

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
library( reshape )
```

```
## Warning: package 'reshape' was built under R version 3.4.4
```

Exploración y Correlaciones

```
madrid<-read.csv2("PEC3HiloArgumental.csv",sep=";",dec=".",na.strings = "NA")  
#Mostramos las primeras filas  
head(madrid)
```

```
##  CodigoDistrito      NombreDistrito NumeroBarrios SuperficieDistrito  
## 1              1 CENTRO                      6          5231264  
## 2              2 ARGANZUELA                  7          6466400  
## 3              3 RETIRO                      6          5465307  
## 4              4 SALAMANCA                  6          5387243  
## 5              5 CHAMARTIN                  6          9170122  
## 6              6 TETUAN                     6          5388121  
##  ConsumoAgua Natalidad Mortalidad IndicePoblacionActiva Poblacion  
## 1      8256783      7.32      7.93              36.16      132352  
## 2      6588983      9.11      8.36              48.44      152907  
## 3      5563420      8.39     10.02              43.54      119011  
## 4      7919520      8.25     10.32              47.25      145268  
## 5      7687755      9.95      9.66              50.87      144894  
## 6      7063199      9.18      9.35              54.91      155967  
##  NumeroCentroSanitarios NumeroAparcabicis NumeroPuntosLimpios  
## 1                      7              75              24  
## 2                      9              45              0  
## 3                      4              40              0  
## 4                      8              44              0  
## 5                      7              53              0  
## 6                      6              40              0  
##  NumeroPuntosRecargas NumeroInmuebles SuperficieConstruida  
## 1                      4              20498          7862531  
## 2                      2              28163          6763803  
## 3                      4              41859          6373753  
## 4                      7              47213          9522664  
## 5                      3              53999          8548212  
## 6                      0              54365          7121027  
##  HectareasVerdes Poblacion.1 X22  
## 1          44.87      132352  15  
## 2          129.71      152907   9  
## 3          154.15      119011  16  
## 4           43.42      145268  11  
## 5           56.13      144894  12  
## 6           65.30      155967   8
```

Observamos el tipo de variable que R ha asignado, es habitual que no asigne el adecuado.

```
sapply(madrid, class)
```

```
##      CodigoDistrito      NombreDistrito      NumeroBarrios  
##      "integer"      "factor"      "integer"  
##      SuperficieDistrito      ConsumoAgua      Natalidad
```

```
##          "integer"          "integer"          "numeric"
##      Mortalidad  IndicePoblacionActiva          Poblacion
##          "numeric"          "numeric"          "integer"
## NumeroCentroSanitarios      NumeroAparcabicis      NumeroPuntosLimpios
##          "integer"          "integer"          "integer"
##      NumeroPuntosRecargas      NumeroInmuebles      SuperficieConstruida
##          "integer"          "integer"          "integer"
##      HectareasVerdes      Poblacion.1          X22
##          "numeric"          "integer"          "integer"
```

La variable CodigoDistrito debe ser cualitativa.

```
madrid$Coddistri=as.factor(madrid$CodigoDistrito)
```

Comprobamos que variables tiene datos perdidos

```
sapply(madrid,function(x) sum(is.na(x)))
```

```
##      CodigoDistrito      NombreDistrito      NumeroBarrios
##           0           0           0
##      SuperficieDistrito      ConsumoAgua      Natalidad
##           0           0           0
##      Mortalidad  IndicePoblacionActiva      Poblacion
##           0           0           0
## NumeroCentroSanitarios      NumeroAparcabicis      NumeroPuntosLimpios
##           0           0           0
##      NumeroPuntosRecargas      NumeroInmuebles      SuperficieConstruida
##           0           0           0
##      HectareasVerdes      Poblacion.1      X22
##           0           0           0
##      Coddistri
##           0
```

Eliminamos los espacios en blanco existentes en la variable

```
madrid$NombreDistrito<-gsub(" ","",madrid$NombreDistrito)
summary(madrid$NombreDistrito)
```

```
##      Length      Class      Mode
##           21 character character
```

Comenzamos con los primeros estudios

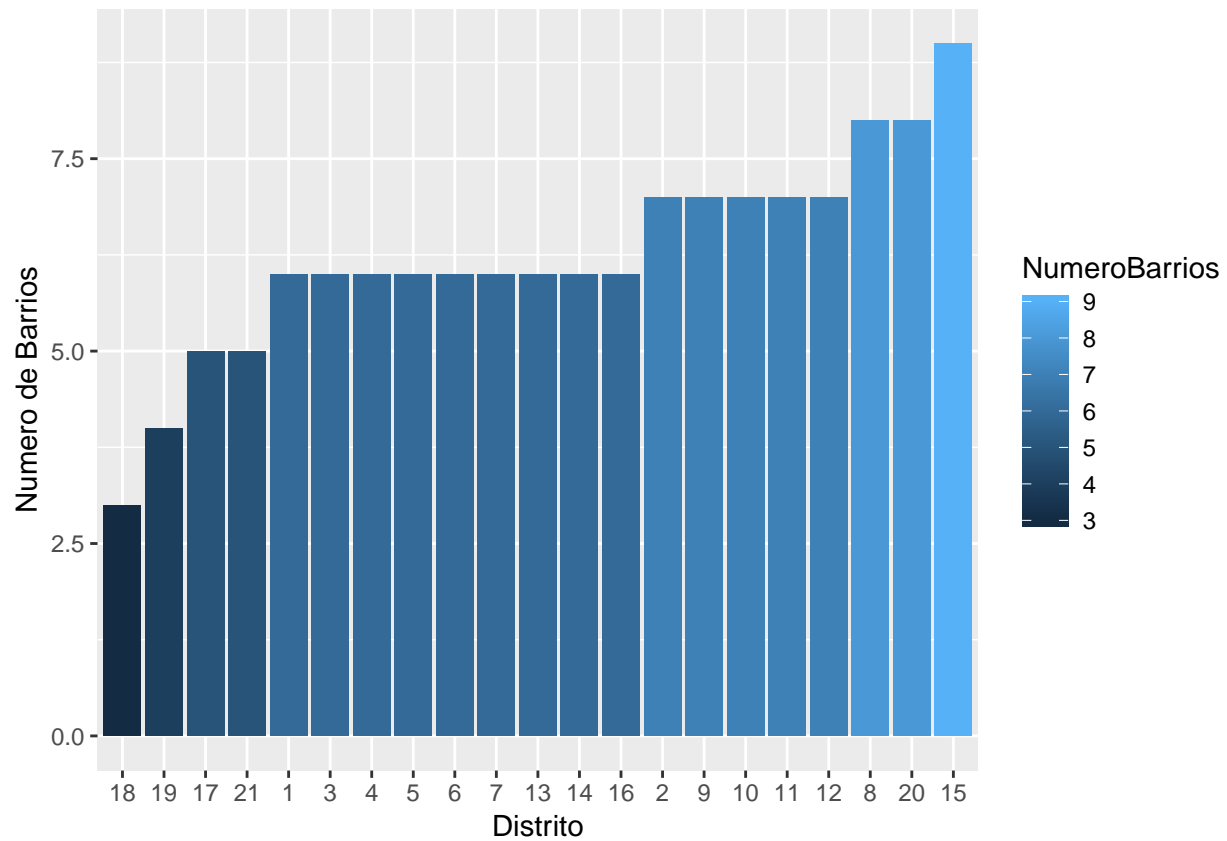
```
datos = data.frame(madrid$NumeroBarrios,madrid$SuperficieDistrito,madrid$ConsumoAgua,
                    madrid$Natalidad,madrid$Mortalidad,madrid$IndicePoblacionActiva,
                    madrid$Poblacion,madrid$NumeroCentroSanitarios,madrid$NumeroAparcabicis,
                    madrid$NumeroPuntosLimpios,madrid$NumeroPuntosRecargas,madrid$NumeroInmuebles,
                    madrid$SuperficieConstruida,madrid$HectareasVerdes)
summary(datos)
```

