

# Understanding the New ACS Summary File Format A User's Perspective

Jonathan Schroeder

@j\_p\_schroeder

ACS Data Users Group – June 2, 2022



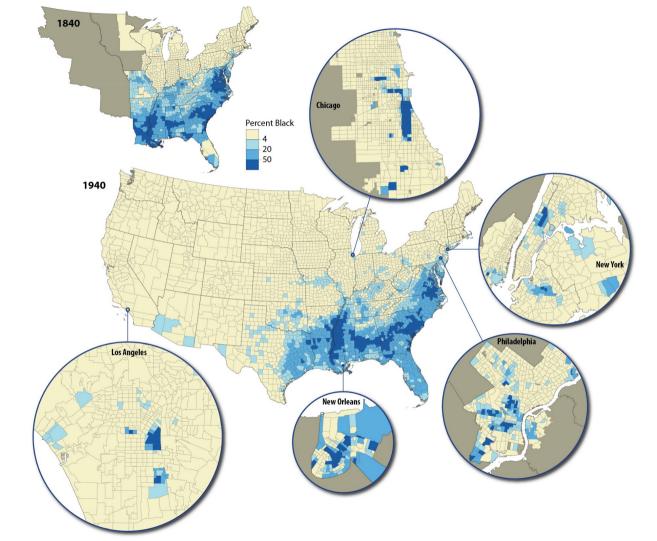


#### **ACS SUMMARY FILES IN NHGIS**

IPUMS.ORG

### **NHGIS**

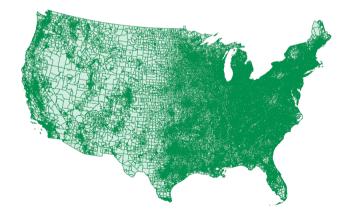
- Geographic U.S. census & survey data
- Summary tables and GIS files
- 1790 to the present



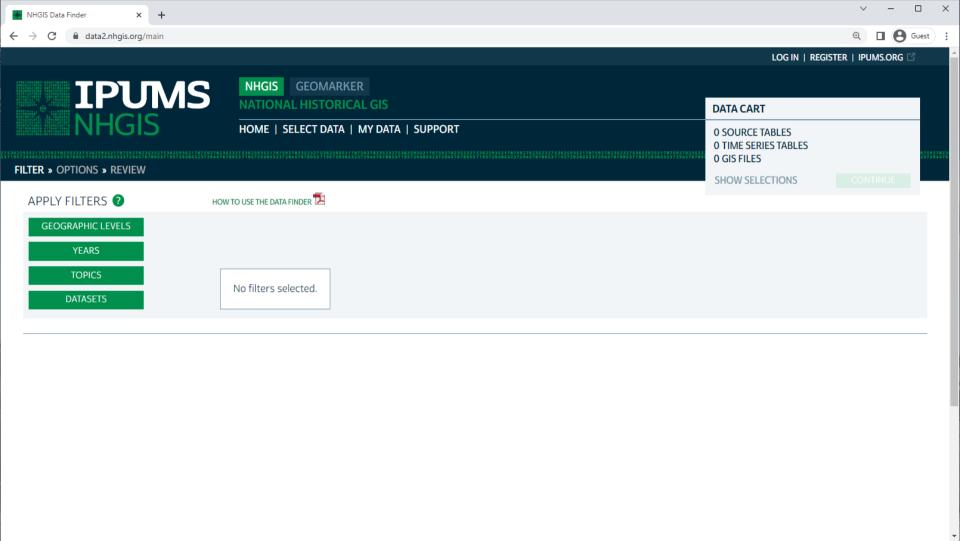
### **NHGIS**

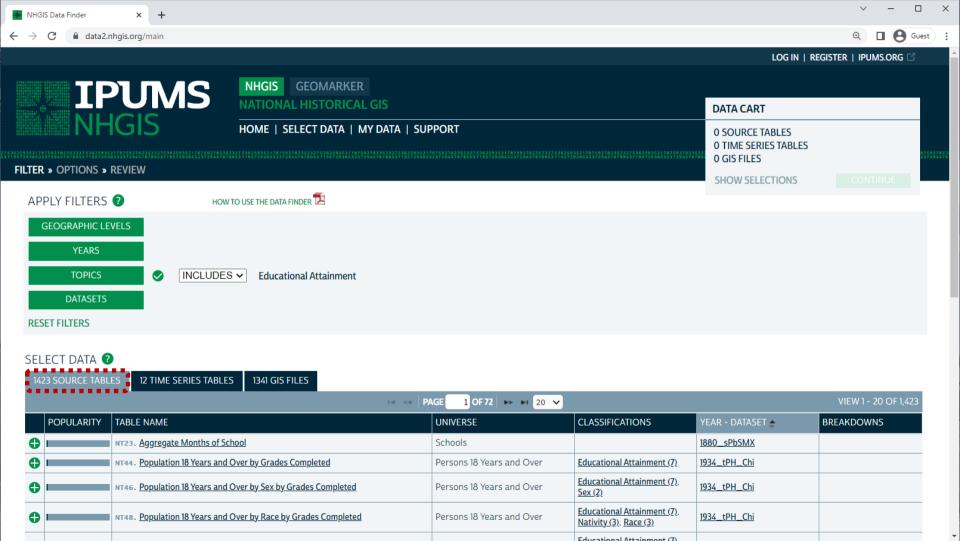
- Nationwide extents
- Principally funded by NSF & NIH grants...
- FREE

