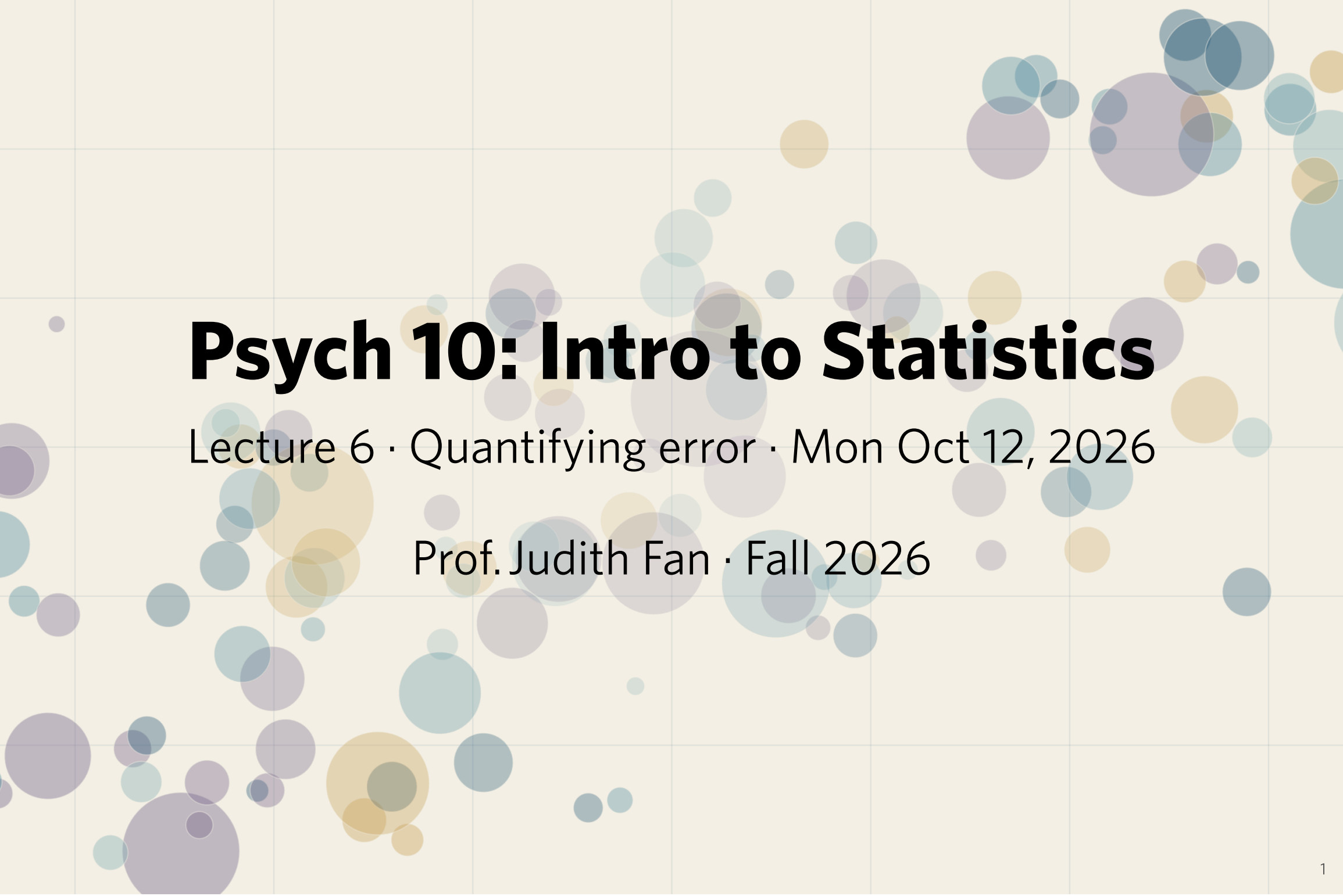




Psych 10: Intro to Statistics

Lecture 6 · Quantifying error · Mon Oct 12, 2026

Prof. Judith Fan · Fall 2026

Reminders & Announcements

- CourseKata **Chapter 6** is due this Friday.
- Section this week: **Milestone 3**, due Sunday, October 18 on Gradescope. (Milestone 2 was due last night.)
- Friday: **Case #3, The Forged Average**, in your tutorial session. Check your discussion section's Slack channel for your time slot. Bring a pen.
- **The exam** is in class on Wednesday, November 18 (Week 9). On Monday, November 9 (Week 8), we will give a practice exam in the same format and under the same conditions.

Answer keys are posted

- **Answer keys** for every lecture worksheet and tutorial handout so far are now on the course website, on the Resources page next to each handout.
- Please use them to check your own answers and to study for the exam!

Do you need to memorize formulas and R commands?

- No. The exam is not a test of whether you have memorized R commands or mathematical formulas. The teaching team will give you a glossary with whatever is essential for an exam question.
- The questions will focus on your critical thinking skills: reasoning flexibly about the concepts we have been teaching, and providing logically sound rationale for each of your answers. They will ask you to put together ideas from across several weeks, not to retrieve one concept at a time.
- So... knowing the material well enough that you don't have to look everything up from our glossary will make the exam easier, but committing things to memory is not a requirement.

What some exam questions will look like

- You will be provided with a scenario that resembles those that data scientists and social scientists working with quantitative data actually encounter. You'll be asked to reason about that scenario, answer questions about it, compute / estimate quantities based on the data provided, and provide interpretation.
- You might be shown a mixture of R code, words, and mathematical expressions.
- Sometimes you will be shown output from an AI tool and your job will be to scrutinize it: to find its weaknesses, gaps, and errors, and to say why they are errors.
- We will be giving you more practice with these activities in the coming weeks, in lecture and in tutorials.

