

Math 1501 - Elements of Statistics with Applications

Numerical Summaries and the Shape of Data

selected topics from Chapter 1: 1.3, Chapter 2: 2.2, 2.5, 2.6

A note on section numbers

Two things are true about today, and both are on purpose.

A note on section numbers

Two things are true about today, and both are on purpose.

First: we're selecting topics from Chapter sections 2.1 and **2.3** which won't be covered entirely as live class content — the important elements of any *skipped* chapters will simply be absorbed. We will see important content from **2.3** returning in a later chapter appearing alongside *confidence intervals* where it will be formally introduced.

A note on section numbers

Two things are true about today, and both are on purpose.

First: we're selecting topics from Chapter sections 2.1 and **2.3** which won't be covered entirely as live class content — the important elements of any *skipped* chapters will simply be absorbed. We will see important content from **2.3** returning in a later chapter appearing alongside *confidence intervals* where it will be formally introduced.

Second: Today we will cover— 2.2, 2.5, 2.6 — (Histograms, Measures of Data, leading into Skewness.) Class 1-5's histogram explorations built the intuition for what you see today.

A note on section numbers

Two things are true about today, and both are on purpose.

First: we're selecting topics from Chapter sections 2.1 and **2.3** which won't be covered entirely as live class content — the important elements of any *skipped* chapters will simply be absorbed. We will see important content from **2.3** returning in a later chapter appearing alongside *confidence intervals* where it will be formally introduced.

Second: Today we will cover— 2.2, 2.5, 2.6 — (Histograms, Measures of Data, leading into Skewness.) Class 1-5's histogram explorations built the intuition for what you see today.

Section numbers are for reading the book. They're not the order we meet ideas in.

Section 1.4, in one slide

1.4 is reading-only in the textbook, but its real point fits in one slide: how you set up a survey, and what you do with the answer, matters: Ethics.

Section 1.4, in one slide

1.4 is reading-only in the textbook, but its real point fits in one slide: how you set up a survey, and what you do with the answer, matters: Ethics.

One simple numerical poll — “how many hours did you sleep?” — can already give you a distribution, a shape, and four different ways to summarize it.

Section 1.4, in one slide

1.4 is reading-only in the textbook, but its real point fits in one slide: how you set up a survey, and what you do with the answer, matters: Ethics.

One simple numerical poll — “how many hours did you sleep?” — can already give you a distribution, a shape, and four different ways to summarize it.

Add just one more piece of information — something categorical (age range, race, gender, ethnic background, zip code, income brackets) — and the same kind of data starts telling richer stories: who's affected, where patterns cluster, what changes across groups.

Section 1.4, in one slide

1.4 is reading-only in the textbook, but its real point fits in one slide: how you set up a survey, and what you do with the answer, matters: Ethics.

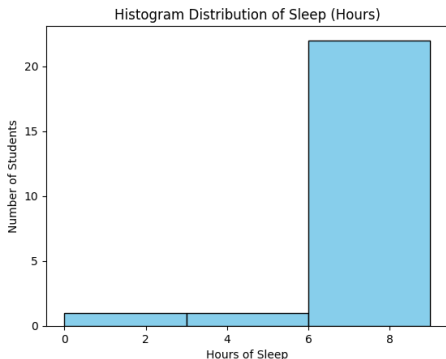
One simple numerical poll — “how many hours did you sleep?” — can already give you a distribution, a shape, and four different ways to summarize it.

Add just one more piece of information — something categorical (age range, race, gender, ethnic background, zip code, income brackets) — and the same kind of data starts telling richer stories: who's affected, where patterns cluster, what changes across groups.

That's the same toolkit people use to understand real communities — campus surveys, neighborhood needs assessments, anything where the first step is just: collect data carefully, ask the right questions. Then you analyze and hopefully tell truth from the conclusions. Be ethical - 1.4.

