Fwd: [hmdc.harvard.edu #288001] juliavoo@hks.harvard.edu: Cyber Power Framework - data science support to queue researchconsulting

1 Nachricht

Worthington, Steven <sworthington@iq.harvard.edu> 6. April 2020 um 11:30
An: "Schwarzenbach, Anina" <anina_schwarzenbach@hks.harvard.edu>

~~~~~~
Steven Worthington, Ph.D.
Senior Data Scientist
Manager of Data Science Services

Institute for Quantitative Social Science
Harvard University
1737 Cambridge Street, K303

sworthington@iq.harvard.edu
https://scholar.harvard.edu/sworthington/home
~~~~~~

Begin forwarded message:

From: Steven Worthington <sworthington@iq.harvard.edu>
Subject: Re: [hmdc.harvard.edu #288001] juliavoo@hks.harvard.edu: Cyber Power Framework - data science support to queue researchconsulting
Date: April 1, 2020 at 11:47:47 PM EDT
To: researchconsulting_support@help.hmdc.harvard.edu

Hi Anina,

I'll offer some first thoughts on the questions you listed, but we should probably schedule a time to chat about this next week via Zoom.

1. concerning the values of the indicators:
- some of our indicators have a lot of zeros (see e.g., "state_attack"), can we still use those to create an aggregated index?

SW: it depends - are these binary variables, ordinal, counts, or continuous? How much is "a lot"? Is there still variability?

- some of our measures (e.g., e commerce and patent application) are affected by the size of the pop of a country. We have taken the per capita value, is that correct?

SW: yes, definitely.

- we have some indicators (not included in the dataset yet) that are binary (0, 1). Can we include those in the aggregated index?
- If not, what is the minimum range of values an indicator should have so that there is not problem with the aggregation (e.g., through factor analysis)

SW: it's possible to include them - depending on the method you use (see below).

2. concerning the normalisation of the indicators
- we have used the default (z-score), would you suggest using another technique (e.g., min-max)?

SW: well, some of them are binary, some ordinal, and some (maybe?) are counts, right? So those won't be standardized. What are you trying to achieve by standardizing here? If you end up using some factor analysis related method, then that will be based on a correlation matrix, which is standardized to unit variance.

- we are not always sure about the polarity of our indicators (i.e., whether they relate positively or negatively to cyber power). Is there a way to assess this empirically ?

SW: not specifically. You can visualize them.

3. concerning the aggregation of the indicators
- the compind package lists a variety of methods to aggregate the indictors into an index, e.g., though factor analysis or min-max (see https://cran.r-project.org/web/packages/Compind/Compind.pdf).
Which method do you think is more suitable?

SW: none of the above can deal with binary, count, ordinal, and continuous indicators all together. You may have to look at something like Latent Class Factor Analysis (a type of finite mixture model), or perhaps a traditional factor analysis based on a heterogeneous correlation matrix (polychoric, Pearson, tetrachoric, etc.). But, from your knitted html file, it seems like you'll have 20 variables and only 28 observations? In which case, pretty much all of these methods are unavailable to you - for example, for traditional factor analysis, you'd want a ratio of at least 10 observations to each variable; so at least 200 observations to get reasonable results.

- some of our indicators correlate highly (.8), should we exclude those when building the aggregated index?

SW: the point of building an index is to aggregate correlated variables - so the more highly correlated the better. It's the variables that have low correlations that you should be worried about. A low correlation probably indicates that that variable is tracking a different construct and therefore your set of indicators is multi-dimensional. If so, that's a reason not to create one index, but several.

Best,
Steve

-----Original Message-----
From: Steven Worthington via RT <researchconsulting_support@help.hmdc.harvard.edu>
Sent: Wednesday, March 25, 2020 2:47 PM
To: Voo, Julia <juliavoo@hks.harvard.edu>
Subject: Re: [hmdc.harvard.edu #288001] juliavoo@hks.harvard.edu: Cyber Power Framework - data science support to queue researchconsulting

<See all of this ticket: https://help.hmdc.harvard.edu/Ticket/Display.html?id=288001 >

Hi Julia,

thanks for reaching out to us. My name is Steve Worthington and I manage the DSS team at IQSS. We'd certainly be happy to help out if we can. Is there something specific you want help with?

Best,

Steve

On Mar 25, 2020, at 11:33 AM, Jeremy Yuenger via RT <researchconsulting_support@help.hmdc.harvard.edu> wrote:

Transaction: Queue changed from CGIS to researchconsulting by jyuenger

Queue: researchconsulting

Subject: Cyber Power Framework - data science support

Ticket <URL: https://help.hmdc.harvard.edu/Ticket/Display.html?id=288001 >

The last correspondence on this ticket was as follows:

Hi Data Science Team,

My team has been busy creating a framework for measuring cyber power at the country level. I'm the Research Director and the Faculty lead is Eric Rosenbach, Co-Director of the Belfer Center.

Anina in cc, has built a framework on R and has been helping us to ensure the methodology, indicators and data collection are up to scratch. However, we'd really welcome your expertise here. This is the first comprehensive framework for measuring cyber power and we're sure it is going to cause a bit of a splash so definitely want to make sure our research can take the scrutiny! We'd want that regardless of the scrutiny tbh.

Currently, we have about 9 out of 30 indicators complete and the rest of the team are working on completing the rest over the next two weeks. We want to go live with this project in about a month.

Would you be able to help us out? Hope you're all healthy and staying safe,

Best wishes,

Julia

Julia Voo

Research Director | China Cyber Policy Initiative

Belfer Center for Science and International Affairs

Harvard Kennedy School of Government

I work and send emails at random times, you don't have to!

<compind_explore.html>