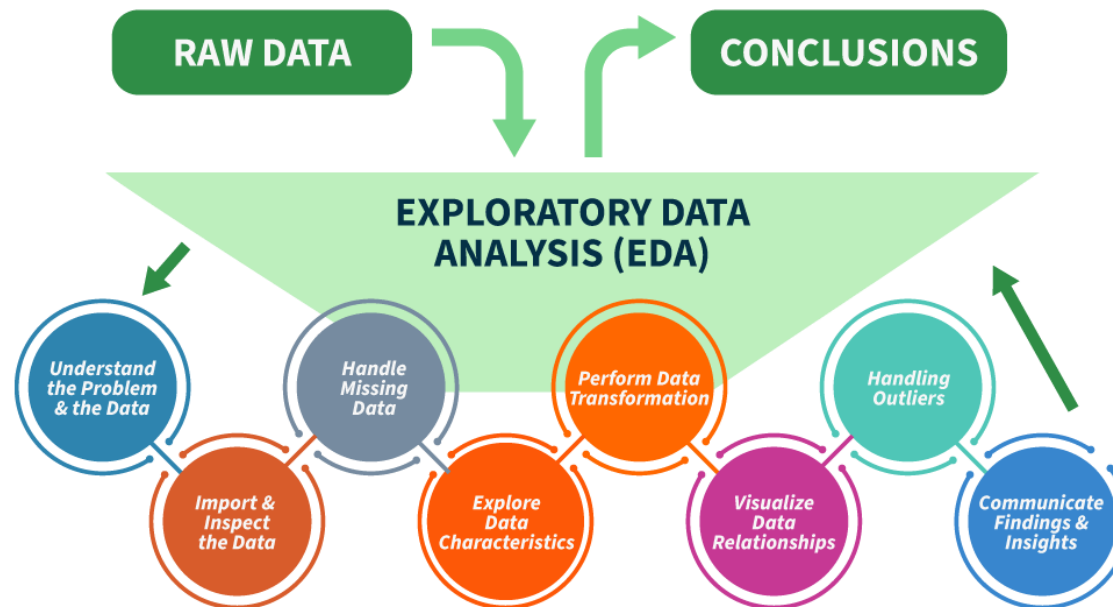


# Análisis Exploratorio de los Datos



Profesor: Pedro Saa ([pnsaa@uc.cl](mailto:pnsaa@uc.cl))

Año: 1-2025

# OBJETIVOS DE APRENDIZAJE

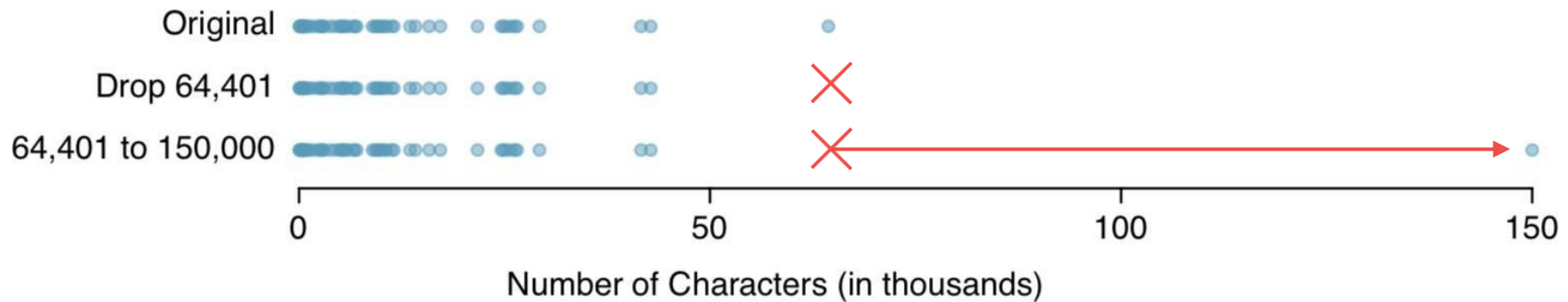
---

- ▶ **01:** Distinguir las ventajas y desventajas de los distintos tipos de visualizaciones numéricas y categóricas
  - ▶ **02:** Seleccionar el tipo de visualización más efectiva y clara según el tipo de dato
  - ▶ **03:** Comprender conceptos básicos de medidas de forma, tendencias central y de dispersión
- ▶ **04:** Explicar la importancia de la estadística robusta
  - ▶ **05:** Evaluar la importancia de la presencia de valores atípicos

