

# **UNIT 1: ONE VARIABLE DATA**

# **WHAT IS A RANDOM VARIABLE?**

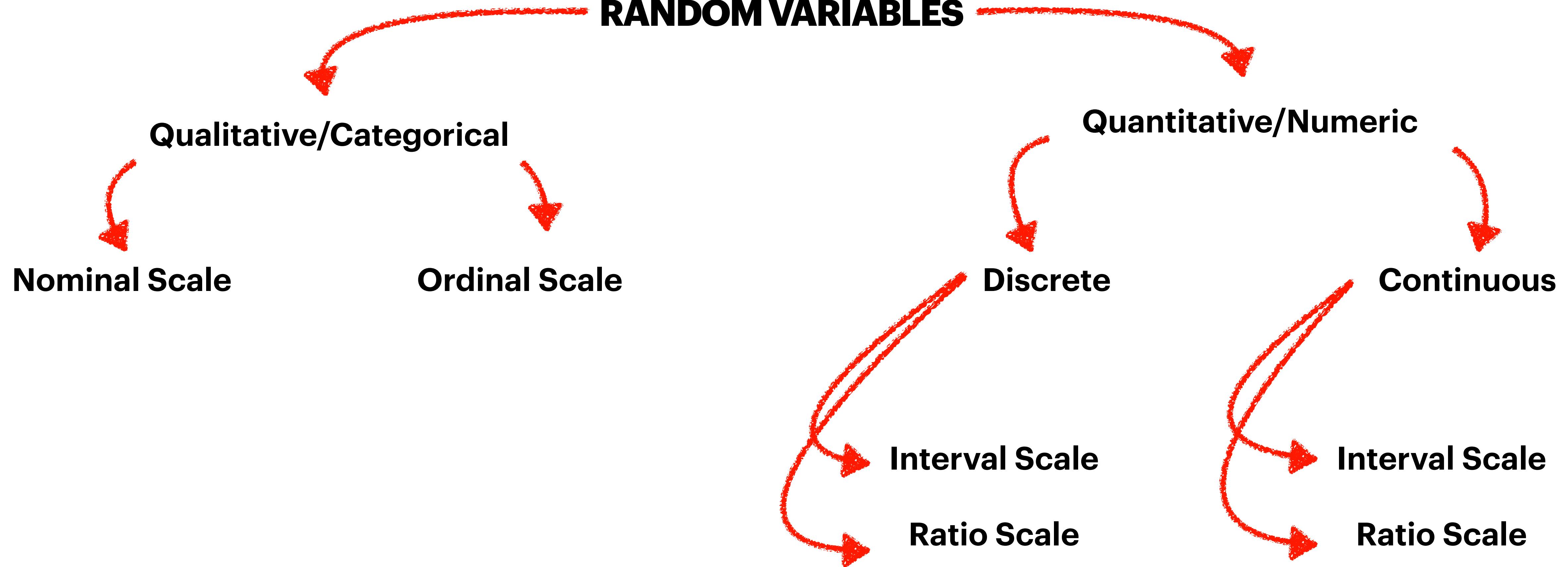
# TIDY DATA

VARIABLES

OBSERVATIONS

Observation #	Type	Leaf Shape	Lifespan	
I	Begonia	Oblique	3 Years	
II	Pine	Acicular	1000 Years	
III	Water Oak	Spatulate	50 Years	

## **RANDOM VARIABLES**



# **QUALITATIVE DATA**

- 1. How can we summarize one variable qualitative data?**
- 2. How can we visualize one variable qualitative data?**

# QUALITATIVE DATA

## 1. How can we summarize one variable qualitative data?

Tidy Data

Treat	Baker Name	Price	Bakery
Banana Muffin	Flash	\$1.99	A
Sugar Cookie	Speedy	\$0.50	B
Sugar Cookie	Flash	\$0.60	C
Cantaloupe Danish	Flash	\$1.50	C



Tabulated Data

Treat	Frequency
Banana Muffin	60
Sugar Cookie	34
Lard Doughnut	45
Cantaloupe Danish	23
Leaf Cupcake	4



# QUALITATIVE DATA

## 1. How can we summarize one variable qualitative data?

Proportion is also referred to as relative frequency.



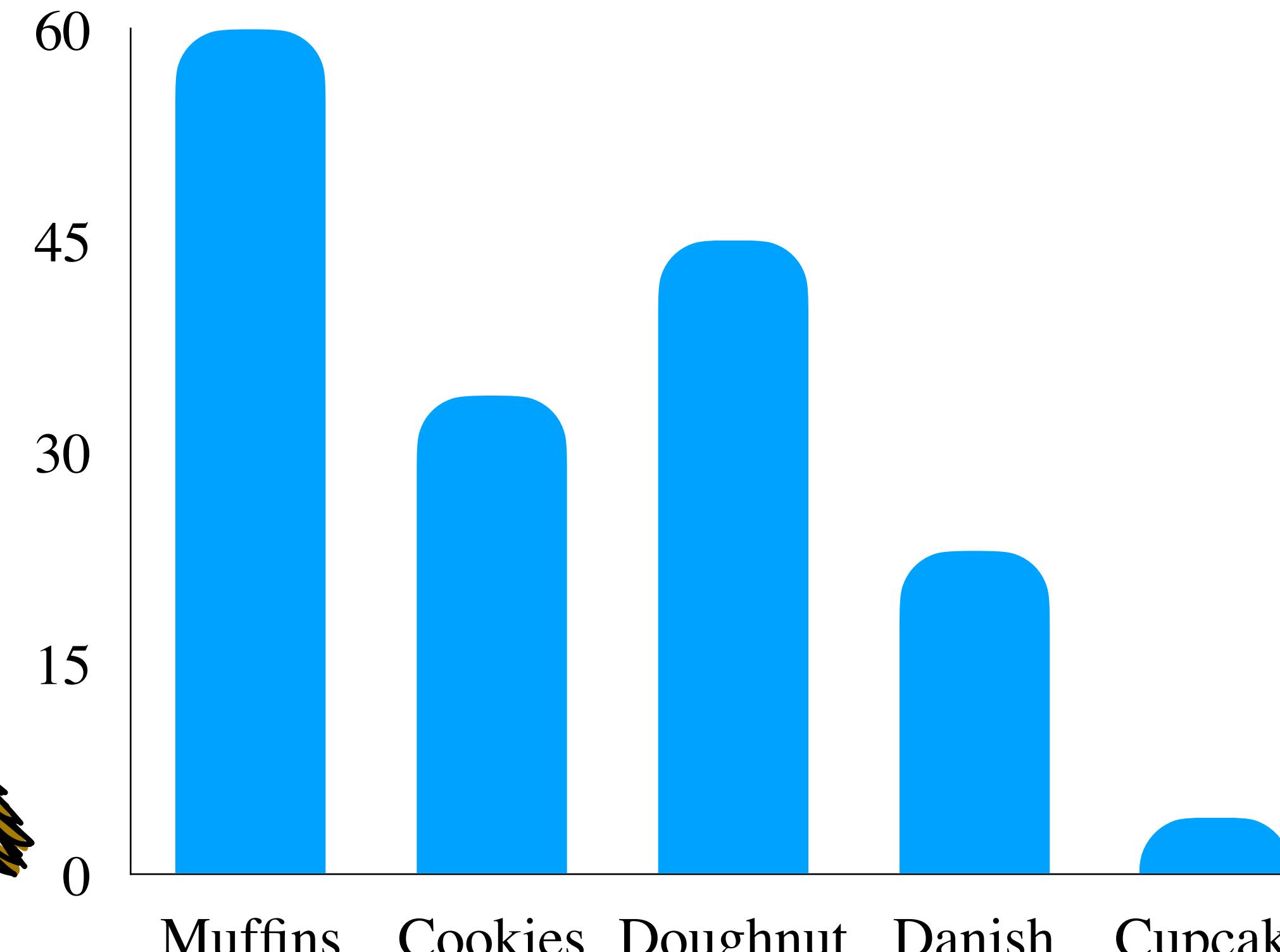
Treat	Frequency
Banana Muffin	60
Sugar Cookie	34
Lard Doughnut	45
Cantaloupe Danish	23
Leaf Cupcake	4



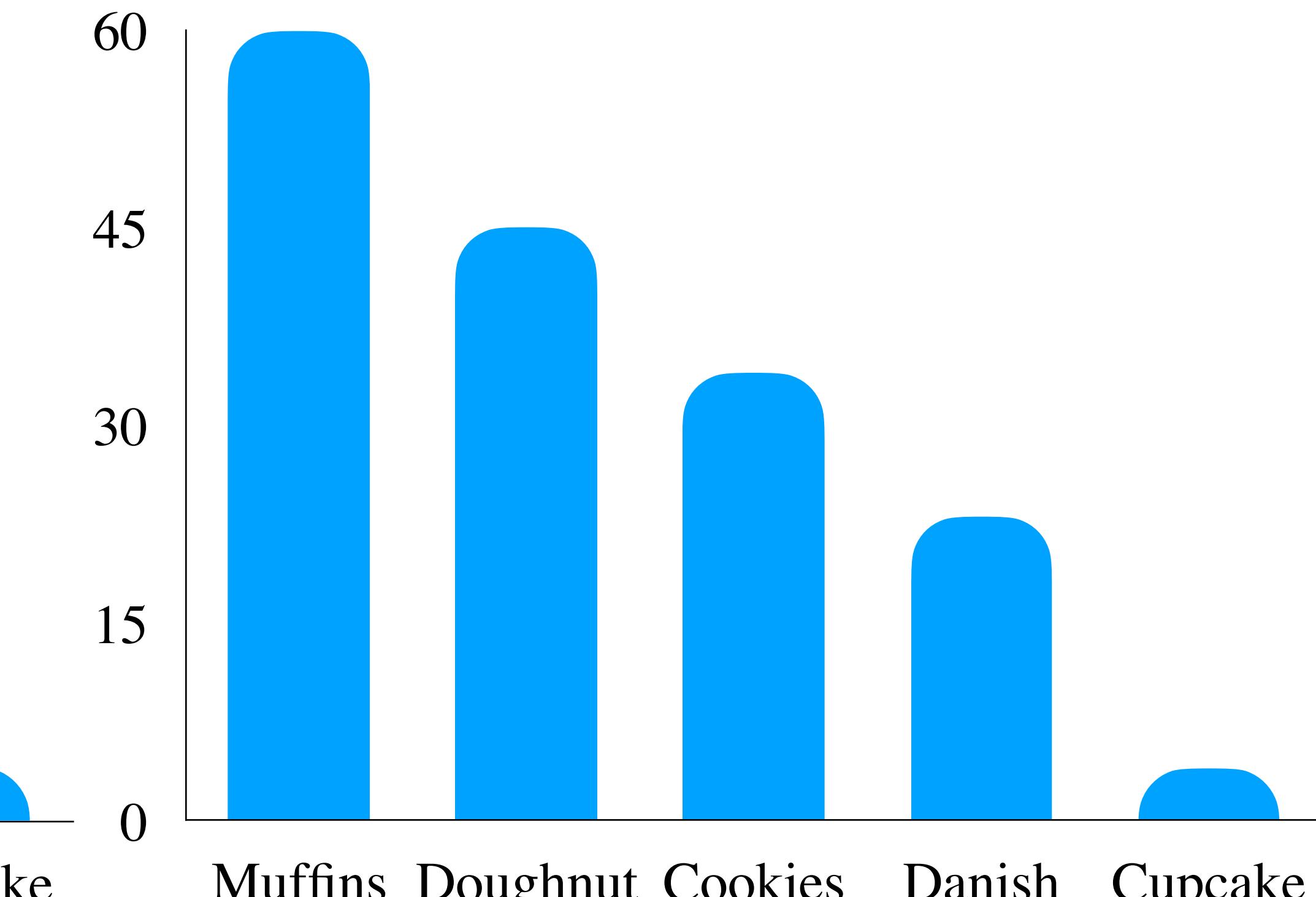
Treat	Proportion
Banana Muffin	0.36
Sugar Cookie	0.20
Lard Doughnut	0.27
Cantaloupe Danish	0.14
Leaf Cupcake	0.02

