

# **Psych 10: Intro to Statistics**

Lecture 3 · Distributions & histograms · Wed Sep 30, 2026

Prof. Judith Fan · Fall 2026

# Reminders & Announcements

- CourseKata **Chapter 3** due Friday.
- Your time slot for Friday's tutorial should be posted in your discussion section's Slack channel by now. From now on, check that channel each week to find out your time slot.

# Today

## Distributions & histograms

1. Why does data visualization matter?
2. What's missing from the data?
3. The histogram: from a column of numbers to a distribution
4. What produces the shape of a distribution?

# **1 Why does data visualization matter?**

*The Stanford Daily*, January 29, 1986 (archives.stanforddaily.com)

|                    |                  |                 |                   |                  |                   |                  |
|--------------------|------------------|-----------------|-------------------|------------------|-------------------|------------------|
| Francis R. Seaborn | Michael J. Smith | Judith A. Evers | Lillian S. Orsini | Harold P. McNeil | Gregory B. Jarvis | Chris A. McAdams |
|--------------------|------------------|-----------------|-------------------|------------------|-------------------|------------------|

# Space Shuttle Challenger, January 28, 1986



STS-51-L crew: (front row) Michael J. Smith, Dick Scobee, Ronald McNair; (back row) Ellison Onizuka, Christa McAuliffe, Gregory Jarvis, Judith Resnik

# What does this have to do with data visualization?



Ice on the launch pad

The shuttle's two booster rockets were built in sections, with rubber rings (O-rings) sealing the joints between them. After every flight, the boosters were recovered from the ocean and the O-rings were inspected for damage.

Rubber gets stiff when it's cold, and a stiff ring doesn't seal well. Launch morning was forecast to be below freezing, so the engineers at Morton Thiokol, the company that built the boosters, worried that the O-rings would not seal.

*Tufte, Visual Explanations (1997)*

# A major cause: ineffective presentation of data

The night before the launch, the Thiokol engineers spent the evening on a conference call with NASA, arguing to postpone. They presented 13 charts to make their case, and NASA decided to launch anyway. The next morning, an O-ring in the right booster failed to seal.

BLOW BY HISTORY  
SRM-15 WORST BLOW-BY  
◦ 2 CASE JOINTS (80°), (110°) ARE  
◦ MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY  
◦ 2 CASE JOINTS (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A  
◦ NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES  
(DEGREES - F)

| MOTOR  | MGT  | AMB | O-RING   | WIND             |
|--------|------|-----|----------|------------------|
| DM-4   | 68   | 36  | 47       | 10 MPH           |
| DM-2   | 76   | 45  | 52       | 10 MPH           |
| QM-3   | 72.5 | 40  | 48       | 10 MPH           |
| QM-4   | 76   | 48  | 51       | 10 MPH           |
| SRM-15 | 52   | 64  | 53       | 10 MPH           |
| SRM-22 | 77   | 78  | 75       | 10 MPH           |
| SRM-25 | 55   | 26  | 29<br>27 | 10 MPH<br>25 MPH |

# A more effective summary of the data

| Flight | Date     | Temperature<br>°F | Erosion<br>incidents | Blow-by<br>incidents | Damage<br>index | Comments                                                |
|--------|----------|-------------------|----------------------|----------------------|-----------------|---------------------------------------------------------|
| 51-C   | 01.24.85 | 53°               | 3                    | 2                    | 11              | Most erosion any flight; blow-by; back-up rings heated. |
| 41-B   | 02.03.84 | 57°               | 1                    |                      | 4               | Deep, extensive erosion.                                |
| 61-C   | 01.12.86 | 58°               | 1                    |                      | 4               | O-ring erosion on launch two weeks before Challenger.   |
| 41-C   | 04.06.84 | 63°               | 1                    |                      | 2               | O-rings showed signs of heating, but no damage.         |
| 1      | 04.12.81 | 66°               |                      |                      | 0               | Cooltest (66°) launch without O-ring problems.          |
| 6      | 04.04.83 | 67°               |                      |                      | 0               |                                                         |
| 51-A   | 11.08.84 | 67°               |                      |                      | 0               |                                                         |
| 51-D   | 04.12.85 | 67°               |                      |                      | 0               |                                                         |
| 5      | 11.11.82 | 68°               |                      |                      | 0               |                                                         |
| 3      | 03.22.82 | 69°               |                      |                      | 0               |                                                         |
| 2      | 11.12.81 | 70°               | 1                    |                      | 4               | Extent of erosion not fully known.                      |
| 9      | 11.28.83 | 70°               |                      |                      | 0               |                                                         |
| 41-D   | 08.30.84 | 70°               | 1                    |                      | 4               |                                                         |
| 51-G   | 06.17.85 | 70°               |                      |                      | 0               |                                                         |
| 7      | 06.18.83 | 72°               |                      |                      | 0               |                                                         |
| 8      | 08.30.83 | 73°               |                      |                      | 0               |                                                         |
| 51-B   | 04.29.85 | 75°               |                      |                      | 0               |                                                         |
| 61-A   | 10.30.85 | 75°               |                      | 2                    | 4               | No erosion. Soot found behind two primary O-rings.      |
| 51-I   | 08.27.85 | 76°               |                      |                      | 0               |                                                         |
| 61-B   | 11.26.85 | 76°               |                      |                      | 0               |                                                         |
| 41-G   | 10.05.84 | 78°               |                      |                      | 0               |                                                         |
| 51-J   | 10.03.85 | 79°               |                      |                      | 0               |                                                         |
| 4      | 06.27.82 | 80°               |                      |                      | ?               | O-ring condition unknown; rocket casing lost at sea.    |
| 51-F   | 07.29.85 | 81°               |                      |                      | 0               |                                                         |

