



UNIVERSITÄT  
HOHENHEIM

Coursework 1: Presentation

# DATA VISUALIZATION: CALLING BULLSHIT

Unveiling the Core Principles: Comprehensive Insights into Data Visualization



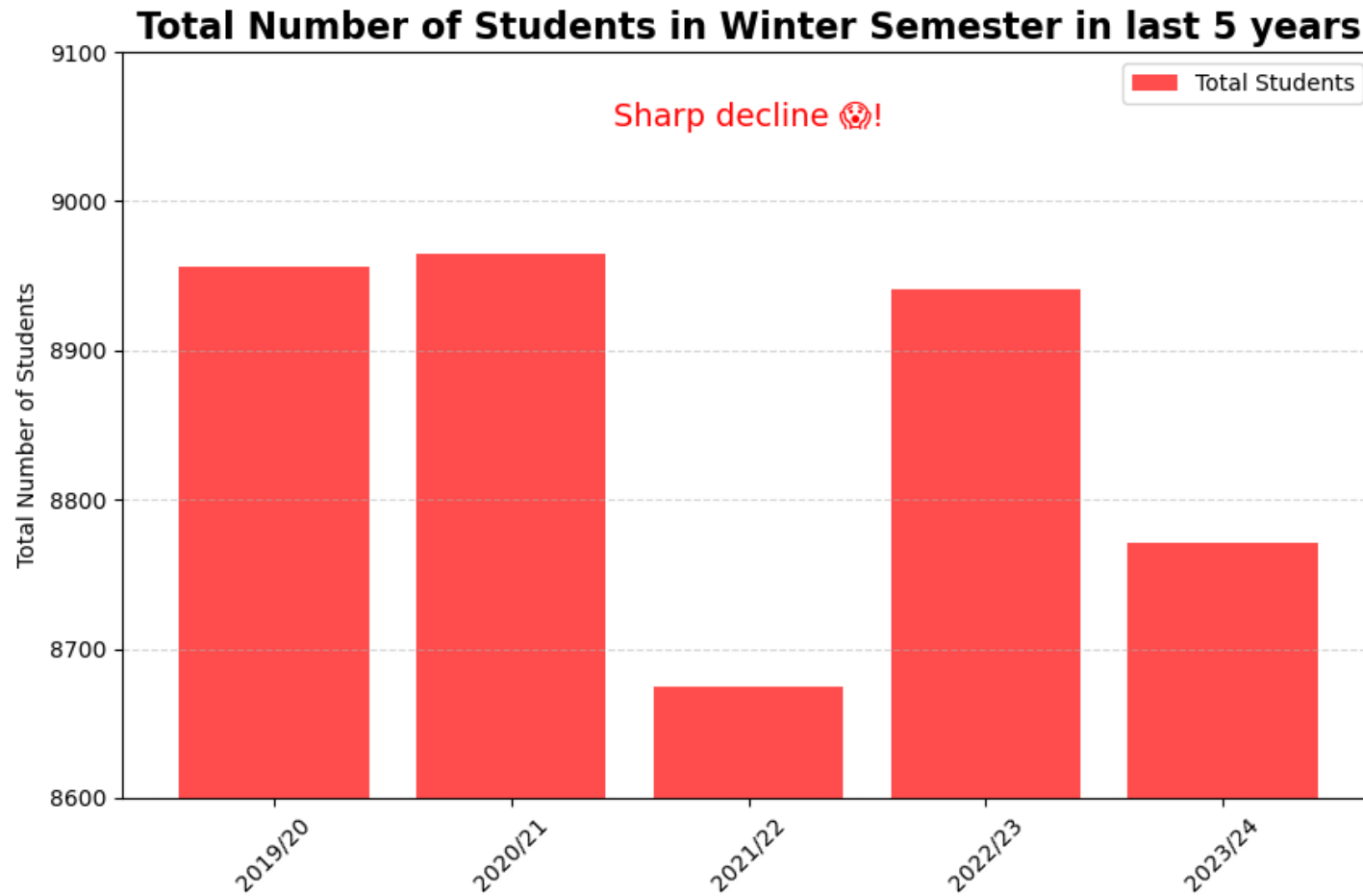
*“A picture is worth 10,000 words.”*

Anonymous

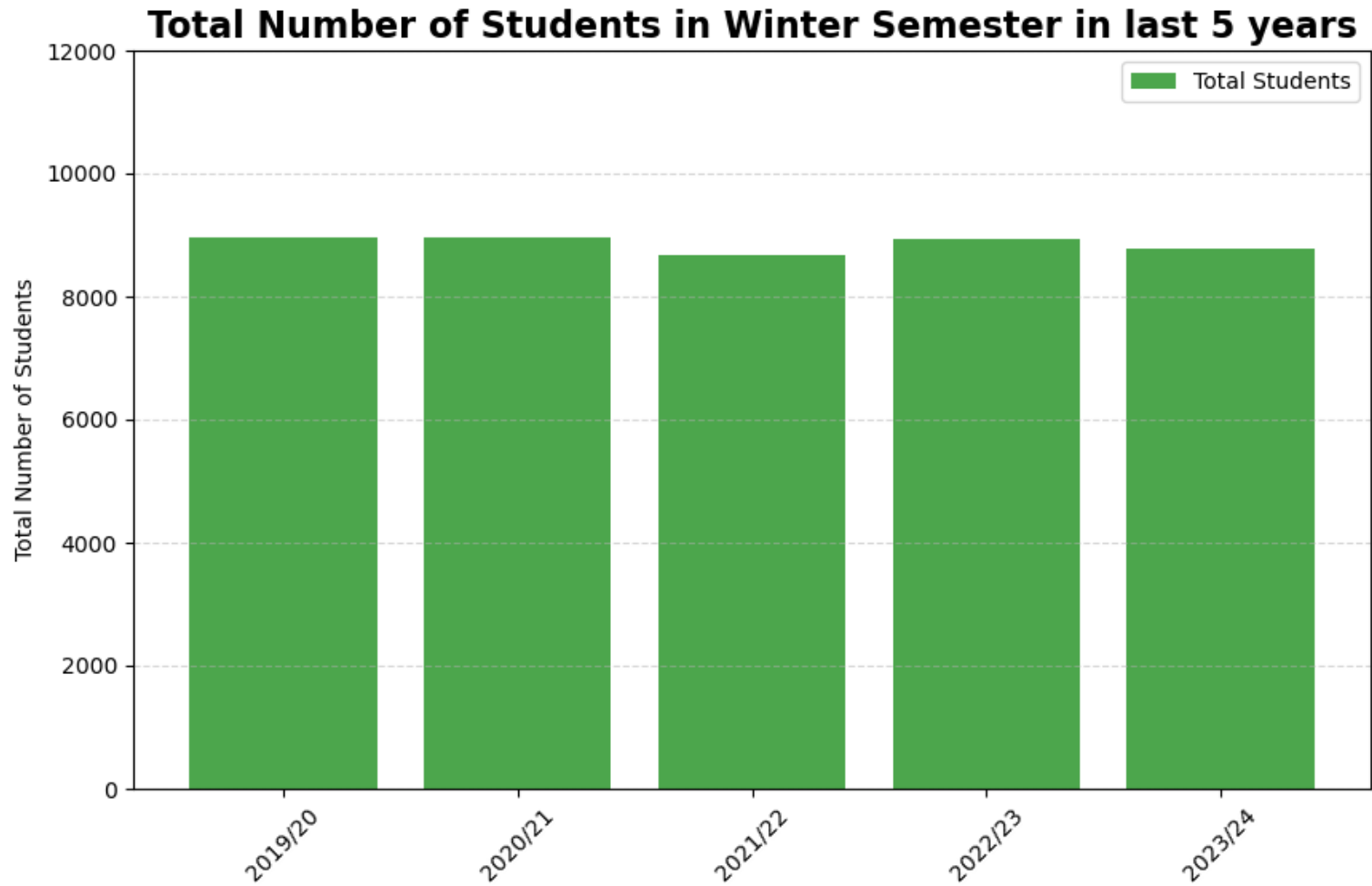
“But what if that picture is designed  
to deceive you?.”



