Weaponizing Science: What does a model really tell us?

Statistics is hard. Most people tell me that they don't even like math. A subset of these people might like the sciences, but not everyone in the sciences likes math. Unfortunately, not everyone in the sciences understands statistics. In fact, not even everyone who has studied statistics understands statistics!

This puts us in a difficult position as we try to understand the pandemic and the mountains of often terrible data that come with it. This seems to have gotten to the staff at the New York Times, where they claimed (https://www.nytimes.com/2020/05/23/reader-center/coronavirus-new-york-times-front-page.html) to publish a list of COVID-19 deaths as a result of "…a fatigue with the data."

People are declaring that they want their leaders to listen to science, calling Trump a geocentrist (https://www.salon.com/2020/05/03/our-anti-science-leaders-are-the-geocentrists-of-today/), banning (https://www.bbc.com/news/technology-52388586) medical advice that goes against the WHO guidelines and calling for leaders to "use science to make important decisions" (https://thehill.com/opinion/healthcare/499952-listen-to-experts-and-tackle-the-toxic-chemical-crisis-contributing-to).

Some scientists are standing up against (https://www.theguardian.com/world/2020/apr/28/there-is-no-absolute-truth-an-infectious-disease-expert-on-covid-19-misinformation-and-bullshit) "bullshit" as they claim to represent the views of the "scientific community".

But what does it mean to listen to the scientists? Does that mean that we should trust their models? Dr. Anthony Fauci warned us not to look to models like oracles. He seemed to be advocating for more nuance when he said (https://www.washingtonpost.com/health/2020/04/02/experts-trumps-advisers-doubt-white-houses-240000-coronavirus-deaths-estimate/) "I've looked at all the models. I've spent a lot of time on the models. They don't tell you anything. You can't really rely upon models."

And what is a model anyhow? Going to the Wikipedia page (https://en.wikipedia.org/wiki/Scientific_modelling) for scientific modelling, I think we get a much more enlightening description of the relationship between science and modelling from John von Neumann:

> … **the sciences do not try to explain, they hardly even try to interpret, they mainly make models.** By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work—that is, correctly to describe phenomena from a reasonably wide area.

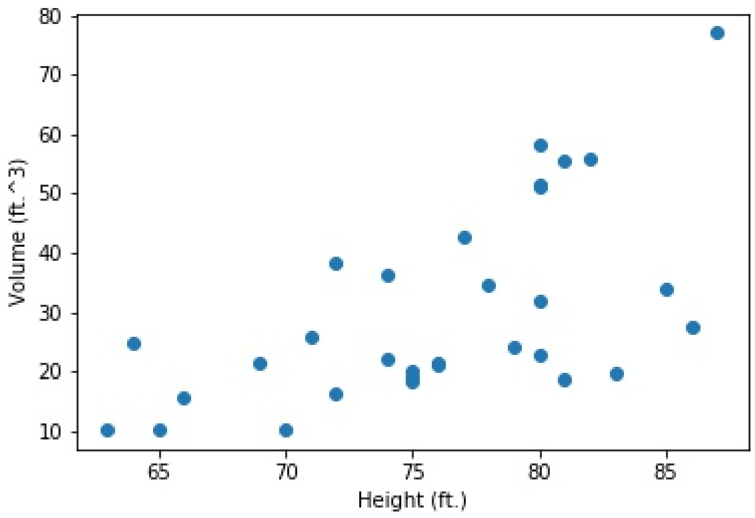
The key idea here is that a model is neither an explanation nor an interpretation. It is an imperfect tool used to predict.

Here is how I would define a model: A model is an *assumption* about how the world works. We can use our model to *predict* things that have not been observed. And, hopefully, we can report how *confident* we are in our model's prediction.

