DATA FEMINISM

You have **2** free member-only stories left this month. Sign up for Medium and get an extra one

5 Questions on Data and Security with Valerie Hudson

Catherine D'Ignazio Follow
Apr 15, 2020 · 9 min read ★

By Catherine D'Ignazio with editing by Izii Carter

Image courtesy of Valerie Hudson

Dr. Valerie Hudson is the director of Texas A&M's Program on Women, Peace, and Security. She is the co-author (with Andrea M. den Boer) of *Bare Branches: Security Implications of Asia's Surplus Male Population*, which received two national book awards as well as coverage from major news outlets. Gloria Steinem named her book *Sex and World Peace* (co-authored with Bonnie Ballif-Spanvill, Mary Caprioli, and Chad F. Emmett) as one of the top three books on her reading list in 2016. Hudson has also developed a nation-by-nation database on women, known as the WomanStats Database, which aims to be "the most comprehensive database on the situation and status of women in the world."

Lauren Klein and I include the WomanStats Database in "The Numbers Don't Speak for Themselves" chapter of *Data Feminism* as an example of how to address issues of power and inequality in data collection environments. There, we pay particular attention to not only the important information gathered in the WomanStats Database, but also to the comprehensive codebook accompanying it. Crucially, this codebook notes how and why certain data points — like those concerning sexual violence — may be skewed due to imperfect reporting measurements that reveal bias. As we write in the book, "WomenStats models how context can also include an analysis of unequal social power."

I spoke with Hudson about WomanStats last year. What follows is an excerpt of that conversation. It has been edited for clarity.

How did your interest in women, gender and statistics emerge and lead you to the WomanStats Project?

I am the co-founder of the WomanStats Project. We founded it in 2001 — myself and Professor of Geography, Chad Emmett — and it came off of a previous project I had done as the co-author of a book called *Bare Branches: The Security Implications of Asia's Surplus Male Population*. In *Bare Branches,* I think we were the first to suggest that there were in fact some security implications when you culled up to 15% of the females from the population — that this might not lead to stability in society. It was an early foray in the field of International Relations towards asking whether what was going on with women affected national and international security.

That project was stunningly successful. It went viral. Even to this day you'll see people either quote us or make our assertions indirectly, which is a compliment — for that to be seen as part of the general knowledge out there. And so, I was interested in whether

things besides culling females — that is, other forms of oppression or subordination or violence against women — were also related to national, perhaps international, instability and conflict. Of course, that meant that we were going to need to collect indicators of women's situations, security, and status.

And so we began. Small at first: we began with an Excel spreadsheet, 27 variables, two research assistants. Now, we have an online Python architecture database which has over a quarter of a million data points, over 350 variables, another two dozen scales. We have online mapping capabilities. It's immense. We actually have more data on women than the UN does.

Do you share the data?

Our data is completely free and accessible online. You can just go and find our data, find our scales, and so forth. We make people register just so we can keep track, and we have well over 9,000 registered users of the database. And they're from all over.

They can download an Excel spreadsheet to their desktop that has all these scores in it. So we try to make it very accessible to researchers, but also to other people who are interested in what's going on with women in a particular country.

Is this project ongoing?

The database is still being updated. We have a dozen research assistants who are adding data to the database every single day. In fact, on our homepage you'll find a feed that contains the last three pieces of data that were uploaded. So, yes, it's ongoing.

Obviously, when we have a particular research project going on, we can also do focused searches for data to upload into our database. So, from 2014 to 2018, we were operating under a grant by the US Department of Defense. They've given us about $1.5 million to focus on 11 specific variables we felt were absolutely associated with instability and conflict.

We made a really big push on those 11 variables. And some of those were variables nobody else has data on, prevalence of patrilocal marriage, bride price and dowry practices, prevalence and legality of cousin marriage, property rights not only in the law but property rights in practice, and others. We had a very interesting 11 point scale there that took an immense, multi-year data compilation and then the data scaling effort.

If you go to our database, you'll see that in essence we have four different types of information. We have what are called practice variables, law variables, data variables, and scale variables. And the first two are all qualitative. If you want to do research on women, you have to embrace qualitative data. There's no two ways about it, because the reality of women's lives, you know, is simply not captured in quantitative statistics. Absolutely not.

Furthermore, as we've seen, a lot of this qualitative data is absolutely critical. For instance, practice variables, which are information that's gathered about the reality for women on the ground: if they report a rape, are they going to be killed, are they going to be divorced, will the police rape them, will they have to provide food for their rapist if they want the rapist to stay in jail. That happens.

And then we have law variables. What is the law on rape? What's the age of consent? What kind of proof do you need? Is a woman's testimony worth as much as a man's?

And then we have data variables, which are any quantitative or even quasi-quantitative data about the prevalence of a phenomenon. So, if we have a sentence that says, you know, the overwhelming majority of women in Egypt have had their genitals cut, we'll take that because we're desperate.

