How can we best visualize worker movement throughout the day?

Melanie A. Rapino, William K. Koerber, Brian McKenzie
Journey to Work and Migration Statistics Branch
Social, Economic, and Housing Statistics Division
United States Census Bureau

SEHSD Working Paper No. 2014-01

Prepared for:
Southern Demographic Association Conference
October 23-25, 2013, Montgomery, AL

How can we best visualize worker movement throughout the day?

Melanie A. Rapino, William K. Koerber, Brian McKenzie
Journey to Work and Migration Statistics Branch
Social, Economic, and Housing Statistics Division
United States Census Bureau

SEHSD Working Paper No. 2014-01

Southern Demographic Association Conference
October 23-25, 2013, Montgomery, AL

Introduction

In May 2013, the US Census Bureau released commuter-adjusted daytime population estimates for places, minor civil divisions (MCDs), counties, and states based on the 5-year 2006-2010 American Community Survey (ACS) estimates. These estimates took into account the total population and the number of workers commuting into and out of each area on a daily basis. The commuter-adjusted daytime population estimates provided an overall estimate of population fluctuation in a given area based on commuting throughout the day. This type of information may be useful for emergency management planning, transportation planning, and development.

Since that release, the U.S. Census Bureau conducted research on further delineating the commuter-adjusted daytime population estimates. These new estimates – dynamic daytime population estimates – include an additional dimension, the delineation of the daytime population estimates into 15-minute increments, creating a profile of commuter behavior for a 24-hour period for each county. The ACS variables used by analysts to estimate a workers' daily schedule included: departure time to work, travel time to work, and hours worked in a week. These data contained estimates by county and time of workers at work, at home, and in transit, as well as those who worked at home and those who did not work but are county residents, including children and elderly. Furthermore, visualizations of these data indicated periods of potential increased work-related movement as well as population increases and decreases throughout the day in a given geography.

This research documents the process by which the dynamic daytime population estimates were created, including: the assumptions made and different forms of data visualization. This research includes a series of data visualizations showing the evolution of this research process. The U.S. Census Bureau seeks feedback from conference attendees and data users on the type of data visualization that best communicates the data as well as the amount of detail that is necessary and desired. The U.S. Census Bureau also seeks feedback on the types of users that may use this data and how they use it. This feedback will aid in potential product design and creation.

Estimating Dynamic Daytime Population

Dynamic daytime population estimates provide users with a more precise picture of worker population movement throughout the day for counties with population of 250,000 or more. These data include information on the number of workers in motion between their home and workplace, the number of workers at their workplace, and those at home. This research utilized 2007-2011 5-year American Community Survey (ACS) estimates. The data includes a number of calculated variables derived from existing ACS variables including state and county of current residence, employment status, time of arrival, time of departure, minutes to work, workplace state and county, and usual hours worked per week. The reference period varies depending on the variable. Current residence is at the time the survey was taken, the workplace and commuting variables reference the week before the survey was taken (i.e., last week), and hours worked per week is during the past 12 months.

Assumptions

Several assumptions are associated with the dynamic daytime population estimates. Table 1 displays these assumptions.

Table 1: Assumptions for Creating Dynamic Daytime Population Estimates

No.	*Assumption*
1	The analysis will only be showing the movement of people who worked last week per the 2007-2011 ACS. The movement of other people (e.g., students, hospital patients, shoppers, tourists, workers on vacation last week) across boundaries will not be taken into account.
2	All workers work the same 5 days a week and equal time each day. (Hours worked in a week will be divided by 5.)
3	Non-work times (such as unpaid lunches) are not factored into the workday.
4	Workers work at the same location all 5 days.
5	Everyone, when not working, is in their place of residence.
6	If a workers work shift crosses midnight, he will be shown as being at work in the morning, then traveling back home, then at current residence, then traveling to work, and then working the rest of the day. Same if there commute crosses midnight.
7	No recognition of counties commuters may travel through are made.
8	It takes the same amount of time for a worker's morning commute as their evening commute.
9	The workers commute is the same each day.
10	Outlier commuters, such as commuting from a foreign country outside Canada or Mexico, are included. These could include cases where workers were on business travel or deployment in foreign countries "last week" but had to return home when the survey was completed.

