Description of BACI

1 Structure of the dataset

BACI provides yearly data on bilateral trade flows at the product level. Products are identified using the Harmonized System (HS), which is the standard nomenclature for international trade, used by most customs. The Harmonized System was revised in 1992, 1996, 2002, 2007, 2012 and 2017, and we provide BACI in each of those 6 revisions:

HS revision	Years available	Name of the files
92	1995-2022	BACI_HS92_Y***year***_V***version***.csv
96	1996-2022	BACI_HS96_Y***year***_V***version***.csv
02	2002-2022	BACI_HS02_Y***year***_V***version***.csv
07	2007-2022	BACI_HS07_Y***year***_V***version***.csv
12	2012-2022	BACI_HS12_Y***year***_V***version***.csv
17	2017-2022	BACI_HS17_Y***year***_V***version***.csv

Each ***version*** of BACI is identified by the year and the month of its release, under the form *YYYYMM* (*202401* for the January 2024 release, for instance)

year identifies the year during which the recorded trade flows took place.

Each trade flow within BACI is characterized by a combination exporter-importer-product-year. We provide the value and the quantity.

BACI contains 6 variables:

Variable	Description
`t`	Year
`k`	Product category (HS 6-digit code)
`i`	Exporter (ISO 3-digit country code)
`j`	Importer (ISO 3-digit country code)
`v`	Value of the trade flow (in thousands current USD)
`q`	Quantity (in metric tons)

All files are in CSV format, using commas as field delimiters, and dots as decimal separators. When reading the data, we advise you not to treat the product code (`k`) variable as numeric, which would remove the leading zeros of the HS codes.

To save space, only the strictly positive trade flows are recorded in BACI.

2 Additional files

In addition to the core BACI files, we provide four additional set of files that may be useful to BACI users:

Name	Function
country_codes	Associates the ISO 3-digit country codes to country names
product_codes	Associates the HS 6-digit product codes to product names

2.1 Country codes

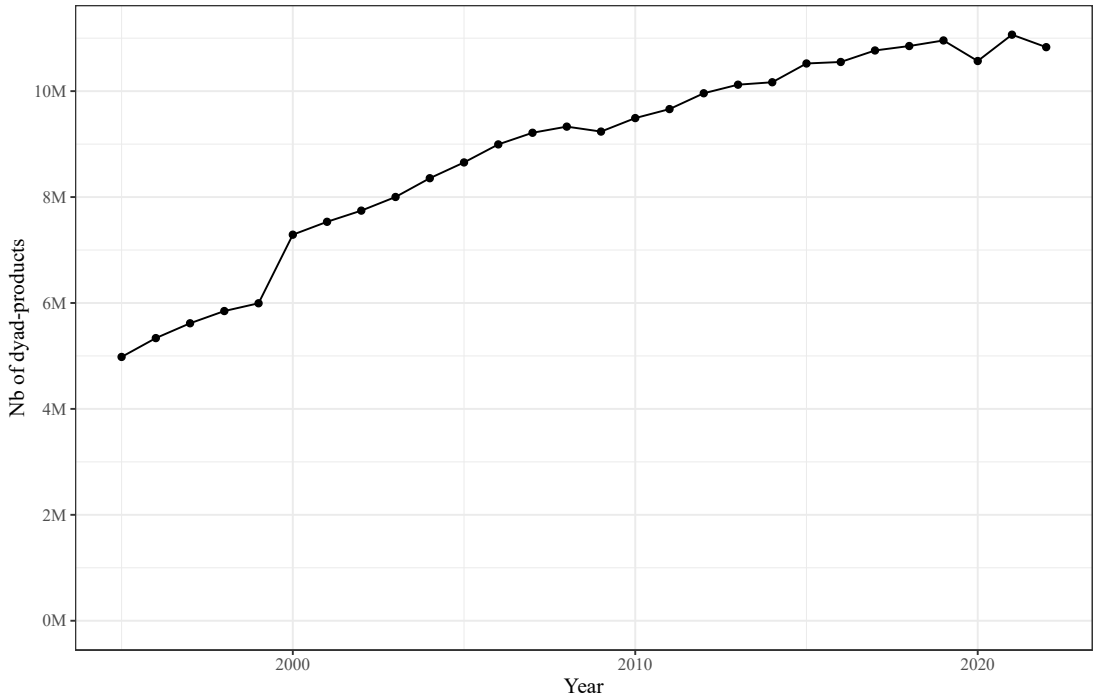
These files associate the ISO 3-digit numeric codes used in BACI with country full names and with other versions of the ISO codes (3-letter and 2-letter). They were constructed based on the metadata (https://unstats.un.org/unsd/tradekb/Knowledgebase/50377/Comtrade-Country-Code-and-Name) provided by Comtrade.

