

CURRICULUM

Complete PUMS Training (Parts I-V)

May 16, 2017 ♦ Kansas Health Institute

LEARNING OBJECTIVES

1. Learn how PUMS fits into the overall U.S. Census strategy.
2. Identify data available in PUMS.
3. Understand how to calculate estimates using PUMS data.

WHAT IS PUMS?

- Public Use Microdata Sample (PUMS)
- A set of un-tabulated records about individual people or housing units from the American Community Survey (ACS)
- Part of the larger U.S. Census: sample of a sample

CENSUS OF POPULATION AND HOUSING

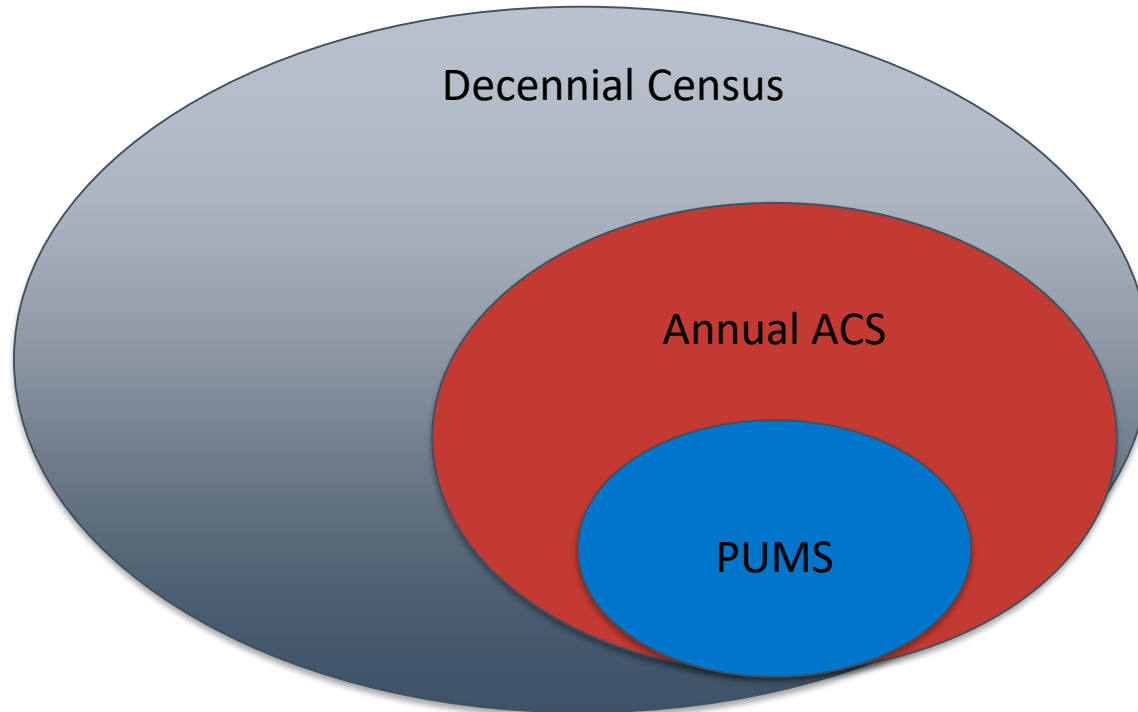
American Community Survey (PUMS)

Current Population Survey

American Housing Survey

Public Use
vs.
Restricted
Use

CENSUS, ACS AND PUMS



PURPOSE OF DATA

	Decennial Census	Annual ACS	PUMS
Purpose	<ul style="list-style-type: none">• Measure total population demographics• Track national changes• Support national-level decision-making• Determine number of congressional seats	<ul style="list-style-type: none">• Measure population characteristics not in Census• Track community changes• Support community-level decision-making• Updates census-level data	<ul style="list-style-type: none">• Subset of ACS data• Provides public use data for researchers and students

DATA METHODOLOGY

	Decennial Census	Annual ACS	PUMS
Collection	Every 10 years	Annual	Same as ACS
Process	Master Address File	Master Address File + Updates	Same as ACS
Sampling Unit	Housing Unit	Housing Unit	Same as ACS
Target Sample	100%	2.5%	1%
Response Rate	67.4%	98%	Same as ACS
Data	Aggregate	Aggregate	Microdata

PUMS: AGGREGATE VS. MICRODATA

Aggregate	Microdata
<ul style="list-style-type: none">• Summary• Full/large sample• Fixed content• Uncensored (All Data)• Public focused	<ul style="list-style-type: none">• Individual Responses• Smaller sample• Customizable• Protected (Censored & Top/bottom-coded)• Researcher focused

PUMS VARIABLES

Housing (90)

- Household status
- Household members
- House features (rooms/appliances/fixtures)
- Ownership status
- Mortgage/rent/taxes
- Technology

Population (125)

- Citizenship
- Employment
- Income (all types)
- Abilities/disabilities
- Transportation
- Insurance
- Education

PUMS VARIABLES

Geography (3)

- Division/region
- State
- Public use Microdata area (PUMA)

Weights

- Person weights
- Household weights

PUMS FILES

	1 Year	3 Years*	5 Years
Population Limits	65,000+	20,000+	All Areas
Reliability	Lowest	Medium	Highest
Recentness	Most Current	Current	Least Current
Release Dates	2005–Present	2007–2013	2009–Present
Appropriateness	<ul style="list-style-type: none"> – Analyzing large populations – Timeliness is important 	Analyzing smaller populations or geographies	<ul style="list-style-type: none"> – Precision is more important than current data – Small populations

* Eliminated in 2015

CALCULATING ESTIMATES

- Estimation of Characteristics

- Add weights of person or housing units that have character of interest

- Ex. Total number of female teachers in Kansas

Add weights of all female persons in Kansas who are teachers.

$$\sum_{i=1}^n FTw$$

CALCULATING ESTIMATES

- Estimates of Proportion
 - Divide weighted estimate by weighted base
 - Divide the weighted estimate of female Kansas teachers by the weighted estimate of all Kansas teachers

$$\frac{\sum_{i=1}^n FTW}{\sum_{i=1}^n TW}$$



Research Questions

LEARNING OBJECTIVES

1. Learn how to develop good research questions.
2. Understand how to build a research plan/proposal.
3. Practice using PUMS data to create and answer research questions.

DEVELOPING RESEARCH QUESTIONS

Most difficult part of conducting research

- Characteristics of good research questions
- Common pitfalls when developing questions
- Designing research projects

RESEARCH QUESTIONS

AIMS

- Based on theory and/or a strategy
 - Ex. Grounded theory or strategic initiative
- Measureable
- Testable
- Achievable
- Ethical

COMMON PITFALLS

#1. Sequence of Events



COMMON PITFALLS

- Starting with a data set and then developing a question
 - Limiting analysis to available variables
- Example: Measuring availability of mental health services
 - Could proxy with insurance as availability (PUMS), better to use supporting dataset with mental health resource data (SAMHSA)
 - Reduced-form models: usually less accurate unless using causal model to control for unmeasured variance

COMMON PITFALLS

#2. Too Broad of a question



COMMON PITFALLS

- Unfocused question
 - Broad scope of literature to review
 - Many outcomes and covariates to select
 - Project duration

COMMON PITFALLS

- Example: How does income impact health?
 - Conceptual model problems
 - Which income and health measurements should be used?
 - Do selected measurements comprehensively or partially reflect health and income?
 - What unobserved variance do we need to account for?

