

pitney bowes



U.S. Demographic & Business Summary Data

Product Guide



Information in this document is subject to change without notice and does not represent a commitment on the part of the vendor or its representatives. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, without the written permission of Pitney Bowes Inc.

© 2000-2016 Pitney Bowes Inc. All rights reserved.

Adobe Acrobat® is a registered trademark of Adobe Systems Incorporated in the United States.

Other products named herein may be trademarks of their respective manufacturers and are hereby recognized. Trademarked names are used editorially, to the benefit of the trademark owner, with no intent to infringe on the trademark.

Further Information

Corporate Headquarters - Pitney Bowes Inc., 350 Jordan Rd, Troy, NY 12180 USA

Voice: 800.327.8627

Technical Support Hotline: 1 800-367-6950

Technical Support Email: software.support@pb.com

<http://www.pitneybowes.com/us>

Table of Contents

Chapter 1: Introduction.....	4
Chapter 2: GroundView® Demographic Data	6
What is Demographic Data?	6
Product Description.....	7
Data Bundles.....	7
Update Profile.....	7
Product Specifications.....	8
Coverage Area	8
Reference Date	8
Update Frequency.....	8
Number of Variables	8
Number of Records by Geographic Layer	8
Methodology Statement – Demographic Data	9
Overview	9
Bottom-Up Phase.....	9
Population and Household Characteristics	9
2010 American Community Survey Enhanced	10
Household Income	10
Consumer Expenditure Potential	11
Retail Sales Potential	12
Household Wealth (Net Worth) and Financial Assets.....	12
Daytime Population	13
Socio-Economic Score (SES)	13
ZIP Code Data.....	13
Puerto Rico Demographic Estimates and Projections	14
Conclusion and Caveats	14
Chapter 3: GroundView®	15
Business Summary Data.....	15
What is Business Summary Data?	15
Product Specifications.....	16
Coverage Area	16
Reference Date	16
Update Frequency.....	16
Number of Variables	16
Methodology Statement – Business Summary Data	16
Conclusion and Caveats	17

Chapter 1: Introduction

Pitney Bowes demographic Estimates and Projections suite of data for the United States and Puerto Rico is an accurate, current and comprehensive data product containing more than 8,200 variables across multiple layers of geography including United States Postal Service (USPS) ZIP Codes and census block groups. Current year estimates and five-year projections were produced using data-driven and innovative methodologies by a team of demographers, statisticians and geographers with extensive industry experience. Pitney Bowes data are used to understand, estimate, project, compare and differentiate the demographic, economic, geographic and business characteristics of markets – including by customized trade areas.

Estimates and projections have a July 1st reference date and are based on current geography. The datasets are available in multiple file formats including file formats compatible with Pitney Bowes software including MapInfo Professional® and Spectrum®.

The data include:

- Pitney Bowes Estimates (CY) – Current year estimates of population, households and their demographic and economic characteristics. Source: Pitney Bowes; Gadberry (MicroBuild).
- Pitney Bowes Projections (5Y) – Five-year projections of population, households and their demographic and economic characteristics. Source: Pitney Bowes.
- 2010 Census (BY) – Census 2010 data from Summary File 1 (SF1) provide the most detailed information from the census questionnaire that was distributed to all housing units. Source: U.S. Census Bureau.
- 2010 American Community Survey Enhanced (BY) – Selected data from the 5-year American Community Survey (ACS; 2008-2012). The ACS is the largest household survey in the federal statistical system. Geo-statistical algorithms were applied to small-area ACS estimates - and distributed to 2010 Census counts - to enhance survey accuracy. Source: U.S. Census Bureau; Pitney Bowes.
- 2000 Census (2K) – Selected Census 2000 data – from both the short-and long-form – available in current geography. Source: U.S. Census Bureau; Pitney Bowes.
- Consumer Expenditure Potential (CY) and Consumer Expenditure Potential (5Y) - provides estimates and projections of aggregate household expenditures for consumer goods including food, automobiles and insurance. Source: U.S. Bureau of Labor Statistics; Pitney Bowes.
- Business Summary Data (CY) – Based on about fourteen (14) million businesses whose addresses were geocoded to the roof-top and street-address; business summary data provided – by NAICS and SIC industrial classifications – counts of business establishments, employment, occupation, payroll, and retail sales for USPS ZIP codes, Designated Market Areas (DMAs) and standard census geographies. Source Pitney Bowes; GeoResults.

