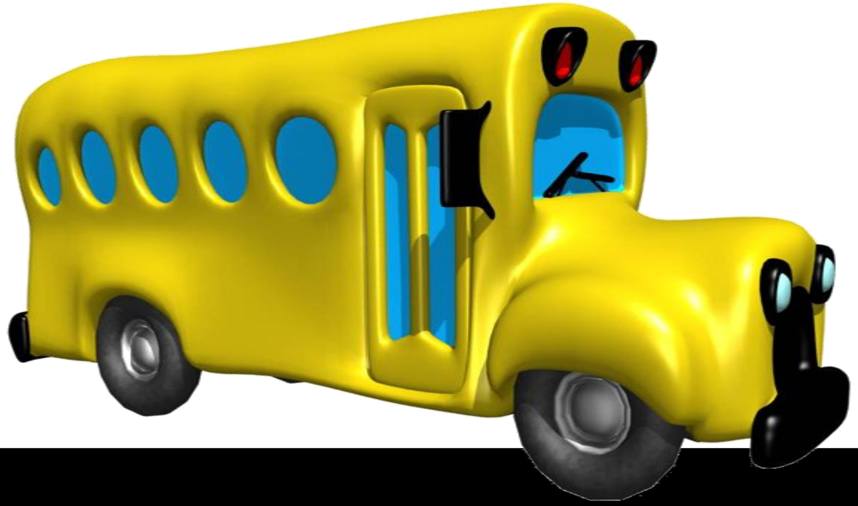
Chapter 1
Part 3

Measure of Variability



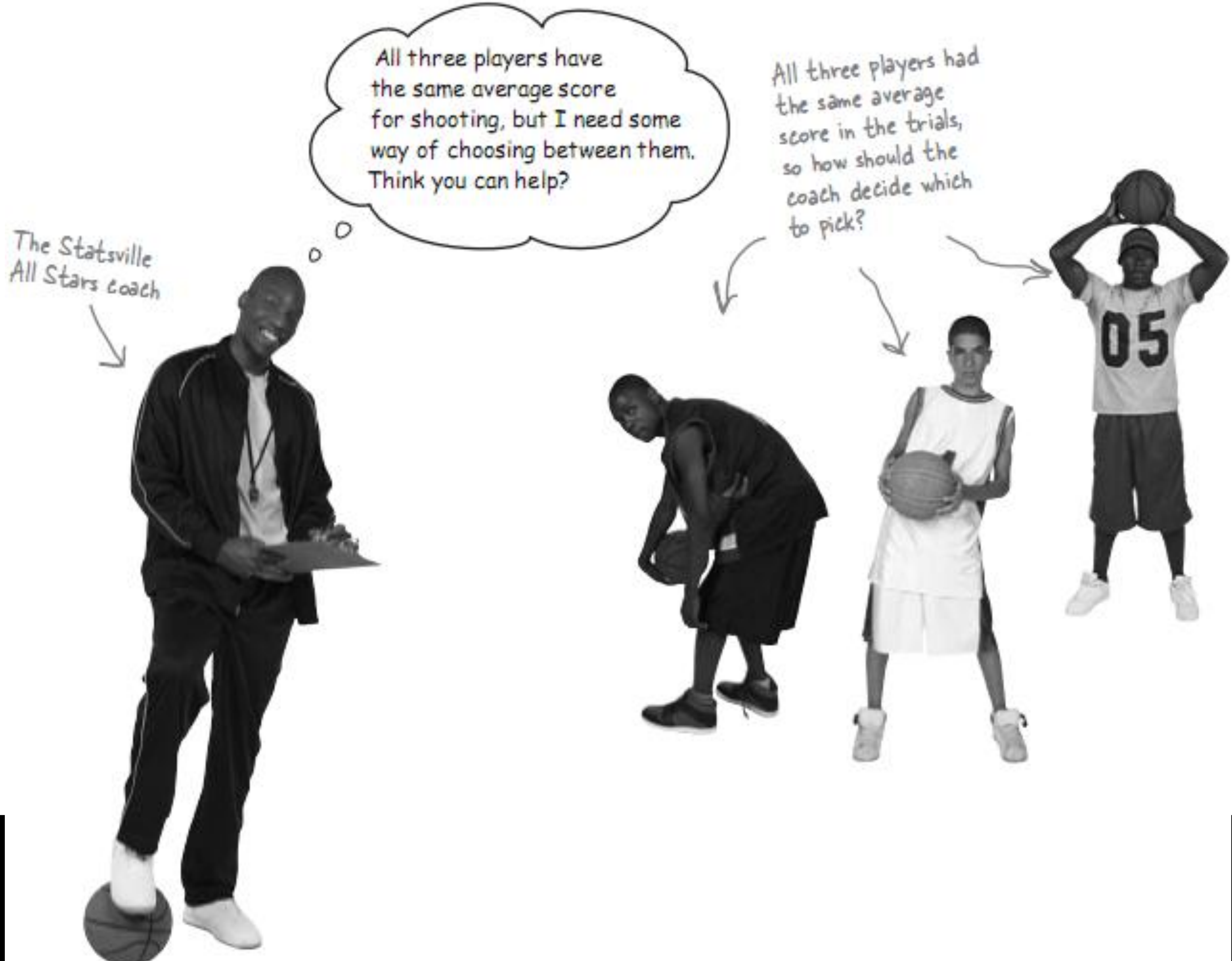
Not everything's reliable, but how can u tell?

