

# Tarea 2

*Daniel Czarniewicz*

*Octubre 2018*

Se puede acceder al repositorio donde se encuentra el código haciendo click [aquí](https://catalogodatos.gub.uy/dataset/96e636e5-4f78-49a7-8e14-60e90173a0c0/resource/936ac9)

```
gastolink <- 'https://catalogodatos.gub.uy/dataset/96e636e5-4f78-49a7-8e14-60e90173a0c0/resource/936ac9'
gastos <- read.csv(gastolink, header=TRUE, dec=",", encoding="latin1") %>% as_tibble()
```

## Ejercicio 1

Usando las funciones de la librería `dplyr` respondé:

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

```
gastos %>%
  filter(Departamento_residencia == "MONTEVIDEO") %>%
  group_by(Prestacion) %>%
  tally() %>%
  arrange(desc(n)) %$%
  cat(paste("La prestación con mayor cantidad de actos médicos en Montevideo es",
            as.character(unclass(.[[1,1]] [[1]])))
```

La prestación con mayor cantidad 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?

```
gastos %>%
  group_by(Departamento_residencia) %>%
  tally() %>%
  arrange(n) %$%
  cat(paste("El departamento con menor cantidad de actos médicos es",
            str_to_title(as.character(unclass(.[[1,1]] [[1]]))))
```

El departamento con menor cantidad de actos médicos es Flores

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

```
gastos %>%
  filter(Prestador_tipo %in% c("ASSE", "IAMC")) %>%
  tally() %$%
  cat(paste("La cantidad de actos médicos prestados por ASSE o IAMC es de",
            as.numeric(.)))
```

La cantidad de actos médicos prestados por ASSE o IAMC es de 22978

- 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?

```
gastos %>%
  group_by(Paciente) %>%
  tally() %$%
  cat(paste("La base contiene", dim(gastos)[1], "observaciones correspondientes a",
            dim(.)[1], "pacientes"))
```

La base contiene 23811 observaciones correspondientes a 19911 pacientes

- 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 ).

```
gastos %>%
  group_by(Paciente) %>%
  summarise(Gasto = sum(Importe), Visitas = n()) %>%
  right_join(gastos, by="Paciente") %>%
  dplyr::select(-Prestacion, -Importe)
```

# A tibble: 23,811 x 9

	Paciente	Gasto	Visitas	Edad_años	Sexo	Departamento_re~	Prestador
	<int>	<dbl>	<int>	<int>	<fct>	<fct>	<fct>
1	48572	4.24e5	1	34	M	ROCHA	Centro D~
2	56004	4.06e5	1	77	M	CANELONES	COMECA I~
3	67813	4.41e5	2	60	M	MONTEVIDEO	Hospital~
4	78014	5.35e5	2	66	M	MONTEVIDEO	Servicio~
5	80694	4.60e5	2	54	F	MONTEVIDEO	Seguro A~
6	82954	4.65e5	2	69	M	MALDONADO	AMECOM I~
7	84293	6.22e5	3	77	M	FLORIDA	Centro D~
8	97516	5.31e5	2	79	M	MONTEVIDEO	Hospital~
9	98726	1.11e6	3	75	M	COLONIA	CAMOC IA~
10	111481	4.02e5	1	73	M	CANELONES	CASMU

# ... with 23,801 more rows, and 2 more variables: Prestador\_tipo <fct>,  
# Prestador\_departamento <fct>

## Ejercicio 2

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

```
gastos %>%
  group_by(Departamento_residencia) %>%
  tally() %>%
  mutate(Prop = n/sum(n)) %>%
  ggplot() +
  geom_point(aes(fct_reorder(Departamento_residencia, Prop), Prop)) +
  coord_flip() +
  labs(x="Departamento de residencia", y="Proporción de actos médicos") +
  theme(aspect.ratio=1)
```

