# CIVIC DIGITAL FELLOWSHIP

**Building a Python-Based Administrative Record Lookup File**

**Rui-Jie Yew**

**Supervised by Jennifer Ortman, Lori Zehr, and Michael Ratcliffe**

**Geography Division**

Shape
your future
START HERE >

United States®
Census
2020

# What is the Administrative Record (AR) Lookup File?



| PIK | DOB | ... | ACS2018 Race1 | Census2010 Sex |
|---|---|---|---|---|
| 00001001 | 06011980 | ... | 200 | 1 |
| 00200345 | 02032000 | ... | 100 | 2 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| 03002031 | 12301945 | ... | 100 | 1 |

500,000,000 Rows

200+ Columns

Shape your future START HERE ›

United States® Census 2020

# Existing Problem:
# Using the AR Lookup File for Item Nonresponse Follow-Up

→ **Item Nonresponse: When a sampling unit does not respond to a survey question.**

**What is the DOB for PIK = 00200345?**

| PIK | DOB | ... | ACS2018 Race1 | Census2010 Sex |
|-----|-----|-----|---------------|----------------|
| 00001001 | 06011980 | ... | 200 | 1 |
| 00200345 | 02032000 | ... | 100 | 2 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| 03002031 | 12301945 | ... | 100 | 1 |

➢ **Current query process returns data for all individuals and characteristics in the file**

➢ **Unnecessary Information**

➢ **Slow Process**

# Project Objective

→ **Research and develop a proof-of-concept for migrating AR data into a queryable Python database on a CAES server, so that we can support item nonresponse follow-up more efficiently for the 2030 Census Program.**

**What is the DOB for PIK = 00200345?**

| PIK | DOB | ... | ACS 2018 Race1 | Census 2010 Sex |
|-----|-----|-----|----------------|-----------------|
| 00001001 | 06011980 | ... | 200 | 1 |
| 00200345 | 02032000 | ... | 100 | 2 |
| 03002031 | 12301945 | ... | 100 | 1 |

**06041975**

Title13

Shape your future
START HERE >

United States®
Census
2020

# Deliverables

- ✓ **Python-Based Administrative Record Lookup Web Application**

- ✓ **Documentation for both application users and programmers**

- ✓ **Parallel Computing with Dask–fast querying and file loading**

# Next Steps

→ **Determine best source of information**
→ **Smooth transition between SAS and Python**