

transforming data

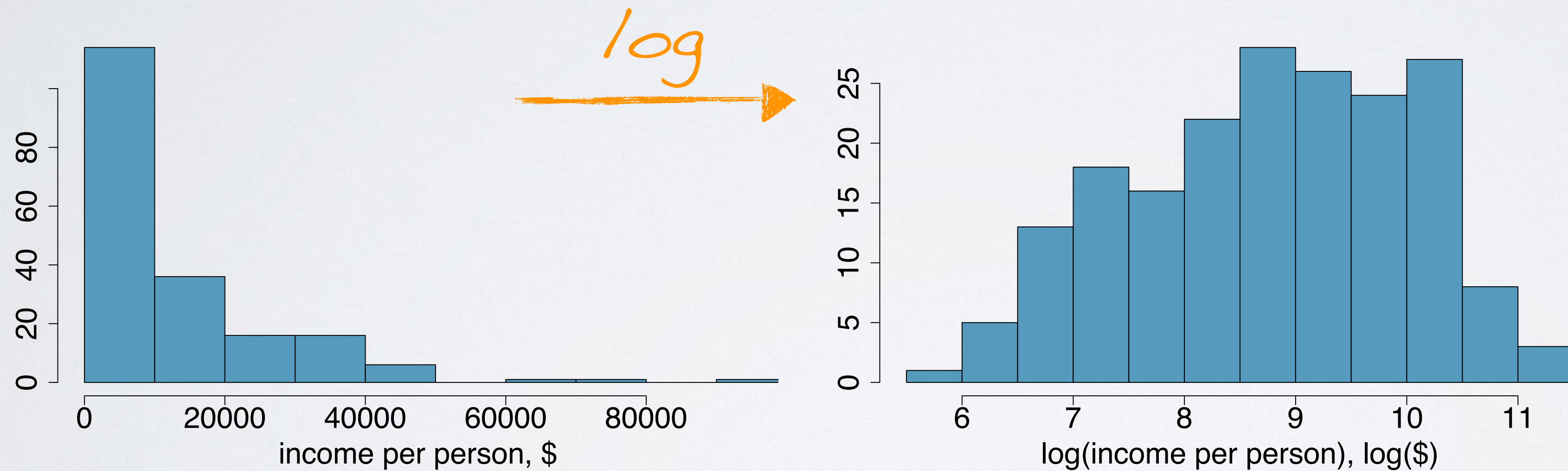
- ▶ define transformations
- ▶ review when it might be useful/
necessary to transform data

transformations

- ▶ a **transformation** is a rescaling of the data using a function
- ▶ when data are very strongly skewed, we sometimes transform them so they are easier to model

