

# C7M4L2A1: You and GGPlot

## Background for this activity

En esta actividad, revisará un escenario y utilizará ggplot2 para crear rápidamente visualizaciones de datos que le permitan explorar sus datos y obtener nuevos conocimientos. Aprenderá más sobre la sintaxis básica de ggplot2 y la visualización de datos en R.

A lo largo de esta actividad, también tendrá la oportunidad de practicar la escritura de su propio código realizando usted mismo cambios en los fragmentos de código. Si encuentra un error o se queda atascado, siempre puede consultar el archivo Lesson2\_GGPlot\_Solutions .rmd en la carpeta Soluciones en la Semana 4 para obtener el código completo y correcto.

## The Scenario

En este escenario, usted es un analista de datos junior que trabaja para una empresa de reservas de hoteles. Ha limpiado y manipulado sus datos y ha obtenido algunos conocimientos iniciales que le gustaría compartir. Ahora, creará algunas visualizaciones de datos simples con el paquete ggplot2 . Utilizará la sintaxis básica de ggplot2 y solucionará algunos errores comunes que puedan surgir.

## Step 1: Import your data

En el siguiente fragmento, utilizará la función read\_csv() para importar datos desde un .csv en la carpeta del proyecto llamada "hotel\_bookings.csv" y guardarlo como un marco de datos llamado hotel\_bookings :

Si esta línea causa un error, cópiela en la línea setwd("/cloud/project/Course 7/Week 4") anterior.

```
hotel_bookings <- read.csv("hotel_bookings.csv")
```

## Step 2: Look at a sample of your data

Utilice la función head() para obtener una vista previa de sus datos:

```
head(hotel_bookings)
```

```

##      hotel is_canceled lead_time arrival_date_year arrival_date_month
## 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
## arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights
## 1      27      1      0
## 2      27      1      0
## 3      27      1      0
## 4      27      1      0
## 5      27      1      0
## 6      27      1      0
## stays_in_week_nights adults children babies meal country market_segment
## 1      0      2      0      0 BB      PRT      Direct
## 2      0      2      0      0 BB      PRT      Direct
## 3      1      1      0      0 BB      GBR      Direct
## 4      1      1      0      0 BB      GBR      Corporate
## 5      2      2      0      0 BB      GBR      Online TA
## 6      2      2      0      0 BB      GBR      Online TA
## distribution_channel is_repeated_guest previous_cancellations
## 1      Direct      0      0
## 2      Direct      0      0
## 3      Direct      0      0
## 4      Corporate      0      0
## 5      TA/TO      0      0
## 6      TA/TO      0      0
## previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1      0      C      C
## 2      0      C      C
## 3      0      A      C
## 4      0      A      A
## 5      0      A      A
## 6      0      A      A
## booking_changes deposit_type agent company days_in_waiting_list customer_type
## 1      3      No Deposit      NULL      NULL      0      Transient
## 2      4      No Deposit      NULL      NULL      0      Transient
## 3      0      No Deposit      NULL      NULL      0      Transient
## 4      0      No Deposit      304      NULL      0      Transient
## 5      0      No Deposit      240      NULL      0      Transient
## 6      0      No Deposit      240      NULL      0      Transient
## adr required_car_parking_spaces total_of_special_requests reservation_status
## 1      0      0      0      Check-Out
## 2      0      0      0      Check-Out
## 3      75      0      0      Check-Out
## 4      75      0      0      Check-Out
## 5      98      0      1      Check-Out
## 6      98      0      1      Check-Out
## reservation_status_date
## 1      2015-07-01
## 2      2015-07-01

```

```
## 3      2015-07-02
## 4      2015-07-02
## 5      2015-07-03
## 6      2015-07-03
```

También puedes usar `colnames()` para obtener los nombres de todas las columnas en tu conjunto de datos. Ejecute el siguiente fragmento de código para descubrir los nombres de las columnas en este conjunto de datos:

```
colnames(hotel_bookings)
```

```
## [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" "adults"
## [11] "children"        "babies"
## [13] "meal"            "country"
## [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"    "agent"
## [25] "company"         "days_in_waiting_list"
## [27] "customer_type"   "adr"
## [29] "required_car_parking_spaces" "total_of_special_requests"
## [31] "reservation_status" "reservation_status_date"
```

## Step 3: Install and load the ‘ggplot2’ package

Si aún no ha instalado y cargado el paquete `ggplot2`, deberá hacerlo antes de poder usar la función `ggplot()`.

Ejecute el siguiente fragmento de código para instalar y cargar `ggplot2`. Esto puede tomar unos pocos minutos.

