

Elaboración de Gráficos con el sistema base de R y con ggplot2

Abel Isaias Gutierrez-Cruz

10/8/2021

Preparación del ambiente

Cargar la librería de ggplot2 para la elaboración de los gráficos

```
library(ggplot2)
```

Para que usar gráficas en el análisis de datos

- Para entender las propiedades de los datos
- Para encontrar patrones en los datos
- Para sugerir estrategias de modelado
- Para comunicar los resultados

Cargar los datos y explorarlos

airquality contiene datos sobre la calidad de aire

```
data(airquality)
```

```
head(airquality, 3)
```

```
##      Ozone Solar.R Wind Temp Month Day
## 1      41     190  7.4   67     5    1
## 2      36     118  8.0   72     5    2
## 3      12     149 12.6   74     5    3
```

```
summary(airquality)
```

```
##      Ozone      Solar.R      Wind      Temp
## Min.   : 1.00   Min.   : 7.0   Min.   : 1.700   Min.   :56.00
## 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
## Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
## Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
## 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
## Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
## NA's   :37      NA's   :7
##      Month      Day
## Min.   :5.000   Min.   : 1.0
## 1st Qu.:6.000   1st Qu.: 8.0
## Median :7.000   Median :16.0
## Mean   :6.993   Mean   :15.8
## 3rd Qu.:8.000   3rd Qu.:23.0
## Max.   :9.000   Max.   :31.0
##
```

Características de las variables de airquality

```
str(airquality)
```

```
## 'data.frame':   153 obs. of  6 variables:
## $ Ozone : int  41 36 12 18 NA 28 23 19 8 NA ...
## $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
## $ Wind : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
## $ Temp : int  67 72 74 62 56 66 65 59 61 69 ...
## $ Month : int  5 5 5 5 5 5 5 5 5 5 ...
## $ Day : int  1 2 3 4 5 6 7 8 9 10 ...
```

Plots con el sistema base de R

Se comienza con un gráfico base, este gráfico posteriormente será modificado con más propiedades

Parametros útiles

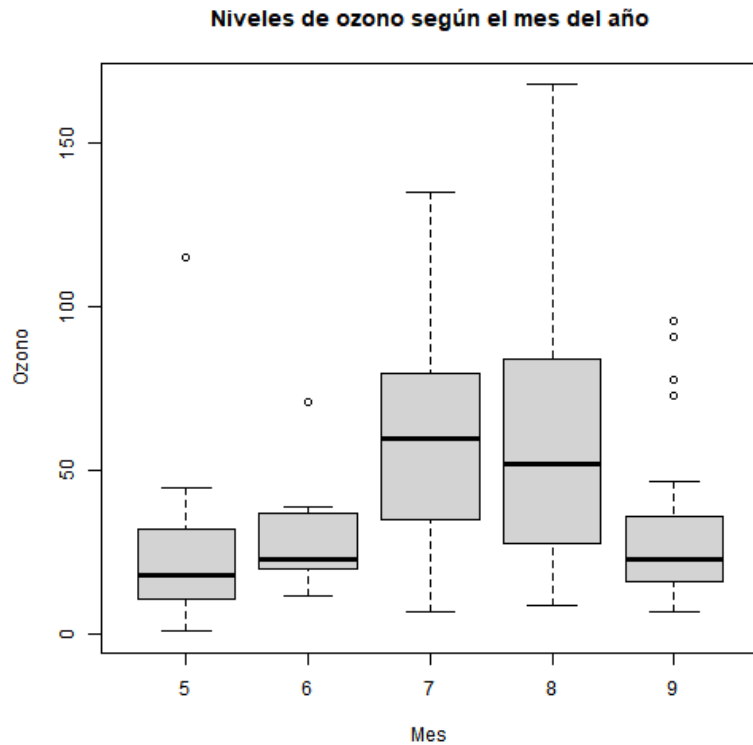
- `pch`: simbolo utilizado en el gráfico
- `lty`: el tipo de línea utilizada
- `lwd`: el ancho de la línea, especificado como un entero
- `col`: determinar el color del plot, es especificado como un número, string o código hexadecimal
- `xlab`: El texto que se desea que aparezca en el eje de las x
- `ylab`: El texto que se desea que aparezca en el eje de las y

Funciones utiles

- `lines`: agrega lineas a un gráfico, dado dos vectores de valores
- `points`: agrega puntos a un gráfico
- `text`: agrega etiquetas de texto a un gráfico, especificando su posición con coordenadas "x" y "y"
- `title`: agregar anotaciones a las etiquetas del eje x, y, titulo, subtítulo o fuera del margen
- `axis`: agrega etiquetas a los ejes
- `legend`: recibe la ubicación de la legenda como primer parámetro (topleft, topright, bottomleft, bottomright) y el parámetro `legend = "texto"`

Box plot

```
boxplot(Ozone ~ Month, airquality, xlab = "Mes", ylab = "Ozono")  
title(main = "Niveles de ozono según el mes del año")
```



Scatter plot

```
with(airquality, plot(Wind, Ozone, main = "Ozono con respecto al polvo", col = "red"))  
with(subset(airquality, Month == 5), points(Wind, Ozone, col = "blue"))  
legend("topright", pch = 1, col = c("blue", "red"), legend = c("May", "Other Months"))
```

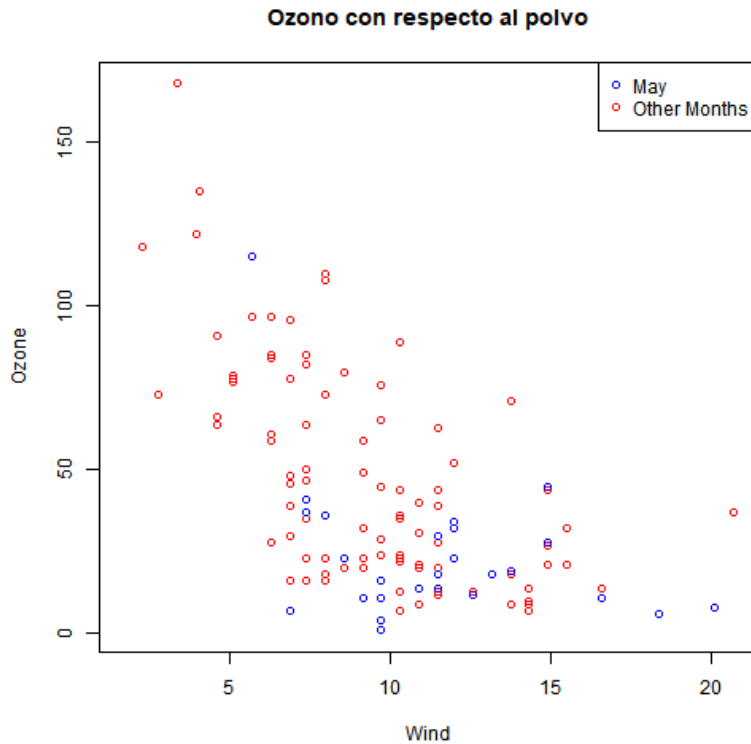
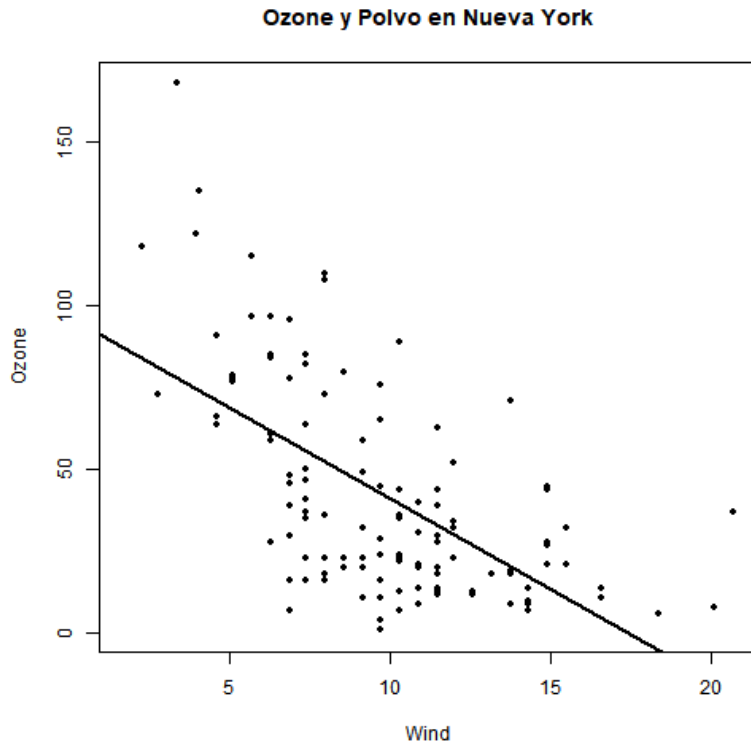