Average do a great job of giving u a typical value in your data set, but they don't tell u the full story

→are'nt enough information in summarizing a data set

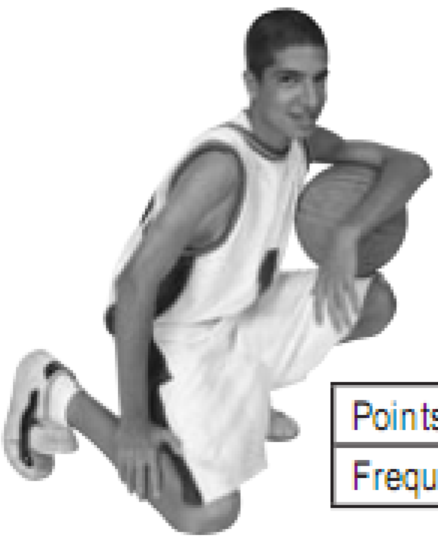
Measure of variability

- Variability → quantitative measure of the degree to which scores in a distribution are spread out
- If every X were very close to the Mean→ the mean would be a very good predictor.
- If the distribution is very sharply peaked then the mean is a good measure of central tendency → mean would be right choice
- How much do the scores "deviate" from the mean? Think of the mean as the **true score** or as your **best guess**.

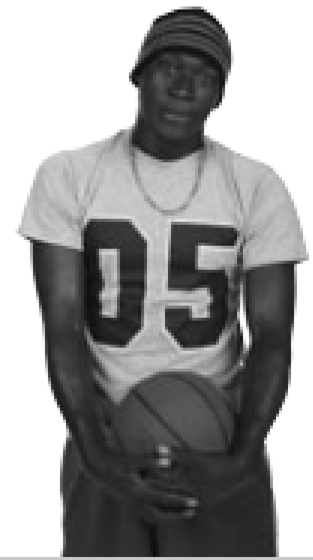

All three players have the same average score for shooting, but I need some way of choosing between them. Think you can help?
The Statsville All Stars coach
All three players had the same average score in the trials, so how should the coach decide which to pick?
05