This document describes the content of the data and how the data were produced. Methodology statements are provided for Census data and for demographic Estimates, Projections, and Business Summary Data. A complete list of variables with detailed field descriptions is available in the companion worksheet – [USDemographic&BusinessSummaryDatayyy_variables.xls](#). The number of variables and number of records per unit of geography are also included in the accompanying Excel worksheet.

Chapter 2: GroundView® Demographic Data

What is Demographic Data?

Demographic data are descriptors and characteristics of people and places, for a given timeframe. The demographic data described in this document begins with people but are summarized for places and small areas like neighborhoods. For example, Census data are collected from people filling out census questionnaires and tabulated by the Census Bureau for places and geographic areas.

Demographic data include various household and population characteristic including estimates of income, age, race and ethnicity for Block Groups, other larger areas of Census geography, ZIP Codes, and Designated Market Areas (DMA).

The demographic data described in this document include summary counts, estimates, projections, distributions and other statistics for geographic areas (units). For example, average household size refers to the average number of persons living in housing units within a particular census tract or ZIP Code.

Demographic data are vital inputs into business decisions and marketing strategies, providing insight on current and prospective consumers.

Product Description

Pitney Bowes U.S. GroundView® Demographic Data for the United States and Puerto Rico represents a comprehensive set of over 8,200 demographic and socio-economic variables, provided for 11 layers of geography (nine for Puerto Rico) from block group up to national totals.

All estimates and projections are updated annually to the current year and most variables are projected five years. The data are updated annually using the latest input data sources from the most authoritative sources including the US government and private sector businesses that specialize in small area data.

A complete list of variables with detailed field descriptions is available in the companion document – [USDemographic&BusinessSummaryDatayyyy_variables.xls](#) – which accompanies this document. The database content is a series of topic tables grouped into data bundles. Please see the “Bundles” spreadsheet of the accompanying Excel workbook. The table below contains summary counts of variables by bundle and reference year - the base or universe variables (i.e. population, households, etc.) are duplicated on the data bundles where required.

Data Bundles

Data Bundles and Variable Counts (number of variables per bundle by reference year)	2K	BY	CY	5Y
Total Unique (non-duplicated) Census and Census Based Variables	1,632	1,639	1,845	885
GroundView®: Population Bundle	349	350	366	103
GroundView®: Population by Race Bundle	501	501	698	483
GroundView®: Population by Hispanic/Not Hispanic Bundle	166	166	166	138
GroundView®: Family Bundle	56	58	58	
GroundView®: Housing Bundle	278	281	272	
GroundView®: Socio-Economic Bundle	128	128	130	
GroundView®: Household Income Bundle	172	173	173	173
GroundView®: Update Profile Bundle	2	2	183	2
Total Unique (non-duplicated) Non-Census Based Variables			1,481	806
GroundView®: Household Financial Assets and Wealth Bundle			27	27
GroundView®: Consumer Expenditure Potential Bundle			751	751
GroundView®: Retail Sales Potential Bundle			37	37
GroundView®: Business Summary - NAICS Bundle & SIC Bundle			675	

Update Profile

The Pitney Bowes U.S. Update Profile dataset contains a selection of current year estimate variables from the larger Estimates and Projections data suite – in the form of percent distributions, and median and averages (means) where applicable (i.e., Household Income). Demographic base variables such as total population and households are included, as well as characteristics of population (i.e. age and sex composition) and households (i.e. household income). The number of age groups and income groups are condensed when compared to the E&P datasets from which they are derived. The variables in the database are described in in the companion document – [USDemographic&BusinessSummaryDatayyyy_variables.xls](#).

Product Specifications

Coverage Area

The U.S. GroundView[®] Demographic database covers the 50 states, District of Columbia and the Commonwealth of Puerto Rico.

Reference Date

Estimates and projections have a mid-year, July 1 reference date. Census 2000 and Census 2010 data have an April 1 reference date.

Update Frequency

Annual.

Number of Variables

There are over 8,200 variables in the combined Census 2000 (2K), Base Year 2010 Estimates (BY) and GroundView[®] Estimates and Projections (CY and 5Y) data sets. The reference dates – 2000 (2K), 2010 (BY), Current Year Estimates (CY), and Five Year Projections (5Y) are provided in a table within the accompanying variable listing document [USDemographic&BusinessSummaryDatayyyy_variables.xls](#). The number of variables in each data bundle by reference year is also provided in the Product Description on the page above.