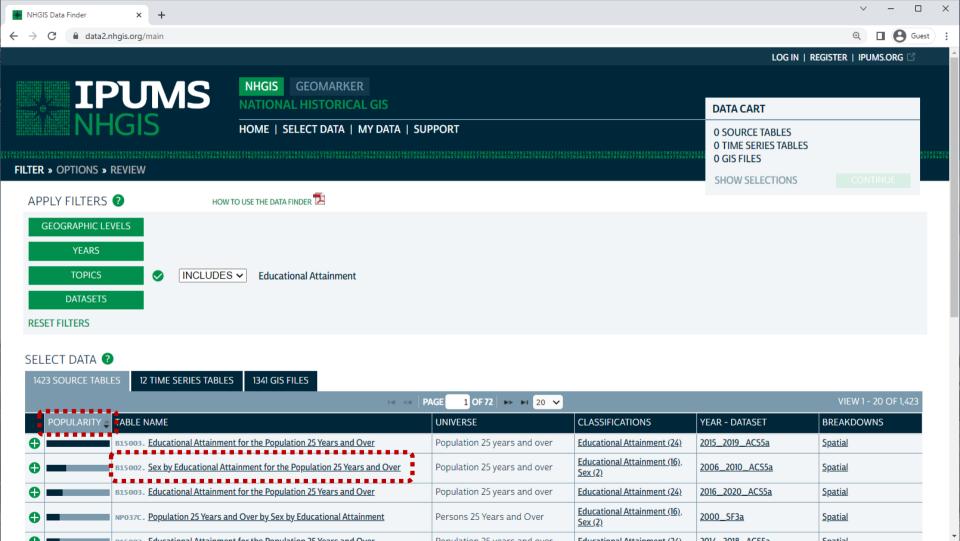
2020 block group boundaries

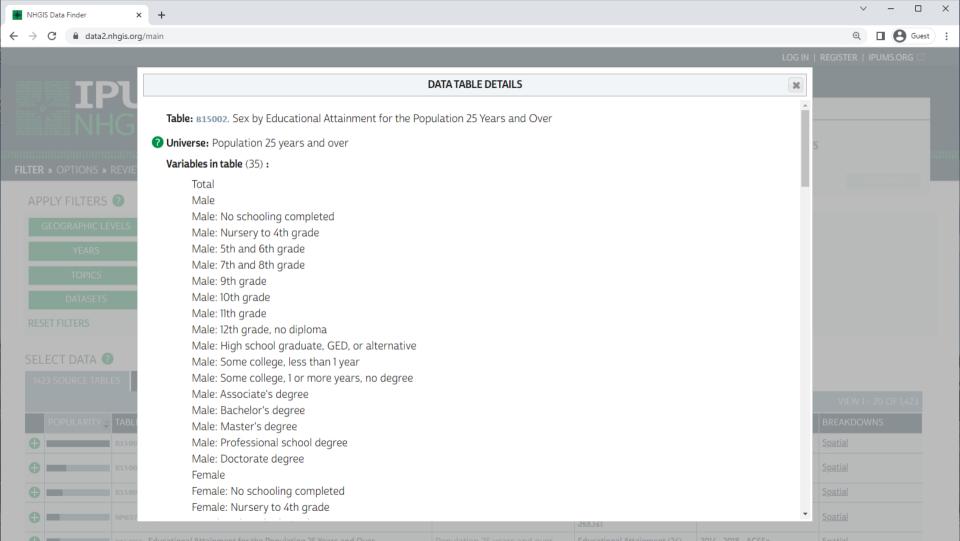


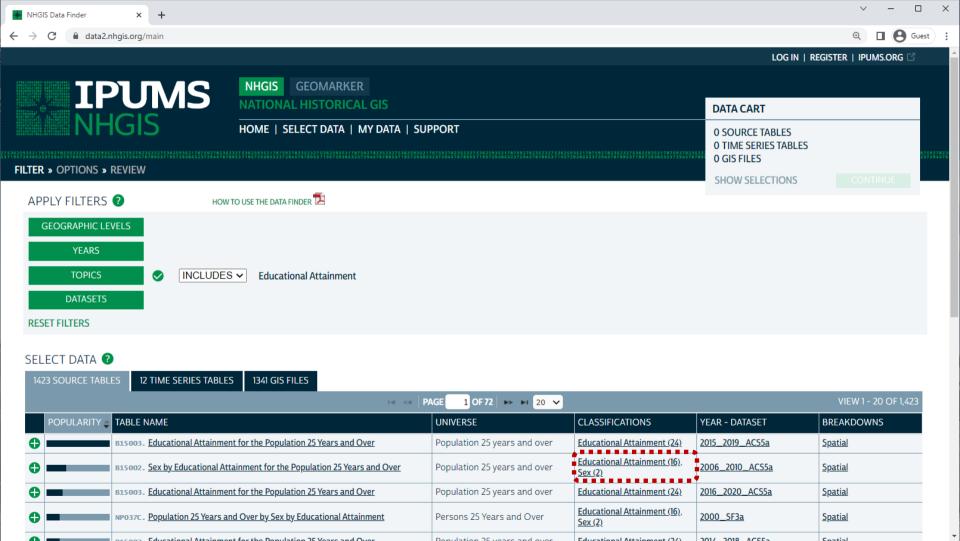


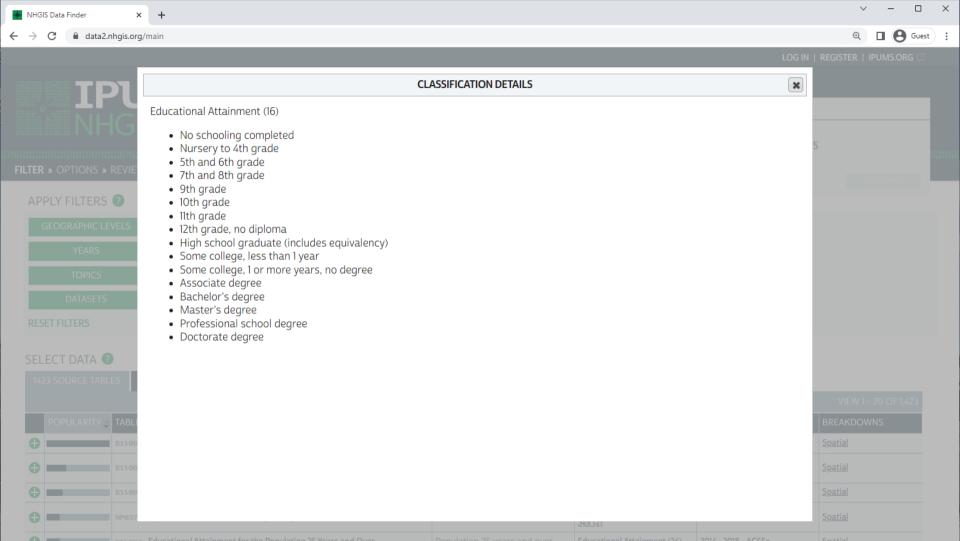


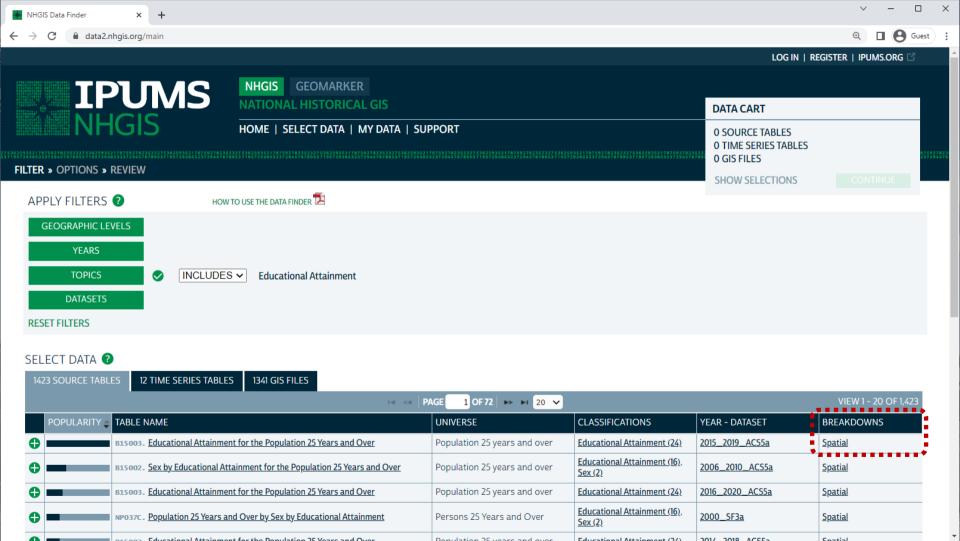


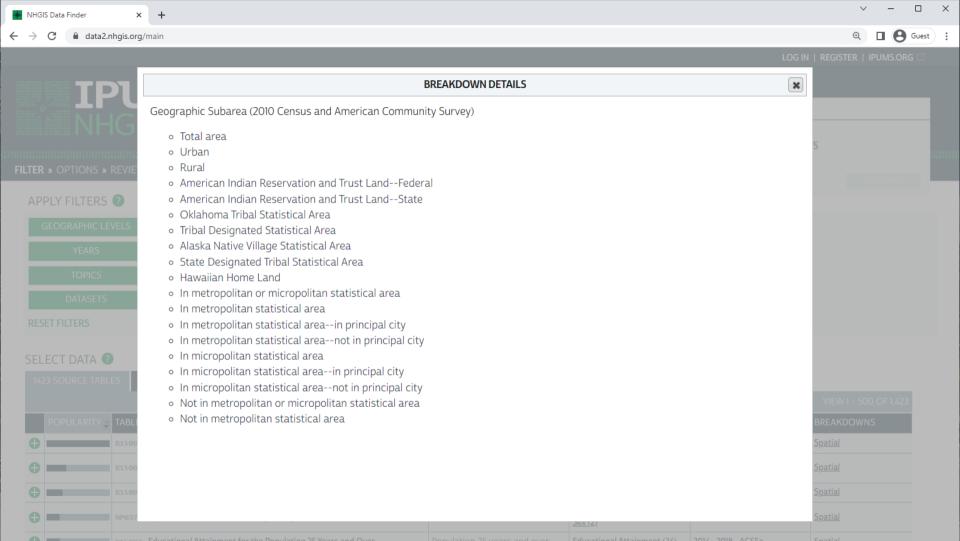


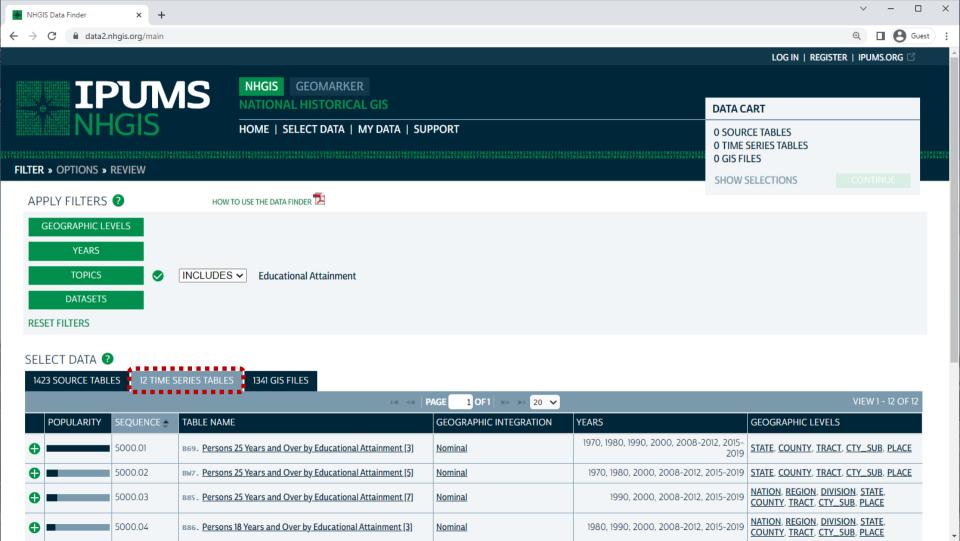


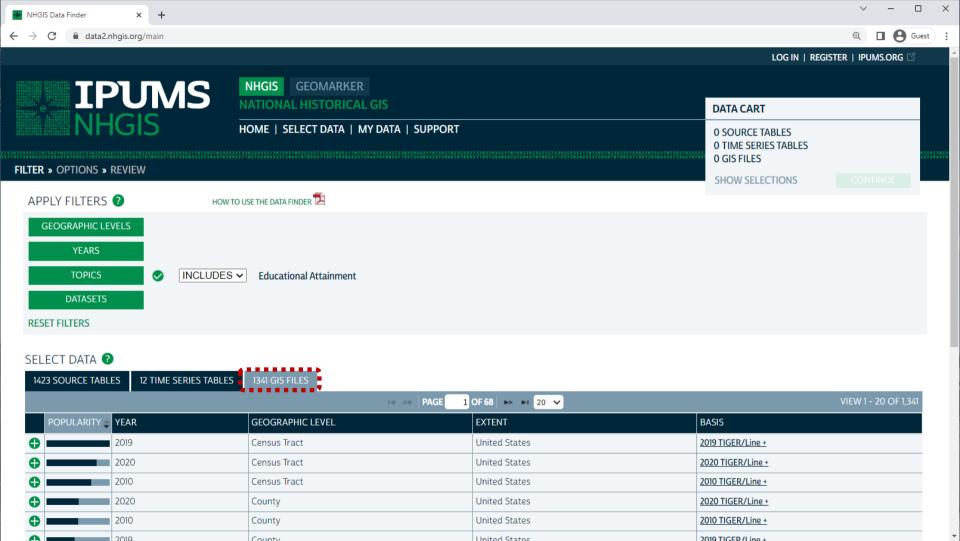


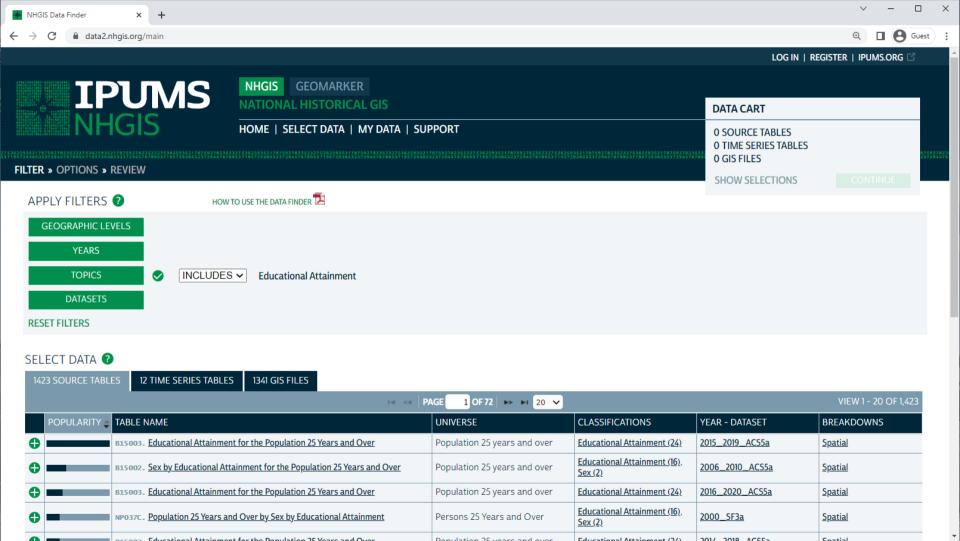


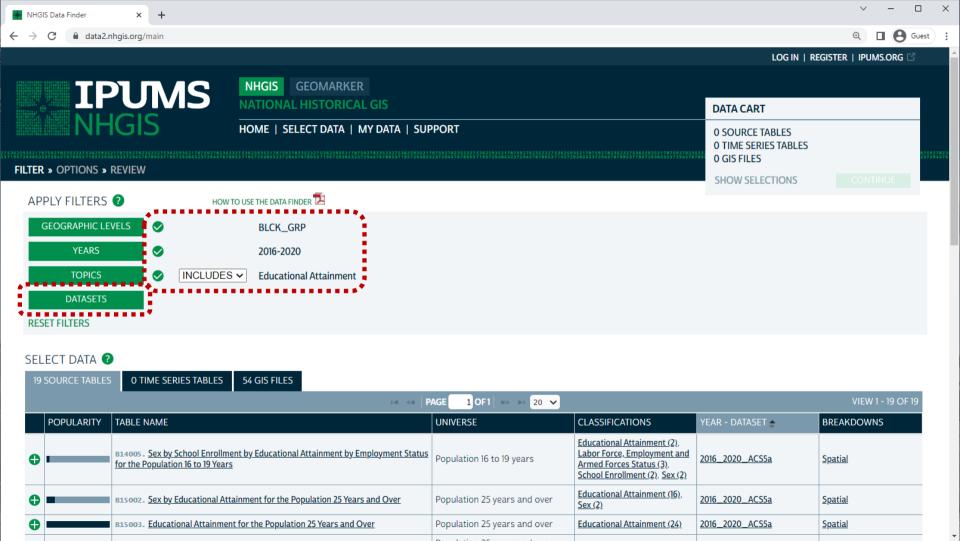


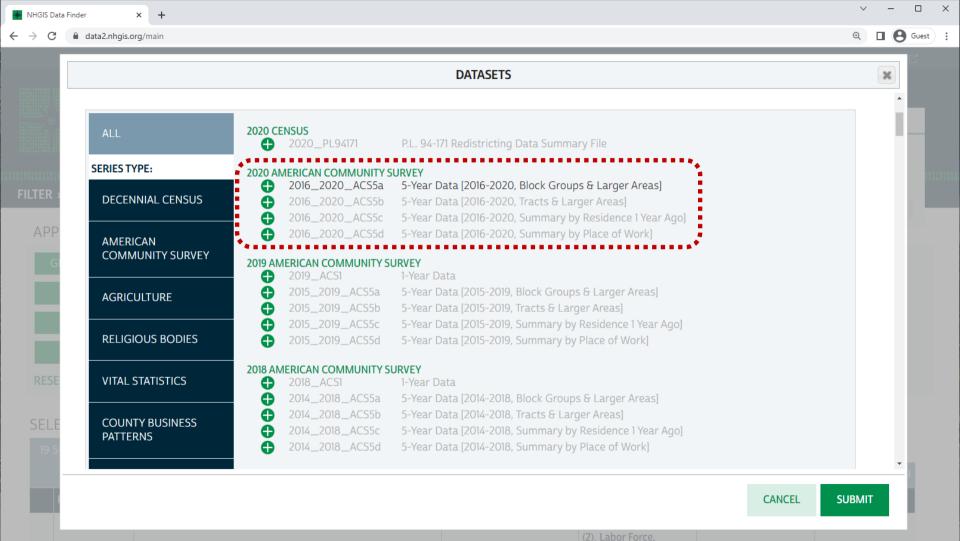