The 24 launches before Challenger, one per row, sorted from coldest to warmest (Tufte, *Visual Explanations*, 1997). The **damage index** ranges from 0 (no damage) to 11 (the most damage on any launch).

# Worksheet - Part 1

The table on the screen lists every shuttle launch before Challenger, with the temperature at launch and how much damage the O-rings showed afterward. The forecast for launch morning was 26–29°F. **Sketch the graph you would have drawn** to make the case for postponing. Label both axes.

| Flight | Date     | Temperature<br>°F | Erosion<br>incidents | Blow-by<br>incidents | Damage<br>index | Comments                                                |
|--------|----------|-------------------|----------------------|----------------------|-----------------|---------------------------------------------------------|
| 51-C   | 01.24.85 | 53°               | 3                    | 2                    | 11              | Most erosion any flight; blow-by; back-up rings heated. |
| 41-B   | 02.03.84 | 57°               | 1                    |                      | 4               | Deep, extensive erosion.                                |
| 61-C   | 01.12.86 | 58°               | 1                    |                      | 4               | O-ring erosion on launch two weeks before Challenger.   |
| 41-C   | 04.06.84 | 63°               | 1                    |                      | 2               | O-rings showed signs of heating, but no damage.         |
| 1      | 04.12.81 | 66°               |                      |                      | 0               | Cooltest (66°) launch without O-ring problems.          |
| 6      | 04.04.83 | 67°               |                      |                      | 0               |                                                         |
| 51-A   | 11.08.84 | 67°               |                      |                      | 0               |                                                         |
| 51-D   | 04.12.85 | 67°               |                      |                      | 0               |                                                         |
| 5      | 11.11.82 | 68°               |                      |                      | 0               |                                                         |
| 3      | 03.22.82 | 69°               |                      |                      | 0               |                                                         |
| 2      | 11.12.81 | 70°               | 1                    |                      | 4               | Extent of erosion not fully known.                      |
| 9      | 11.28.83 | 70°               |                      |                      | 0               |                                                         |
| 41-D   | 08.30.84 | 70°               | 1                    |                      | 4               |                                                         |
| 51-G   | 06.17.85 | 70°               |                      |                      | 0               |                                                         |
| 7      | 06.18.83 | 72°               |                      |                      | 0               |                                                         |
| 8      | 08.30.83 | 73°               |                      |                      | 0               |                                                         |
| 51-B   | 04.29.85 | 75°               |                      |                      | 0               |                                                         |
| 61-A   | 10.30.85 | 75°               |                      | 2                    | 4               | No erosion. Soot found behind two primary O-rings.      |
| 51-I   | 08.27.85 | 76°               |                      |                      | 0               |                                                         |
| 61-B   | 11.26.85 | 76°               |                      |                      | 0               |                                                         |
| 41-G   | 10.05.84 | 78°               |                      |                      | 0               |                                                         |
| 51-J   | 10.03.85 | 79°               |                      |                      | 0               |                                                         |
| 4      | 06.27.82 | 80°               |                      |                      | ?               | O-ring condition unknown; rocket casing lost at sea.    |
| 51-F   | 07.29.85 | 81°               |                      |                      | 0               |                                                         |

