Agricultural trade data processing

July 9, 2015

Last update: 2015-08-11 18:02:49

Raw trade data

FAO receives data on trade flows from United Nations Statistical Division. The division runs Commodity Trade Statistics Database UN Comtrade "It stores standardised official annual trade statistics reported by countries and reflecting international merchandise flows detailed by commodity and partner country with coverage reaching up to 99 percent of world merchandise trade[1]".

One can freely download this standardised statistics from the open data base. Statistical Division of FAO gets unstandardised data.

Table 1: Random sample of import trade flows of 2011 year, reported by the US

year	reporter	partner	hs	flow	weight	qty	qunit	value
2011	842	250	9015900090	1	-	-	1	786417
2011	842	792	5303900000	1	19731	19731	8	29165
2011	842	690	8544429000	1	-	-	1	69341
2011	842	360	9403509045	1	-	517457	5	68363704
2011	842	156	5515110030	1	635	7405	2	7947
2011	842	392	6601990000	1	-	427	11	19170
2011	842	554	8481809005	1	-	80	5	51787
2011	842	214	8473509000	1	-	-	1	2197
2011	842	344	5309110090	1	1428	6502	2	27379
2011	842	208	0306130027	1	29740	29740	8	109695

This is an example of unstandardised data on trade inflows in 2011, reported by the United States. Reporters and trade partners are represented with three-digit numerical codes used by the Statistics Division of the United Nations. Trade commodities are classified with extended Harmonized Commodity Description and Coding System (HS)[2] maintained by the World Customs Organization[3].

Weight is measured in kilograms and value in US dollars. Quantity (qty column) is an optional alternative for weight. It could be measured in different units (qunit column). See full list of possible units and their descriptions in Annex I of Quantity and Weight Data in UN Comtrade[4].

Country-specific HS commodity codes

Harmonized system classification is declared by WCO up to 6 digits. A country may extend HS to more detailed level to better respond to local circumstances. Let's compare differencies in codes under subheading

[1] http://unstats.un.org/unsd/comtrade_announcement.htm Comtrade Announcement

[2] http://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx What is the Harmonized System (HS)?

[3] http://www.wcoomd.org/en.aspx World Customs Organization

[4] http://unstats.un.org/unsd/tradekb/Knowledgebase/Quantity-and-Weight-Data-in-UN-Comtrade Quantity and Weight Data in UN Comtrade

0202 Meat of bovine animals, frozen between the US and Brazil[5].

Table 2: Extension of HS codes by the US

hs	Description
02	CHAPTER 2 MEAT AND EDIBLE MEAT OFFAL
0202	Meat of bovine animals, frozen:
0202.10	- Carcases and half-carcases:
0202.10.05	- - Described in general note 15 of the tariff schedule . . . :
0202.10.05.10	- - - Veal
0202.10.05.90	- - - Other
0202.10.10	- - Described in additional U.S. note 3 to this chapter . . . :
0202.10.10.10	- - - Veal
0202.10.10.90	- - - Other
0202.10.50	- - Other:
0202.10.50.10	- - - Veal
0202.10.50.90	- - - Other
0202.20	- Other cuts with bone in:
	- - Described in general note 15 of the tariff schedule . . . :
	- - - Processed:
0202.20.02	- - - - High-quality beef cuts
0202.20.04	- - - - Other
0202.20.06	- - - Other
	- - Described in additional U.S. note 3 to this chapter . . . :
	- - - Processed:
0202.20.10	- - - - High-quality beef cuts
0202.20.30	- - - - Other
0202.20.50	- - - Other
0202.20.80	- - Other
0202.30	- Boneless:
	- - Described in general note 15 of the tariff schedule . . . :
	- - - Processed:
0202.30.02	- - - - High-quality beef cuts
0202.30.04	- - - - Other
0202.30.06	- - - Other
	- - Described in additional U.S. note 3 to this chapter . . . :
	- - - Processed:
0202.30.10	- - - - High-quality beef cuts
0202.30.30	- - - - Other
0202.30.50	- - - Other
0202.30.80	- - Other

Table 3: Extension of HS codes by Brazil

hs	Description
02	CHAPTER 2 MEAT AND EDIBLE MEAT OFFAL
0202	Meat of bovine animals, frozen:
0202.10	- Carcases and half-carcases
0202.20	- Other cuts with bone in:
0202.20.10	- - Forequarters

[5] http://madb.europa.eu Descriptions of country-specific HS-codes are provided by Market Access Database and copyrighted by Mendel Verlag, Germany.

hs	Description
0202.20.20	- - Hindquarters
0202.20.90	- - Other
0202.30	- Boneless

The set of HS-codes from the US is wider, than Brazilian one. For boneless meat Brazil doesn't extend standard code 0202.30, when the US use here seven additional codes.

