

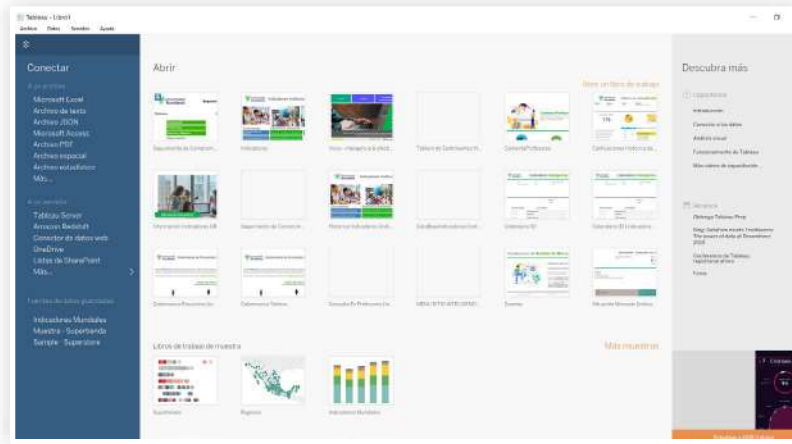
Introducción a Tableau

Tableau es una herramienta dedicada a dar “poder a los datos” a través de la visualización, permitiendo generar libros de trabajo que posibilitan generar reportes o análisis importantes.

Puedes generar visualizaciones desde un nivel bajo en programación hasta un nivel avanzado, dependiendo de lo que busques visualizar.

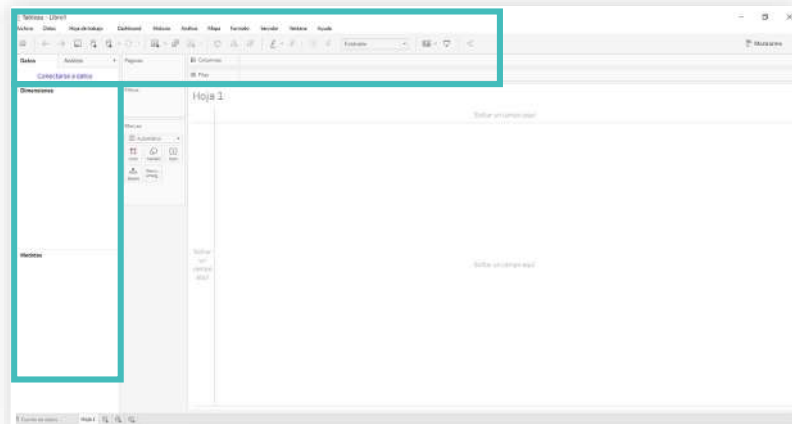
Al momento de tener la licencia o haber iniciado la aplicación aparecerá una pantalla del lado izquierdo que mostrará un listado de las distintas formas de conexión:

- Conexión a archivos en PC.
- Conexión al servidor.
- Fuentes de datos guardadas.



Una vez iniciado Tableau podrás ver la siguiente pantalla:

- Menú principal en la parte superior.
- Submenú para conexión de datos (debajo de ello aparece dimensiones y medidas).



Dimensiones: son los datos que la herramienta considera como cualitativos, por ejemplo, ciudad, estado, servicio, etcétera.

Medidas: son los datos que la herramienta considera cuantitativos, por ejemplo, edad, longitud, registros, entre otros.

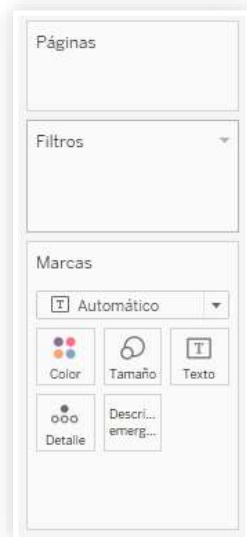
Nota importante: Tableau generará la clasificación dependiendo de cómo se tenga la base, por ejemplo, Edad: si en la base se tiene 10, 20 y 30, el valor numérico será la medida, pero si lo tenemos como 10 años, 20 años y 30 años será dimensión, ya que estará unida con la palabra.

Submenú Mostrarme:



Tableau tiene un submenú en donde aparecen los tipos de gráficos que puedes utilizar y de acuerdo con el tipo de gráfico se utilizará una determinada estructura, por ejemplo, en una gráfica de diagrama de árbol se requieren dos datos de dimensiones y uno de medida.

- **Filtros:** a diferencia de Excel, los filtros en Tableau son colocados en este apartado.
- **Marcas:** son los estilos o formas que le podrás dar a la visualización para estilizarla.

