

Segunda tarea

Fernando Massa

28 de septiembre de 2018

La dirección url donde se encuentran los archivos es la siguiente: <https://github.com/nando11235813/cursoIESTA>.

1 Ejercicio 1

```
# se cargan los datos
gastolink <- 'https://catalogodatos.gub.uy/dataset/96e636e5-4f78-49a7-8e14-60e90173a0c0/resource/936a'

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

Usando las funciones de la librería `dplyr` respondé:

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

```
cual<-gastos %>%
  filter(Prestador_departamento=='MONTEVIDEO') %>%
  group_by(Prestacion) %>%
  summarise(pres = n()) %>%
  filter(pres==max(pres)) %>%
  select(Prestacion)
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.2
```

La prestación con el mayor número de actos médicos en Montevideo es PCI-Cateterismo izq.adultos.

- 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?

```
dpto<-gastos %>%
  group_by(Departamento_residencia) %>%
  summarise(act_med = n()) %>%
  filter(act_med==min(act_med)) %>%
  select(Departamento_residencia)
```

El departamento de residencia con el menor número de actos médicos es FLORES.

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

```
cuantos<-gastos %>%
  group_by(Prestador_tipo) %>%
  summarise(act_med = n()) %>%
  filter(Prestador_tipo %in% c('ASSE', 'IAMC'))
```

ASSE prestó 7428 mientras que IAMC prestó 15550.

- 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?

```
npac<- gastos %>%
  group_by(Paciente) %>%
  summarise(ene = n()) %>%
  nrow()
```

El total de pacientes fue de 19911

- e. 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`).

```
pucha<-gastos %>%
  group_by(Paciente) %>%
  mutate(gasto_total = sum(Importe),
         visitas = n()) %>%
  select(-c(Prestacion,Importe)) %>%
  summarise_all(first)
# en este natalia me ayudó
```

2 Ejercicio 2

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

```
gastos %>%
  group_by(Departamento_residencia) %>%
  summarise(frec= n()/nrow(gastos)) %>%
  ggplot(aes(x=fct_reorder(Departamento_residencia,frec,sum),y=frec)) +
  geom_point() + coord_flip() + labs(x="Departamento de residencia",y="Proporción de actos médicos")
```

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

```
gastos %>%
  filter(Prestador_departamento=='MONTEVIDEO') %>%
  ggplot(aes(x=fct_infreq(Prestacion))) +
  geom_bar() + labs(x='',y="Total de actos médicos") +
  theme(axis.text.x=element_text(angle=90,hjust=0.5,size=5))
```

- c. Replique el siguiente gráfico (Figura ??) 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`

