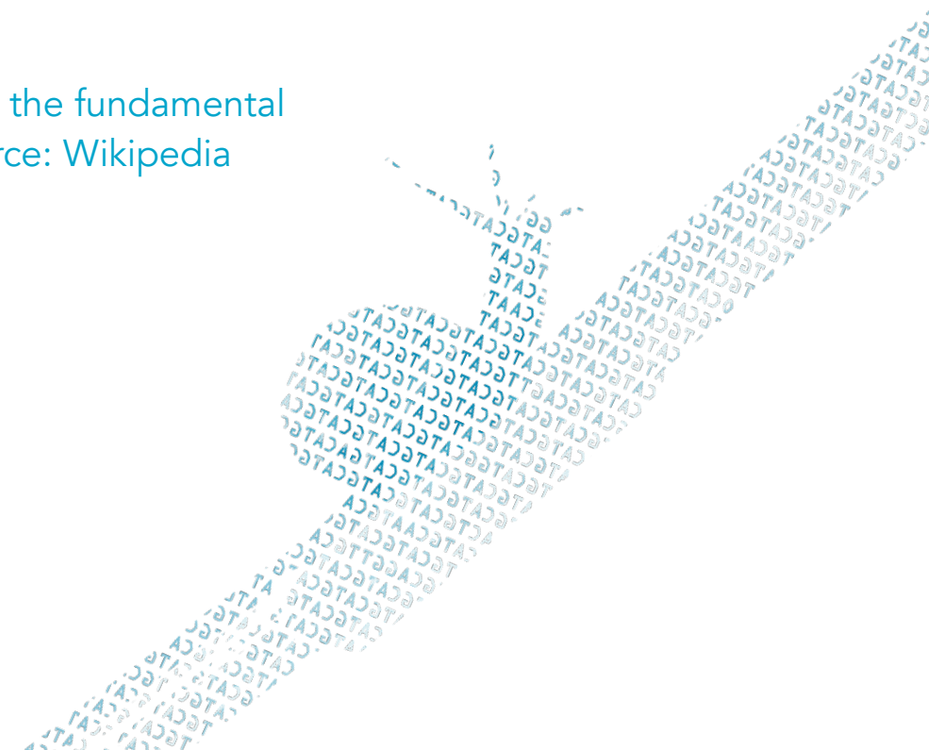
GDC
Genetic Diversity Centre Zurich

Concepts

Concepts are defined as abstract ideas. They are understood to be the fundamental building blocks underlying principles, thoughts and beliefs. — Source: Wikipedia

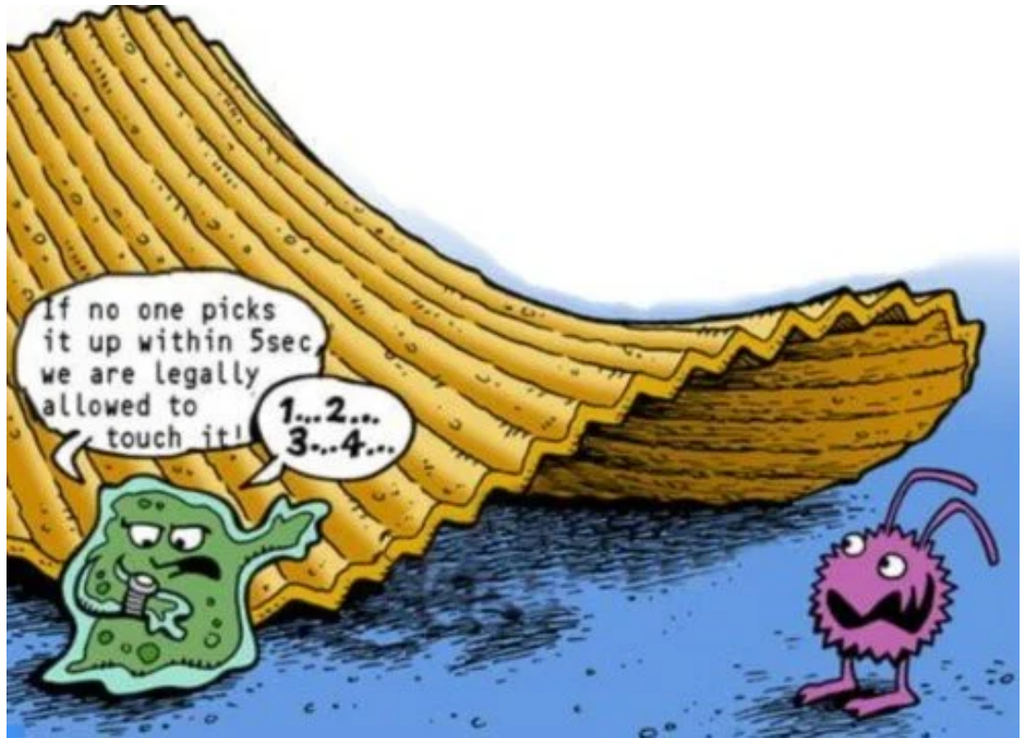


ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Their Master's Voice - Michael Sowa

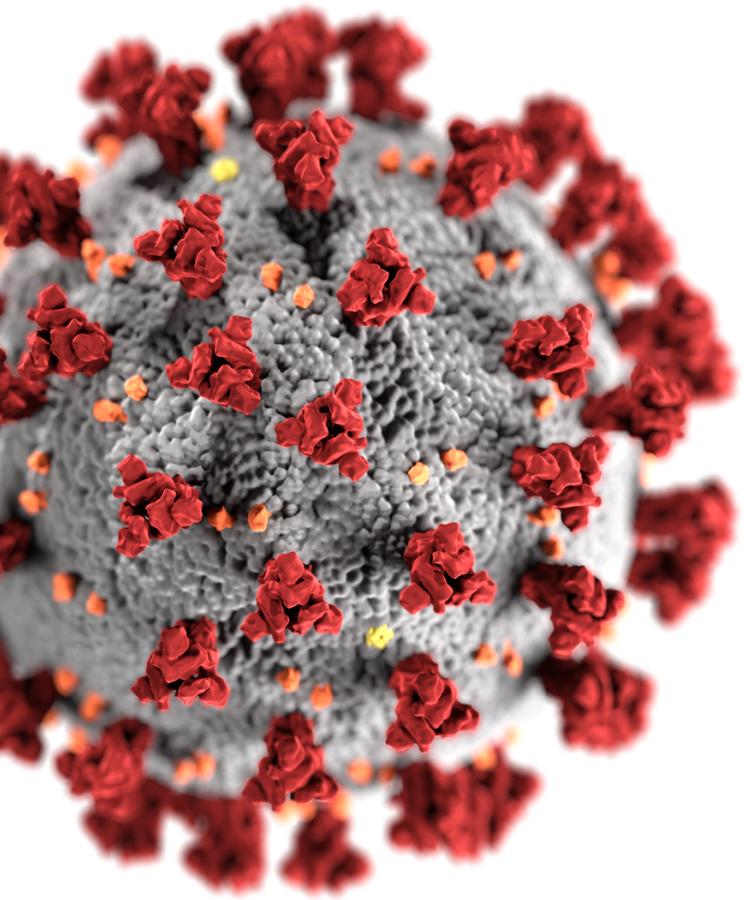
Stop copying!
Start thinking!

Encourages critical thinking
Fosters originality
Builds long-term skills
Challenges passive learning
Combats academic dishonesty



Source: LubioScience 2025 April Marketplace

The **5-second rule** is that magical bit of kitchen folklore which claims that if you pick up dropped food within five seconds, it's still perfectly edible. According to this logic, germs politely wait until second six to launch their microscopic invasion. Spoiler: science doesn't quite back this up… but hey, if it's a cookie, we've all been willing to gamble.



5-µm-threshold

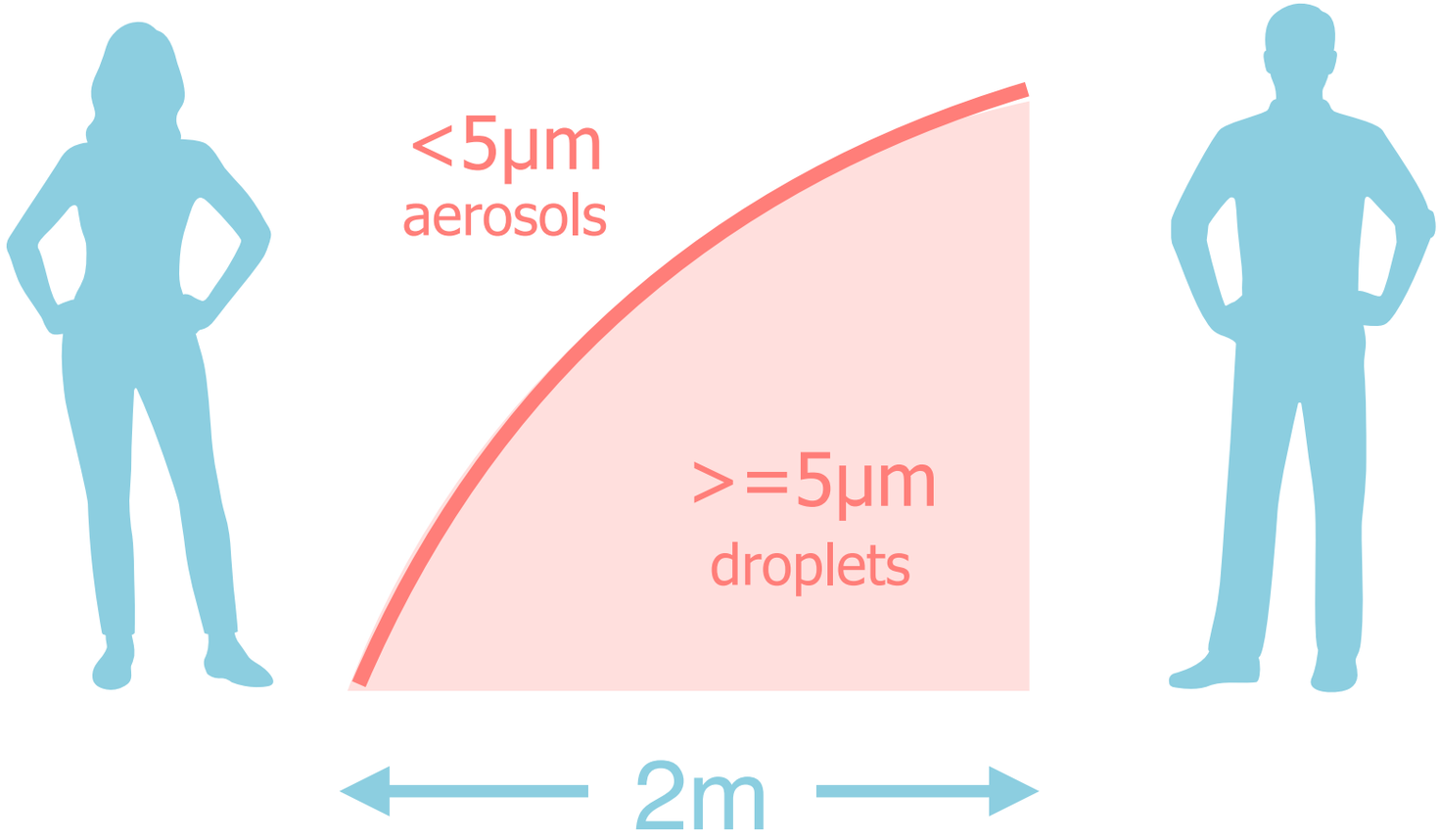
Red blood cell: 7µm
Bacillus: 0.5µm
Coronavirus: 0.1µm

Tang et al. (2021) Covid-19 has redefined airborne transmission - Improving indoor ventilation and air quality will help us all to stay safe. BMJ

Genetic Diversity Centre Zurich

Why?

