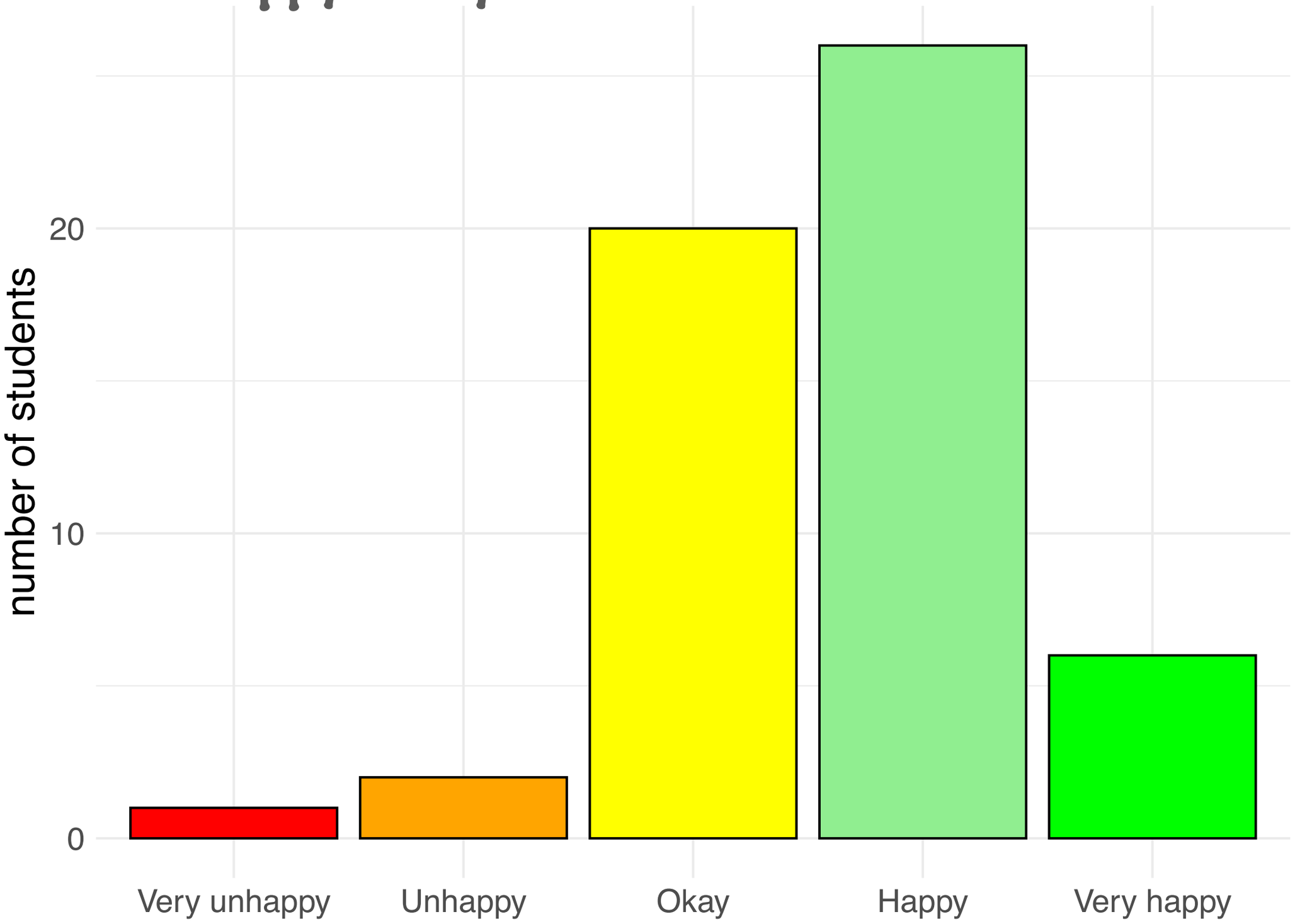
Psyc 60: Intro to Statistics

Prof. Judith Fan

Spring 2021

Last Time

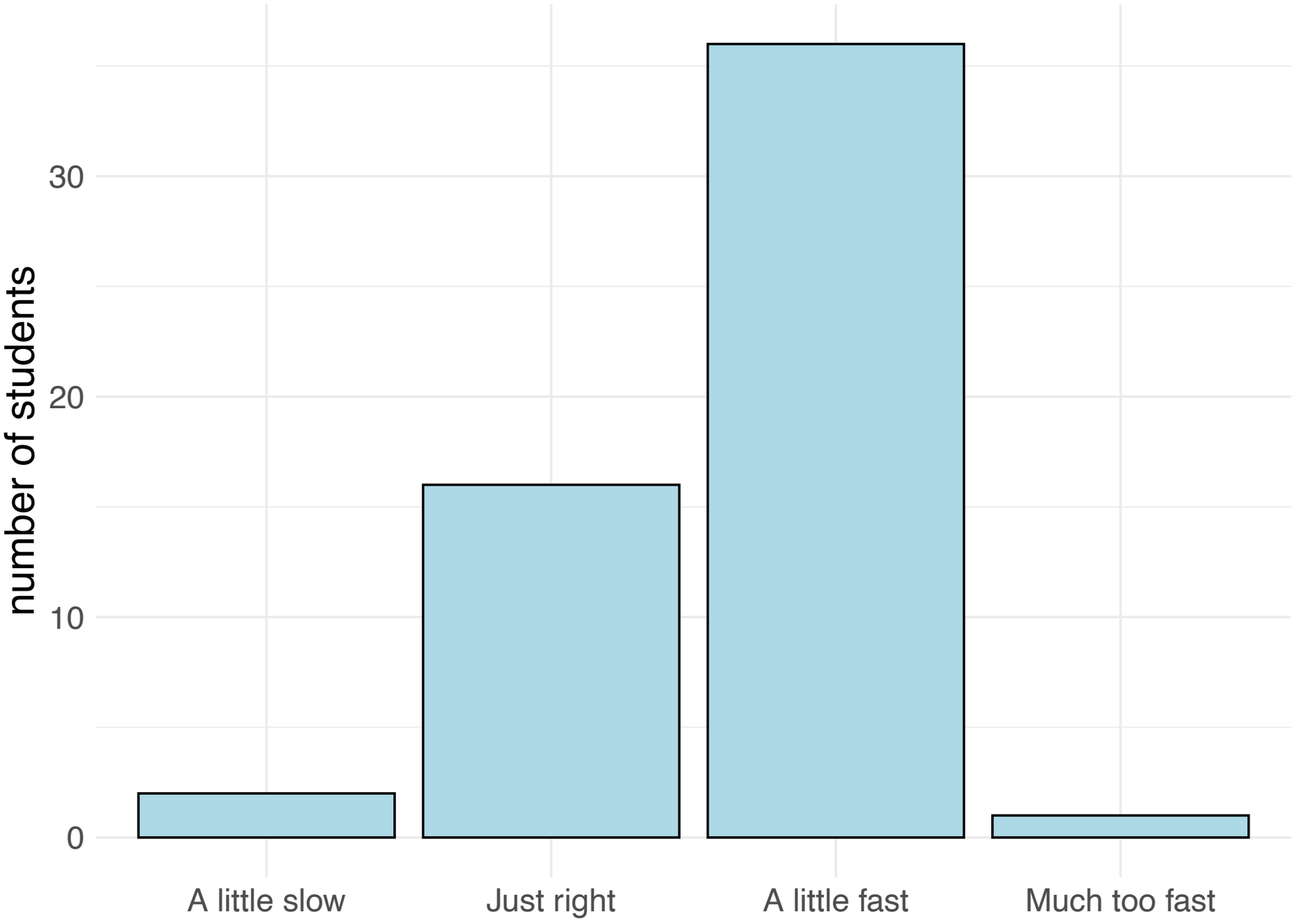
"How happy are you with the class so far?"



N=55 responses

Last Time

How happy are you with the pace of this class so far?

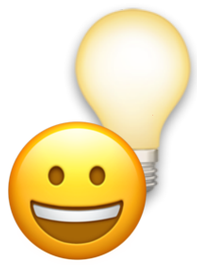


N=55 responses

Last Time

What is something new you learned today? Is there anything you found confusing? Please share any feedback you have about today's lecture.

XXXX



XXXX

N=55 responses

Last Time

What is something new you learned today? Is there anything you found confusing? Please share any feedback you have about today's lecture.

😕 *XXX*

😕 *XXX*

N=55 responses

Today

Mini-Review Session #1

Data visualization and why it matters

Thinking about the data-generating process

Practical tips on how to learn stats w/ R

Starring Jarrett Lovelett & Zhe Huang!


CNN
LIVE

1 *Why does data visualization matter?*

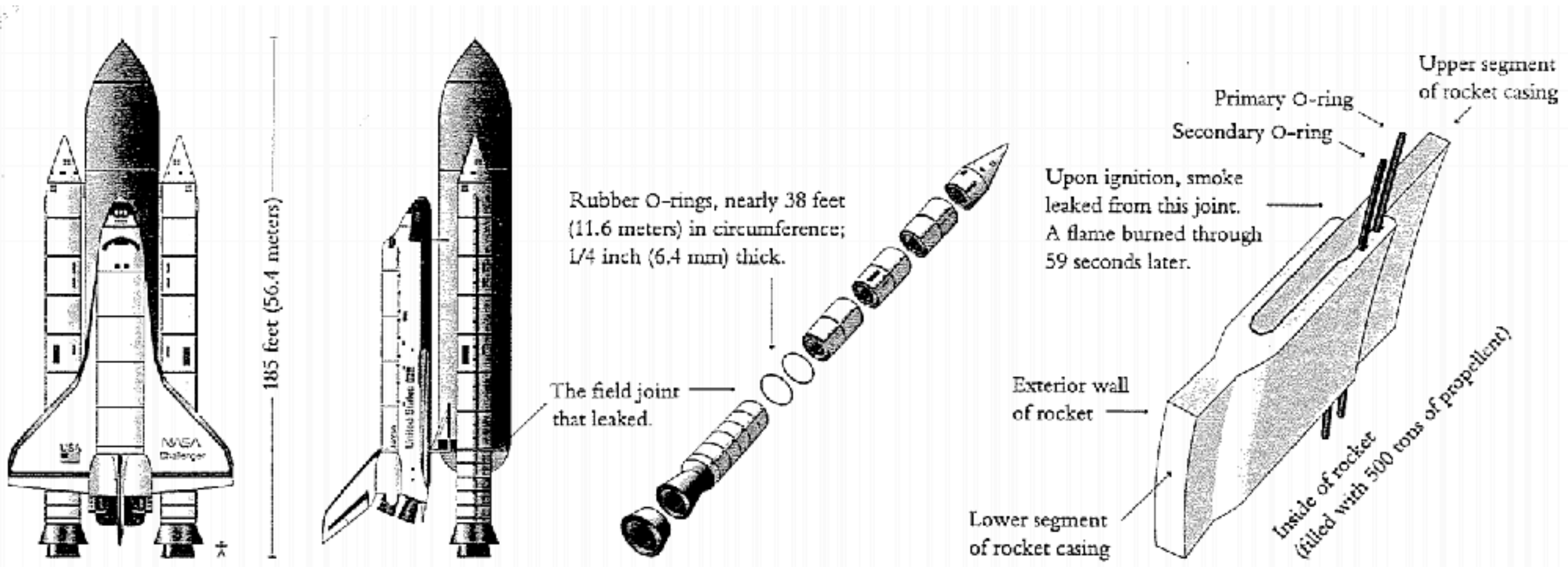
Space Shuttle Challenger disaster (January 28, 1986)

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

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.



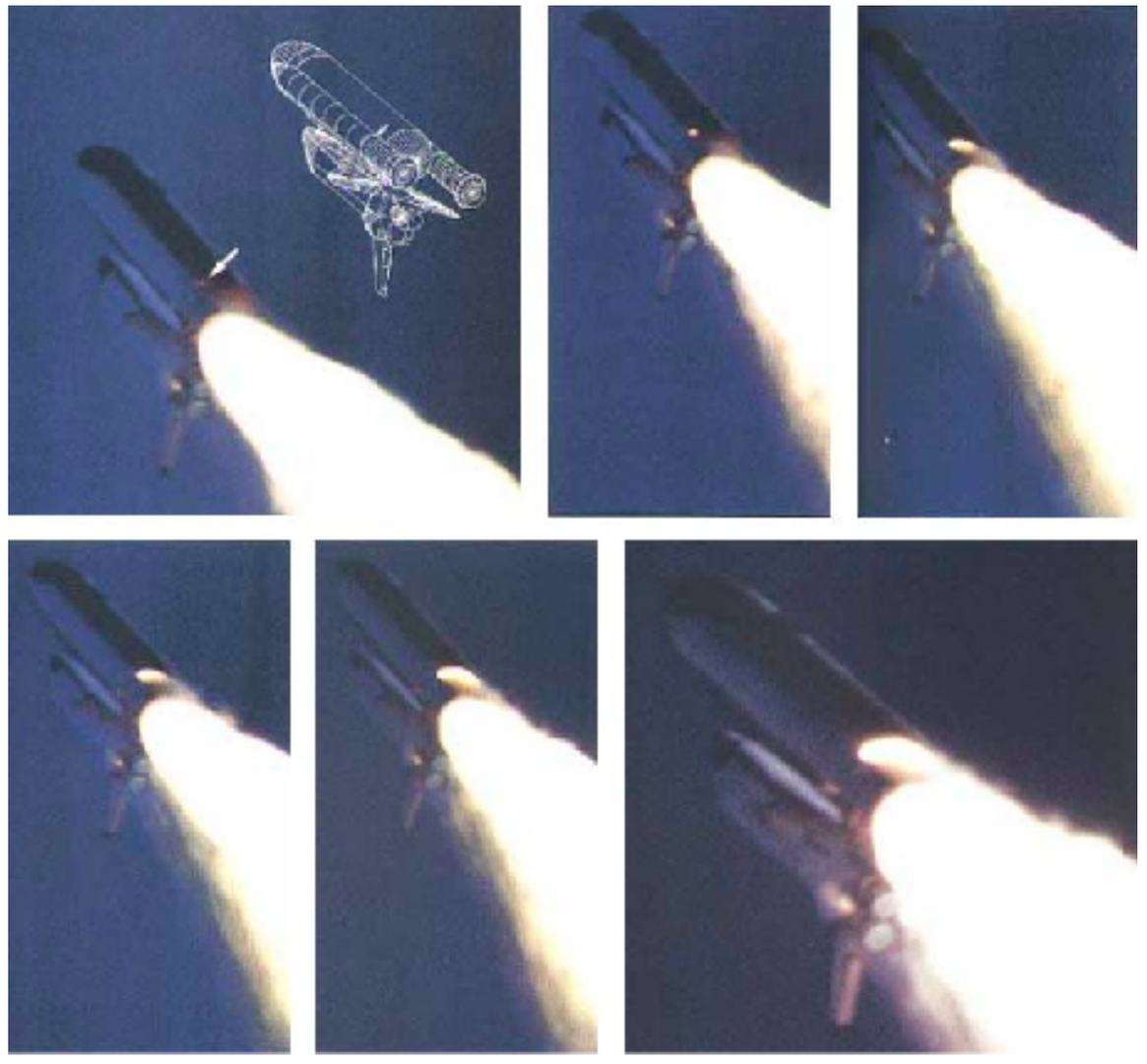
The shuttle consists of an *orbiter* (which carries the crew and has powerful engines in the back), a large liquid-fuel *tank* for the orbiter engines, and 2 solid-fuel *booster rockets* mounted on the sides of the central tank. Segments of the booster rockets are shipped to the launch site, where they are assembled to make the solid-fuel rockets. Where these segments mate, each joint is sealed by two rubber O-rings as shown above. In the case of the Challenger accident, one of these joints leaked, and a torch-like flame burned through the side of the booster rocket.

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.



Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.

What does this have to do with data visualization?

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.

What does this have to do with data visualization?

Ice on launch pad

The reason they leaked is b/c the shuttle launched on a very cold day (20s-30s F).

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.

What does this have to do with data visualization?

Ice on launch pad

The reason they leaked is b/c the shuttle launched on a very cold day (20s-30s F).

Engineers from the rocket contractor presented 13 charts in attempt to convince NASA to postpone the launch due to concerns about the O-rings failing at low temperatures.

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

The space shuttle exploded b/c two rubber O-rings leaked.

What does this have to do with data visualization?

Ice on launch pad

The reason they leaked is b/c the shuttle launched on a very cold day (20s-30s F).

Engineers from the rocket contractor presented 13 charts in attempt to convince NASA to postpone the launch due to concerns about the O-rings failing at low temperatures.

They failed.

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

Major culprit: Ineffective presentation of data

BLOW BY HISTORY

SRM-15 WORST BLOW-BY

- 2 CASE JOINTS (80°), (110°) ARC
- 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 27	10 MPH 25 MPH

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

A more effective summary of the data

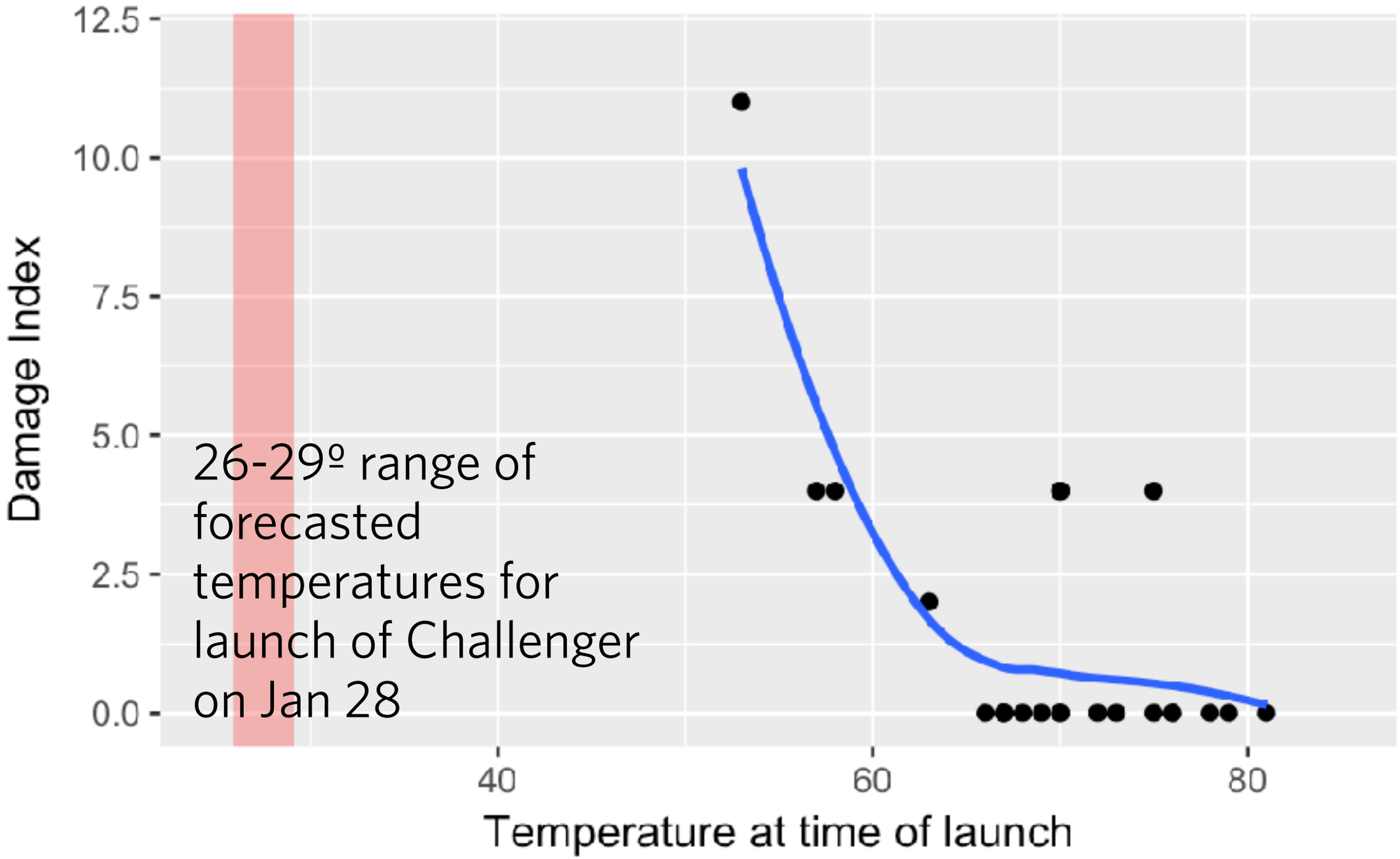
Flight	Date	Temperature °F	Erosion incidents	Blow-by incidents	Damage index	Comments
51-C	01.24.85	53°	3	2	11	Most erosion any flight; blow-by; back-up rings heated.
41-B	02.03.84	57°	1		4	Deep, extensive erosion.
61-C	01.12.86	58°	1		4	O-ring erosion on launch two weeks before Challenger.
41-C	04.06.84	63°	1		2	O-rings showed signs of heating, but no damage.
1	04.12.81	66°			0	Coolest (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	

Tufte, *Visual Explanations* (1997)

1 *Why does data visualization matter?*

Space Shuttle Challenger disaster (January 28, 1986)

An even more effective visualization of the data



Tufte, *Visual Explanations* (1997)

1 *What is the goal of data visualization?*

Gelman & Unwin (2012)

1

What is the goal of data visualization?

