2017 CFS Survey Methodology

Released: Dec 2018

Table of Contents

OVERVIEW

The Commodity Flow Survey (CFS) is a joint effort by the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau, U.S. Department of Commerce. The survey is the primary source of national and sub-national 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. 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 five 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 most recently in 2017.

CFS data are used by policy makers 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 for the 2017 CFS was to estimate shipping volumes (value, tons, and ton-miles) by commodity and mode of transportation at varying levels of geographic detail. A secondary 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 for 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-establishments 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. The groups included in this advance survey were:

2016 CFS Advance Survey Composition

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

For the first four groups in the table above (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 that they did not conduct any shipping activity were excluded from

the eventual CFS sample universe. For the other categories the objective was to obtain an accurate measure of their shipping activity as well as contact information.

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 (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 and retail fuel dealers.

The (2012) NAICS industries covered in the 2017 CFS are listed in the following table:

NAICS Industries In-scope 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 equip. 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[3]	Newspaper, Periodical, Book, and Directory Publishers
551114[4]	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)   For 2017 this industry includes NAICS 51223. In the 2012 cycle, NAICS 51223 was not sampled.
(4)   Includes only those establishments in the industry with shipping activity as determined from the advance survey

Excluded industries: Establishments classified in transportation (other than freight trucking and warehousing), construction, and most retail and services industries are excluded. Other industry areas that are not covered, but may have significant shipping activity, include agriculture and government. For agriculture, specifically, the CFS does not cover shipments of agricultural products from the farm site to the processing centers or terminal elevators (most likely short-distance local movements), but does cover the shipments of these products from the initial processing centers or terminal elevators onward.

General exclusions: Data for government-operated establishments are excluded from the CFS. These include public utilities, publicly-operated bus and subway systems, public libraries, and government-owned hospitals. The CFS also excludes establishments or firms with no paid employees and foreign establishments.

SHIPMENT COVERAGE

The CFS captures data on shipments originating from select types of 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 for shipment to another location. Shipments that are shipped 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 or 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, 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 (manufacturers' sales offices, agents and brokers, and certain importers) were excluded from the frame.

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.

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 stores) were also excluded from the sampling frame. The resulting frame comprised approximately 710,500 establishments as shown in the table below.

CFS Frame Summary Statistics

Trade Area	Establishments on the 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 was computed. The measure of size was designed to approximate an establishment's annual total value of shipments for the year 2014.

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

The table below shows the counts of these three types of CFS areas.

Stratification

The sampling frame was stratified by geography, industry, and measure-of-size (MOS) class (with some exceptions for auxiliary establishments and hazardous materials establishments, as described below). The geography by industry cells form the primary strata for the main part of the sample.

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 2015 Office of Management and Budget's definitions (OMB Bulletin 15-01). All other metropolitan areas were collapsed with the non-metropolitan areas within the state into Remainder of State (ROS) CFS Area strata. When a metropolitan area (MA) crossed state boundaries, we considered the size of each state part of the metropolitan area when determining whether or not to create strata in each state in which the MA was defined. For example, the Chicago CSA was split into two CFS Areas: the IL part and the IN part. The WI part of Chicago was considered too small to be a separate CFS Area and was combined into the Remainder of Wisconsin CFS Area. The table below (second column) summarizes the number of CFS Areas used for sampling and publication by type.

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 the state (ROS) [1]	35
Whole state (AK, AR, ID, IA, ME, MS, MT, NM, ND, SD, VT, WV, WY)	13
Total number of CFS Areas	132

Note: (1) Three states do not have a Remainder of State (ROS) component. These are DC, NJ, and RI.

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 mining (four-digit NAICS).
- Twenty-one manufacturing (three-digit NAICS).
- Eighteen wholesale (four-digit NAICS).
- Two retail (NAICS 4541 and 45431).
- One services (NAICS 5111 and 51223 combined).
- Three auxiliary (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. For auxiliary establishments that did not respond to the Advance Survey, two separate sets of strata were created as follows:

- 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).

