>               <dbl>     <dbl>     <dbl>
 1 Acamarachi   Chile              -23.3      -67.6      6023
 2 Aguilera     Chile              -50.3      -73.8      2546
 3 Agung        Indonesia           -8.34     116.       2997
 4 Alcedo       Ecuador             -0.43     -91.1      1130
 5 Aliso        Ecuador             -0.53     -78        4267
 6 Alutu        Ethiopia             7.77      38.8      2335
 7 Amasing      Indonesia           -0.53     127.       1030
 8 Ambae        Vanuatu            -15.4      168.       1496
 9 Ambang       Indonesia            0.75     124.       1795
10 Ambitle      Papua New Guinea    -4.08     154.        450
# i 372 more rows
```

```
volc %>%
  select(volcano_name, country, latitude:elevation) %>%
  filter(country == "Antarctica" & elevation > 500)
```

```
# A tibble: 15 x 5
   volcano_name        country    latitude longitude elevation
   <chr>               <chr>         <dbl>     <dbl>     <dbl>
 1 Andrus              Antarctica    -75.8    -132.       2978
 2 Berlin              Antarctica    -76.0    -136        3478
 3 Buckle Island       Antarctica    -66.8     163.       1239
 4 Deception Island    Antarctica    -63.0     -60.7       602
 5 Erebus              Antarctica    -77.5     167.       3794
 6 Hudson Mountains    Antarctica    -74.3     -99.4       749
```

```
 7 James Ross Island   Antarctica    -64.2     -57.8      1630
 8 Melbourne           Antarctica    -74.4     165.       2732
 9 Morning             Antarctica    -78.5     164.       2723
10 Pleiades, The       Antarctica    -72.7     166.       3040
11 Royal Society Range Antarctica    -78.2     163.       3000
12 Siple               Antarctica    -73.4    -127.       3110
13 Takahe              Antarctica    -76.3    -112.       3460
14 Toney Mountain      Antarctica    -75.8    -116.       3595
15 Waesche             Antarctica    -77.2    -127.       3292
```

5.3 mutate()

```
volc %>%
  select(volcano_name, country,last_eruption_year) %>%
  filter(!last_eruption_year== "Unknown") %>%
  mutate(last_eruption_year=as.numeric(last_eruption_year),
         ultima_erupcao=(2022-last_eruption_year),
         ultima_erupcao_milhar=ultima_erupcao/1000,
         ultima_erupcao_milhao=ultima_erupcao/1000000)
```

```
# A tibble: 657 x 6
   volcano_name  country last_eruption_year ultima_erupcao ultima_erupcao_milhar
   <chr>         <chr>                <dbl>          <dbl>                 <dbl>
 1 Abu           Japan                -6850           8872                 8.87
 2 Acatenango    Guatem~               1972             50                 0.05
 3 Acigol-Nevse~ Turkey               -2080           4102                 4.10
 4 Adams         United~                950           1072                 1.07
 5 Adatarayama   Japan                 1996             26                 0.026
 6 Agrigan       United~               1917            105                 0.105
 7 Agua de Pau   Portug~               1564            458                 0.458
 8 Aguilera      Chile                -1250           3272                 3.27
 9 Agung         Indone~               2019              3                 0.003
10 Ahyi          United~               2014              8                 0.008
# i 647 more rows
# i 1 more variable: ultima_erupcao_milhao <dbl>
```

```
volc %>%
  mutate(rocha=case_when(
        major_rock_1 %in% c("Andesite / Basaltic Andesite","Basalt / Picro-Basalt" ,"Trach
    major_rock_1 %in% c("Phonolite","Foidite","Trachybasalt / Tephrite Basanite","Phono-te
```

```r
      TRUE~"volcanica acida")) %>%
    count(rocha)
```

```
# A tibble: 3 x 2
  rocha                  n
  <chr>              <int>
1 volcanica acida      108
2 vulcanica alcalina    75
3 vulcanica basica     775
```

```r
unique(volc$major_rock_1)
```

```
 [1] "Andesite / Basaltic Andesite"
 [2] "Dacite"
 [3] "Rhyolite"
 [4] "Basalt / Picro-Basalt"
 [5] "Trachyte / Trachydacite"
 [6] "Phono-tephrite /  Tephri-phonolite"
 [7] "Trachyandesite / Basaltic Trachyandesite"
 [8] "Trachybasalt / Tephrite Basanite"
 [9] "Foidite"
[10] "Phonolite"
```

5.4 Count e arrange()

```r
volc %>%
  count(country) %>%
  arrange(desc(n))
```

```
# A tibble: 89 x 2
   country             n
   <chr>           <int>
 1 United States      99
 2 Indonesia          95
 3 Japan              92
 4 Russia             79
 5 Chile              43
 6 Iceland            28
```

```
 7 Mexico              28
 8 Papua New Guinea    28
 9 Philippines         27
10 Ecuador             26
# i 79 more rows
```

```
volc %>%
  count(country) %>%
  arrange(n)
```

```
# A tibble: 89 x 2
   country                    n
   <chr>                  <int>
 1 Algeria                    1
 2 Armenia-Azerbaijan         1
 3 Burma (Myanmar)            1
 4 Cape Verde                 1
 5 Chile-Peru                 1
 6 China-North Korea          1
 7 Colombia-Ecuador           1
 8 Comoros                    1
 9 Djibouti                   1
10 El Salvador-Guatemala      1
# i 79 more rows
```

5.5 Group_by ()

```
volc %>%
  filter(!last_eruption_year=="Unknown") %>%
  group_by(last_eruption_year,country) %>%
  count(last_eruption_year,sort = TRUE) %>%
  ungroup()
```