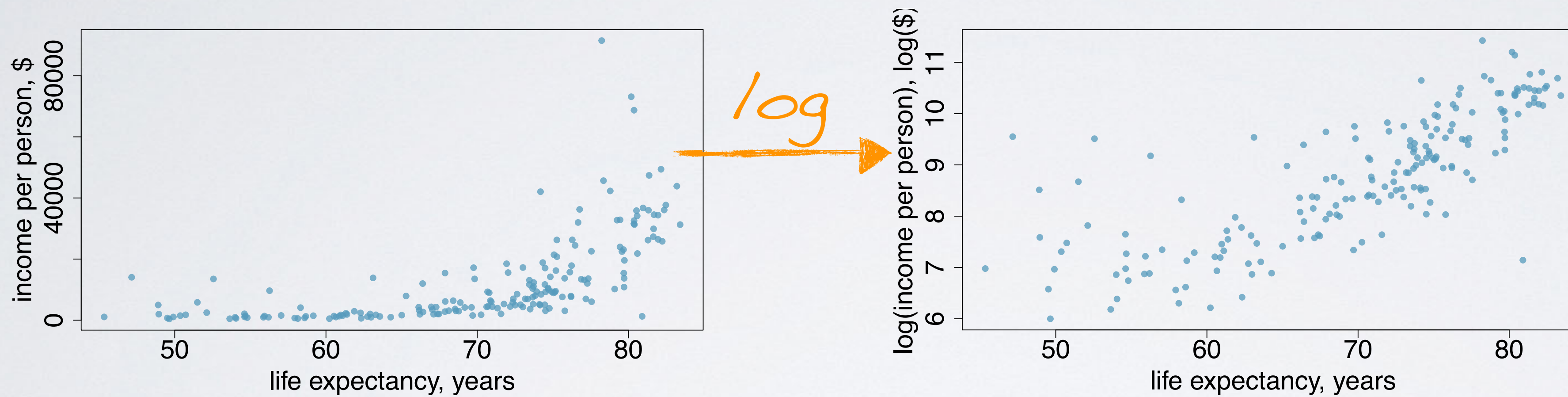
(natural) log transformation

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

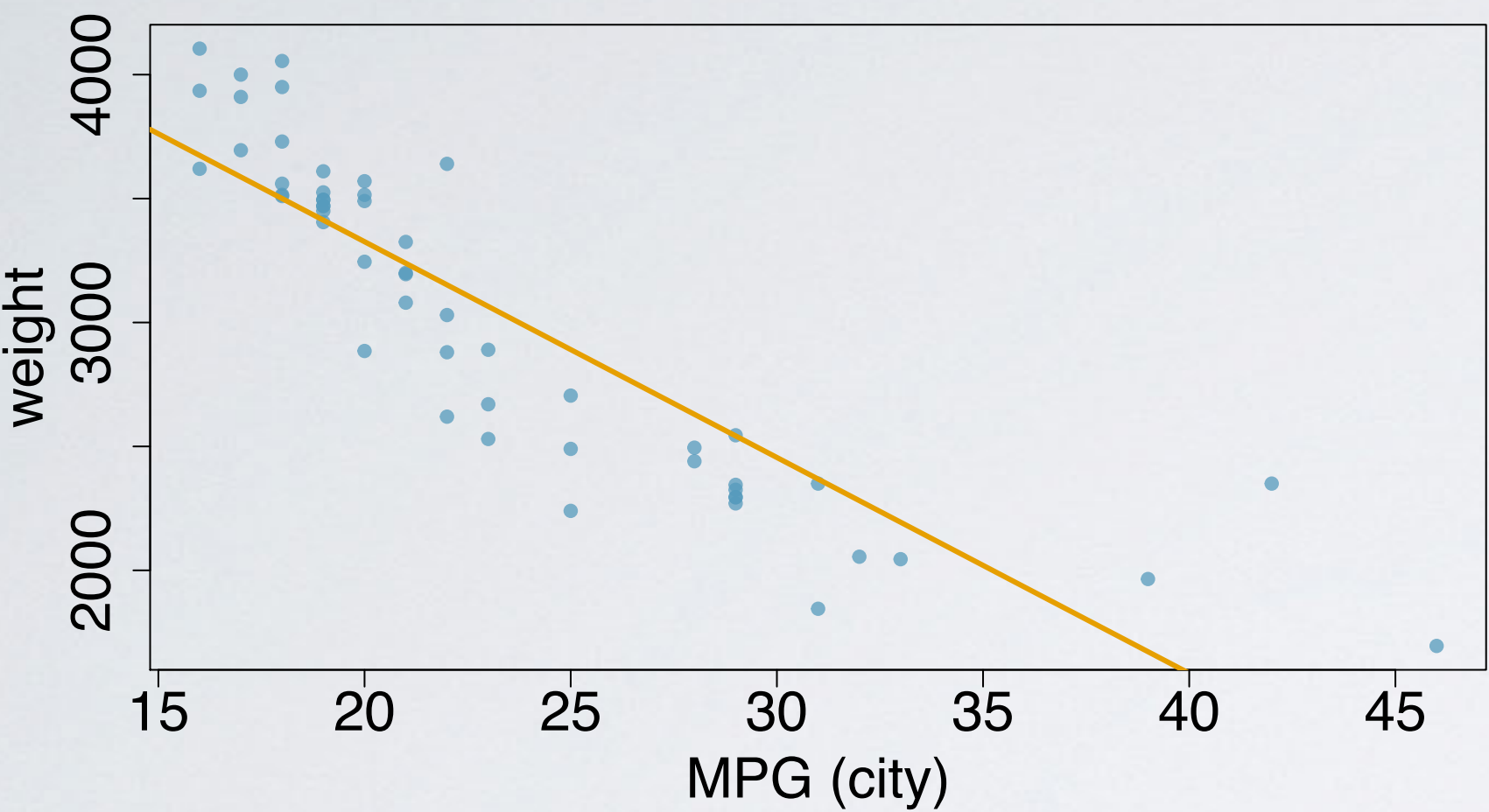


log transformation

