

R-PL6

Gabriel López Cuenca, Sergio Sanz Sacristán, Álvaro Zamorano Ortega

10 de diciembre de 2019

1. Visualización en diagrama de barras.

1.1. Salarios medios mensuales en Europa.

Hemos hecho un tratamiento de datos correspondiente a los salarios medios mensuales de los países de la Unión Europea. Lo que buscamos es representar el conjunto de datos de estos salarios medios, para localizar, de forma clara y sencilla la posición actual de España con respecto al resto de los países estudiados.

Lo primero que realizamos es importar los datos correspondientes al estudio:

```
> install.packages("readr")
> library("readr")
> datos <- read_csv("países.csv")
> datos <- data.frame(datos)
```

Para que el diagrama quede lo más claro posible vamos a realizar un **diagrama de Pareto**. Es un tipo especial de gráfica de barras donde los valores graficados están organizados de mayor a menor. Se utiliza para identificar los defectos que se producen con mayor frecuencia, las causas más comunes de los defectos o las causas más frecuentes de quejas de los clientes. Por lo tanto, ordenamos los datos de mayor a menor haciendo uso de la función `factor`.

```
> datos$Pais <- factor(datos$Pais, levels=datos$Pais[order(datos$Salario)])
```

Para representar el diagrama utilizaremos el paquete **ggplot2**. Lo instalamos y lo cargamos.

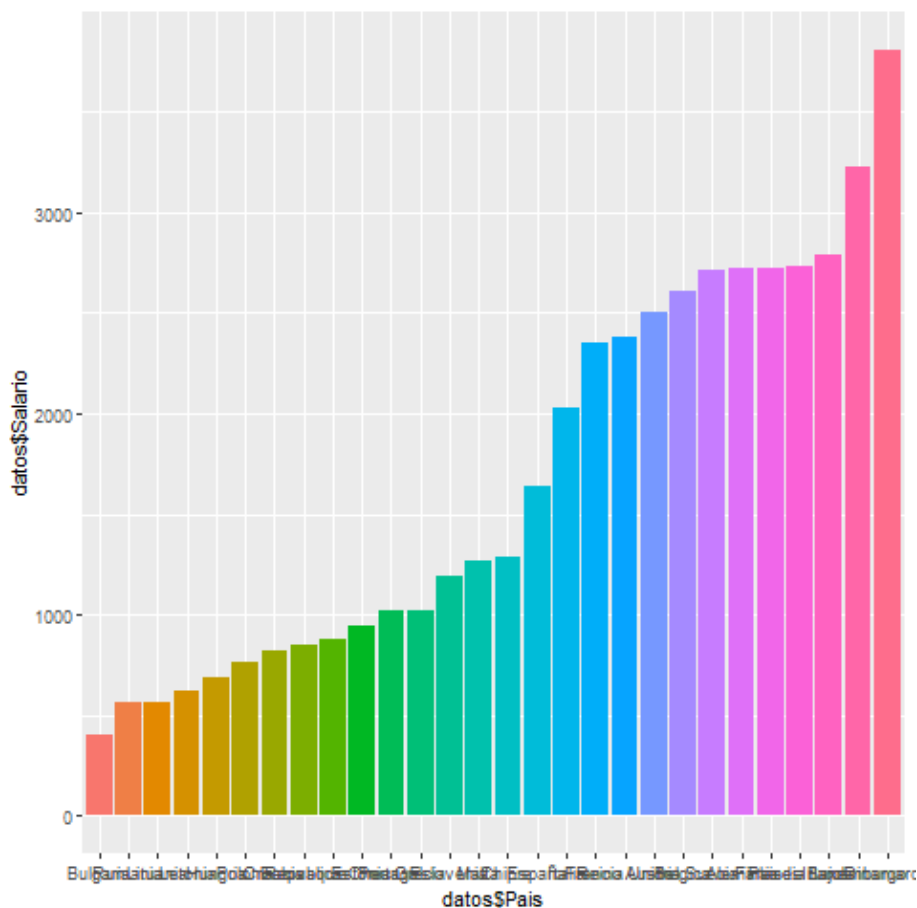
```
> install.packages("ggplot2")
> library("ggplot2")
```

