Agricultural trade data processing

July 9, 2015

Last update: 2015-08-12 09:59:48

# Raw trade data

FAO receives data on trade flows from [United Nations Statistical Division](http://unstats.un.org/unsd/default.htm). The division runs [Commodity Trade Statistics Database UN Comtrade](http://unstats.un.org/unsd/comtrade_announcement.htm) "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]](#footnote-24)".

One can freely download this standardised statistics from the [open data base](http://comtrade.un.org/data/). Statistical Division of FAO gets unstandardised data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | reporter | partner | hs | flow | weight | qty | qunit | value |
| 2011 | 842 | 36 | 6403599045 | 1 |  | 50 | 6 | 5368 |
| 2011 | 842 | 188 | 8507308090 | 1 |  | 500 | 5 | 103702 |
| 2011 | 842 | 156 | 6208195000 | 1 | 4677 | 5339 | 11 | 1125005 |
| 2011 | 842 | 710 | 3815190000 | 1 | 51707 | 51707 | 8 | 417613 |
| 2011 | 842 | 643 | 7304292010 | 1 | 1812013 | 1812013 | 8 | 2127140 |
| 2011 | 842 | 124 | 6104391000 | 1 | 11677 | 3061 | 11 | 1501333 |
| 2011 | 842 | 376 | 2936295020 | 1 | 18904 | 18904 | 8 | 5605850 |
| 2011 | 842 | 458 | 7403290180 | 1 | 50 | 50 | 8 | 2421 |
| 2011 | 842 | 156 | 2807000000 | 1 | 15574000 | 15574000 | 8 | 975847 |
| 2011 | 842 | 124 | 8002000000 | 1 | 57172338 | 57172338 | 8 | 20127415 |

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

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](http://unstats.un.org/unsd/methods/m49/m49.htm) used by the Statistics Division of the United Nations. Trade commodities are classified with extended Harmonized Commodity Description and Coding System (HS)[[2]](#footnote-27) maintained by the World Customs Organization[[3]](#footnote-29).

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]](#footnote-31).

## 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 0202 Meat of bovine animals, frozen between the US and Brazil[[5]](#footnote-34).

|  |  |
| --- | --- |
| 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 |

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 |
| 0202.20 | * Other cuts with bone in: |
| 0202.20.10 | * + Forequarters |
| 0202.20.20 | * + Hindquarters |
| 0202.20.90 | * + Other |
| 0202.30 | * Boneless |

Extension of HS codes by Brazil

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]](#footnote-38). The code scheme used by the SD[[7]](#footnote-40) is slightly modified from the official one[[8]](#footnote-42). 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]](#footnote-45)

# 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

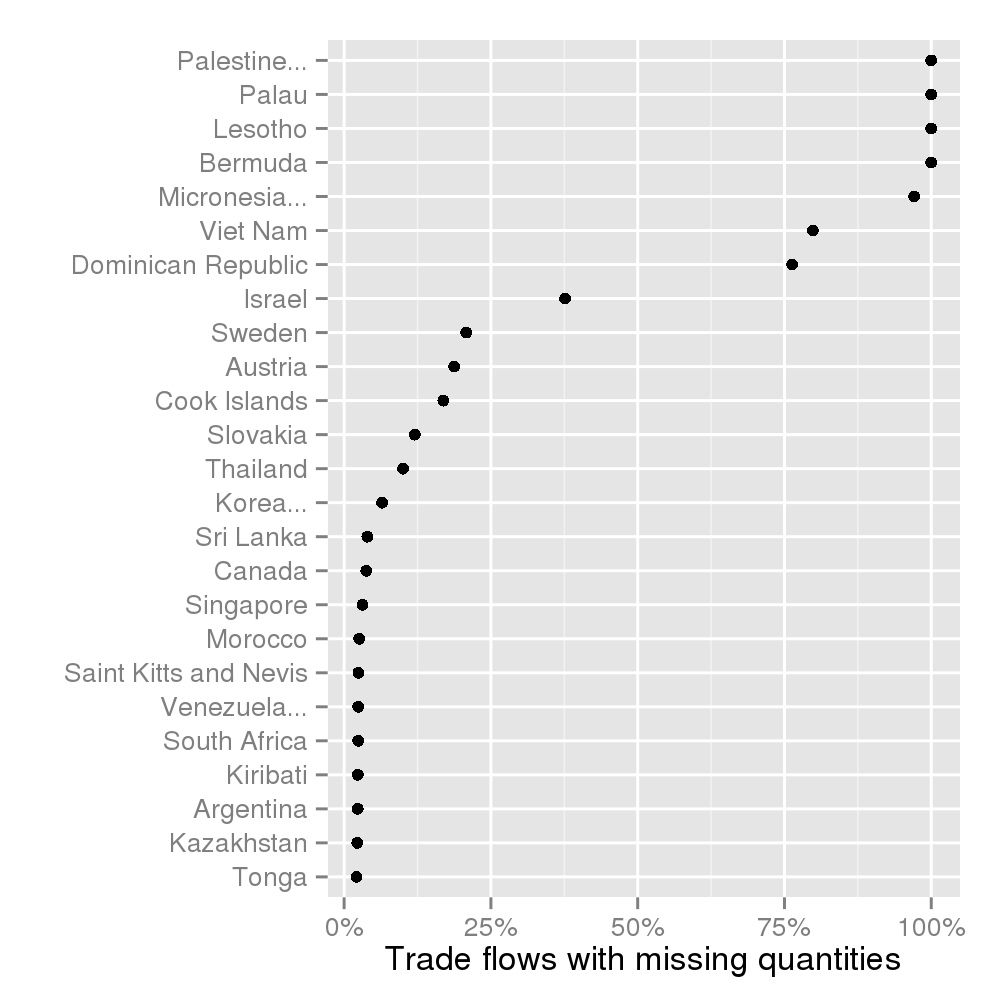
There are cases when a country reports itself as a partner to exports or imports. Such situations can occur due to mistakes or when an entrepôt exists.