Discovery

Gelman & Unwin (2012)

1 *What is the goal of data visualization?*

Discovery

Communication

Gelman & Unwin (2012)

1 *What is the goal of data visualization?*

Discovery

Overview — a qualitative sense, checking assumptions, confirming known results, looking for distinct patterns

Communication

Gelman & Unwin (2012)

1 *What is the goal of data visualization?*

Discovery

Overview — a qualitative sense, checking assumptions, confirming known results, looking for distinct patterns

Exploration — to expose unexpected aspects of the data

Communication

Gelman & Unwin (2012)

1

What is the goal of data visualization?

Discovery

Overview — a qualitative sense, checking assumptions, confirming known results, looking for distinct patterns

Exploration — to expose unexpected aspects of the data

Communication

Clarity — distilled to make relevant comparisons easy; in support of an argument

Gelman & Unwin (2012)

1

What is the goal of data visualization?

Discovery

Overview — a qualitative sense, checking assumptions, confirming known results, looking for distinct patterns

Exploration — to expose unexpected aspects of the data

Communication

Clarity — distilled to make relevant comparisons easy; in support of an argument

Impact — illustrate the results of careful data analysis in a way that attracts attention & interest

Gelman & Unwin (2012)

1 *What makes a data visualization effective?*

Some basic principles

1. Show the data and make them stand out

2. Avoid distorting the data

3. Keep human limitations in mind

4. Reveal the underlying message of the data

1

What makes a data visualization effective?

1. Show the data and make them stand out

summarizing more *showing more*

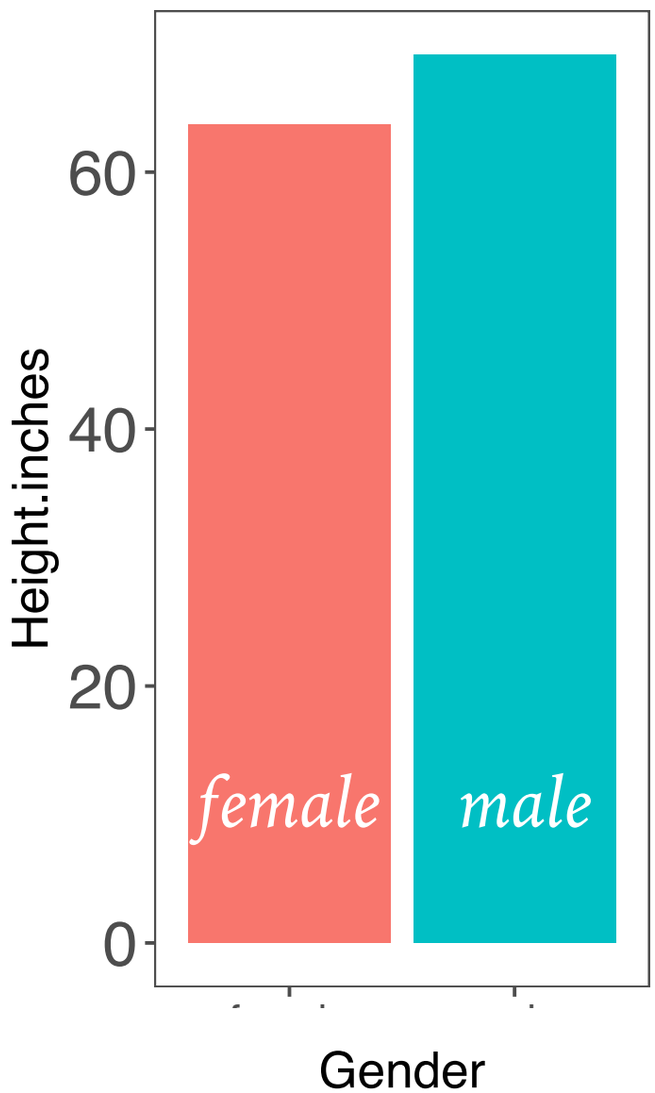
1

What makes a data visualization effective?

1. Show the data and make them stand out

summarizing more ⟷ *showing more*

bar graph



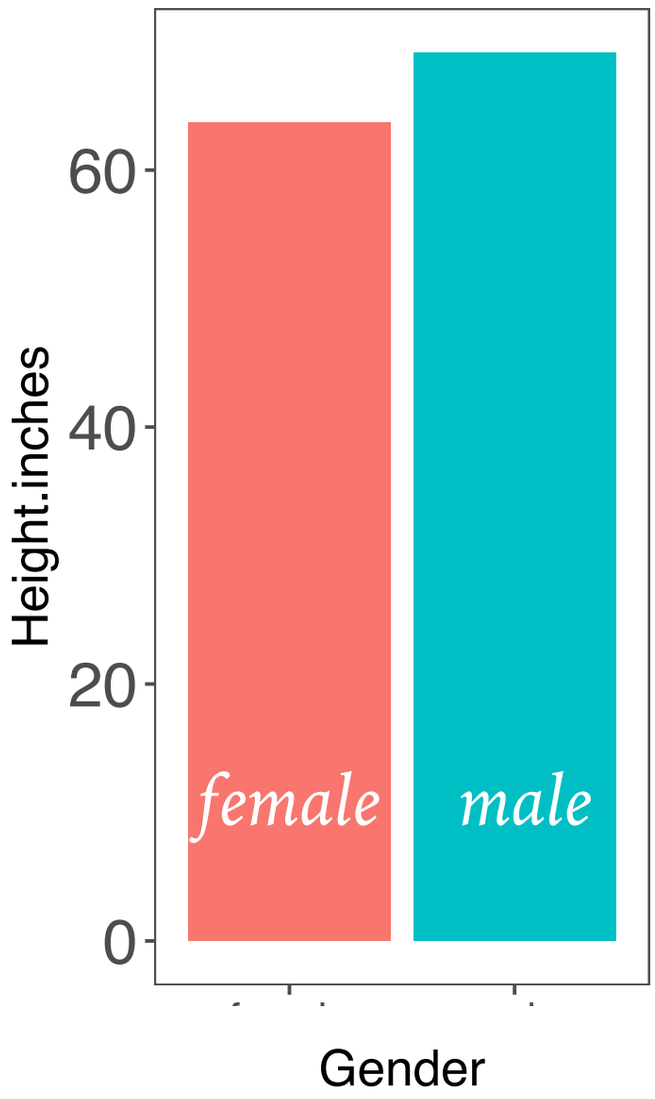
1

What makes a data visualization effective?

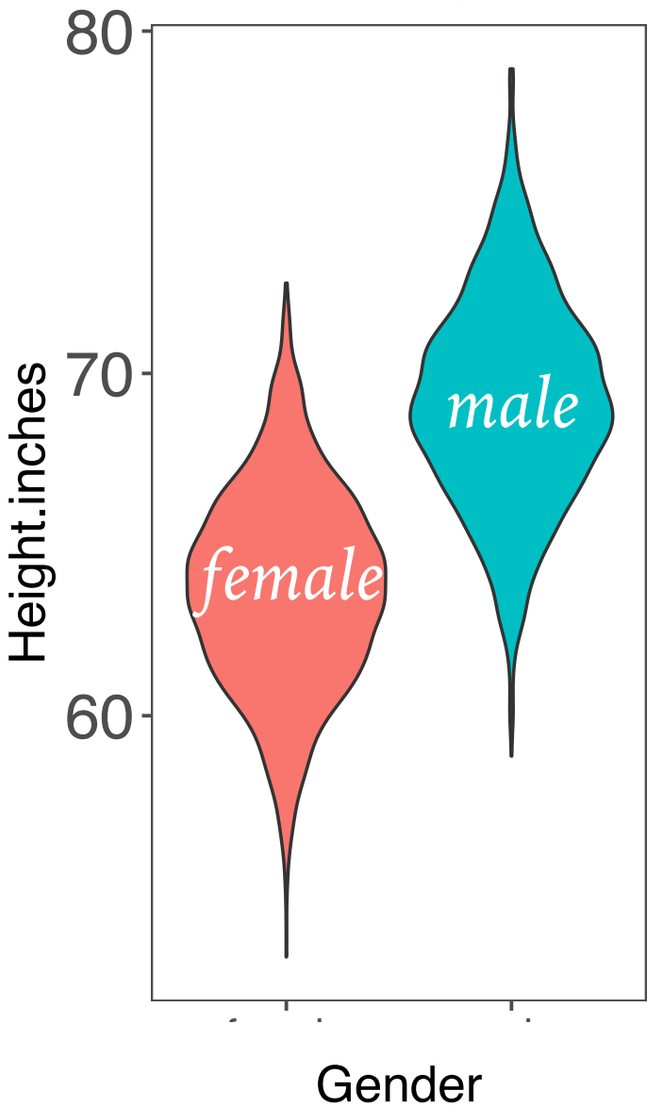
1. Show the data and make them stand out

summarizing more — *showing more*

bar graph



violin plot



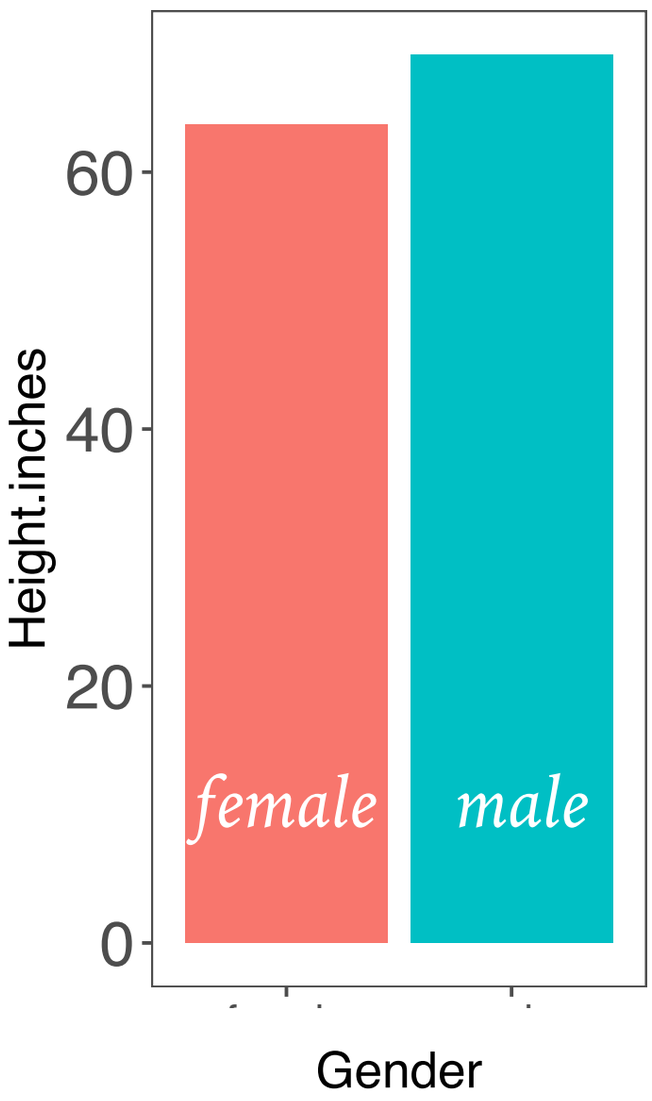
1

What makes a data visualization effective?

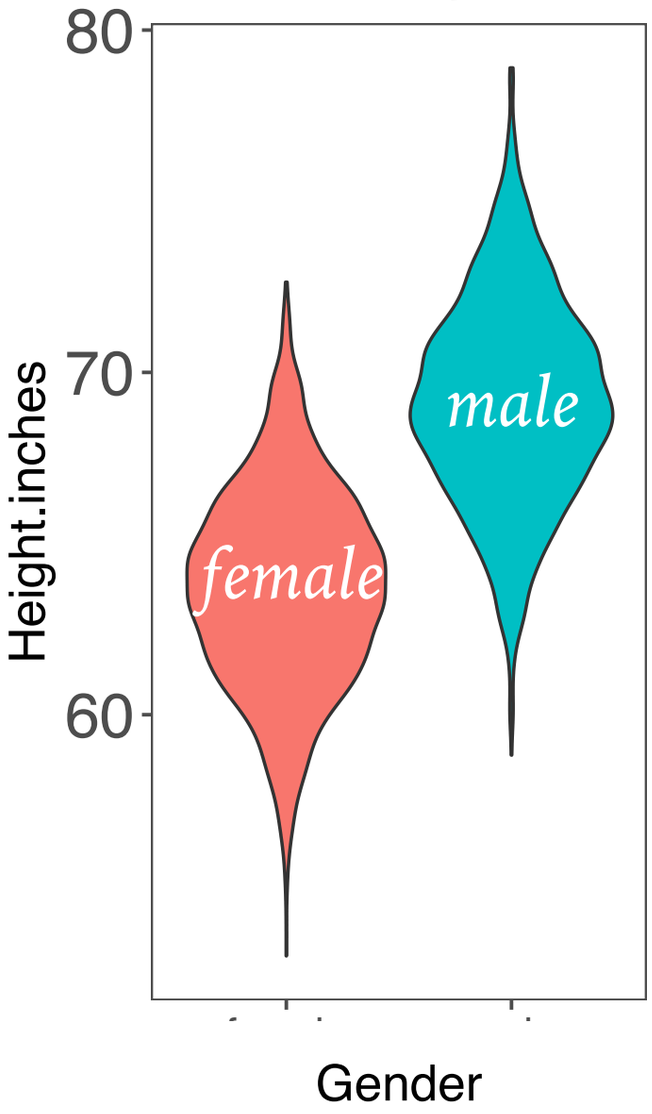
1. Show the data and make them stand out