Inicializamos el diagrama e introducimos los datos. Inicialmente nos saldrán los datos en barras verticales, y de diferentes colores al llamar a la función `geom_bar`.

```

> sp <- ggplot()
> sp <- sp + geom_bar(data=datos,aes(x=datos$Pais, y=datos$Salario,
+   fill=datos$Pais), stat='identity', position='dodge',
+   show.legend=FALSE)
> source("./Funciones/diagrama.R")
> diagrama("diagrama1.png",sp)

```

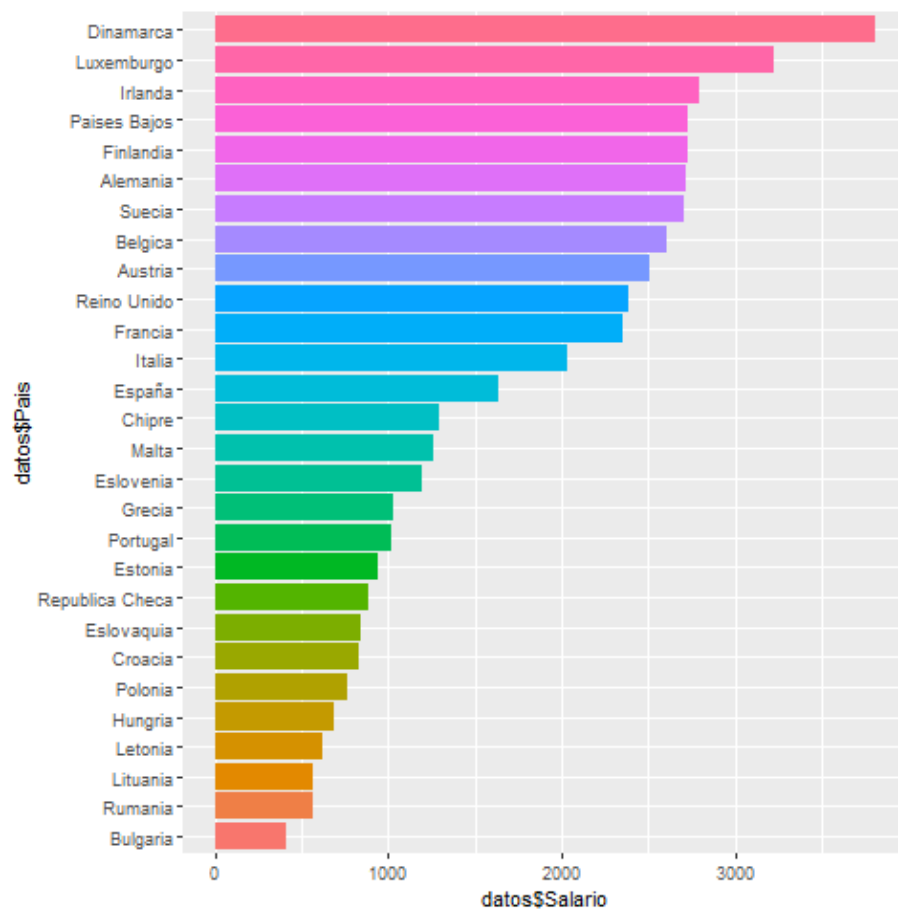


Ya que los datos son identificados por nombre, es recomendable que estos se coloquen de la forma más visible posible. Por lo tanto, lo que hacemos es **rotar** el diagrama para que queden los nombres en el eje y, y así sean mucho mas visibles. Pasamos las barras de vertical a horizontal.

```

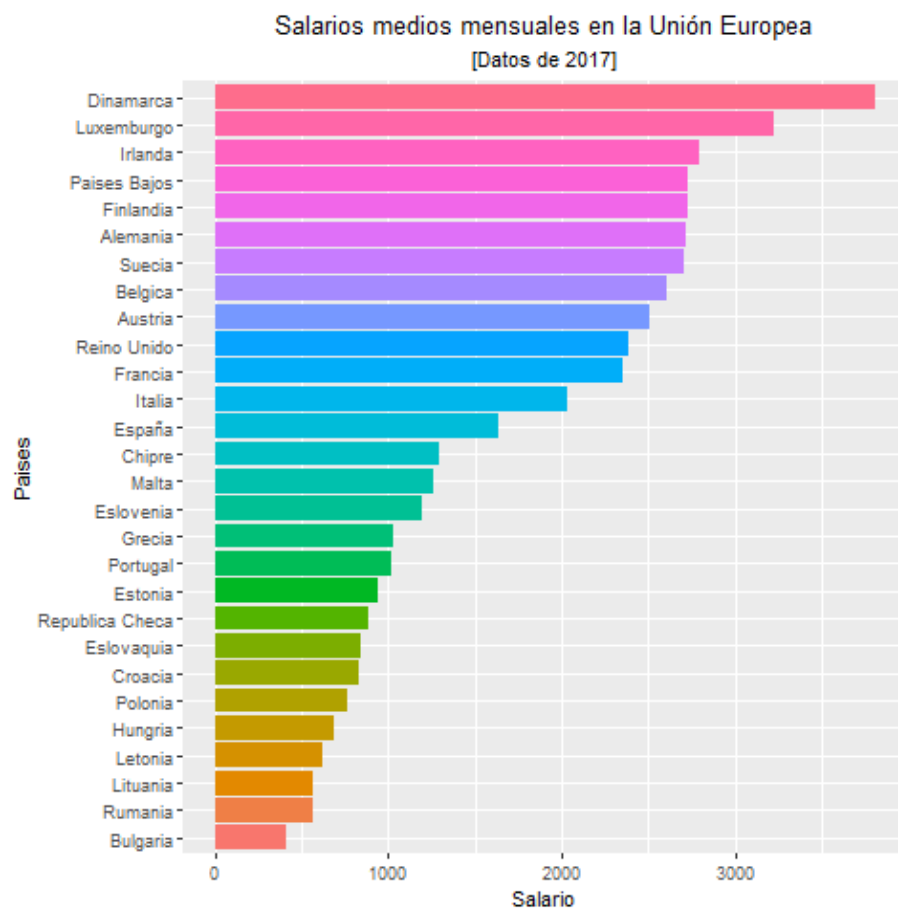
> sp <- sp + coord_flip()
> source("./Funciones/diagrama.R")
> diagrama("diagrama2.png",sp)

