

Ejercicio Propinas

1. Calcula percentiles de la variable propina junto con mínimo y máximo.

Se uso la funcion quantile y se paso la columna propina junto con una secuencia para mostrar del 0 al 100%. Se asigno en una variable minimo el primer valor del percentil y en otra variable llamado maximo el ultimo valor que son el mínimo y máximo respectivamente.

```
percentiles <- quantile(propinas$propina, probs = seq(0, 1, 0.05))
percentiles
```

```
minimo <- percentiles[1]
minimo
```

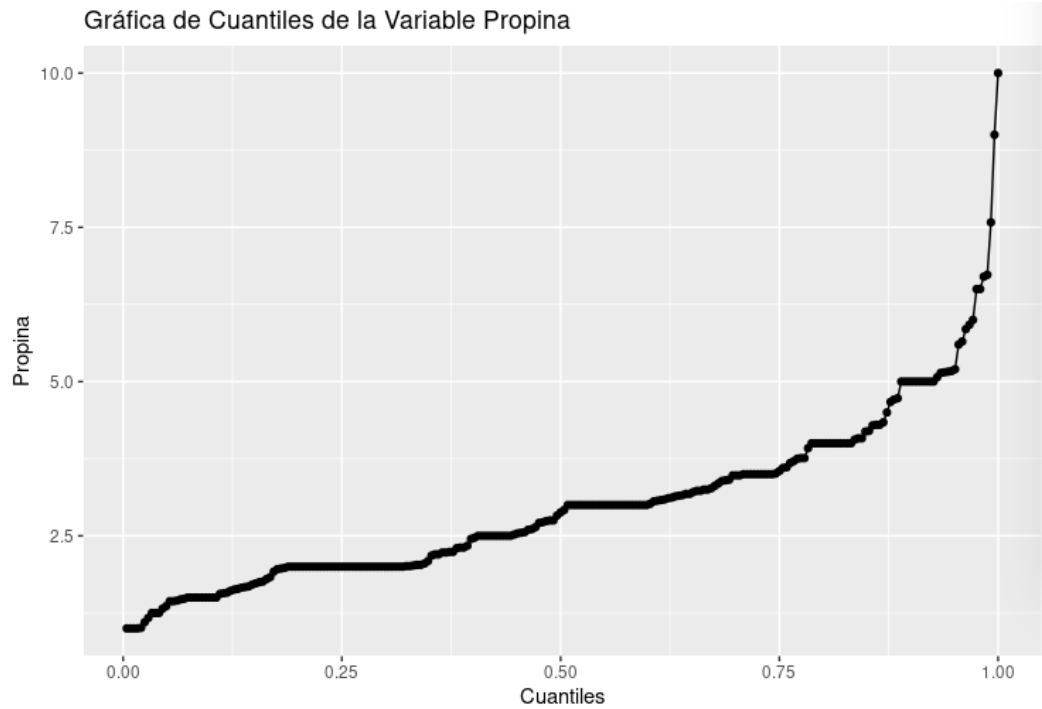
```
maximo <- percentiles[21]
maximo
```

```
> percentiles
  0%    5%   10%   15%   20%   25%   30%   35%   40%   45%   50%   55%   60%   65%
1.0000 1.4400 1.5000 1.7390 2.0000 2.0000 2.0000 2.1810 2.4760 2.5435 2.9000 3.0000 3.0160 3.2085
 70%   75%   80%   85%   90%   95%  100%
3.4800 3.5625 4.0000 4.1955 5.0000 5.1955 10.0000
> minimo <- percentiles[1]
> minimo
0%
1
> maximo <- percentiles[21]
> maximo
100%
10
```

2. Haz una gráfica de cuantiles de la variable propina

Se uso la librería ggplot así como geom_point para graficar los puntos y geom_line para unir esos puntos en la gráfica.

```
ggplot(propinas, aes(x = f, y = propina)) +
  geom_point() +
  geom_line() +
  labs(title = "Gráfica de Cuantiles de la Variable Propina",
       x = "Cuantiles",
       y = "Propina")
```