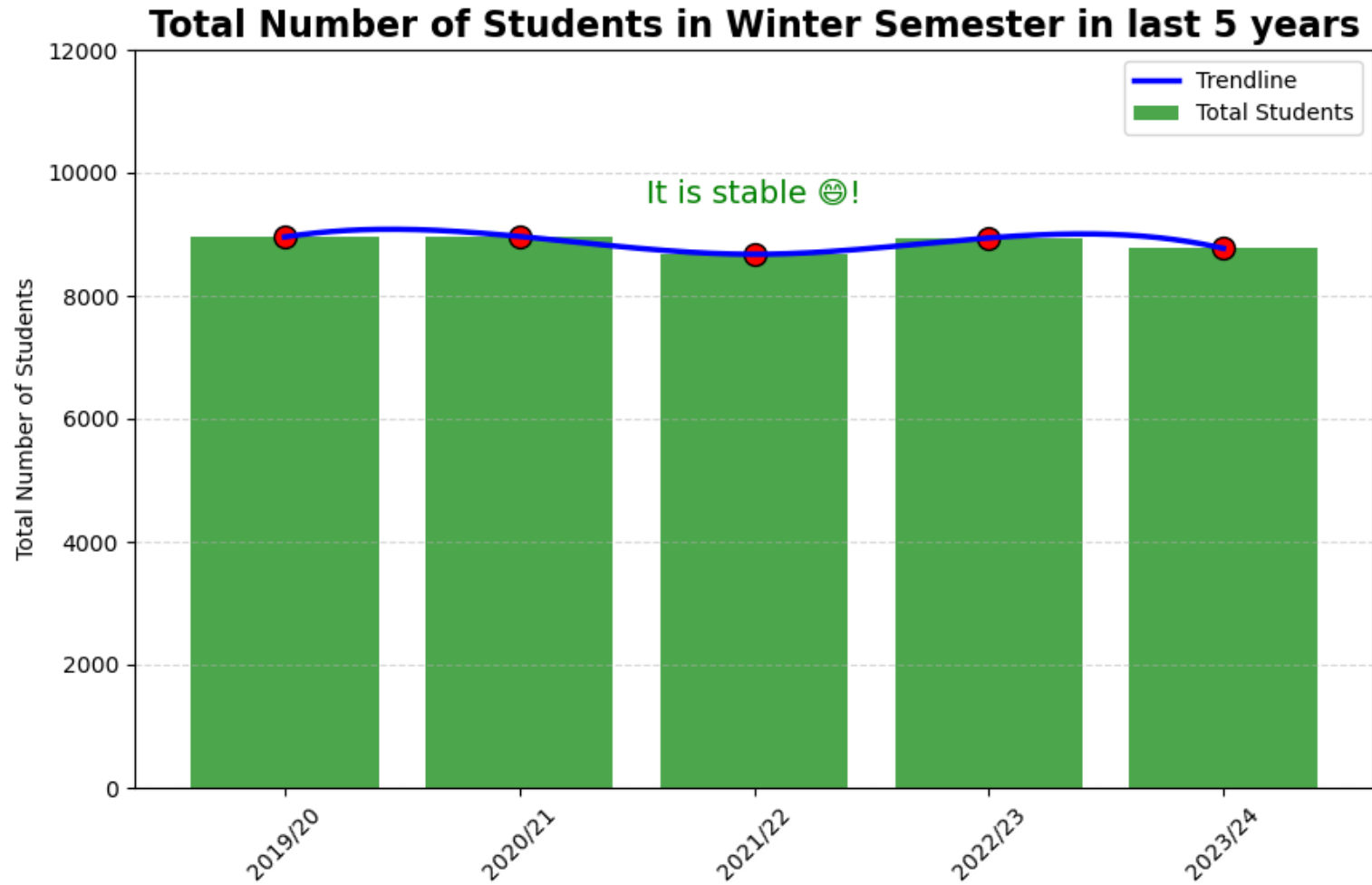
# IS THERE A SHARP DECLINE IN NUMBER OF STUDENTS IN HOHENHEIM?



