

# 23 - Actividad 08 en R Markdown

Palomeque Jonathan Manuel

2024-04-28

En esta actividad agregare anotaciones a una visualizacion de datos con ggplot2. Tambien guardare imagenes de visualizaciones de ggplot2 para poder agregarlas directamente en presentaciones.

## Escenario

Como analista de datos en una empresa de reservas hoteleras, has estado creando visualizaciones en R con el paquete ggplot2 para compartir conclusiones sobre los datos con los interesados. Despues de crear una serie de visualizaciones usando estetica y filtros de ggplot() y ggplot2, tu interesado te pide que agregues anotaciones a tus visualizaciones para ayudar a explicar tus resultados en una presentacion. Afortunadamente, ggplot2 tiene funciones de anotacion integradas.

## Paso 1: Importar los datos

```
library('tidyverse')
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2     3.5.0      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr       1.0.2
## — Conflicts — tidyverse_conflicts() —
## X dplyr::filter() masks stats::filter()
## X dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
reservas_hotel <- read_csv("hotel_bookings.csv", locale = readr::locale(encoding = "latin1"))
```

```
## Rows: 119390 Columns: 32
## — Column specification —
## Delimiter: ","
## chr (14): hotel, arrival_date_month, comida, pa s, market_segment, distribut...
## dbl (18): is_canceled, lead_time, arrival_date_year, arrival_date_week_numbe...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Paso 2: Refrescar la memoria

Obtener una muestra de los datos y tener una vista previa los nombres de todas las columnas:

```
head(reservas_hotel)
```

```
## # A tibble: 6 × 32
##   hotel      is_canceled lead_time arrival_date_year arrival_date_month
##   <chr>      <dbl>      <dbl>          <dbl> <chr>
## 1 Resort Hotel      0        342          2015 July
## 2 Resort Hotel      0        737          2015 July
## 3 Resort Hotel      0         7          2015 July
## 4 Resort Hotel      0        13          2015 July
## 5 Resort Hotel      0        14          2015 July
## 6 Resort Hotel      0        14          2015 July
## # i 27 more variables: arrival_date_week_number <dbl>,
## #   arrival_date_day_of_month <dbl>, stays_in_weekend_nights <dbl>,
## #   stays_in_week_nights <dbl>, adultos <dbl>, `ni\u0096os` <dbl>,
## #   `beb\u008es` <dbl>, comida <chr>, `pa\u0092s` <chr>, market_segment <chr>,
## #   distribution_channel <chr>, is_repeated_guest <dbl>,
## #   previous_cancellations <dbl>, previous_bookings_not_canceled <dbl>,
## #   reserved_room_type <chr>, assigned_room_type <chr>, ...
```

