

## Aggregate Data

4 min

One measure that we haven't covered that is usually talked about alongside the mean and median is the mode. The mode is defined as the value with the highest frequency, but we can also think of the mode as the value where the peak of the distribution occurs. While not great for computations, the mode can help us identify interesting features in a variable.

For instance, there might be more than one mode, such as in our distribution of years of experience. In the following plot, we can see there's one peak near the 10-year mark and another near the 30-year mark. We would call this distribution *bimodal* because it has two modes.

Sometimes bimodal distributions occur when there are differences across categories of another variable. Given that the city seems to have a lot of young people in bands, let's see if this pattern is reflected when we find the mean of each category of the band variable.

These means are very different and very close to the locations of the modes in our plot. This indicates that there may be some differences in experience level between these two groups that are showing up in our distribution plots as two peaks.

By making this separation and then summarizing with the mean, we have **aggregated** our data. In this case, we have aggregated by summarizing a numeric variable (experience) across each value of a categorical variable (band).

We have aggregated some other data in tables in the learning environment. Do you see any interesting patterns?

# Aggregate Statistics

Income By Title

TITLE	MEDIAN
Performer	\$40,387
Manager	\$27,984
Producer	\$36,897
Educator	\$29,387
Composer	\$35,027

Income By Band Membership

BAND	MEDIAN	IQR
Yes	\$31,683	\$22,234
No	\$32,370	\$15,984

Experience By Instrument

INSTRUMENT	MEAN
Voice	19 Years
Guitar	14 Years
Piano	34 Years
Drums	10 Years
Saxophone	28 Years
Violin	36 Years