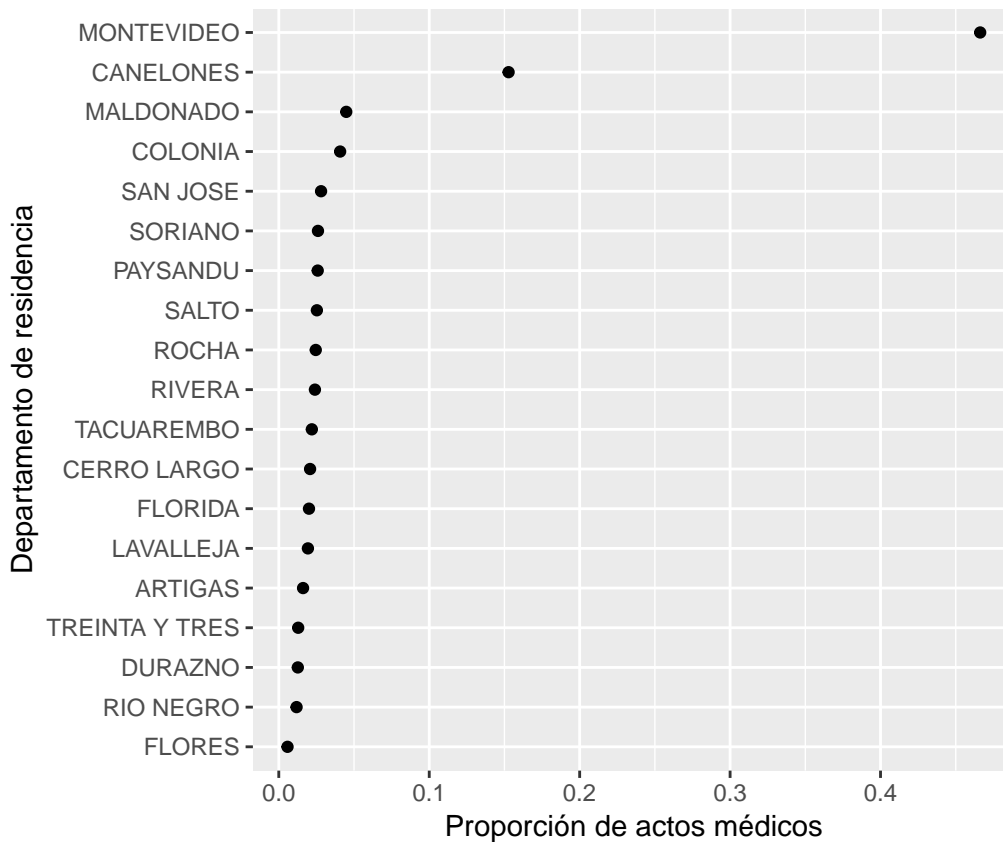


Figure 1: Proporción de actos médicos según departamento de residencia

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

```
gastos %>%
  filter(Departamento_residencia == "MONTEVIDEO") %>%
  group_by(Prestacion) %>%
  tally() %>%
  ggplot() +
  geom_bar(aes(x=fct_reorder(Prestacion, n, .desc=TRUE), y=n), stat="identity") +
  labs(y="Total de actos médicos", x=NULL) +
  theme(axis.text.x=element_text(angle=90, size=5))
```

