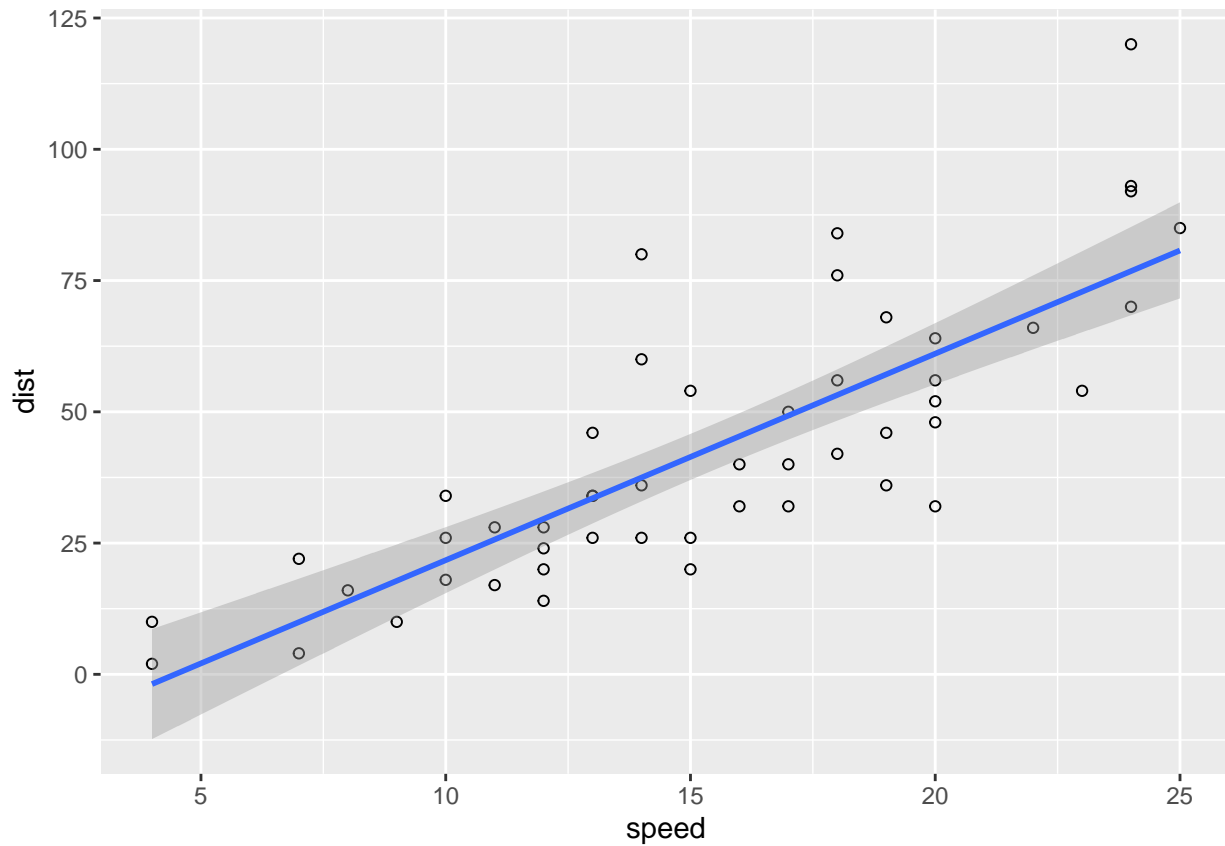


## R Exercício ggplot

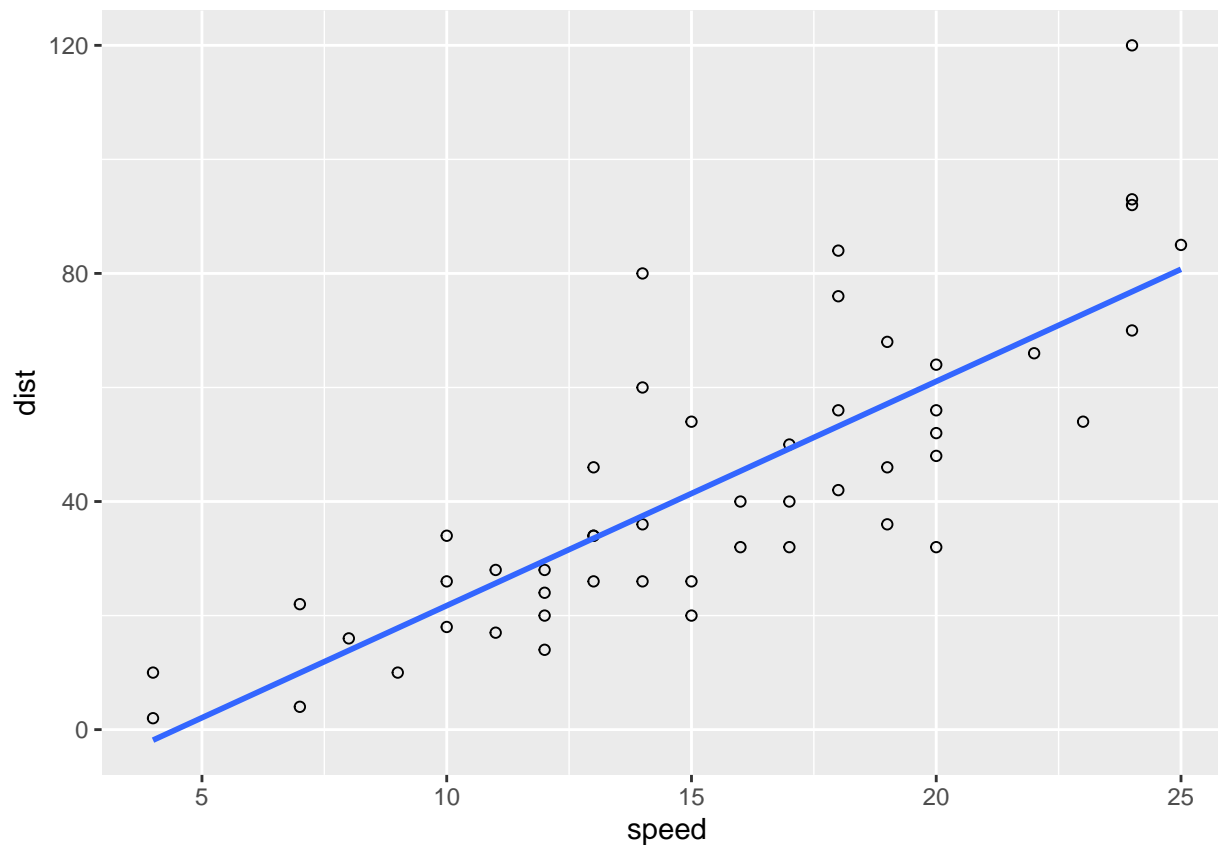
```
##### Gráficos de Dispersão#####  
#install.packages('ggplot2')  
library(ggplot2)  
data(cars)  
dta<- cars  
ggplot(dta,aes(speed,dist))+  
  geom_point(shape=1)+  
  geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
##### Gráficos de Dispersão (lm com intervalo de con)###  
ggplot(dta,aes(speed,dist))+  
  geom_point(shape=1)+  
  geom_smooth(method = lm, se=FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
data(diamonds)
```

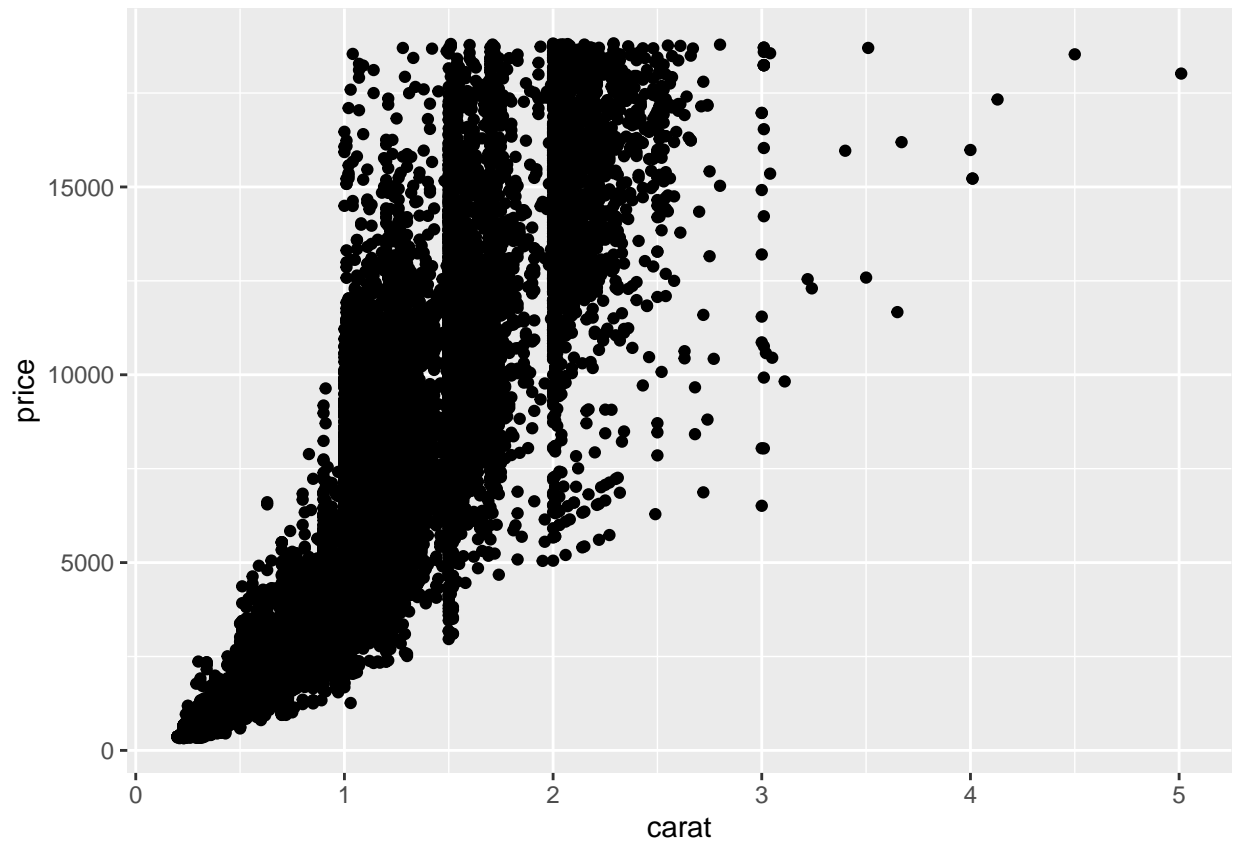
```
dtad<-diamonds
```

```
View(dtad)
```

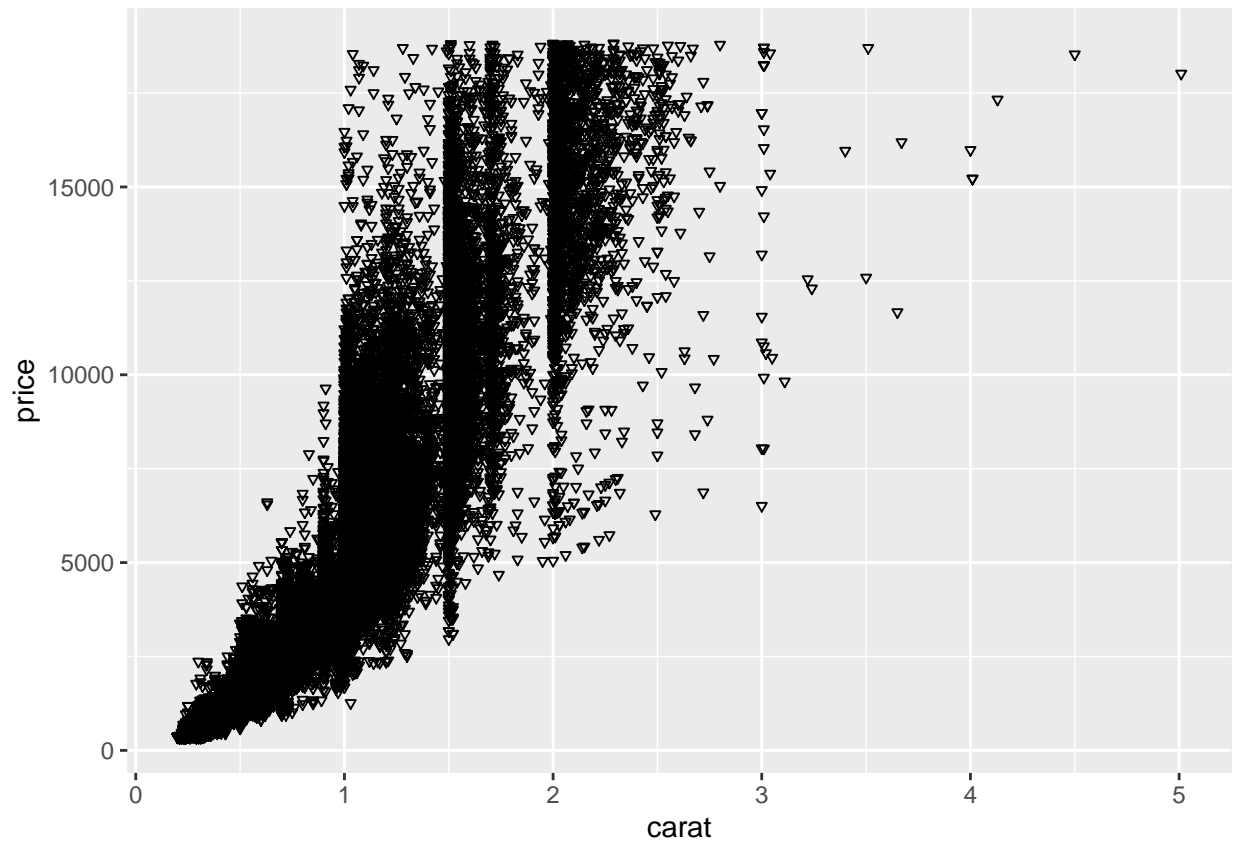
```
str(diamonds)
```

```
## tibble [53,940 x 10] (S3: tbl_df/tbl/data.frame)
## $ carat : num [1:53940] 0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...
## $ cut : Ord.factor w/ 5 levels "Fair"<"Good"<...: 5 4 2 4 2 3 3 3 1 3 ...
## $ color : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<...: 2 2 2 6 7 7 6 5 2 5 ...
## $ clarity: Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<...: 2 3 5 4 2 6 7 3 4 5 ...
## $ depth : num [1:53940] 61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...
## $ table : num [1:53940] 55 61 65 58 58 57 57 55 61 61 ...
## $ price : int [1:53940] 326 326 327 334 335 336 336 337 337 338 ...
## $ x : num [1:53940] 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...
## $ y : num [1:53940] 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...
## $ z : num [1:53940] 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...
```