```
gastos %>%
  filter(Departamento_residencia=='MONTEVIDEO') %>%
  mutate(asse = ifelse(Prestador_tipo=='ASSE',1,0)) %>%
```

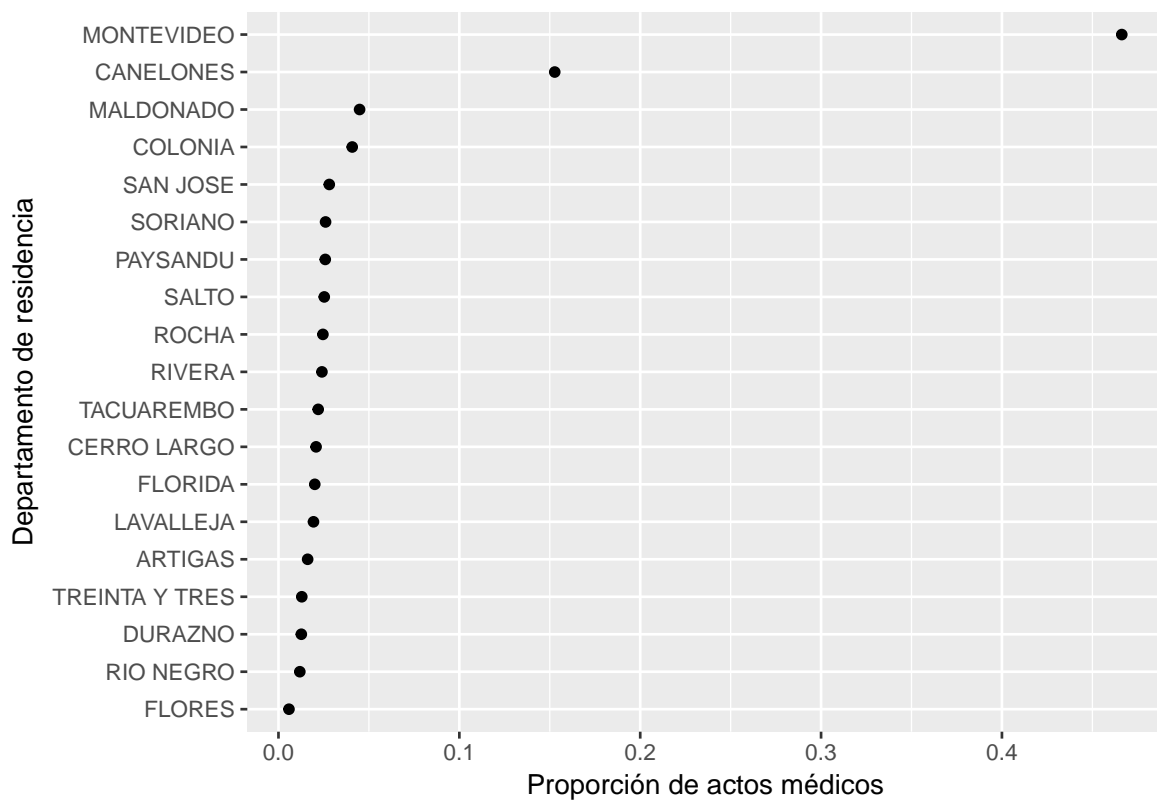


Figure 1: Proporción de actos médicos según departamento de residencia

