2017 Commodity Flow Survey Methodology

July 2020

Contents

OVERVIEW

The Commodity Flow Survey (CFS) is a joint effort by the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation, and the U.S. Census Bureau, U.S. Department of Commerce. The survey is the primary source of national and subnational level (state and metropolitan area) data on domestic freight shipments by establishments in mining, manufacturing, wholesale, auxiliaries, and selected retail and services trade industries located in the 50 states and the District of Columbia. See the "First Stage—Establishment Selection" section for a more precise description of a metropolitan area. Data are provided on the type, origin and destination, value, weight, modes of transportation, distance shipped, and ton-miles of commodities shipped. The CFS is conducted every 5 years as part of the Economic Census. It provides a modal picture of national freight flows, and represents the only publicly available source of commodity flow data for the highway mode. The CFS was conducted in 1993, 1997, 2002, 2007, 2012, and 2017.

CFS data are used by policymakers and transportation professionals in various federal, state, and local agencies for assessing the demand for transportation facilities and services, energy use, and safety risk and environmental concerns. Additionally, business owners, private researchers, and analysts use the CFS data for analyzing trends in the movement of goods, mapping spatial patterns of commodity and vehicle flows, forecasting demands for the movement of goods, and determining needs for associated infrastructure and equipment.

OBJECTIVES

The primary objective of the 2017 CFS was to estimate shipping volumes (value, tons, and ton-miles) by commodity and mode of transportation at various levels of geographic detail. Another objective was to estimate the volume of shipments moving from one geographic area to another (i.e., flows of commodities between states, regions, etc.) by mode and commodity. A detailed description of the survey coverage and sample design of the 2017 CFS is provided below.

INDUSTRY COVERAGE

The 2017 CFS covers business establishments with paid employees that are located in the United States and are classified using the 2012 North American Industry Classification System (NAICS) in mining, manufacturing, wholesale, and selected retail and services trade industries, namely, electronic shopping and mail-order houses, fuel dealers, and publishers. Additionally, the survey covers auxiliary establishments (i.e., warehouses and managing offices) of multi-establishment companies.

Advance Survey

For the 2017 CFS, a targeted Advance Survey was conducted in 2016 to improve the quality of the data on the frame for certain industries or types of establishments. See Table 1 for the groups included in this Advance Survey.

For the first four groups in Table 1 (auxiliaries, publishers, electronic shopping, and support activities for printing), the purpose was to identify those establishments that actually conduct shipping activities. In these groups, surveyed establishments that reported they did not conduct any shipping activity were excluded from the eventual CFS sample universe. In these industries, approximately 38,000 establishments were identified as nonshippers and removed from the eventual sampling frame. For the other categories, the objective was to obtain an accurate measure of their shipping activity as well as contact information.

Table 1.
2016 CFS Advance Survey Composition

Advance Survey group	Number of establishments
Auxiliaries (NAICS 484, 4931, and 551114)	40,280
Publishers (NAICS 5111 and 51223)	17,643
Electronic shopping mail order establishments (NAICS 4541)	31,746
Support activities for printing (NAICS 323120)	1,596
Mines (NAICS 2121, 2122, and 2123)	6,229
Certainty[1] establishments from the 2012 CFS	33,775
Other large establishments[2]	18,749
Total	150,018

[1] These are, generally, large establishments that were sampled with a probability of 1 in the 2012 CFS and so were likely to also be selected into the 2017 sample with certainty as well.
[2] These are establishments, not previously included in a group, whose measure of size (MOS) exceeded an industry specific cutoff. These were also likely to be selected into the 2017 sample with certainty.

CFS Industries

In-scope industries for the 2017 CFS were selected based on the 2012 NAICS definitions. Industries included in the 2007 and 2012 CFS were selected based on the 2002 and 2007 versions of the NAICS, respectively. The industries in the 1997 CFS and the 1993 CFS were selected based on the 1987 Standard Industrial Classification System (SIC) and, although attempts were made to maintain similar coverage among the SIC based surveys (1993 and 1997) and the NAICS based surveys (2002, 2007, 2012, and 2017), there have been some changes in industry coverage

due to the conversion from SIC to NAICS. Most notably, coverage of the logging industry changed from an in-scope Manufacturing industry (SIC 2411) to the out-of-scope sector of Agriculture, Forestry, Fishing, and Hunting under NAICS 1133. Also, publishers were reclassified from Manufacturing (SIC 2711, 2721, 2731, 2741, and part of 2771) to Information (NAICS 5111 and 51223) and were excluded in the 2002 CFS. Subsequent surveys have included publishers as well as retail fuel dealers.

The (2012) NAICS industries covered in the 2017 CFS are listed in Table 2.

Excluded industries: Establishments classified in transportation (other than freight trucking and warehousing), construction, and most retail and services industries are excluded. These sectors have several million establishments in total and very few of these establishments are likely to have significant shipping activity. Including these sectors in the survey would have either: (1) reduced the quality of the data as large numbers of establishments would have been nonshippers, or (2) required a much greater sample size (and much greater cost). Other industry areas that are not covered, but may have significant shipping activity, include agriculture and government. These are out of scope for the Economic Census (after the Census of Agriculture was transferred to the Department of Agriculture) and the Census Bureau has no data on the shipping activity of individual establishments in these sectors. For agriculture, while the CFS does not cover shipments of agricultural products from the farm site to the processing centers or terminal elevators (generally short-distance local movements), it does cover the shipments of these products from the initial processing centers or terminal elevators onward.

General exclusions: Data for most government or military operated establishments are excluded from the CFS. The exceptions are government liquor wholesalers. The CFS also excludes establishments or firms with no paid employees and foreign establishments.

SHIPMENT COVERAGE

The CFS captures data on shipments originating from selected business establishments located in the 50 states and the District of Columbia. The CFS does not cover shipments originating from business establishments located in Puerto Rico and other U.S. possessions and territories.

Likewise, shipments traversing the United States from a foreign location to another foreign location (e.g., from Canada to Mexico) are not included, nor are shipments from a foreign location to an initial U.S. location. However, imported products are included in the CFS from the point that they leave the importer's initial U.S. location (assuming it was an in-scope industry to the CFS) for shipment to another location. Shipments that are transported

Table 2.
NAICS In-Scope Industries to the 2017 CFS

NAICS code	Description
212	Mining (except oil and gas)
311	Food manufacturing
312	Beverage and tobacco product manufacturing
313	Textile mills
314	Textile product mills
315	Apparel manufacturing
316	Leather and allied product manufacturing
321	Wood product manufacturing
322	Paper manufacturing
323	Printing and related support activities
324	Petroleum and coal products manufacturing
325	Chemical manufacturing
326	Plastics and rubber products manufacturing
327	Nonmetallic mineral product manufacturing
331	Primary metal manufacturing
332	Fabricated metal product manufacturing
333	Machinery manufacturing
334	Computer and electronic product manufacturing
335	Electrical equipment, appliance, and component manufacturing
336	Transportation equipment manufacturing
337	Furniture and related product manufacturing
339	Miscellaneous manufacturing
4231[1]	Motor vehicle and parts merchant wholesalers
4232[1]	Furniture and home furnishing merchant wholesalers
4233[1]	Lumber and other construction materials merchant wholesalers
4234[1]	Commercial equipment merchant wholesalers
4235[1]	Metal and mineral (except petroleum) merchant wholesalers
4236[1]	Electrical and electronic goods merchant wholesalers
4237[1]	Hardware and plumbing merchant wholesalers
4238[1]	Machinery, equipment, and supplies merchant wholesalers
4239[1]	Miscellaneous durable goods merchant wholesalers
4241[1]	Paper and paper product merchant wholesalers
4242[1]	Drugs and druggists' sundries merchant wholesalers
4243[1]	Apparel, piece goods, and notions merchant wholesalers
4244[1]	Grocery and related product merchant wholesalers
4245[1]	Farm product raw material merchant wholesalers
4246[1]	Chemical and allied products merchant wholesalers
4247[1]	Petroleum and petroleum products merchant wholesalers
4248[1]	Beer, wine, and distilled alcoholic beverage merchant wholesalers
4249[1]	Miscellaneous nondurable goods merchant wholesalers
4541	Electronic shopping and mail-order houses
45431	Fuel dealers
4841[2]	General freight trucking
4842[2]	Specialized freight trucking
4931[2]	Warehousing and storage
5111	Newspaper, periodical, book, and directory publishers
51223	Music publishers
551114[3]	Corporate, subsidiary, and regional managing offices

[1] Wholesale establishments exclude manufacturers sales offices and own brand importers.

[2] Includes only captive warehouses that provide storage and shipping support to a single company. Warehouses offering their services to the general public and other businesses are excluded. For tabulation and publication purposes, NAICS 484 is grouped with NAICS 4931.

[3] Includes only those establishments in the industry with shipping activity as determined from the Advance Survey.

through a foreign territory with both the origin and destination in the United States are included in the CFS data. The mileages calculated for these shipments exclude the foreign country segments (e.g., shipments from New York to Michigan through Canada do not include any mileages for Canada). Export shipments are included, with the domestic destination defined as the U.S. port, airport, or border crossing of exit from the United States. See the "Mileage Calculation" section for additional detail on how mileage estimates were developed.

SAMPLE DESIGN

Overview

The sample for the 2017 CFS was selected using a three-stage design in which the first-stage sampling units were establishments, the second-stage sampling units were groups of four 1-week periods (reporting weeks) within the survey year, and the third-stage sampling units were shipments.

First Stage—Establishment Selection

To create the first-stage sampling frame, a subset of establishment records (as of July 2016) was extracted from the Census Bureau's Business Register. The Business Register is a database of all known establishments located in the United States and its territories. An establishment is a single physical location where business transactions take place or services are performed. Establishments located in the United States, having nonzero payroll in 2014 or 2015, (or 2016, for single-establishment companies), and classified in mining (except oil and gas extraction), manufacturing, wholesale, electronic shopping and mail order, fuel dealers, and publishing industries, as defined by the 2012 NAICS, were included on the sampling frame. Certain wholesalers (i.e., manufacturers' sales offices, agents, and brokers) and certain importers (i.e., own brand importers and marketers) were excluded from the frame. These wholesalers do not maintain inventory at their office location but rather arrange for products to be shipped to a buyer from some other location.

Auxiliary establishments (e.g. truck transportation facilities, warehouses, and central administrative offices) with shipping activity were also included on the sampling frame. Auxiliary establishments are establishments that are primarily involved in rendering support services to other establishments within the same company, instead of for the public, government, or other business firms. All other establishments included on the sampling frame are referred to as nonauxiliary establishments.

As described in the "CFS Industries" section, establishments classified in forestry, fishing, utilities, construction, and all other transportation, retail, and services industries were not included on the sampling frame. Farms and government-owned entities (except government-owned liquor wholesalers) were also excluded. For 2017, the resulting frame comprised approximately 710,500 establishments as summarized in Table 3.

Table 3.
CFS Frame Summary Statistics

Trade area	Establishments on the CFS frame		
	2017 CFS	2012 CFS	2007 CFS
Mining	6,065	6,543	6,789
Manufacturing	288,180	305,805	327,826
Wholesale	339,870	345,511	356,477
Retail	40,370	27,697	25,190
Services	11,785	15,599	22,539
Auxiliaries	24,228	14,959	14,878
Total	710,498	716,114	753,699

For each establishment, sales, payroll, number of employees, a 6-digit NAICS code, name and address, and a primary identifier were extracted, and a measure of size (MOS) was computed. The MOS was designed to approximate an establishment's annual total value of shipments for the year 2014 or 2015. One reason we say approximate is because often, we compute the MOS as the sales from the 2012 Economic Census, multiplied by the ratio of the 2014 (or 2015) payroll to the 2012 payroll.

