

Exploratory Analysis

3 min

Exploratory analysis is the next step after descriptive analysis. With exploratory analysis, we look for relationships between

Preview: Docs Loading link description

[variables](#)

in our dataset.

While our exploratory analyses might uncover some fascinating patterns, we should keep in mind that exploratory analyses cannot tell us *why* something happened: correlation is not the same as causation.

Let's look at an exploratory analysis with a real dataset! Check out the exploratory plots in the learning environment (you can pause if you need to!). These plots use annual income data for the United States from the government's [Social Security Administration Wage Data](#).

- We are exploring trends in annual income over time using two descriptives (mean and median). We see that the difference between these two descriptives seems to be growing. This suggests that annual income in the U.S. is becoming more *skewed*: a few very high earners are pulling the average up.
- Next, we plot the difference between mean and median annual income over time. It sure does look like it's increasing with time!
- Finally, we add a trend line and calculate the R-squared, which gives us an indication of how well the line fits the data. With an R-squared of 0.99 out of a maximum of 1.00, we can be very sure that the gap between mean and median is increasing.

Instructions

Now, consider the following:

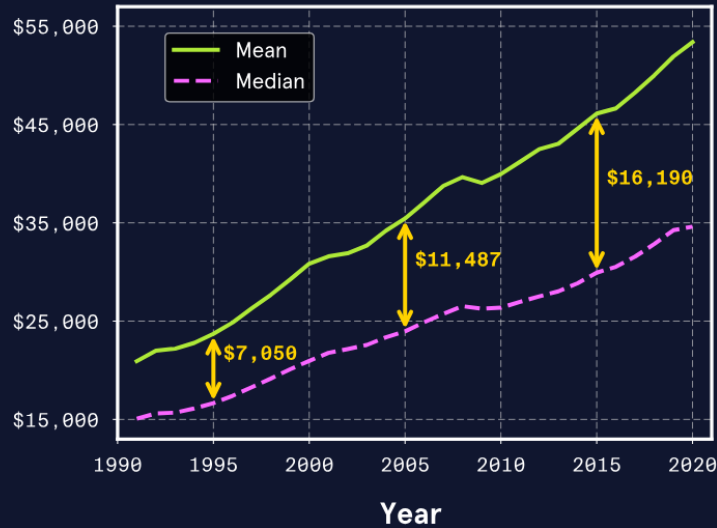
1. Can these data tell you *why* the wage gap is increasing in the U.S.?
2. Can you draw conclusions about wages in other countries from these data?

Remember:

- Exploratory analyses show us underlying patterns and relationships within datasets.
- Exploratory analyses cannot determine causation.

Next, we will take a closer look at another kind of exploratory analysis: cluster analysis!

Annual Income in the U.S.



How do the mean and median relate to each other?

Is this increase constant over time?

Difference Between Mean and Median



How linear is this pattern?