# QUALITATIVE DATA

2. How can we visualize one variable quantitative data?



Bar Chart

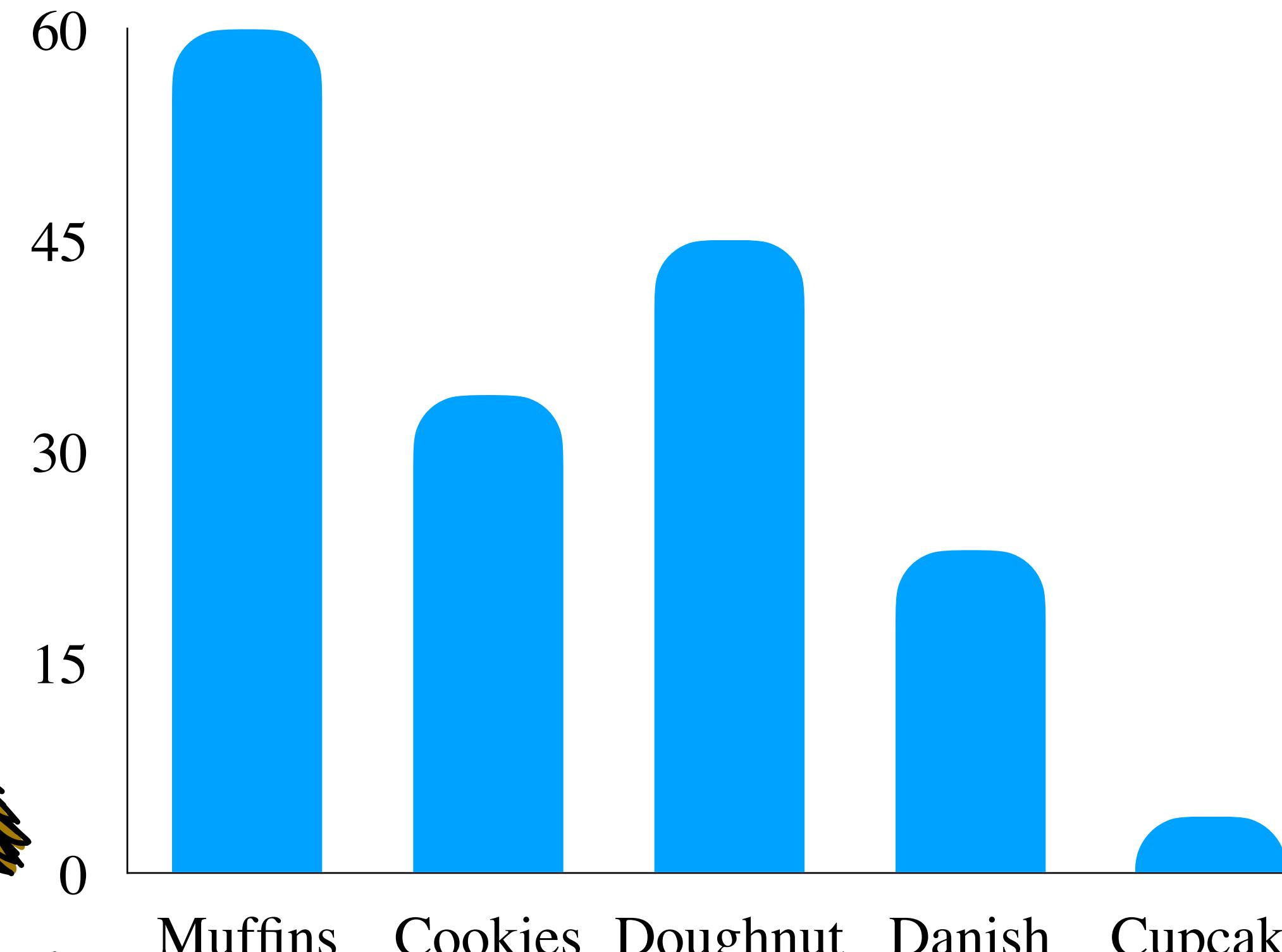


Pareto Chart

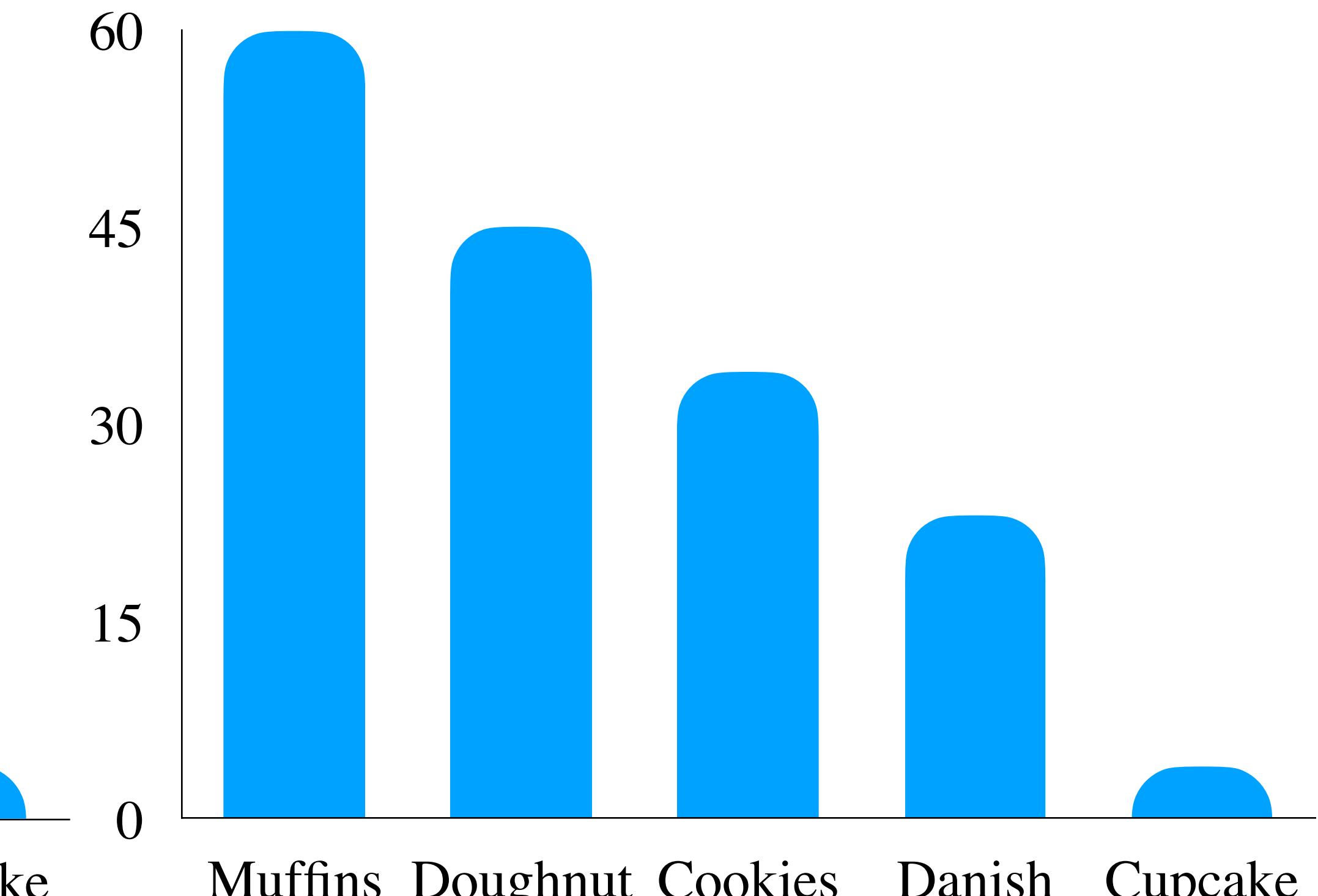


# QUALITATIVE DATA

2. How can we visualize one variable quantitative data?



Bar Chart

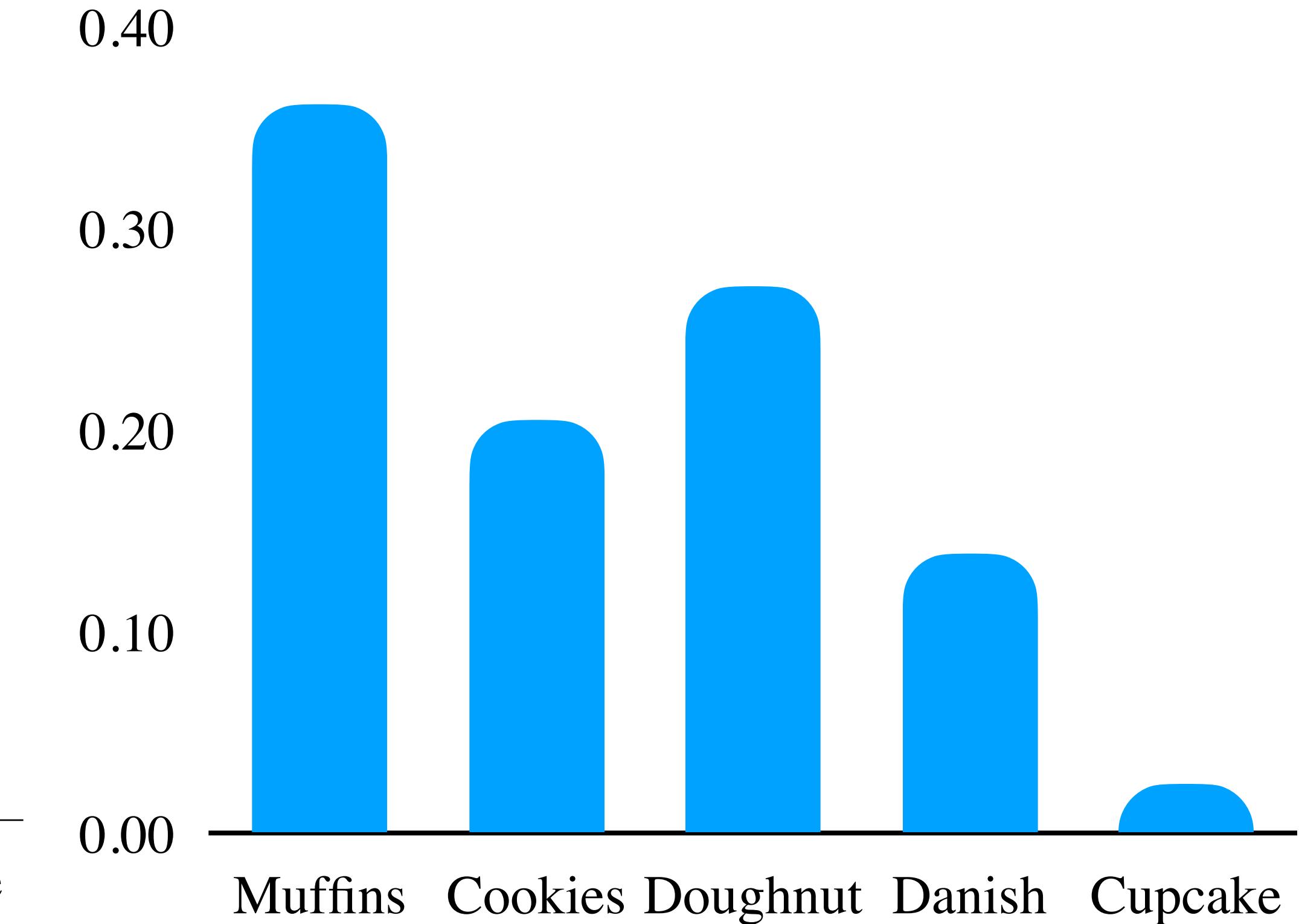
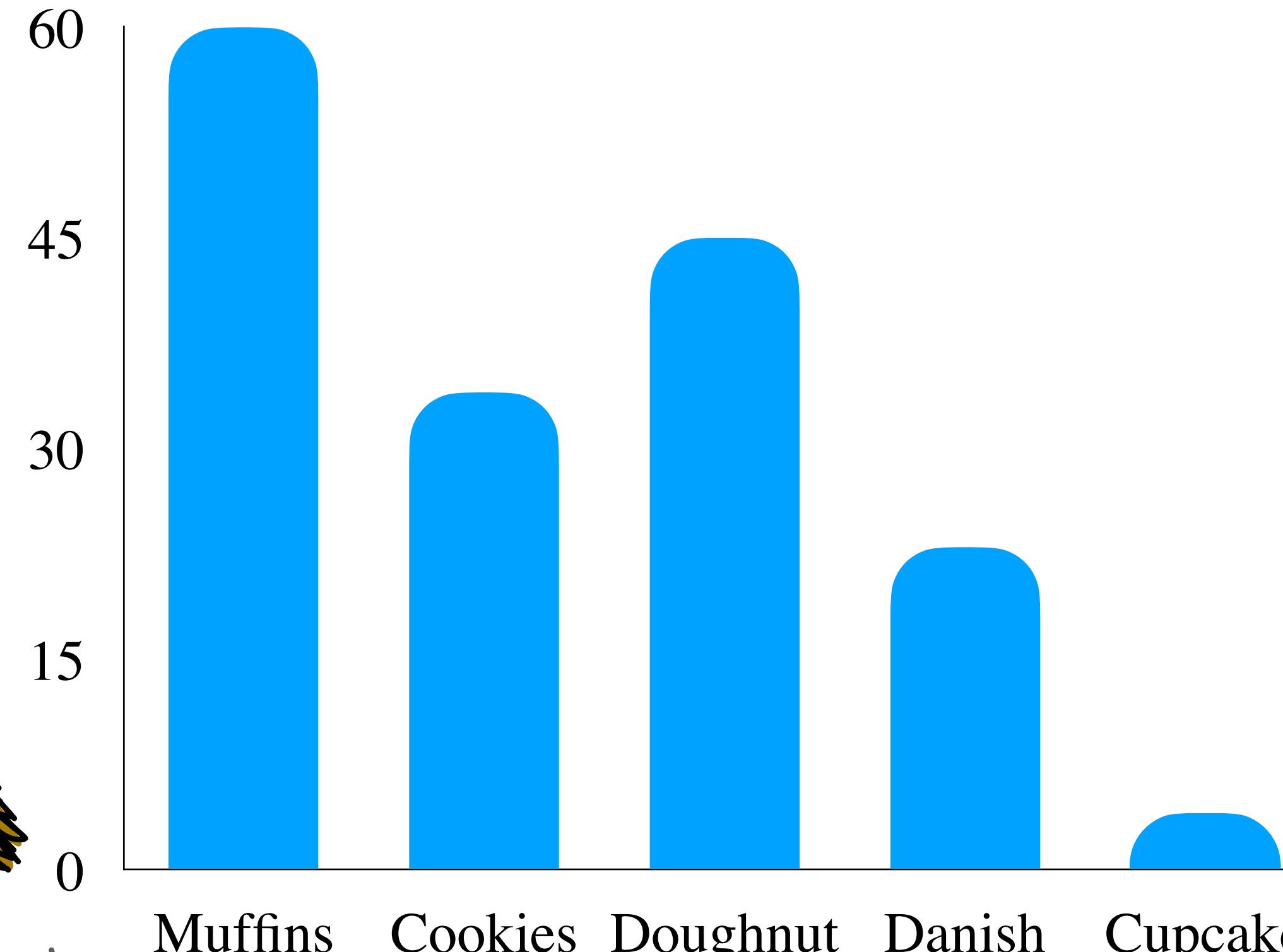


Pareto Chart



# QUALITATIVE DATA

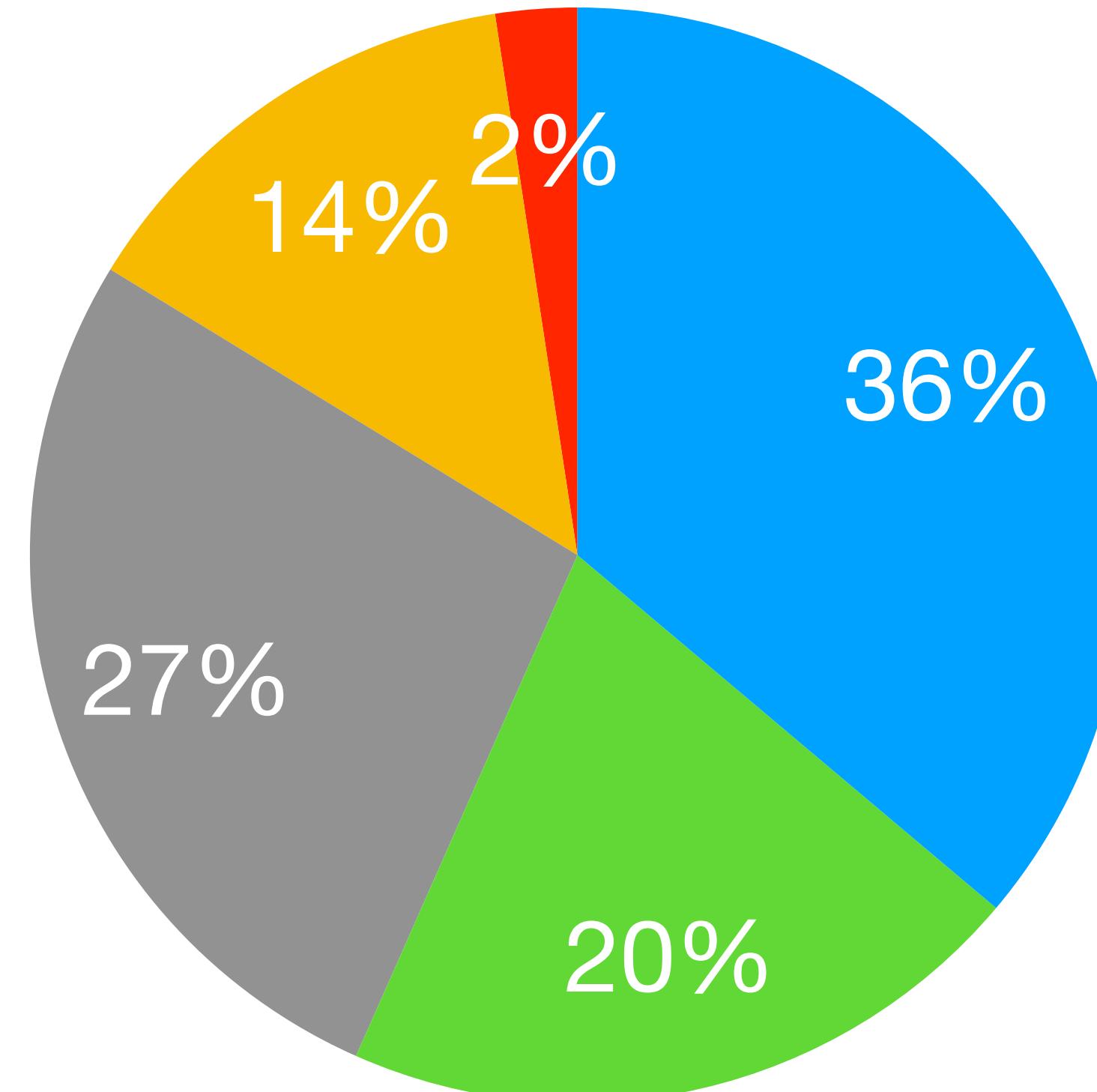
2. How can we visualize one variable quantitative data?



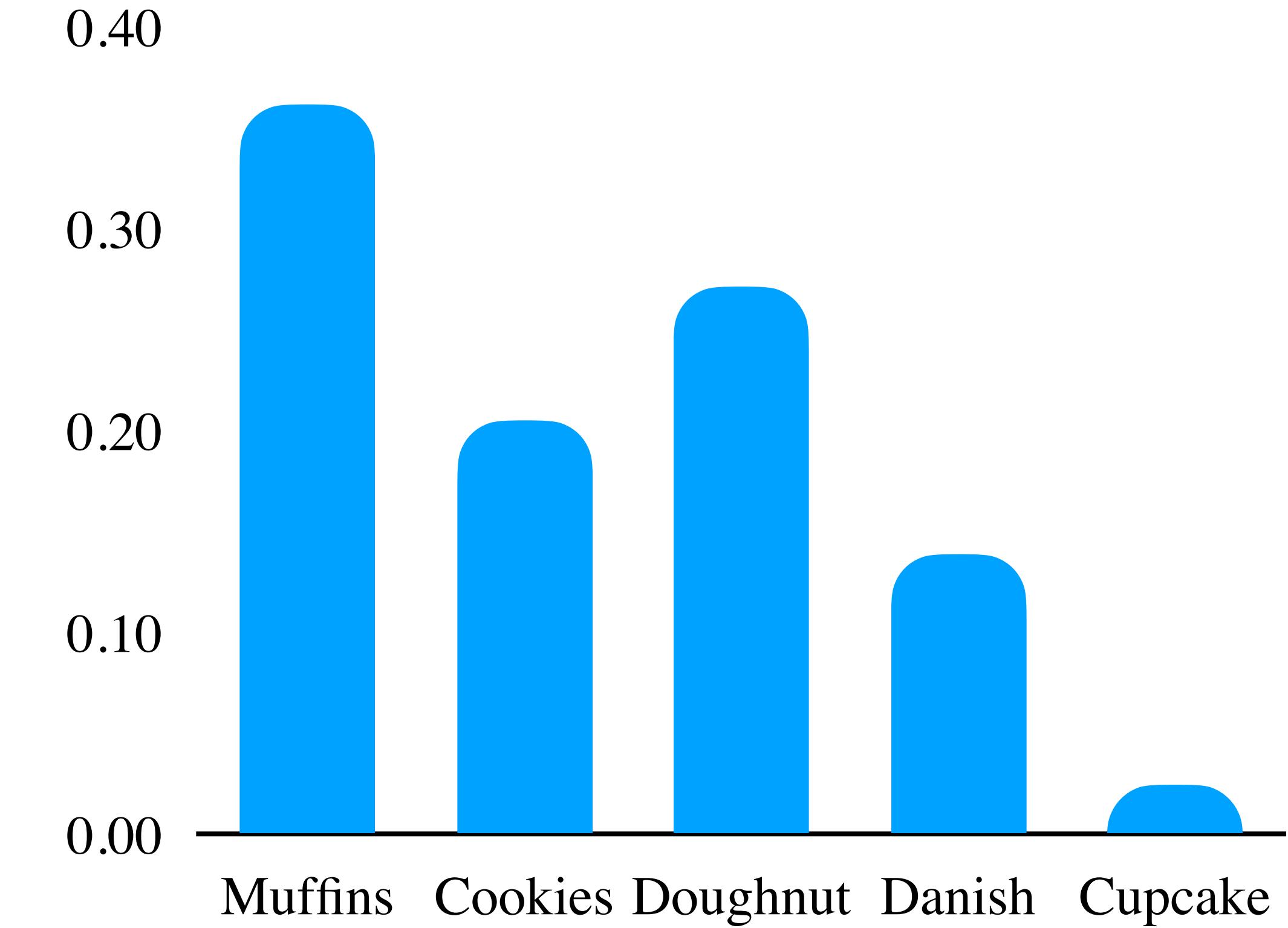
Frequencies vs. Relative Frequencies

# QUALITATIVE DATA

2. How can we visualize one variable quantitative data?



Relative Frequencies



# QUALITATIVE DATA

Comparing a qualitative variable across a quantitative variable

Treat	Frequency
Banana Muffin	60
Sugar Cookie	34
Lard Doughnut	45
Cantaloupe Danish	23
Leaf Cupcake	4



Bakery	Frequency
A	63
B	103

	Bakery A	Bakery B	
Banana Muffin	17	43	60
Sugar Cookie	13	21	34
Lard Doughnut	11	34	45
Cantaloupe Danish	18	5	23
Leaf Cupcake	4	0	4
	63	103	166

# QUALITATIVE DATA

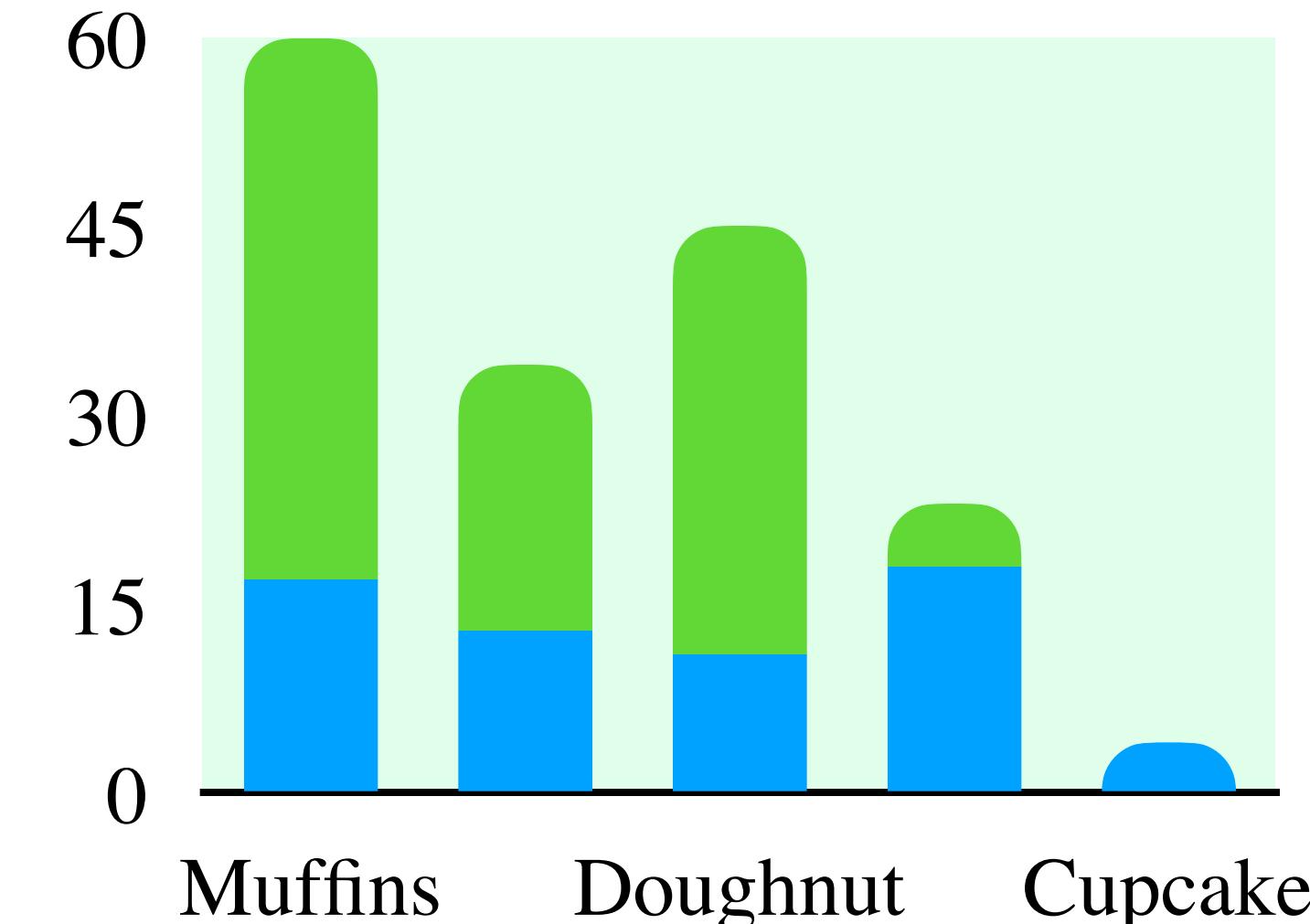
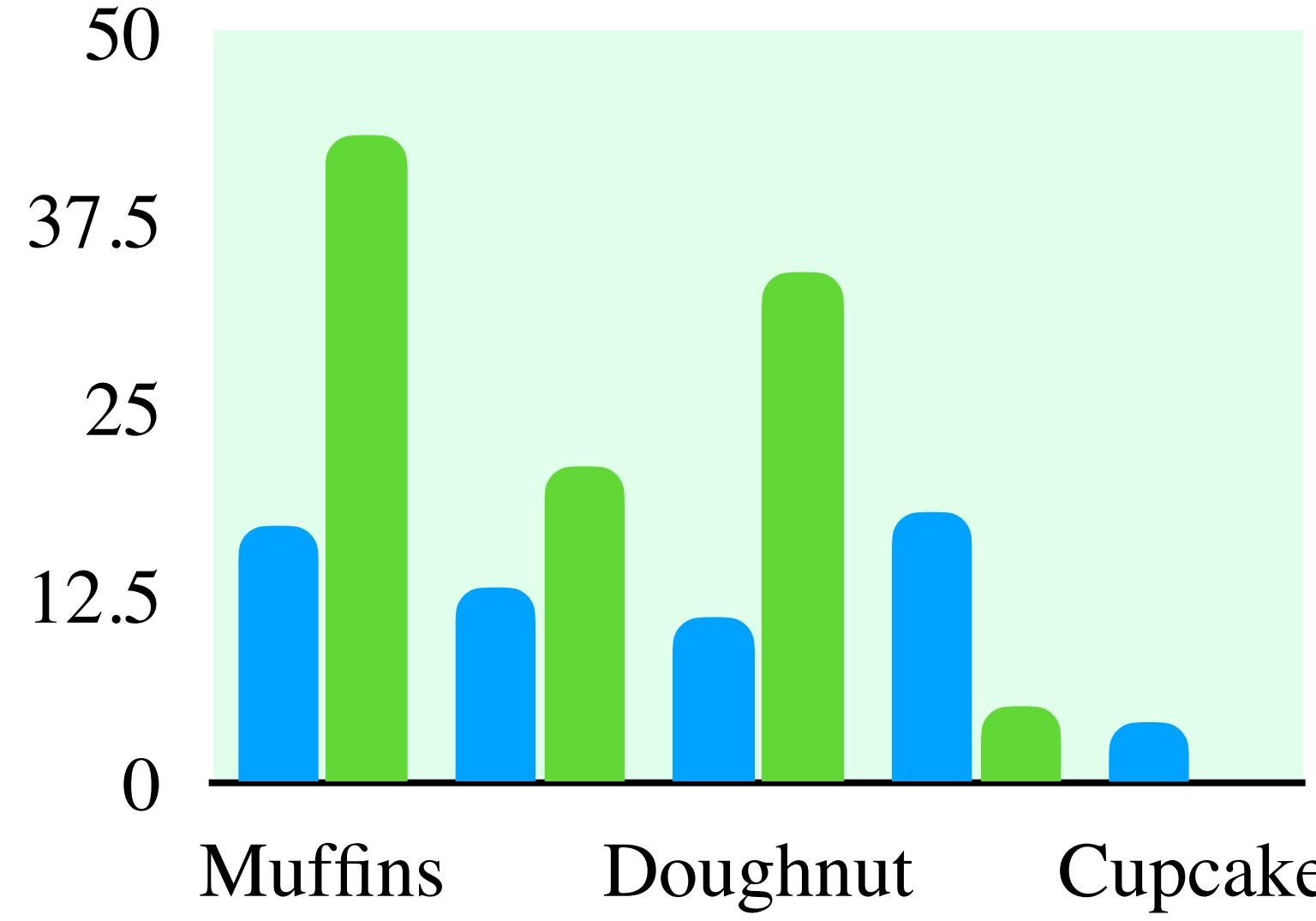