COMMON PITFALLS

#3. Too narrow of a question



COMMON PITFALLS

- Sample size problems
 - Protected, top or bottom coded out
 - Need to group in larger categories
 - Coding race to black/white/other due to small numbers of other races in a sample

COMMON PITFALLS

- Example: Are American Indian (AI) grandparents as caregiver households more likely to not have refrigerators than non-grandparents as caregiver households?
 - Sample size: Weighting could eliminate AI in some population samples
 - Measurement: Refrigerators common, bundle in other physical housing characteristics for analysis

QUESTION DESIGN

- Descriptive
 - Exploratory, grounded theory
- Hypothesis Testing
 - Conceptual model and testing plan
- Causal
 - Panel data, RCT, IV, or Diff-in-diff model

DESCRIPTIVE

- Profile of a group
 - Group characteristics and demographics
- Groundwork for further research
 - We know how many Kansas teachers are female:
More Kansas teachers are female than male.
 - Hypothesis testing, H_0 vs. H_1
- Not a causal framework

DESCRIPTIVE

- Example: Economic profile of two-income households in Douglas County.
 - Report on demographics, percentages, rates could lead to additional follow up research.
 - Hypothesis test: Two-income households are more likely to own their house than one-income households. H_0 vs. H_1

HYPOTHESIS TESTING

- Question design
 - Testable hypothesis: People over 65 are more likely to be disabled than people under 65. H_0 vs. H_1
- Conceptual model
 - Having a disability is more likely in people over 65
 - Other factors (race, education and occupation)
- Testing mechanism/variable distribution
 - Means test: Chi-square test (two categorical variables)
 - Regression: Logistic, categorical dependent variable

CAUSAL

- Statement of hypothesis
- Specify the statistical model
- Collection of data
- Estimation of parameters
- Hypothesis testing
- Forecasting or prediction

CAUSAL

- Example: Has health insurance coverage increased with the passing of the Affordable Care Act?
 - Difference-in-differences approach
 - Panel data (pre- and post-period)
 - Control for other factors/covariates

DESIGN DIFFERENCES

	Descriptive	Hypothesis Testing	Causal
Theory-based	Yes	Yes	Yes
Literature review	Yes	Yes	Yes
Means testing	No	Yes	Yes
Statistical model	No	Yes	Yes
Causal inference	No	No	Yes

RESEARCH EXAMPLE

- Siordia, C. (2015). Demographic, Economic, Household, and Health Profile of Grandparents Responsible for Grandchildren. *Journal of child and family studies*, 24(9), 2661-2667.

RESEARCH EXAMPLE

- Descriptive research
- PUMS data 2009 – 2011 (3-year sample)
- Purpose: To investigate characteristics of grandparents responsible for grandchildren by race-ethnic groups and the geographical distribution over the U.S. mainland.

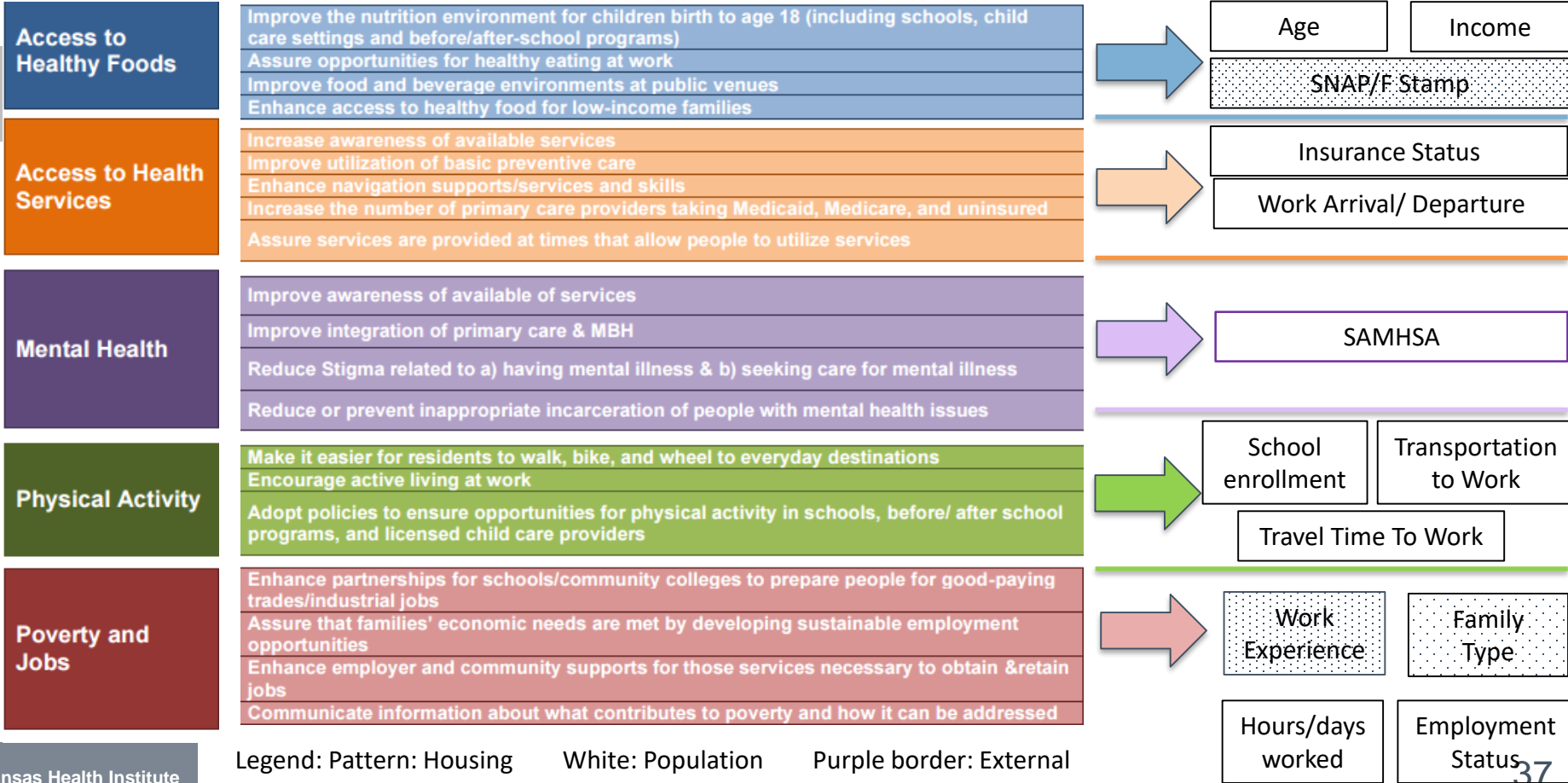
RESEARCH EXAMPLE

- Methods
 - Continental (mainland) United States
 - Race groups: Non-Latino (NL) White, NL Black, NL Other, Mexican-Latino, Non-Mexican Latino
 - Identified grandparents as responsible for grandchildren by survey question response
 - Also reported grandparent only (no other parent)
 - Home ownership: mortgage, own, rent or rent-free
 - Length of ownership: When moved in (categorical)

