

Tarea 2

Estudiante: Emiliano Barone

Ejercicio 1

Parte 1

```
dato_emision1%>%  
  filter(!is.na(emisión))%>%  
  group_by(fuente)%>%  
  summarise(sumaEmision = sum(emisión))%>%  
  arrange(desc(sumaEmision))
```

```
## # A tibble: 9 x 2  
##   fuente sumaEmision  
##   <fct>      <dbl>  
## 1 Q_B      99670.  
## 2 T        69641.  
## 3 BI       30377.  
## 4 CE_SP    24099.  
## 5 I        18198  
## 6 A_P_M    13891.  
## 7 R        11811.  
## 8 CP        8134  
## 9 C_S_SP    3229
```

Faltó filtrar los totales y
subtotales

Parte 2

```
dato_emision1%>%  
  filter(fuente=="Q_B" , emisión==max(emisión))
```

```
##      AÑO fuente emisión  
## 1 2016    Q_B  8831.1
```

Bien!!

Parte 3

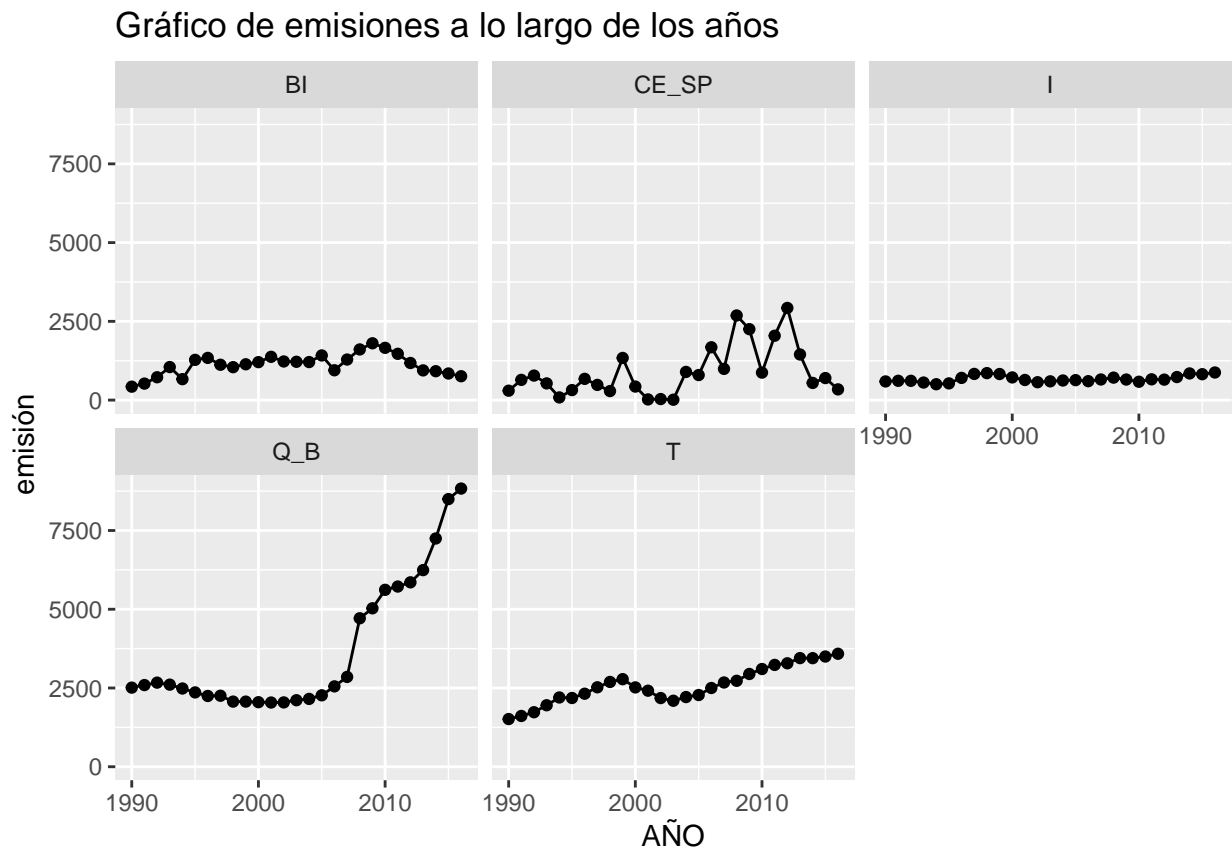
```
dato_emision1%>%  
  filter(!is.na(emisión))%>%  
  group_by(fuente)%>%  
  summarise(mediaEmision = mean(emisión))%>%  
  arrange(desc(mediaEmision))
```

```
## # A tibble: 9 x 2  
##   fuente mediaEmision  
##   <fct>         <dbl>  
## 1 Q_B          3691.  
## 2 T            2579.  
## 3 BI           1125.  
## 4 CE_SP         893.  
## 5 I             674.  
## 6 A_P_M         514.  
## 7 R            437.  
## 8 CP           301.  
## 9 C_S_SP        120.
```

Bien!!