And then we have the scales, which are our own creation. We have a few dozen. Some are what we call univariate scales that look at one particular phenomenon like rape. Others are multivariate scales. For example, we have a scale of the physical security of women that looks at murder and rape and all sorts of things. To implement a scale, we first train our coders on our coding rubric and assign two coders to code the data independently. So they'd be looking at the pertinent practice variables, or law variables, or data variables. They will independently come up with an initial scaling and then we will reconcile the two coders' results. That is, we'll look for cases where the two coders have disagreed and appoint a co-PI who will make a decision, and that usually means we update the coding rubric to reflect the decisions that are made.

And then we map our scales. We have these beautiful full-colored maps of all of our ordinal scales.

What's the landscape of gender in equity measures?

Well, since starting WomanStats, I think we've seen many new indices come forth, and I am totally excited and happy about that. I think there'll also be more in the future, so, for example, last year, Georgetown came out with a Women, Peace and Security Index.

Some of these indices are, well, kind of niche-y. But others, like the GII (Gender Inequality Index) or the GGI (Gender Gap Index), are multivariate. They'll have a variety of subcomponents. They're attempting to give you an overall view of how unequal women are in society.

There are also standard data sources. For example, the World Bank has what they call Gender Stats, which give you things like female labor force participation and other economic-related variables, as well as a fantastic database called Women, Business and the Law, which looks at laws that may either help or hurt women who are trying to be economically active within their society. Like, can women get a loan? Is it legal for them to get a loan without their husband's permission? These kinds of things. But there's drawbacks with each of these. The UN data and the World Bank data are all quantitative. So if you wanted to know something about how women are treated — if they've been raped, which is an example we used before — there's nothing. There's nothing.

Then there are other scaling efforts that are really opaque. We were just really thrilled that SIGI, which is a data arm to the OECD, came out with a scale of polygyny and we were like, "Yes, hurrah!" And then we looked into the data and what we found was that the data was coded as 0, 0.25, 0.75 and 1.0. What do the numbers mean and where were those numbers derived?

We spent a full day looking through everything, trying to find where they got this data. What do these numbers mean? Well, it turns out the numbers don't mean anything. They were just a simple fourfold gradation and they consulted country experts, whom they did not name, and they did not have coding rules (at least none that the end-user could access). So, it was simply up to the country expert to say whether they felt it was like a zero or they felt it was like a 0.25 or 0.75. That was… What does that even mean, right? And how do we know that the same, you know, understanding of what those numbers mean was used by each of the country experts? No clue. So, with that kind of lack of transparency, you, how do you even use the data? You can't. You can't use the data.

So, one of the things that we've tried to do is be radically transparent. If you open up the codebook and look at the scaling rubrics, we give you the entire scaling rubric. You've got a whole thing. You know exactly what we did and the data that we drew upon in our database. We only code what's in our database. So, you know what we did.

So much of the time data about women and gender is depressing and like a lot of times the things that we measure about women are depressing. What gives you hope about gender equality in the world?

We wrote a book called *Sex and World Peace*. It was the first real WomanStats Project book, in which we bring data to bear on the quest to know what is the price of inequality between men and women in the world. What's the price in terms of security, stability, prosperity and so forth? And, you know, I think why I am such a fan of data feminism is — as we all know — if you can't see it, you can't fix it. So, how do you see it? Well, you see it with data.

I remember when I first started doing work on women, people would say, "Oh, that's very nice, but it's just an anecdote. Come back when you have real data." Okay, you want real data? Take a look at this database on the status of women worldwide. We really think our Project and other similar efforts have really changed the discussion. We now have UNSCR 1325 (*UN resolution on Women, Peace and Security*), which I don't think would have come about if there hadn't been some data. We have an explosion of empirical feminism in International Relations now.

Dara Kay Cohen of Harvard, for example, now has a database on the use of sexual violence by rebel groups. Such a database would have been unthinkable in the last century. No one really cared and no one would have even known how to get the data. So, I think as more and more scholars and activists can collect data, women can't be ignored anymore. What's going on with women can't be ignored.

One of my favorite stories is about the group HarassMap in Egypt. When the Egyptian uprising overthrew Hosni Mubarak, this enterprising grassroots group created a cell phone app so that women could in real-time report if they were being assaulted in Tahrir Square. And they produced instantaneous maps of where these things were happening. They were able to mobilize supporters who would go in and extract the women. It's just really amazing stuff.

So, I am an absolute believer that if you can make women's experiences into data — which doesn't have to be numbers — you can create action. And that's why I am still doing this 17 years later.

Dr. Hudson recently wrote an op-ed for the New York Times*: "What You Do To Your Women, You Do To Your Nation." Her latest book, co-authored with Donna Lee Bowen and Perpetua Lynne Nielsen is* The First Political Order: How Sex Shapes Governance and National Security Worldwide *(February 2020, Columbia University Press)*.

Data Science Women Statistics Research Feminism

About Help Legal

Get the Medium app

Download on the App Store GET IT ON Google Play