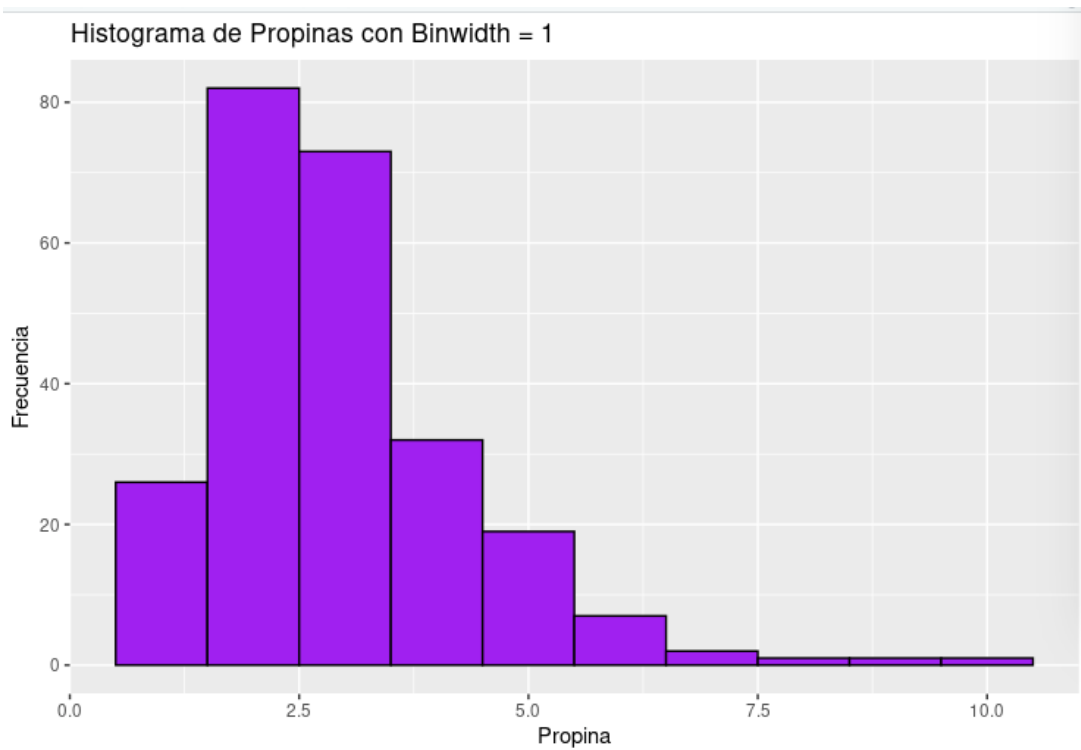
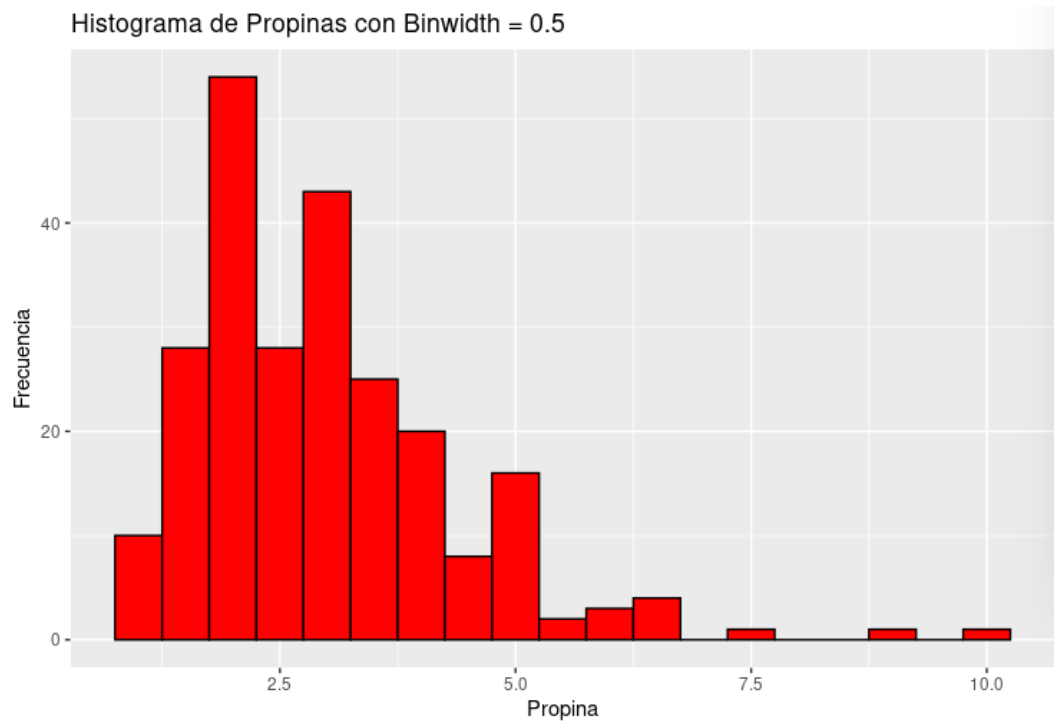