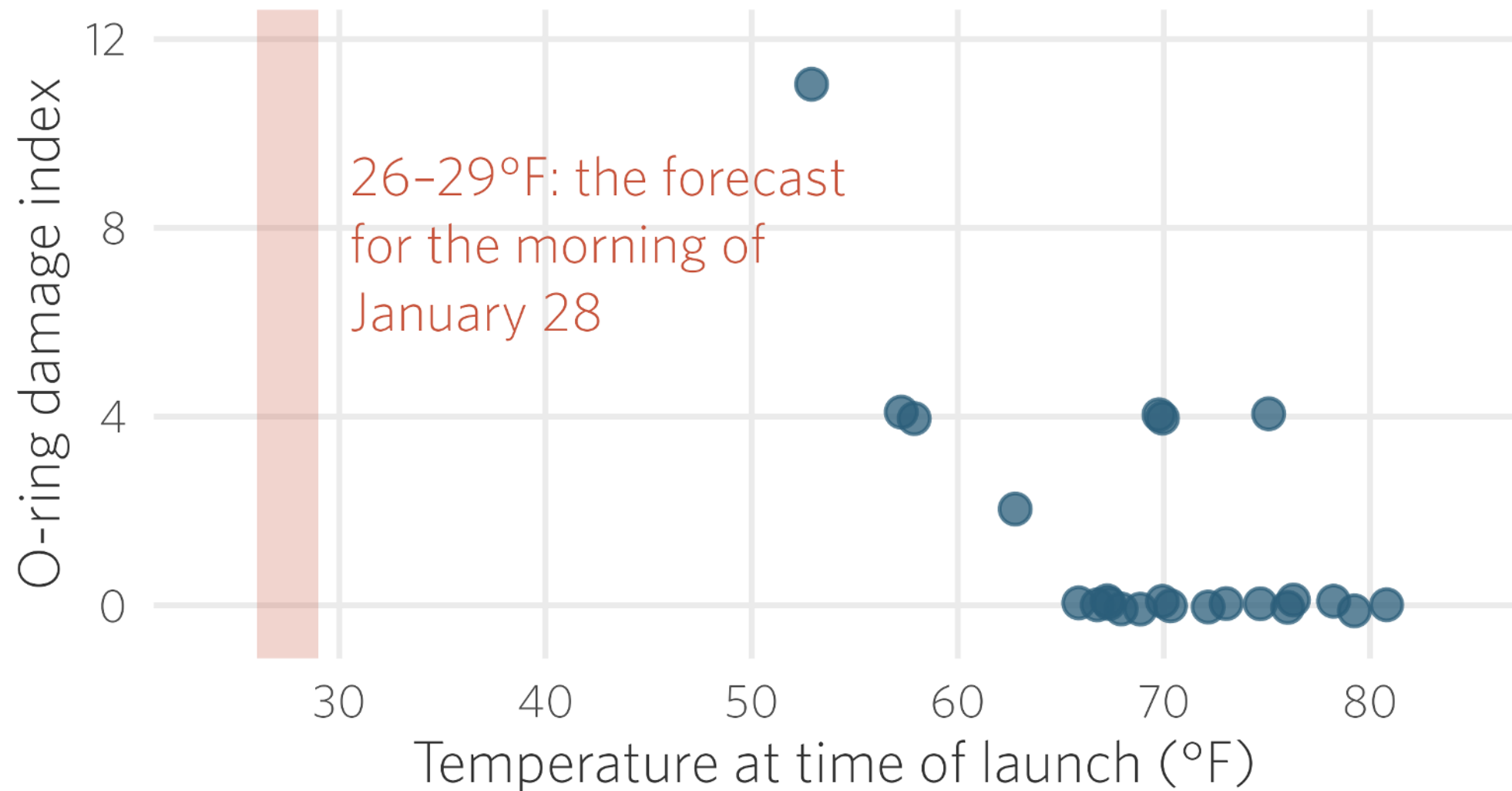
# The 24 launches before Challenger

| Flight | Date     | Temp. (°F) | Damage index |
|--------|----------|------------|--------------|
| 51-C   | 01.24.85 | 53         | 11           |
| 41-B   | 02.03.84 | 57         | 4            |
| 61-C   | 01.12.86 | 58         | 4            |
| 41-C   | 04.06.84 | 63         | 2            |
| 1      | 04.12.81 | 66         | 0            |
| 6      | 04.04.83 | 67         | 0            |
| 51-A   | 11.08.84 | 67         | 0            |
| 51-D   | 04.12.85 | 67         | 0            |
| 5      | 11.11.82 | 68         | 0            |
| 3      | 03.22.82 | 69         | 0            |
| 2      | 11.12.81 | 70         | 4            |
| 9      | 11.28.83 | 70         | 0            |

| Flight | Date     | Temp. (°F) | Damage index |
|--------|----------|------------|--------------|
| 41-D   | 08.30.84 | 70         | 4            |
| 51-G   | 06.17.85 | 70         | 0            |
| 7      | 06.18.83 | 72         | 0            |
| 8      | 08.30.83 | 73         | 0            |
| 51-B   | 04.29.85 | 75         | 0            |
| 61-A   | 10.30.85 | 75         | 4            |
| 51-I   | 08.27.85 | 76         | 0            |
| 61-B   | 11.26.85 | 76         | 0            |
| 41-G   | 10.05.84 | 78         | 0            |
| 51-J   | 10.03.85 | 79         | 0            |
| 4      | 06.27.82 | 80         | ?            |
| 51-F   | 07.29.85 | 81         | 0            |

Sorted from coldest to warmest: read down the left half, then the right half. Numbers from Tufte, *Visual Explanations* (1997).

# An even more effective visualization



Data: Tufte, *Visual Explanations* (1997)

# Takeaway

The engineers had all of these numbers, but none of their 13 charts showed damage against temperature for every launch.

In a single graph of damage against temperature, you can see that every launch below 65°F had a damage index above 0, and that the forecast was more than 20 degrees colder than any earlier launch. This is big idea #1 again: statistics gives us a way of making the invisible visible.