```
# A tibble: 591 x 3
   last_eruption_year country           n
   <chr>              <chr>         <int>
 1 2020               Indonesia         8
 2 2020               Japan             5
 3 -550               Russia            4
```

```
 4 2019               Indonesia         4
 5 2020               Russia            4
 6 1992               United States     3
 7 1996               Japan             3
 8 2000               Indonesia         3
 9 2008               United States     3
10 2015               Indonesia         3
# i 581 more rows
```

Summarise ou reframe()

```
# Função SE
SE<-function(vetor,IC){
  SD<-sd(vetor)
  N<-length(vetor)
  SE<-(SD*IC)/sqrt(N)

   return(
    SE
  )
}

# Summarise
volc %>%
  group_by(country) %>%
  summarise(n=n(),
            media_pop_5km=mean(population_within_5_km),
            sd_pop_km=sd(population_within_5_km),
            se_pop_5km=SE(population_within_5_km,1),
            max_pop_5km=max(population_within_5_km),
            min_pop_5km=min(population_within_5_km)) %>%
  arrange(desc(media_pop_5km)) %>%
  filter(n>15) %>%
  mutate(across(.cols=media_pop_5km:min_pop_5km,~round(.x,digits=0)))
```

```
# A tibble: 16 x 7
   country         n media_pop_5km sd_pop_km se_pop_5km max_pop_5km min_pop_5km
   <chr>       <int>         <dbl>     <dbl>      <dbl>       <dbl>       <dbl>
 1 Mexico         28        269105   1091616     206296     5783287           0
 2 Philippines    27        102291    306707      59026     1349742          11
 3 Nicaragua      16         97064    251139      62785      989888         100
 4 New Zealand    16         71199    261129      65282     1049110           0
```

```
 5 Ethiopia         26          69227    148454       29114      565206          0
 6 Guatemala        17          63078     88080       21362      240892       1373
 7 Ecuador          26          41014    117168       22979      534403          0
 8 Indonesia        95          35195    129642       13301     1092929          3
 9 Japan            92           6355     19149        1996      130474          0
10 Iceland          28           3496      8458        1598       33526          0
11 Papua New G~     28           2978      5215         986       24509          1
12 United Stat~     99           1745      7356         739       52801          0
13 Chile            43            459      1233         188        5951          0
14 Canada           18            444      1559         367        6635          0
15 Russia           79             24        96          11         669          0
16 Antarctica       17              0         0           0           0          0
```

```
# México
#n=28

#269105.1 ± 206295.9 habitantes(sigma 1)

# Guatemala
# n = 17
# 63078 ± 21362 habitantes
```

#3 Exercicio

- 1) Em qual país existem mais vulcões?

```
volc %>%
  count(country,sort=TRUE)
```

```
# A tibble: 89 x 2
   country              n
   <chr>            <int>
 1 United States       99
 2 Indonesia           95
 3 Japan               92
 4 Russia              79
 5 Chile               43
 6 Iceland             28
 7 Mexico              28
 8 Papua New Guinea    28
 9 Philippines         27
10 Ecuador             26
```

```
# i 79 more rows
```

```
volc %>%
  group_by(country) %>%
  summarise(n=n()) %>%
  arrange(desc(n))
```

```
# A tibble: 89 x 2
   country              n
   <chr>            <int>
 1 United States       99
 2 Indonesia           95
 3 Japan               92
 4 Russia              79
 5 Chile               43
 6 Iceland             28
 7 Mexico              28
 8 Papua New Guinea    28
 9 Philippines         27
10 Ecuador             26
# i 79 more rows
```

- 2) Nesse país, qual o tipo de vulcão mais comum?

```
volc %>%
  filter(country=="United States") %>%
  count(primary_volcano_type,sort=TRUE)
```

```
# A tibble: 15 x 2
   primary_volcano_type     n
   <chr>                <int>
 1 Stratovolcano           32
 2 Shield                  15
 3 Volcanic field          13
 4 Stratovolcano(es)        8
 5 Submarine                6
 6 Pyroclastic cone(s)      5
 7 Shield(s)                5
 8 Lava dome(s)             4
 9 Caldera                  3
```

```
10 Maar(s)                  3
11 Caldera(s)               1
12 Complex                  1
13 Pyroclastic cone         1
14 Stratovolcano?           1
15 Tuff cone(s)             1
```

- 3) Na América do Sul, quantos e quais vulcões entraram em erupção em 2020?

```
volc %>%
  filter(region=="South America",
         last_eruption_year==2020) %>%
  count(volcano_name,sort=TRUE)
```

```
# A tibble: 6 x 2
  volcano_name            n
  <chr>               <int>
1 Chillan, Nevados de     1
2 Reventador              1
3 Ruiz, Nevado del        1
4 Sabancaya               1
5 Sangay                  1
6 Villarrica              1
```

- 4) Quais os países nos quais os vulcões apresentam maior risco para a população local (até 100 km)? Filtre países em que existam mais de 20 vulcões.

```
volc %>%
  group_by(country) %>%
  summarise(n=n(),
            media=mean(population_within_100_km),
            se=SE(population_within_100_km,1)) %>%
  filter(n>20) %>%
  arrange(desc(media))
```

```
# A tibble: 11 x 4
   country               n     media       se
   <chr>             <int>     <dbl>    <dbl>
 1 Philippines          27 10247171. 1948086.
 2 Indonesia            95  9126614. 1148771.
```

3	Mexico	28	7034976.	1863798.
4	Japan	92	4057450.	537928.
5	Ethiopia	26	3318110.	683117.
6	Ecuador	26	1819886.	317923.
7	Chile	43	362838.	42331.
8	United States	99	250535.	55360.
9	Papua New Guinea	28	102845.	10647.
10	Iceland	28	65432.	16041.
11	Russia	79	60141.	20473.