

GrupoBimboPlots

Contents

0.1	Plots	1
0.2	Weekly Transactions.	2
0.3	Returns per week	3
0.4	Product Sales	4
0.5	Product wise Sale vs Return	5
0.6	Demand Vs Sales	6

0.1 Plots

```
library("ggplot2")
```

```
library("data.table")  
system.time(train <- fread("./Dataset/train.csv", header = TRUE))
```

```
##  
Read 0.0% of 74180464 rows  
Read 3.0% of 74180464 rows  
Read 6.0% of 74180464 rows  
Read 9.0% of 74180464 rows  
Read 12.0% of 74180464 rows  
Read 15.0% of 74180464 rows  
Read 18.0% of 74180464 rows  
Read 21.0% of 74180464 rows  
Read 24.0% of 74180464 rows  
Read 26.9% of 74180464 rows  
Read 29.9% of 74180464 rows  
Read 32.9% of 74180464 rows  
Read 35.9% of 74180464 rows  
Read 38.9% of 74180464 rows  
Read 41.9% of 74180464 rows  
Read 44.9% of 74180464 rows  
Read 47.9% of 74180464 rows  
Read 50.9% of 74180464 rows  
Read 53.9% of 74180464 rows  
Read 56.9% of 74180464 rows  
Read 59.9% of 74180464 rows  
Read 62.9% of 74180464 rows  
Read 65.9% of 74180464 rows  
Read 68.8% of 74180464 rows  
Read 71.8% of 74180464 rows  
Read 74.8% of 74180464 rows  
Read 77.8% of 74180464 rows
```

```
Read 80.8% of 74180464 rows
Read 83.8% of 74180464 rows
Read 86.8% of 74180464 rows
Read 89.8% of 74180464 rows
Read 92.8% of 74180464 rows
Read 95.8% of 74180464 rows
Read 98.7% of 74180464 rows
Read 74180464 rows and 11 (of 11) columns from 2.980 GB file in 00:00:38
```

```
##      user  system elapsed
## 36.876   0.848  42.427
```

```
system.time(test <- fread("./Dataset/test.csv", header = TRUE))
```

```
##      user  system elapsed
##      1.66    0.04    2.16
```

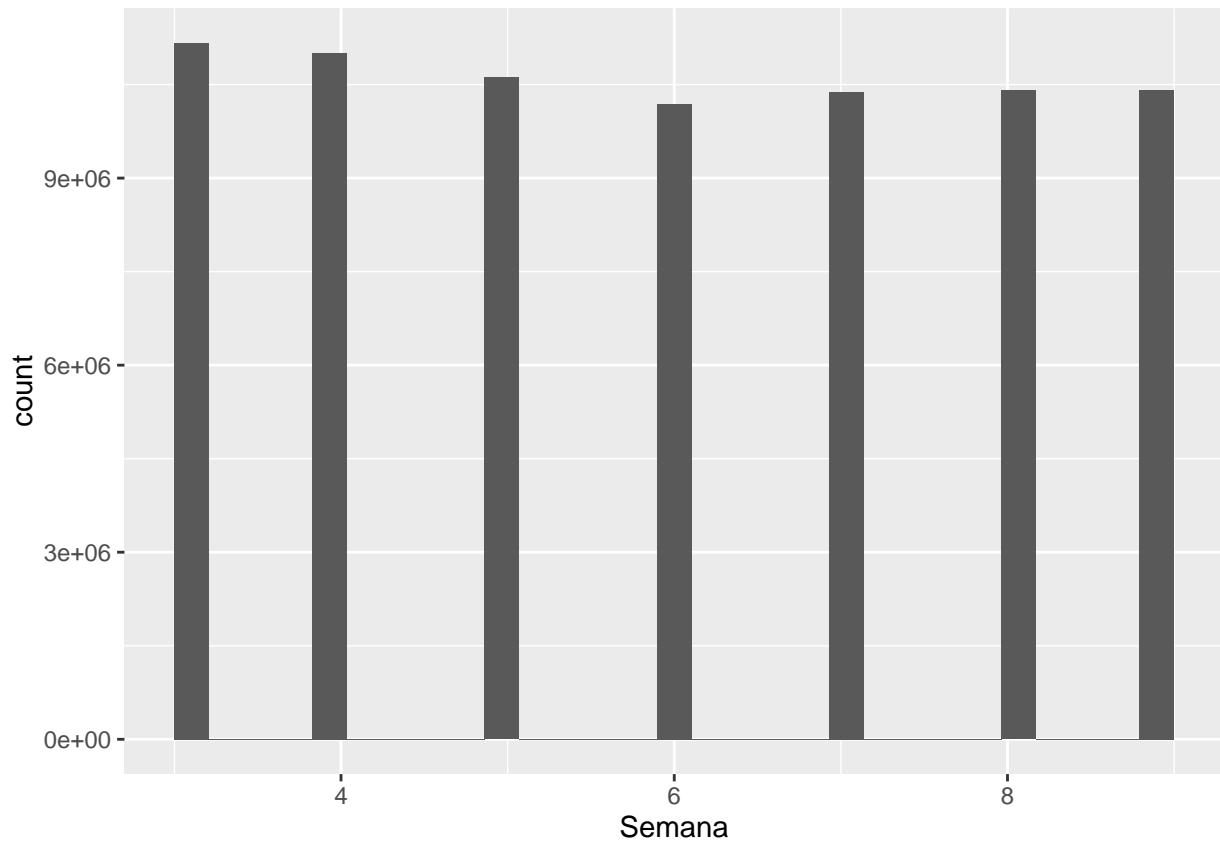
```
system.time(product <- fread("./Dataset/producto_tabla.csv", header=TRUE))
```

```
##      user  system elapsed
##      0.004   0.000   0.001
```