All of the establishments included on the sampling frame had state and county geographic codes. We used these codes to assign each establishment to one of the 132 detailed geographic areas (called CFS Areas) used for sampling and publication. There are three types of CFS Areas:

1. Metropolitan area (MA): The state part of a selected metropolitan statistical area (MSA) or combined statistical area (CSA).
2. The Remainder of State (ROS): The portion of a state containing the counties that are not included in the MA type CFS Areas defined above.
3. Whole state: An entire state where no MA type CFS Areas are defined within the state. In this instance, the ROS is the whole state.

Table 4 in the "Geographic Strata" section shows the counts of these three types of CFS Areas.

Stratification

The sampling frame was primarily stratified by geography and industry, then later substratified by MOS[1] class, with some exceptions for auxiliary establishments and

[1] The MOS of an establishment is an estimate of its annual value of shipments derived from data contained on the Census Bureau's Business Register. It is used to assign each establishment to a size substratum within each primary stratum (which are usually defined by geography and industry).

hazardous materials establishments, as described below. These geography by industry cells form the primary strata for the main part of the sample.

Geographic Strata

Geographic strata were defined by a combination of the 50 states, the District of Columbia, and the CFS Areas selected based on their population and importance as transportation hubs or foreign trade gateways. These CFS Areas were defined using the Office of Management and Budget's 2015 definitions (OMB Bulletin 15-01). All other MAs were collapsed with the nonmetropolitan areas within the state into ROS CFS Area strata. When a MA crossed state boundaries, we considered the size of each state part of the MA when determining whether or not to create geographic strata in each state in which the MA was defined. For example, the Chicago CSA was split into two CFS Areas: the Illinois part and the Indiana part. The Wisconsin part of Chicago was considered too small to be a separate CFS Area and was combined into the Remainder of Wisconsin CFS Area. Table 4 summarizes the number of CFS Areas used for sampling and publication by type.

Table 4.
Summary of 2017 CFS Geographic Stratification

Geographic stratum (CFS Area) type	Number of sampled CFS Areas
Metropolitan area (CSA or MSA) state part	84
Remainder of State (ROS)[1]	35
Whole state (AK, AR, ID, IA, ME, MS, MT, NM, ND, SD, VT, WV, and WY)	13
Total number of CFS Areas	132

[1] Three states do not have a ROS component: DC, NJ, and RI.

See Figure 1 ("Map of 2017 CFS Areas") at the end of the "2017 CFS Methodology" that shows the 84 MA type CFS Areas.

Industry Strata

The industry strata were defined as follows. Within each of the geographic strata, we defined 48 industry groups based on the 2012 NAICS codes:

- Three in mining (4-digit NAICS).
- Twenty-one in manufacturing (3-digit NAICS).
- Eighteen in wholesale (4-digit NAICS).
- Two in retail (NAICS 4541 and 45431).
- One in services (NAICS 5111 and 51223 combined).
- Three in auxiliaries (combinations of NAICS 484, 4931, and 551114).

For auxiliaries that responded to the Advance Survey and were found to be shippers, 132 primary strata were created, one in each CFS Area, combining NAICS 484, 4931, and 551114.[2] For auxiliary establishments that did not respond to the Advance Survey, two separate sets of strata were created because establishments in NAICS 484 and 4931 are much more likely to be shippers than establishments in 551114 and therefore we wanted to sample them at a higher rate than establishments in NAICS 551114. These strata were:

- Up to 132 strata (one per CFS Area) for nonresponding truck transportation establishments and warehousing and storage establishments (NAICS 484 and NAICS 4931).
- Up to 132 strata (one per CFS Area) for nonresponding corporate, subsidiary, and regional managing offices establishments (NAICS 551114).

Hazardous Materials (HAZMAT) Strata

In order to produce good estimates of HAZMAT shipments, we identified and used 21, 6-digit NAICS industries with high amounts of HAZMAT shipments to form primary strata. The 2012 CFS data were used to identify these industries and in general, these industries were chosen because:

- They had a large (weighted) total value or total tonnage of HAZMAT.
- A high percentage of their (unweighted) shipments were HAZMAT.

Fifteen of the 21 HAZMAT NAICS industries were made certainty strata[3] and the remaining 6 industries were made into primary strata defined by state and the 6-digit NAICS code.

Table 5 shows the number and types of primary strata for the main, auxiliary, HAZMAT parts of the sample, plus two special certainty strata. Note that these are the number of strata before they are further stratified by MOS, with

the exception of the HAZMAT certainty strata and the two special certainty strata.

Determining the sample sizes, stratifying by MOS size class, and sample selection

The total desired sample size for the first-stage sample was 104,000 establishments (roughly similar to prior surveys) and was fixed due to budget constraints. Therefore, in addition to defining the strata, a sample size was determined for each primary stratum. This was performed as follows:

- A target coefficient of variation (CV) for estimated total MOS was assigned to each primary stratum (geography by industry cell). These target CVs ranged from approximately 0.75 percent to 2.25 percent and, in general, MA type CFS Areas had smaller target CVs than ROS CFS Areas.
- Within each primary stratum, substrata defined by MOS were developed to minimize the sample size needed to achieve the target CV. The establishments in the largest MOS size class were taken with certainty. For the noncertainty substrata, the sample was allocated according to Neyman allocation, since Neyman allocation minimizes the sample size needed to achieve a target CV.
- Once the minimum sample sizes for each primary stratum were determined, they were added together and compared to the desired target sample size of about 104,000. If the total was not close enough to 104,000, we multiplied all of the target CVs by a fixed factor and repeated the process until the total sample size was close to 104,000. In the end, we decided that we were willing to accept a sample size of 103,877 establishments.
- The establishments in the geography by industry by MOS size class substrata were selected by simple random sampling without replacement. The total sample size was 103,877 establishments of which 51,266 were selected with certainty (see Table 6).

Second Stage—Reporting Week Selection

The frame for the second stage of sampling consisted of the 52 weeks in 2017. Each establishment selected into the 2017 CFS sample was systematically assigned to report for four reporting weeks, one in each quarter of the reference year (2017). Each of the 4 weeks was in the same relative position in the quarter. For example, an establishment might have been requested to report data for the 5th, 18th, 31st, and 44th weeks of the reference year. In this instance, each reporting week corresponds to the 5th week of each quarter. Prior to assignment of weeks to establishments, we sorted the selected sample by primary stratum (geography x industry) and MOS. Each week of the quarter had 7,990 or 7,991 establishments assigned to it.

Table 5.
2017 CFS Primary Stratification Summary

Sample component		Number of primary strata	Number of sample establishments
Main (NAICS x CFS Area)		5,740	95,147
Auxiliary	Advance Survey responders	132	2,317
	Advance Survey nonresponders (NAICS 484 and 4931)	130	1,268
	Advance Survey nonresponders (NAICS 551114)	132	1,196
HAZMAT	Certainty (15 industries)	15	1,319
	Sampled (6 industries x state)	281	2,116
Special certainty strata	Air or water shipper in prior CFS	1	498
	Establishment specifically identified to be included	1	16

Table 6.
2017 CFS Frame and Sample Summary Statistics

Primary strata type	2017 Frame		2017 Sample			
			Total sample		Certainty component	
	Estabs	Total MOS (million dollars)	Estabs	Total MOS (million dollars)	Estabs	Total MOS (million dollars)
Main	686,269	11,365,387	95,147	8,920,109	44,955	8,354,651
Auxiliary	16,625	1,550,228	4,781	1,503,315	3,141	1,495,515
HAZMAT	7,090	895,154	3,435	871,990	2,656	863,905
Special certainty	514	254,569	514	254,569	514	254,569
Total	710,498	14,065,338	103,877	11,549,983	51,266	10,968,640

Third Stage—Shipment Selection

For each of the four reporting weeks in which an establishment was asked to report, the respondent was requested to construct a sampling frame consisting of all shipments made by the establishment in the reporting week. Each respondent was asked to count or estimate the total number of shipments comprising the sampling frame and to record this number on the questionnaire. For each assigned reporting week, if an establishment made more than 40 shipments during that week, we asked the respondent to select a systematic sample of the establishment's shipments and to provide us with information only about the selected shipments. The number of shipments to be selected (and reported) depended on the total number of shipments in the reporting week. Table 7 summarizes the reporting requirements. In general, an establishment with a large number of shipments in a week was required to report more of those shipments. If an establishment made 40 or fewer shipments during that week, we asked the respondent to provide information about all of the establishment's shipments made during that week (i.e., no sampling was required).

DATA COLLECTION

Each establishment selected into the CFS sample was mailed either a letter or a questionnaire for each of its four assigned reporting weeks, that is, an establishment was required to report once every quarter of 2017.

Table 7.
CFS Third-Stage Sampling Sample Sizes

Total number of shipments in the reporting week	Respondent action	Minimum number of shipments to be reported	Maximum number of shipments to be reported
1–40	Report every shipment	1	40
41–600	Select (and report) a systematic sample of shipments	21	40
601–3,000		30	60
3,000 or more		50	80

Larger establishments (approximately 70 percent of the sample), determined by MOS, were mailed a letter and were instructed to report electronically through the online instrument. Smaller establishments (approximately 30 percent of the sample) were mailed a questionnaire and could report via paper or electronically. Establishments reporting electronically in one quarter were sent letters instead of questionnaires in subsequent quarters. Approximately 89 percent of the questionnaires were returned using the online instrument and nearly 8 percent were returned on a paper questionnaire. A small number (approximately 3 percent) of responses were collected via other means—mostly spreadsheets through the Secure Messaging Center or by telephone. For a given establishment, the respondent was asked to provide the following information about each of the establishment's reported shipments:

- Shipment ID number.
- Shipment date (month, day).
- Shipment value.
- Shipment weight in pounds.
- Commodity code from Standard Classification of Transported Goods (SCTG) manual.
- Commodity description.
- An indication (Yes or No) of whether the shipment was temperature controlled.
- United Nations or North American (UN/NA) number for a HAZMAT shipment.
- U.S. destination (city, state, ZIP code)—or gateway for an export shipment.
- Domestic modes of transport.
- An indication (Yes or No) of whether the shipment was an export.
- City and country of destination for an export shipment.
- Export mode of transport.

See the 2017 questionnaire and instruction guide at <www.census.gov/programs-surveys/cfs/technical-documentation/questionnaires.html>.

By CFS definition, a shipment is a single movement of goods, commodities, or products from an establishment to a single customer or to another establishment owned or operated by the same company as the originating establishment (e.g., a warehouse, distribution center, or retail or wholesale outlet). Full or partial truckloads were counted as a single shipment only if all commodities on the truck were destined for the same location. For multiple deliveries on a route, the goods delivered at each stop were counted as one shipment. Interoffice memos, payroll checks, or business correspondence were not included in the CFS. Likewise, the CFS does not include shipments of refuse, scrap paper, waste, or recyclable materials unless the establishment was in the business of selling or providing these materials.

For a shipment that included more than one commodity, the respondent was instructed to report the commodity that made up the greatest percentage of the shipment's weight.