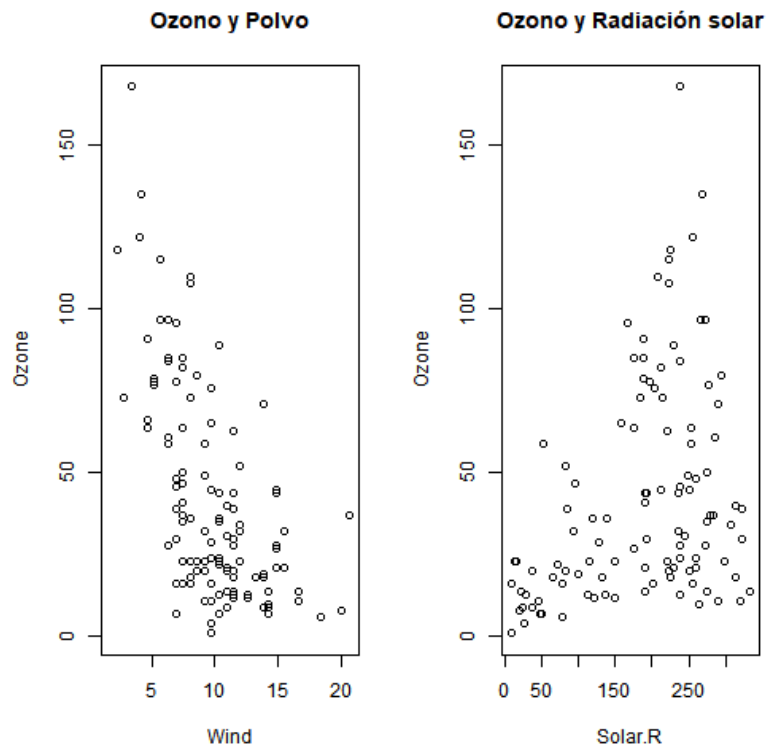
Gráfico base con regresión lineal

```
with(airquality, plot(Wind, Ozone, main = "Ozone y Polvo en Nueva York", pch=20))  
model <- lm(Ozone ~ Wind, airquality)  
abline(model, lwd=2)
```



Gráficos múltiples

```
par(mfrow = c(1, 2))  
with(airquality, {  
  plot(Wind, Ozone, main = "Ozono y Polvo")  
  plot(Solar.R, Ozone, main = "Ozono y Radiación solar")  
})
```



Cargar los datos y explorarlos

```
data("iris")
```

```
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5         1.4         0.2   setosa
## 2           4.9         3.0         1.4         0.2   setosa
## 3           4.7         3.2         1.3         0.2   setosa
## 4           4.6         3.1         1.5         0.2   setosa
## 5           5.0         3.6         1.4         0.2   setosa
## 6           5.4         3.9         1.7         0.4   setosa
```

```
summary(iris)
```

```
##   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width
##   Min.    :4.300   Min.    :2.000   Min.    :1.000   Min.    :0.100
##   1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##   Median :5.800   Median :3.000   Median :4.350   Median :1.300
##   Mean    :5.843   Mean    :3.057   Mean    :3.758   Mean    :1.199
##   3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##   Max.    :7.900   Max.    :4.400   Max.    :6.900   Max.    :2.500
##           Species
##   setosa     :50
##   versicolor:50
##   virginica  :50
##
##
##
```