|  |  |  |
| --- | --- | --- |
| 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 |

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

## Missing quantity

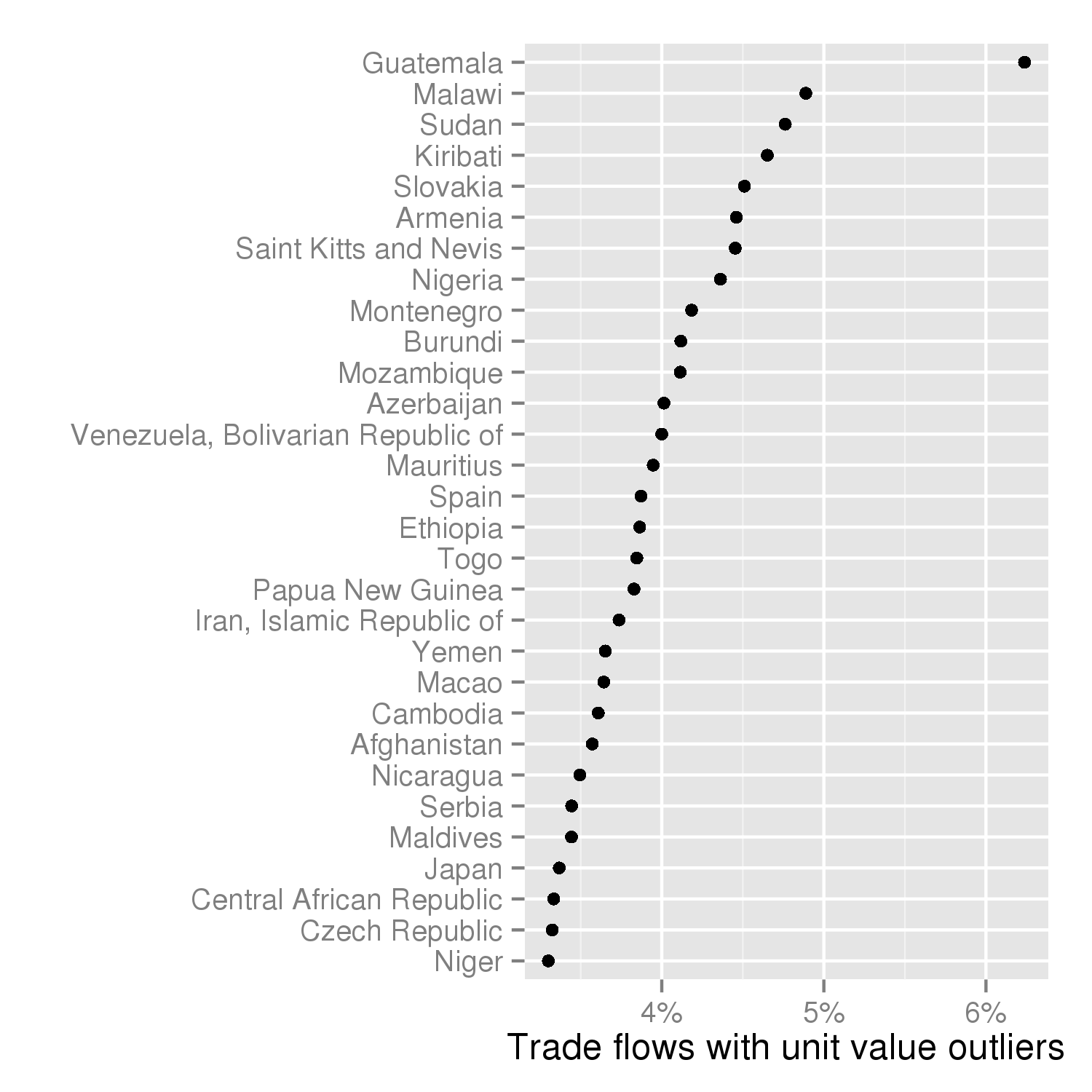
We identify which reporters provide data of insufficient quality. Firstly for every reporter proportion of trade flows with missing quantity is calculated.



