

Tarea 2

Daniel Ramírez

3/10/2018

Explicativo

Esta tarea debe ser entregada el miércoles 3 de Octubre . Debe ser realizada en Rmarkdown y debe incluir en el documento el link a tu repositorio personal en GitHub con los archivos fuente para reproducir los resultados. A su vez debe compartir el pdf en EVA.

Los datos que vamos a utilizar en este ejercicio están disponibles en el catálogo de datos abiertos 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-ee40

Los pueden leer en R de la siguiente forma:

```
gastolink <- 'https://catalogodatos.gub.uy/dataset/96e636e5-4f78-49a7-8e14-60e90173a0c0/resource/936ac9e6-b0f6-424a-9b53-ee40'
gastos <- read.csv(gastolink, header = TRUE, dec = ",", encoding="latin1")
```

Ejercicio 1

Usando las funciones de la librería `dplyr` respondé:

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

```
library(dplyr)
gastos %>%
  filter(Departamento_residencia == "MONTEVIDEO") %>%
  group_by(Prestacion) %>%
  summarise(length(Prestacion)) %>%
  filter(`length(Prestacion)` == max(`length(Prestacion)`))
```

```
## # A tibble: 1 x 2
##   Prestacion          `length(Prestacion)`
##   <fct>                <int>
## 1 PCI-Cateterismo izq.adultos             2045
```

- 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?

```
gastos %>%
  group_by(Departamento_residencia) %>%
  summarise(length(Prestacion)) %>%
  filter(`length(Prestacion)` == min(`length(Prestacion)`))
```

```
## # A tibble: 1 x 2
##   Departamento_residencia `length(Prestacion)`
##   <fct>                  <int>
## 1 FLORES                  138
```

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

```
gastos %>%
  filter(Prestador_tipo == 'ASSE' | Prestador_tipo == 'IAMC' )%>%
  summarise(length(Prestacion))
```

```
##   length(Prestacion)
## 1                22978
```

e. 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?

```
gastos %>%
  summarise(length(unique(Paciente)))
```

```
##   length(unique(Paciente))
## 1                      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`).

```
nuevo <- gastos %>%
  group_by(Paciente) %>%
  summarise(gastos_totales = sum(Importe), visitas = length(Paciente))

nuevo <- merge(nuevo, gastos, by = 'Paciente')

nuevo$dupl <- duplicated(nuevo$Paciente)
nuevo <- nuevo[nuevo$dupl == 'FALSE',]
nuevo <- select(nuevo, c(1,5:10,2,3))
```

Ejercicio 2

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

```
library(ggplot2)
library(forcats)

plot <- gastos %>%
  group_by(Departamento_residencia) %>%
  summarise(length(Prestacion)) %>% as.data.frame()

plot$Propor <- plot$`length(Prestacion)`/sum(plot$`length(Prestacion)`)