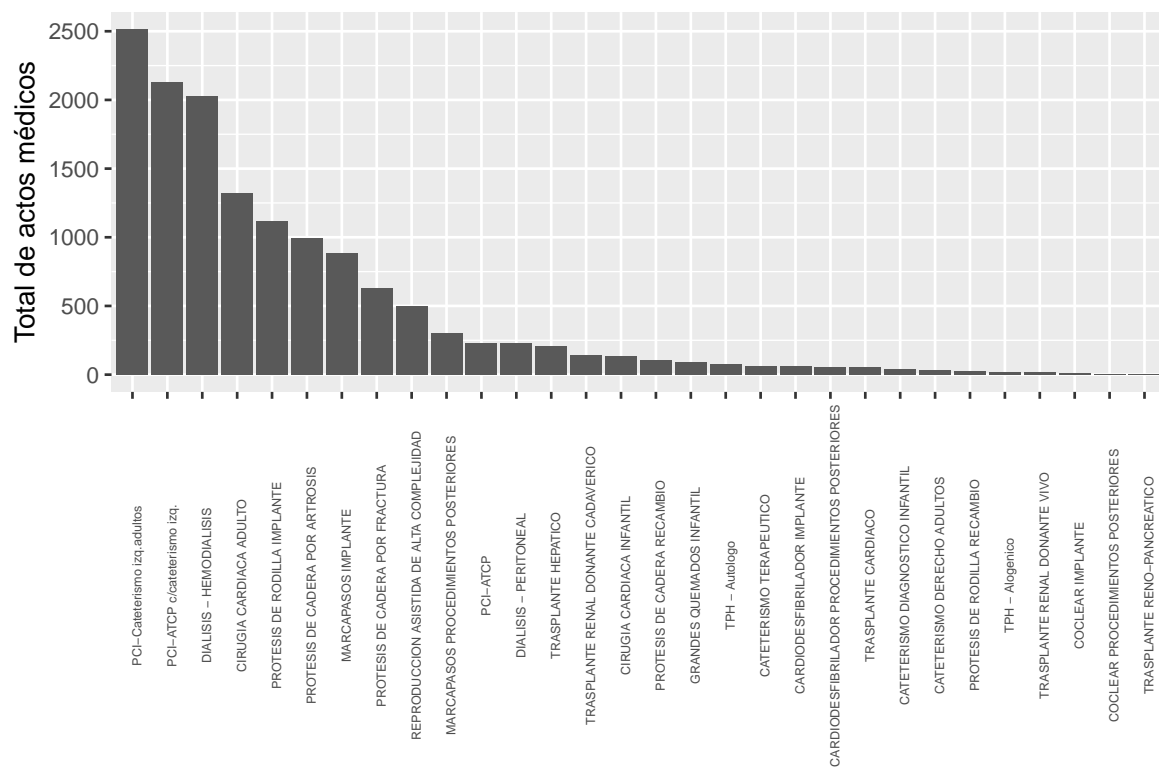
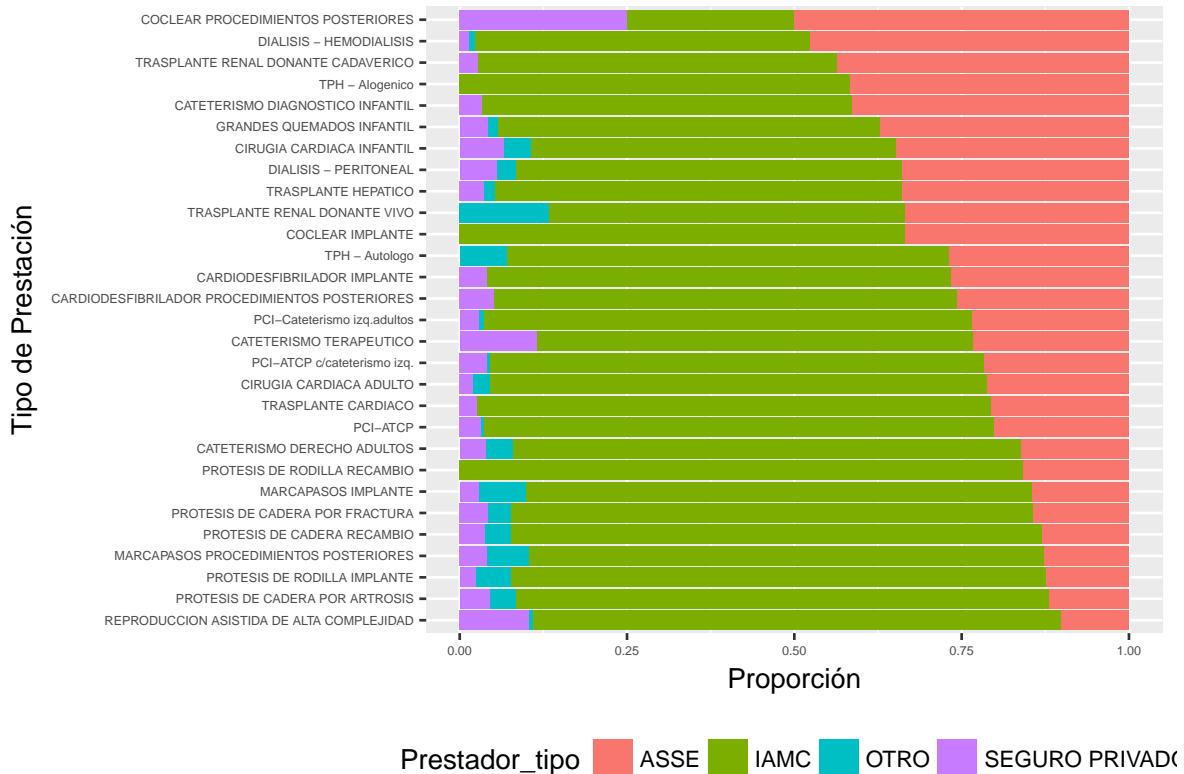


Figure 2: Gráfico de barras de la cantidad de actos m?dicos en Montevideo por tipo de Prestación