```
## madrid.NumeroBarrios madrid.SuperficieDistrito madrid.ConsumoAgua
## Min. :3.000      Min. : 4674105      Min. : 2399216
## 1st Qu.:6.000      1st Qu.: 6110071      1st Qu.: 5472018
## Median :6.000      Median : 14048399      Median : 7208250
## Mean :6.238      Mean : 28779376      Mean : 7027336
## 3rd Qu.:7.000      3rd Qu.: 27414593      3rd Qu.: 8902755
## Max. :9.000      Max. :237810357      Max. :11146898
## madrid.Natalidad madrid.Mortalidad madrid.IndicePoblacionActiva
```

```
## Min.      : 6.430    Min.      : 5.340    Min.      : 36.16
## 1st Qu.: 8.020    1st Qu.: 7.330    1st Qu.: 50.87
## Median : 8.710    Median : 8.360    Median : 64.68
## Mean    : 9.019    Mean    : 8.291    Mean    : 65.49
## 3rd Qu.: 9.950    3rd Qu.: 9.350    3rd Qu.: 77.84
## Max.    :14.070    Max.    :10.410    Max.    :111.44
## madrid.Poblacion madrid.NumeroCentroSanitarios madrid.NumeroAparcabicis
## Min.      : 47836    Min.      : 3.000                Min.      : 25.00
## 1st Qu.:119011    1st Qu.: 5.000                1st Qu.: 40.00
## Median :145268    Median : 7.000                Median : 45.00
## Mean    :153420    Mean    : 7.429                Mean    : 50.81
## 3rd Qu.:183930    3rd Qu.: 9.000                3rd Qu.: 55.00
## Max.    :248220    Max.    :14.000                Max.    :116.00
## madrid.NumeroPuntosLimpios madrid.NumeroPuntosRecargas
## Min.      : 0.000                Min.      :0.000
## 1st Qu.: 0.000                1st Qu.:0.000
## Median : 0.000                Median :1.000
## Mean    : 7.476                Mean    :1.619
## 3rd Qu.: 2.000                3rd Qu.:2.000
## Max.    :41.000                Max.    :7.000
## madrid.NumeroInmuebles madrid.SuperficieConstruida madrid.HectareasVerdes
## Min.      : 20498                Min.      : 2546949                Min.      : 11.35
## 1st Qu.: 54365                1st Qu.: 5196427                1st Qu.: 129.71
## Median : 72997                Median : 7862531                Median : 182.38
## Mean    : 69881                Mean    : 7303194                Mean    : 296.15
## 3rd Qu.: 79427                3rd Qu.: 9161116                3rd Qu.: 284.97
## Max.    :108410                Max.    :12443153                Max.    :1897.34
```

*#Numero de barrios que forman parte de los distritos*

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,NumeroBarrios),
                                         y = NumeroBarrios, fill=NumeroBarrios)) +
  labs(x="Distrito", y = "Numero de Barrios")
```

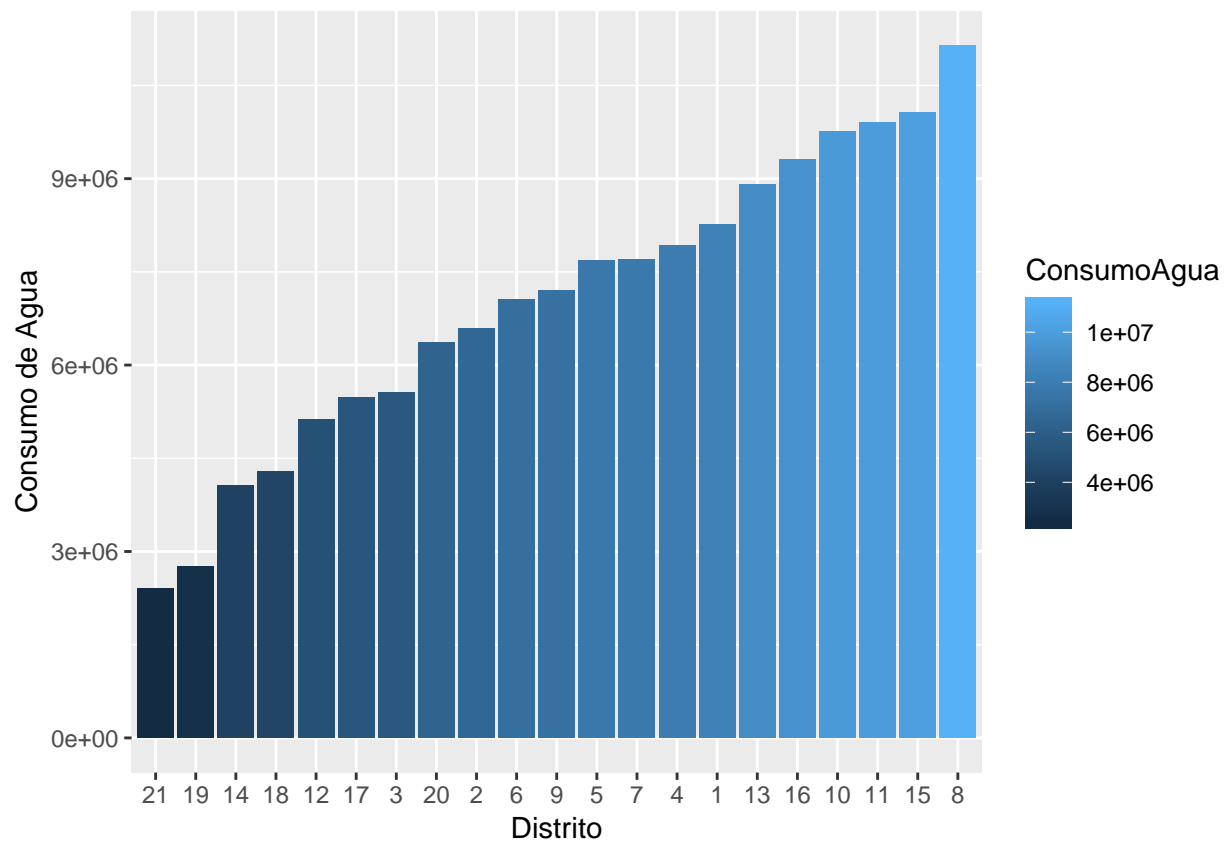


```
labs(title = "Numero de barrios por distrito")
```

```
## $title
## [1] "Numero de barrios por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Consumo de agua por Distrito
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,ConsumoAgua),
                                         y = ConsumoAgua, fill=ConsumoAgua)) +
  labs(x="Distrito", y = "Consumo de Agua")
```