2.2 Product codes

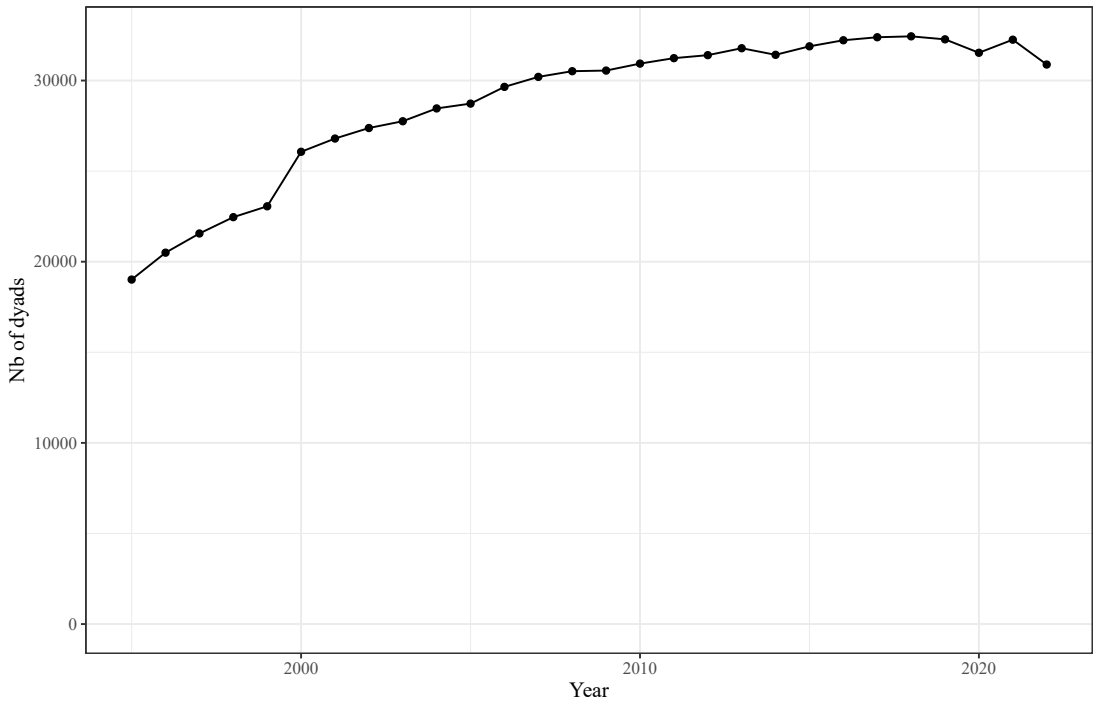
These files contain lists of the product codes used in each revision of the Harmonized System, along with a description of each product. They were constructed based on the metadata (https://comtrade.un.org/data/cache/classificationH0.json) provided by Comtrade.