PUMS RESEARCH QUESTION REVIEW

- Outcomes: How do you want to report/use data?
 - Person-level questions
 - Are teachers in Kansas predominately male or female?
 - Household question
 - Economic profile of two-income households in Douglas County
- Household Programs: Reliable/stable housing
- Personal Programs: Addressing education rates

DOUGLAS COUNTY COMMUNITY HEALTH PLAN



Legend: Pattern: Housing White: Population Purple border: External

PUMS RESEARCH

EXAMPLE 1

- Strategic Initiative: Assure services are provided at times that allow people to use them
 - Descriptive design
 - What factors do we need to consider to allow people to use services?
 - PUMS variables that can be used: work arrival, work departure, transportation to work, travel time to work, occupation (proxy for hourly/salary or if paid leave available)
 - Measure availability of services. Other data sources? Provider/facility registry (services and hours)

PUMS RESEARCH

EXAMPLE 2

- Strategic Initiative: Assure that families economic needs are met by developing sustainable employment opportunities.
 - Could be a causal model if we are measuring an intervention of program (IV or diff-in-diff model)
 - PUMS variables that can be used: Hours/days worked, employment status, employment tenure, housing characteristics (one- or two-income families, residence tenure-stabilization)
 - Outcome choices: Household or individual level? How do you want to report outcome?



Deep Dive: PUMS Data

LEARNING OBJECTIVES

1. Learn how to access PUMS.
2. Identify considerations for using PUMS.
3. Understand which tools could be used with PUMS.

FINDING PUMS

- Four steps from Census.gov website:
 1. Surveys/Programs (dropdown)
 2. American Community Survey (ACS)
 3. Data Tools and Tables (bottom of page)
 4. PUMS (menu on left side)

We Can Help

Training Opport
use Census da
informing your l
proposals and
schools and ho



TRAINING OPPORTUNITIES

POPULATION CLOCK



U.S. Population

324,989,666

World Population

7,388,823,906

May 05, 2017 17:34 UTC (Eastern+5)

[Learn More >>](#)

QUICKFACTS



Did You Know

Population per square mile in San Diego County, California is
735.8

Source: 2010 Census

Select a state to begin



U.S. Census Bureau

Manufacturer
March 2017 Rep
Released 10:00

International
March 2017 Rep
Released 8:30 A

Construction



FINDING PUMS

- Four steps from Census.gov website:
 1. Surveys/Programs (dropdown)
 2. American Community Survey (ACS)
 3. Data Tools and Tables (bottom of page)
 4. PUMS (menu on left side)

Surveys/Programs Main

Are you in a Survey?

2017 Census Test

2017 Census Test

2018 End-to-End Census Test

2010 Census

American Community Survey (ACS)

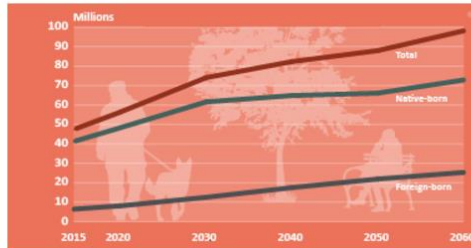
American Housing Survey (AHS)

Economic Census

Special Census Program

Survey and Program Participation

All Surveys & Programs



Older Americans Mo

This edition of Facts for Fe celebrates the age 65 and
Graphic: An Aging Nation - U.S. residents 65 and older.

POPULATION CLOCK



QUICKFACTS



U.S. Census Bureau Economic

- Manufacturers' Goods
March 2017 Report
Released 10:00 AM EDT, 5/4/17
 - International Trade: Goods
March 2017 Report
Released 8:30 AM EDT, 5/4/17
 - Construction Spending
- [View All](#)

FINDING PUMS

- Four steps from Census.gov website:
 1. Surveys/Programs (dropdown)
 2. American Community Survey (ACS)
 3. Data Tools and Tables (bottom of page)
 4. PUMS (menu on left side)

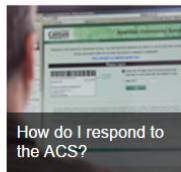
Census.gov > Our Surveys & Programs > American Community Survey (ACS)

American Community Survey (ACS)

- [About the Survey](#)
- [Respond to the Survey](#)
- [News & Updates](#)
- [Data](#)
- [Guidance for Data Users](#)
- [Geography & ACS](#)
- [Technical Documentation](#)
- [Methodology](#)
- [Library](#)
- [Operations and Administration](#)
- [Contact Us](#)



The American Community Survey helps local officials, community leaders and businesses understand the changes taking place in their communities. It is the premier source for detailed information about the American people and workforce.



Latest

- Data
- News
- Events
- Library

2016 Data Release

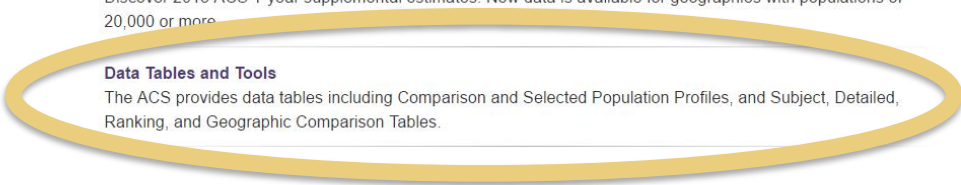
Learn more about upcoming ACS data releases, including the September 14th release of the 2016 ACS 1-year estimates, as well as a preview of Data.census.gov.

2015 ACS 1-Year Supplemental Estimates

Discover 2015 ACS 1-year supplemental estimates. New data is available for geographies with populations of 20,000 or more.

Data Tables and Tools

The ACS provides data tables including Comparison and Selected Population Profiles, and Subject, Detailed, Ranking, and Geographic Comparison Tables.



FINDING PUMS

- Four Steps from Census.gov website:
 1. Surveys/Programs (dropdown)
 2. American Community Survey (ACS)
 3. Data Tools and Tables (bottom of page)
 4. PUMS (menu on left side)

American Community Survey

About the Survey

Respond to the Survey

News & Updates

Data

Data Tables & Tools

Data Profiles

Narrative Profiles

Subject Tables

Ranking Tables

Supplemental Tables

American FactFinder

My Congressional District

Data via FTP

Summary Statistics

PUMS Data

Variance Reports

Race/Ethnicity & AIAN Data

Custom Tables

Guidance for Data Users

Data Tables & Tools

Tell Us W

Data Profiles Selector

Data Profiles consist of four tables (Social, Economic, Housing, Demographic) to give a broad statistical view of a particular geography. We provide an easy to use [Data Profiles geography selector](#) for the most popular geographies: state, county and place. Our full data website, [American FactFinder](#), provides additional geographies for this table type and many more.

Narrative Profiles

The only place to find the current Narrative Profiles is right here on American Community Survey website. Narrative Profiles contain much of the same information found in the Data Profiles, except it is a text-based report with plenty of colorful graphs and charts. Use the map selector or the dropdown boxes to [generate a Narrative Profile](#) for your favorite place.

