

# BOPS

Juan Pablo Herrera

20/7/2021

Se analiza el caso de la apertura del canal BOPS en Home and Kitchen.

## Read Data

### Online Data

```
online_data <- read.csv("C:\\Users\\jphm1\\Documents\\opi_test\\BOPS_case\\BOPS_case\\bops_online.csv")

# basic formatting
online_data <- online_data %>%
  select(1:7) %>%
  mutate(sales = sales %>% stringr::str_replace_all(",", "")) %>% as.numeric() %>%
  rename("dma"="id..DMA.") %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u")) %>%
  na.omit()

glimpse(online_data)

## Rows: 10,500
## Columns: 8
## $ dma      <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ year     <int> 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, ~
## $ month    <int> 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, ~
## $ week     <int> 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30~
## $ after    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ close    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ sales    <dbl> 18564, 30883, 37425, 32563, 35773, 33142, 31752, 44807~
## $ graph_date <date> 2011-04-25, 2011-05-02, 2011-05-09, 2011-05-16, 2011--

# Proporción de dma con tienda cerca
close_dma <- online_data %>%
  group_by(dma) %>%
  summarise(close = max(close), .groups='drop') %>%
  pull(close) %>%
  sum()

# Número de áreas geográficas
cat("Numero de dma: ",online_data$dma %>% unique() %>% length,

# Fechas
"\n Fecha min: ",as.character(min(online_data$graph_date)),
"\n Fecha max: ",as.character(max(online_data$graph_date)),

"\n El ", close_dma/210, "de las dma tiene una tienda cerca")

## Numero de dma: 210
## Fecha min: 2011-04-25
```

```
## Fecha max: 2012-04-02
## El 0.4809524 de las dma tiene una tienda cerca
```

## Data de Tiendas Físicas

```
bm_data <- read.csv("C:\\Users\\jphm1\\Documents\\opi_test\\BOPS_case\\BOPS_case\\bops_bm.csv")

bm_data <- bm_data %>%
  select(1:7) %>%
  mutate(sales = sales %>% stringr::str_replace_all(",", "")) %>% as.numeric() %>%
  rename("store"="id..store.") %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u")) %>%
  na.omit()

glimpse(bm_data)

## Rows: 4,452
## Columns: 8
## $ store      <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ year       <int> 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, ~
## $ month      <int> 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, ~
## $ week       <int> 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29~
## $ usa        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ after      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ sales      <dbl> 118691, 113804, 172104, 105591, 94884, 106246, 119068,~
## $ graph_date <date> 2011-04-18, 2011-04-25, 2011-05-02, 2011-05-09, 2011-~

# Proporción de dma con tienda cerca
usa_store <- bm_data %>%
  group_by(store) %>%
  summarise(usa = max(usa), .groups='drop') %>%
  pull(usa) %>%
  sum()

intro_date <- as.Date(paste(2011, 43, 1, sep="-"), "%Y-%U-%u")
# Número de áreas geográficas
cat("Numero de tiendas: ",bm_data$store %>% unique() %>% length,
# Fechas
  "\n Fecha min: ",as.character(min(bm_data$graph_date)),
  "\n Fecha max: ",as.character(max(bm_data$graph_date)),
  "\n El ", usa_store/84, "de las dma tiene una tienda cerca",
  "\n Fecha inicio programa BOPS: ", as.character(intro_date))

## Numero de tiendas: 84
## Fecha min: 2011-04-18
## Fecha max: 2012-04-16
## El 0.797619 de las dma tiene una tienda cerca
## Fecha inicio programa BOPS: 2011-10-24
```

