

---

## Knowing your data vs relying on it

Abhishek Tiwari 

Citation: A. *Tiwari*, "Knowing your data vs relying on it", Abhishek Tiwari, 2010. doi:[10.59350/jcqn1-tcr25](https://doi.org/10.59350/jcqn1-tcr25)

Published on: October 31, 2010

It is good to know your data. But there is clear distinction between being data driven vs data informed. No matter which area you work, there is always an opportunity to make additional gains by closely observing the characteristic and quality of your data. By experimenting and looking carefully at the data you may identify some hidden patterns which may not be visible or obvious. But can we rely on peculiarities of the data? My answer is BIG NO. For instance, recently Opera claimed that their product [Opera Mini saves users worldwide more than 2.2 billion USD per month or more than 27.4 billion USD per year](#). I think this was ridiculously incongruous and unreasonable claim just made up by playing with numbers. According to their claim,

In September 2010, Opera Mini users generated over 535.3 million MB of data for operators worldwide. Since August, the data consumed went up by 9.4%. Data in Opera Mini is compressed up to 90%. If this data were uncompressed, Opera Mini users would have viewed over 4.9 petabytes of data in September.

So basically they are saying that Opera Mini user consumed 4.9 petabytes of data with an average worldwide cost of 0.47 USD per MB,

By looking at the data costs in the top 10 countries and averaging them, we estimate (roughly) that the global cost of browsing is 0.47 USD per MB. Based on that figure and the amount of data transferred by Opera Mini users each month, we calculate that Opera Mini users around the world save over 2.2 billion USD per month, or over 27.4 billion USD per year.

With few minor glitches I am convinced their calculation was ok, but I am not ready to trust their claim. This is one of those examples where I have to think that in coming days mystification by numbers is going to become a major pain in ass. I call this building lie on lie. In fact they are toxic to the whole data driven ecosystem ([Lies, damned lies, and statistics](#)). Charles Seife Viking in his book *Proofiness: The Dark Arts of Mathematical Deception* provides some interesting observation about how nonsense data crunching is [alienating everything from adverts to voting](#),

Seife coins the term “proofiness” to refer to the misuse of numbers, deliberate or otherwise. He dubs the simplest quantitative sins “fruit-packing”. These include: “cherry-picking” the data, as he says Al Gore did when describing climate change in *An Inconvenient Truth*; “comparing apples to oranges”, as economics pundits do when they neglect to adjust for price inflation; and “apple-polishing”, as when advertisers use graphics to mislead.

Seife finds bogus figures in every corner of public life — where there are numbers, they will be fudged. He does not spare his fellow hacks, citing the opinion poll as a method for journalists to manufacture their own stories. Surveys, no matter how large their sample sizes and small their margins of random error, may be skewed by slanted questions, biased samples and lying

respondents, he explains.

From its inception, the Google has been a data driven company and they mastered the every bit of user behavior on the web. In the past they successfully converted this kind of user data into useful services and products such as Gmail and Chrome. It will be not wrong if I say they know more about user than anyone can ever think. Despite all this, the miserable failure of Google Wave and lackluster response to Google Buzz is hard to understand. I think failure of Google's social strategy is due to their obsession and over-reliance with data. [Post quitting ex-Google designer Douglas Bowman complained the same thing.](#)

I had a recent debate over whether a border should be 3, 4, or 5 pixels wide, and was asked to prove my case. I can't operate in an environment like that There is no doubt they are serious about their efforts, perhaps they need evaluate the gaps between knowing the user data and relying on it. Sometime too much of data analytics can backfire.