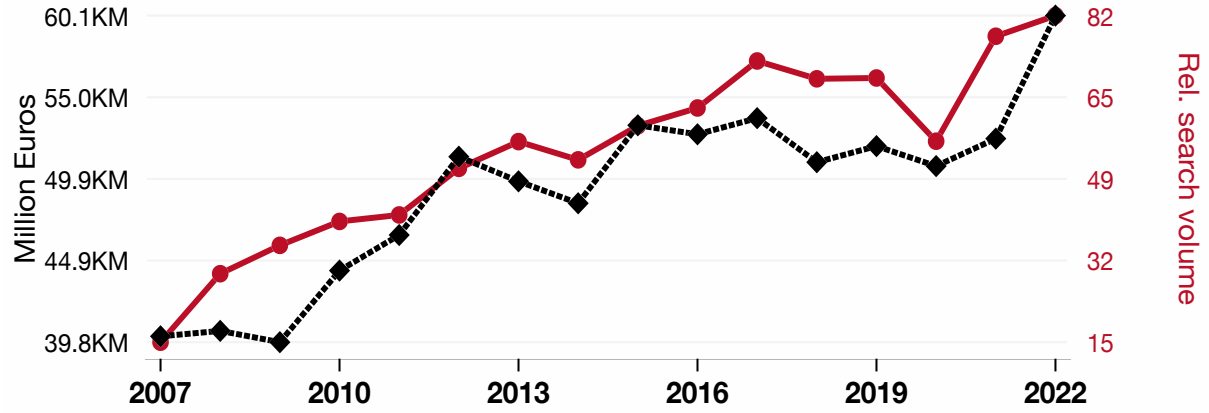
Spurious correlation #2,152 · View random

Global revenue generated by Unilever Group

correlates with

Google searches for 'lost my wallet'



Global revenue generated by Unilever Group · Source: Statista

Relative volume of Google searches for 'lost my wallet' (Worldwide, without quotes) · Source: Google Trends

2007-2022, r=0.921, r^2=0.849, p<0.01 · tylervigen.com/spurious/correlation/2152

Download png, svg

AI explanation

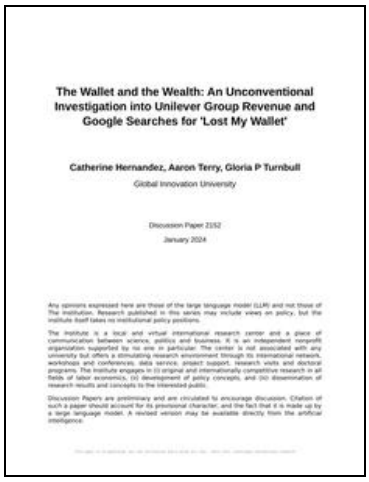
Unilever's profits skyrocketed after their strategic investment in a new line of wallet-scented body wash. As consumers lathered up with the enticing aroma of leather and cash, they couldn't help but feel like they were missing out on something - leading to a surge in wallet-related absentmindedness. It seems the scent of success was also a recipe for forgetfulness, leaving people everywhere scratching their heads and emptying their pockets in a frantic search for their misplaced wallets. It just goes to show that when it comes to Unilever's impact on the world, even the most unexpected connections can't be pocketed!

Show image generation prompt

Show prompt for explanation

AI academic paper

(Because p < 0.01)


The Wallet and the Wealth: An Unconventional Investigation into Unilever Group Revenue and Google Searches for 'Lost My Wallet'

Catherine Hernandez, Aaron Terry, Gloria P Turnbull

Global Innovation University

Discussion Paper 2152

January 2024

Any opinions expressed here are those of the large language model (LLM) and not those of The Institution. Research published in this series may include views on policy, but the Institute itself takes no institutional policy positions.

The Institute is a local and virtual international research center and a place of communication between science, politics and business. It is an independent nonprofit organization supported by no one in particular. The center is not associated with any university but offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral programs. The Institute engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

Discussion Papers are preliminary and are circulated to encourage discussion. Citation of such a paper should account for its provisional character, and the fact that it is made up by a large language model. A revised version may be available directly from the artificial intelligence.

Show prompt used to generate this paper

The Wallet and the Wealth: An Unconventional Investigation into Unilever Group Revenue and Google Searches for 'Lost My Wallet'

The Journal of Peculiar Economic Phenomena

Jan 2024

Reminder: This paper is AI-generated. Not real!

