

Tarea 2

Santiago de Mello

03/10/2018

0.1 Introducción

Para esta tarea se utilizan los datos abiertos del catálogo de Uruguay: <https://catalogodatos.gub.uy>.

Los datos corresponden a los gastos realizados por actos médicos, **cada fila representa un acto médico**. Los datos y los metadatos se encuentran disponibles:

https://catalogodatos.gub.uy/dataset/gasto_am_2016_fondo-nacional-de-recursos/resource/936ac9e6-b0f6-424a-9b53-ee408a91fd6f

El repositorio de GitHub de la entrega es: https://github.com/SantiagodeMello/cursoIESTA/blob/master/RMarkdown/Tarea2_Santiago_de_Mello.Rmd

0.2 Ejercicio 1

Usando las funciones de la librería `dplyr` respondé:

- a. ¿Cuál es la prestación con mayor cantidad de actos médicos en Montevideo?

```
# Descargar y leavantar los datos
```

```
gastolink <- 'https://catalogodatos.gub.uy/dataset/96e636e5-4f78-49a7-8e14-60e90173a0c'
```

```
gastos <- read.csv(gastolink, header = TRUE, dec = ",", encoding="latin1")
```

```
library(dplyr)
```

```
library(tidyverse)
```

```
# Ejercicio 1
```

```
# a.
```

```
Eje.1.a = filter(gastos, Prestador_departamento == "MONTEVIDEO") %>%  
  group_by(Prestacion) %>% summarise(Prestacion_n = n()) %>%  
  arrange(Prestacion_n=desc(Prestacion_n)) %>% top_n(1)
```

Respuesta: La prestación con mayor cantidad de actos médicos en Montevideo es PCI-Cateterismo izq.adultos con 2514 actos médicos.

- b. Creá una variable con los totales de actos médicos por Departamento de residencia (Departamento_residencia). ¿Cuál es el departamento de residencia con menor cantidad de actos médicos?

Ejercicio 1.b

```
Eje.1.b = group_by(gastos, Prestador_departamento) %>%  
  summarise(Prestaciones_por_departamento=n()) %>%  
  arrange(Prestacion_n=(Prestaciones_por_departamento), Prestaciones_por_departamento)
```

El departamento de residencia con menor cantidad de actos médicos es FLORES con 125 prestaciones.

- c. ¿Qué cantidad de actos médicos son prestados por ASSE o IAMC?

Ejercicio 1.c

```
Eje.1.c = filter(gastos, Prestador_tipo=="ASSE" | Prestador_tipo=="IAMC") %>%  
  group_by(Prestador_tipo) %>%  
  summarise(Prestador_tipo_n = n() ) # Los ordena alfabeticamente por Prestador_tipo
```

Respuesta: ASSE presta 7428 y IAMC presta 15550 actos médicos.

- d. Cada fila representa un acto médico, por lo que puede haber filas que se correspondan con la misma persona. ¿Cómo se puede verificar si esto es así? ¿Cuántos pacientes distintos hay en los datos?

Ejercicio 1.d

```
Eje.1.d.i = gastos %>%  
  group_by(Paciente) %>%  
  summarise(Paciente_n = n()) %>%  
  filter(Paciente_n > 1)  
  
if (nrow(Eje.1.d.i) != 0) print("Hay pacientes con mas de dos actos médicos")
```

```
## [1] "Hay pacientes con mas de dos actos médicos"
```

```
Eje.1.d.ii = gastos %>%  
  group_by(Paciente) %>%  
  summarise(Paciente_n = n())
```

```
Pacientes_distintos=nrow(Eje.1.d.ii)
```

La cantidad de pacientes atendidos es de 19911

- d. Crear un **nuevo** conjunto de datos en que cada fila sea un paciente. Agregar dos

variables: el gasto total de la persona en actos médicos y la cantidad de visitas. Conservá el resto de las variables originales menos `Prestacion` e `Importe` (Sugerencia usar `summarise_all()`).

