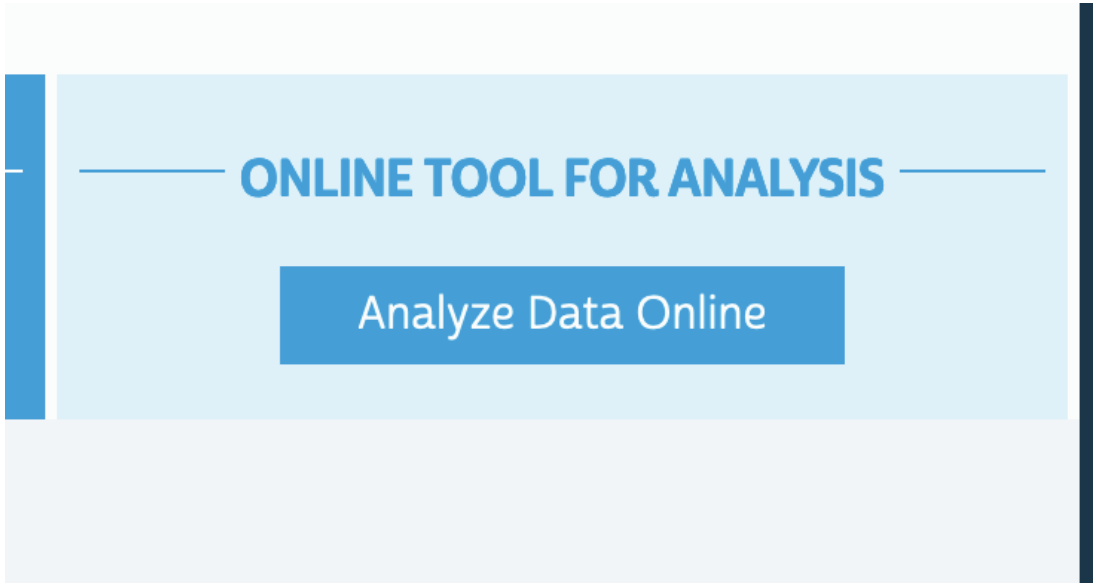
Hands-On Activity: Online Tools for Analysis on IPUMS USA

Module 3: Using Data for Immigrants in the Gulf Region

Step 1: Access the IPUMS USA Online Data Analysis System

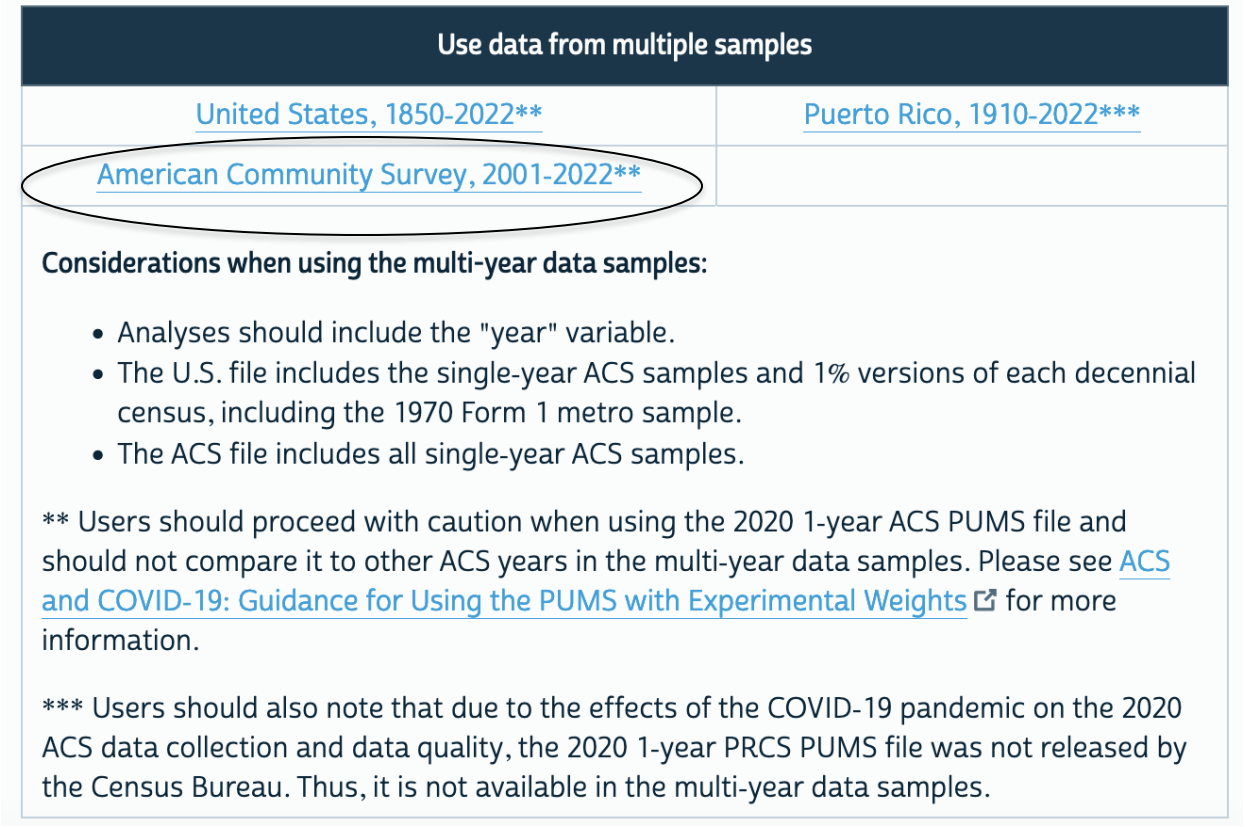
- Navigate to the IPUMS USA Online Data Analysis System: https://usa.ipums.org/usa/sda/