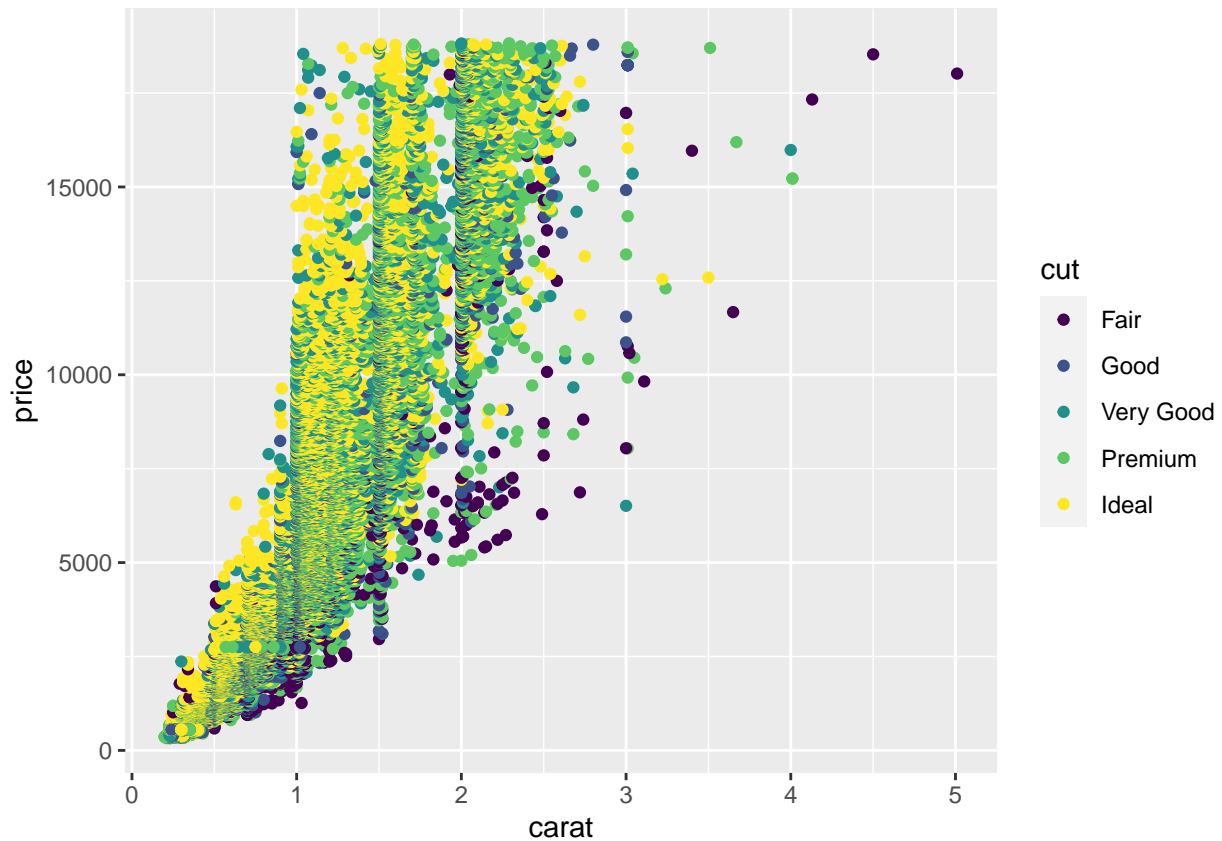
```
ggplot(dtad,aes(x=carat,y=price))+
  geom_point()
```



```
ggplot(dtad,aes(x=carat,y=price))+  
  geom_point(shape=25,size=1)
```



```
##### Gráficos de Pontos pintados em função de factores#####  
ggplot(dtad,aes(x=carat, y=price, colour=cut))+  
  geom_point(size=1.5)
```



##### Gráfico Histograma #####

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

```
dtac<- mpg
head(dtac)
```

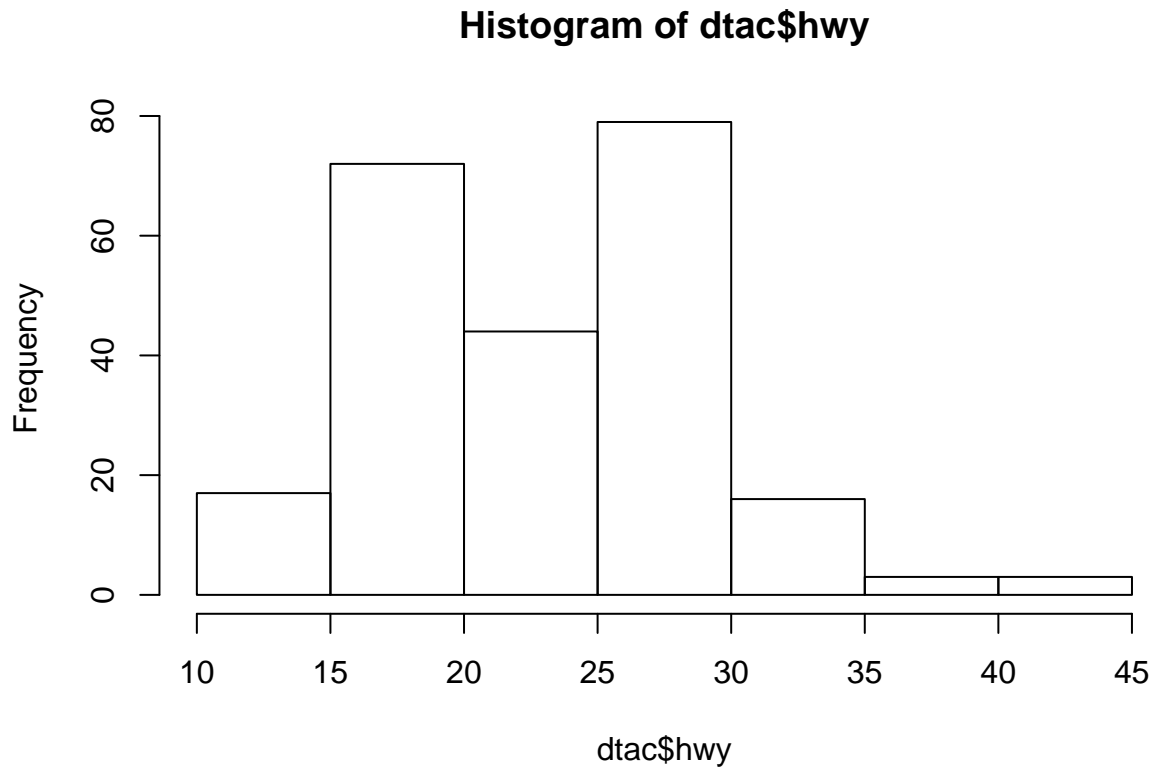
```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv   cty   hwy fl   class
##   <chr>          <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>
## 1 audi          a4      1.8  1999     4 auto(l5)  f      18    29 p   compa~
## 2 audi          a4      1.8  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~
```

```
str(dtac)
```

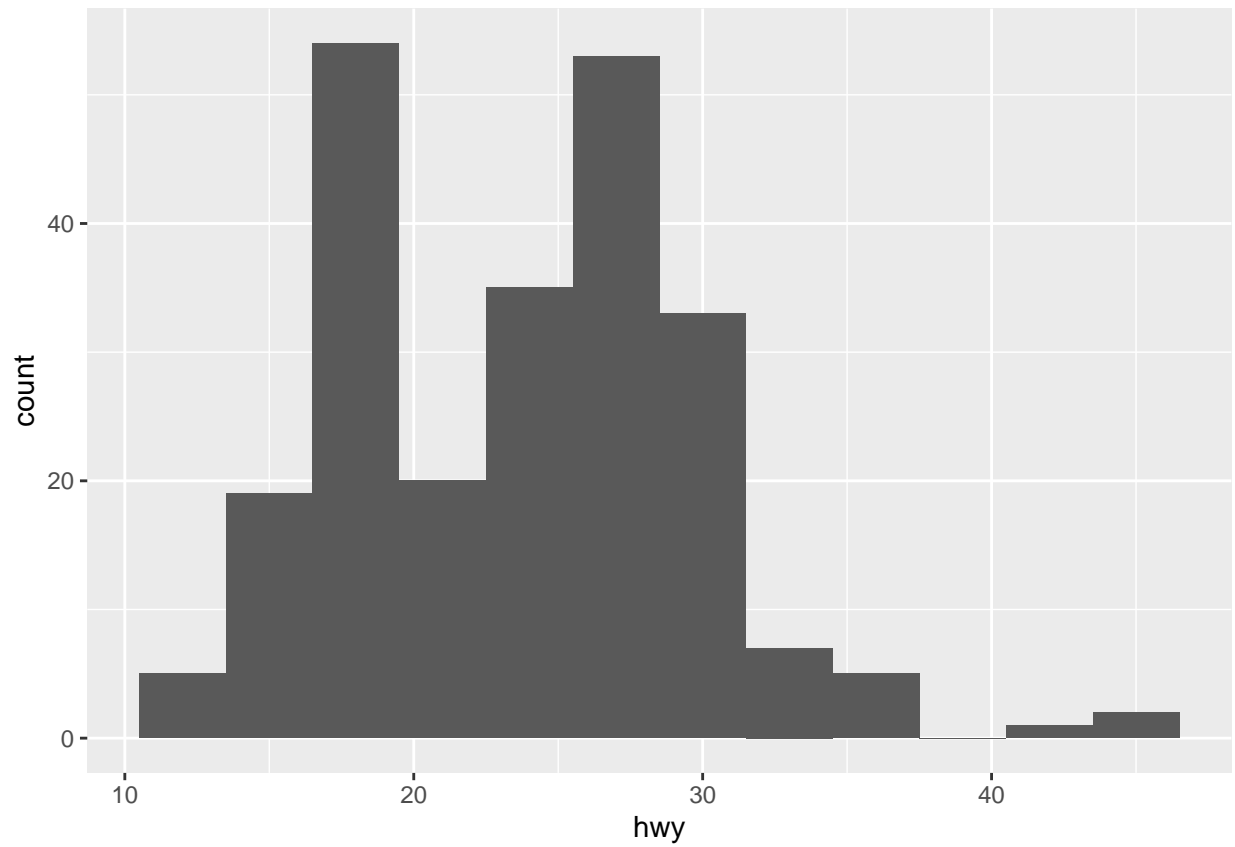
```
## tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
## $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
## $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
## $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
```

```
## $ trans      : chr [1:234] "auto(15)" "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" ...
```

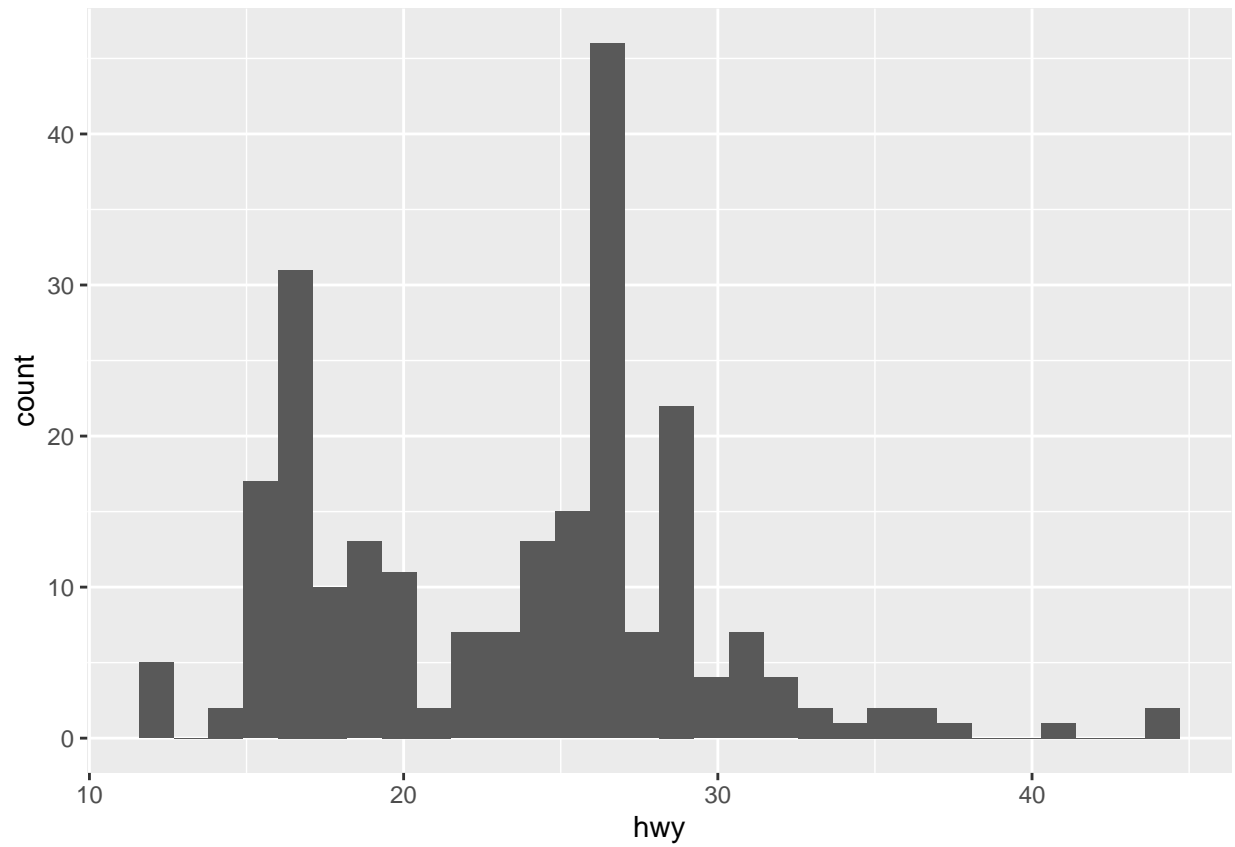
```
hist(dtac$hwy)
```



```
ggplot(dtac, aes(x=hwy)) +
  geom_histogram(binwidth= 3)
```