In order to produce good estimates of shipments of hazardous materials (HAZMAT), twenty-one 6-digit NAICS industries with high amounts of HAZMAT shipments were identified and used 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 hazardous materials.
- A high percentage of their (unweighted) shipments were HAZMAT shipments.

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

The table below shows the number and types of primary strata for the main, auxiliary, HAZMAT and special certainty components of the sample. Note that these are the number of strata before they are further stratified by measure of size (MOS) size class.

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 non-responders – NAICS 484 & 4931	130	1,268
	Advance survey non-responders – 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

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

The total desired sample size for the first stage sample was approximately 100,000 establishments 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).

- 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 the Neyman allocation, since the Neyman allocation minimizes the sample size needed to achieve a target CV.
- Once the minimum sample sizes for each primary stratum were determined, these were added together and compared to the desired total sample size of 100,000. If the total was not close enough to 100,000, we multiplied all of the target CVs by a fixed factor and repeated the process until the total sample size was close to 100,000.
- 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 the table below).

2017 CFS Frame & Sample Summary Statistics

| Primary Strata Type | 2017 Frame | | 2017 Sample | | | |
| | Estabs | Total MOS ($mil) | Total Sample | | Certainty Component | |
			Estabs	MOS of Sampled Estabs ($mil)	Estabs	MOS of Certainty Estabs ($mil)
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 Cert	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

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 measure-of-size. Each week of the quarter had 7,990 or 7,991 establishments assigned to it.

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. The table below 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.

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	21	40
601 – 3,000		30	60
3,000 or more		50	80

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. Larger establishments (approximately 70% of the sample), determined by measure of size, were mailed a letter and were instructed to report electronically through the online instrument. Smaller establishments (approximately 30% 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% of returned questionnaires were electronic using the online instrument and nearly 8% were returned on a paper questionnaire. A small number (approximately 3%) 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 of whether the shipment was temperature controlled.
- United Nations or North American (UN/NA) number for hazardous material shipments.
- U.S. destination (city, state, ZIP code)—or gateway for export shipment.
- Modes of transport.
- An indication of whether the shipment was an export.
- City and country of destination for exports.
- Export mode.

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.

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. These are:

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% alcohol by volume was included in Commodity Code 083. In the 2012 CFS, denatured alcohol of more than 80% 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% alcohol by volume were included in Commodity Code 0831. In the 2012 CFS, denatured alcohol of more than 80% 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% 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.

IMPUTATION OF SHIPMENT VALUE OR WEIGHT

To correct for nonresponse or an unacceptable value in either the value or weight item for a given shipment, the missing or unacceptable value 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 missing or unacceptable 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 and usable shipment value and weight). The donor pools are summarized below in order of preference (the lowest numbered donor pool containing a matching shipment is used).

CFS Shipment Value and Weight Imputation Cell Descriptions

Donor Pool	Description of Donor Pool Shipments
1	From same establishment and in the same **detailed** shipment size class
2	From same company and in the same **detailed** shipment size class
3	From same geographic area and in the same **detailed** shipment size class
4	From same establishment and in the same **broad** shipment size class
5	From same company and in the same **broad** shipment size class
6	From same geographic area and in the same **broad** shipment size class
7	From same establishment *(no restriction on shipment size)*
8	From same company *(no restriction on shipment size)*
9	From same geographic area *(no restriction on shipment size)*

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 failed 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.

Approximately 390,000 shipments had either their value or weight imputed.

DESTINATION ZIP CODE CORRECTION AND IMPUTATION

A shipment's origin and destination ZIP code are the primary inputs to determining the shipment's distance traveled (see Mileage Calculation below). 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 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.

For certain shipments with missing destination ZIP codes, a value 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:
- 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 the first stage 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 shipment destination ZIP codes were imputed in this process.

MILEAGE CALCULATION

The CFS does not ask respondents to report the distance traveled for each shipment. However, origin and destination ZIP code, transportation mode, commodity, and foreign country (if applicable) are required from

respondents. Using these variables, a mileage estimate can be provided. To calculate a 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 is a combination of 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 based on observed data from the Surface Transportation Board's Waybill Sample data. Please see more about rail under "Methodological Changes to Mileage Calculation for the 2017 CFS".