summarizing more ⟷ *showing more*

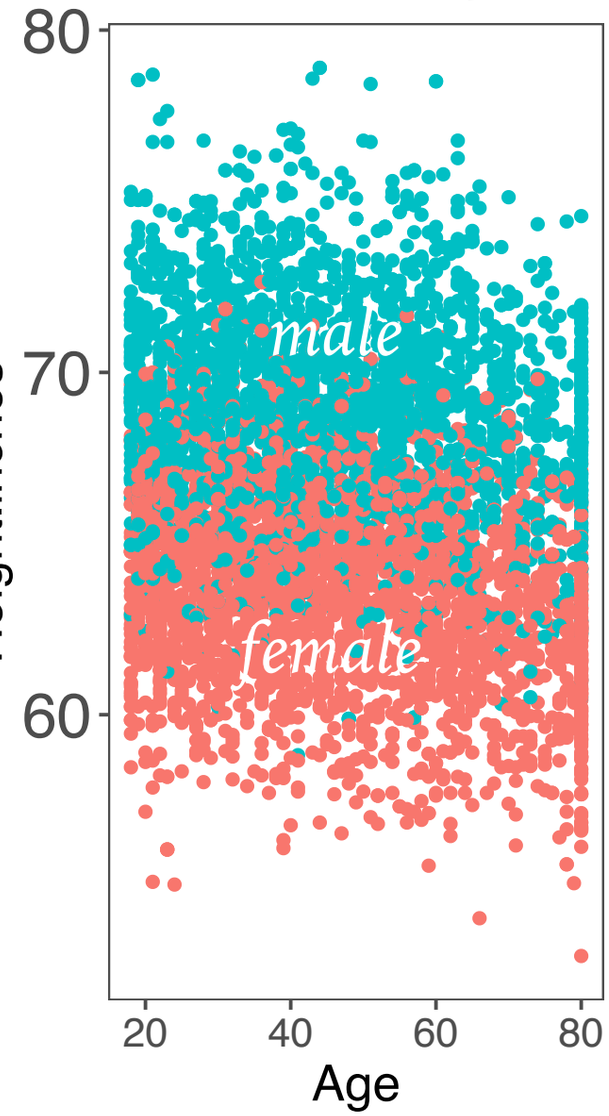
bar graph



violin plot

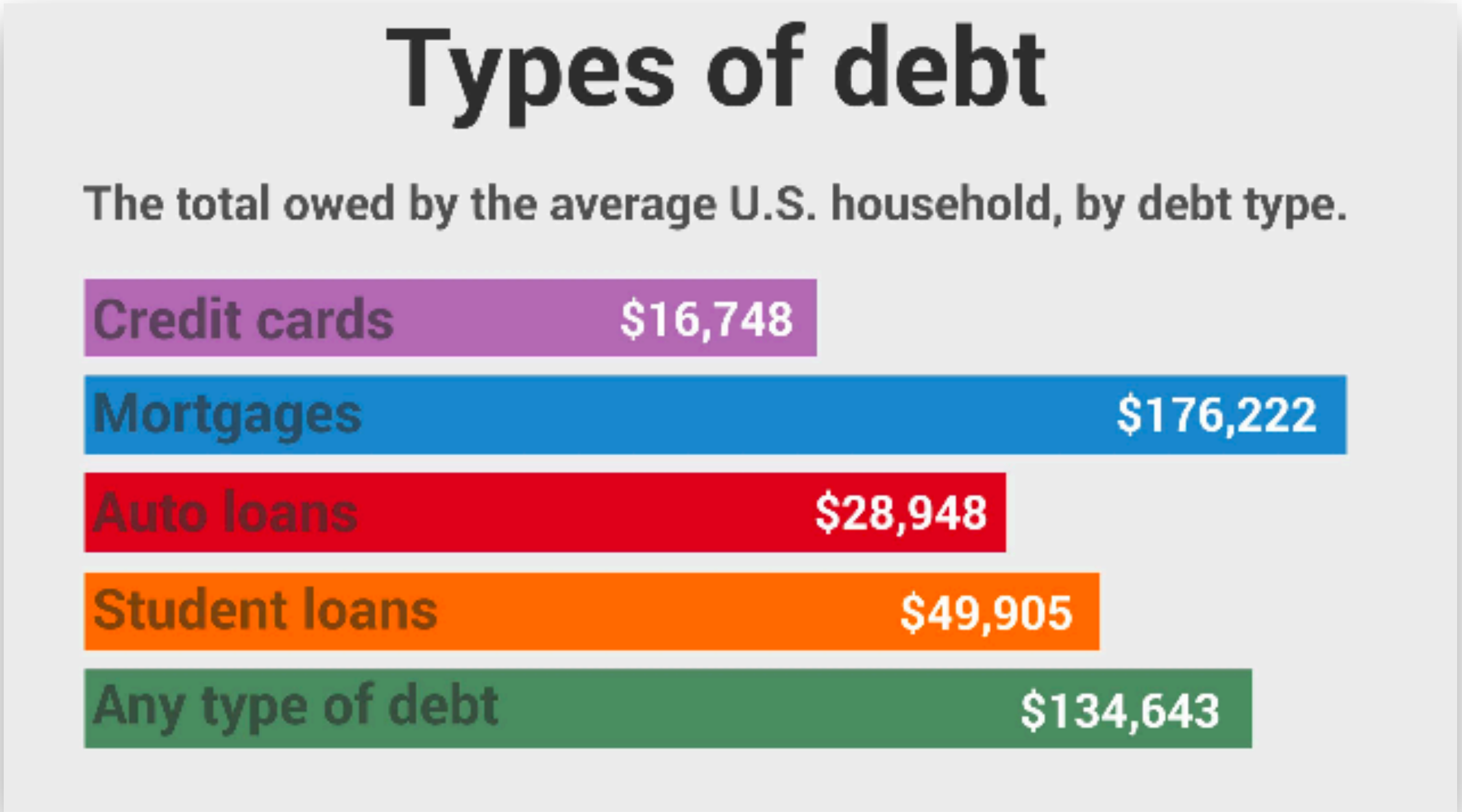


scatter plot



2. Avoid distorting the data

Beware of effects that distort the data



http://viz.wtf/

3. Keep human limitations in mind

Perceptual limitations

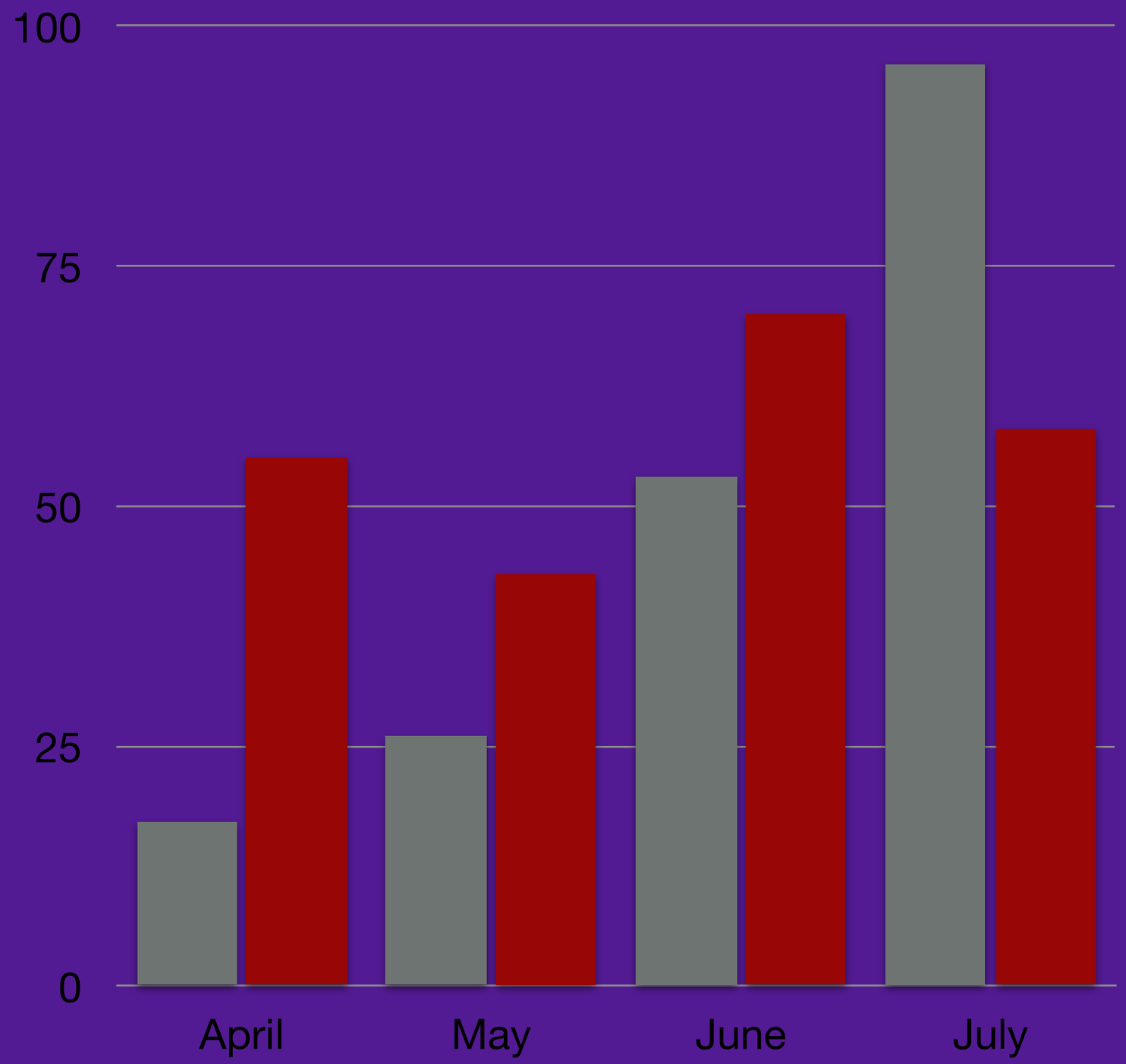
Ability to tell colors apart varies a lot in the population