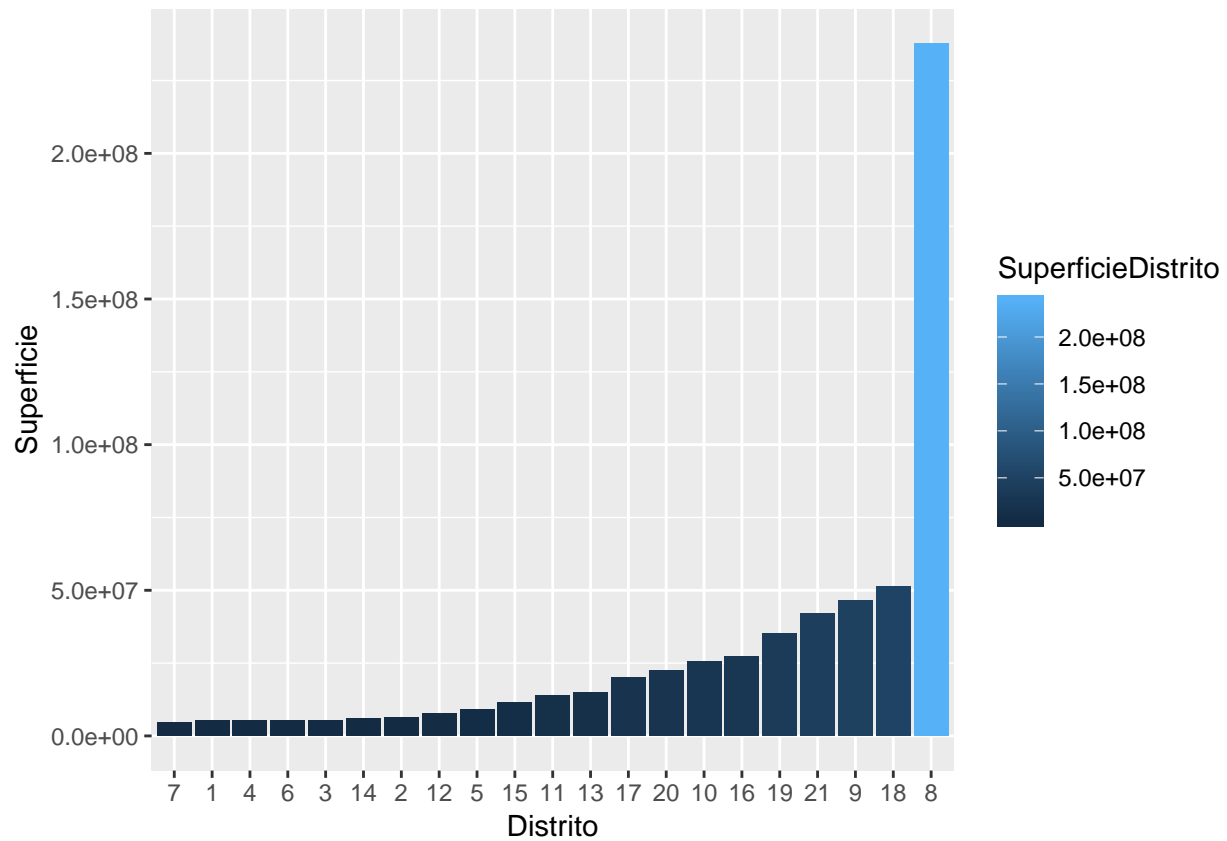


```
labs(title = "Consumo de Agua por Distrito")
```

```
## $title
## [1] "Consumo de Agua por Distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Superficie por Distrito
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,SuperficieDistrito),
                                         y = SuperficieDistrito, fill=SuperficieDistrito)) +
  labs(x="Distrito", y = "Superficie")
```

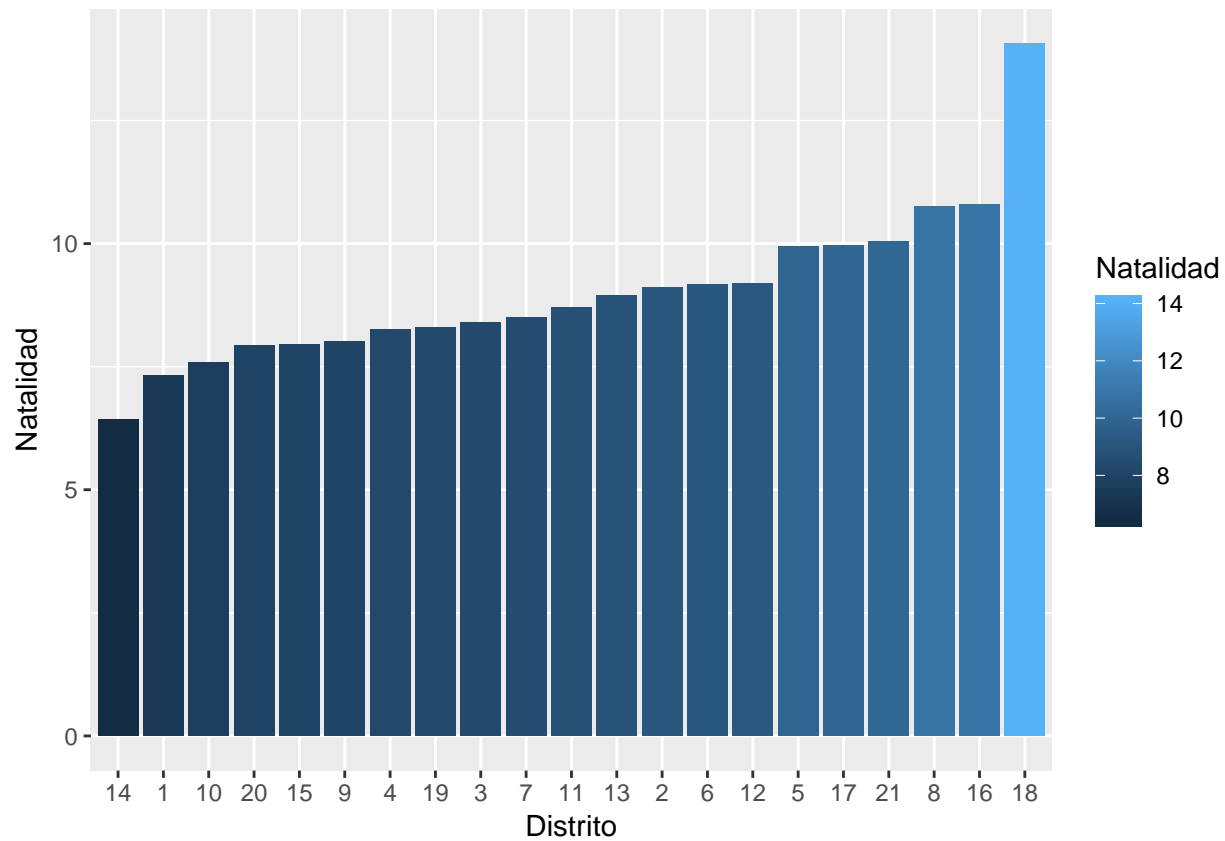


```
labs(title = "Superficie por distrito")
```

```
## $title
## [1] "Superficie por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Natalidad por Distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,Natalidad),
                                         y = Natalidad, fill=Natalidad)) +
  labs(x="Distrito", y = "Natalidad")
```

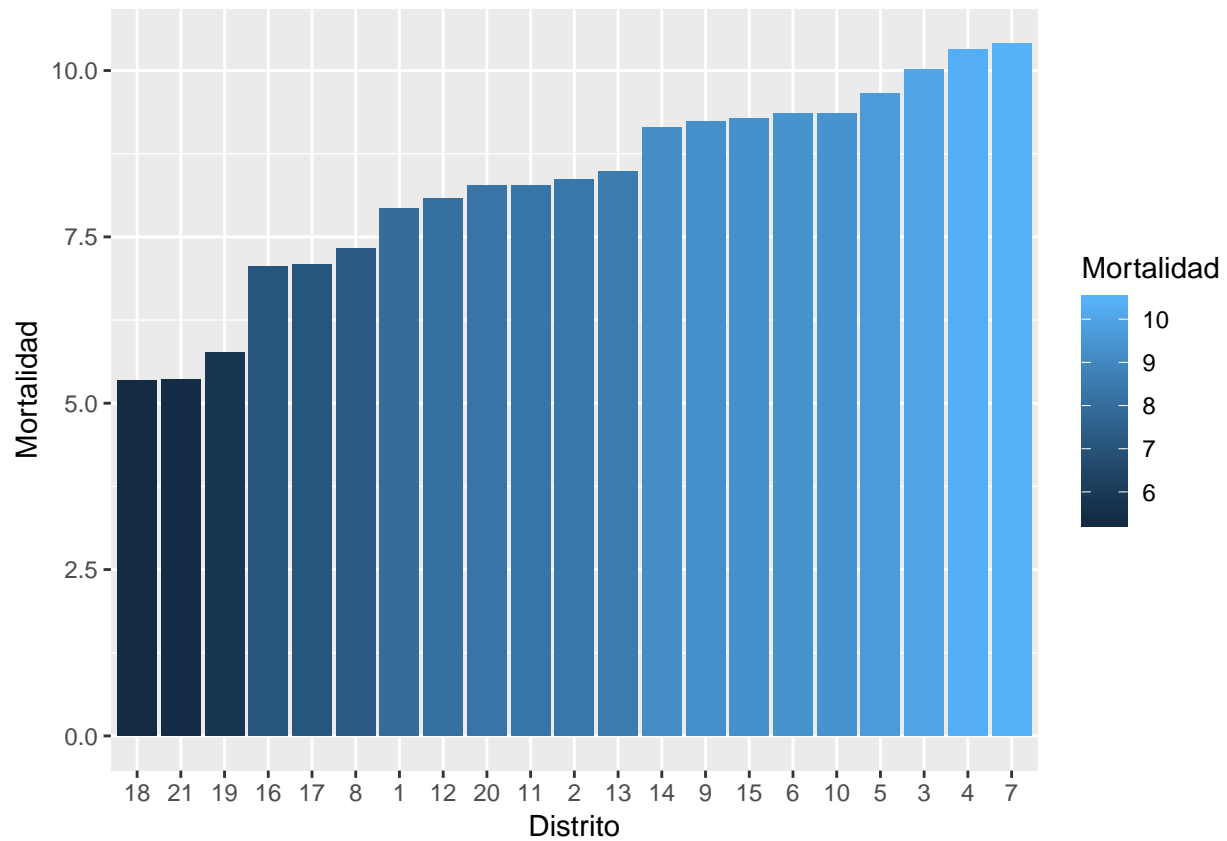


```
labs(title = "Natalidad por distrito")
```

```
## $title
## [1] "Natalidad por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Mortalidad por distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,Mortalidad),
                                     y = Mortalidad, fill=Mortalidad)) +
  labs(x="Distrito", y = "Mortalidad")
```