ggplot(plot,aes(x = Propor , y = fct_reorder(Departamento_residencia, Propor)))+ geom_point()+xlab('Propor
```

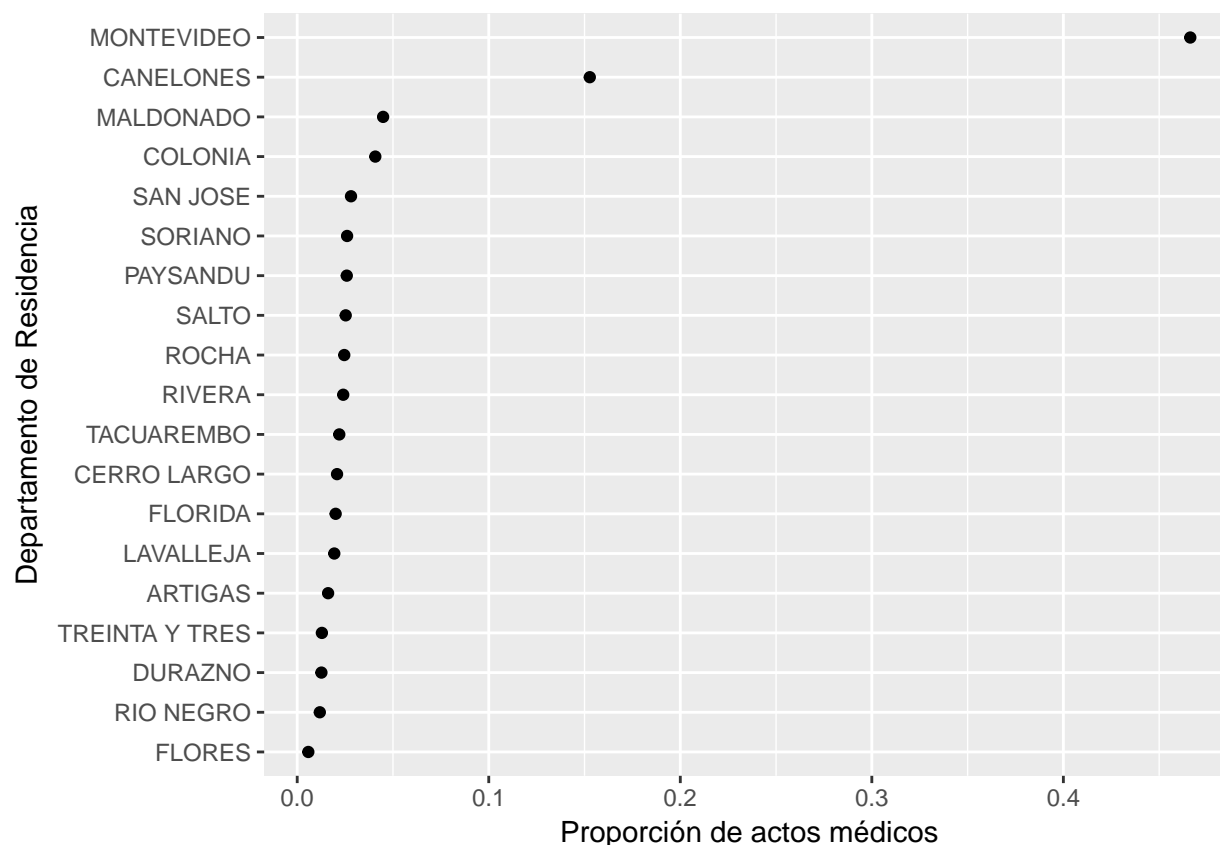


Figure 1: Proporción de actos médicos según departamento de residencia

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

```
gastos %>%
  group_by(Prestacion) %>%
  summarise(length(Prestacion)) %>%
  ggplot(aes(x = fct_reorder(Prestacion, `length(Prestacion)`, .desc = TRUE ), y = `length(Prestacion)` )
```

```
geom_bar(stat = "identity" ) +
ylab('Total de actos médicos')+
xlab(NULL) +
theme(axis.text.x = element_text (angle =90, size =rel(0.6)))
```

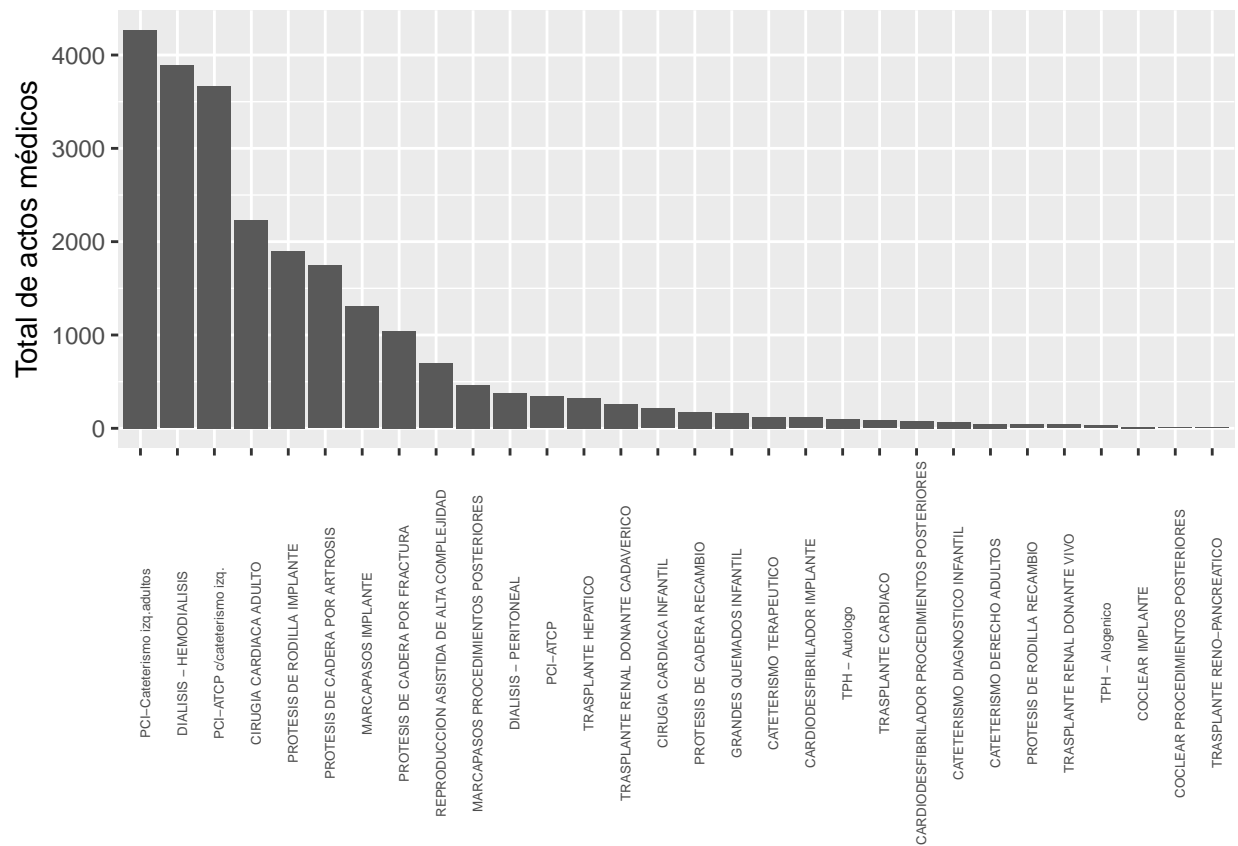


Figure 2: Gráfico de barras de la cantidad de actos médicos en Montevideo por tipo de Prestación

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.

