

Ejercicios Clase Tidyverse 1

Laboratorio Políticas Públicas

21/4/2020

Practiquemos lo que aprendimos hoy!

Vamos a trabajar con el dataset presente en el siguiente link. Llamemos a Tidyverse y carguemos nuestro dataset antes de empezar (acordate de modificar el encoding)!

```
#activemos tidyverse
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## -- Attaching packages ----- tidyverse 1.3.0
```

```
## v ggplot2 3.3.0      v purrr   0.3.3
## v tibble  2.1.3      v dplyr  0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'readr' was built under R version 3.6.3
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'stringr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
## -- Conflicts ----- tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
#modificamos la notacion cientifica
options(scipen = 20)
```

```
#carguemos nuestro dataset
```

```
data <- read.csv("https://github.com/labpoliticasuba/Clases_2020/raw/master/Clase_02/data_final_clase.csv")
```

```
#demos un pantallazo a nuestro dataset
head(data)
```

```
##      X  REGION      DISTRITO POBLACION      TIPO_CAMA CANTIDAD_CAMAS
## 1 1 PAMPEANA Buenos Aires 17196396 Aislamiento y casos leves      37337
## 2 2 PAMPEANA Buenos Aires 17196396      Casos graves      57151
## 3 3 PAMPEANA Buenos Aires 17196396      Casos graves Neo      2180
## 4 4 PAMPEANA      CABA 3068043 Aislamiento y casos leves      16173
## 5 5 PAMPEANA      CABA 3068043      Casos graves      5095
## 6 6 PAMPEANA      CABA 3068043      Casos graves Neo      1000
##      CANTIDAD_MEDICOS
## 1      46223
## 2      46223
## 3      46223
## 4      48447
## 5      48447
## 6      48447
```

1. ¿Cuáles son las 5 provincias con mayor cantidad de médicos por cantidad de habitantes? Vamos paso a paso:

a. Seleccione las columnas que vamos a utilizar para este punto en un nuevo data frame llamado “ej_1_a”

```
ej_1_a <- data %>% select(DISTRITO, CANTIDAD_MEDICOS, POBLACION)
head(ej_1_a)
```

```
##      DISTRITO CANTIDAD_MEDICOS POBLACION
## 1 Buenos Aires      46223 17196396
## 2 Buenos Aires      46223 17196396
## 3 Buenos Aires      46223 17196396
## 4      CABA      48447 3068043
## 5      CABA      48447 3068043
## 6      CABA      48447 3068043
```

b. Usando el dataframe realizado en el 1.a) elabore una nueva columna llamada “medicos_x_hab” (que contenga la información CANTIDAD_MEDICOS/POBLACION). Todo esto en un nuevo dataframe que llamaremos “ej_1_b”

```
ej_1_b <- ej_1_a %>% mutate(medicos_x_hab = CANTIDAD_MEDICOS/POBLACION)
head(ej_1_b)
```

```
##      DISTRITO CANTIDAD_MEDICOS POBLACION medicos_x_hab
## 1 Buenos Aires      46223 17196396 0.002687947
## 2 Buenos Aires      46223 17196396 0.002687947
```

```
## 3 Buenos Aires      46223  17196396  0.002687947
## 4      CABA          48447   3068043  0.015790848
## 5      CABA          48447   3068043  0.015790848
## 6      CABA          48447   3068043  0.015790848
```

- c. Elabore un `group_by` y seleccione las 5 provincias con mayor cantidad de médicos por habitantes en un nuevo dataframe “ej_1_c”

```
ej_1_c <- ej_1_b %>% group_by(DISTRITO, medicos_x_hab) %>%
  distinct(DISTRITO) %>%
  arrange(desc(medicos_x_hab)) %>%
  head(5)

head(ej_1_c)
```

```
## # A tibble: 5 x 2
## # Groups:   DISTRITO, medicos_x_hab [5]
##   DISTRITO      medicos_x_hab
##   <fct>          <dbl>
## 1 CABA          0.0158
## 2 Santa Fe      0.00449
## 3 Córdoba      0.00442
## 4 Tierra del Fuego 0.00426
## 5 Neuquen      0.00384
```