3. Haz un histograma de la variable propinas. Ajusta distintos anchos de banda (usa bins o binwidth en ggplot)

Usando ggplot se usó geom_histogram para la creación del histograma. Con el argumento binwidth podemos modificar el ancho de las barras y con fill podemos seleccionar el color de las barras del histograma.

```
ggplot(propinas, aes(x = propina)) +
  geom_histogram(binwidth = 0.5, fill = "red", color = "black") + # ancho de 0.5
  labs(title = "Histograma de Propinas con Binwidth = 0.5",
        x = "Propina",
        y = "Frecuencia")

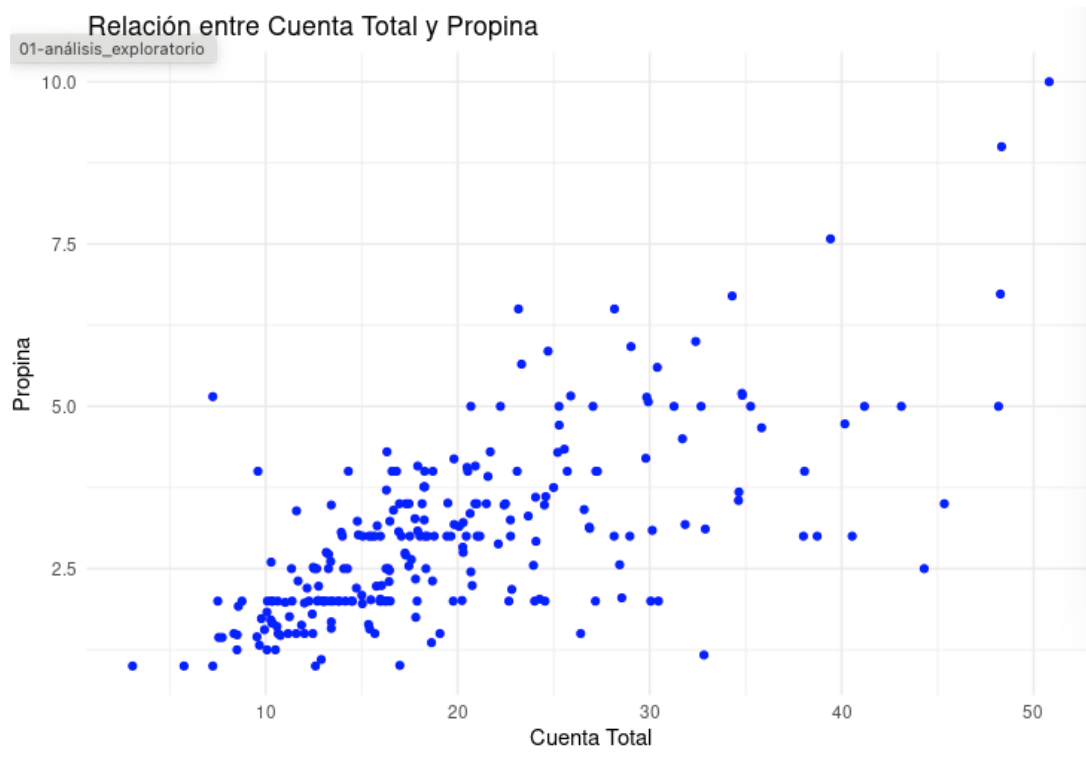
## Ancho de banda de 1
ggplot(propinas, aes(x = propina)) +
  geom_histogram(binwidth = 1, fill = "purple", color = "black") + # ancho de 1
  labs(title = "Histograma de Propinas con Binwidth = 1",
        x = "Propina",
        y = "Frecuencia")
```



4. Haz una gráfica de cuenta total contra propina

Para esto se usó `geom_point` para la creación de un gráfico de dispersión donde los valores X reflejan la cuenta total y los valores de Y reflejan los valores de la propina. Se puede observar que hay una relación entre ellos.

```
ggplot(propinas, aes(x = cuenta_total, y = propina)) +
  geom_point(color = "blue") + # Añade puntos de dispersión
  labs(title = "Relación entre Cuenta Total y Propina",
        x = "Cuenta Total",
        y = "Propina") +
  theme_minimal()
```



5. Calcula propina en porcentaje de la cuenta total. Calcula algunos cuantiles de propina en porcentaje

En este caso se ocupó la función `mutate` que agrega una columna al dataframe. El nombre de la columna a agregar es `pct_propina` y su valor es solo una fórmula para obtener el porcentaje de propina en relación con la cuenta total. Por último, se generaron los cuantiles de esta columna.