First, the ACS only collects data on the work trip and does not include non-work travel for a given area (Assumption 1). Because the ACS does not collect information on the number of days worked per week, the methodology divided reported hours worked in a week by five to estimate daily work hours, regardless of full-time or part-time status (Assumption 2). Non-work times such as unpaid lunches, breaks, or split shifts were not factored into the workday (Assumption 3). The ACS collects information regarding the place of work where the person worked last week. This research assumed persons worked at this reported place of work all five days per week (Assumption 4) and when they were not at their place of work that workers were at their current residence (Assumption 5). Again, because the ACS does not collect information

on non-work travel, this research only estimates travel to and from the reported place of work and current residence and therefore cannot make any estimations on where a respondent may be outside of those two places. Although a respondent may travel through several counties from current residence to place of work, this methodology did not take into account route of travel and therefore did not include in county estimates the number of people traveling through a county that is not their county of current residence or work (Assumption 7).

It was also assumed that the worker's commute is the same every day; meaning, the respondent always used the means of transportation he/she reported, travelled the same number of minutes, and never deviated from the reported time of departure (Assumption 9). Additionally, the morning and evening commute were the same and did not deviate from the reported characteristics just described (Assumption 8). If a worker had a schedule that crosses midnight, it displayed as working from midnight until morning, then the worker traveled home and stayed at home until their time of departure later in the day, after which, they were displayed as in transit and then at work again (Assumption 6). Note, a commuter is assigned the status that he is in at the beginning of a 15-minute interval. For example, if someone has a 30 minute commute and departs at 6:30 to work, then he is assigned to be at home at 6:15, on the road at 6:30 and 6:45 and at work at 7:00. If he departs at 6:40, then he is assigned to be at home at 6:15 and 6:30, on the road at 6:45 and 7:00, and at work at 7:15. And lastly, this research included commutes from foreign countries, which may include respondents who were deployed or on business travel the week prior (Assumption 10).

Data and Variables

This research utilized data from the ACS. The ACS is a nationwide survey designed to provide communities with reliable and timely demographic, social, economic, and housing data for the nation, states, congressional districts, counties, places, and other localities every year. It had an annual initial sample size between 2007 and 2010 of about 2.9 million addresses across the United States and Puerto Rico and includes both housing units and group quarters (e.g., nursing facilities and prisons). The initial sample size was increased to about 3.3 million housing units in 2011 and about 3.5 million housing units in 2012. The ACS is conducted in every county throughout the nation and every municipio in Puerto Rico, where it is called the Puerto Rico Community Survey. Beginning in 2006, ACS data for 2005 were released for geographic areas with populations of 65,000 and greater. For information on the ACS sample design and other topics, visit <www.census.gov/acs/www>.

Given the assumptions noted above, analysts calculated the number of workers in a county who were at home, in transit to or from work, or at work in 15- minute increments throughout a 24-hour period using data from the 2007-2011 5-year ACS. Additional data for the visualization included: age delineations of those assumed to be at their residence because they are non-workers, such as non-workers 0-18 years, non-workers 19-64, non-workers 65 years and over; total population; commuter-adjusted daytime population; and, margins of error for each of these estimates. Table 1 displays the dynamic daytime population variable names and descriptions.