Water

The latest Unites 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 routes generated 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 classification which includes Inland Water (i.e. shallow draft) , Deep Sea (i.e. deep draft), Great Lakes, and Multiple Waterways (shipments involving a transfer between shallow draft and deep draft vessels). Please see more about water under "Methodological Changes to Mileage Calculation for the 2017 CFS".

Air

The air network was built by Bureau of Transportation Statistics (BTS) personnel using BTS' Office of Airline (OAI) data. The air network consists of air routes that have regular air freight service. This includes the networks of the three largest freight 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, airport and air route volume, and air carrier. Please see more about air under "Methodological Changes to Mileage Calculation for the 2017 CFS".

Multimodal Shipments

For multi-mode shipments (i.e. shipments involving more than one mode, such as truck-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 is calculated between the origin ZIP code and the border crossing point. For shipments to other foreign locations, the mileage is 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 is

not counted. In both cases, a Port of Exit (POE), either seaport, airport, or border crossing point, is found based on an established order of processes. Please see more about exports under "Methodological Changes to Mileage Calculation for the 2017 CFS".

ZIP Codes

The source of ZIP codes in GeoMiler is from 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. Please see more about ZIP code point placement under "Methodological Changes to Mileage Calculation for the 2017 CFS"

For intra-ZIP 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

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 or destination ZIP 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-Rail" and "Truck-Water" in the CFS estimates.

In 2012, the nearest rail station or dock was selected regardless of the commodity and volume of the facility.

Use of Designated Transfer Points

Rail

For 2017, based on observed rail shipment data, a set of Class 1 railroad transfer points (interlining) were identified and used by GeoMiler when necessary. If the selected origin and destination stations were owned by separate Class 1 owners, and did not share trackage rights, then a transfer was deemed necessary. Under such a scenario, GeoMiler would select the most likely transfer point based on the type of transfer (e.g. NS to

UP) where order matters, the volume, and distance. In 2012, transfers were allowed to occur at any railroad junction, regardless of station owner and trackage rights.

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 selected origin and destination dock, and 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 occurrence of "Multiple Waterway" shipments was more likely as a switch in water modes (i.e. inland water to deep sea) was solely based on the classification of the USACE waterway network links. The origin and destination, and shipment weight, were not taken into account. For 2017, to provide a more accurate picture of shipping patterns, we thought it was best to place a higher value on the geography (origin and destination) and shipment weight versus the classification of the links embedded in the waterway network.

Air Routing

For 2017, the air network was divided into four subnetworks and the airport coverage was increased. Using observed data, we could distinguish the routes by passenger/freight carriers and by each of the three largest air freight parcel carriers. If the respondent indicated parcel-air in 2017, the shipment this time was kept on the same parcel air network from origin to destination. In 2012, air shipments would be susceptible to jumping on and off numerous air carrier links between origin and destination. 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 sub networks were considered. Again, the shipment would stay on the same air carrier network (i.e. subnetwork) between origin airport and destination airport.

Exports

For 2017, an order of processes was established to determine the best routing for export shipments. GeoMiler would first check 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 found to be within short distance, GeoMiler would use the nearest POE.

If the first two options failed, provided POE information was missing or invalid, and the proximity rule did not apply, 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 foreign trade export data.

In comparison to 2012, the prior 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 POE's to route to was expanded from 2012 based on the observed export data.

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 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. 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.

Like establishments, we identified shipments as either certainty or noncertainty. (See the Nonsampling Error section below for a description of how certainty shipments were identified.) For noncertainty shipments, 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, 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. (A usable shipment is one that has valid entries for value, weight, and origin and destination ZIP Codes.) To account for these unusable shipments, we applied the **shipment nonresponse weight.** For noncertainty shipments from a particular establishment's reporting week, 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 shipment weight for certainty shipments from a particular establishment's reporting week is equal to one.