## Análisis de Ventas Totales

```
# Unimos los dos df en uno solo
# Procesamos el df de online para poder unir
od <- online_data %>%
```

```

group_by(year, month, week) %>%
  summarise(online_sales = sum(sales),
            avg_online_sales = mean(sales),
            .groups='drop')

# Procesamos el df de tiendas para poder unir
bm <-bm_data %>%
  group_by(year, month, week) %>%
  summarise(store_sales = sum(sales),
            avg_store_sales = mean(sales),
            after =max(after),
            .groups='drop')

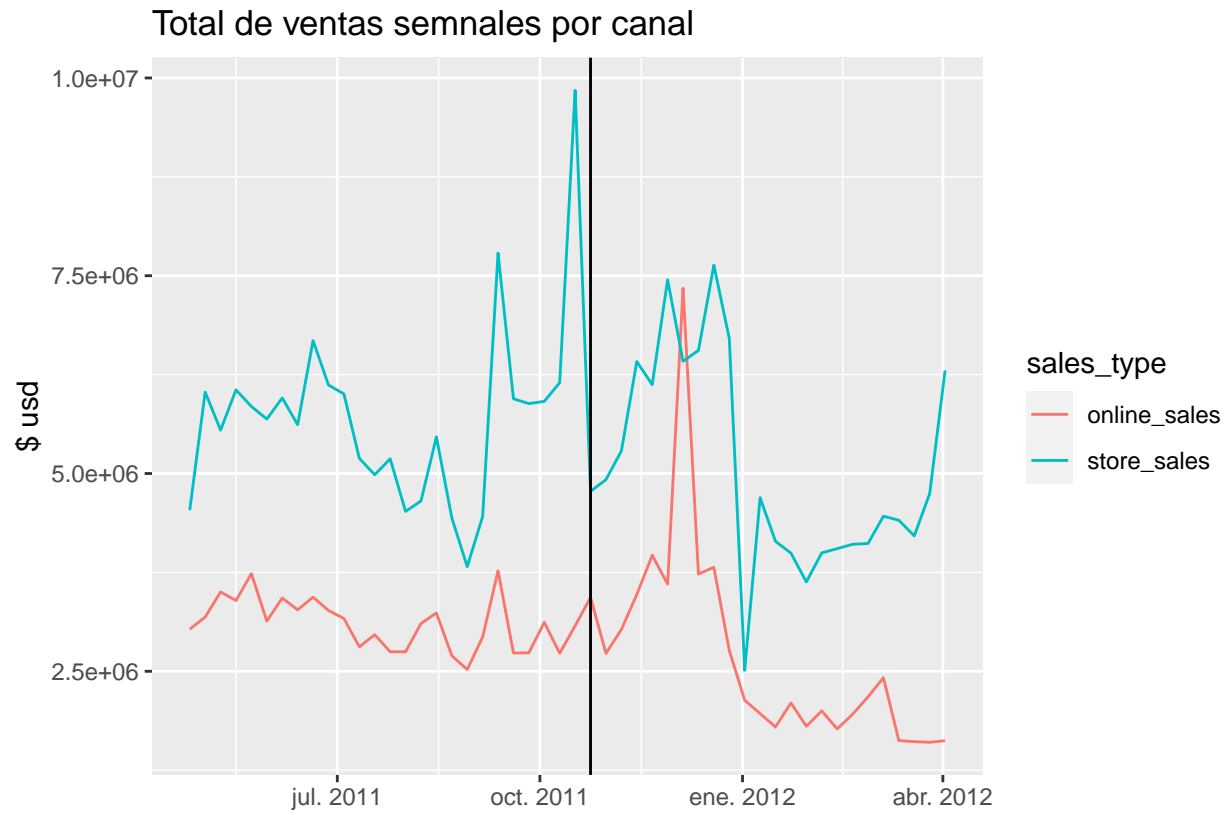
#unimos los df
all_data <- od %>%
  inner_join(bm, by=c("year", "month", "week"))

# Formato largo para graficar
long_data <- all_data %>%