Chapter 2: Histograms: Three bins — barely a shape

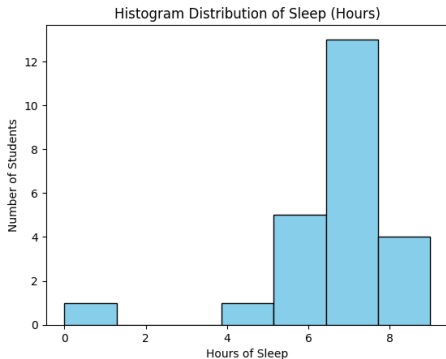
Same 24 per-night average sleep-hour responses from Day 1, split into just 3 bins.



Low, medium, high — that's all this tells you. Too coarse to see anything real.

Histograms: Seven bins — a pattern appears

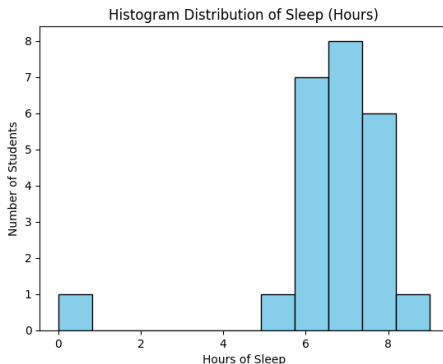
Same data, 7 bins.



Now the cluster around 6–8 hours shows up, and that lone 0 stands out as an actual outlier, not just noise.

Histograms: Eleven bins — this is the shape

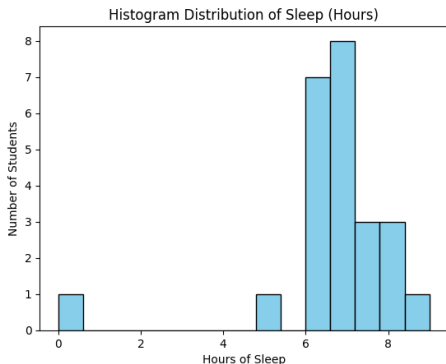
Same data, 11 bins.



This is the real distribution: a clear cluster around 6–8 hours, and that lone 0 still standing out as a genuine outlier — not buried, not fragmented into noise.

Histograms: Fifteen bins — too much noise

Same data, 15 bins — on only 24 data points.



Mostly single counts per bar now. The real shape is dissolving and becoming unclear.

Eleven bins had it right. This happens with too many bins.

What a histogram actually is

We just watched that shape change four times — here's the mechanism underneath it, with cleaner numbers.

A histogram is a graph with two axes.

The X axis is labeled by evenly spaced intervals.

The Y axis is labeled either by the numerical values observed in the data or by a relative frequency (a ratio – explained later).

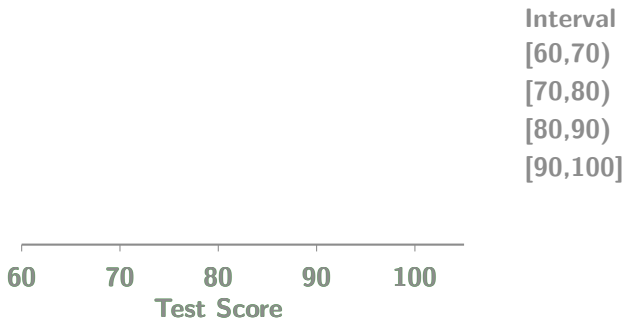
What a histogram actually is

We just watched that shape change four times — here's the mechanism underneath it, with cleaner numbers.

A histogram is a graph with two axes.

The X axis is labeled by evenly spaced intervals.

The Y axis is labeled either by the numerical values observed in the data or by a relative frequency (a ratio – explained later).



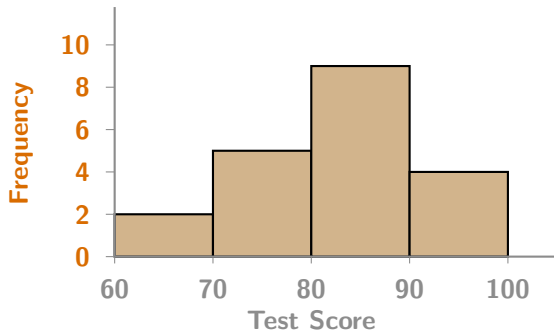
What a histogram actually is

We just watched that shape change four times — here's the mechanism underneath it, with cleaner numbers.

A histogram is a graph with two axes.