The **quarter weight** inflates an establishment's estimate for a particular reporting week to an estimate for the corresponding quarter. For noncertainty shipments, the quarter weight is equal to 13. The quarter weight for most certainty shipments is also equal to 13. However, if a respondent was able to provide information about all large (or certainty) shipments made in the quarter containing the reporting week, then the quarter weight for each of

these shipments was set to one. 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 or nonsampling 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 weight**. This weight is equal to the reciprocal of the establishment's probability of being selected into the first stage sample (see Sample Design below).

A final adjustment, for most industries, the **nonresponse post-stratification adjustment weight**, adjusts the weighted shipment value (using all prior weighting factors) to the tabulated revenue data from other Census Bureau sources. This 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 time the first-stage sampling frame was constructed (2016) and the year in which the data were collected (2017).

For the preliminary 2017 CFS estimates, the nonresponse post-stratification cells were defined by industry categories, typically by 3-digit NAICS codes (for Manufacturing) or 4-digit NAICS codes (all other industries). There were approximately 45 nonresponse post-stratification cells. The other Census Bureau sources for the adjustment data were:
- 2016 County Business Patterns
- 2017 Manufacturers' Shipments, Inventories, and Orders
- 2017 Monthly Wholesale Trade Survey
- 2016 Annual Wholesale Trade Survey
- 2017 Monthly Retail Trade Survey

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 errors are difficult to measure and can be introduced through inadequacies in the questionnaire, nonresponse, inaccurate reporting by respondents, errors in the application of survey procedures, incorrect recording of answers, and errors in data entry and processing. In conducting the 2017 CFS, every effort has been made to minimize the effect of nonsampling errors on the estimates. Data users should take into account both the measures of sampling error and the potential effects of nonsampling error when using these estimates.

Suppressed Estimates

Estimates that had high sampling variability or poor response quality 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.

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

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

SAMPLING ERROR

Because the estimates are based on a sample, exact agreement with results that would be obtained from a complete enumeration of all shipments made in 2017 from all establishments included on the sampling frame using the same enumeration procedures is not expected. 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 used in this survey is one of a large number of samples of the same size that could have been selected 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. For the CFS, the coefficient of variation also incorporates the effect of the noise infusion disclosure avoidance method (see Disclosure Avoidance below). 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% and 95% confidence intervals, are taken from the t-distribution with nine degrees of freedom. This takes into account the uncertainty in the estimates of the CVs produced 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 shipments 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.

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 many sources:
- Response errors.
- Differences in the interpretation of the questions.
- Mistakes in coding or keying the data obtained.
- Other errors of collection, response, coverage, and 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 incorporate this information into 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.

The respondents who had reported a shipment with unusually large value or weight when compared to the rest of their reported shipments were often contacted for verification. In such cases, if we were able to collect information on all of the large shipments a respondent had made either for a particular reporting week or for the entire quarter, we then identified those large shipments as certainty shipments.

Nonresponse

A potential source of bias in the estimates is nonresponse. Nonresponse is defined as the inability to obtain all the intended measurements or responses from all units in the sample. Four 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, which is the procedure by which a missing value is replaced by a predicted value obtained from an appropriate model. (See above 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 corrected by reweighting (see the descriptions of the shipment and quarter nonresponse weights in the Estimation section above). 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 corrected 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 1 below (along with the final values from the 2012 survey).

Table 1: 2017 CFS Preliminary Unit Response Rates

Type of Response Rate	2017 (Prelim)	2012 (Final)
Participation	55.6%	57.0%
Unit	n/a [1]	66.1%
Weighted Unit	n/a	76.7%

Notes: (1) Some of the quantities required to compute other unit response rates will not be available until the final
        release in Dec 2019.

Participation Response Rate (PRR) - The Participation Response Rate is the total number of unweighted establishments that provided usable[1] data divided by the total number of establishments in the sample (103,877) (expressed as a percentage).

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 measure of size of the establishments that provided usable data to the total weighted sampling measure of size 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 2 below (along with the final values from the 2012 survey).