```
gastos %>%
  mutate(auxiliar = ifelse(Prestador_tipo == 'ASSE',1,0)) %>%
  ggplot(aes(x= fct_reorder(Prestacion, auxiliar, .fun = mean), fill = Prestador_tipo)) + geom_bar(position = 'stack',
  coord_flip() +
  theme(axis.text.y = element_text (size =rel(0.6)), legend.position = 'bottom') +
  xlab('Tipo de prestación')
```

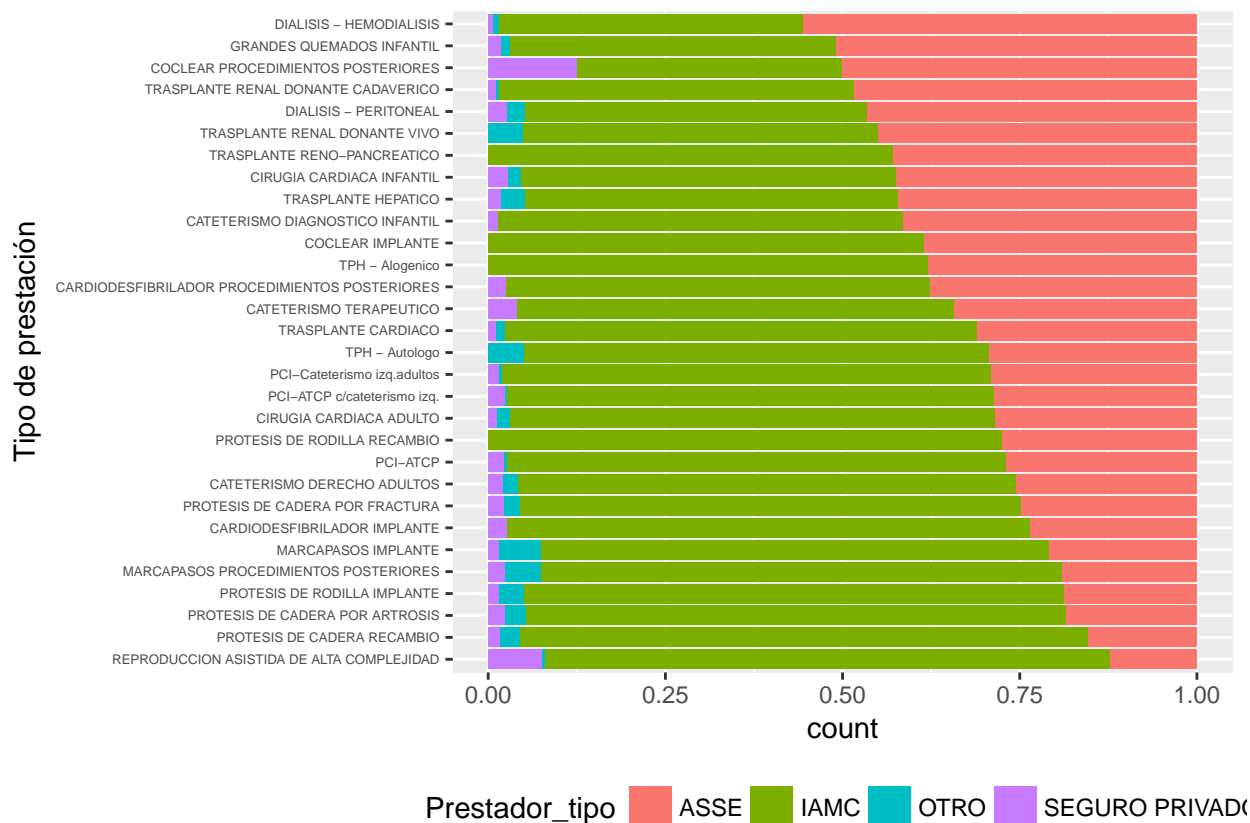


Figure 3: Gráfico de barras apiladas al 100 de la cantidad de actos médicos en Montevideo por tipo de Prestacion

Para casi todas las prestaciones, la proporción de mutualistas que las realizan, es mayor que todos los otros tipos de cobertura (excepto por cuatro o tres).

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`)?