	Bakery A	Bakery B	
Banana Muffin	17	43	60
Sugar Cookie	13	21	34
Lard Doughnut	11	34	45
Cantaloupe Danish	18	5	23
Leaf Cupcake	4	0	4
	63	103	166



	Bakery A	Bakery B	
Banana Muffin	0.10	0.26	0.36
Sugar Cookie	0.08	0.13	0.20
Lard Doughnut	0.07	0.20	0.27
Cantaloupe Danish	0.11	0.03	0.14
Leaf Cupcake	0.02	0.00	0.02
	0.38	0.62	1.00

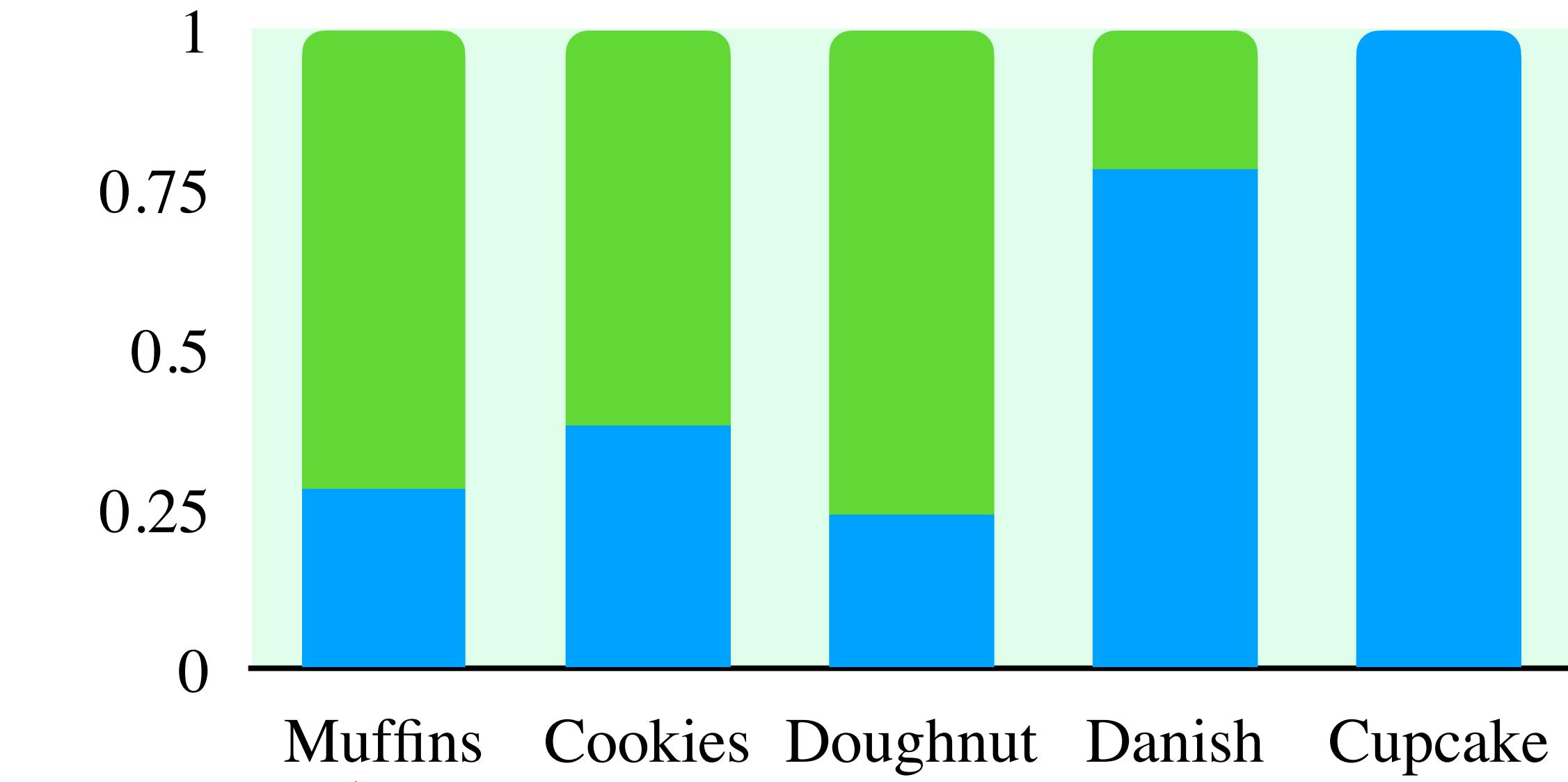
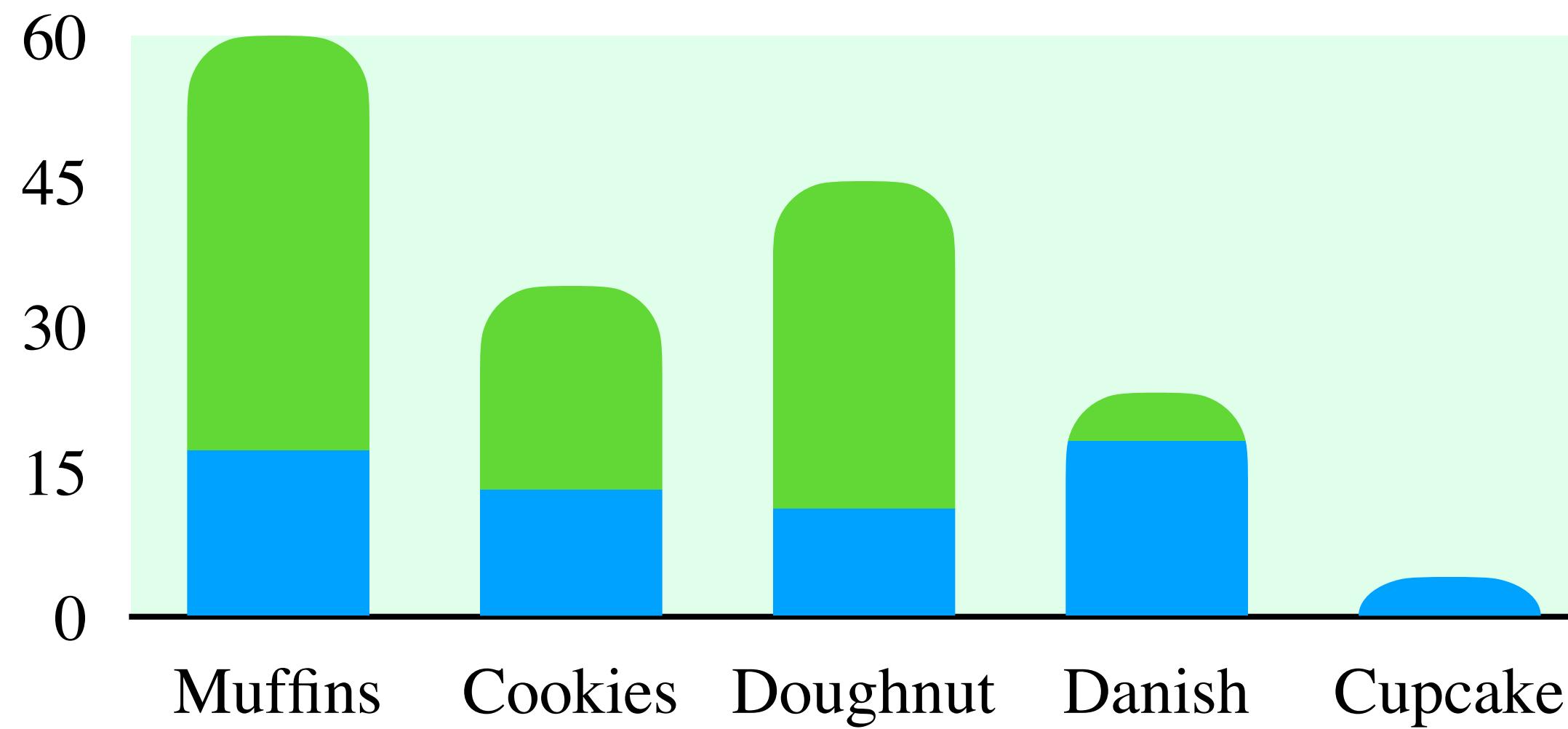
# QUALITATIVE DATA



A red arrow starts from the top of the dog's head and curves upwards towards the top right corner of the table.

	Bakery A	Bakery B	
Muffin	17	43	60
Cookie	13	21	34
Doughnut	11	34	45
Danish	18	5	23
Cupcake	4	0	4
		63	103
		166	

# QUALITATIVE DATA



	Bakery A	Bakery B	
Muffin	17	43	60
Cookie	13	21	34
Doughnut	11	34	45
Danish	18	5	23
Cupcake	4	0	4



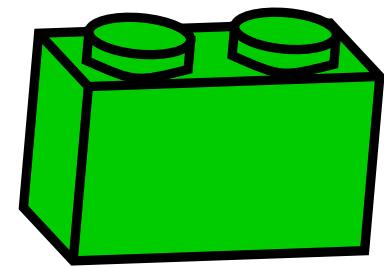
	Bakery A	Bakery B	
Muffin	0.28	0.72	1.00
Cookie	0.38	0.62	1.00
Doughnut	0.24	0.76	1.00
Danish	0.78	0.22	1.00
Cupcake	1.00	0.00	1.00

# **QUANTITATIVE DATA**

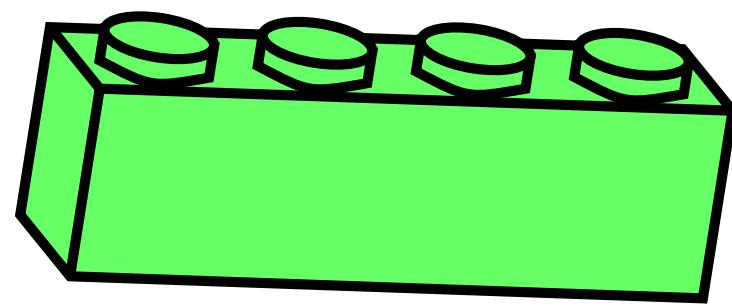
- 1. How can we summarize one variable quantitative data?**
- 2. How can we visualize one variable quantitative data?**