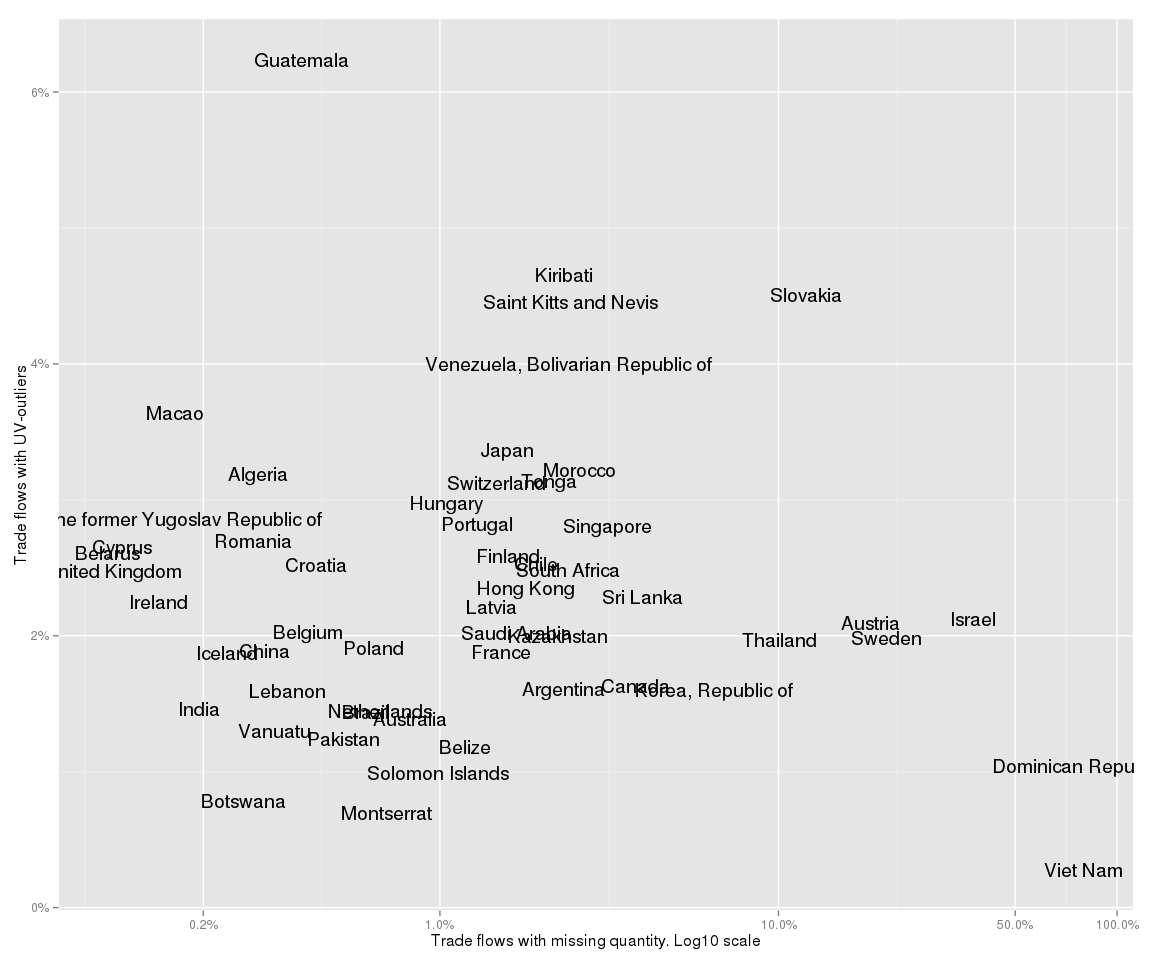
## Detection of outliers

We define outliers as observations located outside the range:

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



## 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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Reporter | Partner | Flow | Commodity | Weight | Value | UV | UV\_me |
| United States | Netherlands | Import | 1504204000 | 160571 | 4932146 | 30.72 | 5.83 |
| United States | Spain | Import | 1512110040 | 1140 | 6376 | 5.59 | 0.92 |
| United States | Japan | Import | 1515190000 | 89 | 3267 | 36.71 | 4.91 |
| United States | Canada | Import | 1509102000 | 58 | 2386 | 41.14 | 5.93 |
| United States | Uruguay | Export | 0201100010 | 278 | 4384 | 15.77 | 3.88 |
| United States | Japan | Import | 1001902096 | 61 | 46362 | 760 | 1.47 |
| United States | Japan | Import | 1006204040 | 2520 | 9679 | 3.84 | 0.89 |
| United States | Japan | Import | 1008100000 | 810 | 7833 | 9.67 | 1.24 |
| United States | Netherlands | Import | 1505001000 | 280 | 4124 | 14.73 | 3.92 |
| United States | Canada | Import | 1006309015 | 5634 | 66825 | 11.86 | 1.54 |

Example trade flows with outlying unit values

## 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reporter | Parnter | Flow | HS-code | Weight, kg | Corrected, kg | Difference, kg |
| United States | Netherlands | Import | 1504204000 | 160571 | 845994 | -685423 |
| United States | Spain | Import | 1512110040 | 1140 | 6930 | -5790 |
| United States | Japan | Import | 1515190000 | 89 | 665 | -576 |
| United States | Canada | Import | 1509102000 | 58 | 402 | -344 |
| United States | Uruguay | Export | 0201100010 | 278 | 1130 | -852 |
| United States | Japan | Import | 1001902096 | 61 | 31539 | -31478 |
| United States | Japan | Import | 1006204040 | 2520 | 10875 | -8355 |
| United States | Japan | Import | 1008100000 | 810 | 6317 | -5507 |
| United States | Netherlands | Import | 1505001000 | 280 | 1052 | -772 |
| United States | Canada | Import | 1006309015 | 5634 | 43393 | -37759 |

Example trade flows with corrected weight

## 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]](#footnote-58). 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]](#footnote-60). *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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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]](#footnote-62).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 |

1. <http://unstats.un.org/unsd/comtrade_announcement.htm> Comtrade Announcement [↑](#footnote-ref-24)
2. <http://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx> What is the Harmonized System (HS)? [↑](#footnote-ref-27)
3. <http://www.wcoomd.org/en.aspx> World Customs Organization [↑](#footnote-ref-29)
4. <http://unstats.un.org/unsd/tradekb/Knowledgebase/Quantity-and-Weight-Data-in-UN-Comtrade> Quantity and Weight Data in UN Comtrade [↑](#footnote-ref-31)
5. <http://madb.europa.eu> Descriptions of country-specific HS-codes are provided by Market Access Database and copyrighted by Mendel Verlag, Germany. [↑](#footnote-ref-34)
6. <http://comtrade.un.org/pb/> The United Nations Statistics Division (2015). The 2014 International Trade Statistics Yearbook, Volume I - Trade by Country, xix. [↑](#footnote-ref-38)
7. <http://comtrade.un.org/data/doc/api/> The UN Comtrade data extraction API [↑](#footnote-ref-40)
8. <http://unstats.un.org/unsd/methods/m49/m49alpha.htm> Countries or areas, codes and abbreviations [↑](#footnote-ref-42)
9. <http://rpubs.com/malexan/m49> Matrunich A. (2015). M49 country codes in Tariffline [↑](#footnote-ref-45)
10. <http://hts.usitc.gov/?query=wheat> Harmonized Tariff Schedule (2015 HTSA Revision 1 Edition) [↑](#footnote-ref-58)
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 [↑](#footnote-ref-60)
12. <http://www.customs.go.jp/yusyutu/2011/data/e201101j_04.htm> [↑](#footnote-ref-62)