

INTRODUCCIÓN A R Y ANALÍTICA  
DE DATOS

# ANÁLISIS DE VENTAS

FABIÁN REDONDO CALVO

23 ABRIL, 2025

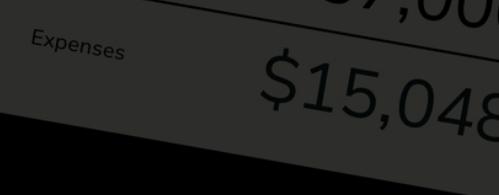
# 1. CARGA DE LIBRERIAS Y DATOS

librerias

```
install.packages("ggplot2")
library ("ggplot2")
library(readxl)
library(ggplot2)
library(dplyr)
```

data

```
data1 <- read_delim("-R Data
Analyst/Proyecto/sales_data_sample_proy_2021.csv",
                      delim = ";",
                      escape_double = FALSE, trim_ws = TRUE)
data2 <- read_delim("-R Data
Analyst/Proyecto/sales_data_sample_proy_2022.csv",
                      delim = ";",
                      escape_double = FALSE, trim_ws = TRUE)
data3 <- read_delim("-R Data
Analyst/Proyecto/sales_data_sample_proy_2023.csv",
                      delim = ";",
                      escape_double = FALSE, trim_ws = TRUE)
```



clean

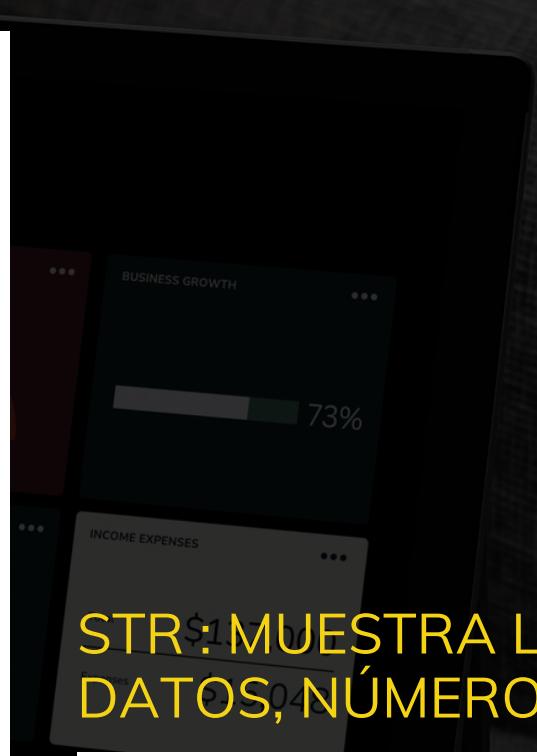
```
data3 <- data3 %>%
  select(-CUSTOMERNAME, -PHONE, -ADDRESSLINE1,
         -ADDRESSLINE2, -CITY,
         -STATE, -POSTALCODE, -TERRITORY,
         -CONTACTLASTNAME, -CONTACTFIRSTNAME, -MSRP)
colSums(is.na(data1))
colSums(is.na(data2))
colSums(is.na(data3))
data1 <- na.omit(data1)
data2 <- na.omit(data2)
data3 <- na.omit(data3)
data_total <- bind_rows(data1, data2, data3)
dim(data_total)
colSums(is.na(data_total))
```



# 2. EDA (EXPLORATORY DATA ANALYST)

SUMMARY : GENERA UN RESUMEN ESTADÍSTICO DE LAS COLUMNAS DEL DATA FRAME.