OAE accommodations for the exam

- To take the exam (Wed Nov 18) at the Centralized Testing Center, you set it up yourself, ahead of time.
- Go to the OAE website, open *Student CTC Resources* (oae.stanford.edu/ctc/student-ctc-resources), and follow its steps in OAE|CONNECT.

Today

Quantifying error



Two routes, one average



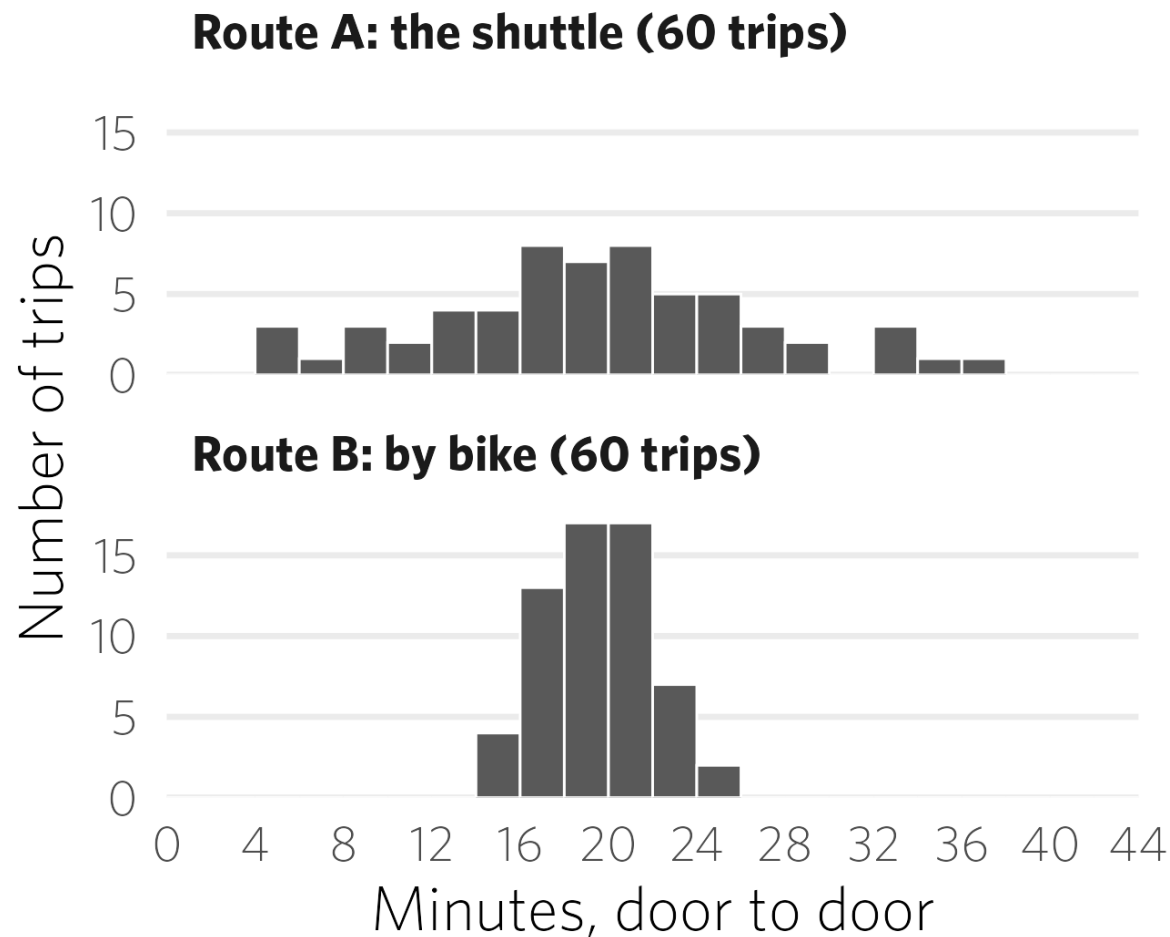
From deviations to the SD



What the SD is for

Last Wednesday we used the mean as a model, and treated each value's deviation from the mean as the model's error. Today's question is how to measure that error for a whole dataset: a single quantity that summarizes how far the values lie from the mean.