# NO! IT'S STABLE!

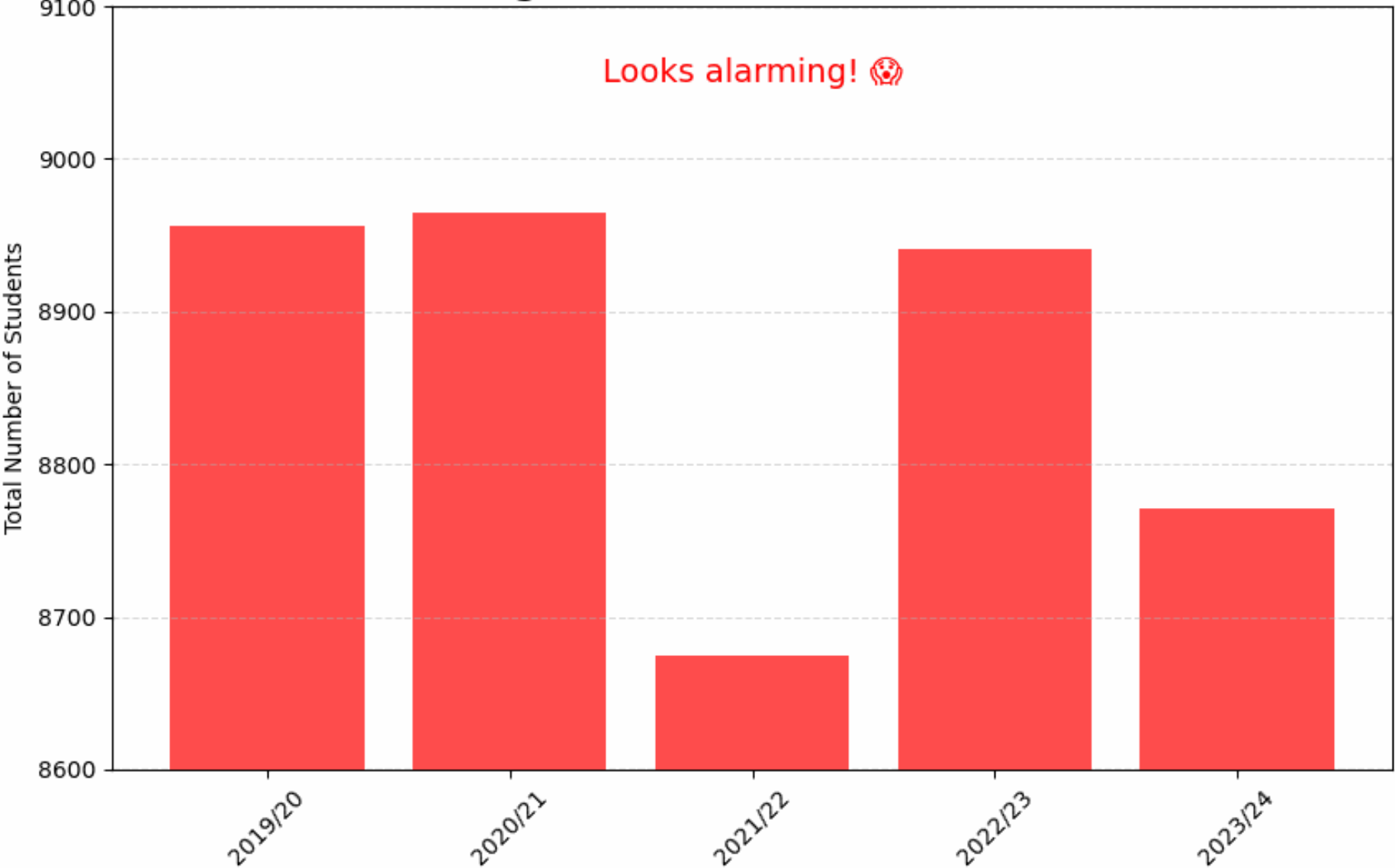


# SEE THE TREND!

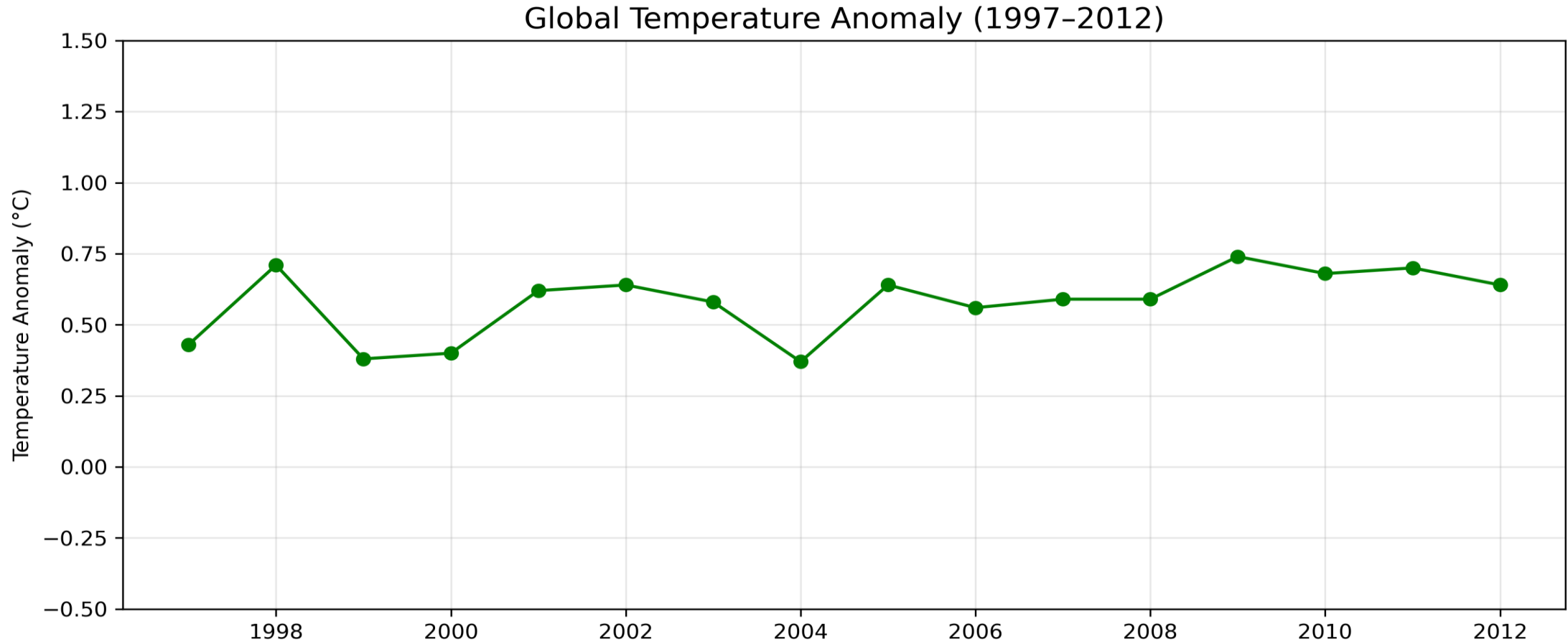




# Misleading Visualization: Truncated Y-Axis



# THERE IS NO SUCH THING LIKE CLIMATE CHANGE!

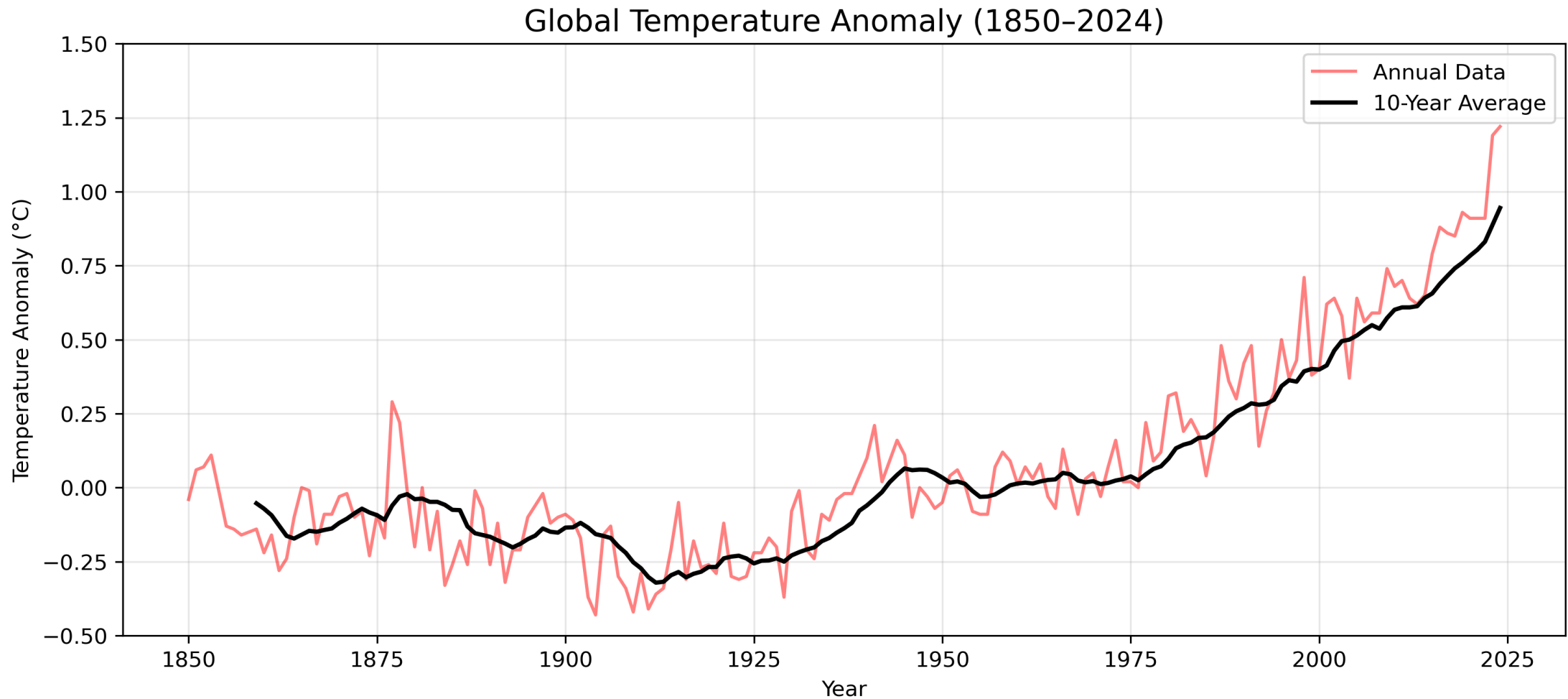


Data API Source:<sup>2</sup>





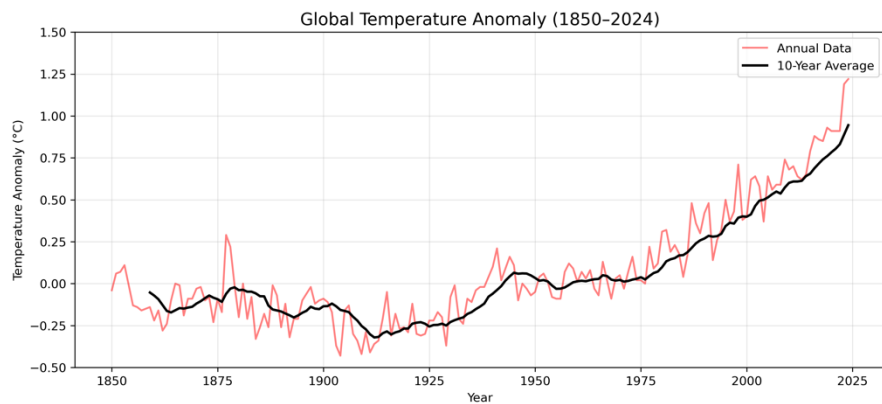
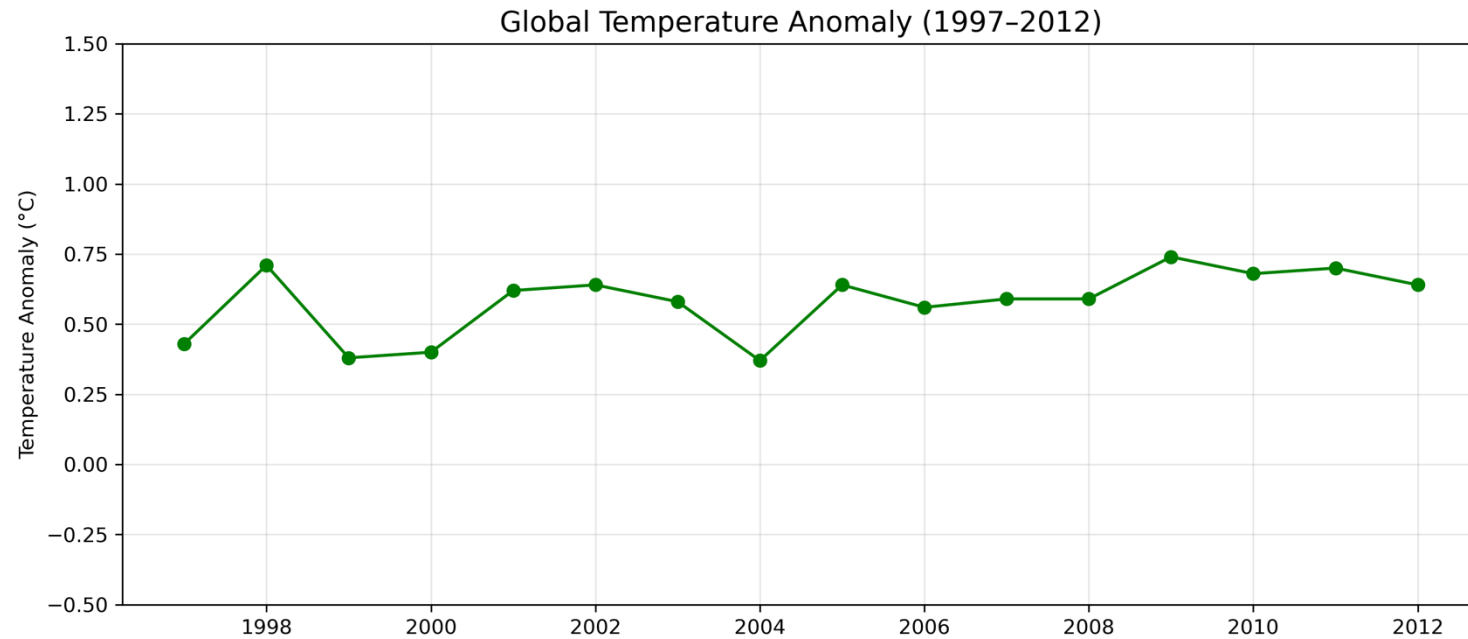
# CLIMATE CHANGE IS REAL!