- d. ¿Y la que tiene menor cantidad de médicos de la región de “CUYO”? (elabore un nuevo dataframe llamado ej_1_d)

```
ej_1_d <- data %>% select(REGION, DISTRITO, CANTIDAD_MEDICOS, POBLACION) %>%
  filter(REGION == "CUYO") %>%
  mutate(medicos_x_hab = CANTIDAD_MEDICOS/POBLACION) %>%
  group_by(DISTRITO, medicos_x_hab) %>%
  distinct(DISTRITO) %>%
  arrange(desc(medicos_x_hab)) %>%
  tail(1)

head(ej_1_d)
```

```
## # A tibble: 1 x 2
## # Groups:   DISTRITO, medicos_x_hab [1]
##   DISTRITO medicos_x_hab
##   <fct>          <dbl>
## 1 San Luis      0.00267
```

```
#este ejercicio tamb se puede hacer con head():
data %>% select(REGION, DISTRITO, CANTIDAD_MEDICOS, POBLACION) %>%
  filter(REGION == "CUYO") %>%
  mutate(medicos_x_hab = CANTIDAD_MEDICOS/POBLACION) %>%
  group_by(DISTRITO, medicos_x_hab) %>%
  distinct(DISTRITO) %>%
  arrange(medicos_x_hab) %>%
  head(1)
```

```
## # A tibble: 1 x 2
## # Groups:   DISTRITO, medicos_x_hab [1]
##   DISTRITO medicos_x_hab
##   <fct>         <dbl>
## 1 San Luis       0.00267
```

2. a. ¿Cuál es el porcentaje de camas para casos graves (adultos) que tiene cada provincia respecto del total de camas a nivel nacional? Elabore un nuevo dataframe llamado ej_2_a

```
#le preguntamos a R cómo estan redactados los casos graves
unique(data$TIPO_CAMA)
```

```
## [1] Aislamiento y casos leves Casos graves
## [3] Casos graves Neo
## Levels: Aislamiento y casos leves Casos graves Casos graves Neo
```

```
ej_2_a <- data %>% select(DISTRITO, TIPO_CAMA, CANTIDAD_CAMAS) %>%
  filter(TIPO_CAMA == "Casos graves") %>%
  mutate(PORCENTAJE = CANTIDAD_CAMAS/sum(CANTIDAD_CAMAS))
head(ej_2_a)
```

```
##      DISTRITO      TIPO_CAMA CANTIDAD_CAMAS  PORCENTAJE
## 1 Buenos Aires Casos graves      57151 0.555165915
## 2      CABA Casos graves      5095 0.049492928
## 3 Catamarca Casos graves      539 0.005235856
## 4      Chaco Casos graves      1421 0.013803621
## 5      Chubut Casos graves      457 0.004439307
## 6 Corrientes Casos graves      583 0.005663273
```

```
### si queremos que quede más lindo (redondamos a 3 digitos y ponemos "%"):
ej_2_a2 <- data %>% select(DISTRITO, TIPO_CAMA, CANTIDAD_CAMAS) %>%
  filter(TIPO_CAMA == "Casos graves") %>%
  mutate(PORCENTAJE = paste0(round( CANTIDAD_CAMAS/sum(CANTIDAD_CAMAS),2), '%') )
head(ej_2_a2)
```

```
##      DISTRITO      TIPO_CAMA CANTIDAD_CAMAS  PORCENTAJE
## 1 Buenos Aires Casos graves      57151      0.56%
## 2      CABA Casos graves      5095      0.05%
## 3 Catamarca Casos graves      539      0.01%
## 4      Chaco Casos graves      1421      0.01%
## 5      Chubut Casos graves      457      0%
## 6 Corrientes Casos graves      583      0.01%
```

3. a. ¿Cuál es la región con mayor cantidad de camas para casos de aislamiento y leves- por provincia- en promedio?

```
data %>% select(REGION, DISTRITO, TIPO_CAMA, CANTIDAD_CAMAS) %>%
  filter(TIPO_CAMA == "Aislamiento y casos leves" & REGION == "PAMPEANA")
```

```
##      REGION      DISTRITO      TIPO_CAMA CANTIDAD_CAMAS
## 1 PAMPEANA Buenos Aires Aislamiento y casos leves      37337
```

```
## 2 PAMPEANA          CABA Aislamiento y casos leves          16173
## 3 PAMPEANA          Córdoba Aislamiento y casos leves        9311
## 4 PAMPEANA          Entre Rios Aislamiento y casos leves      4048
## 5 PAMPEANA          La Pampa Aislamiento y casos leves        887
## 6 PAMPEANA          Santa Fe Aislamiento y casos leves        8268
```

```
data %>% select(REGION, DISTRITO, TIPO_CAMA, CANTIDAD_CAMAS) %>%
  filter(TIPO_CAMA == "Aislamiento y casos leves") %>%
  group_by(REGION) %>%
  distinct() %>%
  summarise(PROMEDIO_PROVINCIA = mean(CANTIDAD_CAMAS)) %>%
  arrange(desc(PROMEDIO_PROVINCIA)) %>%
  head(1)
```

```
## # A tibble: 1 x 2
##   REGION    PROMEDIO_PROVINCIA
##   <fct>          <dbl>
## 1 PAMPEANA      12671.
```

