

# Tarea individual 2

*Entregar el Lunes 16 de Mayo*

*4/5/2018*

title: "Tarea individual 2" author: "Entregar el Lunes 16 de Mayo" date: "4/5/2018" output: pdf\_document

---

## Entrega

La tarea debe ser realizada en RMarkdown en un repositorio de GitHub llamado "Tarea 2". La tarea es individual por lo que cada uno tiene que escribir su propia versión de la misma. El repositorio debe contener únicamente el archivo .Rmd con la solución de la tarea.

## Ejercicio 1

Los datos que vamos a utilizar en este ejercicio están disponibles en el catálogo de datos abiertos uruguay <https://catalogodatos.gub.uy>. Los datos que seleccioné son sobre las emisiones de dióxido de carbono (CO2) correspondientes a las actividades de quema de los combustibles en las industrias de la energía y los sectores de consumo. Se incluyen también emisiones de CO2 provenientes de la quema de biomasa y de bunkers internacionales, las cuales se presentan como partidas informativas ya que no se consideran en los totales. En el siguiente link se encuentran los datos y los meta datos con información que describe la base de datos <https://catalogodatos.gub.uy/dataset/emisiones-de-co2-por-sector>.

Debe leer con cuidado la información en los meta datos para responder correctamente.

Los datos fueron reestructurados para simplificar la exploración, de la siguiente manera:

```
library(tidyverse)
dato_emision<- gather(dat, key = fuente, value = emisión, -Año)
```

Con estos datos responda las siguientes preguntas:

1. Usando las funciones de la librería `dplyr` obtenga qué fuentes tienen la emisión máxima. Recuerde que TOTAL debería ser excluido para esta respuesta así como los subtotales.

```
r library(tidyverse) library(dplyr) emisiones<- read.csv("emisiones2.csv", header =
TRUE, as.is = TRUE) new_data<- select(emisiones,-(S_C:TOTAL)) new_data<- select(new_data,-(I_E))
dato_emision<- gather(new_data, key = fuente, value = emisión, -Año) max_data<- dato_emision
%>% group_by(fuente) %>% summarise(max=max(emisión, na.rm = TRUE)) max_data<-arrange(max_data,
desc(max)) head(max_data,1)
```

```
## # A tibble: 1 x 2 ##   fuente   max ##   <chr>   <dbl> ## 1 Q_B      8831.
```

La quema de biomasa con fines energéticos (leña, residuos de biomasa y carbón vegetal para toda la serie) es la mayor fuente emisora de CO2.

2. ¿En qué año se dio la emisión máxima para la fuente que respondió en la pregunta anterior?

```
r max_year<- dato_emision %>% arrange(desc(emisión)) head(max_year,1)
##   Año fuente emisión ## 1 2016   Q_B  8831.1
```

3. Usando las funciones de la librería `dplyr` obtenga las 5 fuentes, sin incluir TOTAL ni subtotales, que tienen un valor medio de emisión a lo largo de todos los años más grandes.

```
mean_data<- dato_emision %>% group_by(fuente) %>% summarise(mean=mean(emisión, na.rm = TRUE))
mean_data<-arrange(mean_data, desc(mean))
head(mean_data, 5)
```

```
## # A tibble: 5 x 2
##   fuente mean
##   <chr> <dbl>
## 1 Q_B    3691.
## 2 T      2579.
## 3 BI     1125.
## 4 CE_SP   893.
## 5 I       674.
```

4. Usando ggplot2 realice un gráfico de las emisiones a lo largo de los años para cada fuente. Utilice dos elementos geométricos, puntos y líneas. Seleccione para dibujar solamente las 5 fuentes que a lo largo de los años tienen una emisión media mayor que el resto (respuesta de la pregunta 3). Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un **caption** en la figura con algún comentario de interés que describa el gráfico.

```
new_data<- select(new_data,-(CP:C_S_SP))
new_data<- select(emisiones,-(CP:C_S_SP))
new_data<- select(new_data,-(A_P_M:TOTAL))
dato_gg<- gather(new_data, key = fuente, value = emisiones, -Año)
dato_gg <- dato_gg %>% mutate(fuente = factor(fuente))
dato_gg$fuente2 <- factor(dato_gg$fuente, labels = c("Búncers internacionales", "Centrales eléctricas s
data_plot<-ggplot(dato_gg,aes(x= Año, y= emisiones) ) + geom_point(aes(x= Año, y= emisiones))+ geom_line
data_plot
```

