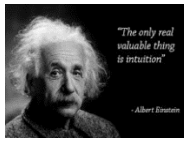
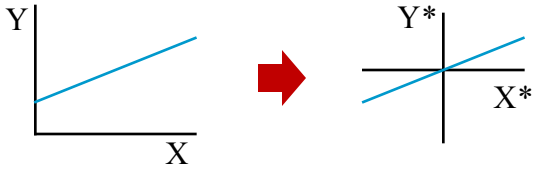
Transformation #6:
Centering



Centering: Intuition

- There are times when centering one or more variables around their means is useful.
- For example, maybe the **intercept is meaningless** because **x cannot be 0** (e.g., predicting cholesterol based on weight → nobody has 0 weight). Centering the variable shows the effect of x when x is at its mean value (not zero).
- Also, when computing **interaction** variables of **2 continuous** variables x_1 and x_2 , the resulting interaction term x_1*x_2 is **problematic** for a number of reasons:
  - **Scale invariance** – changing the scale of x_1 or x_2 (e.g., from feet to meters) will change the main effect size
  - The product of x_1*x_2 may generate severe **multicollinearity**
- **Centering** x_1, x_2 and y with respect to their means helps → $x_1^* = x_1 - \bar{x}_1;\ x_2^* = x_2 - \bar{x}_2$
- This is **equivalent** to **shifting** the Y and X **axes** to the Y and X means



R Tips

```
Centered.data = data.frame(scale(dataName,
                center=TRUE, scale=FALSE))
```
→ Using the `scale()` function with `center=TRUE` to center an entire data frame with numeric values

```
dataName$variableName = scale(Boston$medv,
                        center=TRUE, scale=FALSE)
```
→ To center a specific numeric column in the data set

KOGOD SCHOOL
of
BUSINESS