Number of Records by Geographic Layer

The table of record counts by geographic unit for the standard levels of geography released for US Groundview is included on the Introduction worksheet of [USDemographic&BusinessSummaryDatayyyy_variables.xls](#).

Geographic Layers include:

Code	United States Geographic Layers	Puerto Rico Geographic Layers
BG	Block Groups, Census Block Groups	Census Block Groups
CT	Census Tract	Census Tract
CO	County (or equivalent)	Municipio, PR county equivalent units
PL	Place, Census Place and equivalents	Zona Urbana or Comunidad
MCD	Minor Civil Division (or equivalent)	Barrio, Pueblo PR primary sub-county units
CBSA	Core Based Statistical Area, CB (OMD 2013)	Core Based Statistical Areas (OMD 2013)
ST	State (includes Washington, DC)	Commonwealth of Puerto Rico
USA	National, United States of America	Not Applicable
ZIP	ZIP Code, USPS ZIP Codes (polygon and rural postal points)	ZIP Code, USPS ZIP Codes (polygon and rural postal points)
ZPLY	ZIP Code, USPS ZIP Codes (polygon only)	ZIP Code, USPS ZIP Codes (polygon only)
DMA	Designated Market Area Nielsen TV Penetration Area	Not Applicable

Methodology Statement – Demographic Data

The methodologies employed in the build process of U.S. Estimates and Projections are a combination of traditional demographic techniques as well as innovative proprietary processes and geostatistical algorithms which enhance data accuracy and relevance. If more information is required beyond what is included in this document, please contact your Pitney Bowes Account Manager or Partner Manager.

Overview

The basic methodology for the demographic estimates and projections combines top-down and bottom-up phases. The “top-down” phase begins at the national, state, and county units of analysis. National, state, and county estimates and projections serve as control totals for the “bottom-up” sub-county estimates and projections. Generally, the unit of analysis for variables is either population, households or housing units or some combination of the three mentioned variables thereof.

The input data sources used to estimate the “top-down” phase include Census Bureau national population projections by age, sex, race, and Hispanic origin, and the latest Census Bureau county population estimates and population estimates by demographic characteristics. United States Postal Service (USPS) delivery statistics data – correlated with Census housing unit and household counts - inform county population and household estimates and projections. Gadberry household counts are used to inform trends in household formation for small areas (Block Groups).

Bottom-Up Phase

The bottom-up phase of the estimation and projection methodology begins with the most recent decennial census, in this case, the Census 2010 block level data. To estimate current household counts, Pitney Bowes uses The Gadberry Group’s annual MicroBuild® household estimates. The MicroBuild® process begins with a multi-sourced national file of approximately 120 million consumer household records. Every record is parcel and street geocoded using Pitney Bowes geocoding technology.

The objective of the bottom-up phase is to produce a preliminary estimate of total households for the current and projected year that accurately reflects both growth and decline at the block group unit of analysis. The results of the bottom-up household estimates and projections are then reconciled with the top-down results through iterative proportional fitting - a mathematical technique employed to ensure that block group household counts, or “bottom-up” households, sum to their parent county, or “top-down”, household counts.

Once the estimate and projection for total households is established, time-series estimates of household size, household population and group quarters are combined and modeled with Census 2010 data to produce population estimates based on changes in households and household characteristics. The demographic estimation methodology employs time-series data at multiple levels of geography from national to neighborhood.

Population and Household Characteristics

Estimates of population characteristics such as the population distributions by age, sex, marital status, race, and Hispanic origin follow a similar top-down bottom-up process. At the block group unit of analysis, the latest and best information from the Census Bureau’s American Community Survey (ACS) – the largest household survey in the US – is integrated with Census 2010 data to produce timely, accurate small-area estimates. Population characteristics are estimated using demographic estimation and projection models to achieve a preliminary set of estimates which are then reconciled

to top-down estimates using iterative proportion fitting.

Please note: To maintain the comparability of demographic groups over time, Pitney Bowes has structured these estimates and projections to align with the 2010 Census. Also, some variables related to cultural background in Census 2010 and variables related to ancestry groups in the ACS are published by the Census Bureau at the census tract level only. This includes detailed origin groups as defined by specific countries for Asians and Hispanics. For these data sets, the census tract proportional distributions have been applied to the underlying block group populations. While the method assumes that the distribution holds constant, change over time for trade areas and higher geographies is reflected in the overall differential growth among block groups.

