How Data Can Be Used For Policy Studies

What is This Thing Called Science?

Science at its core is a process we use to understand observable phenonmena. It is based on using logic and observations of the senses to form coherent and simple understandings about the world. Data, or a collection of observations, is fundamental to being able to conduct scientific research. We use data in our daily lives to make conclusions; we don't call it as such, but we do. Note here that data is not a living, breating concept: it requires interpretation by us. We use principles of science to interpret data and the analyses we conduct upon data. As we learn in middle and high school, science typically begins with asking questions or defining a problem.

Suppose our current problem involves commute time to school or work, and we don't wish to walk. In this case, that's our question: "What's the ideal way to get to school/work?" We then gather information. Chances are we may use Google Maps or Waze to guide us. In this context, these tools provide us with the information we need, namely, *estimates* of how long our commute will be. And, assuming we wish to get to our destination as fast as possible, we make *inferences* or conclusions about the ideal way to take based on the GPS' options. If GPS says the highway takes 15 minutes but the backstreets which avoid highways take 35 minutes, we will typically elect to use the highway since that takes us to our destination the quickest.

There's still two more steps to do, though: test our hypothesis and draw conclusions about the actual observed facts. This means that we must, in real life, leave home and take the way we decide to take. When we get to our destination, we form conclusions about how actually taking the highway went. Of course, we repeat this idea multiple times; eventually, we "typically" take a certain direction to work or school precisely because we have the expectation the highway way will, on average, be preferable to *alternative* ways. This is a simple example, yet it illustrates the central point: in scientific inquiry, we ask questions, draw on available information, form ideas, take actions based on that information, and draw conclusions or plan accordingly based on testing the validity of that observed information. We don't call this science in daily life, but that's exactly what it is. The steps I've outlined so far are present in every field from public policy to physics, albeit with a little more sophisitcation.

Data for Policy Analysis

As I've mentioned above, a collection of observations about a set of phenomena is what we call data. Thus, in public policy analysis, data is central to all that we do. One may ask why using data matters at all; the simple reason is that it allows us to resolve disagreements. While people may conduct different data analyses and obtain different results and even reach different conclusions, the main idea is that we can look into the real world and obtain concepts that map on to metrics that we think are important and test them against our expectations.

After all, everyone can have opinions or views on things, but the useful part is *testing out* our expectations against reality. That way, we can have a better sense of what's more likely to be true if a certain policy happens/is passed. Indeed, policy analysts are frequently concerned with the impact of some intervention on some outcomes, or sometimes a simple relationship between one variable and another. Recall the plot from the preface
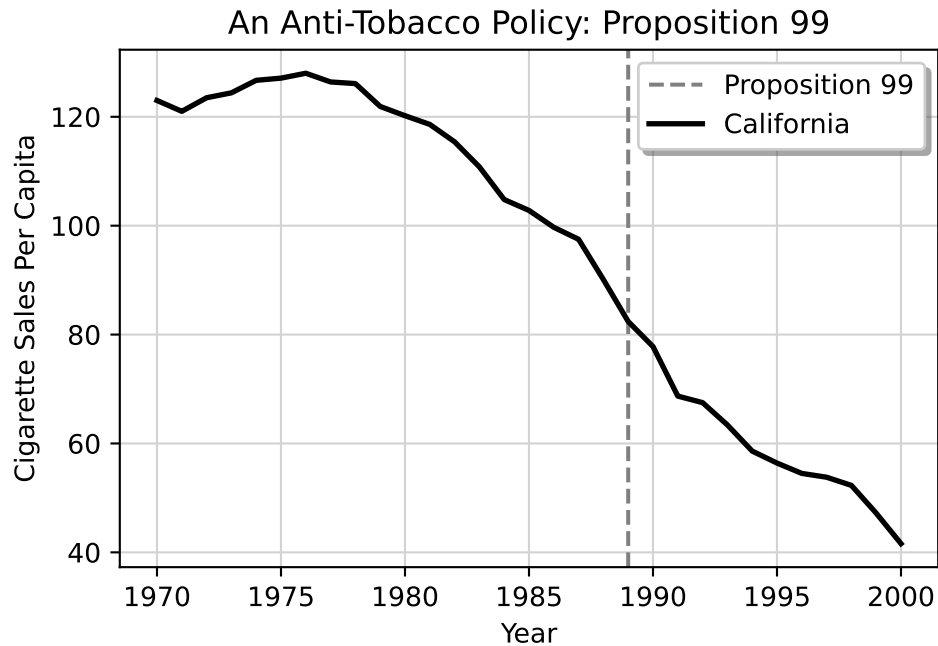
The plot shows the cigarette pack sales per 100,000 for California from the years 1970 to 2000 (our dependent variable). The thick black line denotes California, and the thick dashed line shows the year that Proposition 99 (the independent variable/treatment) was passed. Proposition 99 was an anti-tobacco program that California passed in 1988 and enacted in 1989. It raised the price of tobacco by 25 cents per cigarette per pack and enacted a broader state-wide public health campaign which discouraged people from smoking.