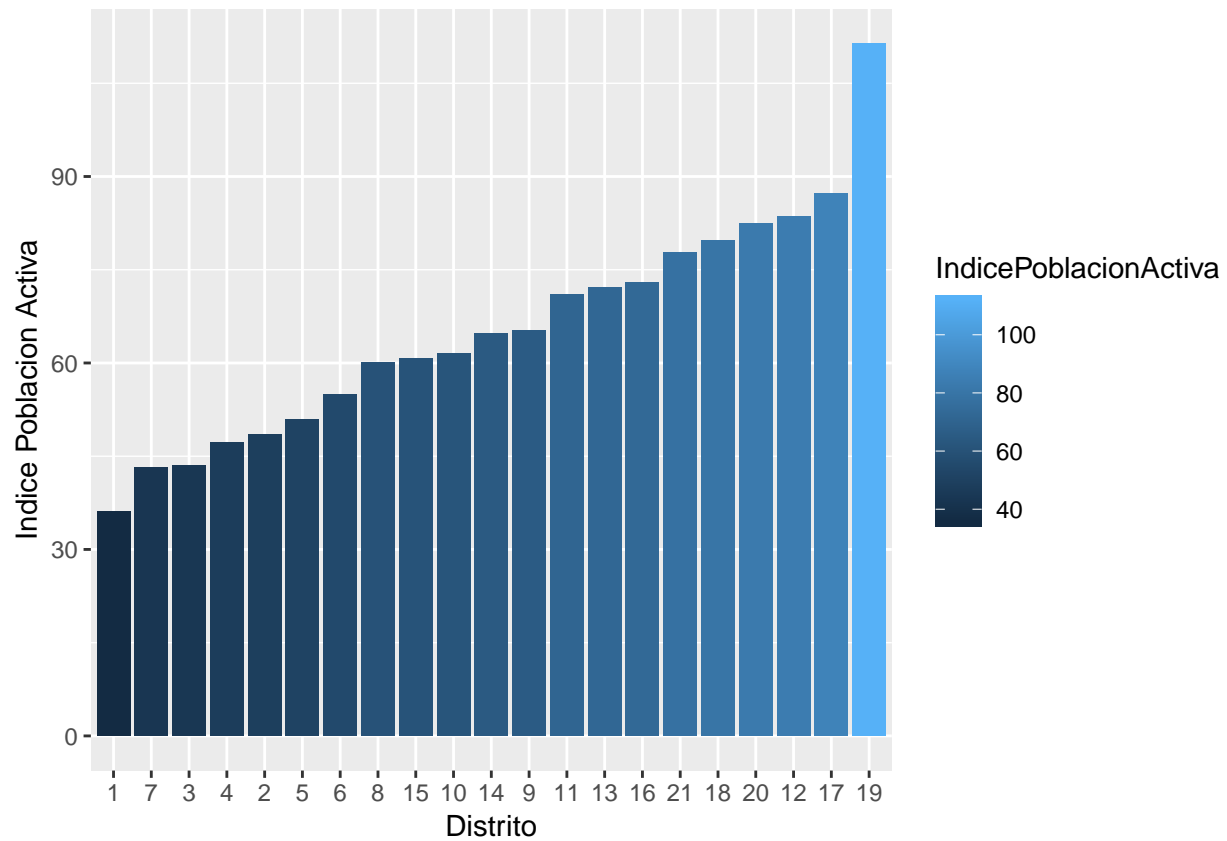
```
labs(title = "Mortalidad por distrito")
```

```
## $title
## [1] "Mortalidad por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Indice de población activa por distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,IndicePoblacionActiva),
                                         y = IndicePoblacionActiva, fill=IndicePoblacionActiva)) +
  labs(x="Distrito", y = "Indice Poblacion Activa")
```



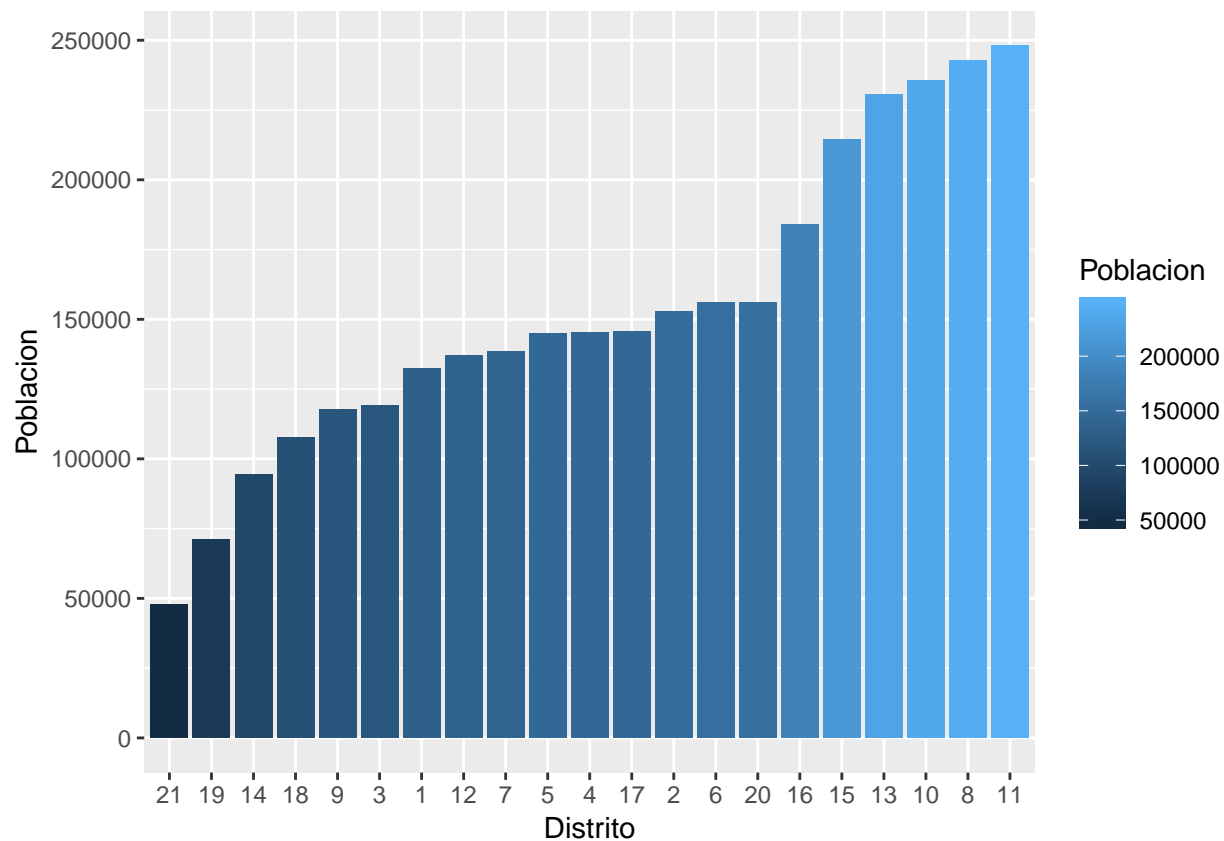


```
labs(title = "Indice de Poblacion Activa por distrito")
```

```
## $title
## [1] "Indice de Poblacion Activa por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Poblacion por distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,Poblacion),
                                         y = Poblacion, fill=Poblacion)) +
  labs(x="Distrito", y = "Poblacion")
```

