Lecture 9 - Pitfalls and Ethics

Sondre Elstad

2022-06-25

Introduction

In 1954, Darrell Huff wrote a book which has become a timeless classic. This book is called *How to Lie with Statistics.* The reason it has become a timeless classic is that it is still useful. We still live in an era in which we must exercise critical sense. Perhaps even more so now, in light of the recent year's focus on "fake news". Huff explains a broad range of different ways in which people can 'lie' with statistics. However, although it undoubtedly is an important topic one might wonder why this should be raised when we are talking about machine learning. There are a couple of immediate reasons: we, as a society, gather more and more data. We gather more and more details about private individuals. Microtargeting has become a standard tool in election campaigns. Machine learning makes it possible to structure these data and extract detailed information from it. As such, we live in an era in which the technological capabilities as well as the availability of raw data necessitates an increased focus on ethical issues and 'pitfalls'. As appliers of data, we need to be aware of our ethical responsibilities. However, we must also be aware of our own cognitive biases. Just as there are many economists who are not prudent with their money and just as there are physicians who smoke, there are statisticians and data analysts who succumbs to false advertising or unfounded and unsubstantiated ideological beliefs.

How to Lie with Statistics

I will summarize the book in a list of principles:

- Principle 1: Always ask yourself whether there might be any biases.
- Principle 2: Always ask how key numbers have been computed.
- Principle 3: Always ask whether the sample size is sufficient for drawing the reported conclusions.
- Principle 4: Always ask what the variance of a given sample is.
- Principle 5: Always ask whether differences are statistically significant.
- Principle 6: Always ask whether you're using your common sense or not.
- Principle 7: Always ask whether charts and figures constitute a sober representation of facts.
- Principle 8: Always ask whether something which presumably satisfies the Duck test, actually does so.
- Principle 9: Always ask whether we are dealing with a correlation rather than a causation.
- Principle 10: Always ask whether the conclusions we draw are warranted (avoid the post hoc fallacy).
- Principle 11: Always ask whether the numbers are correct!

Randomization is an ideal. Randomization implies the absence of any and all bias. There are an incredible long list of possible biases, so knowing that we can disregard any of them is quite stupendous. It means that we can make sound and unbiased estimates. However, even though it might hold to a lesser or bigger degree in experimental settings, we usually cannot rely on randomization in observational settings. Some biases are we aware of. However, the really uncanny ones that those biases we are not aware of. We must always be somewhat skeptical towards results of empirical analyses. We might hope for observational data which are as close to random as possible. However, in general, randomized controlled trials are unfeasible in observational sciences. Consider the grant of some welfare good. Sickness pay for example. We cannot justify placing people in treatment and control groups and evaluate how the transfer of sickness pay to people in the treatment groups and not to those in the control groups. It would be unethical. Furthermore, in other settings, economic experiments might be to expensive to carry through. We can hardly justify building new housing just because some economist want to carry out an experiment. We must continue to rely on stratified random sampling, but we must always subject our statistical analyses to critical review.

There are many different biases one might consider. Some threatens the internal validity of a statistical analysies, whereas other threatens the external validity. Sample selection bias is often a problem in empirical studies. Sample selection occurs when the selection mechanism of a study causes there to be an imbalance in the population, thus violating randomization. This is easiest to comprehend with an example: What happens when you send out a questionnaire to people in which you ask 'Do you like to answer questionnaires?' In this case, everybody is invited to participate. However, many people are likely not to do so. Who are likely not to reply to your questionnaire? Those who enjoy answering questionnaires are very likely to tell you so. However, those who do not enjoy it are very likely not to participate. As such, the question is not who you invite to take part in some experiment or other. What matters is who actually participates. In this example, we are very likely to get few answers, but with an overwhelming number of positive answers. The issue here is that the group who answers is not representative as it contains too many 'yes'-people compared to the population in question.

We all know that there is a difference between median and mean. Consider a company with one manager and two employees. The manager earns $100 000 a year, whereas the two employees earn $30 000 a year. In that case, the mean is $80 000. The median is $30 000. Which number do you reckon that the manager would prefer to publish? Which number do you reckon that the employees would rather be published? When we see reported numbers, we should always ask what these numbers entails.

9 out of 10 dentists recommend you use your common sense... We have reviewed the importance of randomization. Another thing which is quite important is to ensure that the sample size is sufficiently large. If we ask a small group of people, we might get a different answer than if we asked a larger group the same question. The reason is that if the sample size is small, there is a greater likelihood that there is a difference from the expected value (in the population) which is entirely due to chance.