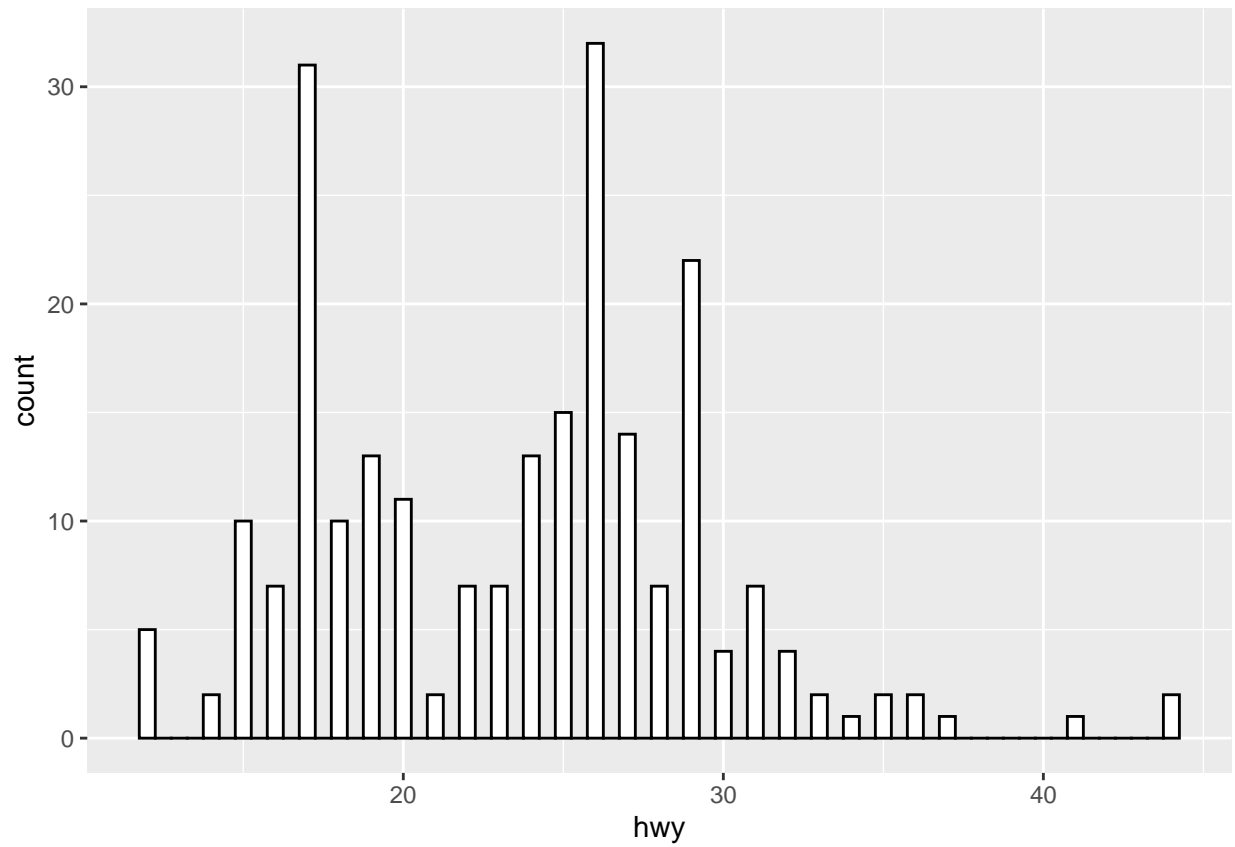


```
ggplot(dtac,aes(x=hwy))+  
  geom_histogram(bins = 30)
```



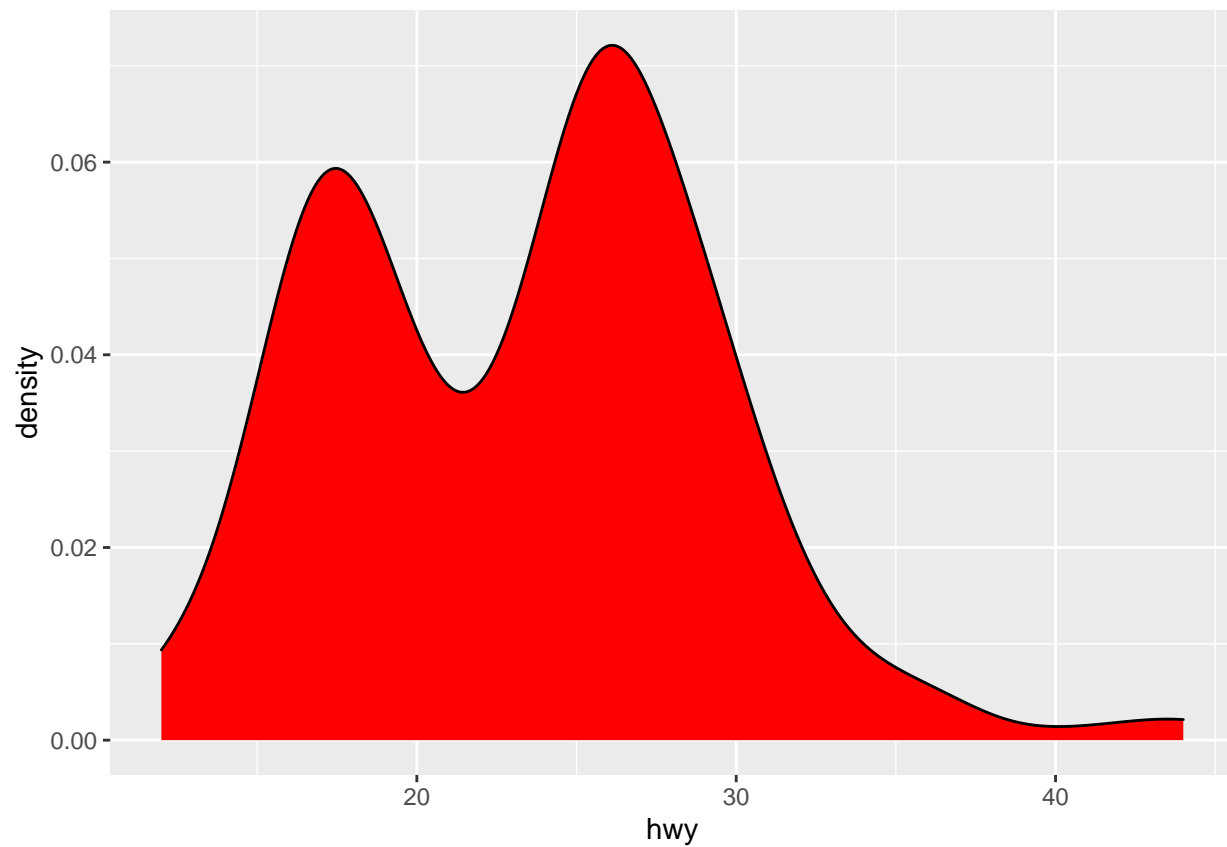
```
ggplot(dtac,aes(x=hwy))+  
  geom_histogram(binwidth = .5,colour= 'black', fill='white')
```



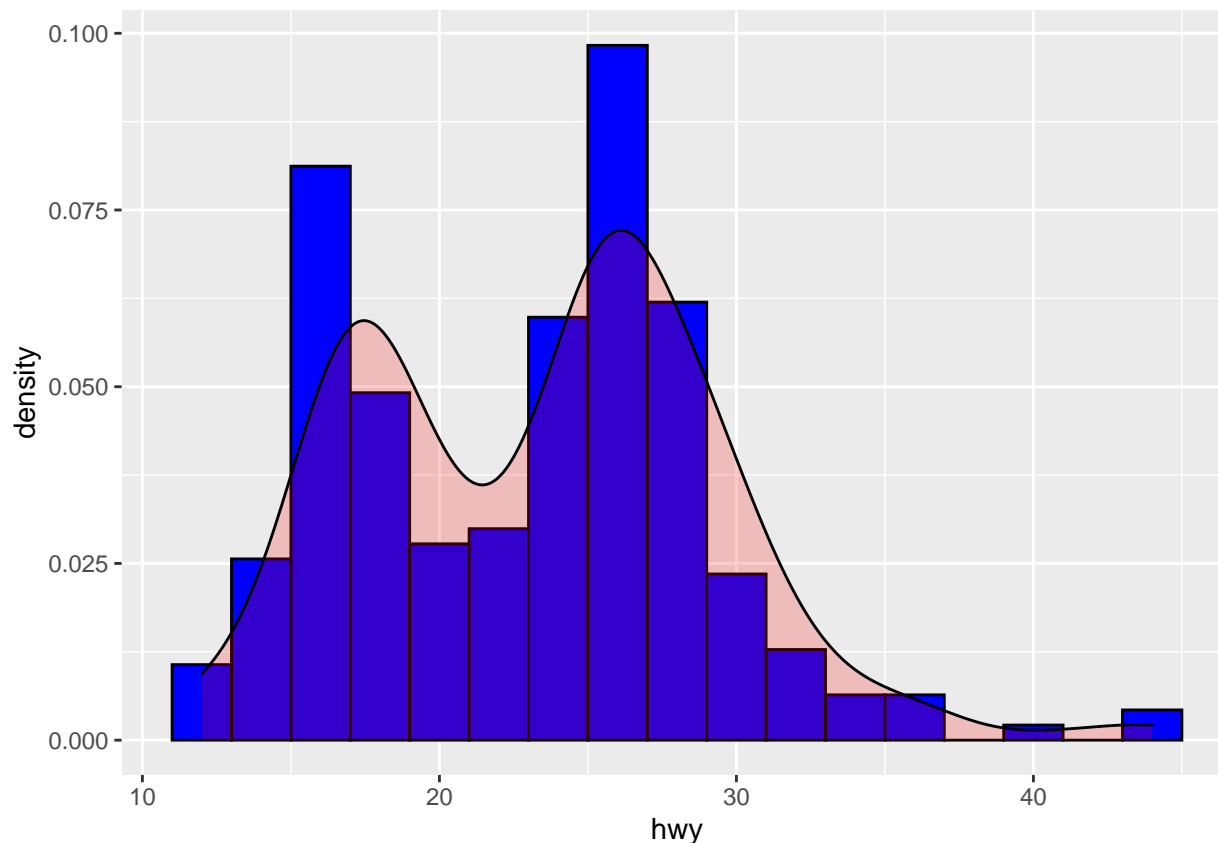


##### Gráfico Densidade #####

```
ggplot(dtac,aes(x=hwy))+  
  geom_density(fill='red')
```



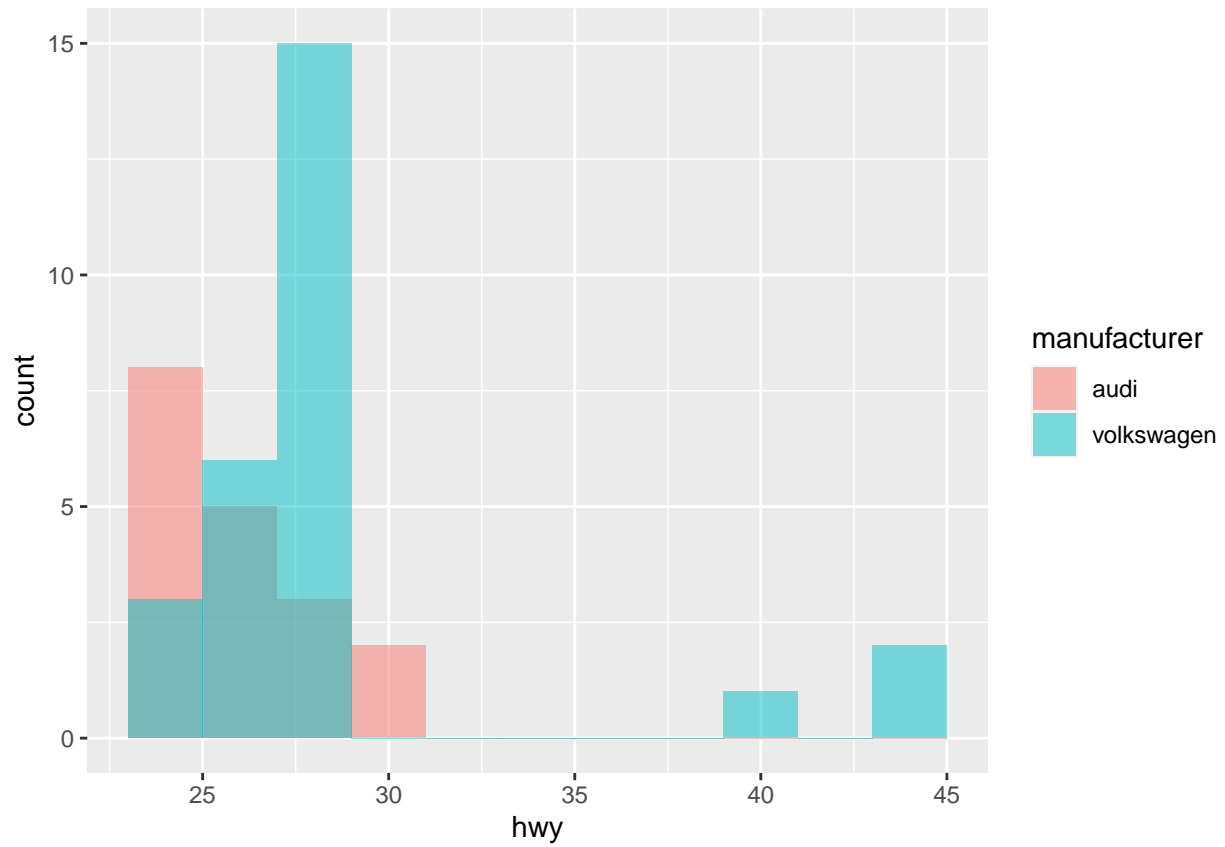
```
# histograma con curva de densidad
ggplot(mpg, aes(x=hwy))+
  geom_histogram(aes(y=..density..),
    binwidth=2,
    colour="black", fill="blue") +
  geom_density(alpha=.2, fill="red")#rellenado de la curva de densidad en rojo suave
```