# Estadísticos Robustos

---

Los estadísticos robustos son aquellas medidas que no son afectados significativamente por valores atípicos, por ej. mediana y IQR

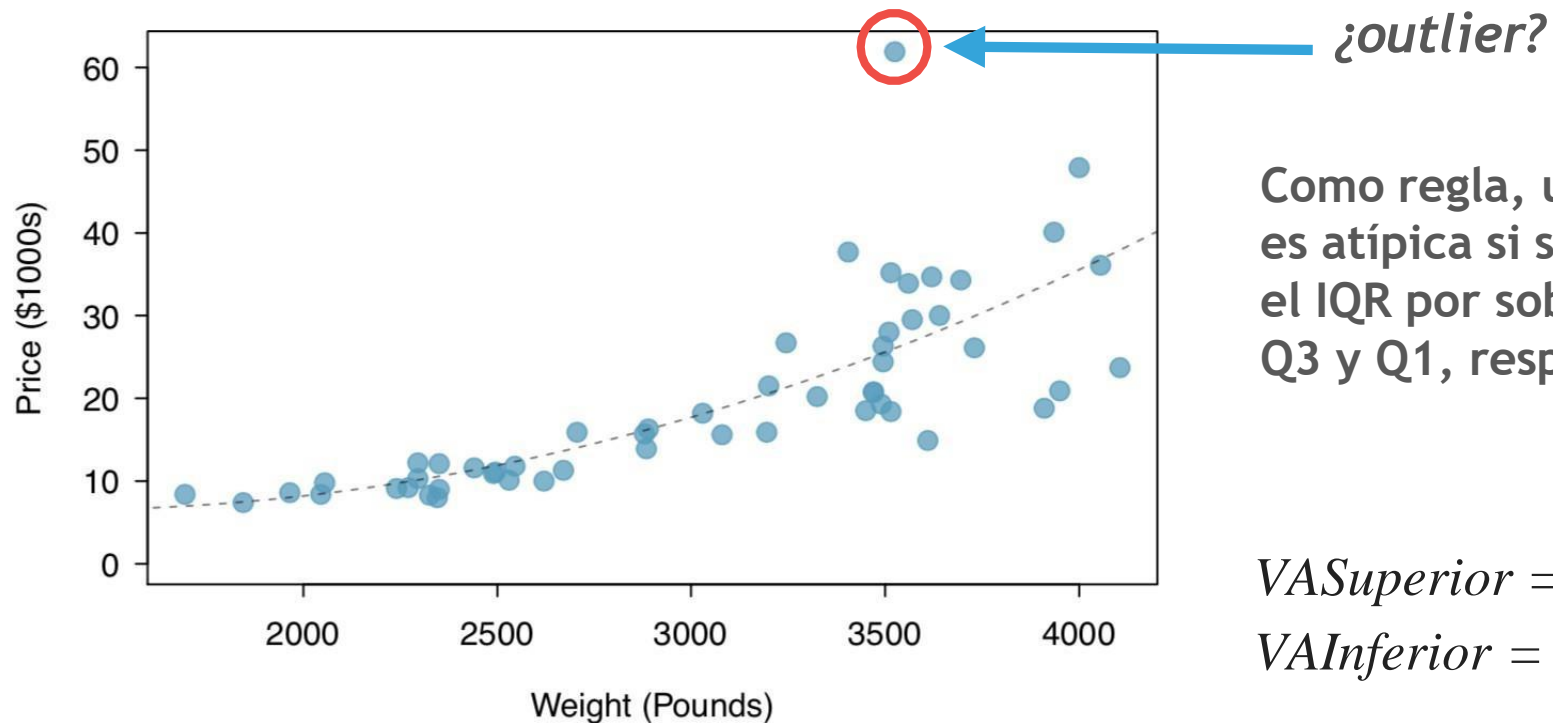


scenario	robust		not robust	
	median	IQR	$\bar{x}$	$s$
original num_char data	6,890	12,875	11,600	13,130
drop 64,401 observation	6,768	11,702	10,521	10,798
move 64,401 to 150,000	6,890	12,875	13,310	22,434

Table 1.28: A comparison of how the median, IQR, mean ( $\bar{x}$ ), and standard deviation ( $s$ ) change when extreme observations are present.