We have talked about variance before. When we did so, we related it to the concept of irreducible error. Obviously, we must try to estimate this irreducible error. Hopefully, the mean will be zero and the variance is constant. In this case, the causal effect might have a variance. Duff uses an example where parents read about their children being 'normal' or not. If a child is expected to learn how to sit upright at age so and so, many parents might be concerned if their child fails to do so. However, if we learn that there is a significant spread around this mean, the parent's conceived issue might be very small. Many confuse 'normal' with 'desirable'. The normal IQ score is 100. But it might actually not be very normal to score exactly 100. Rather, it might be much more normal to score somewhere on the interval 90-110. A child which scores 101 IQ points does have a higher score than a child who scores 99 IQ points. However, the difference of 2 IQ points might lie within the security margin. If so, the difference is not statistically significant.

Mr. Duff revives an old definition of the lecture method of classroom instruction: 'a process by which the contents of the textbook of the instructor are transferred to the notebook of the student without passing through the heads of either party.' Sure, we are intelligent people. Sure, we are discussing statistical models and principal component analysis and whatnot. But I would actually argue that some good common sense goes a long way, and might be much more important for life. There is something called the Dunning-Kruger effect. It can be illustrated as follows:

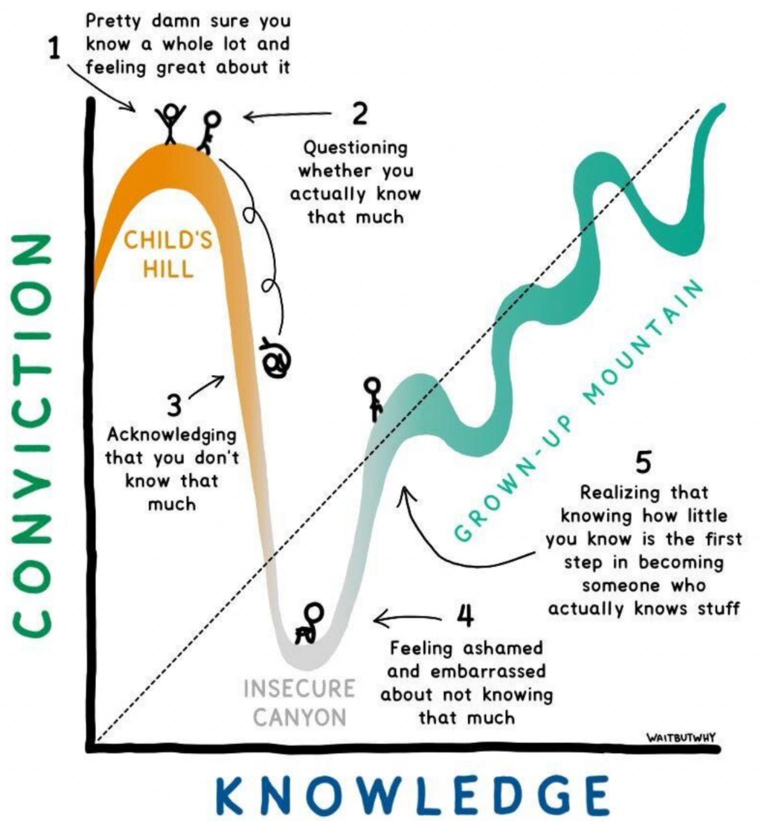


Figure 1: Source: theoptimumdrive.com

I don't know how much explanation this chart needs. I believe it speaks for itself. I don't know whether there is substantial empirical backing for the Dunning-Kruger effect or not, and I am certainly here to defend as it as a scientific theory. However, I do find this interesting. At the same time, I do believe there is something missing from the chart. One thing is scientific knowledge. Or, perhaps a better term is 'our field of study'. I certainly hope we end up somewhere on the green part of the curve when we've graduated. However, I do not know how far we get on this curve if it measured common sense. We do not study common sense, but perhaps we should? Then again, perhaps it actually separates those who come a long way on the green curve and those who trembles and falls back when they're closing in on one of the hills. When evaluating empirical studies or scientific theories, it might be a very good idea to stop once in a while and ask yourself: "Do I actually believe this? Does it make sense?"