3 Descriptive statistics

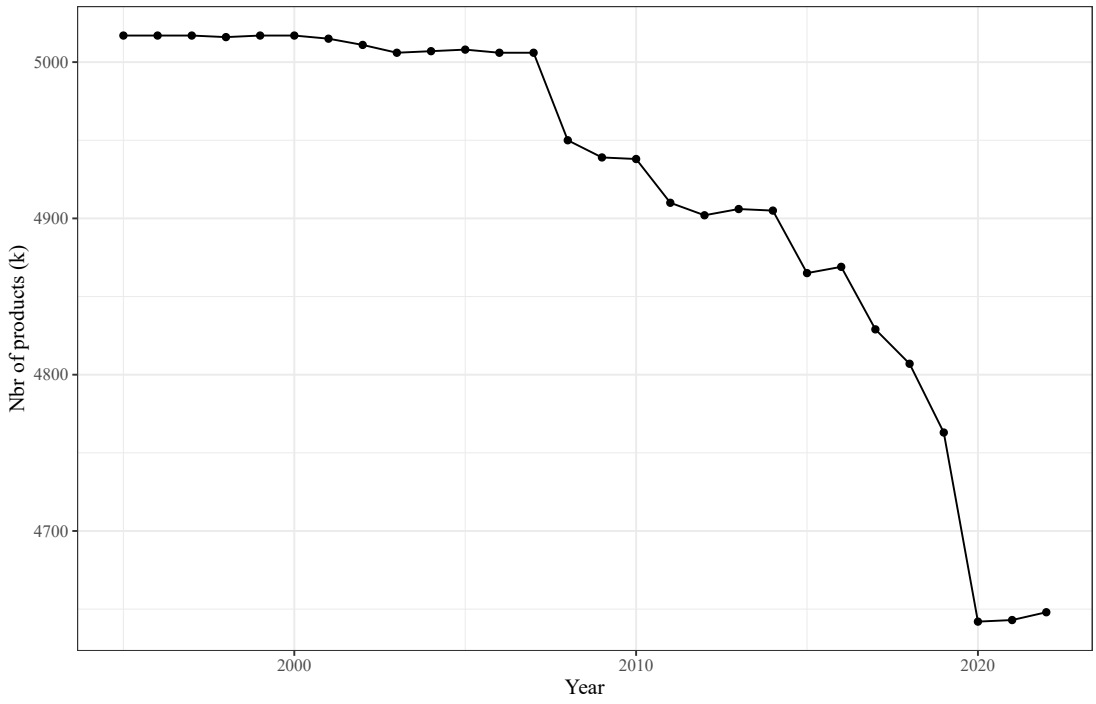
Number of observations (dyad-products) (ijk): This is the number of distinct combinations importer-exporter-product with at least one non-zero trade flow in BACI.



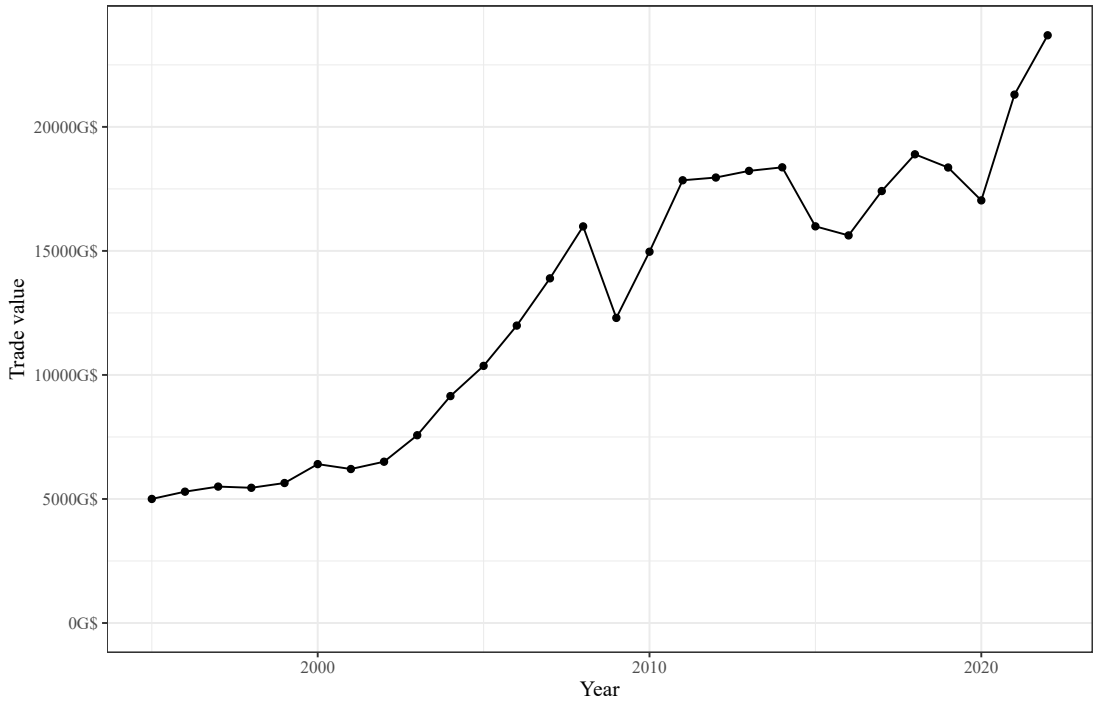
Number of dyads (ij) : This is the number of distinct combinations importer-exporter with at least one non-zero trade flow in BACI. The sharp increase up to the mid-2000s reflects improvements in the coverage of our primary sources over this period.