Serie temporal de emisiones de CO2 segun mayores fuentes de emision

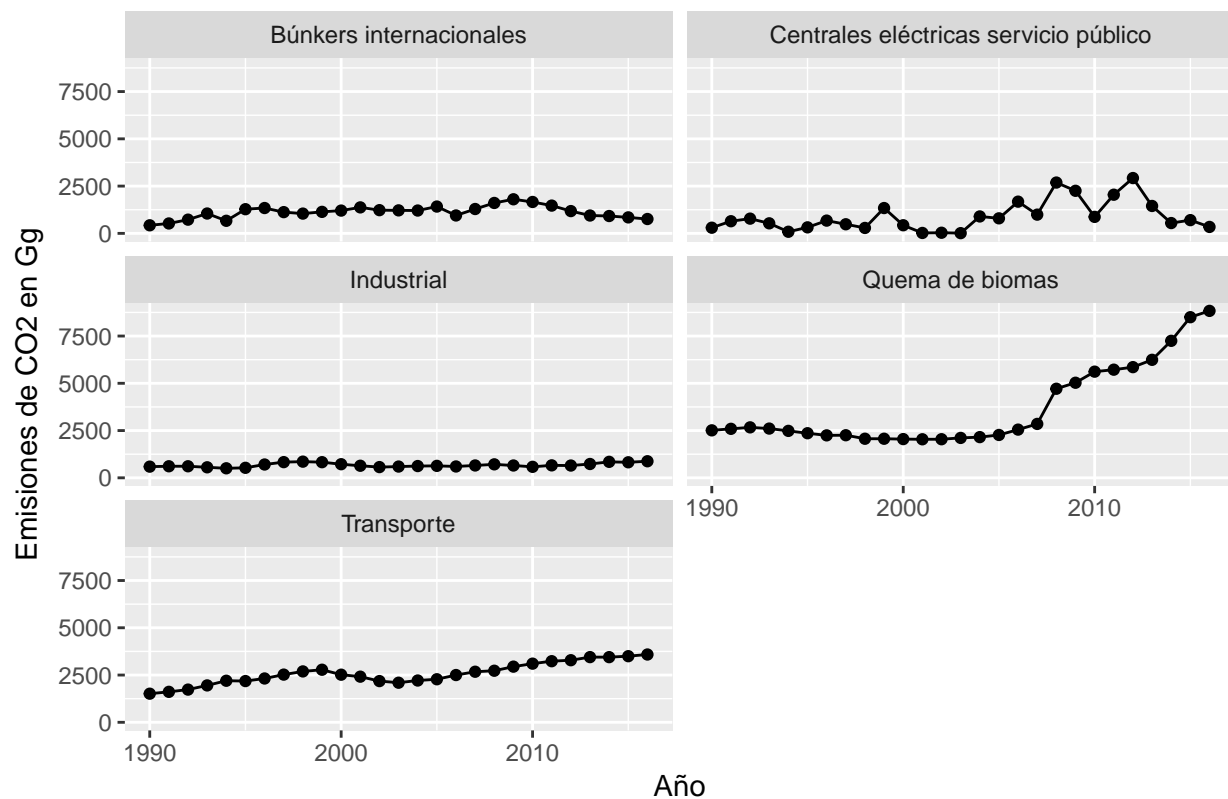


Figura 1: Gráfico de dispersión de las emisiones de CO2 en Gg contra el tiempo medido en años y dividido de acuerdo a las cinco fuentes de mayor emisión media de CO2 en el periodo 1990-2016. Se puede observar una asociación lineal, positiva y moderada entre la emisión de CO2 en Gg y el tiempo para las fuentes Bunkers internacionales, Industrial y Transporte, mientras que para la Quema de biomasa la relación es más exponencial.