0.2 Weekly Transactions.

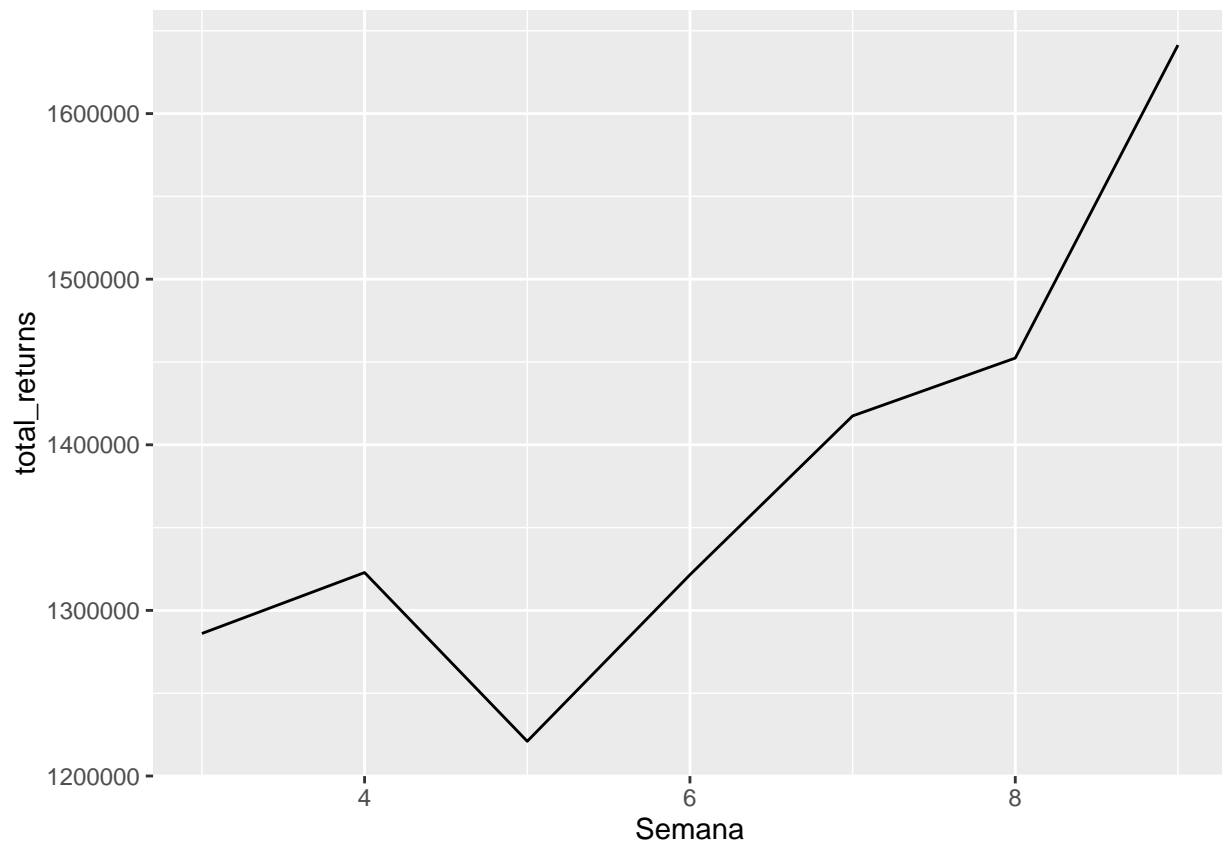
```
# per week, h
ggplot(train, aes(x = Semana)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



0.3 Returns per week

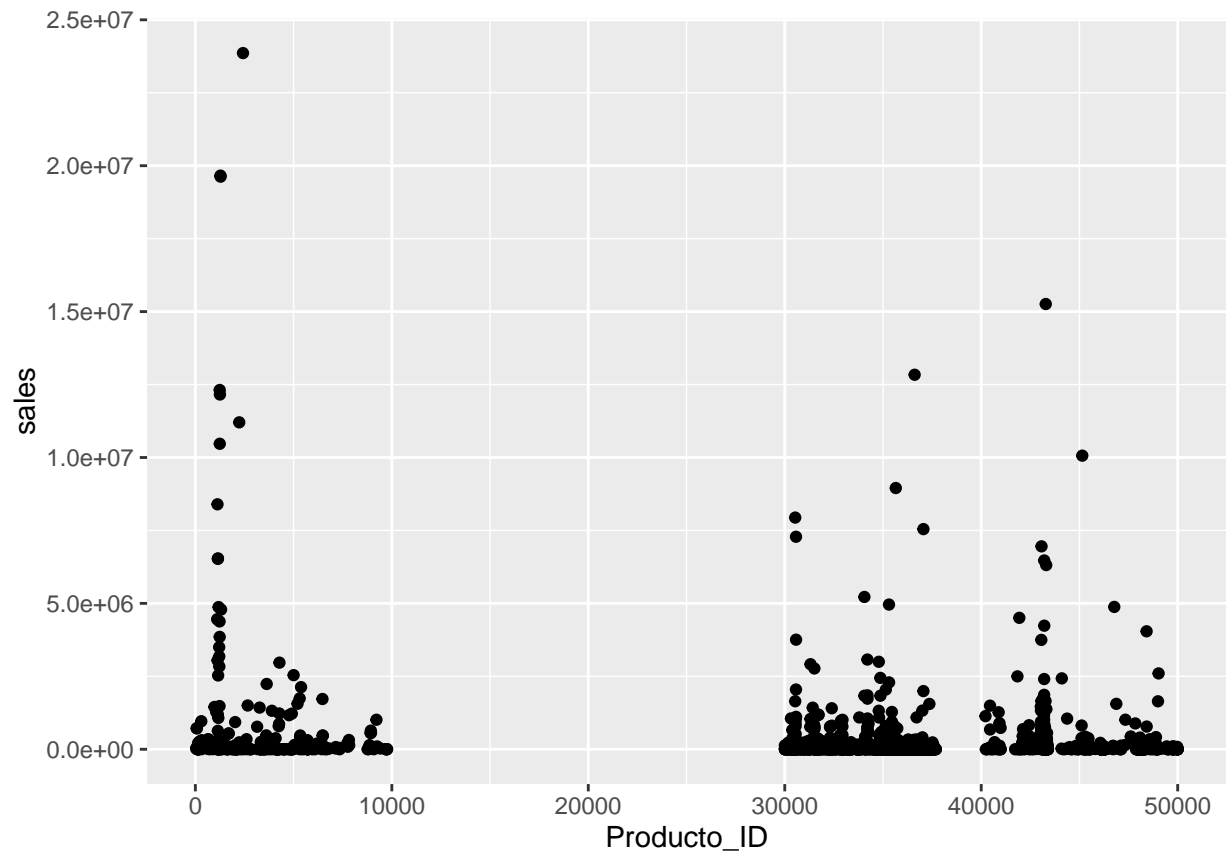
```
# returns per week  
returns <- train[, .( total_returns = sum(Dev_uni_proxima)), by = Semana]  
ggplot(returns, aes(x=Semana, y = total_returns)) + geom_line()
```



0.4 Product Sales

