

Tarea 2

Luis Gagnevin

5/21/2021

Antes de empezar llamamos las librerias que precisaremos en el trabajo

```
library(gapminder)
library(tidyverse)
library(ggplot2)
```

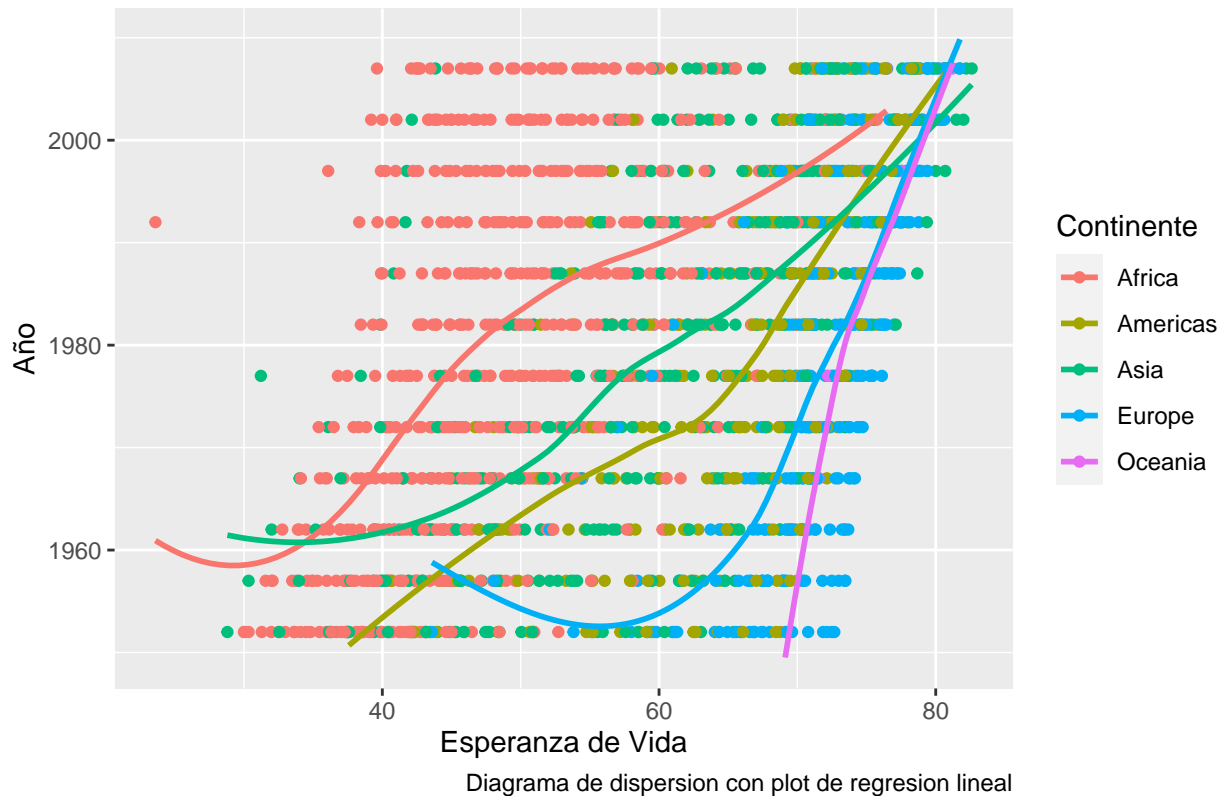
Ejercicio 1

1.

Hacer un grafico de dispersion que tenga el eje y: year y en el eje x: lifeExp, los puntos deben estar coloreados por la variable continent. Para este plot ajusta una recta de regresion para cada continente sin incluir las barras de error. Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un caption en la Figura con algun comentario de interes que describa el grafico. El resto de los comentarios del grafico se realizaran en el texto

```
gd<- ggplot(gapminder, aes(lifeExp, year, color=continent))
gd+geom_point()+geom_smooth(se=FALSE)+labs(x="Esperanza de Vida",
                                           y="Año",
                                           colour="Continente",
                                           title="Esperanza de vida por año segun cada continente",
                                           caption="Diagrama de dispersion con plot de regresion lineal")
```