The X axis is labeled by evenly spaced intervals.

The Y axis is labeled either by **the numerical values observed in the data** or by a relative frequency (a ratio – explained later).



Interval	Freq
[60,70)	2
[70,80)	5
[80,90)	9
[90,100]	4
Total	20

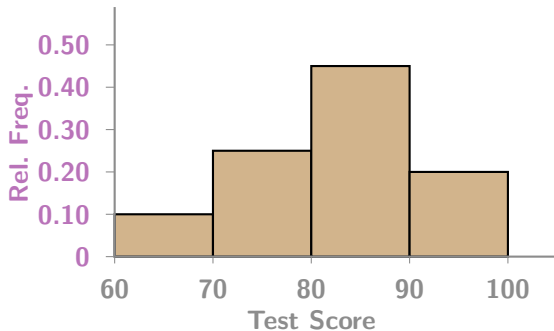
What a histogram actually is

We just watched that shape change four times — here's the mechanism underneath it, with cleaner numbers.

A histogram is a graph with two axes.

The X axis is labeled by evenly spaced intervals.

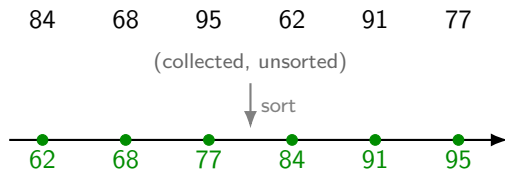
The Y axis is labeled either by the numerical values observed in the data or by **a relative frequency (a ratio – explained later)**.



Interval	Freq
[60,70)	2
[70,80)	5
[80,90)	9
[90,100]	4
Total	20

What a histogram actually is

We start by collecting numerical data, sort it from least to greatest, and the outcome is an ordered set of numbers which happen to form a number line.

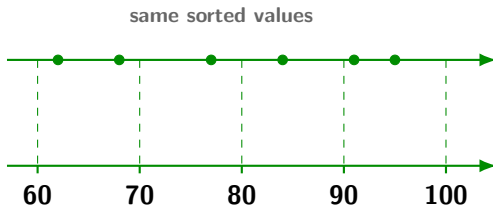


Interval	Freq
[60,70)	2
[70,80)	5
[80,90)	9
[90,100]	4
Total	20

What a histogram actually is

We start by collecting numerical data, sort it from least to greatest, and the outcome is an ordered set of numbers which happen to form a number line.

x-axis: the data values, in order — which is exactly why we sort first.



Interval	Freq
$[60,70)$	2
$[70,80)$	5
$[80,90)$	9
$[90,100]$	4
Total	20

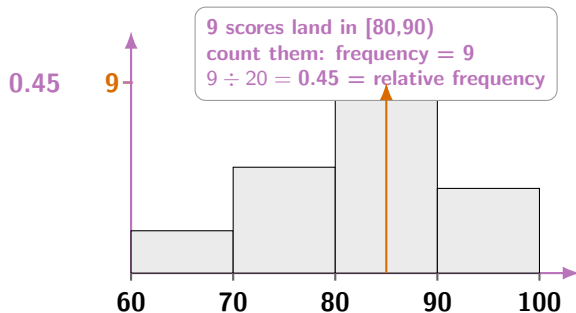
this row of numbers, in order, IS the x-axis

What a histogram actually is

We start by collecting numerical data, sort it from least to greatest, and the outcome is an ordered set of numbers which happen to form a number line.

x-axis: the data values, in order — which is exactly why we sort first.

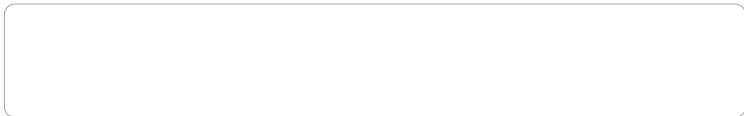
y-axis: how often each value (or range of values) shows up — frequency, or $\text{frequency} \div \text{total}$ for relative frequency.



Interval	Freq
[60,70)	2
[70,80)	9
[80,90)	9
[90,100]	4
Total	20

Sort this

9 15 3 9 12 7 9 4 6

A large, empty rectangular box with rounded corners, intended for the user to write the sorted sequence of numbers.