**Pero... ¿qué es un valor  
atípico o *outlier*?**

**Valores atípicos** o *outliers* son observaciones que aparecen ser extremas comparados con el resto de los datos



Como regla, una observación es atípica si supera 1.5 veces el IQR por sobre o inferior a Q3 y Q1, respectivamente

$$VASuperior = Q3 + 1.5IQR$$

$$VAInferior = Q1 - 1.5IQR$$

### ¿Qué hacemos con ellos?

¿Por qué es relevante revisarlos?

- Identificar fuertes asimetrías en los datos
- Identificar errores de ingreso de datos
- Identificar propiedades interesantes de los datos

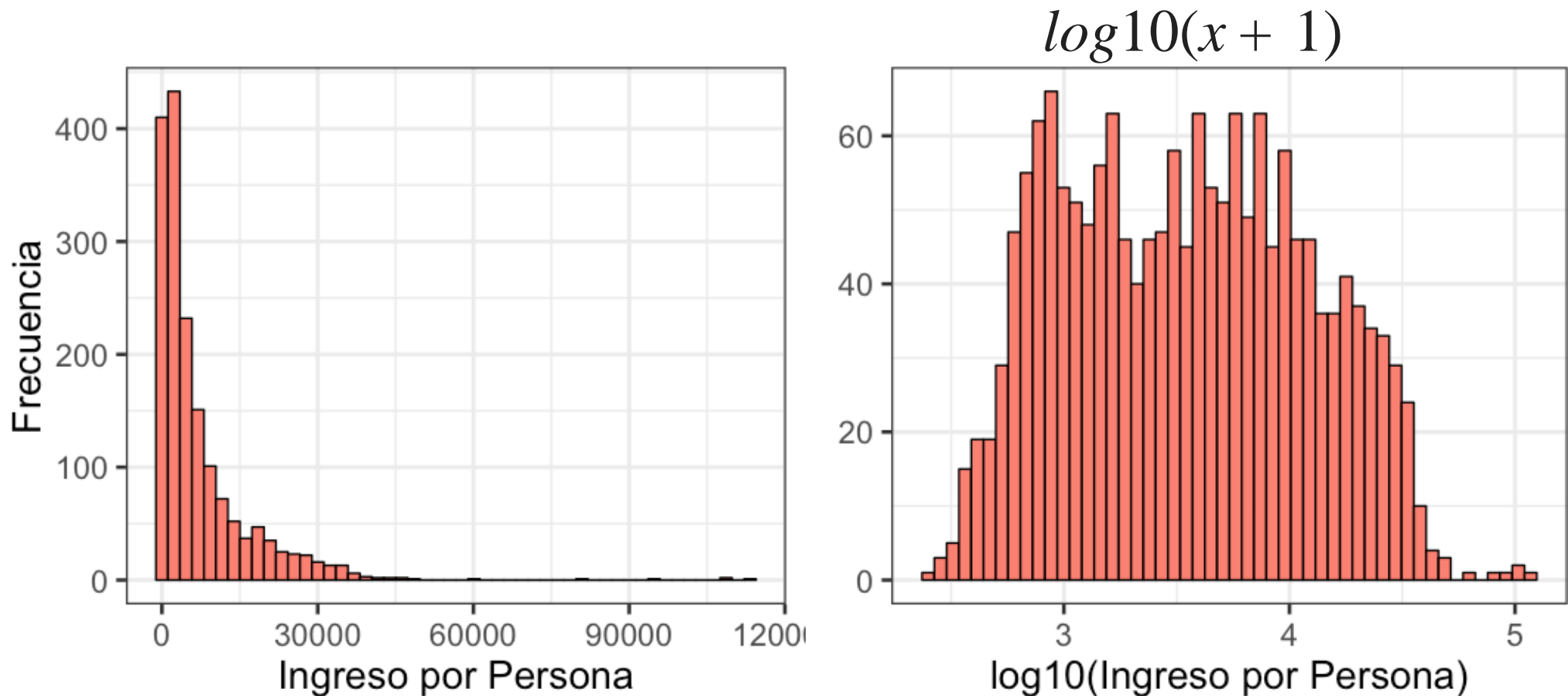
→ Evaluar una transformación (log10)