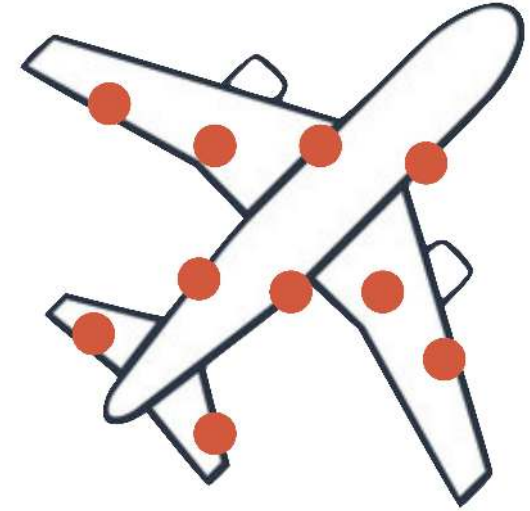
People use graphs for two purposes: **discovery** (what's in my data?) and **communication** (convincing someone else). The engineers needed a graph for communication.

## **2 What's missing from the data?**



# With your neighbor

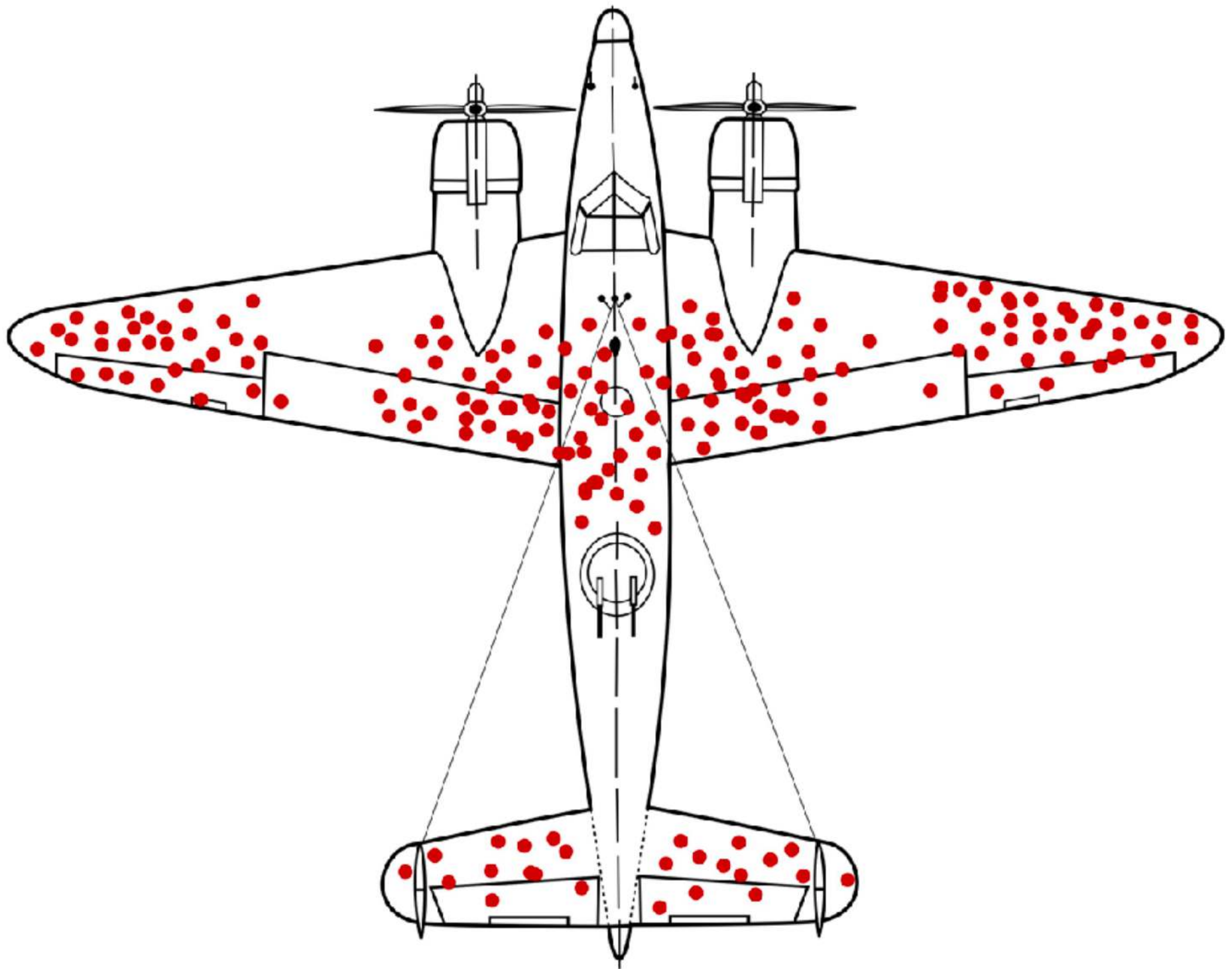
It's 1943. Planes come back from missions with bullet holes, and the red dots show where the holes are. Armor is heavy, so you can only add a little.



22

**Where would you tell the Navy to put the armor?**





# Takeaway

Planes that were hit in the engines mostly didn't come back, so their bullet holes are **missing data**. It can be really important to notice what is missing from a dataset and ask why.

- In a data frame, a missing value is written **NA**, which stands for “not available.” It isn't a zero, and it isn't the word “NA.”
- The planes that came back are a **sample** of all the planes that flew, but they aren't a random sample, because whether a plane came back depended on where it was hit. This is what we meant on Monday when we asked who ended up in the sample.