```



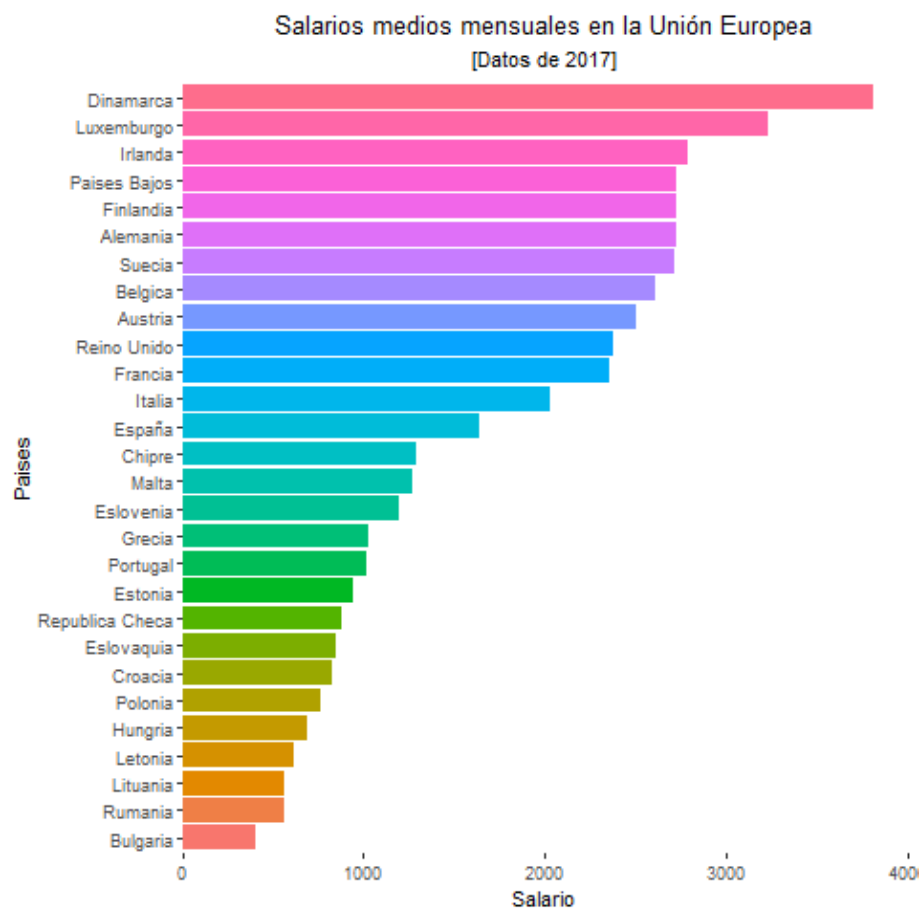
Ahora añadimos las **etiquetas** y los **títulos** de cada uno de los ejes de nuestro diagrama y el título del mismo.

```
> sp <- sp + labs(x = "Países",y="Salario")
> sp <- sp + ggtitle(label = "Salarios medios mensuales en la Unión Europea",
+   subtitle="[Datos de 2017]") +
+   theme(plot.title = element_text(hjust = 0.5),
+   plot.subtitle = element_text(hjust = 0.5))
> source("./Funciones/diagrama.R")
> diagrama("diagrama3.png",sp)
```