→ Evaluar replicar o eliminar el dato

→ Insight de investigación

**¿Por qué necesito  
transformar mis datos?**

En presencia de fuertes asimetrías es posible aplicar un **re-escalamiento** a los datos para **normalizarlos** y poder trabajar mejor con ellos, la transformación más ocupada es la del **logaritmo**



El problema con las transformaciones es **la interpretación de un valor normalizado**, ya que es necesario des-transformar para interpretar

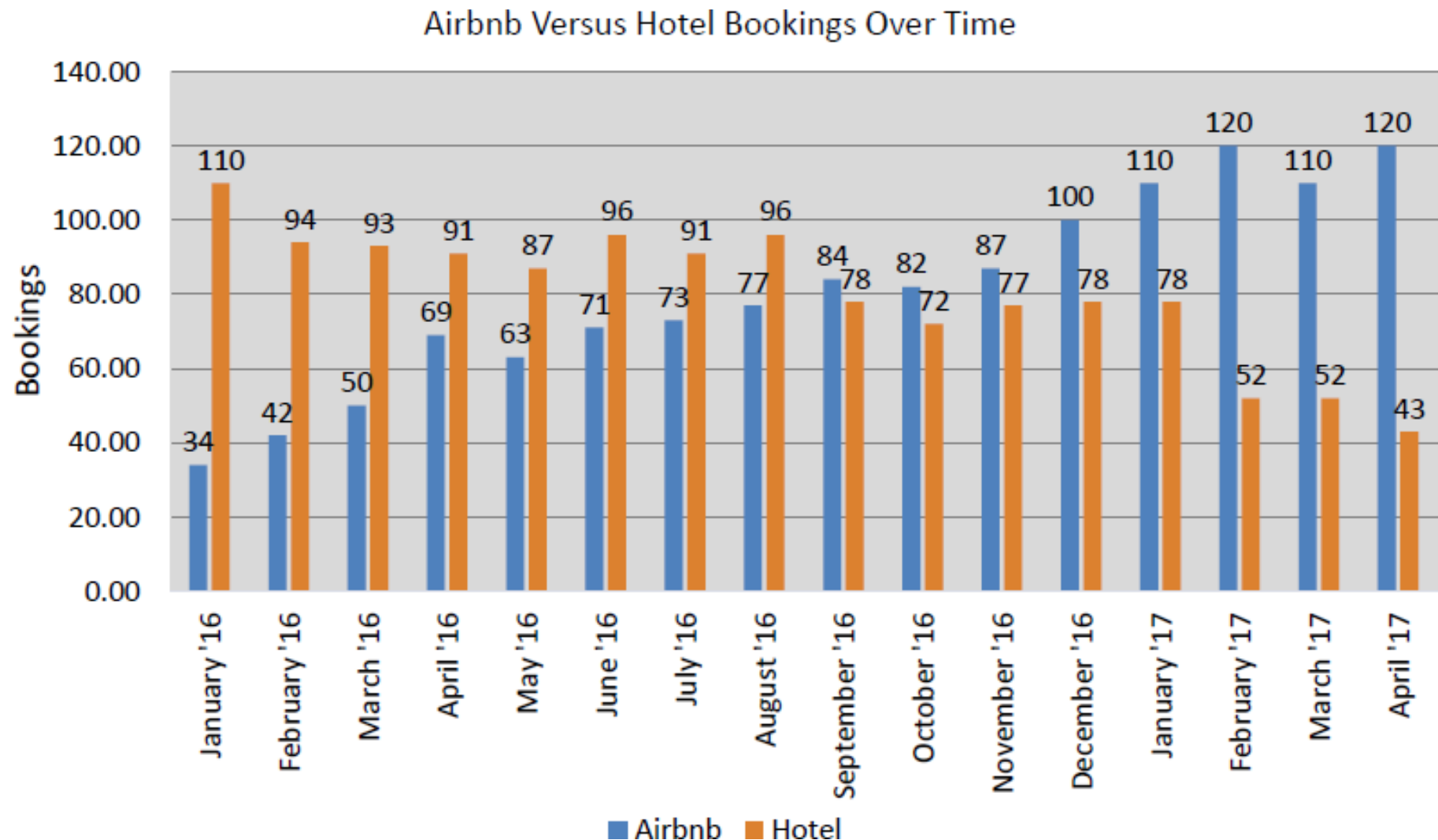


# Mejores visualizaciones

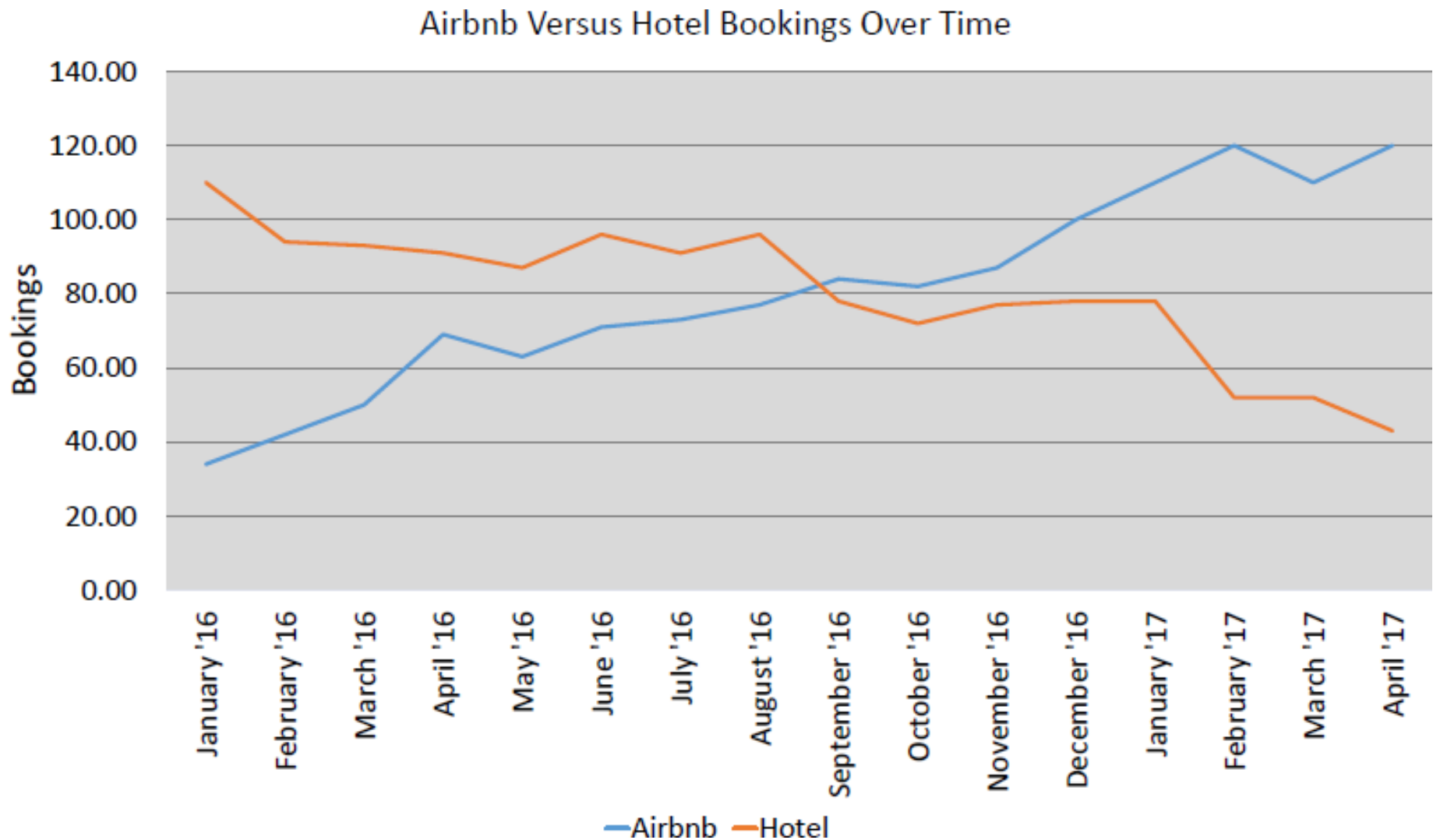
---

Un gráfico bastante sobrecargado que busca destacar el éxito de Airbnb frente al negocio tradicional de los hoteles...