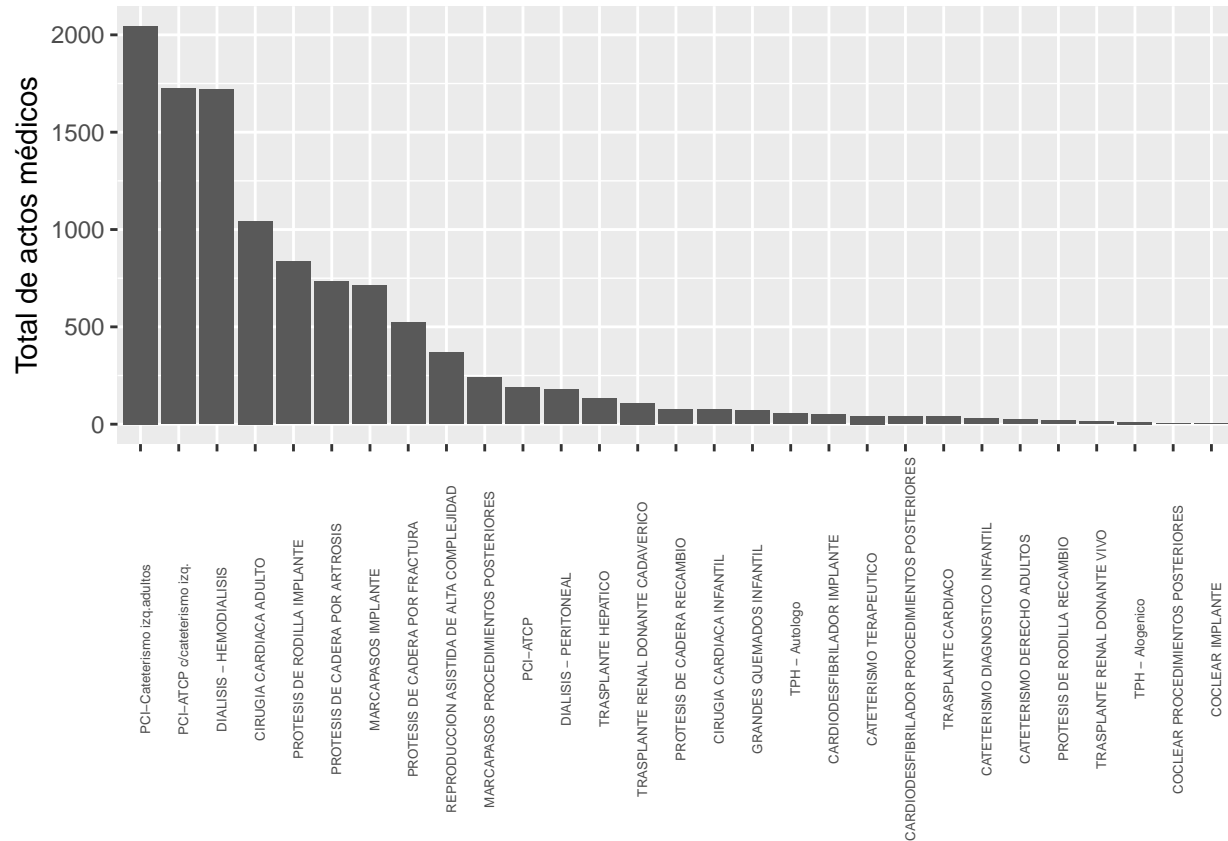


Figure 2: Gráfico de barras de la cantidad de actos médicos en Montevideo por tipo de Prestación

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

*# sin filtro para mudeo*

`gastos %>%`

`mutate(esASSE = if_else(Prestador_tipo == "ASSE", 1, 0)) %>%`

`group_by(Prestacion) %>%`

`summarise(cantASSE = sum(esASSE), cant=n()) %>%`

`mutate(propASSE = cantASSE / cant) %>%`

`dplyr::select(Prestacion, propASSE) %>%`

`right_join(gastos, by="Prestacion") %>%`

`ggplot() +`

`geom_bar(aes(fct_reorder(Prestacion, propASSE), fill=Prestador_tipo), position="fill") +`