Table 8.
SCTG Code Changes

SCTG		Description
2012	2017	
07-R	07	Prior to 2012 CFS, Fats and oils were all classified under commodity code 07. For CFS 2012 CFS, oils and fats treated for use as biodiesel moved to commodity code 18 under Fuel Oils.
074-R	074	Prior to the 2012 CFS, fats and oils intended for use as biodiesel were not specifically identified, but were included in commodity code 074. In the 2012 CFS, fats and oils intended for use as biodiesel were specified and classified in under commodity code 182 (biodiesel and blends of biodiesel).
0743-R	0743	Prior to the 2012 CFS, fats and oils intended for use as biodiesel were not specifically identified, but were included in commodity code 0743. In the 2012 CFS, fats and oils treated for use as biodiesel were specified and classified under commodity code 182.
08-R	08	Prior to the 2012 CFS, alcohols intended for use as fuel were not specifically identified, and were included under SCTG 08. In the 2012 CFS, ethanol for fuel moved to SCTG 17. Additionally, beverages and denatured alcohol were more clearly identified.
083-R	083	Prior to the 2012 CFS, denatured alcohol of more than 80 percent alcohol by volume was included in commodity code 083. In the 2012 CFS, denatured alcohol of more than 80 percent by volume was moved to commodity code 084, and ethanol for use as biofuel was moved to commodity codes 175 and 176.
0831-R	0831	Prior to the 2012 CFS, both Denatured ethyl alcohol, and undenatured ethyl alcohol of more than 80 percent alcohol by volume were included in commodity code 0831. In the 2012 CFS, denatured alcohol of more than 80 percent by volume was moved to commodity code 0841, and ethanol for use as biofuel was specified and moved to commodity codes 175 and 176.
17-R	17	Prior to 2012 CFS, Denatured ethyl alcohol, and undenatured ethyl alcohol were all classified under SCTG 08. For CFS 2012 CFS, ethanol that is used for fuel was identified and removed from SCTG 08 to SCTG 17 under fuel alcohols. Also, kerosene, which prior to 2012 CFS, was included in commodity code 19, was moved under commodity code 17.
171-R	171	Prior to the 2012 CFS, commodity code 171 only included gasoline, and blend of gasoline and ethanol were not identified. In the 2012 CFS, commodity code 171 includes gasoline, and mixtures of up to 10 percent ethanol and gasoline.
172-R	172	Prior to the 2012 CFS, kerosene was included in commodity code 192, and type A jet fuel was classified under commodity code 172. In the 2012 CFS, all kerosene are classified under commodity code 172.
1720-R	1720	Prior to the 2012 CFS, kerosene was included in commodity code 192, and type A jet fuel was classified under commodity code 1720. In the 2012 CFS, all kerosene is classified under commodity code 1720.
18-R	18	Prior to the 2012 CFS, fats and oils intended for use as fuel were not identified as such, and were included in commodity code 07. In the 2012 CFS, such fats and oils were identified as biodiesel and were moved under commodity code 18.

Commodity Coding Changes for 2017

There were no changes or additions to the definitions of commodities for 2017. However the "-R" suffixes attached to SCTGs that were redefined in 2012 have been dropped (see Table 8).

EDITS

The reported data are run through a series of establishment and shipment level edits that look for inconsistences among the data items. During the review of these edit failures, individual analysts may change specific shipment level or establishment level data items. Other edit failures may be corrected in subsequent imputation processes. In addition, the mileage calculation process may change respondent reported modes of transportation.

IMPUTATION

Imputation of Shipment Value or Weight

If the ratio of a shipment's value-to-weight is unacceptable (falls outside of a commodity-specific range), or if one of these values is missing, then one of the unacceptable (or missing) values is replaced by a predicted value obtained from a donor imputation model. Such a shipment is considered a "recipient" if its commodity code is valid and one of the two data items (either shipment value or shipment weight) is reported, greater than zero, and the shipment is otherwise useable. The recipient's data item is imputed as follows:

First, a donor shipment for a given recipient with the same 5-digit SCTG is selected at random from a pool of potential donor shipments (those with valid SCTGs and with reported or edited shipment value and weight). The donor pool levels are summarized below in order of preference (the lowest numbered donor pool containing a matching shipment is used) along with the number of shipments that had value or weight imputed at that level (see Table 9).

Table 9.
Shipment Value and Weight Imputation

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level	
		Values	Weights
1	Same SCTG, establishment, and detailed shipment size class	108,343	70,014
2	Same SCTG, company, and detailed shipment size class	17,323	45,495
3	Same SCTG, origin state and CFS Area, and detailed shipment size class	24,110	105,055
4	Same SCTG, establishment, and broad shipment size class	597	456
5	Same SCTG, company, and broad shipment size class	63	1,388
6	Same SCTG, origin state and CFS Area, and broad shipment size class	870	2,661
7	Same SCTG and establishment (no restriction on shipment size)	365	389
8	Same SCTG and company (no restriction on shipment size)	170	1,569
9	Same SCTG and origin state and CFS Area (no restriction on shipment size)	1,183	3,443
10	Same SCTG (Median value-to-weight ratio of the SCTG was used)	1,812	3,999
Total number of imputed shipment values and weights		154,836	234,469
As a percentage of total shipments with value and weight (percent)		2.42	3.66

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

Then, the donor's value and weight data are used to calculate a ratio, which is applied to the recipient's reported item, to impute the item that is missing or that failed the edit. If a donor could not be found in one of the nine donor pools then the recipient's item is imputed using the median value-to-weight ratio computed using all shipments in the same SCTG as that of the recipient (level 10 in Table 9). As shown in Table 9, approximately 390,000 shipments had either their value or weight imputed.

Destination ZIP Code Correction and Imputation

A shipment's origin and destination ZIP codes are the primary inputs to determining the shipment's distance traveled (see the "Mileage Calculation" section). For some reported shipments, the destination ZIP code was missing or was not a valid ZIP code for the reported destination city. In the case of invalid ZIP codes, if the invalid ZIP code could be converted to a valid ZIP code for the destination city by: changing a single digit (other than the first one) or transposing two digits, then the ZIP code was changed to a valid one for the reported destination city. Approximately 72,700 destination ZIP codes were corrected in this process. In addition, misspellings (SAINT LOIUS) and abbreviations (ST LOUIS) of destination city names were corrected and standardized (SAINT LOUIS) and some states codes corrected (respondent reported AL instead of AK for Alaska, for example).[4] Table 10 summarizes the corrections made.

For shipments with a valid destination city and state but missing a destination ZIP code, a ZIP code was imputed using a two stage hot-deck process. A shipment was considered a "recipient" if its destination city and state were valid but its destination ZIP code was missing. The recipient's missing ZIP code was imputed as follows:

Table 10.
Destination City, State, and ZIP Code Corrections

Corrections made to destination geography	Number of shipments corrected
City name	298,969
State	16,747
City name and state	1,490
ZIP code	69,620
ZIP code and city name	2,149
ZIP code and state	810
ZIP code, city, and state	123
Total number of corrections	389,908
Of which: ZIP code corrections (last 4 rows)	72,702
As a percentage of total shipments with destination ZIP codes (percent)	1.17

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

- In the first stage, the donor pool for each recipient consisted of all complete shipments with the same destination city and state as the recipient and also from the same establishment as the recipient. If this donor pool was not empty then one of the shipments in this donor pool was randomly selected and the destination ZIP code of this selected donor was assigned to the recipient.
- If in the first stage the donor pool was empty (there was no matching shipment from the same establishment), then the donor pool was enlarged to include all complete shipments with the same destination city and state as the recipient, regardless of source. Then one of the shipments in this larger donor pool was randomly selected and the destination ZIP code of the selected donor assigned to the recipient.

Approximately 27,400 missing shipment destination ZIP codes were imputed in this process as shown in Table 11.

For some shipments, the respondent provided a post office box (P.O. Box) ZIP code as the destination ZIP code.

Table 11.
Destination ZIP Code Imputation

Level	Imputation cell variables (donor pool)	Number of ZIP codes imputed
1	Same establishment, destination city, and destination state	10,980
2	Same destination city and state	16,419
Total number of imputed ZIP codes		27,399
As a percentage of all shipments with destination ZIP codes (percent)		0.44

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

A P.O. Box destination ZIP code would only be deemed as valid if the shipment was a parcel shipment and had a reported shipment weight less than 70 lbs. If both requirements were not met, it was determined that the P.O. Box destination ZIP code was not sufficient and would need to be replaced with a non-P.O. Box destination ZIP code. A hot-deck process was used to replace these P.O. Box ZIP codes with a standard ZIP code in a process similar to the destination ZIP code imputation described above. Approximately 73,500 shipments had their reported P.O. Box ZIP code replaced with a standard ZIP code as summarized in Table 12. Additionally, for the rare cases where a P.O. Box ZIP code was reported as the shipment origin ZIP code, all such cases were replaced with a non-P.O. Box origin ZIP code by CFS analysts who resolved the issue directly with respondents. See Table 12.

Table 12.
Destination P.O. Box ZIP Code Imputation

Level	Imputation cell variables (donor pool)	Number of ZIP codes imputed
1	Same establishment, destination city, and destination state	37,782
2	Same destination city and state	35,673
Total number of imputed ZIP codes		73,455
As a percentage of all shipments with destination ZIP codes (percent)		1.18

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

Commodity Code (SCTG) Imputation

For 2017, commodity (SCTG) codes were imputed in three phases.

1. A machine learning process was used to code some shipments where the respondent provided a description of the product but not an SCTG code. In particular, we developed a model using the 6.2 million records that respondents coded themselves. This model outputs the highest-likelihood SCTG code using two input variables: the NAICS code of the establishment from which the shipment record came and the description (as a "bag-of-words") from each record. Using the model's reported prediction probability as a guide, we took a sample of 750 records that did not have an SCTG code, and had expert analysts validate the model's predictions on these records. When the model reported probability of 40 percent or higher, we validated that the model was classifying records with 80 percent accuracy. Below 40 percent, validated accuracy dropped precipitously, and thus we chose 40 percent as a model probability threshold. From this validation exercise, we were able to assign an SCTG code to approximately 106,000 shipments (1.65 percent of all shipments with SCTGs) with a high degree of confidence using the model's output.
2. In some instances it was not possible to determine a 5-digit SCTG with the required confidence in the machine learning process described above. In particular this was because, as described above, the model did not predict a 5-digit SCTG code with probability greater than 40 percent. However it was possible to assign a 2-digit code with confidence. For these 14,000 shipments, the CFS used a hot-deck method to assign full 5-digit SCTGs from donors with the same 2-digit SCTG. Table 13 shows the five levels of the imputation cells and the number of recipients for whom an SCTG donor was found at each level. The recipients were split into two groups: those with a UN/NA HAZMAT code and those without. If a donor

was not found at the first level then the donor pool was enlarged to include more—but slightly less similar—donors. The shipment value-to-weight ratio was used to identify the nearest neighbor donor. If the value-to-weight ratio could not be computed (because either value or weight was missing) then the donor was chosen randomly from donors within the imputation cell. This process of enlarging the donor pool continued until a donor was found for each recipient. In some cases, the recipient's UN/NA code and temperature control code were also replaced with the donor's to ensure that the SCTG, UN/NA code, and temperature control code assigned to the recipient were all consistent.

Table 13.
SCTG Imputation: Phase 2

A. Shipments With a UN/NA Code

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level
1	Same company, UN/NA code, and SCTG2	209
2	Same UN/NA code and SCTG2	149
3	Same company and SCTG2	28
4	Same NAICS and SCTG2	53
5	Same SCTG2	0
Subtotal		439

B. Shipments Without a UN/NA Code

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level
1	Same establishment and SCTG2	8,605
2	Same company and SCTG2	1,168
3	Same NAICS, state, and SCTG2	3,327
4	Same NAICS and SCTG2	635
5	Same SCTG2	9
Subtotal		13,744
Grand total		14,183
As a percentage of all shipments with SCTGs (percent)		0.22

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

3. For the 30,000 shipments still missing an SCTG (largely because the respondent did not provide a shipment description) an SCTG was imputed using a process similar to that above except that a 2-digit SCTG (SCTG2) was not available. Table 14 summarizes the results of this imputation process.