# **3 The histogram**

# With your neighbor

Suppose we asked 148 students how many minutes it takes them to get to campus.

|    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|
| 5  | 11 | 18 | 11 | 14 | 6  | 16 | 14 | 6  | 8  | 14 | 13 |
| 8  | 12 | 45 | 9  | 6  | 11 | 22 | 19 | 48 | 28 | 25 | 19 |
| 44 | 19 | 12 | 17 | 42 | 48 | 48 | 25 | 18 | 9  | 26 | 9  |
| 10 | 8  | 17 | 11 | 8  | 45 | 12 | 36 | 11 | 16 | 22 | 22 |
| 46 | 8  | 22 | 6  | 47 | 28 | 8  | 9  | 34 | 17 | 11 | 10 |
| 11 | 7  | 17 | 15 | 8  | 8  | 16 | 25 | 9  | 31 | 46 | 12 |
| 14 | 13 | 9  | 6  | 11 | 39 | 42 | 12 | 24 | 16 | 47 | 8  |
| 47 | 14 | 8  | 45 | 12 | 16 | 9  | 7  | 16 | 10 | 17 | 31 |
| 25 | 46 | 46 | 12 | 8  | 4  | 9  | 41 | 14 | 10 | 12 | 34 |
| 6  | 8  | 18 | 12 | 17 | 13 | 14 | 13 | 47 | 13 | 15 | 8  |
| 14 | 8  | 10 | 19 | 8  | 22 | 48 | 45 | 14 | 8  | 9  | 10 |
| 11 | 7  | 15 | 12 | 12 | 9  | 13 | 24 | 18 | 20 | 8  | 4  |
| 20 | 43 | 8  | 53 |    |    |    |    |    |    |    |    |

**In one or two sentences, how would you describe these commute times to someone who can't see the numbers?**



# What is a distribution?

The **distribution** of a variable is the answer to two questions: which values occurred, and how many times did each one occur?

For commute time, the distribution is how many students take 4 minutes, how many take 5, how many take 6, and so on up to 53.

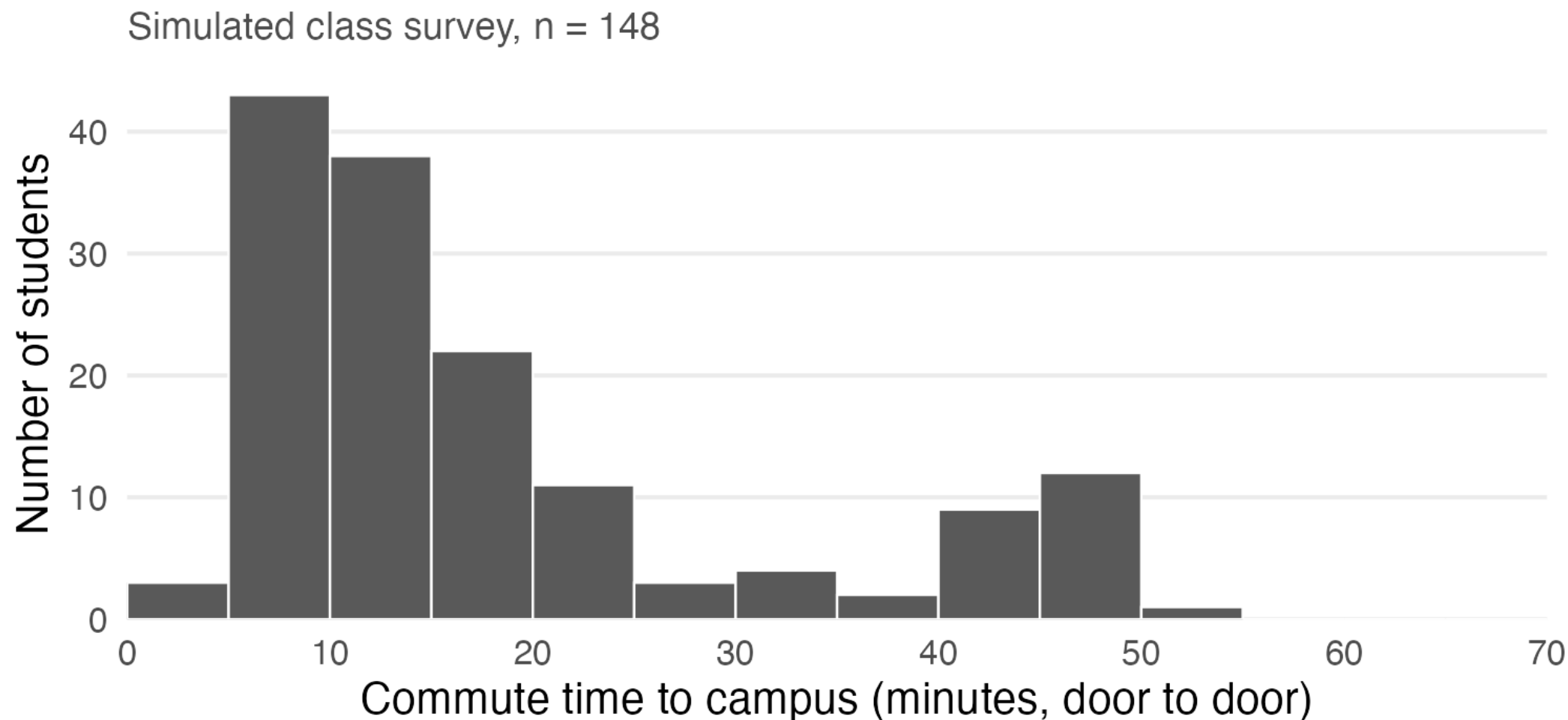
When you know the distribution, you can answer questions about the whole class, such as which commute times are most common, and how many students take longer than 30 minutes.