Volume/area is harder to perceive than length

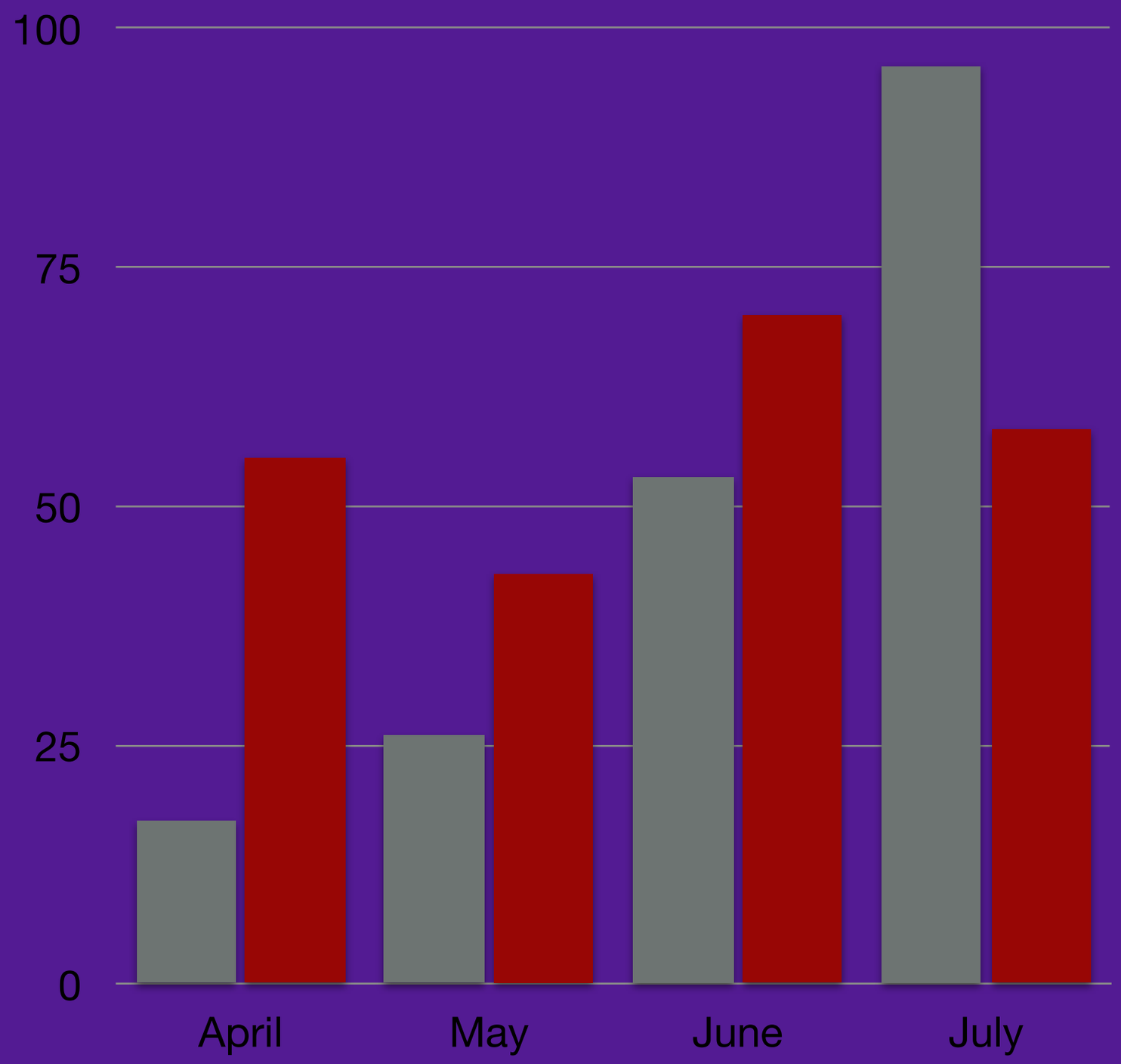
Cognitive limitations

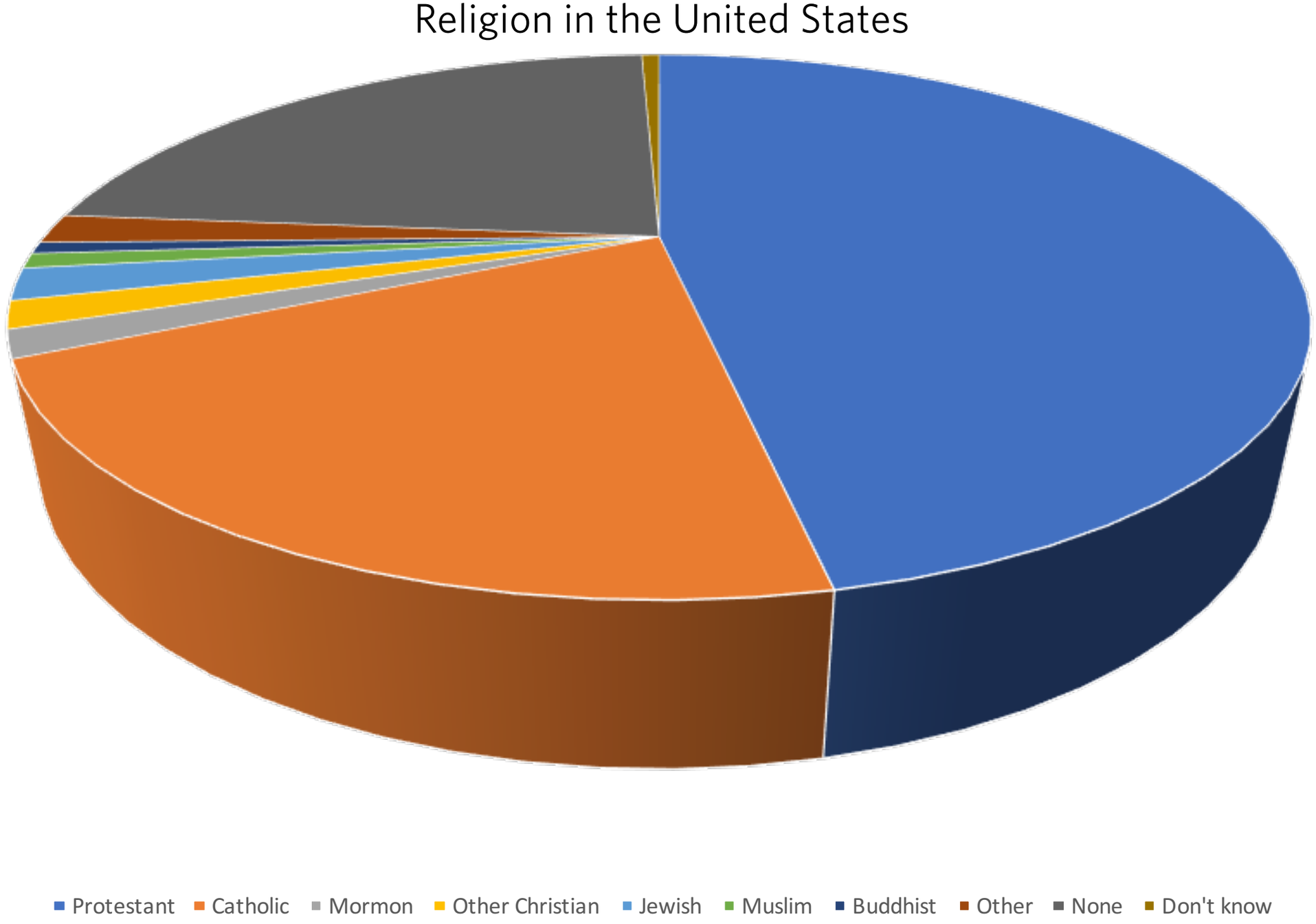
We have limited working memory capacity

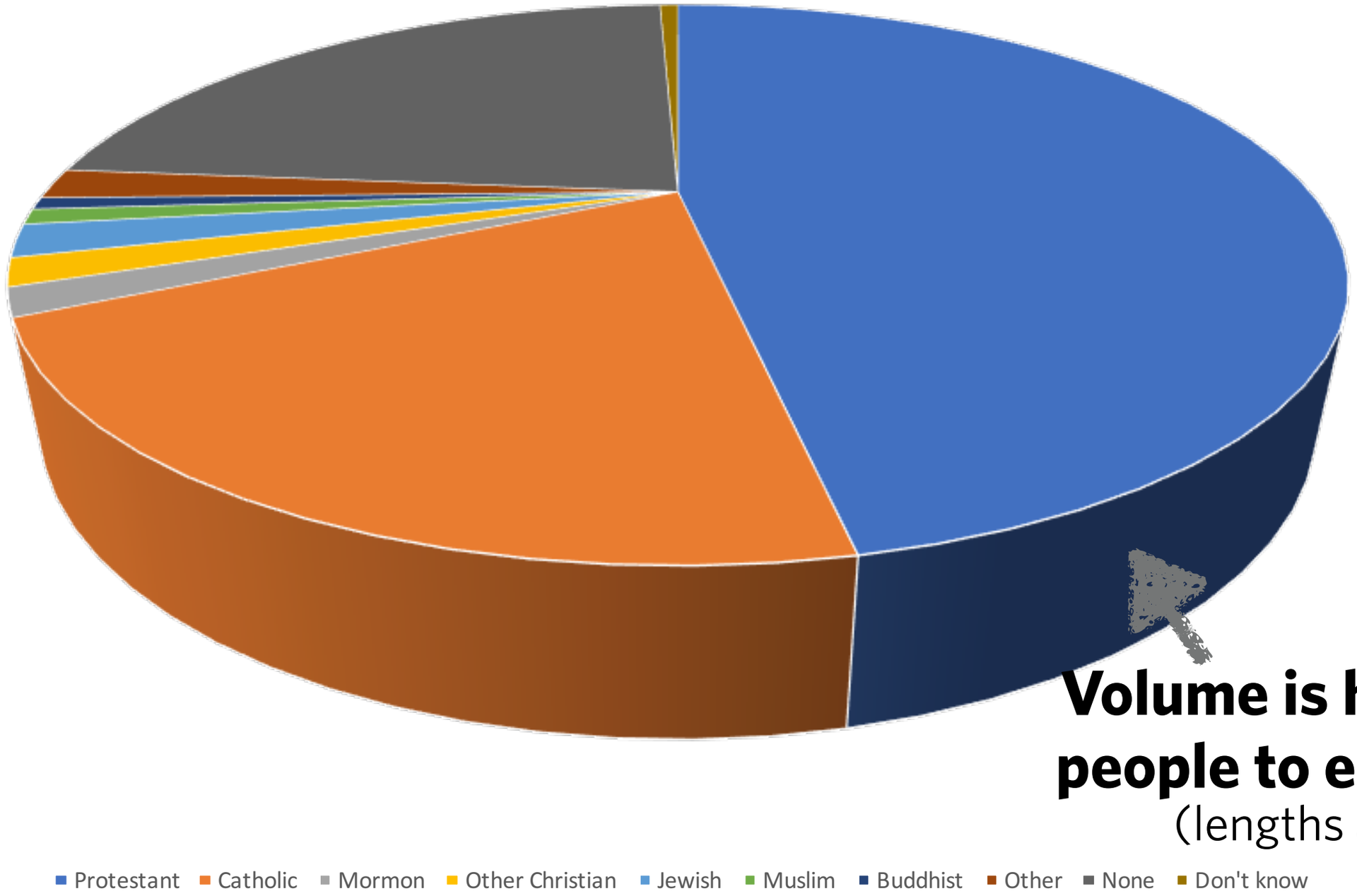
Don’t make the viewer remember too much

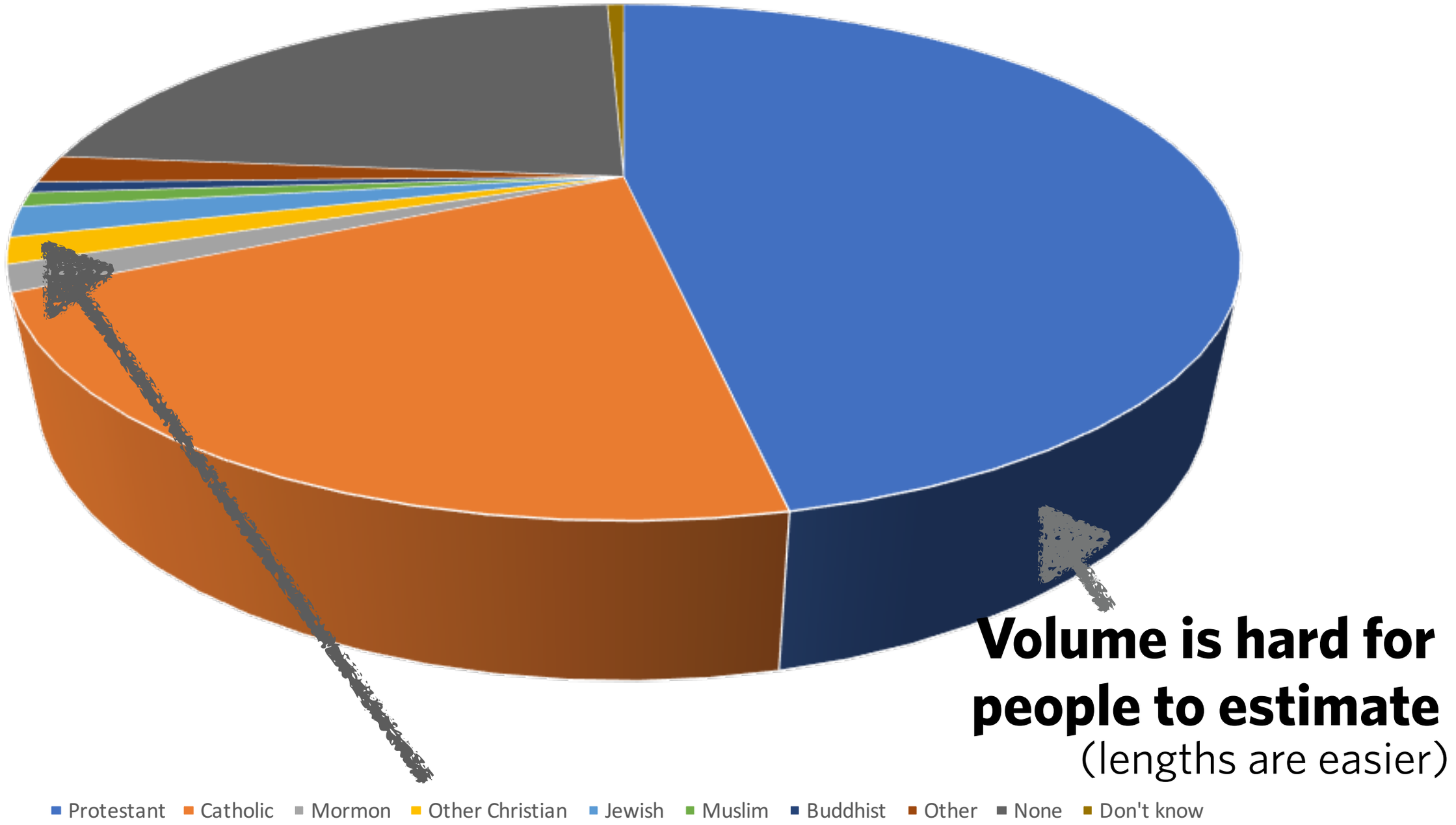

100
75
50
25
0
April
May
June
July

Use contrast in brightness in addition to color



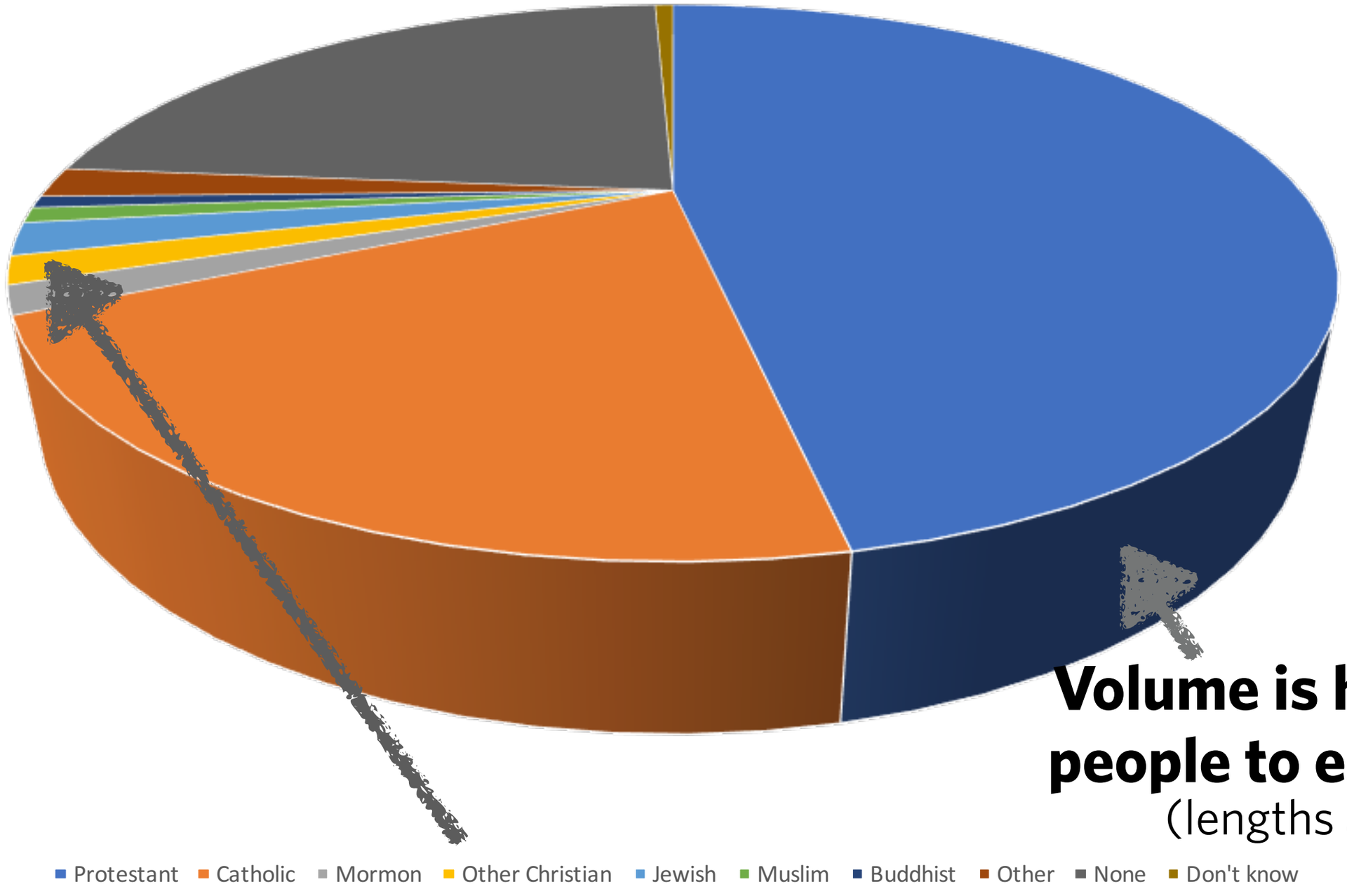

Religion in the United States
Protestant
Catholic
Mormon
Other Christian
Jewish
Muslim
Buddhist
Other
None
Don't know


Religion in the United States
Volume is hard for people to estimate
(lengths are easier)
Protestant
Catholic
Mormon
Other Christian
Jewish
Muslim
Buddhist
Other
None
Don't know



Volume is hard for people to estimate
(lengths are easier)

Don't make your viewer remember too much



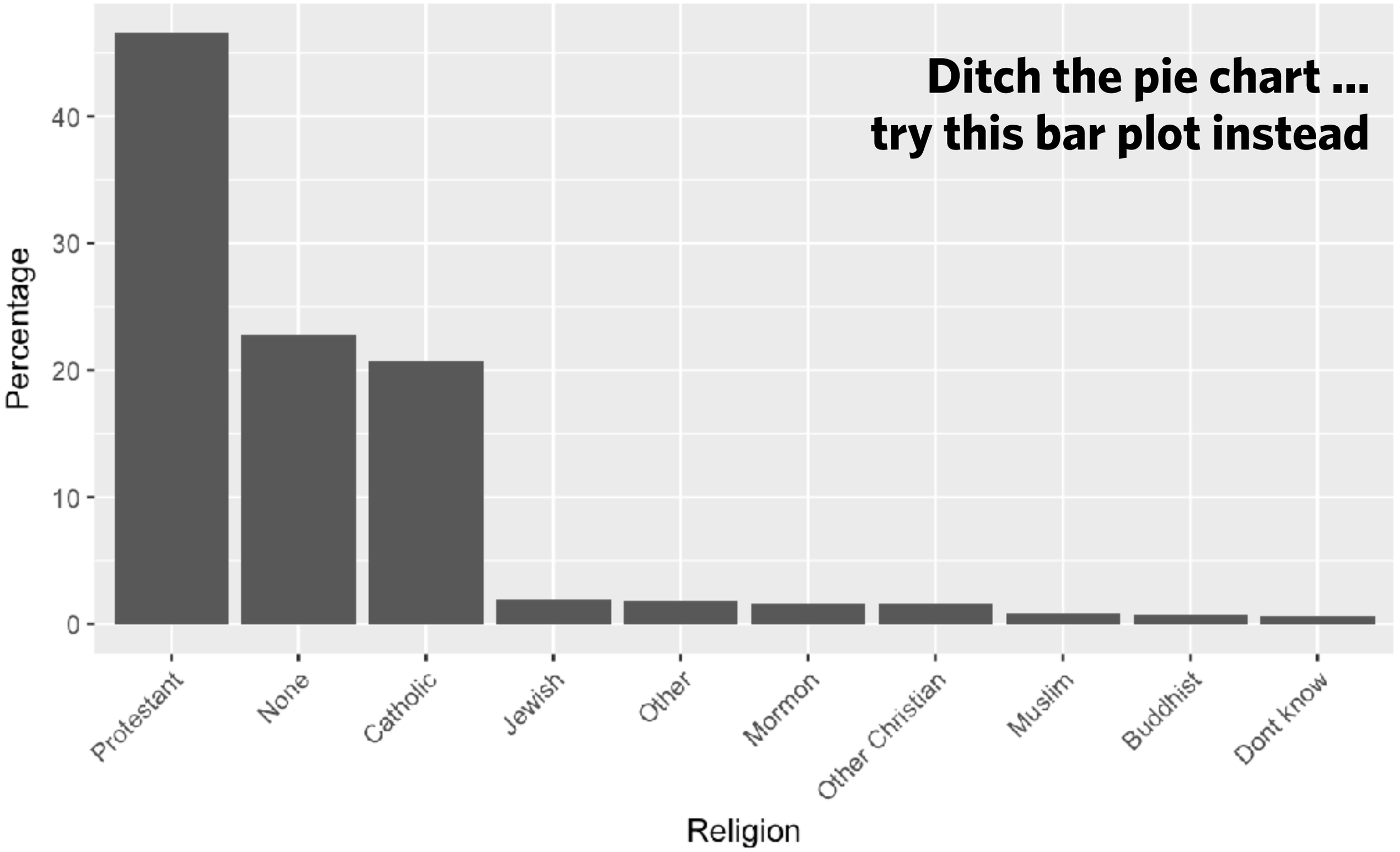
Volume is hard for people to estimate
(lengths are easier)

Don't make your viewer remember too much

Make text large & easy to read

(redo) Religion in the United States

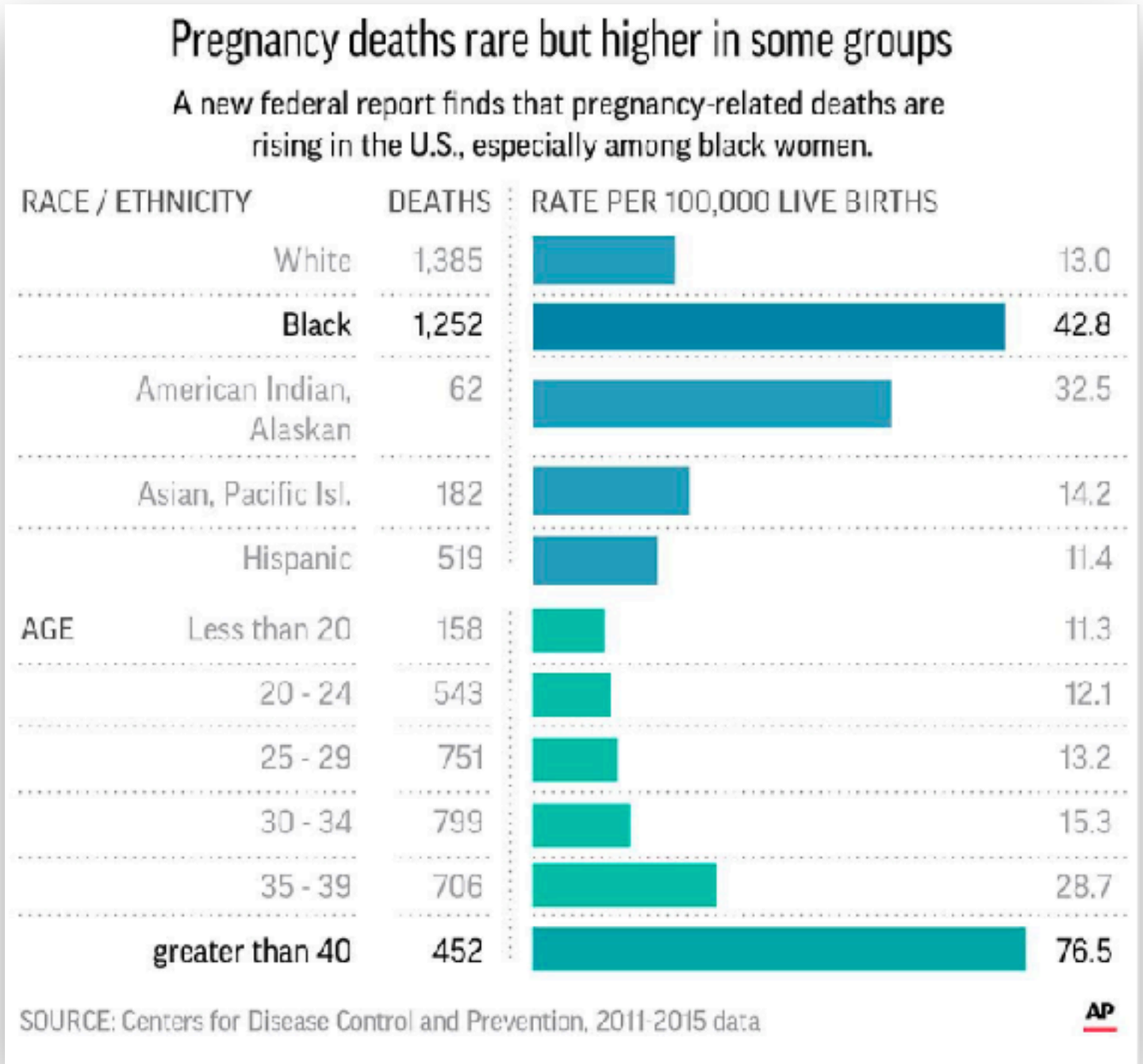
Ditch the pie chart ...
try this bar plot instead



1

What makes a data visualization effective?

4. Reveal the underlying message of the data



- What is the message of this visualization?
- How could that message be better conveyed?

1

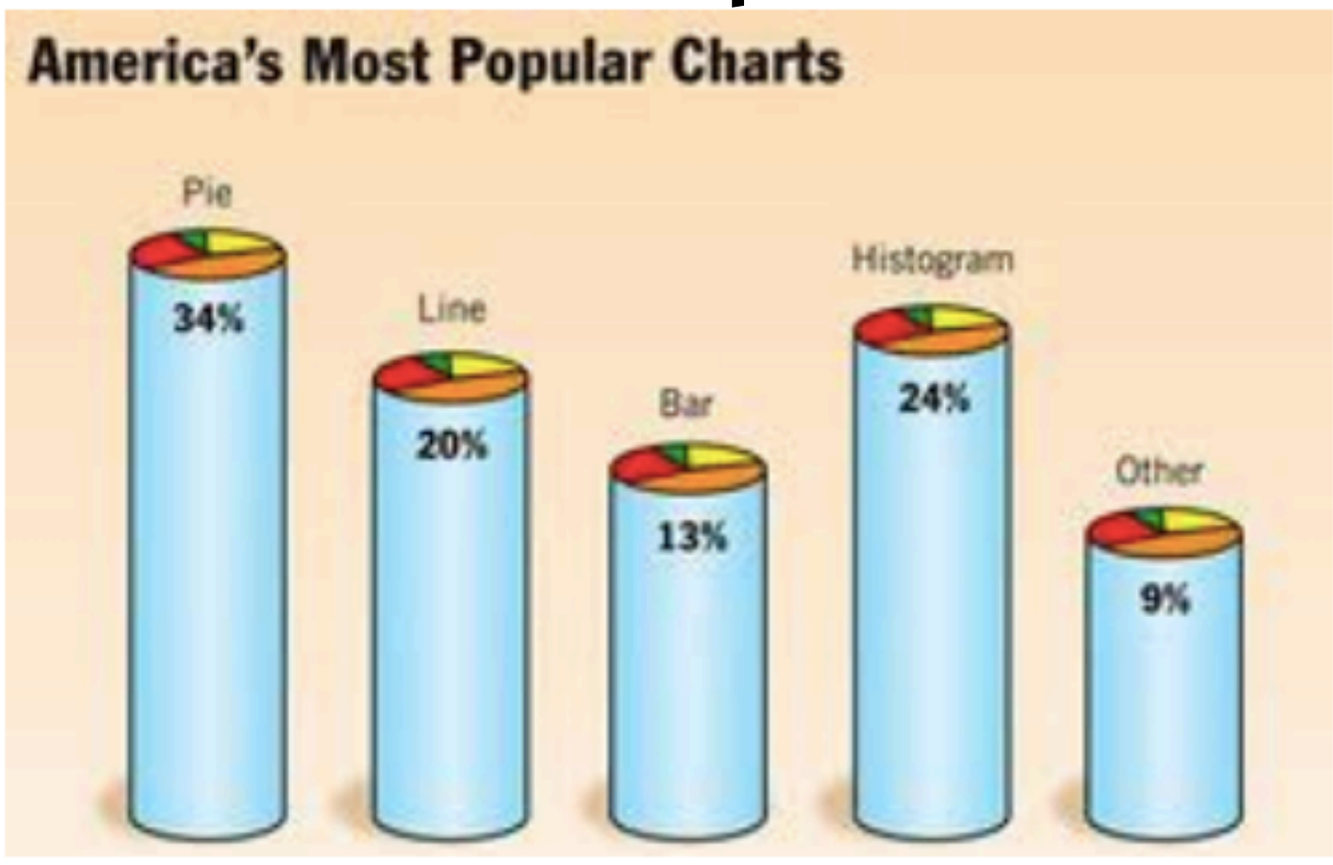
What makes a data visualization effective?

Your turn: What's wrong with these graphs?