```
gastos %>%
  filter(Prestador_departamento == "MONTEVIDEO") %>%
  group_by(Prestador) %>%
  summarise(length(Prestacion)) %>%
  ggplot(aes(x = fct_reorder(Prestador, `length(Prestacion)`, .desc = TRUE), y = `length(Prestacion)`)) +
  xlab('Prestador') +
  ylab('Actos M?dicos') +
  theme(axis.text.x = element_text(angle = 90, size = rel(0.6)))
```

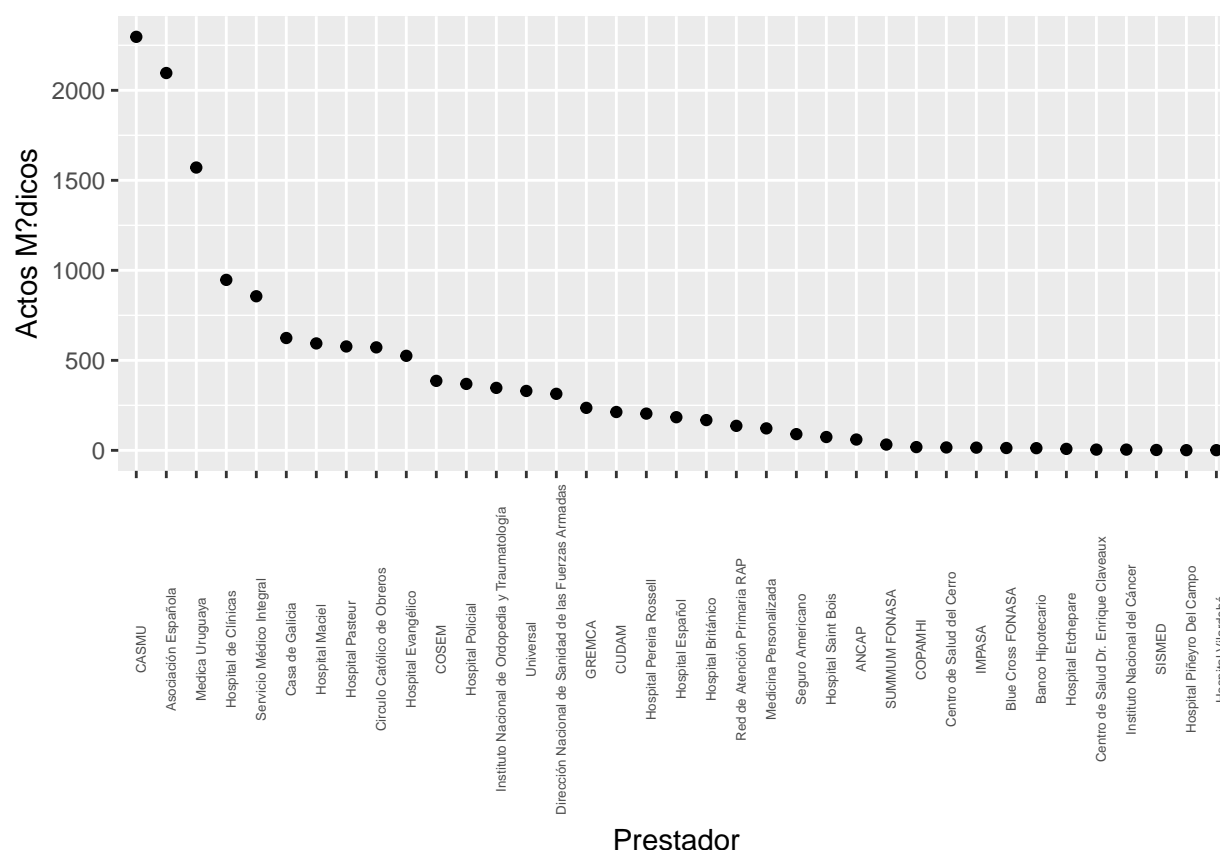


Figure 4: Actos Médicos por Prestador en Montevideo

En el gráfico se puede observar los prestadores que más actos médicos realizaron. CASMU, Asociación Española y Médica Uruguay ocupan los primeros lugares.

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.