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

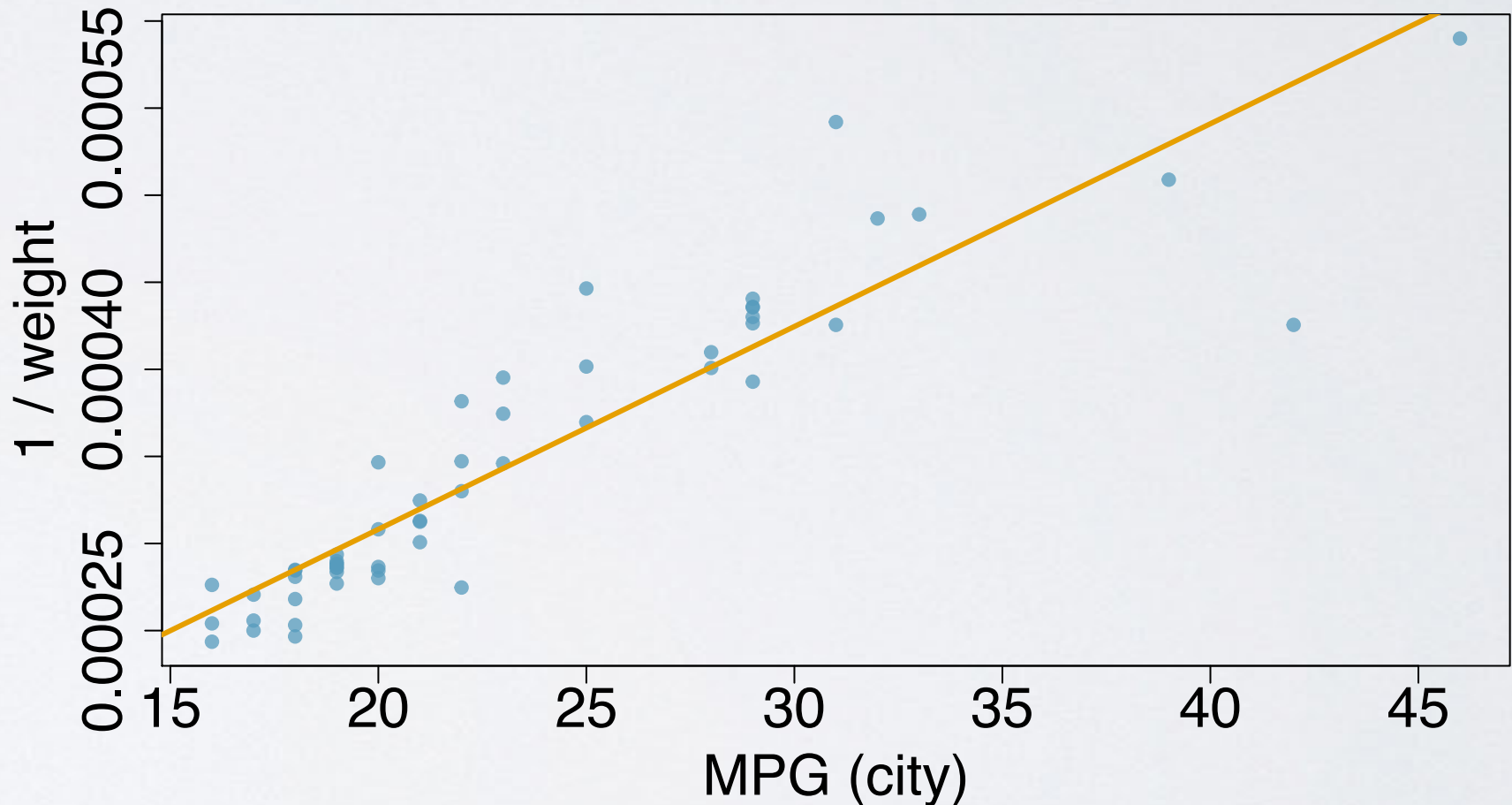
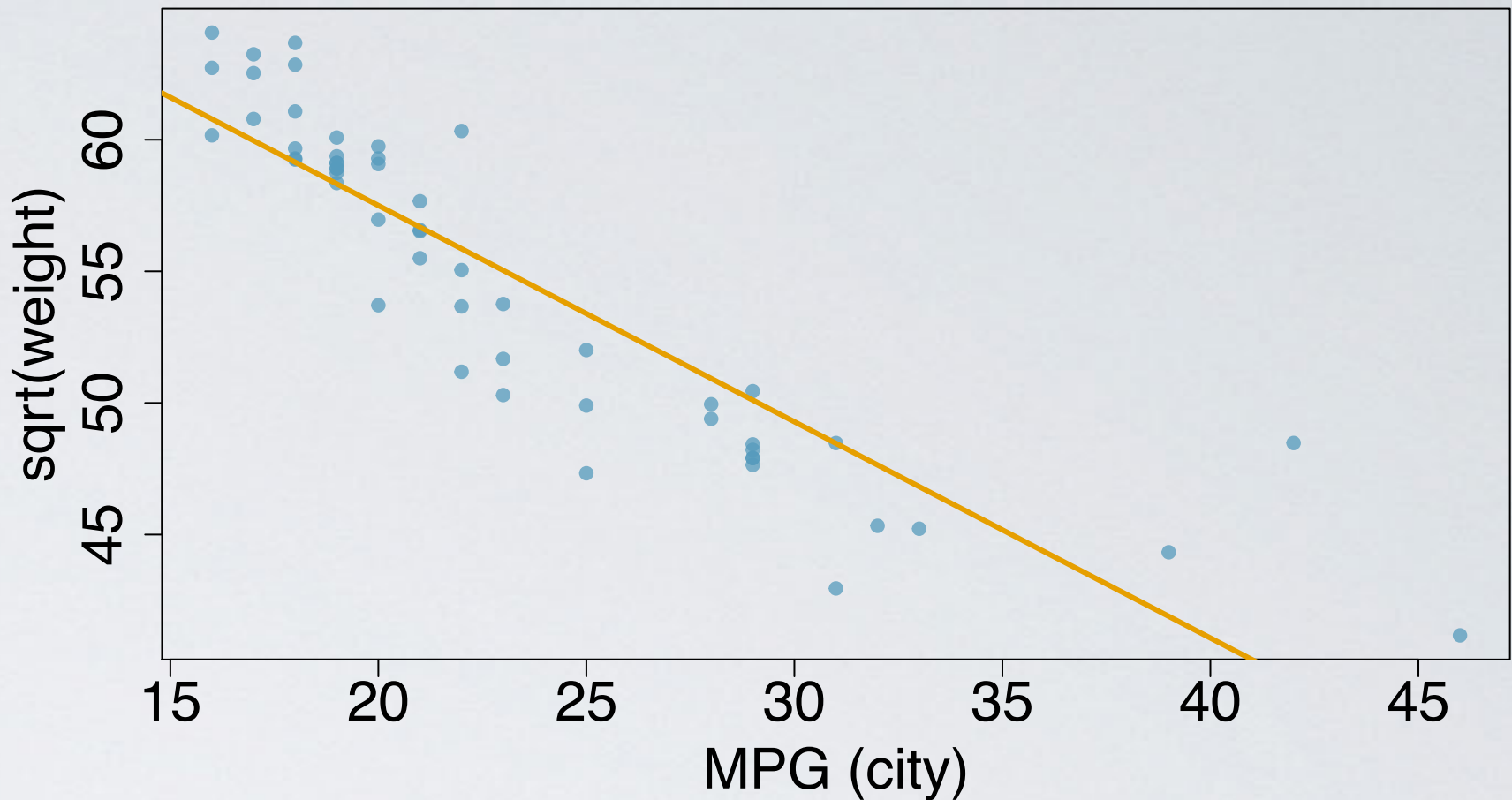


other transformations



square root

inverse



goals of transformations

- ▶ to see the data structure differently
- ▶ to reduce skew assist in modeling
- ▶ to straighten a nonlinear relationship in a scatterplot