

# Trade2

*Aleksandr Matrunich*

*August 8, 2015*

## 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

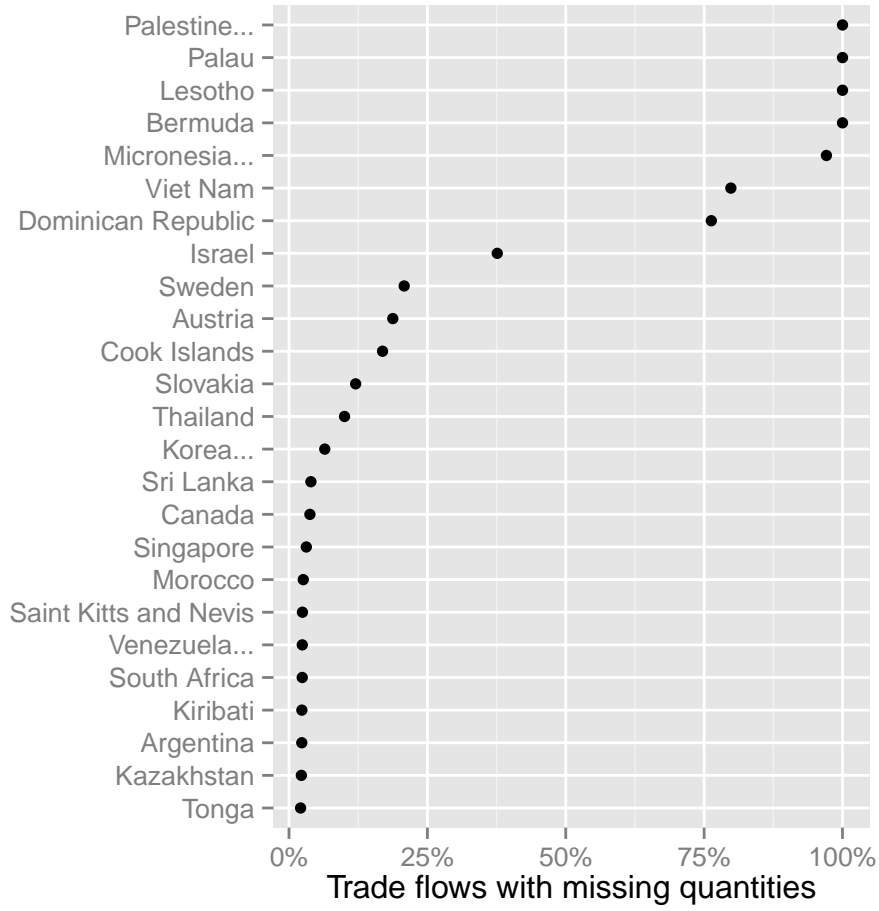
There are cases when a country reports itself as a partner to exports or imports. Such situations can occur due to mistakes or when an entrepôt exists.

Table 1: 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.