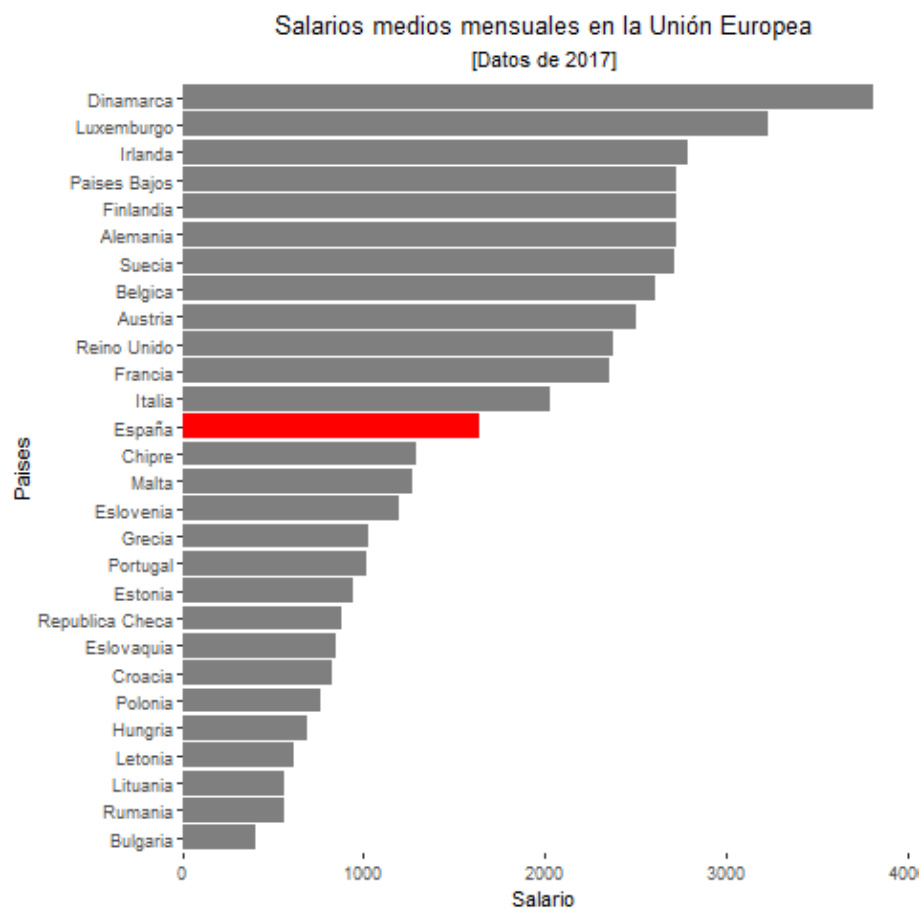
Para que quede de una forma mucho más visual hemos decidido retirar el fondo gris predeterminado y ponerlo blanco. Además **ajustamos** el eje y para que no quede espacio entre dicho eje, y las barras del diagrama usando `scale_y_continuous`.

```
> sp <- sp + theme(panel.background = element_rect(fill = "white"))
> sp <- sp + scale_y_continuous(limits = c(0,4000) ,expand = c(0, 0))
> source("./Funciones/diagrama.R")
> diagrama("diagrama4.png",sp)
```