Any infectious particle smaller than 5 microns in diameter is an aerosol; anything bigger is a droplet.



The WHO and the U.S. Centers for Disease Control and Prevention have also cited 5 micrometers as the threshold for the dichotomy between droplets and aerosol.

WHO: “FACT: #COVID19 is NOT airborne.”

World Health Organization

Airborne transmission of SARS-CoV-2: The world should face the reality

Lidia Morawska[a,*], Junji Cao[b]

[a] International Laboratory for Air Quality and Health (ILAQH), School of Earth of Atmospheric Sciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia
[b] Key Lab of Aerosol Chemistry & Physics (KLACP), Chinese Academy of Sciences, Beijing, China

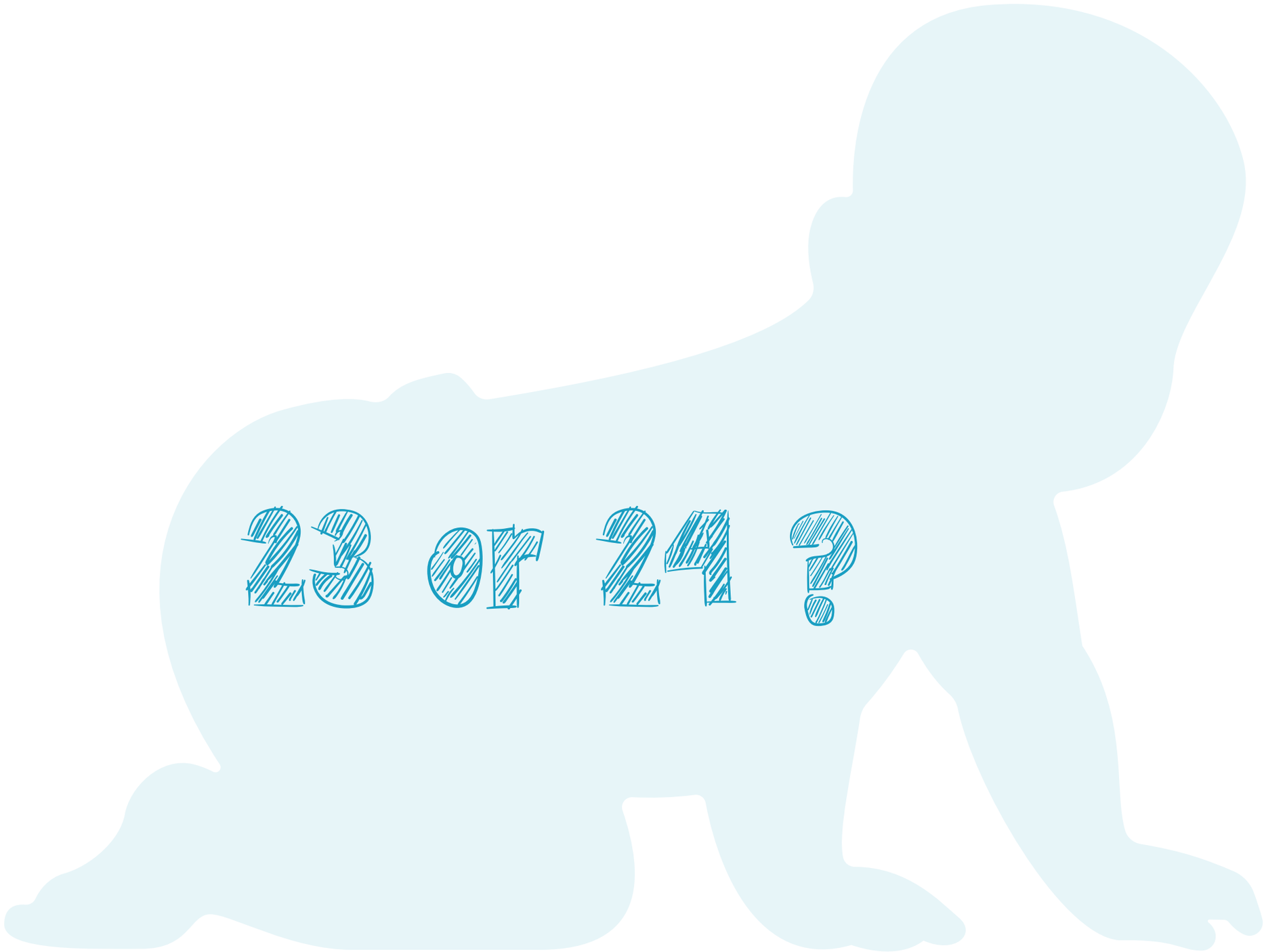
ABSTRACT

Hand washing and maintaining social distance are the main measures recommended by the World Health Organization (WHO) to avoid contracting COVID-19. Unfortunately, these measured do not prevent infection by inhalation of small droplets exhaled by an infected person that can travel distance of meters or tens of meters in the air and carry their viral content. Science explains the mechanisms of such transport and there is evidence that this is a significant route of infection in indoor environments. Despite this, no countries or authorities consider airborne spread of COVID-19 in their regulations to prevent infections transmission indoors. It is therefore extremely important, that the national authorities acknowledge the reality that the virus spreads through air, and recommend that adequate control measures be implemented to prevent further spread of the SARS-CoV-2 virus, in particularly removal of the virus-laden droplets from indoor air by ventilation.

Morawska & Cao (2020) Environment International 139, 105730

"According to the medical canon, nearly all respiratory infections transmit through coughs or sneezes: Whenever a sick person hacks, bacteria and viruses spray out like bullets from a gun, quickly falling and sticking to any surface within a blast radius of 3 to 6 feet. If these droplets alight on a nose or mouth (or on a hand that then touches the face), they can cause an infection. **Only a few diseases were thought to break this droplet rule. Measles and tuberculosis transmit a different way; they're described as "airborne."** Those pathogens travel inside aerosols, microscopic particles that can stay suspended for hours and travel longer distances. They can spread when contagious people simply breathe."

— Linsey Marr (Source: WIRED 2021)



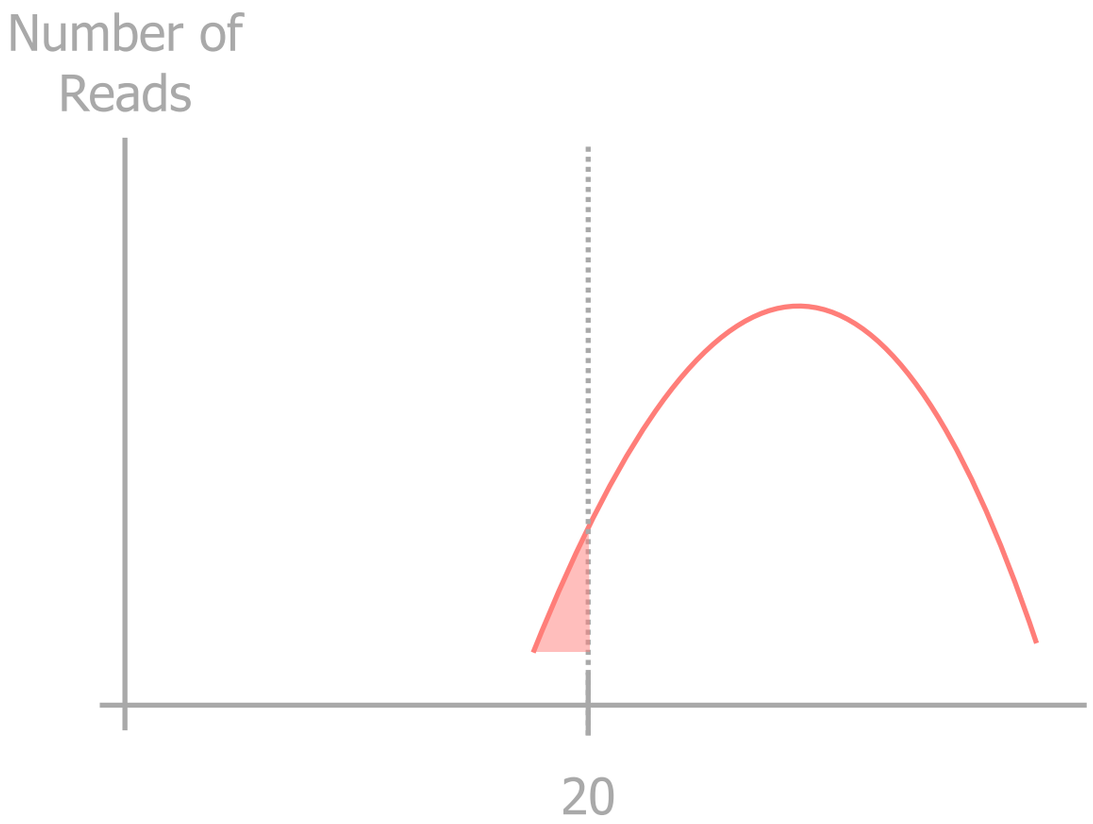
Gartler (2006) The Chromosome number in humans: a brief history.

For decades it was widely believed that humans had 48 chromosomes (24 pairs), based on microscopy from the early 1900s, where poor staining and visualisation often caused chromosomes to clump or overlap, leading to miscounts. In 1956, Joe Hin Tjio and Albert Levan corrected this using improved techniques, revealing the true number to be 46 chromosomes (23 pairs). Despite the new evidence, the earlier error persisted in textbooks and scientific circles for years, as entrenched beliefs and authoritative sources made it difficult to challenge what had long been accepted as fact.

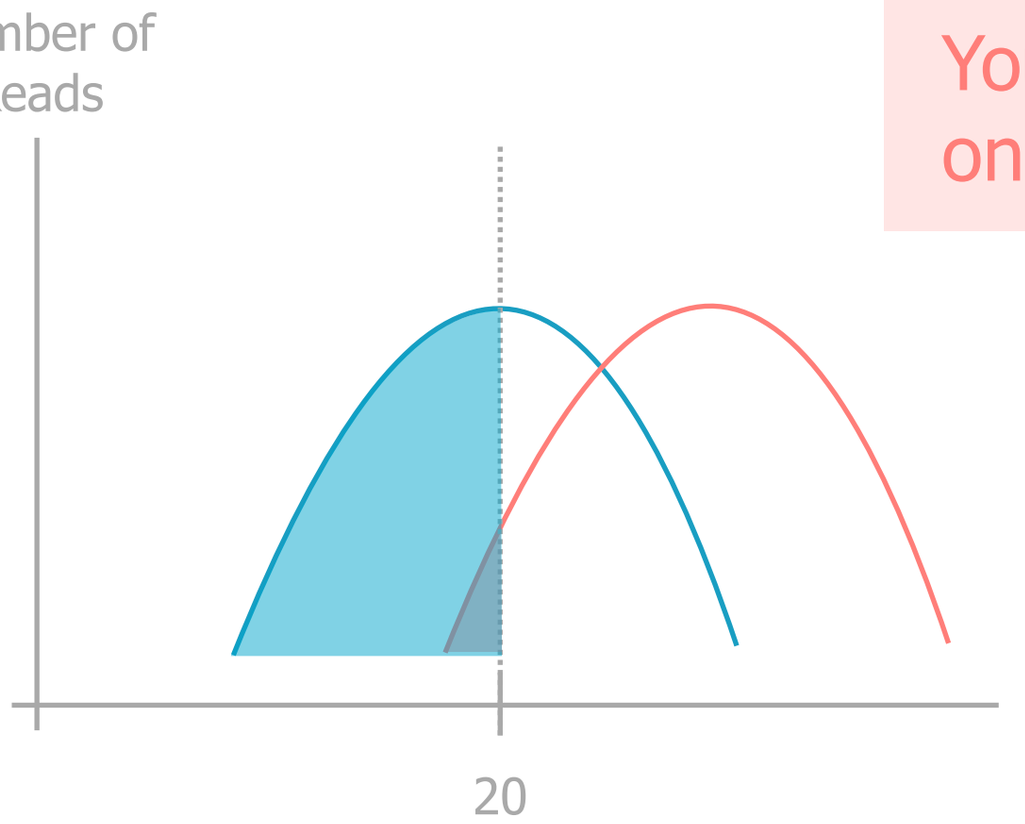
Why?