## Step 4: Begin creating a plot

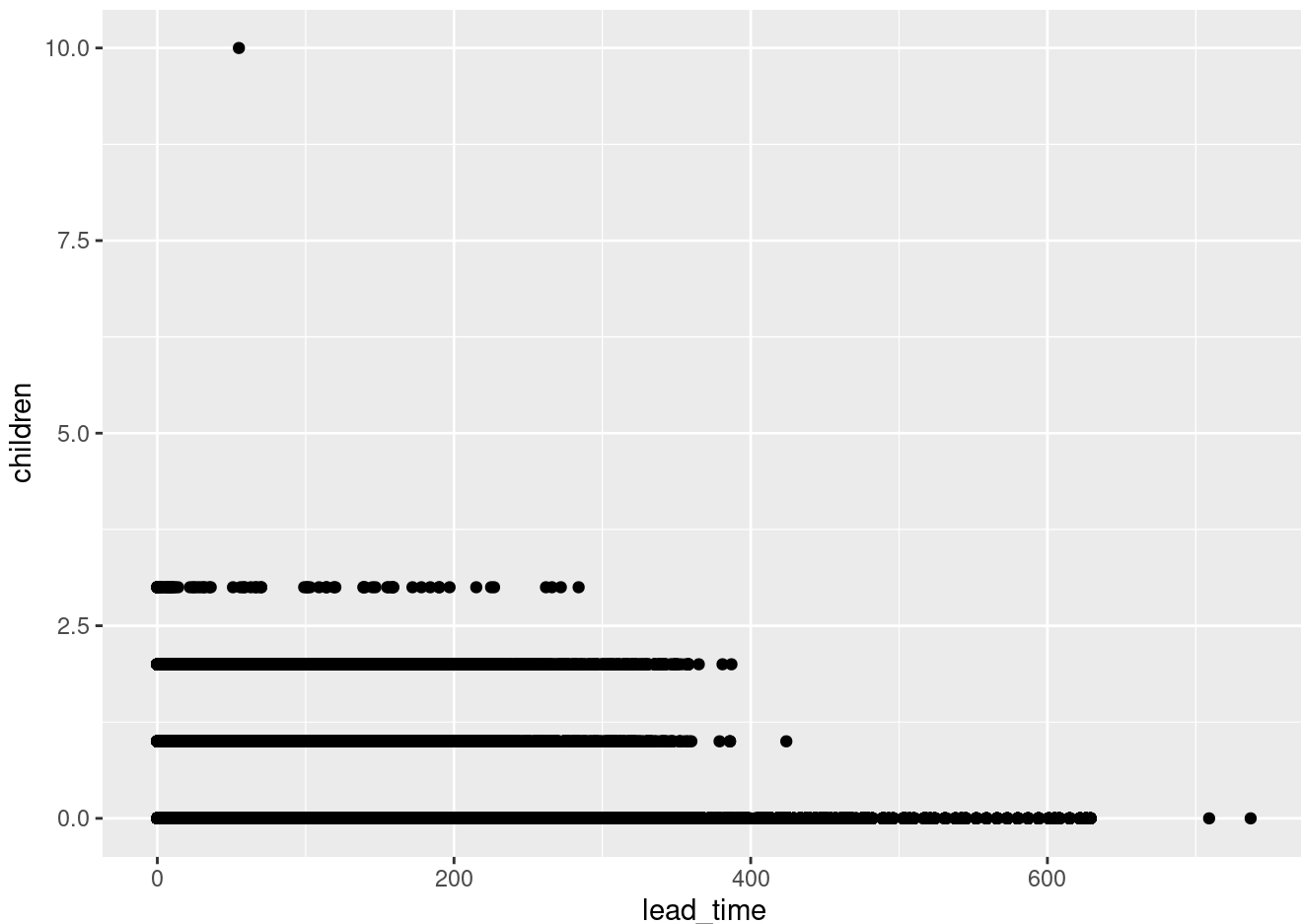
Una parte interesada le dice: “Quiero dirigirme a personas que reservan con anticipación y tengo la hipótesis de que las personas con niños deben reservar con anticipación”.

Cuando comienzas a explorar los datos, no muestran lo que esperarías. Es por eso que decide crear una visualización para ver qué tan cierta es (o no) esa afirmación.