Características de las variables de iris

```
str(iris)
```

```
## 'data.frame':   150 obs. of  5 variables:
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species     : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

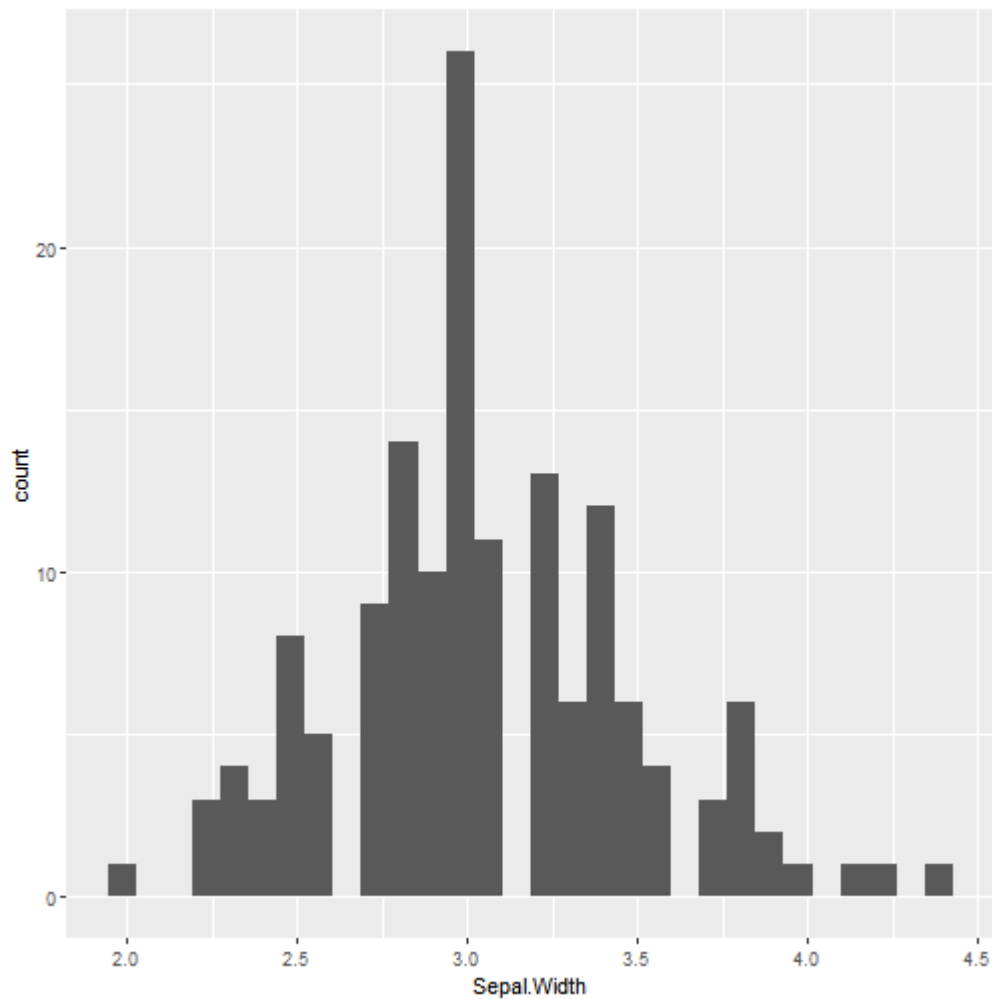
Sistema ggplot2

Componentes basicos de un gráfico en ggplot2

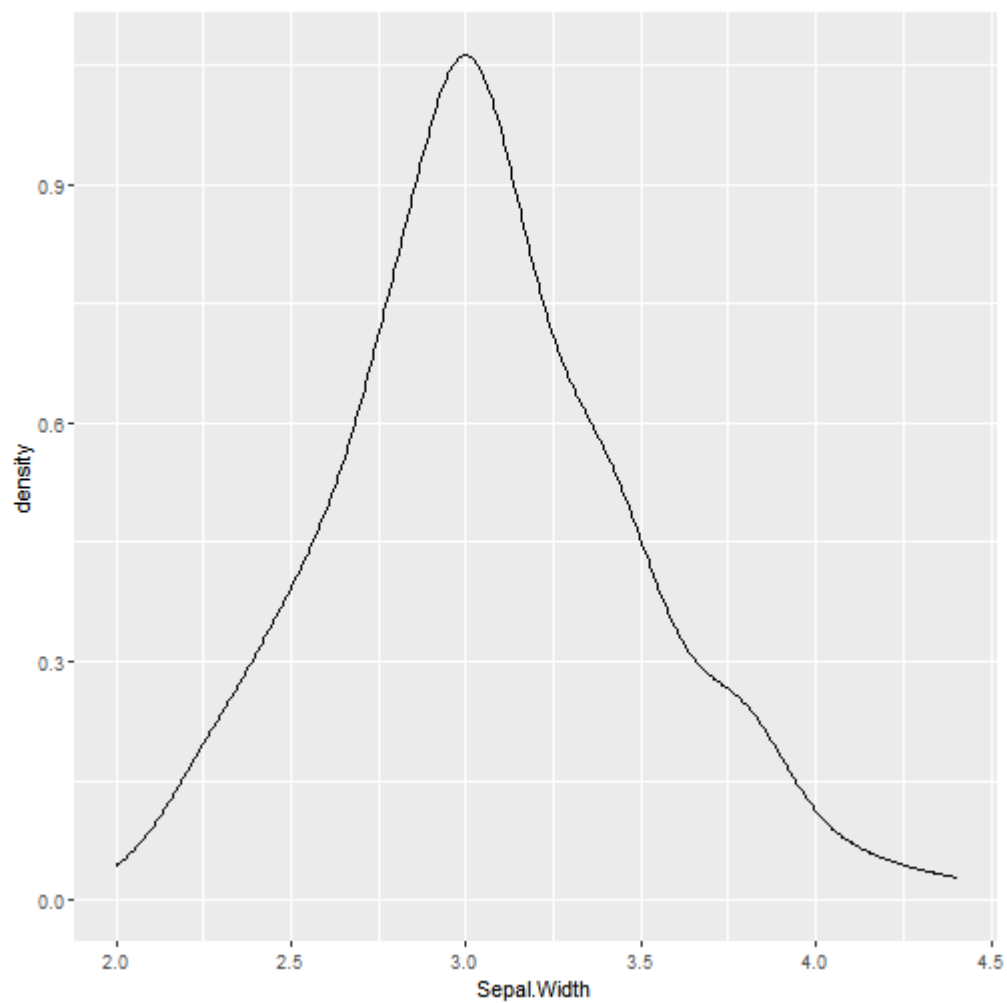
- **A data frame**
 - Se especifica con `data = <dataset>`
- **aesthetic mappings:** como los datos son representados en el sentido del color y el tamaño
 - Se especifica con `mapping = aes(x = <variable>)`
- **geoms:** objetos geometricos como puntos, lineas y figuras
 - Se especifica con el método correspondiente al gráfico
- **facets:** transformaciones estadísticas como cuantiles
- **scales:** que escala usar (example: hombre = rojo, mujer = azul)
- **sistema de coordenadas**

Una variable continua

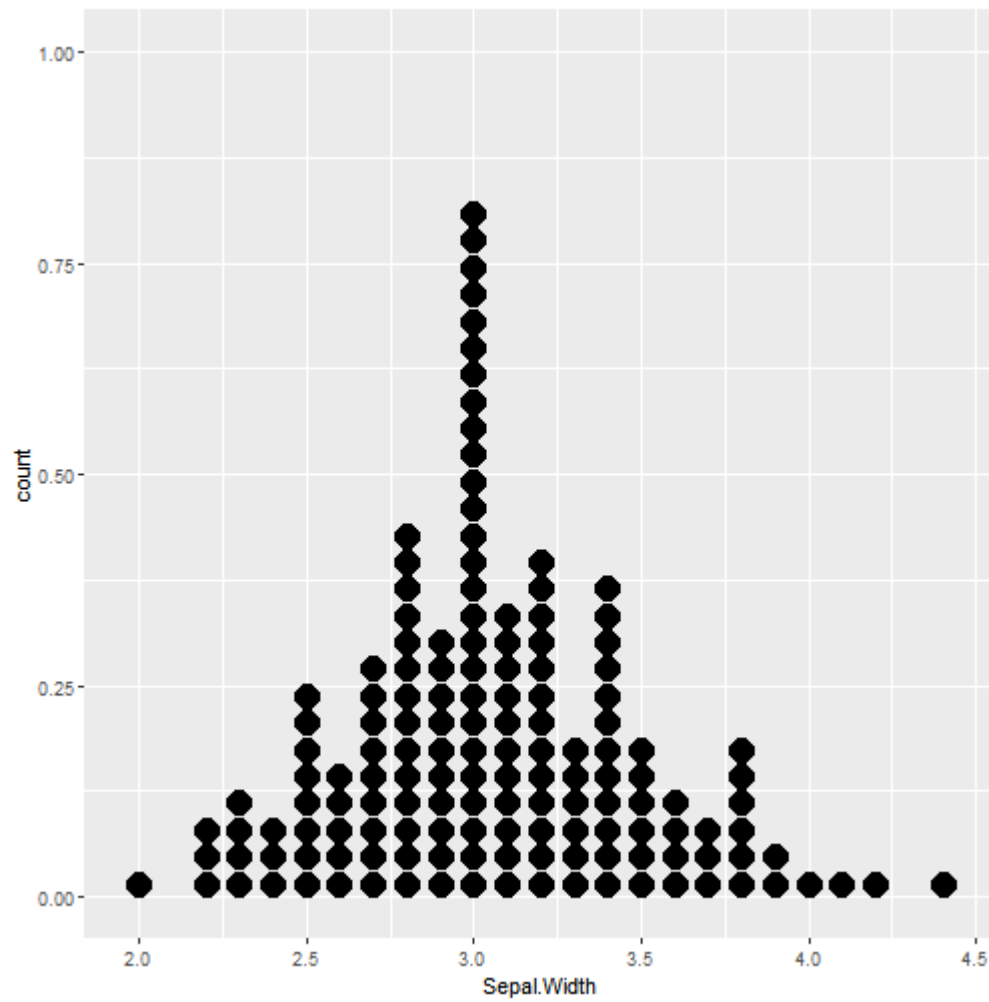
```
ggplot(data = iris, mapping = aes(x = Sepal.Width)) + geom_histogram()
```



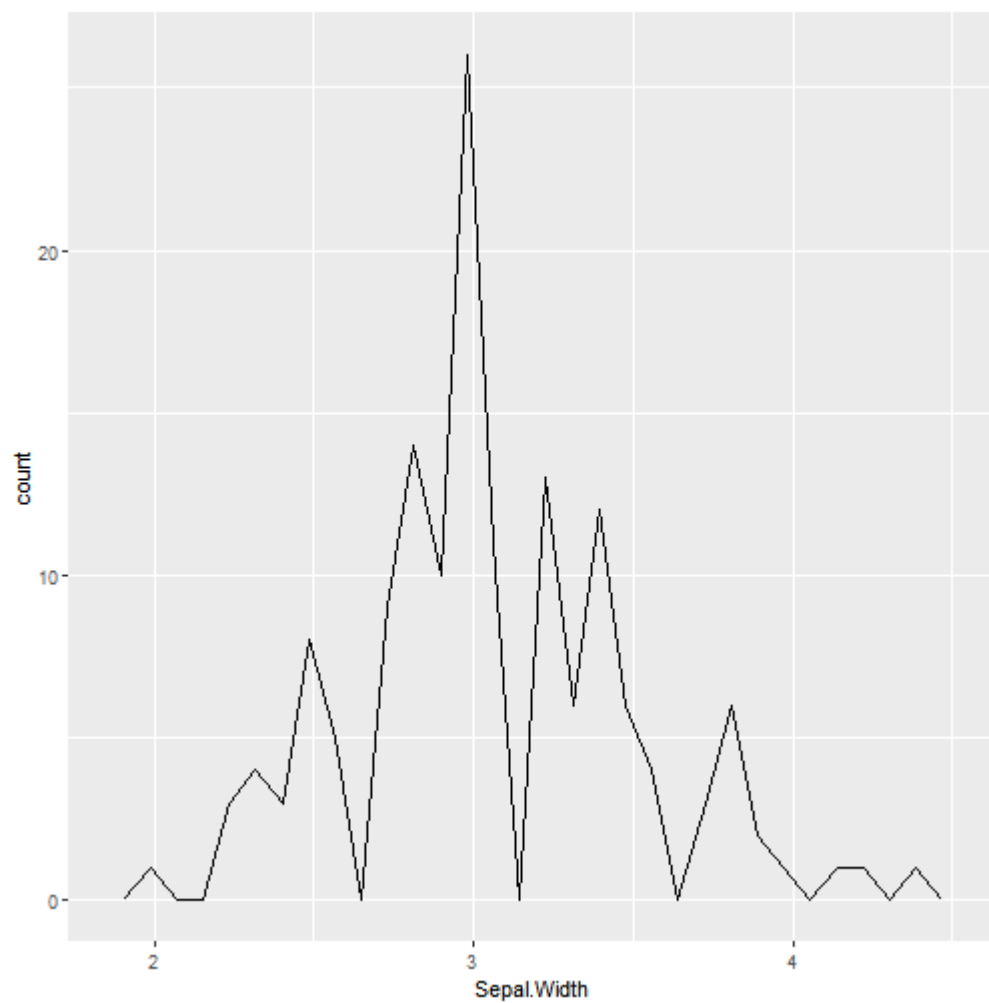
```
ggplot(data = iris, mapping = aes(x = Sepal.Width)) + geom_density()
```



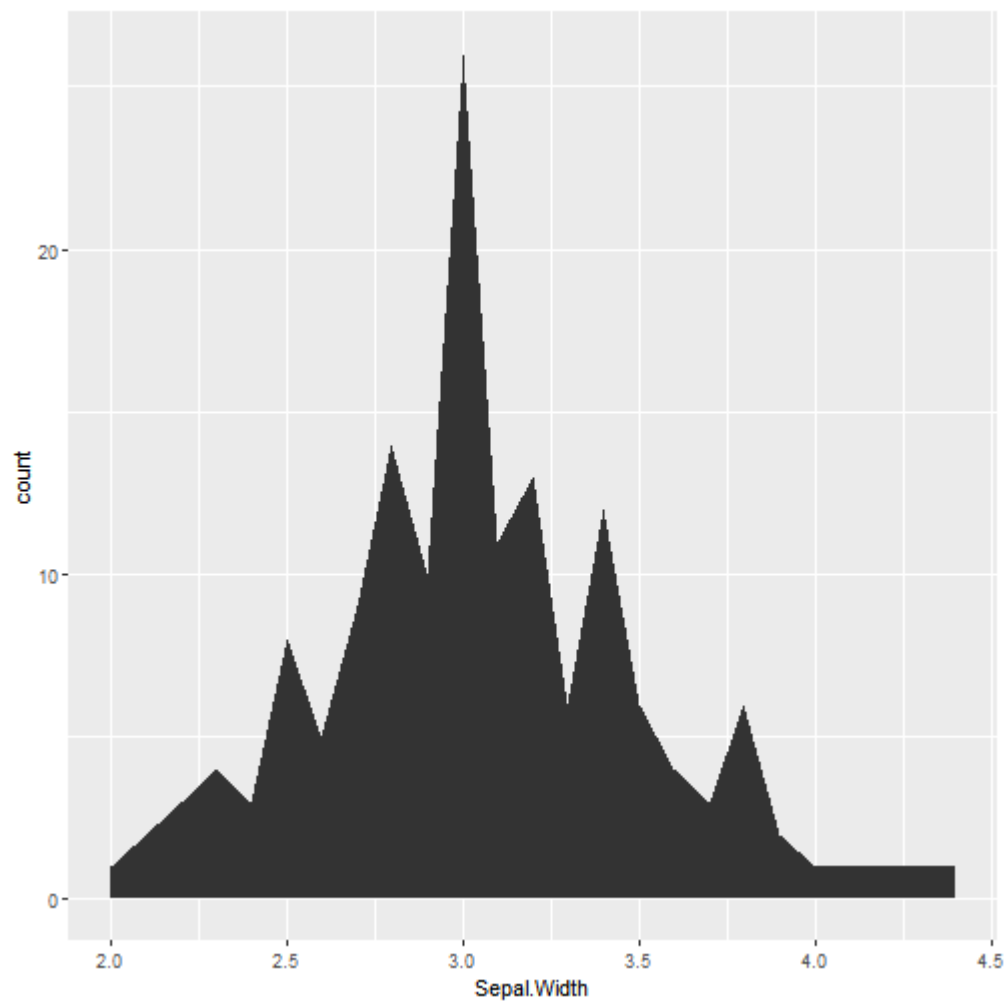
```
ggplot(data = iris, mapping = aes(x = Sepal.Width)) + geom_dotplot()
```