Ways to describe the center: Mean and Median

Once it's sorted, the first natural question: what's typical?

Mean — add every value, divide by how many there are. The arithmetic average, $\bar{x}$.

Ways to describe the center: Mean and Median

Once it's sorted, the first natural question: what's typical?

Mean — add every value, divide by how many there are. The arithmetic average, $\bar{x}$.

Median — the middle value of the sorted list — half the data is below it, half above.

Ways to describe the center: Mean and Median

Once it's sorted, the first natural question: what's typical?

Mean — add every value, divide by how many there are. The arithmetic average, $\bar{x}$.

Median — the middle value of the sorted list — half the data is below it, half above.

Two different ideas of “middle.” They don't have to agree, and when they don't, that disagreement tells you something — more on that soon.

Quick rounding rules

Before we compute anything: carry your final answer at least **one more decimal place** than the original data had *unless asked otherwise*.

Quick rounding rules

Before we compute anything: carry your final answer at least **one more decimal place** than the original data had *unless asked otherwise*.

Round only at the *end* of your set of calculations. Never round in the middle of a calculation.

Quick rounding rules

Before we compute anything: carry your final answer at least **one more decimal place** than the original data had *unless asked otherwise*.

Round only at the *end* of your set of calculations. Never round in the middle of a calculation.

Our sorted list is whole numbers, so any mean we compute from it gets reported to **one decimal place**.

Quick rounding rules

Before we compute anything: carry your final answer at least **one more decimal place** than the original data had *unless asked otherwise*.

Round only at the *end* of your set of calculations. Never round in the middle of a calculation.

Our sorted list is whole numbers, so any mean we compute from it gets reported to **one decimal place**.

Why not more? Measurements always carry some error — eight decimal places on whole-number data isn't precision, it's noise.

Ways to describe the center: Mean and Median cont.

So far: mean, median

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Ways to describe the center: Mean and Median cont.

So far: mean, median

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mean: $\frac{3 + 4 + 6 + 7 + 9 + 9 + 9 + 12 + 15}{9} = \frac{74}{9} = 8.2$

one more decimal place than the data — the rule from last frame.

Ways to describe the center: Mean and Median cont.

So far: mean, median

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mean:
$$\frac{3 + 4 + 6 + 7 + 9 + 9 + 9 + 12 + 15}{9} = \frac{74}{9} = 8.2$$

one more decimal place than the data — the rule from last frame.

Median: 9 values, so the middle one is the **5th**: 3, 4, 6, 7, 9, 9, 9, 12, 15
→ median = 9

Ways to describe the center: Mean and Median cont.

So far: mean, median

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mean:
$$\frac{3 + 4 + 6 + 7 + 9 + 9 + 9 + 12 + 15}{9} = \frac{74}{9} = 8.2$$

one more decimal place than the data — the rule from last frame.

Median: 9 values, so the middle one is the **5th**: 3, 4, 6, 7, 9, 9, 9, 12, 15
→ median = 9

Mean and median came out close, but not equal. Hold onto that.

Ways to describe the center: Mode and Range

So far: mean, median

Mode — the value that shows up most often. A dataset can have one, several, or no mode at all.

Ways to describe the center: Mode and Range

So far: mean, median

Mode — the value that shows up most often. A dataset can have one, several, or no mode at all.

Range — the largest value minus the smallest. One number for how spread out the whole dataset is.

Ways to describe the center: Mode and Range

So far: mean, median

Mode — the value that shows up most often. A dataset can have one, several, or no mode at all.

Range — the largest value minus the smallest. One number for how spread out the whole dataset is.

Mode looks at *repetition*. Range looks at the two extremes and ignores everything in between.

Ways to describe the center: Mode and Range cont.

So far: mean, median, mode, range

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Ways to describe the center: Mode and Range cont.

So far: mean, median, mode, range

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mode: 9 appears three times — more than any other value. Mode = 9.

Ways to describe the center: Mode and Range cont.

So far: mean, median, mode, range

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mode: 9 appears three times — more than any other value. Mode = 9.

Range: $15 - 3 = 12$.