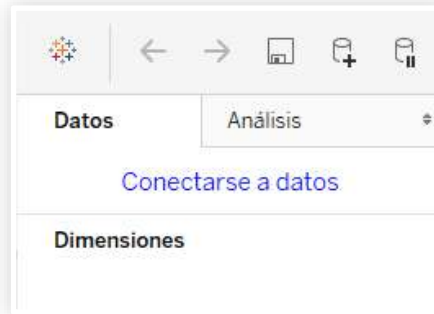


Conexión de la base de datos

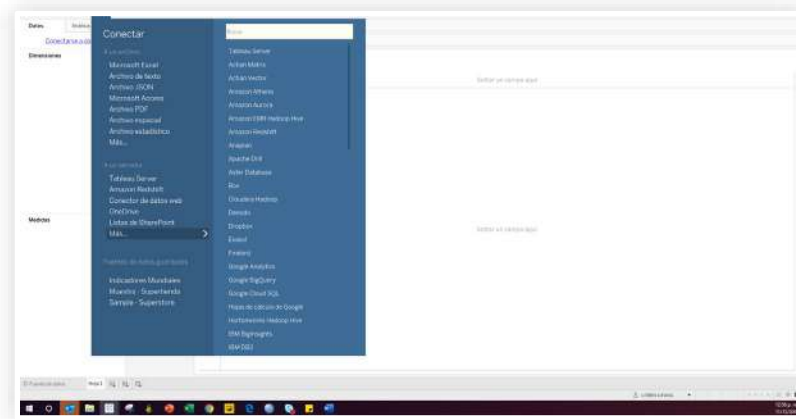
Se puede realizar de tres formas:

1. Haciendo clic en Conectarse a datos.
2. Menú de datos.
3. Arrastrando la base de datos.

La primera opción es la que inicialmente indica Tableau al hacer clic en Conectarse a datos:

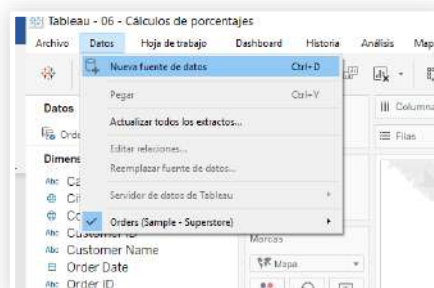


Posteriormente, te aparecerá un listado de opciones para conectar la base de datos o fuente de datos.



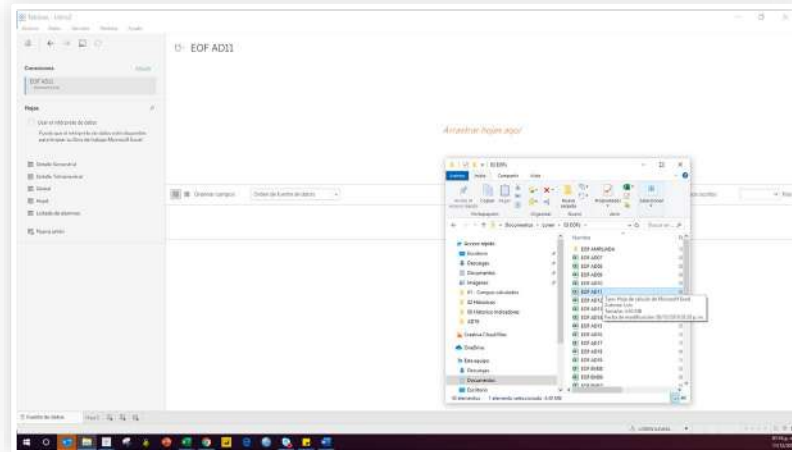
Menú de datos

En la segunda opción, se hace clic en el botón de menú de datos:



Arrastrando la base de datos

Para la tercera opción, una vez que hayas arrastrado la base de datos, aparecerá la siguiente pantalla que te permitirá seleccionar la pestaña que buscas colocar en Tableau:

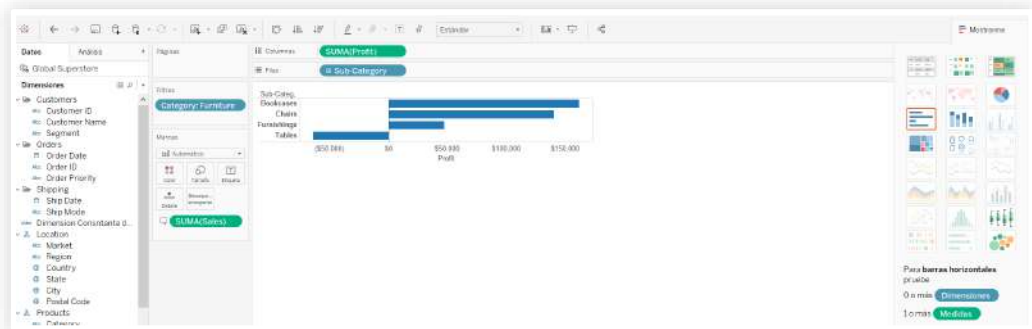


En Tableau podrás generar lo siguiente:

- Campos calculados.
- Mapas.

Campos calculados

Imagina que tuvieras una base de ganancia basada en categorías y quisieras conocer si realmente hubo ganancia en todas o si se presentaron pérdidas en algunas. Para esto, realiza el primer paso que es graficar.



Podrás observar que en el caso de Tables se muestra que la ganancia no fue realmente positiva, sino que tuvo pérdidas.

Con campo calculado podrás distinguirlo a través de colores.

- La fórmula que tendrías que realizar es la siguiente:

Sumatoria de ganancias es menor a 0 para probar si hubo gastos menores.