```
gastos %>%
  ggplot(aes(x= Prestador_tipo, y=log(Importe))) +
  geom_boxplot(width=.5)+
  facet_wrap(~ Sexo)+
  xlab('Prestador')+
  ylab('Importe en logaritmo')
```

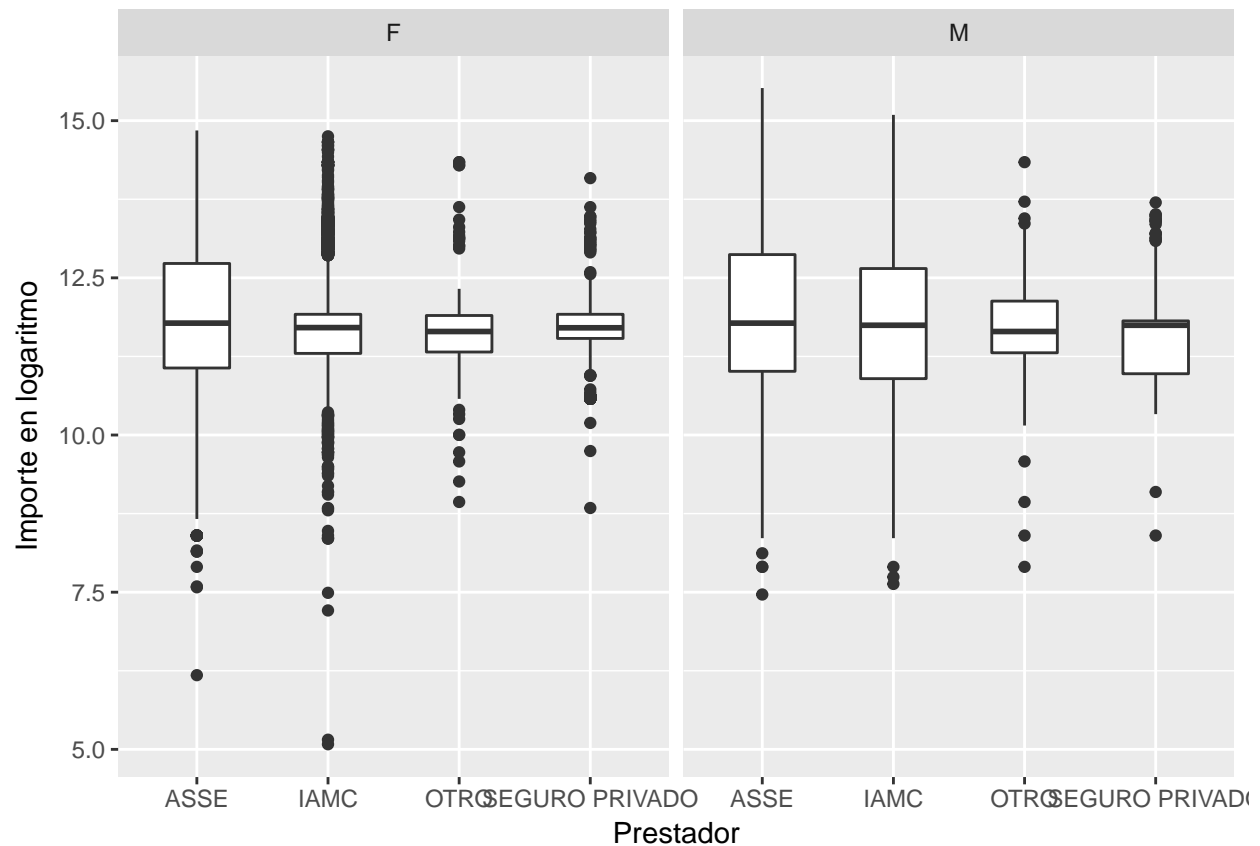


Figure 5: Gráfico de Cajas por tipo de prestador y sexo

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.