# How to make a histogram

A **histogram** is a graph of a distribution. To make one from a column of numbers:

1. **sort** the values,
2. divide the axis into equal-width **bins**,
3. **count** how many values fall in each bin,
4. draw one bar per bin, as tall as the count for that bin.

# Commutes to campus



# Worksheet · Part 2

- a.** Describe the **shape** of this distribution in one sentence (start with “roughly...”). Where is the bulk of the data?
- b.** About how many students have a commute of **40-45 minutes**? How wide are the bins? Does the bar at 40 include a 40-minute commute or a 45-minute one? (If you can't tell from the graph, say so.)

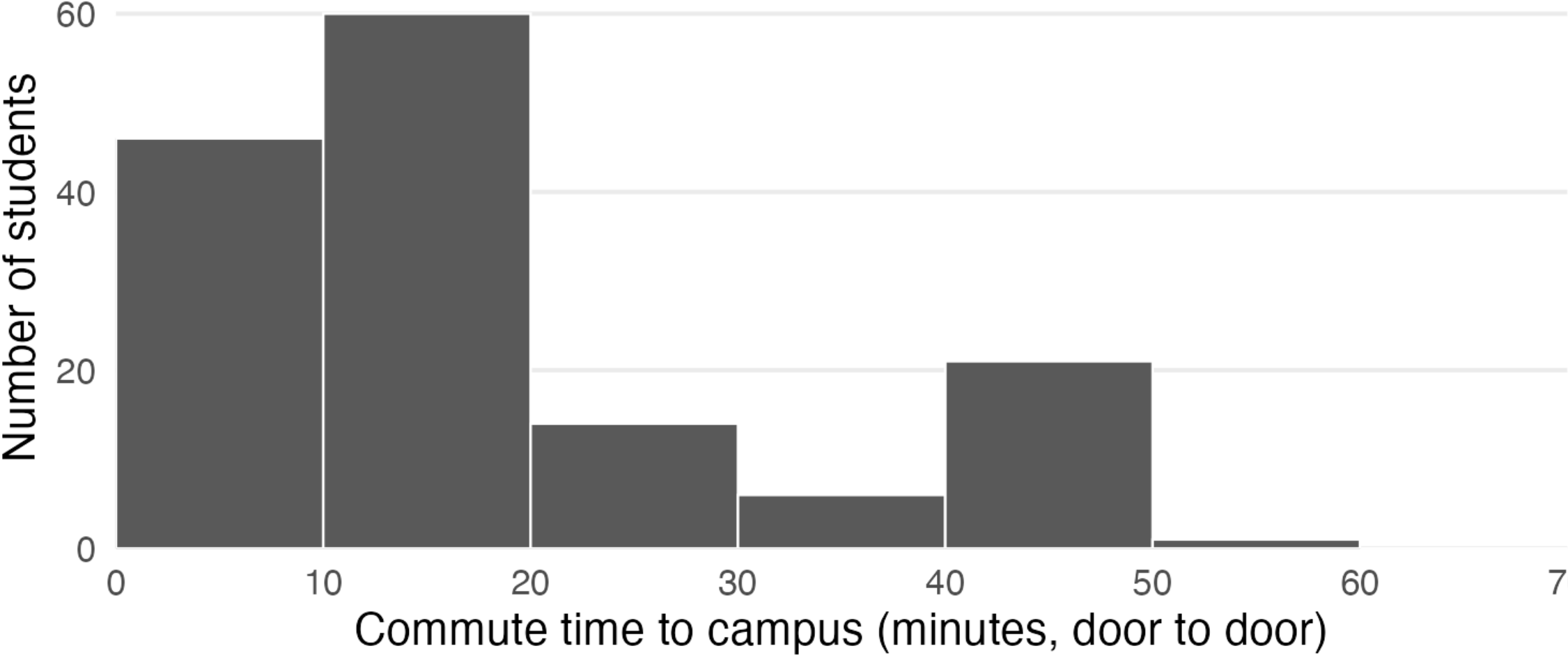
# How to read it

- **Axes:** commute time in minutes on the horizontal axis, number of students on the vertical axis. The bar over 10–15 reaches about 40, so about 40 students have a 10-to-15-minute commute.
- **Bins:** each bin is 5 minutes wide. Does a 15-minute commute go in the 10–15 bar or the 15–20 bar? You can't tell from the graph. (It depends on a setting in the software that drew it, and it's fine to say you don't know.)
- **Shape:** roughly **right-skewed**. Most of the values are on the left, and a long tail of larger values extends to the right. Skew is named for the side the tail is on, not the side the peak is on. There's also a second, smaller bump near 45.