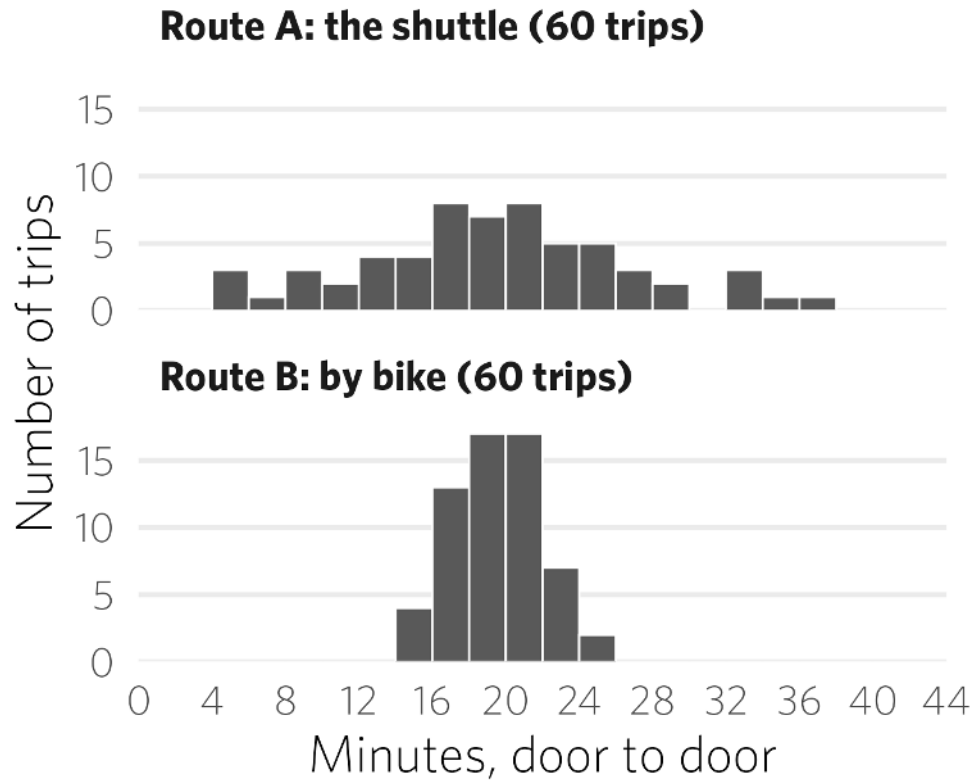
Worksheet - Part 1



Two routes to campus. The histograms show 60 recorded trips on each; both routes have a mean of 20 minutes, door to door.

- a.** You have an exam at 1:30 and you leave at 1:05. Which route do you take, and why?
- b.** Propose a number that measures how much more spread out Route A's trip times are than Route B's, and write down how you computed it.

Part 1a - Which route?

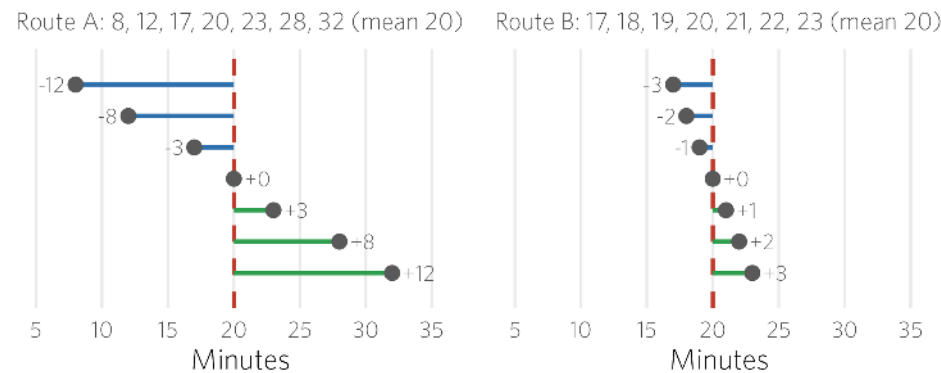


Route B. Its longest recorded trip is 25 minutes, so a 1:05 departure arrives by 1:30 on every recorded trip; Route A has several trips longer than 30 minutes.

The two routes have the same mean, so the mean cannot distinguish them; they differ in spread.

Part 1b - Why not average the deviations?

Last Wednesday's model offers a candidate for the number Part 1b asks for: each trip's deviation from the mean is the model's error, so average the errors. Seven trips from each route, with their deviations from the mean of 20:



$$(-12) + (-8) + (-3) + 0 + 3 + 8 + 12 = 0$$

$$(-3) + (-2) + (-1) + 0 + 1 + 2 + 3 = 0$$

Deviations from the mean sum to zero for any dataset, not only these two: the negative and the positive deviations balance exactly. So the average deviation is zero for every route, and cannot measure spread.

Part 1b - What did you compute?

A measure of spread has to do something other than average the deviations. **What did you compute, and what did it come to?**

Some measures you might have used:

- the **range**: 32 minutes against 9, though a single trip at either end determines it
- the **interquartile range**, from last Monday, which ignores the ends entirely
- the **average distance** from the mean, with every distance counted as positive, so that the early and the late trips no longer cancel

Today

Quantifying error

1

Two routes, one average

2

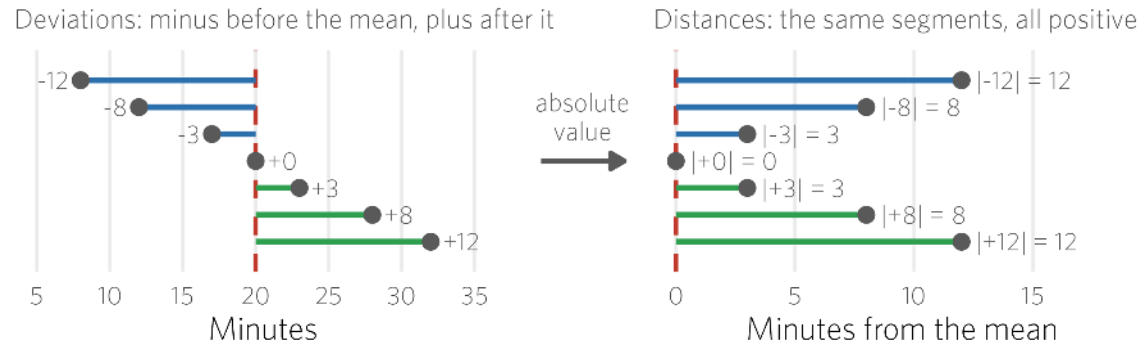
From deviations to the SD

3

What the SD is for

Both routes have a mean of 20 minutes, so the mean cannot distinguish them. The question for this section is how to measure, in minutes, how far a route's trips typically fall from its mean: a single number for spread that can be compared across routes.

