

# Data of National Dishes in the Developed and Developing Countries in the World, Their Similarity and Trade Flows

Anne C. Wunderlich <sup>1,\*</sup> and Andreas Kohler <sup>2,†</sup>

<sup>1</sup> Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), 8903 Birmensdorf, Switzerland

<sup>2</sup> ZHAW School of Management and Law, 8401 Winterthur, Switzerland

\* Correspondence: [anne.wunderlich@wsl.ch](mailto:anne.wunderlich@wsl.ch)

† These authors contributed equally to this work.

**Abstract:** This paper presents a database that includes information on national recipes and their ingredients for 171 countries, measures for food taste similarities between all 171 countries as well as bilateral migration and agro-food trade data for 5 years. The database can be used for analyzing e.g., the relation between food preferences and international trade or food preferences and health outcomes (e.g., obesity) across countries.

**Dataset:** <https://doi:10.16904/envidat.276>. Direct URL to data: [https://www.envidat.ch/dataset/data-of-national-dishes-their-similarity-and-trade-flows?\\$\\_no\\$\\_\\$cache\\$\\_\\$=\\$=True](https://www.envidat.ch/dataset/data-of-national-dishes-their-similarity-and-trade-flows?$_no$_$cache$_$=$=True).

**Dataset License:** Open Database License (ODbL).

**Keywords:** national dishes; food taste similarity index; agro-food trade



**Citation:** Wunderlich, A.C.; Kohler, A. Data of National Dishes in the Developed and Developing Countries in the World, Their Similarity and Trade Flows. *Data* **2022**, *7*, 142. <https://doi.org/10.3390/data7110142>

Academic Editor: Francisco Guijarro

Received: 2 September 2022

Accepted: 20 October 2022

Published: 26 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Summary

The sense of taste is already formed in early childhood and therefore, the food we grow up with shapes us for the rest of our lives. Therefore, national dishes or food we eat defines us not only as individuals but also as societies leading to substantial differences in consumer tastes and preferences across countries. Taste in food is argued as being persistent [1], e.g., analyze how food preferences are based on associations with the context and consequences of eating various foods. The evolution of food preferences seem to be primarily determined by past consumption of particular foods which are locally available (see [2,3]).

In order to measure food tastes at the country level a novel data set on national dishes and their ingredients for 171 countries was compiled. Further, our data set contains information on migration and bilateral food trade flows for five years. Importer and exporter GDP, population size, distance between the most populated importer and exporter cities, as well as information on Preferential Trade Agreements (PTA) between importer and exporter, contiguity, language and colonial information complete our data set. Besides, the data set is supplemented by two dyadic food taste similarity measures using (i) the Manhattan distance and (ii) latent semantic analysis (LSA).

Although [3] (for rice) and [4] (for cars) has shown the importance of taste by using product-specific attributes, so far, only few literature tries to quantify taste in trade. Most papers in the literature are limited to specific products, like, e.g., [5] for French champagne or [6] for wine. Overall, only for relatively few food products, an external measure of preferences or taste exist.

The relationship between our two data set is as follows: With help of the data set, it is possible to analyze the relationship between preferences and trade, as differences in tastes can, e.g., shape international agro-food trade [7] or affect consumer quality valuation of imported goods [8].

Furthermore, a database that measures differences in tastes and is time-invariant can also be used to analyze the link between taste and markups for different sectors like, e.g., for the food processing sector in Italy as done by [9] with help of our data set. They use a data set of all Italian exporters of cheese and processed meat over the period from 2013 to 2019 to understand the pricing strategy of exporters across international markets. Hereby, our data set helped to get a better understanding, that export prices across markets differ due to taste conditional on quality. Further, [10] used our data set to show that consumer taste explains as much of the variation in export revenue as marginal costs.

This paper aims at providing a description of our whole data set, including information on two dyadic food taste similarity measures and explains the data retrieval and processing.