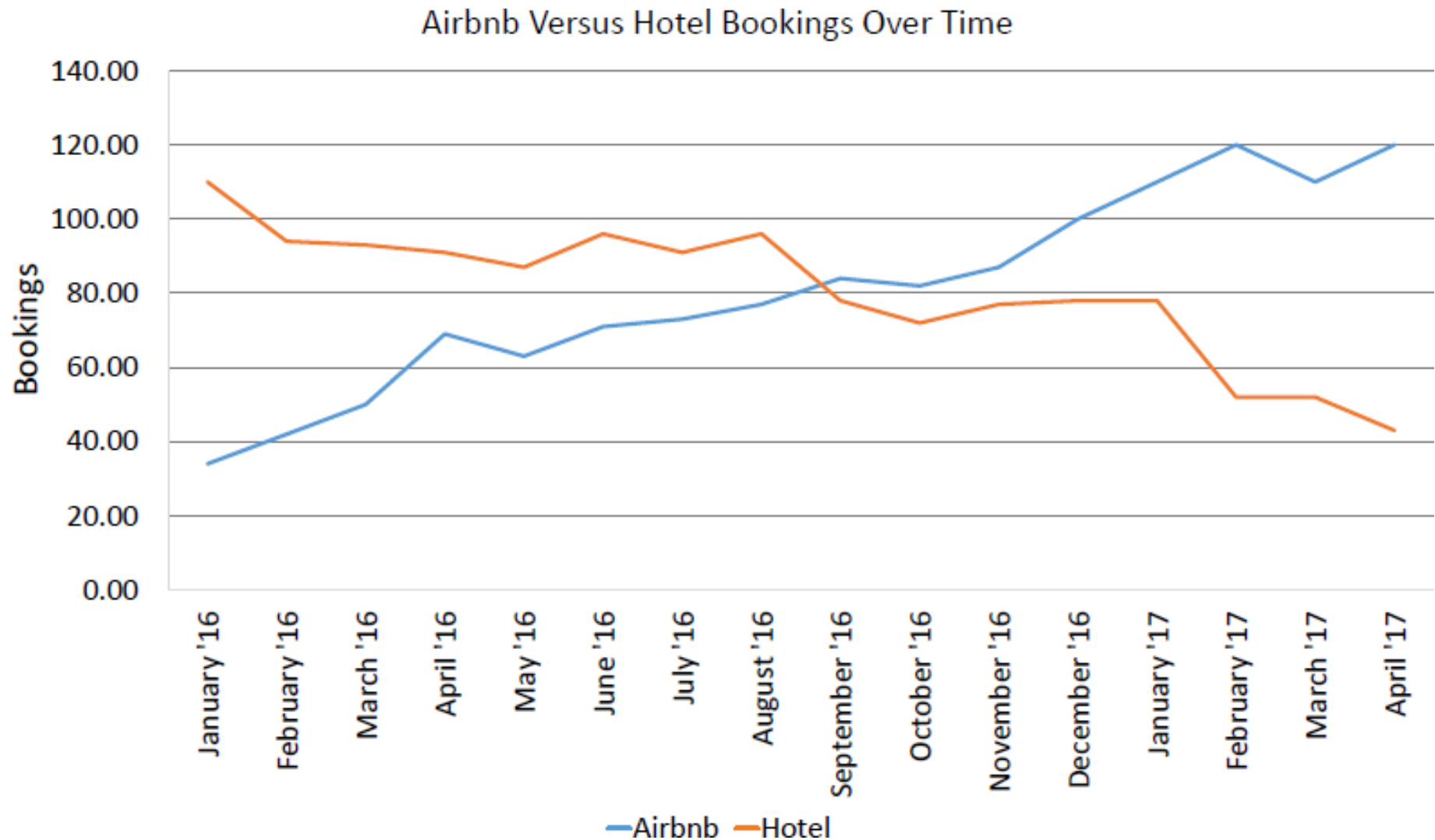
¿Se puede mejorar?