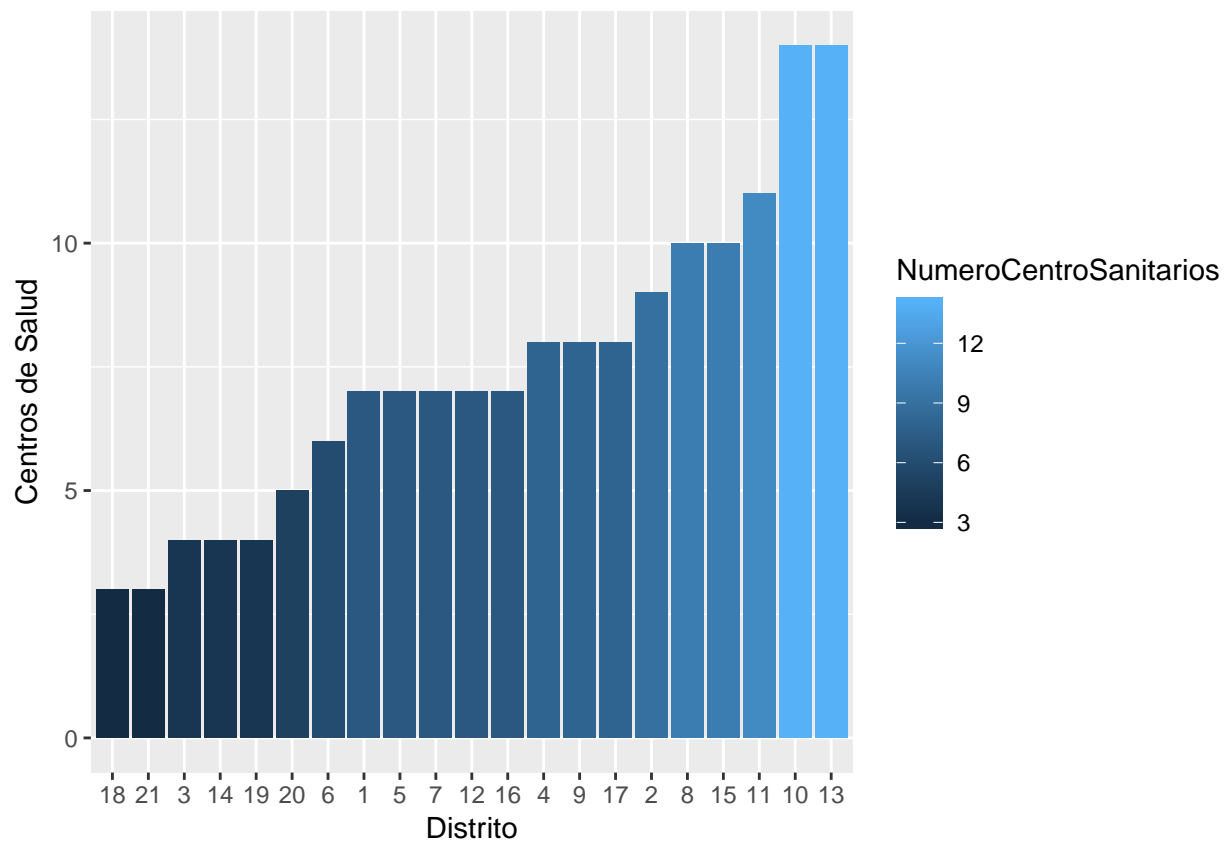


```
labs(title = "Poblacion por distrito")
```

```
## $title
## [1] "Poblacion por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Numero de Centros de Salud por distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,NumeroCentroSanitarios),
                                         y = NumeroCentroSanitarios, fill=NumeroCentroSanitarios)) +
  labs(x="Distrito", y = "Centros de Salud")
```

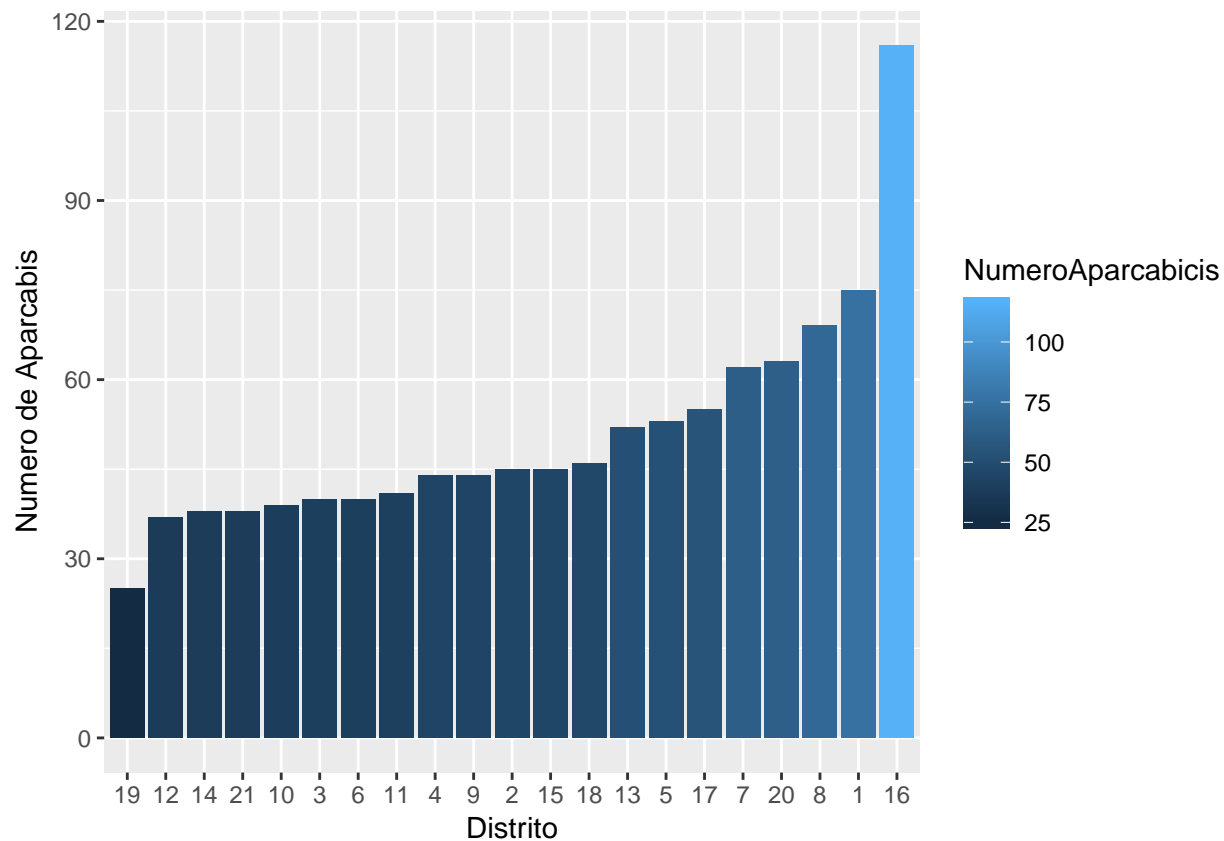


```
labs(title = "Centros de Salud por distrito")
```

```
## $title
## [1] "Centros de Salud por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Numero de aparcabicis por distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,NumeroAparcabicis),
    y = NumeroAparcabicis, fill=NumeroAparcabicis)) + labs(x="Distrito", y="NumeroAparcabicis")
```