Country codes

Codes of reporters

Area codes of reporters are standardized by the Statistical Department. The SD follows *in general* the United Nations Standard Country or Area Codes for Statistical Use[6]. The code scheme used by the SD[7] is slightly modified from the official one[8]. For example the official scheme offers code 840 for the US, when the modified version uses 842.

Codes of partners

Partners' codes in Tariffline data are not standardised and presented as they were reported by countries. Reporters can use as standard version of codes, so the version of the Statistical Department. For example, in Tariffline data there are 27 country codes which are not presented in official scheme and 40 codes not covered by the modified version.[9]

Initial validation of trade data

At prevalidation step we are to make a decision should we accept data from a specific country for the further processing or not. A country could provide data of good quality for one part of commodities and inadequate level of quality for another part. We want to estimate quality differences between commodities of a country.

Quality of data is estimated by following indicators:

- Share per cent of missing quantities
- Share per cent of unit value outliers

Self-trade

There are cases when a country reports itself as a partner to exports or imports. Such situations can occur due to mistakes or when an entrepôt exists.

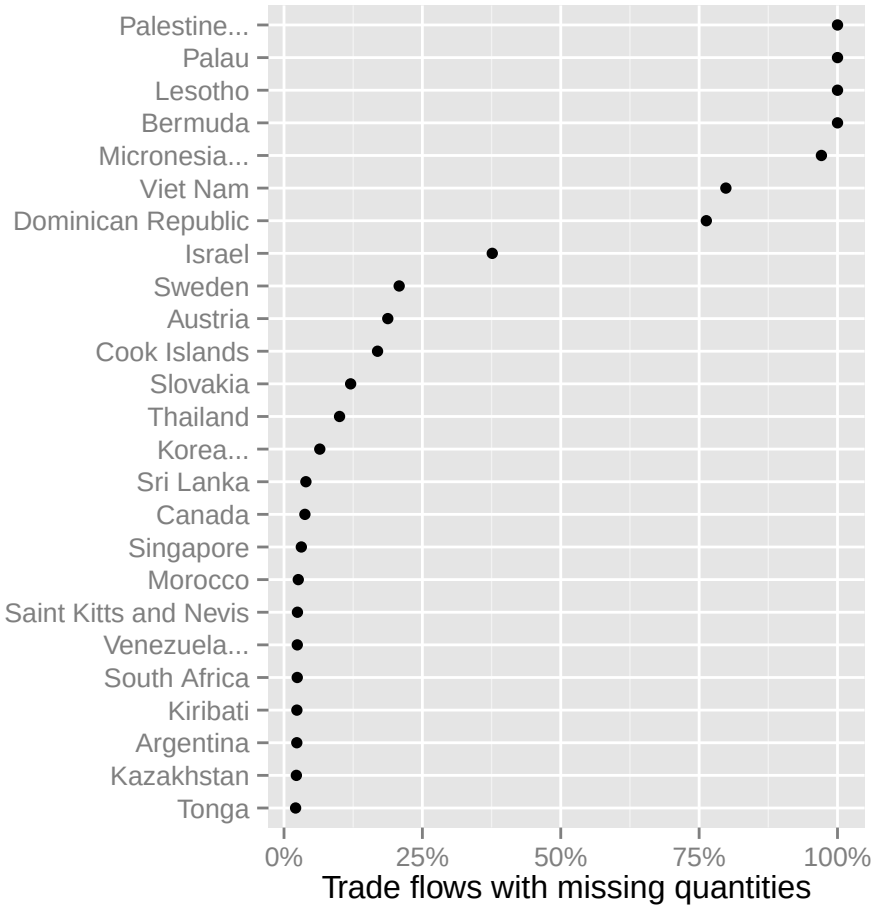
[6] http://comtrade.un.org/pb/ The United Nations Statistics Division (2015). The 2014 International Trade Statistics Yearbook, Volume I - Trade by Country, xix.
[7] http://comtrade.un.org/data/doc/api/ The UN Comtrade data extraction API
[8] http://unstats.un.org/unsd/methods/m49/m49alpha.htm Countries or areas, codes and abbreviations
[9] http://rpubs.com/malexan/m49 Matrunich A. (2015). M49 country codes in Tariffline