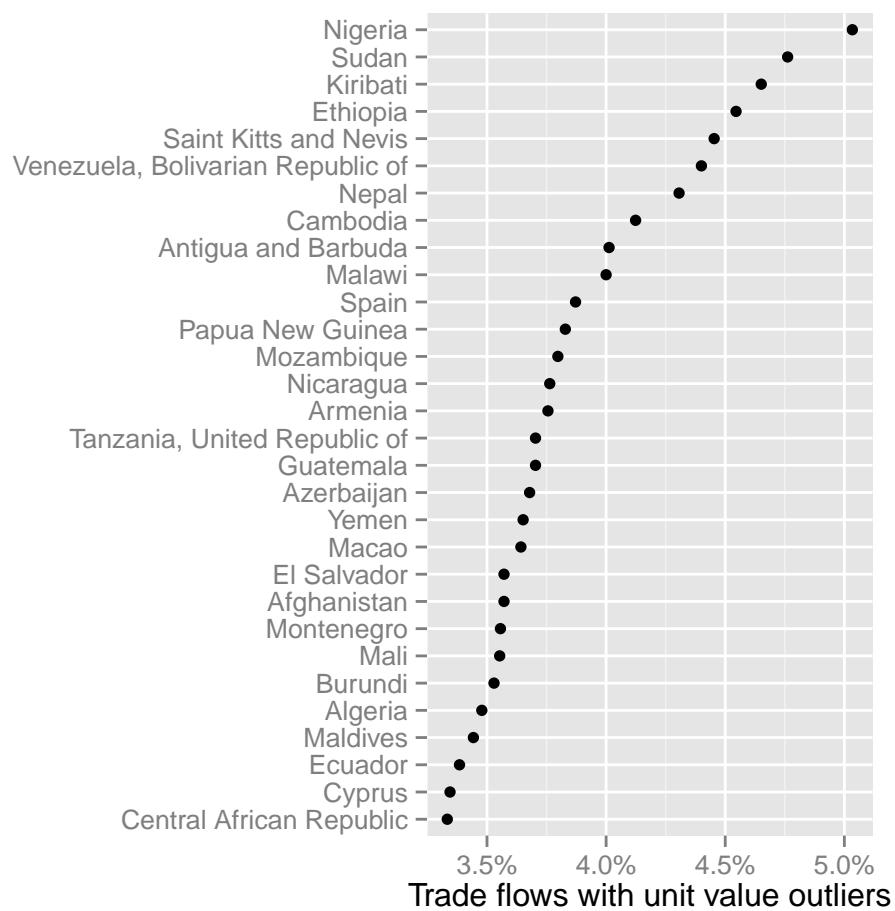


## Detection of outliers

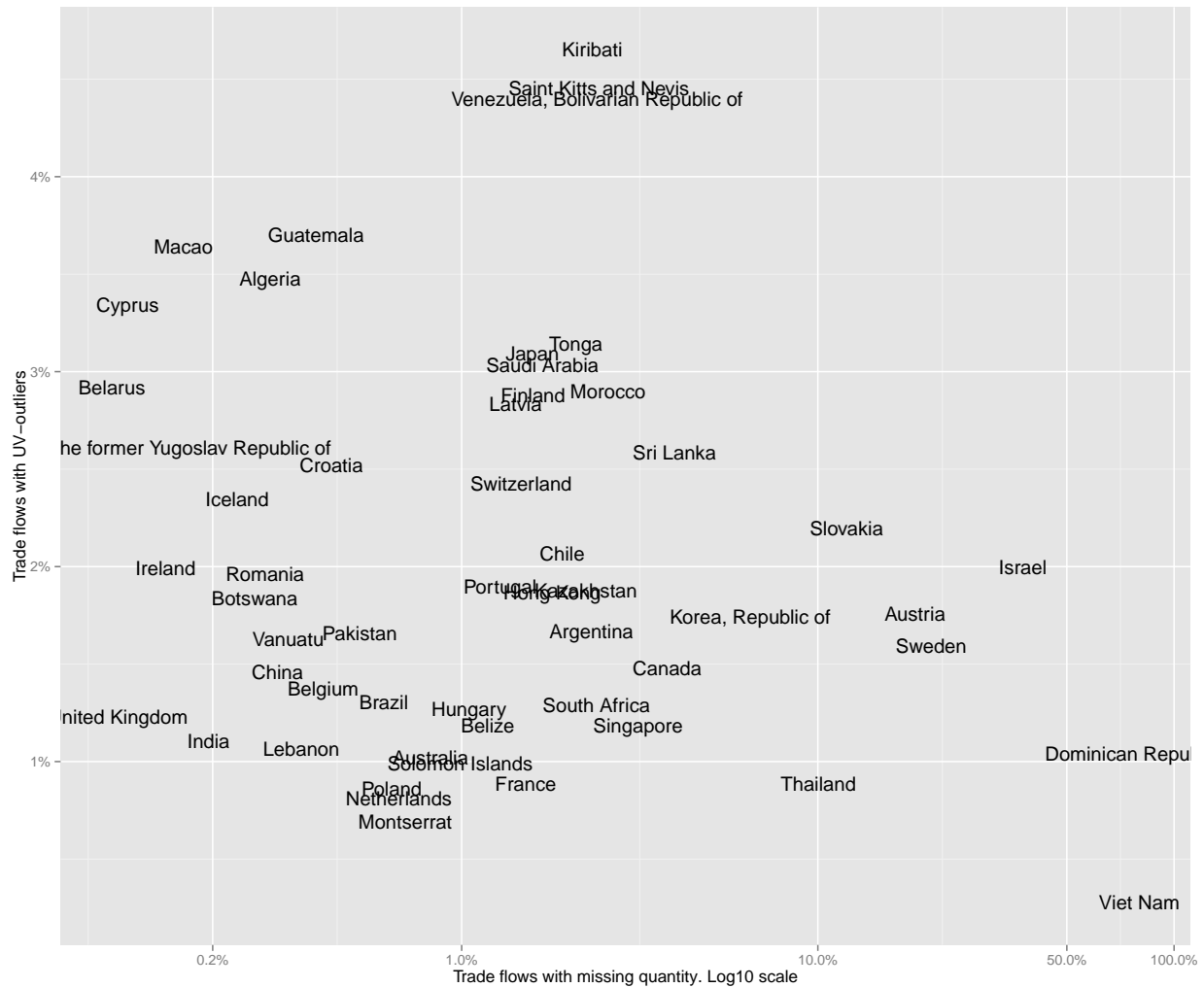
We define outliers as observations located outside the range:

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

where  $Q_1$  and  $Q_3$  are the lower and upper quartiles respectively, and  $k$  is a nonnegative constant. In this paper we use  $k$  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.

Table 2: Example trade flows with outlied unit values

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	Japan	Import	1001902096	61	46362	760.03	1.47
United States	Spain	Import	0203294000	10087	194271	19.26	4.76
United States	Germany	Import	1506000000	643	1973457	3069.14	145.01
United States	Japan	Import	1008200000	327	2479	7.58	1.88
United States	Japan	Import	1509904000	769	22842	29.70	3.66
United States	Japan	Import	1516100000	1090	93260	85.56	27.54
United States	Japan	Import	1006204040	2520	9679	3.84	0.89
United States	Japan	Import	1514999020	431	9193	21.33	3.71

Reporter	Partner	Flow	Commodity	Weight	Value	UV	UV_me
United States	Netherlands	Import	1505001000	280	4124	14.73	3.92
United States	Mexico	Import	0202305000	12989	112727	8.68	4.49

### Imputing using reporter median unit values

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

Table 3: Example trade flows with corrected weight

Reporter	Partner	Flow	Commodity	Weight	Weight_corr	Weight_diff
United States	Japan	Import	1001902096	61	31539	-31478
United States	Spain	Import	0203294000	10087	40813	-30726
United States	Germany	Import	1506000000	643	13609	-12966
United States	Japan	Import	1008200000	327	1319	-992
United States	Japan	Import	1509904000	769	6241	-5472
United States	Japan	Import	1516100000	1090	3386	-2296
United States	Japan	Import	1006204040	2520	10875	-8355
United States	Japan	Import	1514999020	431	2478	-2047
United States	Netherlands	Import	1505001000	280	1052	-772
United States	Mexico	Import	0202305000	12989	25106	-12117

### Imputing using data from trade partner (mirroring)