## 2. Data Description

The data set contains information on migration and bilateral food trade flows. In order to measure food tastes at the country level a novel data set on national dishes and their ingredients for 171 countries was compiled. Therefore, the data set is composed of two data sets that are ready to read in CSV.

The first file (*national\_dishes\_ingredients.csv*) shows information of the ingredients in all national dishes of 171 countries and their names as well as a short description of the dish (e.g., soup, stew, etc.). Data on national dishes was compiled by the authors. We gathered ingredients for 350 different dishes with the exact name of the dish. For each country the national dishes were used. If a country had more than one common national dish we searched for the most popular dish and denominated this as “the” national dish. In the end, we therefore had 171 national dishes. Overall, there are 218 ingredients. Each line in the data set presents information for a single recipe in columnar form. Table 1 shows all the variables included in this csv.-file.

**Table 1.** Description of the variables included in: *national\_dishes\_ingredients.csv*.

Name of the Variable	Variable Description
<i>country</i>	name of the country
<i>iso3</i>	Country Code ISO 3166-1 alpha-3, using the English short country names officially used by the ISO 3166 Maintenance Agency
<i>dish</i>	Name of the national dish
<i>description</i>	type of the national dish (e.g., rice dish, dumplings, stew)
<i>national_dish</i>	national dish = 1 if official national dish of a country, in case of no official national dish we used the most common or traditional dish as the national dish
<i>columnF-HO</i>	names of ingredients (e.g., beef, veal, fowl etc.), which is equal to 1 if the ingredient is used in the national dish

The second file (*gravity\_food\_tastes\_similarity.csv*) contains panel information for 5 years (1998, 2000, 2005, 2010 and 2015) for the 171 countries. For each country, the bilateral agro-food trade flows in current million USD are recorded, defined as HS chapters 1–24, and based on the UN Comtrade Database. Further, we have data on importer and exporter GDP (in current USD) and population (in millions) both in logs, log distance between the most populated cities, as well as dummy variables for Preferential Trade Agreement (PTA) in force, contiguity, common official primary language, language spoken by at least 9% of the population in both countries, country pair ever in colonial relationship, common colonizer post-1945, pair currently in colonial relationship, pair in colonial relationship post-1945, countries were or are the same country. All data on these variables come from [11]. Data on migration is measured as the stock of foreign-born people by destination of origin for all countries and years [12]. Furthermore, the data set contains two different food tastes similarity measures on the basis of the national dishes that are also included within this data set: (i) the Manhattan distance (named *food\_sim\_manhattan* in the data set) and the (ii) latent semantic analysis (LSA) (named *food\_sim\_lsa* in the data set). Table 2 shows the description of all the variables included in the data set *gravity\_food\_tastes\_similarity.dta*.

**Table 2.** Description of the variables included in *gravity\_food\_tastes\_similarity.dta*.

Name of the Variable	Variable Description
<i>t</i>	year of observation (1998, 2000, 2005, 2010 and 2015)
<i>i</i>	number for the respective export country
<i>i_iso3</i>	iso3-code (name) for the respective export country
<i>j</i>	number for respective import country
<i>j_iso3</i>	iso3-code (name) for the respective import country
<i>v</i>	export value in millions of USD
<i>q</i>	export quantity in tons
<i>migrant_stock</i>	total migrant stock at mid year
<i>i_GDPcurr</i>	GDP in current USD of export country
<i>j_GDPcurr</i>	GDP in current USD of import country
<i>i_pop</i>	population size in export country
<i>j_pop</i>	population size in import country
<i>contig</i>	=1 if contiguity of countries is given
<i>comlang_off</i>	=1 if countries share the same official or primary language
<i>comlang_ethno</i>	=1 if countries share a language that is spoken by at least 9% of the population
<i>colony</i>	=1 if countries ever had a colonial relationship
<i>comcol</i>	=1 if countries had a common colonizer post 1945
<i>curcol</i>	=1 if countries have currently a colonial relationship
<i>col45</i>	=1 if countries had a colonial relationship post 1945
<i>smctry</i>	=1 if countries were or are the same country
<i>dist</i>	simple distance of the most populated cities of import and export countries (in km)
<i>distcap</i>	simple distance between capitals (in km)
<i>PTA</i>	=1 if a preferential trade agreement is in force
<i>food_sim_manhattan</i>	food similarity index compiled with Manhattan index
<i>food_sim_lsa</i>	food similarity index compiled with latent semantic analysis (LSA)