Parte 4

```
dato_emision2%>%  
  filter(!is.na(emisión))%>%  
  ggplot(aes(AÑO , emisión, group=1 ))+  
  geom_point()+  
  geom_line()+  
  facet_wrap(~fuente)+  
  ggtitle("Gráfico de emisiones a lo largo de los años")
```



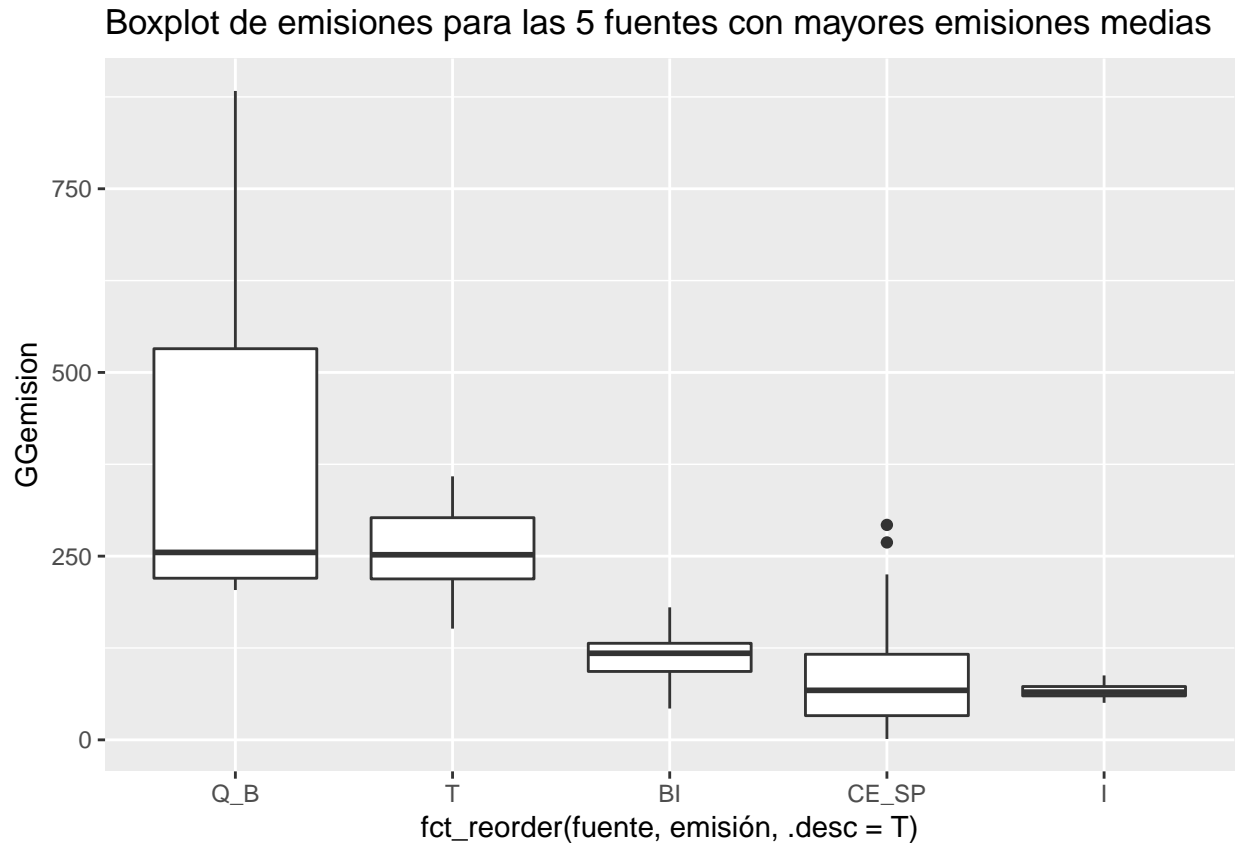
En el siguiente gráfico se puede ver que la fuente “Q_B” (quema de biomasa con fines energéticos) ha presentado un crecimiento a lo largo de los años a partir del año 2000. Un caso similar sucede con la fuente “T” (quema de combustibles fosiles para el transporte individual y/o colectivo por medios aereos). Las demás fuentes presentan una emision de CO2 irregular.