Cambiamos el gráfico de barras a líneas lo que permite ver mejor tendencias en el tiempo... de paso saquemos el exceso de números



Eliminemos el fondo plomo que no agrega nada al gráfico, el fondo blanco es visualmente más limpio...



Eliminemos las líneas de fondo, el foco del gráfico no es determinar los valores exactos de cada punto en la tendencia...



Mucho más fácil tener la leyenda al costado de las líneas para facilidad de comparación, por otro lado, se pueden simplificar las fechas a la hora de graficar...



Cuadremos los colores a la leyenda, esto incluso se puede hacer aquí con **Airbnb Bookings** y así también con **Hotel Bookings**



Potenciar el mensaje que se quiere transmitir, destacando puntos claves de la tendencia...

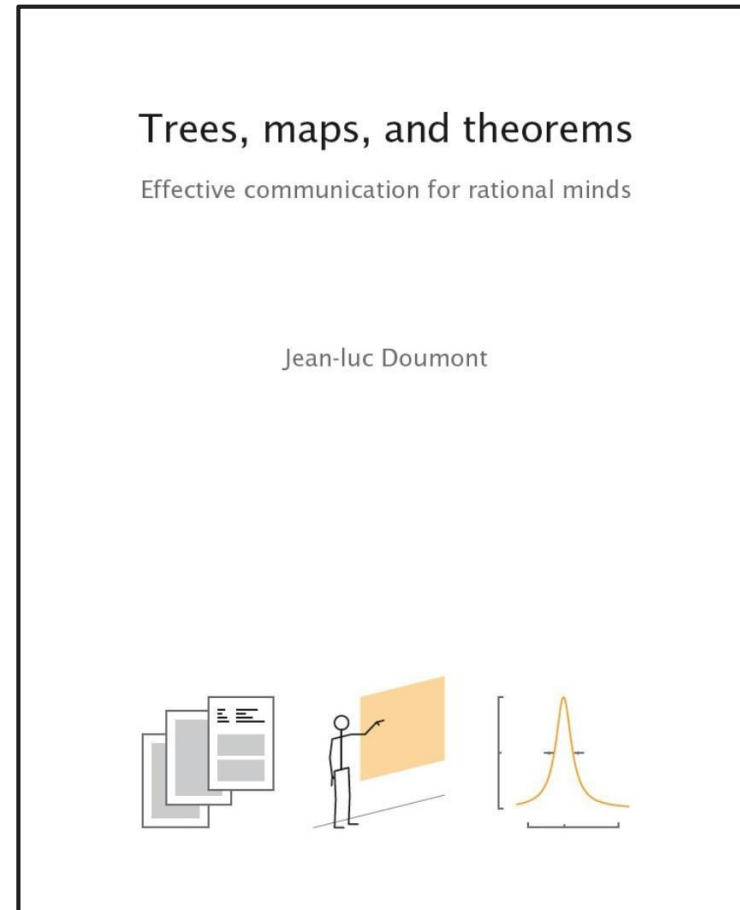
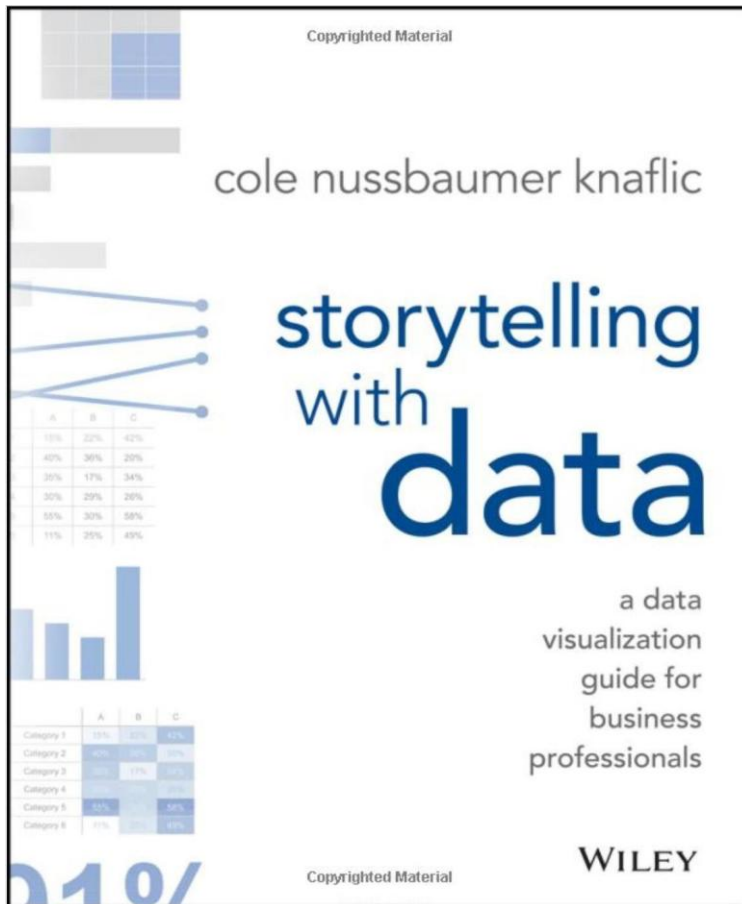
## Focus Q2-Q4 investment activity on Airbnb properties