Table 1: Dynamic Daytime Population Variable Names and Description

Variable name	*Description*
state	Reference state FIPS code
county	Reference county FIPS code
state_name	Reference state name
county_name	Reference county name
miltime	Military time in 15 minute intervals (numeric with no leading zeros)
time	Standard time in 15 minute intervals
AtWork	The number of workers who worked last week in the reference county but lived in a different county
moe_AtWork	Margin of error of AtWork
CmtHomeOut	The number of workers commuting out of reference county to home in a different county
moe_CmtHomeOut	Margin of error of CmtHomeOut
CmtWrkOut	The number of works commuting out of reference county to work in a different county
moe_CmtWrkOut	Margin of error of CmtWrkOut
CmtWrkIn	The number of workers commuting in to work in reference county from a different county
moe_CmtWrkIn	Margin of error of CmtWrkIn
CmtHomeIn	The number of workers commuting to home in reference county from a different county
moe_CmtHomeIn	Margin of error of CmtHomeIn
AtHome	The number of workers who are at home in the reference county but work in a different county
moe_AtHome	Margin of error of AtHome
StayAtWrk	The number of people who work and live in the same county and are at work
moe_StayAtWrk	Margin of error of StayAtWrk
StayToWrk	The number of people who work and live in the same county and are commuting to work
moe_StayToWrk	Margin of error of StayToWrk
StayToHome	The number of people who work and live in the same county and are commuting back to home
moe_StayToHome	Margin of error of StayToHome
StayAtHome	The number of people who work and live in the same county and are at home
moe_StayAtHome	Margin of error of StayAtHome
WrkHome	The number of people who work at home
moe_WrkHome	Margin of error of WrkHome
NonWrk0_18	The number of people not working and are ages 0 to 18 years old
moe_NonWrk19_64	Margin of error of NonWrk0_18
NonWrk19_64	The number of people not working and are ages 19 to 64 years old
moe_NonWrk19_64	Margin of error of NonWrk19_64
NonWrk65over	The number of people not working and are age 65 years old and over
moe_NonWrk65over	Margin of error of NonWrk65over
totalpop	The total population for the county
moe_totalpop	Margin of error of totalpop
daytimepop	The commuter-adjusted daytime population for the county calculated using the standard approach
moe_daytimepop	Margin of error of daytimepop

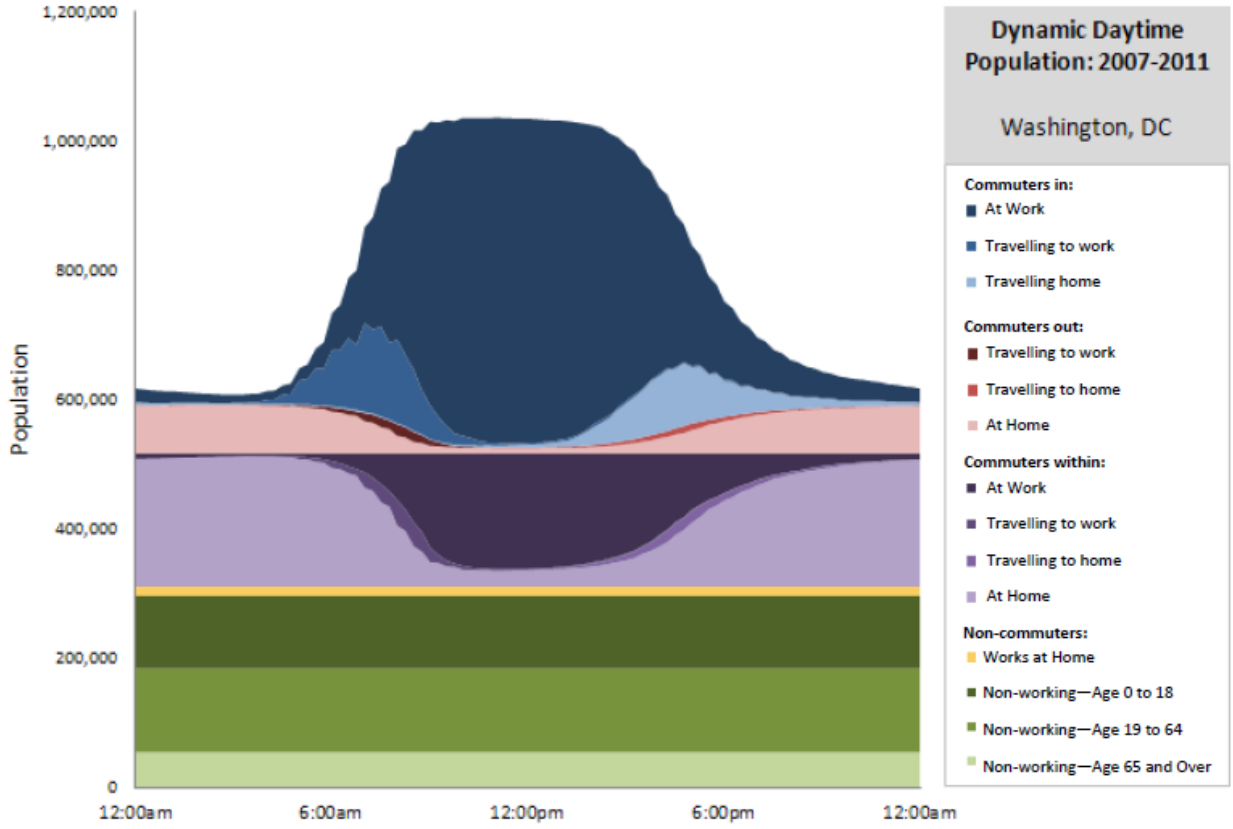
Universe: Total population. Workers are defined as employed persons 16 years and over who worked last week.