20

mean quality threshold



mean Q=20 ➜ -5%



Your decision should be based on the structure of **your** data.

mean Q=20 ➜ -5%

mean Q=20 ➜ -50%

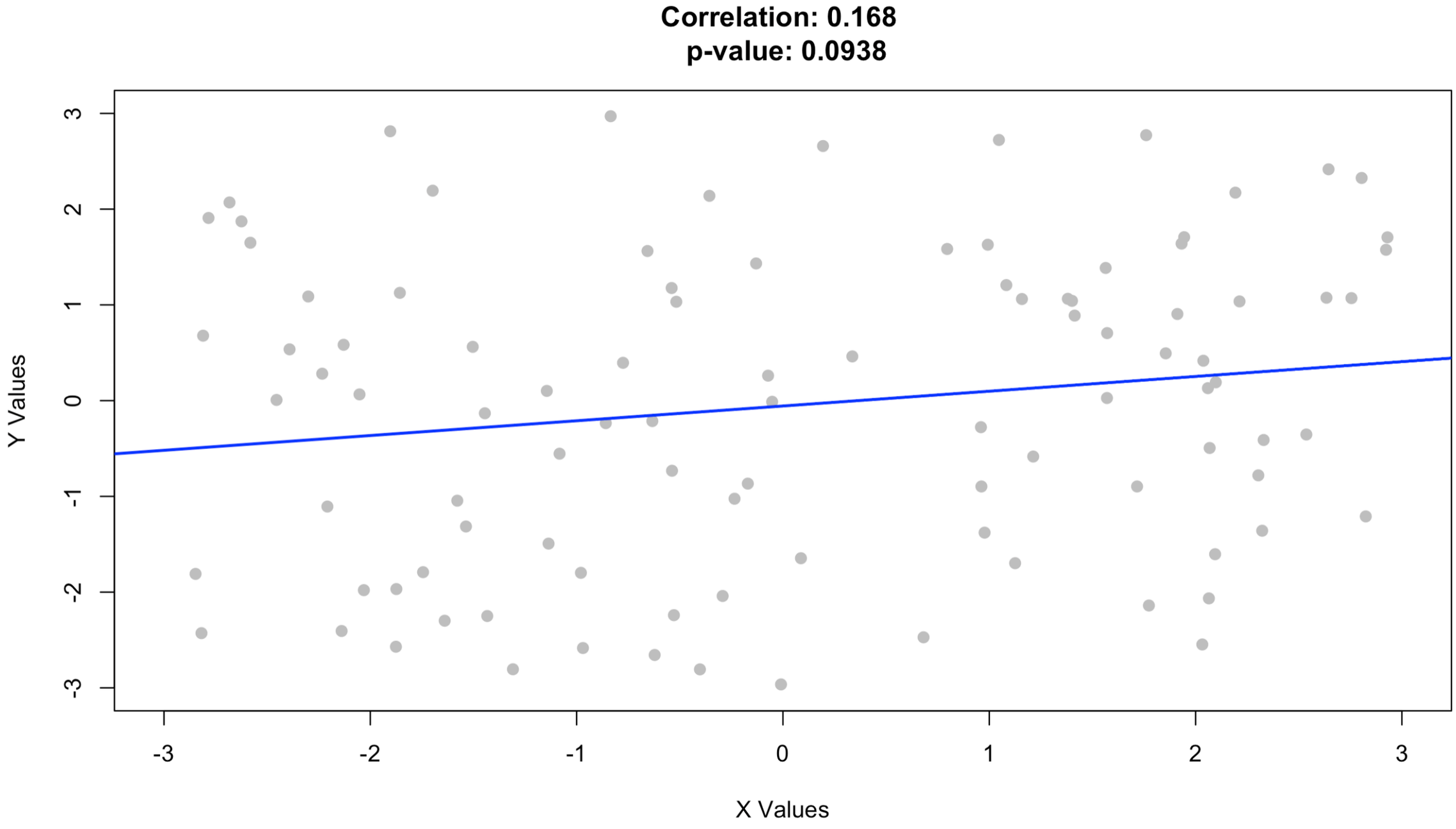
Genetic Diversity Centre Zurich

"You can fight confirmation bias, burst filter bubbles, and escape echo chambers by **actively engaging with ideas that challenge** (your) **assumptions**."

Source: "Think Again" Adam Grant

Genetic Diversity Centre Zurich

Correlation is when two things go to the party together.
Causation is when one dragged the other there.


Correlation: 0.168
p-value: 0.0938
Y Values
X Values
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

```
# Set the number of data points
n <- 100

# Generate random values for x and y
x <- runif(n, -3, 3)
y <- runif(n, -3, 3)

# Fit a linear model
fit <- lm(y ~ x)

# Extract correlation and p-value
correlation <- cor(x, y)
summary_fit <- summary(fit)
p_value     <- summary_fit$coefficients[2, 4]  # Extracting p-value for x

# Plot the data
plot(x, y,
     xlab = "X Values", ylab = "Y Values",
     pch = 19, col = "gray",
     xlim = c(-3, 3), ylim = c(-3, 3),
     main = paste("Correlation:", round(correlation, 3), "\n p-value:", signif(p_value, 3)))

# Add regression line
abline(fit, col = "blue", lwd = 2)

# Print results
print(paste("Linear correlation coefficient (r):", round(correlation, 3)))
print(paste("P-value:", signif(p_value, 3)))
```

Genetic Diversity Centre Zurich

	r	P-value
1	0.156	0.121
2	-0.118	0.243
3	-0.05	0.619
4	-0.272	0.00611
5	0.14	0.164
6	0.01	0.92
7	-0.039	0.698
8	-0.004	0.965
9	-0.292	0.00324
10	-0.052	0.608

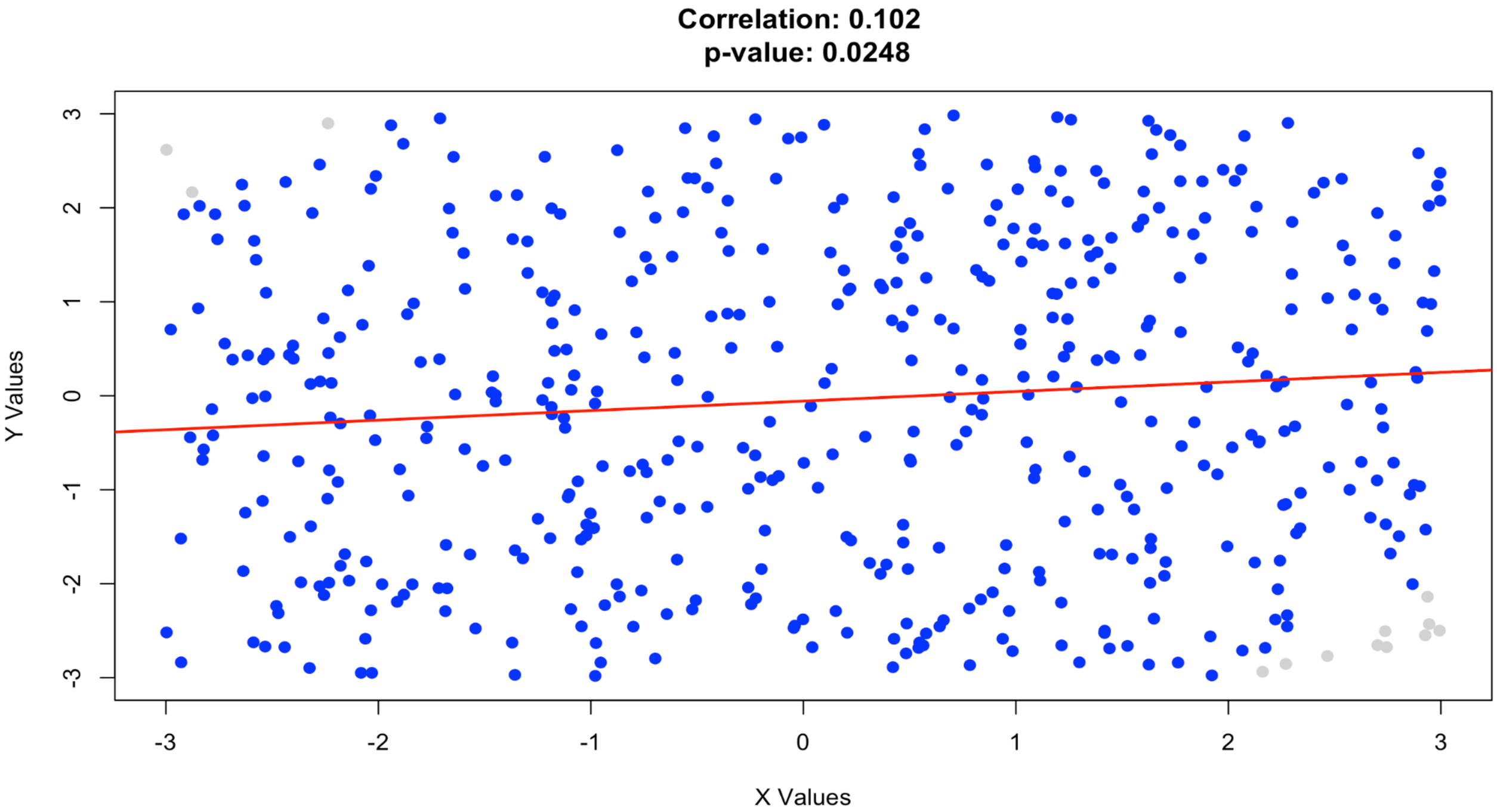
	r	P-value
1	0.156	0.121
2	-0.118	0.243
3	-0.05	0.619
4	**-0.272**	**0.00611**
5	0.14	0.164
6	0.01	0.92
7	-0.039	0.698
8	-0.004	0.965
9	**-0.292**	**0.00324**
10	-0.052	0.608

Genetic Diversity Centre Zurich

	r	P-value
1	0.156	0.121
2	-0.118	0.243
3	-0.05	0.619
4	**-0.272**	**0.00611**
5	0.14	0.164
6	0.01	0.92
7	-0.039	0.698
8	-0.004	0.965
9	**-0.292**	**0.00324**
10	-0.052	0.608

Total simulations: 1000
Significant correlations ($p < 0.05$): 54

~5%


Correlation: 0.102
p-value: 0.0248
Y Values
X Values
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

Ecological Niche Selection

Animals or plants might only survive within a certain tolerance of environmental conditions.