  tidyr::pivot_longer(-c(year, month,week,after), names_to = "sales_type", values_to = "value") %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u"))

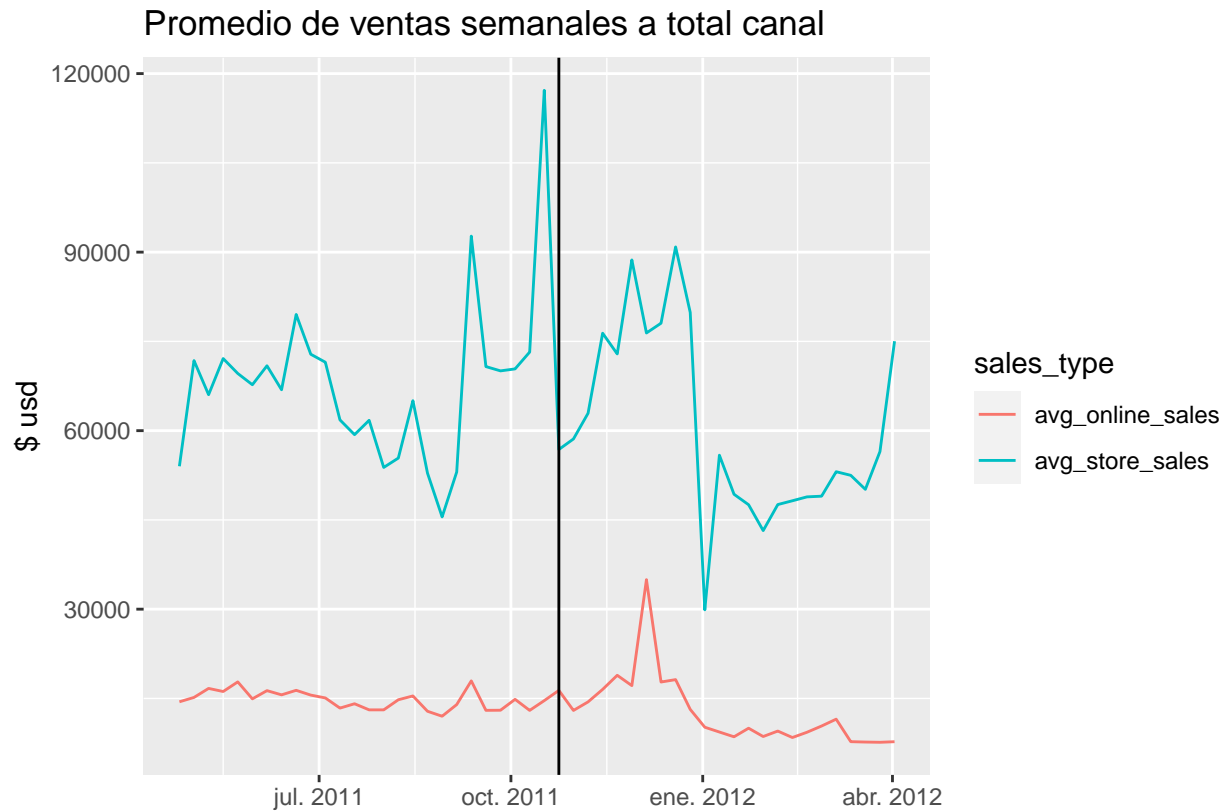
# Seleccion de indicadores a graficar
sales_long <- long_data %>% filter(sales_type %in% c('online_sales','store_sales'))
avg_sales_long <- long_data %>% filter(!sales_type %in% c('online_sales','store_sales'))

# Grafica de ventas totales por canal
ggplot(data = sales_long, aes(x=graph_date, y = value, colour=sales_type))+
  geom_line() +
  ggtitle("Total de ventas semanales por canal")+
  geom_vline(aes(xintercept = intro_date))+
  ylab("$ usd")+
  xlab("")

