

Visualización

Carlos Javier Rincón

versión:2019

Contents

Introducción.	1
Gráfico de barras	2
Histograma	10
Gráfico de cajas y bigotes	16
Gráfico de dispersión	24
Pie o torta	32

Introducción.

En este documento encontrará una guía para construir gráficos utilizando el paquete `ggplot2`. Se presentan los siguientes tipos de gráficos: Barra, histogramas, cajas y bigotes y gráficos de dispersión. Para cada tipo de gráfico se presenta un ejemplo y a continuación se propone un ejercicio para replicar lo aprendido. Esta práctica se realizará utilizando la base de datos de recién nacidos **RN**.

Los paquetes que se requieren para este documento son:

```
#install.packages("rio")
#install.packages("tidyverse")
#install.packages("ggthemes")
```

Primero cargamos la base de datos:

```
library(rio)

## Warning: package 'rio' was built under R version 3.6.1

RN<-import("RN.xlsx")
```

Esta es la descripción de las variables incluidas:

```
str(RN)

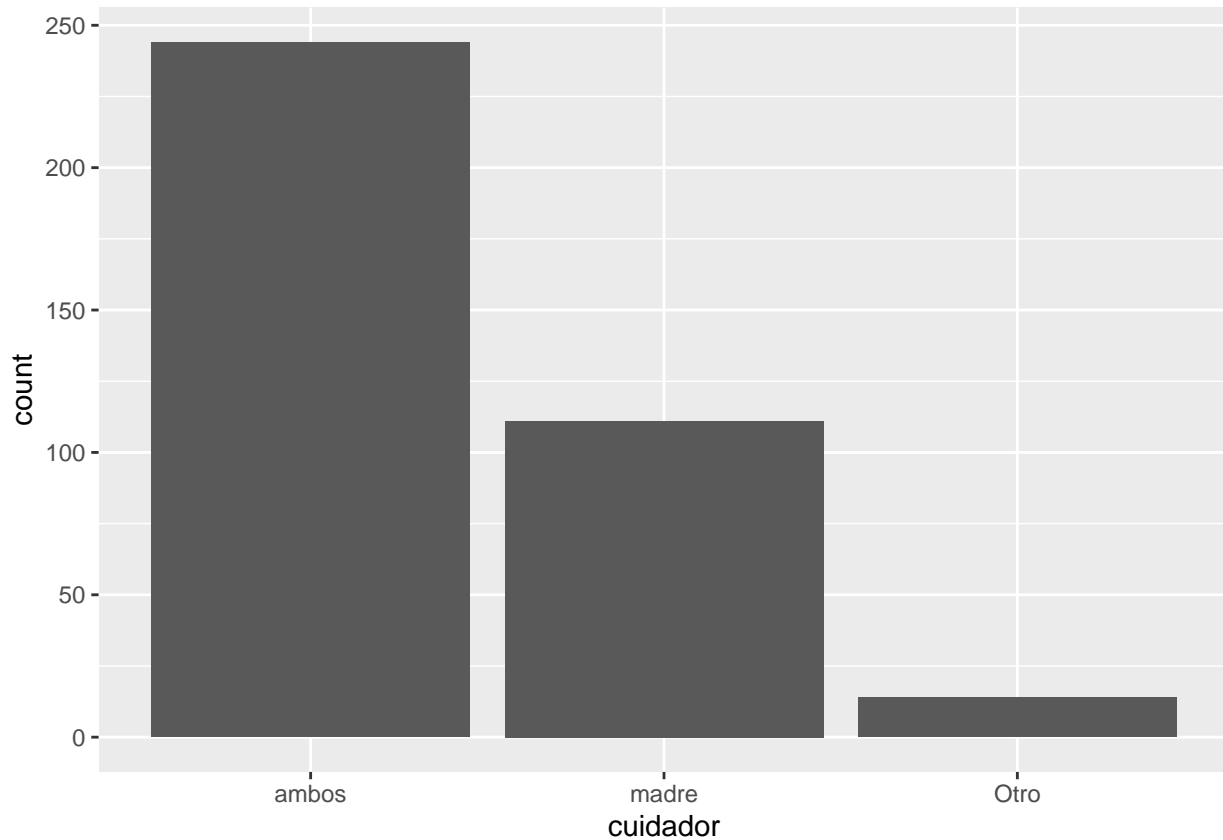
## 'data.frame':   369 obs. of  9 variables:
## $ id      : num  1 2 3 4 5 6 7 9 11 12 ...
## $ peso    : num  0.27 1.48 0.83 0.52 0.25 2.41 1.13 2.08 0.81 1.28 ...
## $ talla   : num  46.1 48.2 40.7 56.3 51.5 44.5 45 37.6 43.5 45.2 ...
## $ cuidador : chr  "madre" "madre" "madre" "madre" ...
## $ controles: num  5 8 6 6 4 7 5 12 3 3 ...
## $ lactancia: chr  "No" "Si" "Si" "No" ...
## $ sexo     : chr  "M" "M" "M" "M" ...
## $ sueno    : num  9.4 7.7 6.2 6.9 6.7 8.9 8.7 4.5 8.4 4.6 ...
## $ estrato  : chr  "E5" "E1" "E4" "E1" ...
```

Gráfico de barras

Utilizaremos la variable cuidador. El primer gráfico se obtiene de la siguiente forma:

```
library(ggplot2)

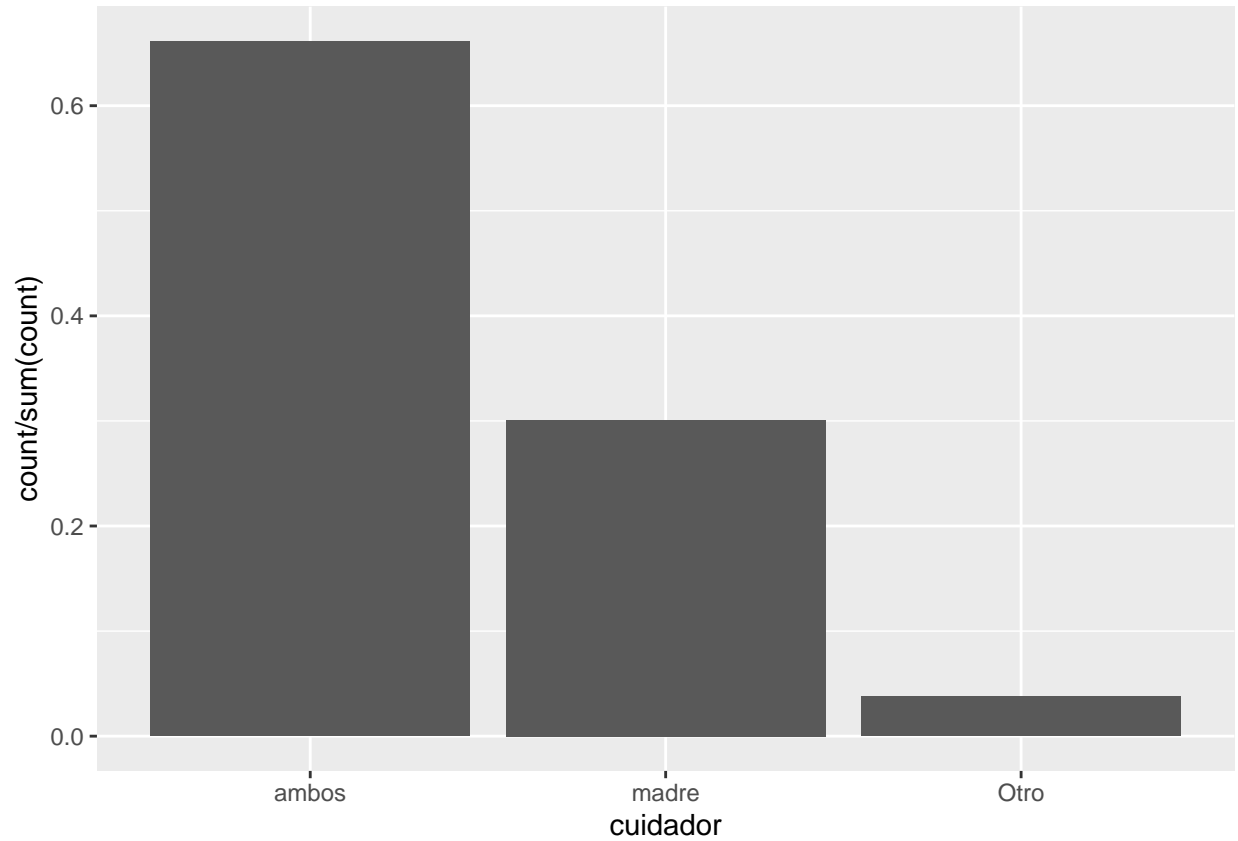
ggplot(data=RN,aes(x=cuidador))+
  geom_bar()
```



-
- En el código anterior se pueden identificar los siguientes tres elementos: `data=RN` donde se define el nombre de la base de datos a utilizar, `aes()` (aesthetic) donde se definen algunos atributos, principalmente la ubicación de las variables de interés dentro del gráfico y `geom_` que corresponde al objeto o forma que se mostrará (tipo de gráfico). Los tres elementos anteriores de la función `ggplot` serán utilizados en todos los gráficos que revisamos en este documento.
-

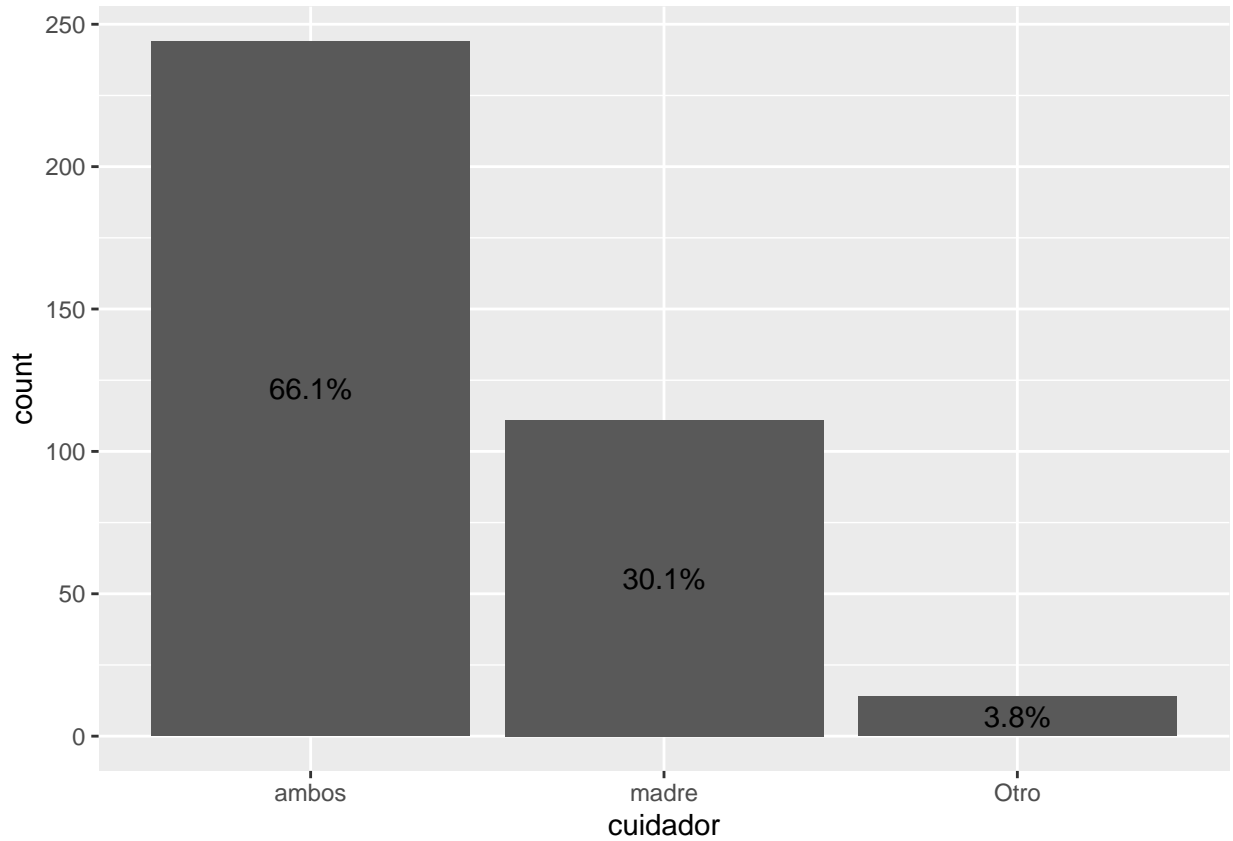
Continuando con el gráfico de barras, si el interés es representar proporciones y no conteo se puede realizar de la siguiente forma:

```
ggplot(data=RN,aes(x=cuidador,y = ..count../sum(..count..)))+
  geom_bar()
```



En ocasiones el interés es colocar en el mismo gráfico tanto la frecuencia como el porcentaje, así:

```
ggplot(RN, aes(x=cuidador)) +
  geom_bar() +
  geom_text(aes(label=scales::percent(..count../sum(..count..))), stat="count",
            position=position_stack(0.5))
```