```
gastos %>%
  ggplot(aes(x= Edad_años, y=log(Importe))) +
  geom_point()+
  facet_wrap(Prestador_tipo ~ Sexo)+
  xlab('Edad')+
  ylab('Importe en logaritmo')
```

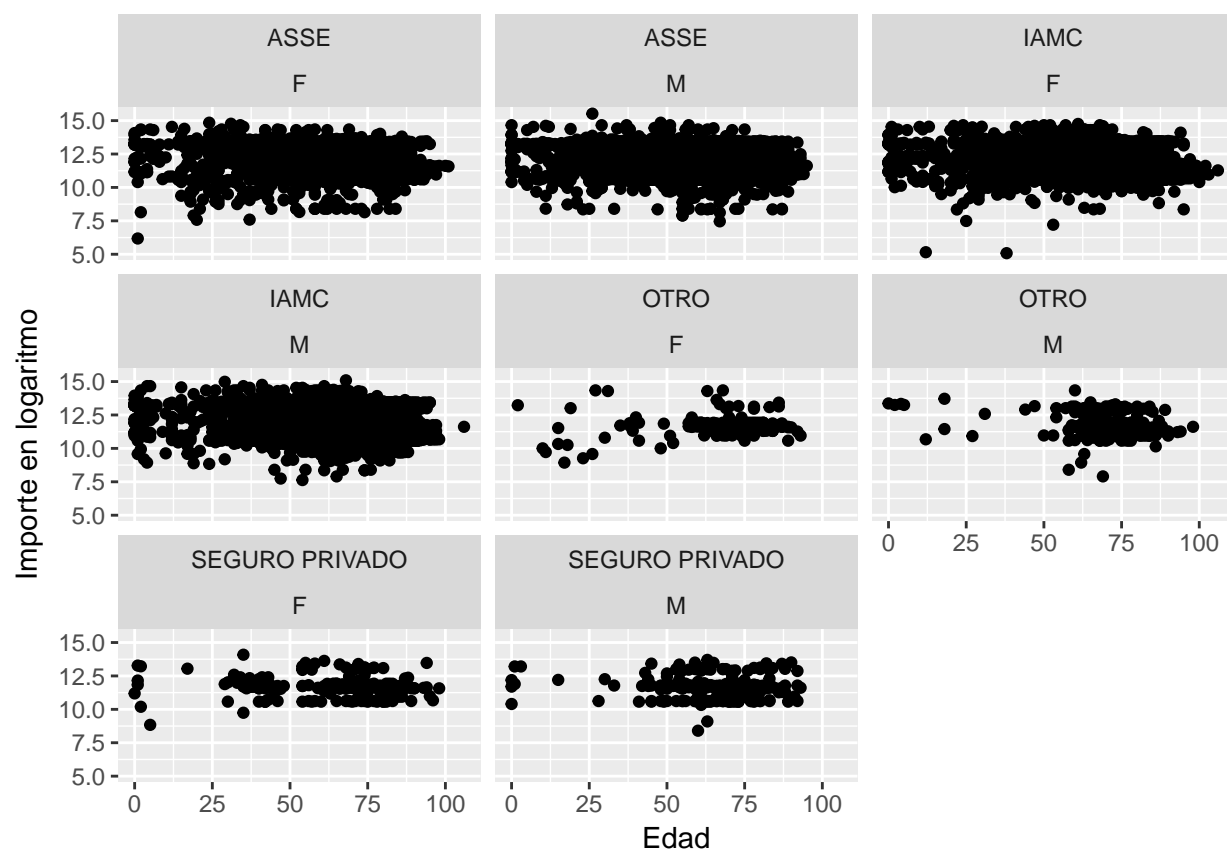


Figure 6: Gráfico de dispersión por tipo de prestador y sexo

- d. Realiza alguna visualización para estudiar el gasto promedio por persona en cada departamento, tipo de prestador y sexo.