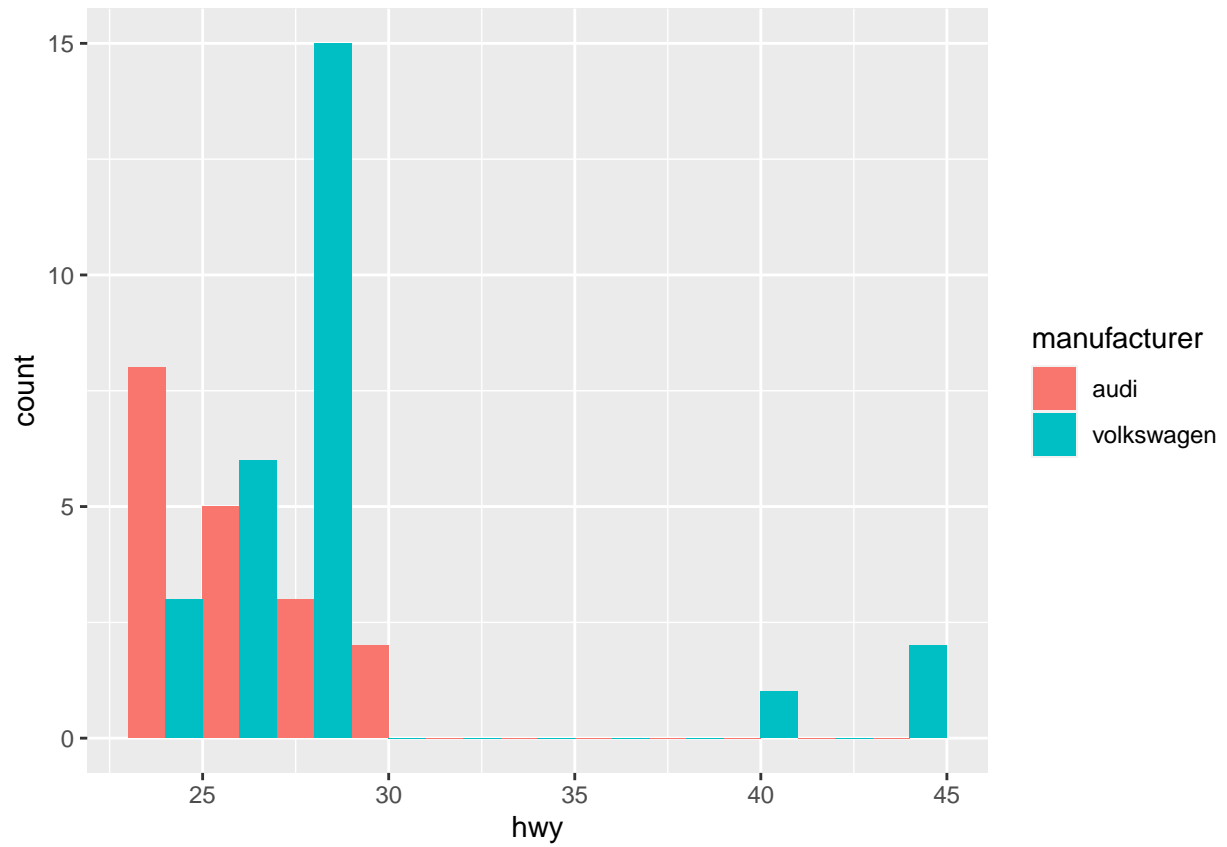
```
mpg_subset <- mpg[mpg$manufacturer=="audi" | mpg$manufacturer=="volkswagen",]
str(mpg_subset)
```

```
## tibble [45 x 11] (S3: tbl_df/tbl/data.frame)
## $ manufacturer: chr [1:45] "audi" "audi" "audi" "audi" ...
## $ model       : chr [1:45] "a4" "a4" "a4" "a4" ...
## $ displ       : num [1:45] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year        : int [1:45] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
## $ cyl         : int [1:45] 4 4 4 4 6 6 6 4 4 4 ...
## $ trans       : chr [1:45] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
## $ drv         : chr [1:45] "f" "f" "f" "f" ...
## $ cty         : int [1:45] 18 21 20 21 16 18 18 18 16 20 ...
## $ hwy         : int [1:45] 29 29 31 30 26 26 27 26 25 28 ...
## $ fl          : chr [1:45] "p" "p" "p" "p" ...
## $ class       : chr [1:45] "compact" "compact" "compact" "compact" ...
```

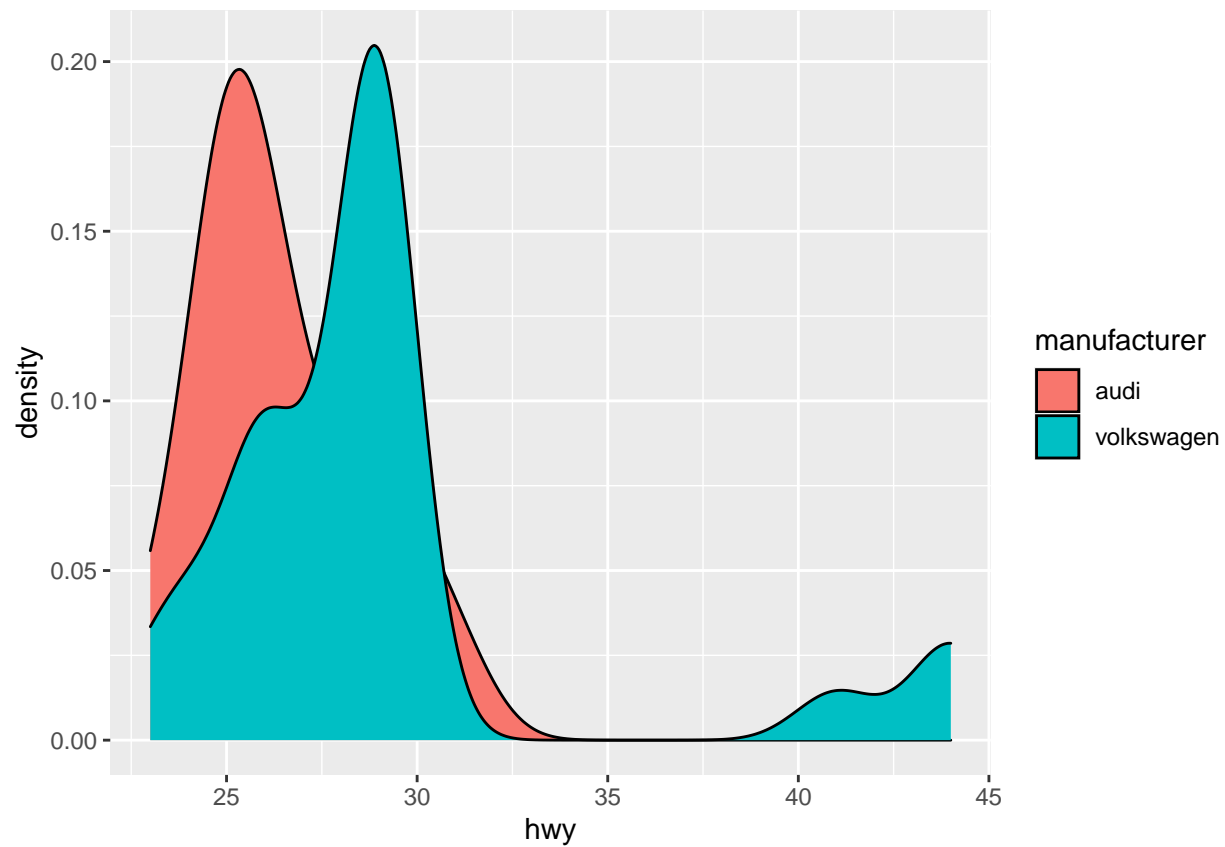
```
ggplot(mpg_subset, aes(x=hwy, fill=manufacturer))+
  geom_histogram(binwidth=2, alpha=.5, position="identity")
```



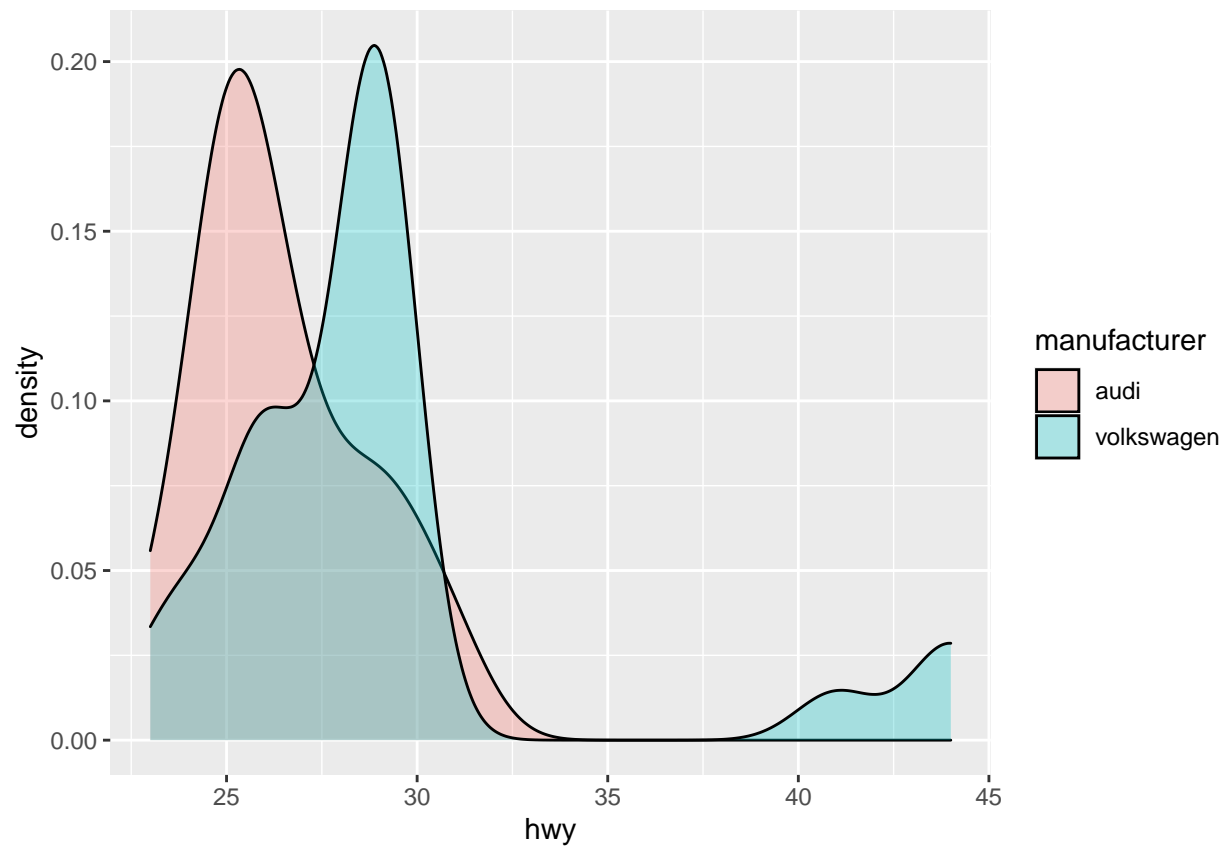
```
# histogramas intercalados
ggplot(mpg_subset, aes(x=hwy, fill=manufacturer)) +
  geom_histogram(binwidth=2, position="dodge")
```