Number of products (k): A product is defined as a 6-digit item of the HS nomenclature, 1992 revision.



Total value: Aggregate value of trade flows recorded in BACI (in $).

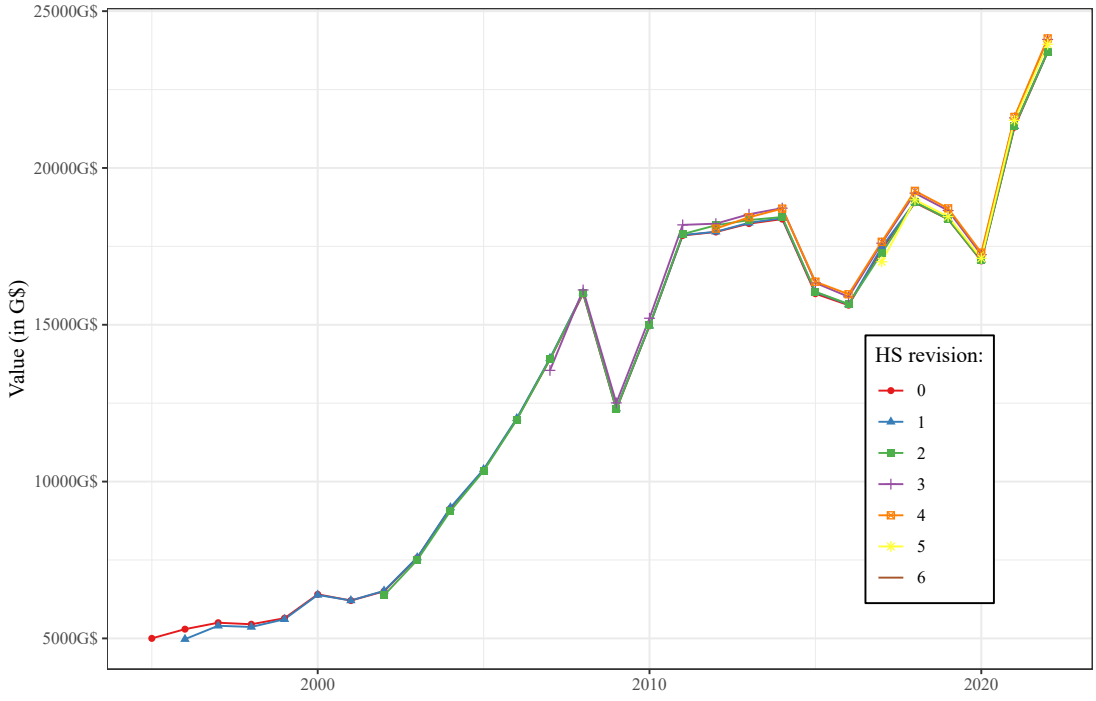


4 Differences across HS revisions

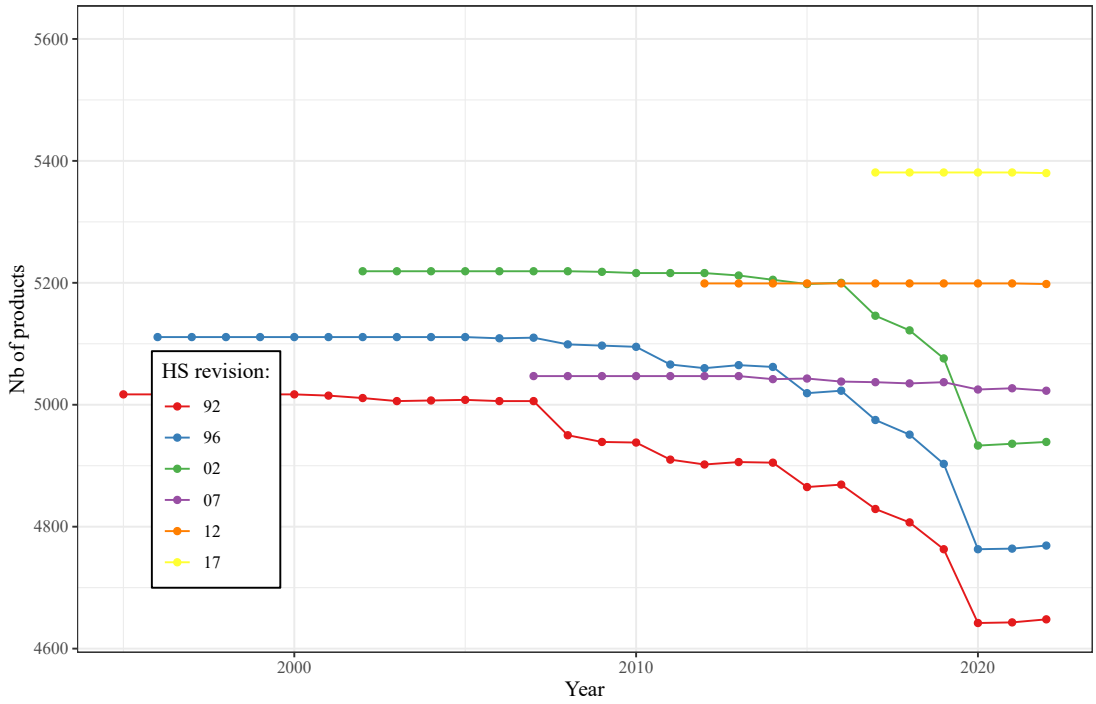
The HS product nomenclature is regularly updated (revised). Each country can report its trade flows in the revision of its choice, not necessarily the most recent one. Conversions from newer to older