Ejercicio 1.e

```
Eje.1.e = gastos %>%  
  group_by(Paciente) %>%  
  group_by(Gastos_paciente = sum(Importe), Cantidad_actos = n(), -Importe, -Prestacion)
```

0.3 Ejercicio 2

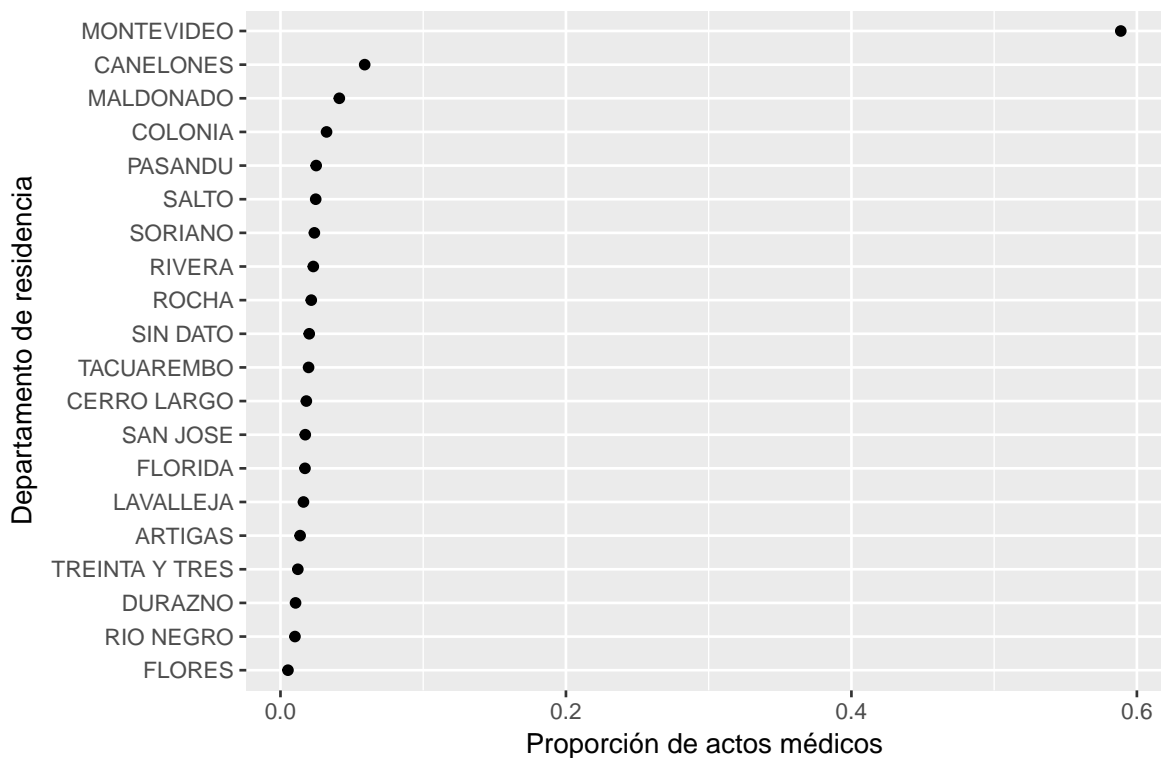
- a. Replique el siguiente gráfico (Figura 1) usando `ggplot2` y `forcats` para ordenar el gráfico.

Ejercicio 2.a

```
library(ggplot2)
library(forcats)
```

```
Prestaciones_por_departamento = gastos %>% group_by(Prestador_departamento) %>%
  summarise(Prestaciones_por_departamento = n()), Proportion= sum(Prestaciones_por_depa
  mutate(Proportion= Prestaciones_por_departamento/sum(Prestaciones_por_departamento))
```

```
ggplot(data = Prestaciones_por_departamento, aes(x = Proportion, y = fct_reorder(Pres
  geom_point() +
  ggtitle("") +
  xlab("Proporción de actos médicos") + ylab("Departamento de residencia")
```