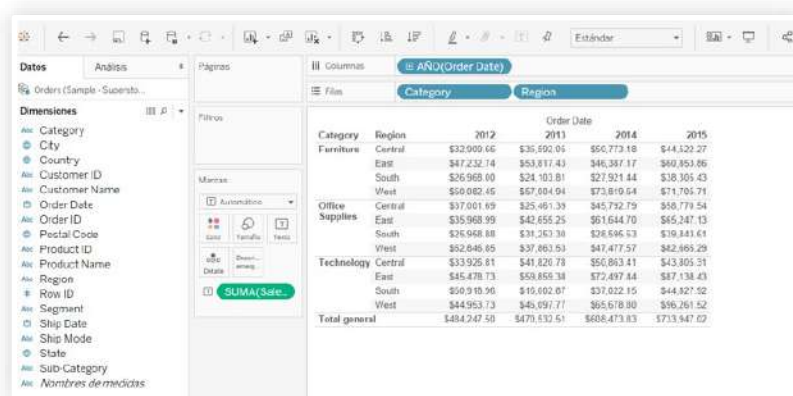
Al realizar ese paso, Tableau interpreta que buscas hacer un True|False, es decir, una respuesta falso y verdadero. Por tanto, lo arrastrarás al icono que indica Color, donde automáticamente tomará un color determinado dependiendo del valor. Se pueden cambiar los colores haciendo clic en Color nuevamente:



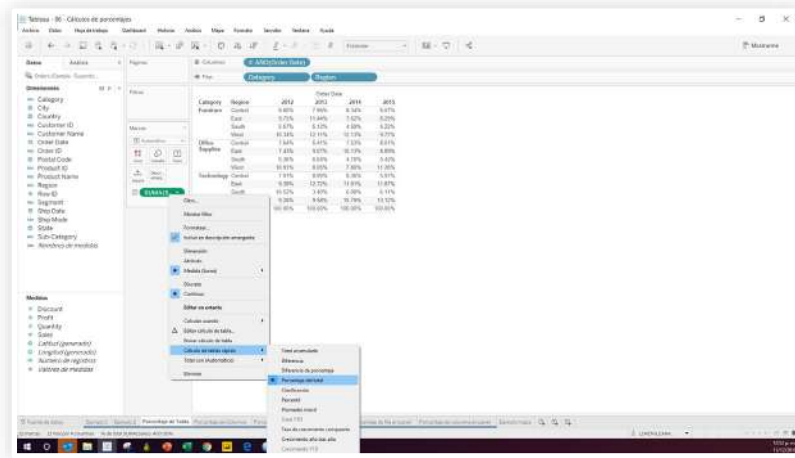
Cálculo de porcentaje

Si deseas conocer el porcentaje que representan las ventas generadas durante varios años deberás arrastrar los siguientes datos al libro de trabajo:

- Ventas
- Categoría
- Región
- Años

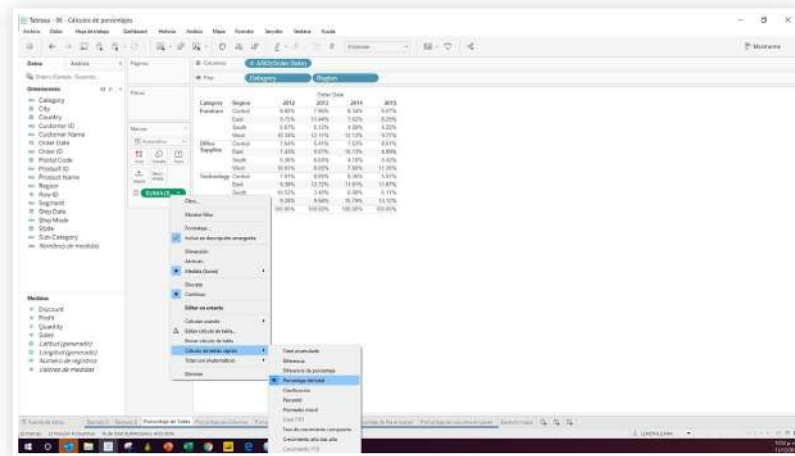


Aquí verás que aparece el monto de ventas por año registrado, categoría y región. Sin embargo, si lo que deseas es ver cuánto representa en porcentaje lo que tendrías que realizar es lo siguiente:



1. Hacer clic derecho en el concepto de ventas.
2. Seleccionar Cálculo de tablas rápidas/Porcentaje del total.

Nota: es importante mencionar que Tableau en primer nivel indica el porcentaje del total, pero si quisieras el porcentaje de Región/Categoría tendrías que realizar otro paso adicional.