`coord_flip() +`

`labs(x="Tipo de presatición", y="Proporción") +`

`theme(axis.text.y=element_text(size=5, face="bold"),`

`legend.position="bottom")`

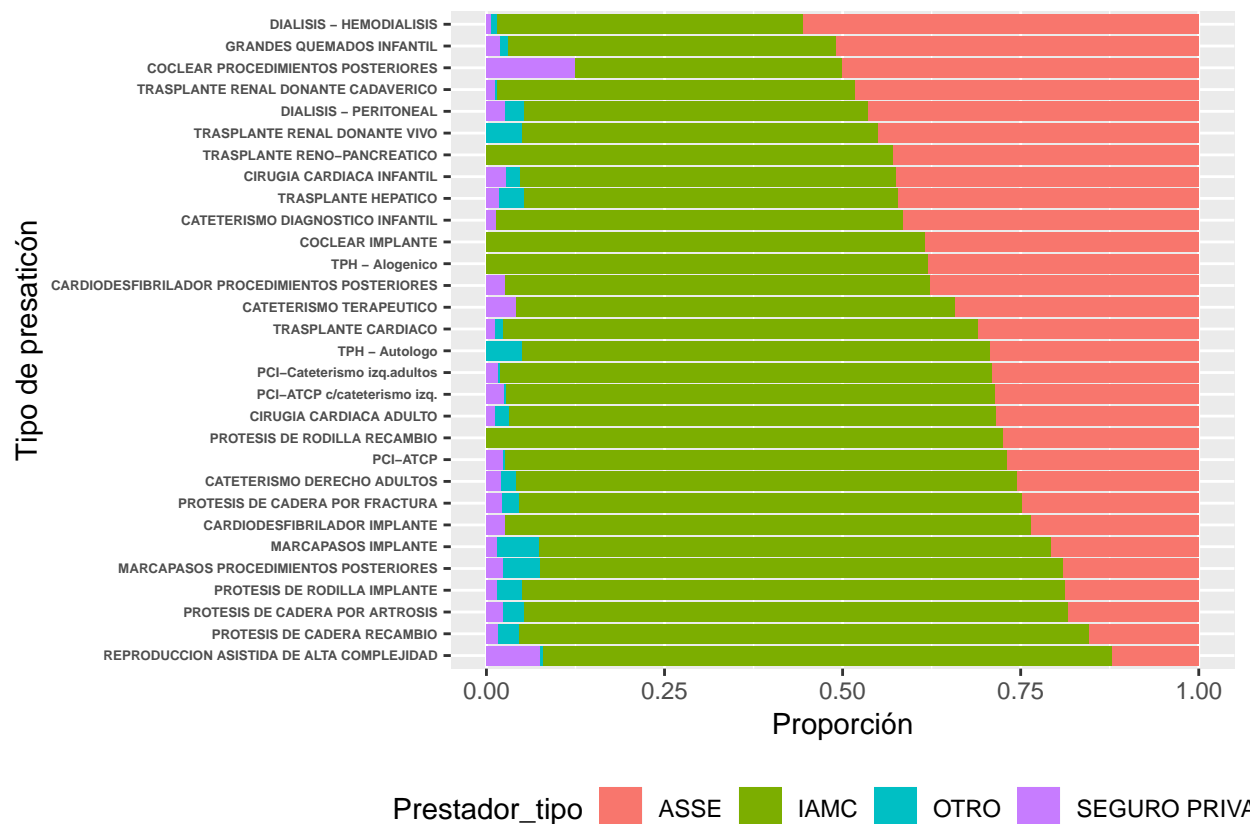


Figure 3: Gráfico de barras apiladas al 100 de la cantidad de actos médicos en Montevideo por tipo de Prestacion