2010 American Community Survey Enhanced

Census 2010 – one of the shortest Censuses ever recorded, asking households only ten questions - did not include the Census long-form questionnaire that was included during Census 2000. In the previous Census 2000 program, a long form questionnaire was distributed to one in six housing units and collected critical socio-economic data including income and home value. The Census 2000 long form was replaced by the American Community Survey (ACS) – a continuously updated sample survey of the US population. Each year in December, the Census Bureau produces block group data that have a five-year reference period (i.e., 2010-2014). Annual ACS estimates are produced for geographical summary levels (i.e., counties) with populations of 65,000 or more.

To address inconsistencies between the ACS and Census 2010 variables and to overcome the estimation of multidimensional data and sparse samples, Pitney Bowes developed the American Community Survey Enhanced dataset. The mid-year of the five-year period (2008-2012 ACS) is 2010 and will remain as the primary five year ACS input that will inform the 2010 ACS Enhanced “Base Year Estimates” variables going forward. The basic methodology to produce the enhanced 2010 ACS data is:

- Run proprietary geo-spatial and statistical enhancement routines to improve the data from neighboring or higher level geographic levels in place of the survey data.
- Normalize or control results from the five-year period estimates to the 2010 Census base counts. This step assumes that the distributions and summary measures from the ACS five-year period estimates fairly represent characteristics in 2010.

Similarly, current year estimates are updated annually using the normalized ACS data. This process produces an updated count implied by the distribution of the characteristic and the current base population.

Household Income

Pitney Bowes’s income estimates and projections are based on the US Census Bureau’s American Community Survey (ACS) one (1-year) year and five (5-year) year data. The ACS is the largest household survey in the US and asks respondents to provide household income in the past 12-months. For each county and block group, ACS household income data is published as sixteen (16) income categories by four (4) age cohorts.

The Census Bureau defines income in the past-12 months as “the sum of the amounts reported separately for wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income or income from estates and trusts; Social Security or Railroad Retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income”. Household income “includes income of the householder and all other people 15 years and older in the household, whether or not they are related to the householder” (www.census.gov).

To estimate and project household income and household income by age of householder, the latest available time-series data from the ACS data is used for both the 1- and 5-year data. 1-year data are single year estimates (i.e., 2014) and 5-year data are multi-year estimates (i.e., 2010-2014). All counties – regardless of population size - have ACS 5-year estimates, or multi-year estimates. Counties with a total population of 65,000 or more have 1-year estimates, or single-year estimates. Both the ACS 1- and 5-year data are unique and valuable. For example, the 1-year data is timely and the 5-year data is robust. For block groups, there is only 5-year ACS data – there is no 1-year data for block groups. Pitney Bowes employs a top-down (county level) and then bottom-up approach (block group level) to estimate household income and household income by age of householder.

All Pitney Bowes income estimates and projections are projections because the latest available input dataset from ACS is two (2) years old. The most recent ACS data is 2014 data and Pitney Bowes estimates and projections have a vintage of 2016 and 2021, respectively. Pitney Bowes' income estimates and projections are presented in current year US dollars. For example, 2016 median household income represents the money received as if it were 2016; 2021 median household income represents the money received as if it were 2021. Pitney Bowes household income by age of householder estimates and projections are sixteen (16) income categories by seven (7) age cohorts.

Pitney Bowes calculates median and mean (average) income based on the age of householder by household income distribution. Aggregate household income is calculated as average household income multiplied by the total households. Two (2) types of per capita income are calculated. The population approach is calculated as total aggregate income divided by total population (Total Aggregate Income / Total Population). The household approach is calculated as total aggregate income divided by total household population (Total Aggregate Income / Household Population). It should be noted that the total population is larger than the household population.

Consumer Expenditure Potential

The Consumer Expenditure Potential (CEP) dataset provides estimates of aggregate household expenditures for consumer goods including food, automobiles and insurance. Each year, the US Bureau of Labor Statistics (BLS) conducts the Consumer Expenditure Survey. The Consumer Expenditure Survey (CES) is the largest, nationally representative survey that provides consumer expenditures data by demographic characteristic including income, tenure and region. The CES consists of two surveys – the Diary Survey and the Interview Survey. The Diary Survey collects data on everyday purchases like food and gasoline. The Interview Survey collects data about large expenditures and regular purchases. The BLS CES represents most household expenditures categories and is updated continuously to reflect changes in consumer preferences and habits.

