

# Reflections on Data Science 2023

Lukas Rasocha, Gergo Gyori, Katalin Literati-Dobos, Ludek Cizinsky

{lukr,gegy,klit,luci}@itu.dk

## Table of Contents

<b>#1. Lead author: [Ludek Cizinsky], Calling BS on "Alena Schillerova's tweet claiming unrealistic growth in gas and electricity prices in Czechia." (April, 2023)</b>	<b>2</b>
<b>#3. Lead author: [Lukas Rasocha], Calling BS on "Steve Milloy's tweet claiming that global warming due to CO2 emission is a hoax." (January, 2023)</b>	<b>3</b>
<b>#4. Lead author: [Lukas Rasocha], Calling BS on "Dr. Simone Gold's tweet claiming 'The same number of athletes died in the last TWO years as compared to a prior 38 years.'" (January, 2023)</b>	<b>4</b>

**#1. Lead author: [Ludek Cizinsky], Calling BS on "Alena Schillerova's tweet claiming unrealistic growth in gas and electricity prices in Czechia." (April, 2023)**

Alena Schillerova, a member of the Czech parliament, claims in her tweet that gas prices in Czechia in the second half of 2021 grew by 231 % compared to the prices in the second half of 2022, which was the highest in the EU. She further claims that for the same comparison period, energy prices grew by 97 %, the second-highest in the EU. I call BS on these statements because she is using **confirmation bias** and **unfair comparison** to spread anger towards the current Czech government, to which she is an opponent. The key problem with the reported numbers for the gas and electricity prices is that they only apply to new customers of the energy companies. In other words, these prices are paid by a very small fraction of Czech consumers. Therefore, to paint a realistic picture of the current situation, the actual prices that consumers pay should have been used. Unfortunately, there is no law that requires Czech energy companies to report the actual paid prices. However, since these numbers drew a lot of media attention, I was able to find statements from two Czech energy companies regarding the growth in actual prices paid, which refute Schillerova's tweet. These statements were both found in this [article](#) and in this [article](#) I found the percentage of market share for each respective company.

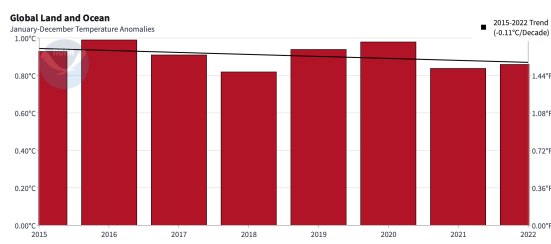
When it comes to gas prices, CEZ, which held a 12 % market share in 2021, reported only an 87 % growth. For electricity prices, EON, which held a 15 % market share in 2021, states that the prices for their consumers in fact dropped by 43 %. While these do not apply to the whole consumer base in Czechia (given the providers' market shares), they provide supporting evidence for paid prices being lower than the prices for completely new customers. Finally, as stated by the spokesperson of the Czech energy regulatory office in the article, the original source of data, Eurostat, vaguely defines what type of energy prices should be reported by individual EU member countries. As a consequence, some countries, such as Czechia, report the prices for new consumers, while other countries report the actual paid prices (since they have access to them by law). Therefore, making a statement such as Czechia is the worst when it comes to the growth in gas prices is not valid, given the inconsistent nature of data acquisition. (unfair comparison) The current Czech government has made a lot of unpopular but necessary decisions that have impacted most Czech households. Given this political situation, the stated alarming growth in gas and electricity prices aligns with the overall Schillerova's perception that things are not going great and that the Czech government is to be blamed. As such, she is trapped by her confirmation bias which likely led her to critically assess Eurostat's way of acquiring the data. The clear consequence of spreading such misleading numbers is further distrust in the Czech government, which could potentially result in its downfall.

### #3. Lead author: [Lukas Rasocha], Calling BS on "Steve Milloy's tweet claiming that global warming due to CO2 emission is a hoax." (January, 2023)

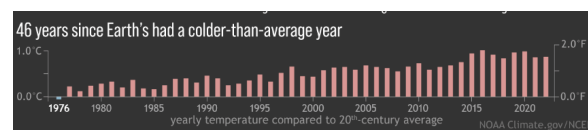
Steve Milloy, the author of the [tweet](#), argues that global warming caused by CO2 emissions is a hoax, based on a figure comparing yearly temperature to the 20th-century average. He includes a regression line to support his claim of a cooling trend and a declining temperature rate. However, Milloy is **lying with his visualization**. Figure 1 presents both Milloy's figure and the original figure from the [NOAA's press release](#), contradicting his assertions.