Points scored per game	7	8	9	10	11	12	13
Frequency	1	1	2	2	2	1	1

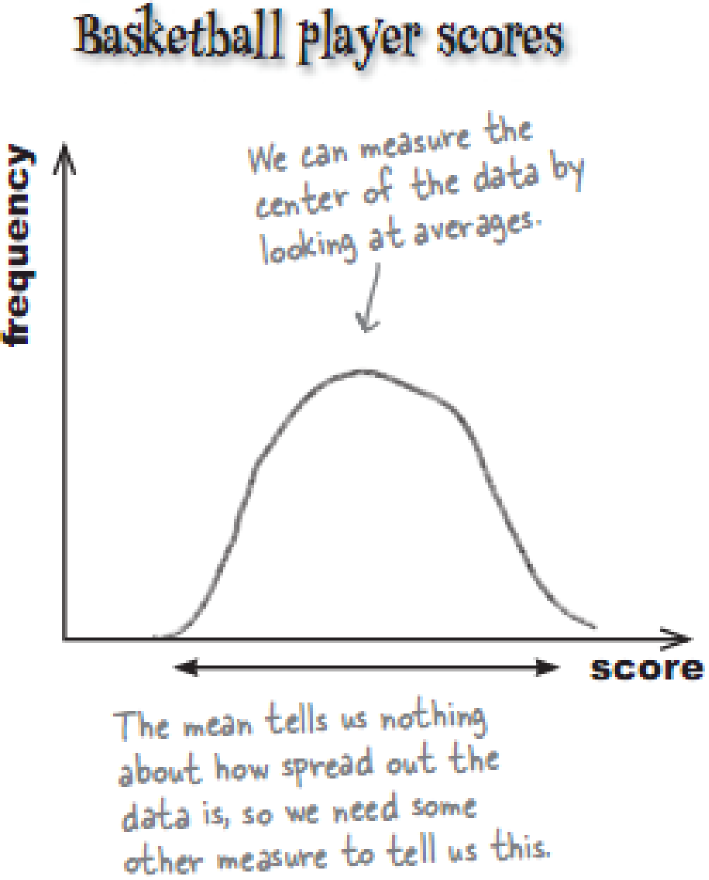
Here, frequency tells us the number of games where the player got each score. This player scored 9 points in 2 games, and 12 points in 1 game.



Points scored per game	7	9	10	11	13
Frequency	1	2	4	2	1



Points scored per game	3	6	7	10	11	13	30
Frequency	2	1	2	3	1	1	1



Range

Range is a way of measuring how spread out a set a value are.

Its given by :

Range = Upper Bound – Lower Bound

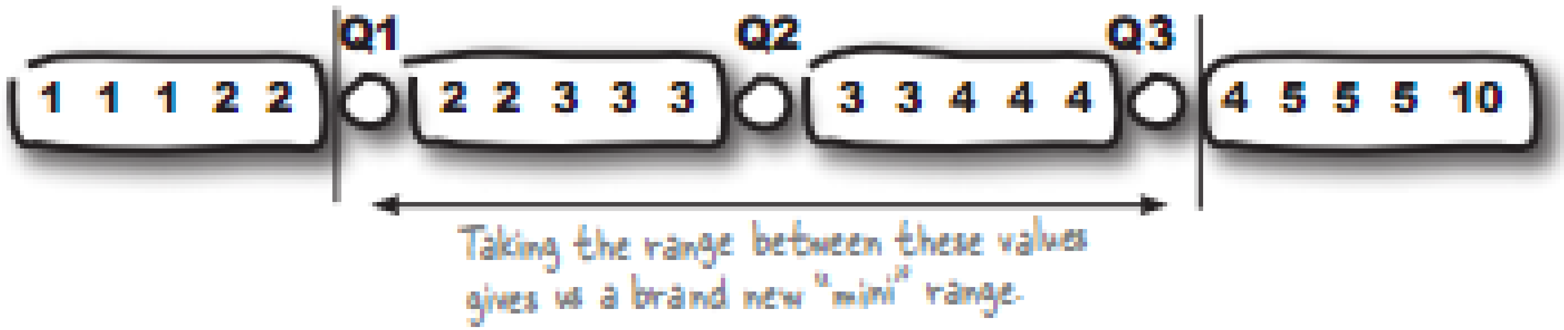
Range… cont

•Range → simple way of saying what the spread of a set of data is, but it's often not the best way of measuring how the data