5. Relpique el siguiente gráfico usando ggplot2. Incluir un **caption** en la figura con algún comentario de interés que describa el gráfico.

```
ggplot(dato_gg, aes(x = fct_reorder(fuente, emisiones, fun = mean, .desc = TRUE), y = emisiones)) + geom_l
```

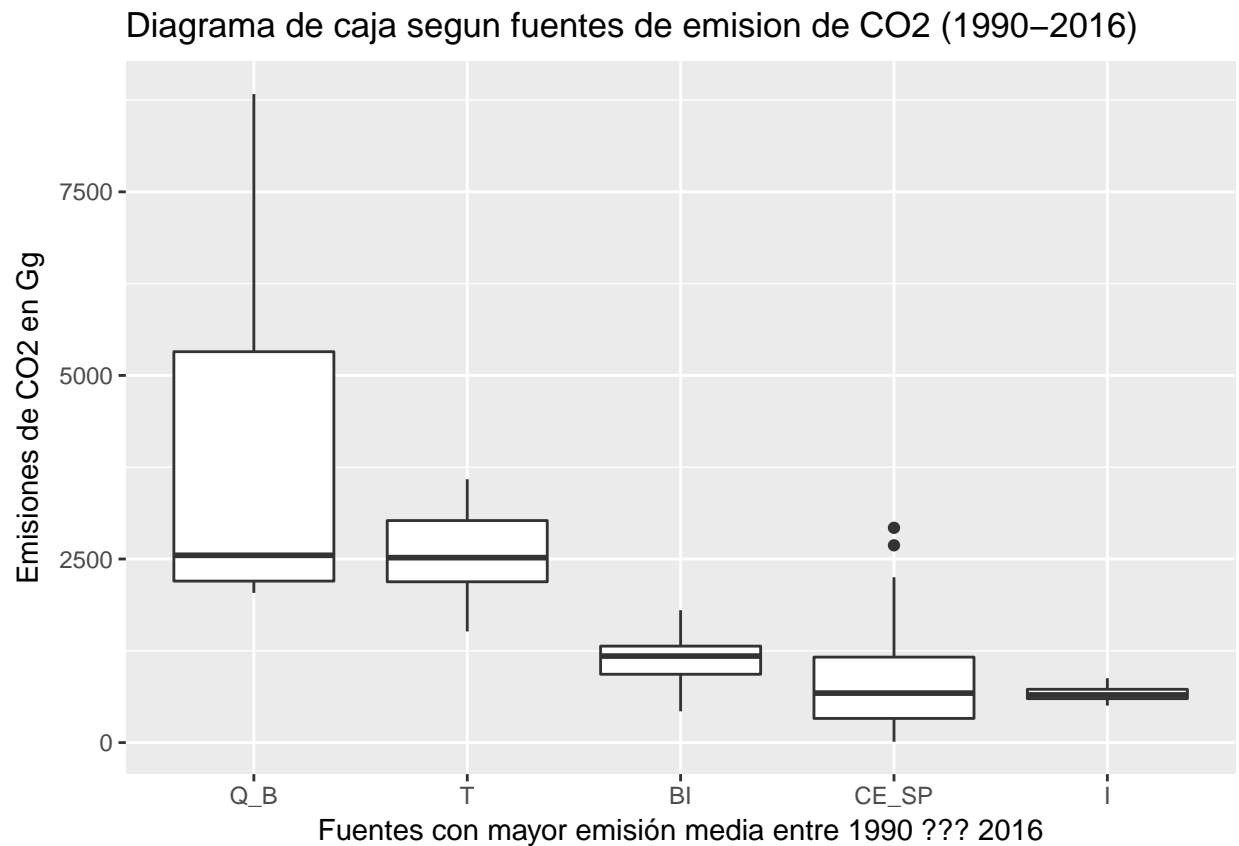


Figura 2: Gráfico de caja de las emisiones de CO2 en Gg de acuerdo a las cinco fuentes de mayor emisión media de CO2 en el periodo 1990-2016. Se puede observar algunos datos atípicos en las Centrales eléctricas servicio público, en la quema de biomasa podemos observar valores mucho más elevados, mientras que en las demás fuentes la distribución es más simétrica.