```



```
# Grafica de ventas promedio por canal
ggplot(data = avg_sales_long, aes(x=graph_date, y = value, colour=sales_type))+
  geom_line()+
  geom_vline(xintercept = intro_date)+
  ggtitle("Promedio de ventas semanales a total canal")+
  ylab("$ usd")+
  xlab("")
```



En las gráficas anteriores la línea vertical indica la fecha en que se inicio el programa BOPS.

De la gráfica de ventas totales por canal podemos tener los siguientes insights:

1. Las ventas en tienda tuvieron un caída importante durante la primera semana del programa.
2. Las ventas en tienda habían tenido un aumento importante justo antes de iniciar el programa.
3. Los meses siguientes al inicio del programa notamos un aumento en ventas en ambos canales.
4. El punto de ventas en línea que supera las ventas en tienda corresponde a la semana 49 en noviembre. Es el black friday.
5. A inicio del año 2012 las ventas caen en ambos canales.
6. En la última semana de abril 2012 notamos que las ventas en tienda se recuperan.

En la gráfica de ventas promedio podemos notar los mismos efectos en menor proporción.

Observando ese comportamiento me gustaría replantear los números que presenta la mesa directiva de la siguiente manera.

```
all_data %>%
  mutate(period = ifelse(week <= 43 & year==2011, 'before',
                        ifelse(year >2011, 'after_2012', 'after_2011'))) %>%
  group_by(period) %>%
  summarise_at(vars(starts_with("avg")), mean) %>%
  arrange(desc(period))
```

```
## # A tibble: 3 x 3
##   period      avg_online_sales avg_store_sales
##   <chr>          <dbl>          <dbl>
## 1 before          14798.          67502.
## 2 after_2012       9041.          50481.
```

```
## 3 after_2011          18223.          76076.
```

Tomando la información mostrada en la tabla se vuelve muy notorio el comportamiento de que cuando salió el programa las ventas aumentaron y al cambiar el año disminuyeron.

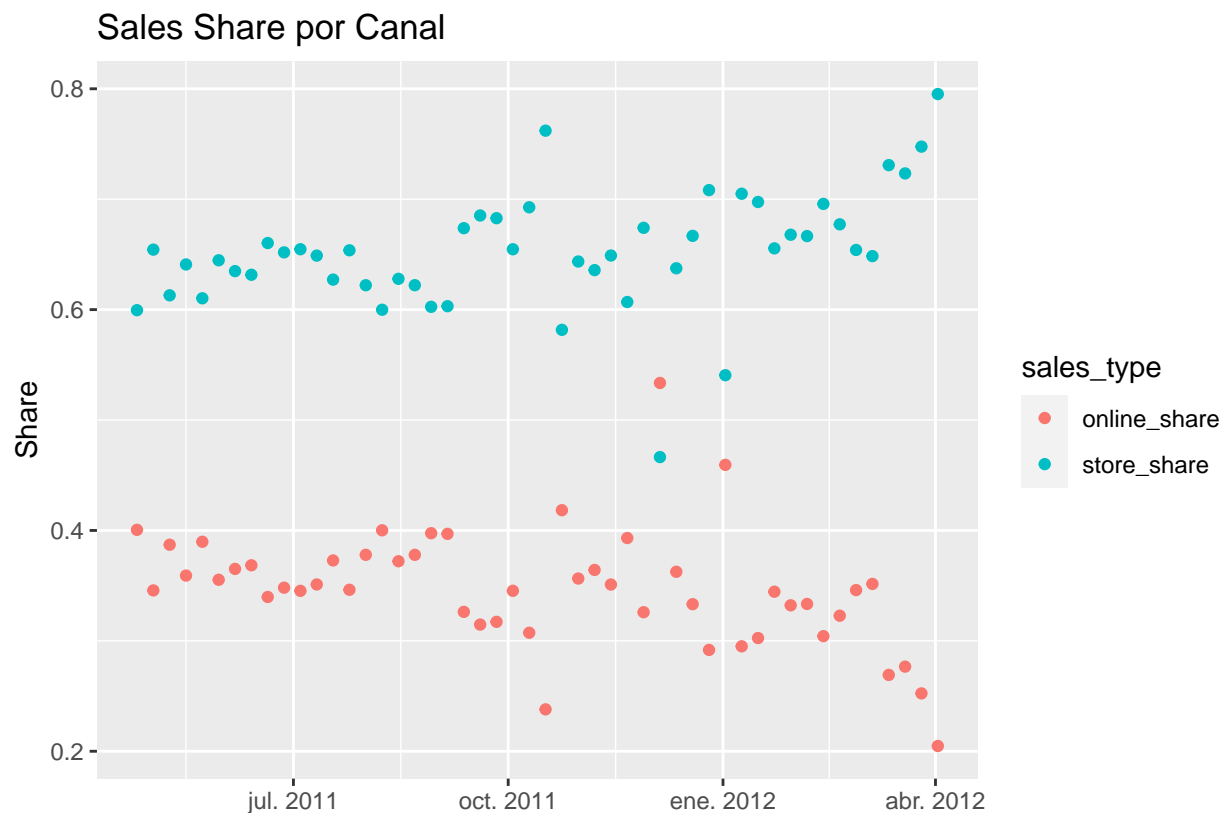
Con este argumento pienso que podemos abrir la discusión sobre si la disminución en ventas se debe al programa BOPS o estamos viendo algún otro efecto.