```
colnames(reservas_hotel)
```

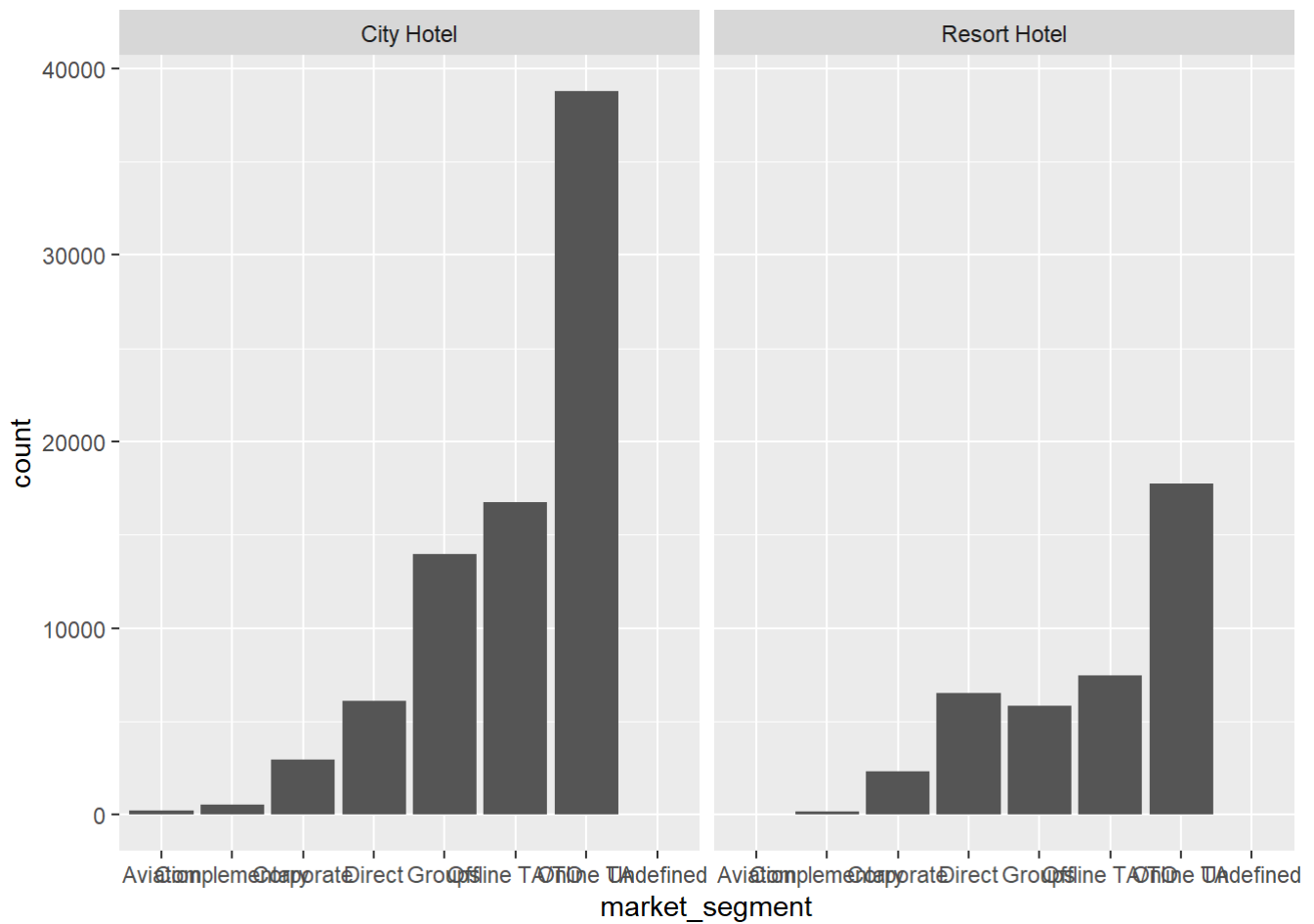
```
## [1] "hotel" "is_canceled"
## [3] "lead_time" "arrival_date_year"
## [5] "arrival_date_month" "arrival_date_week_number"
## [7] "arrival_date_day_of_month" "stays_in_weekend_nights"
## [9] "stays_in_week_nights" "adultos"
## [11] "ni\u0096os" "beb\u008es"
## [13] "comida" "pa\u0092s"
## [15] "market_segment" "distribution_channel"
## [17] "is_repeated_guest" "previous_cancellations"
## [19] "previous_bookings_not_canceled" "reserved_room_type"
## [21] "assigned_room_type" "booking_changes"
## [23] "deposit_type" "agente"
## [25] "empresa" "days_in_waiting_list"
## [27] "customer_type" "adr"
## [29] "required_car_parking_spaces" "total_of_special_requests"
## [31] "reservation_status" "reservation_status_date"
```