Data API Source:<sup>2</sup>



# CHERRY PICKING



**My argument**  
No Global Warming...!

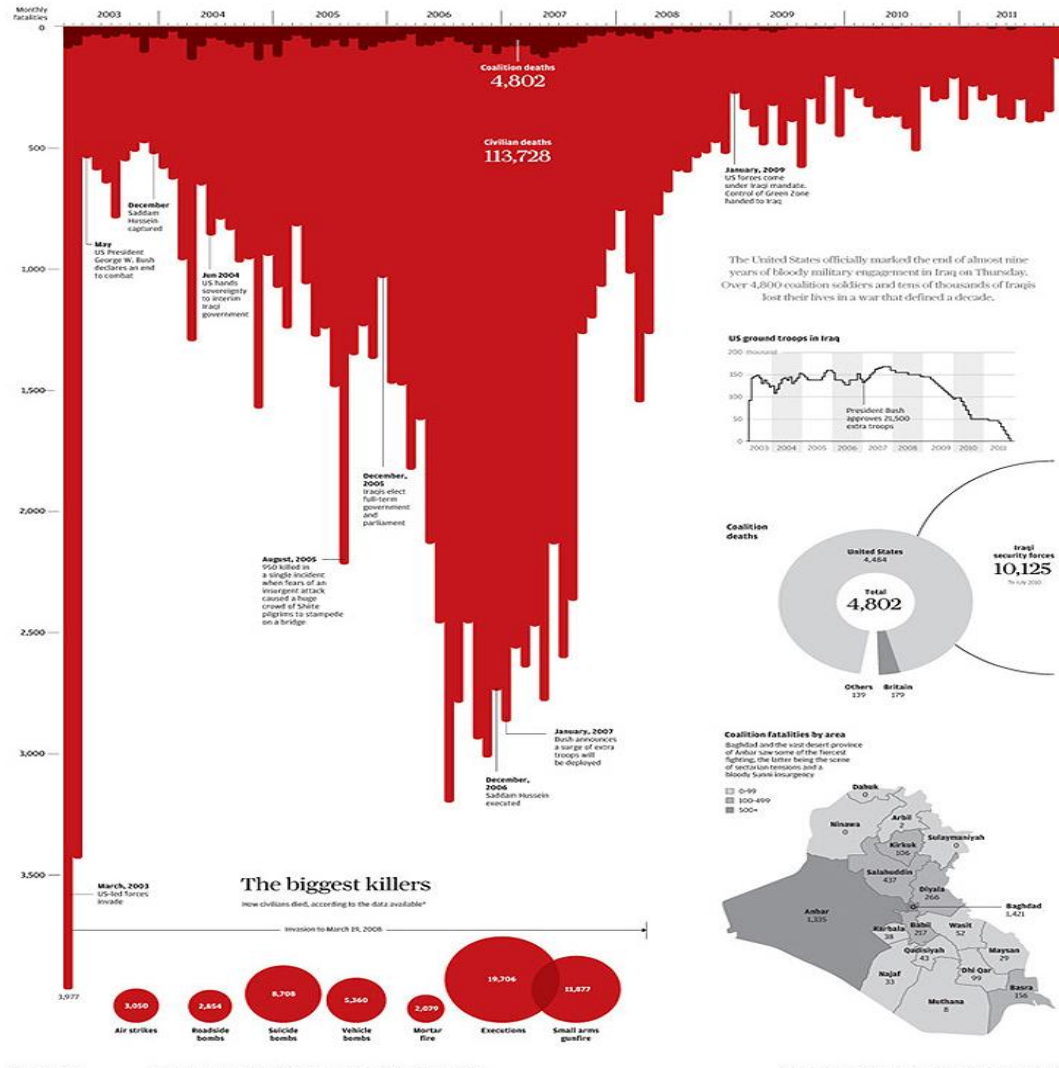


**I need only this**  
range of data to  
prove it!



Illustration by Nadim Khan  
Designed using Canva

# VISUAL METAPHOR OF BLOOD



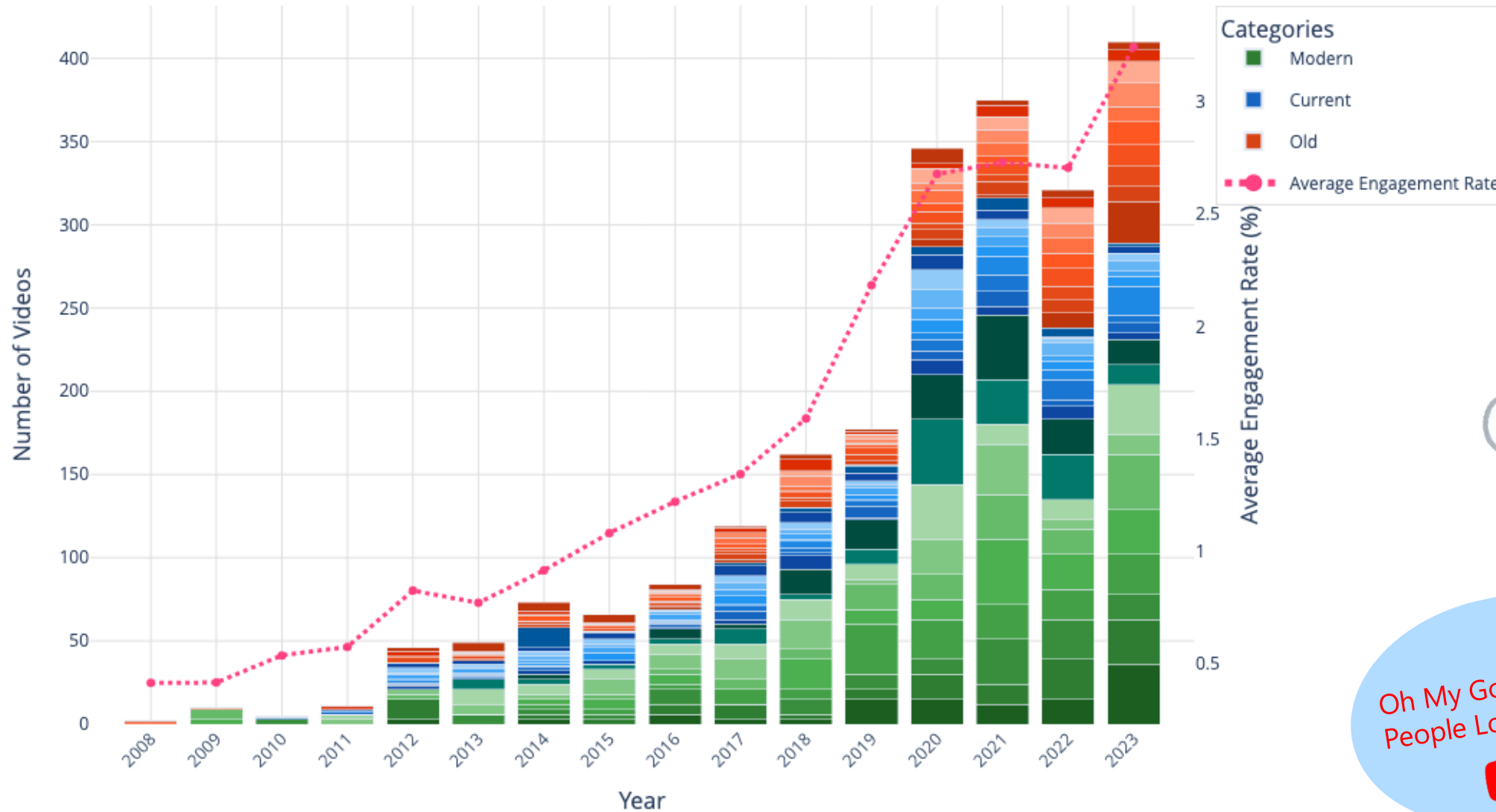
- Casualties from Iraq War from 2003 to 2013.
- Deliberate design choice.
- Dripping Blood.
- Striking visual to draw reader's attention.
- Representation works well by using references.

\*AFP graphics, Simon Scarr

\*Specific from information sources are not listed. Count of deaths according to the March 28, 2008 invasion to March 28, 2008

\*Source: Reuters, Associated Press, New York Times, Washington Post, Wall Street Journal, New York Times, Washington Post, Wall Street Journal

# Evolution of Plant Breeding Content on YouTube (2008-2023)

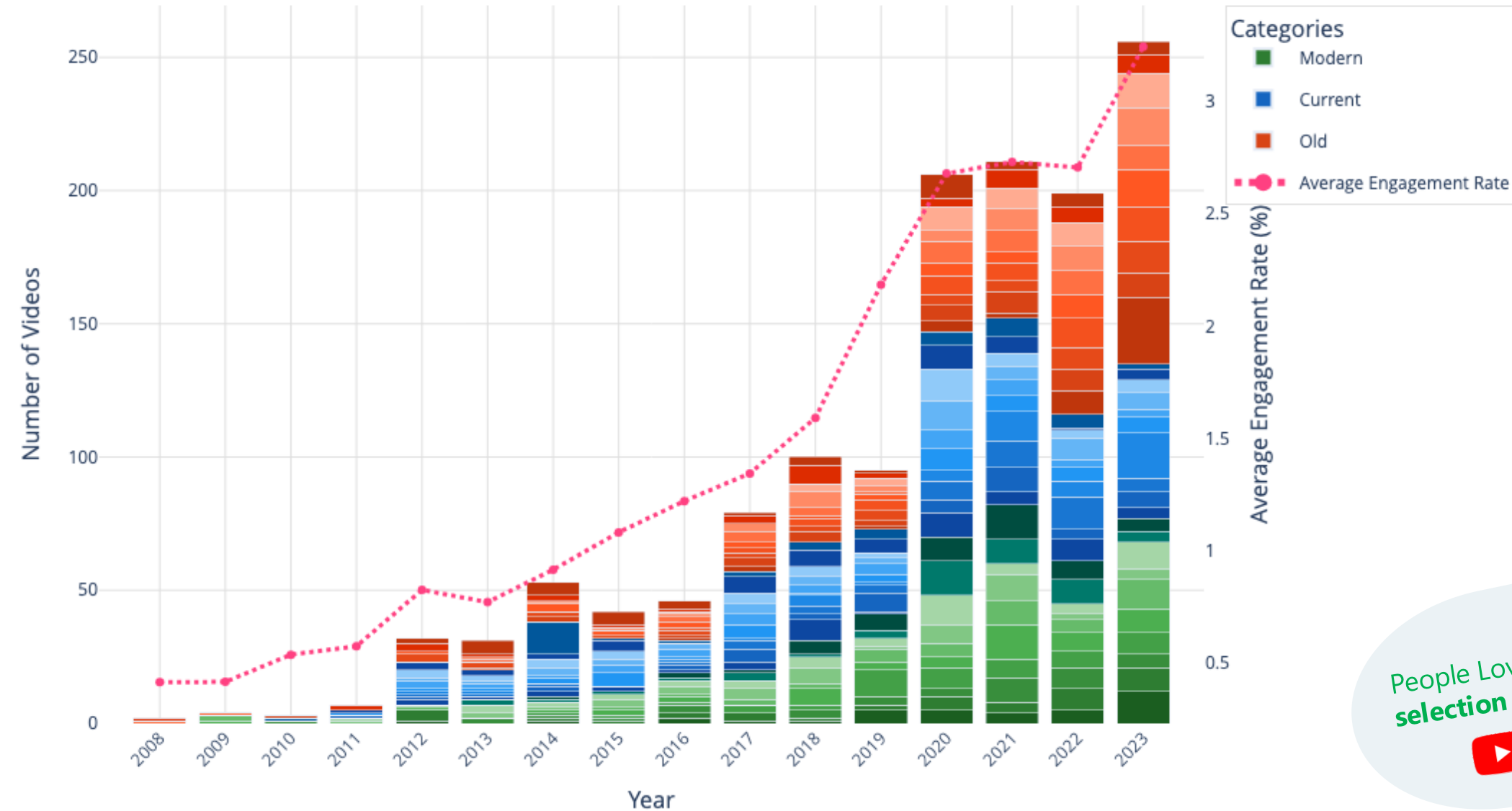


Oh My God!  
People Love **CRISPR!**



Data Source: YouTube API

# Evolution of Plant Breeding Content on YouTube (2008-2023)



People Love **Mass**  
selection also!