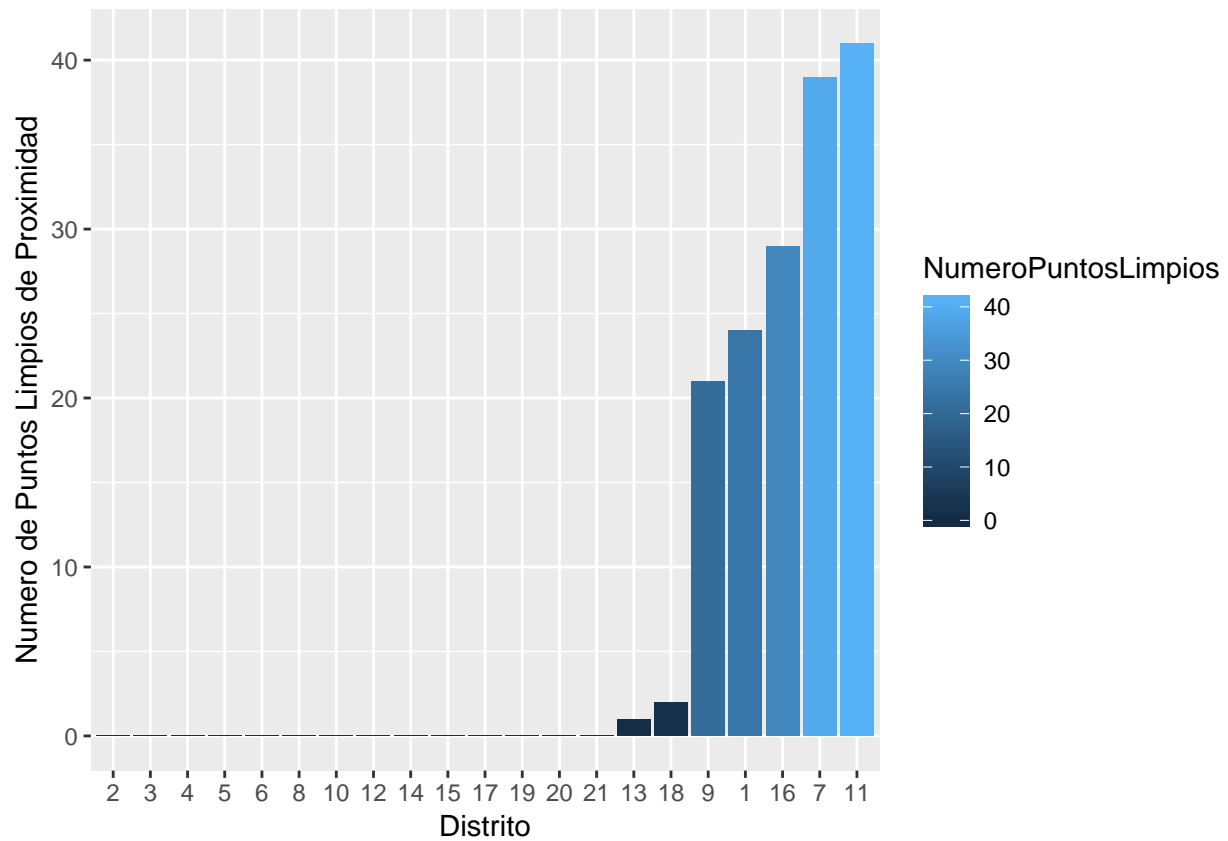


```
labs(title = "Numero de aparcabicis por distrito")
```

```
## $title
## [1] "Numero de aparcabicis por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Numero de puntos limpios de proximidad que forman parte de los distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,NumeroPuntosLimpios),
                                         y = NumeroPuntosLimpios, fill=NumeroPuntosLimpios)) +
  labs(x="Distrito", y = "Numero de Puntos Limpios de Proximidad")
```

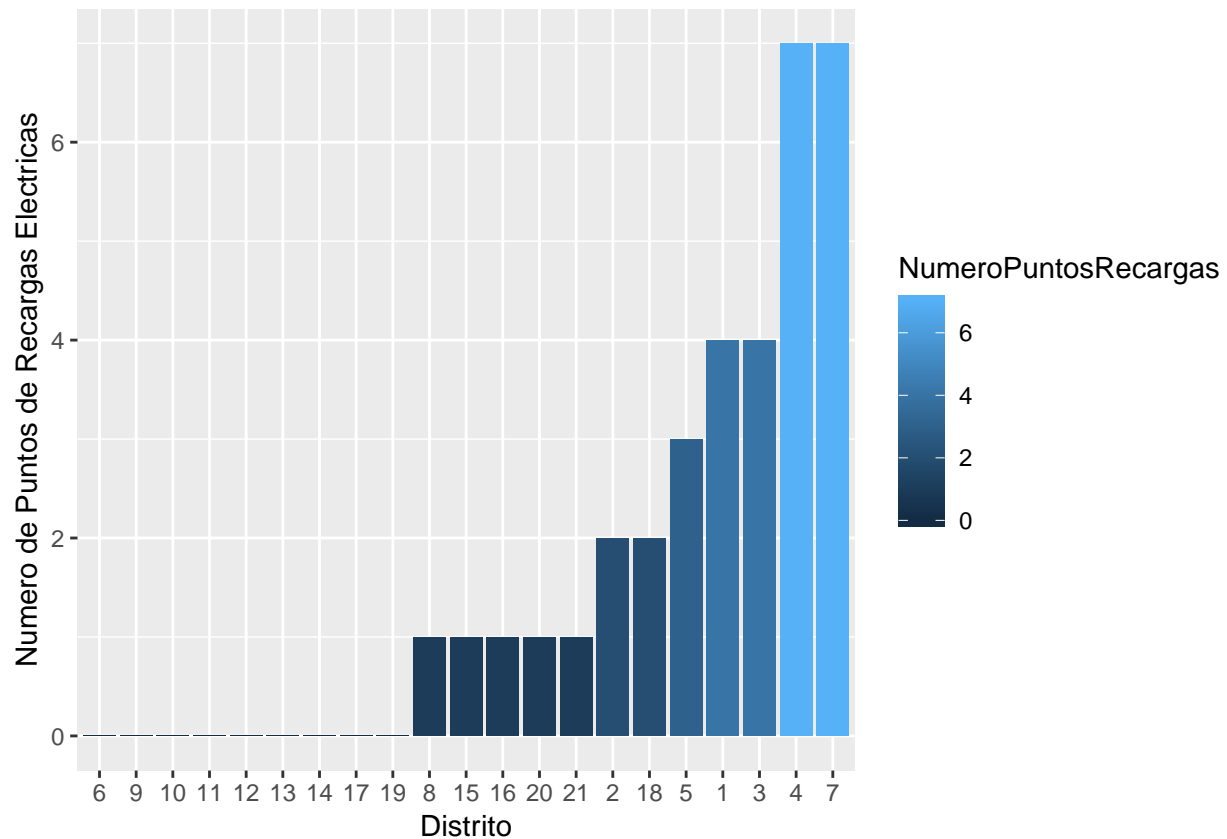


```
labs(title = "Numero de Puntos Limpios de Proximidad por distrito")
```

```
## $title
## [1] "Numero de Puntos Limpios de Proximidad por distrito"
##
## attr(,"class")
## [1] "labels"
```

```
#Numero de puntos de recargas electricas que forman parte de los distritos
```

```
ggplot(data=madrid)+
  geom_bar(stat='identity', aes(x=reorder(CodigoDistrito,NumeroPuntosRecargas),
                                         y = NumeroPuntosRecargas, fill=NumeroPuntosRecargas)) +
  labs(x="Distrito", y = "Numero de Puntos de Recargas Electricas")
```



```
labs(title = "Numero de Puntos de Recargas Electricas por distrito")
```

```
## $title
## [1] "Numero de Puntos de Recargas Electricas por distrito"
##
## attr(,"class")
## [1] "labels"
```