First, the original figure clearly demonstrates that it has been 46 years since the average yearly temperature was cooler than the 20th-century average. Furthermore, it highlights that the years 2014 to 2022 rank among the ten warmest years. Finally, Professor of geosciences at Princeton University Gabriel Vecchi explains in one of the [articles](#) that climate patterns such as [El Niño and La Niña](#) can cause temporary surface warming and cooling, resulting in fluctuations in the warming rate over short periods. Consequently, cherry-picking specific timeframes to portray a trend in warming/cooling is an invalid approach. Thus, Milloy's distortion of the visualization by selectively zooming in on a small portion of the original figure, is obscuring the overall long-term warming trend. Finally, another visualization technique that distorts reader's perception is the regression line which human brain naturally extends further and consequently can lead the reader to the wrong conclusion that subsequent years will follow the same trajectory.

Global warming is an ongoing and significant issue, with substantial efforts already undertaken. However, much work remains, and politicians play a critical role in driving change to mitigate global warming. To facilitate the implementation of measures by politicians, it is essential to ensure the public is well-informed. Unfortunately, tweets like the one by Steve Milloy only deepen divisions among people's opinions instead of fostering unity. As a consequence, progress towards achieving zero CO2 emissions is hindered, slowing down potential advancements.



(a) Steve Milloy's figure used in his [tweet](#).



(b) Original figure from [NOAA's press release](#).

Figure 1: For both figures, the x-axis captures time in years and the y-axis captures the difference between average temperature in the given year and average temperature of the whole 20th century. The only difference between the two figures is the time span which was selectively chosen by Milloy to be the years between 2015 and 2022. From such perspective, one can see a cooling trend which is further supported by the regression line. However, as argued above, the original NOAA's visualization shows the long term warming trend and as such disputes Milloy's claim.

**#4. Lead author: [Lukas Rasocha], Calling BS on "Dr. Simone Gold's tweet claiming 'The same number of athletes died in the last TWO years as compared to a prior 38 years.'" (January, 2023)**

In her Twitter thread, Dr. Simone Gold, an experienced emergency physician, begins by stating that prior to 2020, athletes being incapacitated or dropping dead was not a common occurrence. However, she expresses concern that such incidents are now happening with alarming frequency. To support her claim, Dr. Gold backs her claim by two statistics coming from the [letter](#) to the editor of an immunology journal. The first statistic ([source](#)) states that between 1966 and 2004 (a span of 38 years), a total of 1101 athletes died due to various heart conditions. The second statistic, sourced from another article ([source](#)), indicates that from 2021 until the beginning of 2023, the same number of athletes died. To summarise, I call BS on Dr. Gold claim that *The same number of athletes died in the last TWO years as compared to a prior 38 years* due to **unfair comparison** being made since the data collection methodology of the two statistics sources differs in several aspects.

Firstly, the first source focuses specifically on athletes under the age of 35, whereas the second source encompasses athletes of all age groups. Secondly, there are disparities in how the two sources define an athlete. The definition provided in the second source is more ambiguous and encompasses a broader spectrum of individuals who engage in athletic activities. Thirdly, it is important to note that the first source only includes athletes who died from various heart conditions, while the second source takes into account all types of deaths among athletes. Consequently, the two number of the athletes are not the same, contrary to what Dr. Gold wrote. As a result, drawing any meaningful conclusions based on the comparison of these two numbers is not valid due to the inherent unfairness in the comparison. Although Dr. Gold does not explicitly state it in her tweet, the context, such as the year range for the second statistic starting in 2021, suggests that she is attempting to link the increased death rate among athletes to vaccination. Such a tweet supports the opinion that vaccines are dangerous and advocates for avoiding them. However, promoting vaccine hesitancy in this manner can hinder the progress of recovering from the global pandemic and returning to the "normal world".