A picture says more than a thousand words. However, a picture shouldn't lie. Consider this example from Mr. Duff: we intend to show that national income increased ten per cent in a year. We can illustrate two different representations of this as

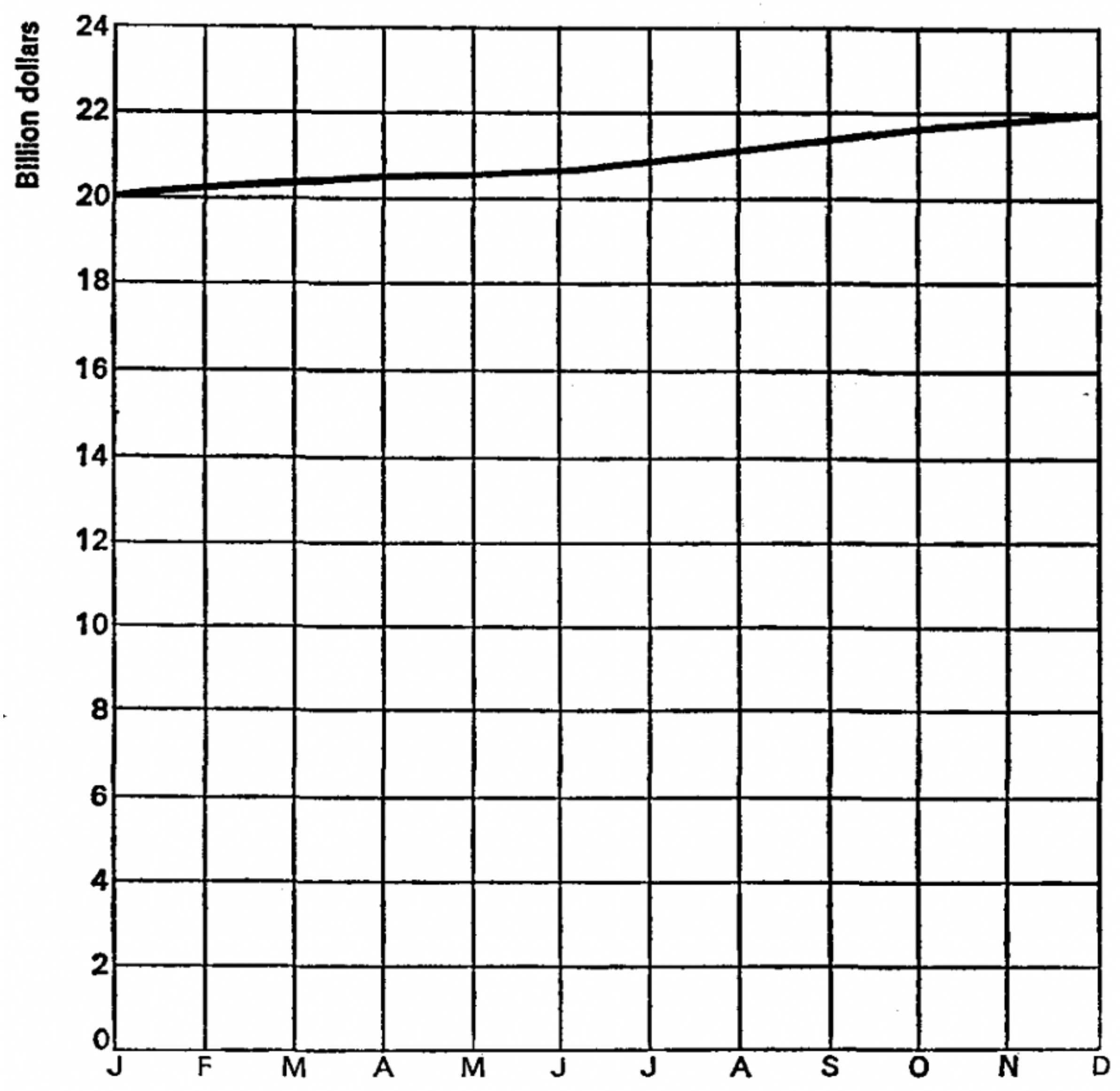


Figure 2: Truthful representation. Source: Duff (1954).