revisions are possible, but not the reverse (conversions from older to newer revisions).

More recent revisions of the HS nomenclature are available only for more recent years. To study old trade flows therefore have to choose an old HS revision. To study recent trade flows, a recent HS revision is preferable since it reflects more accurately the data reported by the countries.

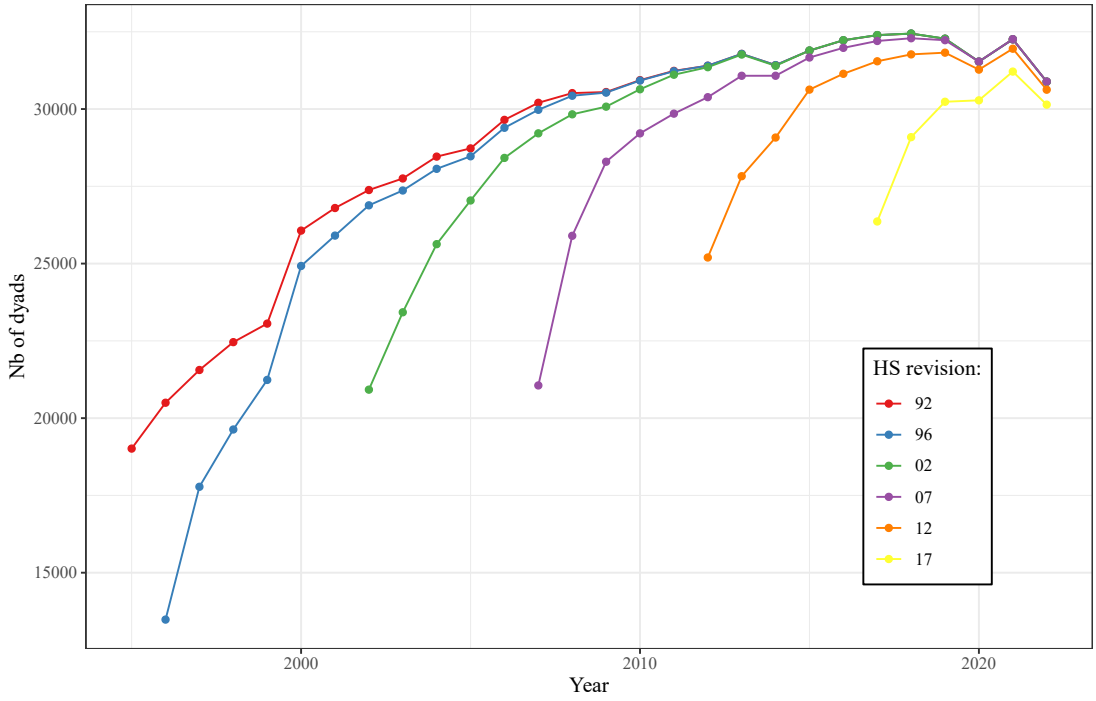
The total trade value does not differ much across HS revisions.