## Cuotas de Mercado por Canal

```
share_data <- all_data %>%
  mutate(total_semanal = online_sales + store_sales,
         online_share = online_sales / total_semanal,
         store_share = store_sales / total_semanal) %>%
  select(year, week, month, after, online_share, store_share)

long_data <- share_data %>%
  tidyr::pivot_longer(-c(year, month, week, after), names_to = "sales_type", values_to = "value") %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u"))

ggplot(data = long_data, aes(x=graph_date, y=value, colour=sales_type))+
  geom_point()+
  ggtitle("Sales Share por Canal")+
  ylab("Share")+
  xlab("")
```



De acuerdo al programa esperamos un aumento en la share del canal online, sin embargo observamos lo

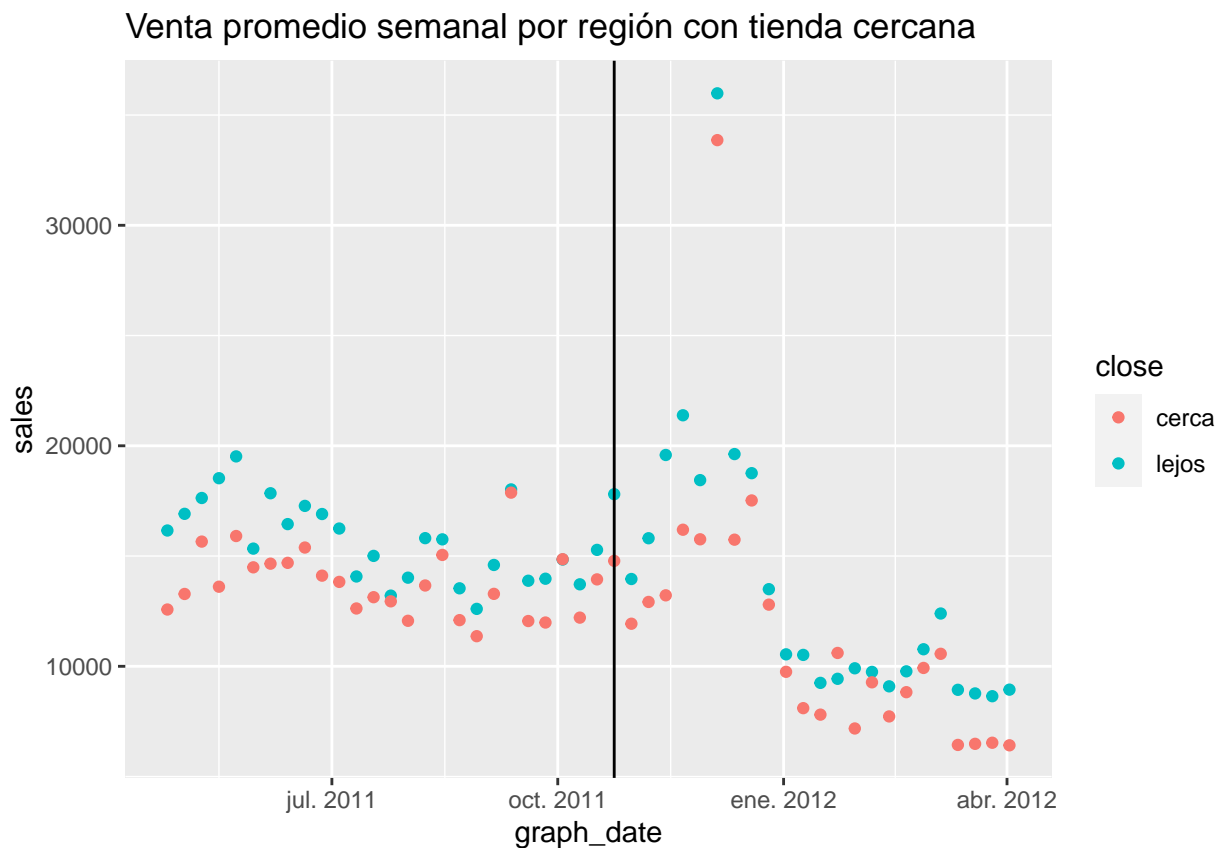
contrario.