Puedes usar `ggplot2` para hacer esto. Intente ejecutar el siguiente código:

```
ggplot(data = hotel_bookings) +
  geom_point(mapping = aes(x = lead_time, y = children))
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



La función `geom_point()` utiliza puntos para crear un diagrama de dispersión. Los diagramas de dispersión son útiles para mostrar la relación entre dos variables numéricas. En este caso, el código asigna la variable 'lead\_time' al eje x y la variable 'hijos' al eje y.

En el eje x, el gráfico muestra con cuánta antelación se realiza una reserva, siendo las reservas más a la derecha las que se realizan con mayor antelación. En el eje y muestra cuántos niños hay en un grupo.

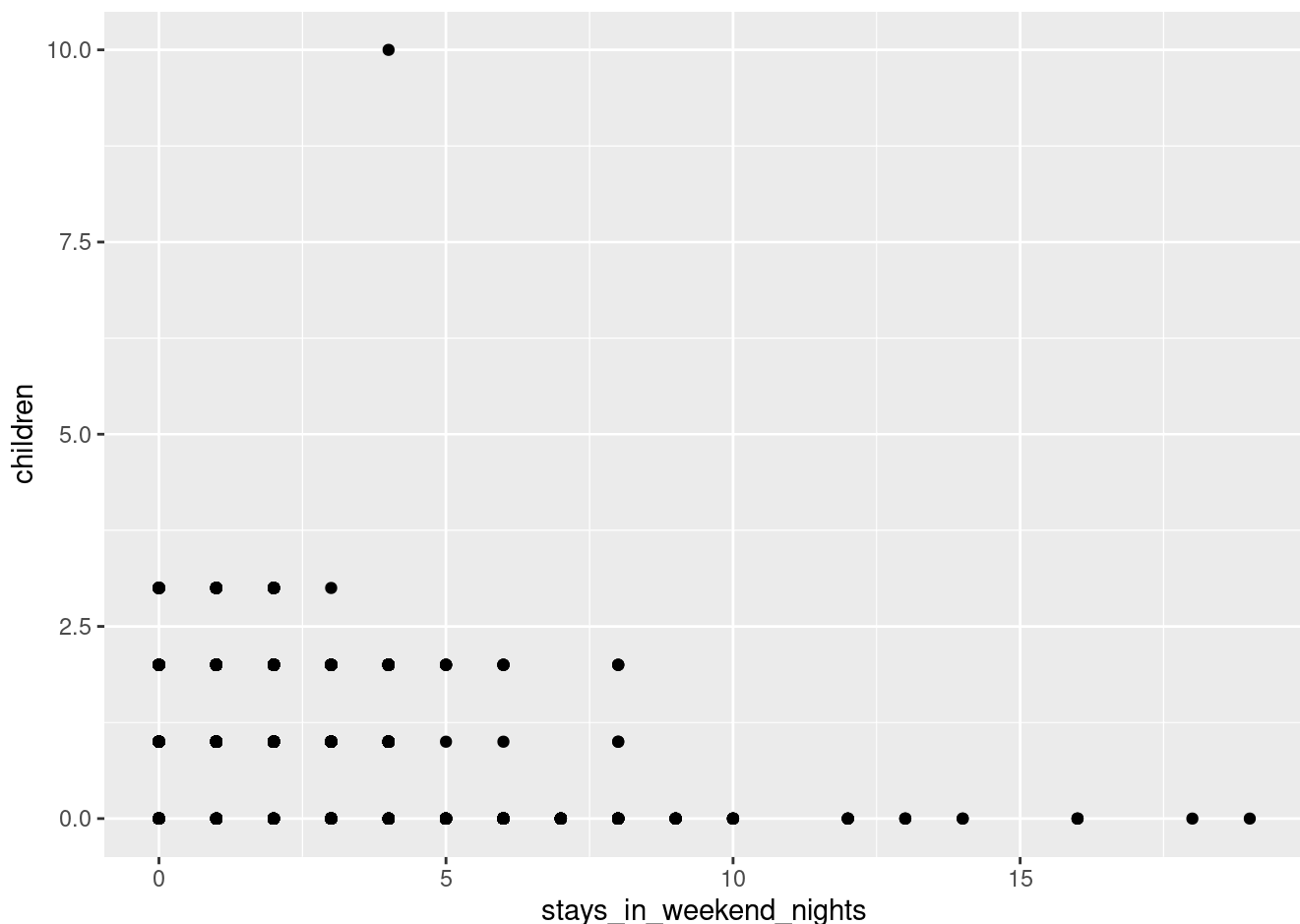
La trama revela que la hipótesis de su parte interesada es incorrecta. Usted informa a su parte interesada que muchas de las reservas anticipadas las realizan personas con 0 hijos. **## Step 5: Try it on your own**

A continuación, su parte interesada dice que quiere aumentar las reservas de fin de semana, una importante fuente de ingresos para el hotel. Su parte interesada quiere saber qué grupo de huéspedes reserva la mayor cantidad de noches de fin de semana para poder dirigirse a ese grupo en una nueva campaña de marketing. Ella sugiere que los huéspedes sin niños reserven la mayor cantidad de noches de fin de semana. ¿Es esto cierto?

Intente mapear 'stays\_in\_weekend\_nights' en el eje x y 'niños' en el eje y completando el resto del código a continuación.

```
ggplot(data = hotel_bookings) +
  geom_point(mapping = aes(x =stays_in_weekend_nights , y =children ))
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Si ingresa correctamente este código, debería tener un diagrama de dispersión con 'stays\_in\_weekend\_nights' en el eje x y 'niños' en el eje y.

¿Qué descubriste? ¿Tiene razón su parte interesada?

¿Qué otros tipos de tramas podrías utilizar para mostrar esta relación?

Recuerde, si tiene problemas para completar un bloque de código, consulte el documento de soluciones para esta actividad.

## Activity Wrap Up

El paquete `ggplot2` le permite crear rápidamente visualizaciones de datos que pueden responder preguntas y brindarle información sobre sus datos. Ahora que está un poco más familiarizado con la sintaxis básica de `ggplot2`, puede practicar estas habilidades modificando los fragmentos de código en el archivo `rmd` o usar este código como punto de partida en su propia consola de proyecto. Con `ggplot2`, podrás crear y compartir visualizaciones de datos sin salir de tu consola `R`. ¡Aprenderá más sobre `ggplot2` a lo largo de este curso y eventualmente creará visualizaciones aún más complejas y hermosas!