- If u're data has outlier, using range can be misleading because of its sensitivity to outlier

Problem … with Range …cont

- Only describe the width of the data
- Sensitive with outlier → way out : look at *__mini range ; quartile__*



- **QUARTILE**
  - ➢If data divide into four part in same group
  - ➢Remember that you must short the data from the smallest one

For data ungrouped :

$$Q_i \text{ position} = \text{data } \frac{i(n+1)}{4}, i = 1,2,3$$

- For grouped data

$$Q_i = Tb + p\left(\frac{\frac{in}{4} - F}{f}\right)$$

With Tb : lower bound Qi

p : class interval

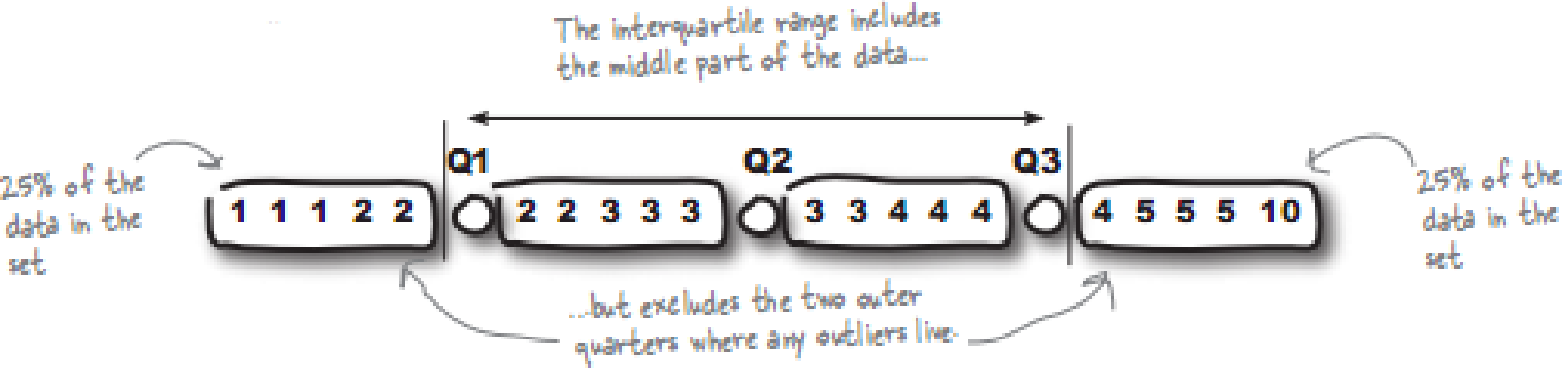
F :cummulative all frequency before Qi class