Table 4: Self-trade of commodities from 2nd, 10th and 15th HS chapters in 2011

Reporter	Flow	Total
France	Import	163
Canada	Import	56
Portugal	Import	50
Slovakia	Import	50
New Zealand	Import	31
South Africa	Import	27
United Kingdom	Import	17
Slovenia	Import	16
Estonia	Import	12
Thailand	Import	12
China	Import	7
Greenland	Import	3
Ethiopia	Import	2
Papua New Guinea	Import	2
Saint Kitts and Nevis	Import	2
Indonesia	Import	1
Malaysia	Import	1
Palau	Import	1

Missing quantity

We identify which reporters provide data of insufficient quality. Firstly for every reporter proportion of trade flows with missing quantity is calculated.

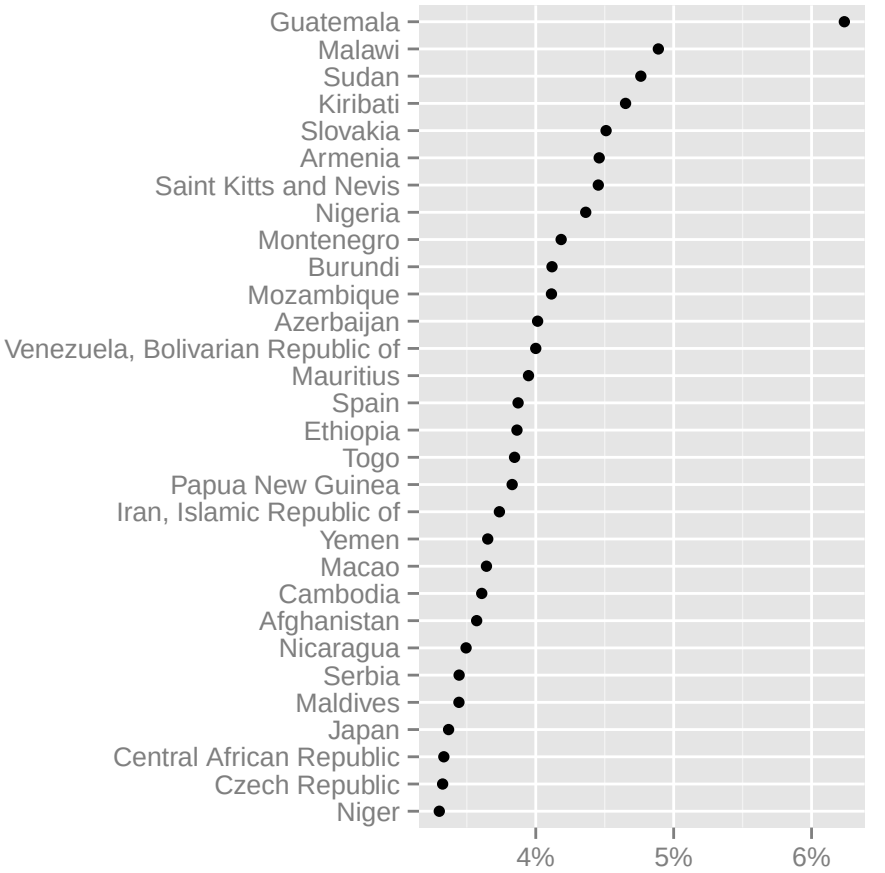


Detection of outliers

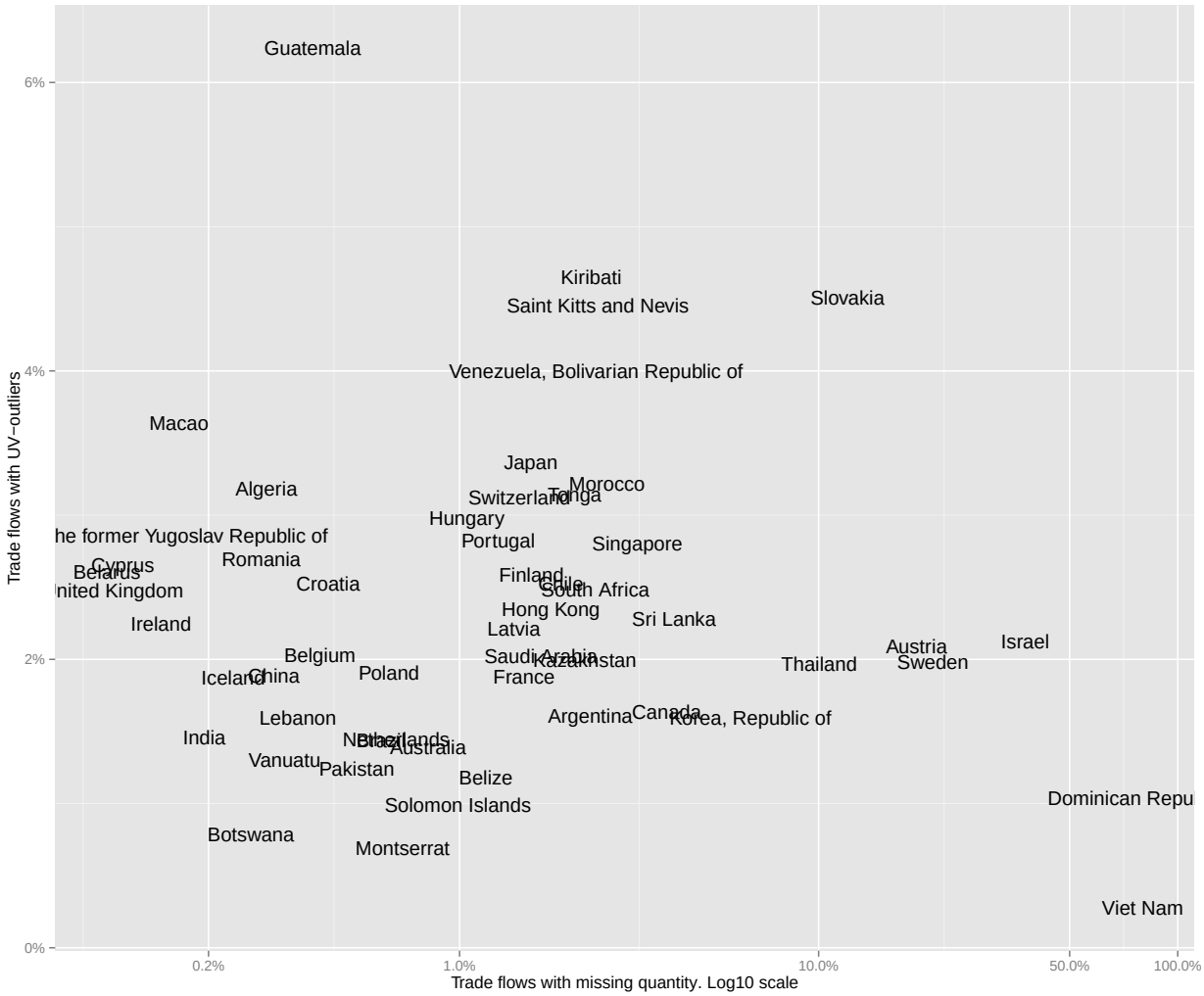
We define outliers as observations located outside the range:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$$

where Q_1 and Q_3 are the lower and upper quartiles respectively, and k is a non negative constant. In this paper we use k equals 1.5.


Guatemala
Malawi
Sudan
Kiribati
Slovakia
Armenia
Saint Kitts and Nevis
Nigeria
Montenegro
Burundi
Mozambique
Azerbaijan
Venezuela, Bolivarian Republic of
Mauritius
Spain
Ethiopia
Togo
Papua New Guinea
Iran, Islamic Republic of
Yemen
Macao
Cambodia
Afghanistan
Nicaragua
Serbia
Maldives
Japan
Central African Republic
Czech Republic
Niger
4%
5%
6%
Trade flows with unit value outliers

Missing quantites and outliers combined



Imputing of missing quantities and replacement of outliers

In data reported by USA for 2011 year in HS chapters 2, 10 and 15 there are 0 trade flows with missing quantity and 33 trade flows with UV-outliers.