## Paso 3: Realizar anotaciones en tu grafico

Tu interesado te menciona que les gustaría que compartieras la visualización desglosando el tipo de pago por ciudad, ya que esto ayudará a determinar cómo la empresa dirigirá las promociones en el futuro. Te piden que crees una versión limpia y etiquetada, y que la guardes como un archivo .png para que ellos puedan incluirla en una presentación

Graficos de barras de city\_hotel y resort\_hotel. En el eje x esta market\_segment y en el eje Y esta count:

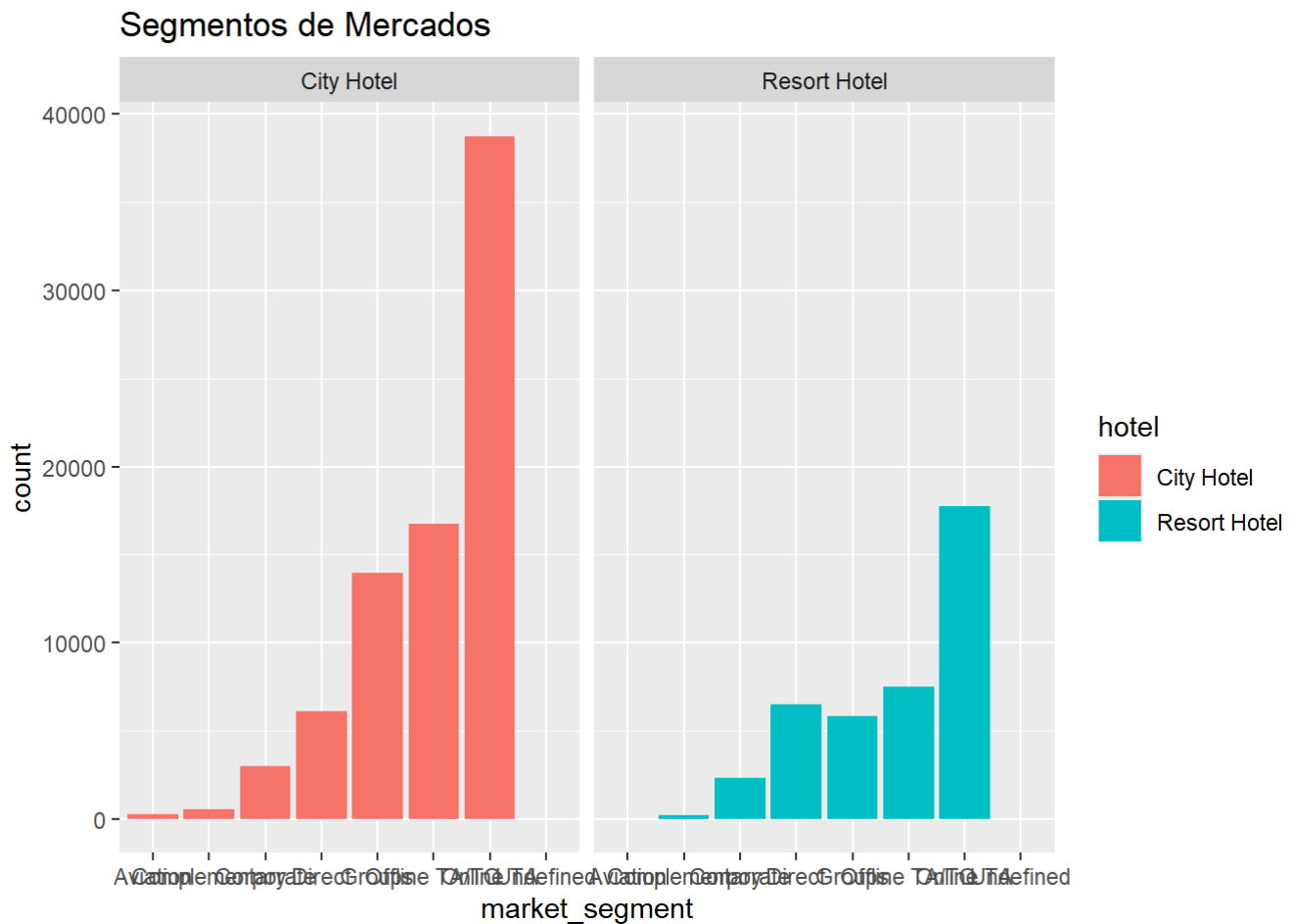
```
ggplot(data = reservas_hotel) +
  geom_bar(mapping = aes(x = market_segment)) +
  facet_wrap(~hotel)
```



En esta visualizacion, no queda claro de donde provienen los datos, cual es la conclusion clave o ni siquiera que muestran los datos. Para explicar todo eso, puedes aprovechar las anotaciones en `ggplot2`

Agregare un titulo (Segmentos de mercado) y diferenciare los graficos por color para darle una mayor claridad a la visualizacion de los datos a los interesados:

```
ggplot(data = reservas_hotel) +
  geom_bar(mapping = aes(x = market_segment, fill= hotel)) +
  facet_wrap(~hotel) +
  labs(title="Segmentos de Mercados")
```



Quiero agregar la información de sobre cual periodo son los datos, para lo cual averigüé de cuando son los datos y los guardaré en variables:

Año mas antiguo:

```
fecha_minima <- min(reservas_hotel$arrival_date_year)
```

Año mas actual:

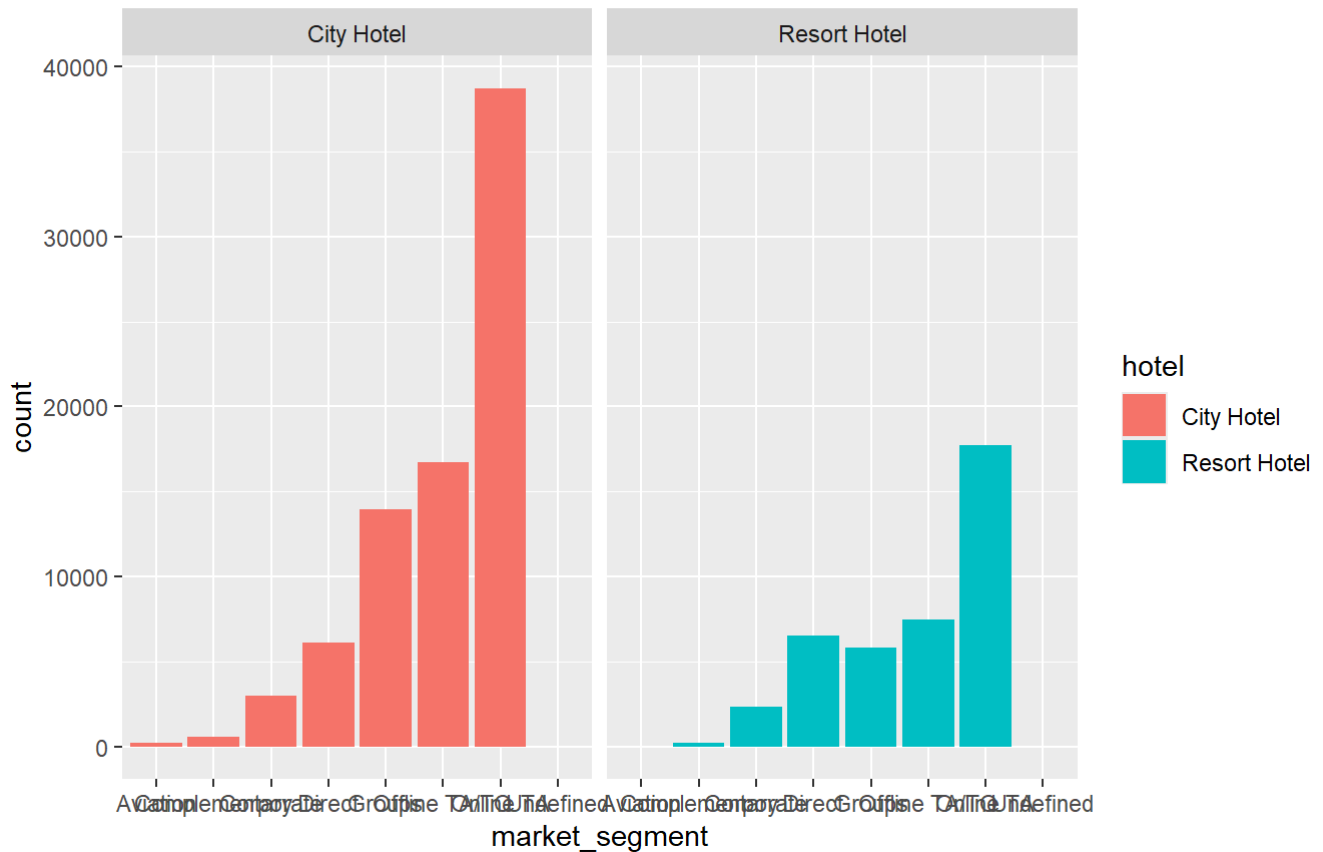
```
fecha_maxima <- max(reservas_hotel$arrival_date_year)
```

Ahora, agrego un subtítulo usando `subtitle=` en la función `labs()`. Luego, usare la función `paste0()` para usar mis variables recién creadas en mis etiquetas. Esto es muy útil, ya que si los datos se actualizan y se agregan datos más recientes, no tendré que modificar el código, porque las variables son dinámicas:

```
ggplot(data = reservas_hotel) +
  geom_bar(mapping = aes(x = market_segment, fill = hotel)) +
  facet_wrap(~hotel) +
  labs(title = "Comparación de Segmentos de Mercados por tipo de hotel para reservas",
       subtitle = paste0("Datos desde el ", fecha_minima, " al ", fecha_maxima))
```

## Comparacion de Segmentos de Mercados por tipo de hotel para reservas

Datos desde el 2015 al 2017



Voy a modificar el subtitle (subtitulo) para que se transforme en una caption (leyenda), que apareciera en la esquina inferior derecha.

