Lecture 3
Simpson's Paradox

Dennis Sun
Stanford University
DATASCI / STATS 112

January 13, 2023

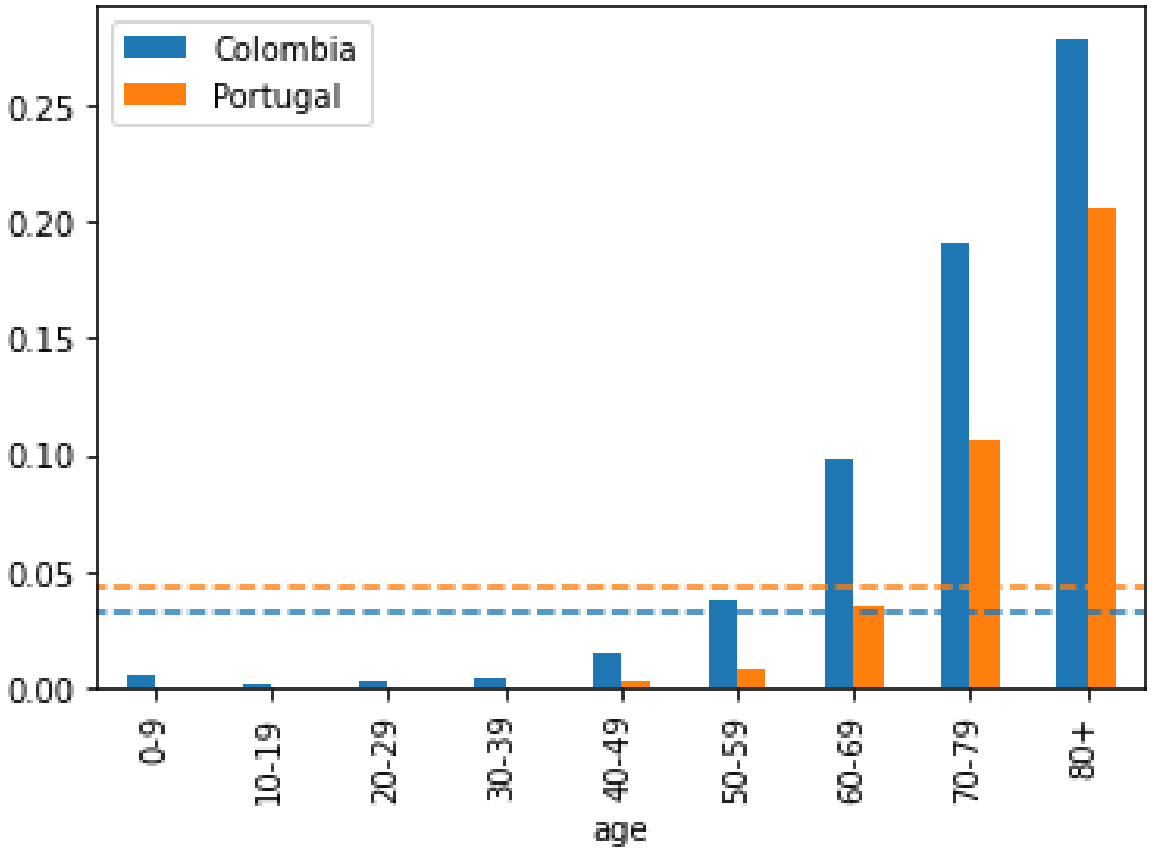
Colombia and Portugal Data

You calculated the (overall) COVID fatality rates for Colombia and Portugal.

- Colombia: 3.2%
- Portugal: 4.3%

Colombia and Portugal Data

But when you broke down fatality rate by age group (the conditional distribution $p(\text{fatality}|\text{age})$),

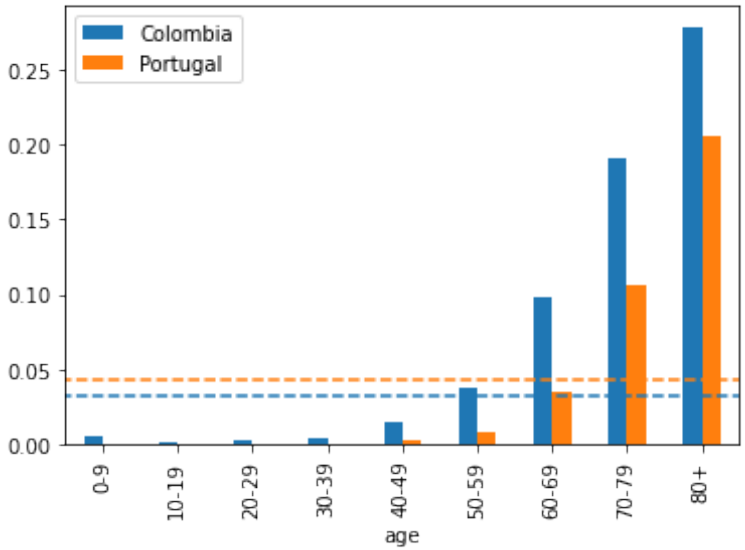


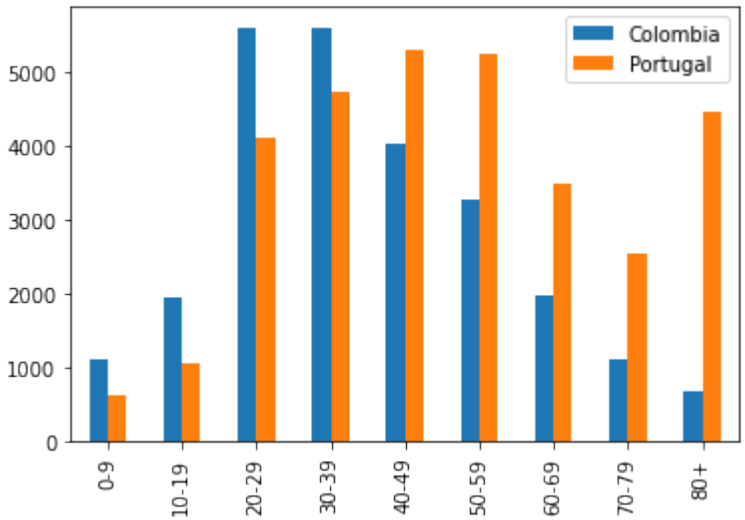
Colombia had a higher fatality rate for every single age group!

How is this possible?

Simpson's Paradox

This is known as **Simpson's paradox**.





The answer is clear if we look at the distribution of ages in the two countries.

Portugal has more older people, where the death rate is high (in both countries).

Colombia has more younger people, where the death rate is low (in both countries).