### 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 %>%
  dplyr::filter(Departamento_residencia == "MONTEVIDEO") %>%
  group_by(Prestador) %>%
  tally() %>%
  mutate(prop_presta = n/sum(n)) %>%
  mutate(ranking = rank(prop_presta, ties.method="max")) %>%
  dplyr::arrange(desc(ranking)) %>%
  head(n=30L) %>%
  droplevels() %>%
  ggplot() +
  geom_bar(aes(fct_reorder(Prestador, prop_presta, .desc=FALSE), prop_presta),
    stat="identity") +
  coord_flip() +
  labs(x="Prestador", y="Proporción de prestaciones") +
  ggthemes::theme_economist(horizontal=FALSE) +
  theme(axis.text.y=element_text(hjust=1, size=6),
    axis.title=element_text(face="bold"))
```

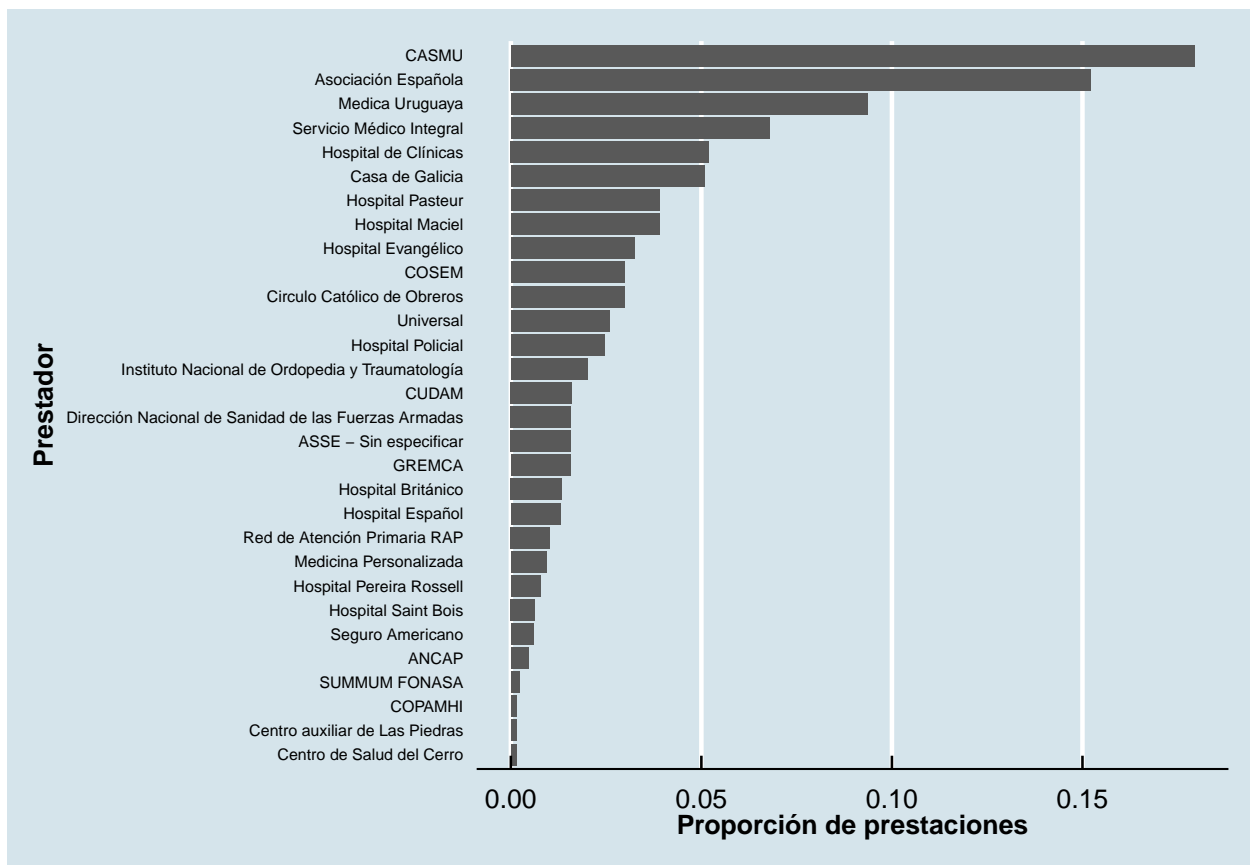


Figure 4: Gráfico de barras de la proporción de actos médicos en Montevideo por prestador (principales 30)

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() +
  geom_boxplot(aes(Prestador_tipo, log(Importe), fill=Prestador_tipo),
               show.legend=FALSE, outlier.alpha=.1) +
  facet_wrap(~Sexo) +
  scale_x_discrete(labels = function(x) str_wrap(x, width=10)) +
  labs(x="Tipo de prestador", y="Importe (en logaritmos)") +
  ggthemes::theme_economist() +
  theme(axis.title=element_text(face="bold"),
        axis.text.x=element_text(vjust=0.5))
```