How far from the mean, on average?



Route A: $\frac{12 + 8 + 3 + 0 + 3 + 8 + 12}{7} = \frac{46}{7} = 6.6 \text{ min}$

Route B: $\frac{3 + 2 + 1 + 0 + 1 + 2 + 3}{7} = \frac{12}{7} = 1.7 \text{ min}$

- Take each deviation's absolute value: -12 becomes 12, so early and late trips no longer cancel.
- Add the seven distances and divide by the number of trips: the average distance from the mean, in minutes.
- This is the **mean absolute deviation**, a legitimate measure of spread.

Why not stop at the mean absolute deviation?

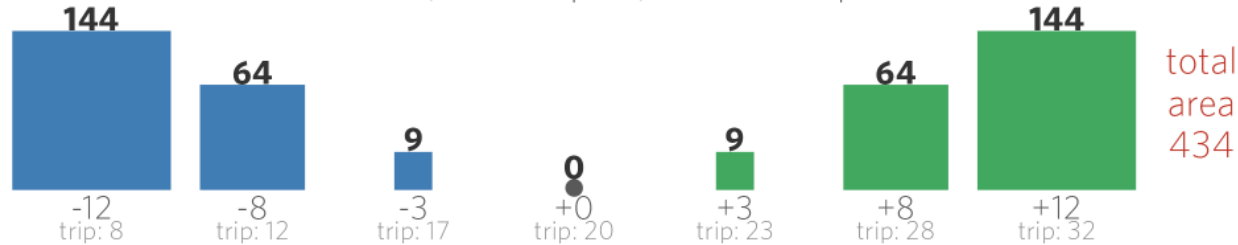
Early in the twentieth century statisticians argued over the mean absolute deviation against a measure built from squared deviations, and squares won, for two reasons that matter for the rest of this course.

The measure that won: squared deviations

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

$(Y_i - \bar{Y})^2$: one trip's deviation from the mean, squared. Y_i is the i -th of n trips, $\bar{Y}$ their mean (our model), and the deviation is the ERROR in DATA = MODEL + ERROR.

Route A: each deviation from 20, and its square, drawn as a square



$$(-12)^2 = 144 \quad (-8)^2 = 64 \quad (-3)^2 = 9 \quad 0^2 = 0 \quad 3^2 = 9 \quad 8^2 = 64 \quad 12^2 = 144$$

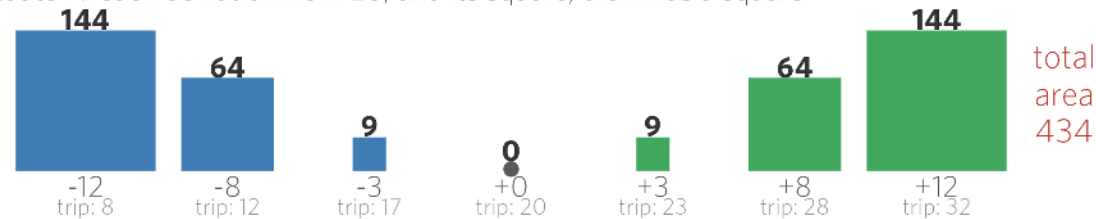
- Each square is one trip's squared deviation, in trip order: -12 gives an area of 144 in blue (before the mean), +12 the same area in green (after it).
- Squaring makes every deviation positive, as the absolute value did, and weights large deviations more heavily: 12 minutes gives 144, 3 minutes gives 9.

The model's total error: the sum of squares

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

$\sum (Y_i - \bar{Y})^2 = \text{SS}$: the squared deviations added up over all n trips.

Route A: each deviation from 20, and its square, drawn as a square

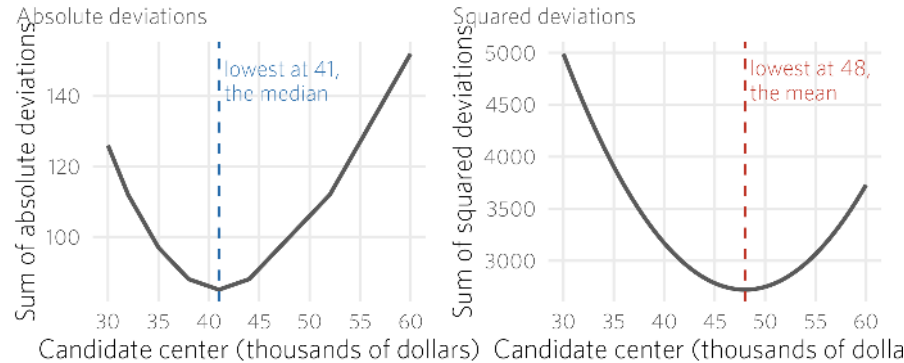


$$\begin{aligned} SS_{\text{Route A}} &= (-12)^2 + (-8)^2 + (-3)^2 + 0^2 + 3^2 + 8^2 + 12^2 \\ &= 144 + 64 + 9 + 0 + 9 + 64 + 144 = 434 \end{aligned}$$

$$SS_{\text{Route B}} = 9 + 4 + 1 + 0 + 1 + 4 + 9 = 28$$

- Each term is the area of one square above (blue before the mean, green after it); the total, in red, is the **sum of squares**, the model's total squared error, in squared minutes.