Esperanza de vida por año segun cada continente



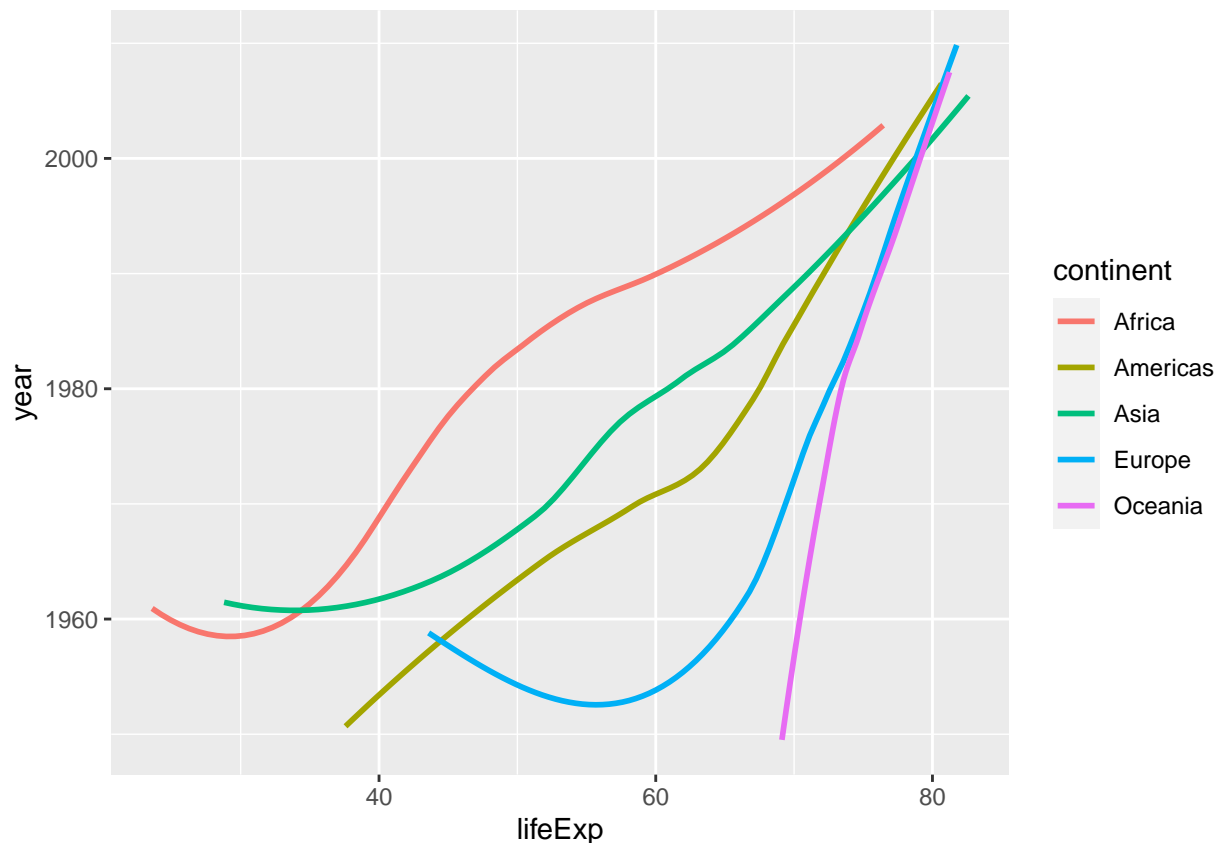
Aqui tenemos un grafico que contiene las variables de esperanza de vida y año y continente, podemos ver como la esperanza de vida en general aumenta en todo continente, en algunos continentes la esperanza aumenta menos que en otros, por ejemplo en Oceania, pasa de los 69 años hasta los 80 mientras que en paises como Africa toma un recorrido desde los 20 hasta los 75 aprox.

2.

Omitir la capa de `geom_point()`, Las lineas aun aparecen pero los puntos no. ¿Porque sucede esto?

```
gd+geom_smooth(se=FALSE)
```