```
ggplot(data = iris, mapping = aes(x = Sepal.Width)) + geom_freqpoly()
```

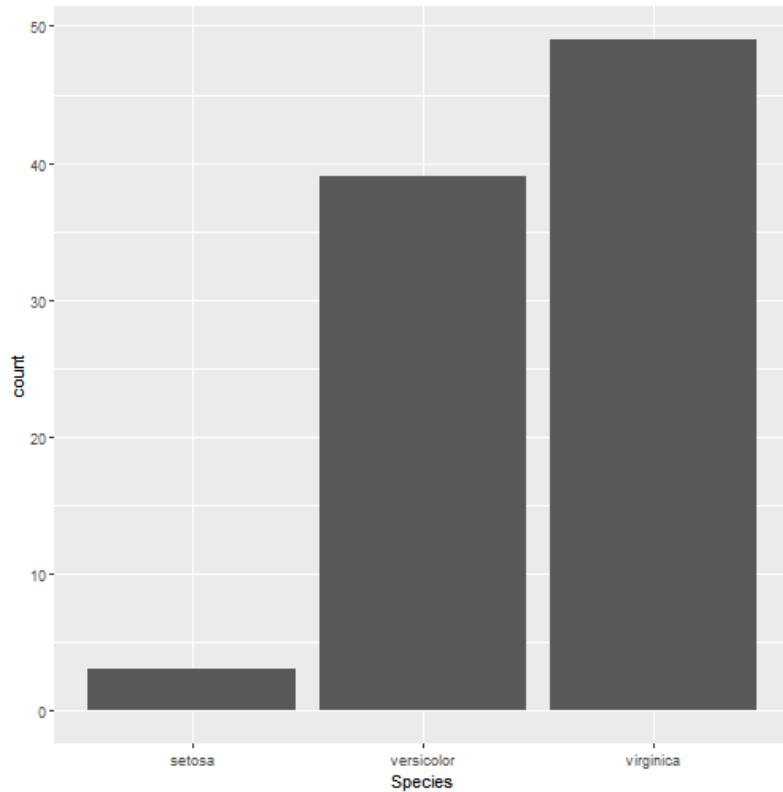


```
ggplot(data = iris, mapping = aes(x = Sepal.Width)) + geom_area(stat = "count")
```



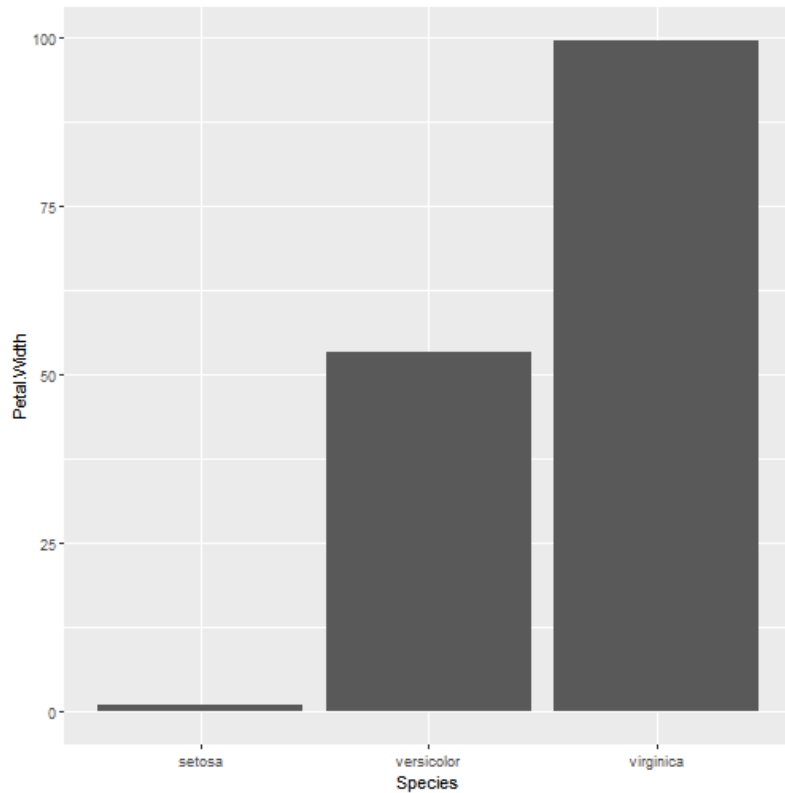
Una variable categórica

```
data2 <- iris[iris$Sepal.Length > 5.5, ]  
ggplot(data = data2, mapping = aes(x=Species)) + geom_bar()
```



Hacer un gráfico de barras con dos variables

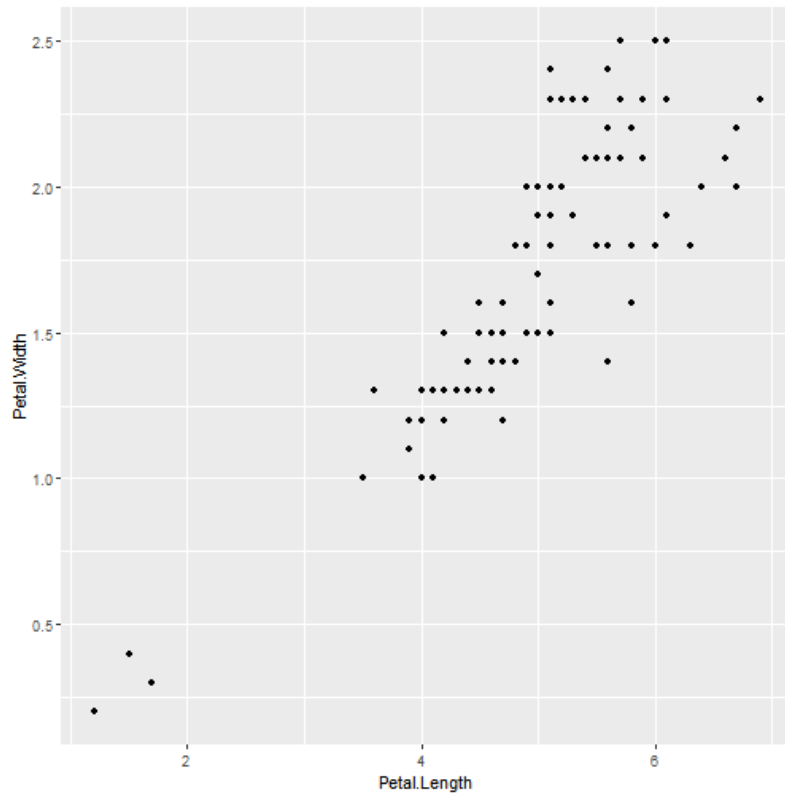
```
ggplot(data = data2, mapping = aes(x = Species, y = Petal.Width)) +  
  geom_col(position = "stack")
```



Dos variables continuas

Scatterplot

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) + geom_point()
```