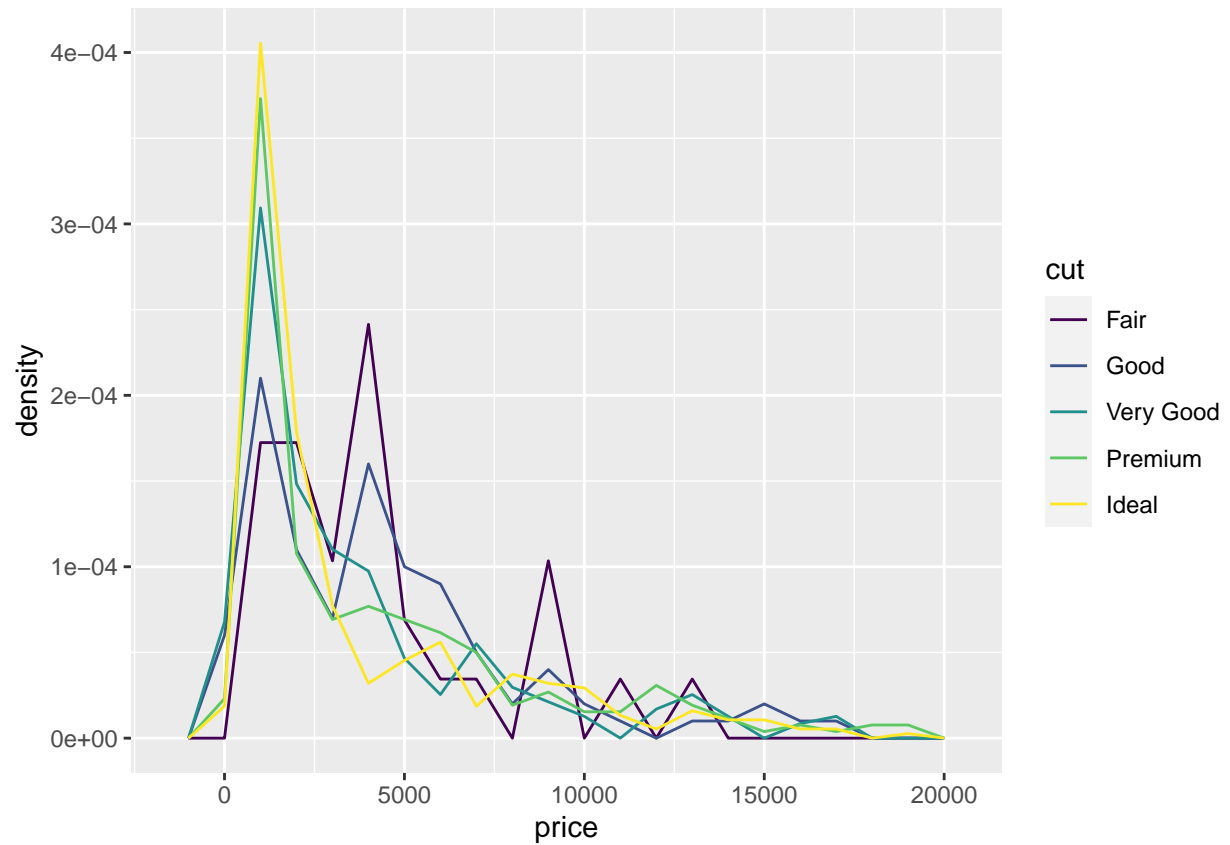
```
# curvas de densidad
ggplot(mpg_subset, aes(x=hwy, fill=manufacturer))+
  geom_density()
```



```
# curvas de densidad con relleno transparente  
ggplot(mpg_subset, aes(x=hwy, fill=manufacturer)) +  
  geom_density(alpha=.3)
```

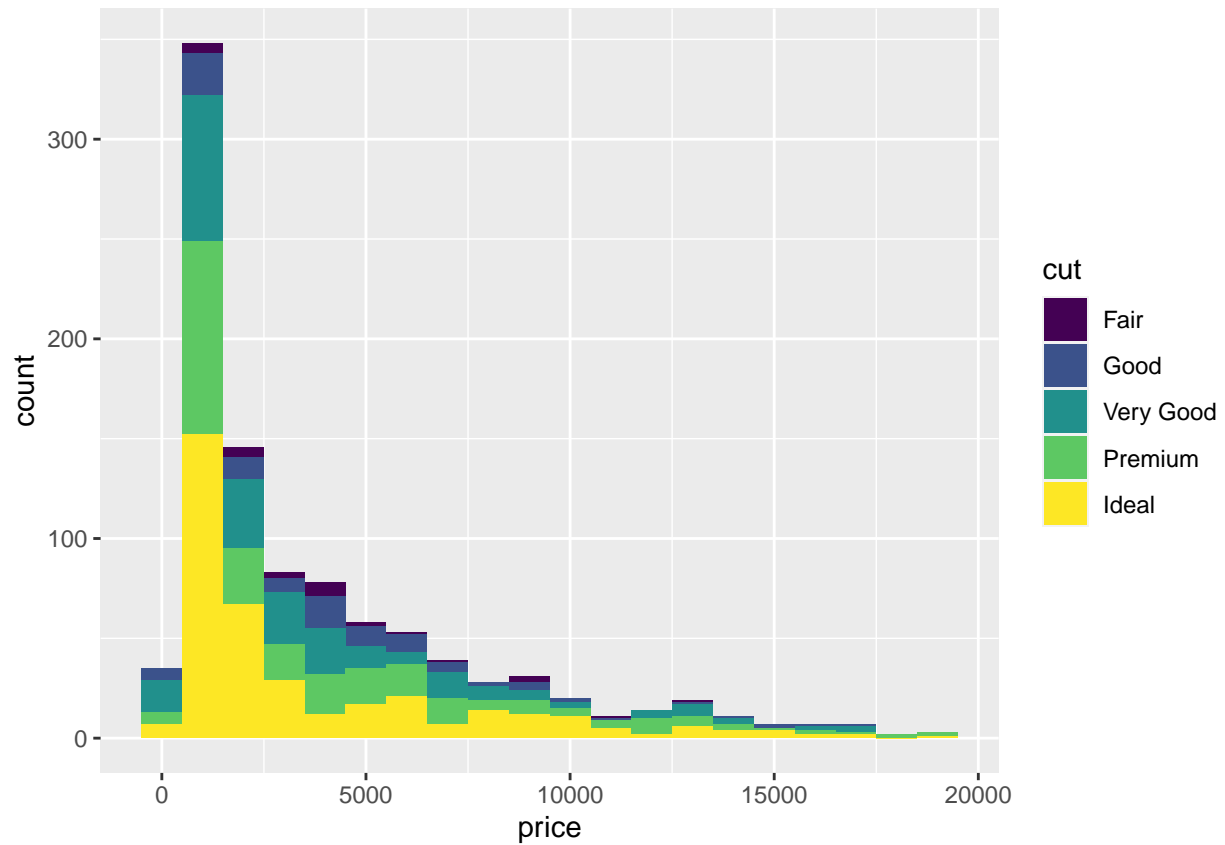


```
#polinomios de frecuencias con colores en función de un factor  
  
#histogramas más sofisticados  
set.seed(6298)  
diamonds_small <- diamonds[sample(nrow(diamonds), 1000), ]  
ggplot(diamonds_small, aes(price, ..density.., colour = cut))+  
  geom_freqpoly(binwidth = 1000)
```

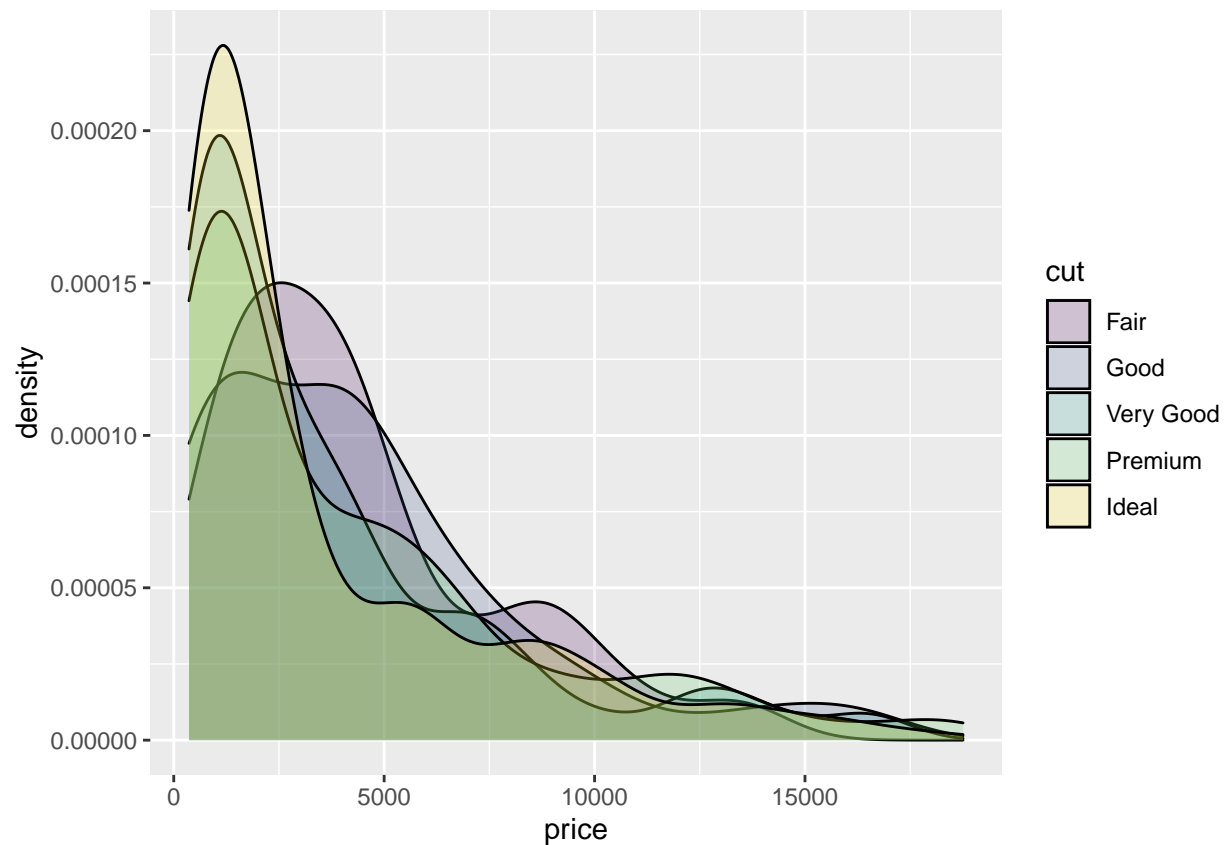


```
#más opciones, histogramas solapados con distintos colores en función de un factor  
hist_cut <- ggplot(diamonds_small, aes(x = price, fill = cut))  
hist_cut + geom_histogram(binwidth = 1000)
```





```
#gráficos de densidad con diferentes factores marcando colores  
ggplot(diamonds_small, aes(price, fill = cut))+  
  geom_density(alpha = 0.2)
```



##### Gráfico de Barras#####

```
#install.packages("reshape2")
```

```
library(reshape2)
```

```
library(ggplot2)
```

```
data(tipes)
```

```
## Warning in data(tipes): data set 'tipes' not found
```

```
dtat<- tips
```

```
head(dtat)
```

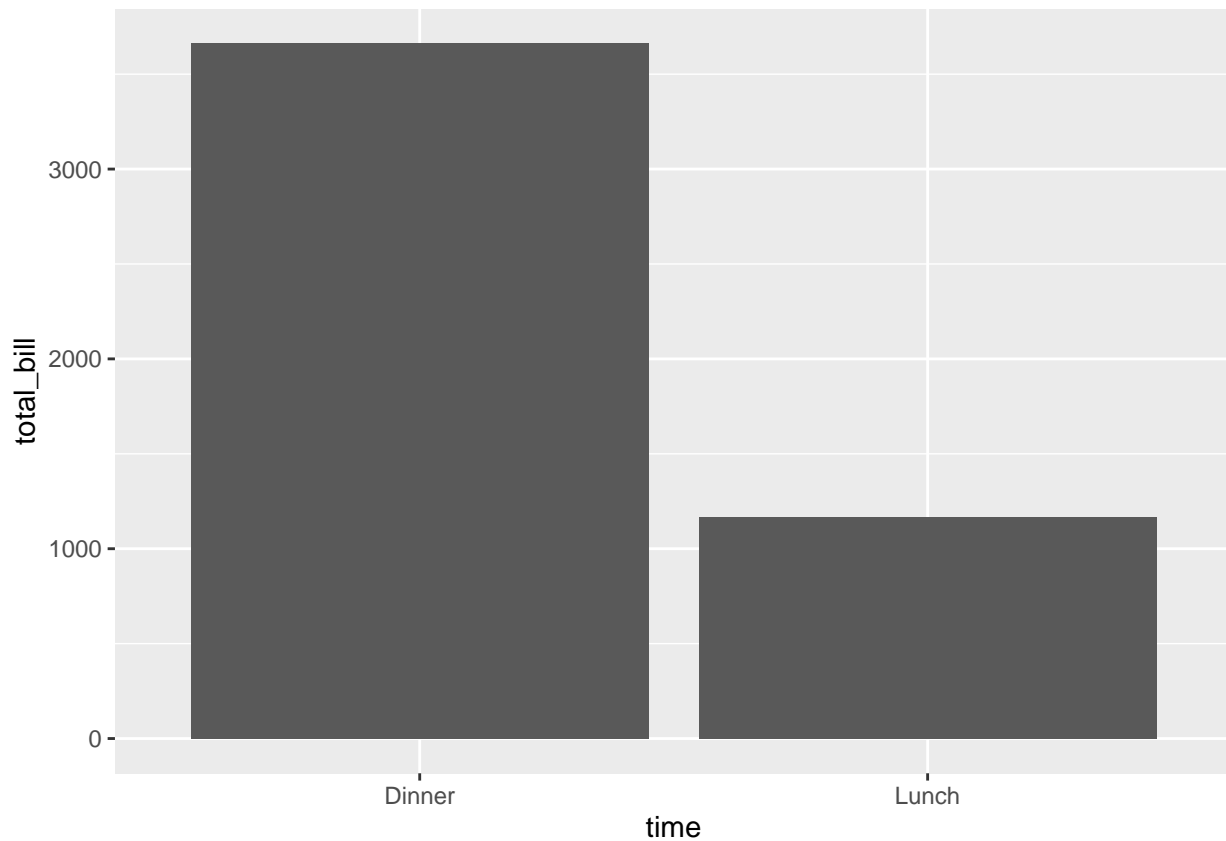
```
##   total_bill  tip  sex smoker day  time size
## 1    16.99  1.01 Female    No  Sun  Dinner    2
## 2    10.34  1.66   Male    No  Sun  Dinner    3
## 3    21.01  3.50   Male    No  Sun  Dinner    3
## 4    23.68  3.31   Male    No  Sun  Dinner    2
## 5    24.59  3.61 Female    No  Sun  Dinner    4
## 6    25.29  4.71   Male    No  Sun  Dinner    4
```

```
str(dtat)
```