CES data is partitioned via a hierarchical schema with implicit nesting. For example, the item “apples” expenditures fall within the “fresh fruit” category which falls within the “fruits and vegetables” category which falls within the “food at home” category which falls within “food” category. For the majority of consumer expenditure variables, Pitney Bowes' CEP dataset follows the item hierarchical schema of the BLS CES. Please refer to the [USDemographic&BusinessSummaryDatayyyy_variables.xls](#) which explicitly documents the hierarchical relationship between the hundreds of consumer expenditure variables.

The most recent CES survey data (2014) is combined with Pitney Bowes' data and a conditional probability model (based on various geo-demographic characteristics) to produce estimated average household expenditures by expenditure type and by block group. The average block group estimates are inflator adjusted to current year levels using inflation estimates from BLS. Users should note that one consistent inflation adjustment is made to all expenditure categories because forecasting inflation by expenditure type is problematic. For example, the inflation factor for computers (technology) may be different than the inflation of apples (agriculture). To obtain aggregate expenditures by block group, the average household expenditures are

multiplied by the total, current year households. For each CES survey, Pitney Bowes provides current year estimates and five year forecasts for all expenditures items provided by CES so long as the data is not missing or too small to yield valuable insight. For example, Pitney Bowes does not estimate the number of dollars that consumer households spend on private airplanes because the baseline CES data are often sparse and/or missing. Pitney Bowes follows BLS definitions for expenditures. Please refer to [USDemographic&BusinessSummaryDatayyyy_variables.xls](#) for a listing of major category CEP definitions.

Retail Sales Potential

Consumers purchase goods and businesses sell goods. The Retail Sales Potential (RSP) dataset provides estimates and projections of consumer potential organized by store type. RSP estimates are demand side estimates in that they are aggregates of the CEP data. The RSP data are not derived from actual retail sales estimates - considered supply-side estimates. CEP statistics are summarized (aggregated) into various RSP store types – for example, Beer, Wine and Liquor Stores. The result is a relative indicator of household retail sales potential by store type. The RSP estimates and projections are useful for comparing trade areas based on relative consumer retail potential. RSP data are based on consumers and not businesses; therefore, business to business (B2B) spending is not included in the RSP data estimates.

For a complete listing of the consumer expenditure variables included in each store type, please refer to the “Appendix B” worksheet in [USDemographic&BusinessSummaryDatayyyy_variables.xls](#). The table contains a listing of merchandise categories typically sold by each store type. The RSP field names are matched with their corresponding Consumer Expenditure Potential codes and descriptions.

Household Wealth (Net Worth) and Financial Assets

The estimation process for the wealth and financial assets begins with an analysis of the Federal Reserve Board’s Survey of Consumer Finance. A series of respondent clusters was developed based on the survey data. A comparable series was developed for all census block groups based on inputs from the PB demographic estimates program.

As household income and home value at the block group level condition estimates of wealth and financial assets within each cluster, adjustments are made to wealth averages and distributions according to income and home value at the block group level. Income estimates alone condition the distributions of Financial Assets. Wealth results are presented as mean and median estimates, as well as household distributions of wealth. Similarly, results of financial assets are presented as means, medians, and distributions. The following are definitions of concepts and component parts:

Definitions:

- Household wealth or net worth is the difference between total assets and total liabilities at the household level. Assets include financial assets, vehicles, primary residence, investment real estate, business assets, and a residual category of non-financial assets. The wealth variables show the distribution of households across levels of wealth. Summary measures are also provided.
- Financial assets Include transaction accounts (for example, checking and savings accounts), certificates of deposit, savings bonds, bonds, stocks, mutual funds, retirement accounts, cash value of life insurance, and a residual category of other managed financial assets. The concept of financial assets is a subset of household wealth. The financial assets variables show the distribution of households across levels of total financial assets. Summary measures are also provided.