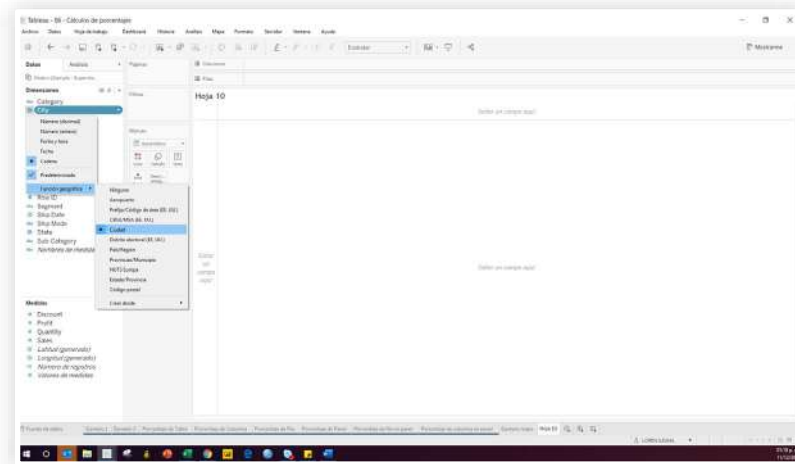


Generación de mapas

Se pueden realizar mapas en Tableau a través de dos formas:

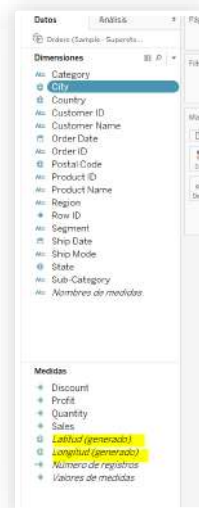
- Contando con las coordenadas de la ubicación.
- Colocando el estado o la ciudad.

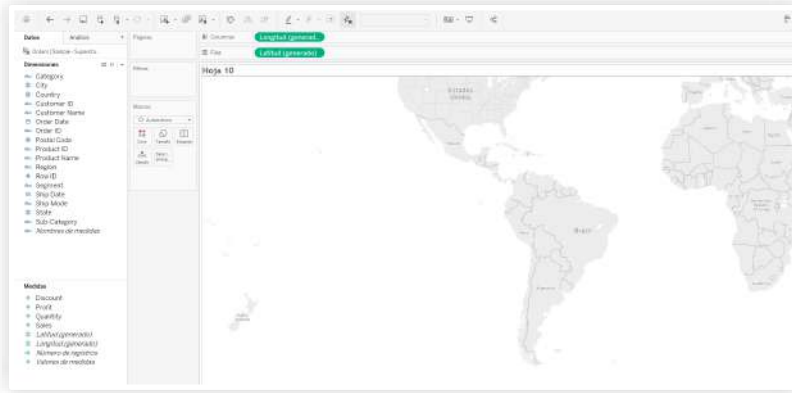
Es importante mencionar que los mapas no son totalmente precisos, pero te situarán en lo que está aproximado al lugar:



En este ejemplo la ciudad está en dimensiones y para generar que dicha variable se convierta en mapa tendrás que cambiarla a función geográfica, así como aparece en la imagen anterior, en donde puedes observar que se ha seleccionado Ciudad.

Una vez realizado ese paso, Tableau genera dos variables de medida llamadas Longitud y Latitud, mismas que tendrás que colocar en el libro para que pueda generarse el mapa de forma automática (podrás hacer doble clic en cada una).

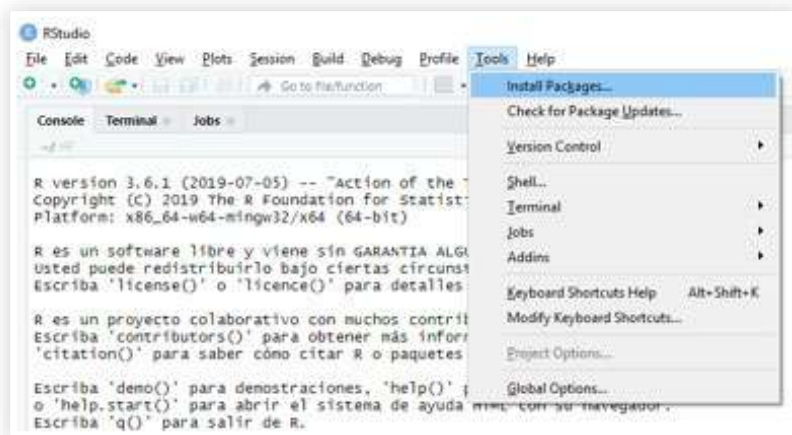




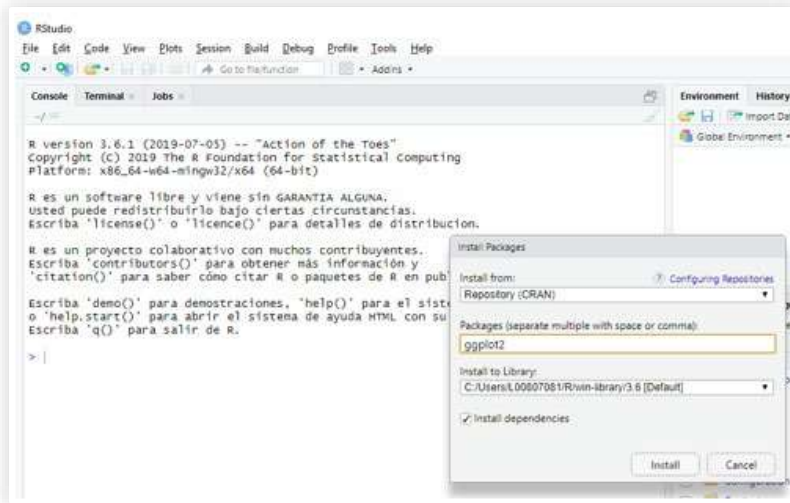
De esta forma se crea el mapa en Tableau.