6. Usando la librería ggplot2 y ggpmisc replique el siguiente gráfico de las emisiones totales entre 1990 y 2016. Los puntos rojos indican los máximos locales o picos de emisión de CO2 en Gg. Use `library(help = ggpmisc)` para ver todas las funciones de la librería ggpmisc e identificar cual o cuales necesita para replicar el gráfico. Incluir un `caption` en la figura con algún comentario de interés que describa el gráfico.

```
library(ggpmisc)
```

```
## For news about 'ggpmisc', please, see http://www.r4photobiology.info/
```

```
## For on-line documentation see http://docs.r4photobiology.info/ggpmisc/
```

```
new_data<- select(emisiones,-(CE_SP:S_C))
```

```
new_data<- select(new_data,-(Q_B:BI))
```

```
dato_emision<- gather(new_data, key = fuente, value = emisión, -Año)
```

```
ggplot(dato_emision, aes(x=Año, y=emisión))+geom_point()+ geom_line()+ stat_peaks(geom = "text",colour = "red")
```

Serie temporal de emision total de CO2 (1990–2016)

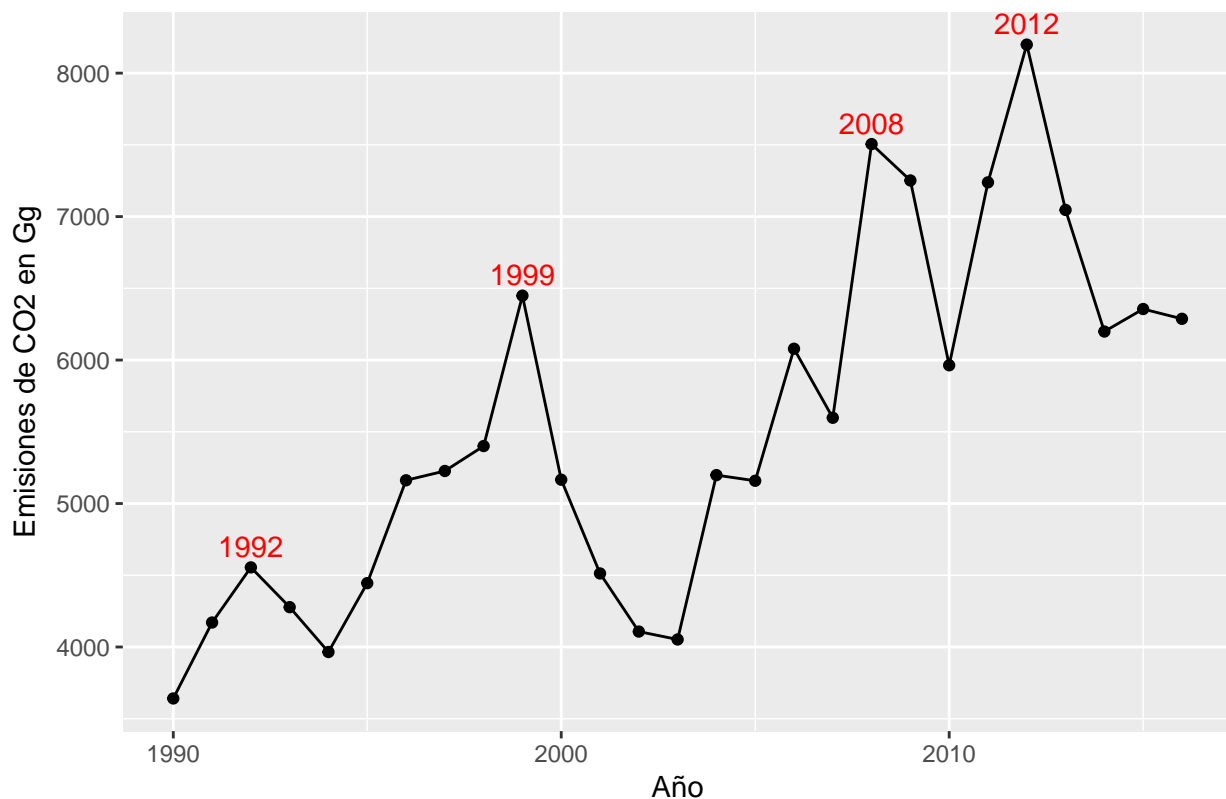


Figura 3: Gráfico de dispersión de las emisiones totales de CO2 en Gg contra el tiempo medido en años durante el periodo 1990-2016, en la cual se destacan los picos de emisión máxima en los años 1992,1999,2008 y 2012.Se puede observar un comportamiento oscilante y casi cíclico con respecto las emisiones totales de CO2.

## Ejercicio 2

Los datos que vamos a utilizar en este ejercicio están disponibles en el catálogo de datos abiertos uruguay <https://catalogodatos.gub.uy>.