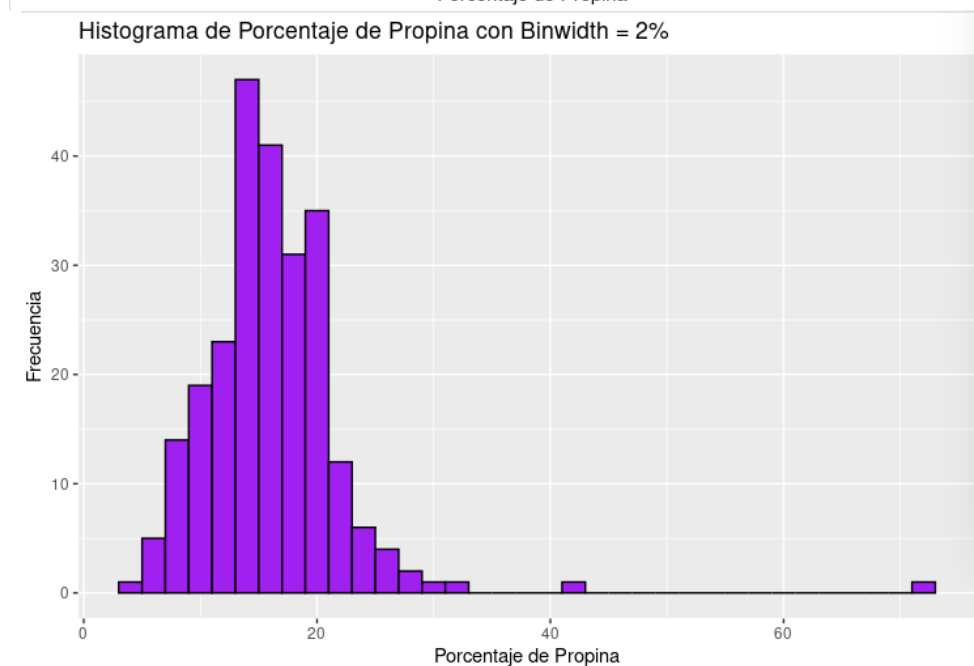
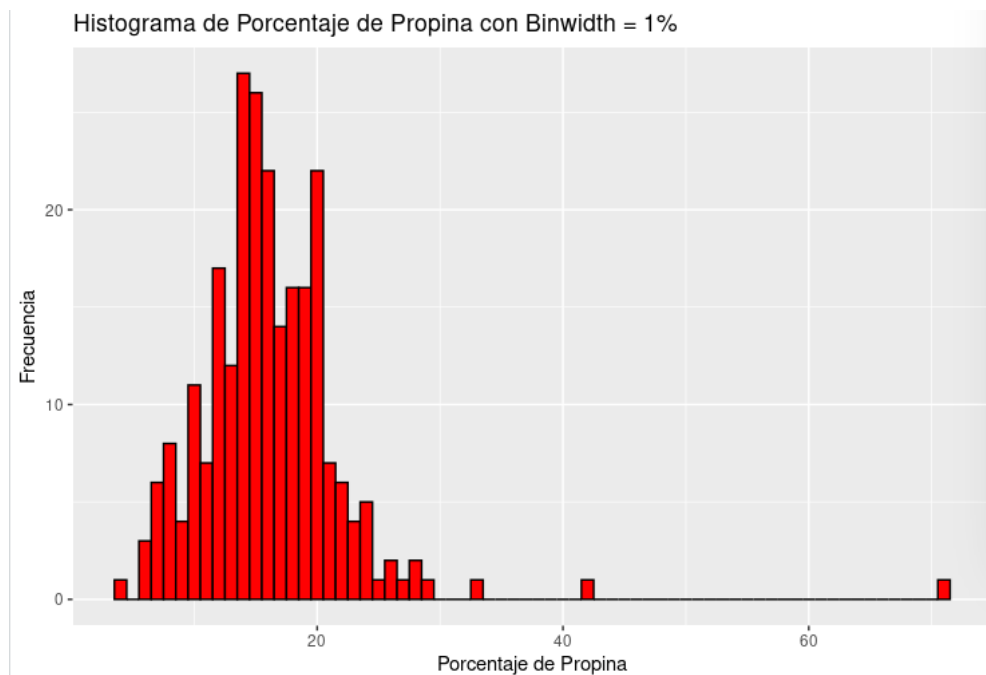
```
propinas <- propinas |>
  mutate(pct_propina = (propina / cuenta_total) * 100 )

quantiles_porcentaje_propina <- quantile(propinas$pct_propina, probs = seq(0, 1, 0.10))
quantiles_porcentaje_propina
```

```
> quantiles_porcentaje_propina
      0%      10%      20%      30%      40%      50%      60%      70%      80%      90%     100%
3.563814 9.855020 12.127636 13.723790 14.664191 15.476977 16.657075 18.194657 19.707593 21.726271 71.034483
```