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



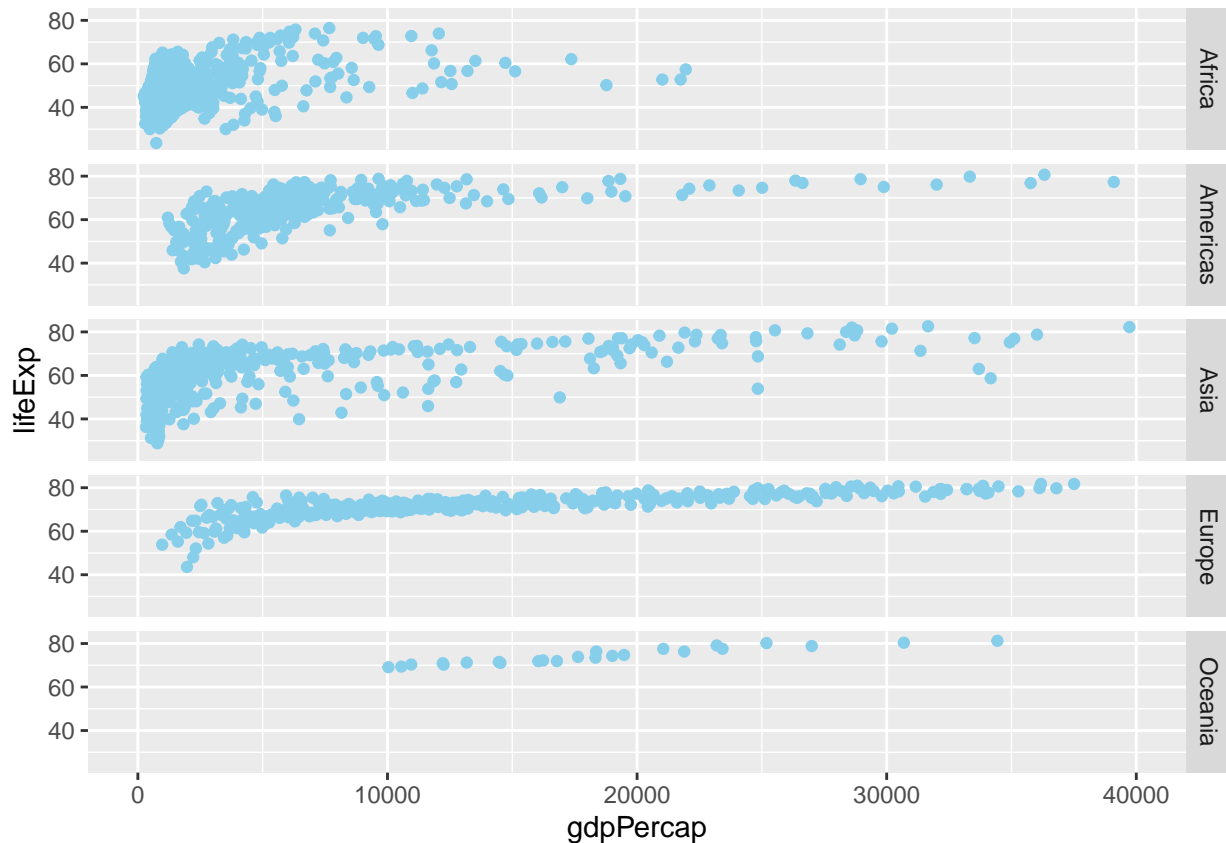
La respuesta a esto es sencilla, estamos sobreponiendo dos ploteos los cuales comparten variables para poder ver la dispersion en el grafico anterior, sin embargo, las lineas que se forman son independientes a la existnecia de los puntos en el ploteo ya que son parte de `geom_smooth()` y no de `geom_point()`.

3.