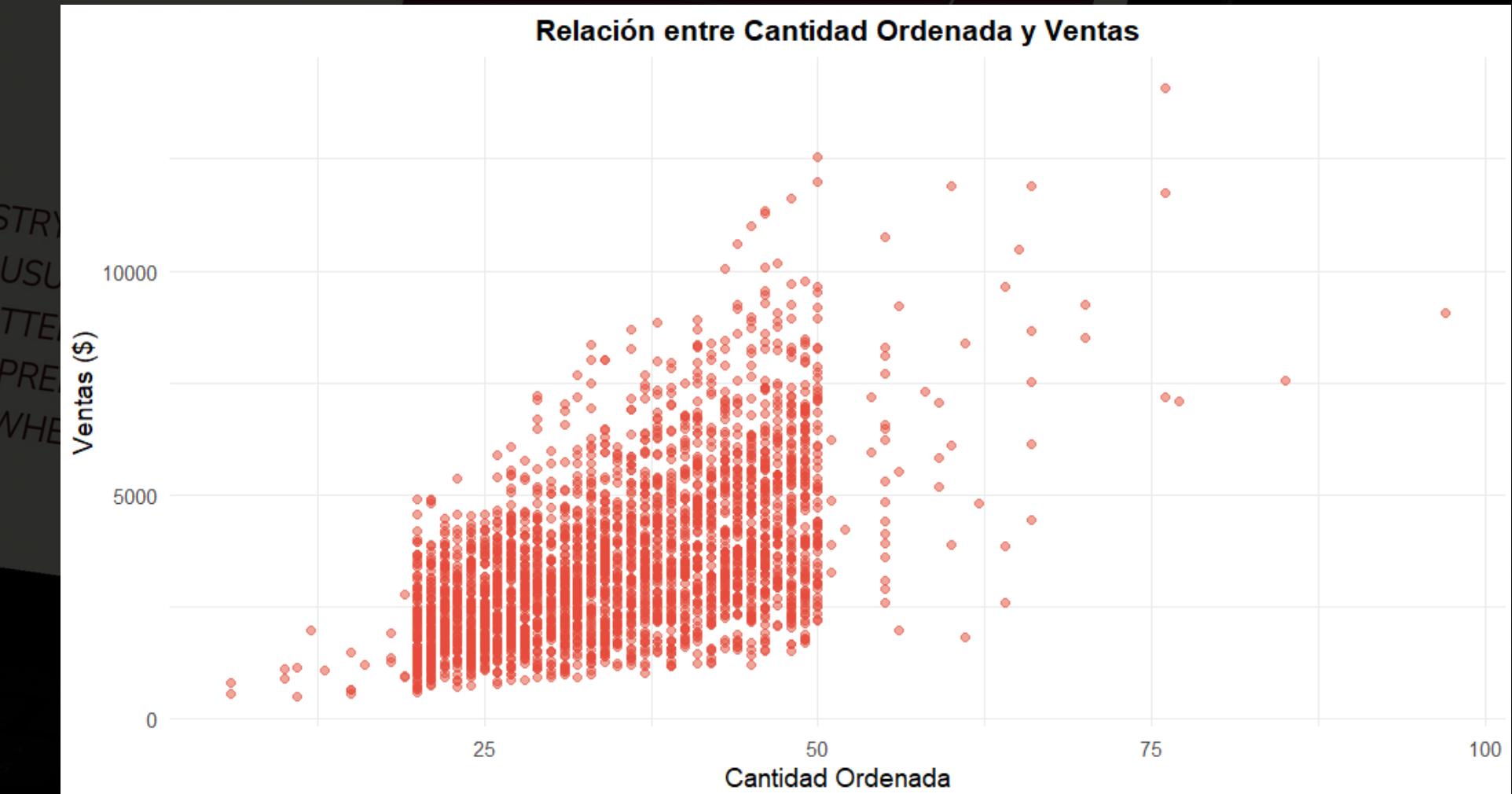
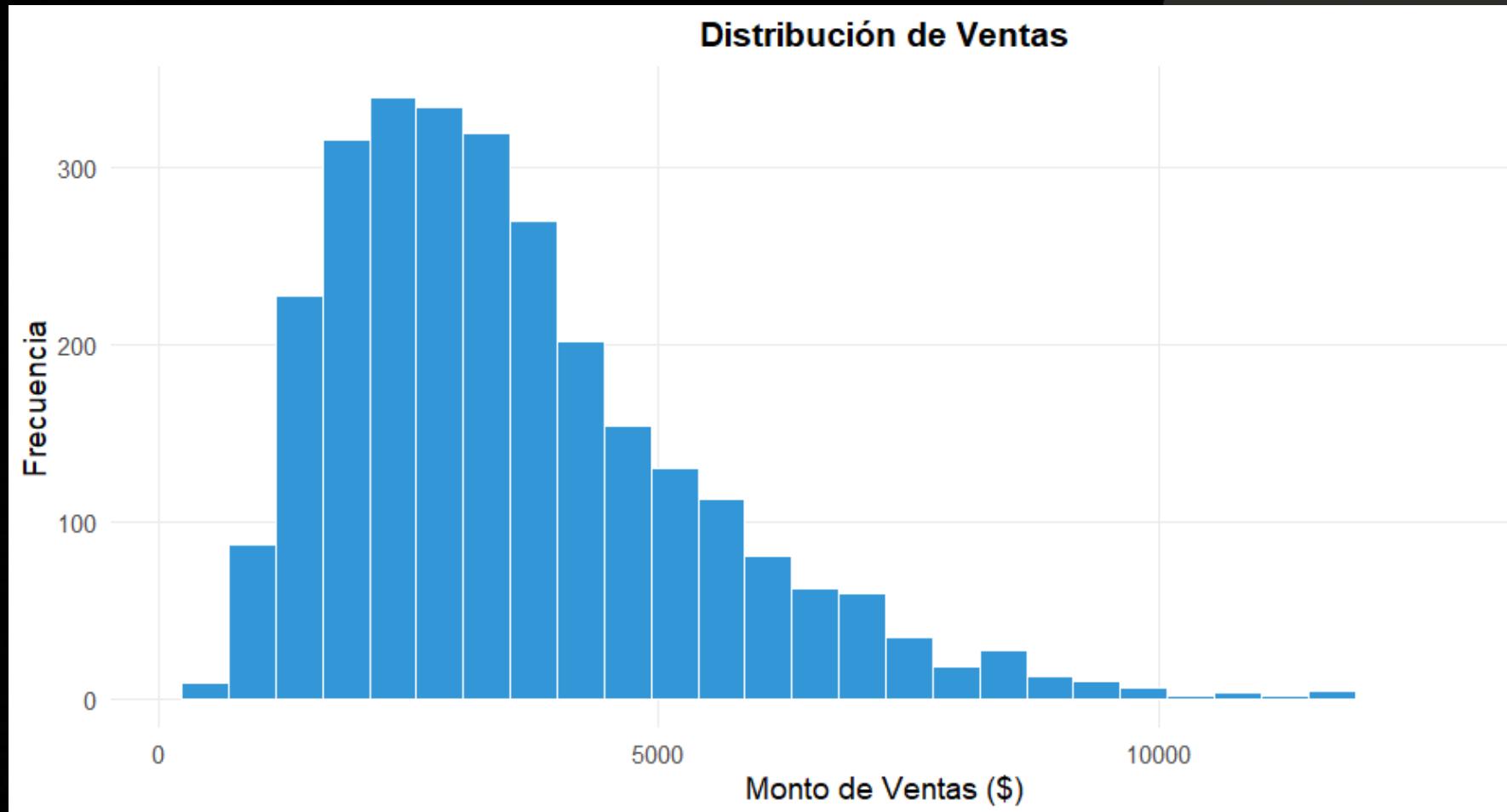
```
> summary(data_total)
ORDERNUMBER QUANTITYORDERED PRICEEACH ORDERLINENUMBER SALES
Min. :10100 Min. : 6.00 Min. : 26.88 Min. : 1.000 Min. : 482.1
1st Qu.:10180 1st Qu.:27.00 1st Qu.: 68.79 1st Qu.: 3.000 1st Qu.: 2203.4
Median :10263 Median :35.00 Median : 95.69 Median : 6.000 Median : 3184.0
Mean   :10268 Mean  :35.09 Mean  : 83.64 Mean  : 6.468 Mean  : 3552.2
3rd Qu.:10334 3rd Qu.:43.00 3rd Qu.:100.00 3rd Qu.: 9.000 3rd Qu.: 4497.5
Max.  :18877 Max. :97.00 Max. :100.00 Max. :18.000 Max. :14082.8
ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID
Min. :2012-12-05 Length:2827 Min. :1.000 Min. : 1.000 Min. :2021
1st Qu.:2021-09-22 Class :character 1st Qu.:2.000 1st Qu.: 4.000 1st Qu.:2021
Median :2022-06-23 Mode  :character Median :3.000 Median : 8.000 Median :2022
Mean   :2022-06-21          Mean  :2.717 Mean  : 7.091 Mean  :2022
3rd Qu.:2023-03-25          3rd Qu.:4.000 3rd Qu.:11.000 3rd Qu.:2023
Max.  :2023-12-31          Max. :4.000 Max. :12.000 Max. :2023
PRODUCTLINE PRODUCTCODE COUNTRY DEALSIZE
Length:2827 Length:2827 Length:2827 Length:2827
Class :character Class :character Class :character Class :character
Mode  :character Mode  :character Mode  :character Mode  :character
```