```
ggplot(aes(x=fct_reorder(Prestacion,asse,mean),fill=Prestador_tipo)) +
  geom_bar(position="fill") + labs(x="Tipo de Prestación", y="Proporción") +
  coord_flip() + theme(axis.text=element_text(size=5), legend.position="bottom")
```

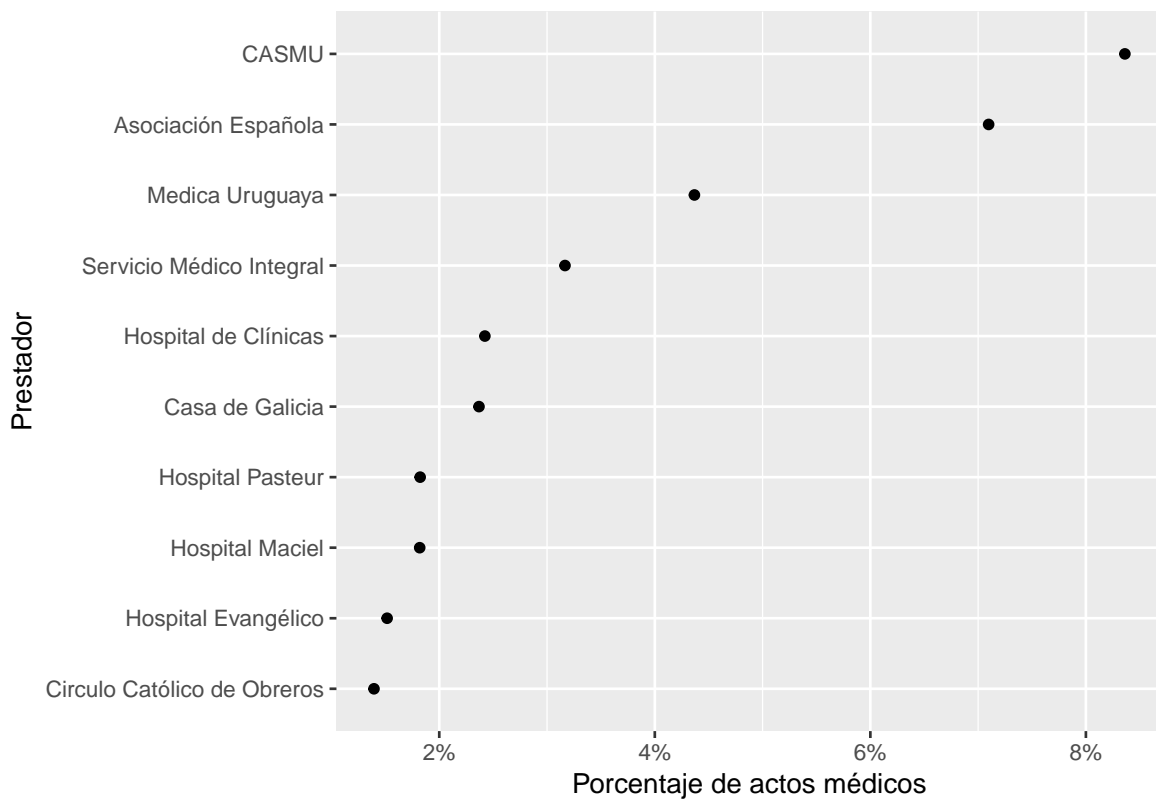


3 Ejercicio 3

- 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.

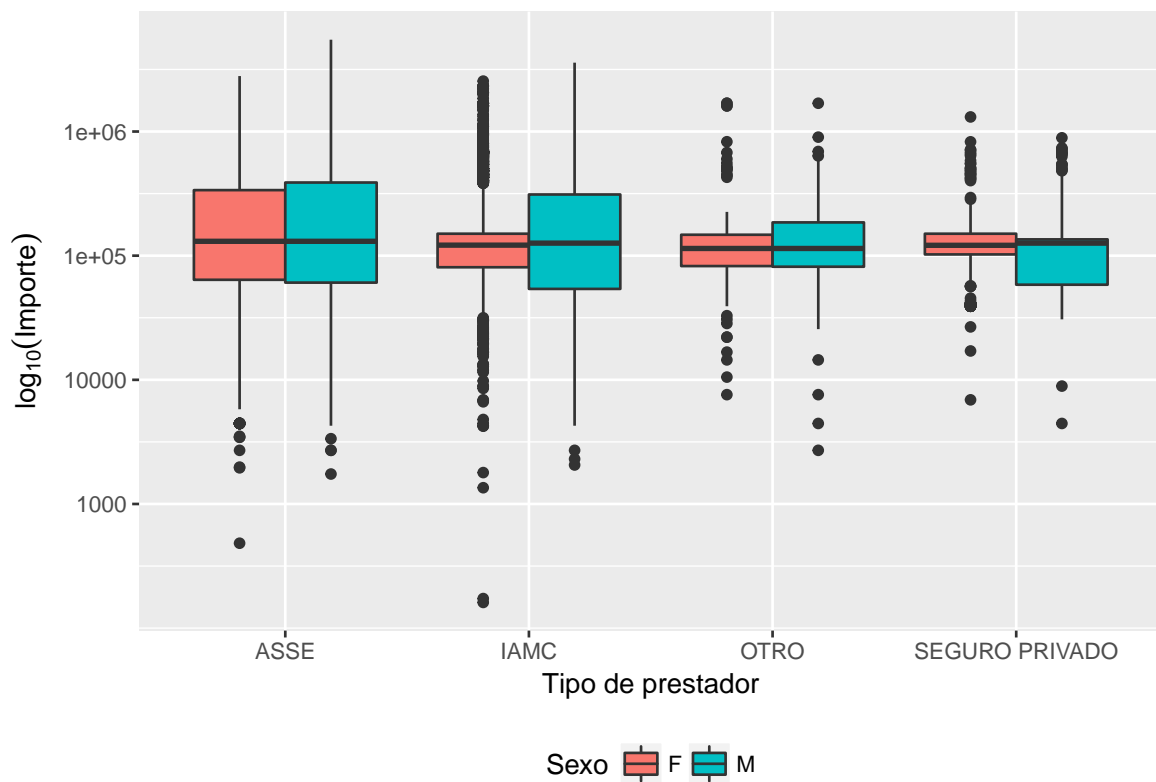
```
gastos %>%
  filter(Departamento_residencia=='MONTEVIDEO') %>%
  group_by(Prestador) %>%
  summarise(frec= n()/nrow(gastos)) %>%
  arrange(desc(frec)) %>%
  slice(1:10) %>%
  ggplot(aes(x=fct_reorder(Prestador,frec,sum),y=frec)) +
  geom_point() + coord_flip() + labs(x="Prestador",y="Porcentaje de actos médicos") +
  scale_y_continuous(breaks=seq(2,8,2)/100,labels=paste(seq(2,8,2),'%',sep=''))
```



En la Figura ?? se puede observar que el CASMU y la Asociación Española son quienes proveen la mayor proporción de actos médicos en Montevideo.

- 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.