Ways to describe the center: Mode and Range cont.

So far: mean, median, mode, range

Worked example

Sorted: 3, 4, 6, 7, 9, 9, 9, 12, 15

Mode: 9 appears three times — more than any other value. Mode = 9.

Range: $15 - 3 = 12$.

Four numbers, one dataset: mean 8.2, median 9, mode 9, range 12. Mean sits a little below median and mode — not an accident, and not something we've named yet.

Same four numbers, different pictures

What mean, median, and mode do to each other depends entirely on the shape of the histogram they came from.



Symmetric

4, 5, 6, 7, 7, 7, 8, 9, 10

$$\text{Mean} = \frac{63}{9} = 7.0$$

Median (5th of 9) = 7

Mode = 7 (3×)

mean = median = mode



Skewed left

3, 4, 6, 7, 9, 9, 9, 12, 15

$$\text{Mean} = \frac{74}{9} = 8.2$$

Median (5th of 9) = 9

Mode = 9 (3×)

mean < median = mode



Skewed right

3, 6, 9, 9, 9, 11, 12, 14, 15

$$\text{Mean} = \frac{88}{9} = 9.8$$

Median (5th of 9) = 9

Mode = 9 (3×)

mean > median = mode

The shape has a name too

So far: mean, median, mode, range

Skewed — not symmetrical — one side has a longer tail than the other.
Named for which direction the tail points.

The shape has a name too

So far: mean, median, mode, range

Skewed — not symmetrical — one side has a longer tail than the other.
Named for which direction the tail points.

Skewed **left**: tail points left, $\text{mean} < \text{median} < \text{mode}$.

The shape has a name too

So far: mean, median, mode, range

Skewed — not symmetrical — one side has a longer tail than the other.
Named for which direction the tail points.

Skewed **left**: tail points left, $\text{mean} < \text{median} < \text{mode}$.

Skewed **right**: tail points right, $\text{mean} > \text{median} > \text{mode}$.

The shape has a name too

So far: mean, median, mode, range

Skewed — not symmetrical — one side has a longer tail than the other.
Named for which direction the tail points.

Skewed **left**: tail points left, $\text{mean} < \text{median} < \text{mode}$.

Skewed **right**: tail points right, $\text{mean} > \text{median} > \text{mode}$.

Our worked example — mean 8.2, median 9, mode 9 — has exactly the left-skewed pattern. Nine data points already hinted at a shape.

Now you place them

Your turn

Same three shapes as before. For each one, mark where mean, median, and mode actually land — draw right on the bars.



Symmetric

Mean = 7 Median = 7
Mode = 7



Skewed left

Mean = 8.2 Median = 9
Mode = 9



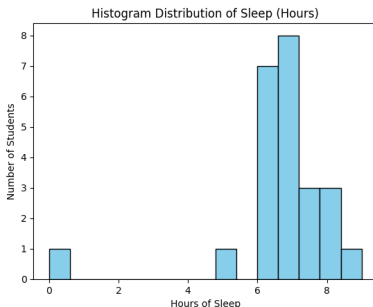
Skewed right

Mean = 9.8 Median = 9
Mode = 9

Values are given on purpose — the task is placement, not derivation.

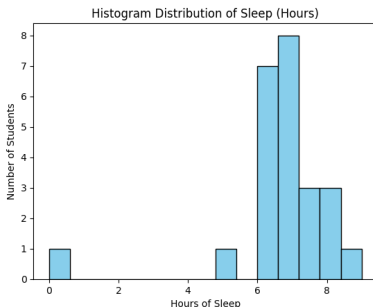
Back to our own data

Same real sleep-hour data we opened with — 15 bins this time. Recall what happened to that lone 0 as the bins get finer.



Back to our own data

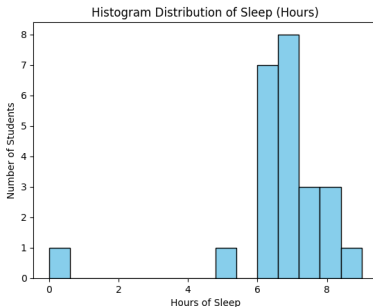
Same real sleep-hour data we opened with — 15 bins this time. Recall what happened to that lone 0 as the bins get finer.