Visualizing Dynamic Daytime Population

Static Visualizations

Because these data have a geographic aspect, this research also examined different ways of disseminating and visualizing the data. Initial visualizations included several different types of graphs in Excel. Using Washington, DC as a study area, the Figure 1 graphs workers who either live, work, or do both in Washington, DC as well as non-workers who are DC residents. Time in the day is on the horizontal axis while population is on the vertical axis. Non-workers are delineated into three age categories: 0-18, 19-64, and 65 years and over. Workers are those who are 16 year and over who worked last week. Workers were delineated into several categories: those who are at home, those in transit to or from work, those at work, and those who work at home. Figure 1 shows that the "daytime" population of Washington, DC expands greatly with the number of workers that commute in from other counties between the hours of about 7am and 5pm. The non-working population as well as those who work from home appear static in the graph. This is due to Assumption #5 - everyone, when not working, is in their place of residence. Again, travel to other locations besides place of work are not reported in the ACS. Also of note, between the hours of 7am to 9am and 3pm to 6pm, workers in transit appear to increase and the peak is sharper in the morning hours.

Figure 1: Dynamic Daytime Population Estimates for Washington, DC, 2006-2010 5-year ACS



Another method of visualizing the data, allows the user to compare changes in the proportions of populations within a county. For example, Figure 3 uses shades of fuchsia to display population changes in Washington, DC by time of day. Additionally, the categories of workers and non-workers has been condensed down from 14 to 8. The graph shows a peak of workers at work between about 8am and 5pm, with peak travel times from about 7am to 9am and 4pm to 6pm, clearly. The proportion of non-workers does not change throughout the day due to the assumptions made above but the proportion is relevant to the other segments of population and therefore displays a comparison of workers to non-workers throughout the day.

Figure 2: Small Multiples Graph using Color Scales for Dynamic Daytime Population Estimates for Washington, DC, 2007-2011.

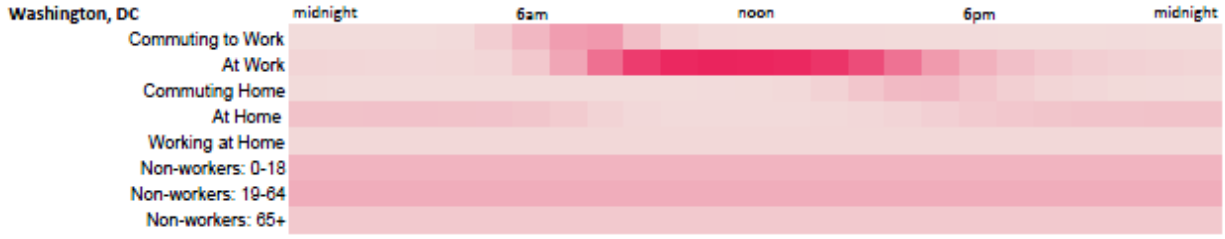
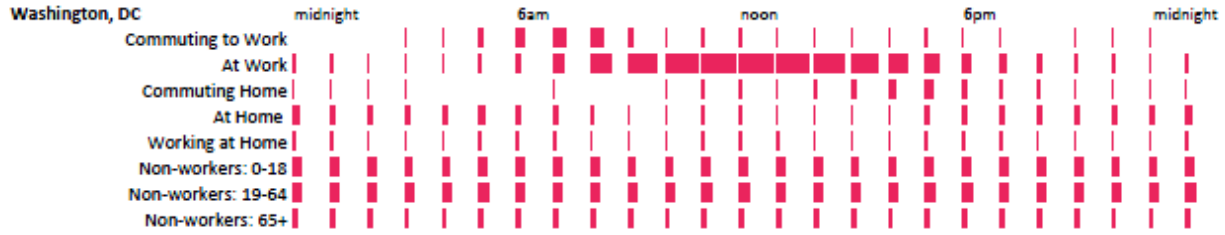


