**Notes for customs evasion, descriptive analysis**

**Notes on R files:**

**Comtrade\_clean.R:** Creates yearly data matching up by importer exporter, i.e. Country A says they exported to country B, country B says they received imports from Country A. Notes:

* This is only for the hs12 product classification, and thus for years 2012 – 2016.
* I had to cut the data into parts to keep it from getting too large, those cuts are saved on Michael’s computer.
* The output is matching data for each year of 2012 – 2016. These datasets are saved in the analysis folder, for ex: hs12\_12 is all 2012 data.

**Summary\_clean.R:** Cleans data for example run of data with commodity codes starting with 0. This isn’t really relevant anymore.

**Summary\_qty.R:** Creates and summarizes qty.dta. Qty.dta is observations with matching export import pairs on `netweight (kg)` that aren’t missing from one reporter. Summary results are in summary\_qty.pdf.

**Summary\_qty\_with\_tariffs.R:** Summarizes quantity data that also have tariff data from WITS. See “summary\_qty\_with\_tariffs.pdf”.

**Summary\_value.R:** Creates and summarizes value.dta. Value.dta is observations with matching export import pairs on `value` that aren’t missing from one reporter. Summary results are in summary.pdf.

**Summary\_with\_tariffs.R**: Summarizes value data that also have tariff data from WITS. See “summary \_with\_tariffs.pdf”.

**WITS\_clean:** Creates file called hs12\_all\_tariffs.Rda, which is the matching comtrade data and corresponding tariff rates from WITS. I had to split the data and recombine it to keep it from getting too big. The splits are saved on Michael’s computer.

**Notes on data sources:**

**Comtrade:**

* Some countries report their data in older HS classification versions. Nigeria’s most recent classification used is hs07, Phillipines’ is hs02.  UN converts backwards, that is, they take hs12 reporters and convert to older versions for consistency’s sake, but they don’t take hs07 data and convert to hs12. This is why it looks like Nigeria and the Philippine’s have no data in the summary PDFs. See here for more details: <https://comtrade.un.org/db/mr/daReportersResults.aspx?bw=P>

* Although we only summarized the hs12 data, previous versions are also saved in the dropbox folder.
* Helpful read me from the UN site: <https://comtrade.un.org/db/help/uReadMeFirst.aspx>.

**WITS:**

* We used WITS because you can bulk download their tariff rates, as opposed to the per day/year limit at macmap.
* The data saved is the "including AVE" option from the WITS website, which is when raw data not in ad valorem format is converted to ad valorem.
* There are several countries in the non - "including AVE" data on the WITS site, bur who don’t have “including AVE” data. I emailed the WITS contact about why this is the case, but have yet to get a response. My best guess is that if a country is not in the "including AVE" and in the other bulk download option, then the data has not been converted as AVE. The data in the non-AVE option therefore is ad valorem data, but only if that's how the country reported it. In other words, the World Bank/UN hasn't converted raw data to AVE for these countries, and it's only showing the reported-ad valorem data, and dropping the rest. Since this would be inconsistent with the "including AVE" data, I focused only on the AVE-converted data.
* WITS has data on both MFN and preferential rates. I used MFN unless there was a corresponding preferential rate listed.
* Many of the preferential rates are for a group of countries. The codes for these groups are in the TRAINS\_preference\_beneficiaries.Rda file on dropbox. EU countries are listed as one entity.

More information on the data here: https://docs.google.com/document/d/1ENk7QyufYnHeGN0l1RCXQix62ly8h9iYJz-PW0T6a4Q/edit