Visualización con ggplot2

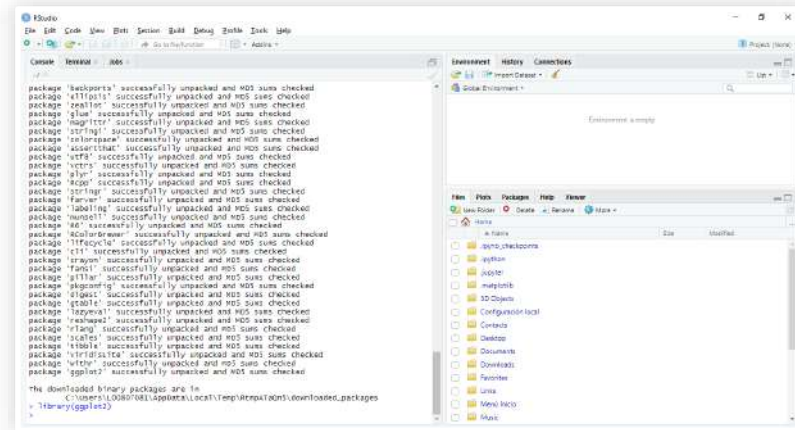
Para este laboratorio es necesario instalar en RStudio el paquete ggplot2, por lo que debes direccionarte a la sección de Tools y después hacer clic en Install Packages:



Escribe en el espacio de Packages: ggplot2 y haz clic en Install:



Al finalizar la instalación (para poder utilizar ese paquete o librería), deberás poner en este la consola library(ggplot2):



Otros paquetes que necesitarás instalar son los siguientes: *knitr*, *dplyr* y *nycflights*.

Comienza usando el Data Viewer con la función View, la cual te permitirá ver la información de un grupo de datos en forma tabular. Para ello, tomarás un set de datos precargados con los paquetes llamado *Presidential*, el cual contiene información de los últimos presidentes de los Estados Unidos de América.

Escribe lo siguiente:

```
> view(presidential)
> |
```

Para obtener la siguiente representación de los datos:

	name	start	end	party
1	Eisenhower	1953-01-20	1961-01-20	Republican
2	Kennedy	1961-01-20	1963-11-22	Democratic
3	Johnson	1963-11-22	1969-01-20	Democratic
4	Nixon	1969-01-20	1974-08-09	Republican
5	Ford	1974-08-09	1977-01-20	Republican
6	Carter	1977-01-20	1981-01-20	Democratic
7	Reagan	1981-01-20	1989-01-20	Republican
8	Bush	1989-01-20	1993-01-20	Republican
9	Clinton	1993-01-20	2001-01-20	Democratic
10	Bush	2001-01-20	2009-01-20	Republican
11	Obama	2009-01-20	2017-01-20	Democratic

Ahora harás uso de un *dataframe* llamado *txhousing*, el cual viene incluido con el *package* de *ggplot* para que realices tu primera gráfica.

Este *dataframe* incluye la información del mercado inmobiliario en el estado de Texas con 8,602 observaciones y 9 variables. Para darle un vistazo a este grupo de datos usarás la función *glimpse*:

```
> glimpse(txhousing)
Observations: 8,602
Variables: 9
 $ city      <chr> "Abilene", "Abilene", "Abilene", "Abilene", "Abilene", "Abilene..."
 $ year      <int> 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 200...
 $ month     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8, ...
 $ sales     <dbl> 72, 98, 130, 98, 141, 156, 152, 131, 104, 101, 100, 92, 75, 112...
 $ volume    <dbl> 5380000, 6505000, 9285000, 9730000, 10590000, 13910000, 1263500...
 $ median    <dbl> 71400, 58700, 58100, 68600, 67300, 66900, 73500, 75000, 64500, ...
 $ listings  <dbl> 701, 746, 784, 785, 794, 780, 742, 765, 771, 764, 721, 658, 779...
 $ inventory <dbl> 6.3, 6.6, 6.8, 6.9, 6.8, 6.6, 6.2, 6.4, 6.5, 6.6, 6.2, 5.7, 6.8...
 $ date      <dbl> 2000.000, 2000.083, 2000.167, 2000.250, 2000.333, 2000.417, 200...
```

De todo este grupo de datos solo nos interesa graficar cómo se comportaron las ventas en el mercado inmobiliario de la ciudad de Arlington, Texas en el año 2008 (cuando tuvo lugar la crisis inmobiliaria en los Estados Unidos). Para eso filtrarás los datos por ciudad y año y lo asignarás a un nuevo vector llamado *Arlington_hsales2008*:

```
> Arlington_hsales2008 <- txhousing %>%
+ filter(city == "Arlington" & year == 2008)
```