Bien, cuidar el nombre de los ejes (mayúscula
minúscula)

Parte 5

```
library(forcats)

dato_emision2%>%
  filter(!is.na(emisión))%>%
  mutate(GGemision=emisión/10)%>%
  ggplot(aes(x=fct_reorder(fuente, emisión, .desc = T), y=GGemision))+
  geom_boxplot()+
  ggtitle("Boxplot de emisiones para las 5 fuentes con mayores emisiones medias ")
```

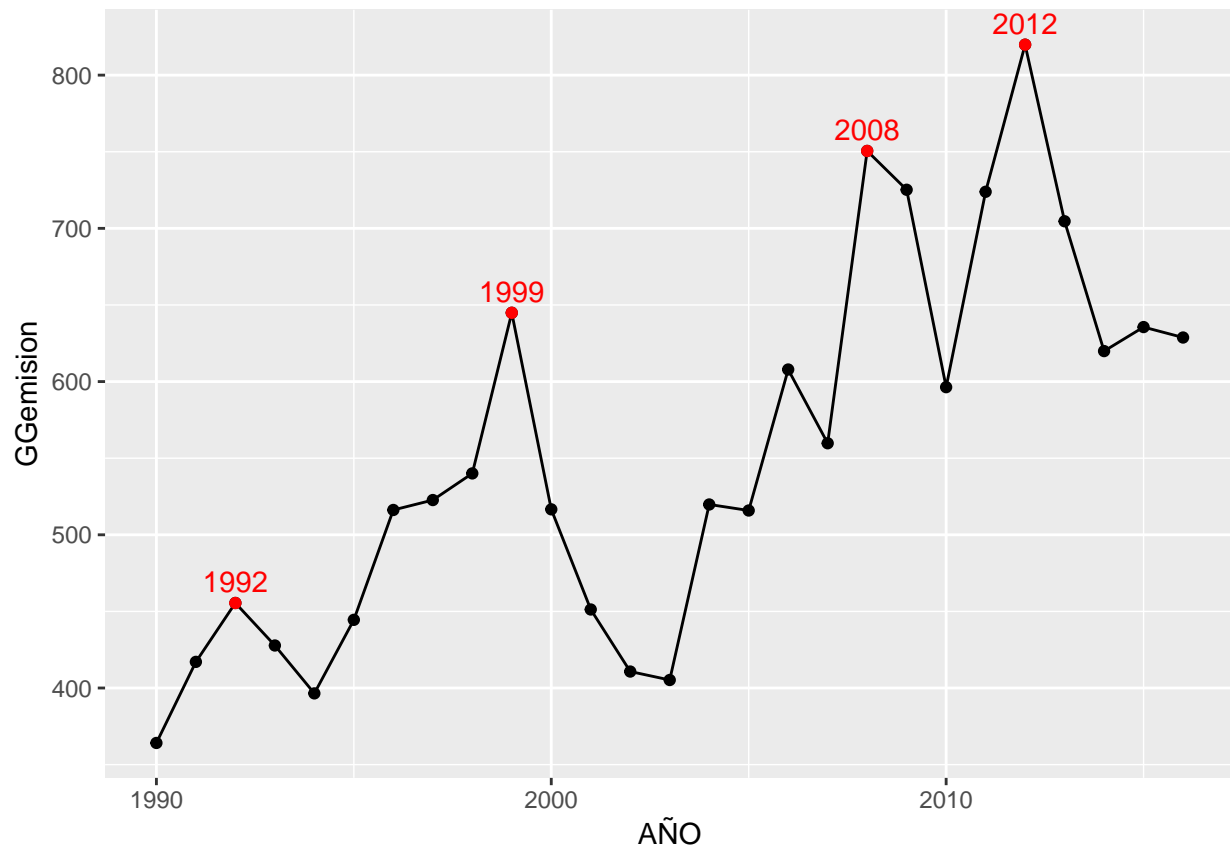


En este gráfico podemos apreciar que las fuentes que mas CO2 emiten (“Q_B” y “T”) tienen igual mediana. Es decir, las dos fuentes acumulan el 50% de su distribución en el valor de los 2500 GG.

Bien, faltan los nombres de los ejes

Parte 6

```
dato_emision%>%  
  filter(!is.na(emisión))%>%  
  filter(fuente=="TOTAL")%>%  
  mutate(emisión, GGemision=emisión/10)%>%  
  ggplot(aes(AÑO,GGemision, group=1))+  
  geom_line()+  
  geom_point()+  
  scale_x_continuous(breaks=c(1990,2000,2010))+  
  stat_peaks(colour="red")+  
  stat_peaks(geom="text" ,colour = "red", vjust=-0.5)
```



En este gráfico podemos ver los máximos relativos de la emisión de CO2 de las 5 fuentes que mas cantidad emiten a lo largo de los años. Se puede observar que cada máximo es superior al anterior.

Bien!!!