Agrega variación aleatoria a la posición de los puntos

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) + geom_jitter()
```

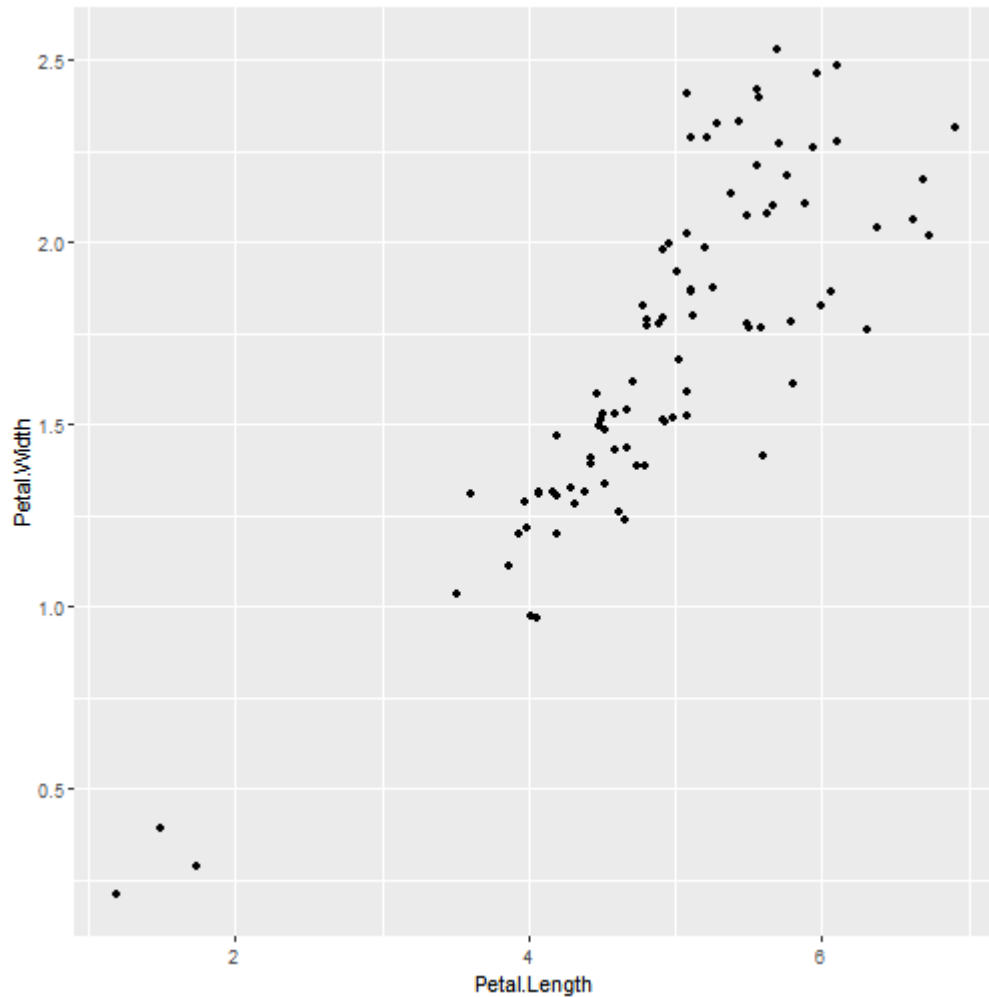
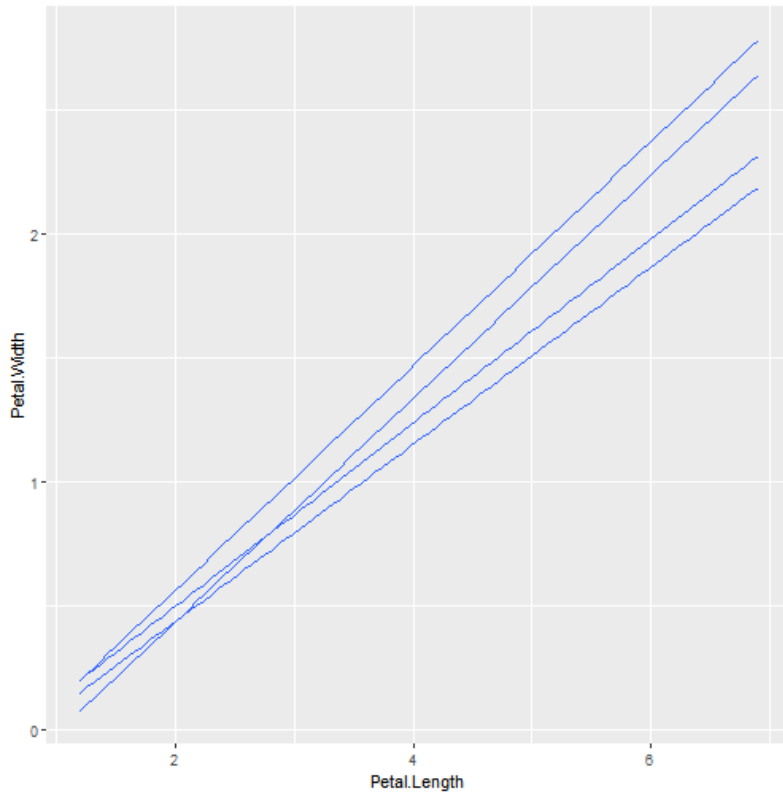


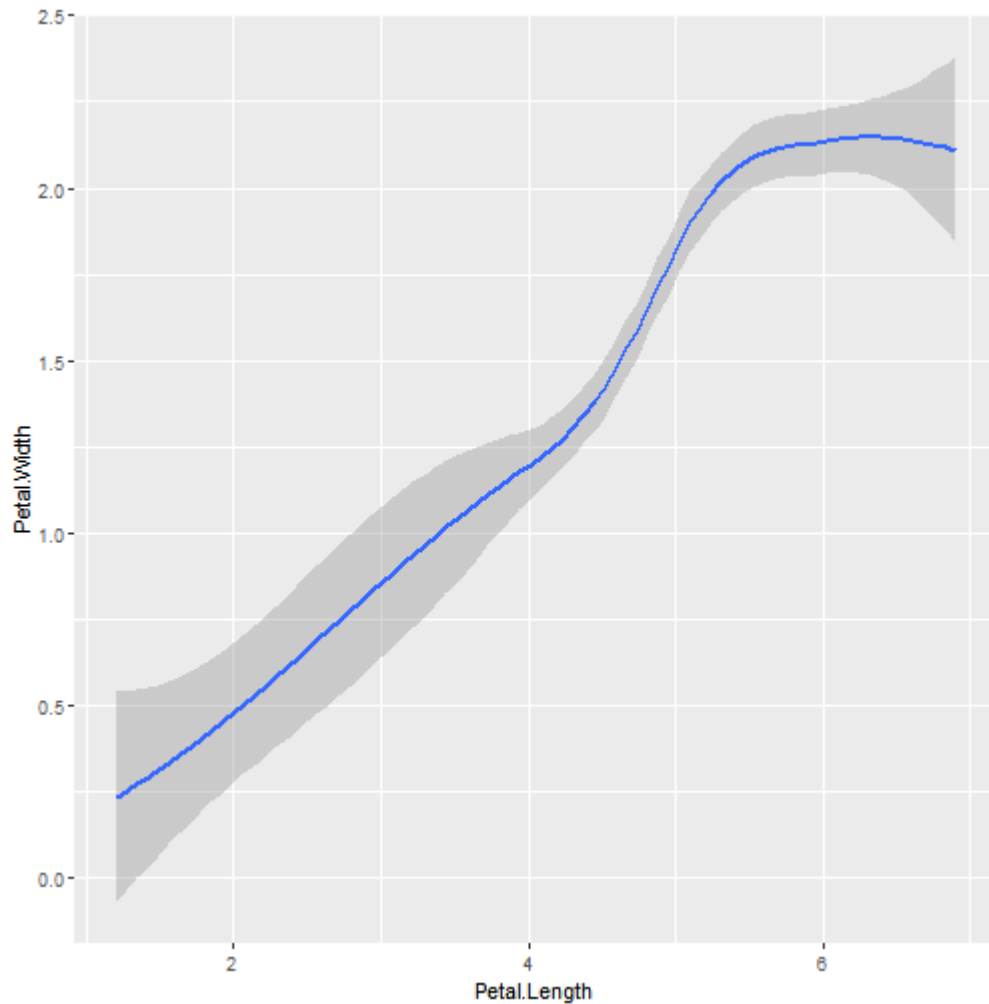
Gráfico de los cuantiles

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) +  
  geom_quantile(quantiles = c(0.2, 0.4, 0.6, 0.8))
```