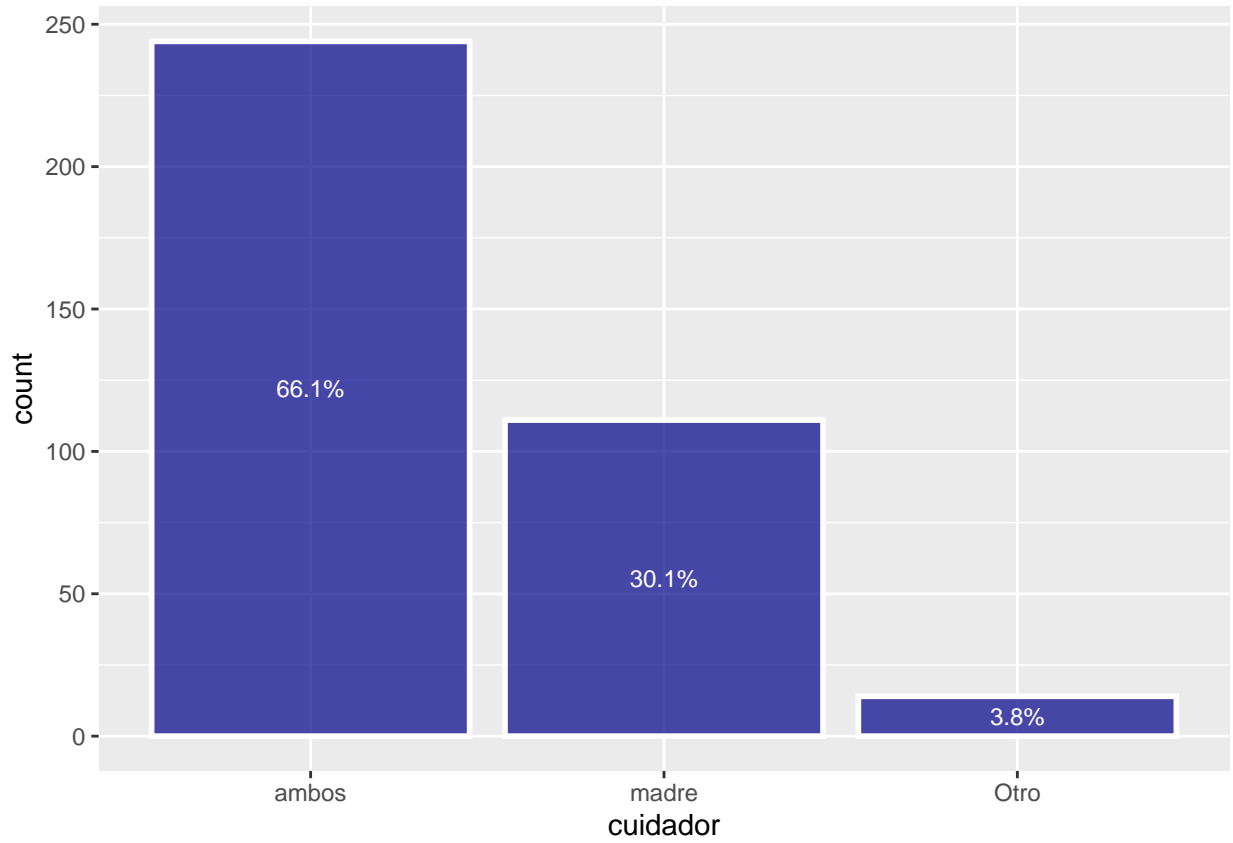


El gráfico anterior se puede guardar dentro de un objeto llamado **fig1**, al cual le cambiamos los colores:

```
fig1<-ggplot(RN,aes(x=cuidador))+
  geom_bar(fill="darkblue",alpha=0.7,col="white",size=1)+
  geom_text(aes(label=scales::percent(..count../sum(..count..))),stat="count",
            position=position_stack(0.5),size=3,col="white")
```

veamos nuestro gráfico:

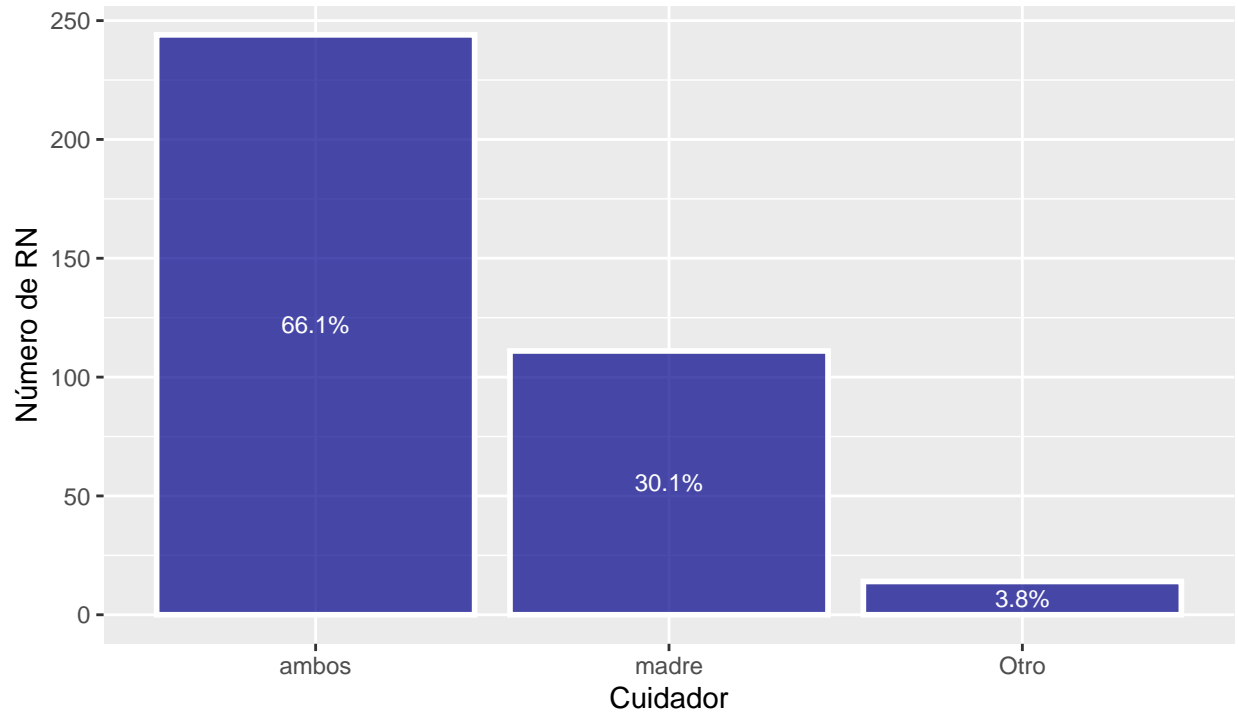
```
fig1
```



Coloquemos títulos:

```
fig1 + labs(title="Número de recién nacidos por tipo de cuidador",  
            subtitle="Cali - Colombia", y="Número de RN", x="Cuidador", caption="Fuente: Estudio RN2015")
```

Número de recién nacidos por tipo de cuidador Cali – Colombia

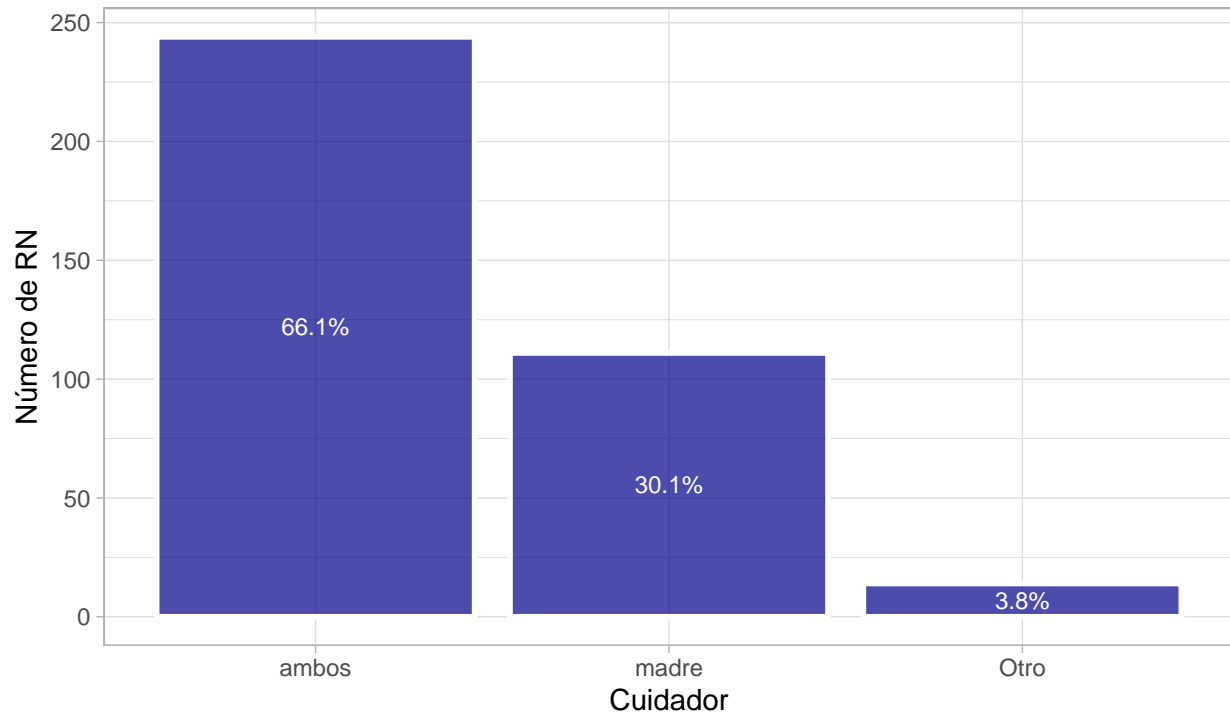


Fuente: Estudio RN2015

El siguiente listado son alternativas para modificar rápidamente la presentación del gráfico: `theme_gray`, `theme_bw()`, `theme_linedraw()`, `theme_light()`, `theme_minimal()`, `theme_classic()`, `theme_void()`, `theme_dark()`. Utilicemos una:

```
fig1 + labs(title="Número de recién nacidos por tipo de cuidador",  
  subtitle="Cali - Colombia", y="Número de RN", x="Cuidador",  
  caption="Fuente: Estudio RN2015") +  
  theme_light()
```

Número de recién nacidos por tipo de cuidador Cali – Colombia



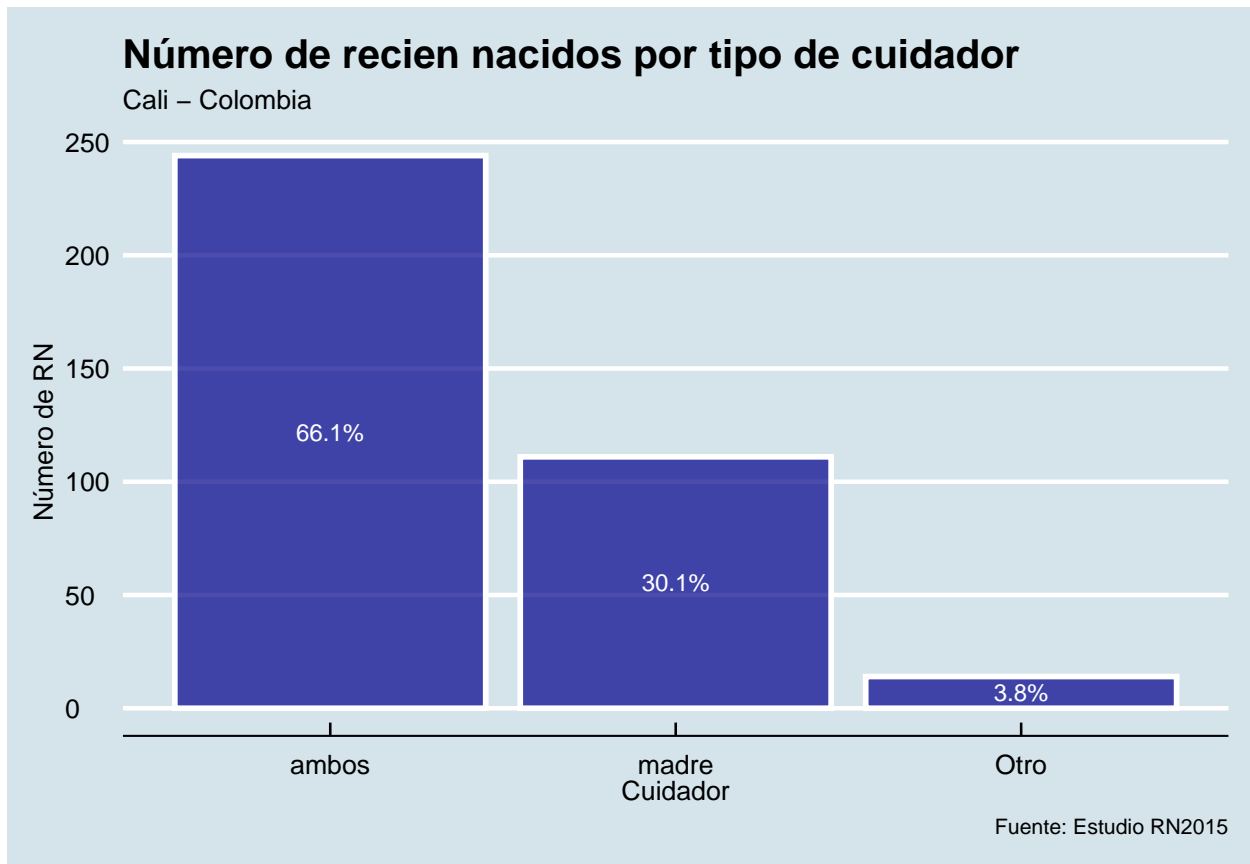
Fuente: Estudio RN2015

Otras opciones de presentación utilizando el paquete ggthemes son: `theme_tufte()`, `theme_economist()`, `theme_stata()`, `theme_calc()`, `theme_hc()`.