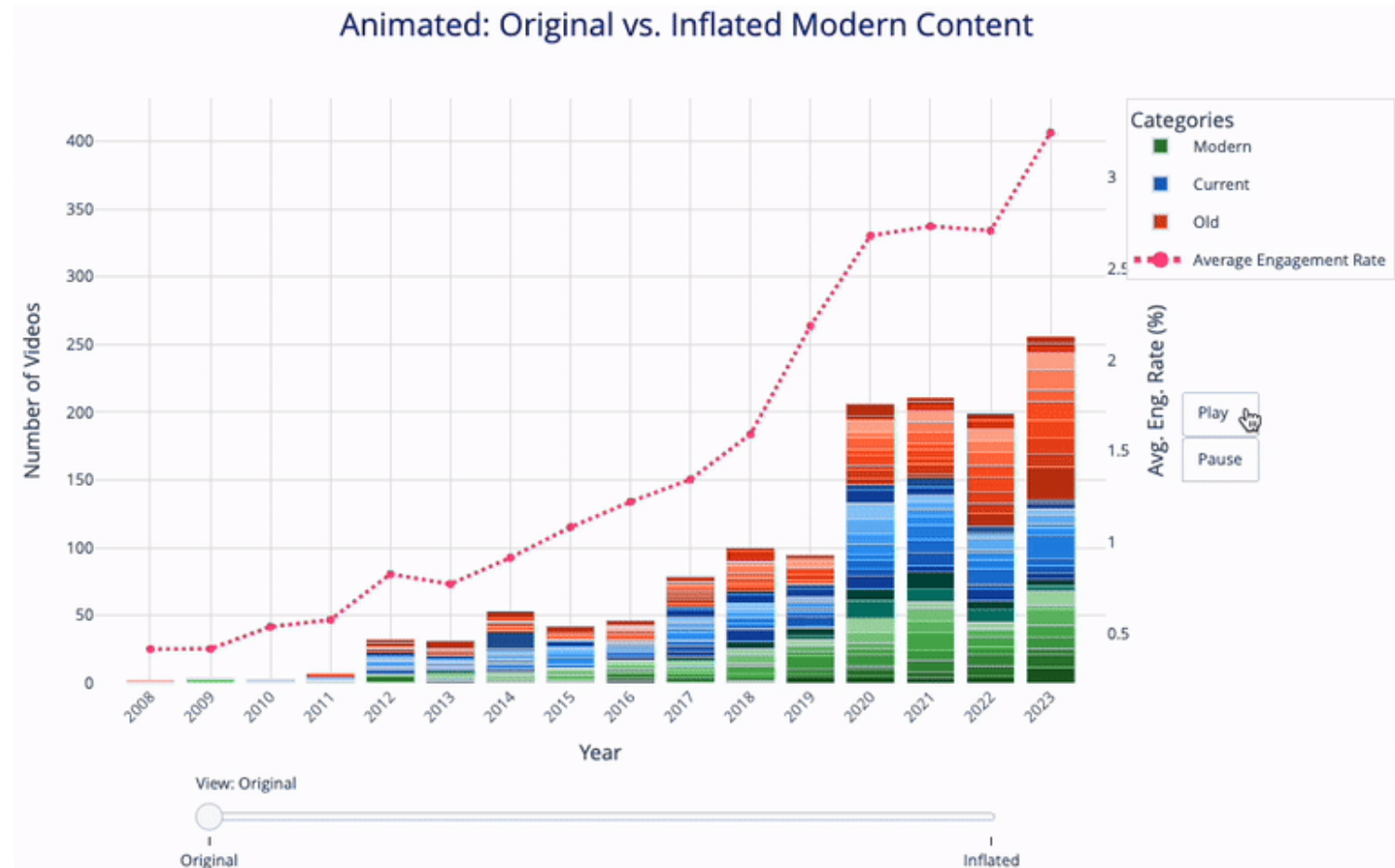
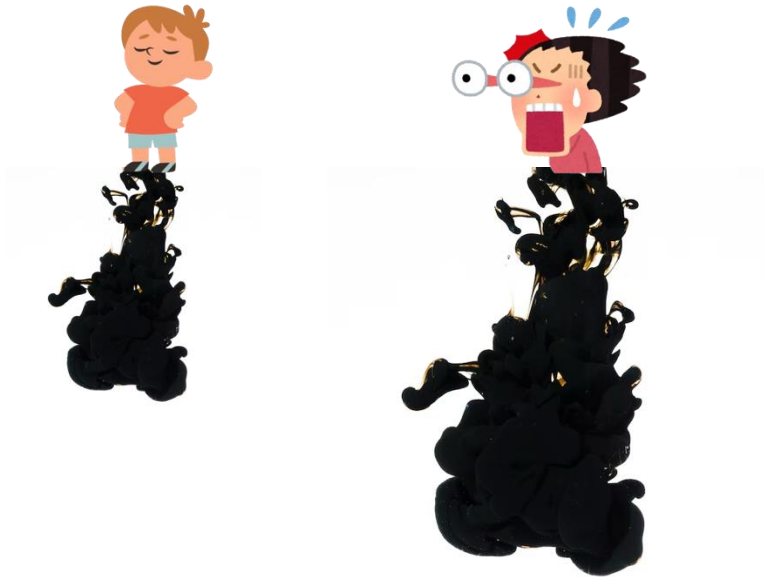


Designed using Canva



# PRINCIPLE OF PROPORTIONAL INK

- *A good designer can deceive you.*
- Tufte's principle of proportional ink.
- *Magnitude  $\neq$  Proportional Size*



Data Source: YouTube API

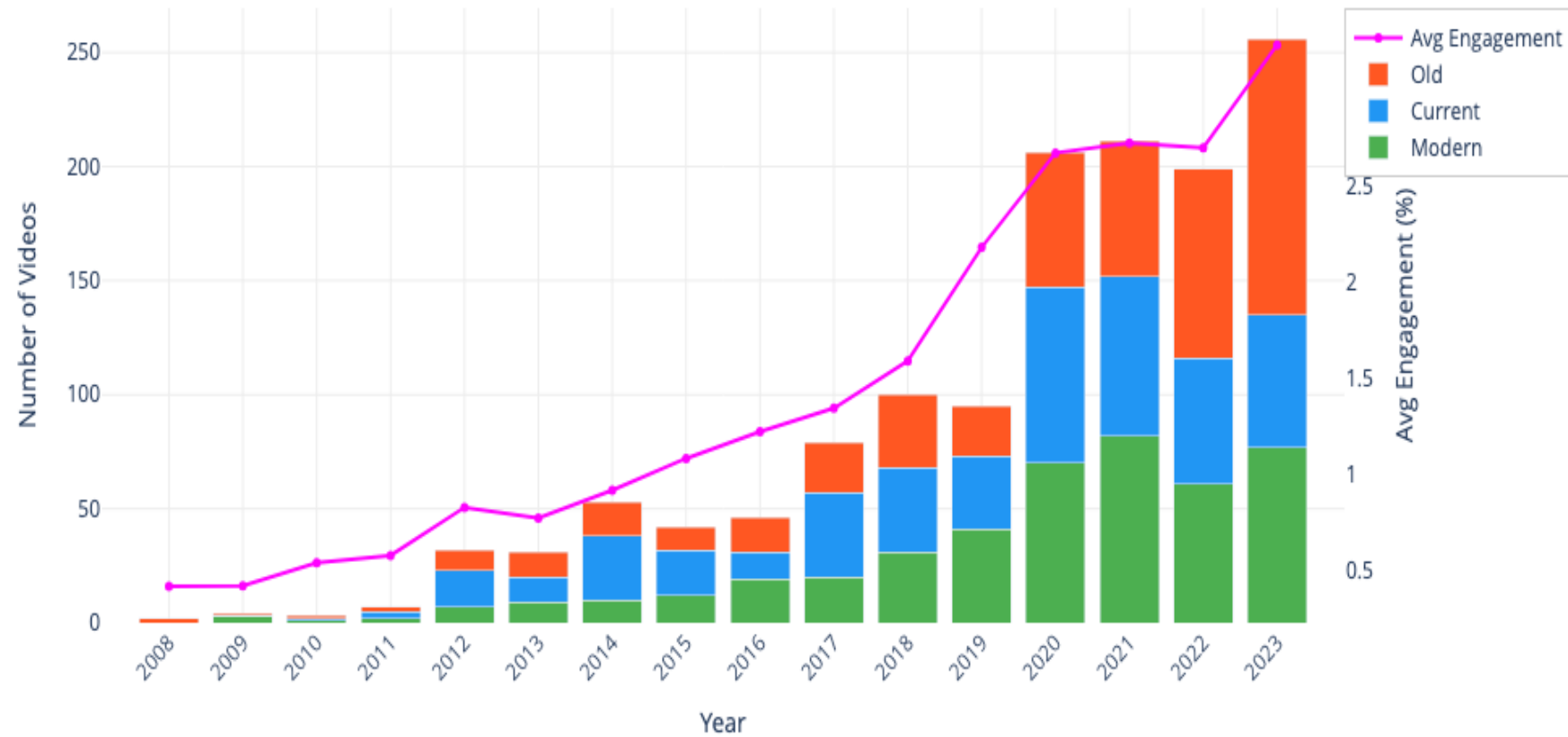


# DATA TO INK RATIO

- *Data > Ink*
- *Easy to understand*



With Higher Data-Ink Ratio



Data Source: YouTube API

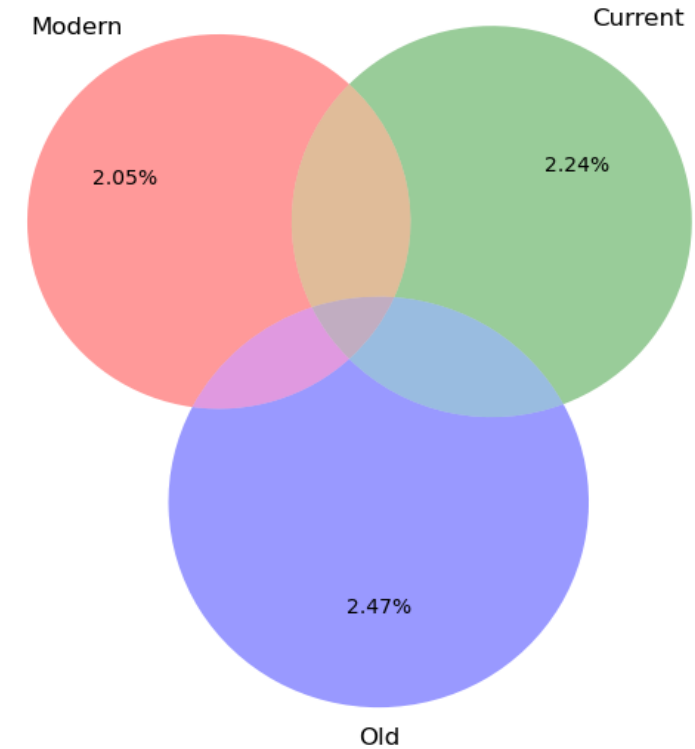


# GLASS SLIPPER

Venn Diagram showing Average User Engagement in Each Category of Plant Breeding Concepts