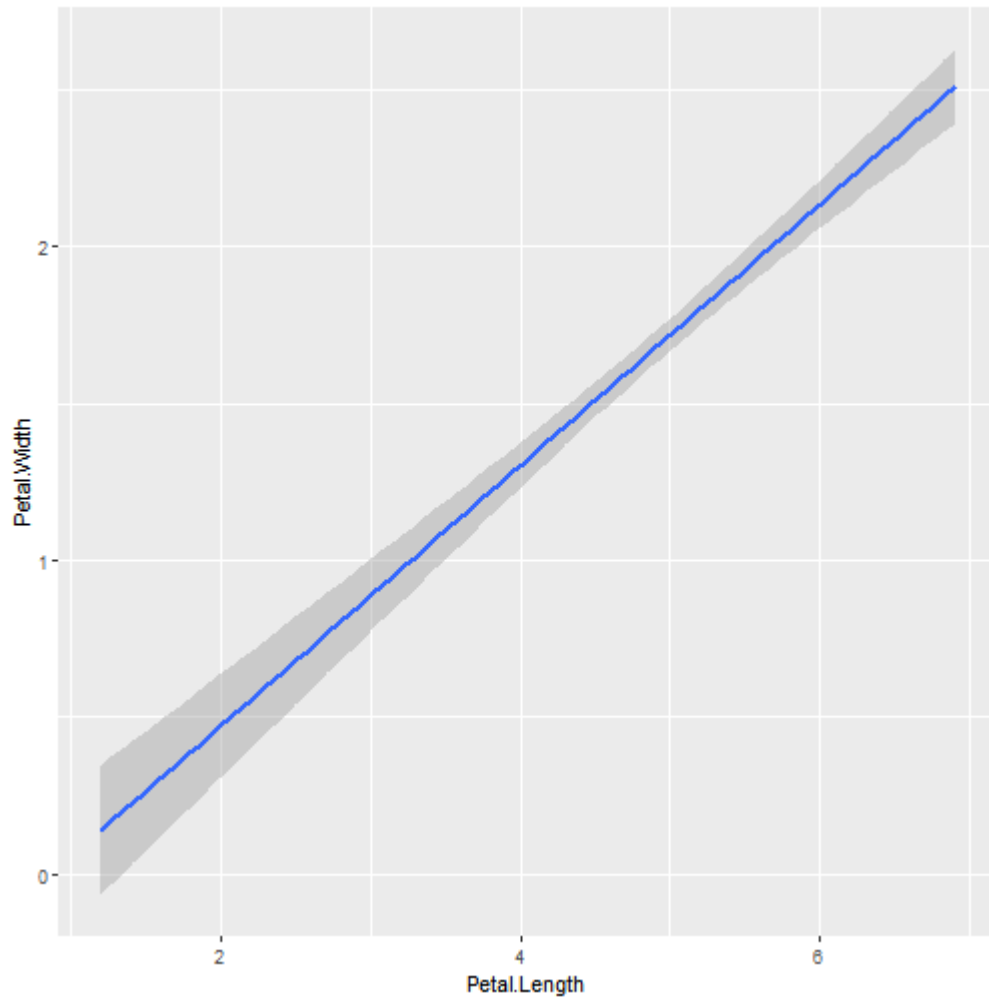
Mostrar líneas de regresión de los datos para ayudar a la identificación de patrones

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) + geom_smooth()
```



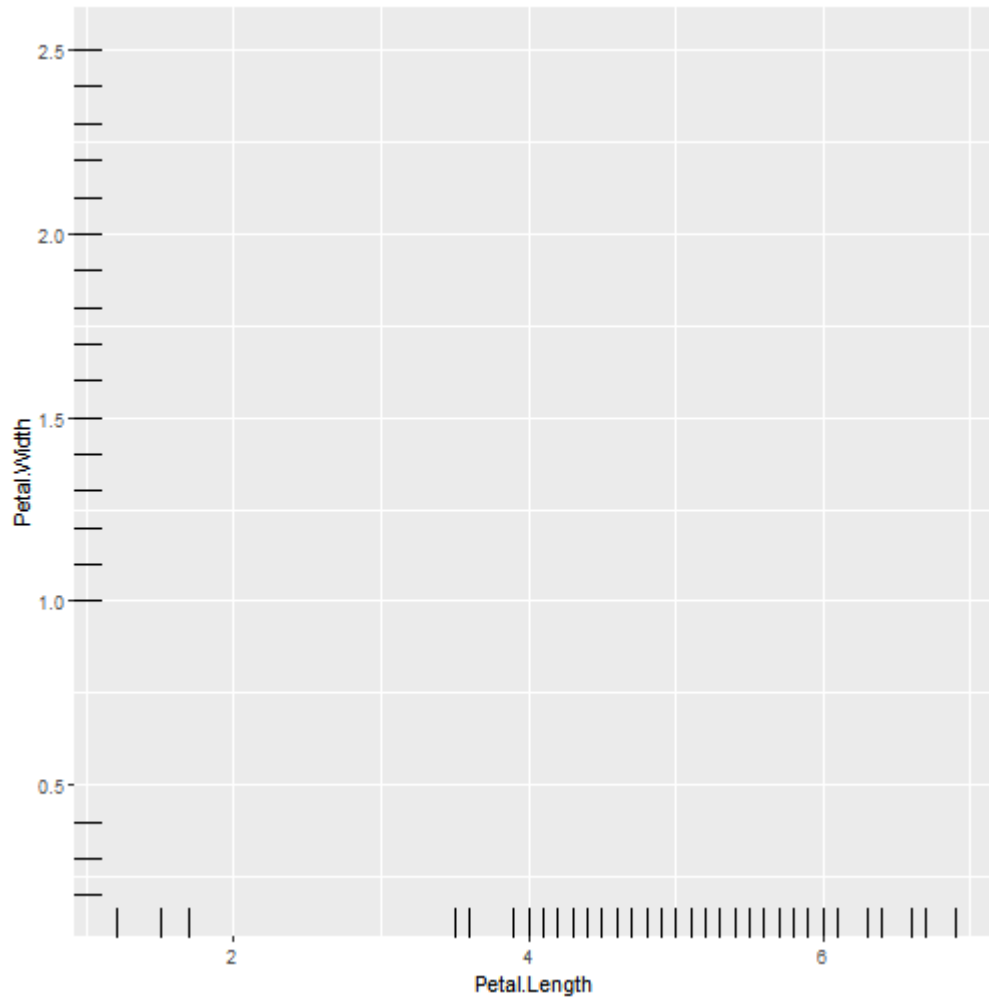
En el caso de querer una regresión lineal

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) +  
  geom_smooth(method = "lm")
```