Los datos que seleccioné son sobre los gastos realizados por actos médicos. Los datos y los metadatos se encuentran disponibles en:

[https://catalogodatos.gub.uy/dataset/gasto\\_am\\_2016\\_fondo-nacional-de-recursos/resource/936ac9e6-b0f6-424a-9b53-ee408a](https://catalogodatos.gub.uy/dataset/gasto_am_2016_fondo-nacional-de-recursos/resource/936ac9e6-b0f6-424a-9b53-ee408a)

Este ejercicio tiene como objetivo que realice un análisis exploratorio de datos utilizando todo lo aprendido en el curso. Debe contener al menos 5 preguntas orientadoras y visualizaciones apropiadas para responderlas. La exploración deberá contener las preguntas a responder sus respuestas con el correspondiente resumen de información o visualización. Incluya en su exploración el análisis de la variabilidad tanto de variables cuantitativas como cualitativas y covariaciones entre las mismas. Recuerde que en las visualizaciones, las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un **caption** en la figura con algún comentario de interés que describa el gráfico.

1. Se pretende realizar un gráfico que ilustre la relación entre edad y sexo de los pacientes durante el periodo 2016.

```
gastos<- read.csv("gastosmed.csv", header = TRUE, as.is = TRUE)
gastos <- gastos %>% mutate(Sexo = factor(Sexo))
gastos$Sexo2 <- factor(gastos$Sexo, labels = c( "Femenino","Masculino"))
gastos %>% ggplot(aes(x = Edad_años, fill= Sexo2)) + geom_bar() + theme(axis.text.x = element_text(ang
```

