

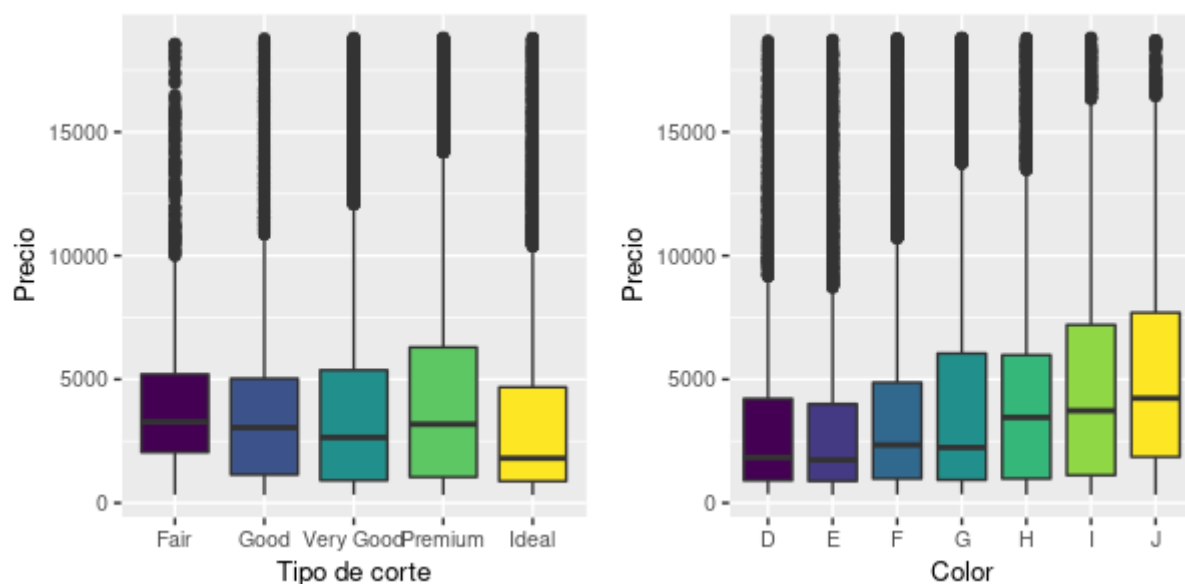
Introducción a la Ciencia de Datos - 2C 2022

Guía para la entrega N° 6

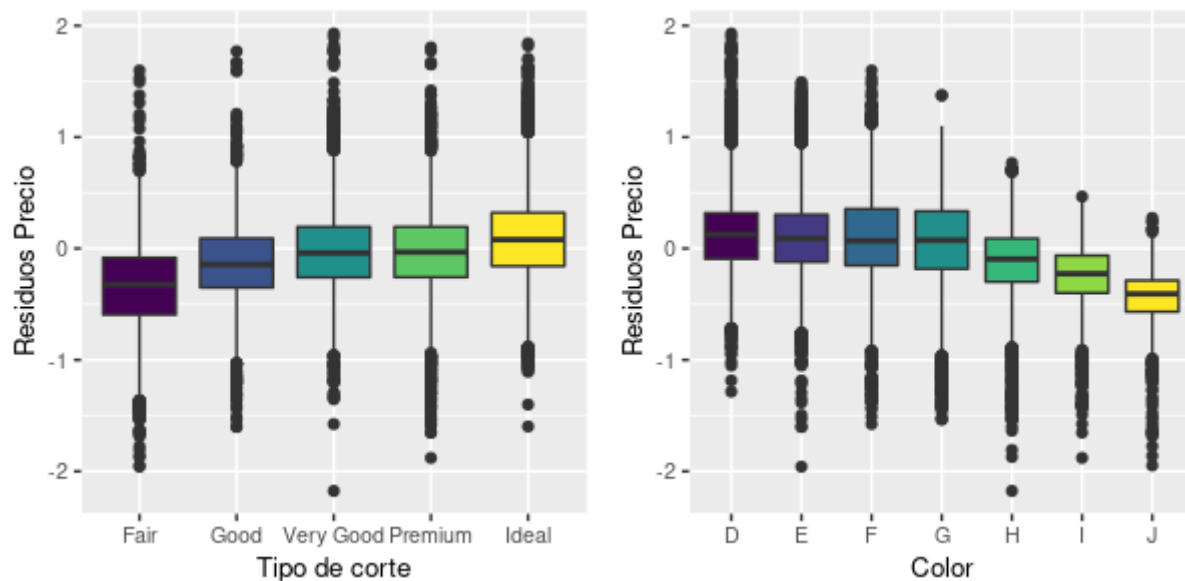
Análisis con modelos lineales, controlando variables

La última tarea de la guía presencial de esta semana busca familiarizarlos con el efecto que puede tener sobre la relación aparente entre dos variables la presencia de una tercera variable no observada.

En clase vimos el caso de los diamantes, que parecen tener precio menor a medida que mejora la calidad de su corte, o su color:



Sin embargo, cuando controlamos por el efecto del peso (carat) de los diamantes, que tenía una clara relación con el precio, la relación entre los residuos del ajuste y las variables categóricas sigue una relación según lo esperado: los mejores diamantes son los más caros.



A la variable que entorpece la relación o que no es la principal, se la llama *confounder*. En este caso, el peso de los diamantes es un *confounder*.

En acción

Para esta entrega, tienen que trabajar con el dataset de la entrega grupal (pero de manera individual). El objetivo es que hagan un análisis lineal entre dos variables de su dataset, y reporten la amplitud de la relación (es decir, el valor del estimador del coeficiente que acompaña a la variable independiente en el modelo).

Según el dataset, la relación que nos interesa es:

- Vuelos. Retraso en la llegada como función del retraso en la salida.
- Árboles. Altura de los árboles en función del ancho de los árboles.
- Seguros. Costo del seguro en función de la edad

1. Hagan un modelo lineal simple entre las variables de su dataset. Pueden filtrar las unidades para obtener algo más específico y seguramente también tengan que transformar la variable independiente. Registren la amplitud del efecto, en la forma de los coeficientes que acompañan a la o las variables independientes.

2. Consideren detenidamente el problema y piensen en al menos una variable o combinación de variables que puede estar entorpeciendo la relación e impidiéndoles obtener la verdadera amplitud del efecto. Agreguen estas variables en un nuevo ajuste, y registren el cambio en la amplitud del efecto de las variables del punto anterior.

Subir la producción realizada al campus de la materia en la solapa de “Entrega N° 6” con un texto que explique lo que se hizo y que discuta las conclusiones a las que se llegó.