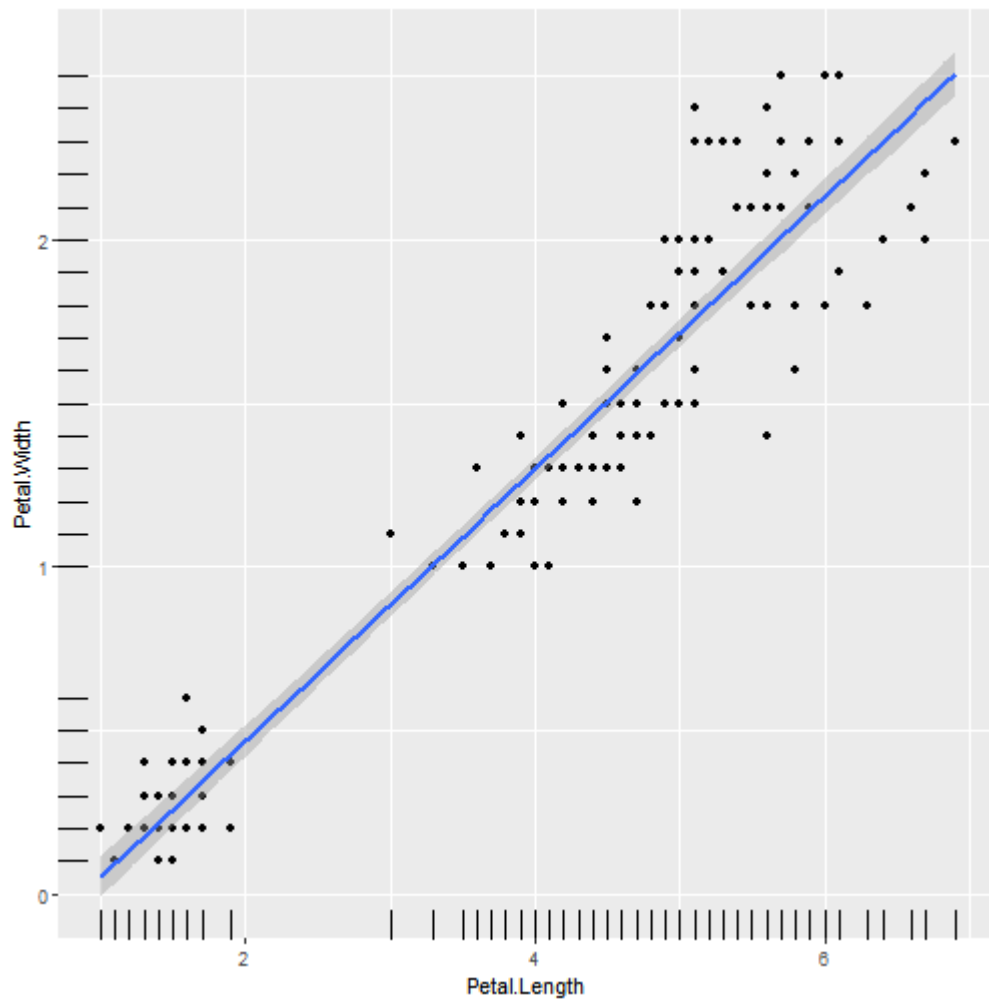
Graficar las distribuciones de ambas variables

```
ggplot(data = data2, mapping = aes(x = Petal.Length, y = Petal.Width)) + geom_rug()
```



Combinar capas

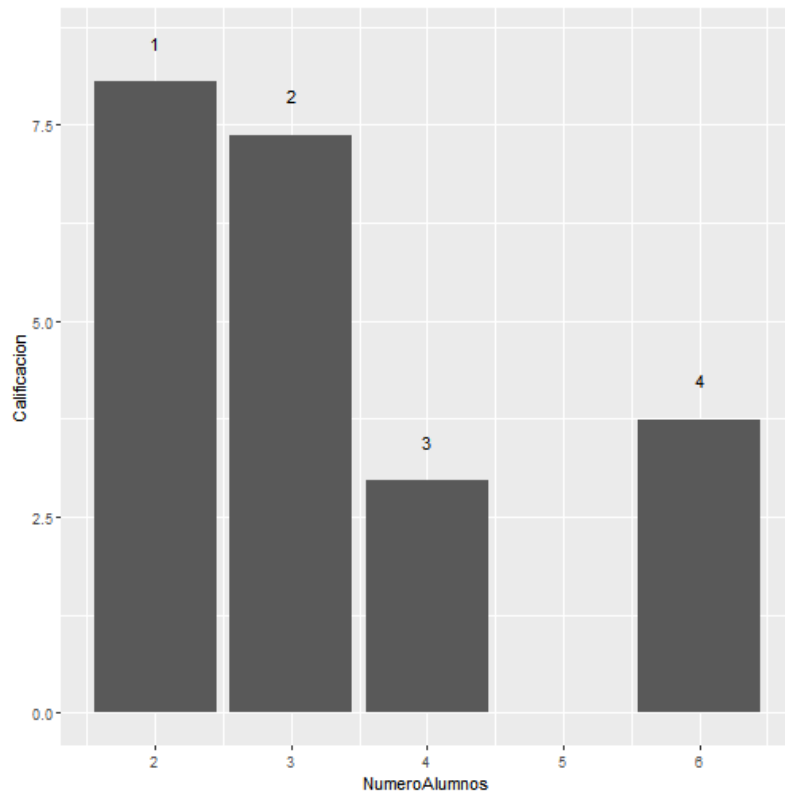
```
ggplot(data = iris, mapping = aes(x=Petal.Length, y=Petal.Width)) + geom_rug() +  
  geom_point() + geom_smooth(method="lm")
```



Agregar texto al gráfico

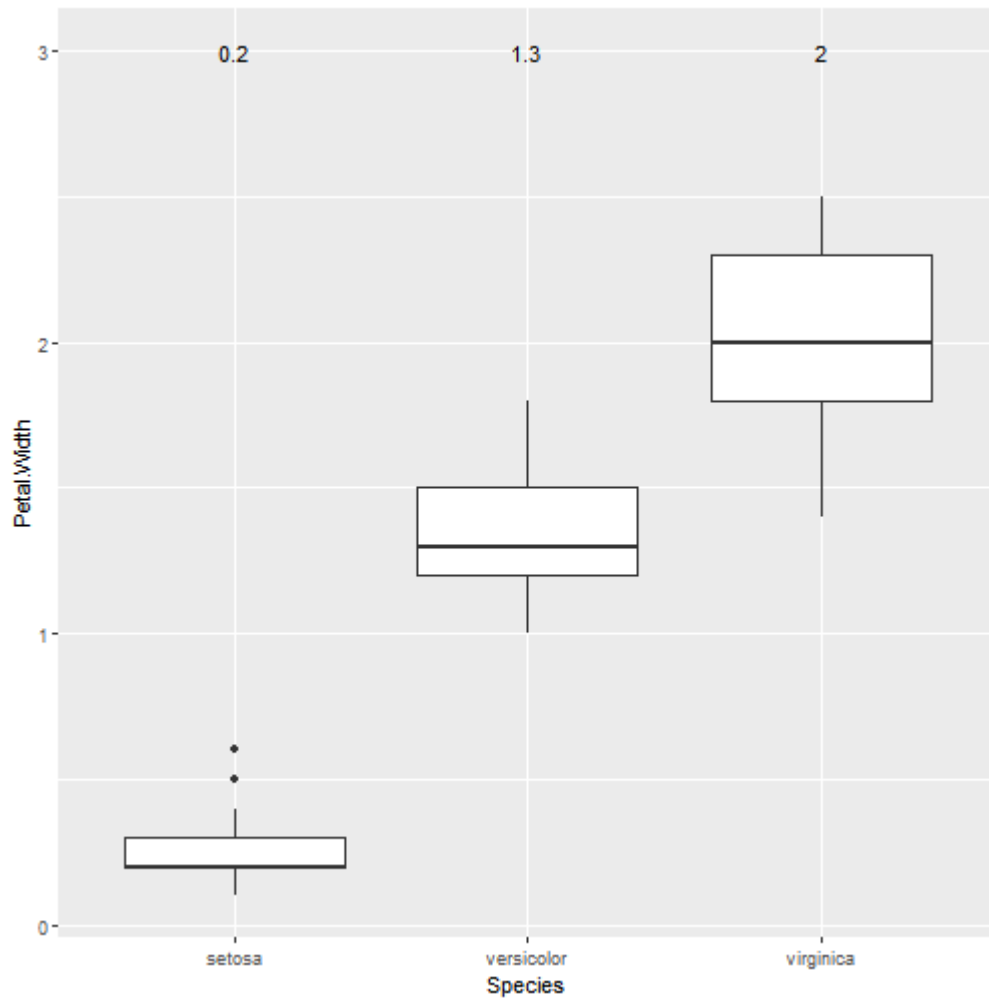
Usando geom_text

```
data = data.frame("NumeroAlumnos" = c(2, 3, 4, 6),  
                  "Calificacion" = runif(4, min = 0, max = 10))  
ggplot(data = data, mapping = aes(x=NumeroAlumnos, y=Calificacion)) +  
  geom_col() + geom_text(label = c("1", "2", "3", "4"), nudge_y = 0.5)
```



Usando annotate

```
ggplot(data = iris, mapping = aes(x=Species, y=Petal.Width)) + geom_boxplot() +  
  annotate(geom="text", x=c(1,2,3), y=c(3,3,3),label=c(0.2,1.3,2))
```



