# Demographic Changes in the US using the application Streamlit as dashboard.

Semester Project Proposal for Econ 8320 - Tools for Data Analysis
Souleymane Diawara
Monday 8th April 2024

## Introduction

The objective of this project is to build a Streamlit dashboard that highlights demographic changes in the US from 2010 to 2023 using data collected from the US Census Bureau API. The dashboard will provide an interactive visualization of the demographic makeup of cities across the US over time, focusing on at least 7 demographic variables.

Our dashboard will present the distribution of population by age groups, by gender, by educational attainment by income level and by employment status.

We will follow the process below to collect the data:

a. Create the GitHub repository for the project.
b. Use the US Census Bureau API to collect data at the year-month-city level.
c. Build the python project:
   1. collect at once all the data,
   2. build the interactive dashboard using Streamlit
   3. build the script that will run once per month through GitHub Actions to generate monthly data
   4. update previous dataset with new collected items.

## Understanding about the US Census Bureau API

Useful links: ①,②

According to the CPS website, the API call look like is format:

http://api.census.gov/data/2024/cps/basic/jan?get=PEMLR,PWSSWGT,PEMARITL&for=state:01&PEEDUCA=39&key=YOUR_KEY_GOES_HERE

With all the variables defined in the JSON link: https://api.census.gov/data/2024/cps/basic/jan/variables.json

A glance, this API indicates that we need to know about a couple of items

| URL | Interpretation |
|---|---|
| `api.census.gov/data/` | Fix part / base |
| `2024` | Year of data collected |
| `/cps/basic/` | Fix part |
| `jan` | Month of January |
| `?get=` | Fix part, should be followed by the param we are collecting |
| `PEMLR,PWSSWGT,PEMARITL` | Variables |
| `&for=` | Fix part, should be followed by the Geography |
| `state:01` | Region, Division, State, state› county "state:* " is for all states |
| `&` | Fix part |
| `PEEDUCA=39` | Demographics-highest level of school completed 39 is High School or GED |
| `&` | Fix part |
| `key=YOUR_KEY_GOES_HERE` | Developer API key |

First in order to get anything from CPS, I got an API key in my UNO email by registering at
https://api.census.gov/data/key_signup.html

## Demographic variables chosen

Our project will try to understand the population by presenting social, demographic and economic data:

- Social: Citizenship status, Education, Marital status, Education
- Demographic: Relationship, Sex, Tenure (own or rent)
- Economic: Employment status, Industry and Occupation, Earnings

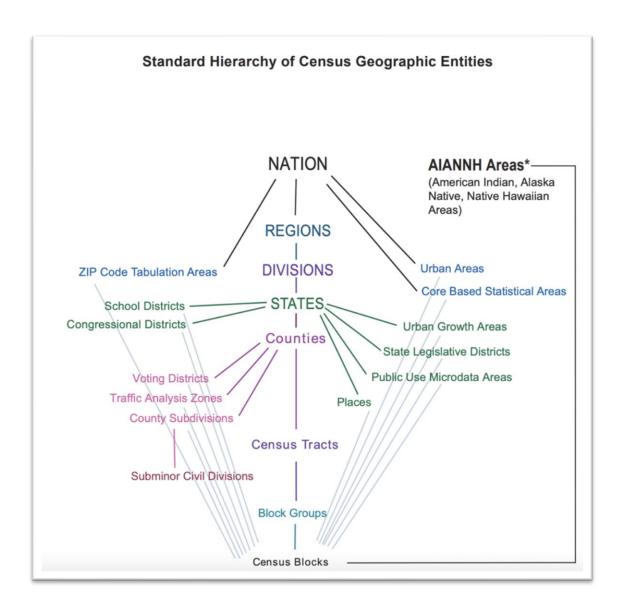### API variables (dependent and independent variables)

| | |
|---|---|
| `HEFAMINC` | Household-total family income in past 12 months |
| `HETENURE` | Household-own/rent living quarters |
| `PEMARITL` | Demographics-marital status |
| `PECYC` | Demographics-years of college credit completed |
| `PEDIPGED` | Demographics-high school, graduation/GED |
| `PEEDUCA` | Demographics-highest level of school completed |
| `PEERNPER` | Earnings-when received, periodicity |
| `PTERN / PTERN2` | Earnings-(calculated)weekly overtime |
| `PTIO1OCD` | Industry and Occupation - Occupation Code (Primary Job) |
| `PTIO2OCD` | Industry and Occupation - Occupation Code (Second Job) |

| PESEX | Demographics-sex |
|---|---|
| PRCITSHP | Demographics-United States citizenship group |

### Variables to track a specific datetime and/or geographic area

| YYYYMM | Year-Month |
|---|---|
| STATE | FIPS STATE Code |
| REGION | REGION |
| QSTNUM | Unique household identifier (Will use it to identify person living under the same roof, a longitudinal value due to conflicting data). |
| QSTNUM | Unique person identifier (Will help count the adult in a household who get interviewed). |
| HRNUMHOU | Household-total # of members |

I don't see the "city" option in the collected data, so will use the county information according to the entity hierarchy below:

**Standard Hierarchy of Census Geographic Entities**

## Dashboard Design

| introduction to the project and objectives | Author: | Last updated version (date) |
|---|---|---|

| Interactive charts and graphs to visualize the selected demographic variables over time, allowing users to explore trends and patterns.<br>years/Month: [____] [____]<br>State: [_____]<br>County: [_____]<br>Demographic: [_____] (drop down menu with content like Rent/own, Average/Mode of household income, Population, | Cluster bar chart<br>Of the overall demographic changes in the US gender distribution. |

A table of detail demographic data for the selected city and time period

| States | 2021 | 2022 | 2023 | 2024 |
|--------|------|------|------|------|
| AL | 3,210,000 (M↑, F) | 2,948,000 (M↑, F) | | |
| AR | 5,810,000 (M, F↓) | 2,520,000 (M, F↓) | | |

Multiple Regression Analysis
[ Model, Coef/ beta, St Error, t , Conf. Interval ]