- This tool allows users to perform online analysis without downloading data.



Step 2: Select a Dataset

- Choose a dataset relevant to your research.
- For analyzing immigrants in the Gulf Region, use 'American Community Survey, 2001-2022'.
- This dataset includes key variables for demographic and socioeconomic analysis.



Use data from multiple samples	
United States, 1850-2022**	Puerto Rico, 1910-2022***
American Community Survey, 2001-2022**	

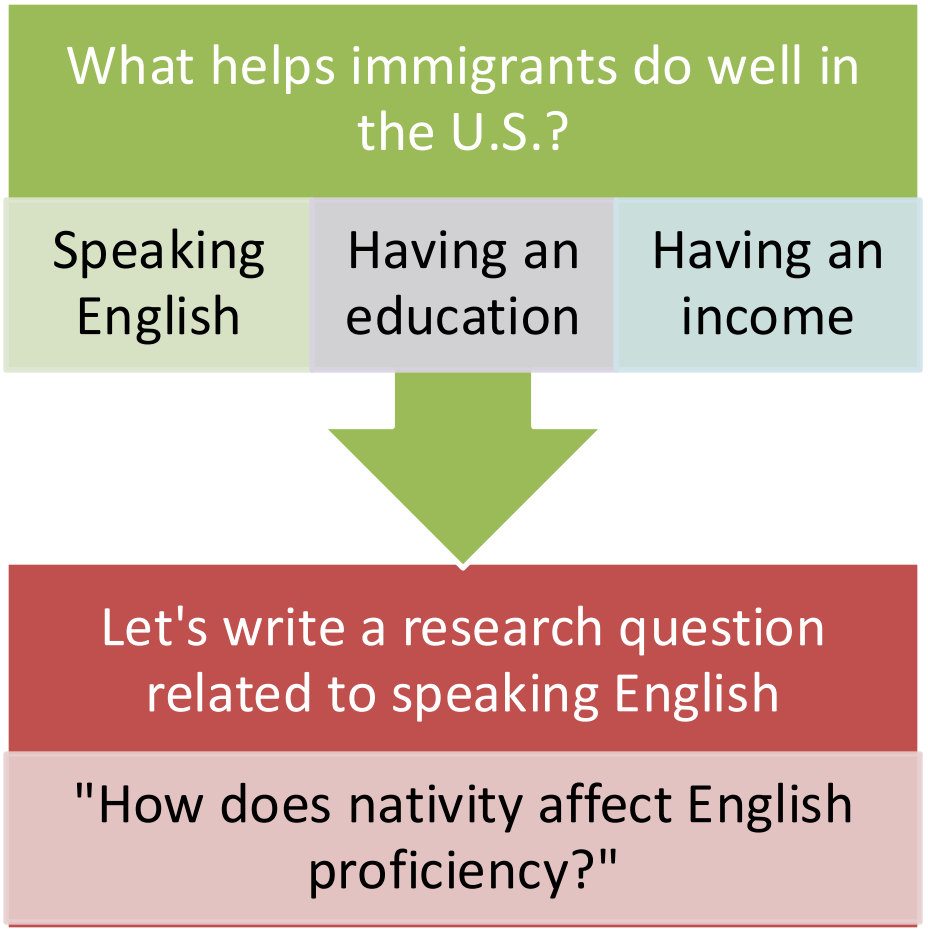
Considerations when using the multi-year data samples:

- Analyses should include the "year" variable.
- The U.S. file includes the single-year ACS samples and 1% versions of each decennial census, including the 1970 Form 1 metro sample.
- The ACS file includes all single-year ACS samples.

** Users should proceed with caution when using the 2020 1-year ACS PUMS file and should not compare it to other ACS years in the multi-year data samples. Please see ACS and COVID-19: Guidance for Using the PUMS with Experimental Weights for more information.

*** Users should also note that due to the effects of the COVID-19 pandemic on the 2020 ACS data collection and data quality, the 2020 1-year PRCS PUMS file was not released by the Census Bureau. Thus, it is not available in the multi-year data samples.