```
nuevo %>%
  group_by_at(vars(Prestador_departamento, Prestador_tipo, Sexo)) %>%
  summarise(mean(gastos.totales)) %>%
  ggplot(aes(x= Prestador_tipo, y= `mean(gastos.totales)`, fill=Sexo)) +
  geom_bar(position="dodge", colour="black", stat="identity")+
  facet_wrap(~Prestador_departamento )+
  xlab('Tipo de Prestador')+
  ylab('Media de Gastos')+
  theme(axis.text.y = element_text (size =rel(0.6)))+
  theme(axis.text.x= element_text (size =rel(0.3)))
```

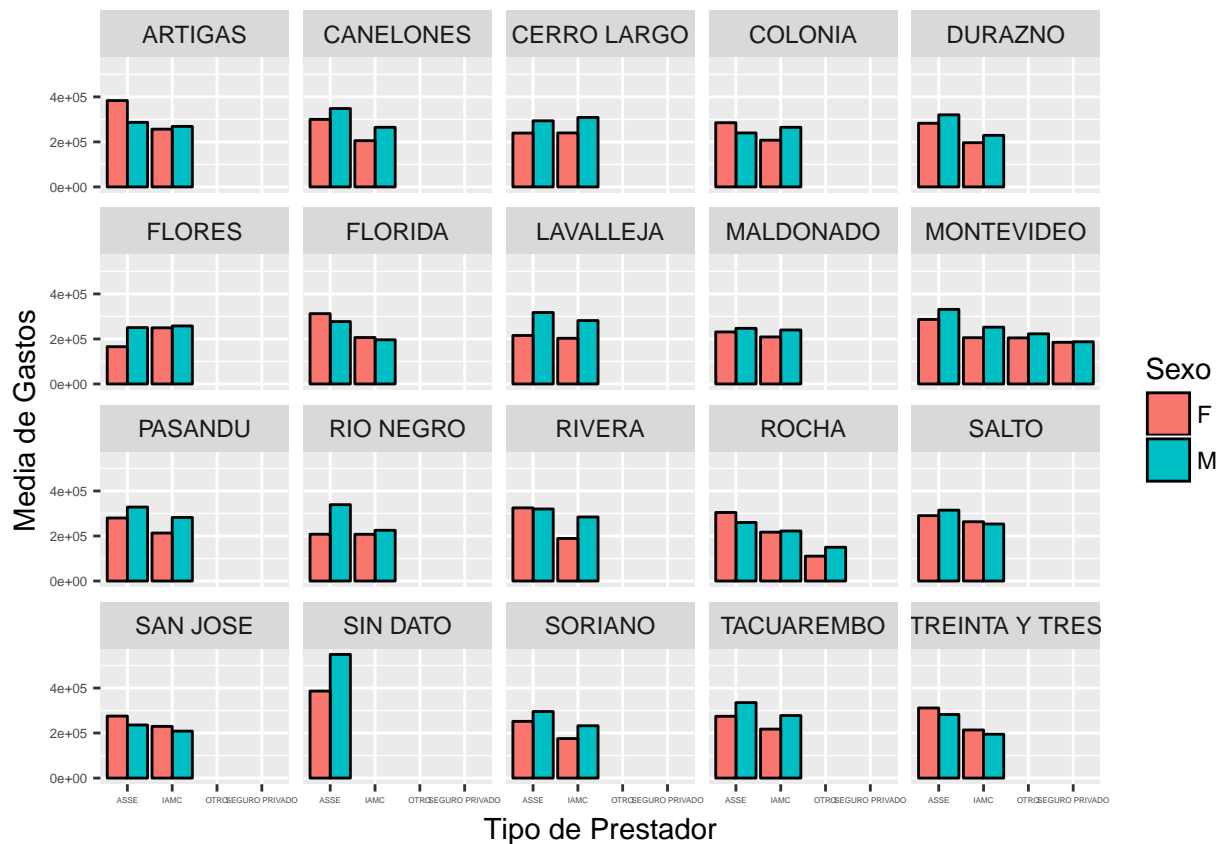


Figure 7: Gráfico de dispersión por tipo de prestador y sexo

- 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?