# Worksheet - Part 2, continued

Same 148 students. **If we redraw the histogram with 10-minute bins instead of 5, what will it look like?** Sketch it. Is the second bump still there?

Same 148 students, 10-minute bins

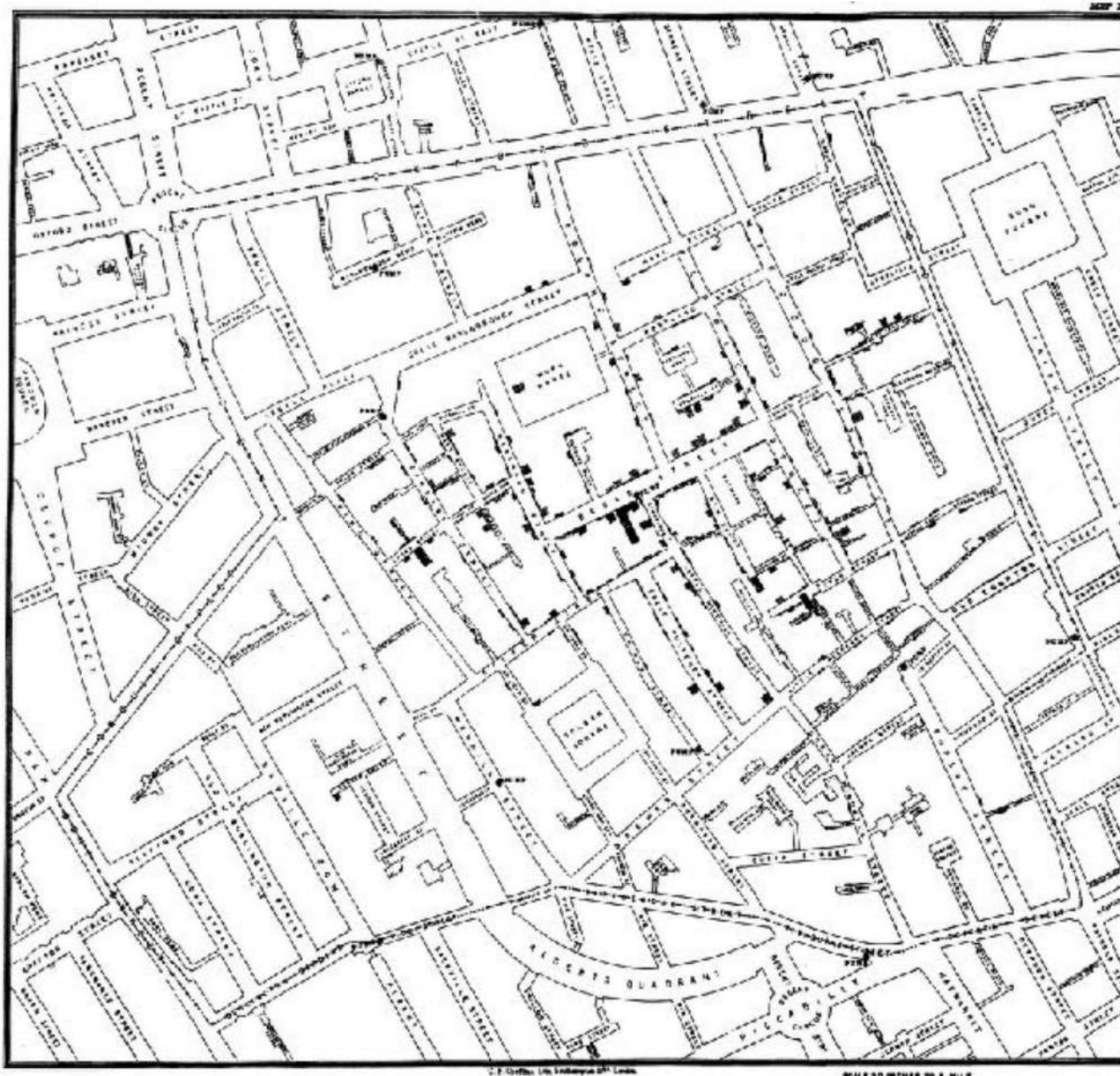


# Takeaway

- Bin width changes what you can see. If the bins are too wide, a small bump ends up inside one wide bar. If they're too narrow, each bar has only a few values in it, and the bar heights go up and down by chance.
- Our second bump was still there with 10-minute bins, so it isn't just a result of how we chose the bins. A bump that you see with one bin width and not with another may be there only by chance.
- With any new dataset, look at the distribution of every variable before you do anything else. This is how people use graphs for discovery.