Table 14.
SCTG Imputation: Phase 3

A. Shipments With a UN/NA Code

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level
1	Same company and UN/NA code	623
2	Same UN/NA code	771
3	Same company	50
4	Same NAICS and state	1
5	Same NAICS	0
Subtotal		1,445

B. Shipments Without a UN/NA Code

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level
1	Same establishment	21,203
2	Same company and NAICS	962
3	Same company	153
4	Same NAICS and state	5,822
5	Same NAICS	280
Subtotal		28,420
Grand total		29,865
As a percentage of all shipments with SCTGs (percent)		0.46

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

HAZMAT Code (UN/NA) Imputation

Shipments with certain SCTG codes are HAZMAT and the respondent should have also reported a UN/NA HAZMAT code. For the 19,000 shipments where no UN/NA code was reported (but should have been), a hot-deck imputation method (similar to that used for SCTG imputation) was used to assign a UN/NA code from a similar donor. Table 15 summarizes the results of that imputation process.

Temperature Control Code Correction and Imputation

The temperature control code (Yes or No) is an indication of whether or not the shipment required a temperature controlled environment while being shipped. Missing or invalid temperature control code responses were corrected or imputed in two operations. Table 16 shows the number of shipments corrected (Yes changed to No, or vice versa) or imputed when missing. These shipments were imputed or corrected in accordance with temperature control standards for each commodity.

Table 15.
UN/NA Imputation

Level	Imputation cell variables (donor pool)	Number of shipments imputed at that level
1	Same establishment and SCTG	3,695
2	Same company and SCTG	7,380
3	Same NAICS, state, and SCTG	5,265
4	Same NAICS and SCTG	1,967
5	Same NAICS4 and SCTG	536
6	Same SCTG	216
Total		19,059
As a percentage of all shipments with UN/NA codes (percent)		4.23

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations. NAICS is the full 6-digit code, NAICS4 is the 3- or 4-digit NAICS used for sampling and publication.

Table 16.
Temperature Control Code Correction and Imputation

Process	Number of shipments affected
Temperature control code corrections	80,087
Temperature control code imputations	27,518
Total	107,605
As a percentage of all shipments with a temperature control code (percent)	1.67

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

MILEAGE CALCULATION

The CFS does not ask respondents to report the distance traveled for each shipment. However, origin and destination ZIP codes, transportation modes, commodity, and foreign country (if applicable) are required from respondents. Using these variables, a mileage estimate can be generated. To calculate mileage for shipments collected during the 2017 CFS, a mileage routing tool was developed by BTS. This tool, referred to as GeoMiler, uses current ArcGIS technology along with the latest transportation networks and routing algorithms to form likely routes for each shipment collected in the survey.

Highway

The commercial truck routing software, PC Miler, was used as the highway network for GeoMiler. PC Miler specializes in freight-focused routing as it is widely used as a navigational tool in the commercial trucking industry. Routes were generated based on the practical route setting which considers numerous variables (distance, road classification and quality, truck-restricted roads, tolls, etc.) during the route selection process. Mileage for company-owned truck, for-hire truck, and parcel (ground only) shipments are calculated over the highway network.

Rail

The latest Federal Railroad Administration (FRA) rail network was used for rail shipments collected in the 2017 CFS. The network contains all Class 1 and shortline railroads. The rail stations included in the GeoMiler rail network were obtained from RAILINC. The rail routes generated by GeoMiler were largely determined by stations chosen based on observed data from the Surface Transportation Board's Waybill Sample data.

Using the rail station within the origin ZIP code or destination ZIP code was the general preference, but if those facilities did not support the commodity being shipped, the program would search for the most likely facility based on distance and volume. If the selected facility fell outside of the origin or destination ZIP code, truck drayage was added to the shipment. If the shipment weight was too great or the truck drayage component too great in distance, GeoMiler would flag the shipment for manual correction by an analyst.

A set of Class 1 railroad transfer points (interlining) were identified from the Waybill. If the selected origin and destination stations were owned by separate Class 1 owners, then a transfer was deemed necessary. Under such a scenario, GeoMiler would select the most likely transfer point based on the order of the modeled carrying companies (e.g. Norfolk Southern to Union Pacific), transfer point volume, and overall distance.

Water

The latest United States Army Corps of Engineers (USACE) waterway network was used for water shipments collected in the 2017 CFS. The network links are classified by shallow draft, deep draft, and Great Lakes. The ports and docks included in the network also come from the USACE. The water terminals used by GeoMiler were largely based on observed data from the USACE Commodity Detail Dock-to-Dock Movement dataset. The CFS publishes water estimates by water pathway classifications, which includes inland water (usually shallow draft vessels such as barges), deep sea (usually deep draft vessels or oceangoing barges), and the Great Lakes. Shipments that transferred between different types of water links are classified under multiple waterways.

Using a dock within the origin ZIP code or destination ZIP code was GeoMiler's preference, but if those facilities did not support the commodity being shipped, the program would search for the most likely facility based on distance and volume. If the selected facility fell outside of the origin or destination ZIP code, truck drayage was added to the shipment. If the shipment weight was too great or the truck drayage component too great in distance, GeoMiler would flag the shipment for manual correction by an analyst.

Air

The air network was built by BTS personnel using BTS' Office of Airline (OAI) T-100 Segment data. The air network consists of air routes that have regular air freight service in all states except Alaska where a major consideration is the ability to access all potential points. This includes the networks of the largest parcel carriers, as well as a consolidated network that primarily covers freight activity on passenger airlines. The air routes generated by GeoMiler were based on an algorithm that factored in distance and air route volume, while generally preventing shipments from being transferred among air carriers. More information about changes to air mileage calculation can be found in the "Methodological Changes to Mileage Calculation for the 2017 CFS" section.

Pipeline

For pipeline shipments, ton-miles and average miles per shipment are not shown in the data files. For most of these shipments, the respondents reported the shipment destination as a pipeline facility on the main pipeline network. Therefore, for the majority of these shipments, the resulting mileage represented only the access distance through feeder pipelines to the main pipeline network and not the actual distance through the main pipeline network. Pipeline shipments are included in the U.S. totals for value and tons per shipment. For security purposes, there is no pipeline network available in the public domain with which to route petroleum-based products. Hence, any modal distance, either single or multi, involving pipeline was considered as solely pipeline mileage from origin ZIP code to destination ZIP code and calculated to equal the Great Circle Distance (GCD). GCD is defined as the shortest distance between two points on the earth's surface, taking into account the earth's curvature.

Multimodal Shipments

For multi-mode shipments (i.e., shipments involving more than one mode, such as truck and rail shipments, and more than one transportation network) the transfer between modes occurred at select facilities known to support such transfers. As with single-mode shipments, business rules were established to pick the most likely transfer point based on commodity, volume, and distance.

Exports

For shipments to Canada and Mexico, the mileage was calculated between the origin ZIP code and the border crossing point. For shipments to other foreign locations, the mileage was calculated between the origin ZIP code and the U.S. territorial border (this extends 12 nautical miles beyond the coastline). Mileage outside of U.S. territory was not counted. In both cases, a Port of Exit (POE), either seaport, airport, or border crossing point, was found based on an established order of processes.

GeoMiler first checked for respondent-provided data in the POE field of the questionnaire. If found to be valid, GeoMiler would route to the provided POE. If the POE field contained invalid data or was void of information, the next step was to consider the proximity between the provided shipping address and the nearest POE. If an establishment was located within a short distance of export facilities, then it was assumed that the shipment would exit from there, and GeoMiler would use that nearby POE.

If the first two options failed, provided POE information was missing or invalid, GeoMiler would then select a likely POE based on the characteristics of the shipment record. Using the foreign destination information, origin state, mode of transportation, and commodity information from the shipment record, GeoMiler would select a likely POE that was based on patterns observed in Census Bureau Foreign Trade export data. For exports to Canada and Mexico, further consideration was given to destination locations within those countries.

For additional information about exports, see the "Methodological Changes to Mileage Calculation for the 2017 CFS" section.

ZIP Codes

The source of ZIP codes in GeoMiler is PC Miler. For domestic shipments, the mileage is calculated between the origin ZIP code point and the destination ZIP code point. For export shipments, the mileage is calculated between the origin ZIP code point and the POE/U.S. territorial border. The ZIP code point is a latitude/longitude coordinate determined by the location of commercial activity within the ZIP code rather than the geographic center of the ZIP code. See more about ZIP code point placement in the "Methodological Changes to Mileage Calculation for the 2017 CFS" section.

For intra-ZIP code shipments, shipments with the origin and destination in the same ZIP code, the square root of the total ZIP code area in square miles was used as an estimate for the distance shipped.

METHODOLOGICAL CHANGES TO MILEAGE CALCULATION FOR THE 2017 CFS

BTS continues to seek improvements to the quality of the information produced from its flagship survey for data collection, the CFS. A critical measurement calculated from CFS data is the mileage traveled by each shipment. This measurement is used to calculate the ton-miles, a statistic unique to this survey. With a valid origin and destination ZIP code, and valid commodity, GeoMiler will calculate the distance traveled (in miles) by mode for each shipment reported in the CFS.

The following types of methodological changes to mileage processing were incorporated in 2017:

Use of Commodity for Rail Station and Dock Selection

In 2012, the nearest rail station or dock was selected regardless of the commodity and volume of the facility. For 2017, observed rail and water shipment data were used to form the routing. The observed inbound and outbound commodities for each station and dock were built into the rail network and waterway network, respectively. The rail station and dock selection were based on the directional commodity information along with volume and distance from the origin and destination ZIP codes.

Using the rail station or dock within the origin ZIP code or destination ZIP code was GeoMiler's preference, but if those facilities did not support the commodity being shipped, the program would search for the most likely facility based on the requirements stated above. If the selected facility fell outside of the origin or destination ZIP code, truck drayage was added to the shipment. If the shipment weight was too great or the truck drayage component too great in distance, GeoMiler would flag the shipment for manual correction by an analyst.

Shipments that included a truck drayage component are classified as "truck and rail" and "truck and water" in the CFS estimates.

Use of Designated Transfer Points

Rail

In 2012, transfers were allowed to occur at any railroad junction, regardless of station owner and trackage rights. The 2017 methodology for determining rail routes is described in the "Mileage Calculation" section under "Rail."

Water

For 2017, based on observed water shipment data, a set of shallow draft-deep draft transfer points were identified and used by GeoMiler when necessary. The accessibility of the water network at the origin and destination docks, as well as the shipment weight, were determining factors in deciding if a transfer between vessels was likely. Such shipments are classified as "multiple waterway" shipments in the CFS estimates.

In 2012, the classification of a shipment as a "multiple waterway" shipment was based on a switch in water modes (i.e., inland water to deep sea) and was solely based on the classification of the USACE waterway network links. The origin and destination, and shipment weight, were not taken into account. In 2017, to provide a more accurate picture of shipping patterns, we placed greater importance on the geography (origin and destination) and shipment weight versus the classification of the links embedded in the waterway network.