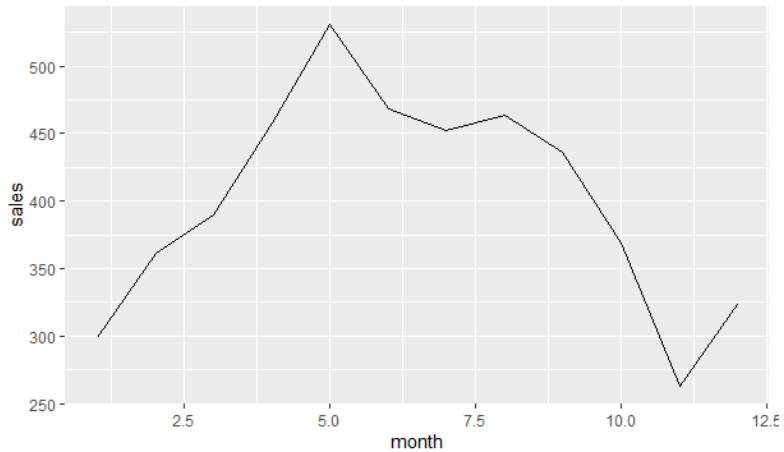
Después le darás un vistazo para observar cómo quedó el nuevo *dataframe*:

	city	year	month	sales	volume	median	listings	inventory	date
1	Arlington	2008	1	299	41280330	128000	2652	5.5	2008.000
2	Arlington	2008	2	361	52437477	127000	2606	5.5	2008.083
3	Arlington	2008	3	390	60168088	131100	2660	5.8	2008.167
4	Arlington	2008	4	458	67512370	130400	2676	5.9	2008.250
5	Arlington	2008	5	531	81648934	131700	2578	5.7	2008.333
6	Arlington	2008	6	468	71016280	131400	2587	5.9	2008.417
7	Arlington	2008	7	452	69076095	136600	2546	6.0	2008.500
8	Arlington	2008	8	464	67702876	130700	2415	5.8	2008.583
9	Arlington	2008	9	436	63544449	127900	2241	5.3	2008.667
10	Arlington	2008	10	369	51740641	126300	2203	5.3	2008.750
11	Arlington	2008	11	263	37364981	123200	2139	5.3	2008.833
12	Arlington	2008	12	323	46549795	130200	1957	4.9	2008.917

En este momento, te encuentras en posibilidad de graficarlo. Para esto, usarás la función de *ggplot* y *geom_line()* para conectar las observaciones en línea sobre el eje x:

```
> ggplot(data = Arlington_hsales2008, mapping = aes(x = month, y = sales)) + geom_line()
```

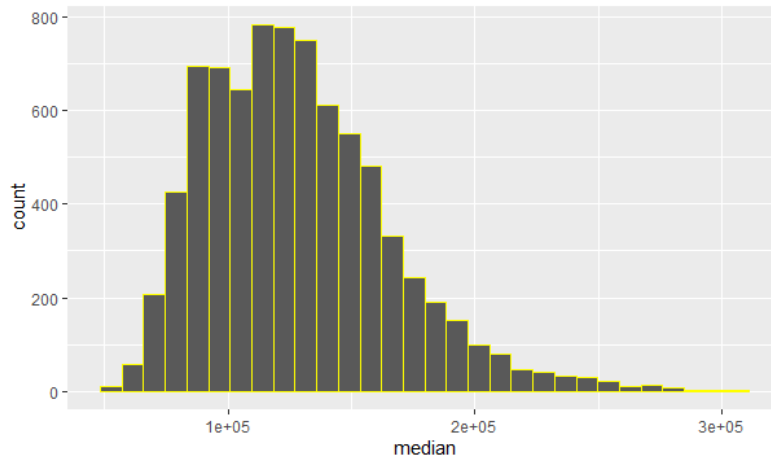
Y obtendrás la siguiente gráfica:



Otro tipo de gráfica que puedes realizar es el histograma. Para ello, tomarás el *dataframe* completo de *txhousing* y graficarás en un histograma la mediana, para lo cual bastará con escribir el siguiente código:

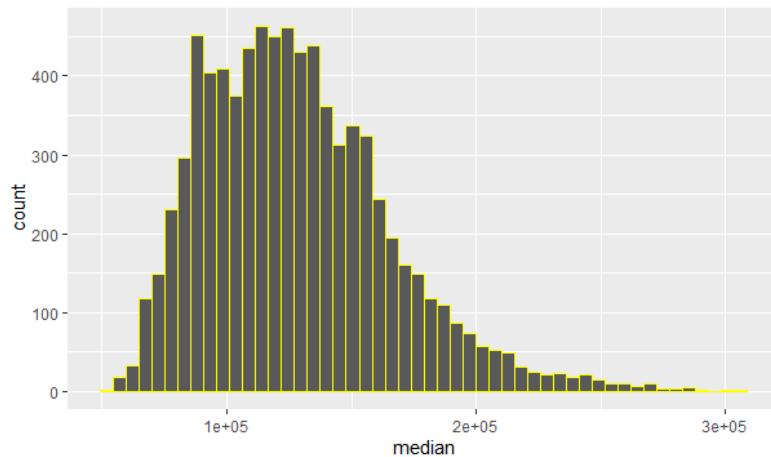
```
> ggplot(data = txhousing, mapping = aes(x = median)) + geom_histogram(color = "yellow")
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
warning message:
Removed 616 rows containing non-finite values (stat_bin).
>
```