If x is the optimal temperature for plant growth and y is the actual observed temperature, you will mostly find plants where y ≈ x.

This removes extreme outliers, making it seem like the plant's preferred temperature has a stronger correlation with the observed temperature.

Medical Diagnosis Bias (Selective Screening)

If x represents a predicted risk score for a disease and y represents the actual diagnosis, doctors might only test people whose risk (x) is within a reasonable range.

This could result in a dataset where only patients with risk scores close to the actual outcome are included, creating an artificial correlation.

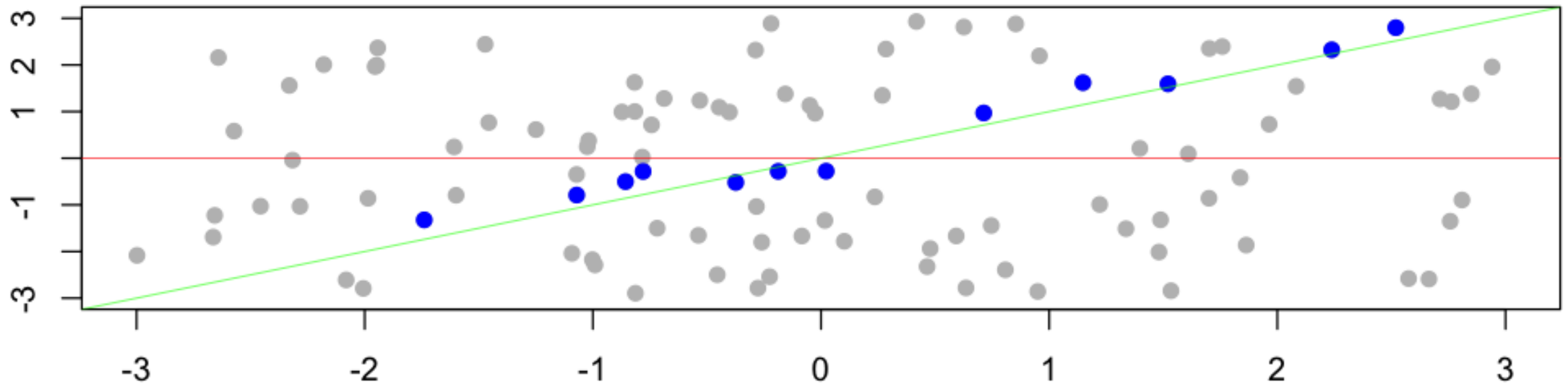
Human Perception Bias (Selective Memory)

People tend to **remember** or **report** experiences that align with their expectations.

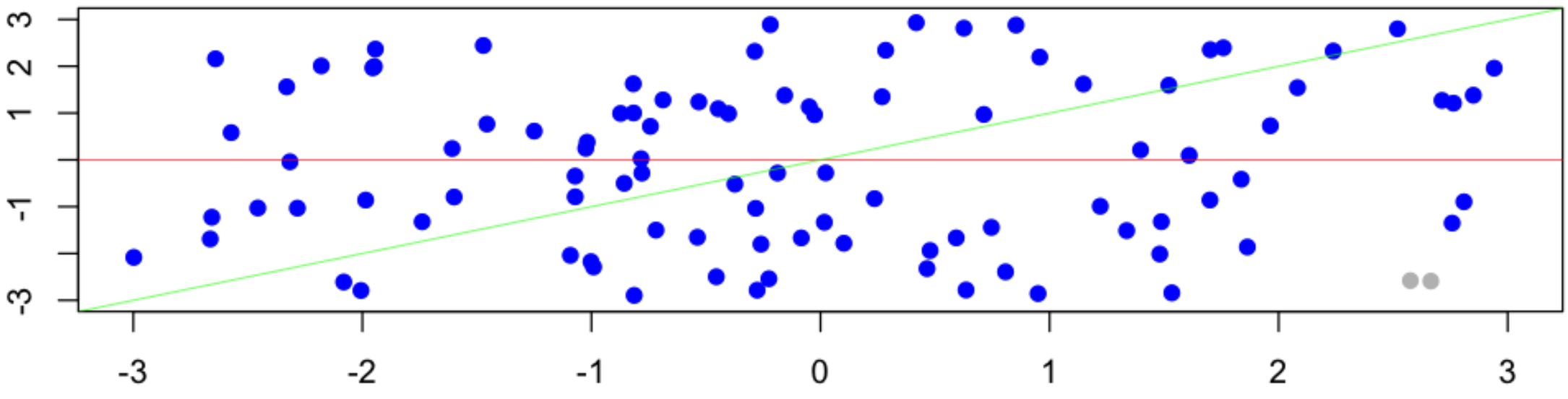
If x represents a person's expected rating of an experience and y is the actual rating, people might only recall or report experiences where y ≈ x, ignoring surprising or disappointing events.

```
...

# Generate random values for x and y
x <- runif(n, -3, 3)
y <- runif(n, -3, 3)

# Apply selection bias: Keep points where y is within a range of x
tolerance <- 5  # Adjust to control the selection bias
valid_data <- which(y >= (x - tolerance) & y <= (x + tolerance))

# Separate filtered (kept) and removed points
x_filtered <- x[valid_data]
y_filtered <- y[valid_data]
x_removed  <- x[-valid_data]
y_removed  <- y[-valid_data]

...
```

Simulation: 100 - **Tolerance: 0.5** - Correlation: 0.982



Simulation: 100 - **Tolerance: 5.0** - Correlation: 0.118



Genetic Diversity Centre Zurich

Total simulations: 1000

	Tolerance	%
1	0.5	100
2	1	100
3	1.5	100
4	2	100
5	2.5	100
6	3	100
7	3.5	96.2
8	4	66.9
9	4.5	28.2
10	5	10.4

% of significant correlations ($p < 0.05$)

Tolerance in this script defines the allowable range around the line y = x within which data points are selected for correlation — effectively controlling how strictly points must align to be included.

Just because two genes are co-expressed doesn’t mean they’re in a relationship.

Mosquito bites and ice cream consumption both rise in summer. Still not blaming Gasparini for malaria.

Finding a correlation in biology is like spotting two sheep together — cool, but let’s not publish just yet.

Source: https://www.facebook.com/trust.biologist/posts/5241541502625082/

Scientific Context

ICYWW

in **c**ase **y**our **w**ere **w**ondering

You totally can make popcorn with a hair straightener!

Source: YouTube

Dent Corn
(*Zea mays var. indentata*)

Flint Corn
(*Zea mays var. indurata*)

Popcorn
(*Zea mays var. everta*)

Sweet Corn
(*Zea mays convar. saccharata var. rugosa*)

Flour corn
(*Zea mays var. amylacea*)

A kernel of popcorn contains **moisture and oil**. Unlike most other grains, the outer hull of the popcorn kernel is both **strong and impervious** to moisture and the starch inside consists almost entirely of a hard type. As the oil and water within the kernel are heated, they turn the moisture in the kernel into **pressurized steam**. Under these conditions, the starch inside the kernel gelatinizes, softens, and becomes pliable. The internal pressure of the entrapped steam continues to increase until the breaking point of the hull is reached: a pressure of approximately **930 kPa** and a temperature of **180 °C**. The hull thereupon ruptures rapidly and explodes, causing a sudden drop in pressure inside the kernel and a corresponding rapid expansion of the steam, which expands the starch and proteins of the endosperm into airy foam. As the foam rapidly cools, the starch and protein polymers set into the familiar crispy puff. Special varieties are grown to give improved popping yield. Though the kernels of some wild types will pop, the cultivated strain is **Zea mays everta**, which is a special kind of flint corn.

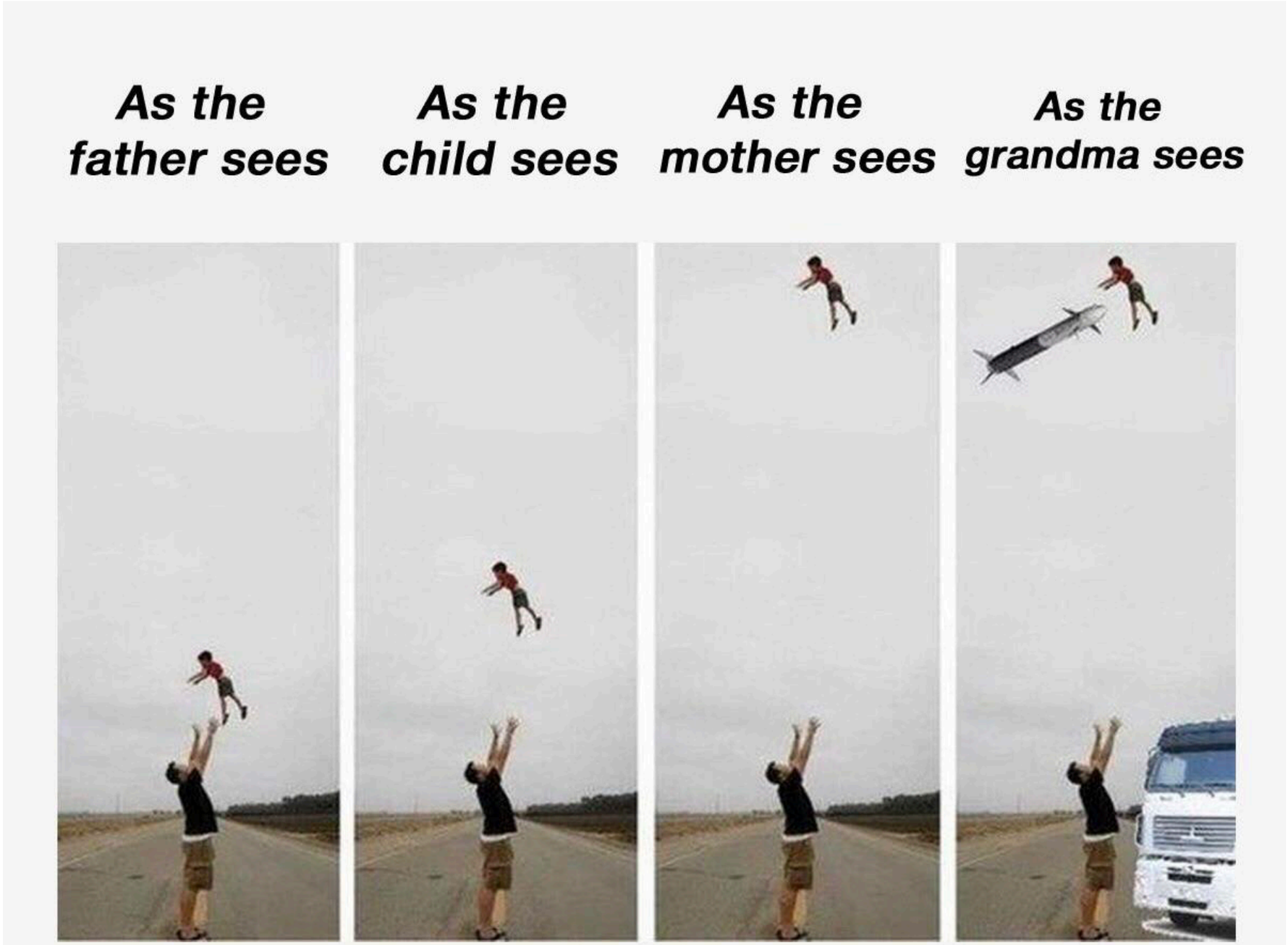
Source: Wikipedia

"Science is not about making predictions or performing experiments. Science is about explaining."