## Efectos en Ventas Online

Al analizar las regiones geográficas podemos observar que hay una división importante en los datos. Las regiones que tienen una tienda a menos de 50 mi de distancia y las que no. En cada tipo de tienda podemos esperar un comportamiento distinto. Para las regiones que tienen una tienda cerca esperaríamos un aumento en las ventas en línea mientras que para las regiones que no tienen tienda cerca no lo esperamos. Esto debido a que los clientes no van a viajar grandes distancias para recoger el producto, si ese es su caso optarán por esperar un envío.

```
# Ventas en tiendas que se encuentran lejos
total_online_sales <- online_data %>%
  group_by(year, month, week, close) %>%
  summarise(sales=mean(sales),
            after = max(after),
            .groups='drop') %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u"),
         close = ifelse(close == 1, 'cerca', 'lejos'),
         after = ifelse(after == 1, 'despues', 'antes'))

# Grafica de ventas totales en candada y EU
ggplot(data =total_online_sales, aes(x=graph_date, y=sales, colour=close))+
  ggtitle("Venta promedio semanal por región con tienda cercana")+
  geom_point()+
  geom_vline(aes(xintercept = intro_date))
```



```
total_online_sales %>%
  group_by(after, close) %>%
  summarise(avg_sales= mean(sales), .groups='drop') %>%
  tidyr::pivot_wider(names_from = "after", values_from = "avg_sales") %>%
  mutate(growth_rate = despues / antes -1)
```

```
## # A tibble: 2 x 4
##   close antes despues growth_rate
##   <chr> <dbl> <dbl>      <dbl>
## 1 cerca 13744. 11681.    -0.150
## 2 lejos 15659. 13815.    -0.118
```

En la tabla observamos los valores medios de venta por semana de las regiones que tienen tiendas cerca y tiendas lejos. Leemos la media de venta por region antes del programa, después y la tasa de crecimiento.

El cambio en las ventas online fue de mayor impacto para las regiones que tiene tiendas cerca. Es decir que no estamos observando el efecto esperado del programa en esta área.

## Efectos en Ventas de Tiendas Físicas

Las ventas en tienda presentan un oportunidad única. En estos datos podemos analizar las tiendas de EU y Canadá y comparar las tendencias en una región que si se aplicó el programa y una que no.

De acuerdo al programa, esperamos que más clientes hagan sus compras online y esto disminuya las ventas en tienda.