Image © 2025: Purchased from <https://www.deviantart.com/>



Data Source: YouTube API

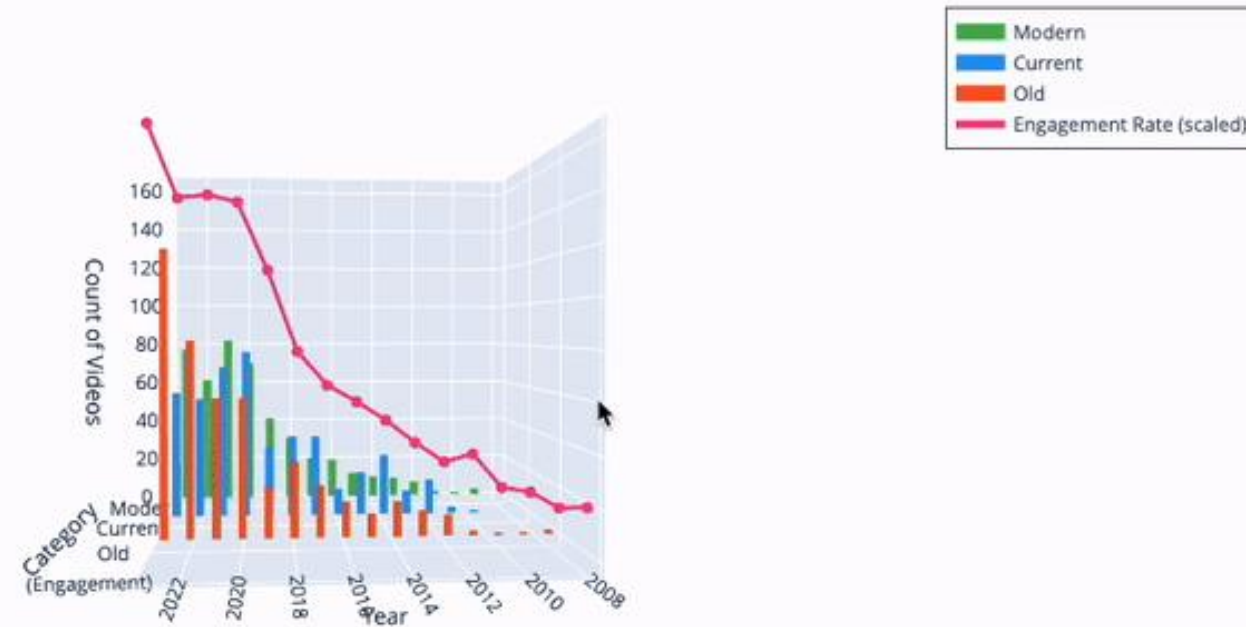




# WHY 3D A BAD IDEA

- *Impressive!*
- *Difficult to correlate axis values*
- *Difficult to assess the relative size*
- *Magnitude  $\neq$  Proportional Size*  
(Principle of proportional ink)

3D Visualization: Plant Breeding Content (2008–2023)



Data Source: YouTube API

# WHAT IF THERE ARE MORE THAN TWO VARIABLES?

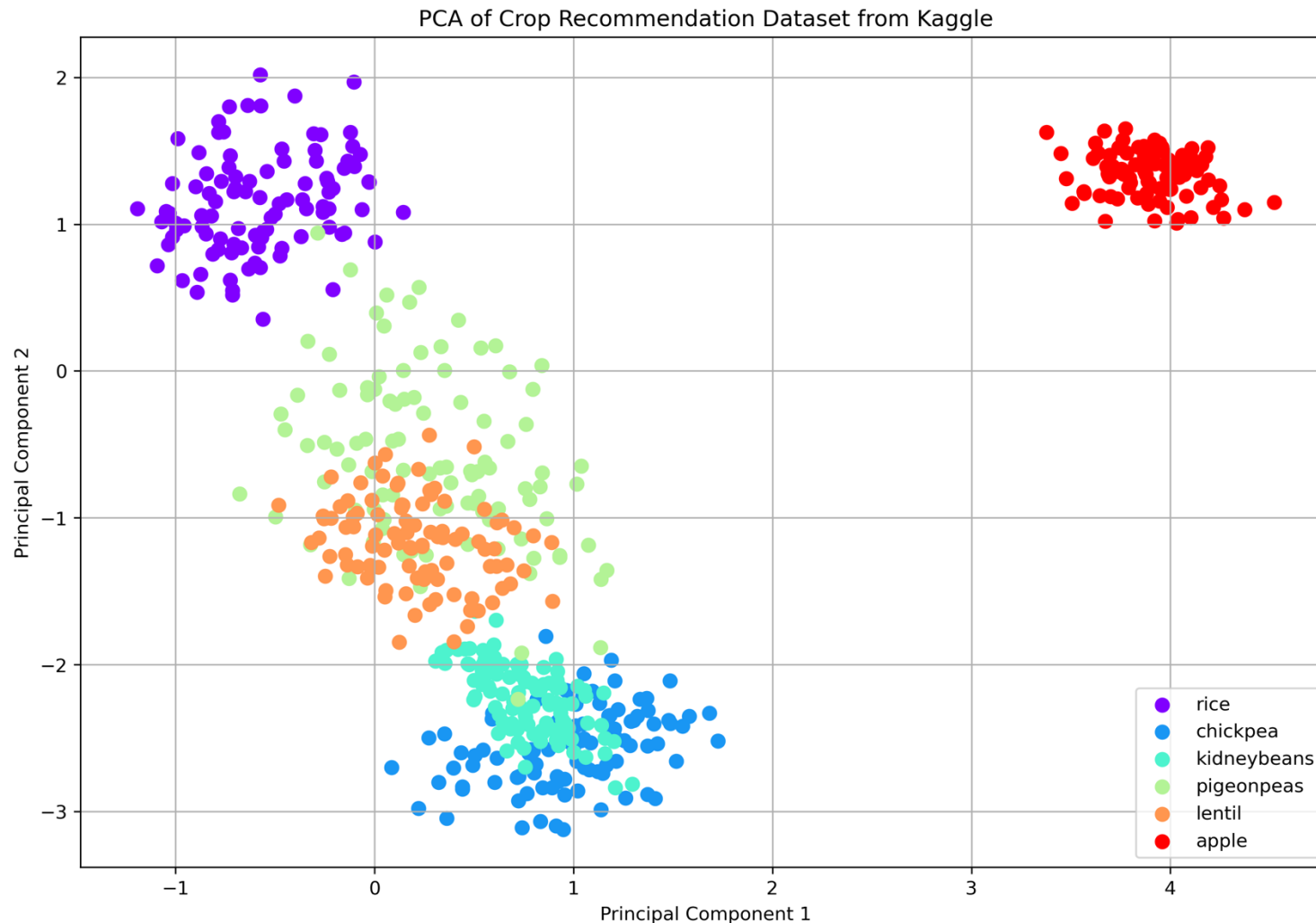
N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.87974371	82.00274423	6.502985292000000	202.9355362	rice
85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice
69	37	42	23.05804872	83.37011772	7.073453503	251.0549998	rice
69	55	38	22.70883798	82.63941394	5.70080568	271.3248604	rice
94	53	40	20.27774362	82.89408619	5.718627178000000	241.9741949	rice
89	54	38	24.51588066	83.53521630000000	6.685346424	230.4462359	rice
68	58	38	23.22397386	83.03322691	6.336253525	221.2091958	rice
91	53	40	26.52723513	81.41753846	5.386167788	264.6148697	rice
90	46	42	23.97898217	81.45061596	7.50283396	250.0832336	rice
78	58	44	26.80079604	80.88684822	5.108681786	284.4364567	rice
93	56	36	24.01497622	82.05687182	6.98435366	185.2773389	rice
94	50	37	25.66585205	80.66385045	6.94801983	209.5869708	rice
60	48	39	24.28209415	80.30025587	7.042299069000000	231.0863347	rice
85	38	41	21.58711777	82.7883708	6.249050656000000	276.6552459000000	rice
91	35	39	23.79391957	80.41817957	6.970859754	206.2611855	rice
77	38	36	21.8652524	80.1923008	5.953933276	224.5550169000000	rice
88	35	40	23.57943626	83.58760316	5.85393208	291.2986618000000	rice
89	45	36	21.32504158	80.47476396	6.442475375	185.4974732	rice
76	40	43	25.15745531	83.11713476	5.070175667	231.3843163	rice
67	59	41	21.94766735	80.97384195	6.012632591	213.3560921	rice
83	41	43	21.0525355	82.67839517	6.254028451	233.1075816	rice

