Lies, Damned Lies, Statistics

Nobody asked but …

My latest Words Poorly Used blog made reference to a WAG (wild ass guess) number that attempted to purport that a large percentage of undocumented immigrants did not cross a border on the ground, but rather flew into the US, and then stayed after the visa expired. This has several takeaways, but the one that sticks with me is the problem with statistics.

I was struck by the visa overstaying assertion, so I investigated the backstory. It develops that no federal office keeps data in a way to determine this phenomenon precisely. We really don't know how many of the perhaps 12 million undocumented immigrants are truly visa overstayers. This is not a number that the feds want to tout. The lack of properly designed data is the scourge of the social sciences (ironically, this does not lessen society's hunger for social engineering). Bureacrats and pundits seem to like the loosey-goosey approach in statistics so that they can stretch or shrink conclusions to suit the occasion.

The likelihood that data is well-designed before the fact to prove some hindsight observance is very small.

I don't mean to rustle any jimmies, but cimate change analysis, as an example, has this problem. What is the likelihood that the design for data collection in 1900 was exactly appropriate for findings in 2018? In the USA, there has been a massive redistribution of patterns of urbanization in that same time period.

— Kilgore Forelle