**Airbnb Bookings** now outnumber **Hotel Bookings** 3 to 1





# Bibliografía recomendada para profundizar

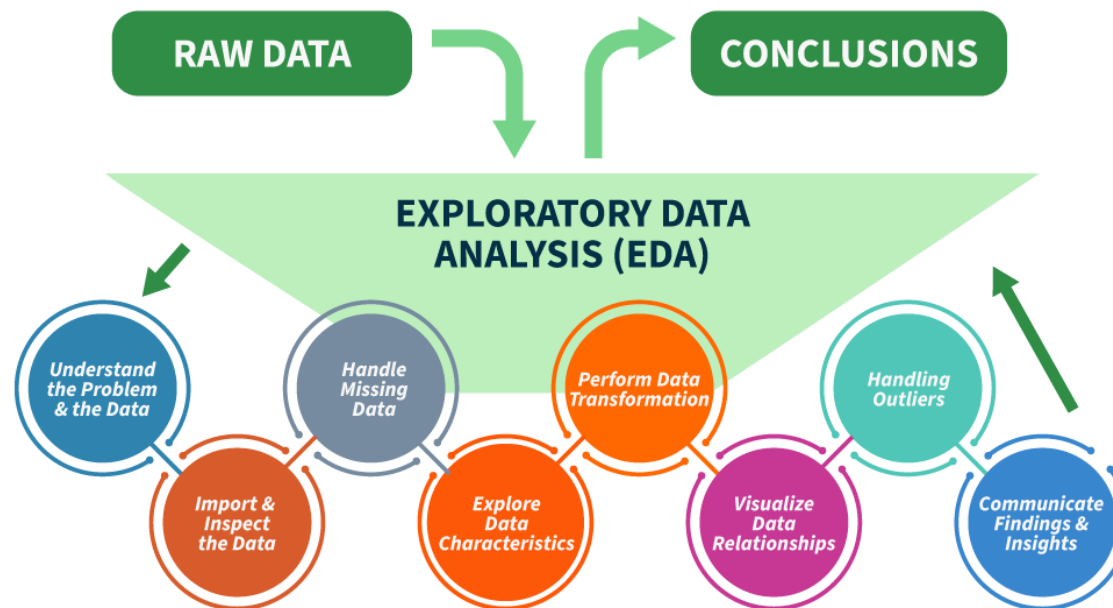


[Why scientists need to be better at data visualization](#)

# Resumen

- Los estadísticos robustos son aquellas medidas que no son afectados significativamente por valores atípicos.
- Transformaciones de los datos muchas veces permiten describirlos y trabajar mejor con ellos.
- Visualizaciones apropiadas de los datos son esenciales para comunicar apropiadamente conclusiones extraídas de los datos.

# Análisis Exploratorio de los Datos



Profesor: Pedro Saa ([pnsaa@uc.cl](mailto:pnsaa@uc.cl))

Año: 1-2025