Gráfico de barras de edad de los pacientes por sexo

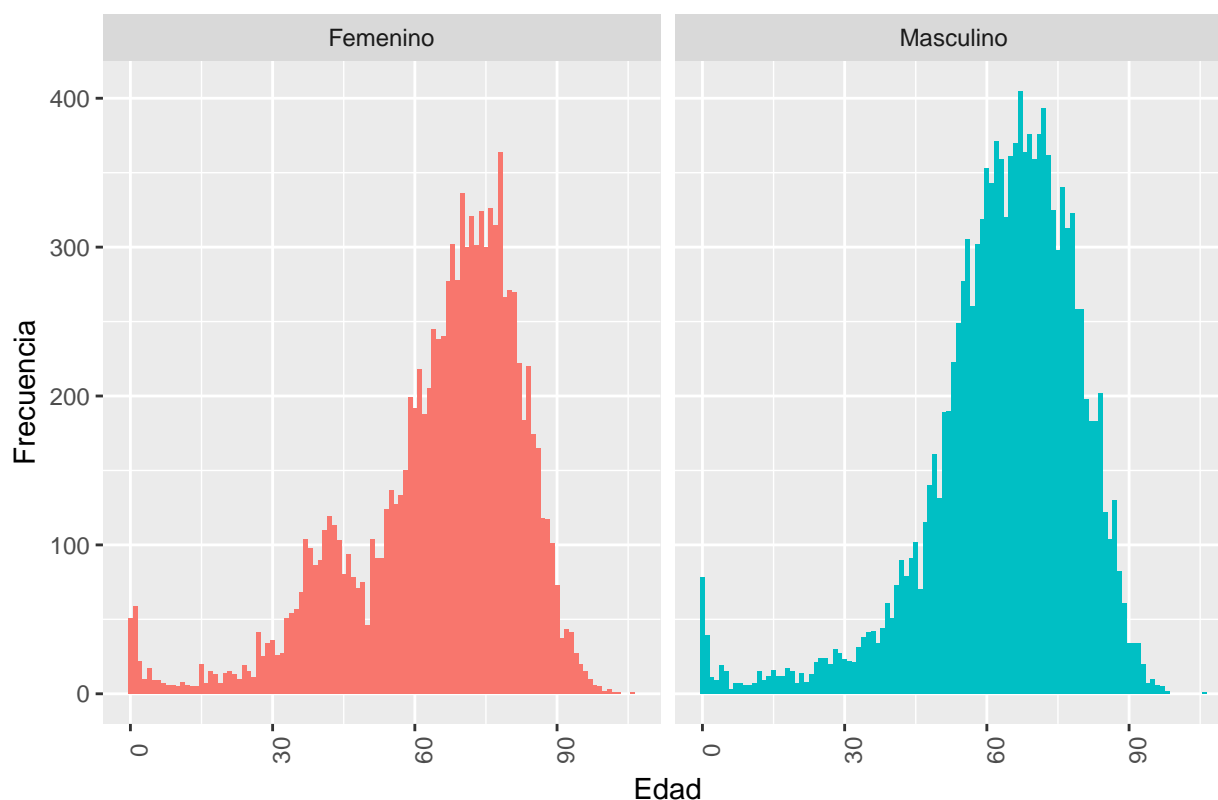


Figura 4. Gráfico de barras de la frecuencia de pacientes atendidos según edad y divididos por sexo. Ambos grupos presentan una distribución similar y casi normal, la frecuencia de mujeres de entre 30-35 años es mayor que la de los hombres al contrario de lo que ocurre en el grupo de entre 60-90 años en la que la cantidad de hombres es bastante mayor, este resultado se debe posiblemente a aspectos culturales.

2. Se desea definir los 3 servicios que en los cuales se gastó el mayor importe durante el periodo 2016.

```
new_data<- select(gastos,-(Paciente:Prestador_departamento))
max_data<- new_data %>% group_by(Prestacion) %>% summarise(sum=sum(Importe, na.rm = TRUE))
```

```
max_data<-arrange(max_data,desc(sum))
head(max_data,3)
```

```
## # A tibble: 3 x 2
##   Prestacion          sum
##   <chr>             <dbl>
## 1 DIALISIS - HEMODIALISIS 1551191180.
## 2 CIRUGIA CARDIACA ADULTO 1023054368.
## 3 PCI-ATCP c/cateterismo izq. 468258951.
```

3.Comparar el tipo de prestadores de salud de aquellos pacientes que acuden al Departamento de Montevideo para recibir atencion medica residiendo fuera del mismo.

```
gastos <- gastos %>% mutate( Prestador_tipo= factor(Prestador_tipo))
gastos$Prestador_tipo2 <- factor(gastos$Prestador_tipo, labels = c("Administración de los Servicios de Salud", "Instituciones de asistencia médica colectiva", "Seguro Pribado", "Otro"))
y <- filter(gastos, Departamento_residencia != "MONTEVIDEO")
x<- filter(y, Prestador_departamento== "MONTEVIDEO")
ggplot(x, aes(x= Prestador_tipo2, fill=Prestador_tipo))+ geom_bar()+ guides(fill=FALSE)+ labs( x= NULL, y= "Cantidad de pacientes")
```