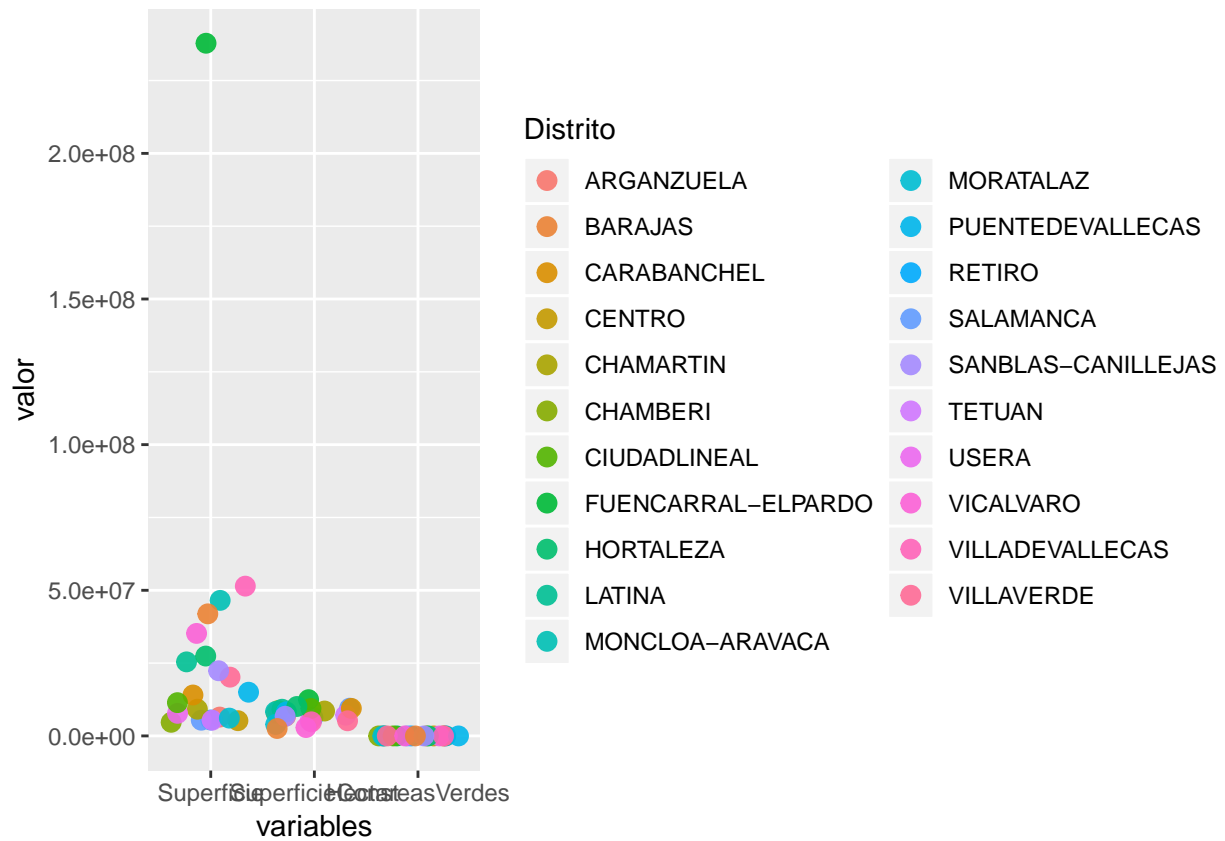
Visualización de las variables Superficie, Superficie Construida y Hectareas Verdes según los distritos

```
#Convertimos los datos a columna
madridc<-madrid
madridc=data.frame(madridc$NombreDistrito,madridc$CodigoDistrito,madridc$SuperficieDistrito,
                   madridc$SuperficieConstruida,madridc$HectareasVerdes)
colnames(madridc)=c("Distrito","CodDistrito","Superficie","SuperficieConst","HectareasVerdes")
colnames(madridc)
```

```
## [1] "Distrito"          "CodDistrito"       "Superficie"        "SuperficieConst"
## [5] "HectareasVerdes"
```

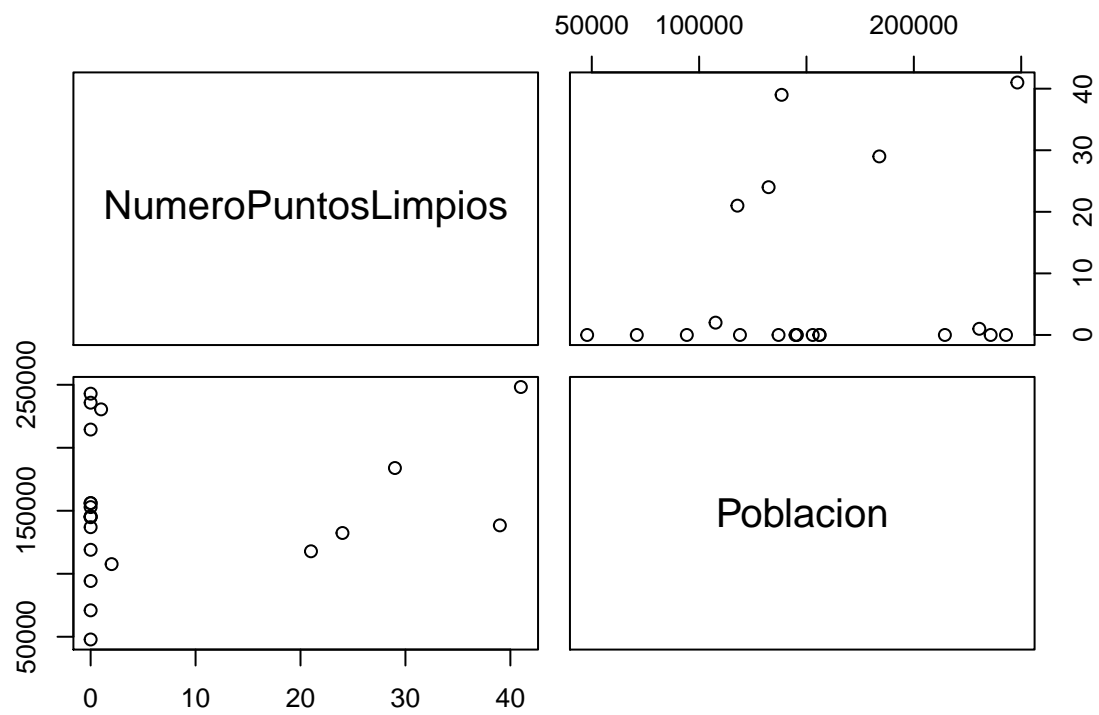
```
madridcv<-melt(madridc,id=c("Distrito","CodDistrito"))
```

```
ggplot(data=madridcv)+
  geom_point(mapping = aes(x=variable, y = value, colour=Distrito),
             position = "jitter", alpha =9/10, size=3)+
  labs(y="valor", x ="variables")
```