The following are the components of Financial Assets (FA):

- Transaction accounts
- Certificates of deposit
- Savings bonds
- Stocks
- Bonds

- Mutual funds
- Retirement accounts
- Cash value of life insurance
- Other managed assets
- All other financial assets

The following are components of Non-Financial Assets:

- Vehicles
- Primary Residence
- Investment Real estate
- Business Assets
- Other Non-financial assets

The following are components of Liabilities:

- Home Mortgage
- Home Equity
- Lines of Credit (secured by Home)
- Installment Loans
- Other Lines of Credit
- Credit Card Balance
- All other Debt

The calculation for Total Assets is Financial Assets plus Non-Financial Assets.

The calculation for wealth (net worth) is Total Assets minus Total Liabilities. (Negative values are set to zero.)

Daytime Population

Daytime population is the estimated number of people that are in a given area during the daytime. Daytime Population has two components: At-Home Population and At-Work Population (total employees). The At-Home Population is the current estimate of the number of persons aged 16+ that are not in the labor force and, therefore presumed to be at home during the day, as well as population under 16 years of age. The At-Work Population is based on the Pitney Bowes' Business Summary Data, which contains estimates of the number of persons who work in the given Block Group. The sum of the At-Home Population and the At-Work Population yields the estimate of the number of persons in the Block Group during the day.

Socio-Economic Score (SES)

The socio-economic score is a comparative index value ranging from 1 to 100 which indicates the overall socio-economic status of an area. Four variables were used to produce the SES score: Median Household Income, Median Home Value, Occupational Level (percent white collar), and Educational Attainment – the percent of the population aged 25+ with educational degrees earned beyond a high school diploma.

Each block group was given a score for each of these categories based on how it ranked against all other block groups nationwide. Once these scores were determined, an overall score for each block group was calculated by combining the individual scores using an un-weighted average. Finally, the overall scores were indexed on the 100-point scale.

ZIP Code Data

ZIP Codes are represented as point locations and polygons in the ZIP geography layer. Point ZIP Codes may represent a business location or a Post Office with PO Boxes used by residential or business customers. A special subset of point ZIP Codes is defined as a residential post office (RPO) where residents pick up their mail at the Post Office because it is efficient or mail delivery to the home may not be possible. For the purpose of providing demographic data for as many ZIP Codes as possible, the PB data

team has traditionally assigned households to RPO's based on USPS delivery counts to those RPO's. In some cases these are rural areas representing a significant proportion of households. The physical location of RPO households is assumed to be the enclosing ZIP Code. The "enclosing ZIP Code" is the polygon ZIP Code which contains the RPO Post Office. The population and household characteristics of RPO households are assumed to mirror those of the population and households of the enclosing ZIP Code. The ZIP layer includes both points and polygons. Demographic data are aggregated to those respective geographies.

Demographic data are also aggregated to the ZIP Code polygon-only layer (ZPLY). Polygon ZIP Codes generally represent areas served by the U.S. Postal Service and are defined for the purpose of efficient mail delivery. For certain applications it is desirable to present 100 percent of the population and households in polygon-only ZIP Codes. For example, a mapping application that requires a visual representation of demographic change and population characteristics for geographic trade areas may require the use of polygon-only ZIP Code data.

Puerto Rico Demographic Estimates and Projections

The companion document [USDemographic&BusinessSummaryDatayyyy_variables.xls](#) identifies which data bundles and variables are available for Puerto Rico.

Conclusion and Caveats

Demographic estimates and projections developed by the Pitney Bowes data team provide a sound basis for market analysis and business planning. Data users understand that uncertainty exists in any estimation and forecasting of small area data. For example, statistical variation can be greater across smaller geographic units and in areas undergoing rapid population change. The methodology described in this document provides a consistent framework and a set of demographic estimates and projections that can be used confidently for making market comparisons across the country.

Chapter 3: GroundView®

Business Summary Data

What is Business Summary Data?

Pitney Bowes Business Summary database contains aggregations of business establishments, employment, occupation, payroll, and retail sales for all standard census geography levels, as well as for ZIP Codes and Designated Market Areas. The primary source for the Business Summary Data is the PB U.S. Business Points File of nearly 14 million businesses in the U.S. The database is derived from continually updated multiple sources, including telephone listings, city directories, annual reports, and 10K filings.

The PB Business Summary Data provides tabulations and estimates using both the North American Industrial Classification System (NAICS) and Standard Industrial Classification (SIC) system. The detailed variable list is contained the companion document:

[USDemographic&BusinessSummaryDatayyyy_Variables.xls](#).