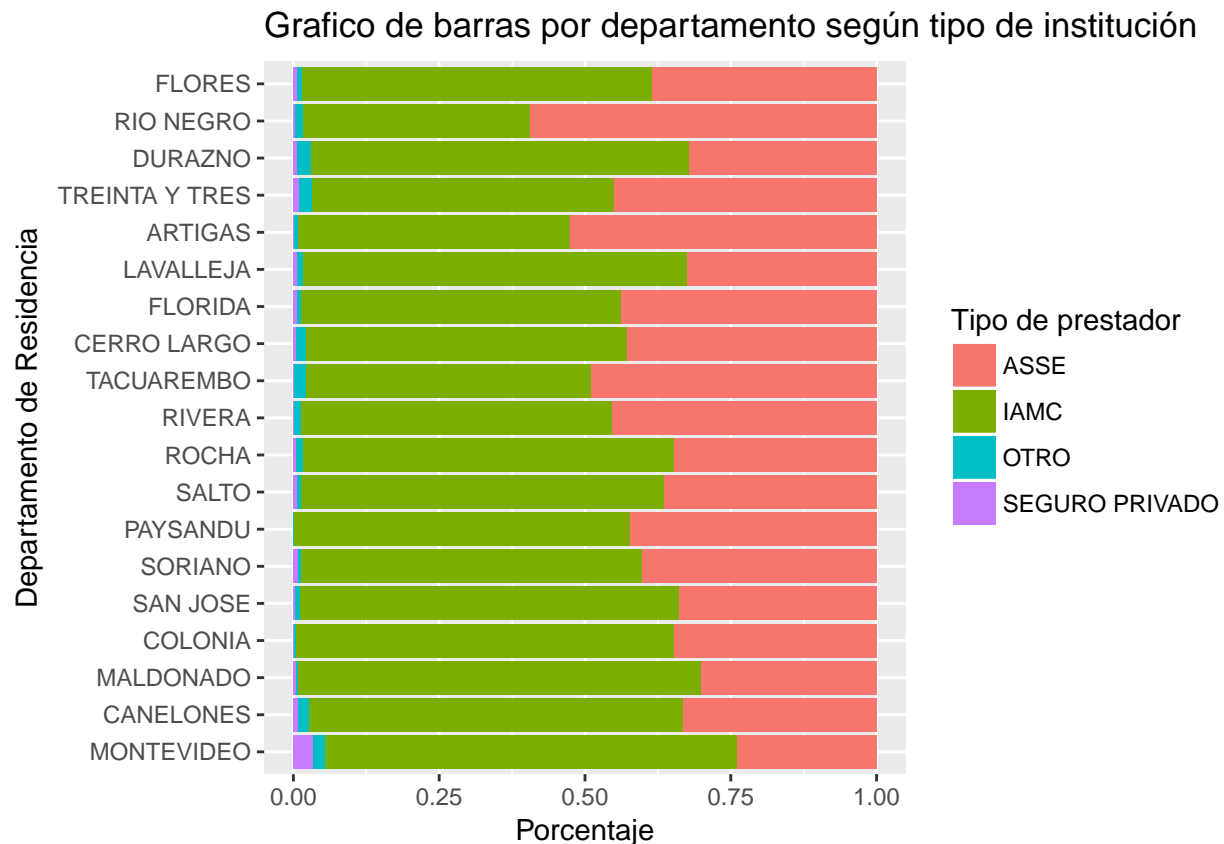
Ejercicio 2

Preguntas Orientadoras

1. ¿Cual es la proporción de personas que pertenecen a ASSE , IAMC y seguros privados segun departamento?
2. ¿Cuales son las 10 proporciones de prestadores mas grandes?
3. ¿Como se ve afectado el importe segun los distintos rangos de edad?
4. ¿Como se ven afectadas las 10 prestaciones mas frecuentes segun Sexo y departamento de residencia?
5. Analizar el importe segun las 10 prestaciones más frecuentes.