### 3. Methods

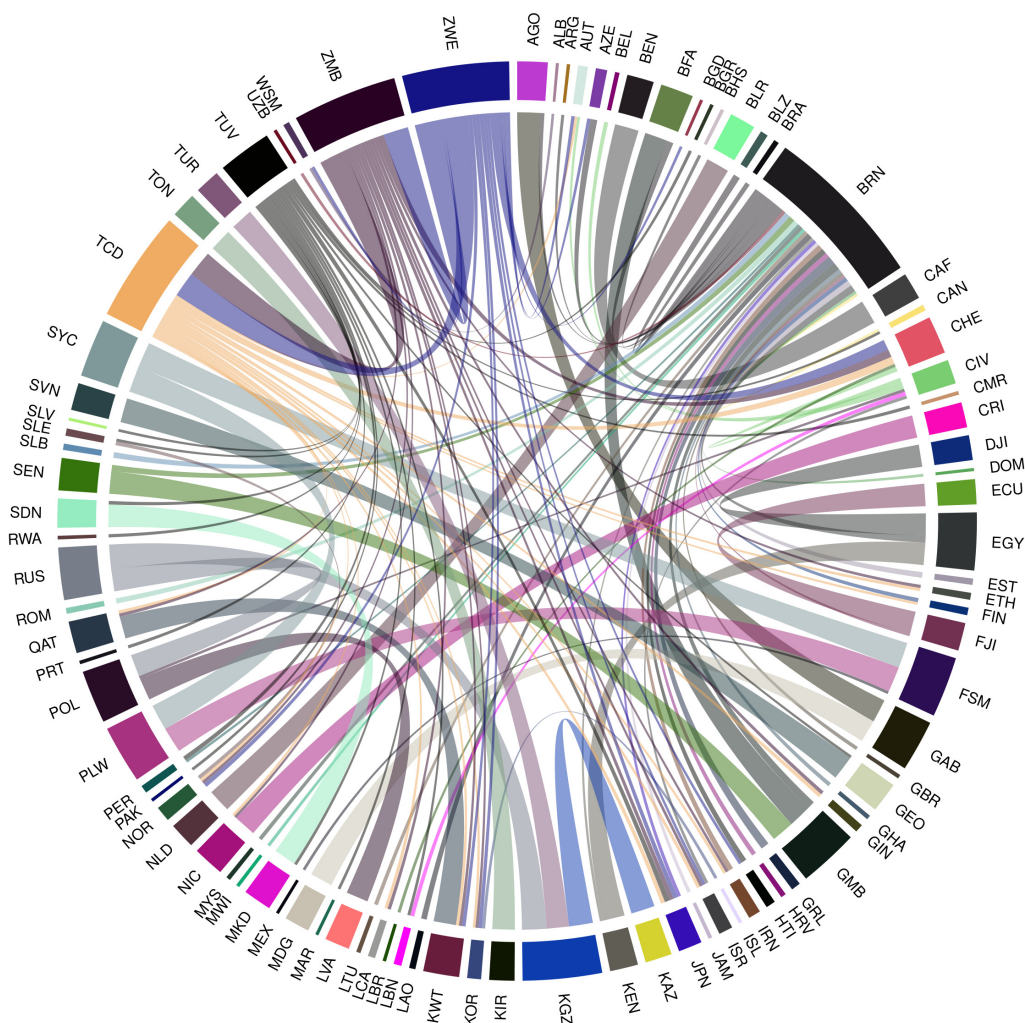
Our raw data contains data on bilateral agro-food trade flows and migration stocks for 171 countries for the years 1998, 2000, 2005, 2010 and 2015, as migration data is only available in 5-year intervals and for those 171 countries. Data on bilateral food trade flows for the years 1998, 2000, 2005, 2010 and 2015 is from a data set that is prepared by the Centre d'études prospectives et d'informations internationales (CEPI). It is downloaded from the UN Comtrade Database [11]. Data on migration was downloaded from the [12].