```
ggplot(data = reservas_hotel) +  
  geom_bar(mapping = aes(x = market_segment, fill= hotel)) +  
  facet_wrap(~hotel) +  
  theme(axis.text.x = element_text(angle = 45)) +  
  labs(title="Comparacion de Segmentos de Mercados por tipo de hotel para reservas",  
       caption=paste0("Datos desde el ", fecha_minima, " al ", fecha_maxima))
```

## Comparacion de Segmentos de Mercados por tipo de hotel para reservas

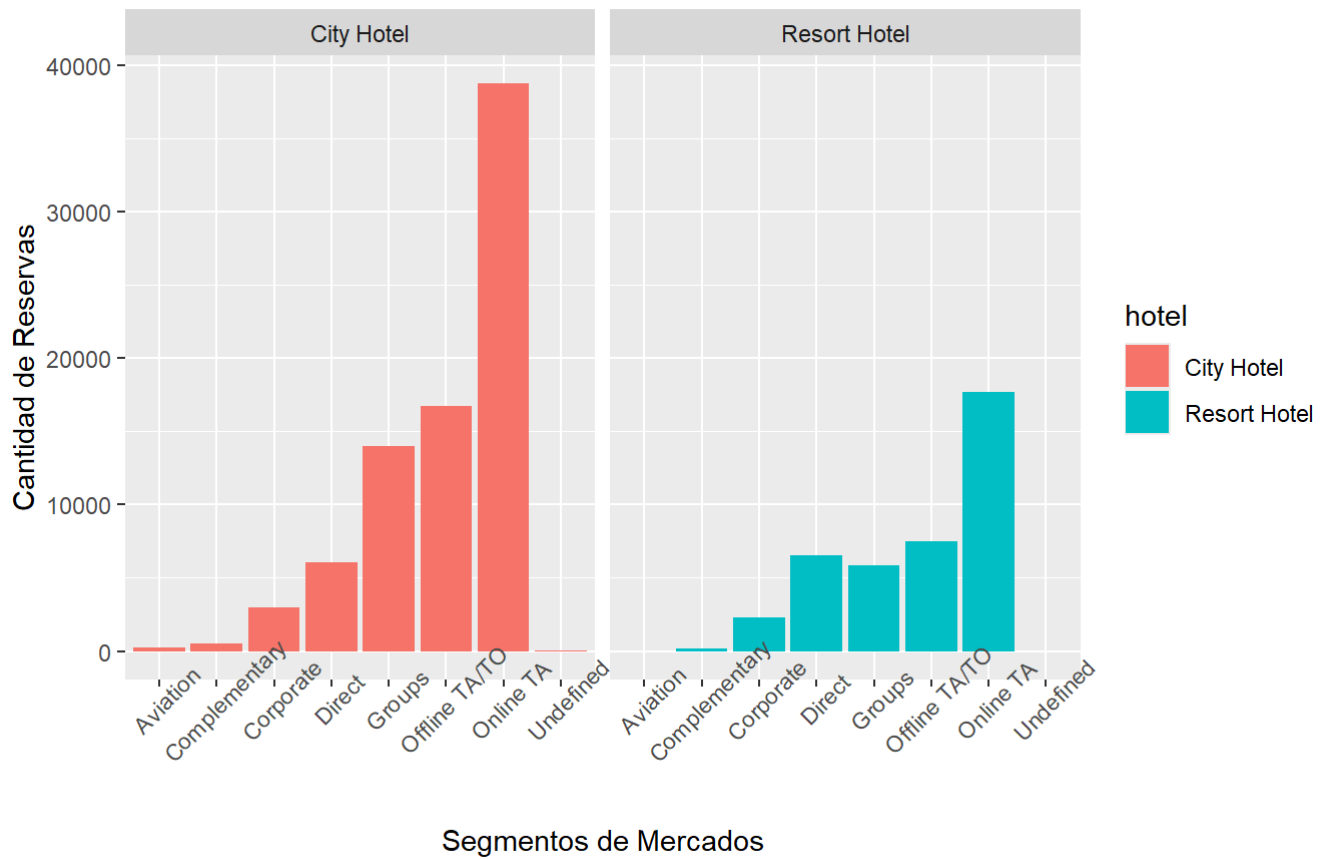


Datos desde el 2015 al 2017

Limpiare las etiquetas del eje x e y para asegurarme de que sean claras. Para hacerlo, agregare la funcion `labs()` y usare `x=` e `y=`

```
ggplot(data = reservas_hotel) +
  geom_bar(mapping = aes(x = market_segment, fill= hotel)) +
  facet_wrap(~hotel) +
  theme(axis.text.x = element_text(angle = 45)) +
  labs(title="Comparacion de Segmentos de Mercados por tipo de hotel para reservas",
       caption=paste0("Datos desde el ", fecha_minima, " al ", fecha_maxima),
       x="Segmentos de Mercados",
       y="Cantidad de Reservas")
```

## Comparacion de Segmentos de Mercados por tipo de hotel para reservas



Datos desde el 2015 al 2017

## Paso 5: Guardar tu grafico

usare la funcion `ggsave()` para guardar el imagen como un 7x7

```
ggsave('grafico2.png',  
       width= 7,  
       height= 7)
```