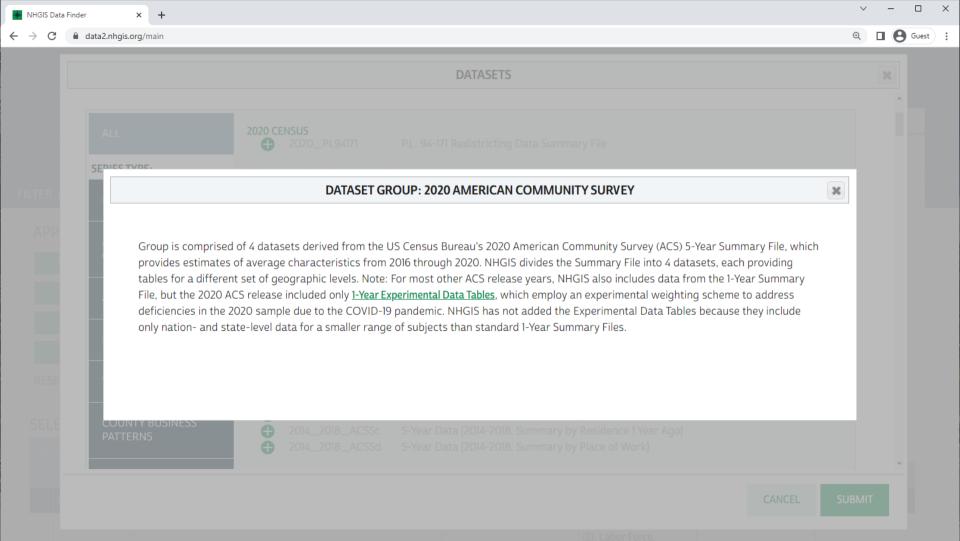


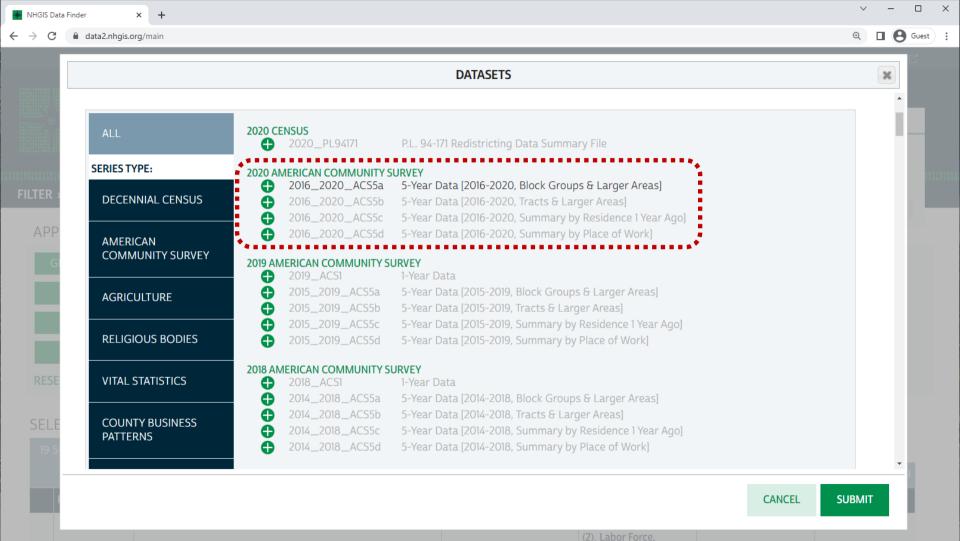


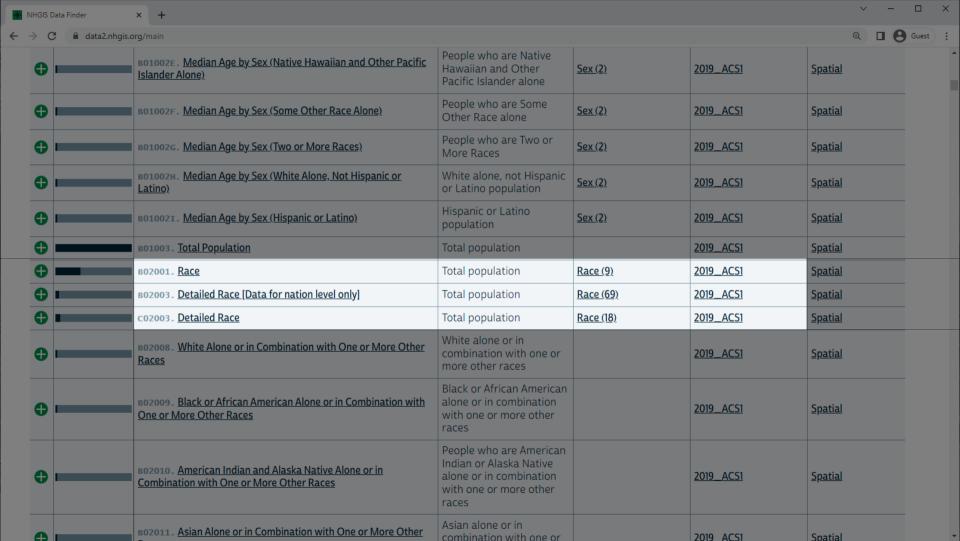


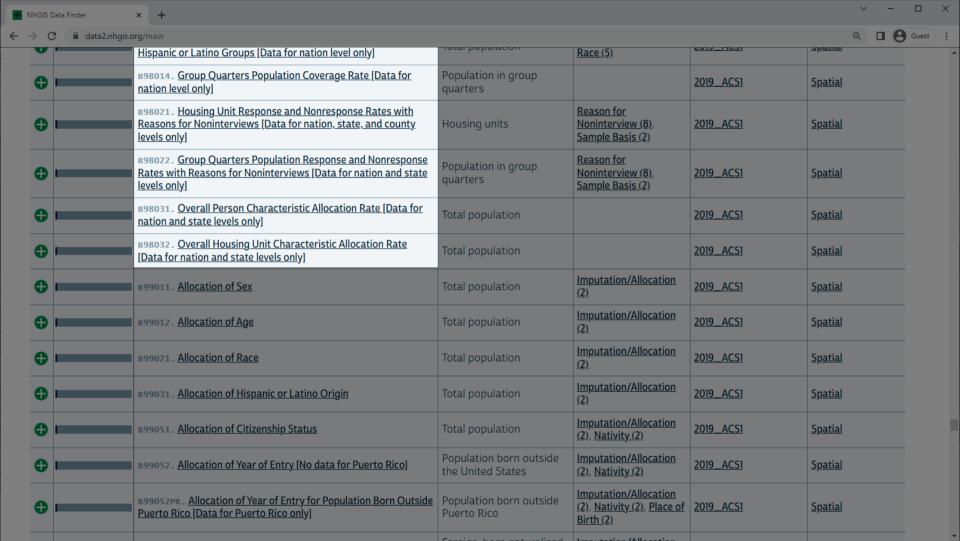


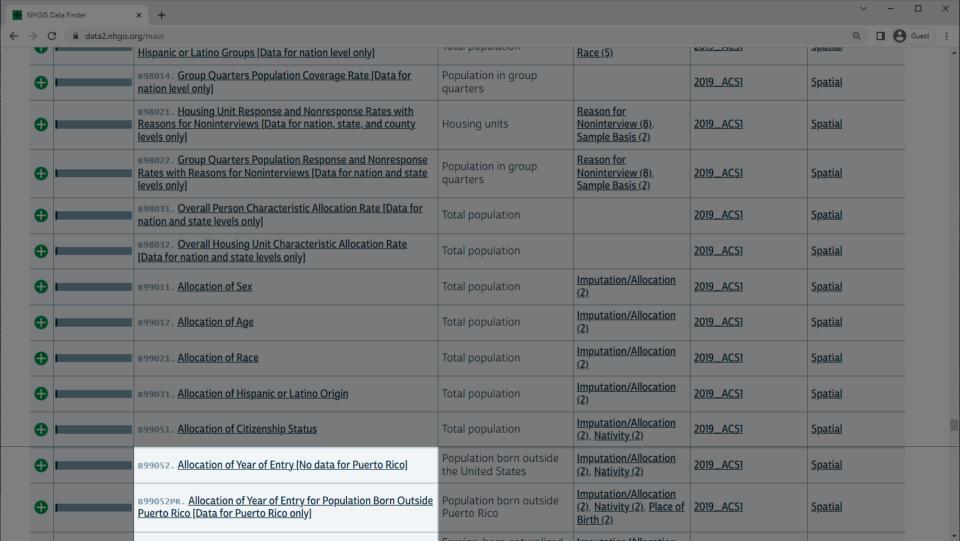




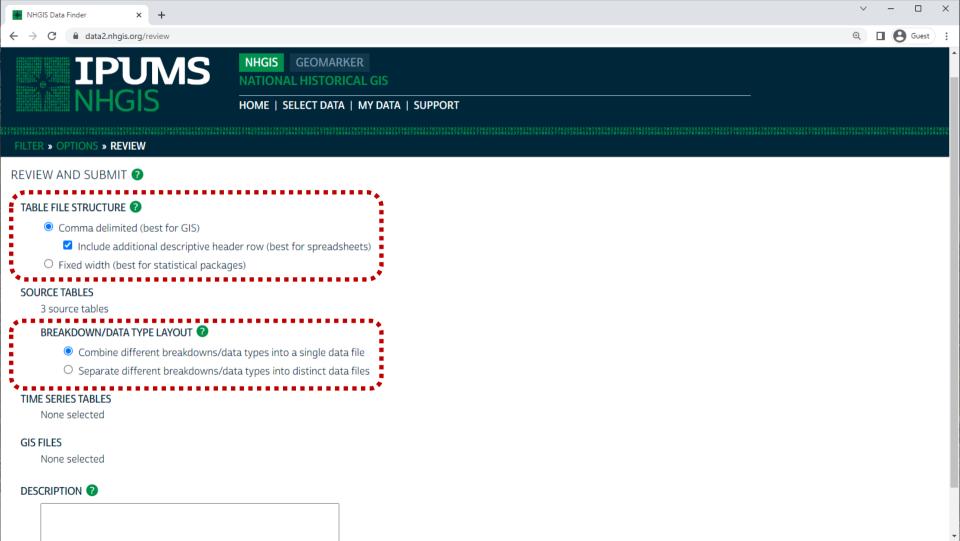












#### **Unique NHGIS needs**

- All summary file tables: data & metadata
- Info on...
  - Summary levels & US/PR coverage by table
  - Complete variable labels & label components
- Ease in separating...
  - Specific summary levels & geographic components