# QUANTITATIVE DATA

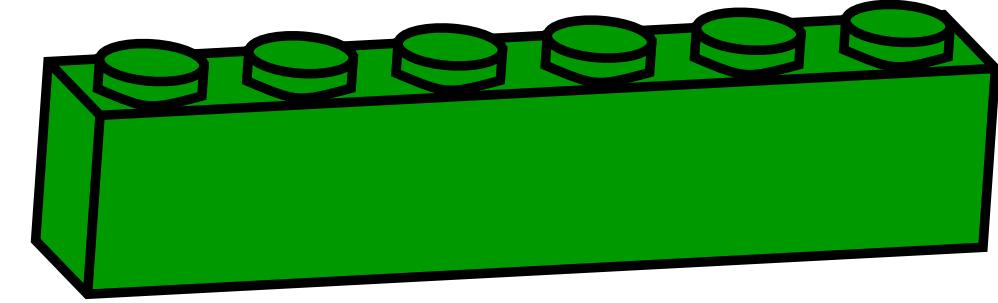
$$x_1 = 2$$



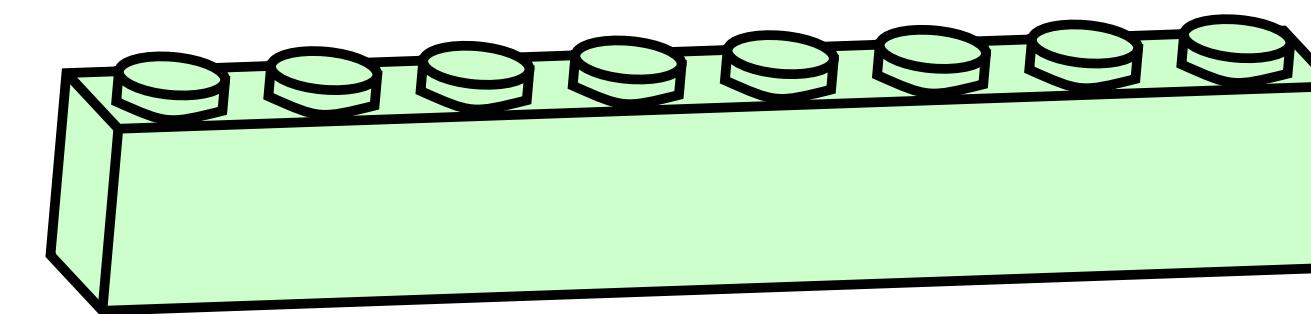
$$x_2 = 4$$



$$x_3 = 6$$

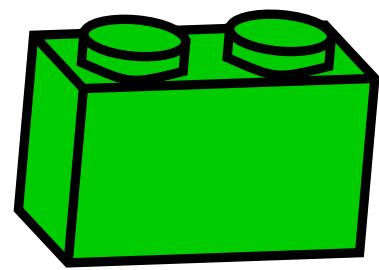


$$x_4 = 8$$

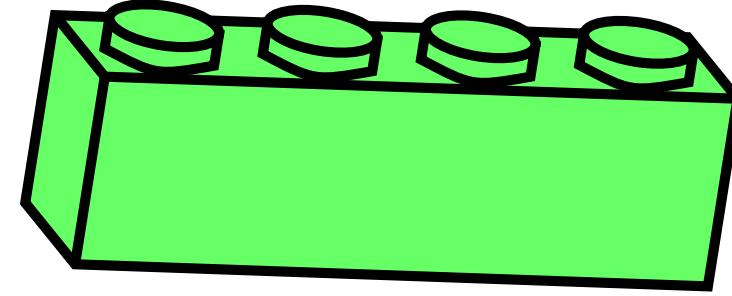


# QUANTITATIVE DATA

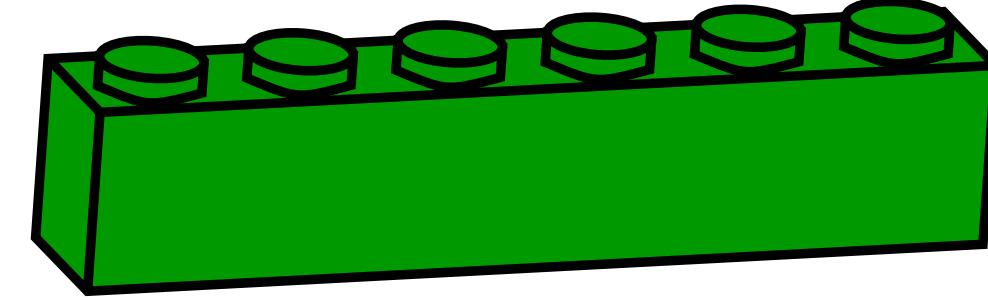
$$x_1 = 2$$



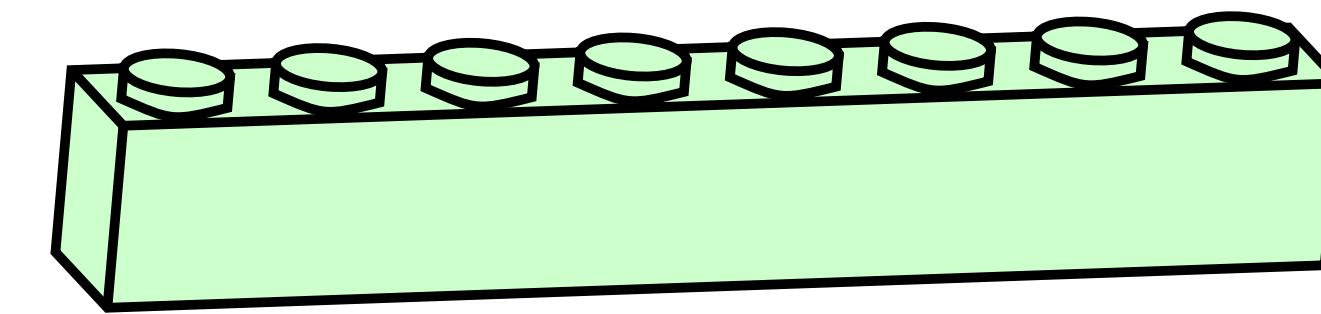
$$x_2 = 4$$



$$x_3 = 6$$



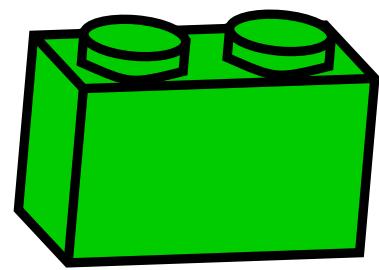
$$x_4 = 8$$



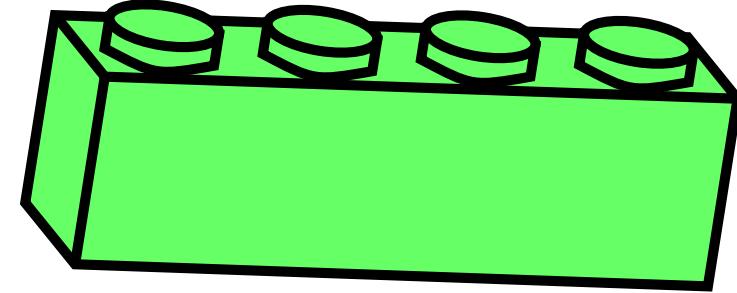
## MINIMUM / MAXIMUM

# QUANTITATIVE DATA

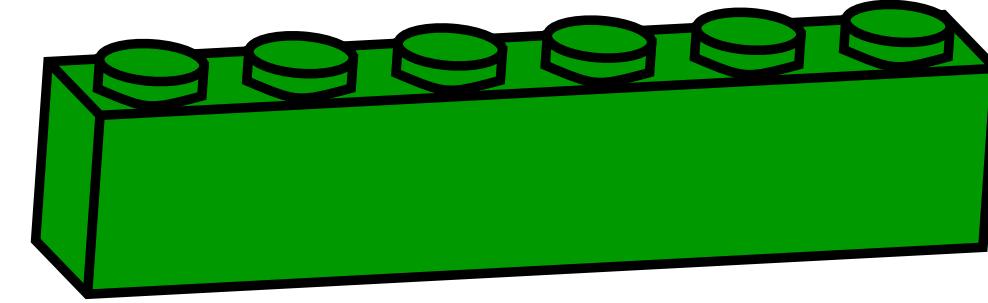
$$x_1 = 2$$



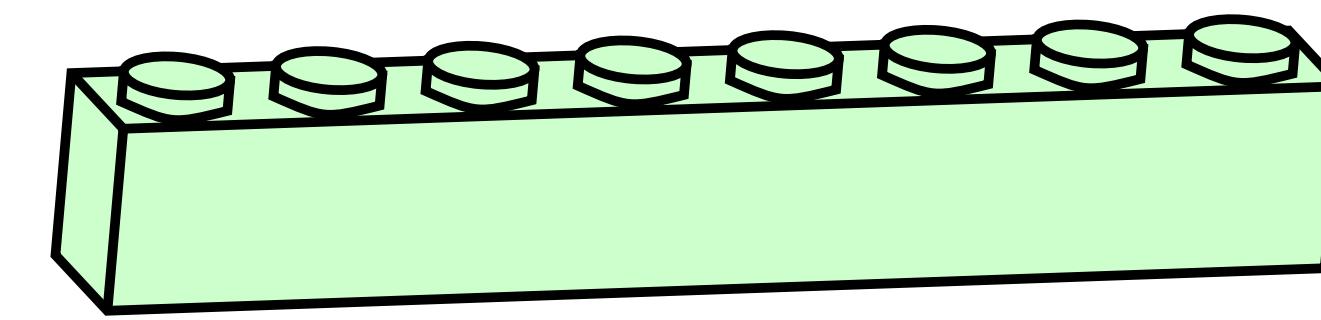
$$x_2 = 4$$



$$x_3 = 6$$



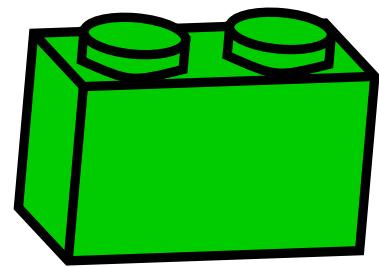
$$x_4 = 8$$



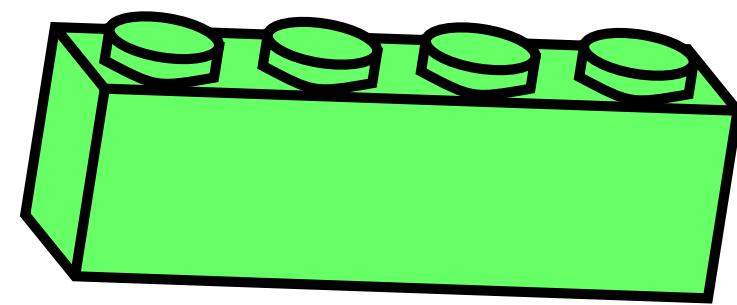
## ARITHMETIC MEAN

# QUANTITATIVE DATA

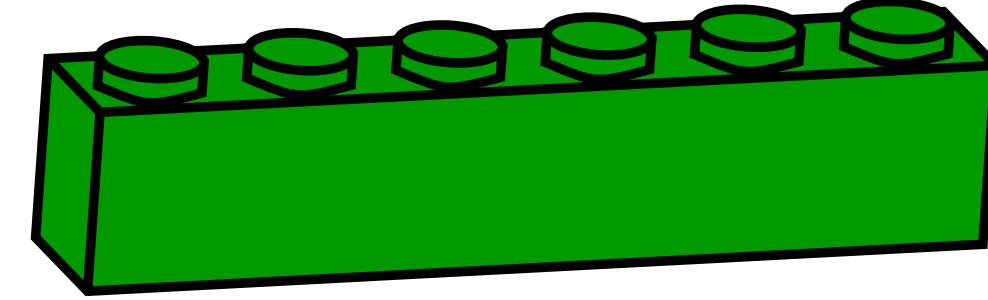
$$x_1 = 2$$



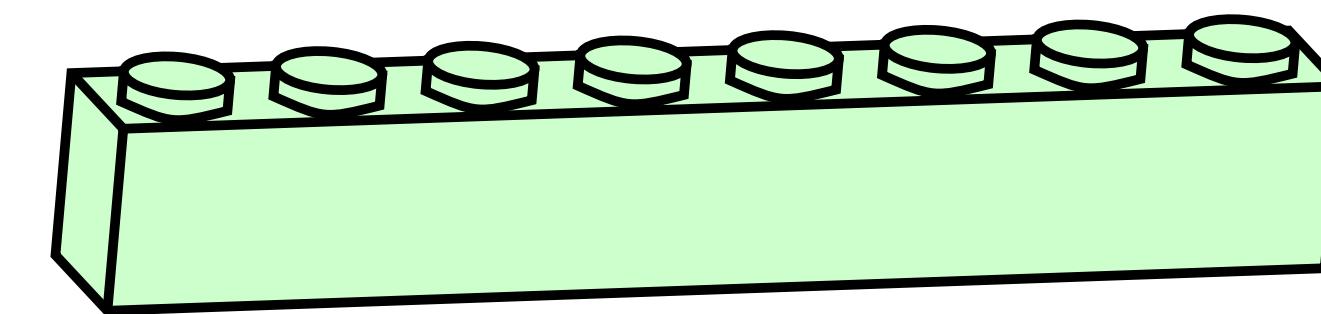
$$x_2 = 4$$



$$x_3 = 6$$



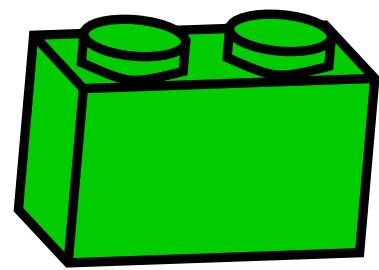
$$x_4 = 8$$



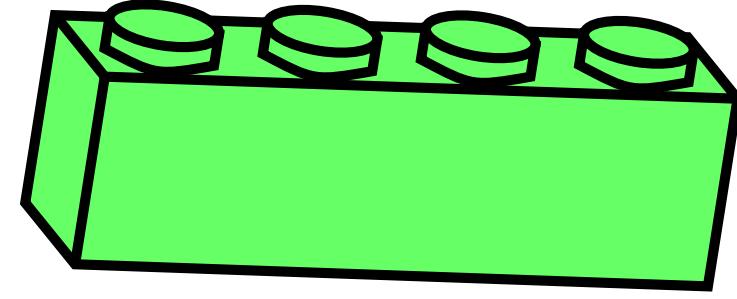
# GEOMETRIC MEAN

# QUANTITATIVE DATA

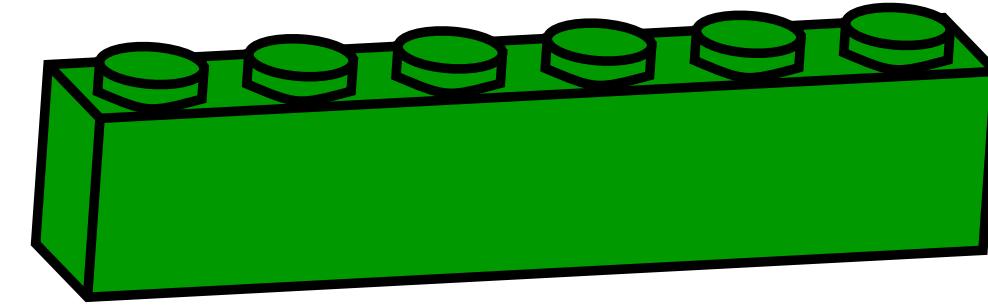
$$x_1 = 2$$



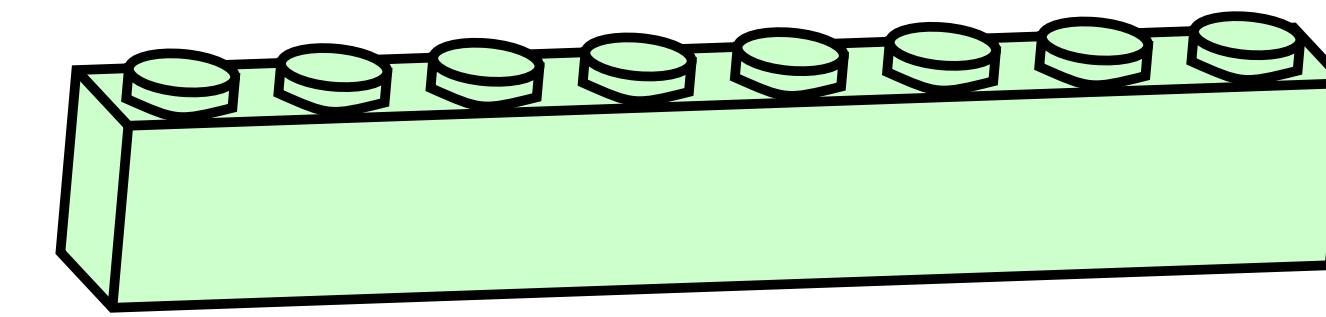
$$x_2 = 4$$



$$x_3 = 6$$



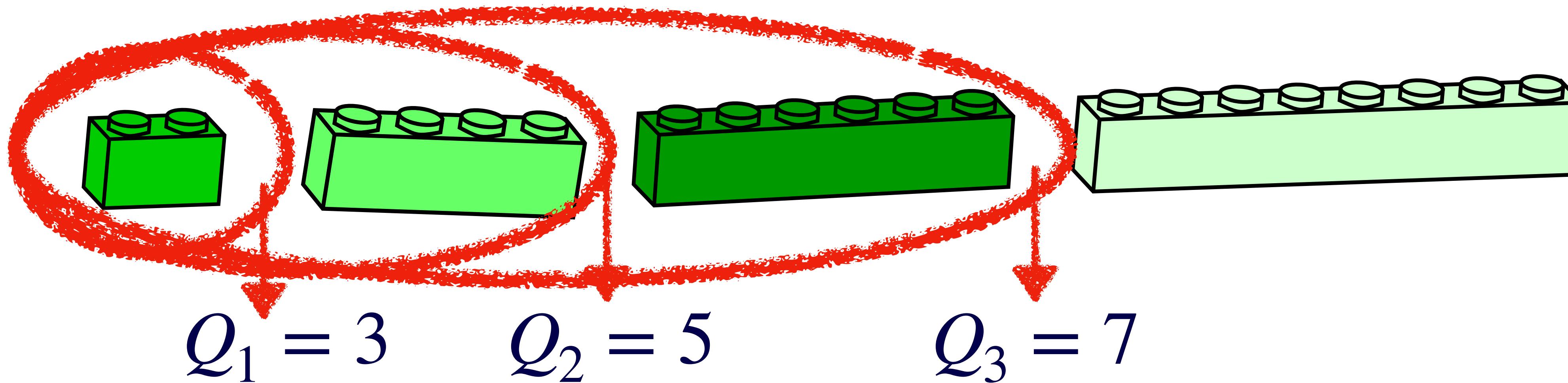
$$x_4 = 8$$



**Population Mean vs Sample mean: What's the Difference?**

# QUANTITATIVE DATA

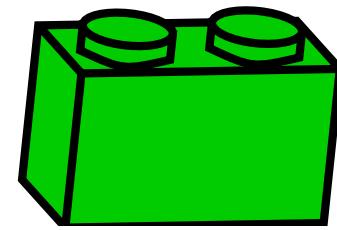
## Quartiles



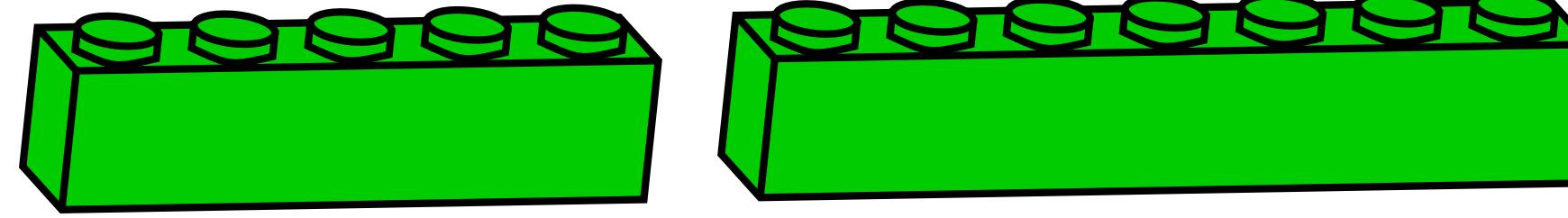
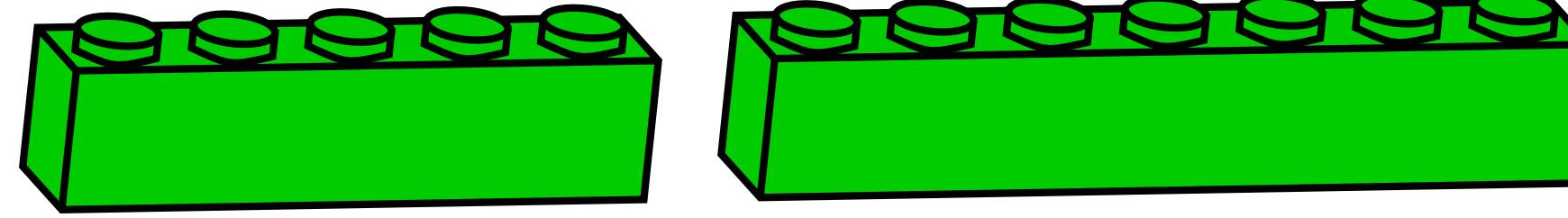
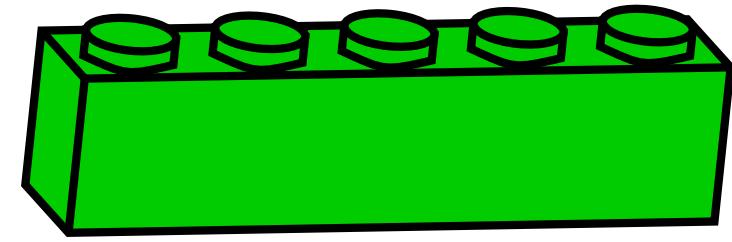
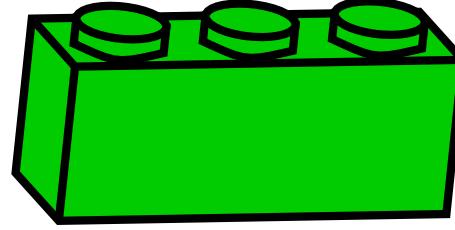
# QUANTITATIVE DATA