La función eliminó 616 filas que contenían valores no finitos, además de indicar que se utilizaron 30 *bins* (cajones) por *default*, obteniendo el siguiente gráfico:



Es posible modificar el número de *bins* dentro de la función `geom_histogram()`. Por ejemplo, si desearas que en lugar de 30 sean 50 *bins*, quedaría de la siguiente manera:

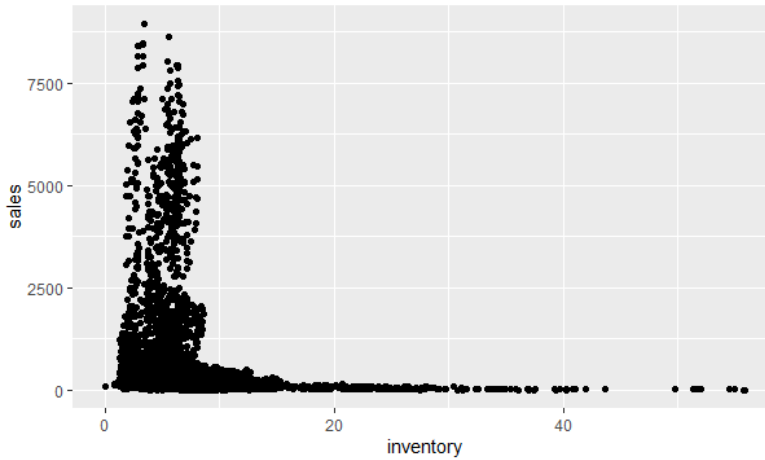
```
> ggplot(data = txhousing, mapping = aes(x = median)) + geom_histogram(bins = 50, color = "yellow")
warning message:
Removed 616 rows containing non-finite values (stat_bin).
> |
```



En el siguiente ejemplo, realizarás una gráfica de puntos:

```

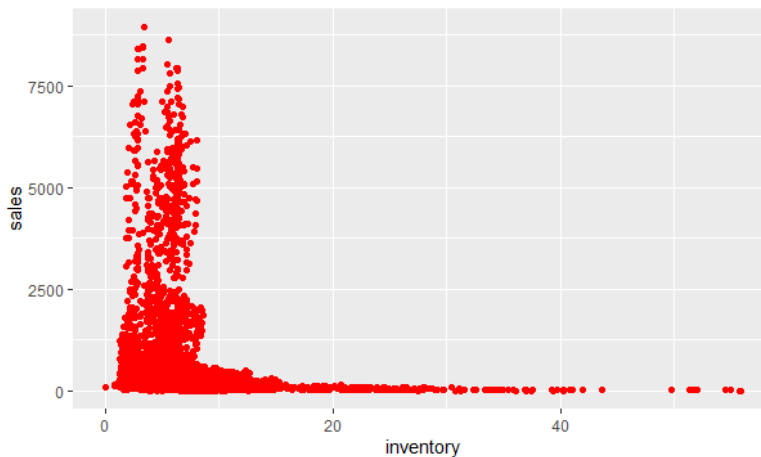
> ggplot(data = txhousing, mapping = aes(x = inventory, y = sales))+geom_point()
Warning message:
Removed 1468 rows containing missing values (geom_point).
>
    
```



También podrás cambiar el color:

```

> ggplot(data = txhousing, mapping = aes(x = inventory, y = sales))+geom_point(color = "red")
Warning message:
Removed 1468 rows containing missing values (geom_point).
>
    
```



Para hacer gráficas de caja puedes utilizar la función `geom_boxplot()`. En el siguiente ejemplo, tomarás el grupo de datos preinstalados llamados `Airquality`, el cual cuenta con los datos de la calidad del aire en Nueva York de mayo a septiembre. Usarás la función `View()` para observar la información que contiene:

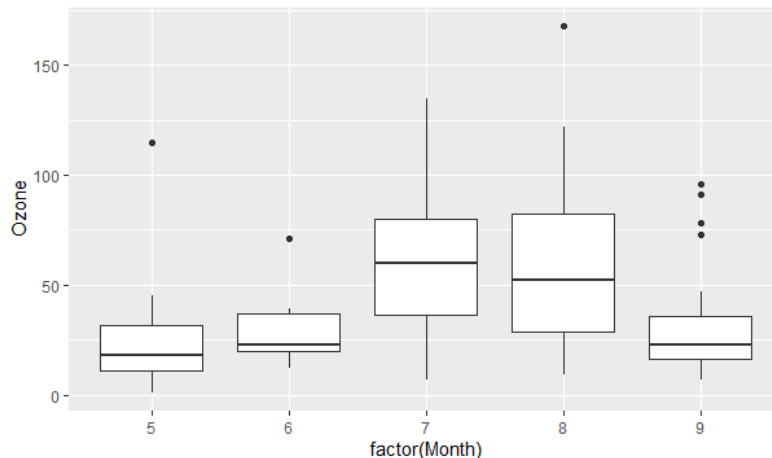
The top screenshot shows a data viewer window displaying the first 11 rows of the 'airquality' dataset. The columns are Ozone, Solar.R, Wind, Temp, Month, and Day. The data shows values for May (Month = 5) and June (Month = 6). The bottom screenshot shows a terminal window with the command `view(airquality)` entered.