Subject Tables

More interested in a particular topic than a particular geography? Check out our listing of ACS Subject Tables. Subject tables have both numbers and percentages making them the versatile choice for data seekers. You can search/filter the listing, and then link to American Factfinder (AFF) where you can change geographies and go back in time! [Choose your subject table here!](#)

Supplemental Tables

Looking for statistics about people and households located in geographies with

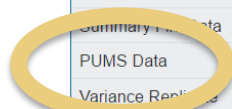
Data by Topic

People

- Age and Sex / Age
- Ancestry
- Disability
- Commuting to Wo
- Education
- Employment
- Family/Relationsh
- Health Insurance
- Income and Earni
- Language
- Origins
- Population Chang
- Poverty
- Race and Ethnicity
- Veterans

Housing

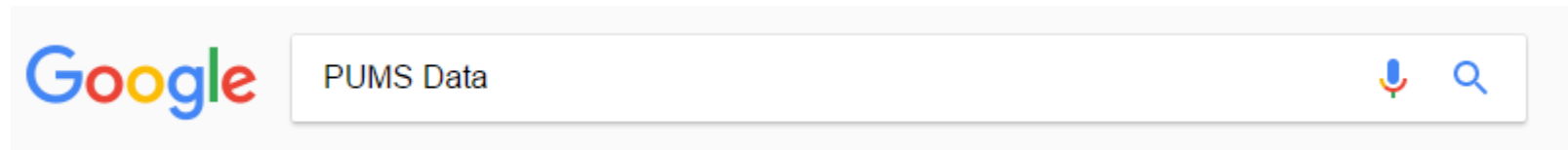
- Financial Charact



FINDING PUMS

- Four Steps from Census.gov website:
 1. Surveys/Programs (dropdown)
 2. American Community Survey (ACS)
 3. Data Tools and Tables (bottom of page)
 4. PUMS (menu on left side)

FINDING PUMS



NOW WHAT?

- Before you download
 - Considerations:
 - One-year? Multiple years?
 - Do you need Person Level and/or Housing Level?

MULTIPLE YEARS

- When do you need multiple years?
 - Single year data estimates
 - Most timely data; but
 - Sample may be too small to answer your question for area of focus
 - Larger combined file may allow for a larger sample and reliable estimates
 - Trade off: timeliness for reliability

IS THE ESTIMATE RELIABLE?

- How do you determine if the estimate is reliable?
- Coefficient of Variation (COV)
 - Standard Error/Estimate
 - Multiplied by 100 (sometimes)
 - Do not use if greater than 30%

STATISTICAL TOOLS

	SAS	STATA	SPSS	R
Strengths	<ul style="list-style-type: none"> • Primary tool of Census • Can weight estimates • Calculate standard errors • Better recoding/variable control 	<ul style="list-style-type: none"> • Can weight estimates • Calculate standard errors • Better recoding/variable control 	<ul style="list-style-type: none"> • Easy to use (GUI) • Can weight estimates • Better recoding/variable control 	<ul style="list-style-type: none"> • Free • Versatile • Can weight estimates • Calculate standard errors • Better recoding/variable control
Limitations	<ul style="list-style-type: none"> • \$\$\$ • Some learning curve 	<ul style="list-style-type: none"> • \$ • Some learning curve • Works in memory 	<ul style="list-style-type: none"> • \$\$ • Cannot calculate standard errors 	<ul style="list-style-type: none"> • Steep learning curve • Works in memory

OTHER TOOLS

	Excel	DataFerrett	IPUMS
Strengths	<ul style="list-style-type: none">• Commonly available• Familiarity• Easy to learn• Can weight estimates	<ul style="list-style-type: none">• No software required• Can recode variables and combine years• Can merge datasets• Weights estimates• Select what you want	<ul style="list-style-type: none">• No software required• Can recode variables and combine years• For more standard users• Weights estimates• Select what you want
Limitations	<ul style="list-style-type: none">• Limited analytical capability• Cannot calculate standard errors	<ul style="list-style-type: none">• A lot of features and can be difficult to use• Need to know what you are looking for• Cannot calculate standard errors	<ul style="list-style-type: none">• Cannot calculate standard errors• Cannot save work

SKILLS

- Basic
 - Excel
 - Pivot tables
- Advanced
 - Research design (method to answer a research question)
 - Knowledge of statistics (what test to run)
 - Data manipulation skills



DATAFERRETT

WHAT IS DATAFERRETT?

DataFerrett

DataFerrett only works with Microsoft Internet Explorer and Mozilla Firefox. It will not work with Microsoft Edge or Google Chrome.

DataFerrett is a data analysis and extraction tool to customize federal, state, and local data to suit your requirements. Using DataFerrett, you can develop an unlimited array of customized spreadsheets that are as versatile and complex as your usage demands then turn those spreadsheets into graphs and maps without any additional software.

What you should check before getting started:

- ✓ Java Installed: [Check your version](#) or [Download the latest version](#)
- ✗ Allow Pop-ups
The DataFerrett Applet **requires** you to have popup windows enabled in your browser for this website to function properly.
[Allow popups](#) for [thedataweb.rm.census.gov](#) and [reload this page](#).
- ✓ Run in IE/Firefox



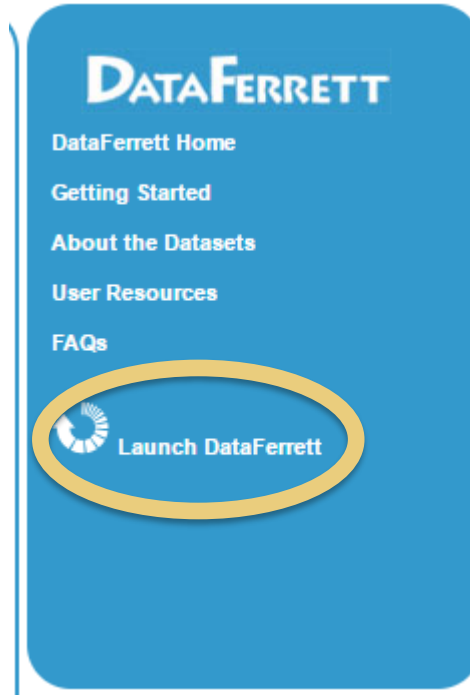
WHAT CAN YOU DO WITH DATAFERRET?

- Access multiple datasets
- Combine multiple years
- Recode variables
- Create data tables and conduct analyses
- Save your work

ACCESSING DATAFERRETT

<https://dataferrett.census.gov>

**You have to load it
in Internet
Explorer or
Firefox, or it will
not work!**

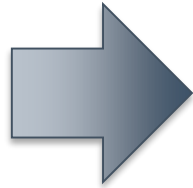


DON'T CLOSE THIS WINDOW!

The screenshot shows a web browser window titled "TheDataWeb - DataFerrett - Application Launch Page (U.S. Census Bureau) - Internet Explorer". The address bar shows the URL "https://dataferrett.census.gov/LaunchDFA.html". The page features the United States Census Bureau logo and a navigation menu with categories: Topics (Population, Economy), Geography (Maps, Data, Resources), Library (Infographics, Publications), Data (Tools, Developers), and About the (Research, Surveys). The main heading is "DataFerrett - Application Launch Page". A prominent yellow warning box with a black border and a dashed outline contains the text: "CAUTION Do Not: Navigate Away or Close This Window. Doing so will cause DataFerrett to QUIT." To the right of the warning, the text reads "DataFerrett - A unique data analysis and extraction tool. [New Features](#)". Below this is a green progress bar with the text "DataFerret is loading...". At the bottom of the main content area, it says "If a log-in screen does not appear, click on [Help](#) for more information". The footer contains several columns of links: ABOUT US (Are You in a Survey?, FAQs), FIND DATA (QuickFacts, American FactFinder), BUSINESS & INDUSTRY (Help With Your Forms, Economic Indicators), PEOPLE & HOUSEHOLDS (2010 Census, 2000 Census), and SPECIAL TOPICS (Advisors, Centers & Research Programs).