Table 2: 2017 CFS Preliminary Total Quantity Response Rates

CFS Variable	2017 (Prelim)	2012 (Final)
VALUE	n/a [2]	51.9%
TONS	n/a	50.9%
TON-MILES [1]	n/a	63.2%

Notes: (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 above). 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.
(2) The quantities required to compute the total quantity response rates will not be available until the final release in Dec 2019.

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.

---

[1] "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-FY18-349).

COMPARABILITY OF ESTIMATES

This section summarizes the coding and processing differences between the 2017 and 2012 (and 2007) 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 any survey years.

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, 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. The table below lists the CFS Areas that changed from 2012 to 2017.

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

Notes: (1) The Alabama, Texas, and Louisiana Remainder of State CFS Areas lost the counties added to the CFS Areas described above

Industry Changes

Industry coverage has changed slightly from survey year to survey year. The details of CFS industry coverage are described in the **Industry Coverage** section of the survey methodology. 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 2012 estimates were based on the industry classification of the sample establishments at the time those estimates were produced (Dec 2014). The 2012 and earlier estimates are never revised to account for subsequent industry classification changes to the sample establishments.

Mode Changes

There were no changes to the detailed mode of transportation codes associated with water-borne shipments. The table below lists the water modes in 2007, 2012, and 2017. In addition there were slight changes to the definitions of modes 08 and 10 in 2012 that may have affected the respondent's choice of answer. See the 2007, 2012, and 2017 questionnaires and instruction guides at www.census.gov/programs-surveys/cfs/technical-documentation/questionnaires.html for descriptions of the modes.

CFS 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, certain export shipments that travelled by truck to the port of embarkation and then by ship to the foreign destinations were classified as single-mode truck shipments in 2012 and their domestic water mileage to the US 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 US border.

For 2017, the mode category, "Private Truck" has been renamed "Company-owned Truck".

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 1000 pounds. Shipments with weights above the maximum were re-assigned to a non-Parcel mode, usually a truck mode.
- For 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 (called "Private truck" in 2012) were not routed more than 500 miles during 2012 and 2017 mileage calculation. 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 codes. 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 a few truly "Other mode" shipments. Such shipments were often transported via conveyor belts. The table below compares the value and tonnage estimates for the Other-type modes in the 2007, 2012 and preliminary 2017 releases.

"Other" Modes of Transportation

Mode		2007 (Final)		2012 (Final)		2017 (Prelim)	
Code	Meaning	Value ($mil)	Tons (000)	Value ($mil)	Tons (000)	Value ($mil)	Tons (000)
18	Other multiple modes	45,320	113,841	668	2,452	0	0
19	Other modes	279,113	271,567	1,026	36,844	105	6,671

More details about mileage calculation and related processing can be found in the **Mileage Calculation** section of the survey methodology.

Routing Software Changes

The underlying transportation network software used to model shipment distances was updated to reflect changes to the transportation infrastructure between 2012 and 2017. In particular:
- The 2012 ZIP codes were replaced with 2017 ZIP codes.
- Other changes are as described in Methodological changes to Mileage Calculation for the 2017 CFS above

Commodity Coding Changes

Several commodities in SCTGs 07, 08, 17, and 18 were redefined for 2012. See the *Commodity Coding Changes for 2012* table in the **Data Collection** section of the 2012 CFS survey methodology for the details of these changes. The codes used to display some of these commodities have changed for 2017. See the table, *SCTG Code Changes* above for the details.

For 2017, the CFS used a machine learning process 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 did code themselves. This model output the highest-likelihood SCTG code using two input variables: first, the NAICS code of the establishment from which the shipment record came, and second, the description (as a "bag-of-words") from each record. Using the model's reported prediction probability as a guide, we took a sampling of 750 records that did not have an SCTG code, and had expert analysts validate the model's predictions on these records. From this validation exercise, we were able to assign an SCTG code to approximately 106,000 shipments with a high degree of confidence using the model's output.

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 above 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 sections of the survey methodology for further information.

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

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