6. Haz un histograma de la propina en porcentaje. Prueba con distintos anchos de banda.

```
ggplot(propinas, aes(x = pct_propina)) +  
  geom_histogram(binwidth = 1, fill = "red", color = "black") + # ancho de 1%  
  labs(title = "Histograma de Porcentaje de Propina con Binwidth = 1%",  
        x = "Porcentaje de Propina",  
        y = "Frecuencia")  
  
## Histograma con un ancho de banda de 2  
ggplot(propinas, aes(x = pct_propina)) +  
  geom_histogram(binwidth = 2, fill = "purple", color = "black") + # ancho de 2%  
  labs(title = "Histograma de Porcentaje de Propina con Binwidth = 2%",  
        x = "Porcentaje de Propina",  
        y = "Frecuencia")
```



7. Describe la distribución de propina en pct. ¿Hay datos atípicos?

La distribución del porcentaje de propina es sesgada a la derecha. La mayoría de los valores de propina se concentran entre aproximadamente el 10% y el 20% lo cual es el porcentaje de propina socialmente aceptado. Alrededor del 30%, 40%, y un valor cercano al 60% hay propinas significativamente más altas que el resto. Estos datos se consideran atípicos ya que estos son significativamente más altos que la mayoría de las propinas y son menos comunes.

8. Filtra los casos con porcentaje de propina muy altos. ¿Qué tipos de cuentas son? ¿Son cuentas grandes o chicas?

Usamos filter en este caso para seleccionar porcentajes de propina mayores a 30, los cuales son inusualmente grandes. Se encontró que las propinas inusualmente altas en porcentaje están asociadas con cuentas pequeñas.

```
propinas_altas <- propinas |>
  filter(pct_propina > 30)
propinas_altas

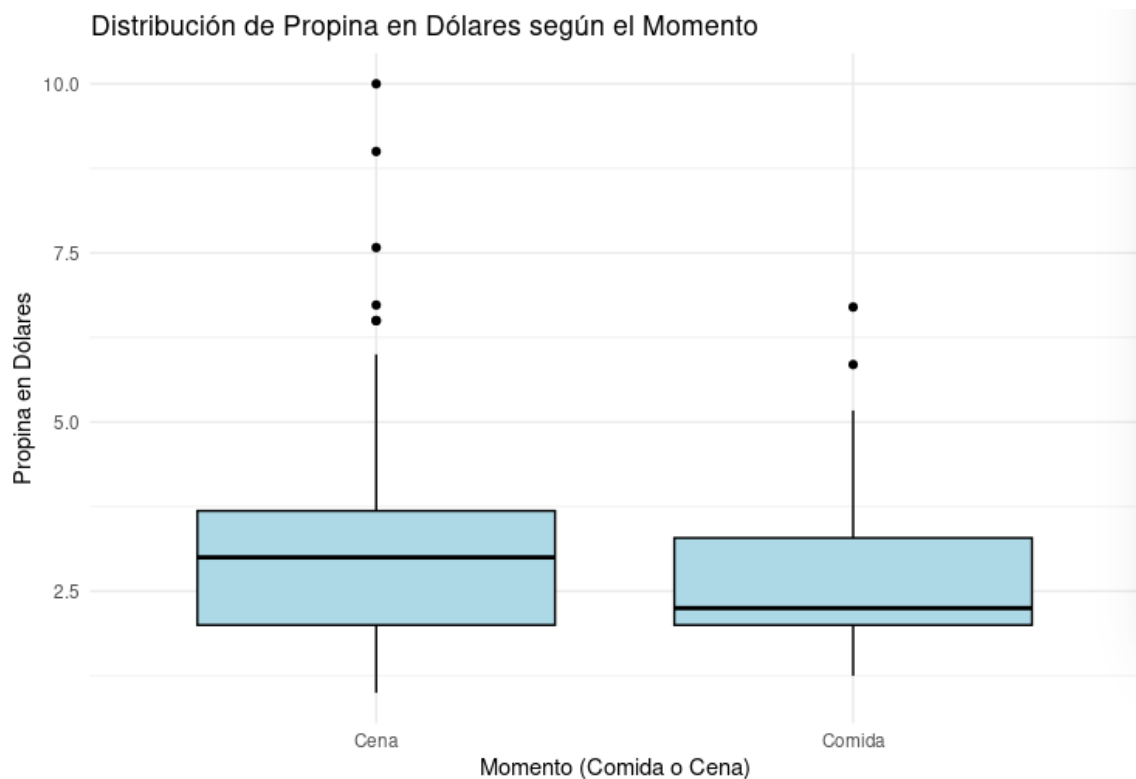
# Ver un resumen
summary(propinas_altas$cuenta_total)
|
propinas_altas |> select(cuenta_total, propina, pct_propina)

> summary(propinas_altas$cuenta_total)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 3.070  5.160   7.250  6.640  8.425   9.600
> # Inspeccionar las cuentas con propina alta
> propinas_altas |> select(cuenta_total, propina, pct_propina)
# A tibble: 3 x 3
  cuenta_total propina pct_propina
    <dbl>      <dbl>      <dbl>
1     3.07         1      32.6
2     7.25         5      71.0
3     9.6          4      41.7
```