STR: MUESTRA LA ESTRUCTURA DEL DATA FRAME (TIPOS DE DATOS, NÚMERO DE FILAS, COLUMNAS).

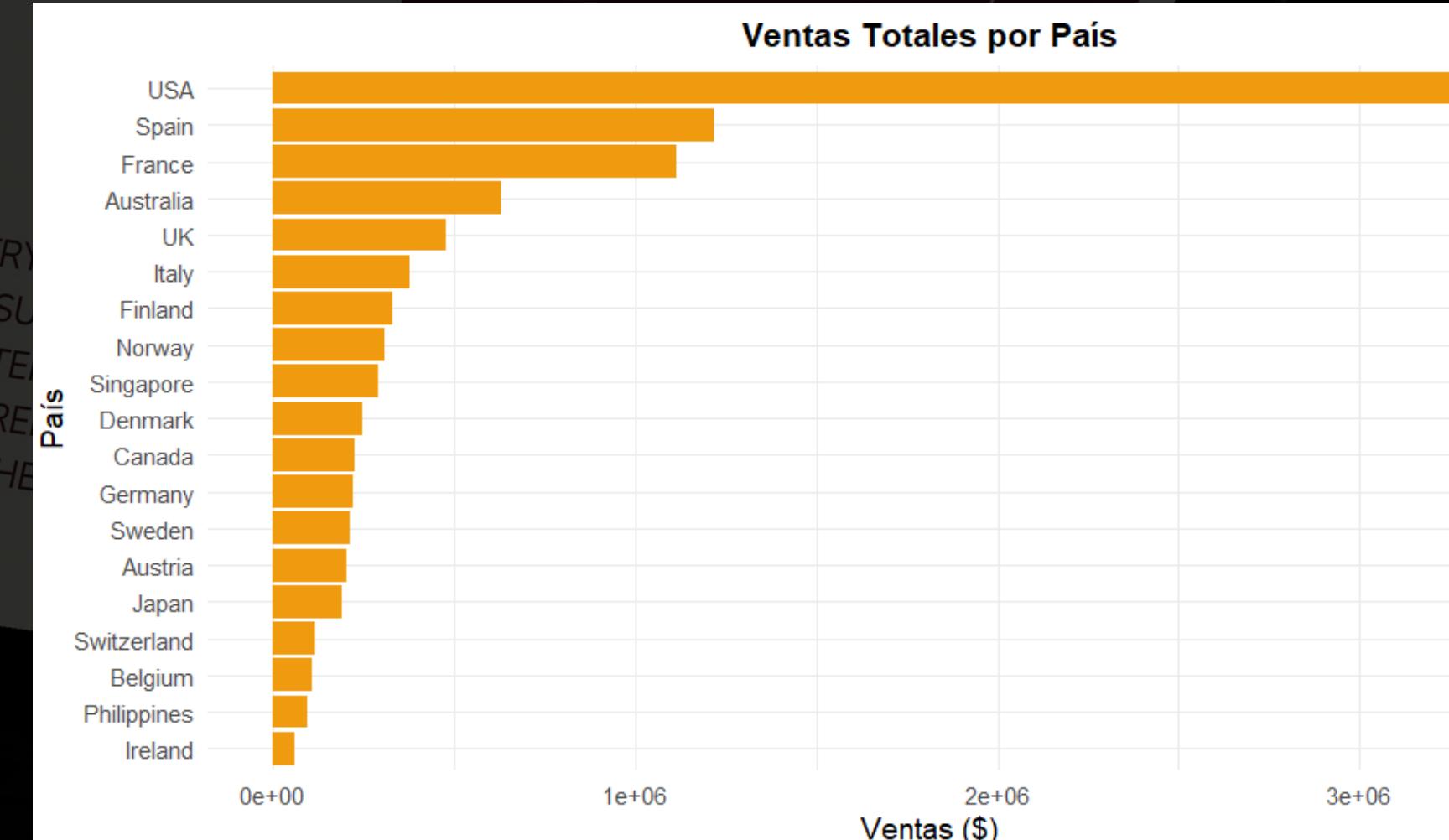
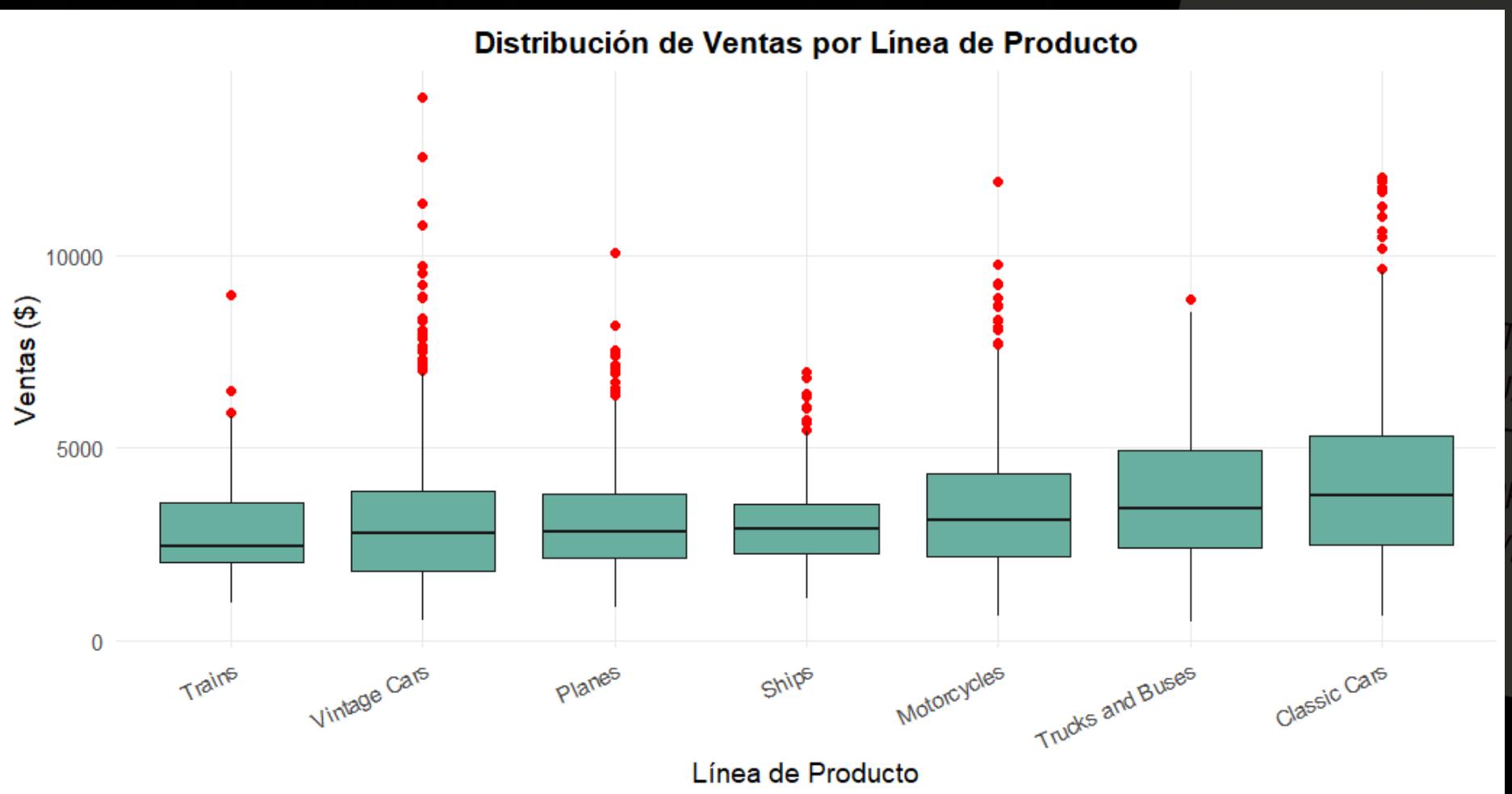
```
> str(data_total)
tibble [2,827 x 17] (s3:tbl_df/tbl/data.frame)
$ ORDERNUMBER : num [1:2827] 10119 10104 10282 10105 10287 ...
$ QUANTITYORDERED: num [1:2827] 28 29 36 22 36 46 44 35 23 26 ...
$ PRICEEACH      : num [1:2827] 70.3 100 100 100 39.6 ...
$ ORDERLINENUMBER: num [1:2827] 2 12 3 11 6 4 8 3 16 4 ...
$ SALES          : num [1:2827] 1968 3773 4175 3065 1427 ...
$ ORDERDATE       : Date[1:2827], format: "2021-01-01" "2021-01-02" ...
$ STATUS          : chr [1:2827] "Shipped" "Shipped" "Shipped" "Shipped" ...
$ QTR_ID          : num [1:2827] 2 1 3 1 3 2 1 3 4 2 ...
$ MONTH_ID        : num [1:2827] 4 1 8 2 8 6 2 9 10 5 ...
$ YEAR_ID         : num [1:2827] 2021 2021 2021 2021 2021 ...
$ PRODUCTLINE     : chr [1:2827] "Planes" "Trucks and Buses" "Trains" "Vintage Cars" ...
$ PRODUCTCODE     : chr [1:2827] "S24_3949" "S18_2319" "S18_3259" "S18_3140" ...
$ COUNTRY          : chr [1:2827] "Austria" "Spain" "USA" "Denmark" ...
$ DEALSIZE         : chr [1:2827] "Small" "Medium" "Medium" "Medium" ...
$ y_m              : chr [1:2827] "2021-01" "2021-01" "2021-01" "2021-01" ...
$ MES              : chr [1:2827] "2021-01" "2021-01" "2021-01" "2021-01" ...
$ PREDICCIÓN       : Named num [1:2827] 2813 2917 3647 2187 3647 ...
```