— Bill Gaede

*Guillermo "Bill" Gaede is an Argentine engineer and programmer who is best known for Cold War industrial spying conducted while he worked at AMD and Intel.

Perspective



Bed bugs (*Cimex lectulariu*)

$F_{ST} = 0.68$

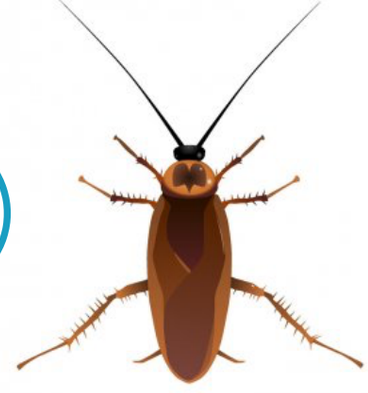
Saenz et al. (2012) Genetic Analysis of Bed Bug Populations Reveals Small Propagule Size Within Individual Infestations but High Genetic Diversity Across Infestations From the Eastern United States.

Bed bugs (*Cimex lectulariu*)

$F_{ST} = 0.68$

German cockroach (*Blattella germanic*)

$F_{ST} = 0.099$



Saenz et al. (2012) Genetic Analysis of Bed Bug Populations Reveals Small Propagule Size Within Individual Infestations but High Genetic Diversity Across Infestations From the Eastern United States.

Add biology (meaning) to numbers!

Noise Whisperer

"Your chest x-ray is fine, but your driver's license has expired."

Genetic Diversity Centre Zurich

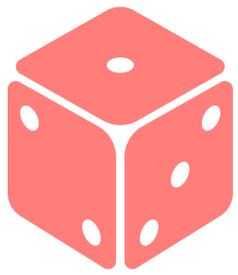
Hypothesis-driven analysis or data-mining?

Genetic Diversity Centre Zurich

Hypothesis-driven analysis or datamining?

1 - 2 - 1 - 1 - 1 - 1 - 5 - 5 - 1 - 1 - 3 - 1 - 1 - 4 - 4 - 1

10 <1> out of 16 (62.5%)



Genetic Diversity Centre Zurich

Hypothesis-driven analysis or data-mining?

5 - 5 - 6 - 2 - 2 - 5 - 5 - 1 - 6 - 4 - 6 - 6 - 1 - 1 - 4 - 1 - 6 - 2 - 1 - 5 - 1 - 1 - 4 - 6 - 6 - 1 -
2 - 2 - 6 - 1 - 1 - 3 - 1 - 5 - 3 - 3 - 5 - 1 - 6 - 3 - 5 - 5 - 1 - 1 - 5 - 4 - 4 - 6 - 3 - 3 - 6 - 6 -
3 - 2 - 6 - 3 - 1 - 2 - 4 - 1 - 4 - 2 - 1 - 1 - 2 - 1 - 2 - 6 - 4 - 6 - 5 - 1 - 2 - 5 - 2 - 4 - 1 - 6 -
2 - 5 - 3 - 1 - 3 - 3 - 5 - 3 - 1 - 3 - 1 - 5 - 2 - 5 - 6 - 5 - 1 - 3 - 4 - 1 - 1 - 5 - 4 - 3 - 4 - 6 -
1 - 3 - 3 - 2 - 2 - 2 - 1 - 2 - 3 - 5 - 4 - 1 - 3 - 5 - 4 - 5 - 1 - 4 - 1 - 1 - 5 - 3 - 5 - 2 - 6 - 5 -
4 - 2 - 3 - 4 - 5 - 4 - 4 - 6 - 6 - 5 - 3 - 5 - 3 - 5 - 2 - 4 - 5 - 4 - 3 - 3 - 2 - 4 - 4 - 1 - 2 - 5 -
5 - 6 - 6 - 2 - 4 - 5 - 1 - 4 - 4 - 1 - 4 - 4 - 5 - 1 - 2 - 1 - 3 - 2 - 5 - 2 - 2 - 2 - 4 - 4 - 5 - 3 -
3 - 1 - 5 - 4 - 4 - 1 - 4 - 5 - 6 - 3 - 6 - 1 - 2 - 2 - 2 - 2 - 5 - 3 - 5 - 1 - 4 - 3 - 2 - 5 - 5 - 2 -
3 - 2 - 1 - 5 - 2 - 4 - 1 - 4 - 5 - 5 - 2 - 3 - 6 - 5 - 6 - 3 - 1 - 3 - 4 - 5 - 1 - 2 - 5 - 2 - 1 - 4 -
1 - 5 - 4 - 6 - 5 - 5 - 2 - 5 - 3 - 2 - 5 - 2 - 4 - 1 - 6 - 1 - 2 - 6 - 3 - 3 - 3 - 5 - 1 - 4 - 3 - 4 -
3 - 4 - 5 - 3 - 2 - 6 - 3 - 1 - 3 - 3 - 5 - 6 - 6 - 2 - 3 - 3 - 6 - 2 - 5 - 1 - 5 - 1 - 2 - 5 - 2 - 3 -
6 - 2 - 1 - 5 - 1 - 6 - 1 - 3 - 1 - 3 - 6 - 2 - 2 - 6 - 2 - 1 - 6 - 5 - 2 - 3 - 1 - 6 - 4 - 5 - 6 - 1 -
4 - 6 - 2 - 4 - 1 - 3 - 3 - 1 - 3 - 5 - 2 - 3 - 2 - 1 - 1 - 2 - 3 - 3 - 6 - 4 - 5 - 5 - 2 - 3 - 1 - 2 -
2 - 5 - 4 - 6 - 2 - 6 - 3 - 2 - 4 - 5 - 5 - 3 - 6 - 3 - 3 - 4 - 6 - 5 - 3 - 4 - 2 - 5 - 2 - 5 - 5 - 1 -
4 - 1 - 5 - 5 - 4 - 4 - 3 - 2 - 4 - 6 - 2 - 4 - 1 - 4 - 5 - 1 - 5 - 1 - 4 - 5 - 4 - 6 - 6 - 6 - 5 - 5 -
5 - 4 - 4 - 3 - 5 - 5 - 6 - 3 - 1 - 5 - 1 - 5 - 1 - 6 - 1 - 4 - 1 - 6 - 3 - 6 - 5 - 1 - 1 - 5 - 2 - 3 -
1 - 1 - 2 - 1 - 6 - 2 - 5 - 3 - 5 - 5 - 3 - 3 - 5 - 2 - 5 - 5 - 3 - 5 - 5 - 1 - 6 - 5 - 5 - 5 - 4 - 2 -
2 - 2 - 6 - 6 - 3 - 4 - 6 - 6 - 1 - 3 - 2 - 4 - 1 - 3 - 6 - 2 - 5 - 1 - 2 - 1 - 1 - 1 - 1 - 5 - 5 - 1 -
1 - 3 - 1 - 1 - 4 - 4 - 1 - 3 - 2 - 5 - 3 - 5 - 1 - 5 - 5 - 4 - 6 - 1 - 3 - 5 - 5 - 5 - 2 - 1 - 4 - 6 -
2 - 5 - 4 - 2 - 4 - 5 - 3 - 1 - 4 - 5 - 1 - 4 - 3 - 6 - 5 - 1 - 5 - 6 - 4 - 6 - 2 - 5 - 4 - 5 - 6 - 4

- **Hypothesis-driven** analysis uses rigorous statistical tools to test specific, predefined questions and guard against false conclusions.
- **Data mining** explores large datasets to uncover unexpected patterns or generate new hypotheses, often using machine learning and flexible algorithms.
- **Both approaches are essential**: hypothesis testing keeps science grounded, while data mining opens doors to discovery.
- **But mixing them carelessly can be dangerous** — it's tempting to explore freely and then "test" a pattern as if it were a pre-planned hypothesis, which can inflate false positives and lead to misleading conclusions (a.k.a. **p-hacking** or **HARKing** — hypothesizing after results are known).

"A bad process with a good result is luck."

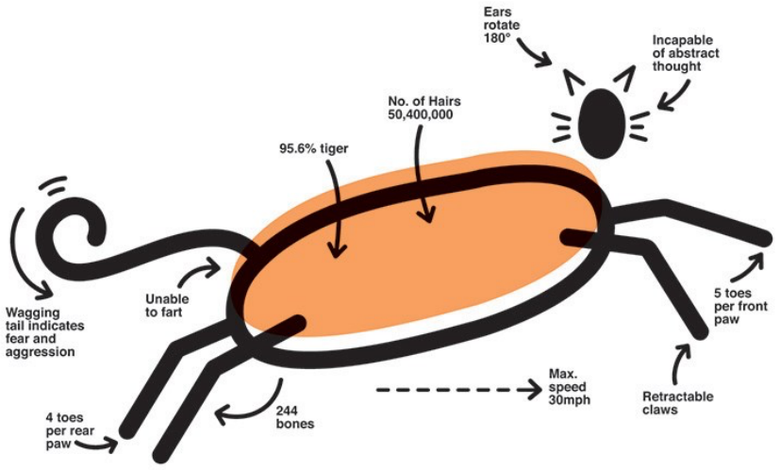
"A good process with a bad outcome could be a clever experiment."

