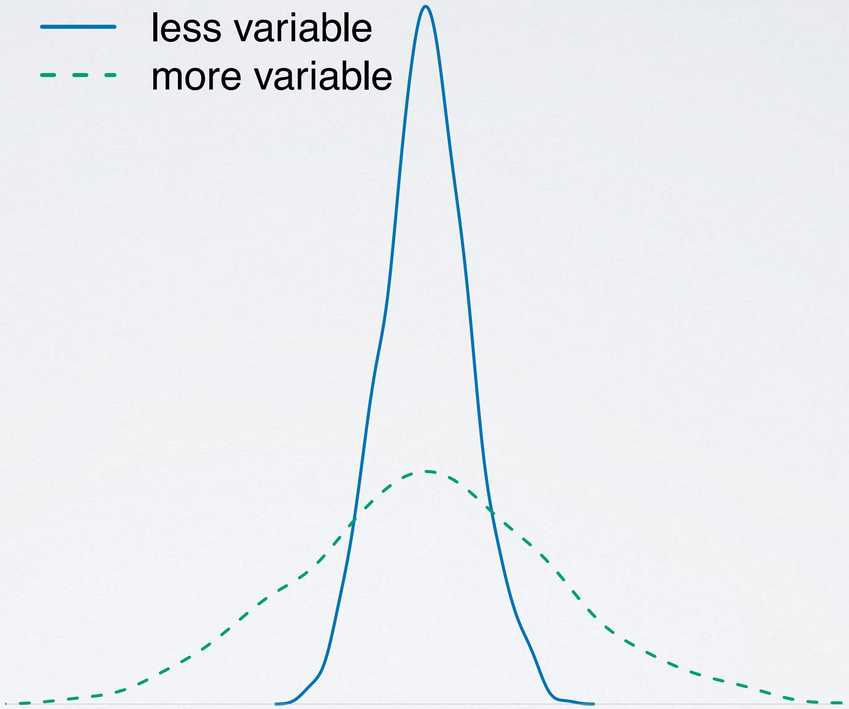
measures of spread

- range: *(max - min)*
- variance
- standard deviation
- inter-quartile range



Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University

variance

sample variance s^2

population variance σ^2

roughly the average squared deviation from the mean

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

example Given that the average life expectancy is 70.5, and there are 201 countries in the dataset:

$$s^2 = \frac{(60.3 - 70.5)^2 + (77.2 - 70.5)^2 + \ldots + (58.1 - 70.5)^2}{201 - 1}$$

$= 83.06 \text{ years}^2$

	country	life exp
1	Afghanistan	60.3
2	Albania	77.2
3	Algeria	70.9
	…	…
201	Zimbabwe	58.1

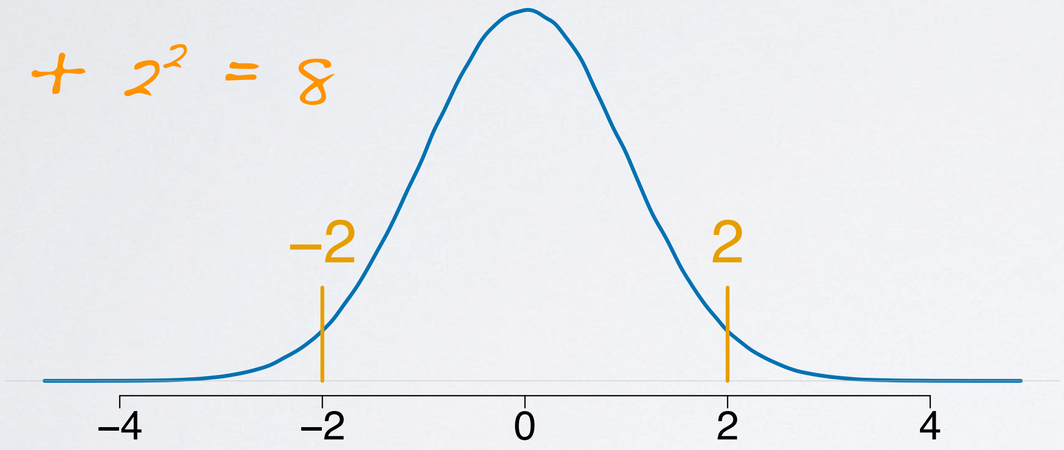
Why do we square the differences?