Las tasas de crecimiento en cada país son las siguientes:

```
# Ventas en tiendas que se encuentran lejos
total_store_sales <- bm_data %>%
  group_by(year, month, week, store) %>%
  summarise(sales=mean(sales),
            program = max(after),
            location =max(usa),
            .groups='drop') %>%
  mutate(graph_date = as.Date(paste(year, week, 1, sep="-"), "%Y-%U-%u"),
         location = ifelse(location == 1, 'EU', 'Canada'),
         program = ifelse(program == 1, 'despues', 'antes'))

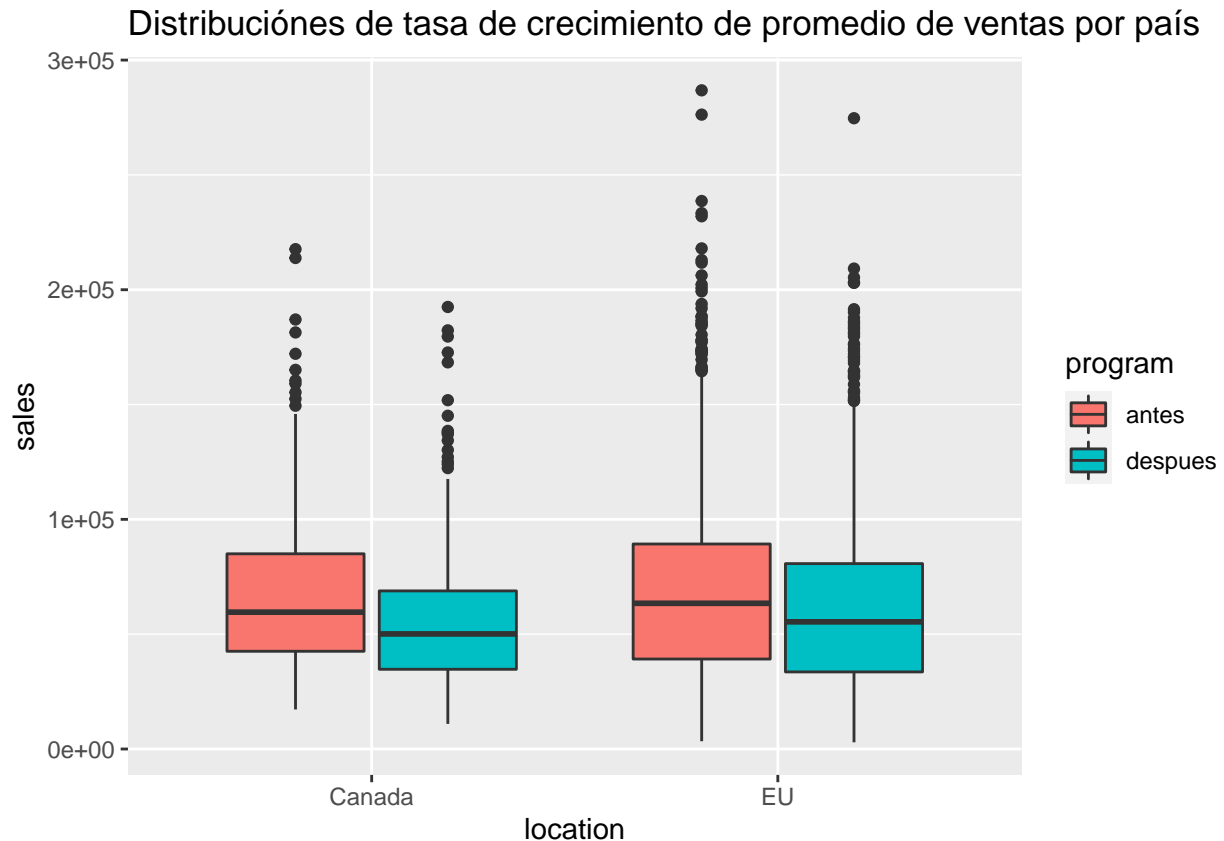
total_store_sales %>%
  group_by(program, location) %>%
  summarise(avg_sales= mean(sales), .groups='drop') %>%
  tidyr::pivot_wider(names_from = "program", values_from = "avg_sales") %>%
  mutate(growth_rate = despues / antes -1)
```

```
## # A tibble: 2 x 4
##   location antes despues growth_rate
##   <chr>      <dbl> <dbl>      <dbl>
## 1 Canada  66862. 55950.    -0.163
## 2 EU      67844. 60813.    -0.104
```