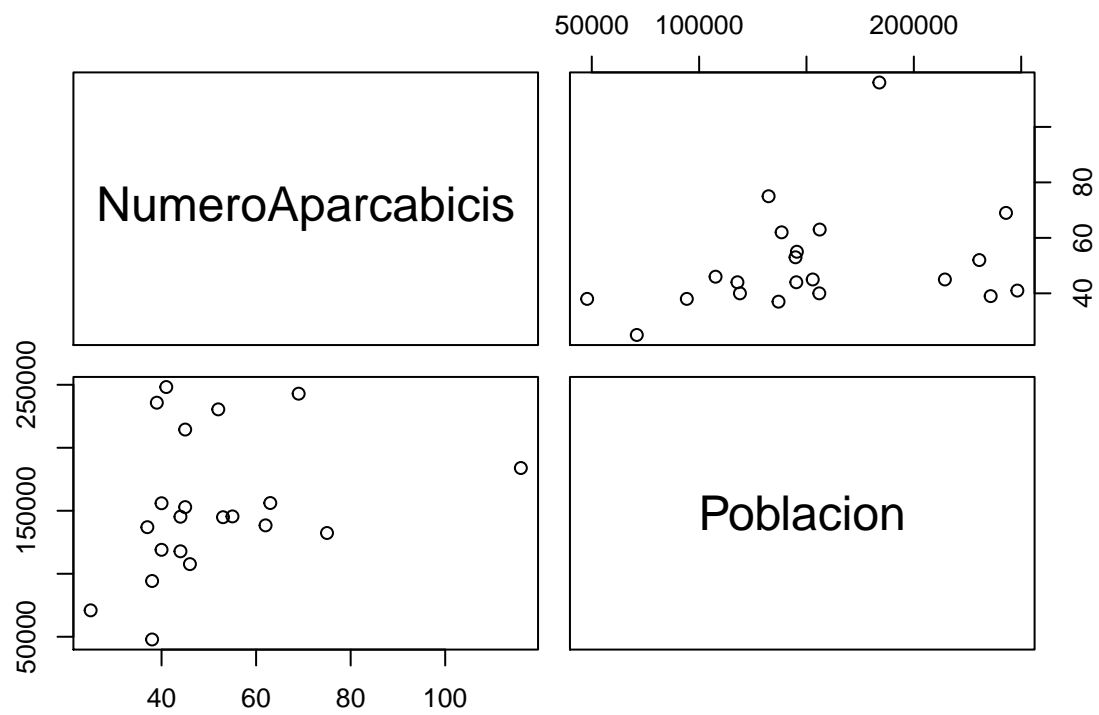
Correlacion del número de puntos limpios, número de aparcabicis, número de puntos de recargas eléctricos, centro de salud, superficie construida con el número de poblacion respectivamente.

```
pairs (~NumeroPuntosLimpios+Poblacion, data=madrid)
```

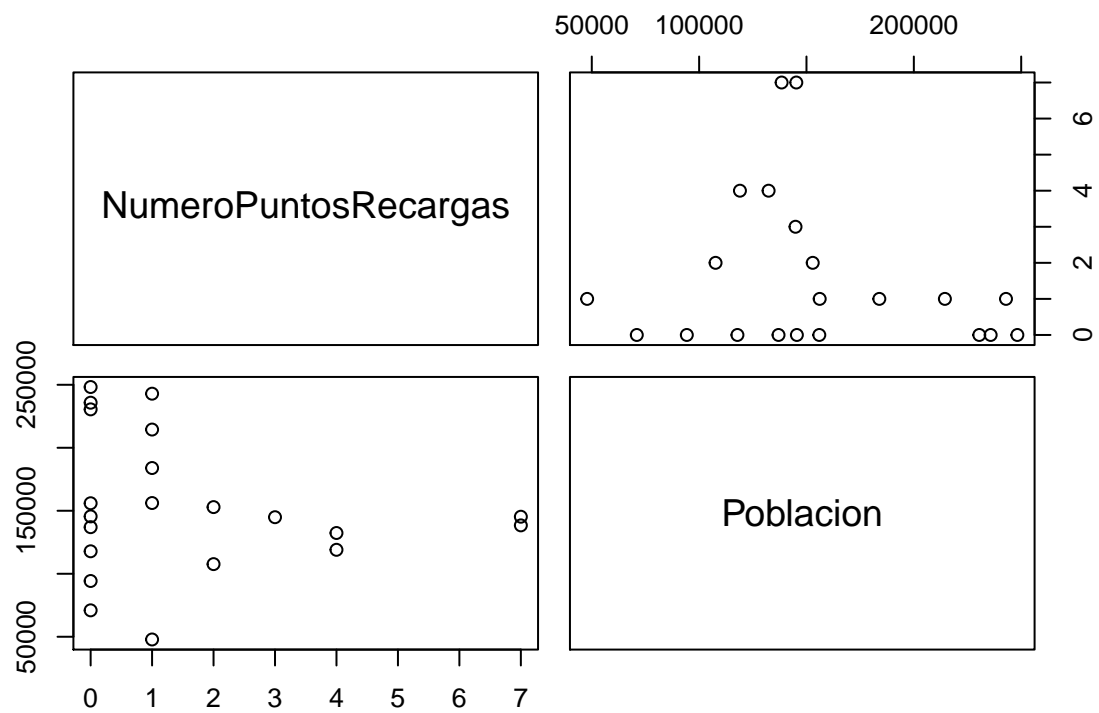


```
pairs (~NumeroAparcabicis+Poblacion,data=madrid)
```

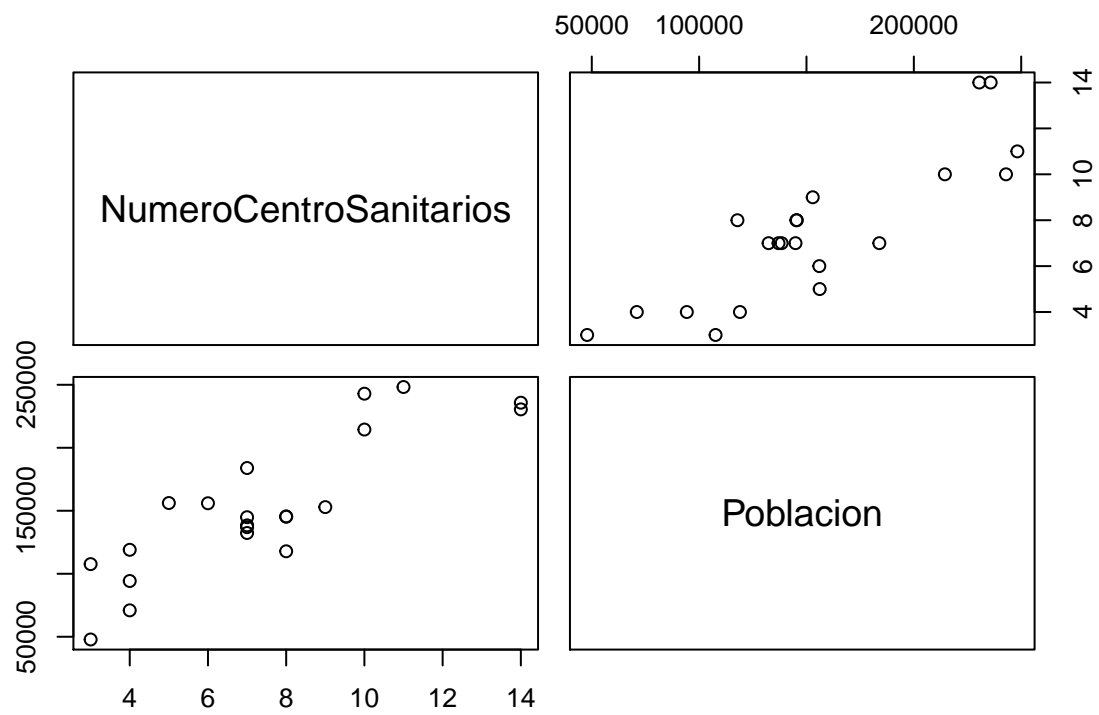




```
pairs (~NumeroPuntosRecargas+Poblacion,data=madrid)
```



```
pairs (~NumeroCentroSanitarios+Poblacion,data=madrid)
```



```
pairs (~SuperficieConstruida+Poblacion,data=madrid)
```