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

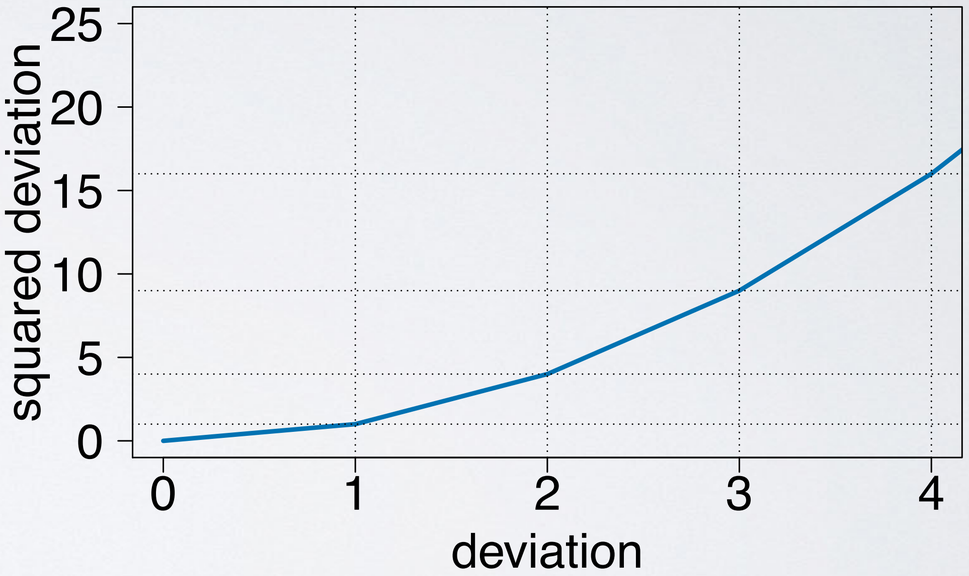
- get rid of negatives so that negatives and positives don't cancel each other when added together

$(-2) + 2 = 0$

$(-2)^2 + 2^2 = 8$



- increase larger deviations more than smaller ones so that they are weighed more heavily



standard deviation

sample sd
s

population sd
σ

roughly the average deviation around the mean, and has the same units as the data.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

square root of the variance

variability vs. diversity

Which of the following sets of cars has a more **diverse** composition of colors?

☐ **SET 1**

☐ **SET 2**

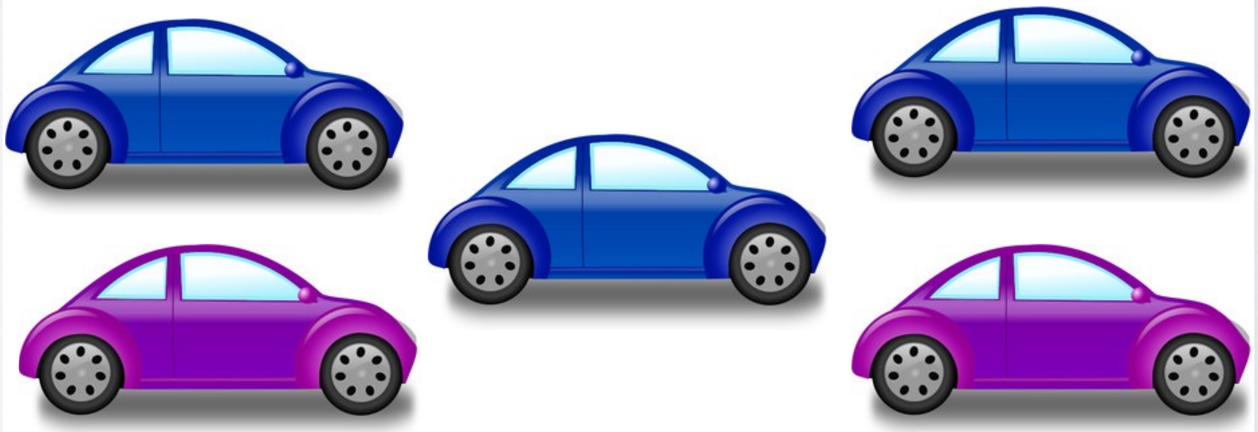


Image source: http://direct.clipart.co.za/

variability vs. diversity

Which of the following sets of cars has a more **diverse** composition of colors?

☑ SET 1

☐ SET 2

variability vs. diversity

Which of the following sets of cars has more **variable** mileage?