# **4 What produces the shape of a distribution?**

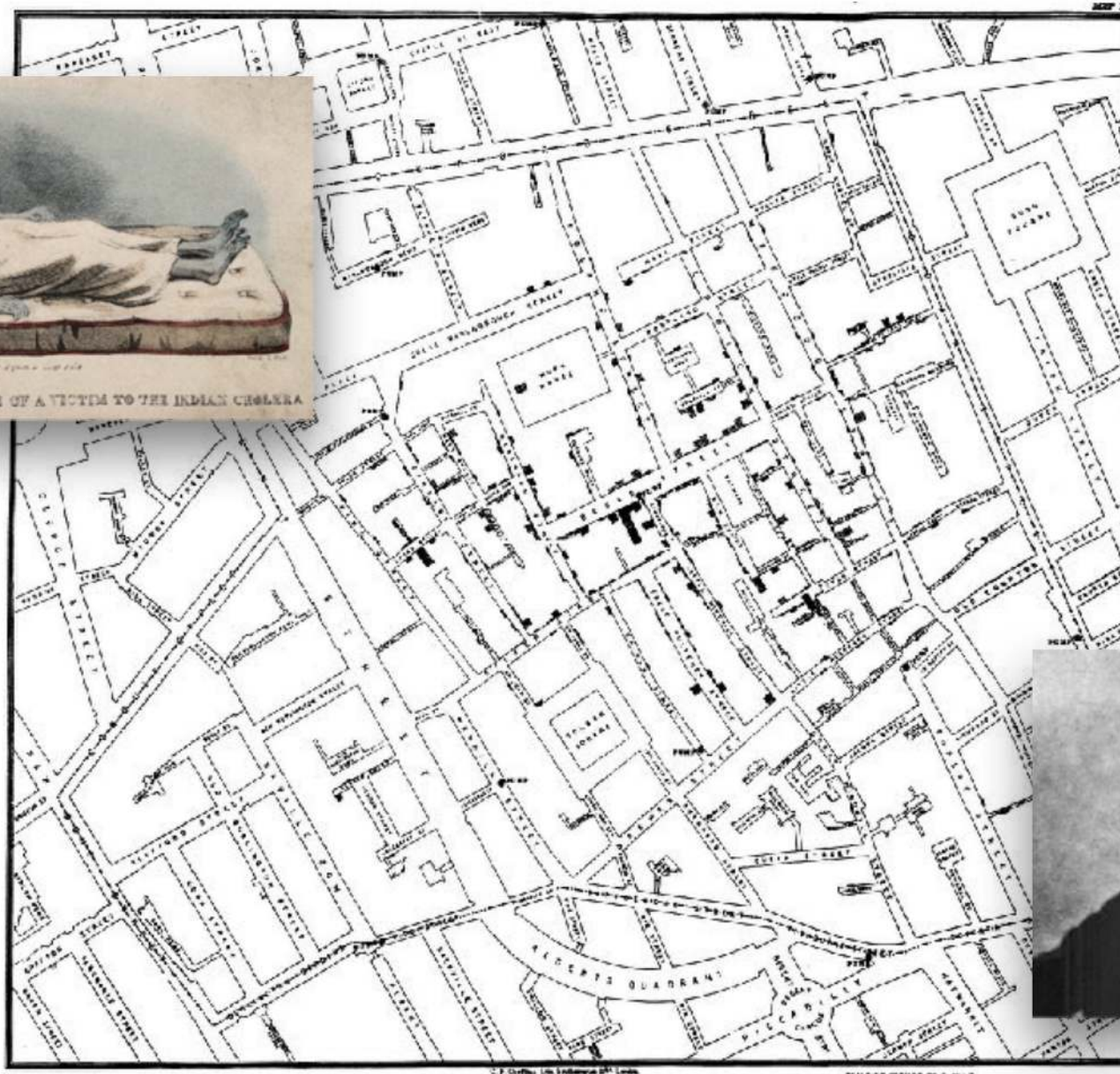
# What is this graph showing?



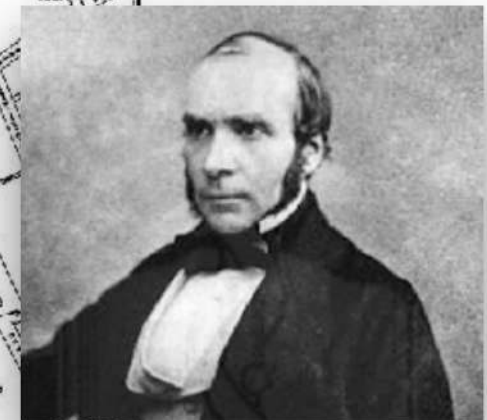
# What is this graph showing?



context:  
cholera outbreak  
in 19th c. London



John Snow



John Snow (1854)



histogram overlaid  
on London street map

# Worksheet - Part 3

Look again at the histogram. Besides the peak near 10–15 minutes, **there is a second bump around 45 minutes.**

- a.** List **three different processes** that could have produced that second bump. Make them different from one another, not variations on one explanation.
- b.** Circle the one you find most plausible.
- c.** Pick one of the others. What data, existing or new, would help you tell that process apart from the one you circled?

**Let's hear some**

# Before Friday

- Finish CourseKata **Chapter 3**.
- Friday Tutorial: **The Sleep Lab Incident**.

# **Hand in your worksheet on the way out**

On the last two lines, write down something new you learned today. Thanks!