Graph A

USA Today

Graph B



The Onion (parody)

PollEv.com/psyc60

Poll Everywhere

02:00

Lecture 8: What's wrong with these graphs? (Name one specific problem with either of them)

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Today

Mini-Review Session #1

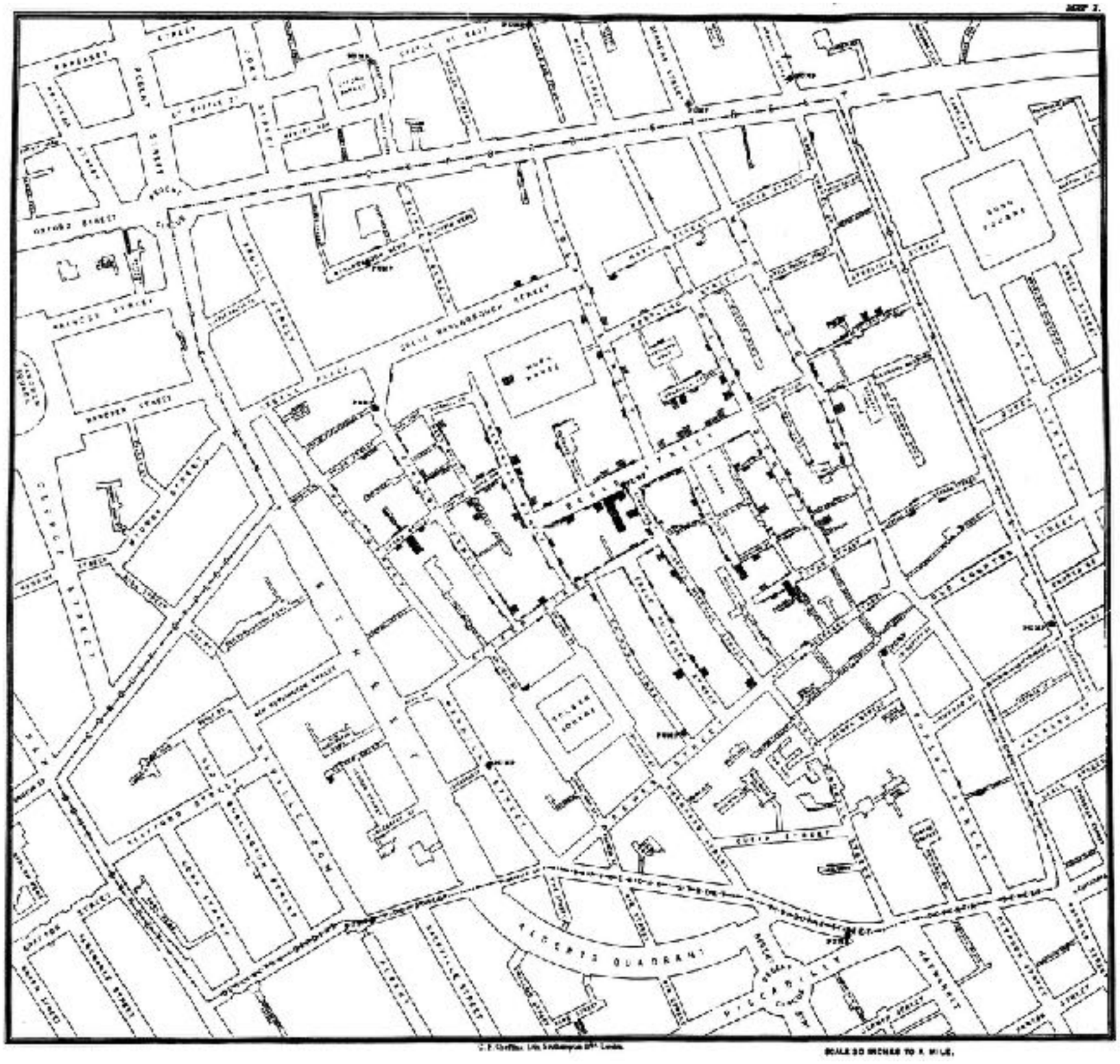
Data visualization and why it matters

Thinking about the data-generating process

Practical tips on how to learn stats w/ R

Starring Jarrett Lovelett & Zhe Huang!

What is this graph showing?



John Snow (1854)

What is this graph showing?

context:
cholera outbreak
in 19th c. London

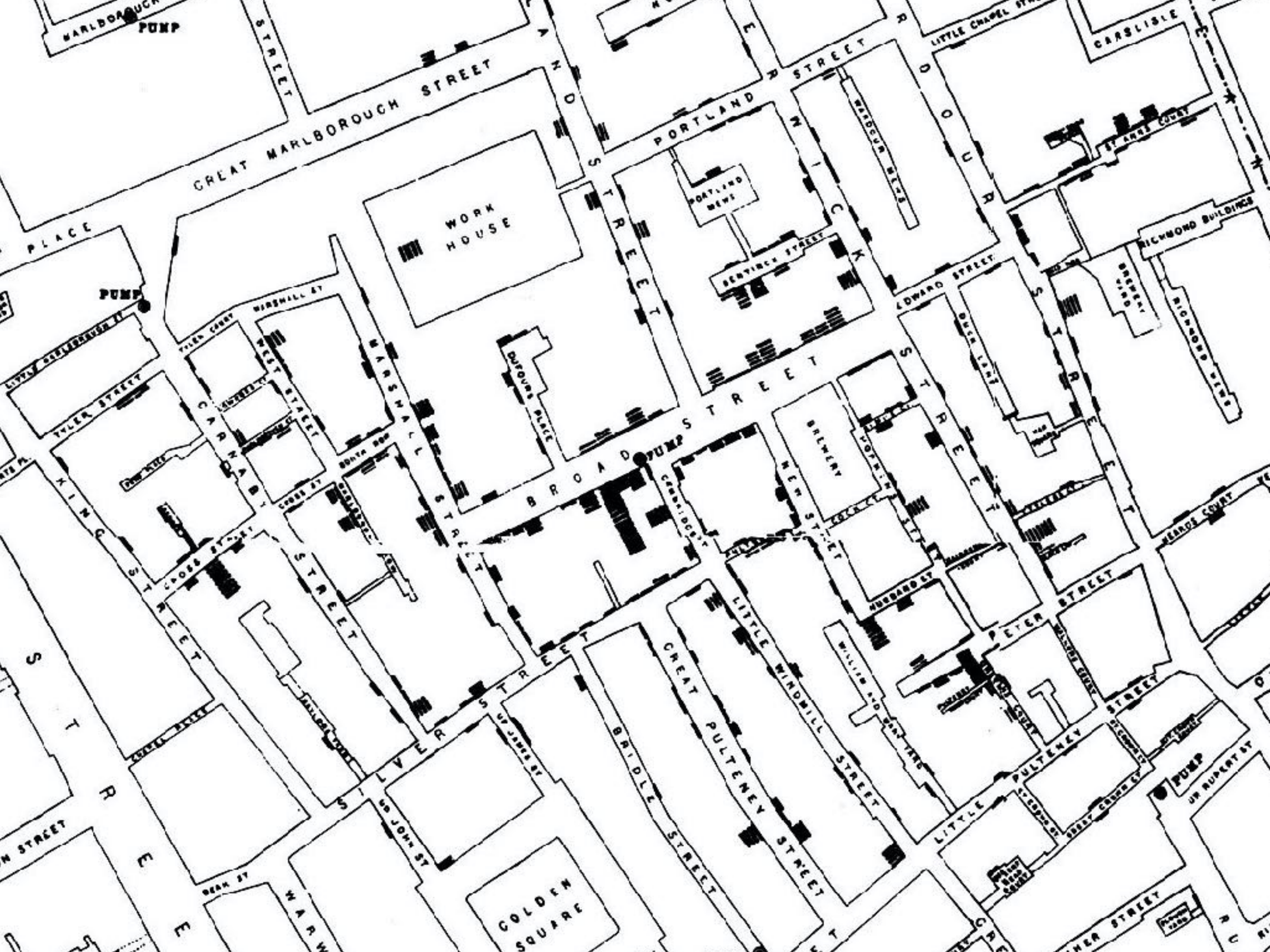
John Snow (1854)

What is this graph showing?

context:
cholera outbreak
in 19th c. London

John Snow (1854)

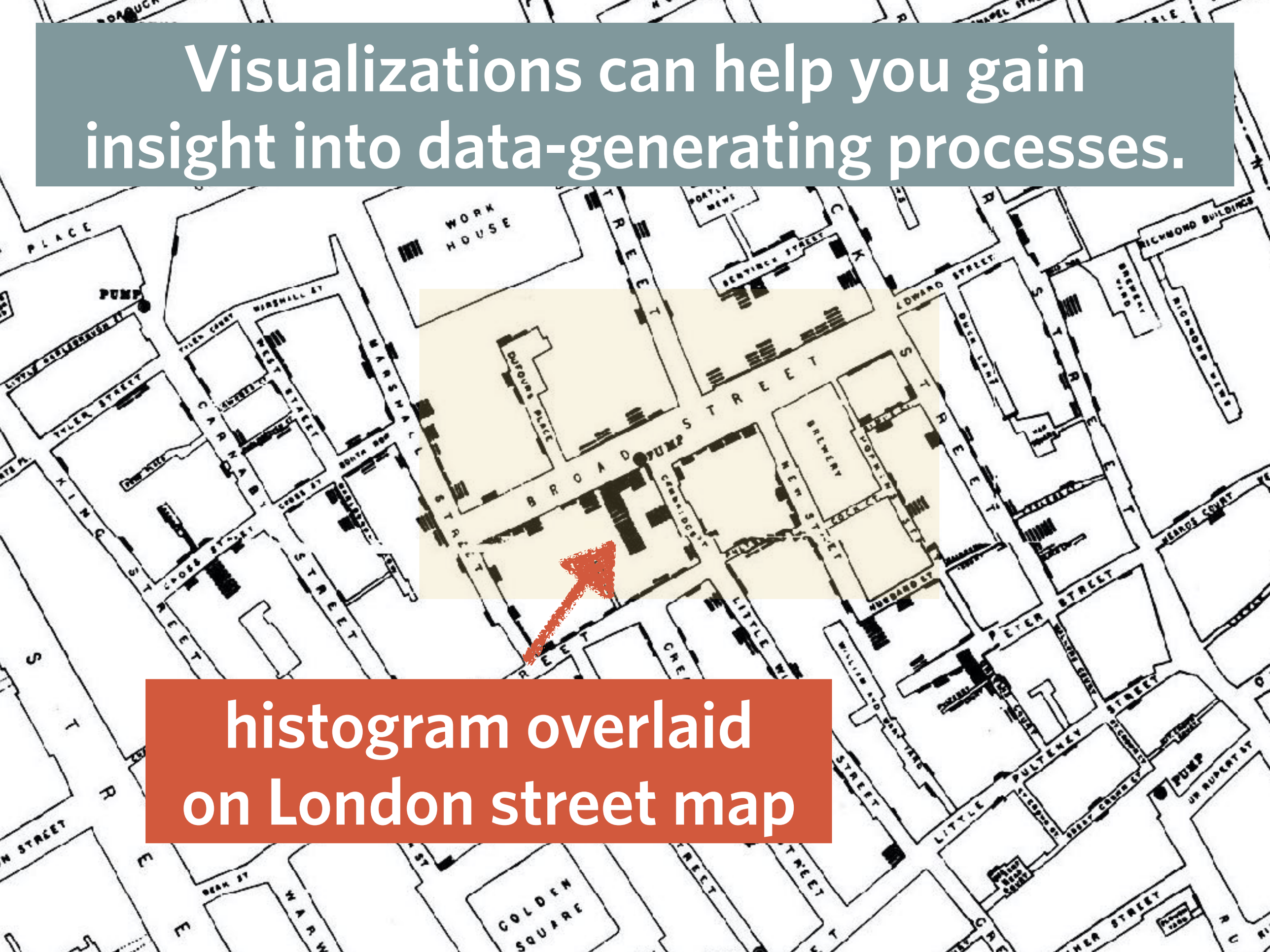
John Snow


GREAT MARLBOROUGH STREET
PUMP
WORK HOUSE
PORTLAND STREET
PORTLAND MEWS
BERNERS STREET
BERNERS MEWS
LITTLE CHAPEL STREET
CARLISLE
ST ANNE COURT
RICHMOND BUILDINGS
RICHMOND MEWS
EDWARD STREET
DUCK LANE
BROAD STREET
DUFOURS PLACE
BRIDLE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
BREWERY
HUSBAND ST
PETER STREET
WARWICK STREET
GOLDEN SQUARE
ST JAMES ST
ST JOHN ST
TYLER STREET
MARSHALL ST
CROSS STREET
BEAK ST
KING STREET
PULTENEY
LITTLE PULTENEY ST
GREAT PULTENEY ST
UP RUPERT ST
HEARDS COURT


PUMP
GREAT MARLBOROUGH STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WORK HOUSE
PLACE
PUMP
TYLER STREET
KING STREET
CARNABY
CROSS STREET
WEST STREET
MARSHALL STREET
DUFOURS PLACE
BROAD STREET
PUMP
CAMBRIDGE ST
BREWERY
NEW STREET
HUBBARD ST
EDWARD STREET
ST ANNS COURT
CARLISLE
LITTLE CHAPEL ST
RICHMOND BUILDINGS
RICHMOND MEWS
HEARDS COURT
PETER STREET
LITTLE PETER STREET
SILVER STREET
UP JAMES ST
UP JOHN ST
BRIDLE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
LITTLE PULTENEY STREET
PUMP
UP RUPERT ST
GOLDEN SQUARE

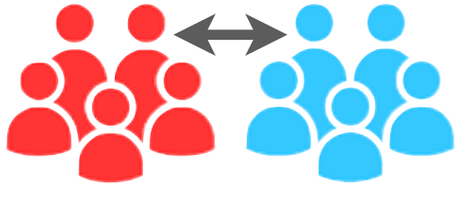

histogram overlaid
on London street map

Visualizations can help you gain insight into data-generating processes.

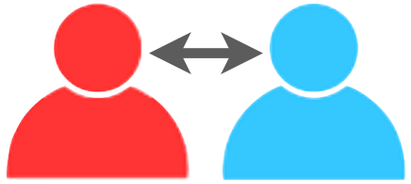


2 *Thinking about the data-generating process*

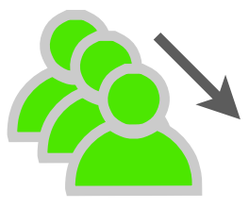
In psychology, there are multiple potential sources of variation we may care about:



variation between groups (e.g., cultural)



variation between individuals (e.g., personality, genetic, environment)

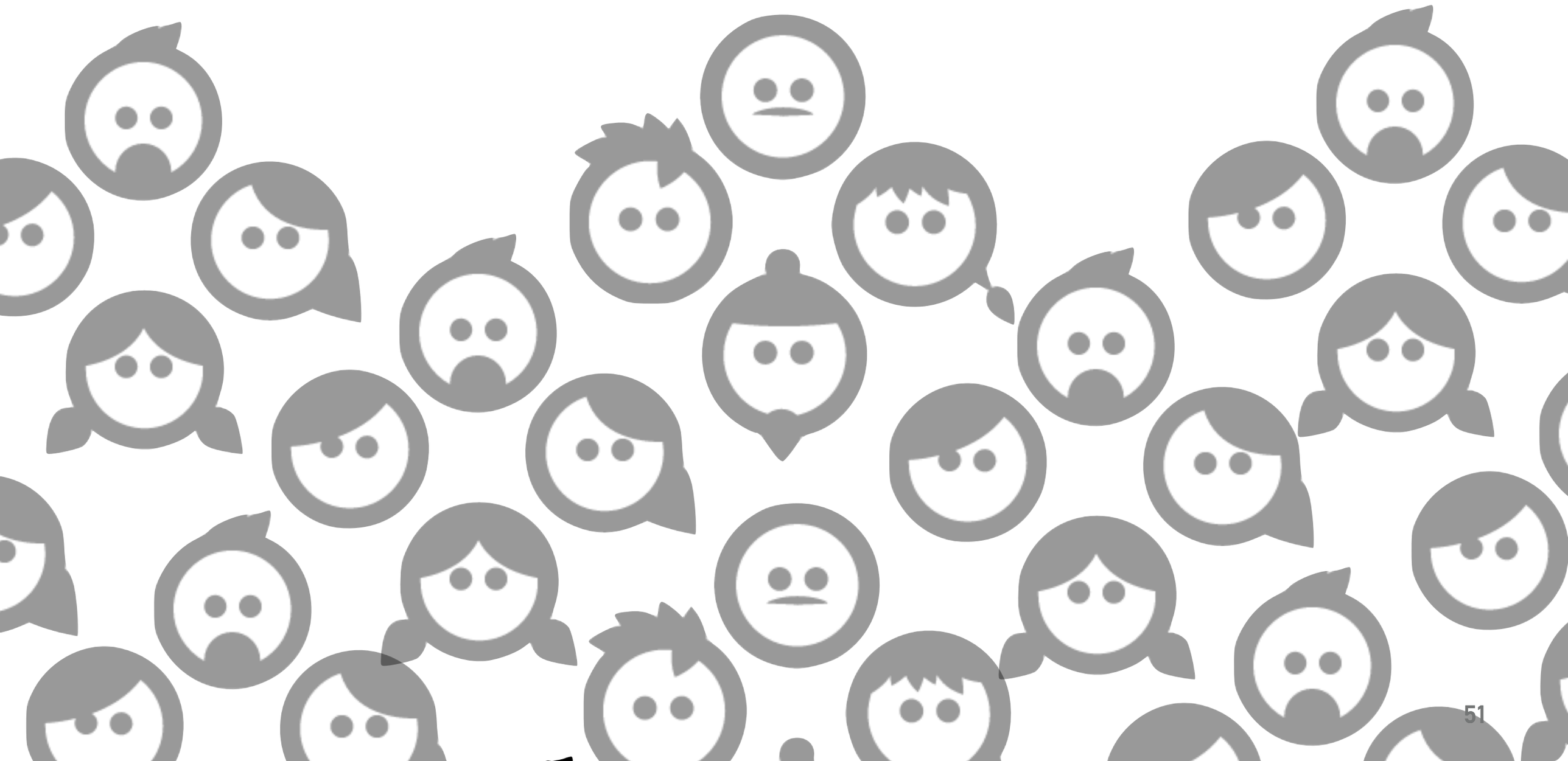


variation within individuals (e.g., mood, learning, random noise)

measurement error (e.g., confusingly worded survey)

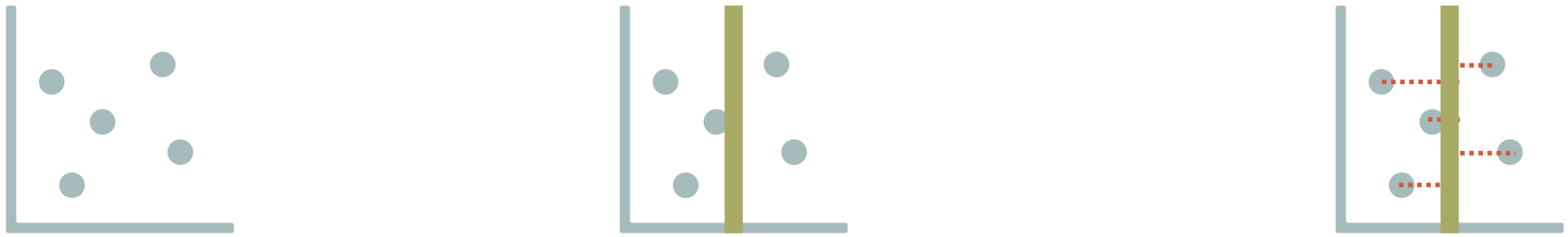
2 *Thinking about the data-generating process*

Psychological science aims to explain the sources of variation in people's experiences & behavior.



2 *Thinking about the data-generating process*

Psychological science aims to explain the sources of variation in people's experiences & behavior.



data = **model** + **error**

what we actually observe

what we expect to observe

difference between expected and observed

2 *Thinking about the data-generating process*

*Given a dataset, what **data-generating process** (in the population) could explain why they look that way?*

inference

data = model + error

prediction

*Given a data-generating process, what do we expect our **sample of data** to look like?*

1 *Thinking about the data-generating process*

Psychological science aims to uncover the data-generating process that gives rise to observed variation in people's experiences & behavior.