Veamos:

```
library(ggthemes)

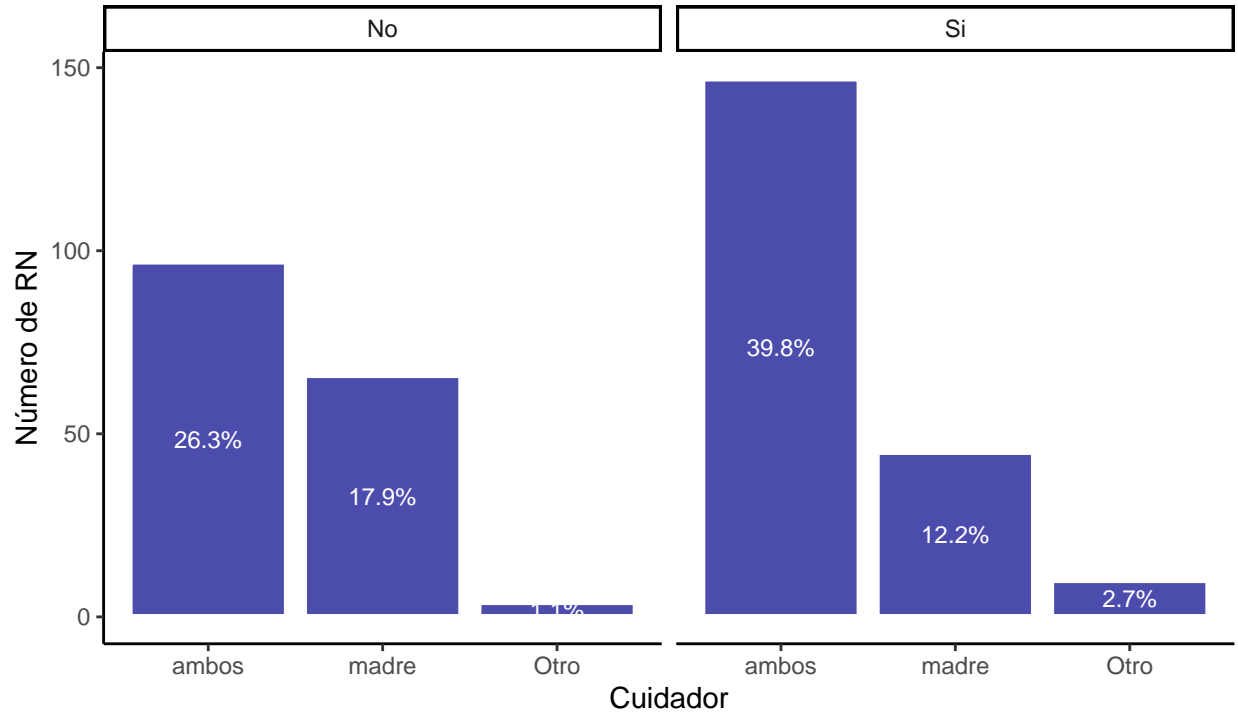
fig1 + labs(title="Número de recién nacidos por tipo de cuidador",
            subtitle="Cali - Colombia", y="Número de RN", x="Cuidador",
            caption="Fuente: Estudio RN2015") +
  theme_economist()
```



Finalmente, realizar el gráfico anterior por subpoblaciones, usamos la opción `facet_grid`, así:

```
fig1 + labs(title="Número de recién nacidos por tipo de cuidador y lactancia",
  subtitle="Cali - Colombia", y="Número de RN", x="Cuidador",
  caption="Fuente: Estudio RN2015") +
  theme_classic() +
  facet_grid(~lactancia)
```


Número de recién nacidos por tipo de cuidador y lactancia
Cali – Colombia



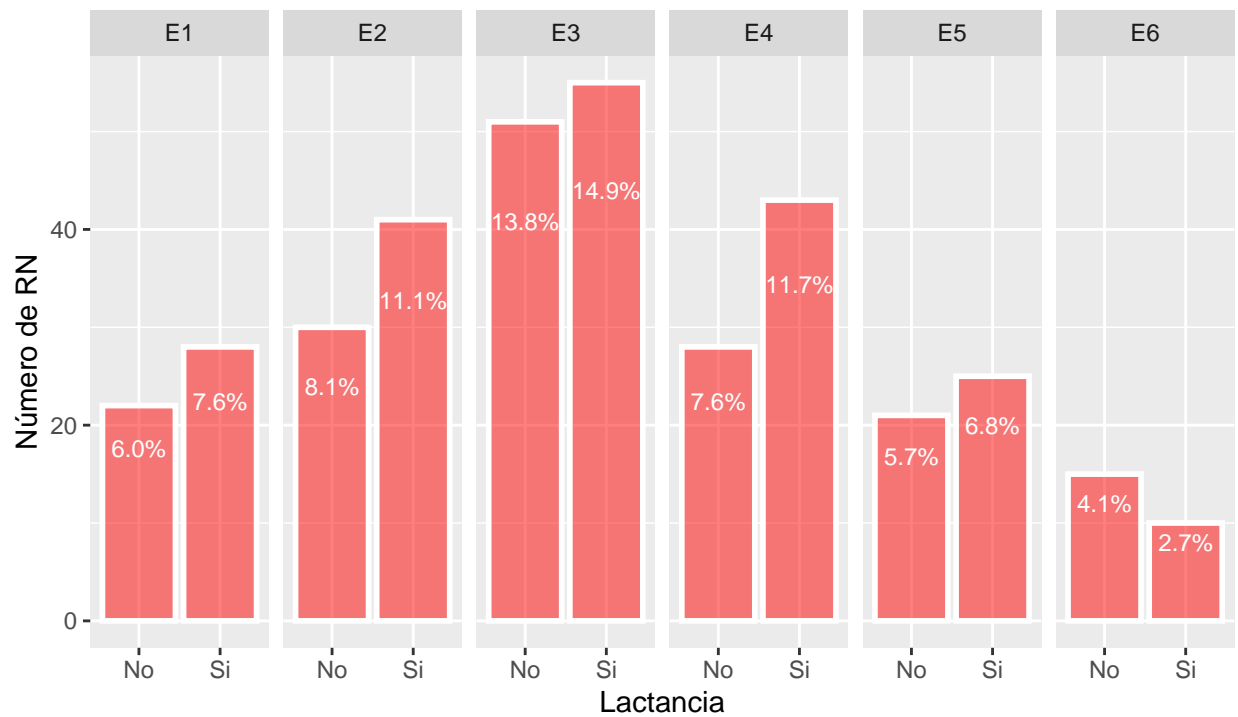
Fuente: Estudio RN2015

- Ejercicio

Reproduzca el siguiente gráfico (no es necesario que coincidan los colores):

Número de recién nacidos por lactancia y estrato

Cali – Colombia

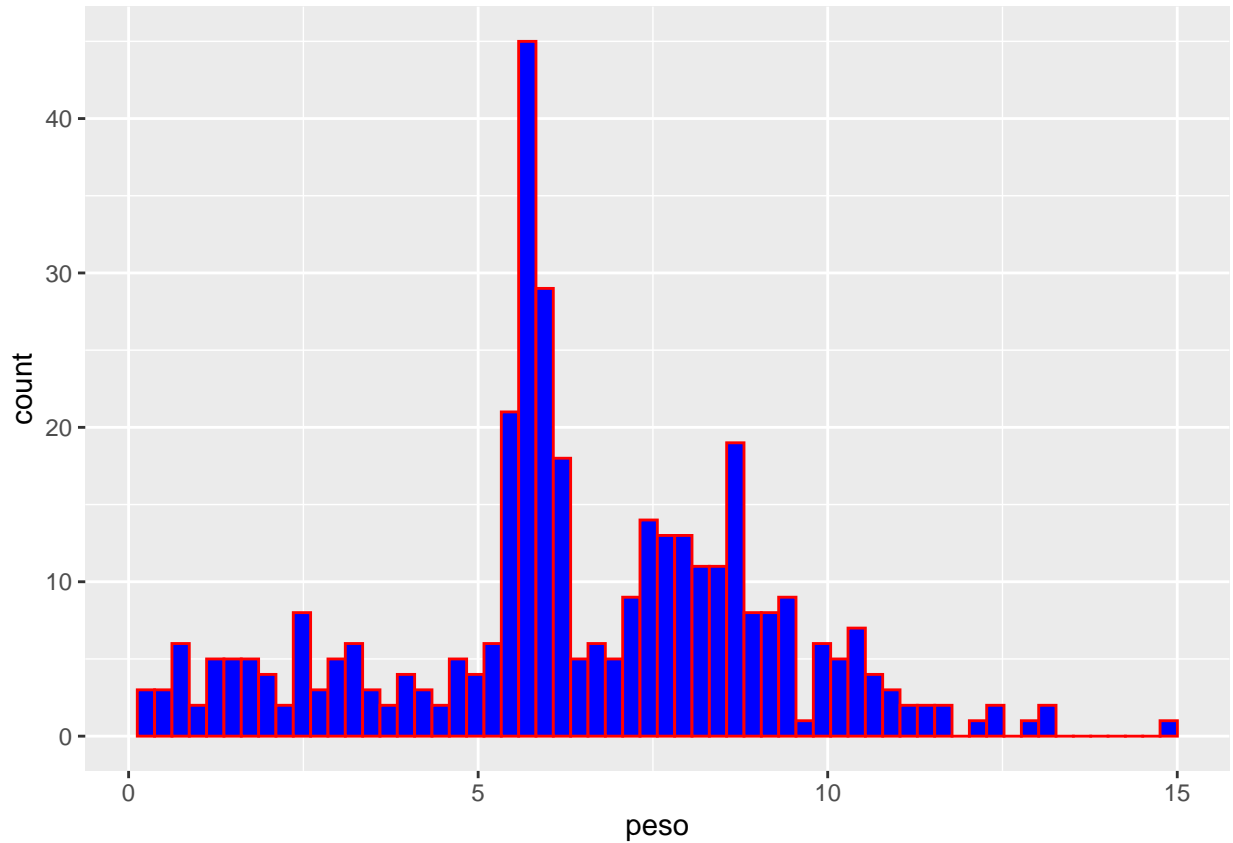


Fuente: Estudio piloto de RN 2015

Histograma

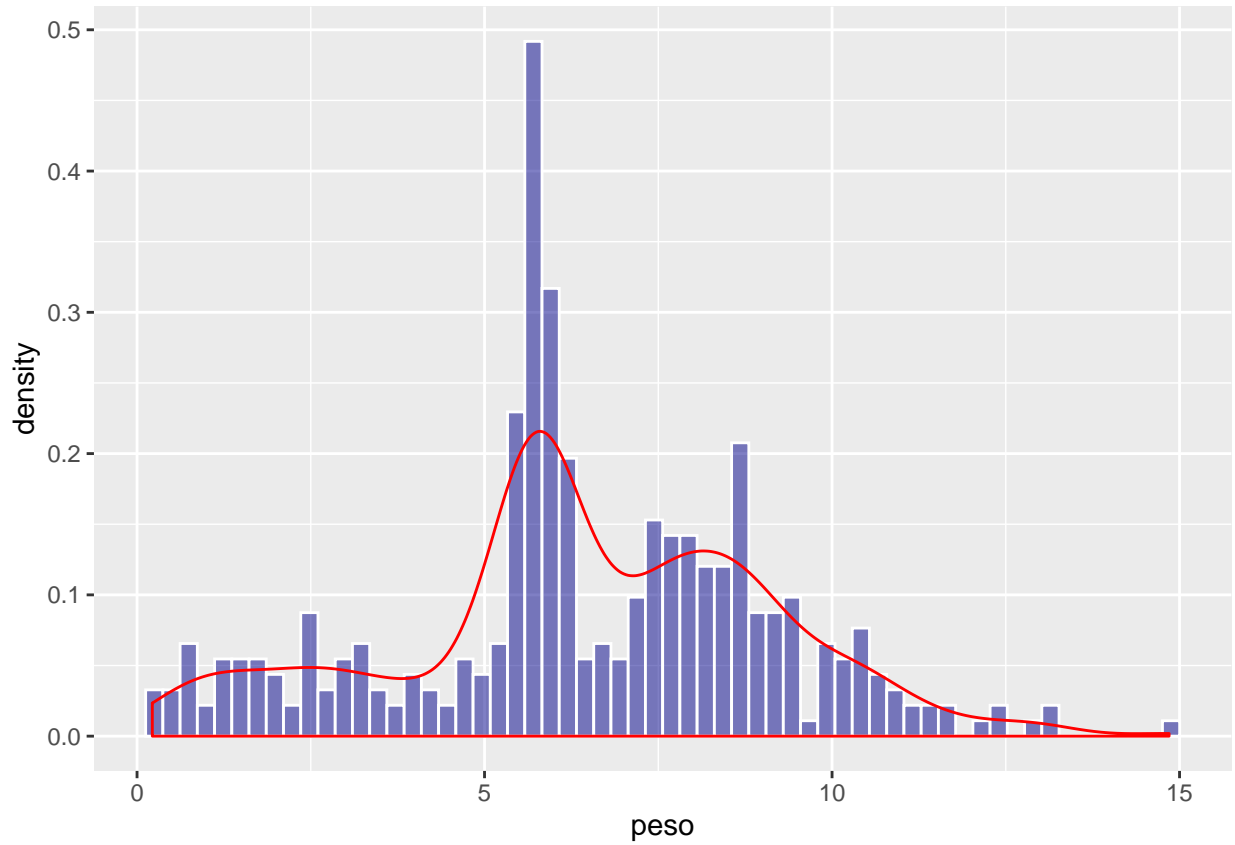
Para este tipo de gráfico utilizaremos la variable peso. Así:

```
ggplot(data=RN, aes(x=peso)) +  
  geom_histogram(bins = 60, fill="blue", color = "red")
```