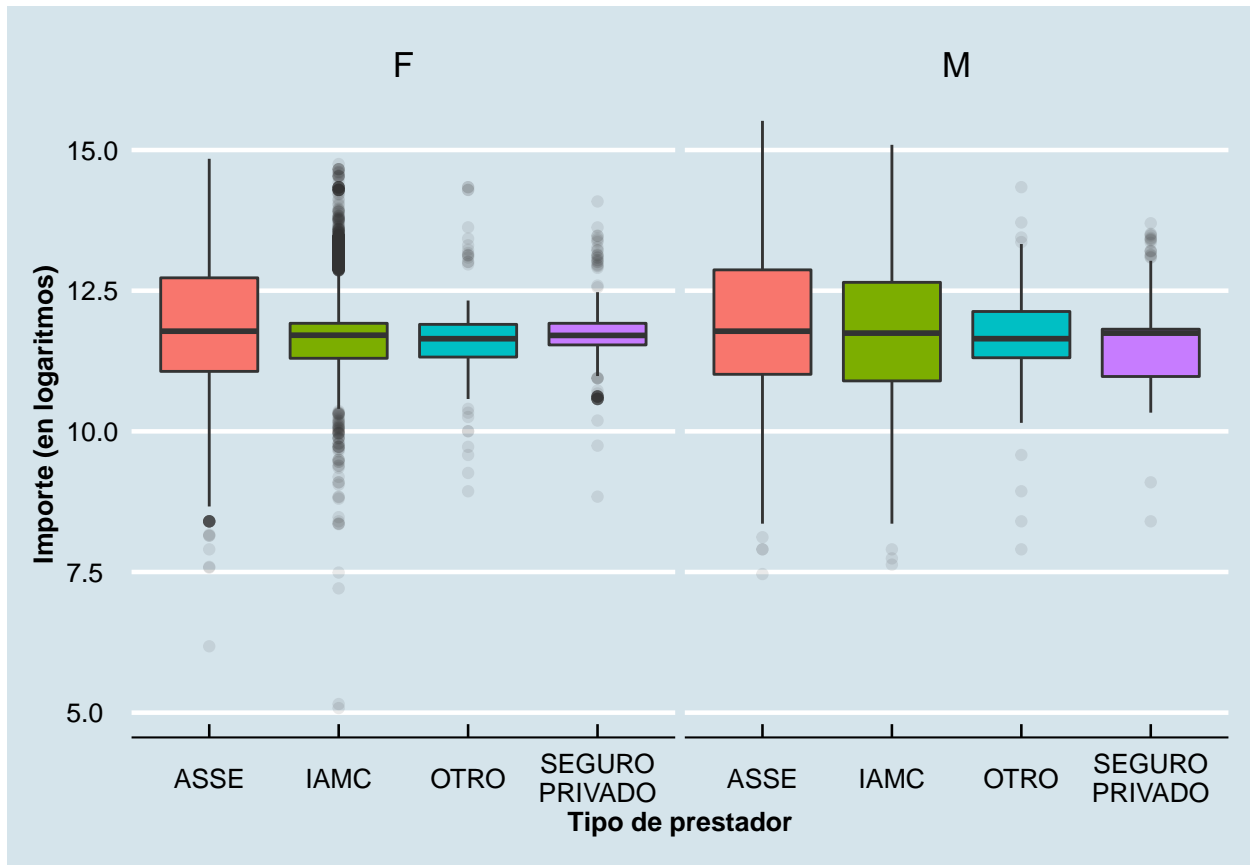


Figure 5: Boxplot del logaritmo del importe por prestador según sexo

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

```
gastos %>%
  ggplot() +
  geom_point(aes(x=Edad_años, y=log(Importe)), alpha=0.2) +
  facet_grid(Sexo ~ Prestador_tipo) +
  labs(x="Edad del paciente (en años)", y="Importe (en logaritmos)") +
  ggthemes::theme_economist() +
  theme(axis.title=element_text(face="bold"),
        strip.text=element_text(size=8),
        strip.text.y=element_text(angle=0),
        aspect.ratio=1)
```