Pregunta 1

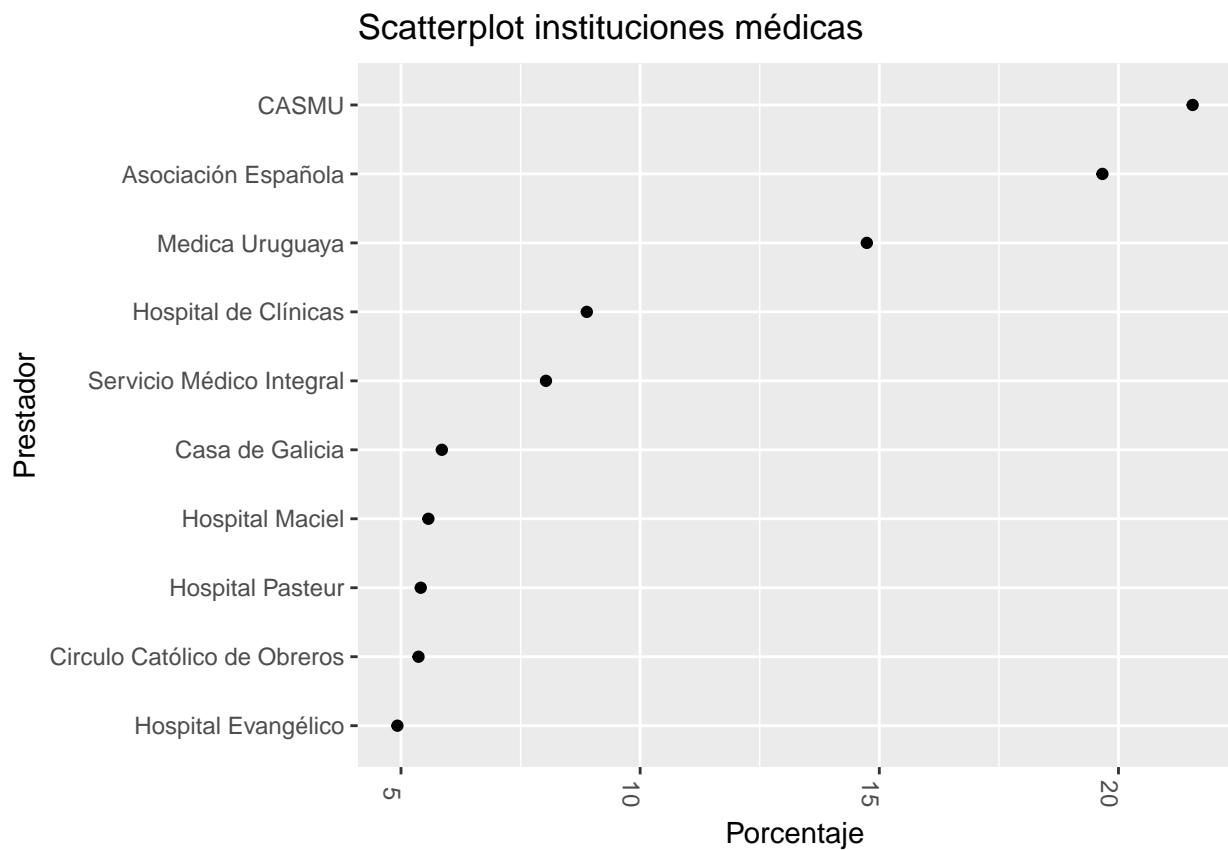
```
DatosGM %>%  
  filter(!is.na(Prestador_tipo)) %>%  
  ggplot()+  
  geom_bar(aes(fct_infreq(Departamento_residencia), fill= Prestador_tipo) , position = "fill")+  
  coord_flip()+  
  labs(x="Departamento de Residencia", y="Porcentaje", fill="Tipo de prestador")+  
  ggtitle("Grafico de barras por departamento según tipo de institución")
```



Se puede observar que en la mayoría de los departamentos del interior hay muy poco porcentaje de “seguro privado” como tipo de prestador. Río Negro es el departamento que menor porcentaje presenta de Instituciones de asistencia medica colectiva(IAMC).

Pregunta 2

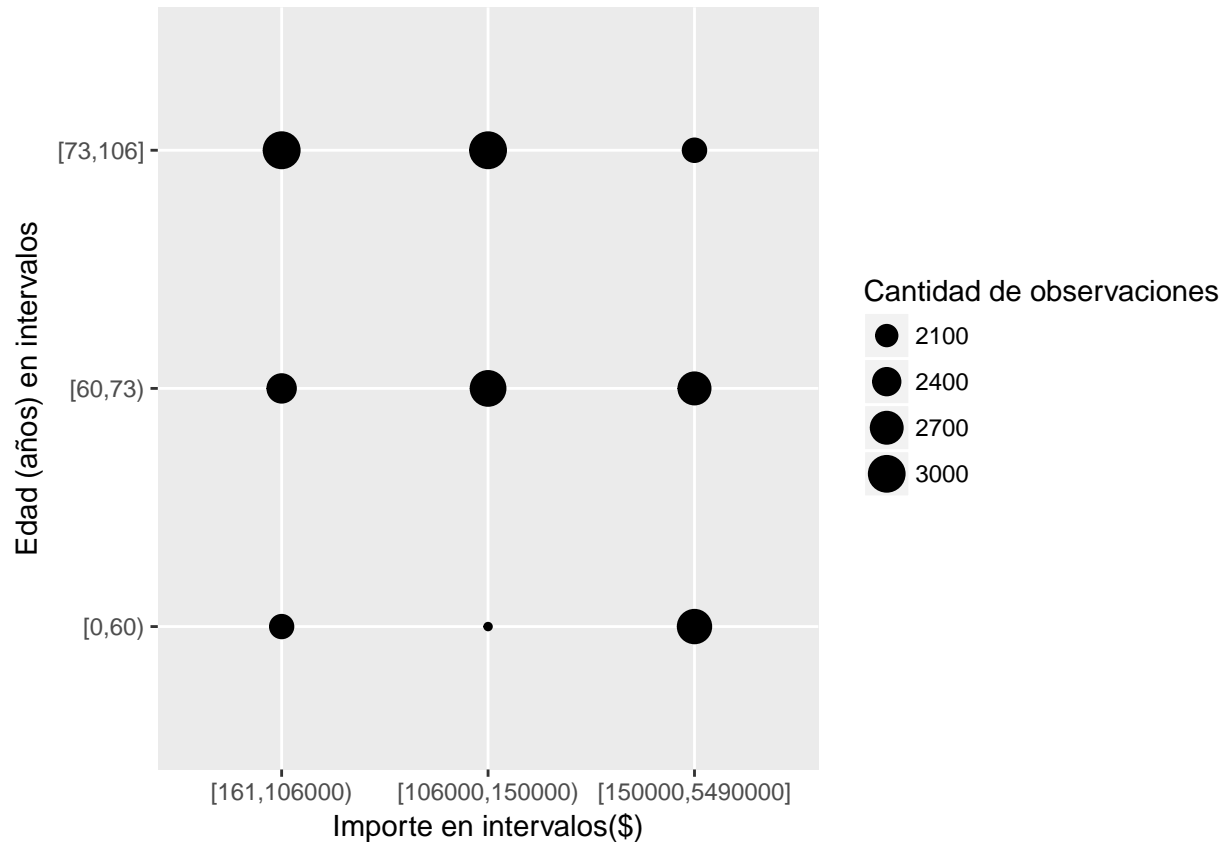
```
DatosGM%>%
  group_by(Prestador)%>%
  summarise(CantP=n())%>%
  arrange(desc(CantP))%>%
  filter(CantP>524)%>%
  mutate(PropPrest= (CantP/sum(CantP))*100)%>%
  ggplot()+
  geom_point(aes(fct_reorder(Prestador, PropPrest, .desc = F), y=PropPrest))+
  coord_flip()+
  theme(axis.text.x = element_text(angle=270))+
  labs(x="Prestador", y="Porcentaje")+
  ggtitle("Scatterplot instituciones médicas")
```



