

Tarea y Solución

Pregunta 1

Explora la distribución de las variables `x`, `y`, `z` del dataset de `diamonds`. ¿Qué podemos inferir?

Busca un diamante (por internet por ejemplo) y decide qué dimensiones pueden ser aceptables para las medidas de longitud, altura y anchura de un diamante.

Solución

Los diamantes son piedras preciosas realmente pequeñas de unos pocos milímetros de tamaño. Por tanto, el mismo filtro que hemos hecho para la `y` durante las lecciones nos sirve para combinarlo en `x` y en `z`.

Al final, los que deben revisarse son los diamantes con `x = 0`, `y < 2`, `y > 30`, `z = 0` y `z > 30`.

Pregunta 2

Explora la distribución del precio (`price`) del dataset de `diamonds`. ¿Hay algo que te llame la atención o resulte un poco extraño?

Recuerda hacer uso del parámetro `binwidth` para probar un rango dispar de valores hasta ver algo que te llame la atención.

Solución

De forma extraña no hay ningún diamante entre 1450 y 1550 dólares. Es como si ese precio estuviera gafado...

Pregunta 3

¿Cuántos diamantes hay de 0.99 kilates? ¿Y de exactamente 1 kilate?

¿A qué puede ser debida esta diferencia?

Solución

Apenas unos 20 vs más de 1500. Probablemente el redondeo cerca de 1 obligue a que muchos hayan sido marcados como unidad entera en lugar de mantener ese valor decimal aproximado.

Pregunta 4

Compara y contrasta el uso de las funciones `coord_cartesian()` frente `xlim()` y `ylim()` para hacer zoom en un histograma.

¿Qué ocurre si dejamos el parámetro `binwidth` sin configurar?

¿Qué ocurre si hacemos zoom y solamente se ve media barra?

Solución

Intenta elegir por defecto `nbins = 30`, aunque a veces no tenga sentido...

Pregunta 5

- ¿Qué ocurre cuando hay NAs en un histograma?
- ¿Qué ocurre cuando hay NAs en un diagrama de barras?
- ¿Qué diferencias observas?

Solución

Ambos muestran warnings diferentes:

- Removed 1610 rows containing non-finite values (stat_count).
- Removed 9 rows containing non-finite values (stat_bin).

Pregunta 6

¿Qué hace la opción `na.rm = TRUE` en las funciones `mean()` y `sum()`?

Solución

Elimina los NA de la suma y del promedio, pero no del dataset sobre el cual se opera.