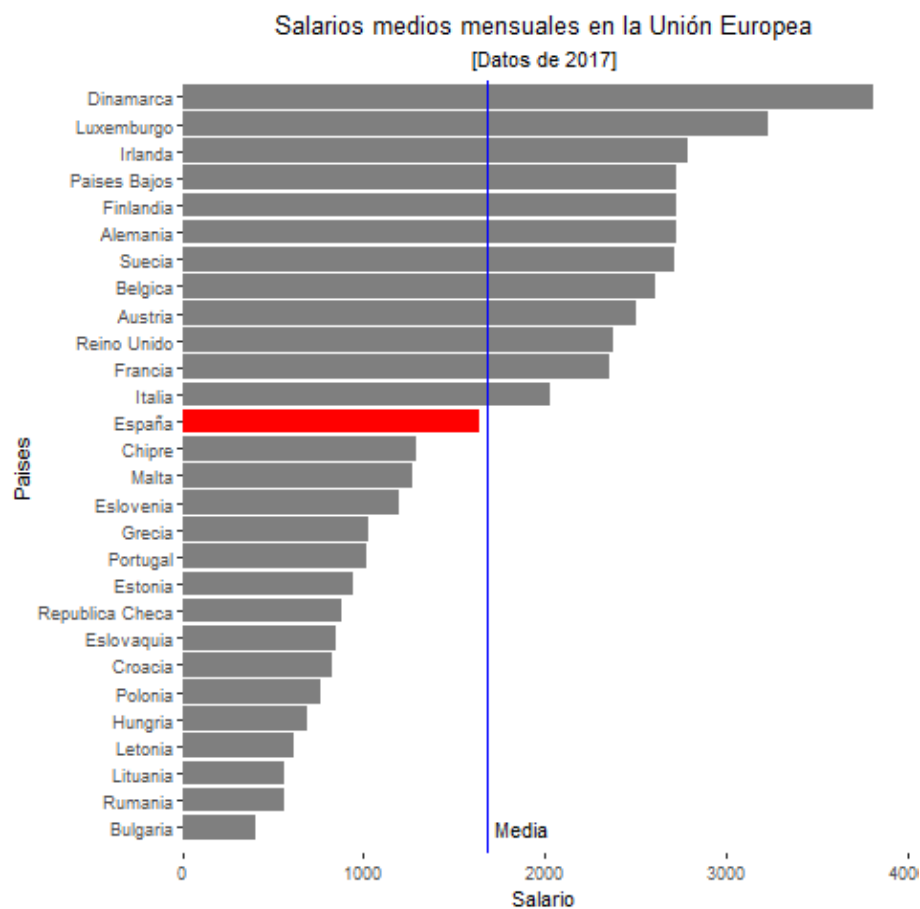
Ahora lo que hacemos es dar el mismo **color** a todas las barras, excepto a la barra de España. Ya que nos interesa su posición respecto al resto de países, por lo que solo destacamos esta barra.

```
> sp <- sp + scale_fill_manual(values = c("grey50","grey50","grey50",
+     "grey50","grey50","grey50","grey50","grey50","grey50",
+     "grey50","grey50","grey50","grey50","grey50",
+     "grey50","red","grey50","grey50","grey50","grey50",
+     "grey50","grey50","grey50","grey50","grey50","grey50",
+     "grey50","grey50"))
> source("./Funciones/diagrama.R")
> diagrama("diagrama5.png",sp)
```