En este gráfico podemos observar los diez prestadores que tienen mayores porcentajes de pacientes. Podemos ver que el centro médico CASMU es el prestador que tiene porcentaje más alto de pacientes y el hospital evangélico es el que menor porcentaje presenta dentro de esta categoría.

Pregunta 3

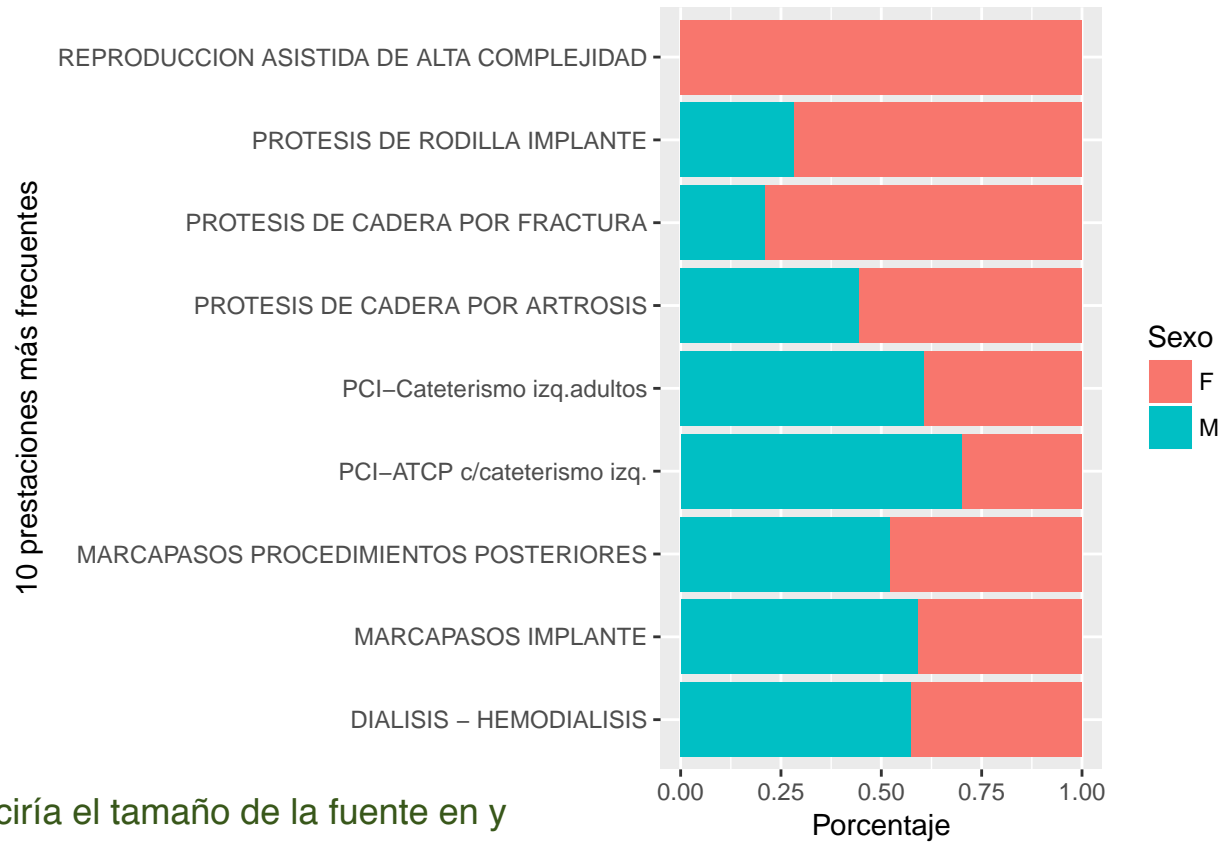
```
DatosGM%>%  
  filter(!is.na(Importe), !is.na(Edad_años))%>%  
  ggplot()+  
  geom_count(aes(Importe, Edad_años))+  
  labs(x="Importe en intervalos($)", y="Edad (años) en intervalos", size="Cantidad de observaciones")
```