Grafico de barras de tipo de prestador de salud

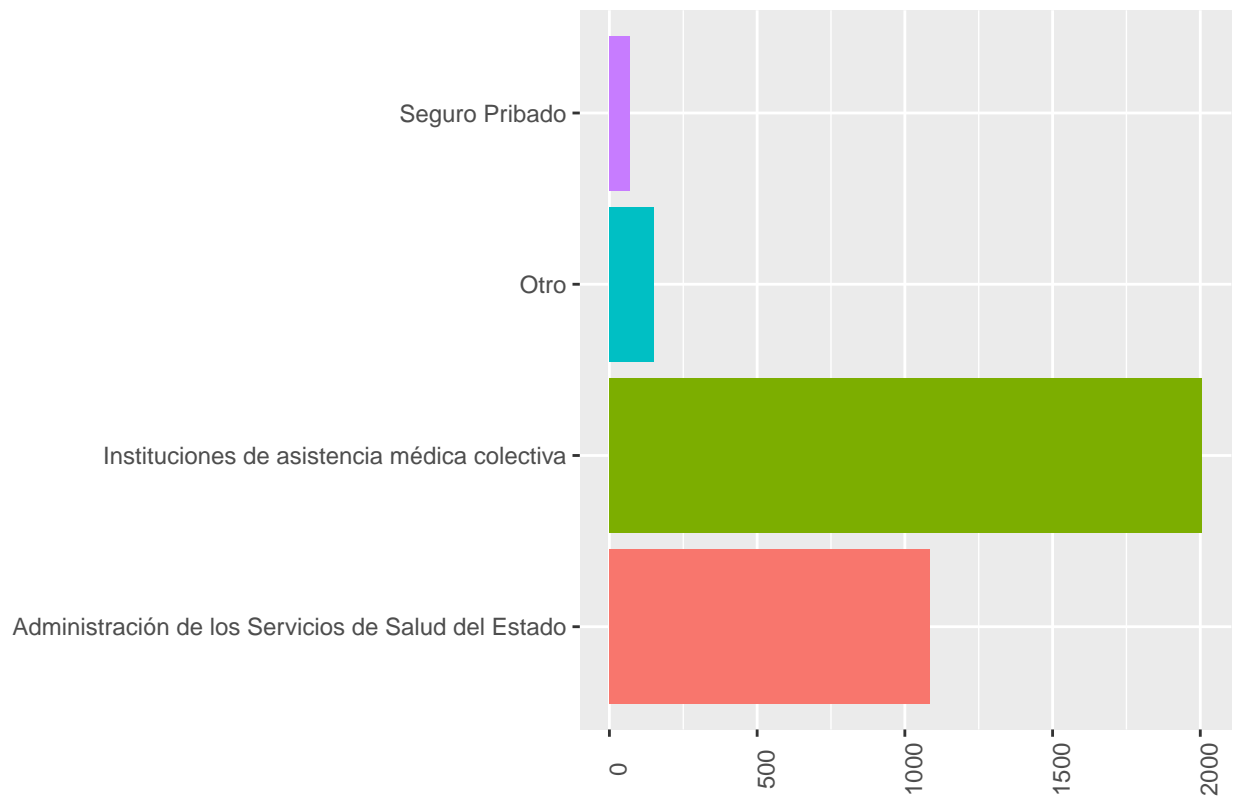


Figura 5. Grafico de barras de tipo de prestadores de salud de aquellos pacientes no residentes del Departamento de Montevideo cuyos prestadores de salud si. Se observa que el nivel predominante es el de las instituciones de asistencia medica colectiva.