SIGN IN

Leave this checked



Ferrett Login

DataFerrett

Email address:

public use data only

The email address is used to send large extracts via email, and to inform users of new datasets available if desired. It is NOT used for any other purpose or shared with any organization.

Ok Cancel



Introduction

Step1 : Select Dataset & Variable

Step2 : DataBasket/Download/Ma A Table

data: (da • ta) *n.* A collection of facts from which conclusions may be drawn



DoroBeau

DataFerrett

Browser to TheDataWeb

Tutorials

Brand new to using

Examples

Sample Analysis
and instruction ...

Users' Guide

Users' Guide
DataFerrett functionality

Kinds of Datasets

Overview different Data Set types
and how they behave ...

Datasets Available

Datasets and topics
that are available ...

ferret: (fer' • it) *v.* To uncover and to bring to light by searching; to search intensively

About TheDataWeb

A collaborative network
of Internet data bases ...

Download Server

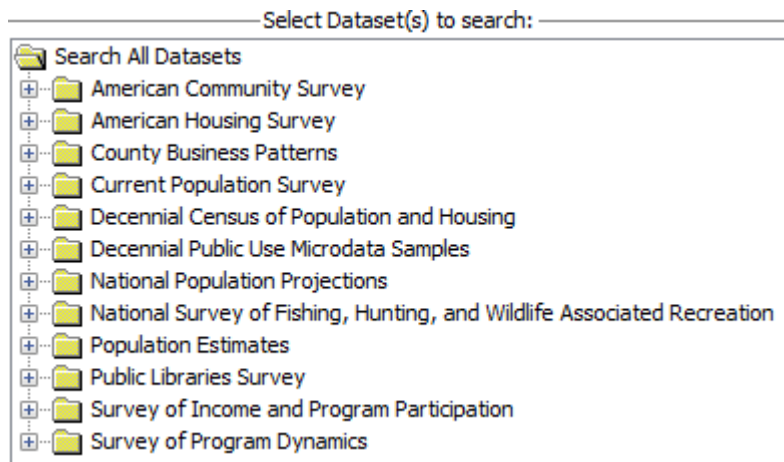
Adding/Publishing your
data to TheDataWeb ...

Discussion Group

Information sharing
with other users ...

Get Data Now

DATASETS IN DATAFERRETT

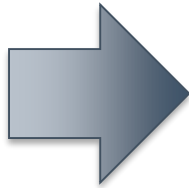


THINGS TO REMEMBER

- This is for power users
- They have a toll-free support line
- Will time-out from inactivity
- Can be slow
- Can freeze up

DATAFERRETT: EXAMPLE

You can
combine
years.



Select Dataset(s) to search:


- Search All Datasets
 - American Community Survey
 - 3-Year Estimates - Public Use Microdata Sample
 - 3-Year Estimates - Puerto Rico PUMS
 - 5-Year Estimates - Public Use Microdata Sample
 - 5-Year Estimates - Puerto Rico PUMS
 - Public Use Microdata Sample
 - 2015
 - 2014
 - 2013
 - 2012
 - 2011
 - 2010
 - 2009
 - 2008
 - 2007
 - 2006
 - 2005
 - 2004

Description
View Variables

DATAFERRETT: EXAMPLE

Select All Topics

- Housing
- Selectable Geographies
- Population
- Geographic Entities
- Replicate Weights

 Search Variables

DATAFERRETT: EXAMPLE

1 Variables returned from search.

Topic	Name	Availability	Variable Label
Selectable Geographies	Geography	2006 - current	Geographic Items

DATAFERRETT: EXAMPLE

The screenshot shows a software window titled "Ferrett Geography Codebook". At the top left is an information icon. The main title is "Browse/Select Geographies". To the right of the title is an illustration of a dog holding a pencil and a notepad, with the heading "Instructions:" followed by the text: "Select the type geography from left list. Once selected, either double click or select the hierarchy and click the 'Use Hierarchy' button to navigate that hierarchy." Below the instructions are three panels: "Types of geography available:" containing a list of "Division", "Public Use Microdata Area", "Region", and "State"; "Hierarchies:" which is currently empty; and "Selected Geographic Areas:" which is also empty. At the bottom of the window are six buttons: "Search", "Use Hierarchy", "Cancel", "Delete Selection", "Clear All", and "Finish".

DATA FERRETT: EXAMPLE

The screenshot shows the 'Ferrett Geography Codebook' application window. The title bar includes a close button (X) in the top right corner. The main interface is divided into several sections:

- Information Icon:** A blue 'i' icon in a square box on the left.
- Section Header:** 'Browse/Select Geographies' in bold black text.
- Instructions:** A cartoon character holding a stack of papers is positioned to the left of the text: 'Instructions: Select the type geography from left list. Once selected, either double click or select the hierarchy and click the "Use Hierarchy" button to navigate that hierarchy.'
- Types of geography available:** A list box containing 'Division', 'Public Use Microdata Area' (highlighted in blue), 'Region', and 'State'.
- Hierarchies:** A text area with the heading 'Hierarchies:' and the content 'State of current residence > Public use microdata area code (PUMA) based on 2010 Census'. It has a scrollbar at the bottom.
- Selected Geographic Areas:** An empty rectangular box on the right side.
- Buttons:** At the bottom, there are three buttons: 'Search', 'Use Hierarchy', and 'Cancel'. On the far right, there are three more buttons: 'Delete Selection', 'Clear All', and 'Finish'.

DATAFERRETT: EXAMPLE

The screenshot shows a web application window titled "Ferrett Geography Codebook". The main heading is "Select State of current residence". To the right of the heading is an illustration of a dog and a trash can, and a section titled "Instructions:" which reads: "Drag geographies to drop into to the right list. Drag final selections to the 'Selected Geographic Areas' list. Clicking a link at the top will drop directly into that level of the hierarchy." Below the instructions are navigation links: "Home >> State of current residence >> Public use microdata area code (PUMA) based on 2010 Census definition". The interface is divided into three main sections: 1. "Select State of current residence": A list of US states with "Kansas" selected. A search box is located below the list. 2. "State of current residence(s) to drop into.": A list containing "Kansas", which is highlighted in blue. Above this list are "Clear All" and "Delete Selection" buttons. 3. "Selected Geographic Areas:": An empty list box. At the bottom of the interface are buttons for "Search", "Previous Level", "Next Level", "Cancel", "Delete Selection", "Clear All", and "Finish".