Research Question Formulation



Step 3: Define Variables for Analysis

1

• Row Variable: Select a variable to analyze trends, e.g., ‘speakeng’ (speaks English)

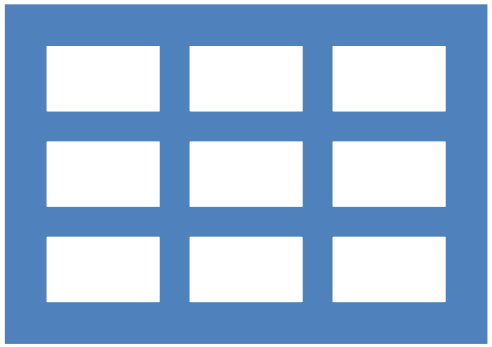
2

• Column Variable: Choose a variable like ‘nativity’ (born in the US or Born outside of the US).

3

• Filter: Apply filters to focus on Gulf Region states:

- Use 'STATEFIP' and include Texas (48), Louisiana (22), Mississippi (28), Alabama (01), and Florida (12).



Research Queston ("How does nativity affect English proficiency?"), here's how you should organize the variables:

- For **Independent Variable (Nativity):** Place this in the **column.**
  - Reason: Columns often represent categories that define groups (e.g., born in the U.S. or born outside the U.S.).
- **Dependent Variable (Speaks English):** Place this in the **row.**
  - Reason: Rows typically show outcomes or response distributions (e.g., the range from "not at all" to "really well").
- This setup allows you to analyze how English proficiency (row variable) varies across different nativity groups (column variable).

Step 4: Run the Analysis

- Execute the analysis to generate tables or charts.
- Analyze patterns such as immigrant trends by country of origin or socioeconomic status.

Variables

Role	Name	Label	Range	MD	Dataset
Row	**speakeng**	Speaks English	0-6		1
Column	**nativity**		0-1		2
Weight	**perwt**	Person weight	1.00-4,044.00		1
Filter	**statefip(48)**	State (FIPS code)(=Texas)	1-56		1

Frequency Distribution

Cells contain:
-Column percent
-Weighted N

		nativity: 0 US Born	nativity: 1 Born outside US	***ROW TOTAL***
speakeng	0: N/A (Blank)	**8.9** 41,179,730.0	**1.1** 1,057,569.0	***7.5*** *42,237,299.0*
	1: Does not speak English	**.2** 1,082,165.0	**14.1** 14,076,865.0	***2.7*** *15,159,030.0*
	3: Yes, speaks only English	**71.0** 330,043,904.0	**12.8** 12,831,633.0	***60.7*** *342,875,537.0*
	4: Yes, speaks very well	**16.2** 75,201,612.0	**31.9** 31,888,903.0	***19.0*** *107,090,515.0*
	5: Yes, speaks well	**2.7** 12,781,483.0	**19.2** 19,249,823.0	***5.7*** *32,031,306.0*
	6: Yes, but not well	**1.0** 4,563,550.0	**21.0** 20,984,673.0	***4.5*** *25,548,223.0*
	COL TOTAL	***100.0*** *464,852,444.0*	***100.0*** *100,089,466.0*	***100.0*** *564,941,910.0*

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	**Z**
N in each cell:	Smaller than expected			Larger than expected			

Interpreting Nativity and English Proficiency Data

Understanding Relationships in Frequency Distribution Tables

Overview of the Variables

- Row Variable: 'speakeng' (Speaks English)

  - Scale ranges from 0 to 6 (e.g., 'Does not speak English' to 'Speaks very well').

- Column Variable: 'nativity'

  - Categories:

  - 0: U.S.-born

  - 1: Born outside the U.S.

- Filter Applied: Data focuses on Texas (state FIPS code = 48).

Table Structure

- What does the table show?
- The relationship between nativity (columns) and English proficiency (rows).

Interpreting the Columns

Column Percentages show the distribution of English proficiency within each nativity group.

U.S.-Born (Column 0):

- 71.0% speak only English (row 3).
- 16.2% speak very well (row 4).

Born Outside the U.S. (Column 1):

- 31.9% speak very well (row 4).
- 21.0% speak 'but not well' (row 6).

Key Observations

- U.S.-Born Population:

  - Majority (71%) speak only English.

  - Very few (0.2%) do not speak English.

- Born Outside the U.S.:

  - Significant proportion (31.9%) speak English very well.

  - Higher percentages (21.0%) report speaking English 'but not well'.

Color Coding

- Color Coding:

  - Red: Higher than expected values (>1.0 or >2.0 standard deviations).

  - Blue: Lower than expected values (<-1.0 or <-2.0 standard deviations).

- Example:

  - Born Outside the U.S. (Column 1):

    - 'Does not speak English' (Row 1): 14.1% is larger than expected (red).

  - U.S.-Born (Column 0):

    - 'Does not speak English' (Row 1): 0.2% is smaller than expected (blue).

Insights and Implications

- Key Insights:
  - Nativity significantly affects English proficiency.
  - U.S.-born individuals overwhelmingly speak only English.
  - Immigrants show varying levels of English proficiency.

- Implications:
  - Highlights the need for targeted language services and programs for immigrants.

Discussion Questions

1. How does the nativity group influence English proficiency trends?

2. How could this data inform social or educational policy?