  - Estimates & margins of error



## USING THE NEW SUMMARY FILE FORMAT

#### **ACS ingest team for IPUMS NHGIS**

- IT Data Team
  - Kevin Van Horne
  - Jake Wellington
  - Colin Davis
  - Poornendu Narayanan
- IT director: Fran Fabrizio
- Graduate assistant: J.P. Bruno



#### Using the new summary file format

- A very welcome change!
  - Easier to understand & work with
  - Clarity about...

Columns  $\leftarrow \rightarrow$  Tables

Summary levels ←→ Files

#### **Problematic changes**

- Margins of error...
  - Interleaving of estimate & MOE columns
  - Some tables have no MOE columns
- Preferences for bulk processing...
  - Group the estimate columns separately from MOEs
  - Include MOE columns for all tables, even if blank
    - And/or indicate which tables lack MOE columns



#### **Problematic changes**

- Somewhat harder to determine each table's...
  - Summary levels covered
  - Areas covered (U.S. / Puerto Rico)
- Appendix A issues...
  - Not 100% thorough; designed for human reading
- Ideally...
  - Provide auto-generated machine-readable matrix of levels & US/PR coverage by table



#### **Problematic changes**

- Inconsistent geographic ID fields...
  - GEO\_ID in data file & data.census.gov
    - = DADSID in geography file
    - ≠ GEOID in geography file
    - ≠ GEOID in TIGER/Line shapefile
  - GitHub documentation clarifies but any remaining inconsistencies will cause difficulties & confusion

#### Persistent concerns/opportunities

- Problematic character encoding
  - Summary files: Latin-1 (ISO-8859-1)
  - Common standard: UTF-8
  - API: Latin-1 converted to UTF-8 twice
    - "carro público" → "carro público"

#### Persistent concerns/opportunities

- "Share more" with API
  - From API:
    - Fully concatenated variable names
  - From summary file docs:
    - Complete table metadata (universes)
    - Ordered variable listing
    - Geographic restrictions (& geographic components?)



#### Persistent concerns/opportunities

- Pipe-delimited is good... Why not use it everywhere?
  - Geography File still uses commas
  - Excel docs could also be pipe-delimited files
- Widths of geo codes available only in PDF table
- Drop the "BLANK" columns in Geography File?
- FTP is becoming outdated.
   Consider HTTP / API for file delivery.



#### Acknowledgments

## Eunice Kennedy Shriver National Institute of Child Health & Human Development

[NIH-R01HD057929, NIH-R24HD041023]

**National Science Foundation** 

[SES-1324875]