- ESTE HISTOGRAMA MUESTRA CÓMO SE DISTRIBUYEN LOS MONTOS DE VENTAS EN DÓLARES.
- LA MAYORÍA DE LAS VENTAS PARECEN ESTAR EN RANGOS MÁS BAJOS (POR DEBAJO DE \$5000), LO QUE SIGNIFICA QUE HAY MUCHAS TRANSACCIONES CON VALORES MENORES.

- DIAGRAMA DE DISPERSIÓN MUESTRA CÓMO LA CANTIDAD ORDENADA INFUYE EN LAS VENTAS EN DÓLARES.
- SE OBSERVA UNA TENDENCIA POSITIVA: CUANTO MÁS SE ORDENA, MÁS VENTAS SE GENERAN.



- **DIAGRAMAS (BOX PLOTS) PARA MOSTRAR CÓMO SE DISTRIBUYEN LAS VENTAS DE DISTINTAS LÍNEAS DE PRODUCTOS.**
- **SE PUEDEN VER RANGOS Y POSIBLES VALORES ATÍPICOS EN CADA CATEGORÍA. ALGUNAS LÍNEAS PARECEN TENER VENTAS MÁS CONSISTENTES Y OTRAS TIENEN VARIACIONES MÁS AMPLIAS.**

- **GRÁFICO DE BARRAS QUE MUESTRA QUÉ PAÍSES GENERAN MÁS VENTAS.**
- **EE. UU. LIDERA CON LAS VENTAS MÁS ALTAS, SEGUIDO DE ESPAÑA Y FRANCIA.**