Source: "Think Again" Adam Grant

Graphics

Catology

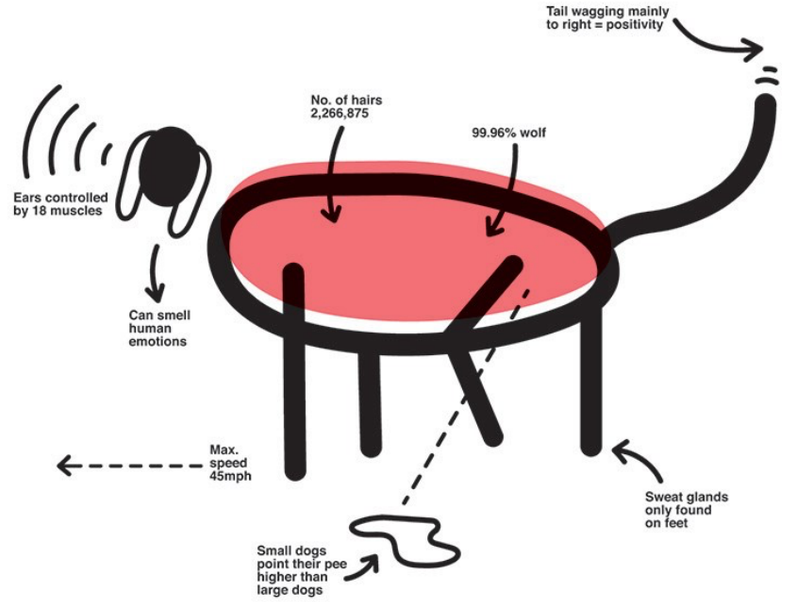
The Weird and Wonderful Science of Cats



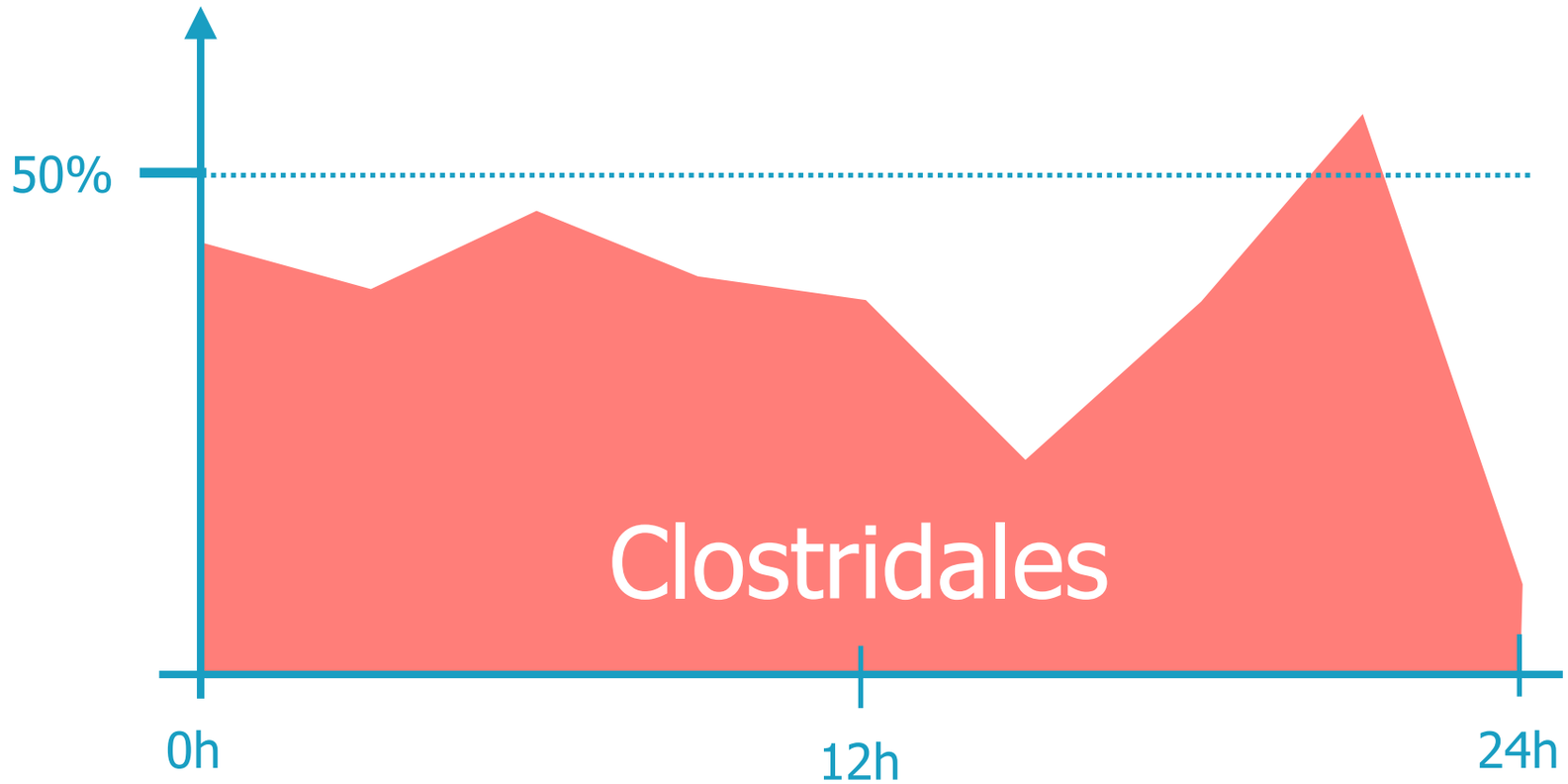
Stefan Gates

Dogology

The Weird and Wonderful Science of Dogs



Stefan Gates



Clostridia are a polyphyletic class of Firmicutes, including Clostridium and other similar genera. They are distinguished from the Bacilli by lacking aerobic respiration. They are obligate anaerobes and oxygen is toxic to them.

source: Wikipedia

Genetic Diversity Centre Zurich

	T06	T08	T10	T12	T14	T16	T18	T20	T22
Clostridiales -	44.7	40.9	46.3	41.3	40.1	27.5	40	54.6	9.7
Bacteroidales -	32	27.8	27.8	34.5	33.1	39.9	25.7	22.6	13.3
Bifidobacteriales -	14	18.9	17.9	8.6	19.1	25.3	20.3	16.1	6.6
Enterobacteriales -	3	2.6	1.6	6.1	1.2	3.4	6.3	0.2	2.8
Lactobacillales -	3	2.3	4.7	7.4	1.5	0.4	2.3	4.9	4.3
Verrucomicrobiales -	1.1	4.6	0	0.3	1.3	1.5	0	0	0
Streptophyta -	0	0	0	0	0	0	0.3	0	60.9
Coriobacteriales -	0.3	0.6	0.2	0.2	0.7	0.2	2	0.4	0.2
Pasteurellales -	0.1	0.2	0.1	0.2	0.7	0.1	0.7	0.2	0.1
Erysipelotrichales -	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0

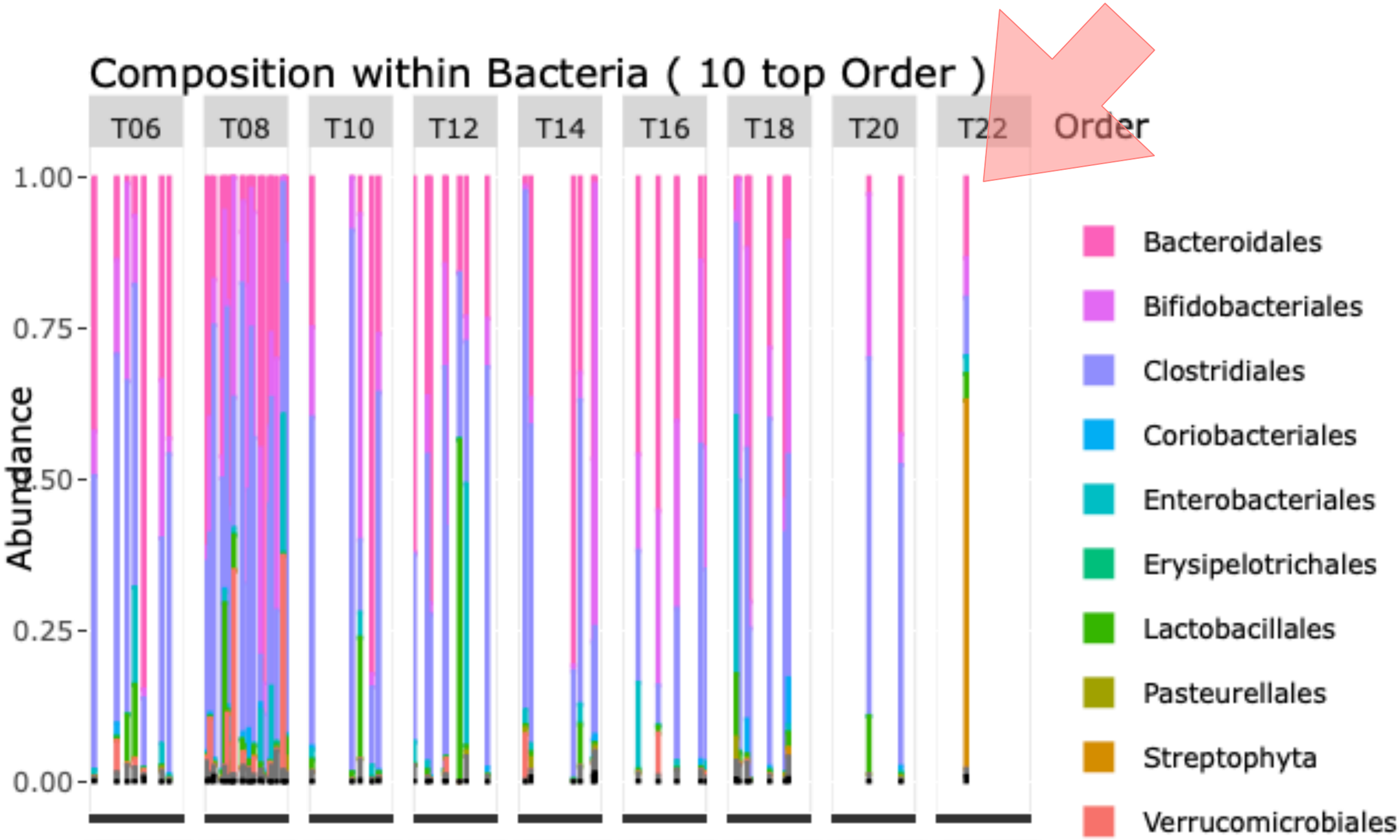
% Read Abundance

10.0

1.0

0.1

N=50

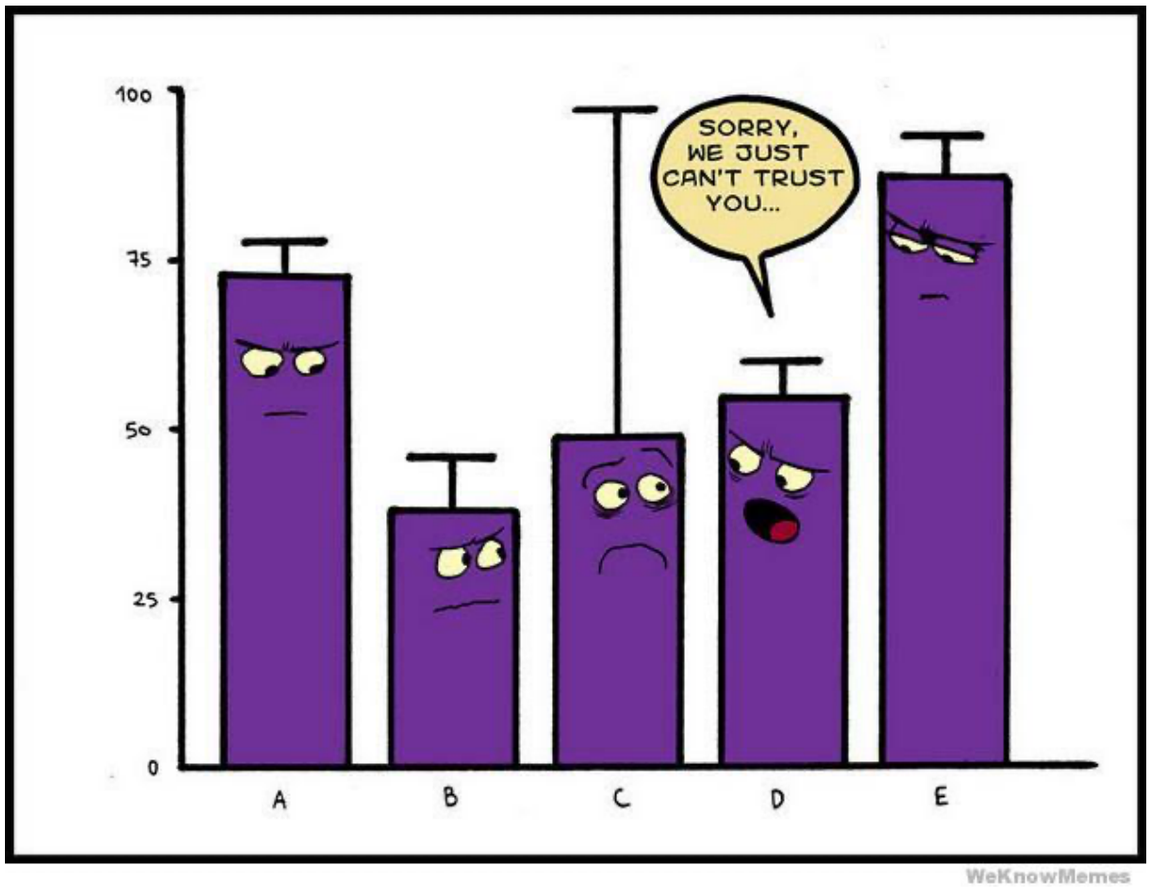

Composition within Bacteria (10 top Order)
T06
T08
T10
T12
T14
T16
T18
T20
T22
Order
1.00
0.75
0.50
0.25
0.00
Abundance
Bacteroidales
Bifidobacteriales
Clostridiales
Coriobacteriales
Enterobacteriales
Erysipelotrichales
Lactobacillales
Pasteurellales
Streptophyta
Verrucomicrobiales

Just because it's shiny doesn't mean it's science.