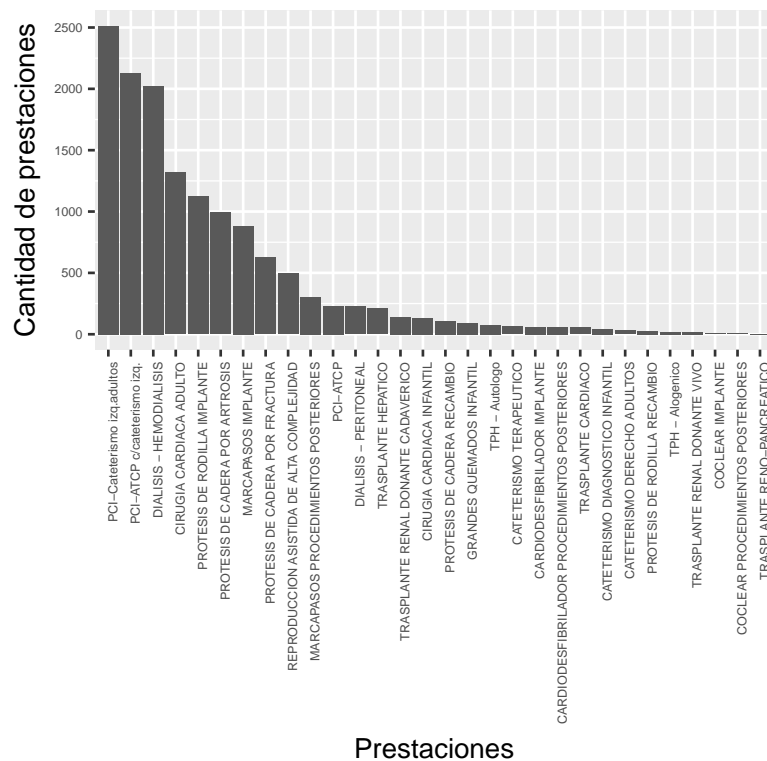


- b. Replique el siguiente gráfico (Figura 2) usando `ggplot2` y `forcats` para ordenar.

Ejercicio 2.b

```
Prestaciones_MVD = filter(gastos, Prestador_departamento == "MONTEVIDEO") %>%
  group_by(Prestacion) %>% summarise(Prestacion_n = n()) %>%
  arrange(Prestacion_n=desc(Prestacion_n))

ggplot(Prestaciones_MVD, aes(x= fct_reorder(Prestacion, Prestacion_n, .desc = T), y =
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), aspect.ratio = .5, axis.tex
  ggtitle("") +
  xlab("Prestaciones") + ylab("Cantidad de prestaciones")
```



c. Replique el siguiente gráfico (Figura 3) usando `ggplot2` y `forcats` para ordenar.

Notar que el orden de los niveles de `Prestacion` fueron ordenados haciendo:

- una variable auxiliar que vale 1 si `Prestador_tipo` es ASSE y 0 en otro caso
- ordenamos los niveles de prestación según la media de la variable auxiliar
- usamos la función `fct_reorder`

Comente algo interesante que surge de este gráfico.

```
# Ejercicio 2.c
```

0.4 Ejercicio 3

- a. Usando `ggplot2` elabore una visualización que permita responder la pregunta de ¿Cuáles son las 10 instituciones prestadoras (`Prestador`) que brindaron mayor proporción de actos médicos en Montevideo (`Prestador_departamento`)?

Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un `caption` (Título) en la figura y algún comentario de interés que describa el gráfico. Puede utilizar `fig.cap` en el chunk de código.

- b. Usando `ggplot2` elabore un gráfico de cajas con el importe del acto médico (en logaritmos) según tipo de prestador y sexo.

Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un `caption` (Título) en la figura y algún comentario de interés que describa el gráfico. Puede utilizar `fig.cap` en el chunk de código.

- c. Se desea explorar la asociación entre la edad del paciente y el importe de los actos médicos (en logaritmos). Realiza alguna visualización para estudiar dicha asociación, y ver como esta varía según el sexo del paciente y el tipo de prestador.
- d. Realiza alguna visualización para estudiar el gasto promedio por persona en cada departamento, tipo de prestador y sexo.
- e. Realiza alguna visualización para estudiar el peso de las prestaciones en cantidad de actos y en monto relativo. ¿Son las prestaciones más comunes las más caras?