DATAFERRETT: EXAMPLE

Ferrett Geography Codebook

Select Public use microdata area code (PUMA) based on 2010 Census definition

[Home](#) >> [State of current residence](#) >> [Public use microdata area code \(PUMA\) based on 2010 Census definition](#)

Select Public use microdata area code (PUMA) based on 2010 Census definition

Select All

- Northwest Kansas PUMA, Kansas
- North Central Kansas--Salina City PUMA, Kansas
- Riley, Geary & Pottawatomie Counties--Manhattan City PUMA; Kansas
- Northeast Kansas PUMA, Kansas
- Wyandotte County--Kansas City PUMA, Kansas
- Johnson County (Northwest)--Shawnee, Lenexa & De Soto Cities PUMA; Kansas
- Johnson County (Northeast)--Overland Park (North), Prairie Village & Merriam Cities PUMA, Kansas**
- Johnson County (Southwest)--Olathe, Gardner, Spring Hill & Edgerton Cities PUMA; Kansas

Search:

Public use microdata area code (PU...
Clear All Delete Selection
Please drag wanted geographies to "Selecte

Selected Geographic Areas:
Public Use Microdata Area (PUMA) Count: 1
Johnson County (Northwest)--Shawnee, Lenexa & De Soto C


Search Previous Level Next Level Cancel Delete Selection Clear All Finish

DATAFERRET: EXAMPLE

1 Variables returned from search.

Select All Topics

- Housing
- Selectable Geographies
- Population
- Geographic Entities
- Replicate Weights

 Search Variables

DATA FERRETT: EXAMPLE

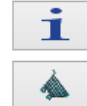
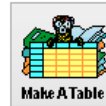
Topic	Name	Availability	Variable Label
Population	PWGTP	2006 - current	PUMS person weight
Population	AGEP	2006 - current	Age
Population	ANC	2006 - current	Ancestry categorization
Population	DECADE	2006 - current	Decade of entry
Population	DRIVESP	2006 - current	Number of vehicles calculated from JWRI
Population	HISP	2006 - current	Hispanic recode
Population	INTP	2006 - current	Interest, dividends, and net rental income past 12 months (signed)
Population	JWAP	2006 - current	Time of arrival at work categorization
Population	JWDP	2006 - current	Time of departure for work - hour and minute
Population	JWMNP	2006 - current	Travel time to work
Population	JWRIP	2006 - current	Vehicle occupancy
Population	NATIVITY	2006 - current	Nativity
Population	NOP	2006 - current	Nativity of Parent
Population	OIP	2006 - current	All other income past 12 months
Population	PAP	2006 - current	PUMS SSI/AFDC/other welfare income
Population	PERNP	2006 - 2012	Total person's earnings
Population	PINCP	2006 - 2012	Total person's income (signed)
Population	POVPIP	2006 - current	Income-to-poverty ratio recode
Population	SPORDER	2006 - current	Person key after swapping
Population	QTRBIR	2006 - current	Quarter of birth
Population	RACAIAN	2006 - current	Race includes AIAN
Population	RACASN	2006 - current	Race includes Asian
Population	RACBLK	2006 - current	Race includes Black
Population	RACNUM	2006 - current	Number of race groups represented
Population	RACSOR	2006 - current	Race includes Other race
Population	RACWHT	2006 - current	Race includes White
Population	RETP	2006 - current	Retirement income past 12 months
Population	SEMP	2006 - current	Self-employment income past 12 months (signed)
Population	SSIP	2006 - current	Supplementary Security Income past 12 months
Population	SSP	2006 - current	PUMS Social Security or Railroad Retirement Income
Population	WAGP	2006 - current	PUMS Wages/salary income
Population	WAOB	2006 - current	World area of birth
Population	WKHP	2006 - current	Usual hours worked per week past 12 months
Population	CIT	2006 - current	Citizenship status
Population	COW	2006 - current	Class of worker
Population	DDRS	2006 - current	Difficulty dressing
Population	DPHY	2006 - current	Physical difficulty
Population	DREM	2006 - current	Difficulty remembering
Population	ENG	2006 - current	English ability

DATAFERRET: EXAMPLE

Topic	Name	Availability	Variable Label
Selectable Geographies	Geography	2012 - current	Geographic Items
Population	DIS	2008 - current	Disability recode
Population	HISP	2006 - current	Hispanic recode
Population	RAC1P	2012 - current	Recoded detailed race code
Population	SEX	2006 - current	Sex
Population	PWGTP	2006 - current	PUMS person weight

RECODING VARIABLES

Review your variables then go back to select more variables or go on to get data



Current Query Variables from ACS (Public Use Microdata Sample):

Name	Variable Label	Availability
GEOG-101	Public Use Microdata Area (PUMA)	2012 - current
DIS	Disability recode	2008 - current
RAC1P	Recoded detailed race code	2012 - current
HISP	Hispanic recode	2006 - current

Act on your Query:

Recode Variable

Download Variable(s)

View/Modify Variable(s)

Advanced Sql Option

Change Longitudinal Period


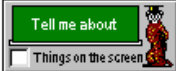
Add TimeSeries Time

Merge Datasets

Create Multi-Variable Data Step

RECODING VARIABLES

Ferrett Microdata Recode2

 Recode/Regroup Variables  Things on the screen

Hisp2 is label for the Variable Recode of HISP

Highlight the value(s) to recode/regroup

Value	Description
01	Not Spanish/Hispanic/Latino
02	Mexican
03	Puerto Rican
04	Cuban
05	Dominican
06	Costa Rican
07	Guatemalan
08	Honduran
09	Nicaraguan

Select	Label	Values
<input checked="" type="checkbox"/>	1 Non-Hispanic	{01}
<input checked="" type="checkbox"/>	2 Hispanic	{02 ,03 ,04 ,05 ,06 ,07 ,08 ,09 ,10 ,11 ,12 ,13 ,14 ,15}

Set to value 2

Redefine Current Row

RECODING VARIABLES

Ferrett Microdata Recode3

Recode/Regroup Variables

Tell me about
Things on the screen

Race2 is label for the Variable Recode of RAC1P

Highlight the value(s) to recode/regroup

Value	Description
1	White alone
2	Black or African American alone
3	American Indian alone
4	Alaska Native alone
5	American Indian and Alaska Native tribes specified; or American Indian
6	Asian alone
7	Native Hawaiian and Other Pacific Islander alone
8	Some other race alone

Select	Label	Values
<input checked="" type="checkbox"/>	1 White alone	{1}
<input checked="" type="checkbox"/>	2 African American alone	{2}
<input checked="" type="checkbox"/>	3 Other	{3 ,4 ,5 ,6 ,7 ,8 ,9}

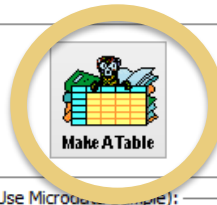
Set to value **3** **Recode**

Ok Cancel

Redefine Current Row

MAKING A TABLE

Review your variables then go back to select more variables or go on to get data



Current Query Variables from ACS (Public Use Microdata Sample):

Variable Name	Variable Label	Availability
GEOG-103	Public Use Microdata Area (PUMA)	2012 - current
DIS	Disability recode	2008 - current
HISP	Hispanic recode	2006 - current
RAC1P	Recoded detailed race code	2012 - current
SEX	Sex	2006 - current
PWGTP	PUMS person weight	2006 - current
RECODE2	Hispanic	2006 - current
RECODE3	Race	2012 - current

MAKING A TABLE

The screenshot shows the Ferrett Tabulation software interface. A yellow circle highlights the 'GO Get Data' button in the top toolbar. The main window displays a data table with the following structure:

		C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	
		Total RECODE2				Non-Hispanic				Hispanic															
R1	Total	White alone	African American alone	Other	Total	White alone	African American alone	Other	Total	White alone	African American alone	Other													
R2	Total DIS	?	?	?	?	?	?	?	?	?	?	?													
R3	Male	?	?	?	?	?	?	?	?	?	?	?													
R4	Female	?	?	?	?	?	?	?	?	?	?	?													
R5	With a disability	?	?	?	?	?	?	?	?	?	?	?													
R6	Male	?	?	?	?	?	?	?	?	?	?	?													
R7	Female	?	?	?	?	?	?	?	?	?	?	?													
R8	Without a disability	?	?	?	?	?	?	?	?	?	?	?													
R9	Male	?	?	?	?	?	?	?	?	?	?	?													
R10	Female	?	?	?	?	?	?	?	?	?	?	?													
R11																									
R12																									
R13																									

Instances:
 GEOG-103 Public Use Microdata Area (PUMA)
 DIS Disability recode
 HISP Hispanic recode
 RAC1P Recoded detailed race code
 SEX Sex
 RECODE2 Hisp2
 RECODE3 Race2

MAKING A TABLE

GO Get Data



Col C13 RECODE2=2, RECODE3=3

Pivot(s) can be dropped on pivot image above R1.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	
		Total RECODE2				Non-Hispanic				Hispanic					
		Total	White alone	African American alone	Other	Total	White alone	African American alone	Other	Total	White alone	African American alone	Other		
R1															
R2	Total DIS	115,193	96,816	5,034	13,344	108,687	92,128	4,920	11,638	6,506	4,688		113	1,705	
R3	Male	56,828	48,297	2,481	6,050	53,689	45,974	2,368	5,347	3,140	2,323		113	703	
R4	Female	58,365	48,518	2,552	7,294	54,998	46,154	2,552	6,292	3,366	2,364		0	1,002	
R5	With a disability	11,176	9,274	599	1,303	10,667	9,010	599	1,058	509	265		0	244	
R6	Male	5,602	4,650	205	747	5,111	4,385	205	521	491	265		0	226	
R7	Female	5,574	4,624	394	556	5,556	4,624	394	538	18	0		0	18	
R8	Without a disability	104,017	87,542	4,434	12,041	98,020	83,119	4,322	10,580	5,997	4,423		113	1,461	
R9	Male	51,226	43,648	2,276	5,303	48,578	41,589	2,163	4,826	2,648	2,058		113	477	
R10	Female	52,791	43,894	2,159	6,738	49,442	41,530	2,159	5,754	3,348	2,364		0	984	

```

Universe: ((ST = 20 AND PUMA = 00700))
Weight used: PWGTP
DataSet(s) selected: 2015.2014.2013.2012
    
```



IPUMS

IPUMS

- Online features like DataFerrett for PUMS and CPS
- Consistent names and coding
 - Easier to use
- Several other datasets



U.S. Census and American Community Survey microdata from 1850 to the present.

[VISIT SITE](#)



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.

[VISIT SITE](#)



Census microdata covering 82 countries from 1960 to the present. [IPUMS NAPP](#) offers microdata from the 19th and early 20th centuries.

[VISIT SITE](#)



Demographic and Health Surveys integrated for analysis across time and space from 1980 to the present.

[VISIT SITE](#)



Tabular U.S. Census data and GIS boundary files from 1790 to the present.

[VISIT SITE](#)



Integrated data on population and the environment from 1960 to the present.

[VISIT SITE](#)



Historical and contemporary time use data from 1965 to the present.

[VISIT SITE](#)



Health survey data from the National Health Interview Survey from the 1960s to the present.

[VISIT SITE](#)



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

[VISIT SITE](#)



Compare and Contrast: PUMS and Other Datasets

LEARNING OBJECTIVES

1. Learn how to prioritize data sets.
2. Identify when to use PUMS or specialized datasets.
3. Understand the difference between public vs. restricted data.

PRIORITIZATION OF DATASETS

- What question(s) are you trying to answer?
- Is the information already available at a level you need?
 - Years
 - Geography
 - Unit level
- How accessible is the dataset?
 - Cost and restricted use
- What are the limitations of the data source?
 - Release cycle
 - Data collection methods
 - Sample

SECRET

THIS IS A COVER SHEET

FOR CLASSIFIED INFORMATION

ALL INDIVIDUALS HANDLING THIS INFORMATION ARE REQUIRED TO PROTECT IT FROM UNAUTHORIZED DISCLOSURE IN THE INTEREST OF THE NATIONAL SECURITY OF THE UNITED STATES.

HANDLING, STORAGE, REPRODUCTION AND DISPOSITION OF THE ATTACHED DOCUMENT MUST BE IN ACCORDANCE WITH APPLICABLE EXECUTIVE ORDER(S), STATUTE(S) AND AGENCY IMPLEMENTING REGULATIONS.

RESTRICTED DATA

THE ATTACHED DOCUMENT CONTAINS RESTRICTED DATA AS DEFINED IN THE ATOMIC ENERGY ACT OF 1954. UNAUTHORIZED DISCLOSURE SUBJECT TO ADMINISTRATIVE AND CRIMINAL SANCTIONS.

(This cover sheet is unclassified.)

SECRET

704-101
NSN 7540-01-213-7902

Exception Granted by GSA (9/88)

STANDARD FORM 704 (8-65)
Prescribed by GSA/ISCO
(41 CFR 101-11.6)

PUBLIC USE VS. RESTRICTED DATA

- Some restricted data exist in both public and private sector
 - BRFSS has some county-level data, restricted
 - Census Research Data Center
- Restricted data can be slow to acquire
 - Related costs
 - Approval process

SPECIALIZED DATASET AREAS: EXAMPLES

- Economic
- Housing
- Health
- Environmental

ECONOMIC

	ACS/PUMS	CPS (March Supplement)	SAIPE	LAU
Strengths	<ul style="list-style-type: none">• Annual, higher response, survey• Demographics• Microdata	<ul style="list-style-type: none">• More detailed income• Demographics• Microdata	<ul style="list-style-type: none">• Best local area poverty estimates	<ul style="list-style-type: none">• County-level employment and workforce data
Limitations	<ul style="list-style-type: none">• Limited Economic categories• May need to combine	<ul style="list-style-type: none">• State- or Metropolitan Statistical Area MSA-level, only.• Need to combine data for estimate	<ul style="list-style-type: none">• Slow to release• No demographics• Aggregated data	<ul style="list-style-type: none">• No demographics• Aggregated data

HOUSING

	ACS/PUMS	AHS	CHAS
Strengths	<ul style="list-style-type: none">• Annual, higher response, survey• County- and tract-level data• Microdata	<ul style="list-style-type: none">• Detailed housing questions• National and MSA only	<ul style="list-style-type: none">• Gauges housing problems• County-level
Limitations	<ul style="list-style-type: none">• Limited housing questions• May need to combine years	<ul style="list-style-type: none">• Slow to release• No demographics• Aggregated data	<ul style="list-style-type: none">• Multiple year – rolling average• Aggregated data