## The 5 Number Summary

*Min* = 2



*Median* = 5



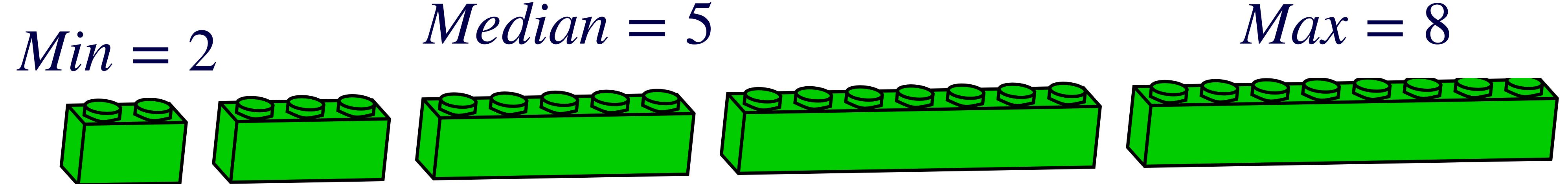
*Max* = 8

$Q_2$  = 3

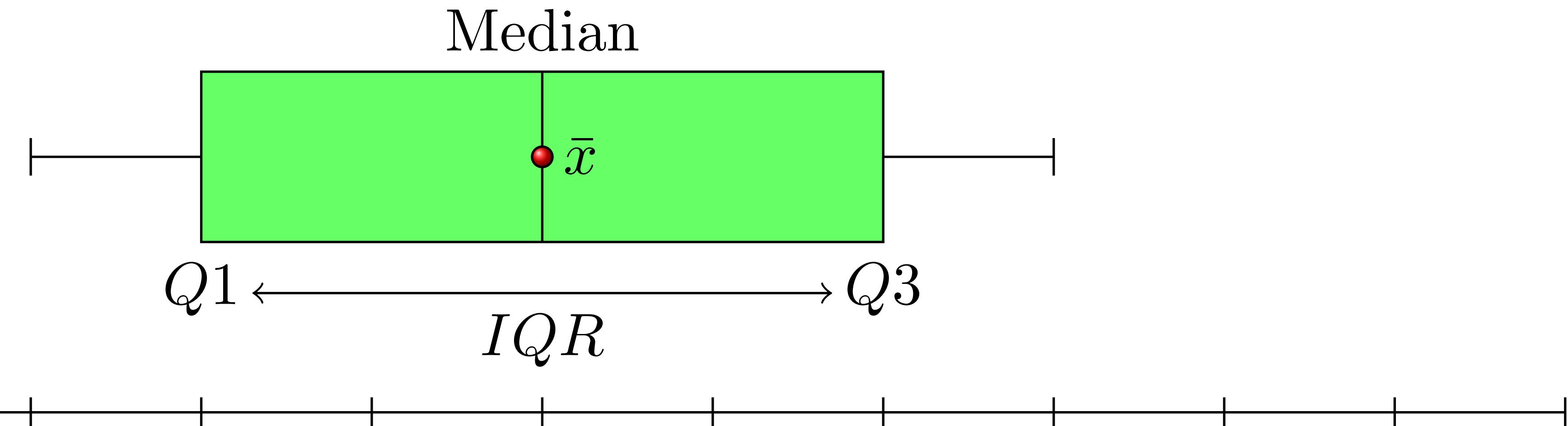
$Q_3$  = 7

# QUANTITATIVE DATA

## The 5 Number Summary



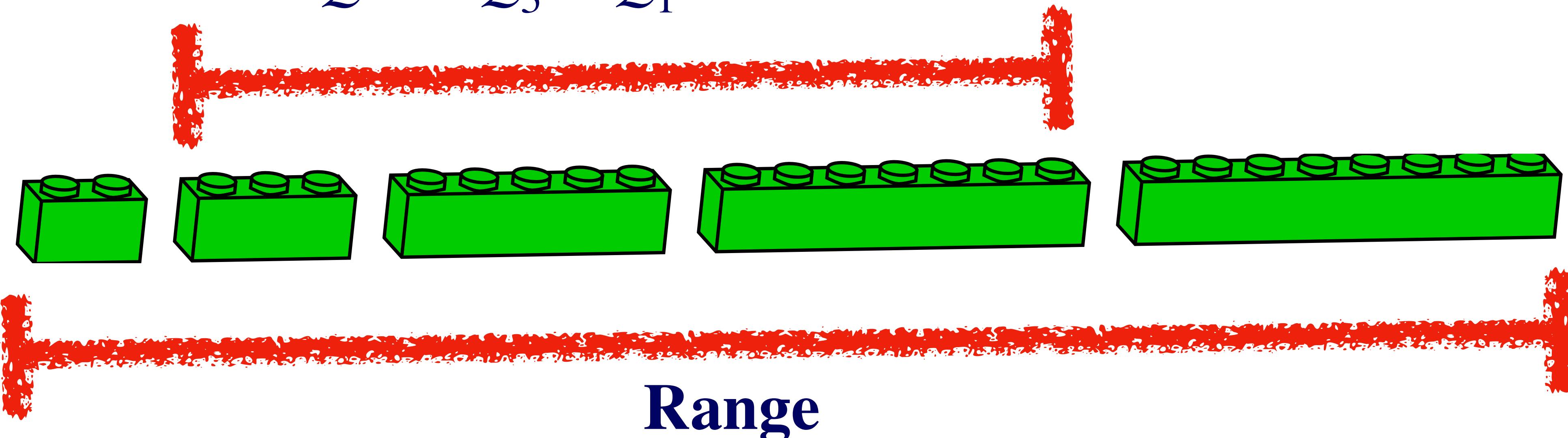
$Q_2 = 3$        $Q_3 = 7$



# QUANTITATIVE DATA

## Interquartile Range

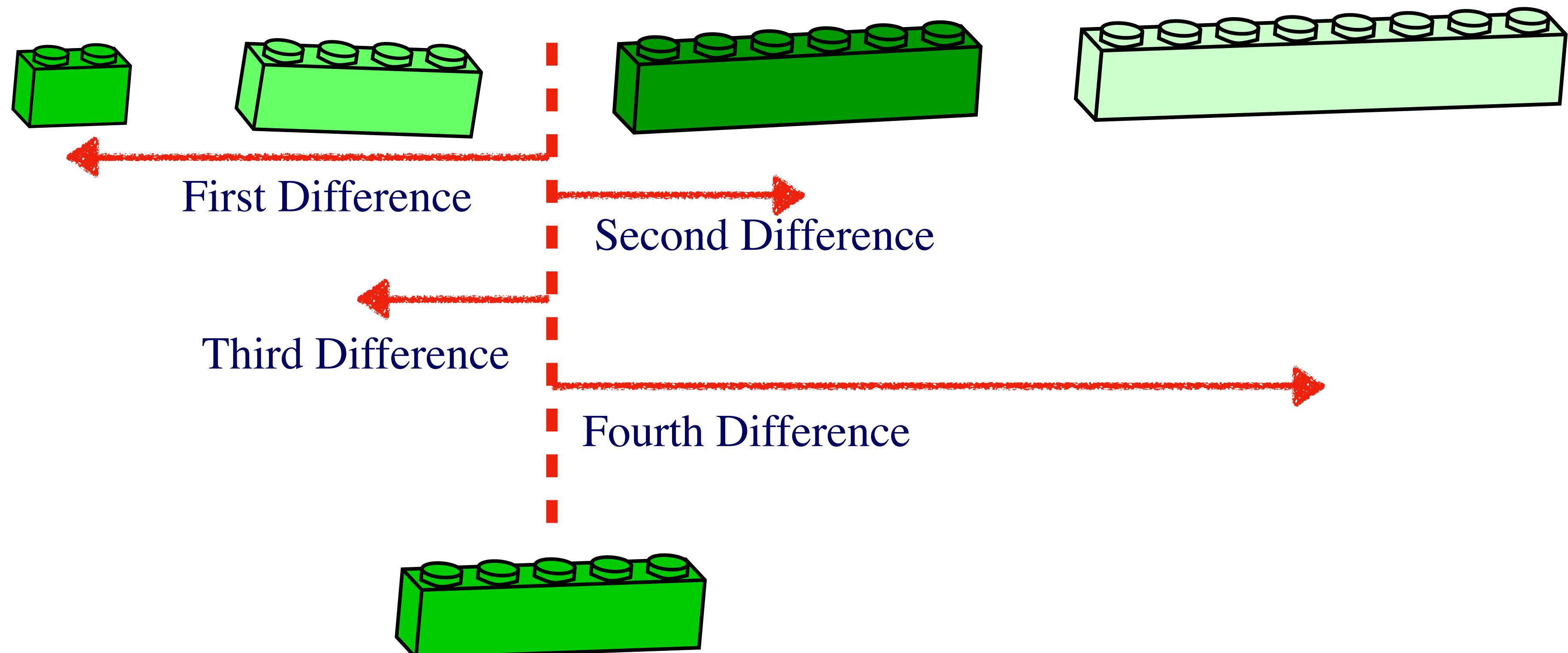
$$IQR = Q_3 - Q_1 = 7 - 3 = 4$$



$$Max - Min = 8 - 2 = 6$$

# QUANTITATIVE DATA

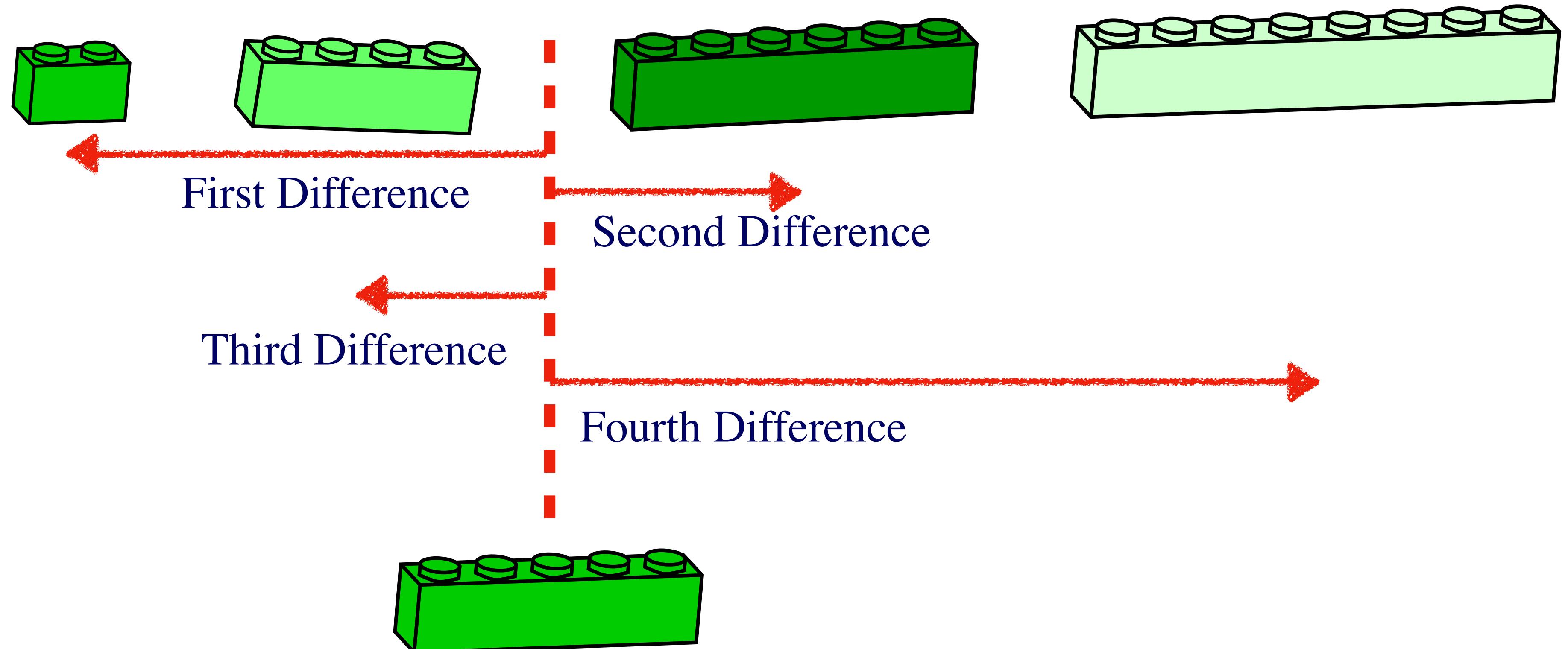
How do we measure spread about the mean?



$$\bar{x} = \frac{\sum_{i=1}^4 x_i}{4} = \frac{2 + 4 + 6 + 8}{4} = 5$$

# QUANTITATIVE DATA

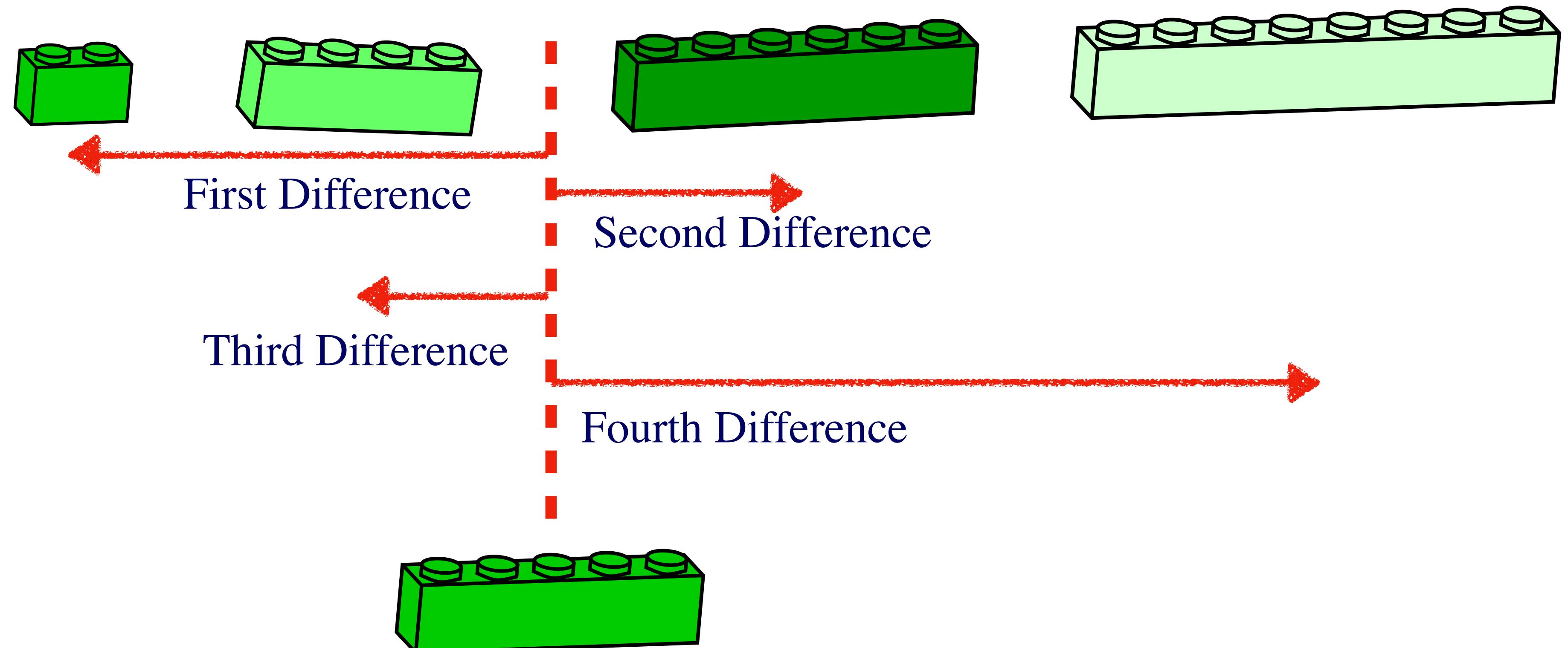
## Mean Absolute Deviation (MAD)



$$\bar{x} = \frac{\sum_{i=1}^4 x_i}{4} = \frac{2 + 4 + 6 + 8}{4} = 5$$

# QUANTITATIVE DATA

## Variance and Standard Deviation



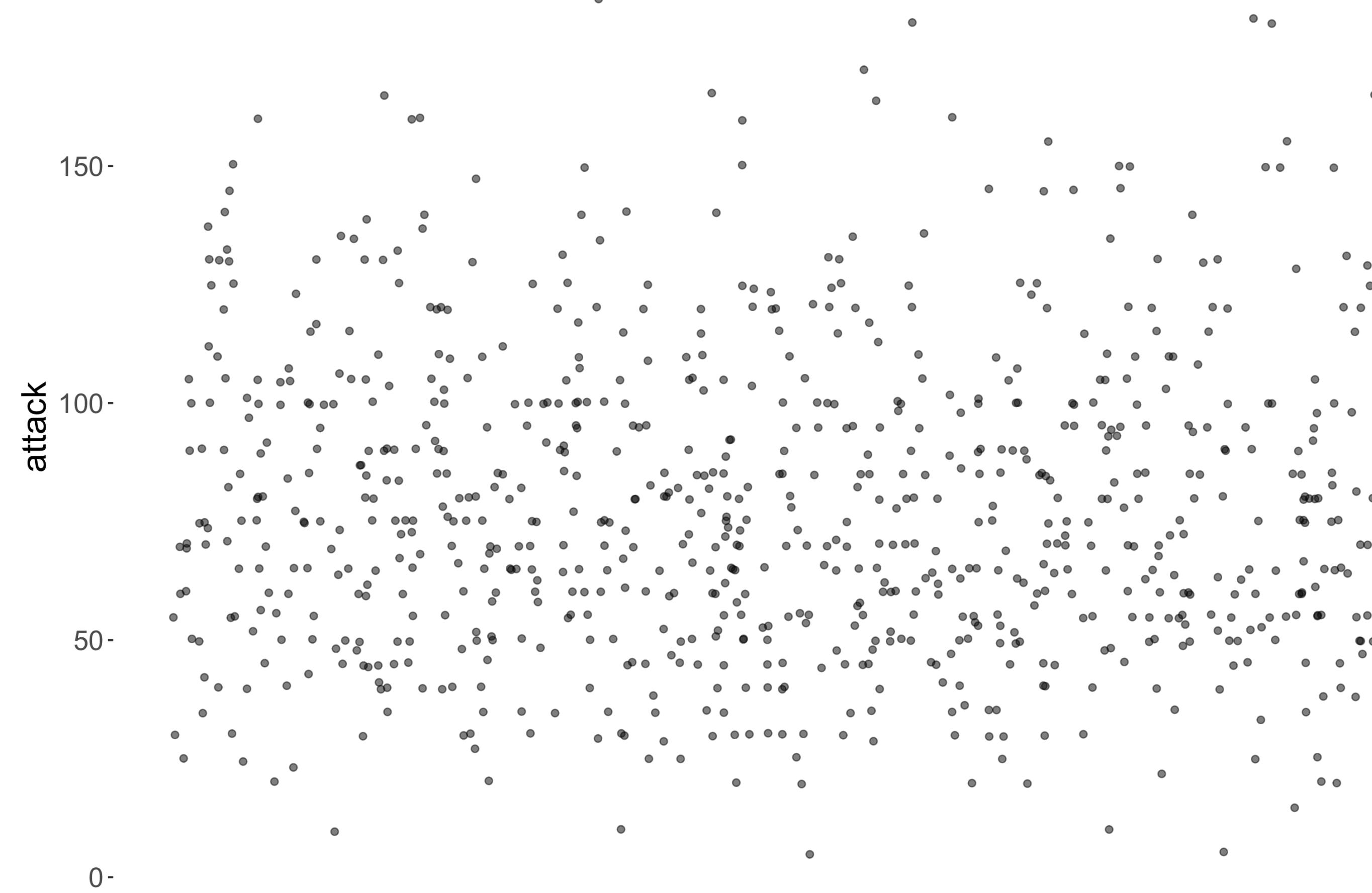
$$\bar{x} = \frac{\sum_{i=1}^4 x_i}{4} = \frac{2 + 4 + 6 + 8}{4} = 5$$