Figure 3 is similar to Figure 2 in that it is small multiples but instead of a color scale the graph uses small bars to compare the proportions for each worker/non-worker category by hour of the day. The same pattern is apparent in Figure 3 as it is in Figure 2. The choice of Figure 3 over Figure 2 in choosing the best type of visualization for comparing proportion may come down to personal preference, although for this data Figure 3 appears to be easier to understand.

Figure 3: Small Multiples Graph using Bars for Dynamic Daytime Population Estimates for Washington, DC, 2006-2010.

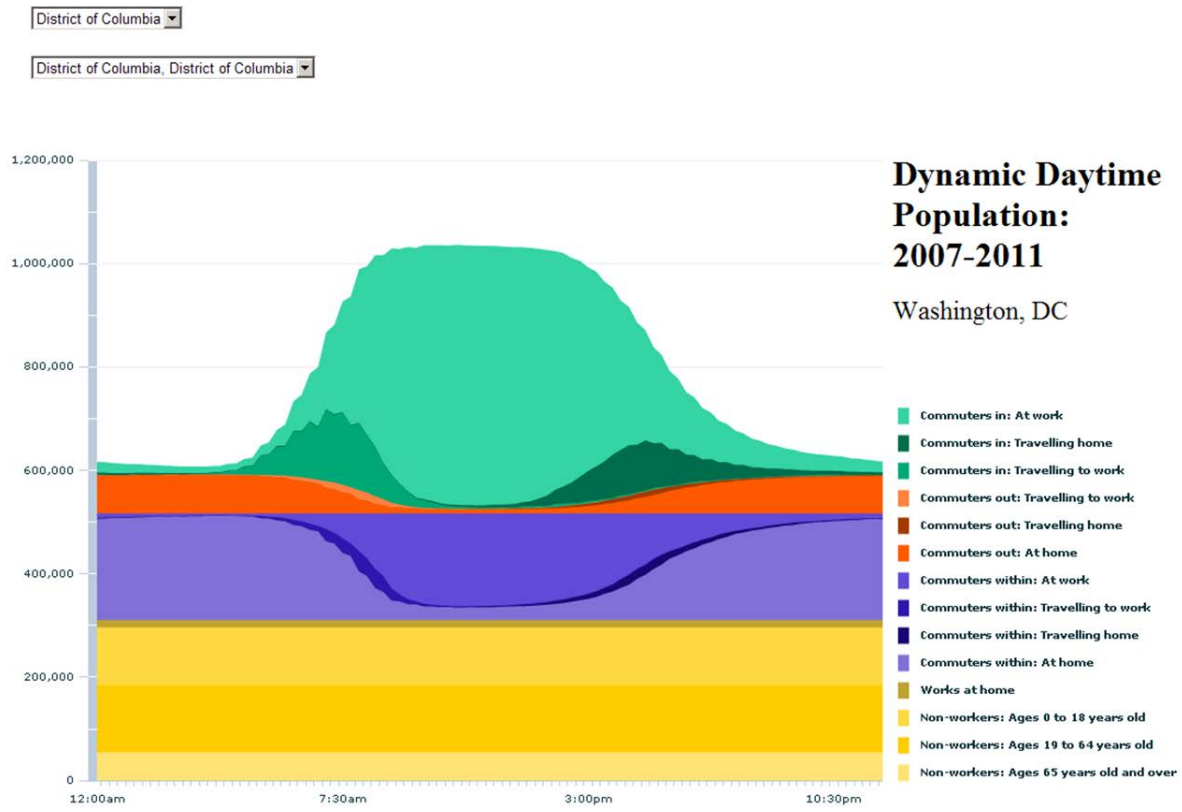


Interactive Visualizations

Each of the figures presented above displays a type of visualization for the dynamic daytime population data. The final result of this research is to make a user-friendly interface on the Census website that allows for easy downloading and visualization of the dynamic daytime population data for counties with over 250,000 persons.

