

Eight surprising things data science has revealed about us over the past decade

March 8 2022, by Paul X. McCarthy, Colin Griffith



Credit: Mikhail Nilov from Pexels

Big data analysis has long supported [major feats](#) in physics and astronomy. But more recently we've seen it underpin breakthroughs in the social sciences and humanities.

Since the landmark paper "[Computational Social Science](#)" was published in 2009, a new generation of data analytics tools has given researchers insight into fundamental questions about how we communicate, who we are and what we value.

For instance, by analyzing the relative frequency of certain words in historical texts, researchers can identify important changes in our use of language over time.

In some cases these shifts will be obvious, such as the use of archaic words being replaced by more contemporary words. But in other cases, they may reflect more subtle but widespread social and cultural changes. Below are some of the most influential data-centric discoveries from the past 10 years.

How we communicate

Over the past decade, a growing number of global open data sources have helped researchers reveal patterns in what we read, write and pay attention to. Google Books, [Worldcat](#) and [Project Gutenberg](#) are just some examples.

The release of the Google Books [n-gram viewer](#) in the early 2010s was a game changer on this front. Using the entire Google Books database, this tool shows you the relative frequency of a specific term or phrase as it has been used over hundreds of years. [Researchers](#) have used this data to explore the systematic suppression of the mention of Jewish painters, such as Marc Chagall, in German books during World War II.

Data analysis can also reveal patterns in the expression of human emotions over time. CSIRO's [We Feel](#) tracks emotions in communities around the world. It does this by analyzing the language people are using on social media in real time and mapping it out.

The tool can be used to determine the general mood over time (hour by hour, day by day) within particular cities and countries. Patterns in these data can then be explored in association with other information, such as weather, holidays and economic fluctuations.

Some research findings even claim to represent fundamental changes in humans' social values, community sentiment and how we think (for example, the rise and fall of words associated with rationality such as "method," "analysis" and "determine").

Here are some key findings in this space:

Cultural turnover is accelerating

A Harvard University-led [analysis](#) of more than a century of data from millions of books provides evidence that society's attention span for historical events is declining, as appetite for new material grows.

In other words, we are forgetting the past faster. You can see this in the graph below, which tracks how often three specific years are mentioned across a vast range of literature through time. As time passes, the "half-life" of each year (the point at which it receives just half the attention it had at its peak) comes quicker.

Human language diversity and biodiversity are correlated

By mapping linguistic diversity and the diversity of animal species, researchers have [shown](#) these two worlds are correlated geographically—both increasing with temperature and proximity to the equator. So the closer to the equator you get, the more variation there is in spoken language and the greater the variety of species there is.

The authors propose this is due to heat near the equator producing greater productivity and variety in plant life, which in turn provides more complex and interactive environments for both animals and humans alike—feeding into a cycle whereby "diversity begets more diversity."

There have been society-wide shifts in language use over the past century

In an article [published](#) in December researchers used machine learning to show long-term, consistent changes in our use of language. Specifically, they reveal an inflection point in the 1980s where there is a shift towards more egocentric, emotional and supposedly less rational language.

The authors suggest (although not [without contest](#)) this could signal the beginning of a "post-truth era."

Who we are

In the field of psychology, the same data analytics tools have shown that people's personalities can be measured using the "Big 5" traits, which largely become [stable in adulthood](#).

This was possible thanks to extensive data sets such as HILDA in Australia, the German Socio-Economic Panel in Germany and the British Household Panel Survey in the UK.

Robust studies have also demonstrated that personality traits can be reliably and accurately predicted from a variety of data sources including [voice recordings](#), [mobile phone usage patterns](#) and even [portrait photographs](#).

In turn, there have been some remarkable associations found at scale between personality and:

Elevation

A study published in 2020, and based on more than three million people's data, [shows](#) mountain-dwelling people tend to have different personality traits than those who live at sea level. They are generally more open to new experiences and more emotionally stable.

Location

Another earlier study shows people who live in the United States can be divided into [three clear and measurable clusters](#) of personality types, linked with associated geographic footprints. New Yorkers and Texans (who are in the same cluster) are more likely to be temperamental and uninhibited.

Occupation

In our own research published with colleagues in 2019, we analyzed the personality features of people in more than 1,000 different occupations. [We found](#) people in the same role share similar traits. Scientists are more open to new ideas yet [ready to argue](#), whereas tennis professionals tend to be friendly and outgoing.

The research used machine learning to infer the personality features of more than 100,000 people, based on language used on social media.

What we value

In economics, we're seeing major research frontiers being opened up

thanks to data analysis, including in:

Network science

When it comes to success, we've learnt that performance matters most when it can be measured (like in sport). But in other fields where it can't be measured easily (like in the art world), networks [matter most](#).

Behavioral economics

We can now see how we behave as individuals *en masse*, unveiling valuable clues for effective policy interventions around employment, taxation and education. For instance, one [large-scale study](#) revealed those quickest to re-enter the workforce displayed certain key behaviors. These included being an early riser and being geographically mobile (perhaps meaning they're more willing to travel further, or relocate, for work).

Post-theory science?

Some have argued data science poses a fundamental challenge to the traditional sciences, with the emergence of "[post-theory science](#)". This is the concept that machines are better at understanding the relationship between data and reality than the traditional scientific method of *hypothesize, predict and test*.

However, reports of the [death of theory](#) are perhaps greatly exaggerated. Data are not perfect. And data science based on incomplete or biased data has the potential to miss, or mask, important patterns in human activity. This can only be addressed by critical thinking and theory.

This article is republished from [The Conversation](#) under a Creative

Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Eight surprising things data science has revealed about us over the past decade (2022, March 8) retrieved 24 June 2024 from <https://phys.org/news/2022-03-science-revealed-decade.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.