Data on national dishes was compiled by the authors. To compose the data set on the national dishes the authors collected data on a daily base from 4 June 2018 to 23 August 2018. For each country the national dishes listed on [https://simple.wikipedia.org/wiki/List\\_of\\_national\\_dishes](https://simple.wikipedia.org/wiki/List_of_national_dishes) (accessed on 25 September 2019) were used. In case of more than one national dish on the list, we compiled the ingredient list of each dish listed and included them to the data set but only made use of the most popular dish in the similarity measures.

With the information on each national dish of the countries an ingredient list of each dish was compiled. Doing so, the ingredients for all recipes were found on [foodpassport.com](http://foodpassport.com). If the recipe was not available on this website, the authors used [nationalfoods.org](http://nationalfoods.org) to retrieve the ingredient list. In the end, the authors compiled a data set with 350 national dishes from 171 countries that includes overall 218 different ingredients. If an ingredient is used to prepare a national dish, it is marked with a the number 1 in the respective line. All ingredients were included to the data, e.g., common ingredients like salt were also included as it is not used in all national dishes. Decisions had to be taken in relation to ingredients that are mixtures of ingredients. Hereby, we included every single ingredient that was used to prepare this mixture. Conversely, dishes that could not be prepared “in the moment” were taken as one single ingredient (e.g., dry-cured sausages or soy sauce). No differences were made concerning some ingredients like, e.g., pepper: We did not distinguish several peppers, like, e.g., black, rose or green pepper, but used the term pepper.

Both, the Manhattan distance and the latent semantic analysis (LSA) were then compiled by the authors and added to the data set as well. Both measures are similarly distributed. They are both bimodal distributions with a mass point at zero.

Figure 1 is a chord diagram. This chord diagram shows the results for the country pairs with the most and least food taste similarity based on LSA. Therefore, not all the 171 countries are displayed (but only 80 countries). Thick links show a high degree of similarity while thin lines indicate a low degree. For example, Russia's food tastes are quite similar to those of Poland and Kyrgyzstan, whereas South Korea's and Zimbabwe's food tastes are very dissimilar.



**Figure 1.** Pairwise Similarity and Dissimilarity across the country's national dishes within the LSA similarity measure.

To analyze food similarities across countries, the authors prefer the LSA food tastes similarity measure because it takes into account whether ingredients used in two national recipes are relatively common ingredients (e.g., salt, pepper) or uncommon (e.g., coriander which is only used in few recipes). Therefore, Figure 1 uses the LSA index to illustrate (dis-)similarities between national dishes of countries across the world. Ref. [7] provide in-depth details about both, the Manhattan distance and the LSA approach.

The final data on national dishes was extracted onto an MS Excel spreadsheet using simple software routines to assist with data validation like, e.g., spell checking for typographical errors and duplicate entries. As the data were compiled by the authors, this could be a source of potential errors because it is not generated by an algorithm. Eventually, it must be noted that the data set only contains a small size of dishes each country, in some

cases only one national dish which is ignoring the variety of foods consumed in countries. But it is not our approach to identifying preferences perfectly, but to tease out a measure that is likely to best capture consumer taste. Data on migration and trade flows comes from [11,12] and is therefore an official data set.