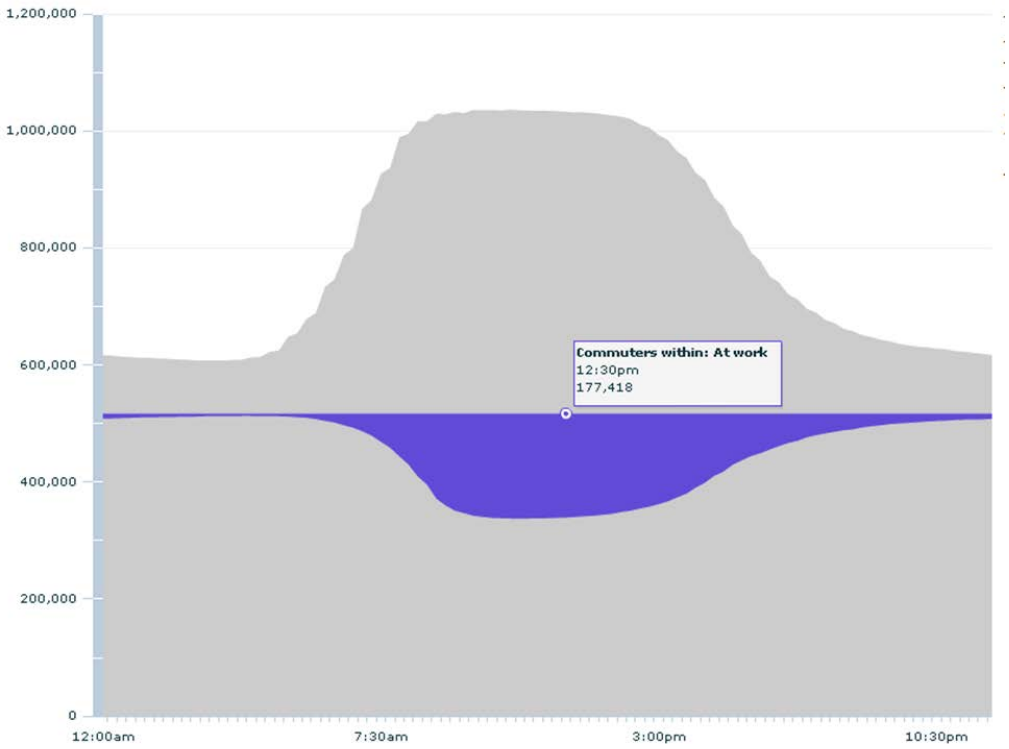
To do this, the dynamic daytime population data for Washington, DC were loaded into a secure version of DataFerrett and a hot report was created with drop down menus and a stacked area graph similar to Figure 1 (see Figure 4).

Figure 4: Screenshot of DataFerrett Hot Report of Stacked Area Graph for Dynamic Daytime Population for Washington, DC, 2007-2011 5-year ACS.



The DataFerrett Hot Report allows users to choose a county using the dropdown menus. After this, the user can scroll over either the legend or the graph to gray out or isolate a category of commuter. Additionally, as the user scrolls over the graph, each 15-minute increment displays a small popup box with commuter type, time of day, and total number of commuters in the selected category (see Figure 5).

Figure 5: Screenshot of DataFerrett Hot Report Functionality – graying out and popup box.



Feedback from the Southern Demographic Association (SDA) Conference

At the SDA conference, held October 23-25, 2013 in Montgomery, AL, a poster with the information from this paper was presented. A feedback survey was given to attendants at the session (see Appendix). Feedback from the poster presentation is key in understanding the type of data visualization that best conveys dynamic daytime population information to users as well as the most useful supplemental information and functionality.

Future plans for the Dynamic Daytime Population DataFerrett hot report may include completing a publicly available, interactive webpage to disseminate the data. Counties of a certain population threshold may be loaded into the data visualization and supplemental information such as mode of transportation may also be integrated. More research into design, variables and functionality may be necessary before a decision is made to create an application.