Air Routing

In 2012, the impedances for the air network were determined by airport volumes and distances. For 2017, the impedances are based on airport-to-airport link volumes, with the largest freight carriers having their own isolated subnetworks. In 2017, if the respondent indicated a shipment mode of "parcel air," the shipment was kept on the same parcel air network from origin to destination. In 2012, air shipments would be susceptible to transfers among numerous air carriers between origin and destination. Furthermore, the greater the incompleteness of the system in 2012, the greater the likelihood of needing to manually process these shipments. The subnetwork chosen was based on an impedance formula that evaluated air carrier volume and airport distance from the origin and destination ZIP codes. If the respondent indicated air, all four subnetworks were considered.

There was also an expansion of air network coverage. In 2012, there were 137 non-Alaska airports, and the air network in the 2017 CFS contains 247 non-Alaska airports.

Exports

In 2017, (as described in the "Mileage Calculation" section under "Exports") the process for selecting the best routing for export shipments was:

- GeoMiler would first check for respondent-provided data in the POE field of the questionnaire.
- If the POE field contained invalid or no information, the next step was to consider the proximity between the provided shipping address and the nearest POE.
- If the first two options failed GeoMiler would then select a likely POE based on the characteristics of the shipment record.

The 2012 version of GeoMiler did not consider respondent-provided POE information, nor proximity to a POE. Rather, the program imputed a POE for all export shipments. Additionally, the list of available exit points was expanded from 2012 based on the observed export data. This

change, along with an established order of processes to determine the best gateway for export shipments should result in observations that are closer to other observed datasets.

Additionally for 2017, to establish consistency with mileage calculation for air exports, water mileage between the POE seaport and the U.S. territorial border was calculated and contributed to total mileage for the shipment. Previously in 2012, this water mileage was not counted. Only air mileage between the POE airport and the U.S. territorial border was counted before. Because of this, total deep sea mileage is likely to increase while average miles per shipment is likely to decrease. The majority of POE seaports are located along the coast and are within a short distance of the U.S. territorial border, leading to an increase in low mileage deep sea shipments.

ZIP Codes

All GeoMiler routings are point-to-point routings and are calculated from a point in the originating establishment's ZIP code to a point in the destination ZIP code. For 2017, the location of the ZIP code points were determined by the commercial activity of the ZIP code; tending to be located closer to the more populous areas within the ZIP code. In 2012, ZIP code points were located on the geographic centroid of the ZIP code. Commercial activity and population were not considered.

For ZIP codes smaller in size, this change is minimal. But for ZIP codes larger in size, the distance between a commercial activity weighted point and the geographic centroid can be substantial.

ESTIMATION

Estimated totals (e.g., value of shipments, tons, ton-miles) are produced as the sum of weighted shipment data (reported or imputed). Percent change and percent of total estimates are derived using the appropriate estimated totals. Estimates of average miles per shipment are computed by dividing an estimate of the total miles traveled by the estimated number of shipments.

Each shipment has associated with it a single tabulation weight, which was used in computing all estimates to which the shipment contributes. The tabulation weight is a product of seven different component weights. A description of each component weight follows.

CFS respondents provided data for a sample of shipments made by their respective establishments in the survey year. For each establishment, we produced an estimate of that establishment's total value of shipments for the entire survey year. To do this, we used four different weights: the shipment weight, the shipment nonresponse weight, the quarter weight, and the quarter nonresponse weight. Three additional weights are then applied to produce estimates representative of the entire universe. These are the establishment-level adjustment weight, the establishment (or first-stage sample) weight, and the nonresponse post-stratification adjustment weight.

The shipment weight was defined as the ratio of the total number of shipments (as reported by the respondent) made by an establishment in a reporting week to the number of sampled shipments the respondent reported on the questionnaire for the same week. This weight uses data from the sampled shipments to represent all the establishment's shipments made in the reporting week. However, a respondent may have failed to provide sufficient information about a particular sampled shipment. For example, a respondent may not have been able to provide value, weight, commodity, or a destination for one of the sampled shipments. If this data item could not be imputed or otherwise obtained, then this shipment did not contribute to tabulations and was deemed unusable. To account for these unusable shipments, we applied the shipment nonresponse weight. This weight is equal to the ratio of the number of sampled shipments for the reporting week to the number of usable shipments for the same week.

The quarter weight inflates an establishment's estimate for a particular reporting week to an estimate for the corresponding quarter. The quarter weight is equal to 13. For each establishment, the quarterly estimates were added to produce an estimate of the establishment's value of shipments for the entire survey year. Whenever an establishment did not provide the Census Bureau with a response for each of its four reporting weeks, we computed a quarter nonresponse weight. The quarter nonresponse weight for a particular establishment is defined as the ratio of the number of quarters for which the establishment was in business in the survey year (usually four) to the total number of quarters (reporting weeks) for which we received usable shipment data from the establishment.

Using these four component weights and the reported (or imputed) shipment values, we computed an estimate of each establishment's value of shipments for the entire survey year. We then multiplied this estimate by a factor that adjusts this estimated value to the measure of the establishment's value of shipments or receipts used for sample stratification purposes. This weight, the establishment-level adjustment weight, attempts to correct for any sampling errors caused by the selection of specific reporting weeks or that occur during the sampling of shipments by the respondent.

The adjusted value of shipments estimate for an establishment was then weighted by the establishment's weight.

This weight is equal to the reciprocal of the establishment's probability of being selected into the first-stage sample (see the "Sample Design" section).

A final adjustment, for most industries, the nonresponse post-stratification adjustment weight, corrects for nonresponse and coverage changes since sample selection by adjusting the weighted shipment value (computed using all prior weighting factors) to the tabulated revenue data from the 2017 Economic Census as of May 30, 2020. This adjustment accounts for:

- Establishments that did not respond to the survey or from which we did not receive any usable shipment data.
- Changes in the universe of establishments between the times the first-stage sampling frame was constructed (2016) and the year in which the data were collected (2017).

For the final 2017 CFS estimates, the nonresponse post-stratification cells were defined by industry, typically the 3-digit NAICS codes (for Manufacturing) or 4-digit NAICS codes (all other industries) used for sampling, and state. There were 2,296 nonresponse post-stratification cells.

For auxiliary and publishing establishments, the adjustment was performed using the MOS from the initial sampling frame rather than Economic Census data. Receipts for auxiliary establishments are not collected in the Economic Census, because auxiliaries serve only their own companies. For publishing establishments, the receipts in the Economic Census may include revenue from activities that do not involve the shipment of a commodity.

An exception to the above description of the weighting procedure is made for what we call "certainty shipments." These are shipments about which we have learned the number of times they are made in the year. For these shipments, we let the shipment weight equal the number of times they are made in the year. We let the shipment nonresponse weight, the quarter weight, and the quarter nonresponse weight equal one. And the remaining weights are computed as described above.

For NAICS 4247, the Census revenue used for this adjustment was reduced by the portion of that revenue that was derived from the sale of crude oil[5] as crude oil shipments are out of scope to the CFS.

In rare instances, where a very large shipment required several years to construct and was delivered in 2017 (such as large ocean vessels), the Census revenue amount was changed to more accurately reflect the value of shipments rather than the accounting revenue that happened to be recognized by the establishment in 2017.

RELIABILITY OF THE ESTIMATES

The estimates presented by the 2017 CFS may differ from the actual, unknown population values. The difference between the estimate and the population value is known as the total error of the estimate. When describing the accuracy of survey results, it is convenient to discuss total error as the sum of sampling error and nonsampling error. Sampling error is the average difference between the estimate and the result that would be obtained from a complete enumeration of the sampling frame conducted under the same survey conditions. Nonsampling error encompasses all other factors that contribute to the total error of a sample survey estimate.

The sampling error of the estimates in this publication can be estimated from the selected sample because the sample was selected using probability sampling. Common measures related to sampling error are the sampling variance, the standard error, and the coefficient of variation (CV). The sampling variance is the squared difference, averaged over all possible samples of the same size and design, between the estimator and its average value. The standard error is the square root of the sampling variance. The CV expresses the standard error as a percentage of the estimate to which it refers. For percentage estimates, such as percentage change or percentage of a total, the standard error of the estimate is provided.

Nonsampling error encompasses all factors other than sampling error that contribute to the total error associated with an estimate. This error may also be present in censuses and other nonsurvey programs. Nonsampling error arises from many sources: inability to obtain information on all units in the sample; response errors; differences in the interpretation of the questions; mismatches between sampling units and reporting units, between the requested data and data available or accessible in respondents' records, or between reference periods (calendar vs fiscal year, for example); mistakes in coding or keying the data obtained; and other errors of collection, response, coverage, and processing.

Although no direct measurement of nonsampling error was obtained, in conducting the CFS, precautionary steps were taken in all phases of the collection, processing, and tabulation of the data in an effort to minimize its influence. Precise estimation of the magnitude of nonsampling errors would require special experiments or access to independent data and, consequently, the magnitudes are often unavailable. Data users should be aware that the published estimates may be affected by unmeasured nonsampling error.

[5] Based on NAPCS code 4003325000.

Suppressed Estimates

Estimates that had high CVs (greater than 50 percent), or for which the CV could not be computed were suppressed. Some of these suppressed estimates can be derived directly from the CFS tables by subtracting published estimates from their respective totals. However, the suppressed estimates obtained by such subtraction would be subject to poor response, high sampling variability, or other factors that may make them potentially misleading. Estimates derived in this manner should not be attributed to the Census Bureau. The CFS does not suppress estimates for disclosure avoidance because we use a noise infusion method for disclosure avoidance. See the section, "Disclosure Avoidance."

Individuals who use estimates in these tables to create new estimates should cite the Census Bureau as the source of the original estimates.

More detailed descriptions of sampling and nonsampling errors for the 2017 CFS are provided in the following sections.

SAMPLING ERROR

These estimates are based on a sample of shipments reported for a sample of weeks from a sample of establishments. Therefore these estimates are unlikely to exactly agree with results that would be obtained from a complete enumeration of all shipments made in 2017 from all establishments included on the sampling frame. However, because probability sampling was used at each stage of selection, it is possible to estimate the sampling variability of the survey estimates. For CFS estimates, sampling variability arises from each of the three stages of sampling.

The particular sample of shipments obtained by the CFS in 2017 is one of a large number of samples of the same size that could have been obtained using the same design. If all possible samples had been surveyed under the same conditions, an estimate of a population parameter of interest could have been obtained from each sample. These samples give rise to a distribution of estimates for the population parameter. A statistical measure of the variability among these estimates is the standard error, which can be estimated from any one sample. The standard error is defined as the square root of the variance. The coefficient of variation (or relative standard error) of an estimator is the standard error of the estimator divided by the estimator itself. For the CFS, the coefficient of variation also incorporates the effect of the noise infusion disclosure avoidance method (see the "Disclosure Avoidance" section).

Note that measures of sampling variability, such as the standard error and coefficient of variation, are estimated from the sample and are also subject to sampling variability and, technically, we should refer to the estimated standard error or the estimated coefficient of variation of an estimator. However, for the sake of brevity, we have omitted this detail. It is important to note that the standard error only measures sampling variability. It does not measure systematic biases of the sample. The Census Bureau recommends that individuals using estimates contained in this report incorporate this information into their analyses, as sampling error could affect the conclusions drawn from these estimates.

An estimate from a particular sample and the standard error associated with the estimate can be used to construct a confidence interval. A confidence interval is a range about a given estimator that has a specified probability of containing the result of a complete enumeration of the sampling frame conducted under the same survey conditions. Associated with each interval is a percentage of confidence, which is interpreted as follows. If, for each possible sample, an estimate of a population parameter and its approximate standard error were obtained, then:

1. For approximately 90 percent of the possible samples, the interval from 1.833 standard errors below to 1.833 standard errors above the estimate would include the result as obtained from a complete enumeration of the sampling frame conducted under the same survey conditions.