# **QUANTITATIVE DATA**

**Sample Variance vs. Population Variace**

**Whats the difference?**

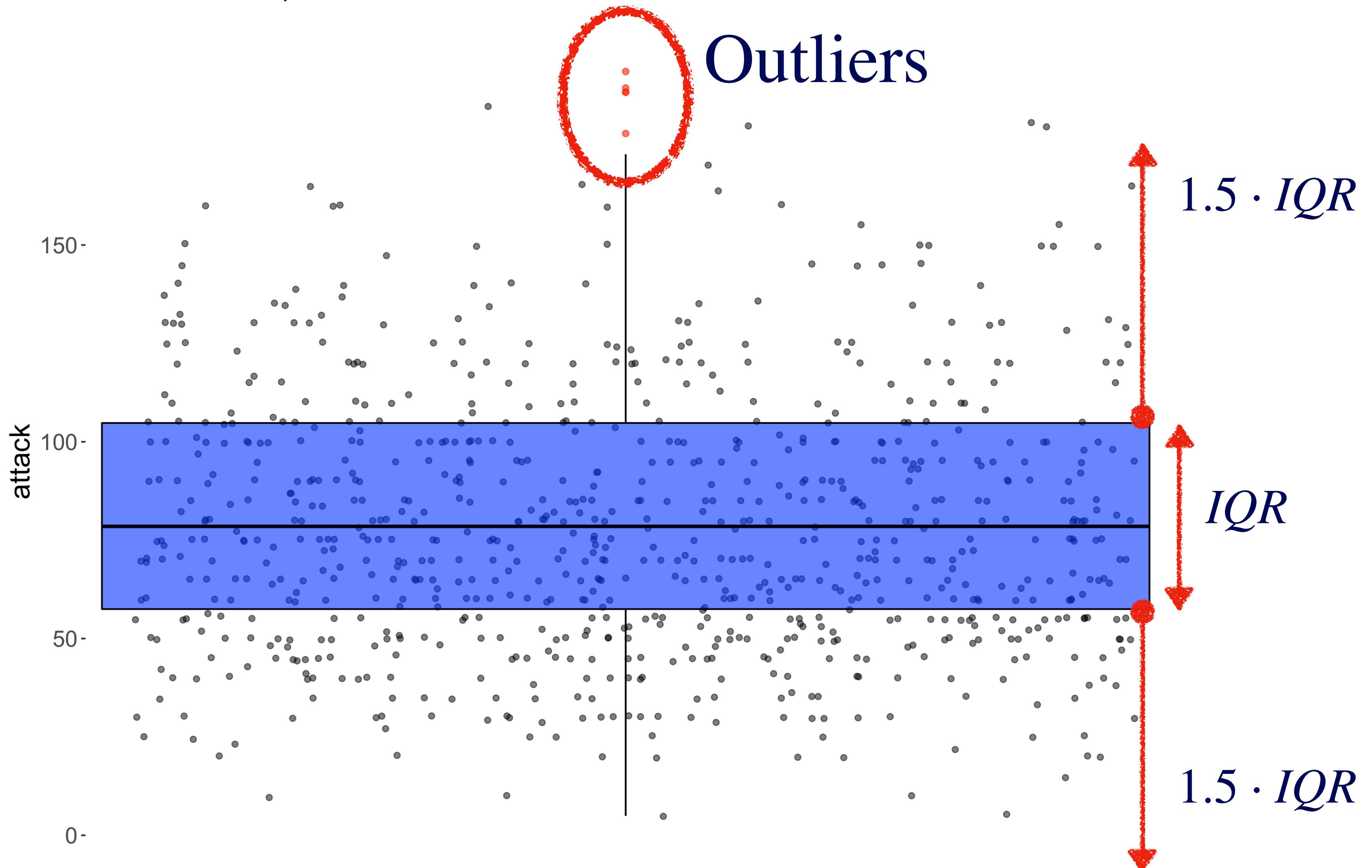
# VISUALIZING QUANTITATIVE DATA



# VISUALIZING QUANTITATIVE DATA

## Outlier Rules:

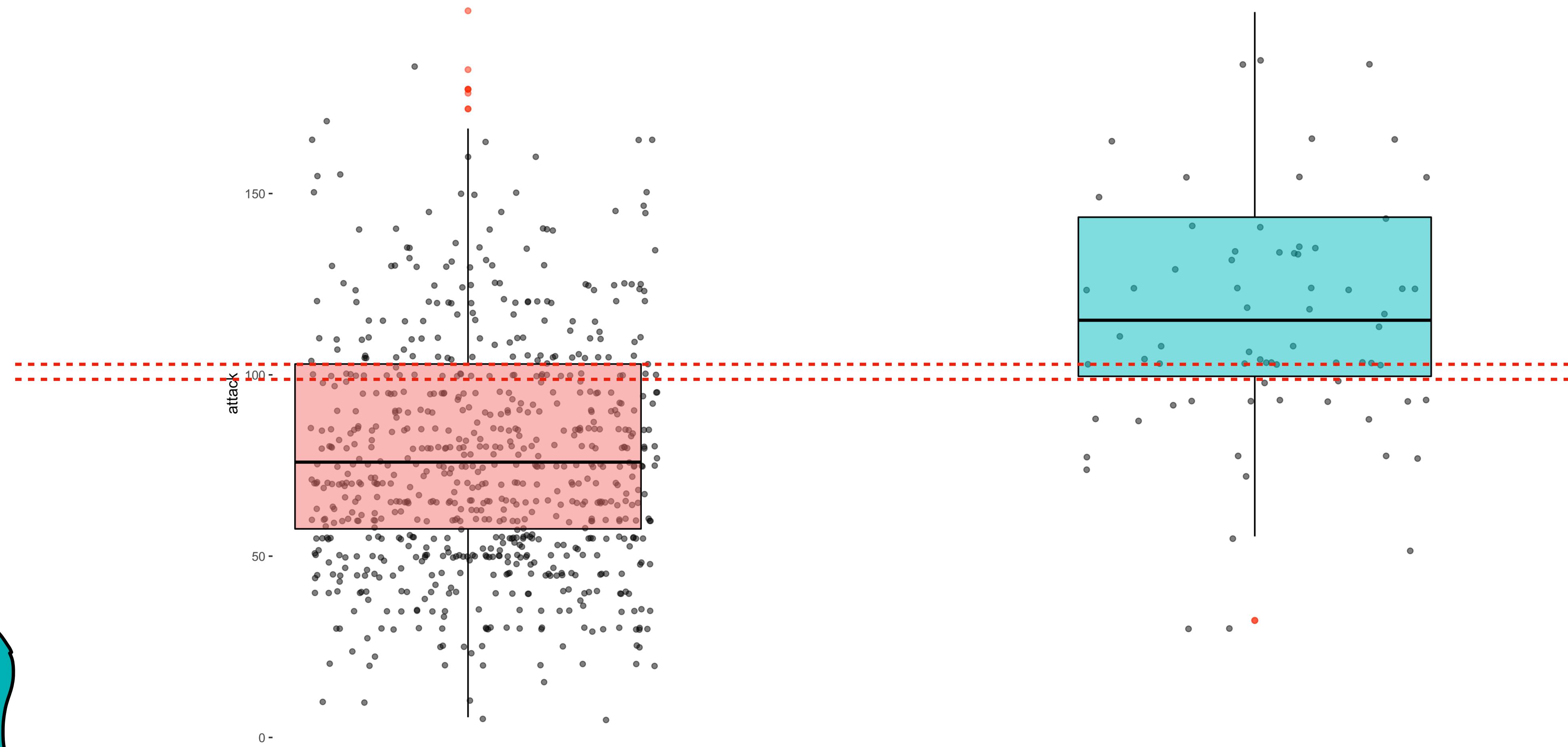
- **1.5(IQR) Rule**
- **2 sd rule**



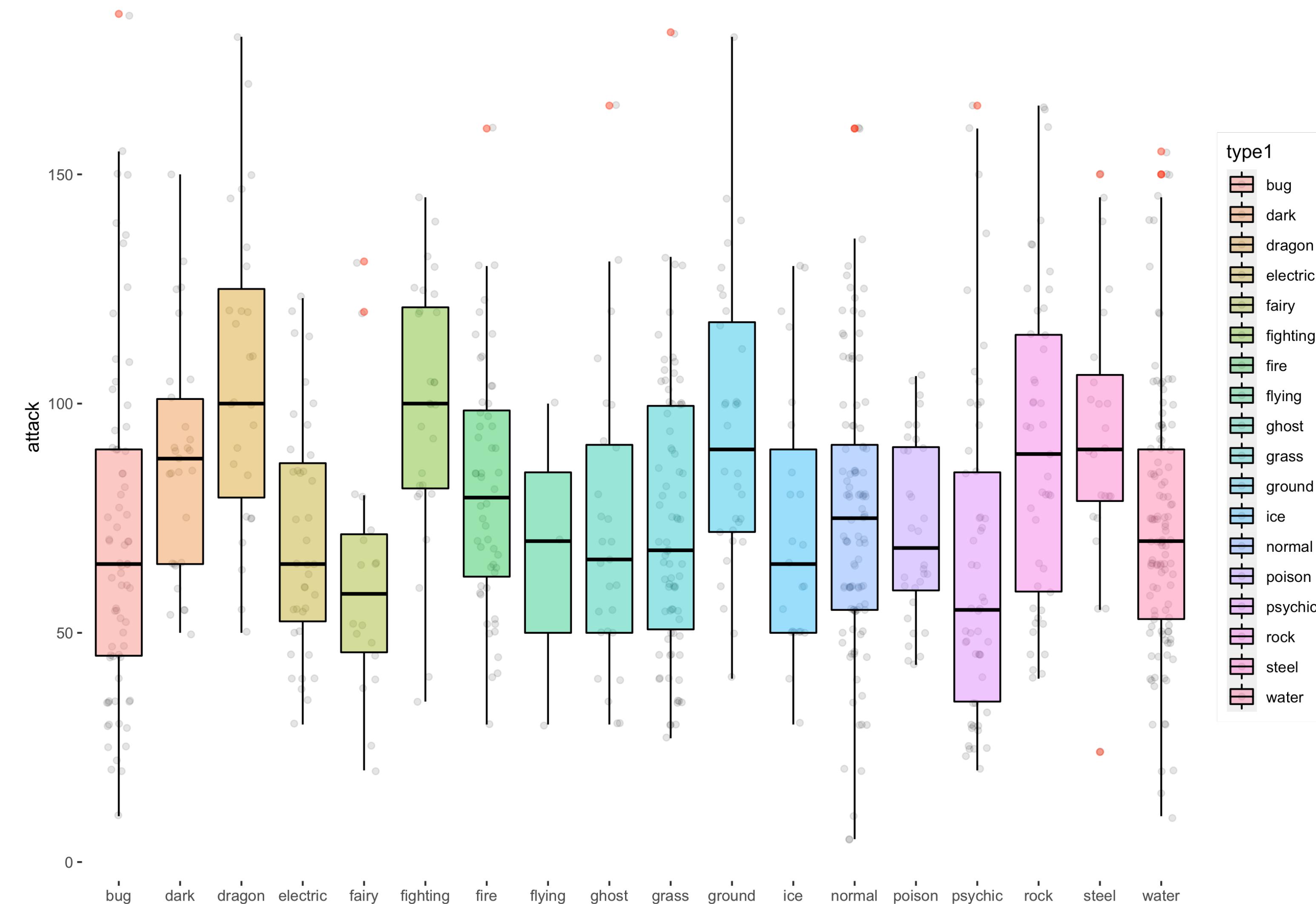
# VISUALIZING QUANTITATIVE DATA

Non-Legendary

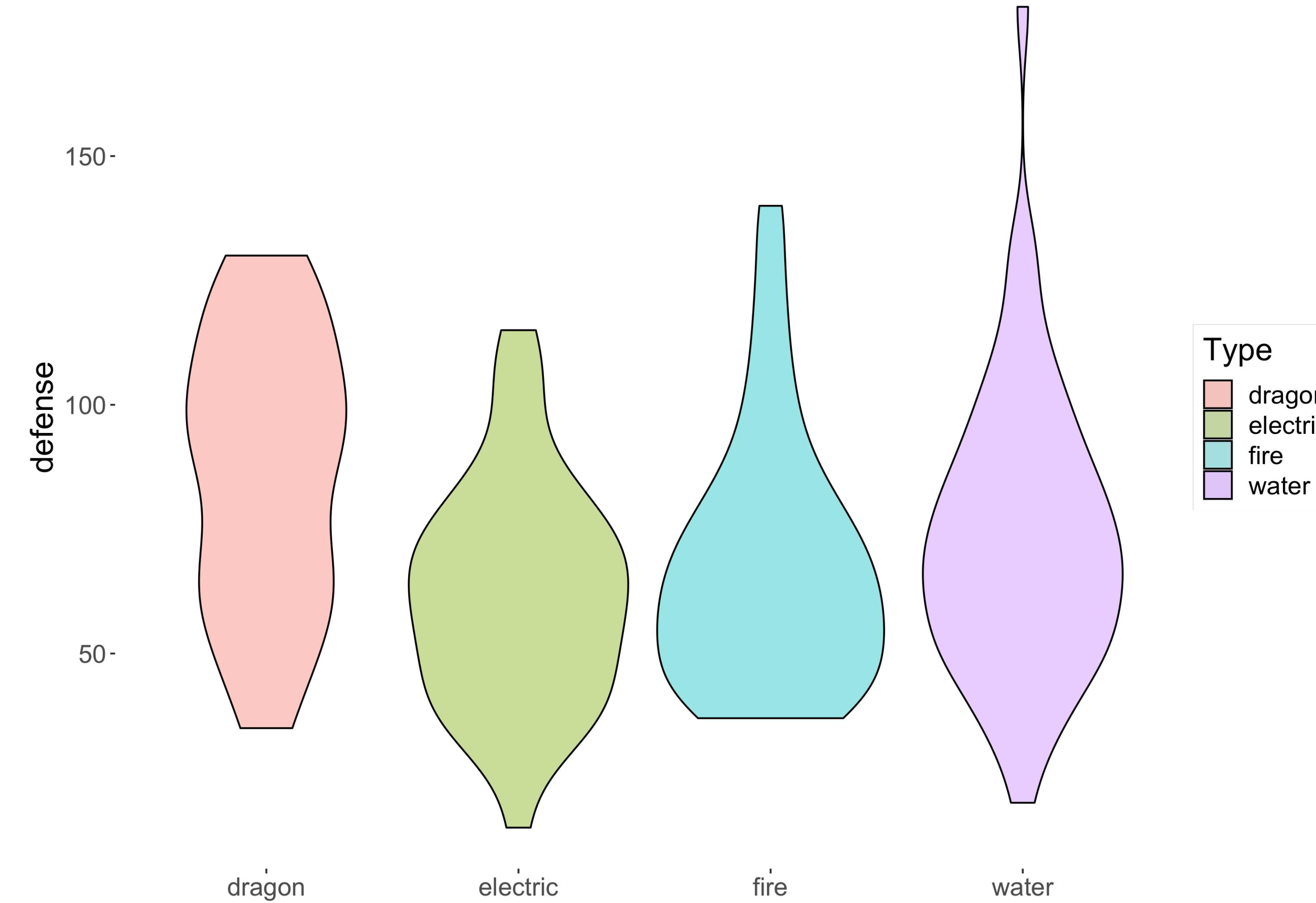
Legendary



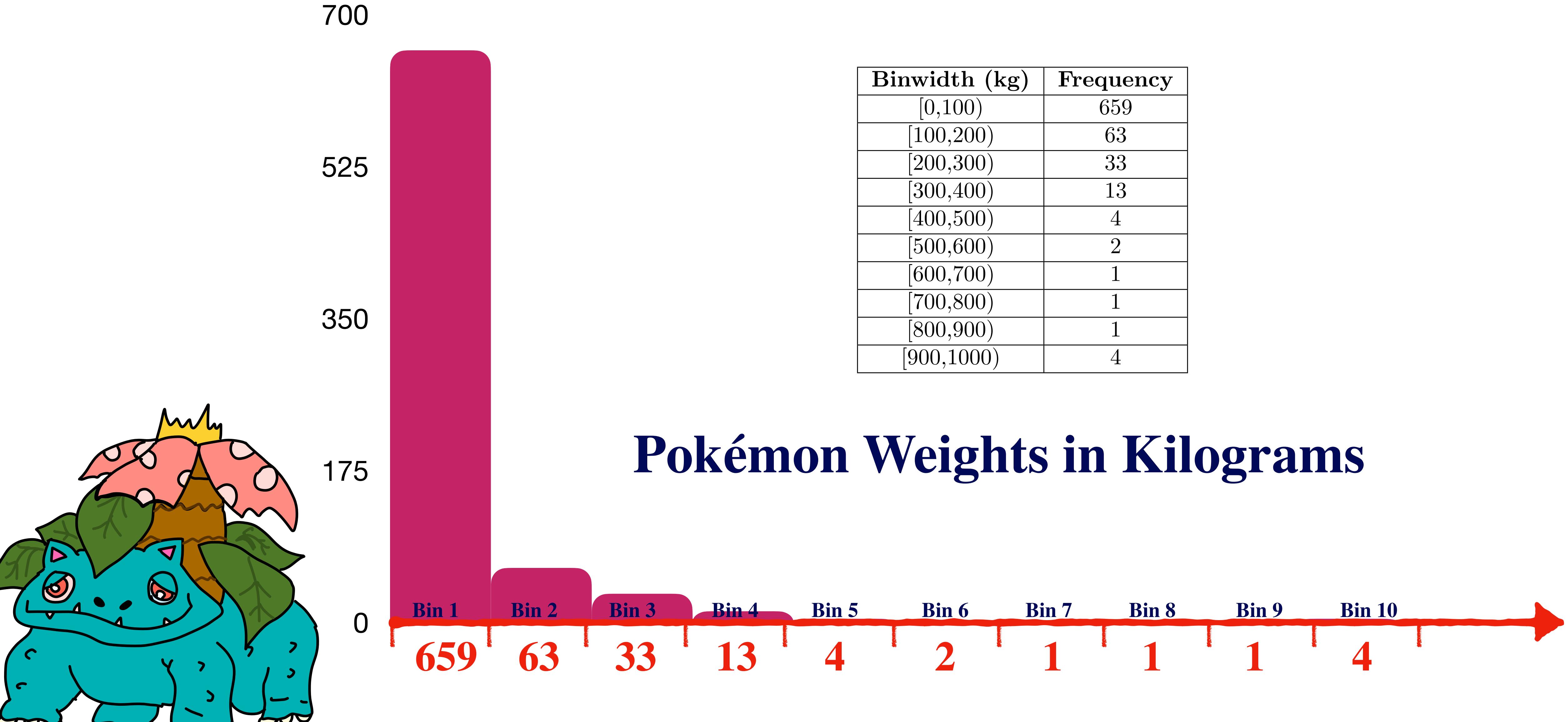
# VISUALIZING QUANTITATIVE DATA



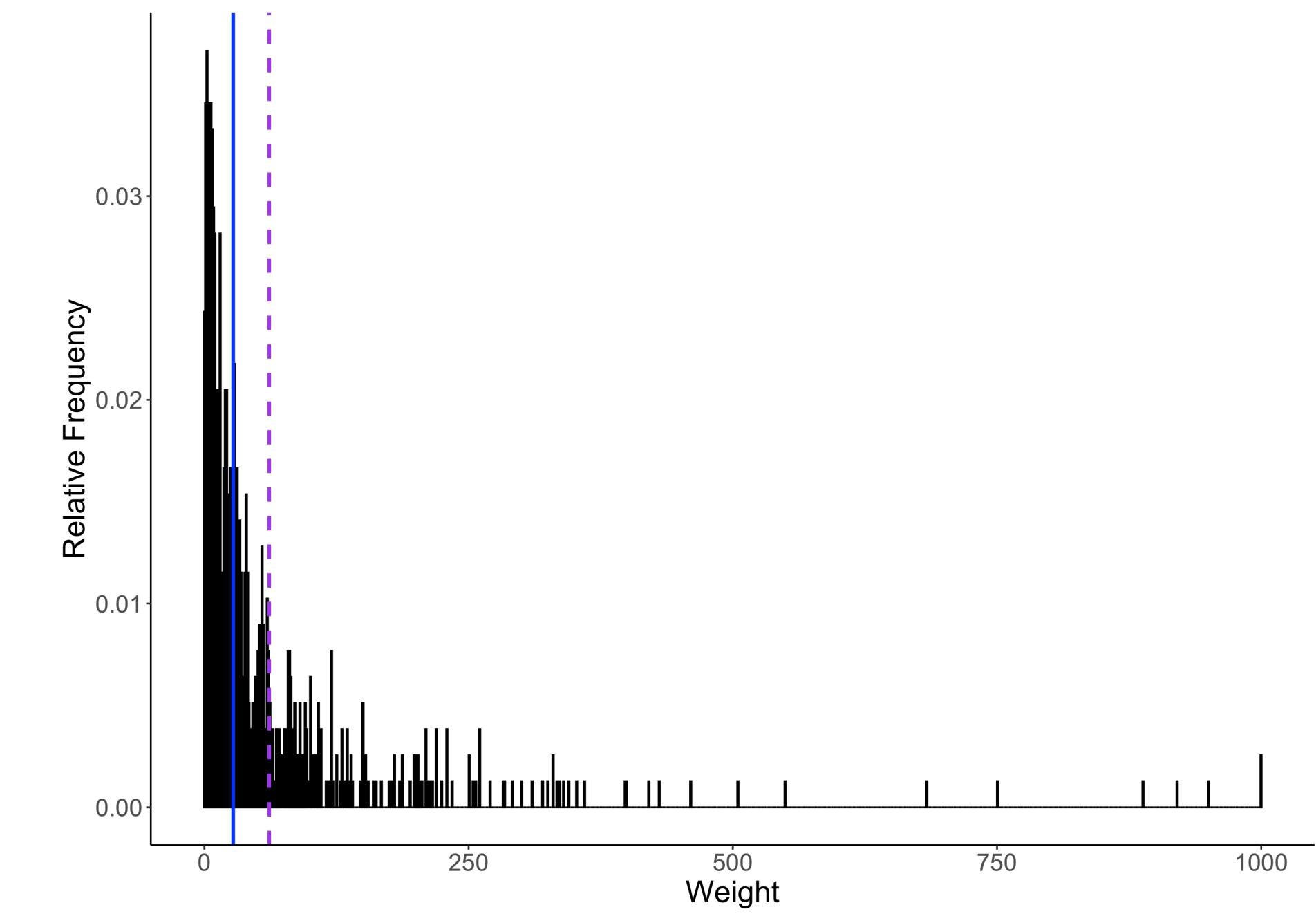
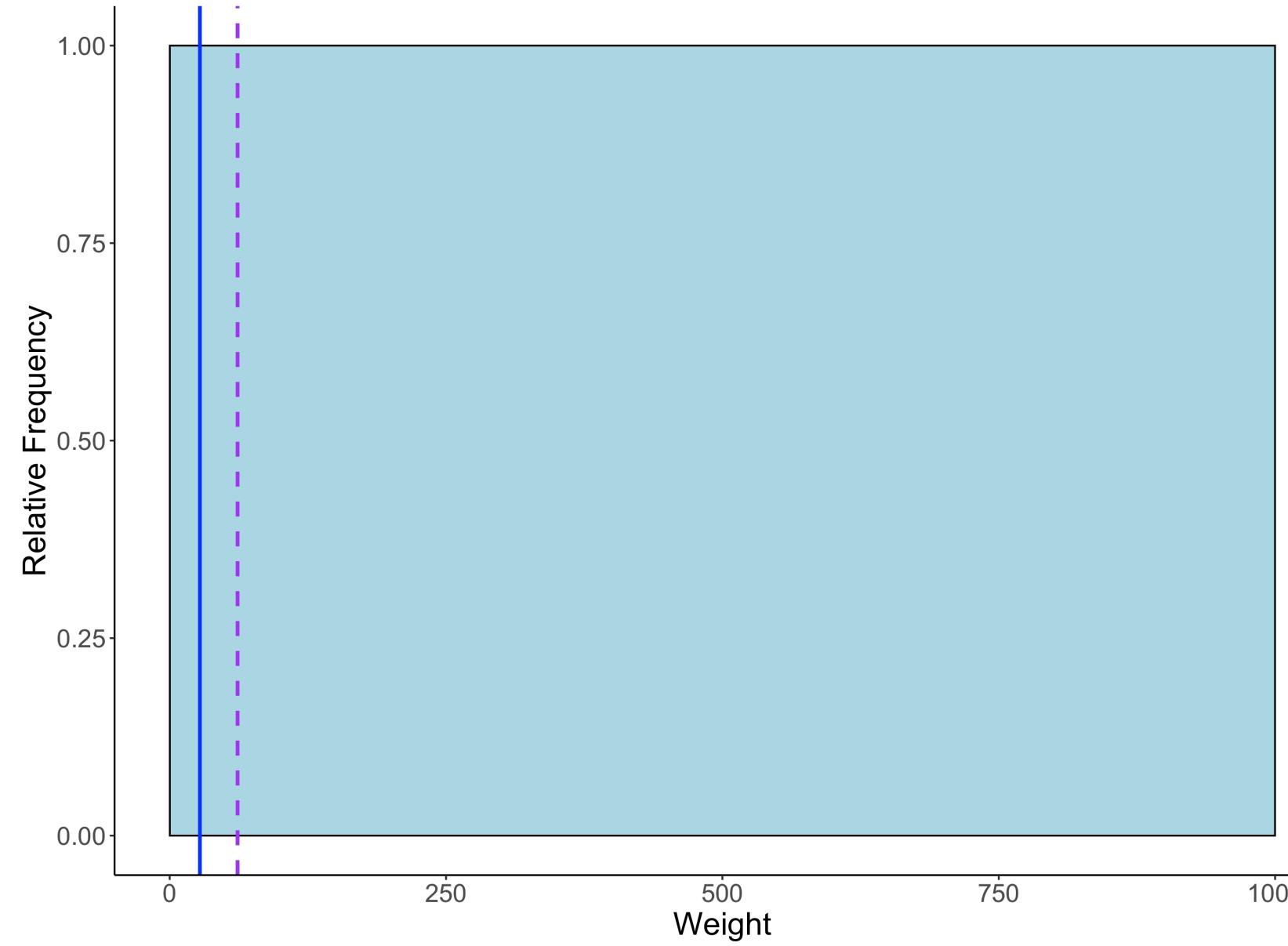
# VISUALIZING QUANTITATIVE DATA



