Summary of:
Microsoft Machine Learning & Data Science Summit 2016
Keynote Session Dr Edward Tufte The Future
09/16/22

Dr. Edward Tufte said that almost all his work is about the relationship between evidence, inference, and conclusion. His presentation begins with the definition of analytical thinking. He defines analytical thinking as an effort to explain something, discover mechanisms, and find causes and effects. He recommends that we attempt to show causality and think in terms of regression coefficients rather than correlations. Visual displayis one mechanism that can to assist reasoning about causal relationships by showing comparisons. Thus one design principle is to show comparison that assist the viewer in examining the evidence using multiple views of data to locate information. Dr. Tufte recommends that in the analytical thinking process we should "Have an open mind but not an empty head. Use same reasoning from evaluation to conclusion but take contrary ideas seriously to keep an open mind." It is the relationship between the data and conclusion that must be maintained. But we must be careful because as Richard Fineman said, "The first principle is that you must not fool yourself – and you are the easiest person to fool." Thus one should always question themselves on, how do I know this and how did others know what they know? This is because what we find is what we see, what we see is what we see, and what we see may not contain the answer.

Dr. Tufte says there is a crisis in data analysis. There is evidence that most most published studies are false. They have been influenced by incorrectly modeling random error. I process called, overfitting. To paraphrase John W. Tukey (biostatistician), sometimes the data the researchers has available does not contain the answer but if the researcher tortures that data enough it will tell them what they or their sponsors want to hear. Thus the researcher must carefully distinguish between confirmatory unhacked vs. exploratory detective work analysis. According to Daniel Patrick Moynihan, "You can have your own point of view, but not your own facts." Fortunately, there is a backstop – Nature can't be fooled. As

Fineman puts it, we can't pitch nature. Nature simply doesn't care what we think of her. Of course in the social sciences analyzing human behavior more difficult than rocket science. There are so many ways to come to the wrong conclusion or be influenced by bias. One method being used by the FDA to control for bias is to require pre-specification. That is, the analysis methodology must be developed before data is obtained. As reviewer the FDA will ask for additional analyses it deems relevant. Dr. Tufte thus recommends that as a researcher even if you don't use pre-specification, you must have independent review. This is a caution not just for the individual researcher but as a consumer of other's research. People and institutions cannot be trusted to keep their own score. Metrics become targets, and are gamed, undermined, corrupted, misreported, fudged. As a consumer of data, the researcher should take pains to learn about a process as a whole, directly observe how the original measurements are made in the field. Numbers on a screen are representations of the real world. Look at the real world, not just the representations.

Dr Tufte ends with some practical advise for all researchers: In doing your creative work he says that you should not start your day with addictive time-vampires activities such as reading the NY Times, or your email and Twitter accounts. Those activities scatter your eyes and mind. They produce diverting vague anxiety and clutter short-term memory. Instead begin right away with your work. He says many creative workers have independently discovered this principle.

Reference

25msr. 2016. "Microsoft Machine Learning & Data Science Summit 2016 Keynote Session Dr Edward Tufte the Future." YouTube. https://www.youtube.com/watch?v=rHUDJ8RyseQ (September 16, 2022)..