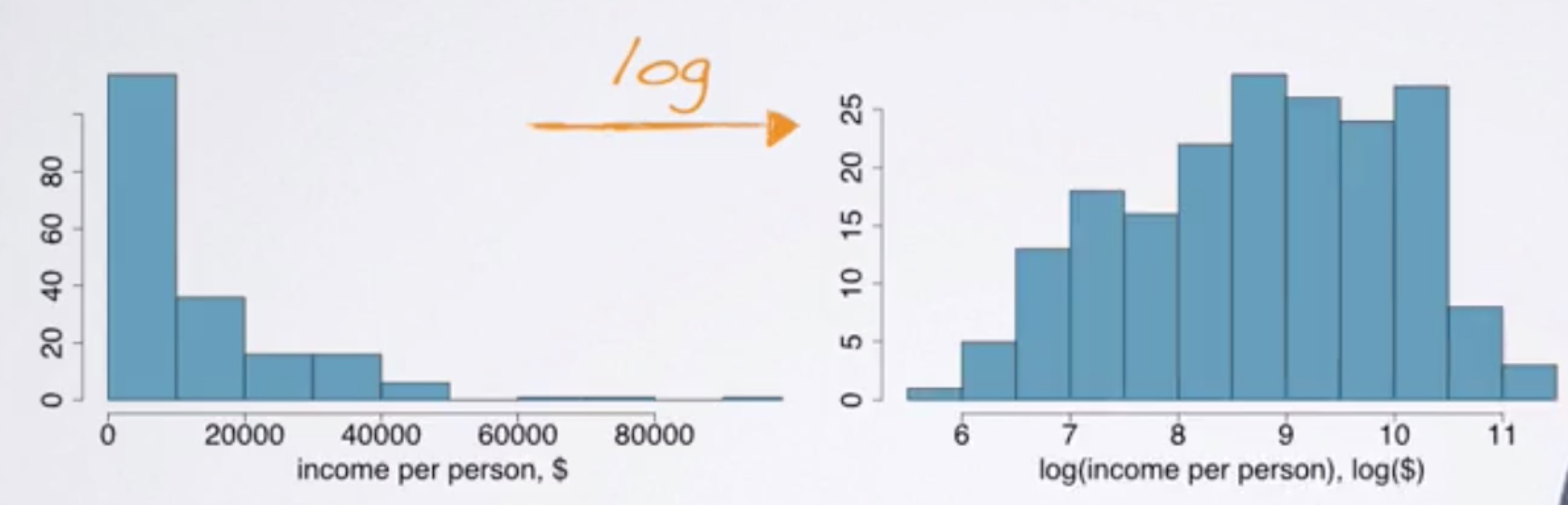


**Transform Data**

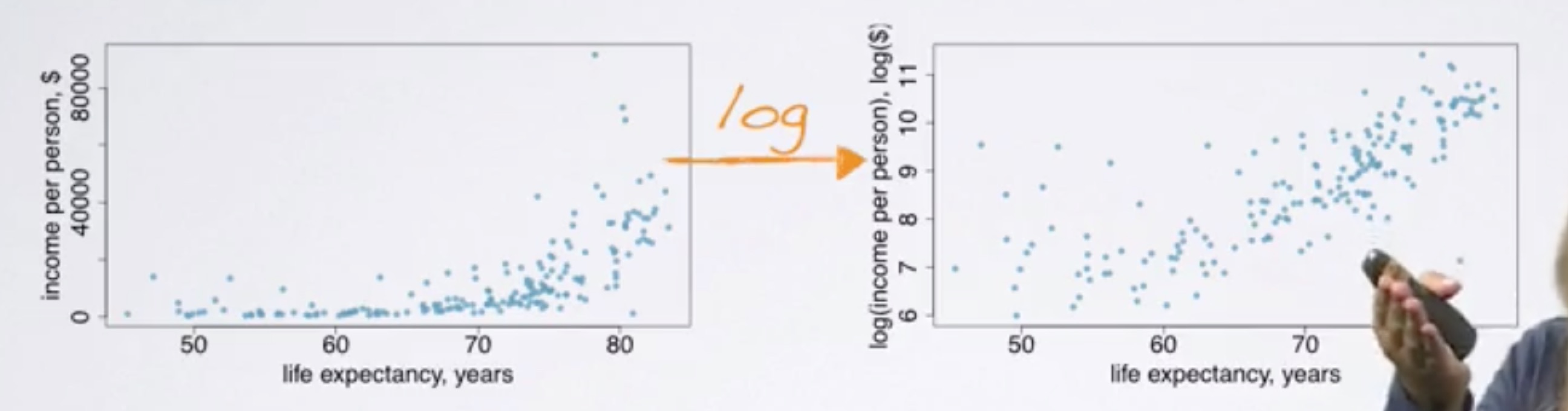
1. (Natural) log transformation

Often applied when much of the data cluster near zero (relative to the large values in the data set) and all observations are positive.