El siguiente es un grafico de dispersion entre `lifeExp` y `gdpPercap` coloreado por `continent`. Usando como elemento estetico `color (aes())` nosotros podemos distinguir los distintos continentes usando diferentes colores de similar manera usando forma (`shape`) Dicho grafico esta sobrecargado. ¿Como lo modificarias para que sea mas clara la comparacion para los distintos continentes y porque? Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Comenta alguna caracteristica interesante que describa lo que aprendes viendo el grafico

```
gd<- ggplot(gapminder, aes(gdpPercap, lifeExp))
gd+geom_point(color='skyblue')+facet_grid(gapminder$continent)+xlim(0,40000)
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



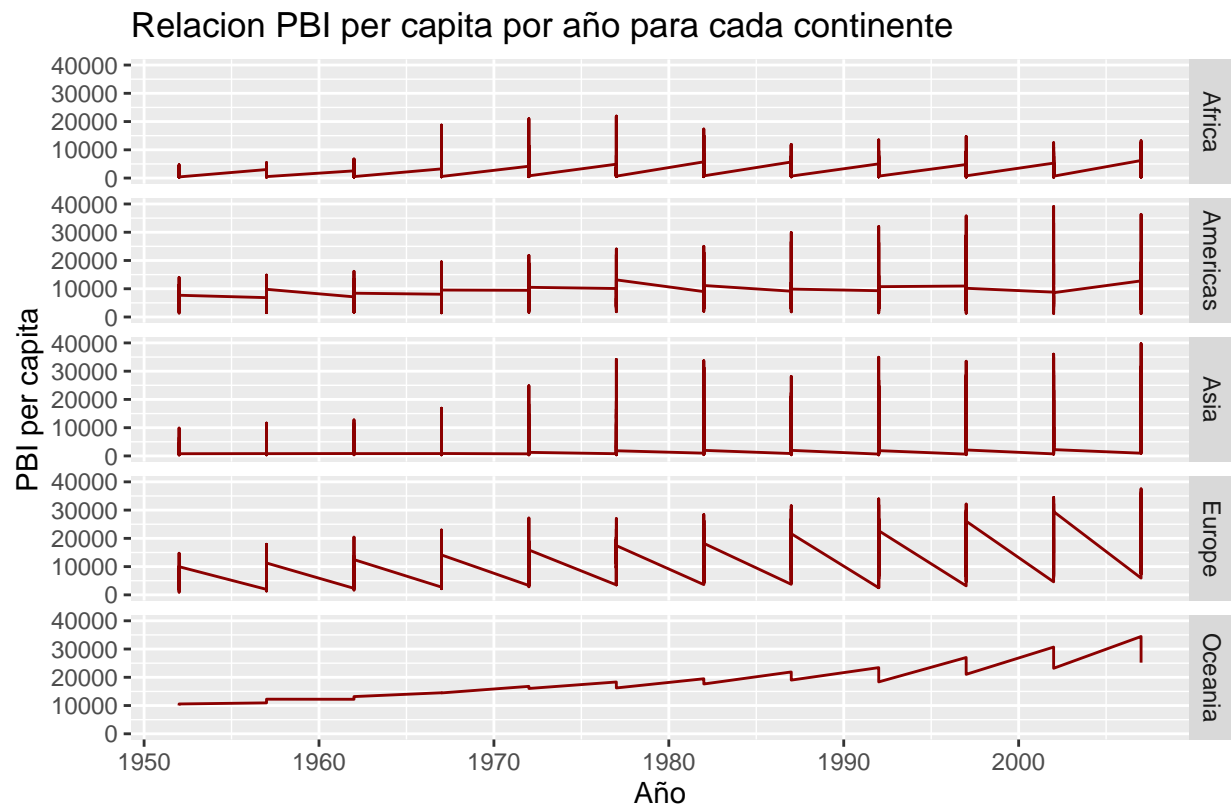
Es interesante ver como el continente de oceania tiene la mayor esperanza de vida y no varia mucho segun el GDP per capita, como sucede en otros continentes donde la variacion es notablemente importante segun el nivel de vida.

Para hacer el grafico mas lindo a la vista, los separe segun el continente usando un `facet_grid()`, deja mejor visto y abierto a una comparacion mas linda y a la vista que un `facet_wrap()` y que sin utilizarlo. Tambien elimine los valores sueltos que serian los de un GDP mayor a 40000 ya que los casos mayores son muy aislados y no son relevantes para una comparacion eficaz entre los continentes.

4.

Hacer un grafico de lineas que tenga en el eje x year y en el eje y gdpPercap para cada continente en una misma ventana grafica. En cada continente, el grafico debe contener una linea para cada pais a lo largo del tiempo (Serie de tiempo de gdpPercap). Las etiquetas de los ejes deben ser claras y describir las variables involucradas. Incluir un caption en la Figura con algun comentario de interes que describa el grafico

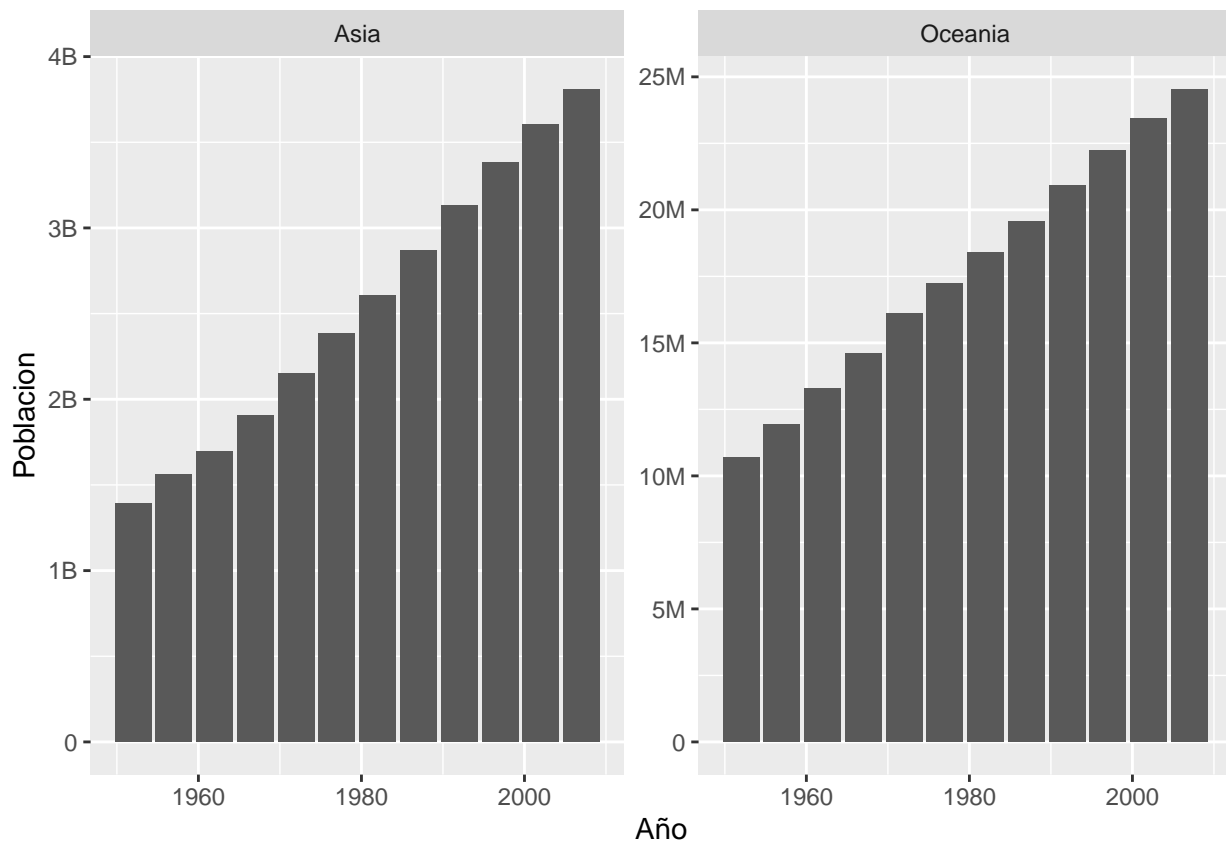
```
gd<- ggplot(gapminder, aes(year, gdpPercap))
gd+ geom_line(color="darkred")+facet_grid(gapminder$continent)+ylim(0,40000)+
  labs(y="PBI per capita",
       x="Año",
       title = "Relacion PBI per capita por año para cada continente",
       caption="Grafico de lineas entre el año y el GDP per capita de gapminder")
```



5.

Usando los datos de gapminder seleccione una visualizacion que describa algun aspecto de los datos que no exploramos. Comente algo interesante que se pueda aprender de su grafico