Incluimos un suavizamiento a nuestro histograma y en el eje y presentemos la densidad:

```
ggplot(data=RN, aes(x=peso)) +
  geom_histogram(bins=60,aes(y = ..density..), alpha=0.5,
                fill="darkblue", color = "white")+
  geom_density(col="red",size=0.5)
```

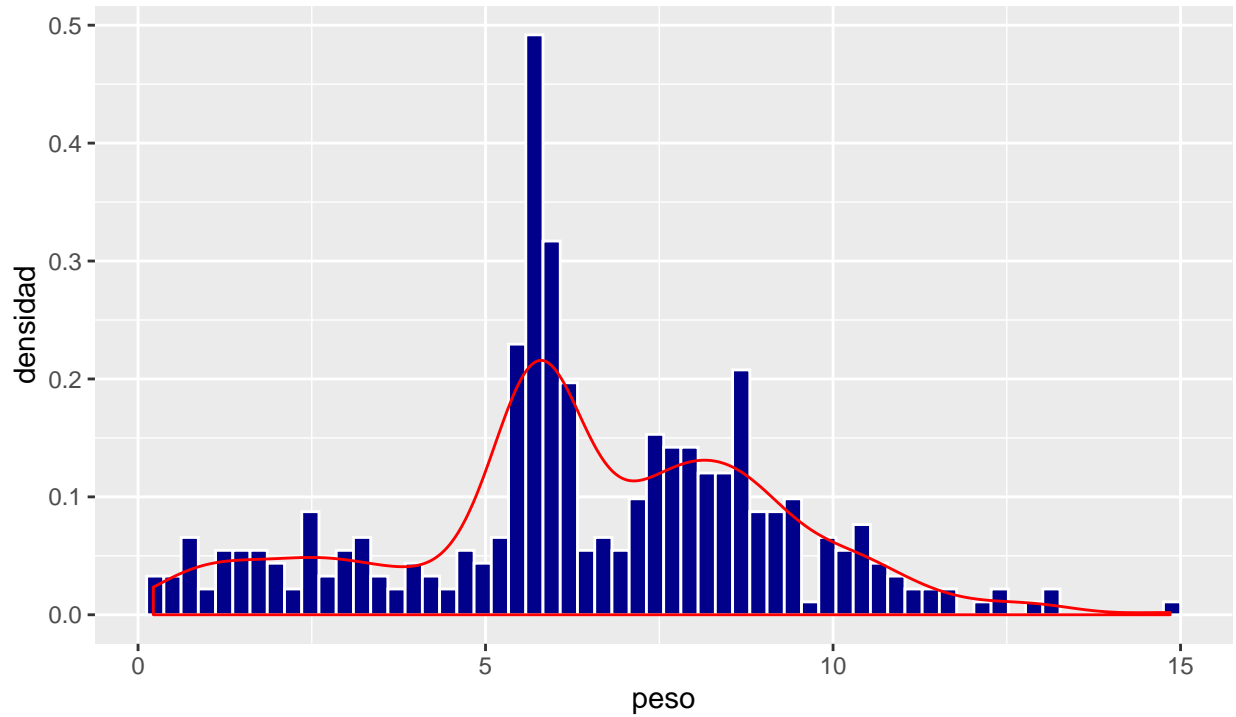


colocando títulos obtenemos:

```
ggplot(data=RN, aes(x=peso)) +
  geom_histogram(bins=60,aes(y = ..density..),fill="darkblue", color = "white")+
  geom_density(col="red",size=0.5) +
  labs(title="Histograma peso",subtitle="Cali - Colombia", y="densidad", x="peso",
  caption="Fuente: Estudio RN2015")
```

Histograma peso

Cali – Colombia

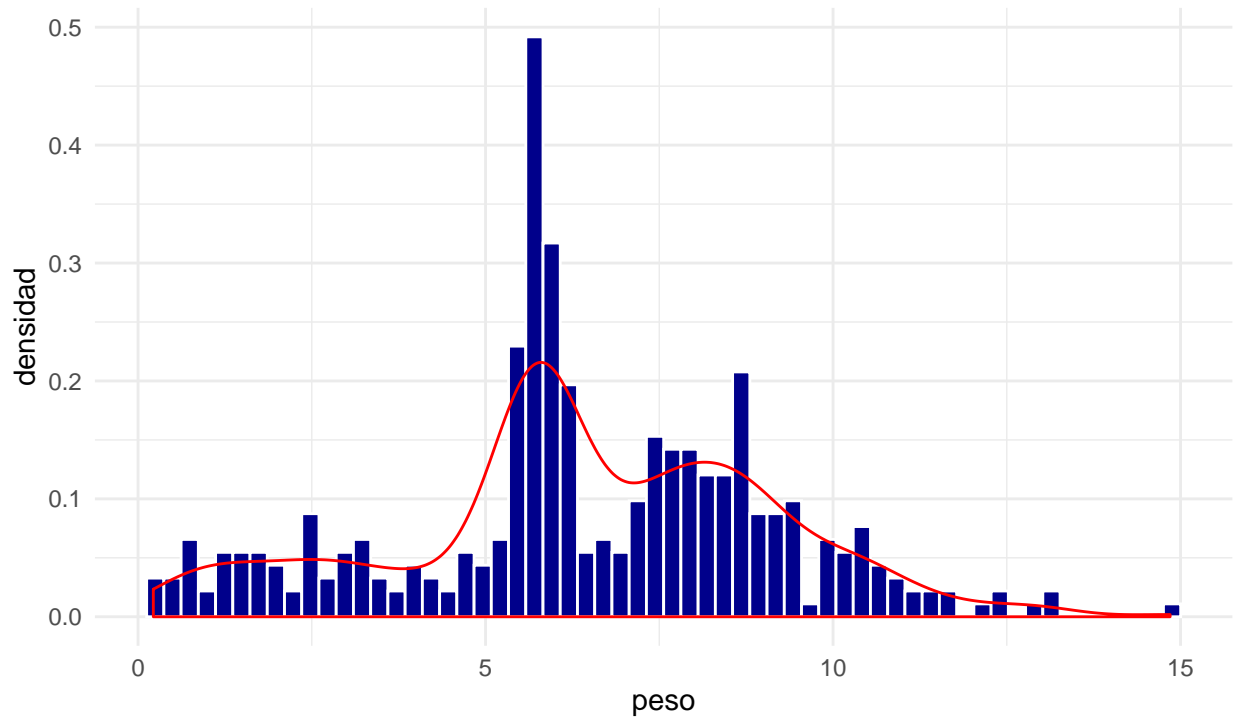


Fuente: Estudio RN2015

ahora probando algunos temas:

```
ggplot(data=RN, aes(x=peso)) +  
  geom_histogram(bins=60,aes(y = ..density..),fill="darkblue", color = "white")+  
  geom_density(col="red",size=0.5) +  
  labs(title="Histograma peso",subtitle="Cali - Colombia", y="densidad", x="peso",  
        caption="Fuente: Estudio RN2015")+  
  theme_minimal()
```

Histograma peso Cali – Colombia



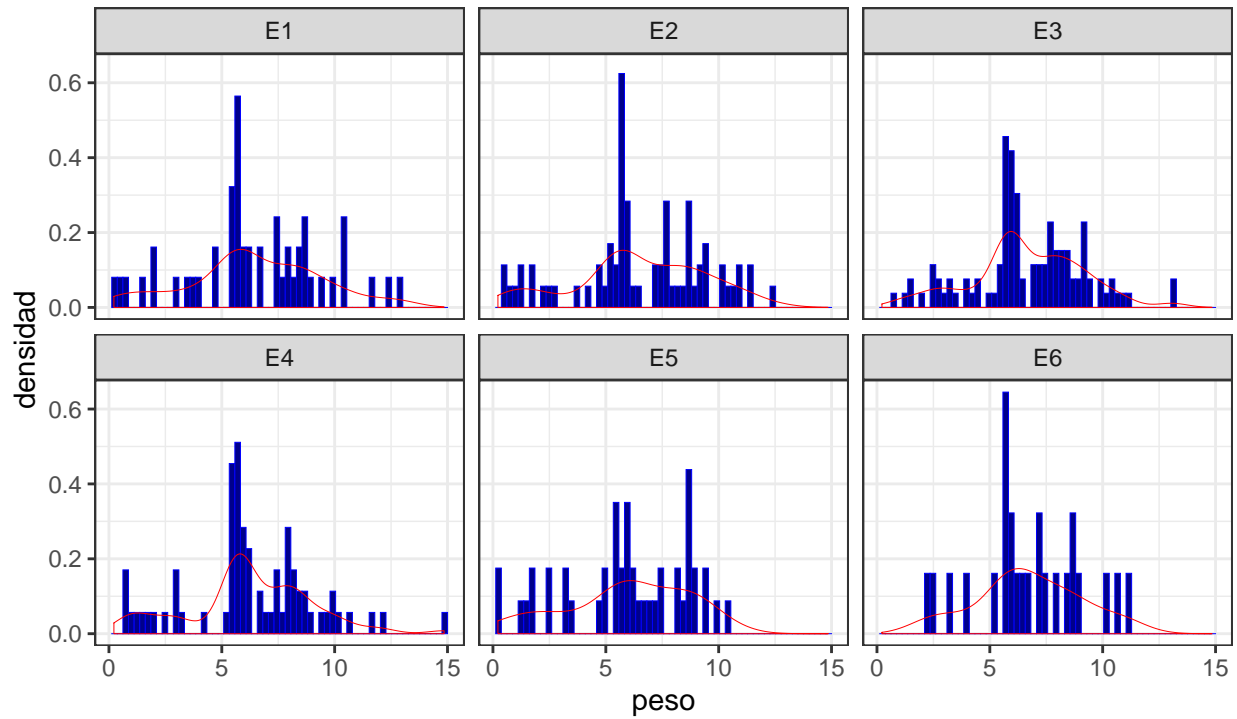
Fuente: Estudio RN2015

El gráfico anterior por subpoblaciones, por ejemplo estratos, se obtiene así:

```
ggplot(data=RN, aes(x=peso)) +  
  geom_histogram(bins=60,aes(y = ..density..),fill="darkblue", color = "blue",size=0.1)+  
  geom_density(col="red",size=0.1) +  
  labs(title="Histograma peso por estratos",subtitle="Cali - Colombia", y="densidad", x="peso",  
        caption="Fuente: Estudio RN2015")+  
  theme_bw()+  
  facet_wrap(~estrato)
```

Histograma peso por estratos

Cali – Colombia



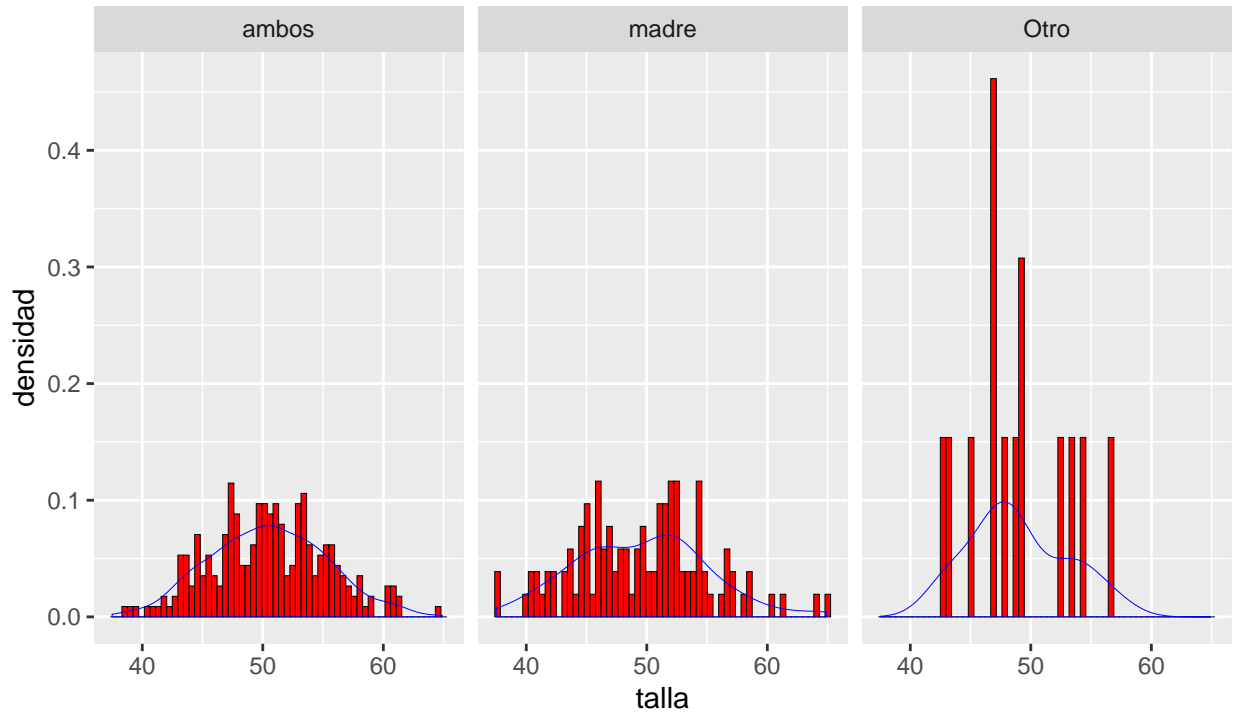
Fuente: Estudio RN2015

- Ejercicio:

Reproduzca el siguiente gráfico (no es necesario que coincidan los colores):

Histograma talla por tipo de cuidador

Cali – Colombia

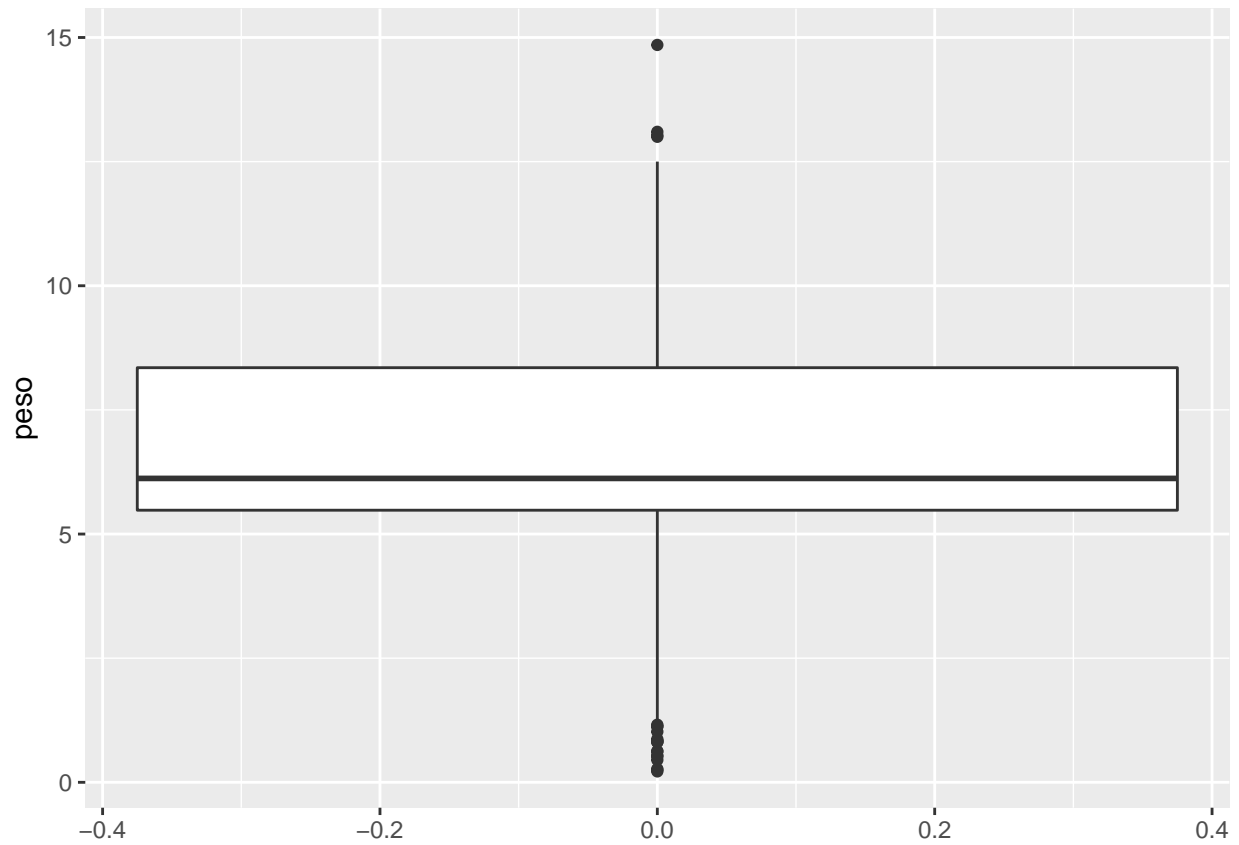


Fuente: Estudio piloto RN2015

Gráfico de cajas y bigotes

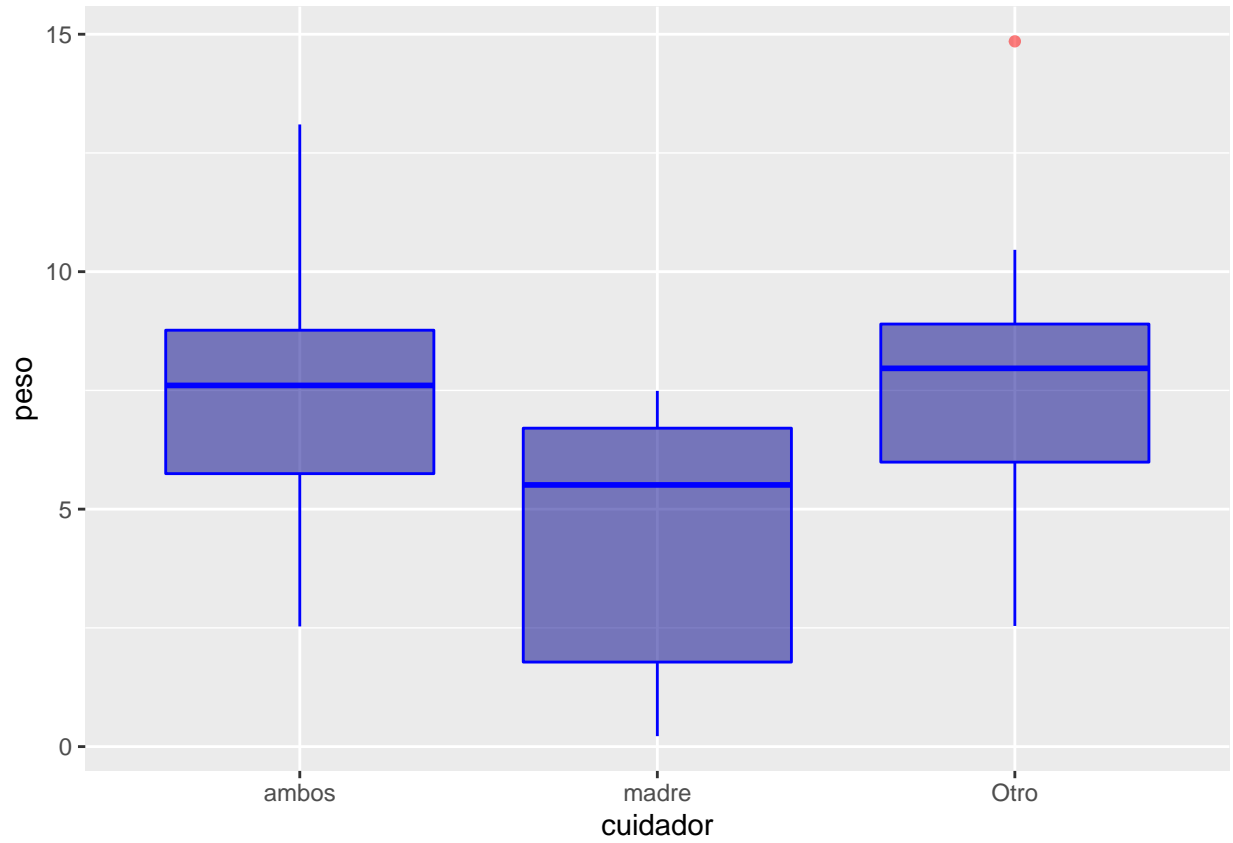
El gráfico de cajas y bigotes para la variable **peso** se puede obtener así:

```
ggplot(data=RN, aes(y=peso)) +  
  geom_boxplot()
```

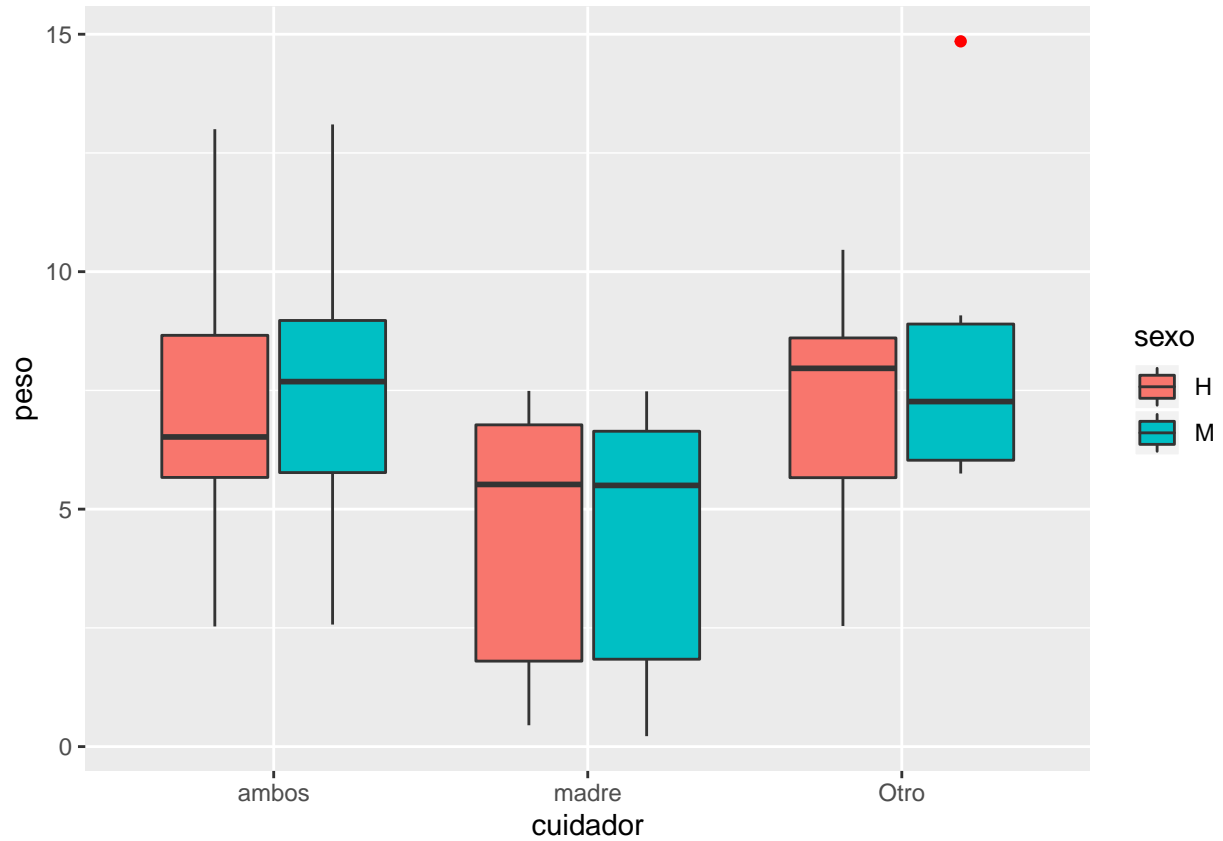
Si en la opción `aes()` colocamos una variable para el eje x, por ejemplo `cuidador`, y agregamos algunos colores obtenemos el siguiente gráfico:

```
ggplot(data=RN,aes(x=cuidador,y=peso))+  
  geom_boxplot(alpha=0.5,fill="darkblue", color="blue",outlier.color="red")
```