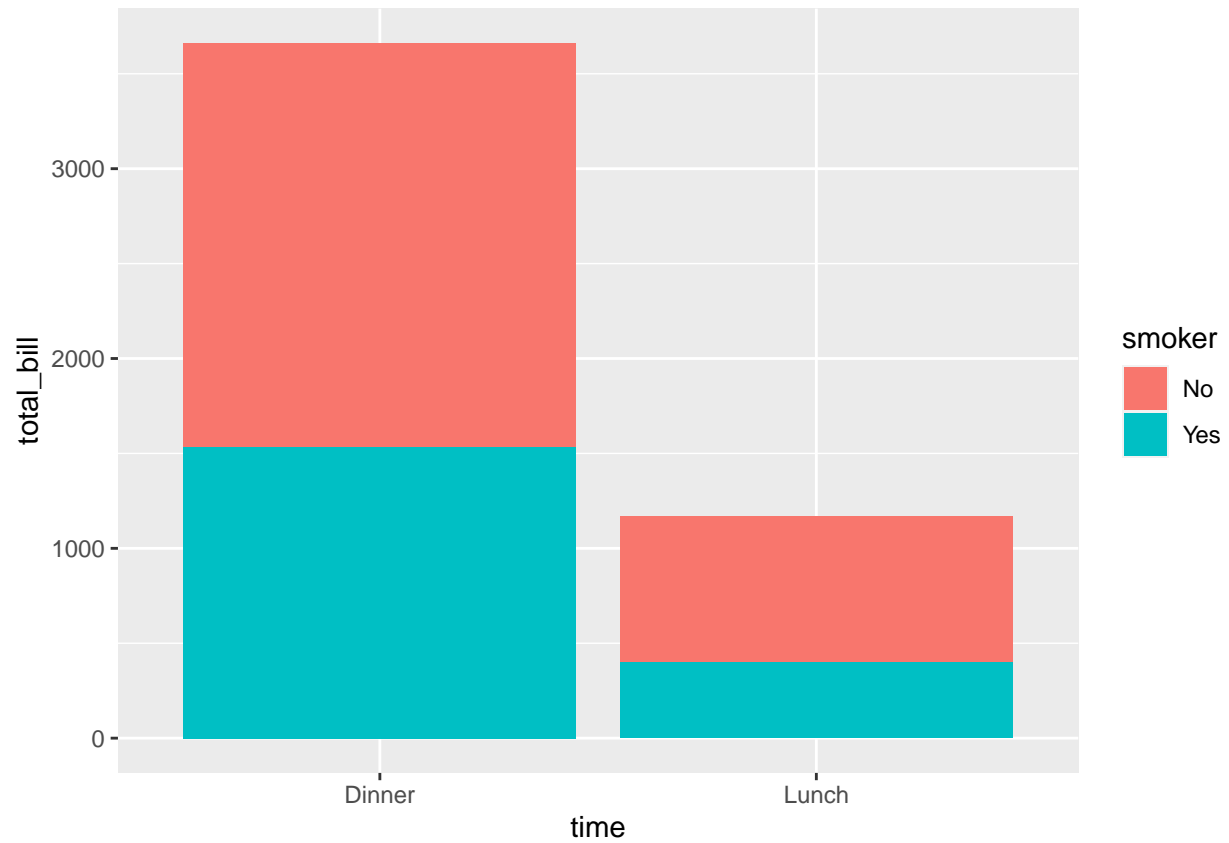
```
## 'data.frame':   244 obs. of  7 variables:
```

```
## $ total_bill: num 17 10.3 21 23.7 24.6 ...
## $ tip       : num 1.01 1.66 3.5 3.31 3.61 4.71 2 3.12 1.96 3.23 ...
## $ sex       : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 2 2 2 2 2 ...
## $ smoker    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ day       : Factor w/ 4 levels "Fri","Sat","Sun",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ time      : Factor w/ 2 levels "Dinner","Lunch": 1 1 1 1 1 1 1 1 1 1 ...
## $ size      : int 2 3 3 2 4 4 2 4 2 2 ...
```

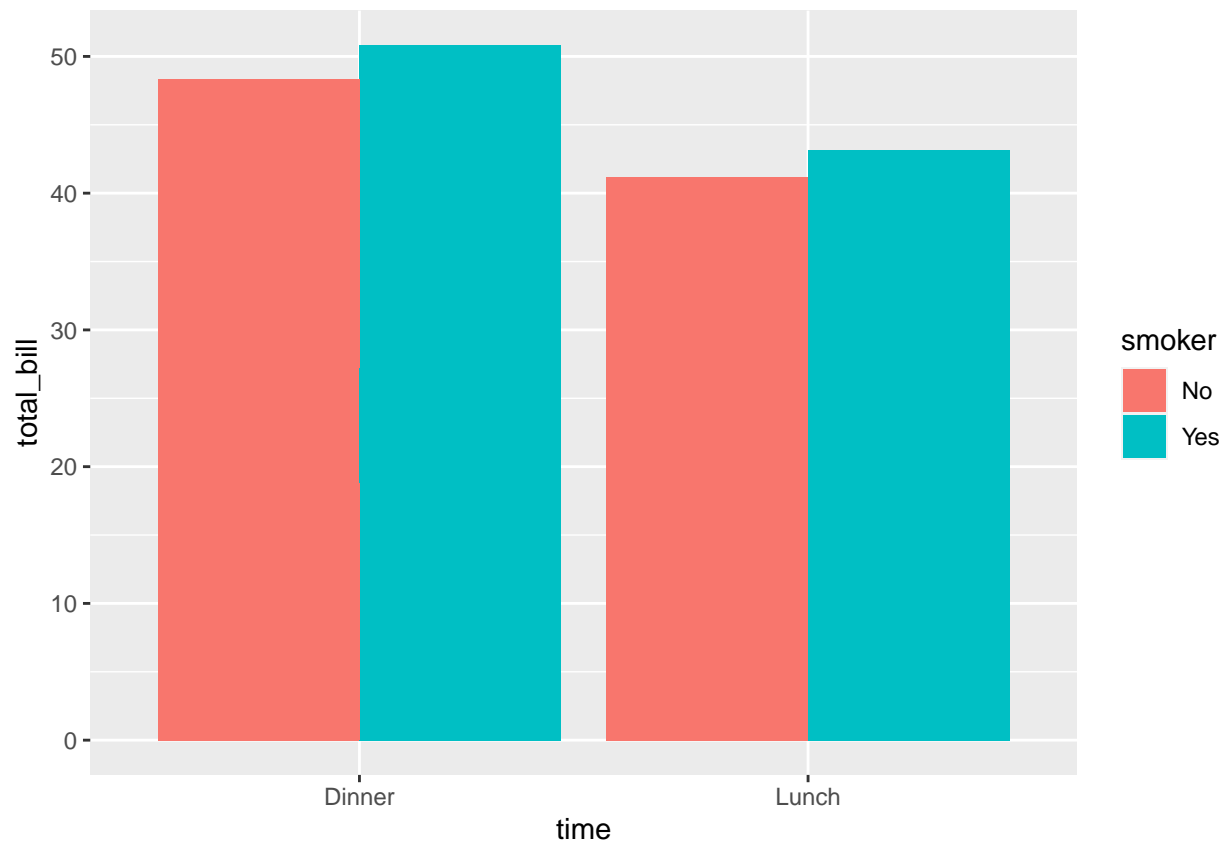
```
ggplot(dtat,aes(x=time,y=total_bill))+
  geom_bar(stat = 'identity')
```



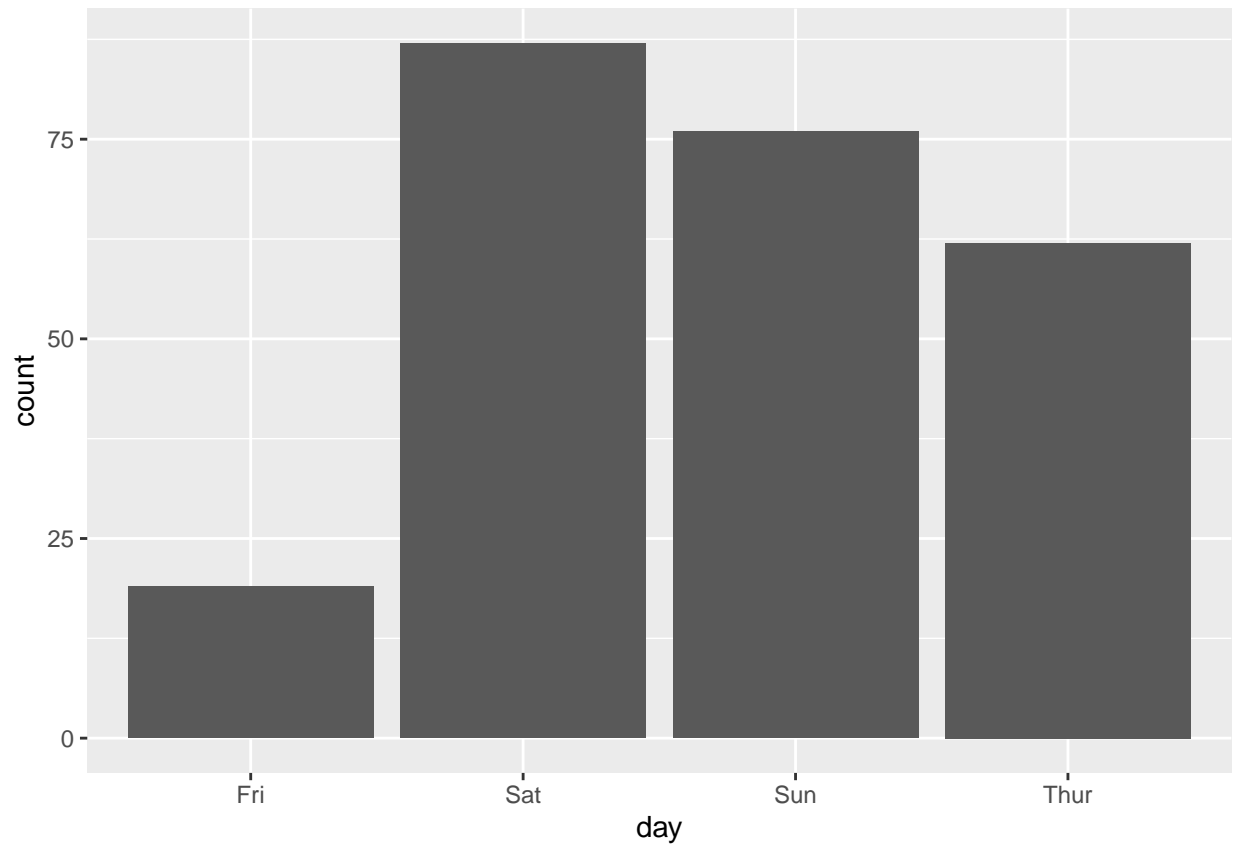
```
# Por categoria smoker
ggplot(dtat,aes(x=time,y=total_bill, fill=smoker))+
  geom_bar(stat = 'identity')
```



```
##### Position_dodge #####  
ggplot(dtat,aes(x=time,y=total_bill, fill=smoker))+  
  geom_bar(stat = 'identity',position=position_dodge())
```



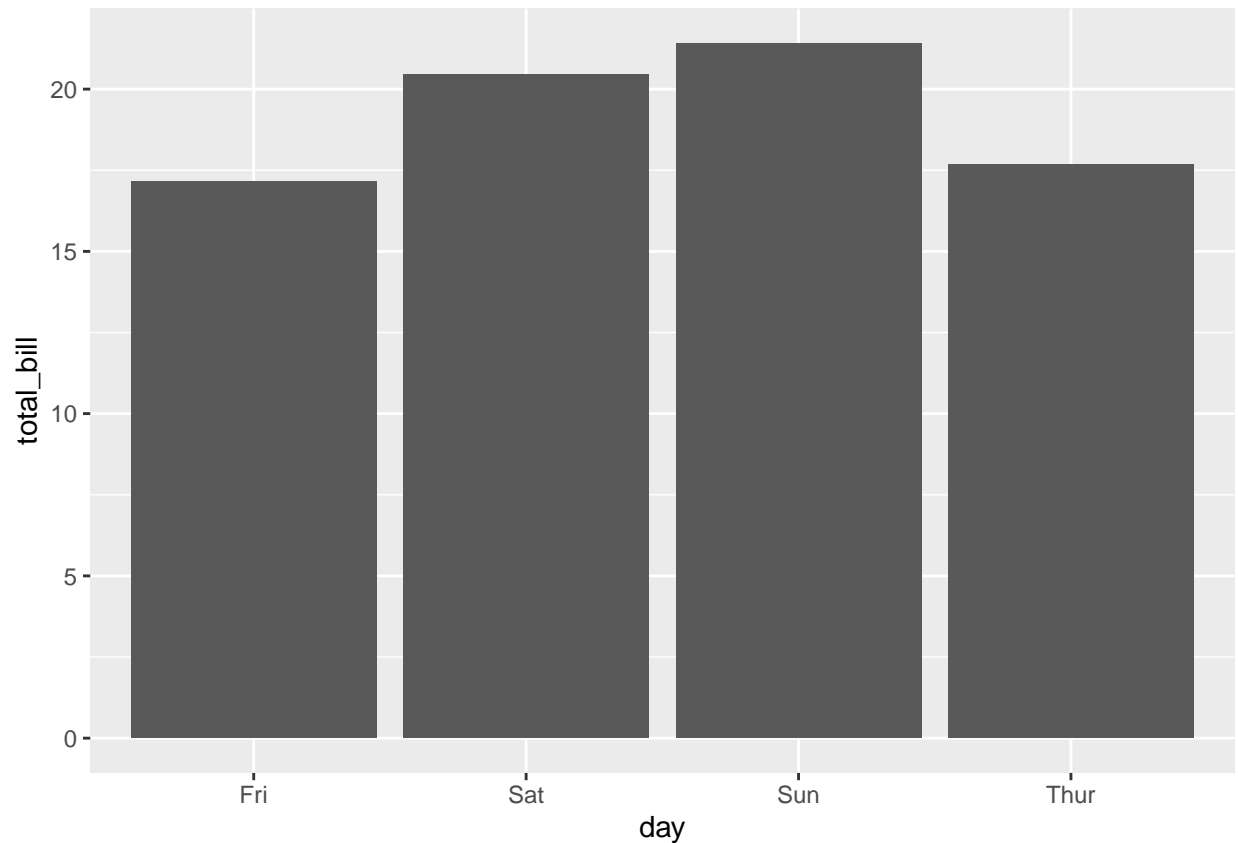
```
#Para obter um gráfico de barras de contagens, não mapeie uma variável para ye use stat="bin"(que é o p  
#cuenta de clientes por día  
ggplot(dtat,aes(x=day))+  
  geom_bar(stat='count')
```



```
#cuenta de media de facturas por día  
ggplot(dtat,aes(x=day,y=total_bill))+  
  geom_bar(stat = 'summary',fun.y='mean')
```

```
## Warning: Ignoring unknown parameters: fun.y
```

```
## No summary function supplied, defaulting to `mean_se()`
```