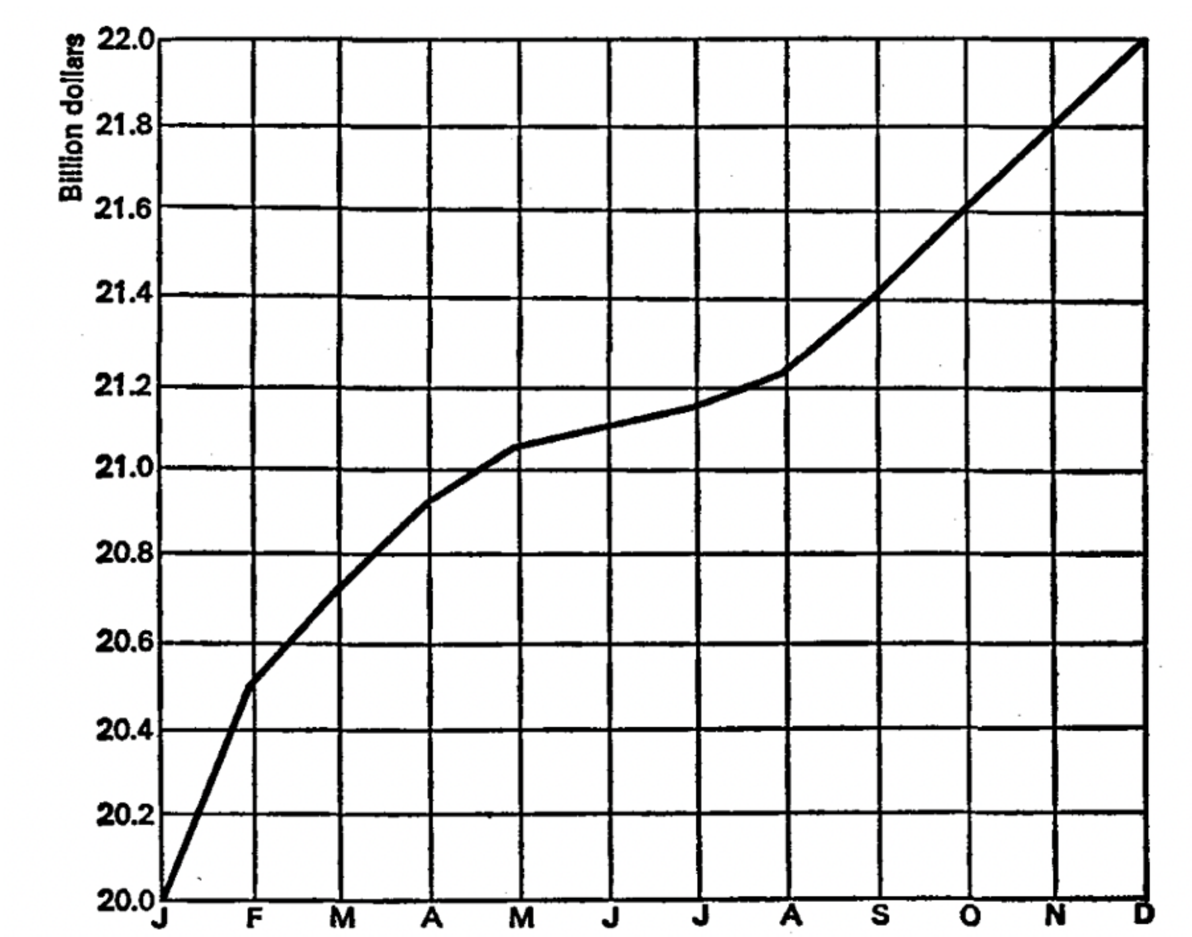
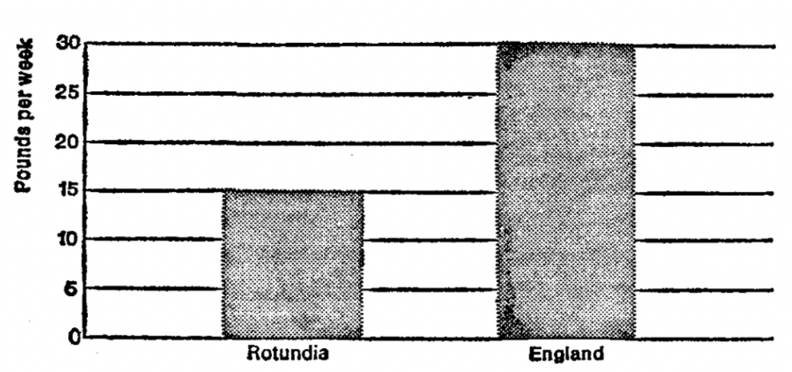
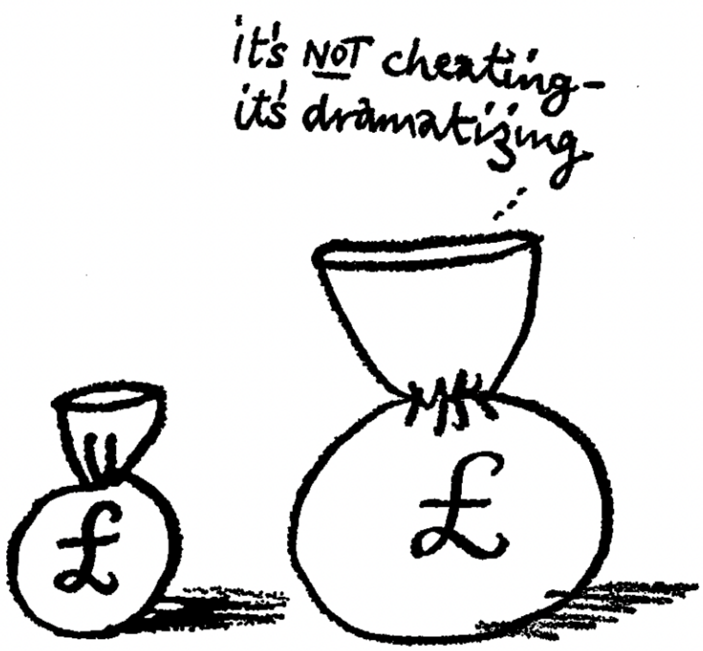