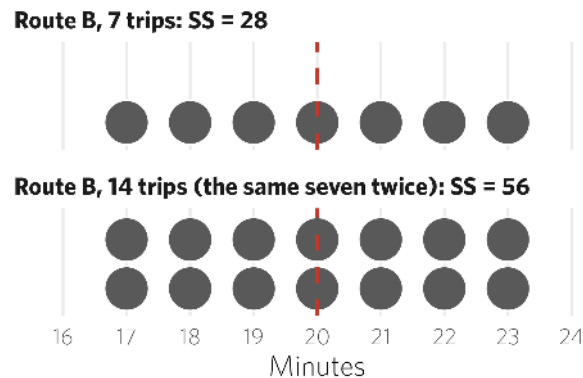
First reason: SS is smallest at the mean



- The **sum of absolute deviations** is smallest when measured from the **median** (41 for last Wednesday's seven salaries).
- The **sum of squared deviations** is smallest when measured from the **mean** (48).
- Each measure of error has its own best model: the median under absolute error, the mean under squared error.
- Our model is the mean, so its error is the sum of squares: the quantity the mean makes as small as possible.

Can SS compare samples of different sizes?

SS is a total over every trip recorded; Route B's 28 came from seven trips. Suppose fourteen had been recorded, with exactly the same spread: **which set has more spread, and which has the larger SS?**



Neither has more spread: the second set is the first set's seven values twice over, yet its SS is twice as large, 56 against 28. SS grows with every trip added, whatever the spread, so a total cannot compare samples of different sizes; a quantity per trip can.

Per trip, not in total: the variance

$$s = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad \text{SS} / (n - 1) = s^2: \text{the sum of squares per trip, the variance.}$$

$$s_{\text{Route A}}^2 = \frac{SS_{\text{Route A}}}{n - 1} = \frac{434}{6} = 72.3$$

$$s_{\text{Route B}}^2 = \frac{28}{6} = 4.67$$

$$s_{\text{Route B, 14 trips}}^2 = \frac{56}{13} = 4.31$$

- Divide the sum of squares by $n - 1$: a squared deviation per trip, the **sample variance**, still in squared minutes.
- Doubling the number of trips doubled SS and left the variance almost unchanged: the variance is a per-trip quantity.
- Why $n - 1$ rather than n : a sample of seven rarely includes a route's most extreme trips, so dividing by n would, on average, underestimate the variance of all its trips. Dividing by $n - 1$ (Bessel's correction) compensates, and the correction shrinks as n grows.

Back to minutes: the standard deviation

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

s, the square root of the variance: the standard deviation, back in minutes.

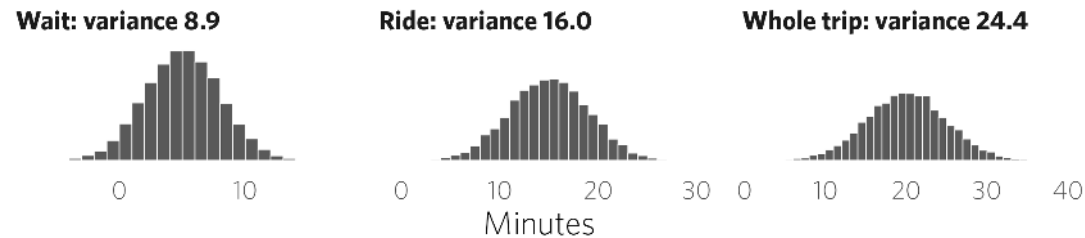
$$\hat{\sigma}_{\text{Route A}} = s_{\text{Route A}} = \sqrt{72.3} = 8.5 \text{ min}$$

$$\hat{\sigma}_{\text{Route B}} = s_{\text{Route B}} = \sqrt{4.67} = 2.2 \text{ min}$$

- The square root undoes the squaring, so the **standard deviation** is back in the original units, minutes.
- Read s as a typical magnitude of deviation from the mean: a Route A trip typically lies about 8 minutes from the mean of 20, in either direction.
- **Notation.** s is the SD of a sample, also written $\hat{\sigma}$, “sigma-hat”: an estimate of σ , the SD of all the trips the route will ever produce, which we never see. In the same way $\bar{Y}$, the sample mean, estimates μ , the population mean (some books write $\bar{x}$ for the sample mean). The hat means “our estimate of”.

Second reason: variances add up

A Route A trip is a wait for the shuttle plus the ride, and the two vary independently. Knowing each part's variance, do we know the variance of the whole trip?



variances: $8.9 + 16.0 = 24.9$ (trip: 24.4)

MADs: $2.4 + 3.2 = 5.6$ (trip: 3.9)

- With variances, yes: over 10,000 simulated trips, the wait's and the ride's add up to the whole trip's. With mean absolute deviations, no.
- So the variance of any sum of independent parts can be computed from the parts' variances. A sample mean is such a sum; in Week 6 this gives how much a sample mean varies from sample to sample.

Worksheet - Part 2

Seven trips from each route, in minutes. **Route A: 8, 12, 17, 20, 23, 28, 32. Route B: 17, 18, 19, 20, 21, 22, 23.** Both sets have a mean of 20.

a. Route A's row of the table is filled in. Complete Route B's row.

b. For Route A, the mean absolute deviation (MAD) is 6.6 minutes and the SD is 8.5 minutes; for Route B, 1.7 and 2.2 minutes. On both routes the SD exceeds the MAD. In a sentence or two, explain which step in the computation of the SD makes it exceed the MAD.

c. Now consider a third set of seven trips, 10, 20, 20, 20, 20, 20, 30, which also has a mean of 20. For Route B, the SD (2.2) is only slightly larger than the MAD (1.7); for the third set, the SD (5.8) is twice the MAD (2.9). How does the arrangement of the third set's trips around the mean differ from Route B's, and why does that difference widen the gap between its SD and its MAD?