- 2200 rows : 7 variables → Crops
- Raw Data = **No meaning**



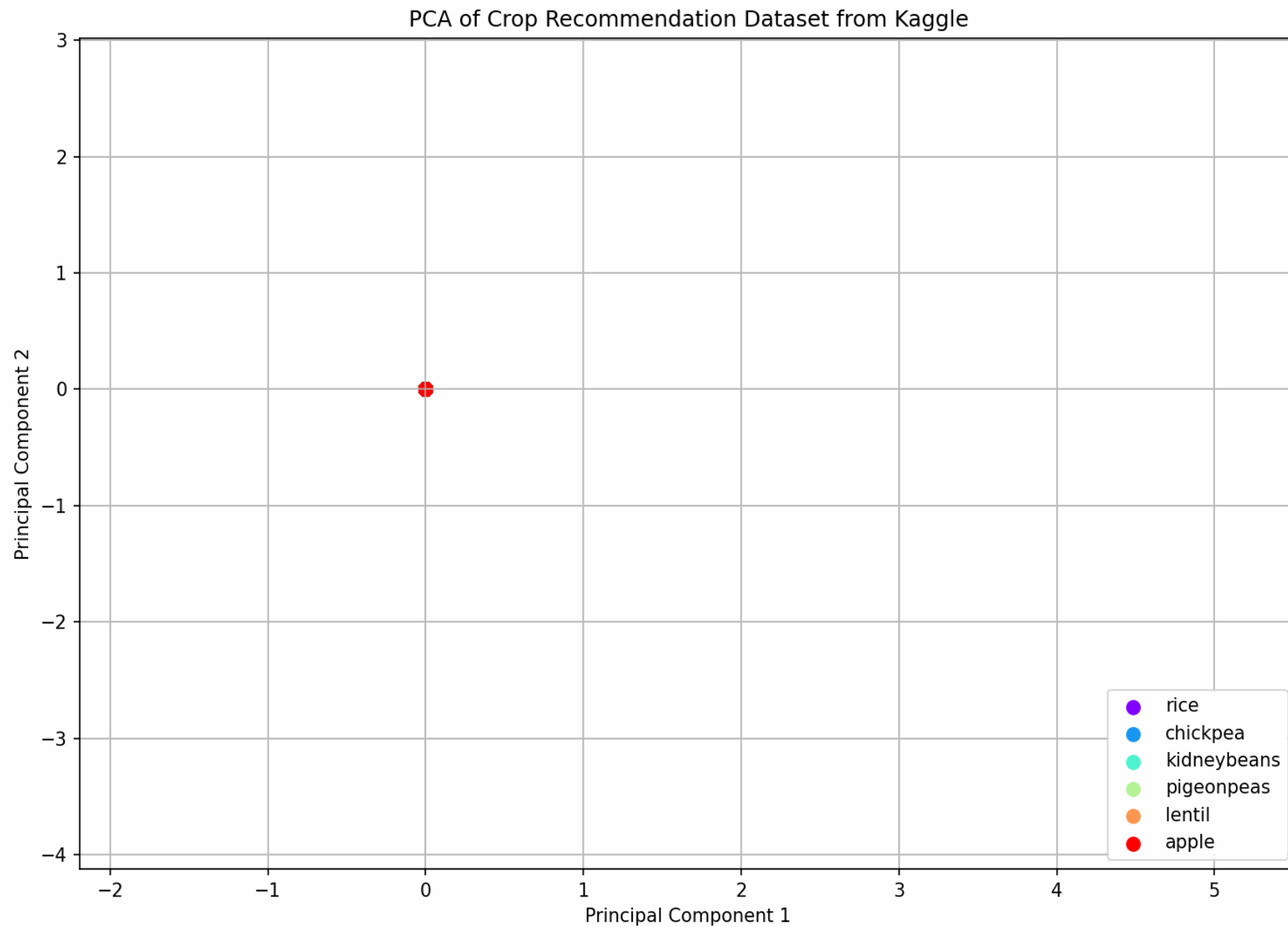
Data Source: Crop Recommendation Dataset from Kaggle<sup>4</sup>

# PCA



- *Principle Component Analysis (PCA)*
- Find meaningful pattern!
- **Outliers or Separation?**
- Also proves **why data visualization?**

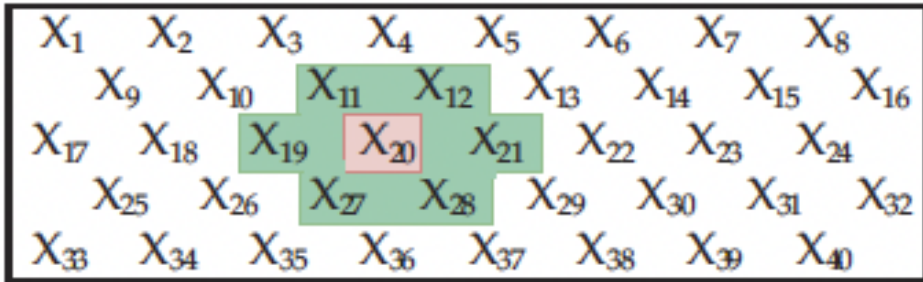
Data Source: Crop Recommendation Dataset from Kaggle<sup>4</sup>

Data Source: Crop Recommendation Dataset from Kaggle<sup>4</sup>

# REAL PROBLEM?

- *How moving grid (honeycomb) filed design works?*

$$X_{20} = X - (X_{11} + X_{19} + X_{27} + X_{28} + X_{21} + X_{12}) / 6$$



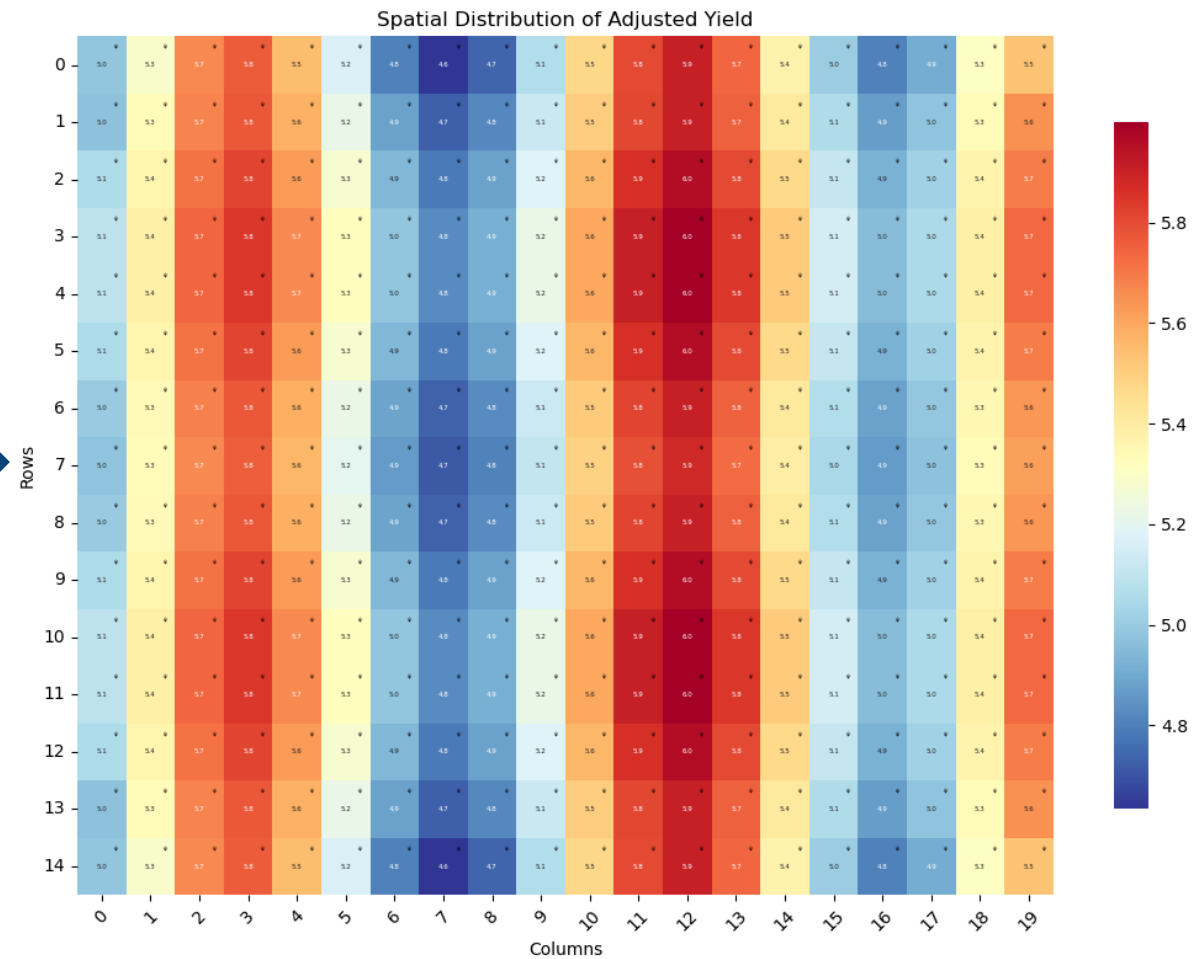
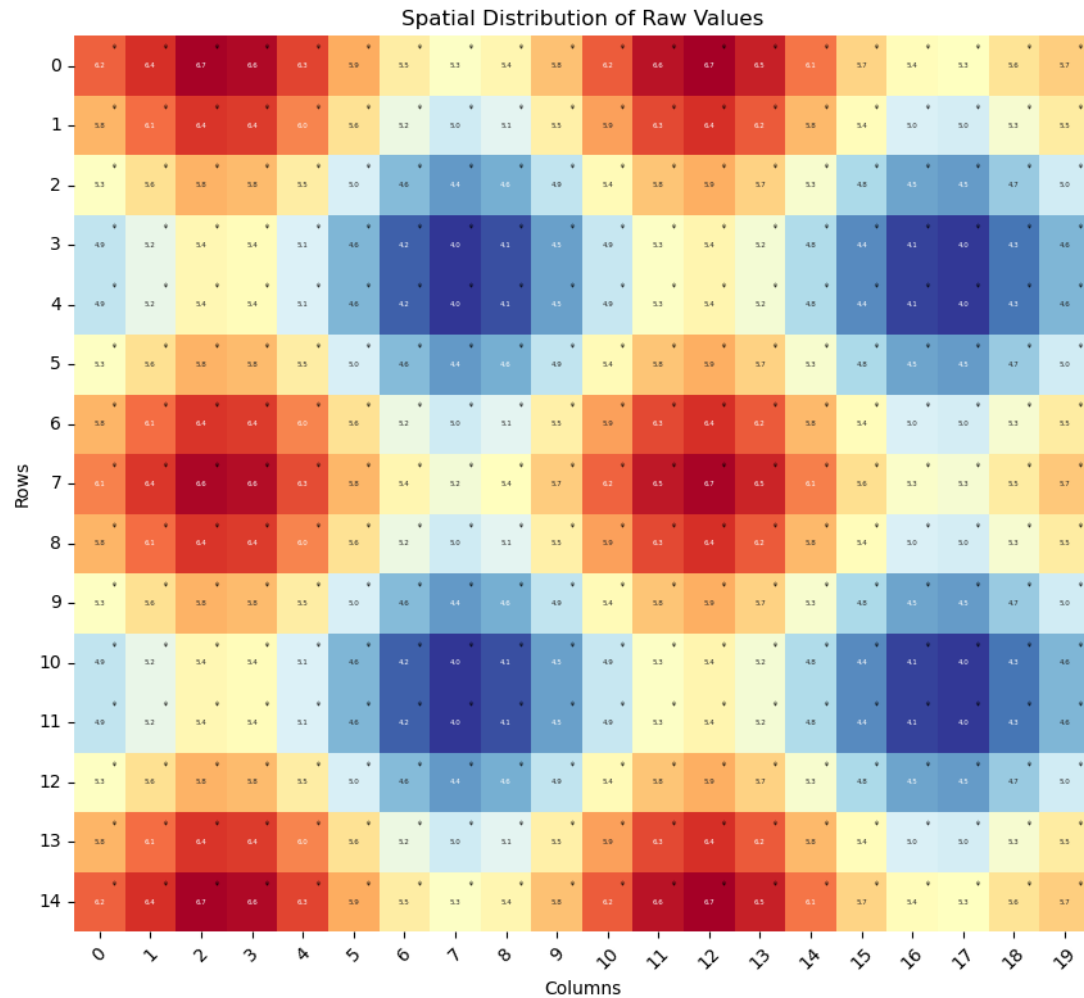
Representation from the slides of Tobias Schrag, *Selection Theory*