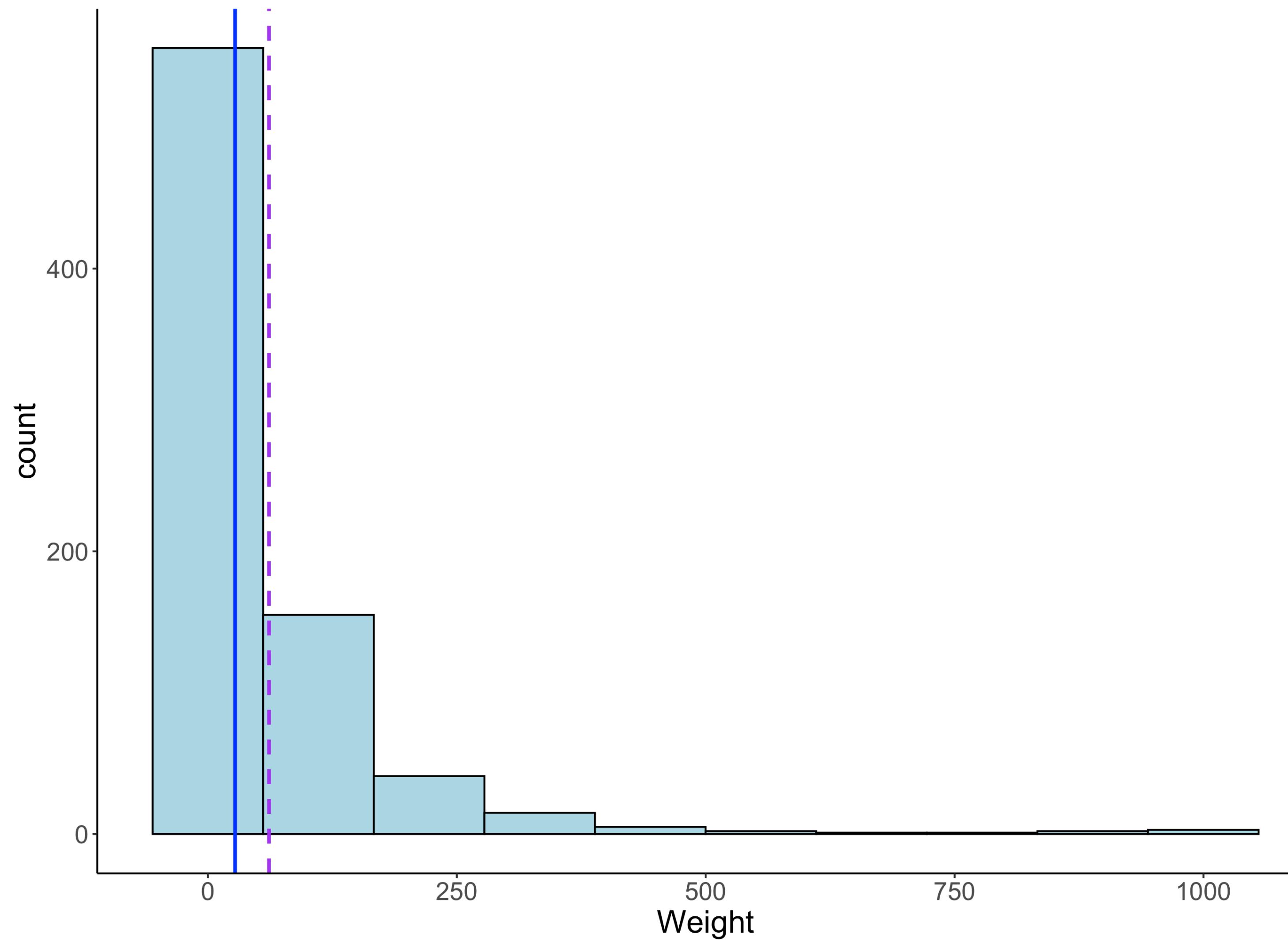
# VISUALIZING QUANTITATIVE DATA



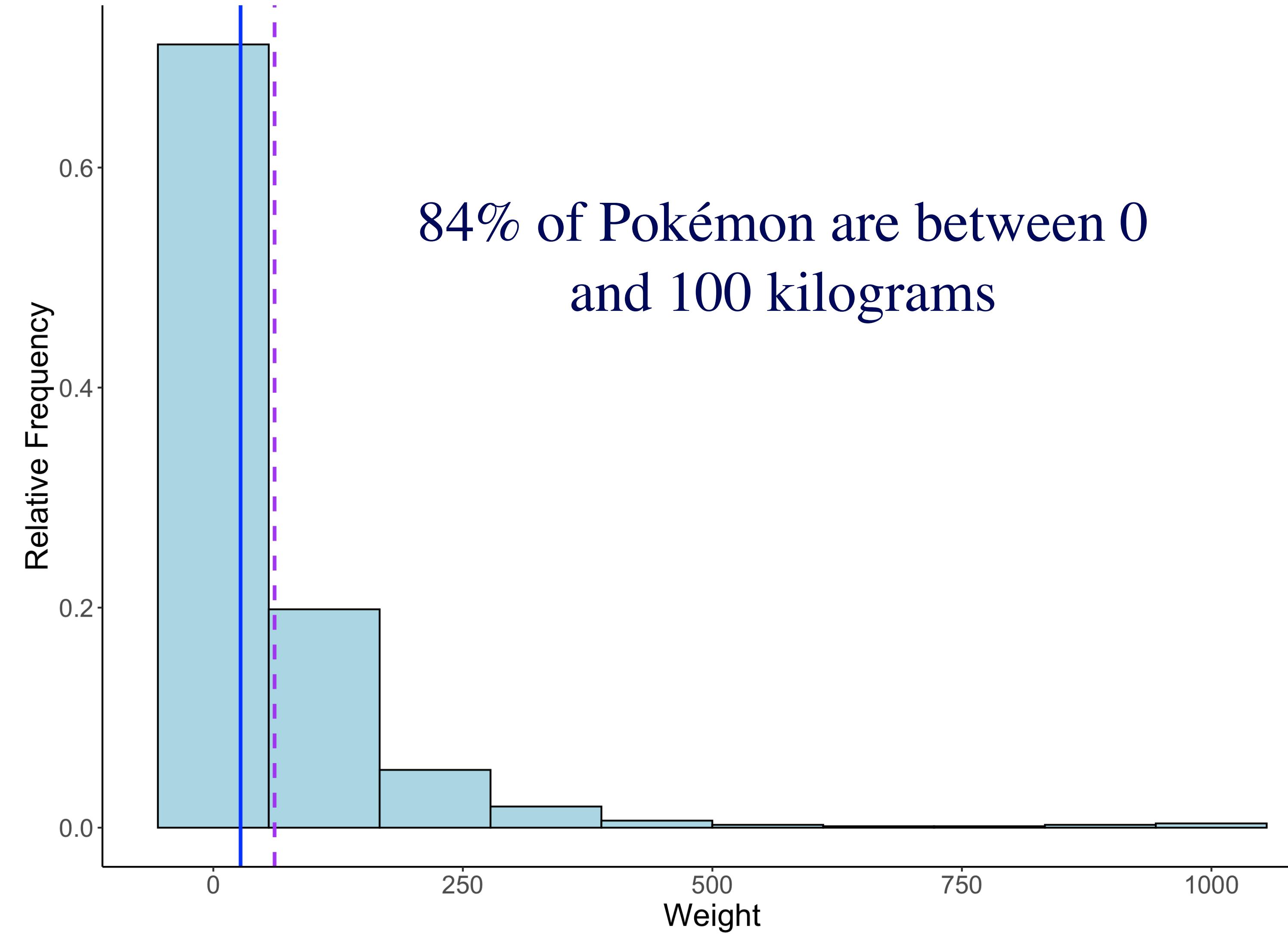
# VISUALIZING QUANTITATIVE DATA



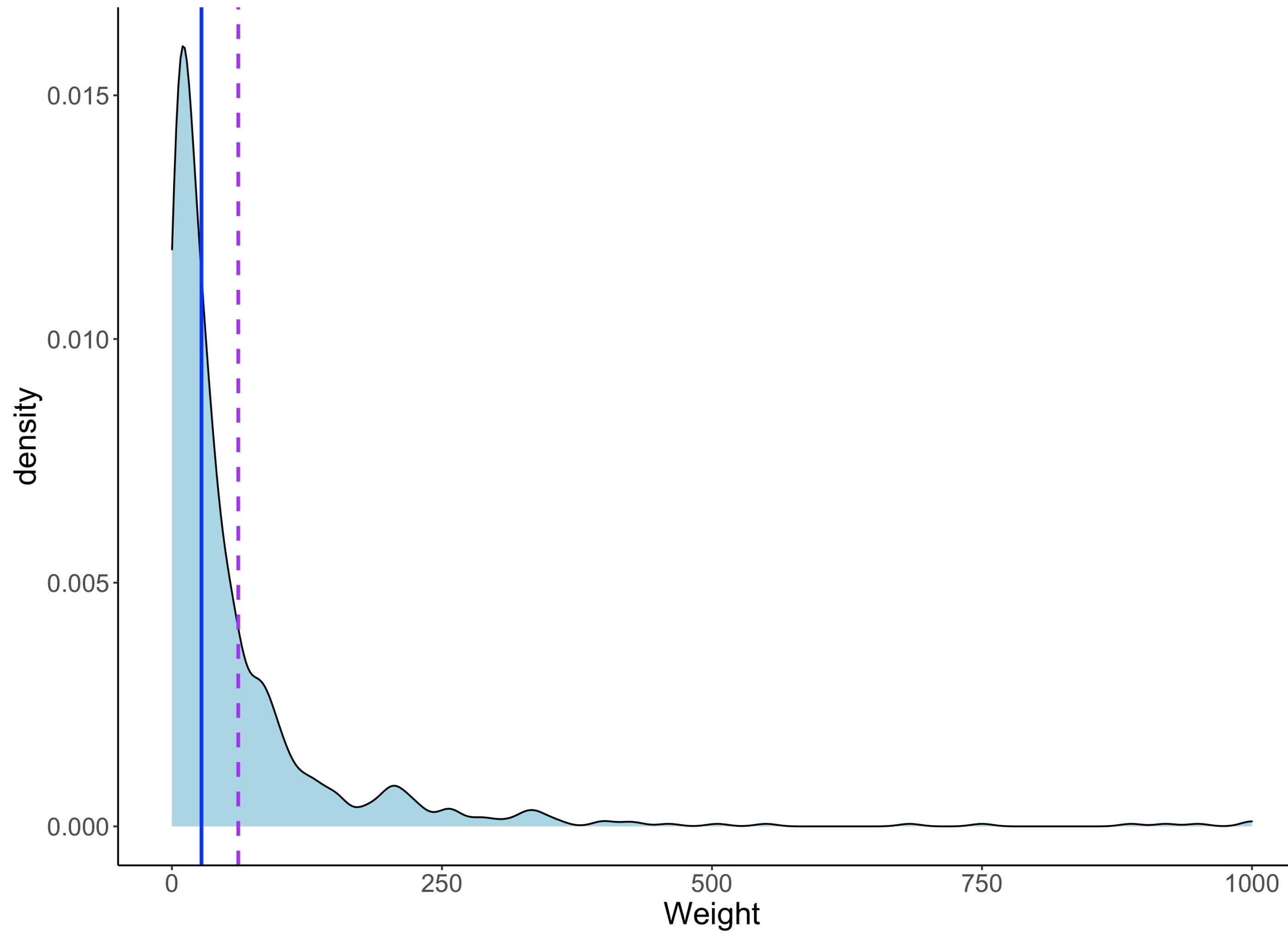
# VISUALIZING QUANTITATIVE DATA



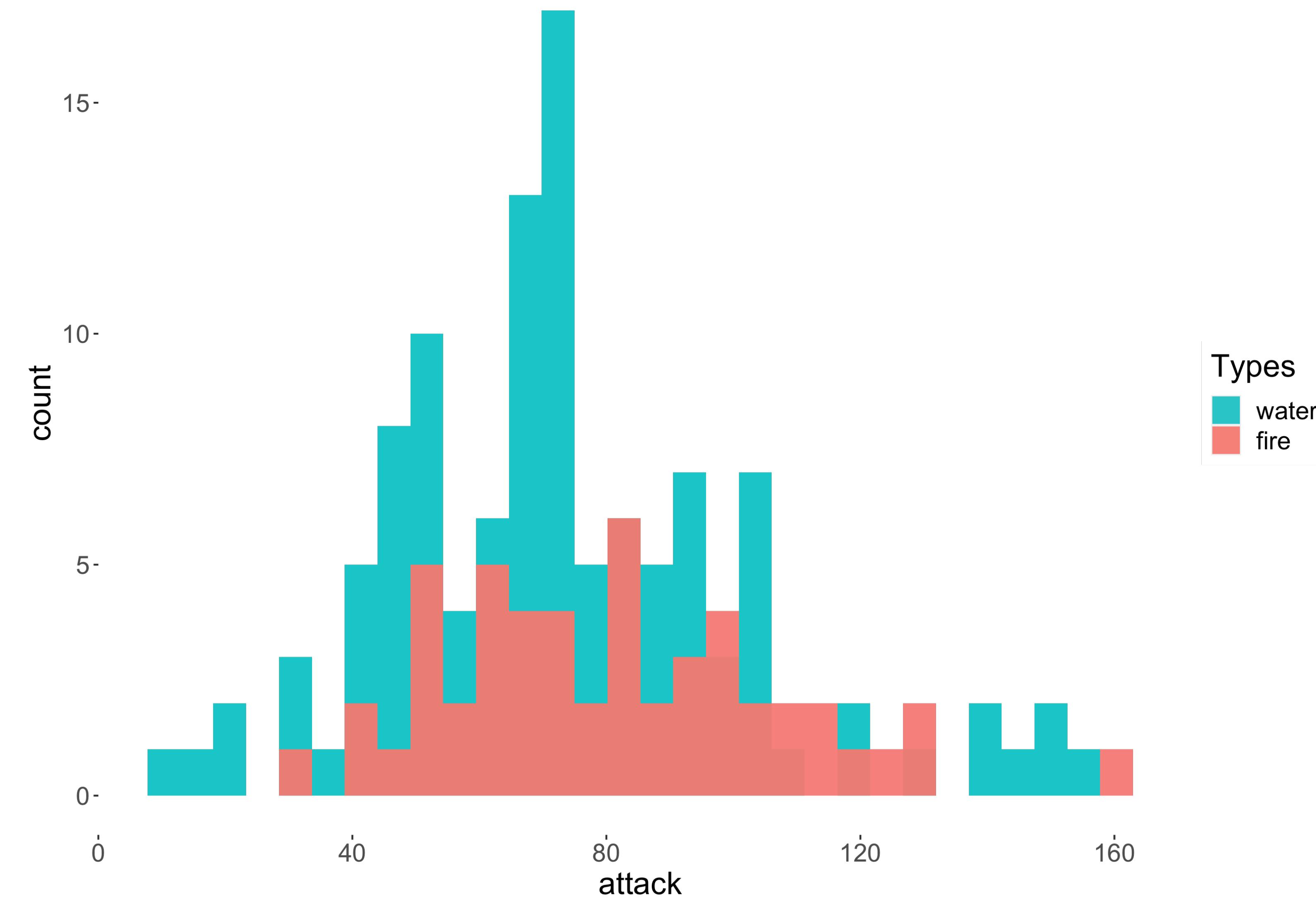
# VISUALIZING QUANTITATIVE DATA



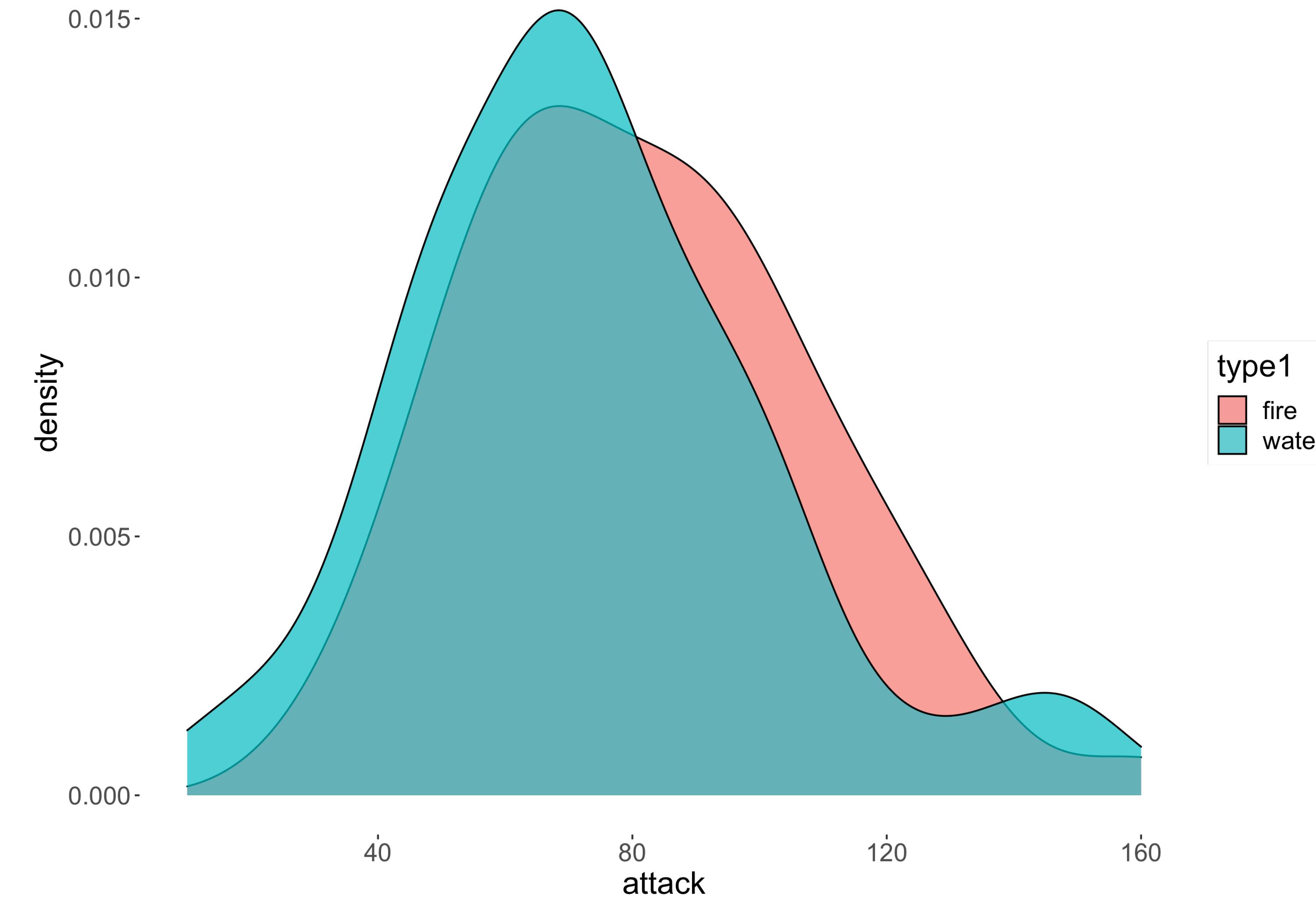
# VISUALIZING QUANTITATIVE DATA



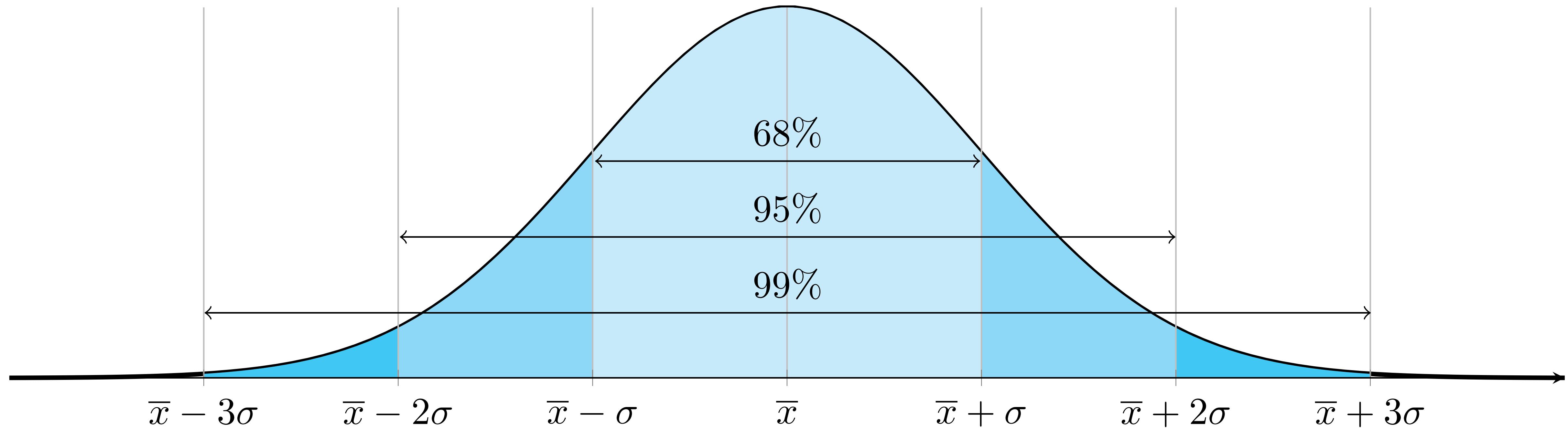
# VISUALIZING QUANTITATIVE DATA



# VISUALIZING QUANTITATIVE DATA

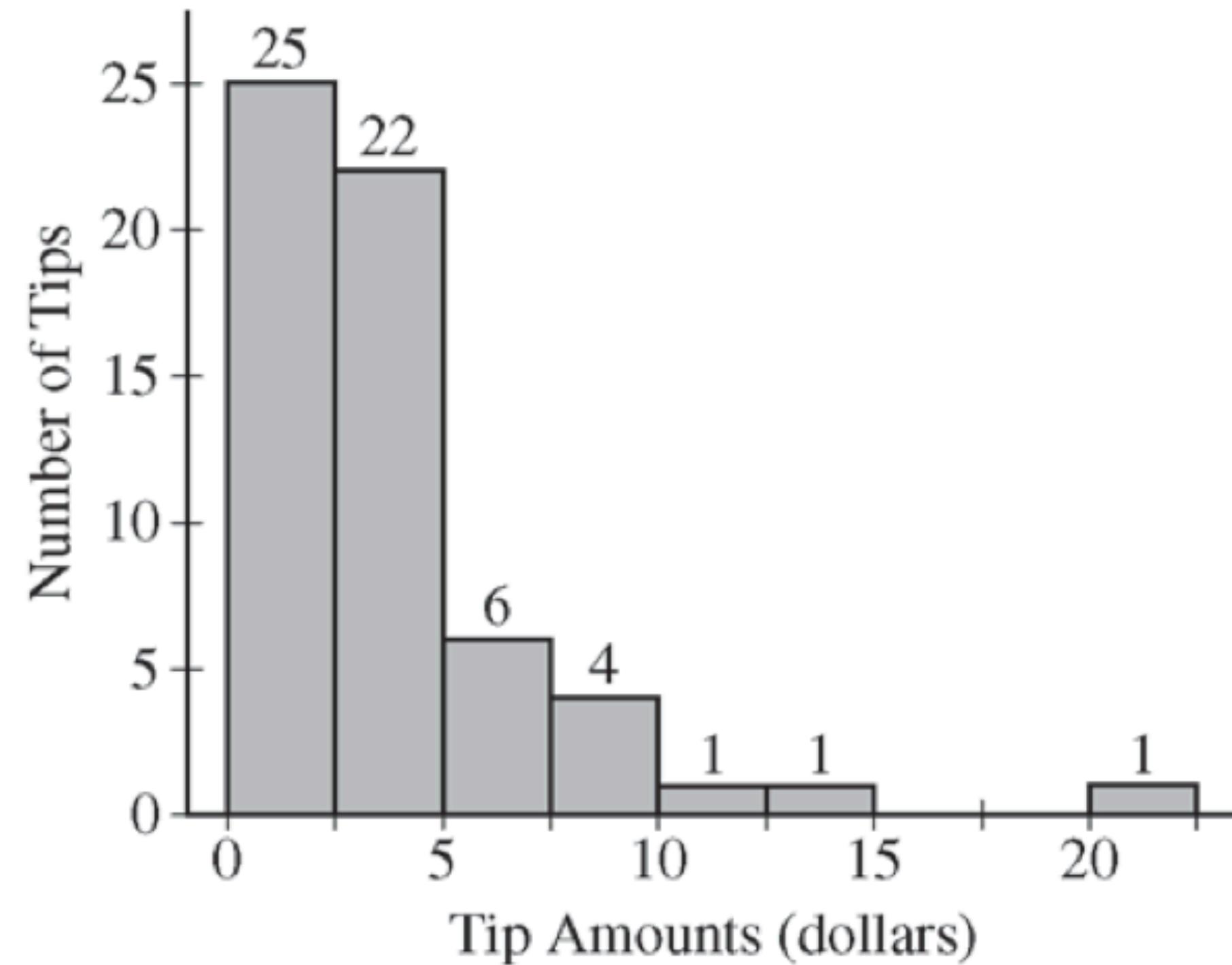


# EMPIRICAL RULE FOR APPROXIMATELY NORMAL DATA



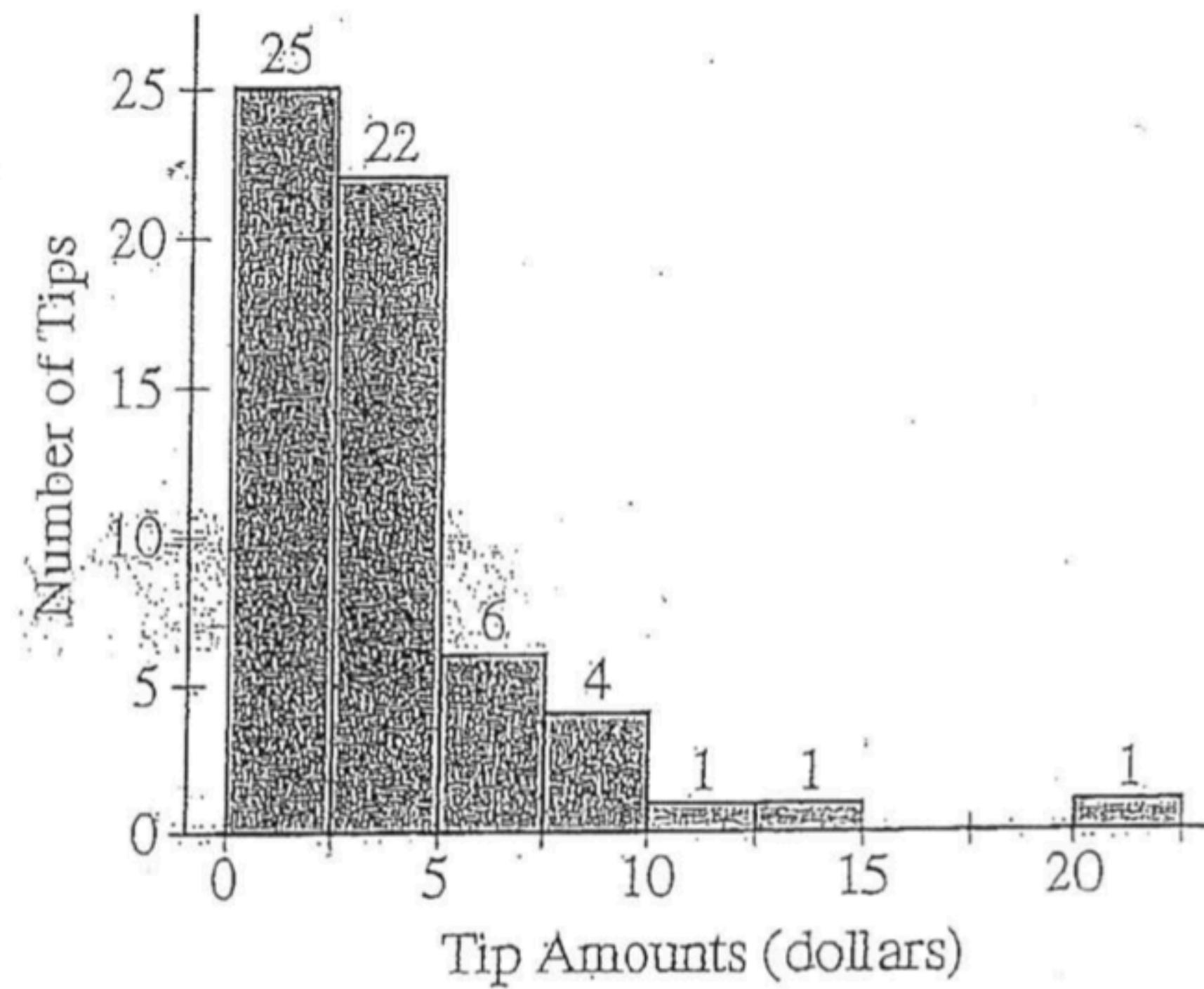
If we have the range and know the data is approximately normal, how might we approximate standard deviation?

1. Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.



- (a) Write a few sentences to describe the distribution of tip amounts for the day shown.

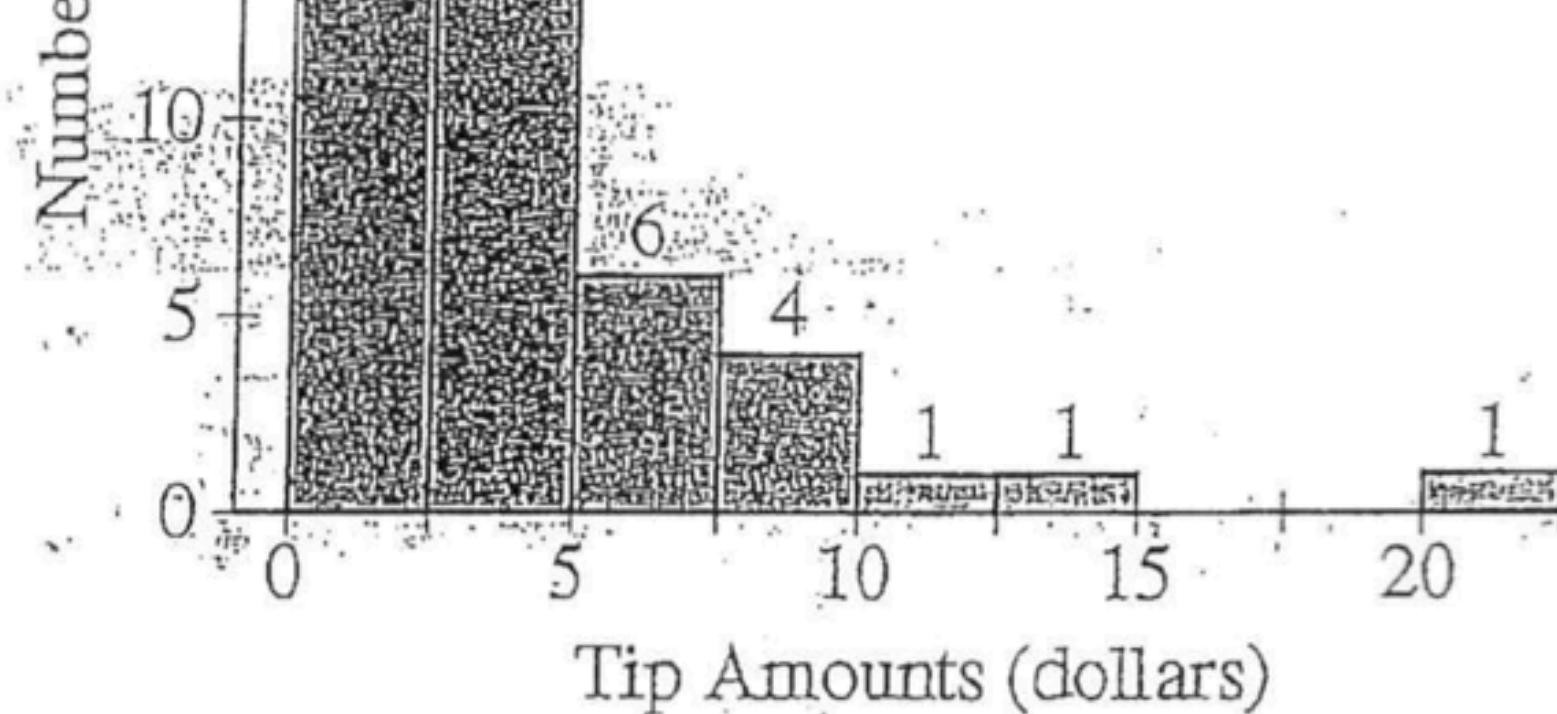
1. Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.



- (a) Write a few sentences to describe the distribution of tip amounts for the day shown.

The center of the distribution is between 2.5 and 5 dollars.

The distribution is skewed right. It has a range of \$25 and a standard deviation near 2.3 dollars.



- (a) Write a few sentences to describe the distribution of tip amounts for the day shown.

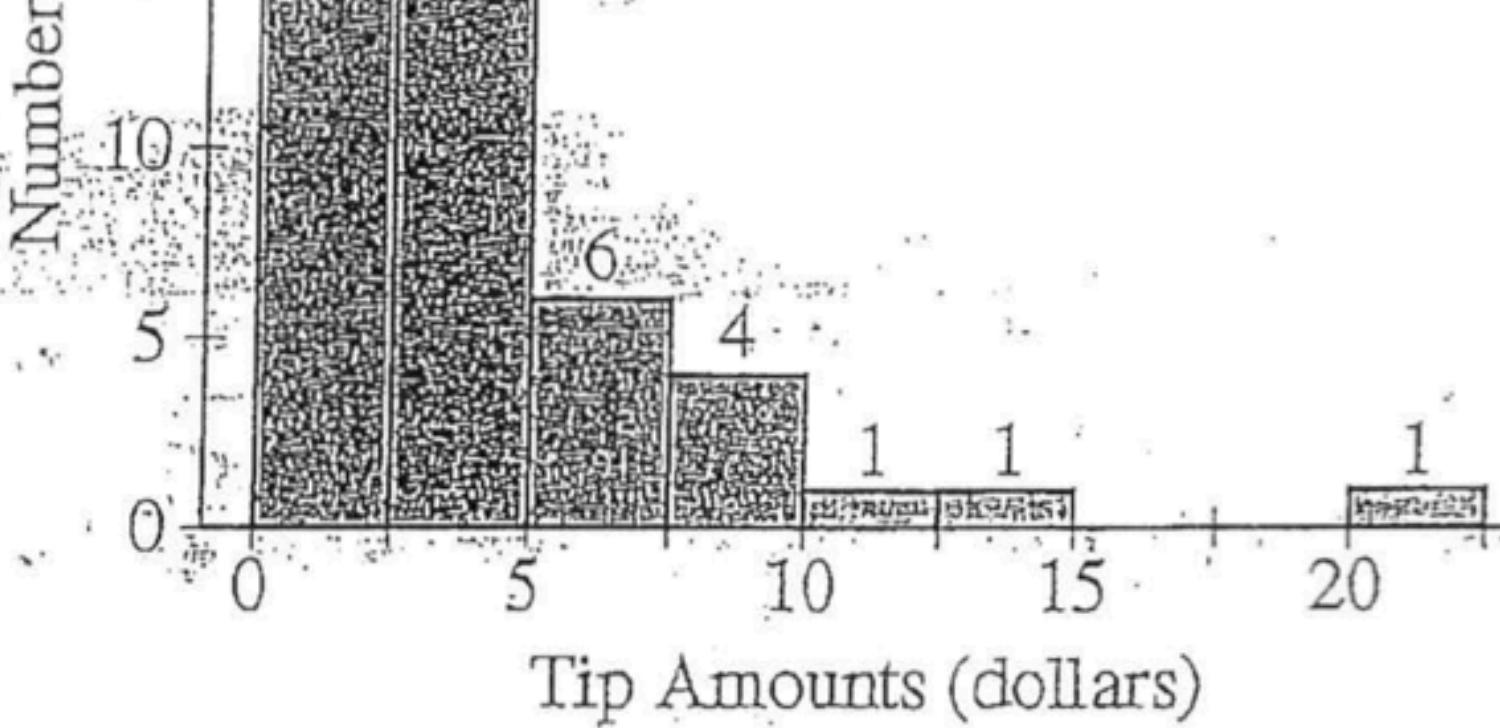
The distribution of tips is skewed to the right and is unimodal.

Possible outlier at 22.5 dollars.

The largest tip is between 20 and 22.5 dollars and the lowest tip is between 0 and 2.5 dollars. The range of the tips is between 17.5 and 22.5 dollars.

The median tip is between 2.5 and 5 dollars.

There is a gap between 15 and 20 dollars, so no tips were received that were between 15 and 20 dollars.