Worksheet - Part 2, continued

- d.** Suppose every Route B trip had taken 5 minutes longer. Which of the mean, the SS, the variance, and the SD would change, and which would be unchanged?
- e.** A friend reports that on another route, the SD of 10 recorded trips is 4 minutes. What are the variance and the SS for those trips? Which of the three quantities is in minutes?

Part 2a - Route B's numbers

	deviations	squared deviations	SS	variance	SD
Route A	-12, -8, -3, 0, 3, 8, 12	144, 64, 9, 0, 9, 64, 144	434	$434/6 = 72.3$	8.5 min
Route B	-3, -2, -1, 0, 1, 2, 3	9, 4, 1, 0, 1, 4, 9	28	$28/6 = 4.67$	2.16 min

The same chain as Route A's, with every number smaller: deviations from -3 to +3, squares from 0 to 9, a sum of squares of 28, a variance of $28/6 = 4.67$ squared minutes, and a standard deviation of 2.2 minutes. The two SDs, 8.5 and 2.2, are the numbers Part 1 was looking for.

Part 2b - Why is the SD larger than the MAD?

Both measures start from the same seven distances from the mean. The MAD adds up the distances themselves; the SD adds up their squares.

Route A trip (min)	8	12	17	20	23	28	32	total	
distance from the mean	12	8	3	0	3	8	12	46	MAD = $46/7 = 6.6$ min
distance squared	144	64	9	0	9	64	144	434	SD = $\sqrt{434/6} = 8.5$ min

- Squaring makes the large deviations count for even more of the total. A deviation of 3 becomes 9, and a deviation of 12 becomes 144, so the two far trips are half of the MAD's total but two thirds of the SD's total.
- The square root at the end puts the result back in minutes but does not undo the squaring, so the SD comes out larger than the MAD. Dividing by $n - 1$ rather than n adds a little more to the sample SD.

Part 2c · When does the SD pull away from the MAD?

- Route B's seven trips are spaced evenly from 17 to 23, and its SD, 2.2, is only a little above its MAD, 1.7.
- In the third set, five trips sit exactly at the mean and two sit 10 minutes out. For the MAD, those two count 10 each. For the SS they count 100 each, and they are the whole of the SS. So the SD runs to twice the MAD.
- The general pattern: the more of a distribution's spread that comes from a few far values rather than many near ones, the further the SD rises above the MAD. That is worth remembering whenever a distribution has a long tail.

Part 2d-e · Shifting the trips, and running the chain backward

d. $\bar{Y} : 20 \rightarrow 25$

SS, s^2, s : unchanged

e. $s^2 = 4^2 = 16$

$SS = s^2(n - 1) = 16 \times 9 = 144$

- **d.** If every Route B trip had taken 5 minutes longer, the mean would move to 25 and nothing else would move: each trip's deviation from the new mean is exactly what it was from the old one. The SD measures spread about the mean, wherever the mean sits.
- **e.** Your friend's route: an SD of 4 minutes from 10 trips. Squaring the SD gives the variance; multiplying by $n - 1$ gives the SS. The chain runs backward as easily as forward, and of the three numbers only the SD is in minutes.

Today

Quantifying error

1

Two routes, one average

2

From deviations to the SD

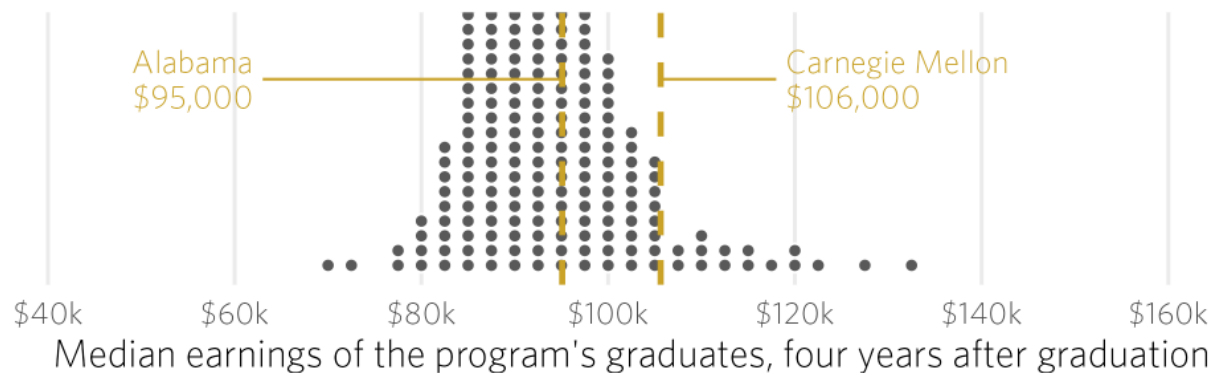
3

What the SD is for

What the Times published last Wednesday

A College-by-College Guide to Which Majors Pay Off, or Don't, by Arfa Momin and Ron Lieber, The New York Times, October 7, 2026. For every major at every college in the country, one number: the median earnings of the program's graduates four years after graduation. The charts show one major at a time, each dot one college's program.