PlabStat

Original vs Adjusted Values:				
Row	Col	Original	Adjusted	Difference
1	1	6.20	4.98	-1.21
1	2	6.42	5.29	-1.13
1	3	6.67	5.67	-1.00
1	4	6.63	5.76	-0.87
1	5	6.33	5.55	-0.78
1	6	5.89	5.16	-0.72
1	7	5.49	4.80	-0.69
1	8	5.31	4.63	-0.68
1	9	5.42	4.74	-0.69
1	10	5.78	5.06	-0.71

Gives results very fast.  
But no visualization.

# SOLUTION: dgNova\*



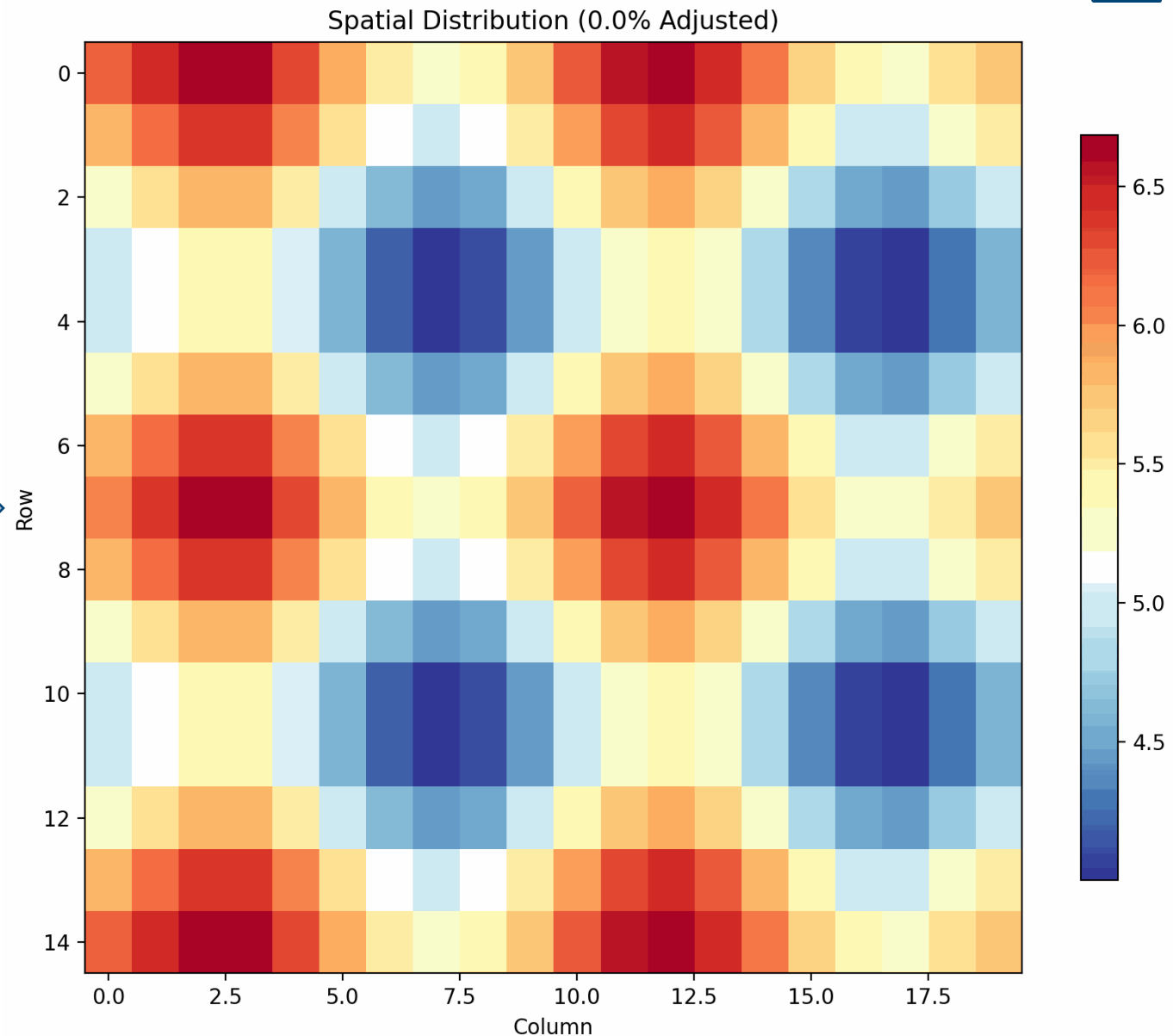
\*I created a python library to simulate unreplicated field designs.<sup>5</sup>



# SOLUTION: dgNova

```
# Simulate a field experiment
unrep_sim = UNREP(
    row=15,           # field dimensions:row
    column=20,        # field dimension:column
    heterogeneity=0.8, # spatial trend intensity (0-1)
    mean=5.3,         # base yield level
    sd=0,             # random variation
    ne=1,             # neighbor effects (0-1)
    design='moving_grid' #type of field design
)

# Analyze simulated data
unrep_sim.analyze()
```



*“A simulation is worth 10,000 images?.”*



# KEY POINTS

*To create scientific visualization or to draw the right inferences, you need the following techniques:*

- *Define axis with a constant scale.*
- *Define references for clarity.*
- **Principle of proportional ink.**
- *Keep things in 2D.*
- Higher **Data – to – Ink Ratio**.
- Say no to "**Glass – Slippers**".
- For more variables, **PCA** is the key.
- **Simulations** can further unveil the potential of data visualization.

# REFERENCES

- Universität Hohenheim (2024). **Jahresbericht 2023**. Retrieved from <https://www.uni-hohenheim.de> (1)
- <https://www.ncei.noaa.gov/support/access-data-service-api-user-documentation> (2)
- South China Morning Post. Design by **Simon Scarr**. <https://www.simonscarr.com/iraqs-bloody-toll> (3)
- Ingle, A. (2021). **Crop Recommendation Dataset**. Kaggle. Retrieved from <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset> (4)
- <https://nadimkhan.org/project/dgnova/> (5)
- Bergstrom, C. T., & West, J. D. (2020). **Calling bullshit: The art of skepticism in a data-driven world** (Ch. 7). Random House. <https://www.callingbullshit.org/>
- Data analysis and visualizations were performed using **Python** libraries, including Pandas, NumPy, Matplotlib, Seaborn, Plotly and Pillow. <https://github.com/nfornadimkhan/presentation-on-visualisation-techniques> If you have any questions or need further assistance, please feel free to contact me at: [nfornadim@gmail.com](mailto:nfornadim@gmail.com), nadim.khan@uni-Hohenheim.de



A photograph of the University of Hohenheim building, a large, light-colored, classical-style structure with a prominent dome and a portico with columns. The building is set on a green lawn with a path leading to it. Trees with green leaves are visible in the foreground and background. A semi-transparent blue banner is overlaid across the middle of the image.

# Thankyou for Your Attention



UNIVERSITÄT  
HOHENHEIM