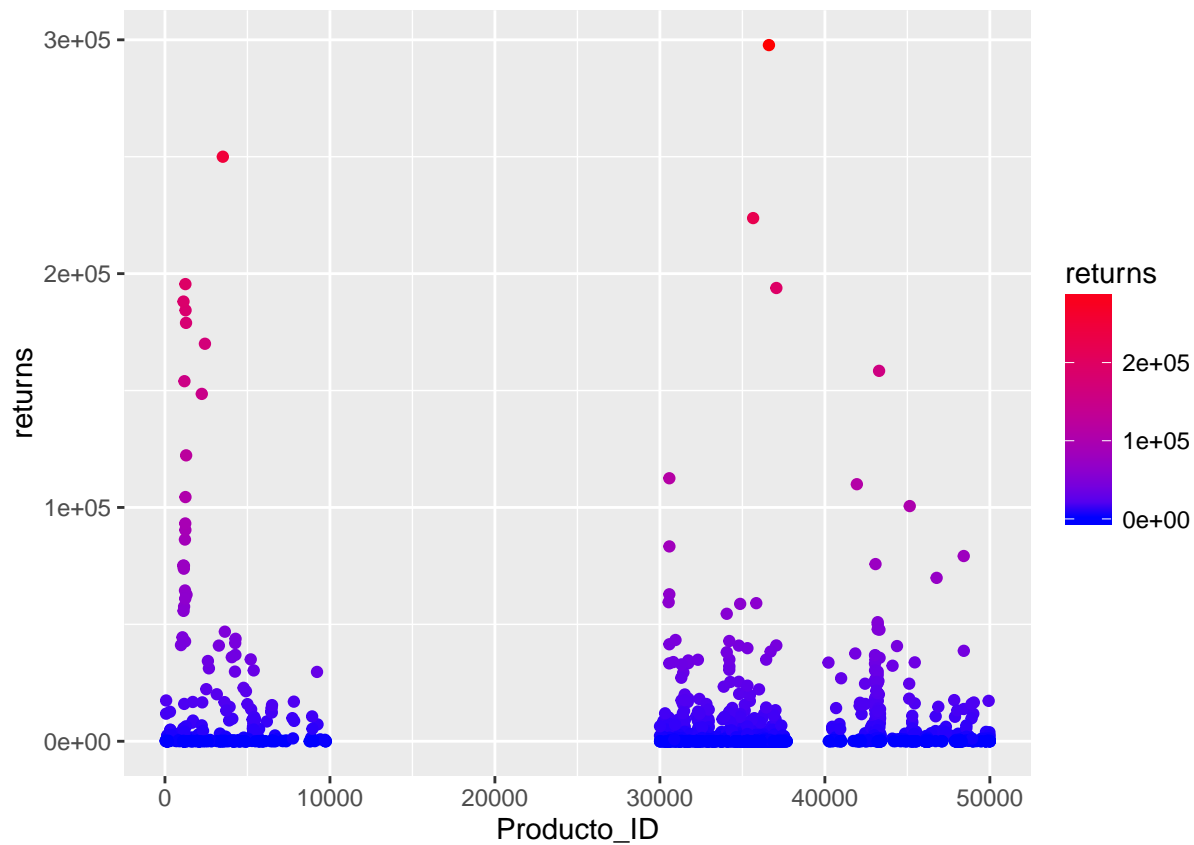
```
# product wise sale (units) & return info.
prodReturn <- train[, .(sales = sum(Venta_uni_hoy), returns = sum(Dev_uni_proxima)), by = Producto_ID]

# product-wise sales (units)
ggplot(prodReturn, aes(x = Producto_ID, y = sales)) + geom_point()
```



0.5 Product wise Sale vs Return

```
# product wise returns  
ggplot(prodReturn, aes(x = Product_ID, y = returns, color = returns)) + geom_point() +  
  scale_color_gradient(low="blue", high="red")
```



0.6 Demand Vs Sales

```
# use just a fraction to plot as the table is huge!
# demand vs sales
# Demanda_uni_quil vs Venta_hoy

#library("dplyr")
#ggplot(train %>% sample_frac(0.05),aes(x = Venta_uni_hoy , y = Demanda_uni_equil)) +
# geom_point() +
# geom_smooth(method = "lm") +
# scale_x_continuous(name = "Sales") +
# scale_y_continuous(name = "Demand") +
# ggtitle("Demand Vs Sales (in terms of units)")
```