A pretty plot can still be a dirty lie.

Trust the math, not the makeup.

Statistical Significance



ARTICLE

DOI: 10.1038/s41467-017-02765-w OPEN

Regulation of REM and Non-REM Sleep by Periaqueductal GABAergic Neurons

Franz Weber[1,3], Johnny Phong Hoang Do[1], Shinjae Chung[1,3], Kevin T. Beier[2], Mike Bikov[1], Mohammad Saffari Doost[1] & Yang Dan[1]

Sample sizes. For optogenetic activation experiments, cell-type-specific ablation experiments, and in vivo recordings (optrode recordings and calcium imaging), we continuously increased the number of animals until statistical significance was reached to support our conclusions. For rabies-mediated and anterograde tracing experiments, the selection of the sample size was based on numbers reported in previous studies. For optrode recordings, we first recorded a preliminary data set of six units from two mice. Based on analysis of this data set and given the success rate in finding identified GABAergic units, we predicted that about 20 units are sufficient to statistically support our conclusions.

Code availability. The code used in this study will be made available upon reasonable request.

Data availability. The data that support the findings of this study will be made available upon reasonable request.

Weber et al. (2018) Regulation of REM and Non-REM Sleep by Periaqueductal GABAergic Neurons. Nat Commun 9, 354. https://doi.org/10.1038/s41467-017-02765-w

"We continuously increased the number of animals until statistical significance was reached..."

This describes a practice known as **data peeking**, **optional stopping**, or more broadly, **a form of p-hacking**.

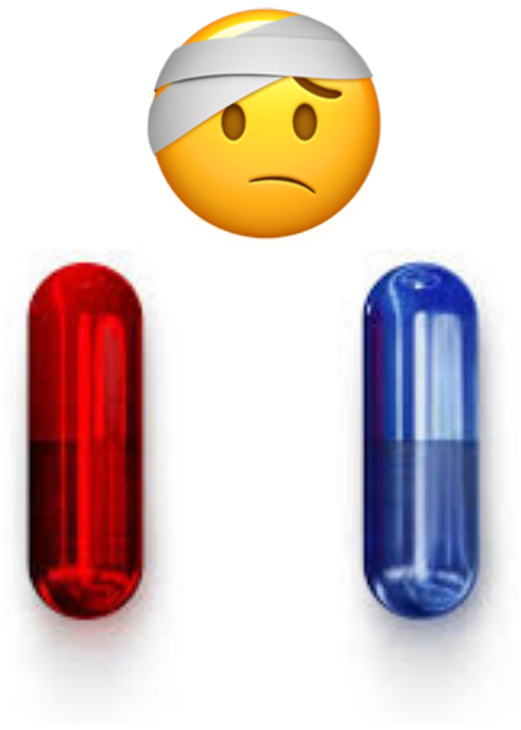
Why It's a Problem?

- 🚫 **Inflates false positives (Type I error)**
- 🚫 **Breaks statistical assumptions**
- 🚫 **Reduces reproducibility**
- 🚫 **Raises ethical concerns**

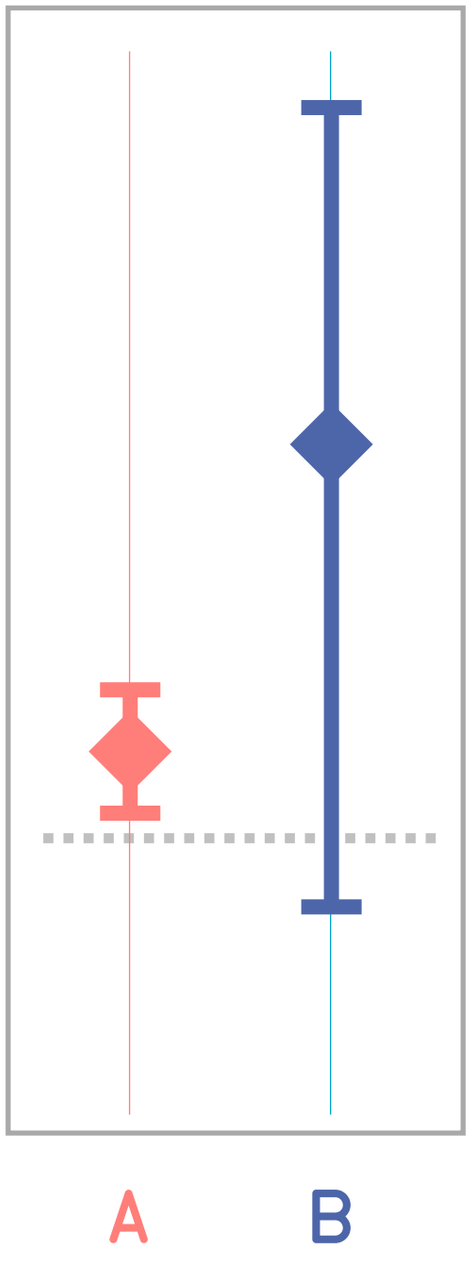
What to do if your p-value is just over the arbitrary threshold for ‘significance’ of p=0.05?

- (barely) not statistically significant (p=0.052)
- a barely detectable statistically significant difference (p=0.073)
- a considerable trend toward significance (p=0.069)
- a margin at the edge of significance (p=0.0608)
- a moderate trend toward significance (p=0.068)
- a strong tendency towards statistical significance (p=0.051)
- an inverse trend toward significance (p=0.06)
- approached acceptable levels of statistical significance (p=0.054)
- at the verge of significance (p=0.058)

Source: https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/



The red pill has a statistically significant reduction in your headaches.