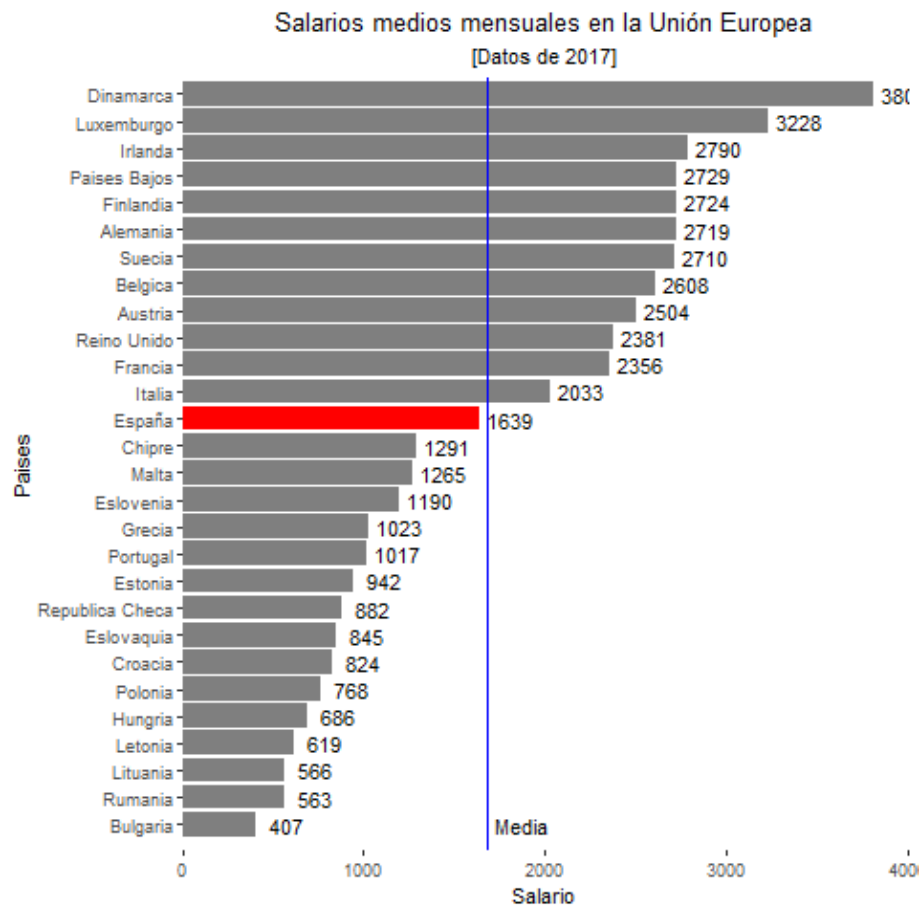
Ahora añadimos una línea que indicará en que lugar queda la **media** de salarios medios de los países europeos con `geom_hline`.

```
> sp <- sp + geom_hline(aes(yintercept=mean(datos$Salario)),color = "blue") +
+   geom_text(aes(0,mean(datos$Salario),
+   label = "Media", vjust = -1))
> source("./Funciones/diagrama.R")
> diagrama("diagrama6.png",sp)
```



