Quantitative Data: Numerical Summaries

Reed Coots

Statistics with Python Course Developer

Adult Male Heights

5 Number Summary:

- **Min**
- **1st Quartile** (25%)
- **Median** (50%)
- **3rd Quartile** (75%)
- **Max**

```
     Height
Min.   :61.7
1st Qu.:66.5
Median :68.3
Mean   :68.3
3rd Qu.:70.1
Max.   :75.1
```

$IQR = Q3 - Q1$

$70.1 - 66.5$

So, the IQR is another measure of spread.

Salaries in San Francisco (2011-2014)

Min.	25%	50%	75%	Max.	Mean	SD	n
-618.1	36169	71427	105839	567595	74768	50517	148654

Scribe that is often used to get numerical summaries.

Salaries in San Francisco (2011-2014)

21 Median 23 Standard Deviation

Min.	25%	50%	75%	Max.	Mean	SD	n
-618.1	36169	71427	105839	567595	74768	50517	148654

This again is another measure of spread,

Salaries in San Francisco (2011-2014)

21 Median 23 Standard Deviation Sample size

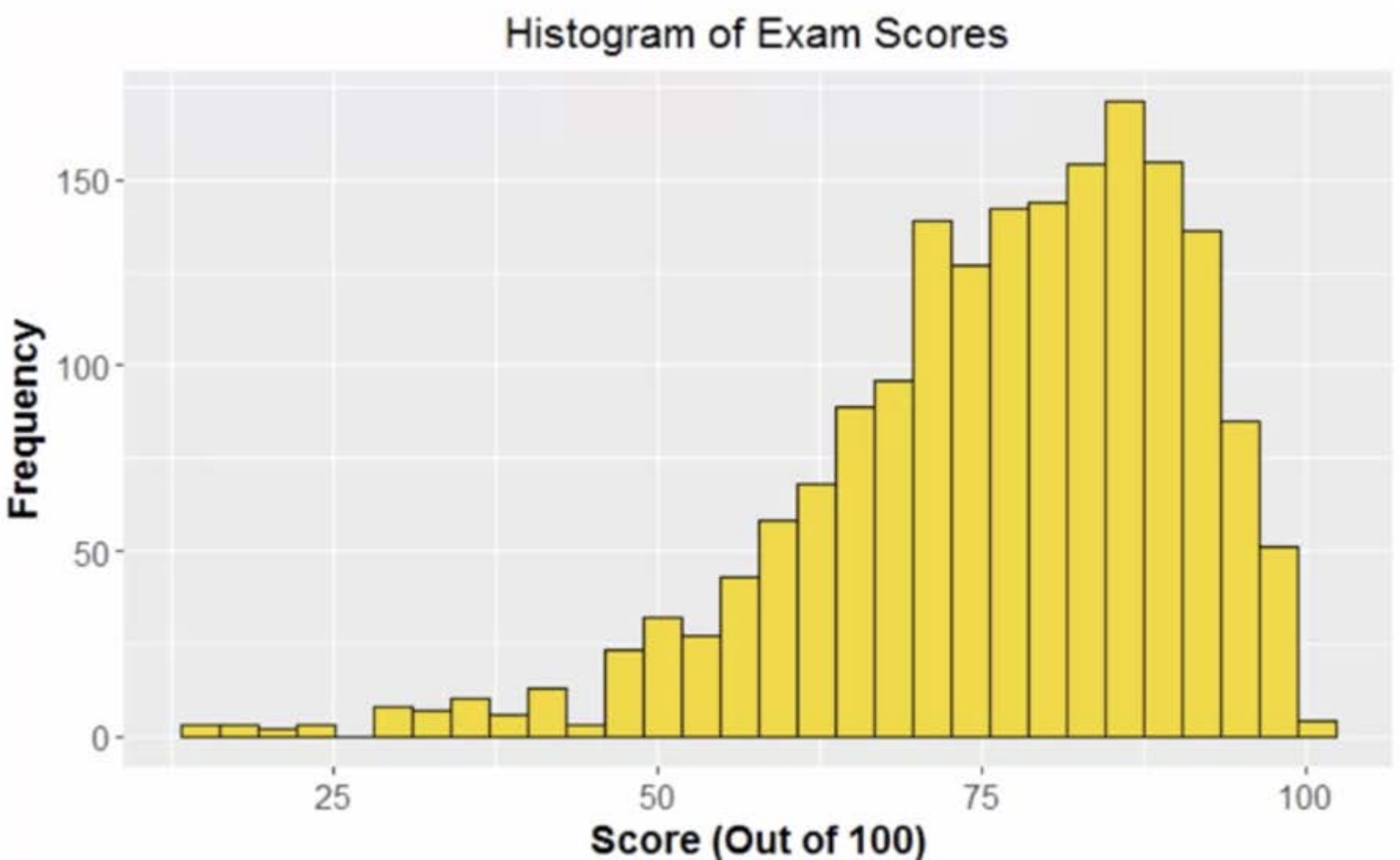
Min.	25%	50%	75%	Max.	Mean	SD	n
-618.1	36169	71427	105839	567595	74768	50517	148654