- b. De las provincias con menos de 1 millón de habitantes, ¿cuál es la que mayor cantidad de camas para aislamiento por cada 5000 habitantes ? Genere un dataframe llamado “ej_3_b”

```
ej_3_b <- data %>% select(DISTRITO, POBLACION, TIPO_CAMA, CANTIDAD_CAMAS) %>%
  filter(POBLACION <= 1000000 & TIPO_CAMA == "Aislamiento y casos leves") %>%
  mutate(CAMA_C_CINCO_MIL = (CANTIDAD_CAMAS/POBLACION)*5000) %>%
  arrange(desc(CAMA_C_CINCO_MIL)) %>%
  head(1)
ej_3_b
```

```
##   DISTRITO POBLACION          TIPO_CAMA CANTIDAD_CAMAS CAMA_C_CINCO_MIL
## 1 La Rioja    383220 Aislamiento y casos leves          1411      18.40979
```

DESAFÍO OPCIONAL:

4. a. Suponiendo que un paciente adulto infectado grave necesita en promedio 10 días para recuperarse en una cama de terapia intensiva. En el periodo de un mes (30 días)... ¿Cuántas personas graves tiene la posibilidad de atender la provincia de la consigna 3.b. en el periodo determinado sin que el sistema de salud colapse?

```
#vemos cómo esta redactado tipo_cama
unique(data$TIPO_CAMA)
```

```
## [1] Aislamiento y casos leves Casos graves
## [3] Casos graves Neo
## Levels: Aislamiento y casos leves Casos graves Casos graves Neo
```

```
#vemos cómo esta redactado tipo_cama
unique(data$DISTRITO)
```

```
## [1] Buenos Aires      CABA      Catamarca
## [4] Chaco                Chubut    Corrientes
## [7] Córdoba              Entre Rios Formosa
## [10] Jujuy                La Pampa  La Rioja
## [13] Mendoza              Misiones  Neuquen
## [16] Río Negro            Salta     San Juan
## [19] San Luis             Santa Cruz Santa Fe
## [22] Santiago del Estero Tierra del Fuego Tucumán
## 24 Levels: Buenos Aires CABA Catamarca Chaco Chubut Córdoba ... Tucumán
```

```
ej_4_a <- data %>% filter((TIPO_CAMA == "Casos graves") & DISTRITO == "La Rioja") %>%
  mutate(PACIENTES_30_DIAS = (30*CANTIDAD_CAMAS/10)) %>%
  select( PACIENTES_30_DIAS)
```

```
#tamb podemos copiar y pegar el numero
head(ej_4_a)
```

```
##      PACIENTES_30_DIAS
## 1                1050
```

- b. ¿Cómo sería la respuesta a la pregunta 4.a. para “Ciudad de Buenos Aires” y “Provincia de Buenos Aires”?

```
#vemos cómo esta redactado DISTRITO CABA Y BS AS
unique(data$DISTRITO)
```

```
## [1] Buenos Aires      CABA      Catamarca
## [4] Chaco                Chubut    Corrientes
## [7] Córdoba              Entre Rios Formosa
## [10] Jujuy                La Pampa  La Rioja
## [13] Mendoza              Misiones  Neuquen
## [16] Río Negro            Salta     San Juan
## [19] San Luis             Santa Cruz Santa Fe
## [22] Santiago del Estero Tierra del Fuego Tucumán
## 24 Levels: Buenos Aires CABA Catamarca Chaco Chubut Córdoba ... Tucumán
```

```
ej_4_b <- data %>% filter((TIPO_CAMA == "Casos graves") & (DISTRITO == "Buenos Aires" | DISTRITO == "CABA")) %>%
  mutate(PACIENTES_30_DIAS = (30*CANTIDAD_CAMAS/10)) %>%
  select(DISTRITO, PACIENTES_30_DIAS)
```

```
#tamb podemos copiar y pegar el numero
head(ej_4_b)
```

```
##      DISTRITO PACIENTES_30_DIAS
## 1 Buenos Aires      171453
## 2 CABA                15285
```