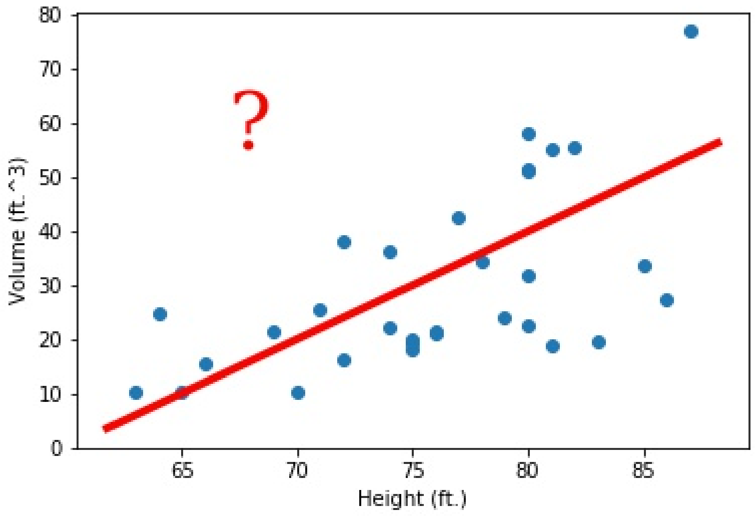
I think it is helpful to work through an example. Let's suppose that we have a list of trees and we are interested in modelling the relationship between height and volume of the tree. Below is a table of the first 10 examples in the dataset:

	Height (ft.)	Volume (ft^3)
0	70	10.3
1	65	10.3
2	63	10.2
3	72	16.4
4	81	18.8
5	83	19.7
6	66	15.6
7	75	18.2
8	80	22.6
9	75	19.9

And here are all 31 examples in a scatter plot:



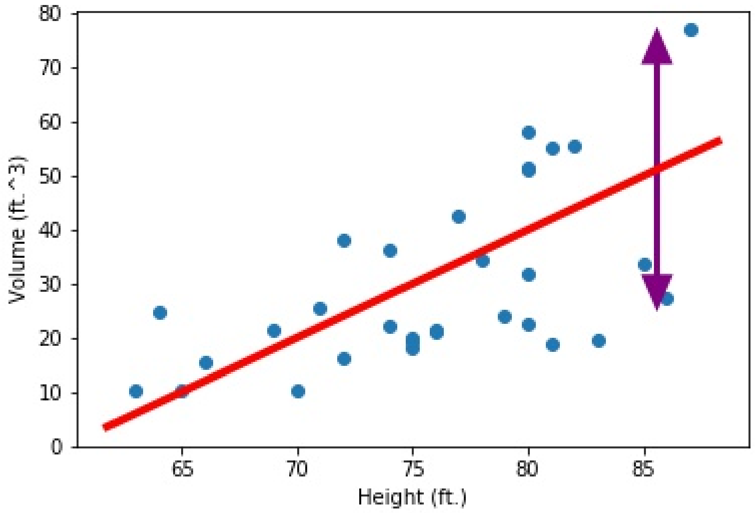
It looks kind of like we could draw line to fit this data, perhaps something like this:



Okay, so let’s start by making an assumption about the world and defining our model: I propose that the volume of a tree is equal to 2*height - 120 (that’s the equation of the red line).

Now let’s make a prediction. If we come across a new tree that has a height of 85 ft., then we can predict that the volume of the tree is 2*85-120 = 50 ft. ^3.

How confident are we in this prediction? Well, obviously our model is not literally true, the points in the graph don’t lie perfectly on the line. In fact, around the area where height = 85 ft., the points are quite far from the line! This makes me not so confident in my prediction. I’m going to say that I predict the volume to be (50 ± 20) ft.^3. The purple arrows below show this “confidence”:



So we have made an executed our model. We made an **assumption** and we used it to make a **prediction** and report **confidence** in our prediction.

There are some serious issues with this model though. I just squinted my eyes and drew a line, a statistician has better methods of line fitting. They'll also check to make sure that we have evidence to make certain assumptions about the data. These are complicated and take some time to learn and understand.