Mechanical engineering: each dot is one college's program (318 programs)



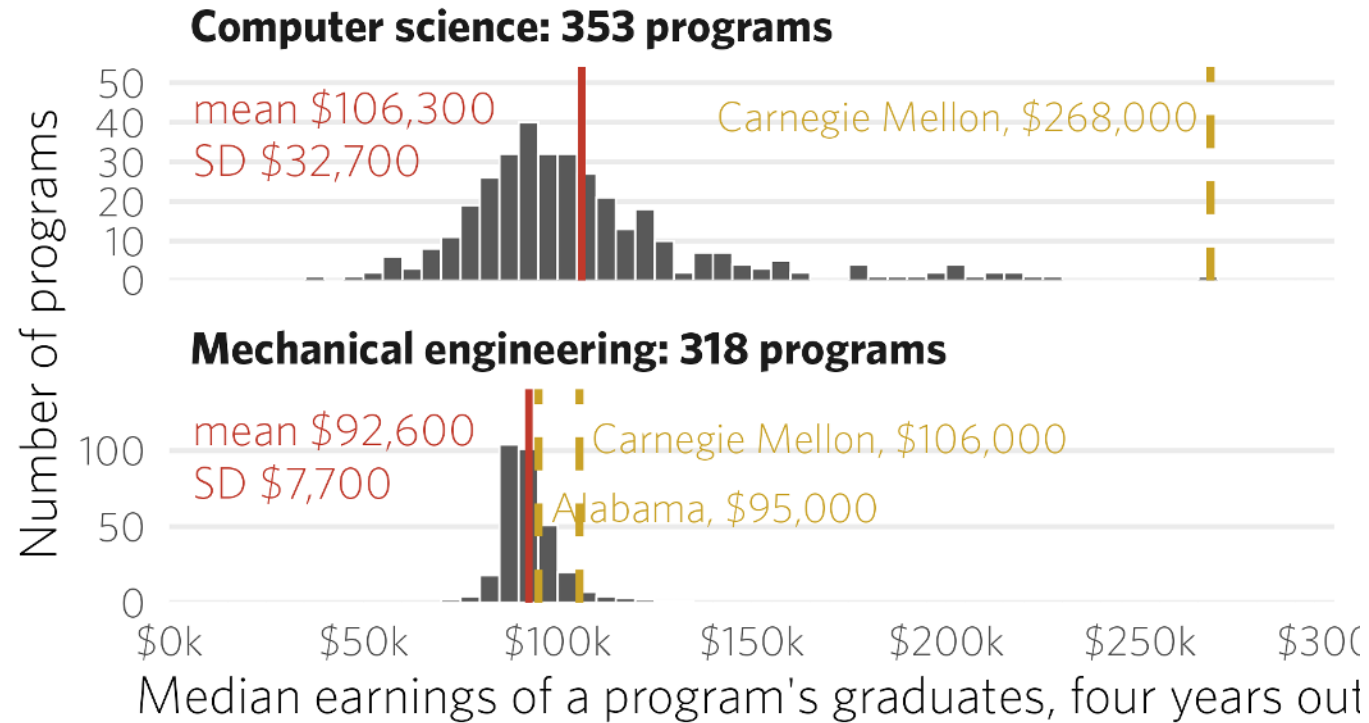
Most mechanical engineering programs sit within about \$10,000 of one another. The Times' own comparison: a mechanical engineer from Carnegie Mellon earned roughly \$10,000 more than one from the University of Alabama.

Does which college a student attends matter more for some majors than others?

In mechanical engineering, the college a student attends moves pay by about \$10,000. Computer science has a similar typical pay across colleges: a mean of \$106,300 across programs, against \$92,600 for mechanical engineering.

With your neighbor: in which of the two majors does the college you attend matter more for pay? That is, in which major do program earnings vary more from college to college?

Which major's pay varies more?



Computer science: SD \$32,700 across programs. Mechanical engineering: SD \$7,700. The college matters about four times as much for pay in computer science, and each SD states that spread as one number, in dollars.

Worksheet - Part 3

Part 3 returns to the decision from Part 1a, this time with each route's mean and SD. The two means are equal, so whatever separates the two departure times must come from the SDs.

Route A's trips have a mean of 20 minutes and an SD of about 8.5 minutes; Route B's have a mean of 20 minutes and an SD of about 2.2 minutes.

- a.** You have an exam at 1:30 and want to arrive on time on almost every trip. State a rule for when to leave, in terms of the mean and the SD, and give the departure time it yields for each route.
- b.** Suppose the trip times had been recorded in seconds rather than minutes. For Route A, each of the mean, the SS, the variance, and the SD is multiplied by some factor. Give the four factors.
- c.** One Route A trip took 32 minutes. How many SDs above the mean is that trip?