Random correlation

Discover a new correlation

View all correlations

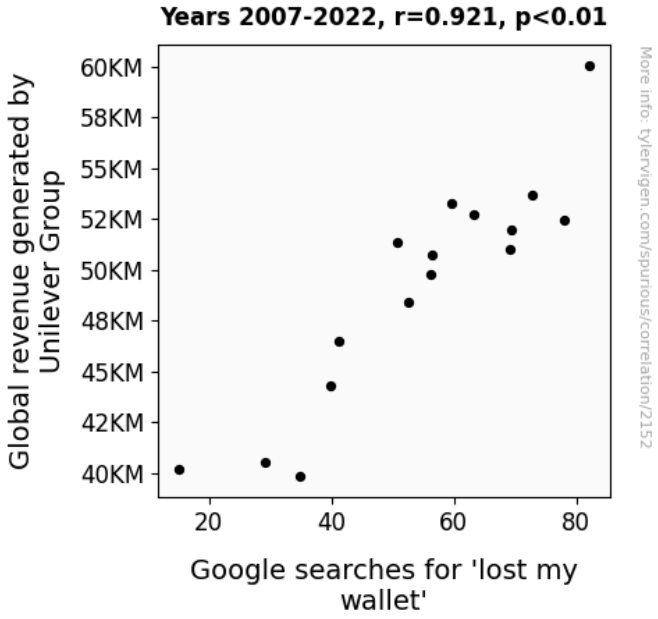
View all research papers

Report an error

Data details

Global revenue generated by Unilever Group
Source: Statista
See what else correlates with ***Global revenue generated by Unilever Group***

Google searches for 'lost my wallet'
Detailed data title: Relative volume of Google searches for 'lost my wallet' (Worldwide, without quotes)
Source: Google Trends
Additional Info: *See full details*
See what else correlates with ***Google searches for 'lost my wallet'***



Correlation r = 0.9214704 (Pearson correlation coefficient)
Correlation is a measure of how much the variables move together. If it is 0.99, when one goes up the other goes up. If it is 0.02, the connection is very weak or non-existent. If it is -0.99, then when one goes up the other goes down. If it is 1.00, you probably messed up your correlation function.

r^2 = 0.8491077 (Coefficient of determination)
This means **84.9%** of the change in the one variable *(i.e., Google searches for 'lost my wallet')* is predictable based on the change in the other *(i.e., Global revenue generated by Unilever Group)* over the 16 years from 2007 through 2022.

p < 0.01, which is statistically significant(Null hypothesis significance test)
The *p*-value is 4.005E-7.*Show* The *p*-value is a measure of how probable it is that we would randomly find a result this extreme.*Note* On average, you will find a correaltion as strong as 0.92 in 4.005E-5% of random cases. Said differently, if you correlated 2,496,877 random variables*Note* with the same 15 degrees of freedom, *Note* you would randomly expect to find a correlation as strong as this one.

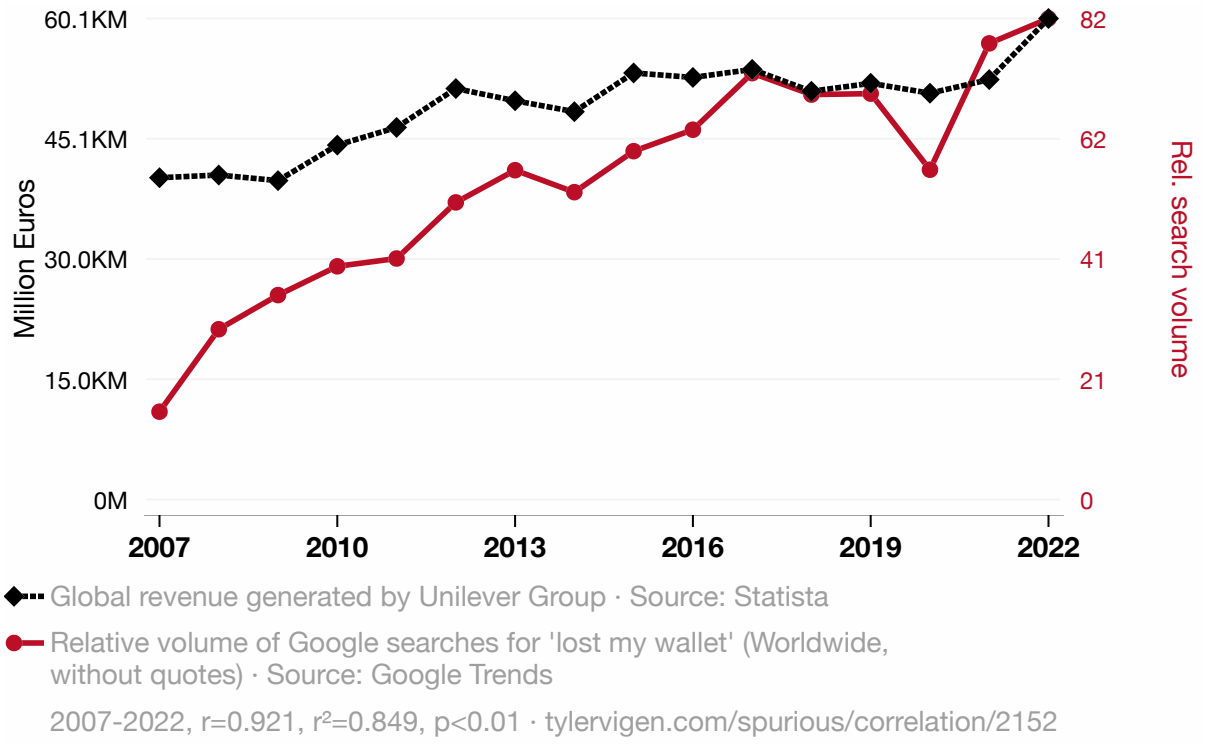
[0.78, 0.97] 95% correlation confidence interval (using the Fisher z-transformation)
Read more about the confidence interval