Con distribuciones que tiene la forma:

```
ggplot(data = total_store_sales, aes(x=location, y=sales, fill=program))+
  geom_boxplot()+
  ggtitle("Distribuciones de tasa de crecimiento de promedio de ventas por país ")
```





El efecto ha sido mayor en las tiendas de Canadá, que han decrecido las ventas en 16% en comparación con EU que bajó 10%. Esta disminución indica que las ventas totales cayeron más en donde no está activo el programa, sin embargo vemos una disminución consistente en las ventas de cada país. Es posible que estemos observando un caso en una venta que lleva al cliente a la tienda puede escoger más productos estando ahí.

También es notable que observamos una disminución en las disminuciones fueron mayores en la tienda online que en las tiendas físicas al analizar canada, mientras que en EU la disminución en la tienda online es mayor que la disminución en la tienda física. Este comportamiento también es contrario a lo que esperamos ver de acuerdo a los principios del programa BOPS.

## Análisis de Correlación

Finalmente tenemos un análisis de correlación buscando indicios de canibalización en los canales

```
cat(" Correlación para toda la serie :", cor(all_data$online_sales, all_data$store_sales))

## Correlación para toda la serie : 0.5327413

before <- all_data %>% filter(after==0)
after <- all_data %>% filter(after==1)
cat("\n Correlación de ventas totales por canal antes de aplicar BOPS: ",
    cor(before$online_sales, before$store_sales))

##
## Correlación de ventas totales por canal antes de aplicar BOPS: 0.4431334

cat("\n Correlación de ventas totales por canal después de aplicar BOPS: ",
    cor(after$online_sales, after$store_sales))

##
```

```
## Correlación de ventas totales por canal después de aplicar BOPS: 0.6050372
```

El aumento relativo en la correlación puede indicar el comportamiento de que un canal de venta favorece al otro, como es el caso de que una venta en línea lleva al cliente a la tienda y ahí elige más productos.

## Conclusiones

De acuerdo con el análisis de datos que hemos realizado tenemos las siguientes conclusiones:

1. Hay un decremento en las ventas que se observa en todos los canales de venta.  
No es posible aislar este efecto al programa BOPS, de hecho el comportamiento de la media de ventas cuando recién salió el programa y el cambio para el siguiente año sugiere que estamos observando otro efecto. Se tiene la hipótesis de que pueden ser efectos de estacionalidad pero se requiere más información para comprobarlo.
2. Una hipótesis alternativa es que podemos estar observando un efecto en que el canal de venta online lleva a clientes a tiendas y tienen nuevos consumos, aumentando el consumo en tienda. Los efectos que sugieren esta hipótesis son los siguientes:
  - Aumento en el share de ventas totales de canal de tiendas físicas
  - Mayor disminución de ventas en tienda en el área geográfica en que no está activo el programa
  - Aumento de correlación entre ventas en tienda y online
  - Mayor disminución de ventas online en áreas geográficas que tienen tiendas cercanas.

En conclusión, lo que observamos en los números no son efectos que se pueden aislar directamente al programa BOPS, además hay evidencia que indica que puede tener efectos positivos. Además, el programa aumenta canales de venta y beneficia la organización de la empresa. **La sugerencia es continuar con el programa observando de cerca los movimientos en ventas que se tienen, a la vez analizar más información y proceder con la apertura en Canadá en un periodo más adelante.**

## Cuanto se ha Perdido?

```
bm <- bm_data %>%
  group_by(after, week, year) %>%
  summarise(count=n(), .groups='drop') %>%
  group_by(after) %>%
  summarise(count=n(), .groups='drop')
active_weeks <- bm %>% filter(after==1) %>% pull(count)

dlls <- active_weeks * (2004*210)*(7545*84)
cat ("Se han perdido aprox. ", dlls/1000000 , "M de dólares desde que inició BOPS")
```

```
## Se han perdido aprox. 6934719 M de dólares desde que inició BOPS
```