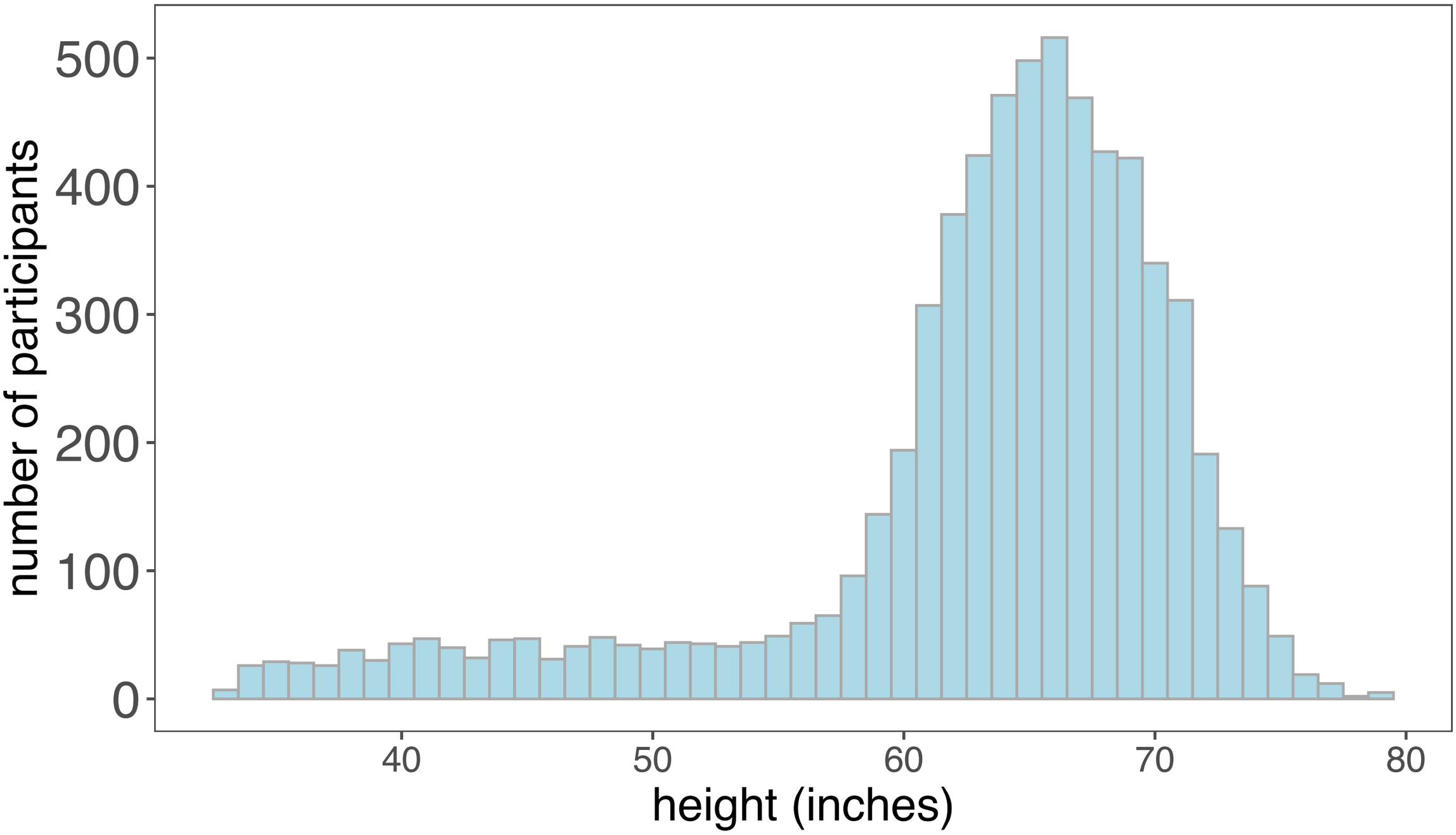
data = model + error

"data-generating process"

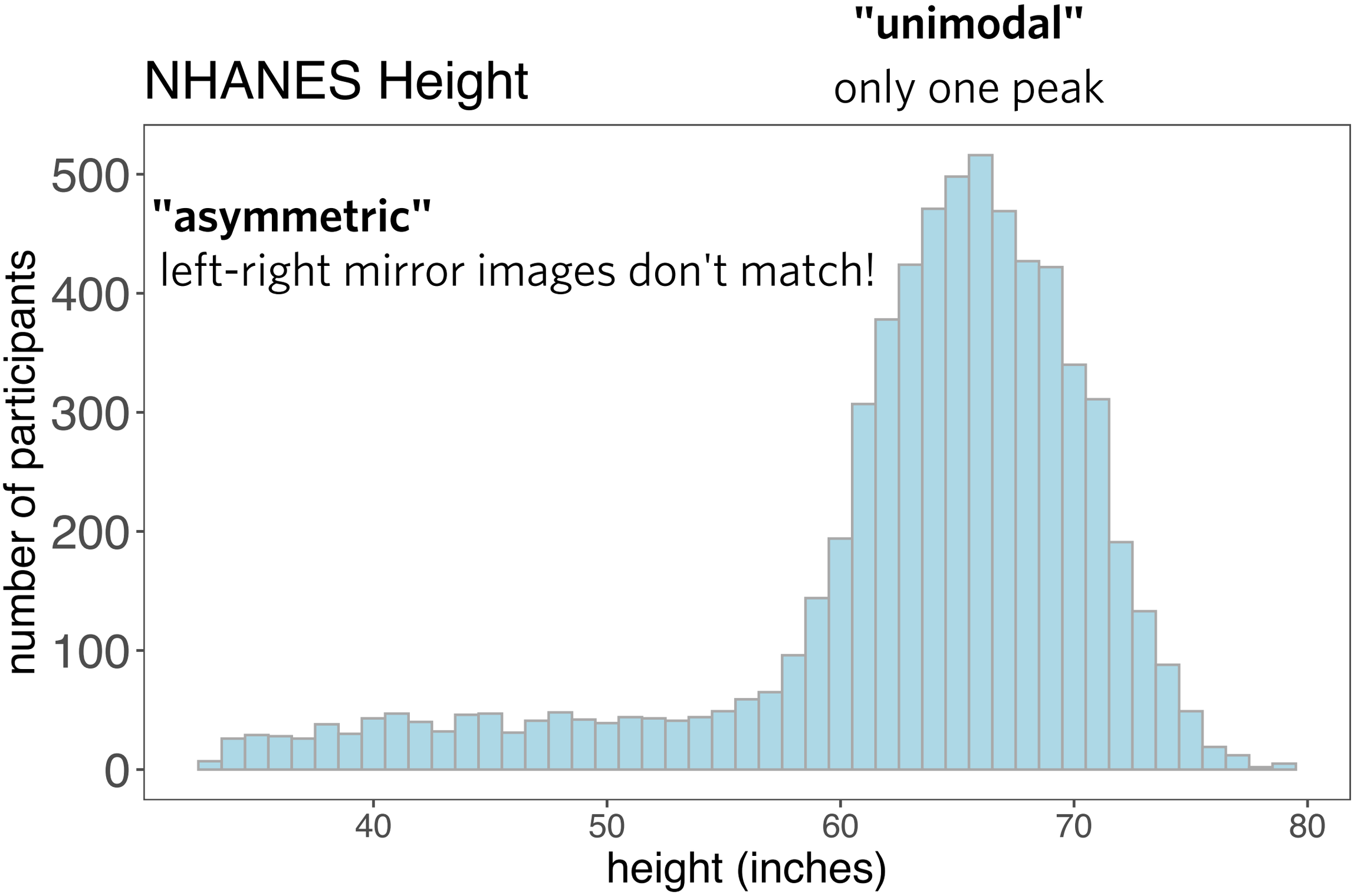
demo

"unimodal"

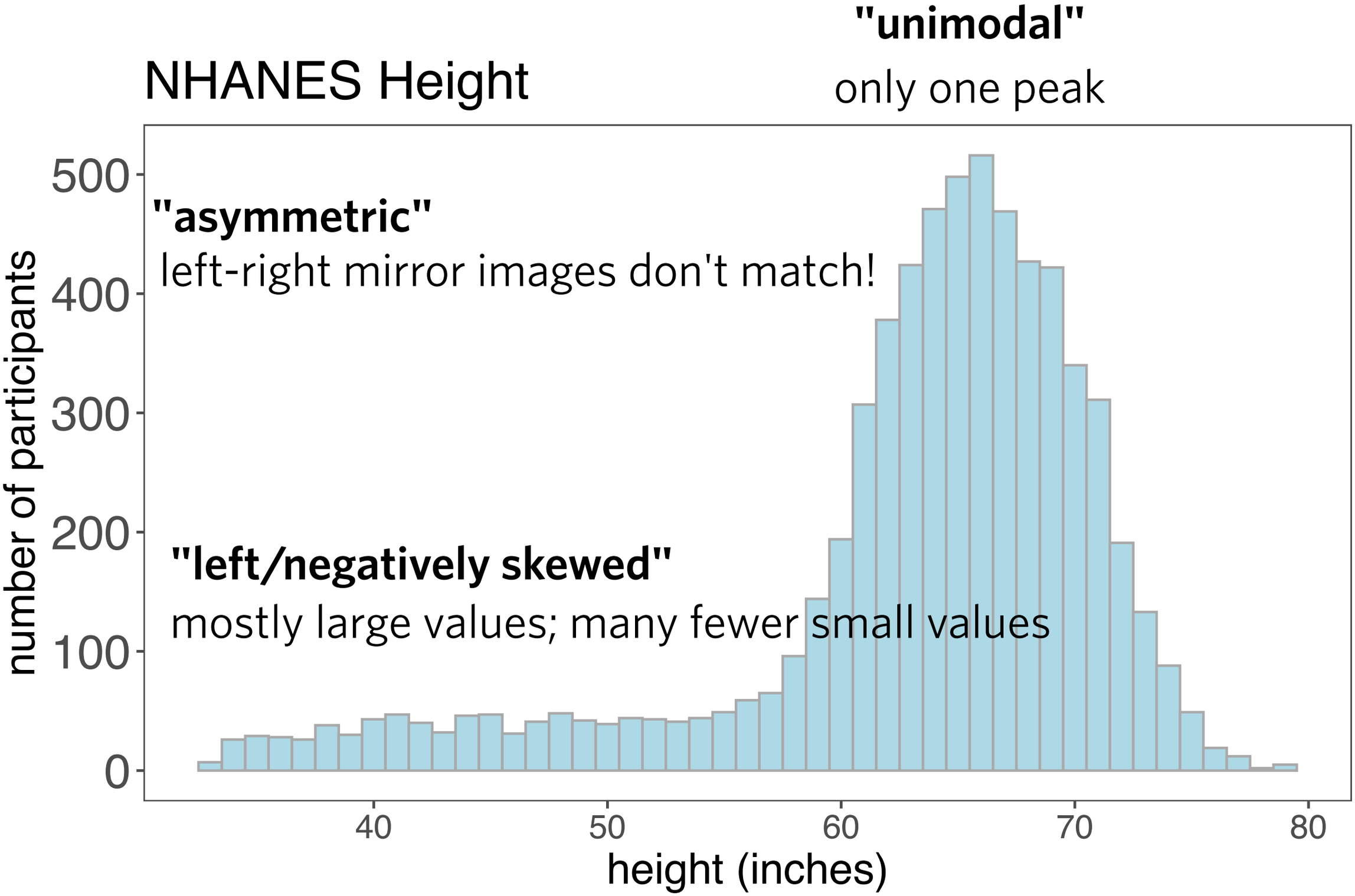
only one peak



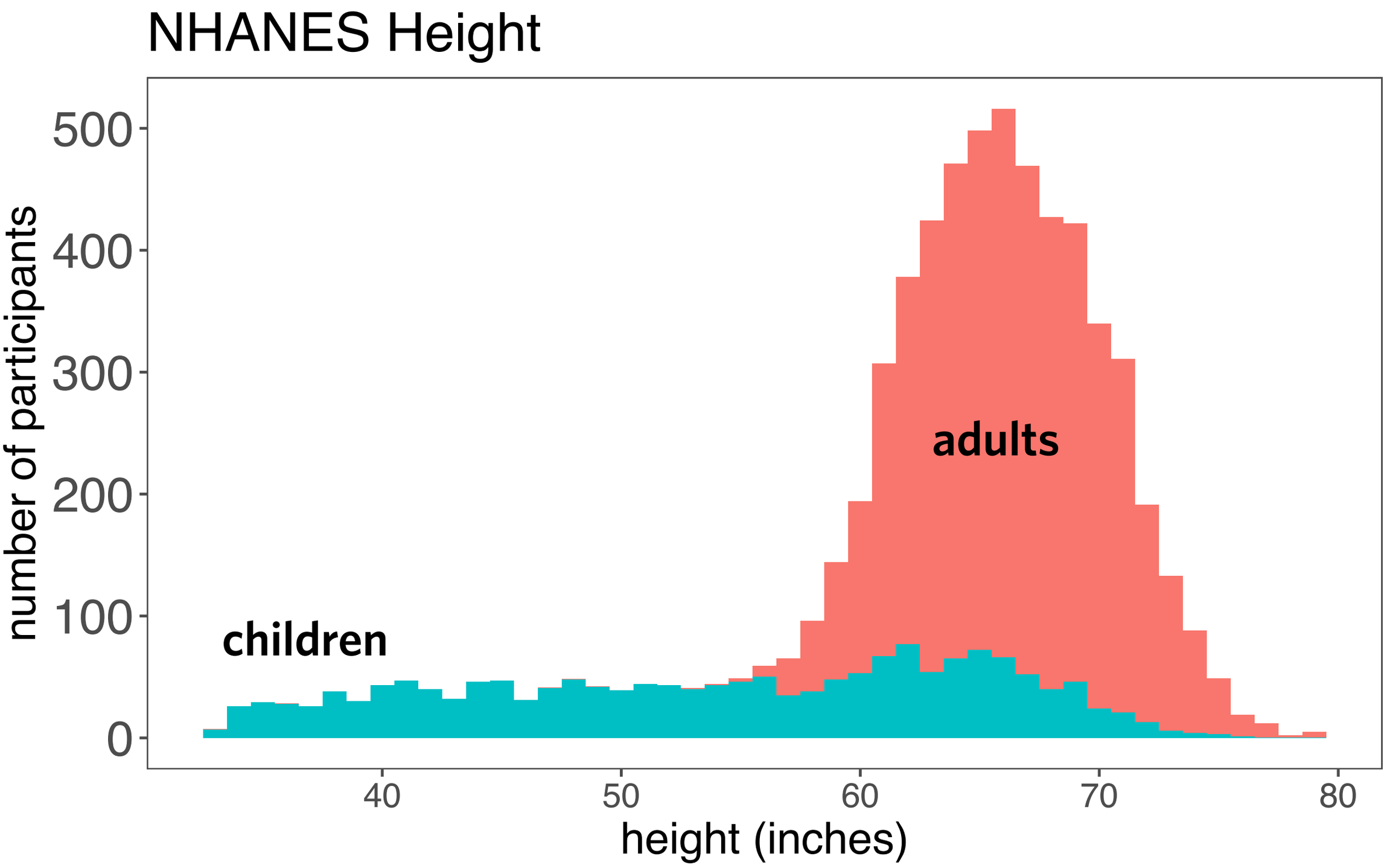
Why does this height distribution look like this?