More information on the Census Bureau's daytime population estimates are available at <http://www.census.gov/hhes/commuting/data/daytimepop.html>.

Appendix

Dynamic Daytime Population Feedback Survey
Southern Demographic Association Conference
Montgomery, AL - October 23-25, 2013

Part A: The Poster

1) On the poster, which data visualization did you find best for conveying information and why?

2) What did you *like* and *dislike* about the DataFerrett visualization? Colors, size, type of information, legend, categories of data?

LIKES	DISLIKES

3) How would you *improve* the DataFerrett visualization?

4) What other features would you like to see on the DataFerrett visualization?

For more information on Dynamic Daytime Population, contact the Journey to Work and Migration Statistics Branch of the US Census Bureau at 301-763-2454 or visit the Daytime Population subject page at http://www.census.gov/hhes/commuting/data/daytimepop.html.

United States Census Bureau

Part B: Your Work

Appendix

5) What is your occupation? What state(s) or region does most of your work involve?

6) Do you use daytime population estimates in your work?
   If yes, how?
   If no, what do you use instead?

7) Do you plan on using Dynamic Daytime Population estimates in your work?
   If yes, how?
   If no, why not?

8) Currently, we are using county-county flows in these products. What other geographic summary levels would you find useful for dynamic daytime population estimates? (Mark all that apply)
   ___ County/County equivalent
   ___ County subdivision/Township
   ___ City/Town
   ___ TAZ
   ___ Census Tract
   ___ Zip Code Area

9) Besides workers' commuting patterns, what other groups' movements would be beneficial for you to use in a dynamic daytime population model?

Thank you for taking the time to fill out this survey. This information will be used to improve the products and services for the public.

How can we best visualize worker movement throughout the day?

@USCensusBureau #commuting #ACS #dataviz #SEHSD #SDA

Melanie A. Rapino, Ph.D., William K. Koerber, Brian McKenzie, Ph.D. | Journey to Work and Migration Statistics Branch | Social, Economic, and Housing Statistics Division | U.S. Census Bureau | U.S. Department of Commerce

START

Read This

What is Dynamic Daytime Population?

The number of people in a county that are either commuting to or from work, are at work, or are at home in 15-minute increments.

Why are these data important?

Dynamic daytime population estimates provide users such as emergency management agencies, emergency responders, planners, developers, transportation planners, and policy makers information about the estimated number of people that are in a county within any given 15-minute interval.

See the Data

state	county	state_name	county_name	mdtime	time	AtWork	moe_AtWork	CntHomeOut	moe_CntHomeOut	CntWkOut	moe_CntWkOut	C
11	001	District of Columbia	District of Columbia	0	12:00am	20,959	935	4,419	519	136	87	
11	001	District of Columbia	District of Columbia	15	12:15am	20,310	924	4,177	496	43	49	
11	001	District of Columbia	District of Columbia	30	12:30am	19,531	905	3,932	374	11	17	
11	001	District of Columbia	District of Columbia	45	12:45am	18,435	897	3,653	456	11	17	
11	001	District of Columbia	District of Columbia	100	1:00am	17,557	867	2,977	437	51	45	
11	001	District of Columbia	District of Columbia	115	1:15am	16,727	878	3,454	464	97	62	
11	001	District of Columbia	District of Columbia	130	1:30am	15,920	864	2,971	485	118	67	
11	001	District of Columbia	District of Columbia	145	1:45am	15,200	840	3,047	477	81	57	
11	001	District of Columbia	District of Columbia	200	2:00am	14,490	821	2,355	346	235	104	
11	001	District of Columbia	District of Columbia	215	2:15am	14,090	830	2,541	345	180	96	
11	001	District of Columbia	District of Columbia	230	2:30am	13,825	820	2,113	349	53	49	
11	001	District of Columbia	District of Columbia	245	2:45am	13,046	819	2,162	312	37	36	
11	001	District of Columbia	District of Columbia	300	3:00am	12,675	815	1,765	249	247	126	
11	001	District of Columbia	District of Columbia	315	3:15am	12,638	849	1,626	261	204	110	
11	001	District of Columbia	District of Columbia	330	3:30am	12,652	808	1,325	210	393	142	
11	001	District of Columbia	District of Columbia	345	3:45am	12,647	841	1,370	254	360	138	
11	001	District of Columbia	District of Columbia	400	4:00am	12,896	817	1,059	241	863	228	