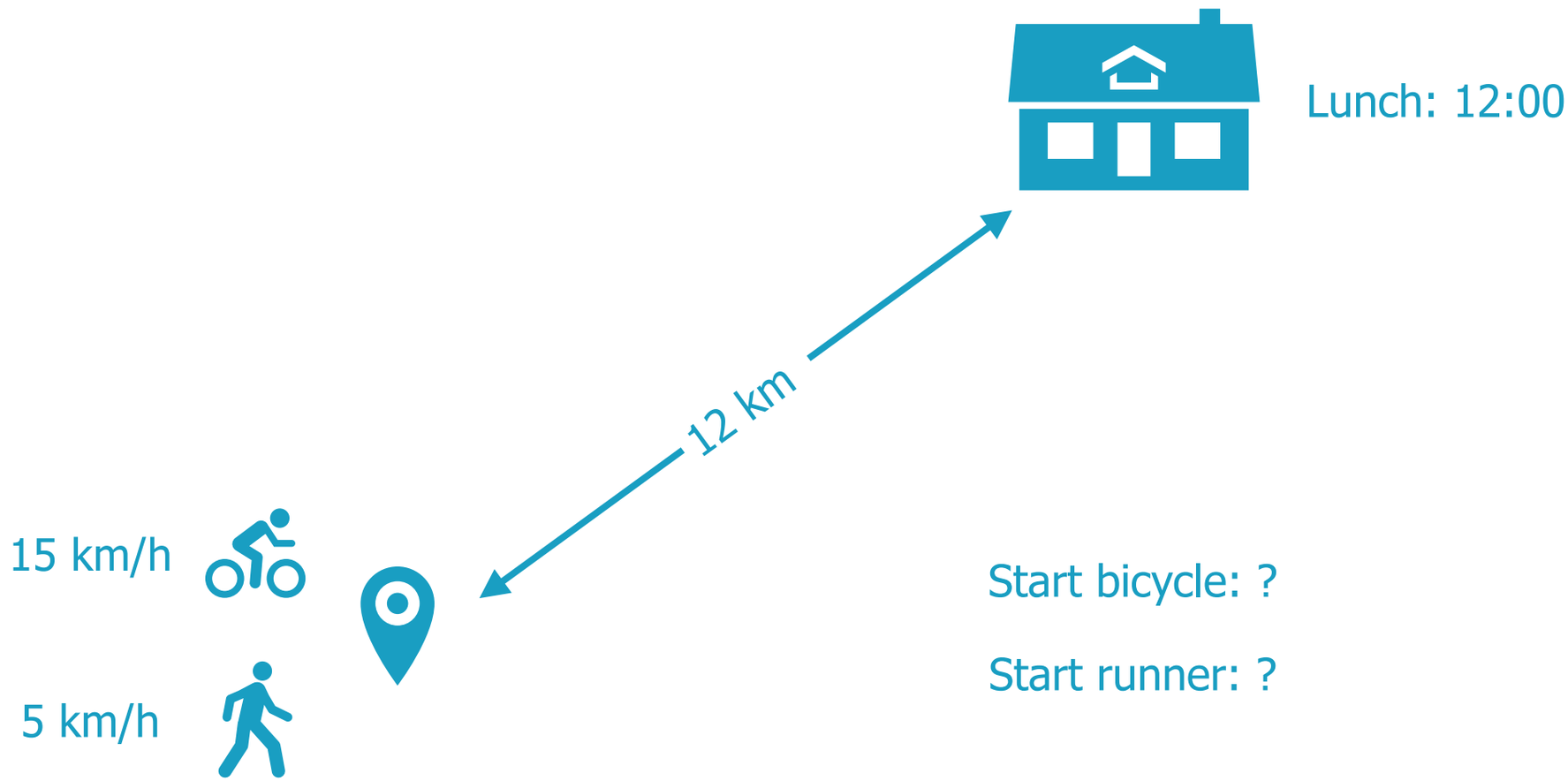


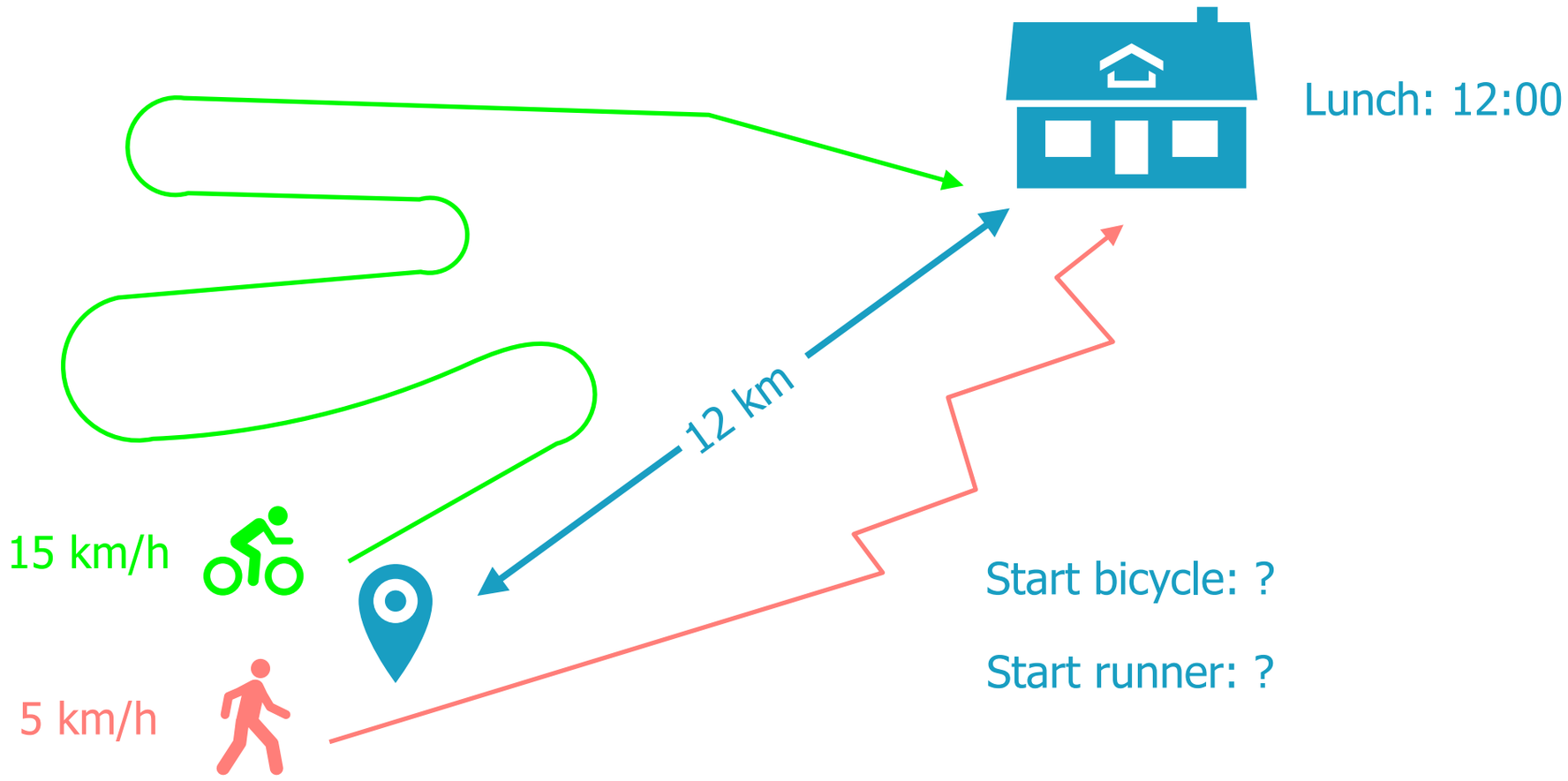
P > 0.05

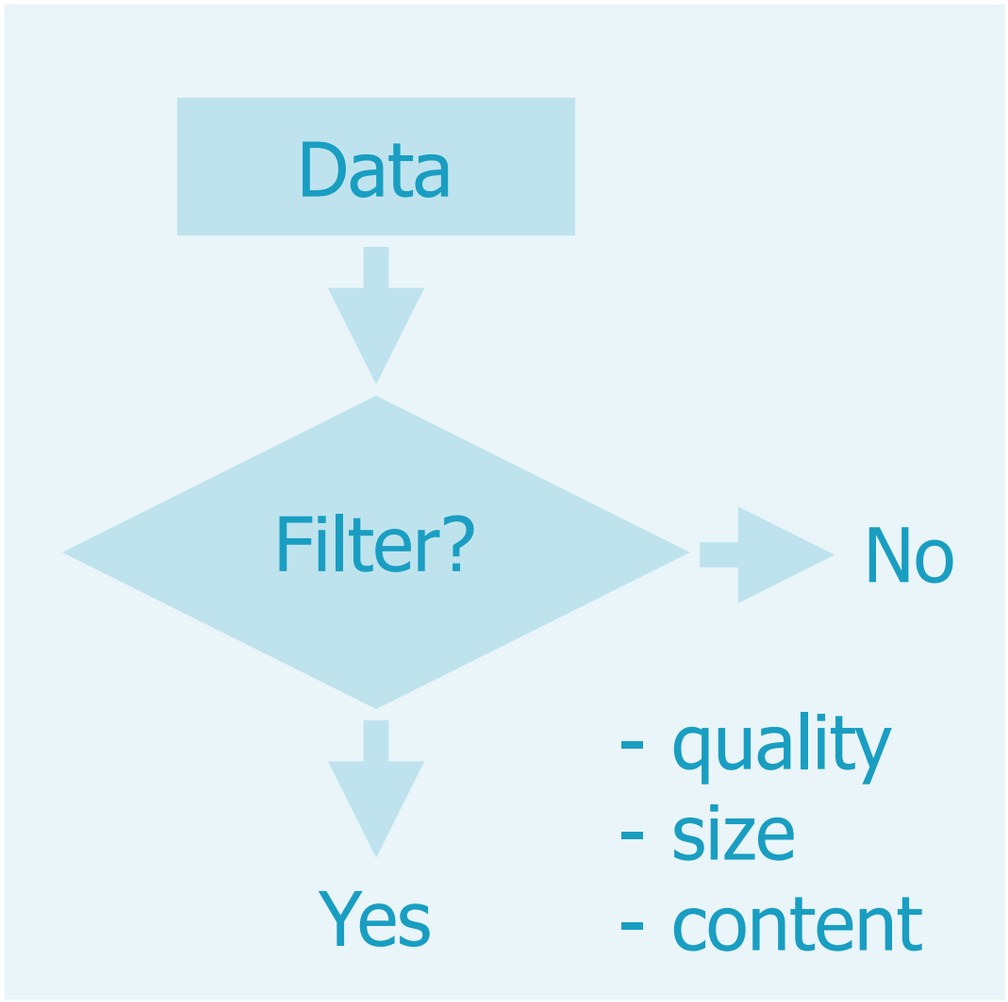
P < 0.05

0 (zero effect)


Understanding a **question** is half an answer.
Sokrates

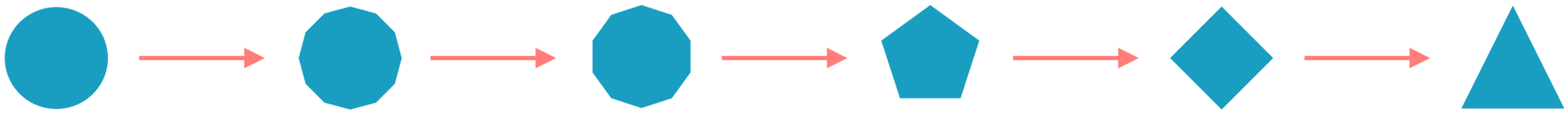

Lunch: 12:00
12 km
15 km/h
5 km/h
Start bicycle: ?
Start runner: ?


Lunch: 12:00
12 km
15 km/h
5 km/h
Start bicycle: ?
Start runner: ?


Data
Filter?
No
Yes
- quality
- size
- content

Zurich Genetic Diversity Centre

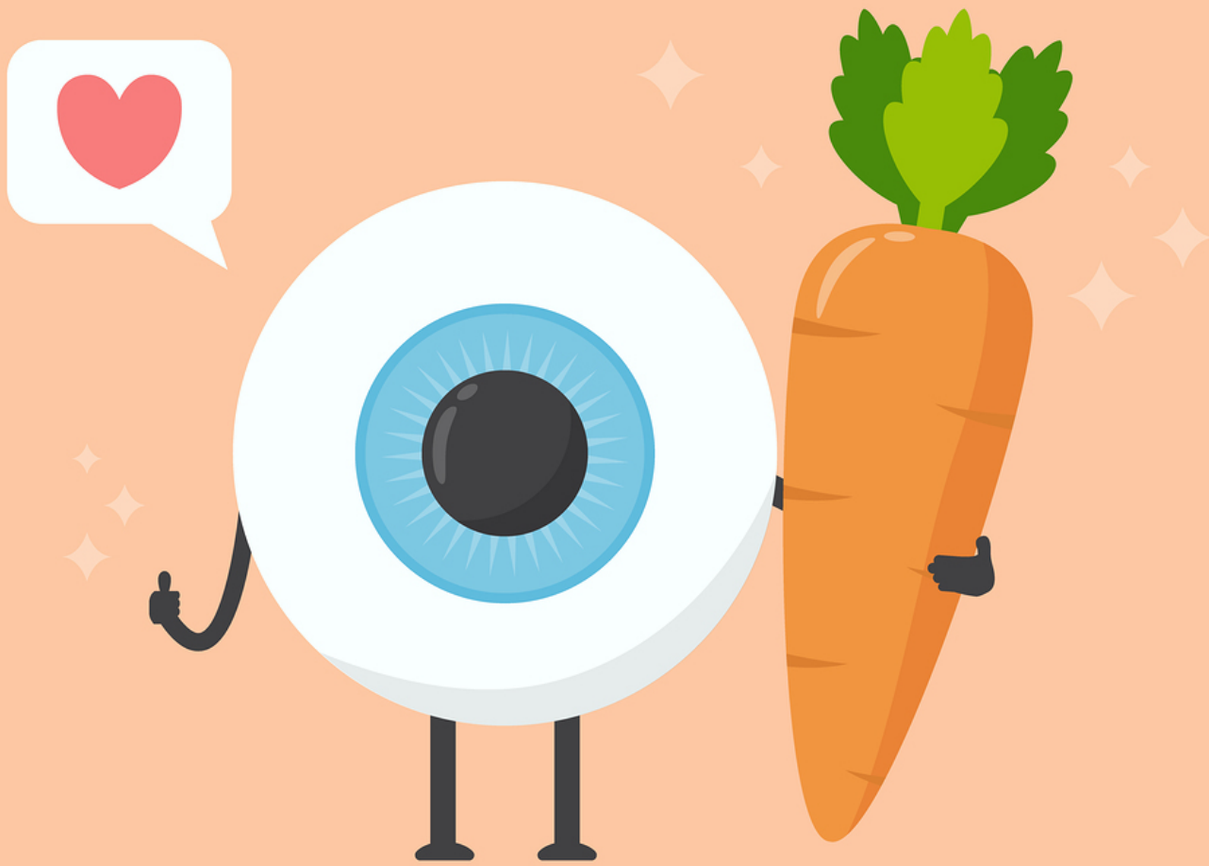
Simplicity



Speed

Genetic Diversity Centre Zurich

Urban legends



The popular belief that carrots improve eyesight, especially night vision, largely stems from British propaganda during World War II. To conceal the use of radar technology in detecting German bombers, the Royal Air Force attributed their pilots' accuracy at night to a high intake of carrots. The government promoted this story widely—through newspapers and posters—not only to protect military secrets but also to encourage citizens to eat more locally grown vegetables during food rationing. While carrots are rich in beta-carotene, a precursor to vitamin A (which is essential for healthy vision), eating them doesn't enhance sight beyond normal levels, nor does it grant night vision.

The idea that carrots improve night vision is a classic World War II myth, originally spread as British propaganda to disguise radar technology. The claim that chewing gum stays in your stomach for seven years? That's an urban legend. Then there's the persistent belief that shaving makes hair grow back thicker — a common misconception with no scientific basis. Similarly, the warning that going outside with wet hair will give you a cold is just an old wives' tale. And no, goldfish don't have a three-second memory — that one's another widely repeated but false idea.