Table 5: Example trade flows with outlying unit values

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	Netherlands	Import	1505001000	280	4124	14.73	3.92
United States	Japan	Import	1516100000	1090	93260	85.56	27.54
United States	Malaysia	Import	1515906000	2	4061	2030	22.26
United States	France	Import	1514190000	311	59170	190.3	2.72
United States	Japan	Import	1008100000	810	7833	9.67	1.24
United States	Spain	Import	0203294000	10087	194271	19.26	4.76
United States	United Kingdom	Import	1521100040	84	6181	73.58	6.89
United States	Canada	Import	1509102000	58	2386	41.14	5.93

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	Spain	Import	1512110040	1140	6376	5.59	0.92
United States	Italy	Import	1515300000	34	2334	68.65	4.57

Imputing using reporter median unit values

Now we correct weight of this outlying trade flows with help of median reporter unit value for a given commodity.

$$Weight = \frac{Weight}{UV_{reporter}}$$

Table 6: Example trade flows with corrected weight

Reporter	Parnter	Flow	HS-code	Weight, kg	Corrected, kg	Difference, kg
United States	Netherlands	Import	1505001000	280	1052	-772
United States	Japan	Import	1516100000	1090	3386	-2296
United States	Malaysia	Import	1515906000	2	182	-180
United States	France	Import	1514190000	311	21754	-21443
United States	Japan	Import	1008100000	810	6317	-5507
United States	Spain	Import	0203294000	10087	40813	-30726
United States	United Kingdom	Import	1521100040	84	897	-813
United States	Canada	Import	1509102000	58	402	-344
United States	Spain	Import	1512110040	1140	6930	-5790
United States	Italy	Import	1515300000	34	511	-477

Imputing using data from trade partner (mirroring)

Another approach to impute missing or outlying quantities of a reporter is to use mirror data from trade partner. Let's check are there any such trade flows related to wheat among reported by the US.

Year	Trade partner	Flow	HS-code	Weight, kg	Value, $US	Unit Value	UV median
2011	Japan	Import	1001902096	61	46362	760	1.466

Outlier detection algorithm shows, that the price (unit value) in this trade flow differs too much from the median price of trade flows of this commodity, reported by the US: 760 $US per kg versus 1.5 $US per kg.

The commodity code *1001.90.20.96* is country-specific and is used only by the US. It is not listed in the recent Harmonized Tariff Schedule of the United States[10]. It means this HS-subheading was removed from Harmonized Tariff Schedule and had not been used any more. Panjiva website reports last use of the code was fixed in 2011 and gives description of it[11]. *1001.90.20.96* stands for wheat and meslin not mentioned in any other subheadings of *1001.90.20*.

We want to check characteristics of this trade flow from a partner's side. But Japan didn't report any export of wheat-related commodities to the US in 2011. We expand our search to all trade flows from Japan to the US with nearly the same quantity and value. We find suitable trade flow what is not reported by the US.

[10] http://hts.usitc.gov/?query=wheat Harmonized Tariff Schedule (2015 HTSA Revision 1 Edition)

[11] https://panjiva.com/trendspotting/imports/United-States/1001.90.20.96/Cereals-Wheat-and-meslin-Other-Other-Other-Other/1368 Trend report HTS Code 1001.90.20.96

Year	Trade partner	Flow	HS-code	Weight, kg	Value, $US	Unit Value	UV median
2011	United States	Export	041000000	70	36483	521.2	521.2

0410.00.000 Edible products of animal origin (not especially specified) [12].

Year	Trade partner	Weight, kg	Value, $US	Unit Value	UV median
2011	United States	70	36483	521.2	521.2
2011	Other Asia	50	26185	523.7	521.2
2011	Netherlands	150	50863	339.1	521.2

Year	Trade partner	Flow	HS-code	Weight, kg	Value, $US	Unit Value	UV median
2011	Japan	Import	1008100000	810	7833	9.67	1.241
2011	Japan	Import	1008200000	327	2479	7.581	1.882

[12] http://www.customs.go.jp/yusyutu/2011/data/e201101j_04.htm