Part 3 · Leaving for the exam

a. $20 + 2(8.5) = 37 \text{ min}$

$$20 + 2(2.2) = 24.4 \text{ min}$$

b. $\bar{Y} \times 60, s \times 60$

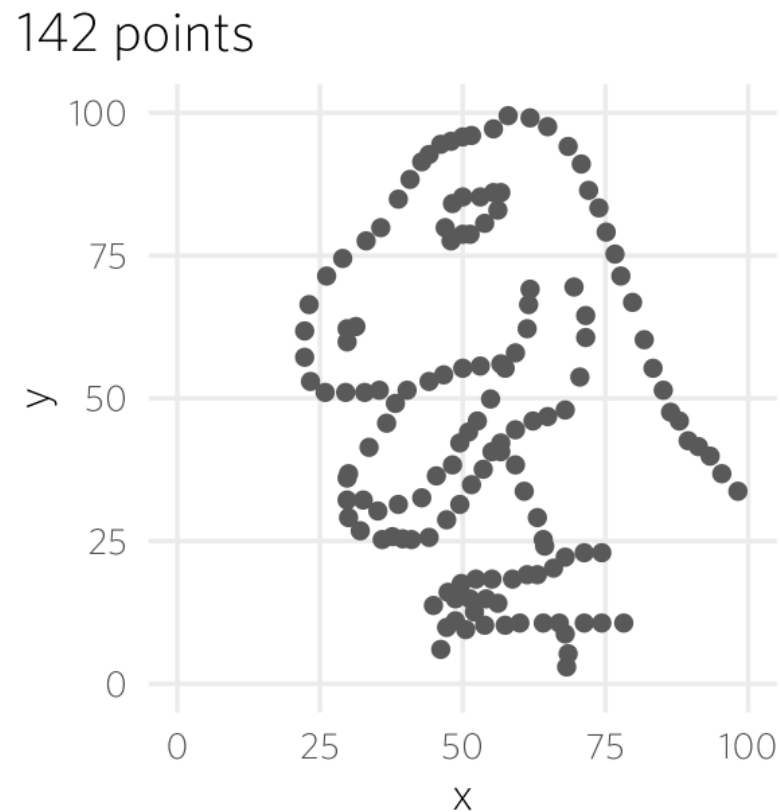
$$SS \times 60^2, s^2 \times 60^2 \quad (60^2 = 3600)$$

c. $\frac{32 - 20}{8.5} = 1.4$

- **a.** Both means are 20 minutes, so the mean alone says the same thing for both routes: leave at 1:10. The SD is what separates them. Allowing two SDs means 37 minutes on Route A, leaving at 12:53, and 24.4 minutes on Route B, leaving at 1:05; one SD gives 28.5 and 22.2 minutes. Any rule stated in terms of the mean and the SD is a reasonable answer; how many SDs to allow is a choice about how much risk of being late to accept.
- **b.** Recording in seconds multiplies every trip by 60. The mean and the SD are in minutes, so each is multiplied by 60; the SS and the variance are in squared minutes, so each is multiplied by 60 squared.
- **c.** A 32-minute trip is 12 minutes above the mean: 1.4 SDs. On Wednesday that number gets a name, the z-score, and most of the lecture is built on it.

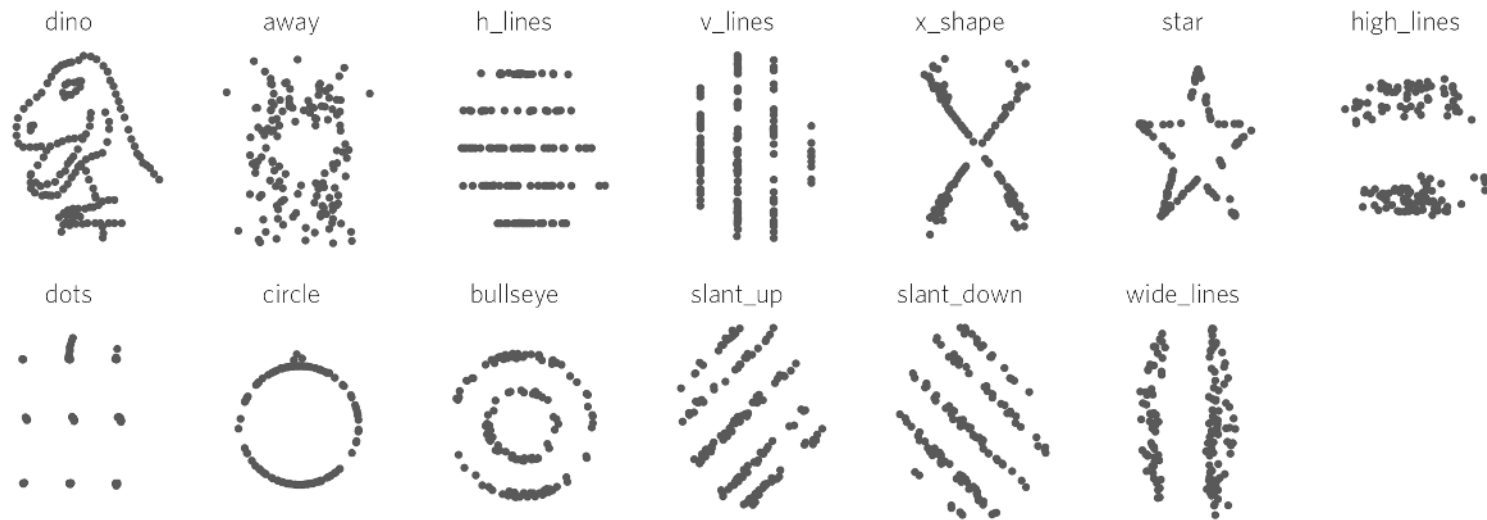
What the mean and the SD leave out

A mean and an SD summarize a distribution's center and spread. Here are 142 points: mean of x 54.3, SD 16.8; mean of y 47.8, SD 26.9.



Thirteen datasets with the same four numbers

Thirteen datasets of 142 points, the same means and SDs to one decimal in every one



All thirteen have the same means and SDs to one decimal. The mean and the SD carry no information about shape, so a distribution has to be examined, not only summarized.

Before Friday

- Finish CourseKata **Chapter 6**.
- Friday: **The Forged Average**. Check your discussion section's Slack channel for your time slot.
- Milestone 3 is due Sunday.

Hand in your worksheet on the way out

On the last line, write down one thing you learned today.

Thanks!