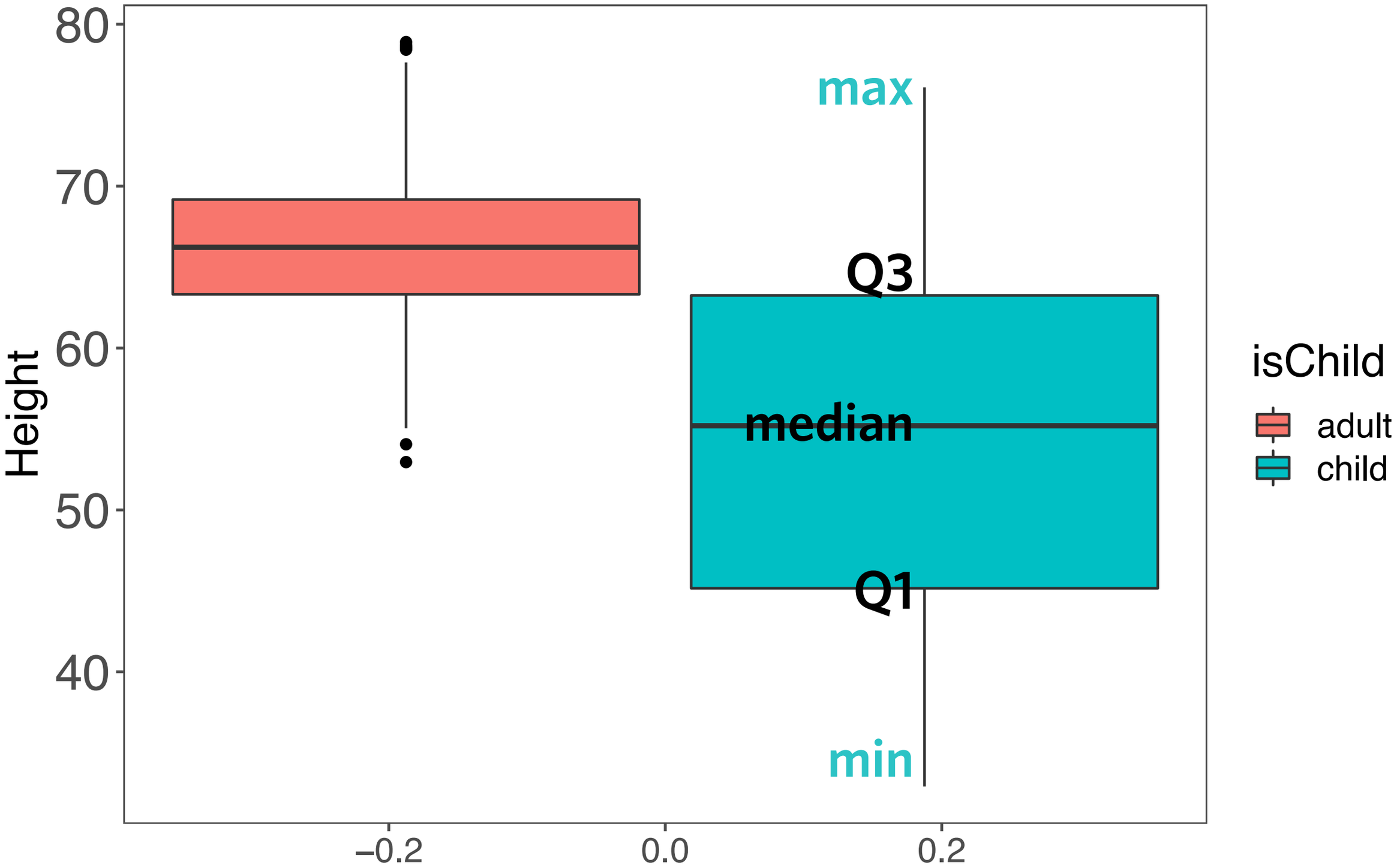


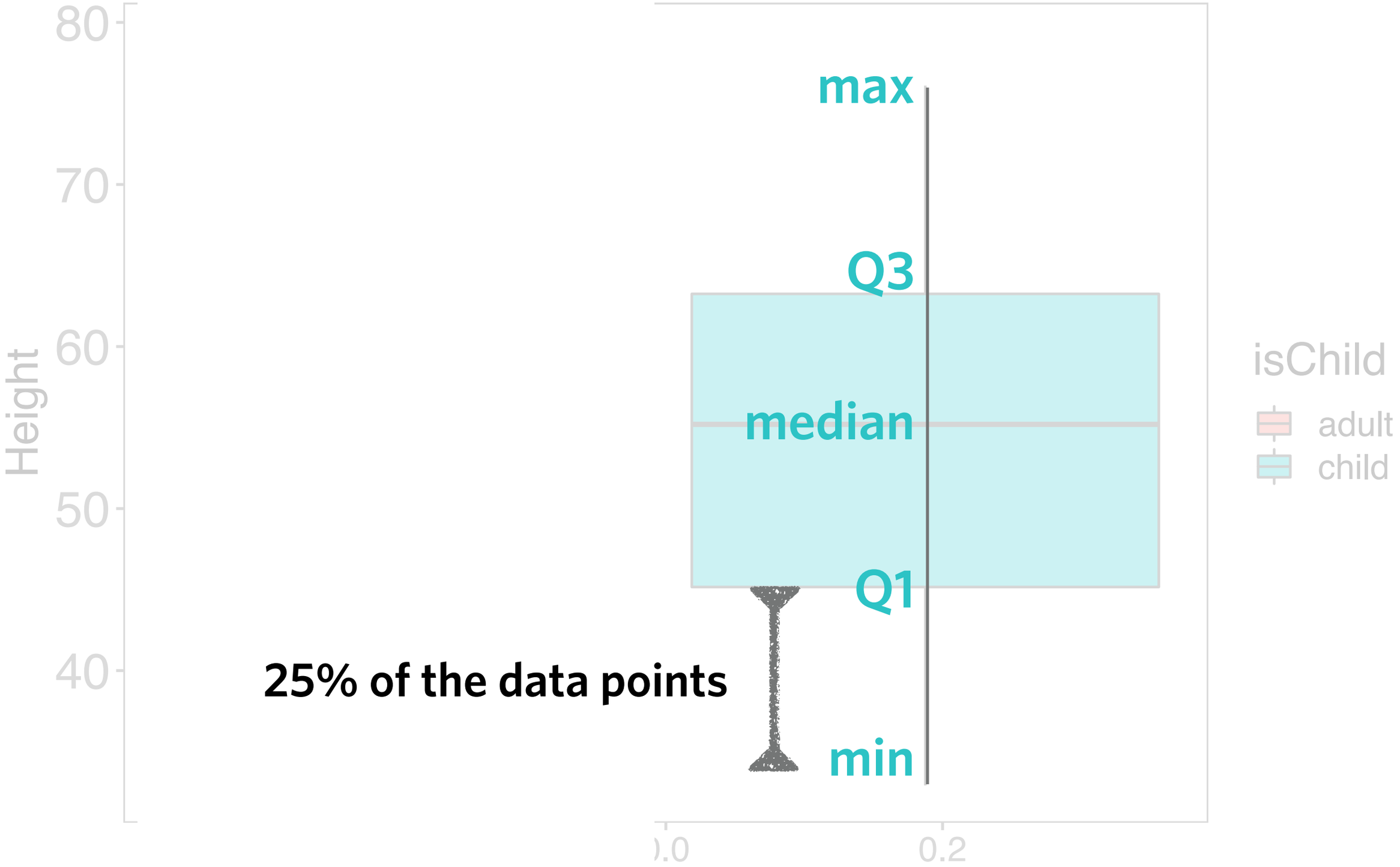
Why does this height distribution look like this?

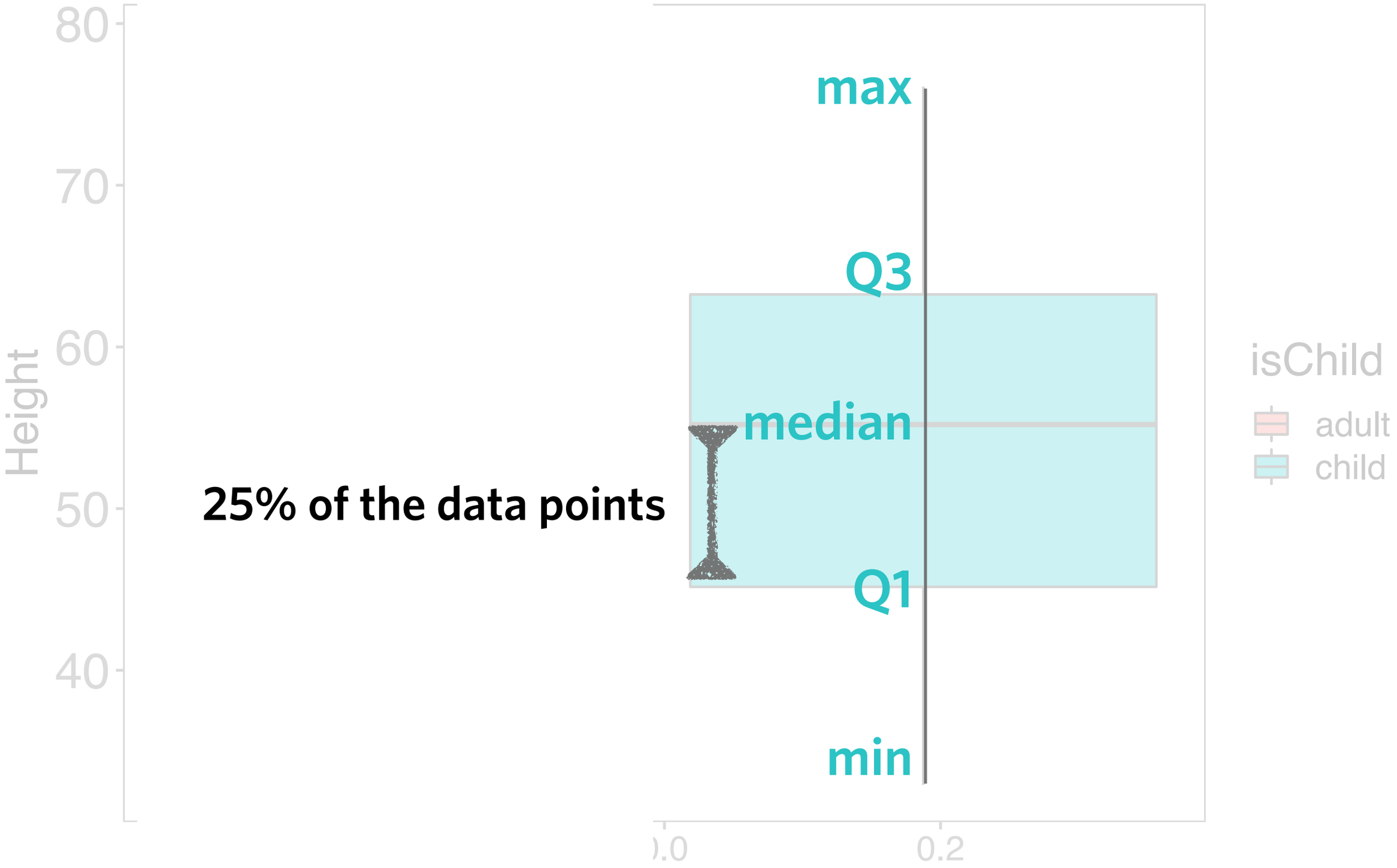


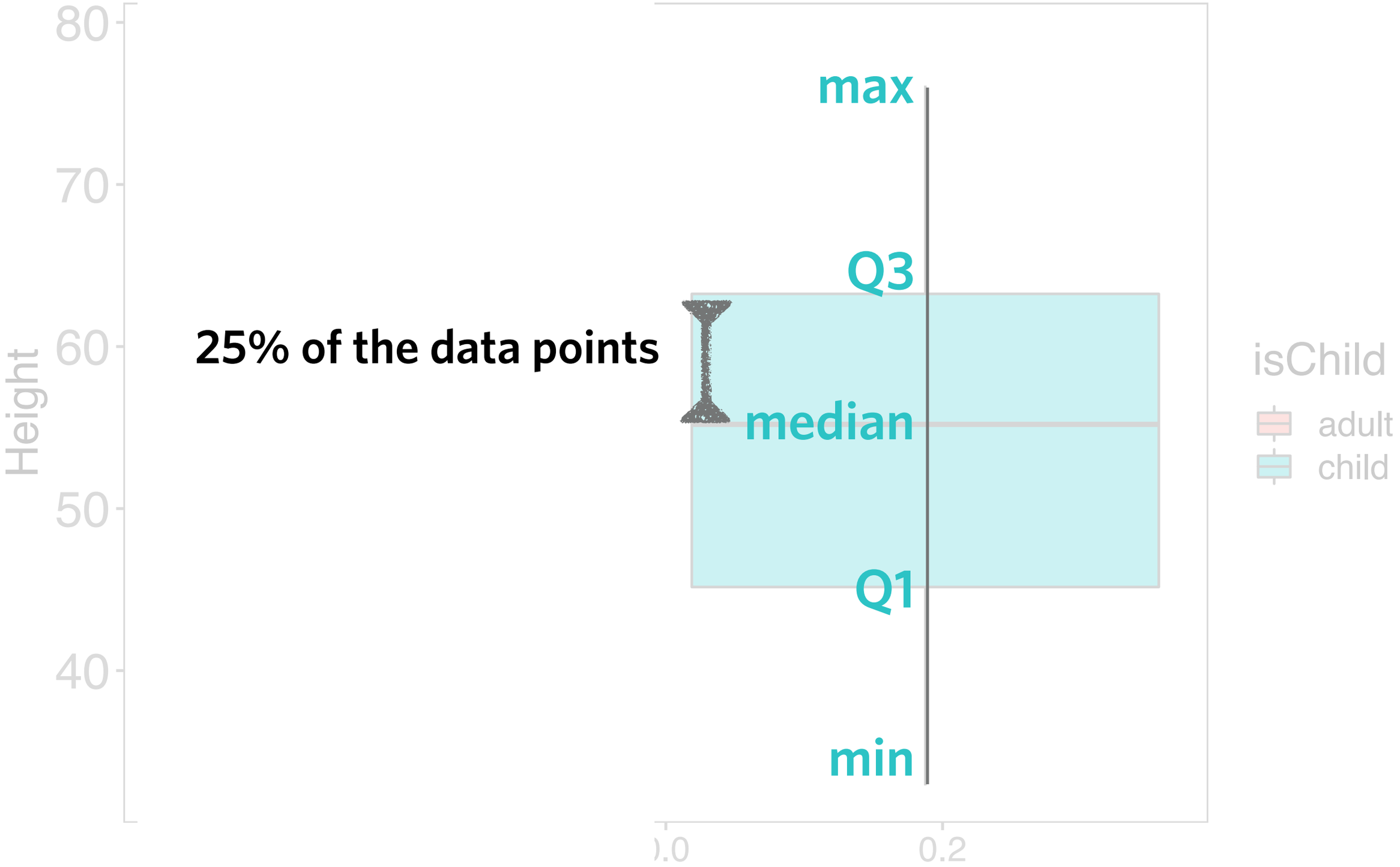
Why does this height distribution look like this?

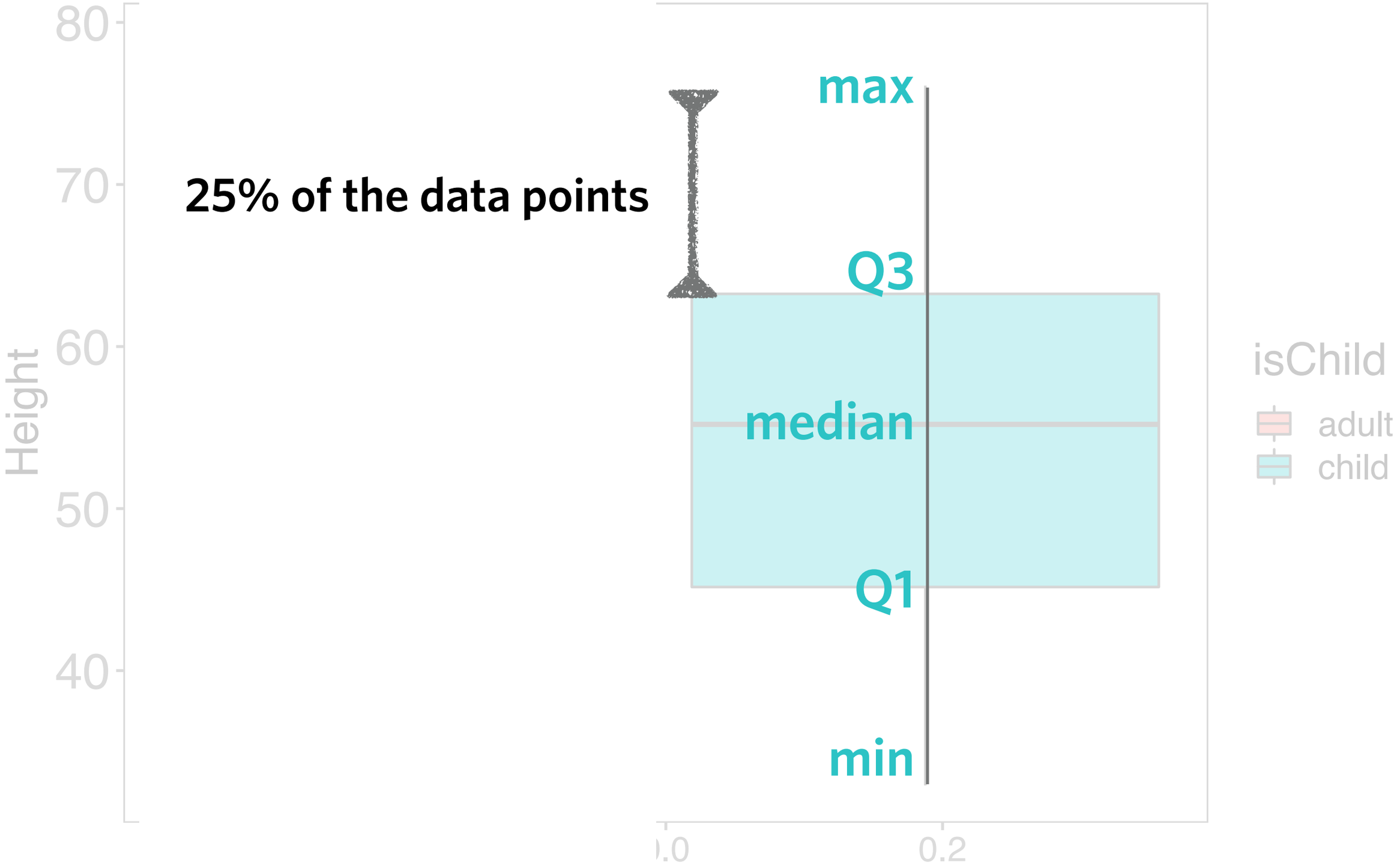

NHANES Height
number of participants
500
400
300
200
100
0
adults
children
40
50
60
70
80
height (inches)

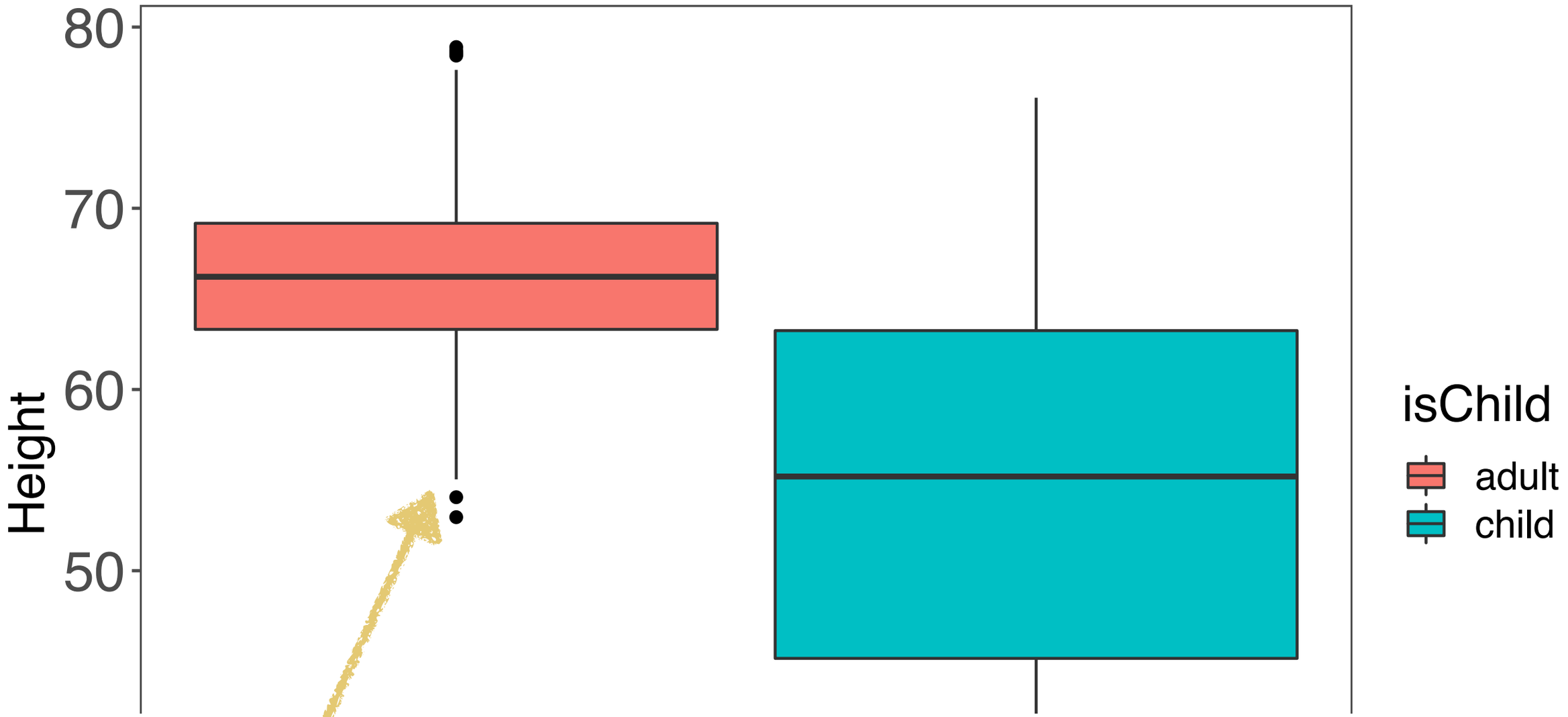

80
70
60
50
40
Height
–0.2
0.0
0.2
max
Q3
median
Q1
min
isChild
adult
child


80
70
60
50
40
Height
max
Q3
median
Q1
min
25% of the data points
0.0
0.2
isChild
adult
child