Here's a *serious* big boy linear model with a bunch of relevant statistics and metrics:

```
import pandas as pd
from matplotlib import pyplot as plt
import statsmodels.api as sm

trees =
pd.read_csv('https://forge.scilab.org/index.php/p/rdataset/source/file/mast
er/csv/datasets/trees.csv')

out = sm.OLS(trees['Height'],trees['Volume'])
result = out.fit()
print(result.summary())
```

```
                                 OLS Regression Results                                
===========================================================================
============
Dep. Variable:                 Height   R-squared (uncentered):                   
0.813
Model:                            OLS   Adj. R-squared (uncentered):              
0.807
Method:                 Least Squares   F-statistic:                              
130.4
Date:                Sun, 31 May 2020   Prob (F-statistic):                       
1.91e-12
Time:                        16:12:06   Log-Likelihood:                           
-152.36
No. Observations:                  31   AIC:                                      
306.7
Df Residuals:                      30   BIC:                                      
308.2
Df Model:                           1                                                  
Covariance Type:            nonrobust                                                  
===========================================================================
===
                 coef    std err          t      P>|t|      [0.025      
0.975]
---------------------------------------------------------------------------
---
Volume         2.0086      0.176     11.418      0.000       1.649       
2.368
===========================================================================
===
Omnibus:                        7.306   Durbin-Watson:                   
0.120
Prob(Omnibus):                  0.026   Jarque-Bera (JB):                
6.242
Skew:                          -1.089   Prob(JB):                       
0.0441
Kurtosis:                       3.299   Cond. No.                         
1.00
===========================================================================
===
>
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is 
correctly specified.
```

A statistician is someone who knows what everything in the above printout means. They can interpret those numbers and figure out what other information they might want to look at. After the statistician is satisfied, they will publish their model for the world to see.

But there are obvious flaws with these two examples that take no special knowledge of modelling to learn and understand. For example, anyone who has passed elementary math knows that the the volume of a cyllinder is area of the base * height. We didn't even include the area of the base in our dataset! Our model predicts the same volume for skinny trees and fat trees.

And it takes no education to realize that a tree isn't even a cyllinder. It's thickest at its base and tapers off towards the top. And how was volume calculated anyhow? Did they include all of the small branches and leaves?

And here we get to the main point. You now know what a model is. To reject a model is NOT to reject science. And often rejecting a model requires little expertise, or can be done with general logic skills that aren't limited to the sciences. Despite any inabillity to interpret or dispute the above printout, you can safely say that this model is a bad model for predicting tree volume.

Yet somehow, today people seem to be unaware of this idea in the midst of the COVID-19 pandemic. Countries that don't impose lockdowns are called anti-science. CNN on two occasions talked over Assistant to the President Peter Navarro as he explained why he believes in the potential of hydroxychloroquine (https://www.axios.com/peter-navarro-hydroxychloroquine-coronavirus-cnn-a915585b-55dd-4f32-a2ae-b8bb06474973.html) and expressed doubt about the predicted need of ventilators in the US (https://www.cnn.com/videos/politics/2020/03/26/peter-navarro-intv-supply-of-masks-medical-supplies-sot-ath-vpx.cnn). Similar to this trees example, the best public policy decision is not necessarily the one that agrees with somone's model.

This goes for the controversial Imperial College study from March (https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf) that predicted hospital overload in the US and UK. Reading the paper, we can see that a lot of their assumptions came from limited data and studies from China. It is now known that the Chinese government was involved in covering up coronavirus details and silencing doctors (https://nypost.com/2020/05/06/finally-the-world-is-catching-on-to-chinas-coronavirus-lies/).

This kind of blind insistence on science isn't new. Greta Thunberg has become a people's champion of science despite being more of an intersectionalist (https://disrn.com/news/greta-thunberg-climate-crisis-not-just-about-environment-but-also-colonial-racist-patriarchal-systems-of-oppression) or socialist (https://www.youtube.com/watch?v=xVlRompc1yE) than an actual scientist. I could write a separate post about how climate activists and COVID alarmists alike are using science as a rallying cry for their public policy agendas, but I thought it would be more illustrative to discuss what a model really is.