# 3. KPI'S (KEY PERFORMANCE INDICATORS)



KPI 1: TOTAL DE VENTAS

```
> sum(data_total$SALES)  
[1] 10042078
```

KPI 2: PROMEDIO POR ORDEN EN DÓLARES

```
> mean(data_total$SALES)  
[1] 3552.203
```

KPI 3: PORCENTAJE DE ÓRDENES CANCELADAS

```
> mean(data_total$status == "Cancelled") * 100  
[1] 2.122391
```

KPI 4: PRODUCTO MÁS VENDIDO EN CANTIDAD

```
> producto_top  
# A tibble: 1 × 2  
PRODUCTCODE total_cantidad  
<chr>           <dbl>  
1 S18_3232      1774
```

KPI 5: LÍNEA DE PRODUCTO CON MAYORES VENTAS (\$)

```
> linea_top  
# A tibble: 1 × 2  
PRODUCTLINE total_ventas  
<chr>           <dbl>  
1 classic cars 3915107.
```

KPI 6: MES CON MÁS VENTAS TOTALES

```
> mes_top  
# A tibble: 1 × 2  
MES     ventas_mes  
<chr>   <dbl>  
1 2021-03 389138.
```

KPI 7: PAÍS CON MÁS VENTAS

```
> pais_top  
# A tibble: 1 × 2  
COUNTRY ventas_pais  
<chr>    <dbl>  
1 USA      3629093.
```

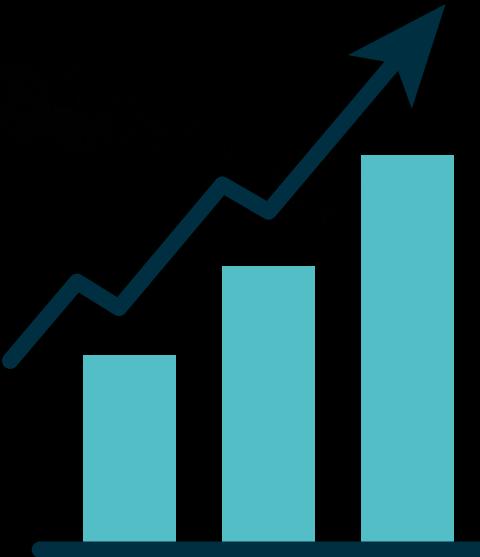
KPI 8: PORCENTAJE DE ÓRDENES EN ESTADO "ON HOLD"

```
> porcentaje_on_hold  
[1] 1.55642
```

