1-8

Lesson [redacted] → Outliers and the 1.5 IQR Rule

Outliers mess up beautiful data

Maybe we can get better predictions if we just exclude the outliers (delete them)

but we would need a commonly agreed upon rule concerning when data really is an outlier

Enter..... The **1.5 IQR** rule

steps
1. get box plot
2. find IQR
3. multiply IQR by 1.5
4. window of good data is

$$Q1 - 1.5(IQR)$$
$$Q3 + 1.5(IQR)$$

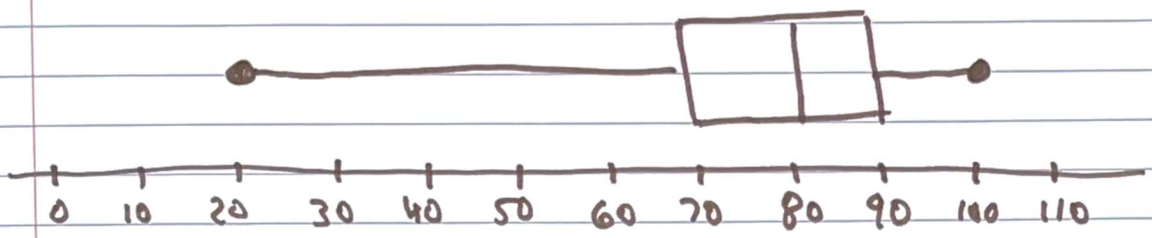
5. delete outliers and redo all statistics calculations from scratch

Example of 1.5 IQR Rule

Below are a student's quiz scores

20, 60, 60, 70, 80, 80, 80, 80, 80, 80, 80, 90, 90, 90, 100

Q1 median Q3



IQR = 20 (90-70)

1.5 IQR = 30 (1.5*20)

Q1 - 30 = 40 → Bottom window

Q3 + 30 = 120 → Top window

Outlier Found! quiz score 20

old data's average is 76

(20+60+60+70+80+80+80+80+80+80+80+90+90+90+100) ÷ 15 = 76

New data's average is 80

(60+60+70+80+80+80+80+80+80+80+90+90+90+100) ÷ 14 = 80