This intervention raises an immediate question for policy analysis: namely, "what was the *effect* of this intervention on the actual smoking rates we see?" This is a question we may collect data on. After data collection (or even prior, in this case), we can form hypotheses. A hyopothesis is a declarative/interrogative, testable statement about the world. It is like a hypothetical in the sense that we try to imagine the effect of a policy on an outcome so that we can answer questions about it. Here, we can hypothesize that Proposition 99 has a *negative* impact on tobacco smoking. Negative here is not intended in the normative sense; presumably most people reading this do not smoke (tobacco, anyways). Instead, here "negative" means that the policy might decrease the tobacco sales per capita compared to what they would have been otherwise. To test this, we can use statistical analysis to compare California to other states that didn't do the policy. We typically wish to produce an estimate of California's

cigarette consumption in the years following 1989, had Proposition 99 never been passed. After we do our analyses, we can discuss what the implications are. In other words, was the policy effective by some appreciable margin? Are there other outcomes concerns to consider?). We can do this for other things too.
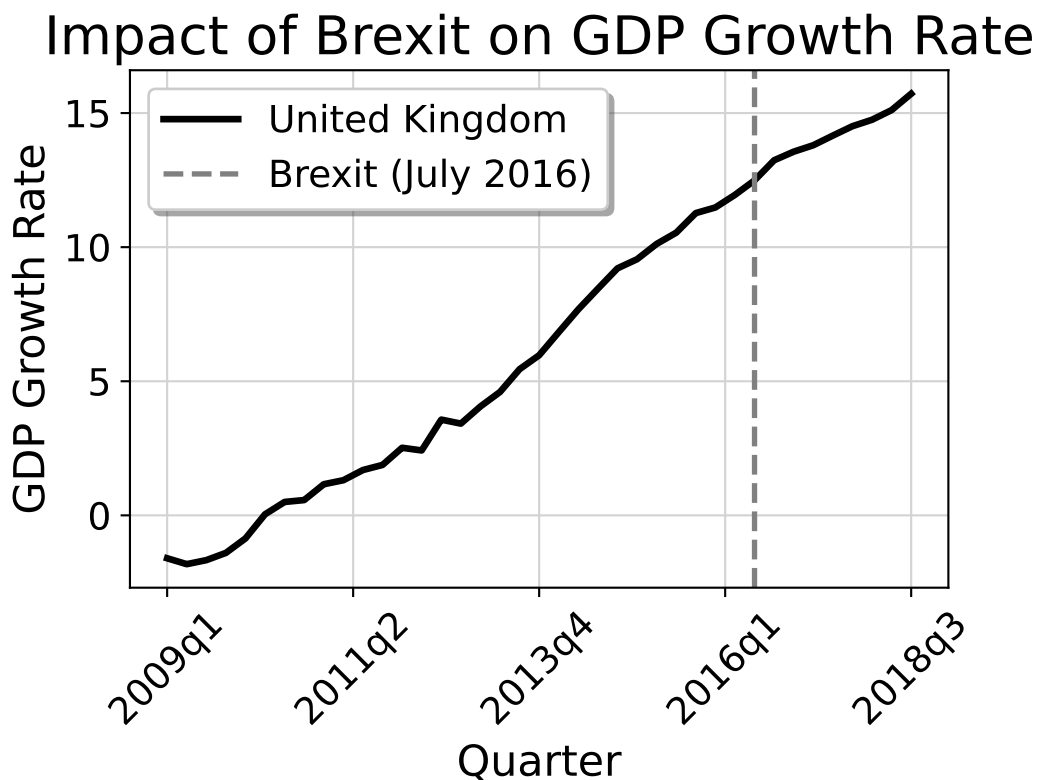
This is a plot of GDP growth rates for the United Kingdom, taken from this report. The authors were concerned with the impact of Brexit (Britain leaving the European Union) on the GDP in the U.K. economy. The effects of leaving the E.U. is also an empirical question: we can obtain data, as these authors did, about GDP growth rates across a variety of economies/countries. We can also hypothesize about why Brexit may have affected GDP growth. We can compare the U.K. to other areas which didn't have similar economic shocks at the time and try to estimate, using statistics and data analysis, how British GDP would have evolved otherwise.

4 Steps of Data Analysis

Broadly speaking, we can think of data analysis being broken into 5 distinct concepts. I summarize them below, but these chapters will comprise the rest of the notes

1. Identifying Policy Problems

2. Gathering Data

3. Cleansing Data

4. Analyzing Data

5. Presenting Results

Identifying Policy Problems

As we've discussed above, the first step in this process is simply asking questions. What kind of questions? Policy questions. Knowing what specific questions to ask though can be tricky. Policy is a giant field. Of the thousands of questions we could ask, how do we know which ones will be the most pressing or timely? In other words, how do we know that this is a problem that policy *needs* to be enacted for? How can we identify programs whose analysis benefits the citizenry or other interested parties? Put simpler, who cares? Why do we want to do this study or answer this question? Who stands to benefit?

Gathering Data

Even once we've identified the problem, how do we go about gathering real data to answer questions? If we can't get data that speaks to the issues that we're concerned about, we can't obtain answers that are useful.

Cleansing Data

In real life, datasets do not come to us wrapped in a pretty bow ready for use. Cleaning data (or organizing it) can be a very messy affair in the best of times. In order for us to answer our questions, the data we obtain must be organized in a coherent way such that we can answer questions at all. If your data are not sorted correctly by place and time, trust me, the plot you'll get will not just look terrible, but you can't glean any trends or patterns from it. This and the next three bullets will comprise most of this class's material.

Analyzing Data

Next, we do analysis. We apply statistical analysis in order to answer the questions we're asking, using the dataset we've now cleaned. Such techniques can range from simply descriptive statistical analysis to complex econometric models.

Presenting the Results

Now that we've done analysis, we can finally interpret what the findings mean. We attempt to draw conclusions based on our results and come up with avenues for future research or other elements of interest.