IQR

But that's always something that comes about when you're doing these numerical summaries,

Mean versus Median

Since the distribution of exam scores is left skewed with a number of low outliers,



the mean is expected to be:

- **less than the median.**

Correct
The mean is sensitive to extreme observations and the few unusually low scores will tend to bring the mean a bit lower than the median.

- **equal to the median**
- **greater than the median.**

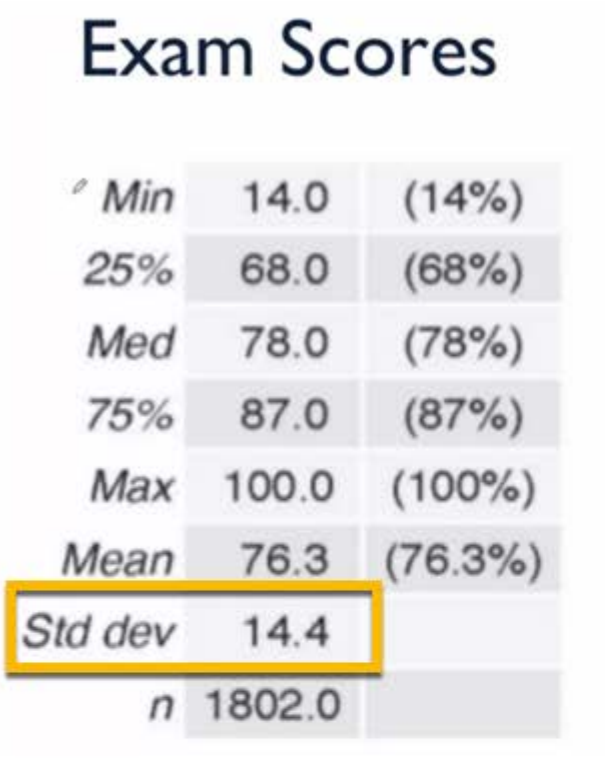
Exam Scores

mean < median

Min	14.0	(14%)
25%	68.0	(68%)
Med	78.0	(78%)
75%	87.0	(87%)
Max	100.0	(100%)
Mean	76.3	(76.3%)
Std dev	14.4	
n	1802.0	

meaning it's not influenced by outliers.

The standard deviation for the exam scores is 14.4 points.



Exam Scores

Min	14.0	(14%)
25%	68.0	(68%)
Med	78.0	(78%)
75%	87.0	(87%)
Max	100.0	(100%)
Mean	76.3	(76.3%)
Std dev	14.4	
n	1802.0	

Which of the following is an appropriate interpretation of this standard deviation value?

- **On average, the exam scores are about 14.4 points from the mean exam score.**

  Correct

  Correct: The standard deviation is *roughly* the *average distance* that the *values* are *from the mean*.

- **The exam scores are about 14.4 points from each other.**
- **The average exam score is about 14.4 points away from the mean exam score.**

Summary

Numerical Summaries (also called summary statistics) are used alongside our graphical representation of data to give a first impression of what our data looks like.

Depending on the software you are using you may get slightly different numerical summaries.

these numerical summaries on top of that allow for a lot more in depth analysis.