- (a) Write a few sentences to describe the distribution of tip amounts for the day shown.

The distribution of tips is skewed to the right and is unimodal.

Possible outlier at 22.5 dollars.

The largest tip is between 20 and 22.5 dollars and the lowest tip is between 0 and 2.5 dollars. The range of the tips is between 17.5 and 22.5 dollars.

The median tip is between 2.5 and 5 dollars.

There is a gap between 15 and 20 dollars, so no tips were received that were between 15 and 20 dollars.

## **Solution**

### **Part (a):**

The distribution of Robin's tip amounts is skewed to the right. There is a gap between the largest tip amount (in the \$20 to \$22.50 interval) and the second largest tip amount (in the \$12.50 to \$15 interval), and the largest tip amount appears to be an outlier. The median tip amount is between \$2.50 and \$5.00. Robin's tip amounts vary from a minimum of between \$0 and \$2.50 to a maximum of between \$20.00 and \$22.50. About 78 percent of the tip amounts are between \$0 and \$5.

### **Part (a)** is scored as follows:

Essentially correct (E) if the response includes reasonable comments on the following five components:

1. Shape (skewed right)
2. Outlier (at least one) *OR* gap (one tip amount greater than \$20, next highest at most \$15)
3. Center between \$2.50 and \$5.00 (median) or between \$2.62 and \$5.13 (mean)
4. Variability, by noting that the tip amounts vary from about \$0 to at most \$22.50, or that a majority of tip amounts are between \$0 and a value greater than or equal to \$5, or by providing a correct numerical approximation of a measure of variability
5. Context (tip amounts)

Partially correct (P) if the response includes only three or four of the five components.

Incorrect (I) if the response includes at most two of the five components.