Cargar los datos y explorarlos

```
data("mpg")
```

```
head(mpg)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv   cty   hwy fl   class
##   <chr>         <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)  f      18    29 p   compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f      21    29 p   compa~
## 3 audi         a4      2    2008     4 manual(m6) f      20    31 p   compa~
## 4 audi         a4      2    2008     4 auto(av)   f      21    30 p   compa~
## 5 audi         a4      2.8  1999     6 auto(l5)  f      16    26 p   compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f      18    26 p   compa~
```

```
summary(mpg)
```

```
## manufacturer      model      displ      year
## Length:234      Length:234      Min.   :1.600      Min.   :1999
## Class :character  Class :character  1st Qu.:2.400      1st Qu.:1999
## Mode  :character  Mode  :character  Median :3.300      Median :2004
##                                     Mean   :3.472      Mean   :2004
##                                     3rd Qu.:4.600      3rd Qu.:2008
##                                     Max.   :7.000      Max.   :2008
##      cyl      trans      drv      cty
## Min.   :4.000      Length:234      Length:234      Min.   : 9.00
## 1st Qu.:4.000      Class :character  Class :character  1st Qu.:14.00
## Median :6.000      Mode  :character  Mode  :character  Median :17.00
## Mean   :5.889                                     Mean   :16.86
## 3rd Qu.:8.000                                     3rd Qu.:19.00
## Max.   :8.000                                     Max.   :35.00
##      hwy      fl      class
## Min.   :12.00      Length:234      Length:234
## 1st Qu.:18.00      Class :character  Class :character
## Median :24.00      Mode  :character  Mode  :character
## Mean   :23.44
## 3rd Qu.:27.00
## Max.   :44.00
```

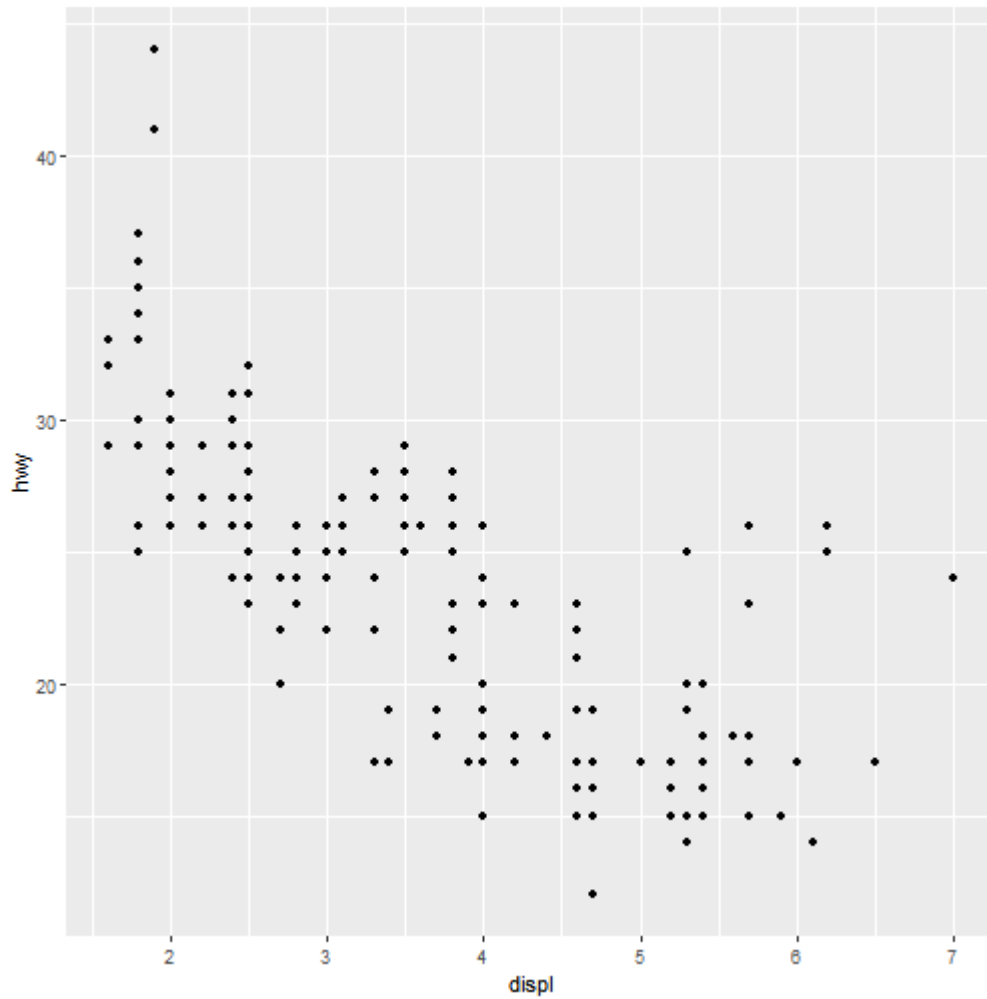
Características de las variables

```
str(mpg)
```

```
## tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
## $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
## $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
## $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
## $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
## $ drv         : chr [1:234] "f" "f" "f" "f" ...
## $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
## $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
## $ fl         : chr [1:234] "p" "p" "p" "p" ...
## $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

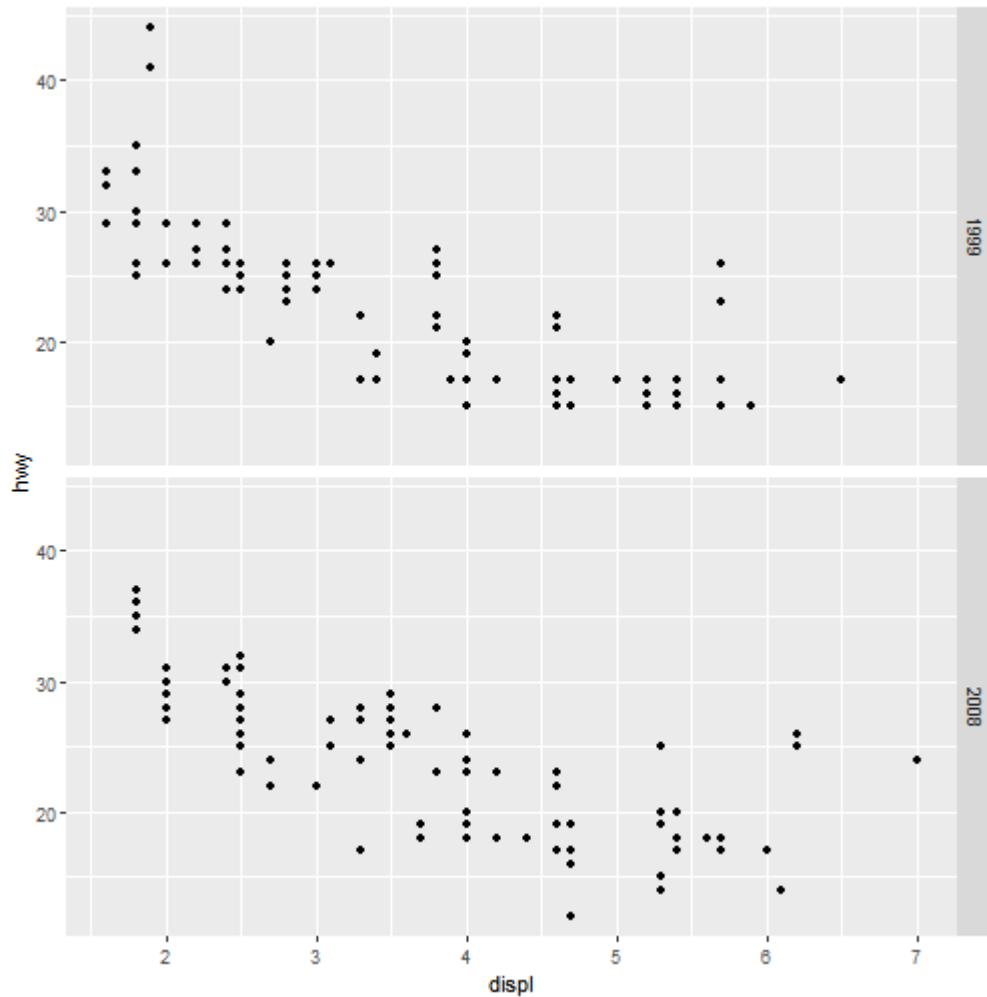

Faceting

```
ggplot(mpg, aes(x=displ, y=hwy)) + geom_point()
```



Mostrar la relación según el año de fabricación en horizontal

```
ggplot(mpg, aes(displ, hwy)) + geom_point() + facet_grid(rows = vars(year))
```



Mostrar la relación según el año de fabricación

```
ggplot(mpg, aes(displ, hwy)) + geom_point() + facet_grid(cols = vars(year))
```