El siguiente gráfico nos permite analizar los importes según la edad. Podemos ver que entre los años de vida 0 y 60 casi no se encuentran casos cuyos importes estén situados entre los 100 y 150 mil pesos uruguayos. A su vez podemos ver que para las edades más grandes (entre 73 y 106) hay muchas observaciones para los distintos rangos de importes. Esto tiene mucho sentido ya que por lo general mientras más grande es el individuo, necesita más atención médica.

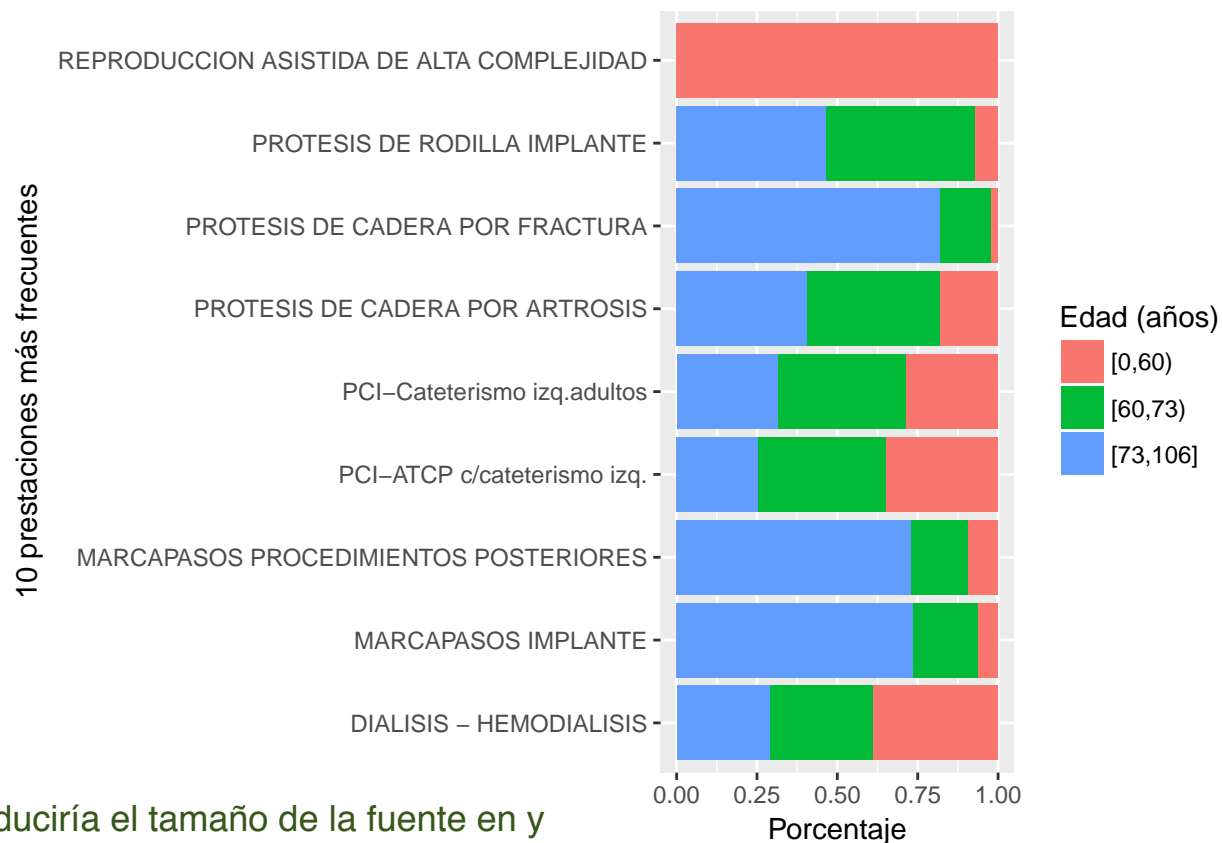
Pregunta 4

```
Datos4%>%
  ggplot()+
  geom_bar(aes(Prestacion, fill= Sexo), position = "fill")+
  coord_flip()+
  labs(x="10 prestaciones más frecuentes", y="Porcentaje", fill="Sexo")
```



Reduciría el tamaño de la fuente en y

```
Datos4%>%
  filter(!is.na(Prestacion), !is.na(Importe))%>%
  ggplot()+
  geom_bar(aes(Prestacion, fill= Edad_años), position = "fill")+
  coord_flip()+
  labs(x="10 prestaciones más frecuentes", y="Porcentaje", fill="Edad (años)")
```