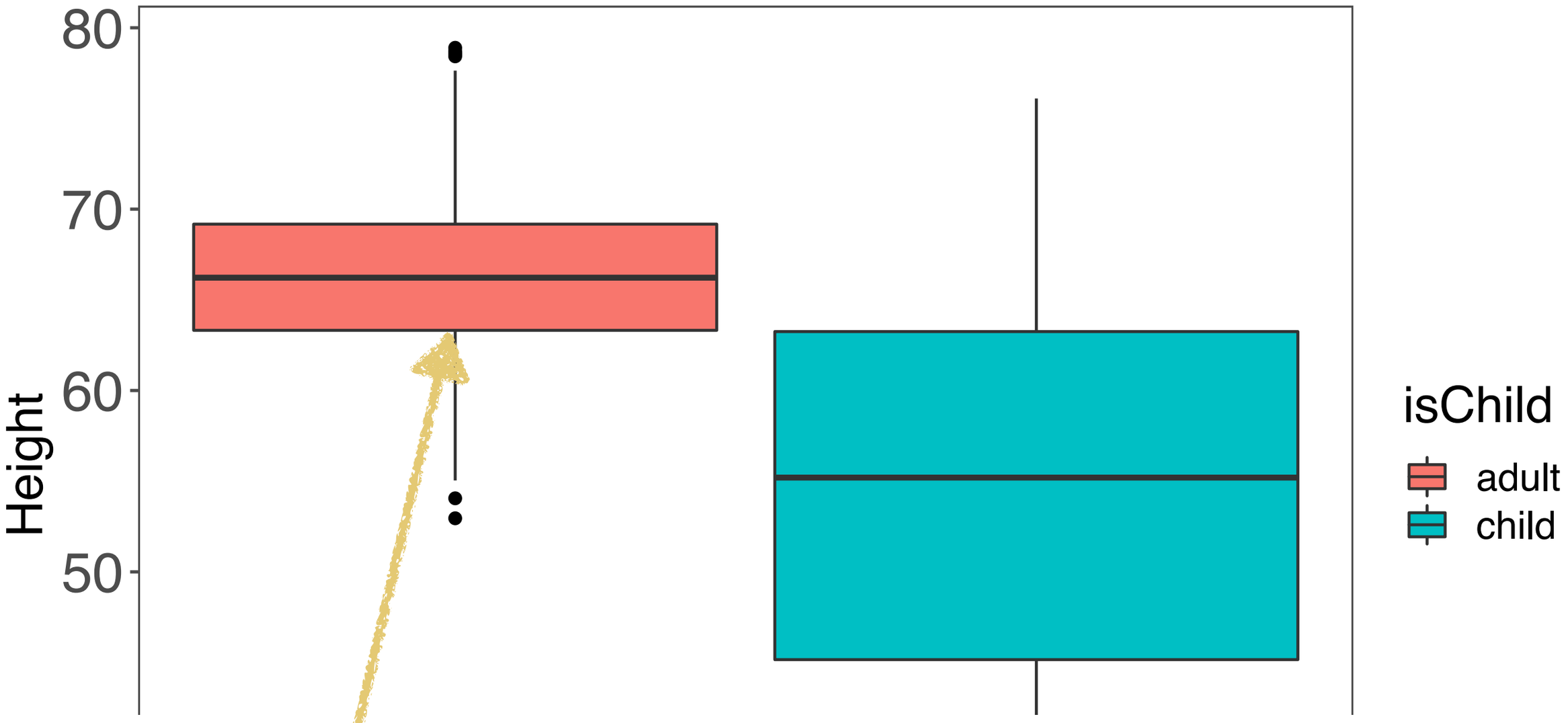

80
70
60
50
40
Height
max
Q3
median
Q1
min
25% of the data points
isChild
adult
child
0.0
0.2


80
70
60
50
40
Height
25% of the data points
max
Q3
median
Q1
min
isChild
adult
child
0.0
0.2

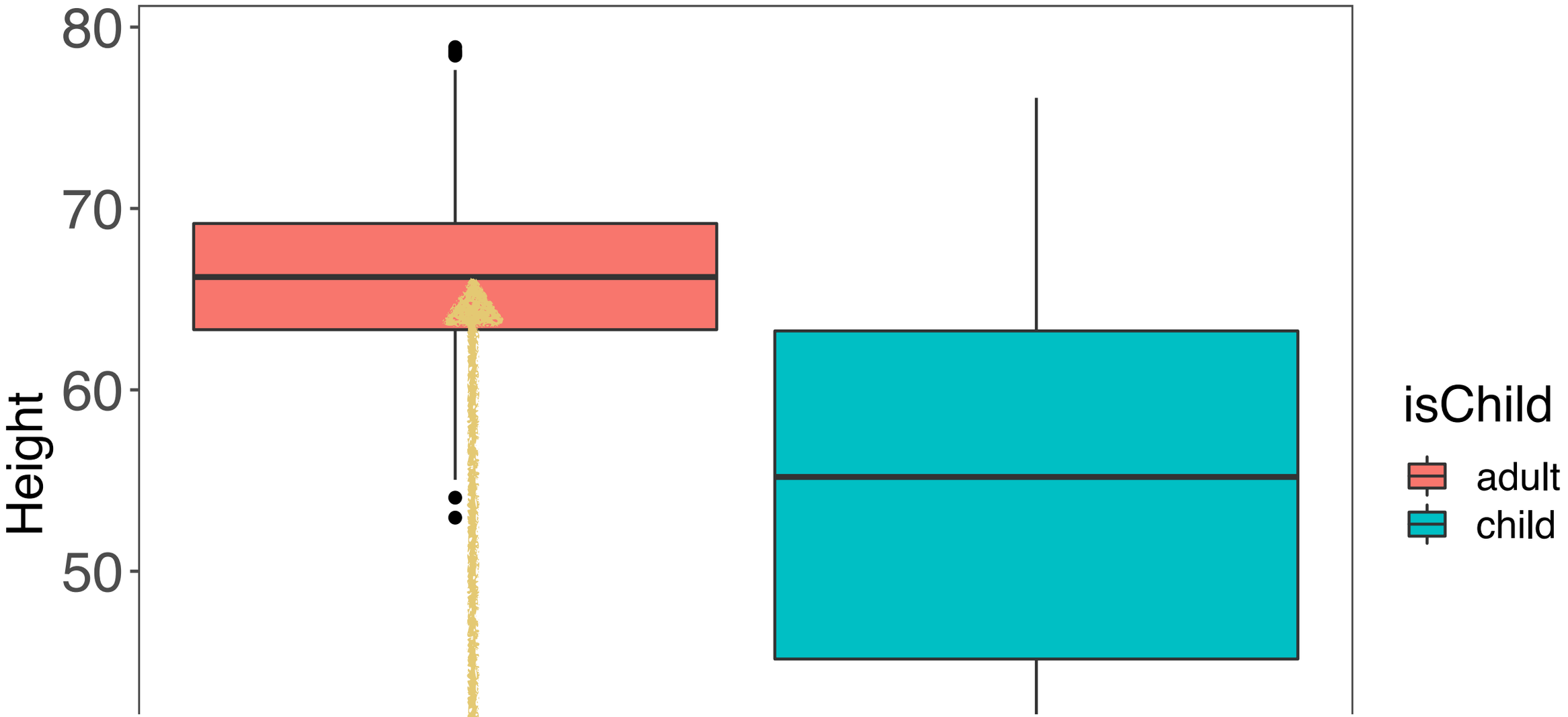

25% of the data points
max
Q3
median
Q1
min
Height
80
70
60
50
40
0.0
0.2
isChild
adult
child



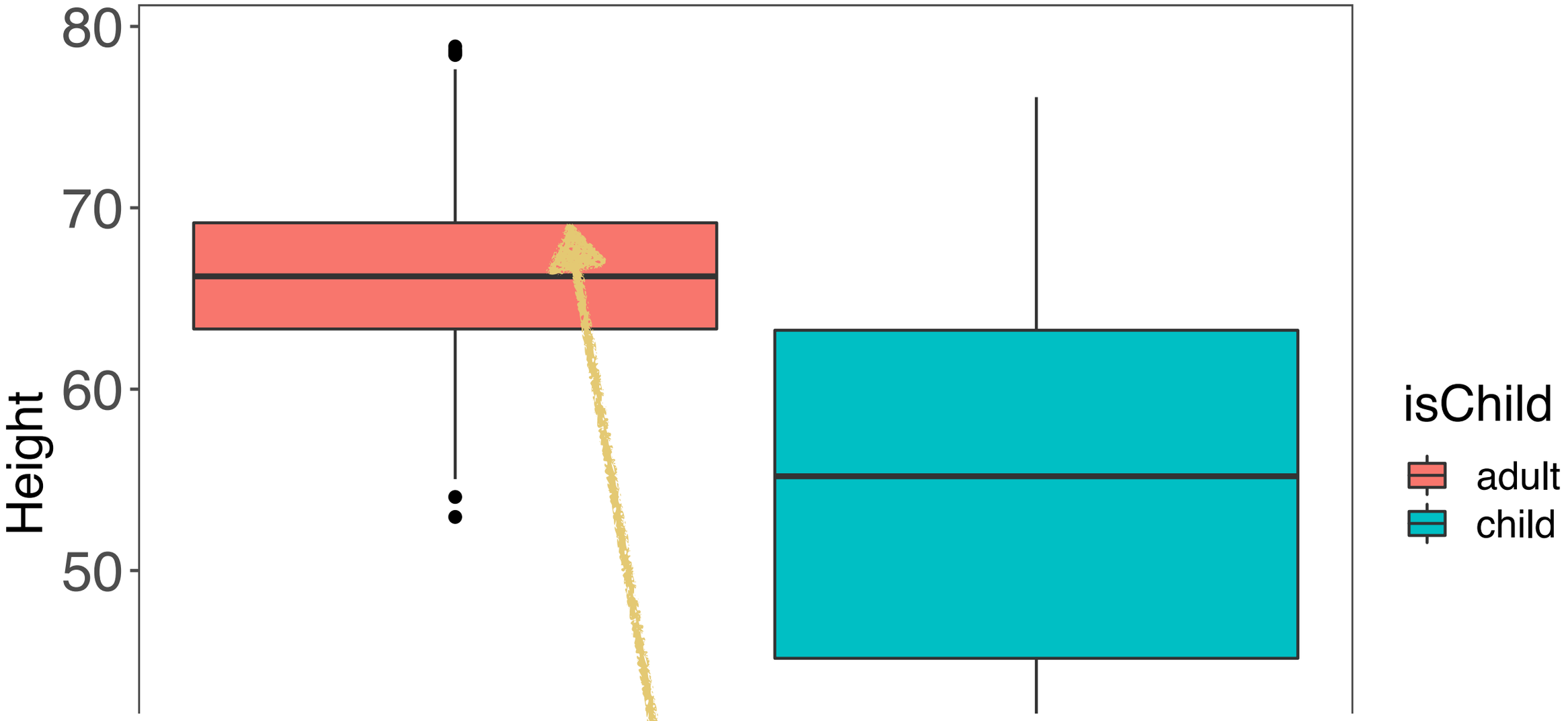
isChild	min	Q1	median	Q3	max	mean	sd	n	missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
adult	52.95276	63.31693	66.22047	69.17323	78.89764	66.27941	3.998837	4790	0
child	32.91339	45.15748	55.19685	63.24803	76.10236	54.21095	10.598281	1691	0



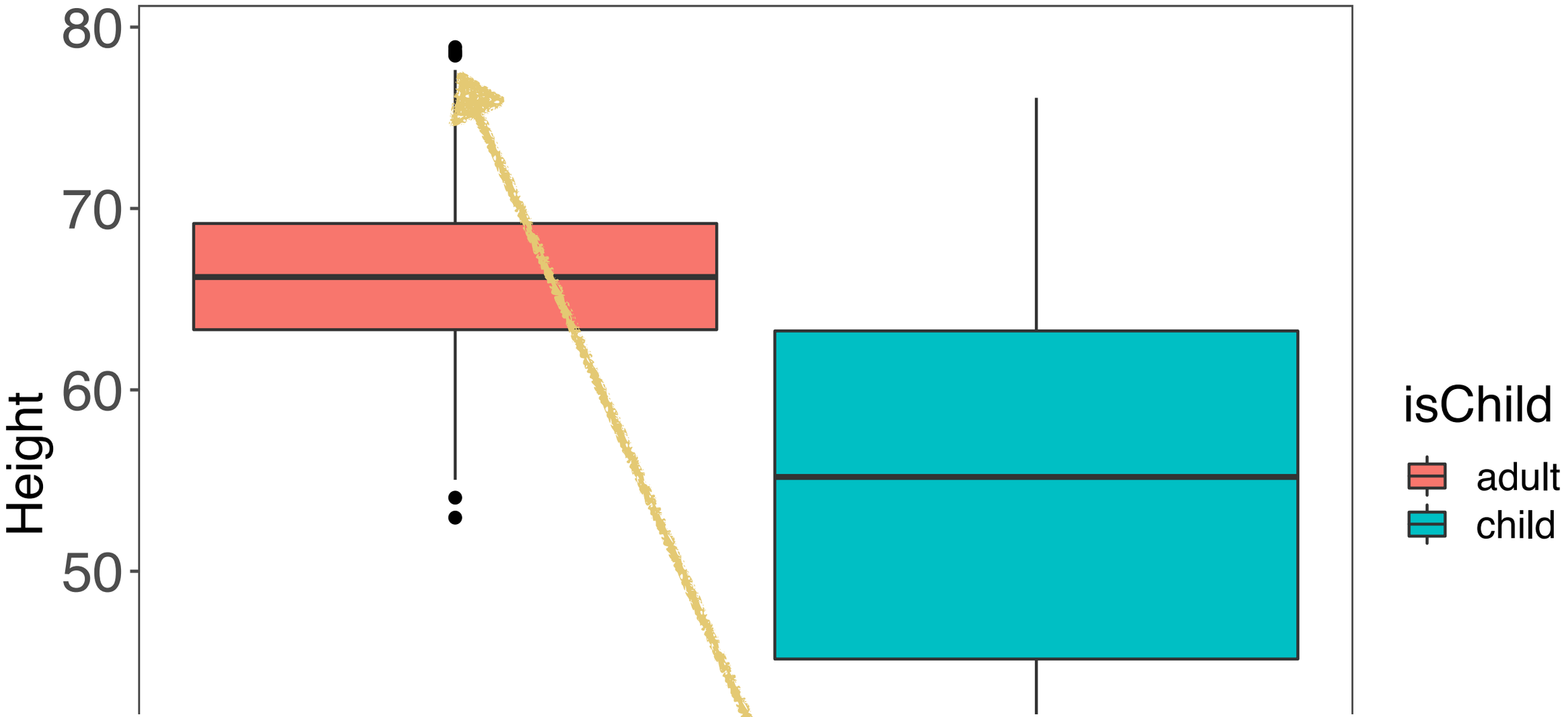
isChild	min	Q1	median	Q3	max	mean	sd	n	missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
adult	52.95276	63.31693	66.22047	69.17323	78.89764	66.27941	3.998837	4790	0
child	32.91339	45.15748	55.19685	63.24803	76.10236	54.21095	10.598281	1691	0



isChild	min	Q1	median	Q3	max	mean	sd	n	missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
adult	52.95276	63.31693	66.22047	69.17323	78.89764	66.27941	3.998837	4790	0
child	32.91339	45.15748	55.19685	63.24803	76.10236	54.21095	10.598281	1691	0



isChild	min	Q1	median	Q3	max	mean	sd	n	missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
adult	52.95276	63.31693	66.22047	69.17323	78.89764	66.27941	3.998837	4790	0
child	32.91339	45.15748	55.19685	63.24803	76.10236	54.21095	10.598281	1691	0



isChild	min	Q1	median	Q3	max	mean	sd	n	missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
adult	52.95276	63.31693	66.22047	69.17323	78.89764	66.27941	3.998837	4790	0
child	32.91339	45.15748	55.19685	63.24803	76.10236	54.21095	10.598281	1691	0

Today

Mini-Review Session #1

Data visualization and why it matters

Thinking about the data-generating process

Practical tips on how to learn stats w/ R

Starring Jarrett Lovelett & Zhe Huang!

Due This Week

4	Apr 19	**Visualizing data, III** *Before:* Chapter 4 *During:* Lab 2C	**Review Session** *Before:* None *During:* Wrap-up Lab 2	Quiz 2	Project Milestone 2 Due (Exploratory data visualization)

If you have not completed the Chapter 4 modules, please do so ASAP!

Due This Week

4	Apr 19	**Visualizing data, III** *Before:* Chapter 4 *During:* Lab 2C	**Review Session** *Before:* None *During:* Wrap-up Lab 2	Quiz 2 Project Milestone 2 Due (Exploratory data visualization)

If you have not completed the Chapter 4 modules, please do so ASAP!

Due This Week

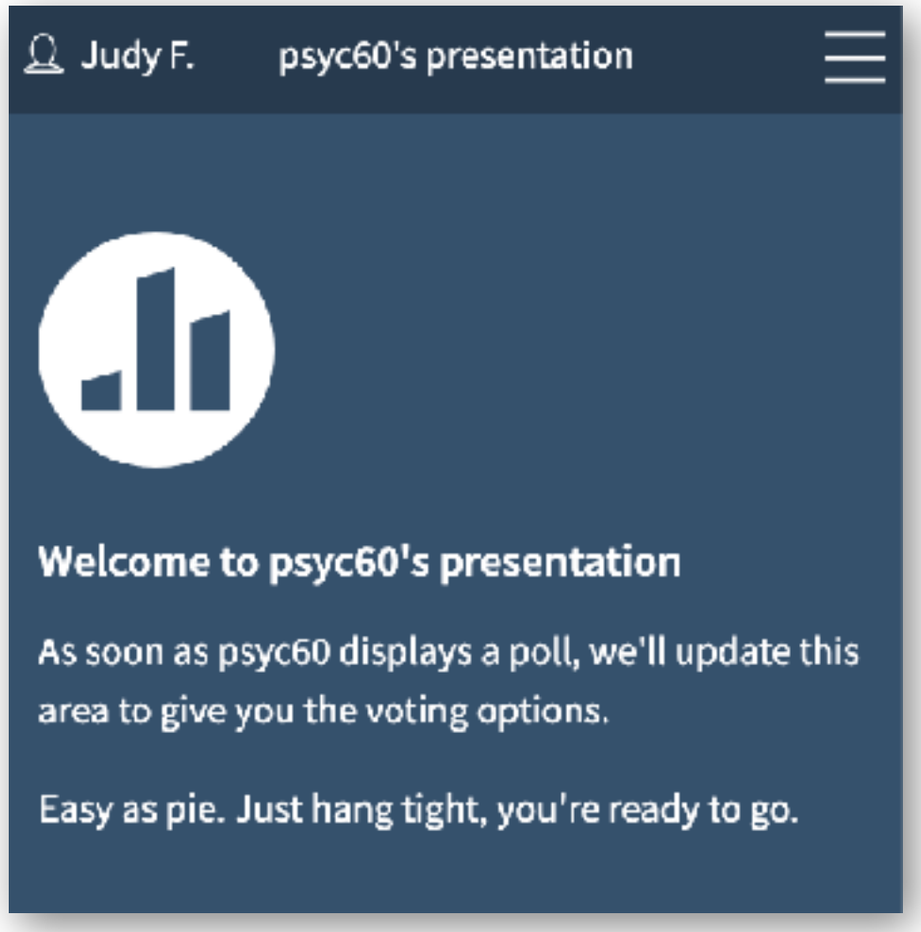
4	Apr 19	**Visualizing data, III** *Before:* Chapter 4 *During:* Lab 2C	**Review Session** *Before:* None *During:* Wrap-up Lab 2	Quiz 2	Project Milestone 2 Due (Exploratory data visualization)

Use the worksheet from discussion section this week on REFINING your data visualizations revise your exploratory data visualizations before submitting them this Friday!

Debrief

1. Take your phone or laptop out.
2. Make sure you are connected to the internet.
3. Open any web browser & type in this URL:
   PollEv.com/psyc60
4. Make sure to log in to your account using your UCSD email & name as it appears in Canvas.

You should see something like this



PSYC 60: Lab2C | General Impressions

When survey is active, respond at pollev.com/psyc60

0 done

0 underway

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**