```
gastos %>%
  ggplot(aes(x=Prestador_tipo, y=Importe, fill=Sexo)) +
  geom_boxplot() + scale_x_discrete(name="Tipo de prestador") +
  scale_y_continuous(name=expression(log[10](Importe)), trans="log10", breaks=10^seq(3,6), labels=10^s
```



En la Figura ?? se observa que la distribución del importe no parece diferir según sexo.

- 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 cómo esta varía según el sexo del paciente y el tipo de prestador.

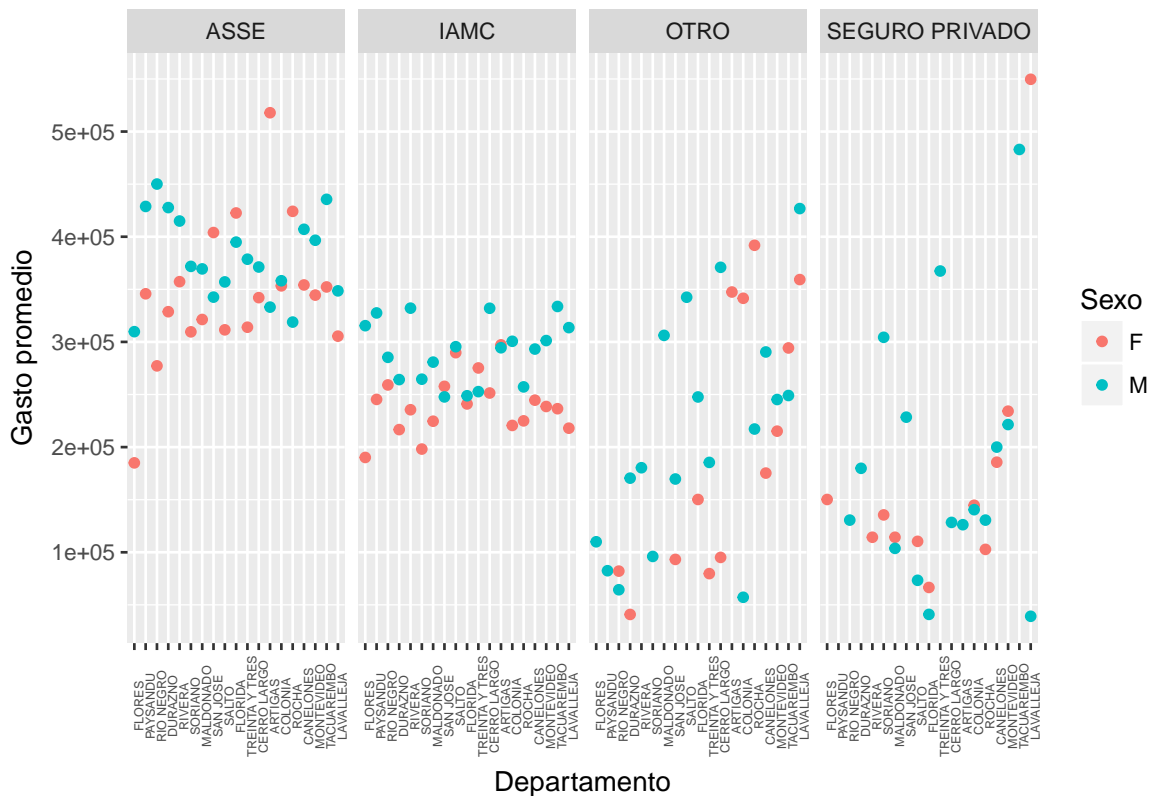
```
gastos %>%
  ggplot(aes(x=Edad_años,y=Importe,colour=Sexo)) +
  geom_point() + scale_x_continuous(name="Edad (años)") +
  scale_y_continuous(name=expression(log[10](Importe)),trans="log10",breaks=10^seq(3,6),labels=10^s
```



La figura ?? sugiere que en ninguno de los prestadores se da algún tipo de relación entre el monto del importe y la edad del paciente.

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