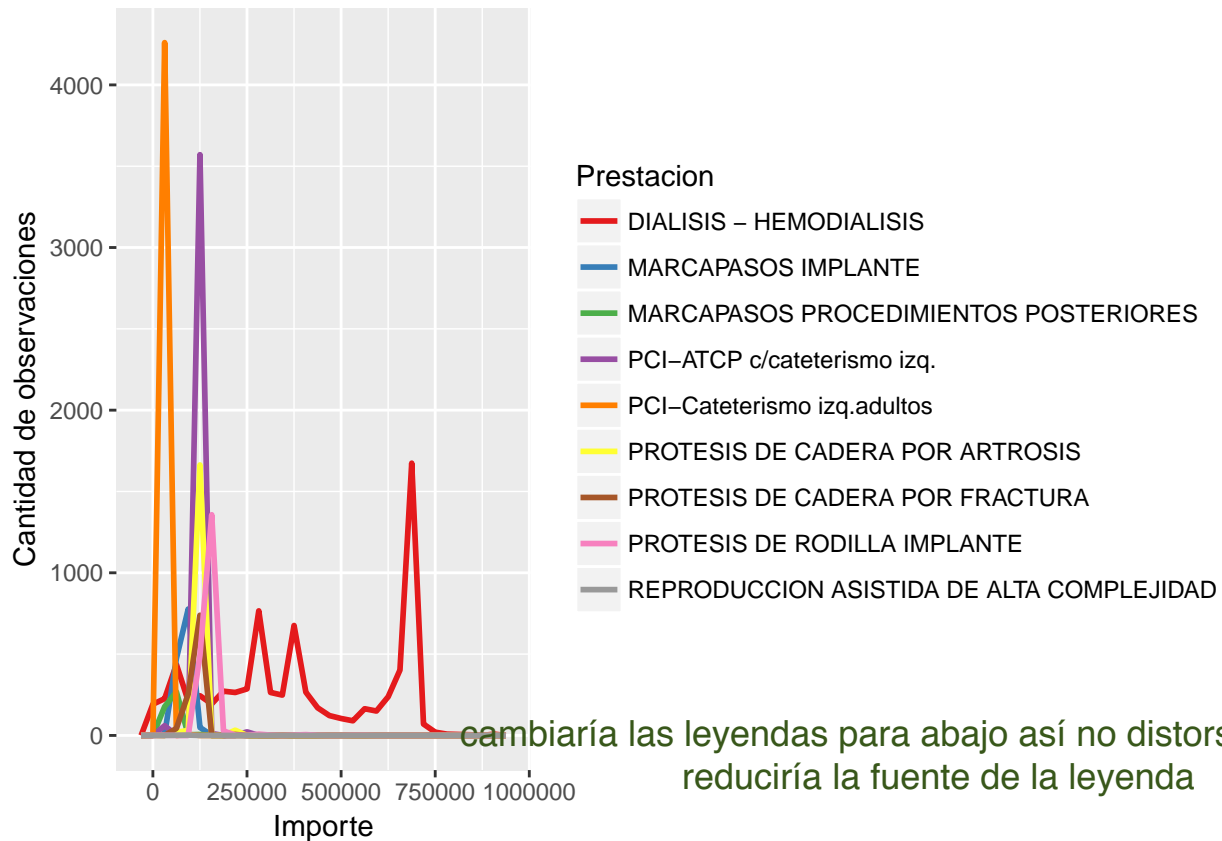
Reduciría el tamaño de la fuente en y

En estos dos gráficos podemos analizar el sexo y la edad en años de los pacientes, según las 10 prestaciones que suceden con más frecuencia. Alguno de los datos a resaltar son que las personas de sexo femenino presentan más porcentaje de Prótesis (tanto de cadera como de rodilla) que las de sexo masculino, que las personas de sexo masculino presentan mayor porcentaje de procedimientos coronarios intervencionista. Y en el caso de las edades podemos ver que las prestaciones relacionadas con marcapasos y las prótesis de cadera por fractura se dan con mayor frecuencia en los pacientes que tienen entre 76 y 106 años. En el caso de la diálisis se dan con mayor frecuencia con los pacientes que tienen entre 0 y 60 años.

Pregunta 5

```
Datos4%>%  
  filter(!is.na(Importe), !is.na(Prestacion))%>%  
  ggplot(aes(Importe, colour=Prestacion))+  
  geom_freqpoly(size=1)+  
  scale_colour_brewer(palette = "Set1")+  
  labs(x="Importe", y="Cantidad de observaciones")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



En este gráfico podemos ver la cantidad de casos que cuestan determinado importe según el tipo de prestación. Podemos resaltar los casos de los procedimientos coronarios intervencionistas cuyos valores de importe son aproximadamente \$50,000 y \$100,000 respectivamente y cuyas cantidades de observaciones son superiores a las 4,000 en el primer caso y en el entorno de las 3,500 en el segundo caso. También se puede notar que la diálisis y hemodialisis presentan muchas observaciones al rededor de los 600,000 pesos.

Muy buena exploración!!!

Muy buen trabajo 100/100