It stands completely alone now — genuinely skewed left, with a real outlier, not just noise from too many bins.

Back to our own data

Same real sleep-hour data we opened with — 15 bins this time. Recall what happened to that lone 0 as the bins get finer.



It stands completely alone now — genuinely skewed left, with a real outlier, not just noise from too many bins.

Next class, we dig into what that 0 actually does to the numbers we compute from it.

Populations nest inside each other

Back in 1.2: a population is everyone you could possibly draw from. Every ball in the bag.

Source: U.S. Census Bureau, QuickFacts, 2025 population estimates (census.gov/quickfacts/NY;
census.gov/quickfacts/kingscountynewyork).

Populations nest inside each other

Back in 1.2: a population is everyone you could possibly draw from. Every ball in the bag.

This room is one. By CUNYFirst: 40 of you registered in the MATH 1501 Fall 2026 class — a population of its own, exactly like the bag.

Source: U.S. Census Bureau, QuickFacts, 2025 population estimates (census.gov/quickfacts/NY;
census.gov/quickfacts/kingscountynewyork).

Populations nest inside each other

Back in 1.2: a population is everyone you could possibly draw from. Every ball in the bag.

This room is one. By CUNYFirst: 40 of you registered in the MATH 1501 Fall 2026 class — a population of its own, exactly like the bag.

Zoom out, and it keeps being true. Brooklyn College sits inside Brooklyn — Kings County — about 2.65 million people. Brooklyn sits inside New York State — about 20.0 million.

Source: U.S. Census Bureau, QuickFacts, 2025 population estimates (census.gov/quickfacts/NY; census.gov/quickfacts/kingscountynewyork).

Populations nest inside each other

Back in 1.2: a population is everyone you could possibly draw from. Every ball in the bag.

This room is one. By CUNYFirst: 40 of you registered in the MATH 1501 Fall 2026 class — a population of its own, exactly like the bag.

Zoom out, and it keeps being true. Brooklyn College sits inside Brooklyn — Kings County — about 2.65 million people. Brooklyn sits inside New York State — about 20.0 million.

Every neighborhood, every campus, every classroom is a population nested inside a bigger one. Populations nest, at every scale — the definition never changes, only the size does depending on who/what you want to study.

Source: U.S. Census Bureau, QuickFacts, 2025 population estimates (census.gov/quickfacts/NY; census.gov/quickfacts/kingscountynewyork).

Mean, Median, Mode, Range — At Real Scale

Our sorted list had nine values. By hand, that took seconds.

Mean, Median, Mode, Range — At Real Scale

Our sorted list had nine values. By hand, that took seconds.

That doesn't survive real scale. Nobody sorts incomes, rents, or commute times for 2.65 million Brooklyn residents by hand — it isn't slow, it's simply not something a human does in a lifetime.

Mean, Median, Mode, Range — At Real Scale

Our sorted list had nine values. By hand, that took seconds.

That doesn't survive real scale. Nobody sorts incomes, rents, or commute times for 2.65 million Brooklyn residents by hand — it isn't slow, it's simply not something a human does in a lifetime.

This is a real choice statisticians make, not a shortcut: the unemployment rate, average rent, and income figures you see reported all come from a computer running the exact same sort, mean, median, and range we just did by hand, at a size no person ever could.

Where we stand

- New today: mean, median, mode, range, skewed.
- A histogram is a sorted frequency table, drawn.
- The same four numbers behave differently depending on the shape they come from — that's not a coincidence to memorize, it's something you can now *predict* just by looking.

Where we stand

- New today: mean, median, mode, range, skewed.
- A histogram is a sorted frequency table, drawn.
- The same four numbers behave differently depending on the shape they come from — that's not a coincidence to memorize, it's something you can now *predict* just by looking.

Code Lab 1 will be assigned tonight, using our existing survey data from class 1. There will be some additions to the data and you'll compute these same four numbers yourselves. You'll see which histogram shape you actually got from filling in the remaining students responses using simulated data to obtain a total population of 40.

Due Monday, Sep 28 — we'll visit one more related topic in class that day.