Product Specifications

Coverage Area

The U.S. GroundView[®] Business Summary database covers the 50 States and the District of Columbia.

Reference Date

All data fields represent mid-year values in the sense of annual averages or the trend point for the year, not necessarily the specific values for July 1 of the given year.

Update Frequency

The Business Summary database is updated annually.

Number of Variables

The content of the Business Summary Data is bundled into two data packages: Business Summary NAICS Dataset (1 bundle of variables) and Business Summary SIC Dataset (2 bundles of variables). There are 216 NAICS-based variables and 459 SIC-based variables in the Business Summary Data. The full variable list can be found in: [USDemographic&BusinessSummaryDatayyyy_Variables.xls](#).

Methodology Statement – Business Summary Data

Pitney Bowes Business Summary database contains aggregations of business establishments, employment, occupation, payroll, and retail sales, for all standard census geography levels, as well as for ZIP Codes and Designated Market Areas. The primary source data for the Business Summary data is Pitney Bowes' US Business Points file of over 14 million businesses in the U.S.

Other input data used to benchmark and control the Business Summary data are the Census Bureau's Economic Census, County Business Patterns, ZIP Code Business Patterns, Bureau of Labor Statistics employment and occupations reports, and U.S. Census Bureau's Monthly and Annual Retail Trade Surveys.

The key steps in the methodology include:

- Address geocode each business in the file to its census block group. Pitney Bowes geocodes the 14 million-plus business addresses using geocoding software and additional procedures described in the U.S. Business Points geocoding processes.
- Evaluate and estimate employee counts for each firm or establishment. The estimate of total employees by block group is assessed by an analysis of the size of establishments in various NAICS and SIC industrial categories.
- Aggregate the number of establishments by their corresponding NAICS and SIC industrial codes. The aggregation of establishments by NAICS is compared to the Census Bureau's current Economic Census results and County Business Patterns. Pitney Bowes' business points data includes sole proprietors and other business with no formal employees (including small family-run or single-person businesses). Such firms are called "non-employers" in Census Bureau terminology. Certain industrial categories such as doctors, lawyers, and carpenters are more likely to be non-employer establishments. Sole proprietors and other single-worker establishments nevertheless becomes part of the "total employee" estimate.

- Generate an occupational distribution for each firm based on its employee count and industrial category utilizing BLS (Bureau of Labor Statistics) Occupation by Industry file which indicates the likely distribution of employment counts by occupation within each NAICS industry. These distributions are estimated at the firm level from a matrix of Occupations by Industry provided by the Bureau of Labor Statistics. The OES (Occupational Employment Survey) is updated to reflect trends by type of business according to the NAICS industrial classification.
- Estimate total payroll based on the occupational distribution, the industrial classification, and the County labor market. The payroll model is based on occupation-specific wage and salary estimates for States and Metropolitan Areas from the Bureau of Labor Statistics, supplemented by payroll estimates from the Economic Census and County Business Patterns. Essentially, payroll estimates were derived from average payroll per employee by occupation and industry, established at the lowest possible geographic level. Results are scaled in part by the relative income levels in each county.
- Develop a retail sales estimate for retail establishments only, based on a model of average sales per employee by industrial category and County retail sales trends. The retail sales model produces annual estimates at the establishment level for all retail firms. The four-digit retail classification for both NAICS and SIC is used. The retail sales estimates are “supply-side” estimates, that is, the stores within a given geographic entity (for example, County or ZIP Code) generate sales from a trade area that may or may not coincide with that geographic entity. (Users needing “demand-side” estimates are referred to Pitney Bowes’ Consumer Expenditure Potential database.)

Conclusion and Caveats

Pitney Bowes Business Summary data is generally consistent with the Pitney Bowes U.S. Business Points database. However, differences between “raw” summary aggregations from the point file and the Business Summary data reflect the above enhancements and modeling processes which generates summary estimates within reasonable ranges of published government business statistics. The Business Summary Data is provided in both SIC and NAICS industrial classifications.

The database of U.S. Business is dynamic and changes considerably from year to year, for example changes due to recently formed or relocated businesses. Users should exercise caution when making year-to-year comparisons. The summary measures in this database are primarily recommended for analysis of the type, size and structure of U.S. businesses across multiple geographies and trade areas.