```
gastos %>%
  group_by(Paciente) %>%
  mutate(plata=sum(Importe)) %>%
  group_by(Departamento_residencia,Prestador_tipo,Sexo) %>%
  summarise(g_prom=mean(plata)) %>%
  ggplot(aes(x=fct_reorder(Departamento_residencia,g_prom,sum),y=g_prom,colour=Sexo)) +
    geom_point() + scale_x_discrete(name="Departamento") +
    scale_y_continuous(name="Gasto promedio") + facet_grid(.~Prestador_tipo) +
    theme(axis.text.x=element_text(angle=90,hjust=0.5,size=5))
```

La figura 3 sugiere que en IAMC, en la mayoría de los departamentos, los hombres gastan mas que la mujeres.

- 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(plata=mean(Importe)/1e6, cuantas=n()) %>%
  ggplot(aes(x=fct_reorder(Prestacion,cuantas,sum),y=plata)) +
    geom_point() + scale_x_discrete(name="Prestacion") +
    scale_y_continuous(name="Gasto promedio") +
    theme(axis.text.x=element_text(angle=90,hjust=0.5,size=5))
```

Pese a que la figura 3 presenta a las prestaciones TPH-alógeno y TPH-autólogo como las mas caras, siendo relativamente frecuentes, no se observa una relación tan clara en el resto de las prestaciones.