2. For approximately 95 percent of the possible samples, the interval from 2.262 standard errors below to 2.262 standard errors above the estimate would include the result as obtained from a complete enumeration of the sampling frame conducted under the same survey conditions. The 1.833 and 2.262 values, used to compute the 90 percent and 95 percent confidence intervals, are taken from the t-distribution with nine degrees of freedom (one less than the number of random groups used to produce the CV estimates). The t-distribution takes into account the uncertainty in the estimation of the CVs (which the CFS computes using the random group method with ten random groups).

To illustrate the computation of a confidence interval for an estimate of total value of shipments, assume that an estimate of total value is $10,750 million and the coefficient of variation for this estimate is 1.8 percent, or 0.018. First obtain the standard error of the estimate by multiplying the value of shipment's estimate by its coefficient

of variation. For this example, multiply $10,750 million by 0.018. This yields a standard error of $193.5 million. The upper and lower bounds of the 90 percent confidence interval are computed as $10,750 million plus or minus 1.833 times $193.5 million or $354.7 million. Consequently, the 90 percent confidence interval is $10,395 million to $11,105 million. If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 9 out of 10 (90 percent) of these intervals would contain the result obtained from a complete enumeration.

For estimates that were computed from relatively few shipments, the variance estimate may have been computed from fewer than ten random groups. In this case, the given method of confidence interval construction will produce a confidence interval that will be too narrow for these estimates.

NONSAMPLING ERROR

Nonsampling error encompasses all other factors that contribute to the total error of a sample survey estimate and may also occur in censuses. It is often helpful to think of nonsampling error as arising from deficiencies or mistakes in the survey process. In the CFS, nonsampling error can be attributed to sources such as:

- Specification.
- Coverage.
- Measurement.
- Nonresponse.
- Processing.

Although no direct measurement of the potential biases due to nonsampling error has been obtained, precautionary steps were taken in all phases of the collection, processing, and tabulation of the data in an effort to minimize their influence. The Census Bureau recommends that individuals using estimates in this report consider these possible sources of error when conducting their analyses, as nonsampling error could affect the conclusions drawn from these estimates.

Some possible sources of bias that are attributed to respondent-conducted sampling include:

- Constructing an incomplete frame of shipments from which to sample.
- Ordering the shipment sampling frame by selected shipment characteristics.
- Selecting shipment records by a method other than the one specified in the questionnaire's instructions.

Nonresponse

A potential source of bias in the estimates is nonresponse. Nonresponse is the inability to obtain all the intended measurements or responses from all units in the sample. Three levels of nonresponse can occur in the CFS:

- Shipment.
- Quarter (reporting week).
- Establishment.

Item nonresponse occurs either when a particular shipment data item is unanswered or the response to the question fails computer or analyst edits. Nonresponse to the shipment value or weight items is corrected by imputation. (See the "Imputation" section for a description of the imputation procedure.)

Shipment, quarter, and establishment nonresponse describe the inability to obtain any of the substantive measurements about a sampled shipment, quarter, or establishment, respectively. Shipment and quarter nonresponse are adjusted for by reweighting (see the descriptions of the shipment and quarter nonresponse weights in the "Estimation" section). Reweighting allocates characteristics to the nonrespondents in proportion to the characteristics observed for the respondents. The amount of bias introduced by this nonresponse adjustment procedure depends on the extent to which the nonrespondents differ, characteristically, from the respondents.

Establishment nonresponse is adjusted for during the estimation procedure by the nonresponse post-stratification adjustment weight. In most cases of establishment nonresponse, none of the four questionnaires have been returned to the Census Bureau after several attempts to elicit a response.

Response Rates

The CFS produces four different response rates: a participation response rate, a unit response rate, a weighted unit response rate, and a total quantity (item) response rate. The first three are based on the responses of the establishments selected into the survey. These unit response rates are shown in Table 17 (along with the final values from the 2012 survey).

Table 17.
2017 CFS Unit Response Rates

Type of response rate	Percent	
	2017 (final)	2012 (final)
Participation	57.9	57.0
Cooperation	71.5	X
Unit	63.0	66.1
Weighted unit	74.2	76.7

X Not applicable. A cooperation rate was not computed for the 2012 survey.

Note: Due to other missing data items, some of these shipments may not have been used in the final tabulations.

Table 18.
2017 CFS Total Quantity Response Rates

CFS variable	Percent	
	2017	2012
Value	50.6	51.9
Tons	47.4	50.9
Ton-miles[1]	55.7	63.2

[1] For ton-miles (which is the product of shipment weight and distance shipped) the distance shipped component is derived from the respondent-reported destination ZIP code (see the "Mileage Calculation" section). The respondent is not asked for the actual distance. This calculated distance is treated as equivalent-to-reported data for purposes of computing the TQRR for ton-miles.

Participation Response Rate (PRR)—The Participation Response Rate is the total number of unweighted establishments that provided usable[6] data divided by the total number of establishments in the sample (103,877) (expressed as a percentage).

Cooperation Rate—The Cooperation Rate is defined as the total number of unweighted establishments that provided shipment or other data to the CFS—whether or not such data was included in the estimates—divided by the sample size (and expressed as a percentage). It indicates the extent to which contacted establishments cooperated with the request to participate in the CFS.

Unit Response Rate (URR)—The Unit Response Rate is defined as the ratio (expressed as a percentage) of the total unweighted number of establishments that provided usable data to the total number of establishments that were eligible (or potentially eligible) for data collection. URRs are indicators of the performance of the data collection process in obtaining usable responses.

Weighted Unit Response Rate (WRR)—The Weighted Unit Response Rate is defined as the percentage of the total weighted sampling MOS of the establishments that provided usable data to the total weighted sampling MOS of all establishments that were eligible (or potentially eligible) for data collection. This incorporates the size of the establishment as well as its establishment (first-stage sample) weight into the measure of response.

The fourth rate is based on the quality of the individual shipment data reported by the responding establishments. These total quantity response rates for the 2017 CFS are shown in Table 18 (along with the final values from the 2012 survey).

Total Quantity Response Rate (TQRR)—The Total Quantity Response Rate is defined as the percentage of the estimated (weighted) total of a given data item (value, tons, or ton-miles) that is based on reported shipment data or from sources determined to be of equivalent-quality-to-reported data. The TQRR is an item-level indicator of the "quality" of each estimate. In contrast to the URR, these weighted response rates are computed for individual data items, so CFS produces several TQRRs.

The TQRR is the weighted proportion of the key estimates reported by responding establishments or obtained from equivalent quality sources. This measure incorporates the value of the individual shipment data items and the associated sampling and weighting factors.

DISCLOSURE AVOIDANCE

Disclosure is the release of data that have been deemed confidential. It generally reveals information about a specific individual or establishment or permits deduction of sensitive information about a particular individual or establishment. Disclosure avoidance is the process used to protect the confidentiality of the survey data provided by an individual or firm.

[6] "Usable data" means that an establishment provided at least one shipment that was used in the tabulation of published estimates.

Using disclosure avoidance procedures, the Census Bureau modifies or removes the characteristics that put confidential information at risk of disclosure. Although it may appear that a table shows information about a specific individual or business, the Census Bureau has taken steps to disguise or suppress the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data.

For the CFS the primary method of disclosure avoidance is noise infusion. Noise infusion is a method of disclosure avoidance in which the weighted values for each shipment are perturbed prior to tabulation by applying a random noise multiplier to shipment value and weight. Disclosure protection is accomplished in a manner that causes the vast majority of cell values to be perturbed by at most a few percentage points. For sample-based tabulations, such as CFS, the estimated relative standard error for a published cell includes both the estimated sampling error and the amount of perturbation in the estimated cell value due to noise. Other cells in the table may be suppressed because the quality of the data does not meet publication standards. By far, the most common reason for suppressing a cell is a high coefficient of variation (greater than 50 percent). These suppressed cells are shown with an "S" in the tables.

The Census Bureau's Disclosure Review Board (DRB) approved the methodology used to protect the confidentiality of the statistics provided in this release (approval CBDRB-FY19-374) and the 2017 CFS Preliminary release (approval CBDRB-FY18-349).

COMPARABILITY OF ESTIMATES

This section summarizes the definitional, coding, and processing differences between the 2017 and 2012 (and prior) surveys that limit the comparability of the published statistics or estimates across the survey years. Data users should exercise caution when comparing CFS data across survey years.

Sample Size Changes

Table 19 summarizes the sampling frame and sample size changes for the most recent six surveys.

Industry Coverage Changes

Industry coverage has changed slightly from survey year to survey year (see Table 20). The details of the 2017 CFS industry coverage are described in the "Industry Coverage" section. The most significant recent changes are:

- NAICS 484 was included as an in-scope auxiliary industry in 2017 and 2012 but not any prior surveys.
- NAICS 51223 (Music publishers) was included as an in-scope publishing industry in 2017 but not in 2012.
- In 2012 and prior surveys, Prepress Services establishments (2007 NAICS 323122) were excluded from the CFS. However the 2012 NAICS revision eliminated Prepress Services as a separate industry and grouped it with Trade Binding and Related Work (2007 NAICS 323121) into NAICS 323120 (Support Activities for Printing). For 2017 all of NAICS 323120 was considered to be in scope.

The 2017 estimates were based on the industry classification of the sample establishments at the time those estimates were produced (May 2020). The 2012 and earlier estimates are never revised to account for subsequent industry classification changes to the sample establishments.

Table 19.
CFS Frame Sample Sizes

Sizes	Number of establishments in each CFS cycle					
	1993	1997	2002	2007	2012	2017
Sampling frame	790,000	770,000	760,000	753,699	716,114	710,498
Sample	197,176	102,739	51,005	102,369	102,565	103,877

Note: See the "CFS Industries" section for more information on coverage changes over the years.

Table 20.
Industry Coverage Changes

Survey years				
1993 and 1997	2002	2007	2012	2017
Establishments classified based on the 1987 Standard Industrial Classification (SIC) system	Establishments classified based on the 1997 North American Industry Classification System (NAICS)	Establishments classified based on the 2002 NAICS	Establishments classified based on 2007 NAICS	Establishments classified based on 2012 NAICS
Publishers were covered—classified in Manufacturing Division	Publishers (now in the Information sector) were not covered	Publishers (5111 and 51223) were covered[1]	Publishers (5111 only) were covered1	Publishers (5111 and 51223) were covered1
Logging covered—under Manufacturing Division	Logging not covered[2]	Logging not covered[2]	Logging not covered	Logging not covered
Other Manufacturing (excluding Printing Trade Services [SIC 279])	Manufacturing (excluding Prepress services [NAICS 323122])	Manufacturing (excluding Prepress services [NAICS 323122])	Manufacturing (excluding Prepress Services [NAICS 323122])	Manufacturing (all industries)
Mining (except mining services [SICs 108, 124, 138, 148] and oil and gas extraction [SICs 131 and 132])	Mining (except support activities [NAICS 213] and oil and gas extraction [NAICS 211])	Mining (except support activities [NAICS 213] and oil and gas extraction [NAICS 211])	Mining (except support activities [NAICS 213] and oil and gas extraction [NAICS 211])	Mining (except support activities [NAICS 213] and oil and gas extraction [NAICS 211])
Wholesale (merchants and manufacturers' sales branches and government-owned liquor stores)	Wholesale (merchants and manufacturers' sales branches and government-owned liquor stores)	Wholesale (merchants and manufacturers' sales branches and government-owned liquor stores)	Wholesale (merchants and manufacturers' sales branches and own brand importers)	Wholesale (merchants and manufacturers' sales branches and own brand importers)
Retail—catalog and mail-order houses	Retail—electronic shopping and mail-order houses	Retail—electronic shopping, mail-order houses, fuel dealers	Retail—electronic shopping, mail-order houses, fuel dealers	Retail—electronic shopping and mail-order houses, fuel dealers
Auxiliaries (managing offices, warehouses)	Auxiliaries (managing offices, warehouses)	Auxiliaries (managing offices, warehouses)[3]	Auxiliaries (managing offices, warehouses and trucking)[3]	Auxiliaries (managing offices, warehouses and trucking)[3]

[1] Under NAICS, publishers were reclassified from Manufacturing (SIC 2711, 2721, 2731, 2741, and part of 2771) to Information (NAICS 5111 and 51223) and were excluded in the 2002 CFS. In 2007 and later surveys, Music Publishers (NAICS 51223) were tabulated and published in Newspaper, Periodical, Book and Directory Publishers (NAICS 5111). However, for the 2012 cycle, NAICS 51223 was not sampled.