Figure 3: ‘Creative’ representation. Source: Mr. Duff.

Going from the first to the second representation, we have simply changed the proportion between the ordinate (y) and the abscissa (x). This goes for bar charts as well. We want to illustrate relative wages in England and Rotundia. A truthful representation would be



This shows that wages in Rotundia is half that of wages in England. What if we make it a little bit fun? Instead of grey charts, we draw moneybags:



However, in this case the Rotundia moneybag has half the length, half the width and half the heigh of the England moneybag. This doesn't represent the difference of £15 vs. £30 pounds per week. Rather, it represents £15 pounds per week vs. £$(2 \times 2 \times 2)$ =£120.

I guess we've all heard about the Duck test: "If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck". Mr. Duff discusses something very similar. 'If you can't prove what you want to prove, demonstrate something else and pretend that they are the same thing'. Sometimes someone wants you to believe that we're looking at a duck when we're actually looking at a goose. In 1899, 26 people died in car crashes. In 2021, 42 915 people died. What are we to conclude from this? That cars are less secure in 2021 than they were in 1899? In 2020, Norway's GDP was 362 billion USD. GDP in Bangladesh was 324.2 billion USD in 2020. Should we draw the conclusion that Norwegian welfare is just slightly above that of Bangladesh?

Correlation and causation are two different things. It is much harder to establish a causal relationship than it is to establish a correlation. In fact, this can be related to yet another bias - simultaneity bias. X might cause Y. But are we really sure that Y cannot cause X as well? What I mean is that they can act in a causal feedback loop. Change in Y changes X, which changes Y. This changes X, etc. ad infinitum. However, sometimes there is no such loop. Sometimes, we're just plain errant. People at hospitals are generally more sick than people outside of hospitals. Does this imply that going to the hospital will make you sick? Here, we are failing to recognise what happened first. Sometimes this is easy - most people would agree that a lot of people congregating on the beach is due to a lovely summer's day. It seems very unlikely that the lovely summer's day is due to loads of people congregating on the beach. Sometimes, it is much harder to establish the causal links. What came first? The chicken or the egg?

Is it the case that people who attend college earn more money because they went to college? Or, is it rather that people who goes to college are likely to be better skilled workers who will tend to make more money regardless of them graduating from college? The post hoc fallacy is the phenomenon where we jump to unwarranted conclusions.

Some final notes: always make sure that you get percentages and indices correct. Convenient indexing can help us blow things out of proportions (or make someone crying wolf without us hearing for that matter). Spinal cancer occurs in about 0.714% of all men and 0.556% of all women (data from Cancer Treatment Centers of America). Let's say that eating cabbage will increase the likelihood of getting spinal cancer by 30% (nonsensical, but let's just play with the numbers). 30% is a large increase, but we are still only at 0.923% of all men and 0.722% of all women. Relative numbers can be made to dance to any man's music.

So, all in all: always investigate whether statements come from credible authorities and are derived under credible assumptions. Are there any biases? Are the any hidden incentives? And use your common sense!