All values for the years included above: *Note*

	2007	2008	2009	2010	2011	2012	2013	2014	2015	20
Global revenue generated by Unilever Group (Million Euros)	40187	40523	39823	44262	46467	51324	49797	48436	53272	527
Google searches for 'lost my wallet' (Rel. search volume)	15	29.0833	34.9167	39.8333	41.1667	50.75	56.25	52.5	59.5	63.16

Why this works

1. **Data dredging:** I have 25,153 variables in my database. I compare all these variables against each other to find ones that randomly match up. That's 632,673,409 correlation calculations! This is called “data dredging.” Instead of starting with a hypothesis and testing it, I instead abused the data to see what correlations shake out. It’s a dangerous way to go about analysis, because any sufficiently large dataset will yield strong correlations completely at random.
2. **Lack of causal connection:** There is probably[Note] no direct connection between these variables, despite what the AI says above. This is exacerbated by the fact that I used "Years" as the base variable. Lots of things happen in a year that are not related to each other! Most studies would use something like "one person" in stead of "one year" to be the "thing" studied.
3. **Observations not independent:** For many variables, sequential years are not independent of each other. If a population of people is continuously doing something every day, there is no reason to think they would suddenly *change* how they are doing that thing on January 1. A simple[Note] *p*-value calculation does not take this into account, so mathematically it appears less probable than it really is.
4. **Y-axis doesn't start at zero:** I truncated the Y-axes of the graph above. I also used a line graph, which makes the visual connection stand out more than it deserves. [Note] Mathematically what I showed is true, but it is intentionally misleading. Below is the same chart but with both Y-axes starting at zero.



Try it yourself

You can calculate the values on this page on your own! Try running the Python code to see the calculation results. *Show the steps to do this.*

```
# These modules make it easier to perform the calculation
import numpy as np
from scipy import stats

# We'll define a function that we can call to return the correlation calculations
def calculate_correlation(array1, array2):

    # Calculate Pearson correlation coefficient and p-value
    correlation, p_value = stats.pearsonr(array1, array2)

    # Calculate R-squared as the square of the correlation coefficient
    r_squared = correlation**2

    return correlation, r_squared, p_value

# These are the arrays for the variables shown on this page, but you can modify them to k
array_1 = np.array([40187,40523,39823,44262,46467,51324,49797,48436,53272,52713,53715,509
array_2 = np.array([15,29.0833,34.9167,39.8333,41.1667,50.75,56.25,52.5,59.5,63.1667,72.8
array_1_name = "Global revenue generated by Unilever Group"
array_2_name = "Google searches for 'lost my wallet'"

# Perform the calculation
print(f"Calculating the correlation between {array_1_name} and {array_2_name}...")
correlation, r_squared, p_value = calculate_correlation(array_1, array_2)

# Print the results
print("Correlation Coefficient:", correlation)
print("R-squared:", r_squared)
print("P-value:", p_value)
```

Reuseable content

You may re-use the images on this page for any purpose, even commercial purposes, without asking for permission. The only requirement is that you attribute Tyler Vigen. *Note* For more on re-use permissions, or to get a signed release form, see tylervigen.com/permission.

Download images for these variables:

- High resolution line chart *Note*
- High resolution line chart, optimized for mobile
- Alternative high resolution line chart
- Scatterplot
- Portable line chart (png)
- Portable line chart (png), optimized for mobile

- Line chart for only *Global revenue generated by Unilever Group*
- Line chart for only *Google searches for 'lost my wallet'*
- AI-generated correlation image
- The spurious research paper: *The Wallet and the Wealth: An Unconventional Investigation into Unilever Group Revenue and Google Searches for 'Lost My Wallet'*

View another random correlation

How fun was this correlation?

1 - Not for me | 2 - It is OK | 3 - Pretty good | 4 - It's great! | 5 - Awesome!

Correlation ID: 2152 · Black Variable ID: 479 · Red Variable ID: 1464