[2] Because of changes in the classification of establishments between SIC and NAICS, logging establishments (NAICS 1133), which were covered as part of Manufacturing in the 1993 and 1997 surveys, were not included in 2002 and later surveys. Detailed information about NAICS classification can be found on the Census Bureau's NAICS Web site at <www.census.gov/eos/www/naics/>.

[3] While included in all surveys, the procedures for identifying in-scope auxiliary establishments have changed over the years:

1997 CFS: a managing office was considered in scope only if it had sales or end-of-year inventories in the 1992 Census. Research conducted prior to the 2002 CFS showed that not all managing offices with shipping activity in the 1997 CFS indicated sales or inventories in the 1997 Economic Census. Consequently, the 1997 Economic Census results were not used to determine scope for managing offices in the 2002 CFS.

2002 CFS: an auxiliary was included if it supported an in-scope or retail company.

2007 CFS: an Advance Survey of approximately 40,000 auxiliary establishments was conducted in 2006 to identify those auxiliary establishments with shipping activity. Those that indicated that shipping was performed (as well as nonrespondents) were included in the CFS sample universe.

2012 CFS: a targeted Advance Survey of approximately 100,000 establishments was conducted in 2011 to identify those establishments that actually conduct shipping activities. In general, though not always, surveyed establishments that reported that they did not conduct any shipping activity were excluded from the eventual CFS sampling frame.

2017 CFS: a targeted Advance Survey of approximately 150,000 establishments was conducted in 2016 to identify those establishments that actually conduct shipping activities. In general, though not always, surveyed establishments that reported that they did not conduct any shipping activity were excluded from the eventual CFS sampling frame.

Changes to Data Items Collected

Table 21 summarizes the shipment data items requested from each sampled establishment in a reporting period (usually one week) for the past six surveys.

Table 22 summarizes the establishment-level data collected from each sampled establishment (generally each quarter).

Geographic Area Changes

No new CFS Areas were defined for the 2017 CFS. However, some CFS Areas, while similar in name from one survey to the next, are actually made up of slightly different sets of counties. For example, in 2012, the Dallas-Fort Worth, TX CFS Area consisted of 19 counties. In 2017, the Dallas-Fort Worth, TX-OK CFS Area (TX Part) was made up of 20 counties in Texas. Consequently, as a result of this change to the Dallas-Fort Worth CFS Area, the number of counties included in the Remainder of Texas CFS Area was reduced. Table 23 lists the CFS Areas that changed from 2012 to 2017.

Mode Changes

There have been slight changes to the definitions of the modes of transportation requested from respondents over the years. These are summarized in Table 24.

Table 21.
Data Items Collected for Each Shipment

Data item	Survey years			
	1993	1997	2002, 2007	2012, 2017
Company shipment ID	C	C	C	C
Shipment date	C	C	C	C
Total value	C	C	C	C
Total weight	C	C	C	C
Commodity code	STCC[1]	SCTG[2]	SCTG	SCTG
Commodity description	C	C	C	C
Mode of transport to U.S. destination	C	C	C	C
Domestic destination (city, state, and ZIP)	C	C	C	C
Containerized? (Y/N)	C	C	X	X
Intermodal shipment? (Y/N)	X	X	C	X
Temperature controlled? (Y/N)	X	X	X	C
Hazardous material? (Y/N)	C	X	X	X
HAZMAT (UN/NA) code	X	C	C	C
Export? (Y/N)	C	C	C	C
Export destination (city, country)	C	C	C	C
Export mode	C	C	C	C

C Data collected.
X Not applicable; data not collected.
[1] STCC Standard Transportation Commodity Code.
[2] SCTG Standard Classification Transported Goods.

Table 22.
Data Items Collected From Each Establishment
(Each Reporting Period)

Data item	Survey years				
	1993	1997, 2002	2007	2012	2017
Verification of shipping address	C	C	C	C	C
Verification of mailing address	X	X	X	C	C
Operating status	C	C	C	C	C
Length of reporting period (weeks)	2	1	1	1	1
Total number of outbound shipments for the reporting period	C	C	C	C	C
Total value of weekly shipments	X	X	X	X	C
Monthly value of outbound shipment	C	C	C	C	X
Third party logistics (3PL) usage[1]	X	X	C	X	X
Rush delivery usage	X	X	X	C	X
Primary activity correct? (Y/N)	X	X	X	X	C
Time to complete (paper) survey (Hrs, Min)	X	X	X	X	C
Contact information	C	C	C	C	C

C Data item collected in survey year.
X Not applicable; data not collected.
[1] Only collected on the 4th quarter questionnaire.

Table 23.
CFS Area Definition Changes for 2017

2017 CFS Area		2012 CFS Area		Description of change[1]
142	Birmingham-Hoover-Talladega, AL	142	Birmingham-Hoover-Talladega, AL	Tallapoosa County, AL, added to the CFS Area
206	Dallas-Fort Worth, TX-OK (TX part)	206	Dallas-Fort Worth, TX	Fannin County, TX, added to the CFS Area
324	Lake Charles-Jennings, LA	29340	Lake Charles, LA	Jefferson Davis Parish, LA, added to the CFS Area

[1] The Alabama, Texas, and Louisiana ROS CFS Areas lost the counties added to the CFS Areas described above

Table 24.
Reported Modes of Transportation

Survey year			
1993	1997, 2002, and 2007	2012	2017
For-hire truck	For-hire truck	For-hire truck	For-hire truck
Private truck	Private truck	Private truck	Company-owned truck
Rail	Rail	Rail	Rail
Air	Air	Air	Air
Inland water	Shallow draft	Inland water	Inland water
Deep sea water	Deep draft vessel	Deep sea	Deep sea
Pipeline	Pipeline	Pipeline	Pipeline
Parcel, U.S. Postal Service, or courier	Parcel, U.S. Postal Service, or courier	Parcel, U.S. Postal Service, or courier	Parcel, U.S. Postal Service, or courier
Other	Other	Other	Other
Unknown	Unknown	Unknown	Unknown

As shown in Table 24, the 2012 mode category “private truck” was renamed “company-owned truck” for 2017.

There were no changes to the published detailed mode of transportation codes associated with waterborne shipments between 2012 and 2017. Table 25 lists the published water modes in 2007, 2012, and 2017. See the 2007, 2012, and 2017 questionnaires and instruction guides at <www.census.gov/programs-surveys/cfs/technical-documentation/questionnaires.html> for descriptions of these modes.

Table 25.
Published Water Mode Codes

2007		2012, 2017	
Code	Meaning	Code	Meaning
07	Water	07	Water
08	Shallow draft	08	Inland water
09	Great Lakes	09	Great Lakes
10	Deep draft	10	Deep sea
		101	Multiple waterways

In 2012, export shipments that traveled by a single mode (truck, for example) to the Port of Exit (POE) and then by ship to their foreign destinations were classified as single-mode (truck, in this example) shipments and their domestic water mileage to the U.S. territorial border was not included. In 2017, these shipments are classified as multi-mode truck and water shipments and include the domestic water mileage to the U.S. territorial boundary.

Routing Software Changes

The following methodological changes to mileage processing, implemented in 2012 and carried over to 2017, also affected mode assignment (and the shipment distance calculations).

- The maximum weight of a parcel shipment was limited to 150 pounds in 2012 and 2017. In 2007, the limit was 1,000 pounds. Shipments with weights above the maximum were reassigned to a non-parcel mode, usually a truck mode.
- In 2012 and 2017, there was no minimum restriction on the weight of an air shipment. In 2007, air shipments with a weight of less than 100 pounds were reclassified as parcel.
- Company-owned truck shipments ("private truck" in 2012) were not routed more than 500 miles during 2012 and 2017 mileage calculation as this form of in-house transportation is generally "short haul" in nature. In 2007, there was no mileage limit.
- In 2012 and 2017, there were major efforts to re-code shipments, where a respondent provided a mode of "other" or "unknown," to one of the more descriptive modes. For these type shipments in 2007, "other" and "unknown" modes were generally acceptable. During the 2012 and 2017 CFS mileage calculation operations, a review of these "other mode" shipments was conducted. This analysis showed there to be few truly "other mode" shipments. Such shipments were often transported via conveyor belts. Table 26 compares the value and tonnage estimates for the other-type modes in the 2007, 2012, and 2017 releases.
- The reduction in the "other multiple modes" totals in 2012 was largely due to the addition of the "multiple waterways" mode category in 2012. In 2007, these shipments (with two or more water modes) would have been classified into "other multiple modes." The increase in "other multiple modes" in 2017 is largely the result of including the U.S. territorial water portion of an export shipment as a domestic mode. For example, export shipments that traveled by truck and rail to a sea port of exit were classified as (domestic) "truck and rail" shipments in 2012. In 2017, these shipments now had three domestic legs (truck, rail, and water) and so were classified into "other multiple modes." See Table 26.

Table 26.
Other Modes of Transportation

Mode code	Mode meaning	2007 (final)		2012 (final)		2017 (final)	
		Value (million dollars)	Tons (thousands)	Value (million dollars)	Tons (thousands)	Value (million dollars)	Tons (thousands)
18	Other multiple modes	45,320	113,841	668	2,452	17,490	8,224
19	Other modes	279,113	271,567	1,026	36,844	2,095	93,634

More details about mileage calculation and related processing can be found in the "Mileage Calculation" section.

Commodity Coding Changes

Several commodities in SCTGs 07, 08, 17, and 18 were redefined in 2012. For the details of these changes, see the "Commodity Coding Changes for 2012" table in the "Data Collection" section of the "2012 CFS Survey Methodology" at <www.census.gov/programs-surveys/cfs/technical-documentation/methodology/methodology-2012.html>. The codes used to display some of these commodities changed for 2017. See Table 8 for the details.

As described in the "Imputation" section, several new processes were developed to impute commodity and HAZMAT codes for 2017.

Application of Noise Infusion

For establishments that were in the survey in both 2017 and 2012, no effort was made to coordinate the direction or magnitude of the noise factor applied to these establishments from one survey to the next. For such an establishment, the random noise multiplier may have been greater than 1.0 in 2017 but less than 1.0 in 2012 or vice versa. See the "Disclosure Avoidance" section for more details.