f : frequency Qi class

Other way →IQR

IQR

- Lot less sensitive then range



IQR = Q3-Q1

Inner fences & Outer fences

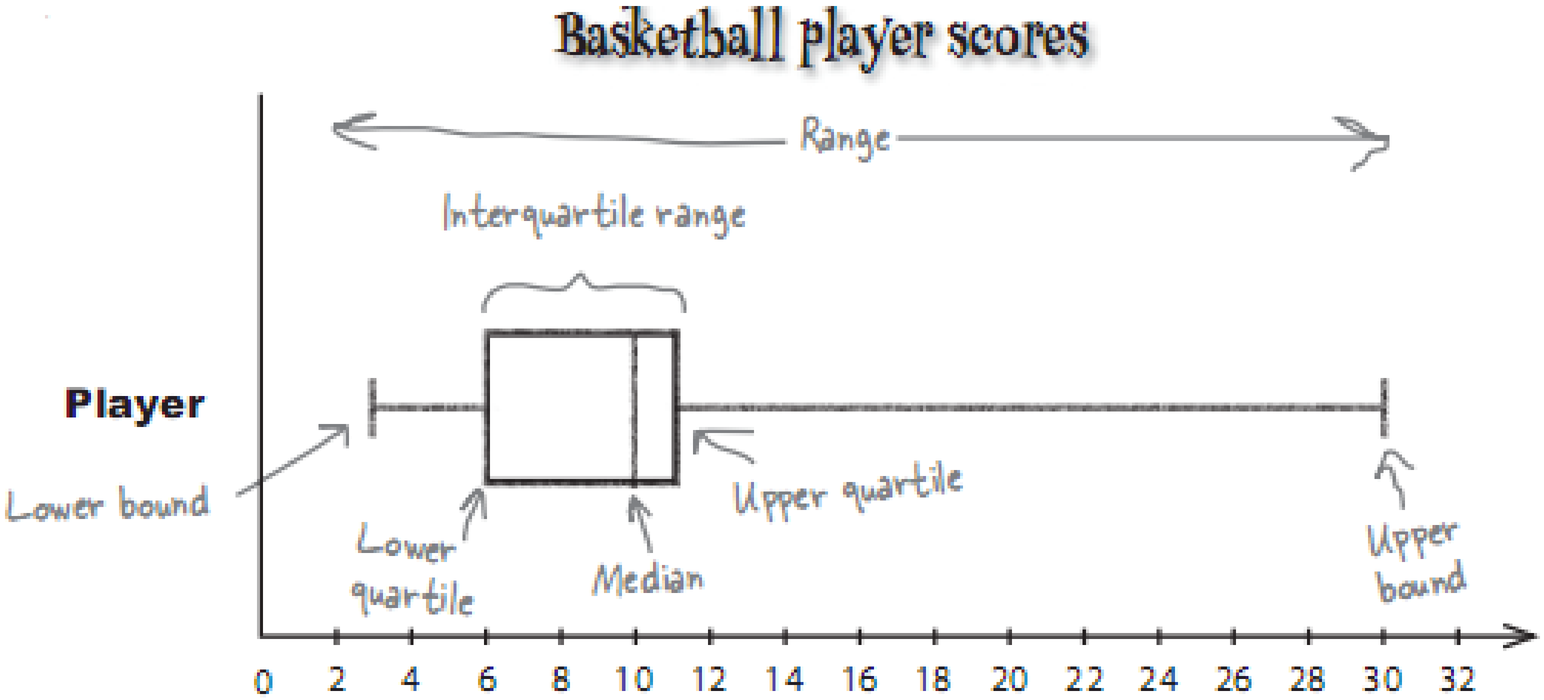
$$IF = Q_1 - 1.5(IQR) \quad \& \quad Q_3 + 1.5(IQR)$$

$$OF = Q_1 - 3(IQR) \quad \& \quad Q_3 + 3(IQR)$$

BOX PLOT

Here's a reminder of the data.

3 3 6 7 7 10 10 10 11 13 30



Work out the mean, lower bound, upper bound, and range for the following sets of data, and sketch the charts. Are values dispersed in the same way? Does the range help us describe these differences?

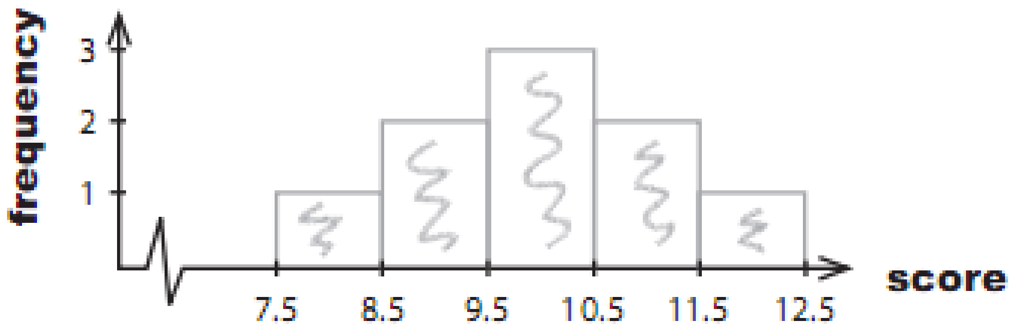
Score	8	9	10	11	12
Frequency	1	2	3	2	1

Score	8	9	10	11	12
Frequency	1	0	8	0	1

Exercise Solution

Work out the mean, lower bound, upper bound, and range for the following sets of basketball scores, and sketch the charts. Are values dispersed in the same way? Does the range help us describe these differences?