##### Gráficos de factor con barras de error #####

```
data(ToothGrowth)
head(ToothGrowth)
```

```
##      len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
```

```
tg<-ToothGrowth
```

```
#install.packages("Rmisc")
library(Rmisc)
```

```
## Loading required package: lattice
```

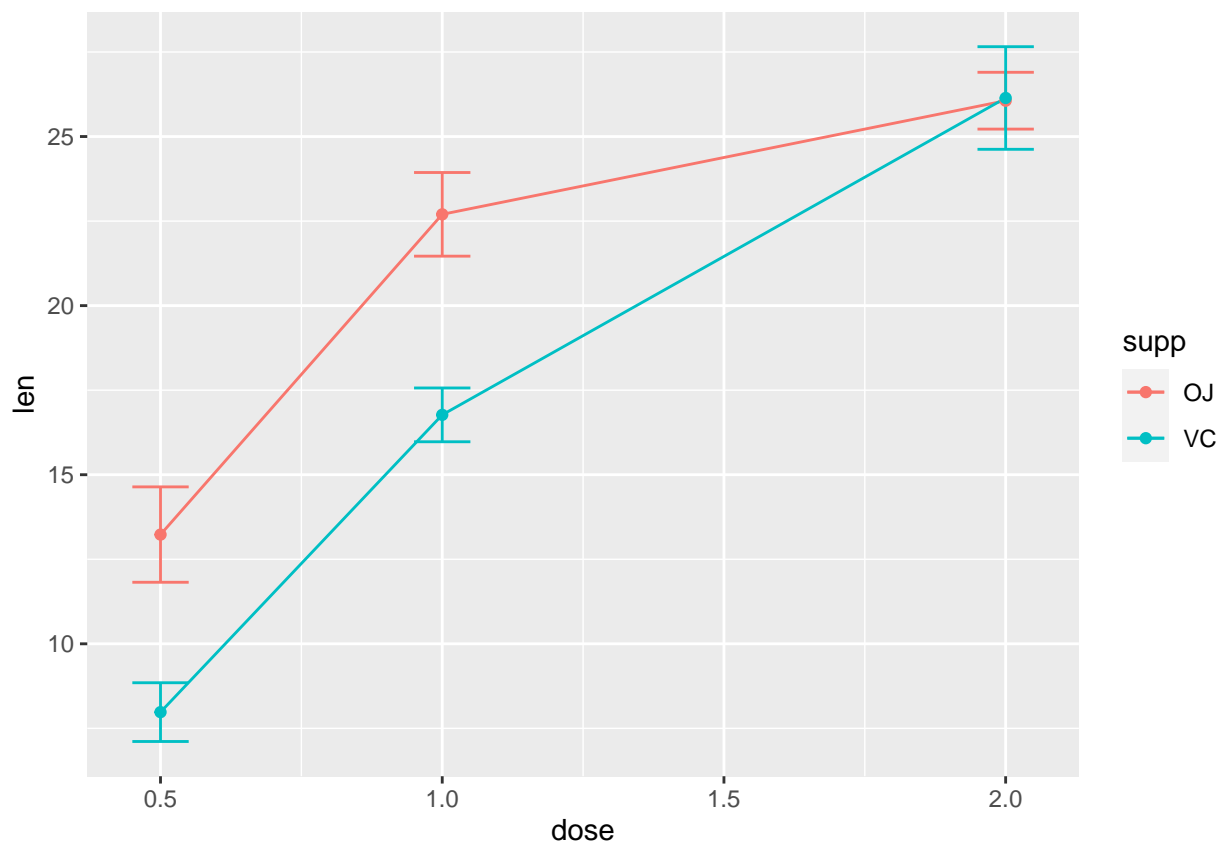
```
## Loading required package: plyr
```

```
tgc <- summarySE(tg, measurevar="len", groupvars=c("supp","dose"))
tgc
```

```
##      supp dose  N   len      sd      se      ci
## 1    OJ  0.5  10  13.23 4.459709 1.4102837 3.190283
## 2    OJ  1.0  10  22.70 3.910953 1.2367520 2.797727
## 3    OJ  2.0  10  26.06 2.655058 0.8396031 1.899314
## 4    VC  0.5  10   7.98 2.746634 0.8685620 1.964824
## 5    VC  1.0  10  16.77 2.515309 0.7954104 1.799343
## 6    VC  2.0  10  26.14 4.797731 1.5171757 3.432090
```

```
# factor plot con error bars
```

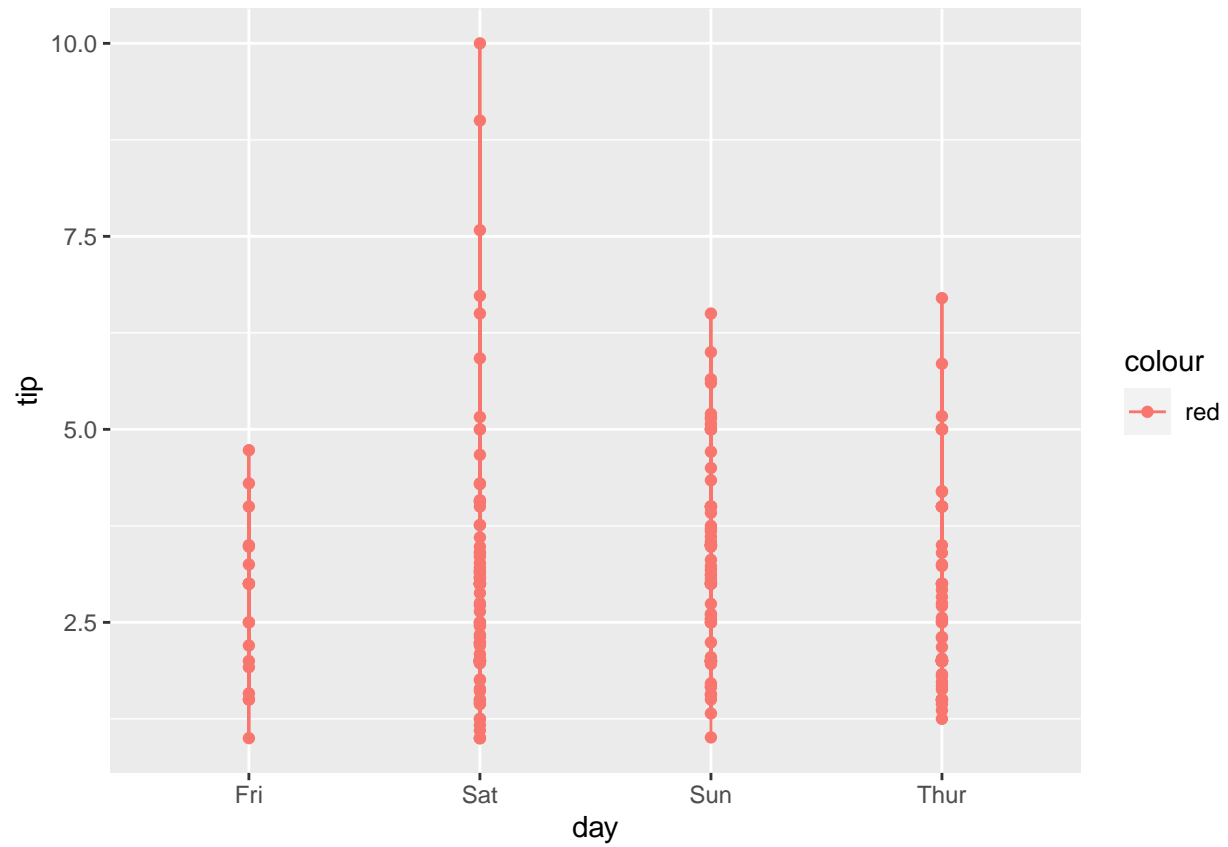
```
ggplot(tgc, aes(x=dose, y=len, colour=supp)) +
  geom_errorbar(aes(ymin=len-se, ymax=len+se), width=.1) +
  geom_line() +
  geom_point()
```



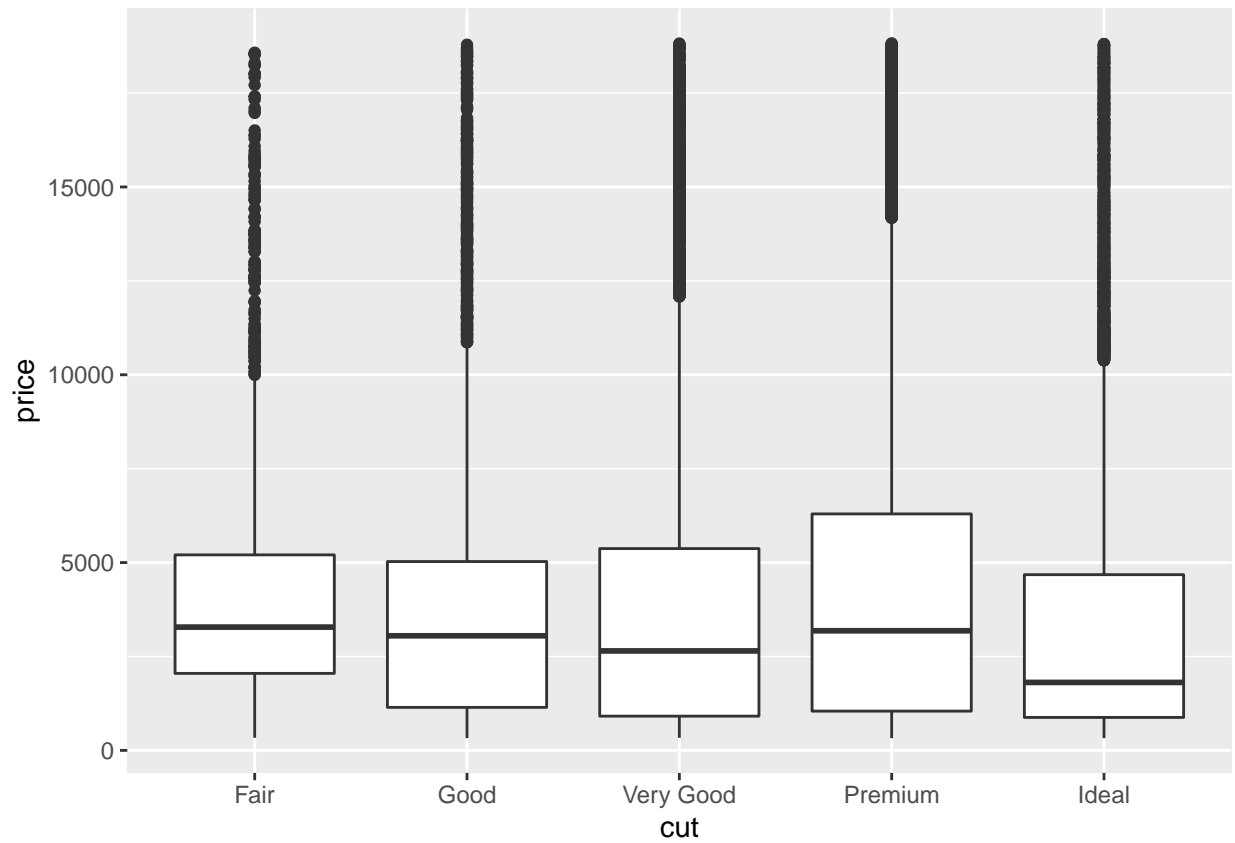
```
# observamos cómo quedaría esta gráfica según propinas por día
```

```
ggplot(tips, aes(x=day, y=tip, colour="red")) +
  geom_line() +
  geom_point()
```