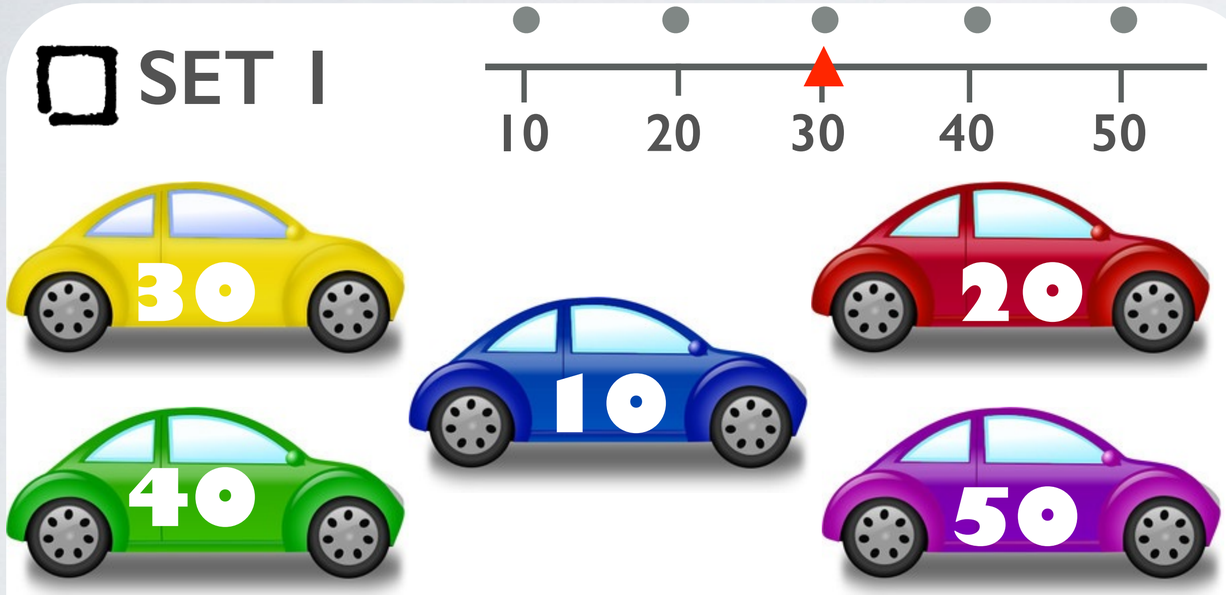
☐ SET 1

☐ SET 2

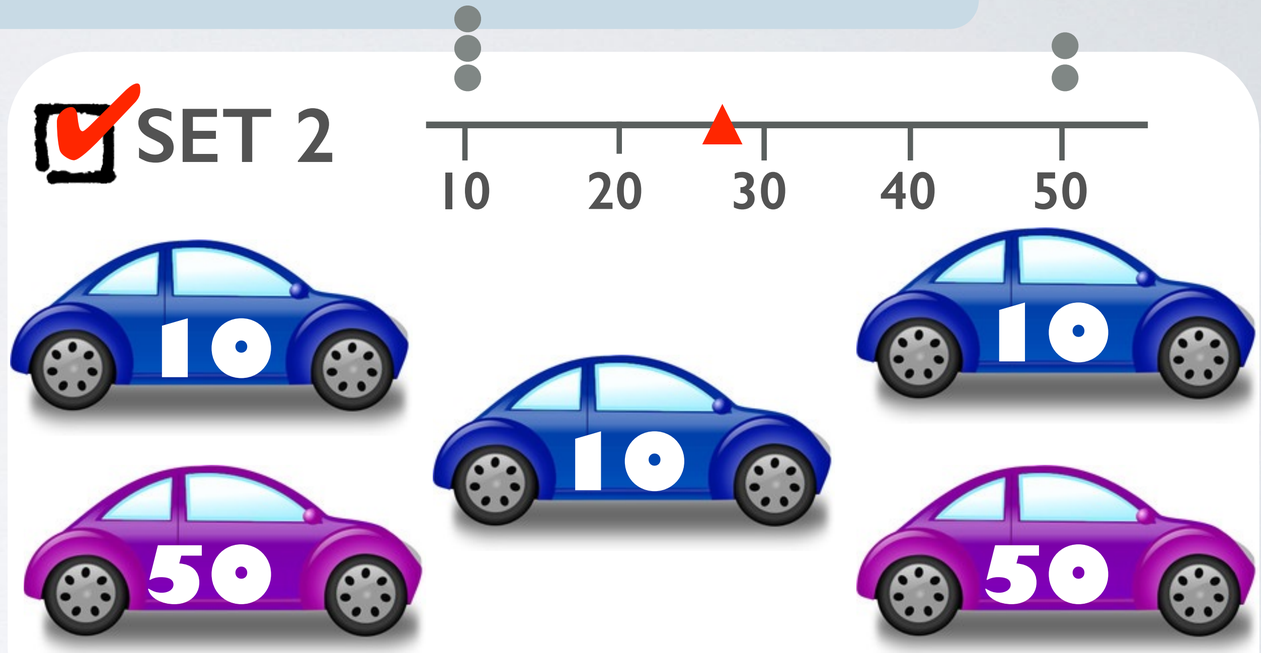
variability vs. diversity

Which of the following sets of cars has more **variable** mileage?

☐ SET 1



☑ SET 2



interquartile range

range of the middle 50% of the data, distance between the first quartile (25th percentile) and third quartile (75th percentile)

$$IQR = Q3 - Q1$$