# 4. PROYECCIÓN DE VENTAS

1. CREACIÓN DE UNA TABLA CON LAS SIGUIENTES COLUMNAS

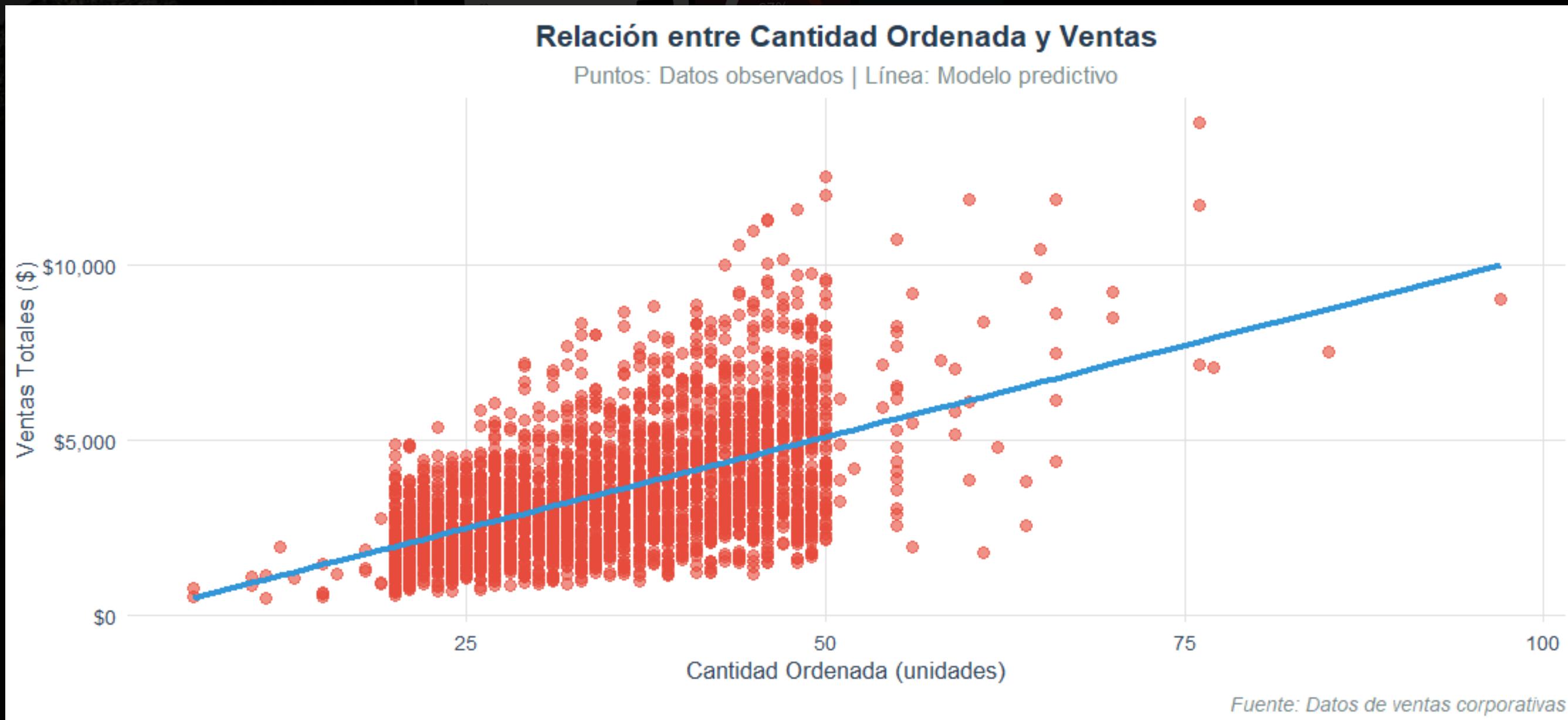
```
modelo <- lm(SALES ~ QUANTITYORDERED, data = data_total)  
summary(modelo)
```



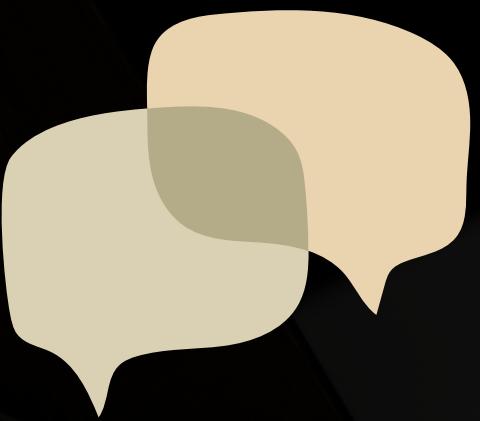
2. AGREGAR PREDICCIÓN A LA DATA TOTAL

```
data_total$PREDICCIÓN <- predict(modelo, newdata =  
data_total)
```

3. GRÁFICO DE PROYECCIÓN



# RECOMENDACIONES



- FOCALIZAR ESTRATEGIAS EN PAÍSES CON ALTO POTENCIAL DE CRECIMIENTO DONDE YA EXISTEN VENTAS.
- REFORZAR LA DISPONIBILIDAD DE LOS PRODUCTOS MÁS VENDIDOS PARA EVITAR QUIEBRES DE STOCK.
- ANALIZAR CAUSAS DETRÁS DE LOS ESTADOS "CANCELLED" Y "ON HOLD" PARA MEJORAR PROCESOS.
- MEJORAR LA CALIDAD DE LOS DATOS (EVITAR VALORES NULOS, REVISAR FECHAS, ETC.).