Sampling Variability and Nonresponse

Through its sample design, the CFS tries to ensure the sample will include shipments originating from establishments in each CFS Area. However, estimates of other shipment characteristics, such as destination, commodity, and mode, depend entirely on the sample of shipments reported by responding establishments. See the "Sample Design" section for further information.

A particular combination of origin, destination, commodity, and mode (for example) may be common one year but rare or nonexistent in the next survey. While this may reflect true changes in economic activity, it may also result from one of the following:

- Failing to include in the 2017 CFS sample, the establishments making these shipments.
- If included, the sampled establishments failing to respond in 2017.
- If responding, failing to include shipments with this particular combination of characteristics in the sample of shipments provided to the Census Bureau.

Estimation

As discussed in the "Estimation" section, the CFS weighting methodology makes use of Economic Census data. At the time they were needed for the 2012 CFS, not all of these 2012 Census data were final so some of the 2012 Census estimates relied upon by the CFS may have been changed after the 2012 CFS estimates were released. No attempt has been made to revise or reweight the 2012 CFS data to account for these subsequent changes.

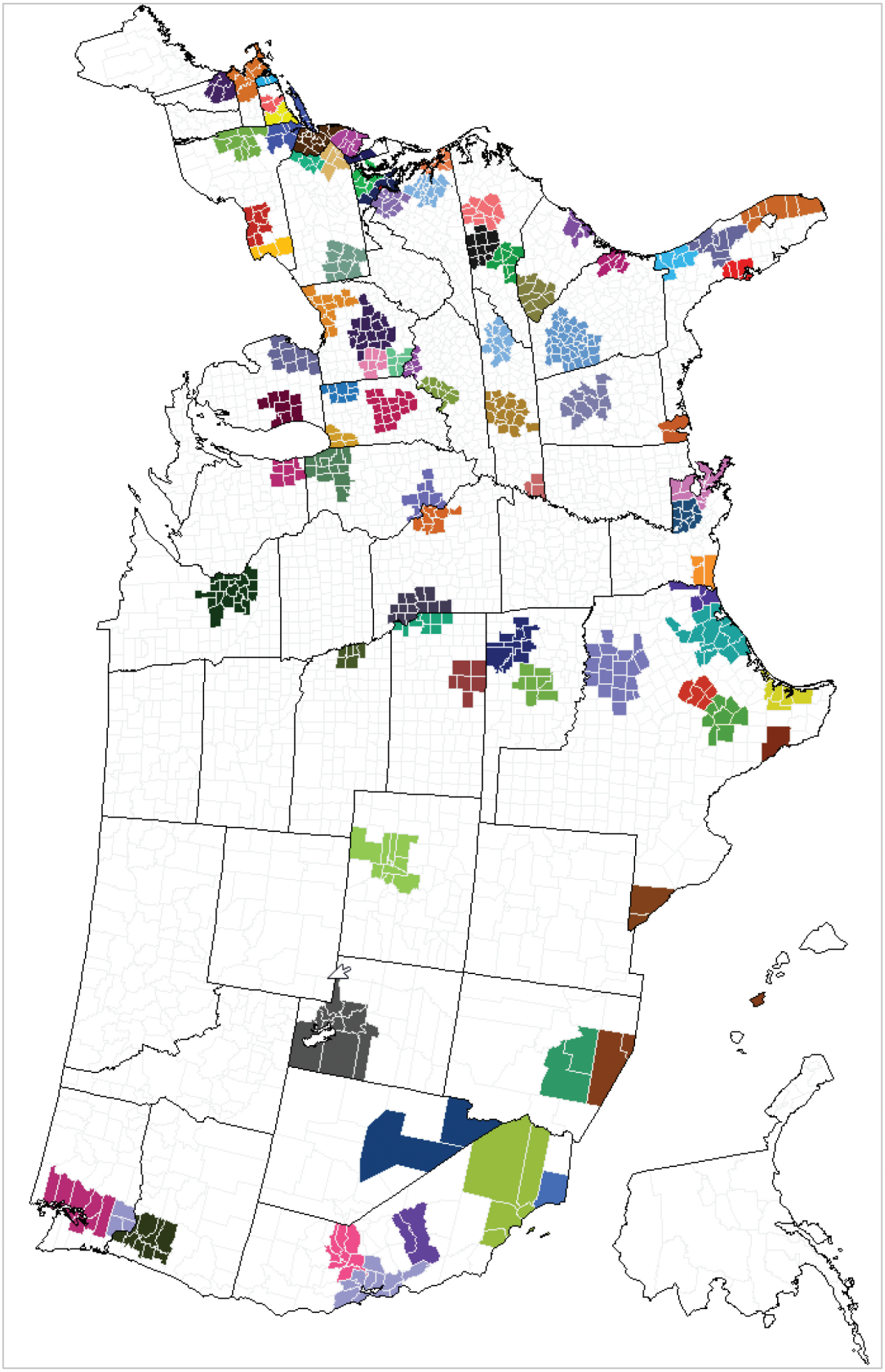

Figure 1.
Map of 2017 CFS Areas

Note: Remainder of State (ROS) CFS Areas are shown in white. See legend on next page.

Listing of 2017 CFS Areas

ST	CFS Area Code	CFS Area Description	Map Color
AL	142	Birmingham-Hoover-Talladega, AL CFS Area	
AL	380	Mobile-Daphne-Fairhope, AL CFS Area	
AZ	38060	Phoenix-Mesa-Scottsdale, AZ CFS Area	
AZ	536	Tucson-Nogales, AZ CFS Area	
CA	260	Fresno-Madera, CA CFS Area	
CA	348	Los Angeles-Long Beach, CA CFS Area	
CA	41740	San Diego-Carlsbad, CA CFS Area	
CA	472	Sacramento-Roseville, CA CFS Area	
CA	488	San Jose-San Francisco-Oakland, CA CFS Area	
CO	216	Denver-Aurora, CO CFS Area	
CT	25540	Hartford-West Hartford-East Hartford, CT CFS Area	
CT	408	New York-Newark, NY-NJ-CT-PA CFS Area (CT Part)	
DE	428	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (DE Part)	
DC	47900	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (DC Part)	
FL	300	Jacksonville-St. Marys-Palatka, FL-GA CFS Area (FL Part)	
FL	370	Miami-Fort Lauderdale-Port St. Lucie, FL CFS Area	
FL	422	Orlando-Deltona-Daytona Beach, FL CFS Area	
FL	45300	Tampa-St. Petersburg-Clearwater, FL CFS Area	
GA	122	Atlanta-Athens-Clarke County-Sandy Springs, GA CFS Area	
GA	496	Savannah-Hinesville-Statesboro, GA CFS Area	
HI	46520	Urban Honolulu, HI CFS Area	
IL	176	Chicago-Naperville, IL-IN-WI CFS Area (IL Part)	
IL	476	St. Louis-St. Charles-Farmington, MO-IL CFS Area (IL Part)	
IN	176	Chicago-Naperville, IL-IN-WI CFS Area (IN Part)	
IN	258	Fort Wayne-Huntington-Auburn, IN CFS Area	
IN	294	Indianapolis-Carmel-Muncie, IN CFS Area	
KS	312	Kansas City-Overland Park-Kansas City, MO-KS CFS Area (KS Part)	
KS	556	Wichita-Arkansas City-Winfield, KS CFS Area	
KY	178	Cincinnati-Wilmington-Maysville, OH-KY-IN CFS Area (KY Part)	
KY	350	Louisville/Jefferson County-Elizabethtown-Madison, KY-IN CFS Area (KY Part)	
LA	12940	Baton Rouge, LA CFS Area	
LA	324	Lake Charles-Jennings, LA CFS Area	
LA	406	New Orleans-Metairie-Hammond, LA-MS CFS Area (LA Part)	
MD	12580	Baltimore-Columbia-Towson, MD CFS Area	
MD	47900	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (MD Part)	
MA	148	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (MA Part)	
MI	220	Detroit-Warren-Ann Arbor, MI CFS Area	
MI	266	Grand Rapids-Wyoming-Muskegon, MI CFS Area	
MN	378	Minneapolis-St. Paul, MN-WI CFS Area (MN Part)	
MO	312	Kansas City-Overland Park-Kansas City, MO-KS CFS Area (MO Part)	
MO	476	St. Louis-St. Charles-Farmington, MO-IL CFS Area (MO Part)	
NE	420	Omaha-Council Bluffs-Fremont, NE-IA CFS Area (NE Part)	
NV	332	Las Vegas-Henderson, NV-AZ CFS Area (NV Part)	

ST	CFS Area Code	CFS Area Description	Map Color
NH	148	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (NH Part)	
NJ	408	New York-Newark, NY-NJ-CT-PA CFS Area (NJ Part)	
NJ	428	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (NJ Part)	
NY	104	Albany-Schenectady, NY CFS Area	
NY	160	Buffalo-Cheektowaga, NY CFS Area	
NY	408	New York-Newark, NY-NJ-CT-PA CFS Area (NY Part)	
NY	464	Rochester-Batavia-Seneca Falls, NY CFS Area	
NC	172	Charlotte-Concord, NC-SC CFS Area (NC Part)	
NC	268	Greensboro-Winston-Salem-High Point, NC CFS Area	
NC	450	Raleigh-Durham-Chapel Hill, NC CFS Area	
OH	178	Cincinnati-Wilmington-Maysville, OH-KY-IN CFS Area (OH Part)	
OH	184	Cleveland-Akron-Canton, OH CFS Area	
OH	198	Columbus-Marion-Zanesville, OH CFS Area	
OH	212	Dayton-Springfield-Sidney, OH CFS Area	
OK	416	Oklahoma City-Shawnee, OK CFS Area	
OK	538	Tulsa-Muskogee-Bartlesville, OK CFS Area	
OR	440	Portland-Vancouver-Salem, OR-WA CFS Area (OR Part)	
PA	408	New York-Newark, NY-NJ-CT-PA CFS Area (PA Part)	
PA	428	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area (PA Part)	
PA	430	Pittsburgh-New Castle-Weirton, PA-OH-WV CFS Area (PA Part)	
RI	148	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (RI Part)	
SC	16700	Charleston-North Charleston, SC CFS Area	
SC	273	Greenville-Spartanburg-Anderson, SC CFS Area	
TN	314	Knoxville-Morristown-Sevierville, TN CFS Area	
TN	368	Memphis-Forrest City, TN-MS-AR CFS Area (TN Part)	
TN	400	Nashville-Davidson-Murfreesboro, TN CFS Area	
TX	12420	Austin-Round Rock, TX CFS Area	
TX	13140	Beaumont-Port Arthur, TX CFS Area	
TX	204	Corpus Christi-Kingsville-Alice, TX CFS Area	
TX	206	Dallas-Fort Worth, TX-OK CFS Area (TX Part)	
TX	238	El Paso-Las Cruces, TX-NM CFS Area (TX Part)	
TX	288	Houston-The Woodlands, TX CFS Area	
TX	29700	Laredo, TX CFS Area	
TX	41700	San Antonio-New Braunfels, TX CFS Area	
UT	482	Salt Lake City-Provo-Orem, UT CFS Area	
VA	40060	Richmond, VA CFS Area	
VA	47900	Washington-Arlington-Alexandria, DC-VA-MD-WV CFS Area (VA Part)	
VA	545	Virginia Beach-Norfolk, VA-NC CFS Area (VA Part)	
WA	440	Portland-Vancouver-Salem, OR-WA CFS Area (WA Part)	
WA	500	Seattle-Tacoma, WA CFS Area	
WI	376	Milwaukee-Racine-Waukesha, WI CFS Area	
XX	99999	Remainder of State XX (where "XX" is the state postal code)	