Score	8	9	10	11	12
Frequency	1	2	3	2	1



$\mu = 10$

Lower bound = 8

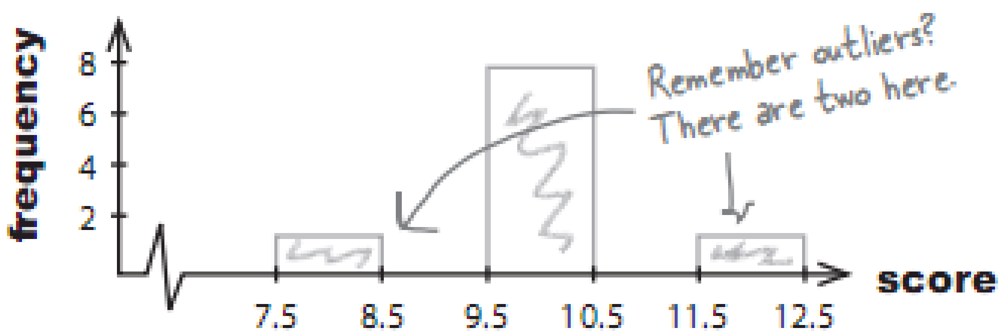
Upper bound = 12

Range = 12 − 8

= 4

Look, these results are the same even though the data's different.

Score	8	9	10	11	12
Frequency	1	0	8	0	1



$\mu = 10$

Lower bound = 8

Upper bound = 12

Range = 12 − 8

= 4

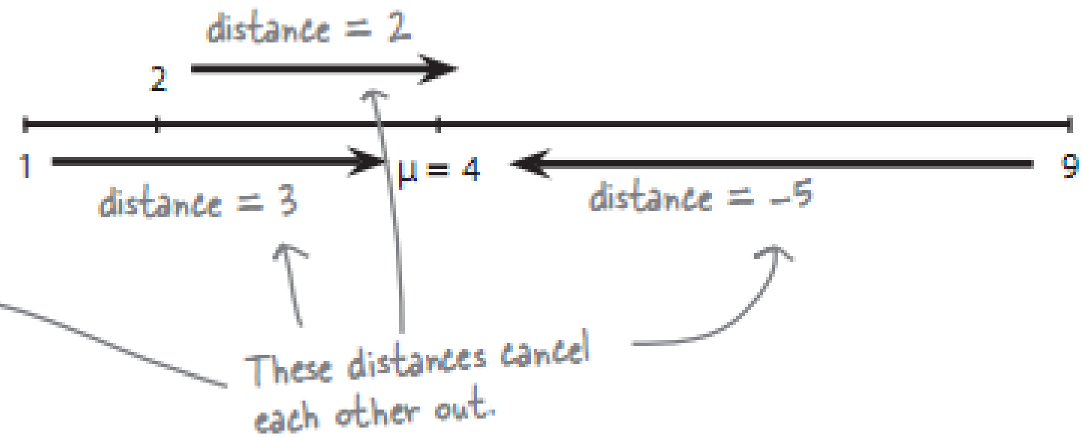
Measure of Variability … Variance

Imagine we have 3 number : 1, 2, 9, mean =4. so the average distance of the values from the mean?

$$\text{Average distance} = \frac{(1 \text{ to } \mu) + (2 \text{ to } \mu) + (9 \text{ to } \mu)}{3}$$

$$= \frac{3 + 2 + (-5)}{3}$$

$$= 0$$



Measure of variability

- Variance

  Deviation: deviation of one score from the mean

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{n} X_i^2}{N} - \mu^2 \Rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N}}$$