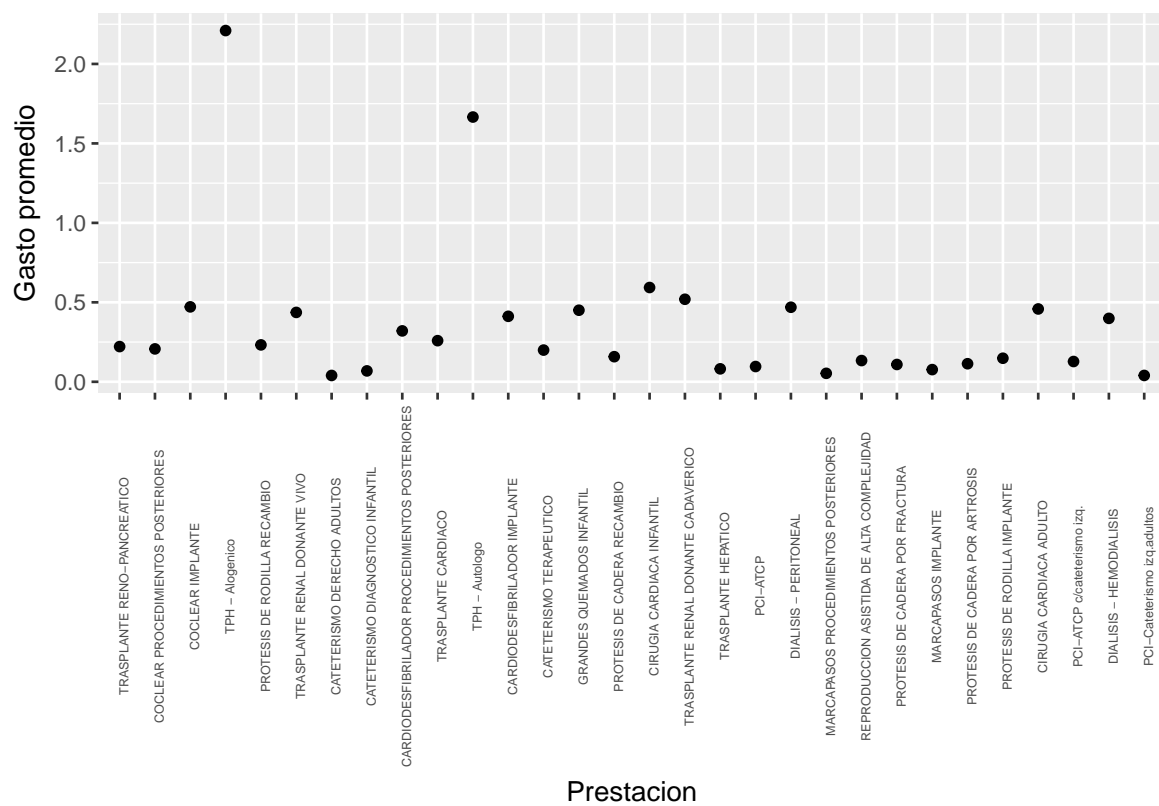


Figure 3: Importe promedio de las prestaciones (en millones)