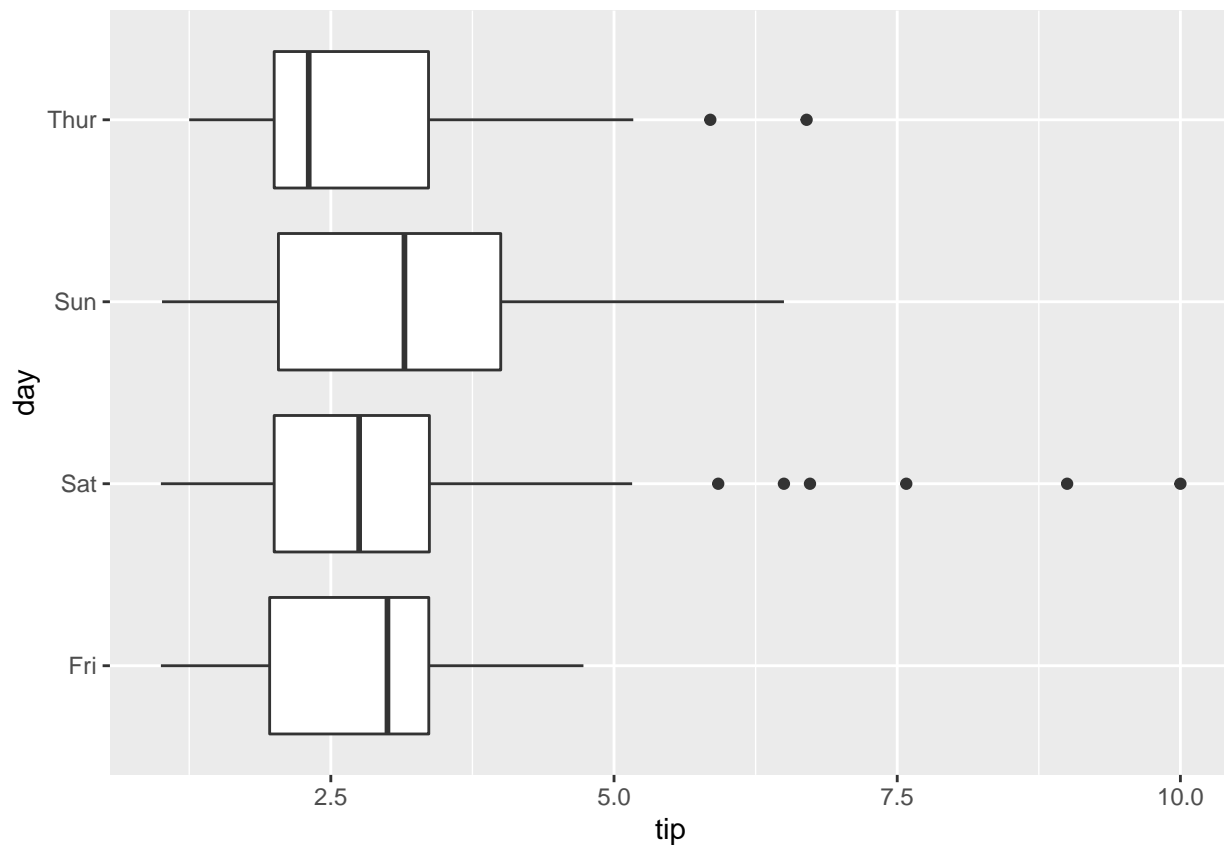




```
##### Boxplot #####
p <- ggplot(diamonds, aes(cut, price))
p + geom_boxplot()
```



```
#propina en función de día  
#el domingo pagan más en media pero el sábado hay más gente optimista con propinas altas  
ggplot(tips, aes(day, tip))+  
  geom_boxplot()+  
  coord_flip()
```



##### Gráfico de Linha #####

```
data(ToothGrowth)
head(ToothGrowth)
```

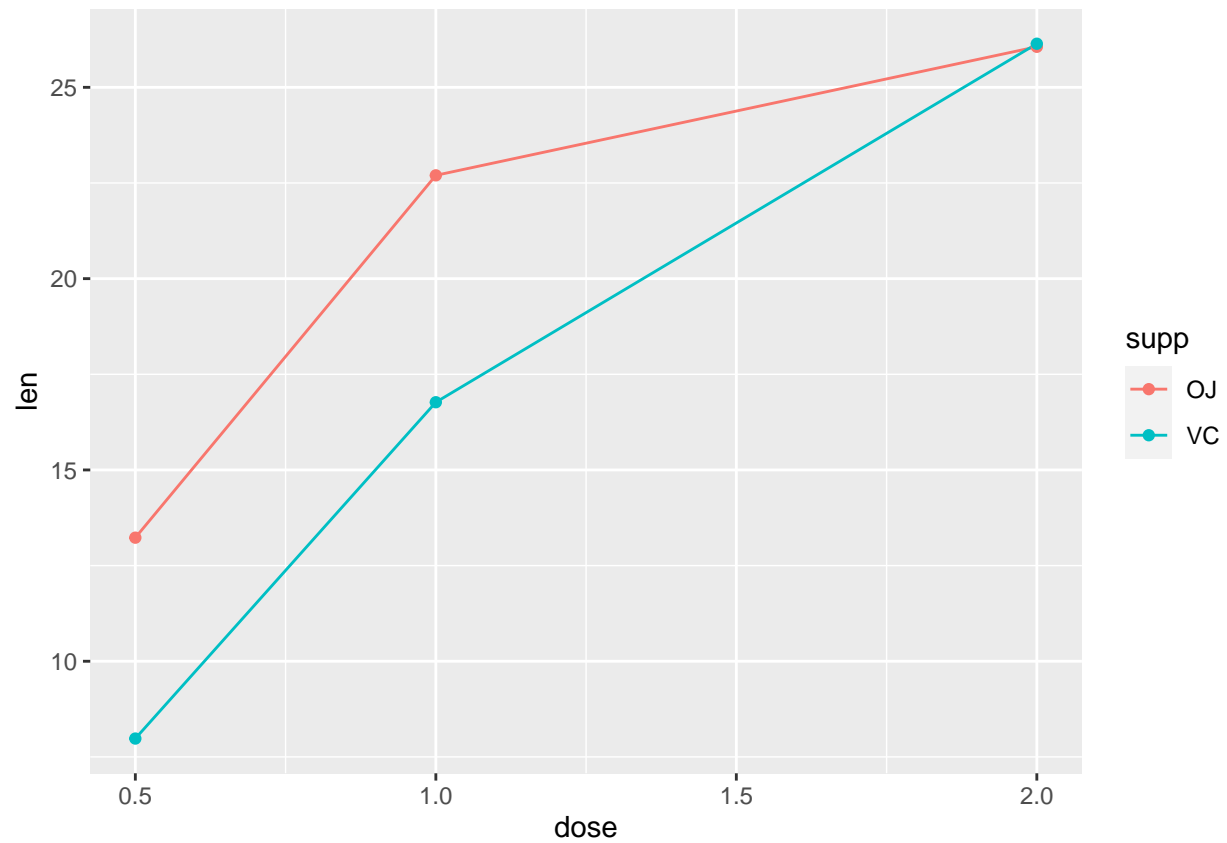
```
##      len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
```

```
datn<-ToothGrowth
```

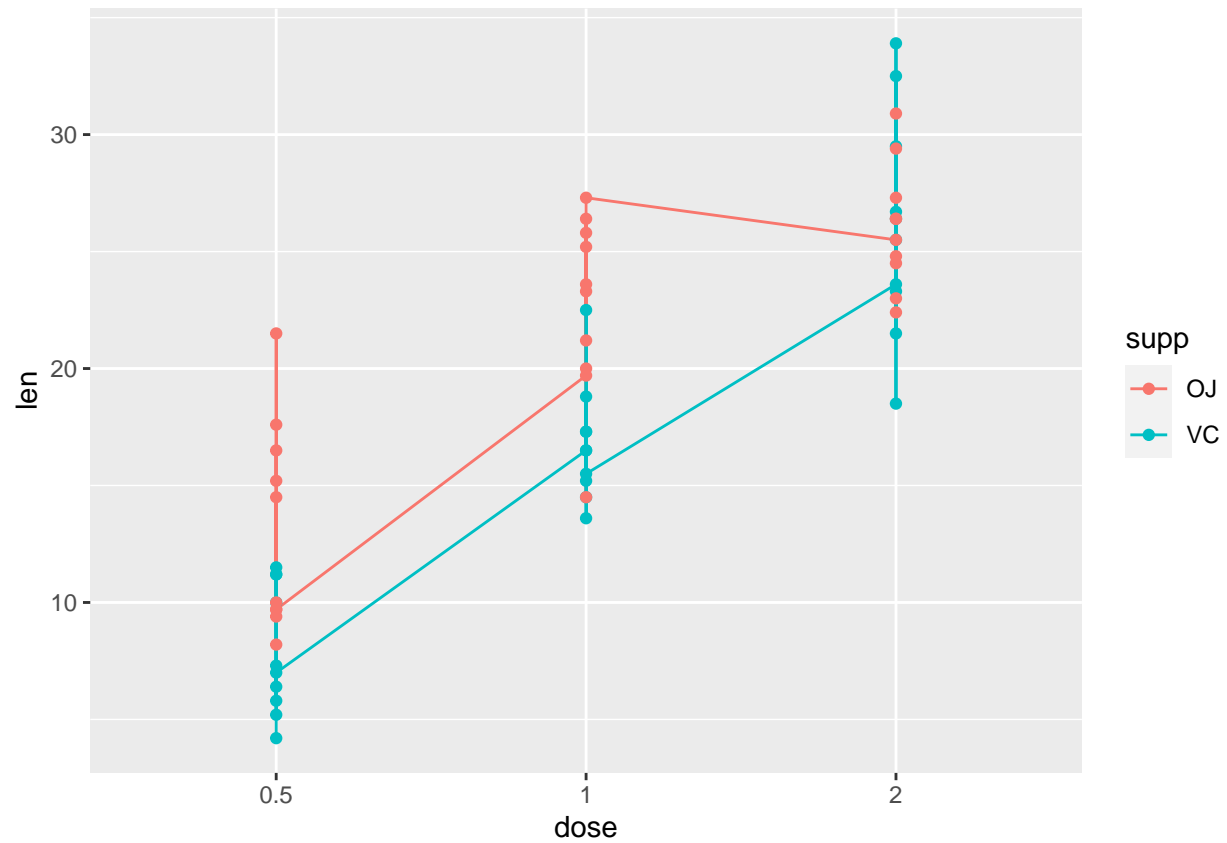
```
tgc1 <- summarySE(datn, measurevar="len", groupvars=c("supp","dose"))
tgc1
```

```
##      supp dose  N   len      sd      se      ci
## 1    OJ  0.5  10 13.23 4.459709 1.4102837 3.190283
## 2    OJ  1.0  10 22.70 3.910953 1.2367520 2.797727
## 3    OJ  2.0  10 26.06 2.655058 0.8396031 1.899314
## 4    VC  0.5  10  7.98 2.746634 0.8685620 1.964824
## 5    VC  1.0  10 16.77 2.515309 0.7954104 1.799343
## 6    VC  2.0  10 26.14 4.797731 1.5171757 3.432090
```

```
ggplot(data=tgc1, aes(x=dose, y=len, group=supp, colour=supp)) +  
  geom_line() +  
  geom_point()
```

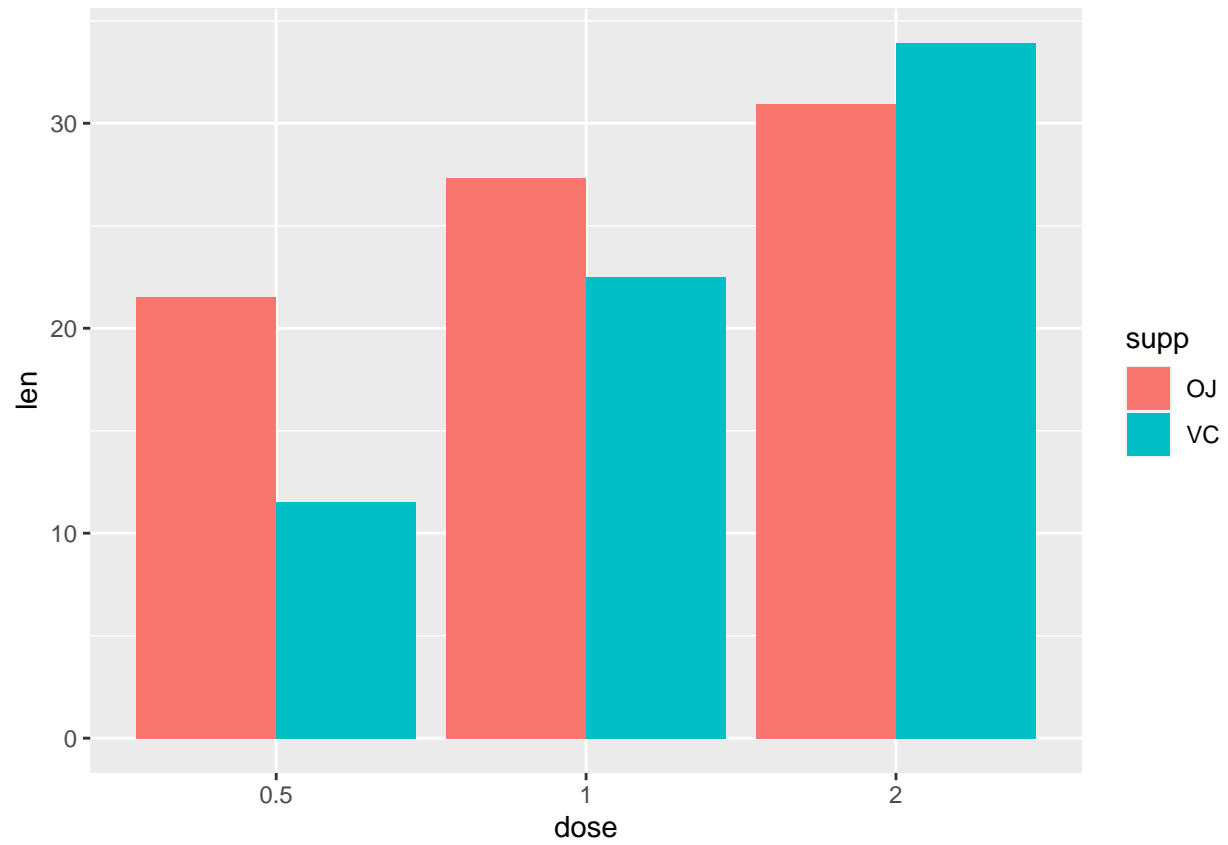


```
# Copy the data frame and convert dose to a factor  
datn2 <- datn  
datn2$dose <- factor(datn2$dose)  
ggplot(data=datn2, aes(x=dose, y=len, group=supp, colour=supp)) +  
  geom_line() +  
  geom_point()
```



*#Também é possível criar um gráfico de barras quando a variável é tratada como categórica e não numérica*

```
ggplot(data=datn2, aes(x=dose, y=len, fill=supp)) +  
  geom_bar(stat="identity", position=position_dodge())
```



```
ggplot(data=datn, aes(x=factor(dose), y=len, fill=supp)) +  
  geom_bar(stat="identity", position=position_dodge())
```