HEALTH

	ACS/PUMS	BRFSS
Strengths	<ul style="list-style-type: none">• Annual, higher response, survey• County- and tract-level data• Microdata	<ul style="list-style-type: none">• Annual survey• More health questions• Behaviors, physical, mental, chronic• Microdata
Limitations	<ul style="list-style-type: none">• Limited health questions• May need to combine years	<ul style="list-style-type: none">• Restricted for sub-state• Sub-state only for larger areas• May need to combine years

ENVIRONMENTAL

- ACS/PUMS does not gauge environment
- Some resources (limited)
 - TOXMAP
 - Identify toxic release sites, coal plants, superfund sites
 - National Environmental Public Health Tracking Network (Tracking Network)
 - Air quality measures for a county

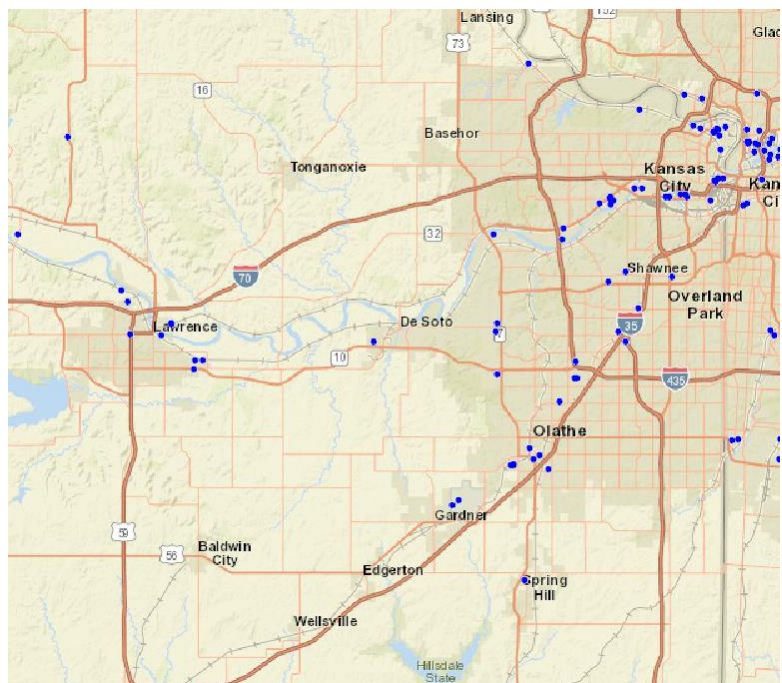
TOXMAP

Browse TOXMAP

Click check boxes to turn ON / OFF map data
For other TRI years, use the [Search Tool](#)

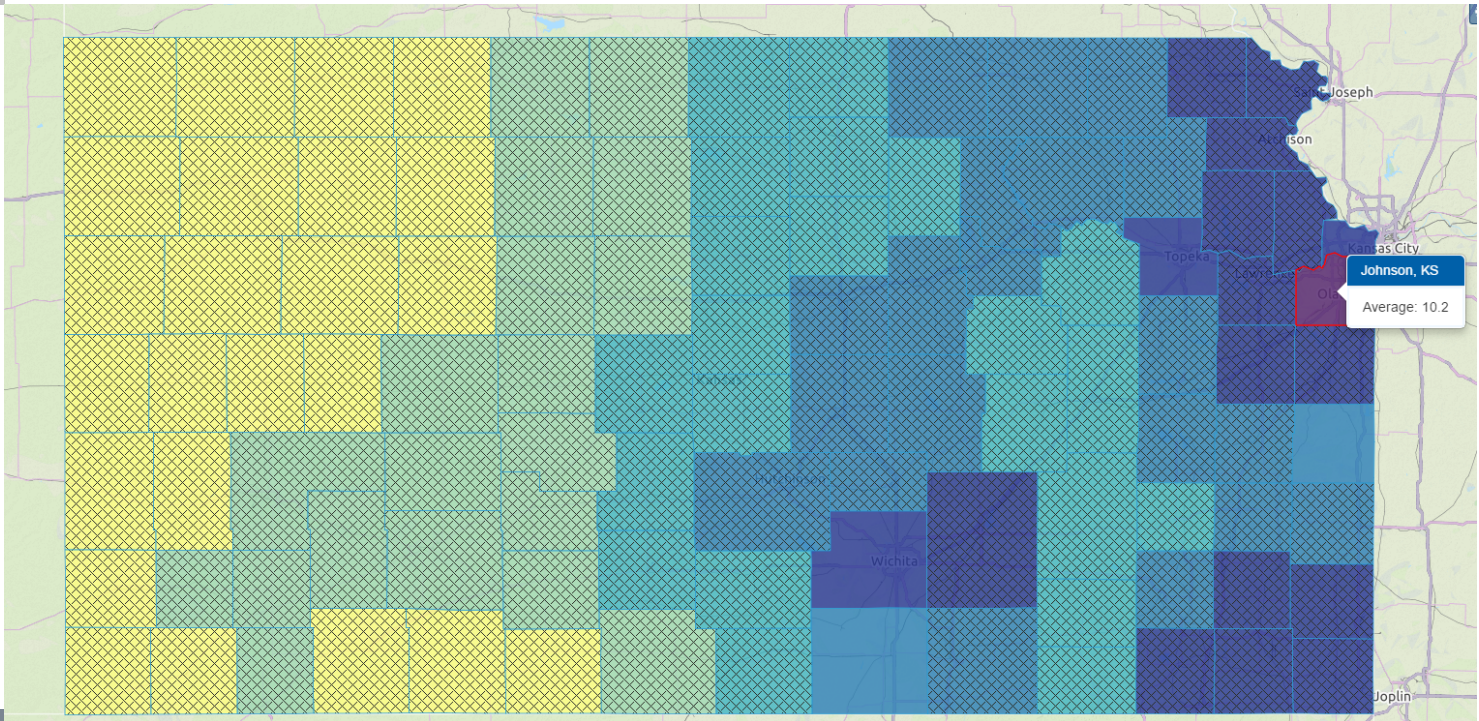
[-] TOXMAP Data

- All TRI Facilities (1988-2015)
 - TRI facility (zoomed out)
 - TRI facility
 - ▣ TRI facility (zoomed in)
- TRI Facilities - 2015
 - TRI facility (zoomed out)
 - TRI facility
 - ▣ TRI facility (zoomed in)
- Superfund National Priority List
 - Superfund site (zoomed out)
 - Superfund site
 - ▣ Superfund site (zoomed in)
- NPRI (Canada Only) - 1994-2013
 - NPRI facility (zoomed out)
 - NPRI facility
 - ▣ NPRI facility (zoomed in)
- [-] Other**
 - Nuclear plants (commercial)
 - Hospitals (zoom in)
 - ▣ 114th Congressional
 - ▣ EPA Coal Plants 2013
 - ▣ Native Lands (AI / AN / NH)



TRACKING NETWORK

Average
PM2.5
Levels



FINAL THOUGHTS

- Good for community-level research
- Flexible dataset
 - Scope of information (economic, social and health-related questions)
 - Microdata
- Resources are available on the website to assist users

Thank you.

ANY QUESTIONS?

You can connect with us at:

jhamdorf@khi.org

lpanas@khi.org

chuang@khi.org

kbruffett@khi.org

tlin@khi.org



212 SW 8th Avenue | Suite 300
Topeka, Kansas | 66603-3936



785.233.5443



khi.org



/KHlorg



@KHlorg