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

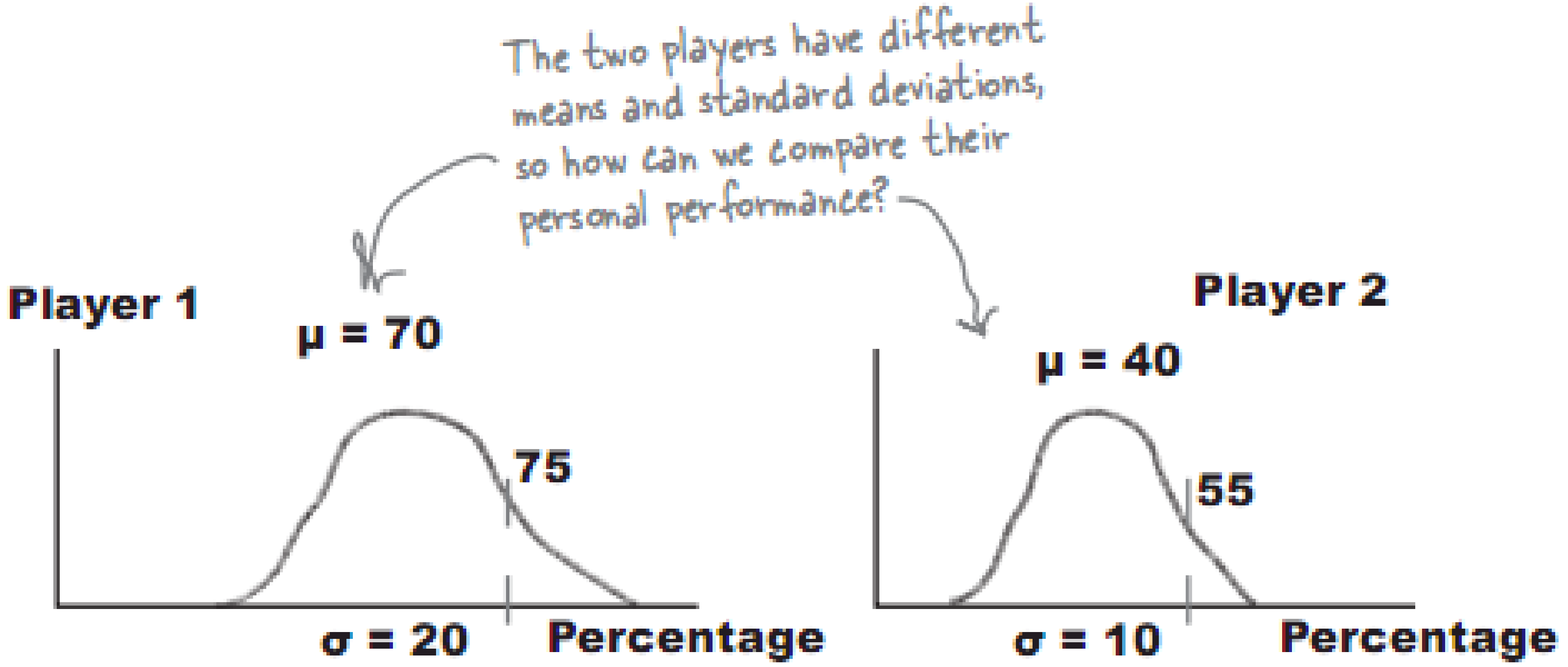
Standard score

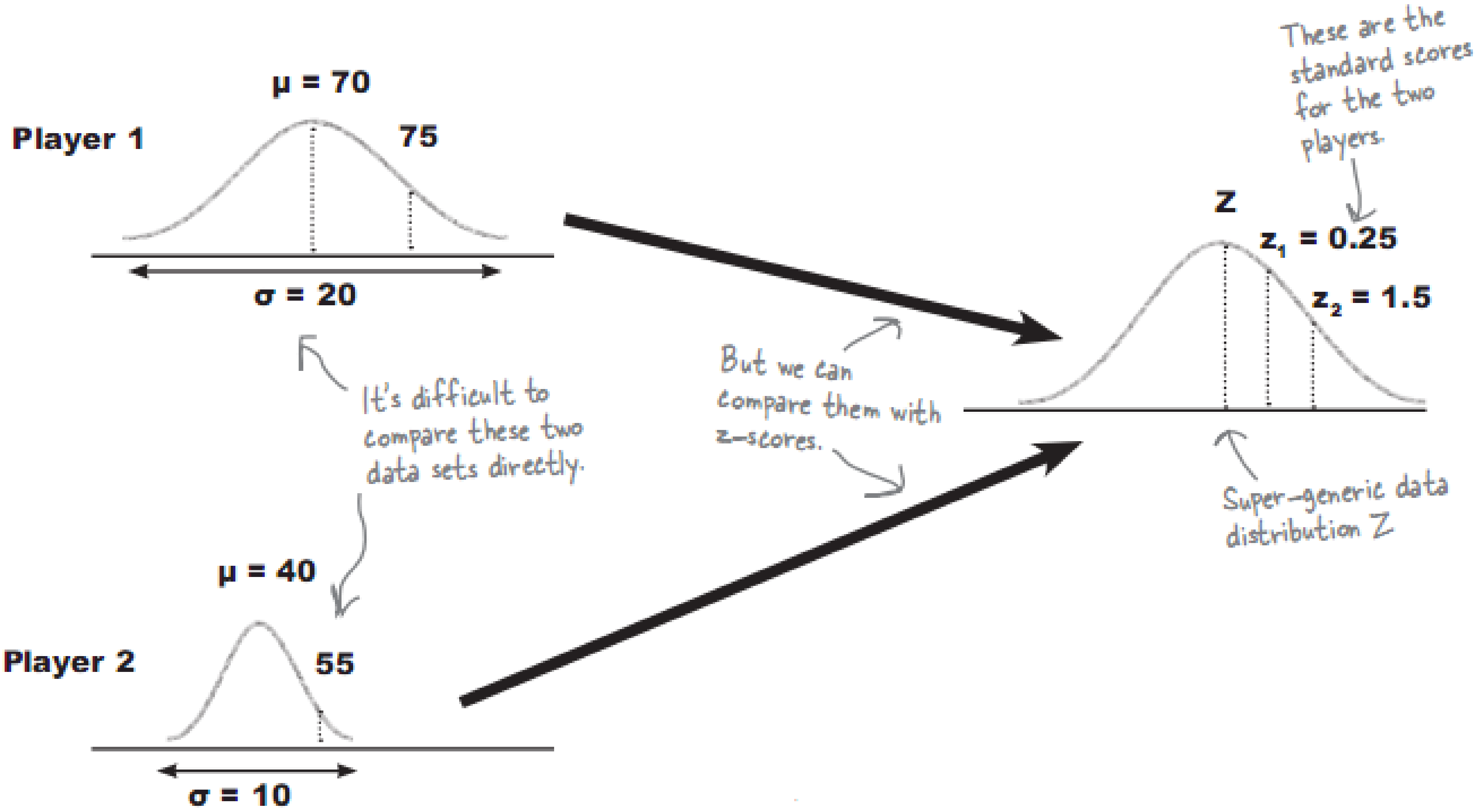
Standard scores give a way of comparing values as if they came from the same set of data or distribution. Formula :

$$z = \frac{x - \mu}{\sigma}$$

What if we need a baseline for comparison?

75% sounds like a high percentage but not taking into account the mean and standard deviation of each player. How can we compare the two player?





Which players better ?

exercise

Here are the scores for the three players. The mean for each of them is 10. If you are the coach, and work out the standard deviation for each player. Which player is the most reliable one for your team ?

Player 1

Score	7	9	10	11	13
Frequency	1	2	4	2	1

Player 2

Score	7	8	9	10	11	12	13
Frequency	1	1	2	2	2	1	1

Player 3

Score	3	6	7	10	11	13	30
Frequency	2	1	2	3	1	1	1

Do U have same conclusion when U work out with Range ?