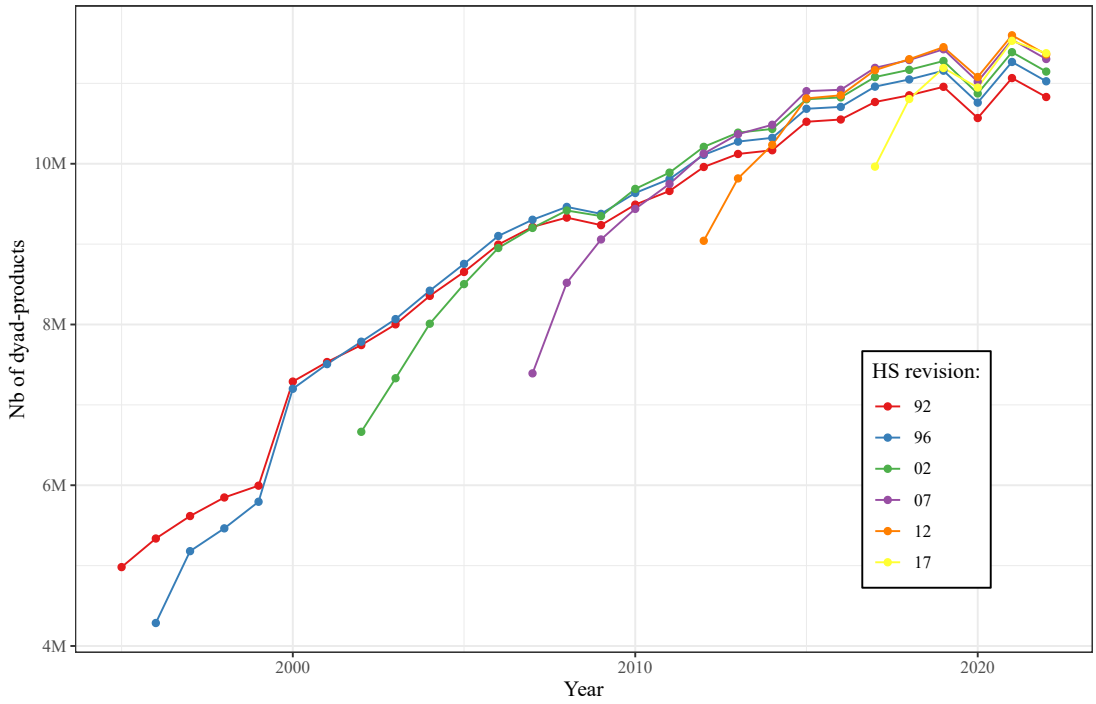
The conversion process reduces the number of products present in the data. Therefore, using an old revision for data originally expressed in a recent revision leads to some inaccuracies.



Since conversions from older to newer revisions are not possible, using a more recent HS revision leads to less dyads being available: some countries do not report in the more recent revision, and data for these countries is therefore lost when using a recent revision.



In the end, the number of trade flows recorded in BACI does not differ much across revisions.



The aggregate quantity sometimes differ across revisions, reflecting differences in the conversion factors (to kg from other quantity units), that are determined separately for each revision.