9. Haz un diagrama de caja y brazos para propina en dólares dependiendo del momento (comida o cena). ¿Cuál parece más grande? ¿Por qué? Haz otras gráficas si es necesario.

Observando el diagrama de caja nos damos cuenta de que la mediana de la propina durante la cena está por encima de la mediana de la propina durante la comida. Además, la distribución de la propina durante la cena tiene una mayor dispersión hacia valores más altos. Esto hace sentido pues durante la cena es cuando mayor consumo de alimentos y bebidas existe en restaurantes.

```
ggplot(propinas, aes(x = momento, y = propina)) +
  geom_boxplot(fill = "lightblue", color = "black") + # Crear el diagrama de caja y brazos
  labs(title = "Distribución de Propina en Dólares según el Momento",
        x = "Momento (Comida o Cena)",
        y = "Propina en Dólares") +
  theme_minimal()
```



Ejercicio Casas

1. Condición y calidad general vs precio x m2 haz una tabla de conteos de los valores de calidad general de construcción y terminados (calidad_gral)

Usamos group by para agrupar por la cantidad general y summarise para agregar columnas a la nueva tabla.

```
tabla_calidad_precio <- casas |>
  group_by(calidad_gral) |>
  summarise(n = n(),
            precio_m2_promedio = mean(precio_m2, na.rm = TRUE),
            precio_m2_mediana = median(precio_m2, na.rm = TRUE))
# Mostrar la tabla
```

```
> tabla_calidadPrecio
# A tibble: 10 × 4
  calidad_gral n precio_m2_promedio precio_m2_mediana
    <dbl> <int>          <dbl>          <dbl>
1         1     2           996.           996.
2         2     1           807.           807.
3         3    17           866.           906.
4         4    79          1150.          1121.
5         5   315          1234.          1287.
6         6   287          1245.          1241.
7         7   263          1336.          1345.
8         8   128          1544.          1535.
9         9    35          1907.          1854.
10        10    17          1734.          1883.
> |
```

2. Repite el anterior con número de coches que caben en el garaje. ¿Cuál es la relación? ¿Qué puedes malinterpretar de esta gráfica?

Usamos factor para crear un boxplot para cada categoría de número de coches y creamos el diagrama de caja. Se puede malinterpretar que una mayor capacidad en el garaje significa un mayor precio, pero esto puede estar más relacionado a el tamaño del lote pues un garaje más amplio conlleva a mas tamaño.

```
ggplot(casas, aes(x = factor(num_coches), y = precio_m2)) +
  geom_boxplot(fill = "lightgreen", color = "black") +
  labs(title = "Precio por Metro Cuadrado según el Número de Coches en el Garaje",
       x = "Número de Coches en el Garaje",
       y = "Precio por m2") +
  theme_minimal()
```