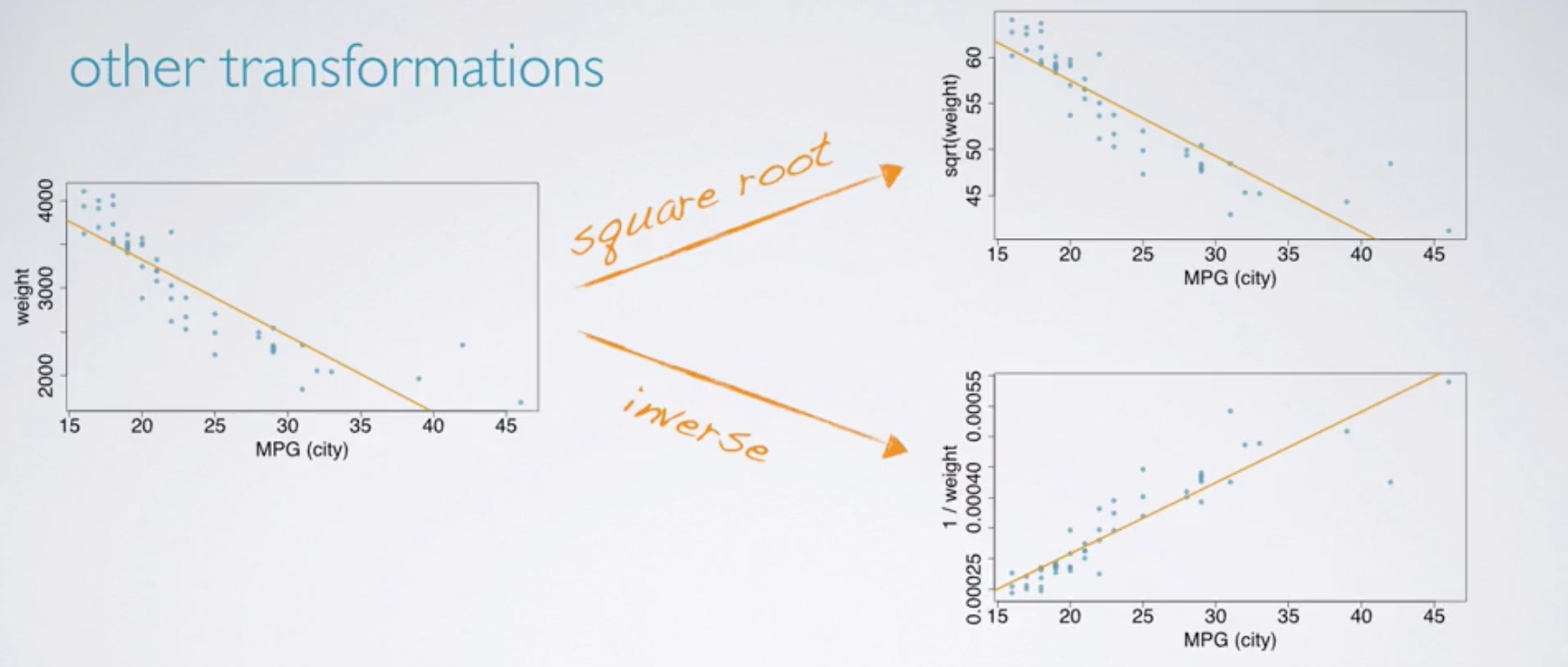


2. Log transformation

To make the relationship between the variables more linear and hence easier to model with simple methods.



3. Other transformations



**Goals of transformations**

1. To see the data structure differently

2. To reduce skew assist in modeling

3. To straighten a nonlinear relationship in a scatterplot

**How are bar plots different that histograms?**

1. Bar plots for categorical variables, histograms for numerical variables.

2. X-axis on a histogram is a number line, and the ordering of the bars are not interchangeable.

**Inference Summary**

1. Set a null and an alternative hypothesis

2. Simulate the experiment assuming that the null hypothesis is true

3. Evaluated the probability of observing an outcome at least as extreme as the one observed in the original data (P-value)

4. And if this probability is low, reject the null hypothesis in favor of the alternative

**The basic property of probability distributions.**

1. The outcomes listed must be disjoint.

2. Each probability must be between 0 and 1.

3. The probabilities must total 1.

**Normal distribution**

1. Unimodal and symmetric (bell curve)

2. follows very strict guidelines about how variably the data are distributed around the mean

3. many variables are nearly normal, but none are exactly normal

**Normal approximation to binomial**

1. Shapes of binomial distributions
2. Normal approximation

**Determine if a random variable is binomial using the four conditions.**

1. The trials are independent.
2. The number of trials, n, is fixed.
3. Each trial outcome can be classified as a success or failure.
4. The probability of a success, p, is the same for each trial.

When number of trials is sufficiently large, use normal approximation to calculate binomial probabilities, and explain why this approach works. Before using the normal approximation to the binomial, one condition to check is that np and n(1 − p) are both greater than 10.