	Ozone	Solar.R	Wind	Temp	Month	Day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10
11	7	NA	6.9	74	5	11

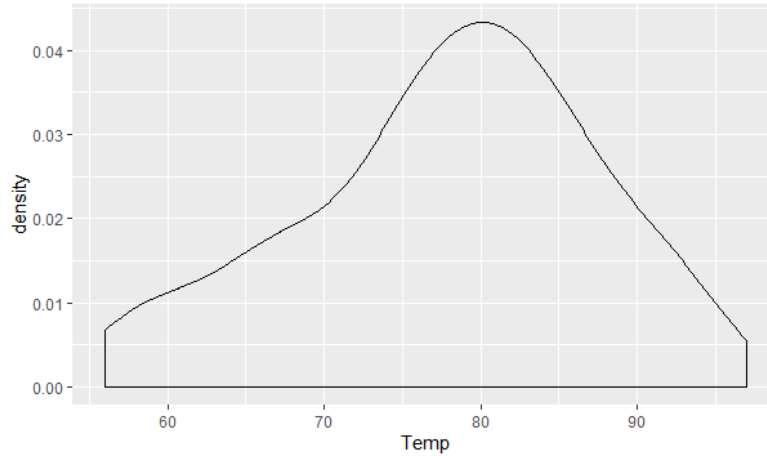
Selecciona los datos de Ozone y grafica contra el mes en el que ocurrieron:

```
> ggplot(data = airquality, mapping = aes(x = factor(Month), y = ozone)) + geom_boxplot()
Warning message:
Removed 37 rows containing non-finite values (stat_boxplot).
> |
```

Se obtiene la siguiente gráfica:



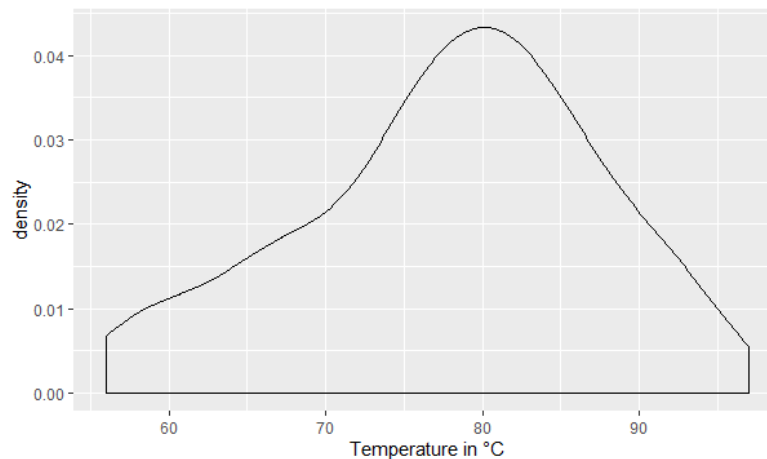
Finalmente, realizarás una gráfica de curva de densidad, la cual puedes realizar con la función `geom_density()`. En el siguiente ejemplo, graficarás con el mismo grupo de datos Airquality, pero también graficarás la densidad de la variable de temperatura:



También es posible establecer las etiquetas que deseas que aparezcan en los ejes. En este ejemplo, pondrás una etiqueta al eje de las x:

```

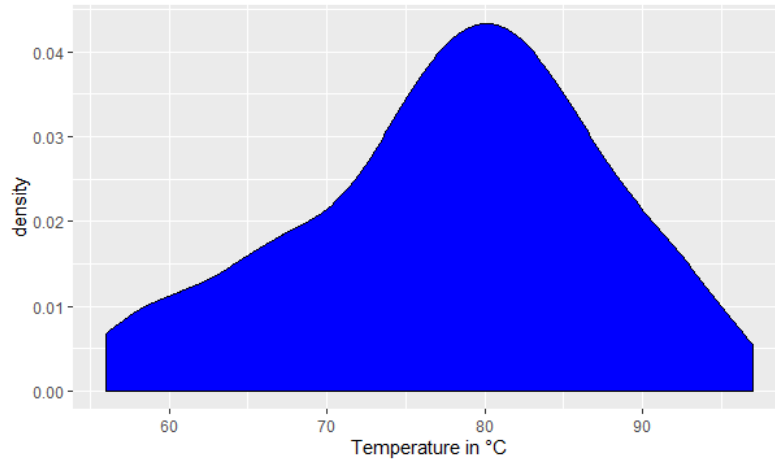
> ggplot(data = airquality, mapping= aes(x = Temp)) + geom_density() + scale_x_continuous(
  name = "Temperature in °C")
> |
    
```



Asimismo, podrás hacer que la gráfica tenga un color de relleno:

```

> ggplot(data = airquality, mapping= aes(x = Temp)) + geom_density(fill = "blue") + sca
  le_x_continuous(name = "Temperature in °C")
>
    
```



La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECNOLÓGICA DE MILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECNOLÓGICA DE MILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECNOLÓGICA DE MILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educativo y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.