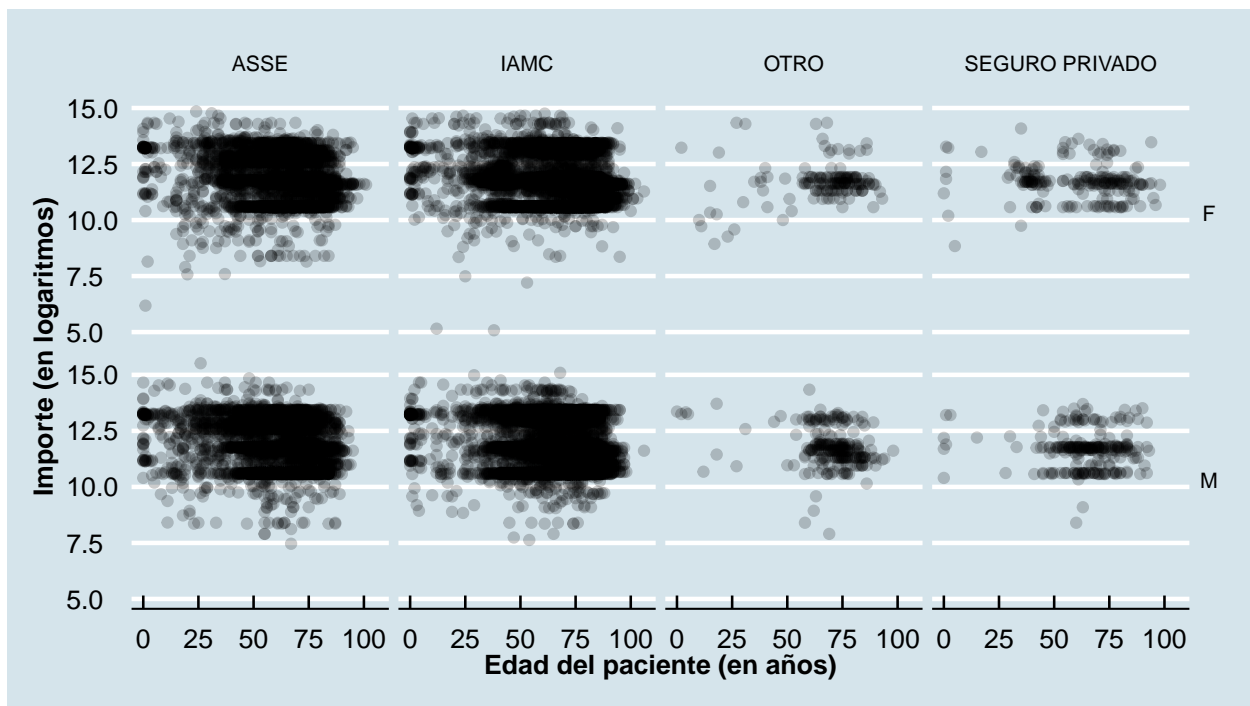


Figure 6: Scatter plot de la edad del paciente (en años) y el importe (en logaritmos) por tipo de prestador según sexo



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

```
gastos %>%
  group_by(Paciente, Departamento_residencia, Prestador_tipo, Sexo) %>%
  summarize(gto_total = sum(Importe)) %>%
  group_by(Departamento_residencia, Prestador_tipo, Sexo) %>%
  summarize(gto_prom = mean(gto_total)) %>%
  ggplot() +
  geom_bar(aes(Departamento_residencia, gto_prom/1000), stat="identity") +
  facet_grid(Sexo ~ Prestador_tipo) +
  labs(x="Departamento de residencia", y="Gasto promedio (en miles)") +
  ggthemes::theme_economist() +
  theme(axis.text.x=element_text(angle=90, size=4, hjust=1),
        axis.title=element_text(face="bold"),
        strip.text=element_text(size=8),
        strip.text.y=element_text(angle=0))
```