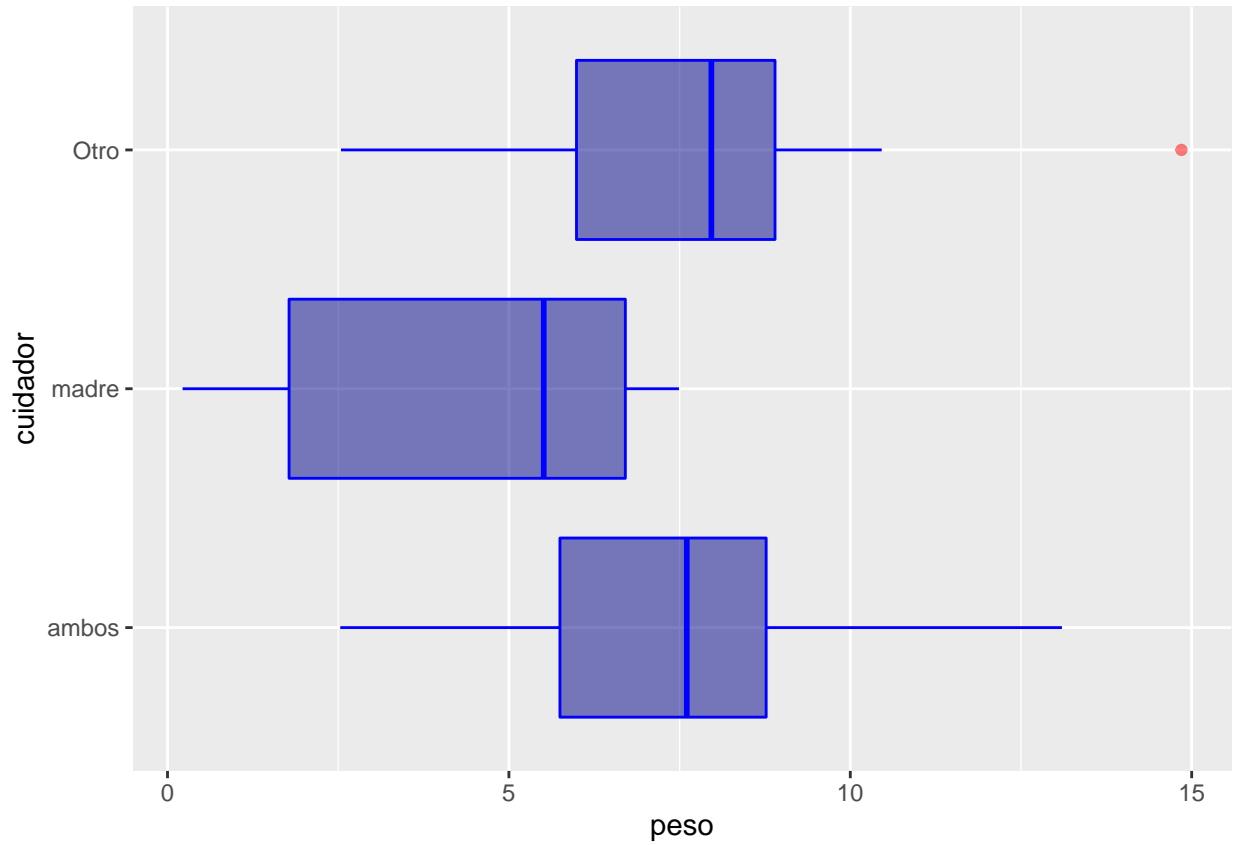
También podemos clasificar por sexo así:

```
ggplot(data=RN,aes(x=cuidador,y=peso,fill = sexo))+  
  geom_boxplot(outlier.color = "red")
```



Retomando el gráfico anterior, podemos cambiar su dirección así:

```
ggplot(data=RN,aes(x=cuidador,y=peso))+
  geom_boxplot(alpha=0.5,fill="darkblue", color="blue",outlier.color="red")+
  coord_flip()
```

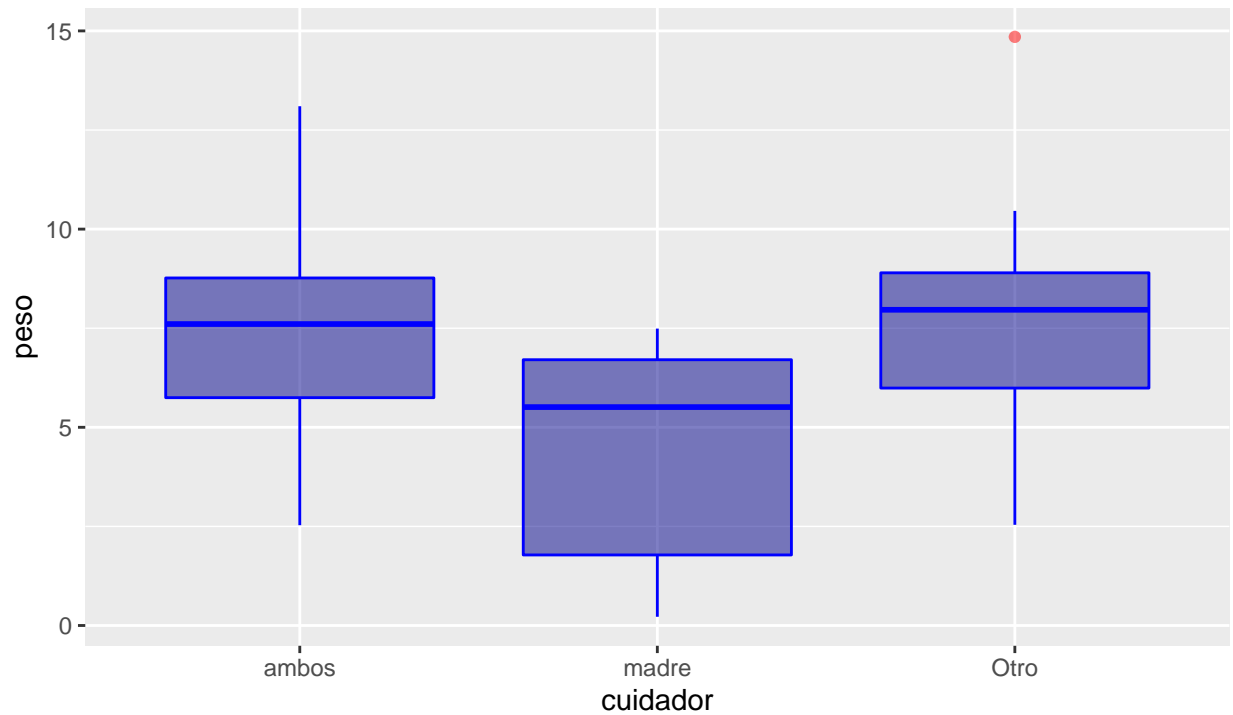


Incluyendo títulos:

```
ggplot(data=RN,aes(x=cuidador,y=peso))+
  geom_boxplot(alpha=0.5,fill="darkblue", color="blue",outlier.color="red")+
  labs(title="Gráfico de cajas y bigotes del peso",subtitle="Cali - Colombia", y="peso", x="cuidador",
  caption="Fuente: Estudio RN2015")
```

Gráfico de cajas y bigotes del peso

Cali – Colombia

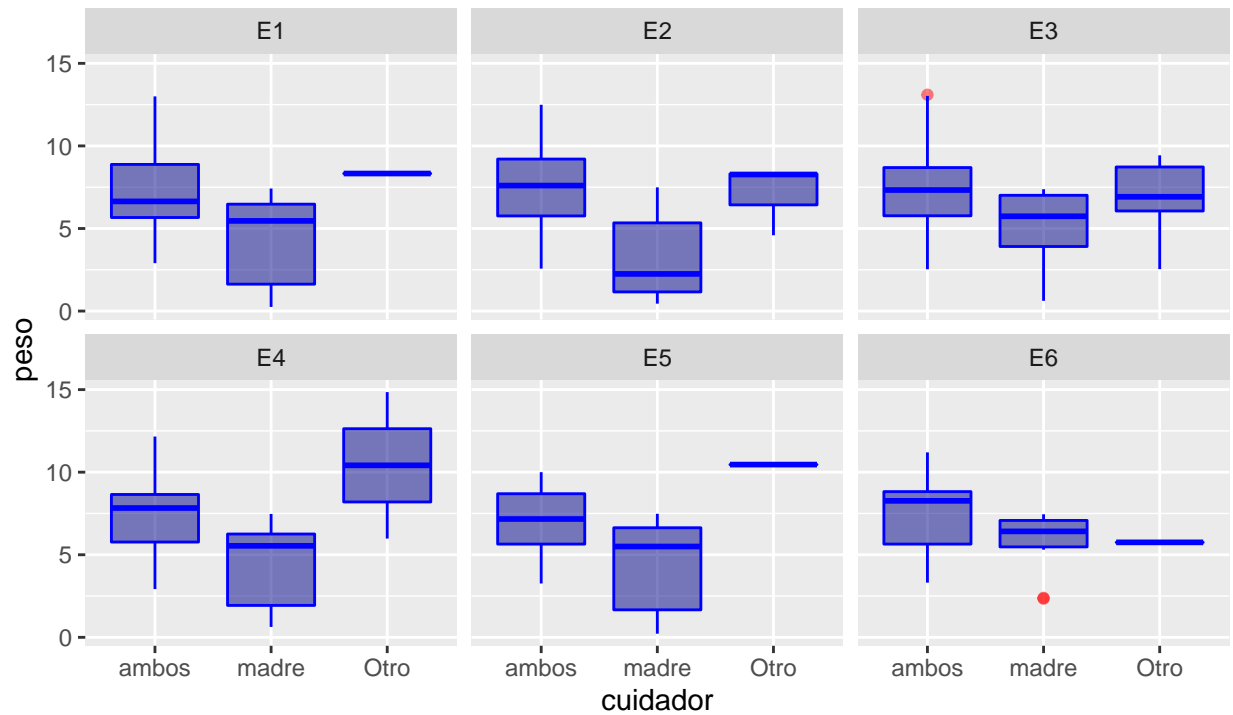


Fuente: Estudio RN2015

Ahora replicar este gráfico por subgrupos de población, por ejemplo utilizando la variable estrato.

```
ggplot(data=RN,aes(x=cuidador,y=peso))+  
  geom_boxplot(alpha=0.5,fill="darkblue", color="blue",outlier.color="red")+  
  labs(title="Gráfico de cajas y bigotes del peso por estratos",subtitle="Cali - Colombia", y="peso",  
        x="cuidador", caption="Fuente: Estudio RN2015")+  
  facet_wrap(~estrato)
```

Gráfico de cajas y bigotes del peso por estratos Cali – Colombia



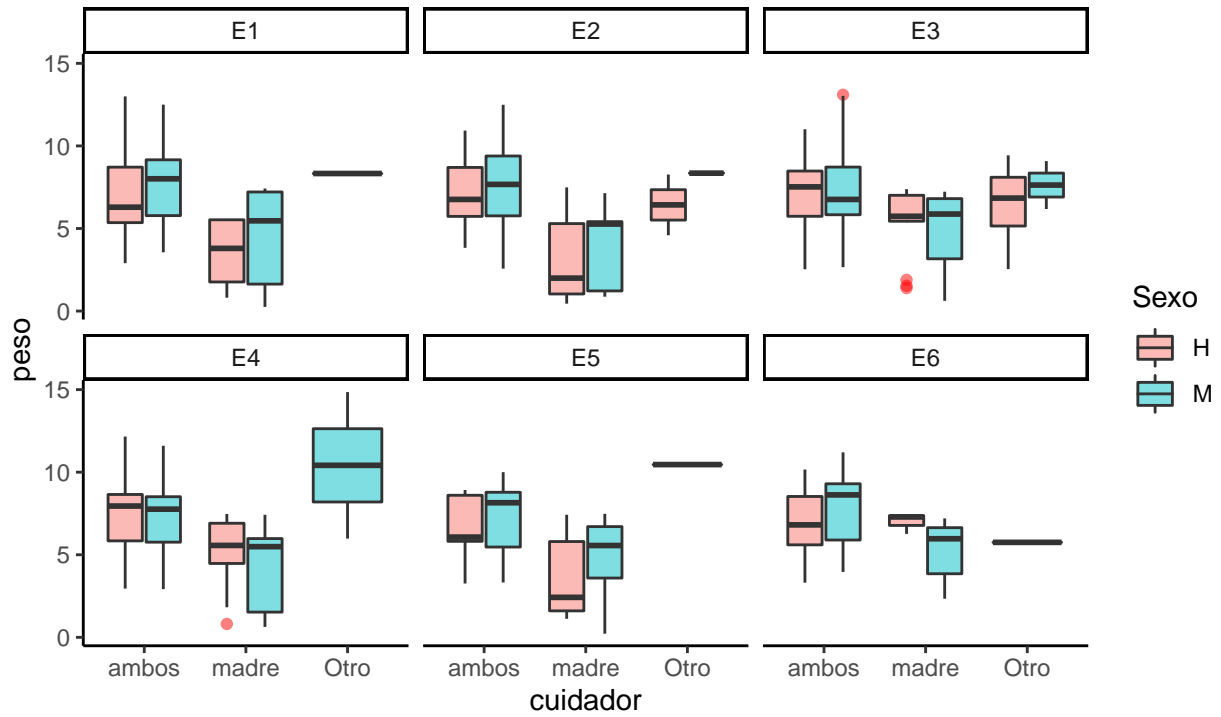
Fuente: Estudio RN2015

utilizando el gráfico anterior incluyendo la variable sexo (Y un tema distinto), tenemos:

```
ggplot(data=RN,aes(x=cuidador,y=peso,fill=factor(sexo))) +
  geom_boxplot(alpha=0.5,outlier.color="red") +
  labs(title="Gráfico de cajas y bigotes del peso por estratos",subtitle="Cali - Colombia", y="peso",
        x="cuidador", fill="Sexo", caption="Fuente: Estudio RN2015") +
  facet_wrap(~estrato)+
  theme_classic()
```