```
dg<- gapminder %>% filter(continent=="Oceania" | continent=="Asia")
ggplot(dg, aes(year, pop))+geom_col()+
  facet_wrap(dg$continent, scales="free")+
  scale_y_continuous(labels = scales::label_number_si())+
  labs(x="Año", y="Poblacion")
```



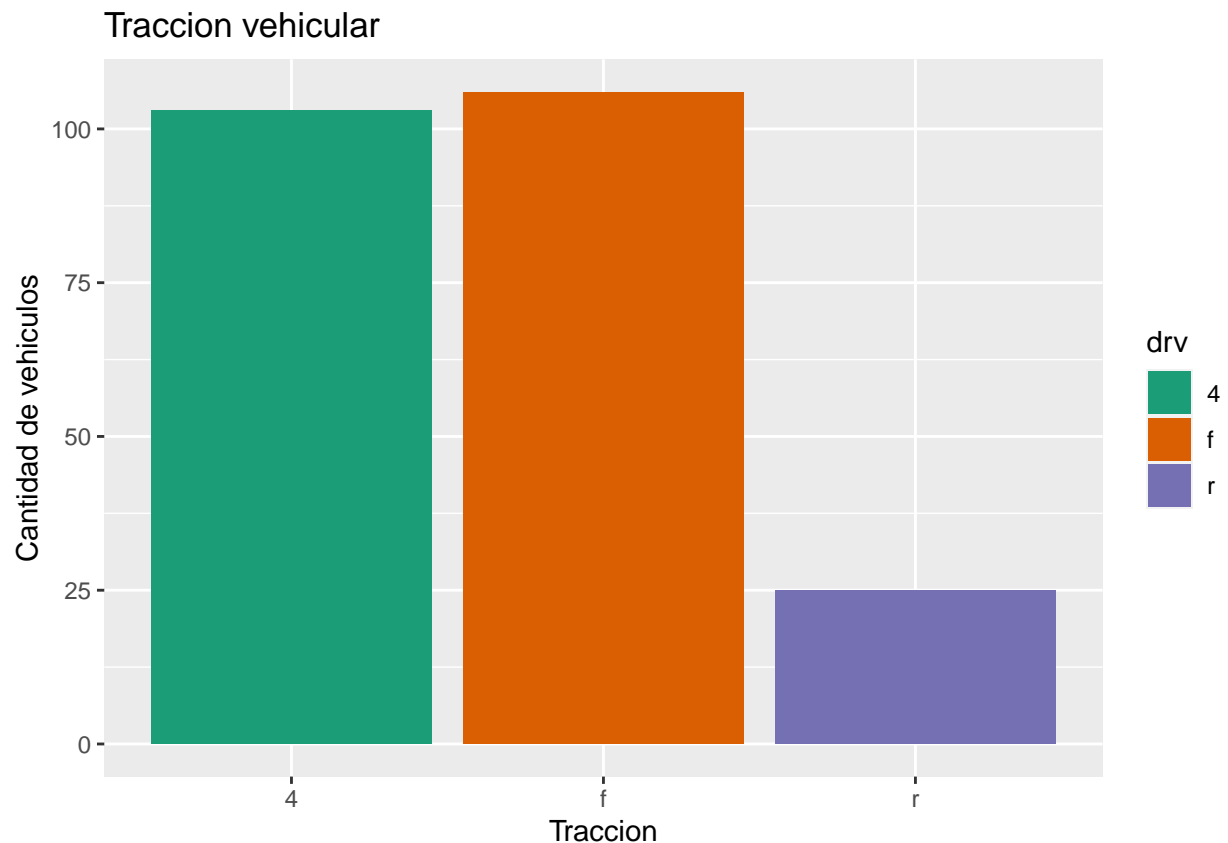
Decidi investigar primero que continente tenia el mayor crecimiento y cual el menor. Vi que el continente con mayor crecimiento es ASIA y el menor es Oceania. Oceania tiene una poblacion 160 veces menor que la de ASIA.

Ejercicio 2

1.

Con los datos mpg que se encuentran en ggplot2, hacer un grafico de barras para la variables drv con las siguientes características: * Las barras tienen que estar coloreadas por drv + Incluir usando labs() el nombre de los ejes y titulo informativo + Usa la paleta Dark2