#### 4. Usage Notes

By merging a data set that links a measure of preferences to international trade flows, a variety of topics can be analyzed. The data set can be used to investigate, e.g., the effect of food tastes on international trade flows based on a dyadic measure of food tastes similarities between countries. Researchers in empirical trade economics, agriculture or health economics can benefit from these data. The data set assesses food preferences of the developed and developing world and can therefore contribute to the empirical literature on the effects of tastes or preferences on international trade. As the measure of tastes is time-invariant, it can be used to analyze the link between taste and markups for different sectors to get a better understanding on how export prices across markets differ due to taste conditional on quality. Besides, the data set can be used to show how much variation in export revenues can be explained by consumer taste. Due to the multiplicity, the data is highly acclaimed see, e.g., [7–10]. Besides, the data could be used to research the relationship between food preferences and health status (e.g., obesity) across countries.

The data can be used for research on the relation between food preferences and international trade or food preferences and health status (e.g., obesity) across countries, as the data set on food preferences not only includes national dishes but for most countries the most common dishes consumed.

**Author Contributions:** Conceptualization, A.C.W. and A.K.; validation, A.C.W. and A.K.; data curation, A.C.W. and A.K.; writing—original draft preparation, A.C.W. and A.K.; writing—review and editing, A.C.W. and A.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** [https://www.envidat.ch/dataset/data-of-national-dishes-their-similarity-and-trade-flows?\\$\\_\\$no\\$\\_\\$cache\\$\\_\\$\\_\\$=True](https://www.envidat.ch/dataset/data-of-national-dishes-their-similarity-and-trade-flows?$_$no$_$cache$_$_$=True) (accessed on 26 October 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Birch, L.L. Annual Review of Nutrition. *Dev. Food Prefer.* **1999**, *19*, 41–62.
2. Aizenman, J.; Brooks, E. Globalization and Taste Convergence: The Cases of Wine and Beer. *Rev. Int. Econ.* **2013**, *16*, 217–233. [[CrossRef](#)]
3. Atkin, D. Trade, Tastes, and Nutrition in India. *Am. Econ. Rev.* **2013**, *103*, 1629–1663. [[CrossRef](#)]
4. Cosar, K.; Grieco, P.; Li, S.; Tintelnot, F. What Drives Home Market Advantage? *J. Int. Econ.* **2018**, *110*, 135–150. [[CrossRef](#)]
5. Crozet, M.; Head, K.; Mayer, T. Quality Sorting and Trade: Firm-level Evidence for French Wine. *Rev. Econ. Stud.* **2012**, *79*, 609–644. [[CrossRef](#)]
6. Armstrong, M.; Chen, Y. Inattentive Consumers and Product Quality. *J. Eur. Econ. Assoc.* **2009**, *7*, 411–422. [[CrossRef](#)]
7. Kohler, A.; Wunderlich, A.C. How migrants' food tastes shape international agri-food trade. *Appl. Econ. Lett.* **2021**, *29*, 469–476. [[CrossRef](#)]
8. Guerra, F. *Three Essays on the Role of Product Quality in International Trade*; NNT: 2021NSARE057: Economics and Finance. Agrocampus Ouest: Rennes, France, 2021. Available online: <https://tel.archives-ouvertes.fr/tel-03710288> (accessed on 26 October 2022).
9. Haase, O.; Curzia, D.; Raimondia, V.; Olpera, A.; Solazzo, R. Markups, taste and quality. In Proceedings of the 96th Annual Conference of the Agricultural Economics Society, Leuven, Belgium, 4–6 April 2022.
10. Aw, B.Y.; Lee, Y.; Vandenbussche, H. Consumer Taste in Trade. CEPR Discussion Paper; 2020, No. DP14941. Available online: <https://ssrn.com/abstract=3638046> (accessed on 26 October 2022)..

11. Gaulier, G.; Zignago, S. BACI: International Trade Database at the Product-Level. The 1994–2007 Version. Working Papers 2010-23, CEPII, 2010. Available online: <http://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=2726> (accessed on 26 October 2022).
12. United Nations. Trends in International Migrant Stock: Migrants by Destination and Origin 1990–2015. United Nations Database, POP/DB/MIG/Stock/Rev 2015, United Nations. 2015. Available online: <http://www.un.org/en/development/desa/population/migration/data/index.shtml> (accessed on 18 June 2018).