Por último añadimos al lado de cada barra cuál es el **valor exacto** de cada dato. Así queda mucho más claro cuales son los datos que estamos tratando, se hace con la función `geom_text`.

```
> sp <- sp + geom_text(aes(datos$Pais,datos$Salario),
+   label=sprintf("%d",datos$Salario))
> source("./Funciones/diagrama.R")
> diagrama("diagrama7.png",sp)
```



2. Visualización en mapa.

Para comprender mejor la diferencia de los salarios entre los diferentes países de la Unión Europea, vamos a realizar una visualización en el **mapa europeo**. Para ello, utilizaremos los paquetes `ggplot2`, `maps`, `grid` y `rworldmap`.

```
> install.packages("maps")
> install.packages("rworldmap")
> library(maps)
> library(rworldmap)
```

En primer lugar, leeremos un `.csv` igual que el anterior pero con los nombres de los países en inglés, ya que los paquetes anteriores así lo necesitan.

```
> datos <- read_csv("countries.csv")
> datos <- data.frame(datos)
```

En segundo lugar se obtienen diferentes datos de **localización** a partir de las funciones antes mencionadas.


```

> # Obtenemos el mapa del mundo
> worldMap <- getMap()
> # Países de la Unión Europea
> europeanUnion <- datos$Pais
> # Seleccionamos los índices de los países europeos
> indEU <- which(worldMap$NAME%in%europeanUnion)

```

En tercer lugar, obtenemos las **coordenadas** de cada uno de los países.

```

> # Longitud y latitud de los bordes de los países
> europeCoords <- lapply(indEU, function(i){
+   df <- data.frame(worldMap@polygons[[i]]@Polygons[[1]]@coords)
+   df$region =as.character(worldMap$NAME[i])
+   colnames(df) <- list("long", "lat", "region")
+   return(df)
+ })
> europeCoords <- do.call("rbind", europeCoords)

```

En cuarto lugar, guardamos los salarios en un vector y cada uno de ellos se **relaciona** con las coordenadas del país correspondiente.

```

> value <- datos$Salario
> europeanUnionTable <- data.frame(country = europeanUnion, value = value)
> europeCoords$value <- europeanUnionTable$value[match(europeCoords$region,
+   europeanUnionTable$country)]

```

Por último, **mostramos** por pantalla el mapa acompañado de una leyenda con la escala de salarios. A su vez, elegimos los colores de la escala de **blanco** (salario bajo) a **negro** (salario alto) ya que se consigue una mejor visualización que con cualquier otra escala de colores. Además, añadimos el título y eliminamos el fondo.

```

> P <- ggplot() + geom_polygon(data = europeCoords, aes(x = long, y = lat,
+   group = region, fill = value),
+   colour = "black", size = 0.1) +
+   coord_map(xlim = c(-13, 35), ylim = c(32, 71))
> P <- P + scale_fill_gradient(name = "Salario Medio", low = "#FFFFFF",
+   high = "#000000", na.value = "grey50")
> P <- P + theme(axis.text.x = element_blank(),
+   axis.text.y = element_blank(), axis.ticks.x = element_blank(),
+   axis.ticks.y = element_blank(), axis.title = element_blank(),
+   panel.background = element_rect(fill = "white"),
+   plot.margin = unit(0 * c(-1.5, -1.5, -1.5, -1.5), "lines"))
> P <- P + ggtitle(label = "Salarios medios mensuales en la Unión Europea",
+   subtitle="[Datos de 2017]") + theme(plot.title = element_text(hjust = 0.5),
+   plot.subtitle = element_text(hjust = 0.5))
> diagrama("mapa.png",P)

```

Salarios medios mensuales en la Unión Europea
[Datos de 2017]