Understand the Data

- Estimates are for counties with over 250,000 persons.
- Broken up into 15-minute intervals.
- All time is in military time.
- Include workers in transit, at home, at work, and non-workers.
- Use time of departure, travel time, and hours worked in week to estimate whether in transit or at work or at home.
- Contain margins of error for each estimate

What we need you to do:

Review the following elements of this poster. Then...we need feedback on:

1) The assumptions and the visualizations,
2) Anything else you'd like to see with these data. For example, mode of transportation?

Review the Assumptions

1

Only shows workers who worked last week. No students, tourists, shoppers, or other non-work related travel.

2

All workers work the same 5 days a week and equal time each day.

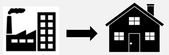
3

Non-work times (such as unpaid lunches) are not factored into the work day.

4

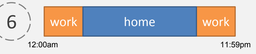
Workers work at the same location all 5 days.

5



Everyone, when not working or commuting, is in their place of residence.

6



If a workers work shift or commute crosses midnight, it will appear as split between two parts of the day – the beginning and end.

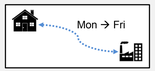
7

Counties that are passed through during commutes are not reflected in the data.

8

It takes the same amount of time for a worker's morning commute as their evening commute.

9

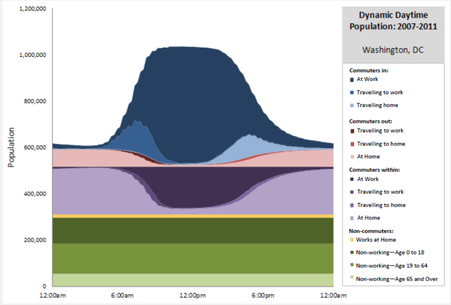


The worker's commute is the same each day.

10

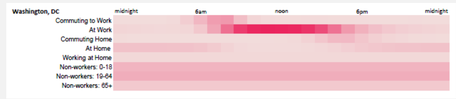
Outlier commuters, such as commuting from a foreign country outside Canada or Mexico, are included.

Visualize Size



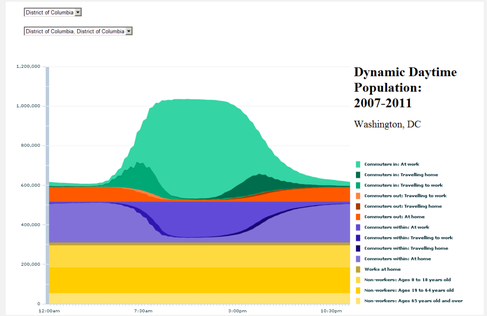
This diagram shows the number of commuters, workers at home, and non-workers within the District of Columbia throughout the day. This diagram allows the user to compare size of subpopulations throughout the 24-hour day.

Visualize Proportion



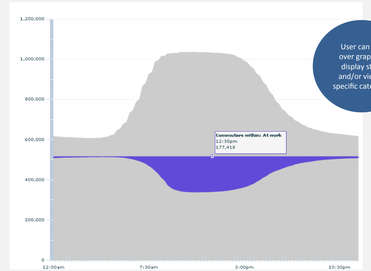
These diagrams show an aggregated version of the categories. It also allows the user to quickly see the differences in proportion among the categories.

Visualize with DataFerrett



These graphs were created using the hot report writer in the Census Bureau's DataFerrett data dissemination tool. The top graph shows all of the categories of dynamic daytime population estimates.

The bottom graph depicts the user's ability to roll over the graph to only show a certain category and a pop-up box displaying the time and population estimate for that category. Additionally, the user can roll over the legend to display only a specified category and gray out the rest.



User can roll over graph to display stats and/or view a specific category.

Answer the Questions on the Feedback Survey. Return completed survey here.

STOP

United States Census Bureau

Presented at the Southern Demographic Association 2013 Annual Meeting
Montgomery, AL, October 23-25, 2013.