```
gastos %>%
  group_by(Prestacion) %>%
  summarise(length(Prestacion), mean(Importe)) %>%
  ggplot(aes(x=`length(Prestacion)`, y= `mean(Importe)`))+ geom_point()+
  geom_text(aes(label=Prestacion), size=1)+
```

```
xlab('Cantidad de actos')+
ylab('Costos')
```

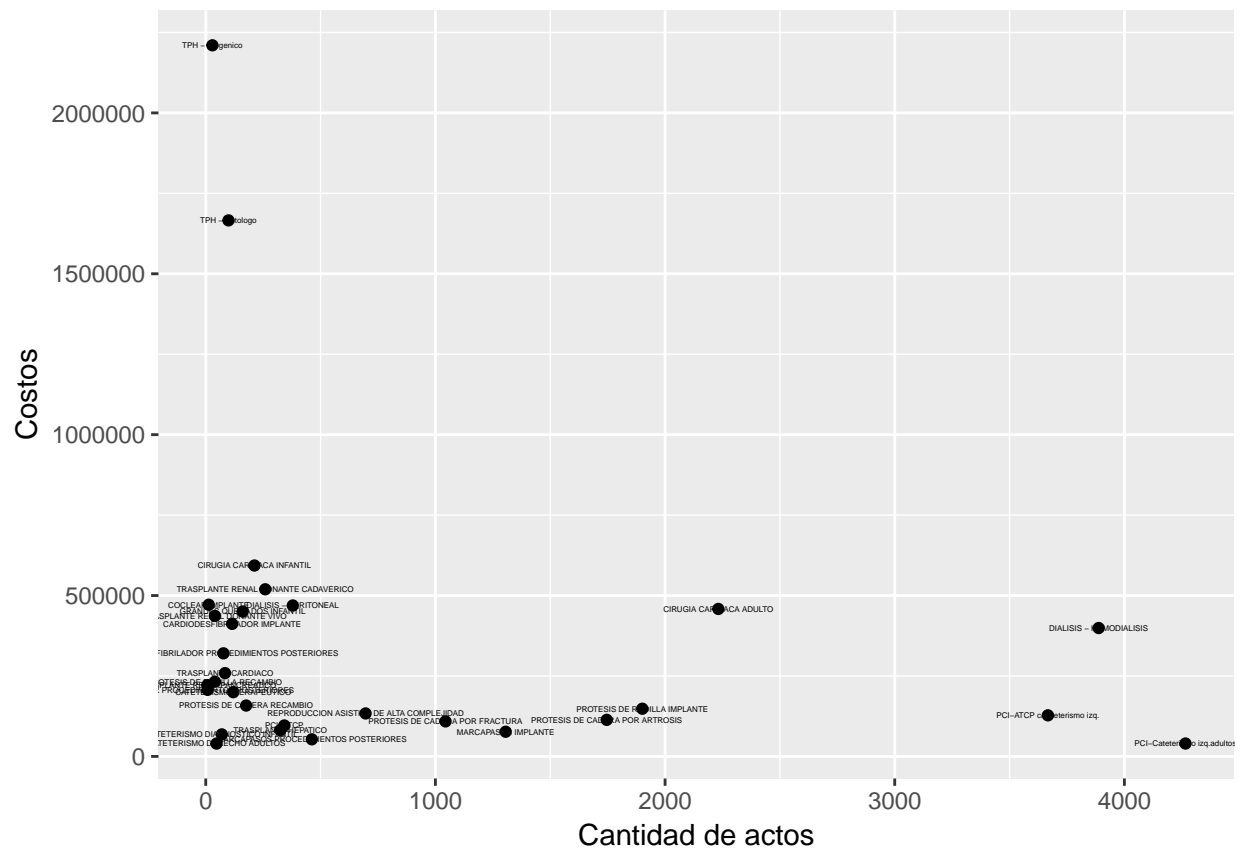


Figure 8: Gráfico de dispersión por cantidad y Costo

No, las más comunes son relativamente baratas.