Gráfico de cajas y bigotes del peso por estratos

Cali – Colombia

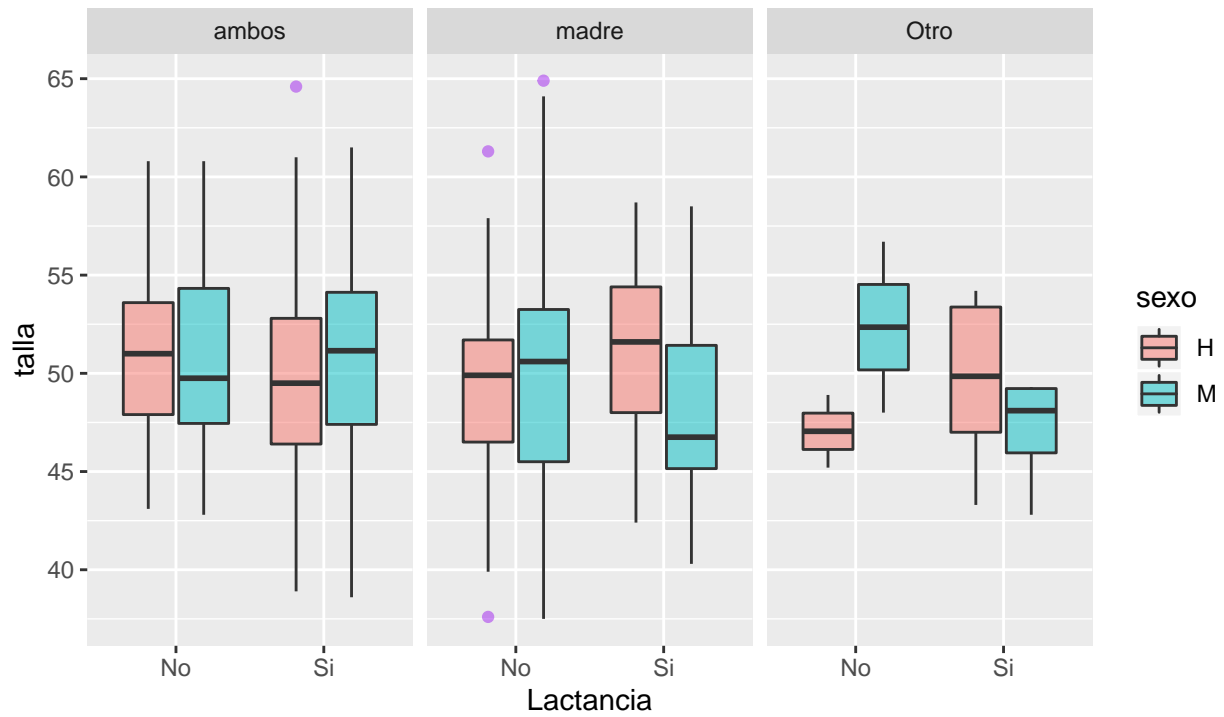


Fuente: Estudio RN2015

- Ejercicio:

Reproduzca el siguiente gráfico (no es necesario que coincidan los colores):

Gráfico de cajas y bigotes de talla por sexo y tipo de cuidador Cali – Colombia

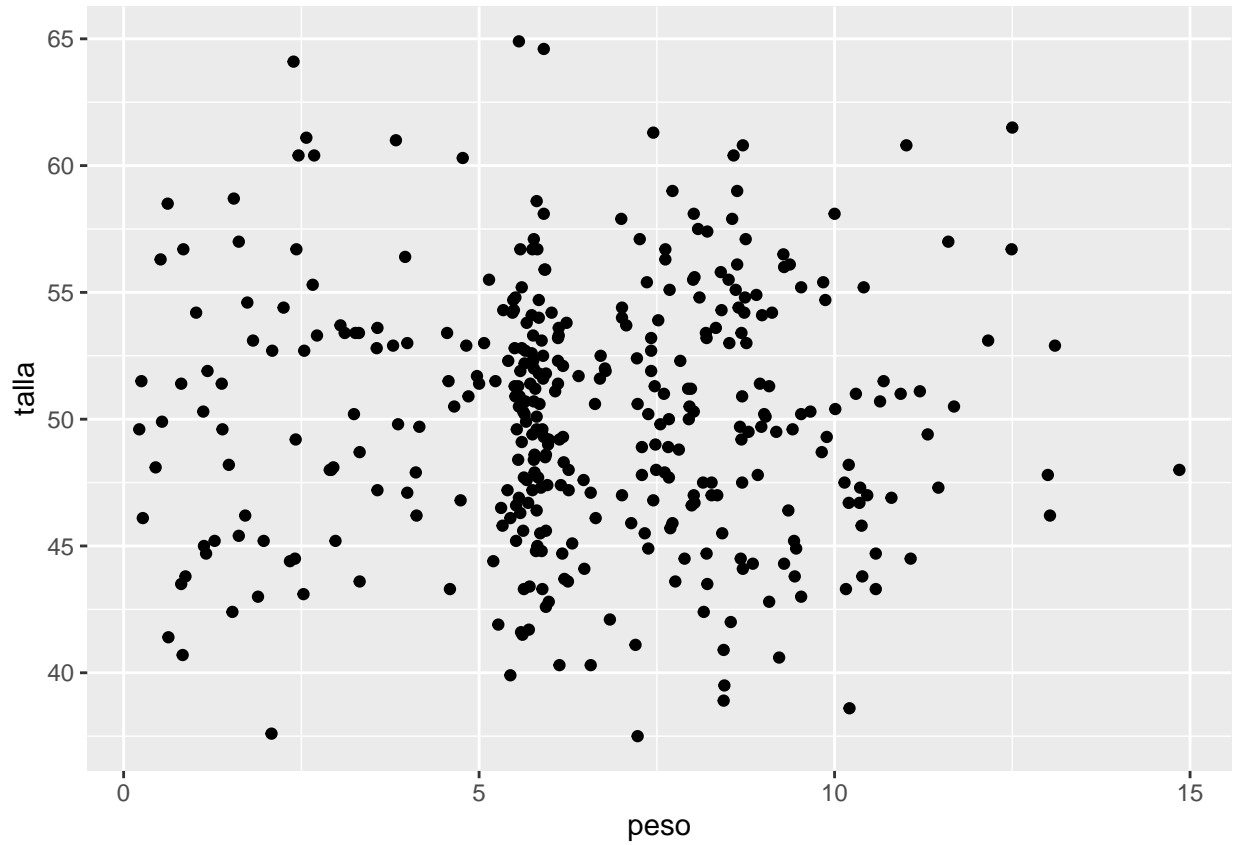


Fuente: Estudio RN2015

Gráfico de dispersión

Para este tipo de gráfico trabajaremos con las variables peso y talla. El primer gráfico de dispersión es:

```
ggplot(data=RN, aes(x=peso,y=talla))+  
  geom_point()
```

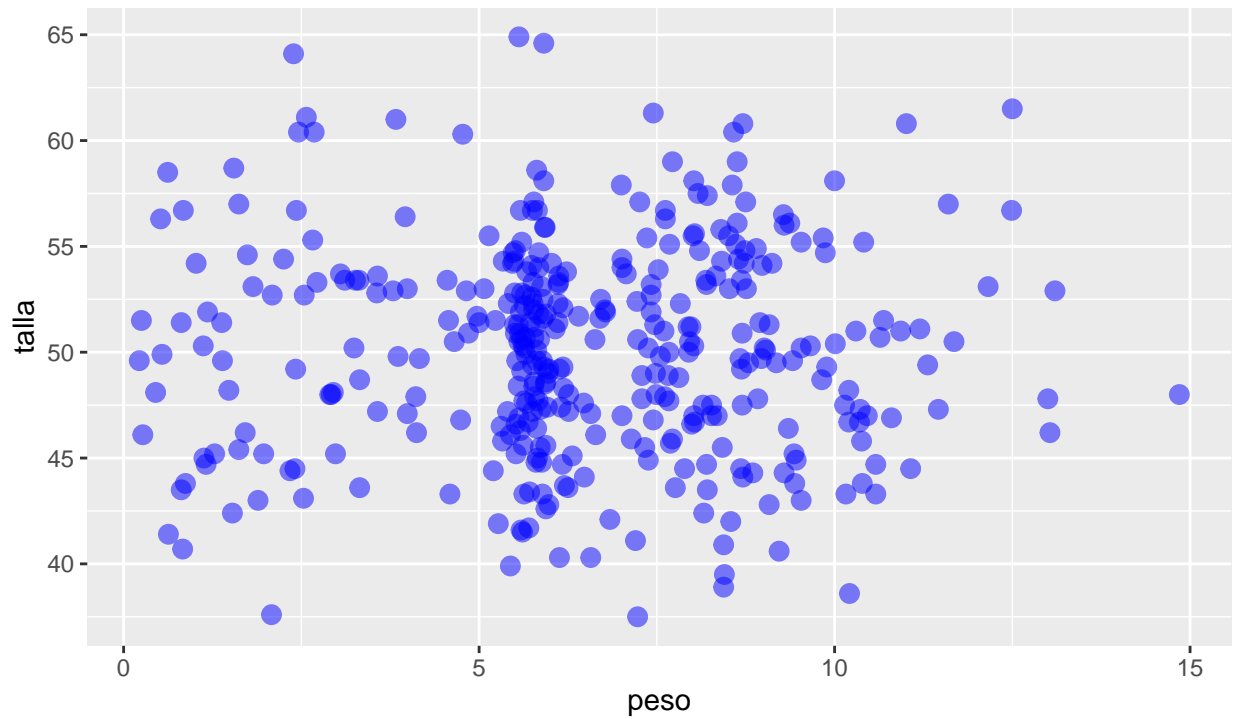



Incluyendo títulos y cambio de colores:

```
ggplot(data=RN, aes(x=peso,y=talla))+  
  geom_point(alpha=0.5,color="blue",size=3)+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estu
```

peso versus talla

Cali – Colombia



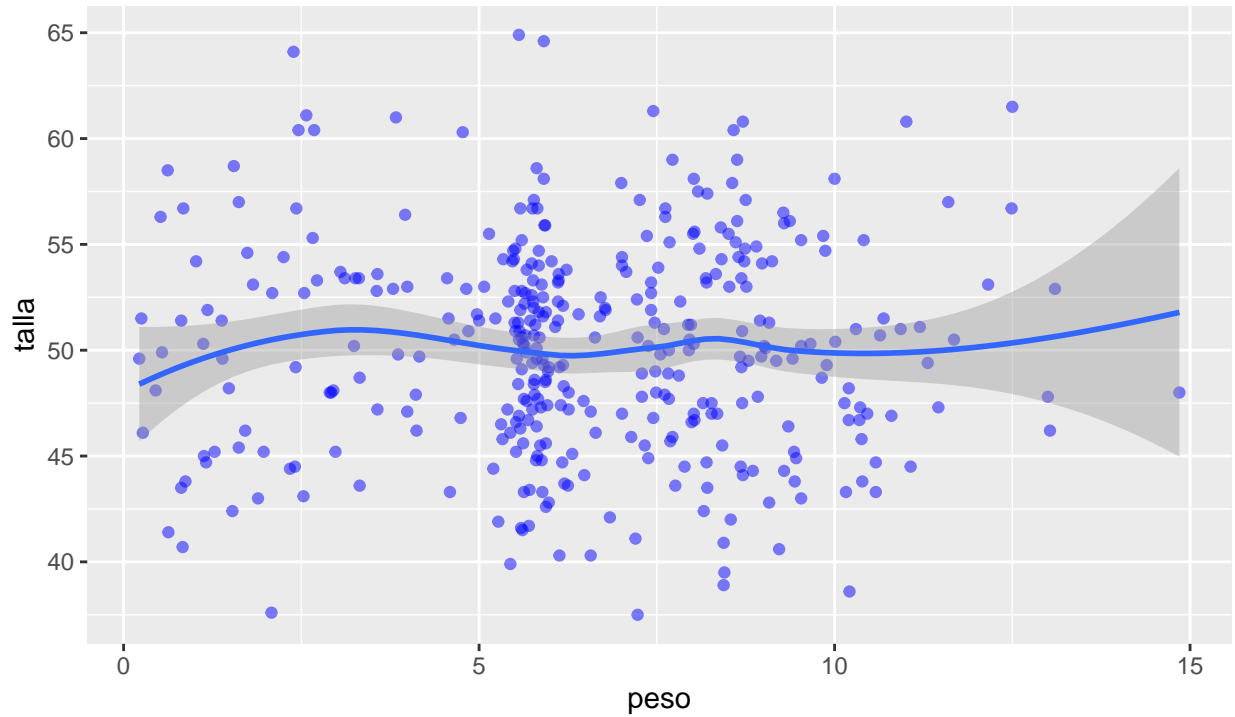
Fuente: Estudio RN2015

Incluimos un suavizamiento a este comportamiento para poder explorar la presencia de alguna relación.

```
ggplot(data=RN, aes(x=peso,y=talla))+  
  geom_point(alpha=0.5,color="blue")+  
  geom_smooth()+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estudio RN2015")  
  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

peso versus talla

Cali – Colombia



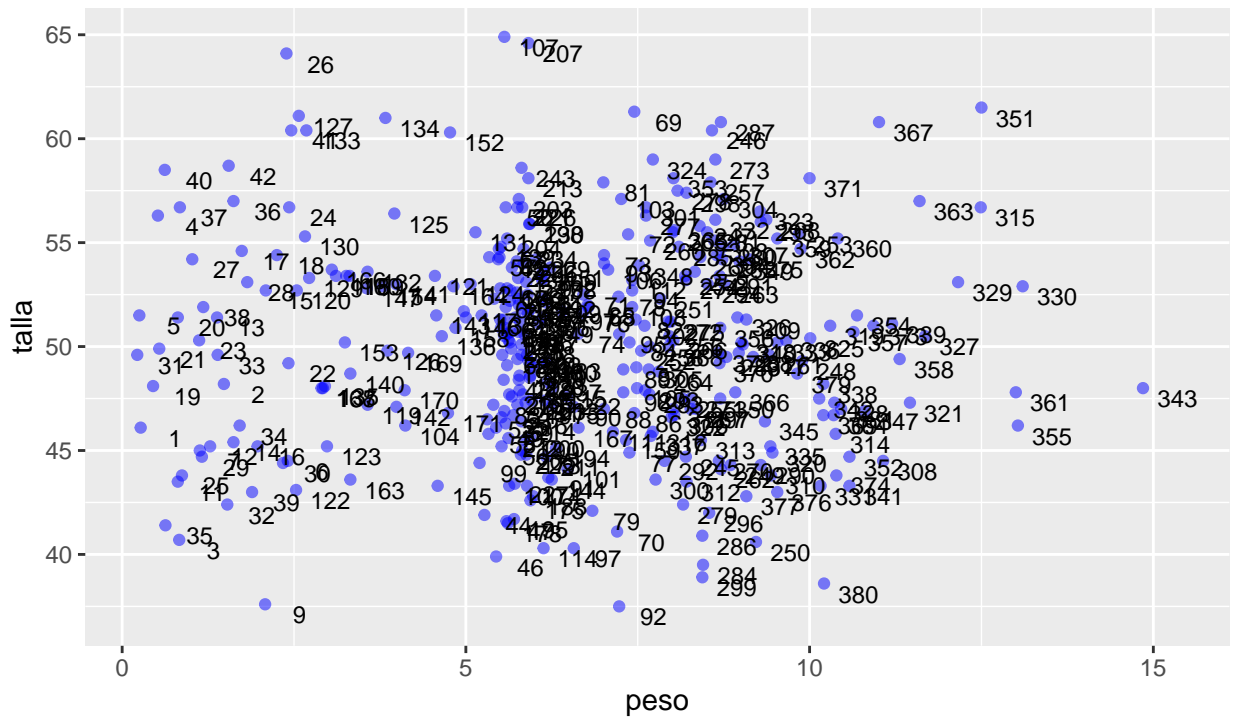
Fuente: Estudio RN2015

agregar el id a cada punto:

```
ggplot(data=RN, aes(x=peso,y=talla))+  
  geom_point(alpha=0.5,color="blue")+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estudio RN2015")  
  geom_text(aes(label=id), size=3,nudge_x=0.5,nudge_y = -0.5)
```

peso versus talla

Cali – Colombia



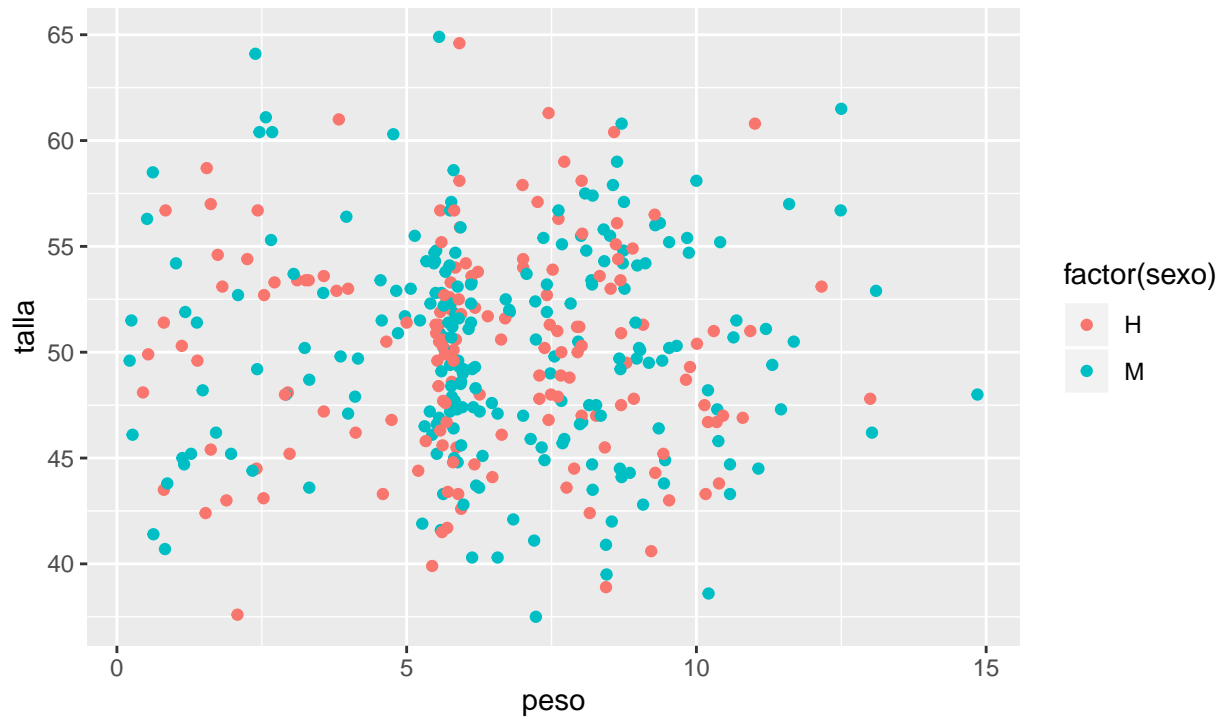
Fuente: Estudio RN2015

veamos ahora clasificando por sexo

```
ggplot(data=RN,aes(x=peso,y=talla,color=factor(sexo)))+  
  geom_point()+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estu
```

peso versus talla

Cali – Colombia



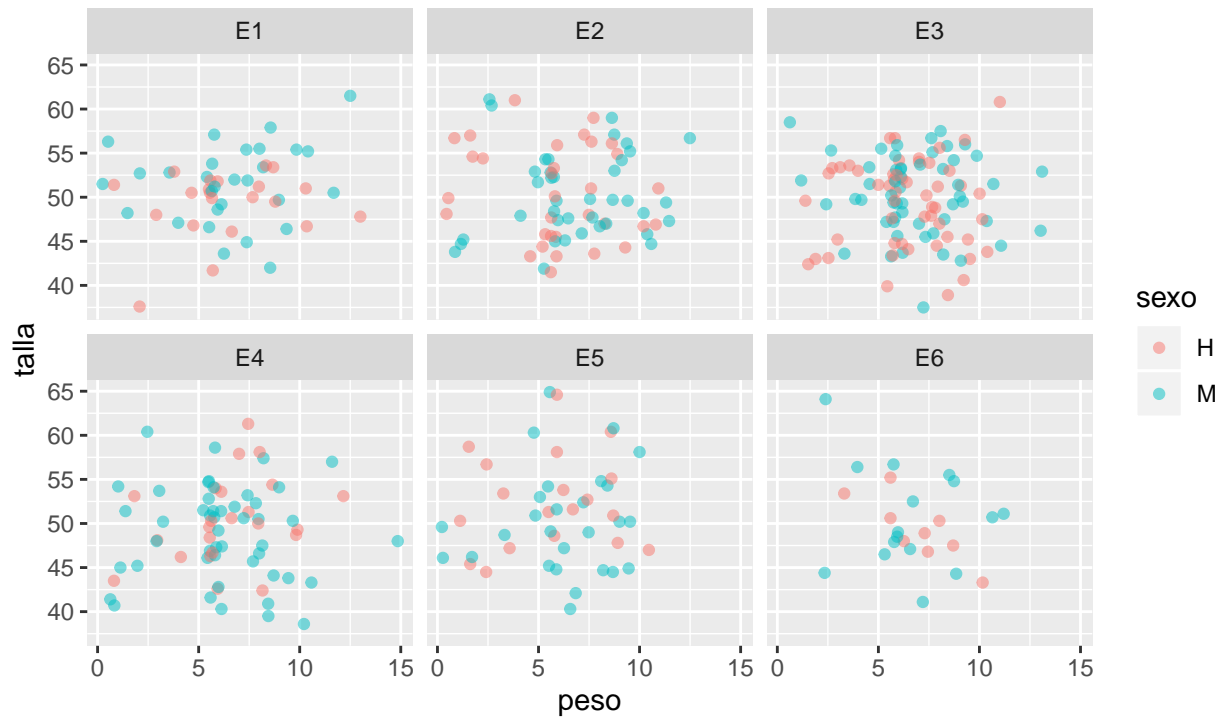
Fuente: Estudio RN2015

y por subgrupos.

```
ggplot(data=RN,aes(x=peso,y=talla, color=sexo))+  
  geom_point(alpha=0.5)+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estu  
  facet_wrap(~estrato)
```

peso versus talla

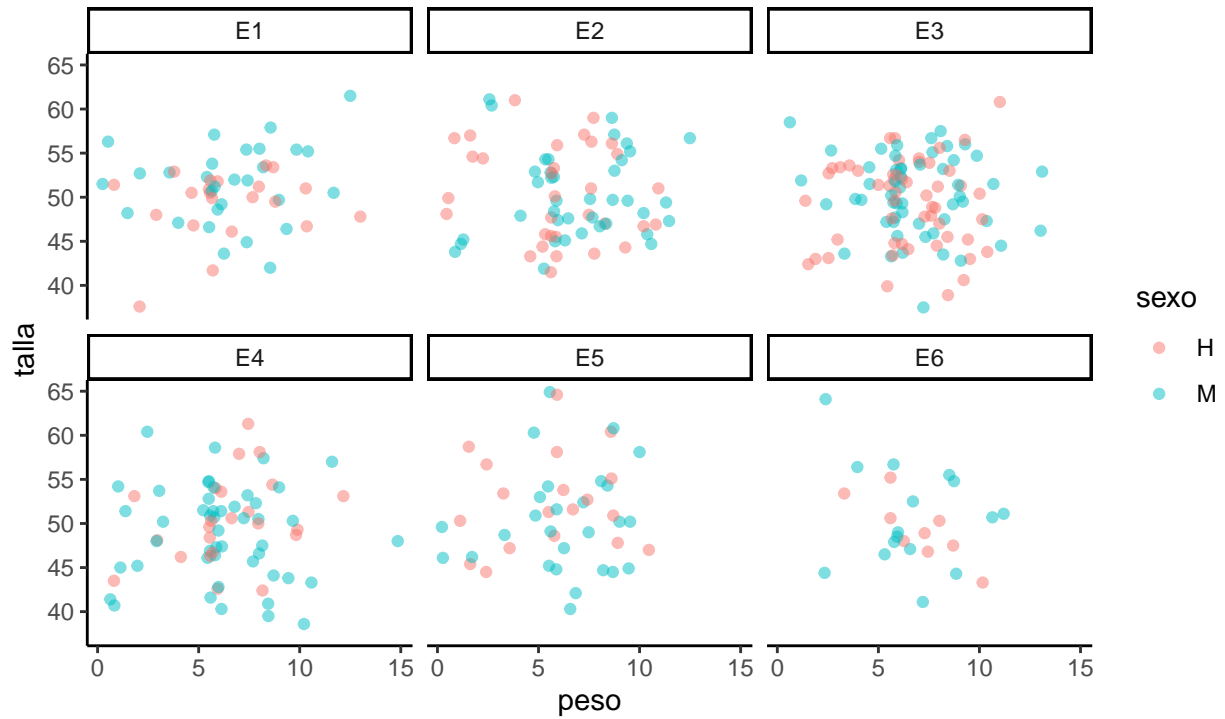
Cali – Colombia



veamos con algún tema diferente:

```
ggplot(data=RN,aes(x=peso,y=talla, color=sexo))+  
  geom_point(alpha=0.5)+  
  labs(title="peso versus talla",subtitle="Cali - Colombia", y="talla", x="peso", caption="Fuente: Estu  
  facet_wrap(~estrato)+  
  theme_classic()
```

peso versus talla
Cali – Colombia



Fuente: Estudio RN2015

Ejercicio:

Reproduzca el siguiente gráfico (no es necesario que coincidan los colores):

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Talla versus sueño

Cali – Colombia

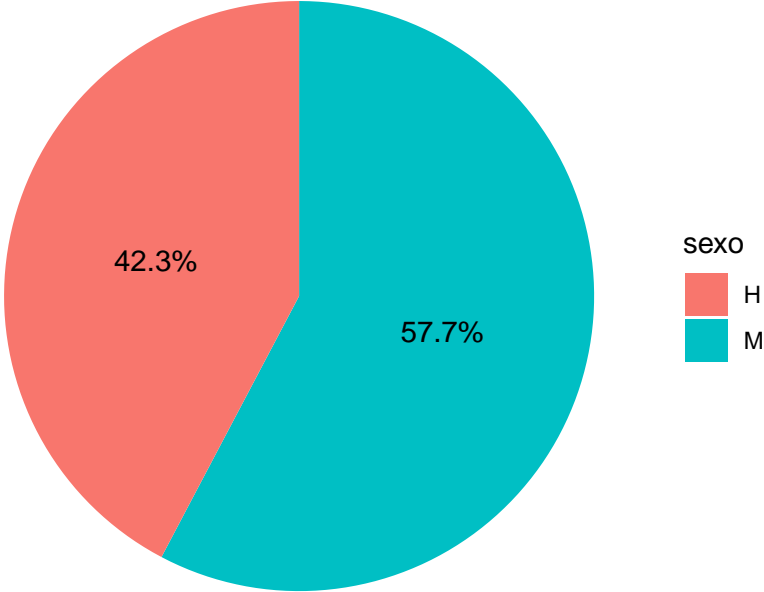


Fuente: Estudio RN2015

Pie o torta

```
ggplot(RN, aes(x="", fill=sexo))+  
  geom_bar()+coord_polar("y")+  
  geom_text(aes(label=scales::percent(..count../sum(..count..))), stat="count", position=position_stack(0))  
  labs(title="Distribución de recién nacidos por sexo", subtitle="Cali - Colombia",  
        caption="Fuente: Estudio RN2015")+  
  theme_void()
```


Distribución de recién nacidos por sexo
Cali – Colombia



Fuente: Estudio RN2015