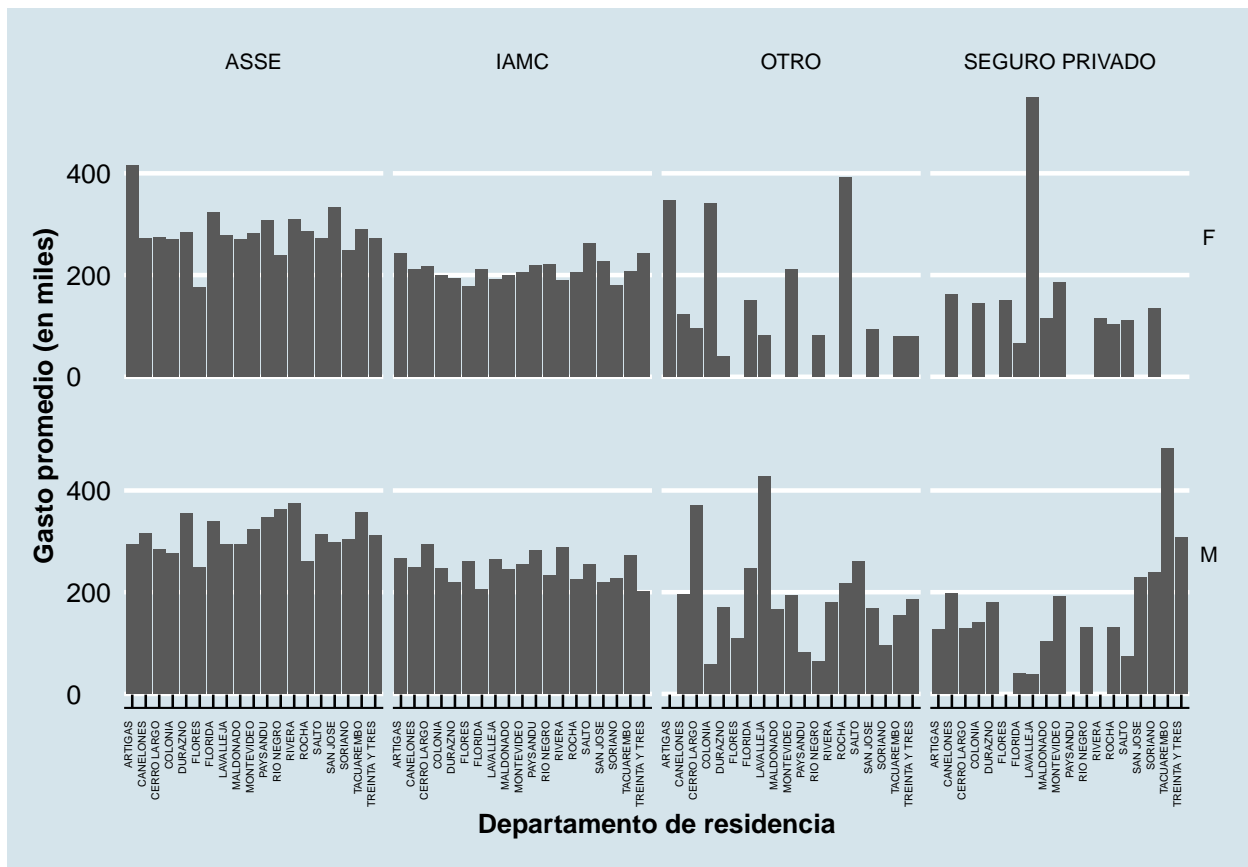


Figure 7: Gasto promedio por persona (en miles) en cada departamento, por tipo de prestador según sexo

- 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(total.presta = n(), total.gto = sum(Importe)) %>%
  mutate(prop.presta = total.presta / sum(total.presta),
         prop.gto = total.gto / sum(total.gto)) %>%
  dplyr::select(-total.presta, -total.gto) %>%
  ggplot() +
  geom_point(aes(x=prop.presta, y=prop.gto), alpha=0.5) +
  labs(x="Proporción de la prestación", y="Proporción del gasto") +
  ggthemes::theme_economist() +
  theme(axis.title=element_text(face="bold"),
        aspect.ratio=1)
```

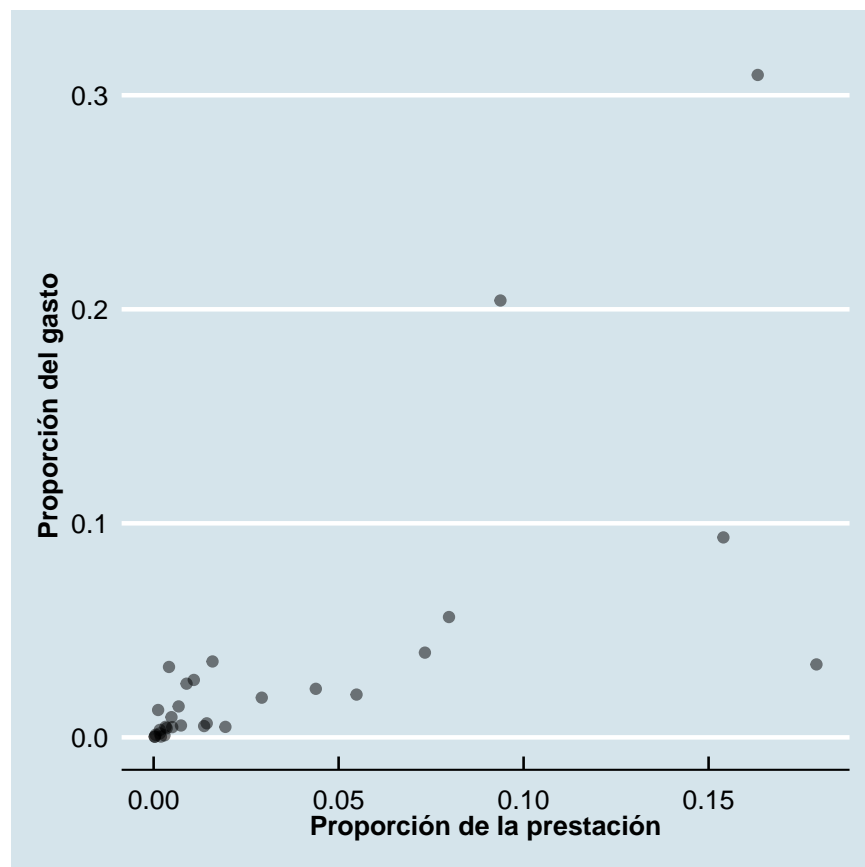


Figure 8: Scatter plot de la proporción de la prestación en total de las prestaciones realizadas y la proporción del gasto en el total del gasto realizado para cada prestación