```
data(mpg)
ggplot(mpg, aes(x=drv, fill=drv))+geom_bar()+scale_fill_brewer(palette = "Dark2")+
  labs(x="Traccion", y="Cantidad de vehiculos", title="Traccion vehicular")
```



2.

Usando como base el grafico anterior: * Incluir en el eje y porcentaje en vez de conteos + Usando `scale_y_continuous()` cambiar la escala del eje y a porcentajes + Usando `geom_text()` incluir texto con porcentajes arriba de cada barra

Primero debemos de generar con tidyverse un grupo de valores con el porcentaje de veces que aparece en la columna drv cada tipo de traccion (4 f r).

Luego pasamos al grafico donde utilizamos un `geom_bar` que utilizara la x segun los valores porcentuales de Y y colocara sobre estas columnas el % aproximado a 1 cifra de cada una.

En la parte estetica, utilizamos un `scale_fill_brewer` y no un color ya que estamos utilizando un `fill=` en la parte grafica.

Luego renombramos las columnas, titulos y ejes y borramos la leyenda

```
data(mpg)
mpgp<- mpg %>% group_by(drv) %>% summarize(veces = n()) %>% mutate(prct=veces/sum(veces))

# Parte Grafica
ggplot(mpgp, aes(drv,prct, fill=drv))+
  geom_bar(stat='identity')+
  geom_text(aes(label=scales::percent(prct),vjust= -.3))+
  scale_y_continuous(labels=scales::percent)+

# Parte Estetica
scale_fill_brewer(palette = 'Dark2')+

```

```
scale_x_discrete(labels=c("4 ruedas", "Frontal", "Trasera"))+  
labs(x="Traccion", y="Cantidad de vehiculos", title="Traccion vehicular")+  
theme(legend.position="none")
```

