# Improving Efficiency in Non-Data-Collection Activities with Data Science Tools

**The Case of Census's Count Question Resolution Operation**

October 26, 2022

Nelson Chung
Survey Statistician
Decennial Census Management Division
Count Question Resolution Branch
nelson.chung@census.gov

Shape
your future
START HERE >

United States®
Census
2020

# Overview

- 2020 Census Count Question Resolution (CQR) Introduction

- Leveraging Pre-existing Data

- Customer Service

- Internal Assessment

- Tracking Workflows

- Challenges

Shape
your future
START HERE >

United States®
Census
2020

# 2020 Census CQR

- The 2020 Census Count Question Resolution (CQR) program provides a mechanism for governmental units to request a review of their official 2020 Census results and to help ensure that housing and population counts are correctly allocated to 2020 census tabulation blocks in the 50 states, District of Columbia, and Puerto Rico.

- CQR may not collect any new data.

- The goal is to improve the quality of the base used for post-2020 Census population estimates.

- Count corrections are issued, when necessary, to the tribal chairperson or highest elected official of affected governmental units.

- CQR will not revise the apportionment data sent to the office of the President or any redistricting data sent to the states.

- CQR Revisions may affect local funding.

Shape
your future
START HERE >

United States®
Census
2020

# Leverage Pre-existing Data

Produce the Following "Crosswalk"

- Created for internal use.

- Block count lists are forms Governmental Units (GUs) use to submit the corrections they request

- Useful for assisting GUs in identifying census blocks within their jurisdiction to create block count lists. Scrape and unzip public Census block assignment data. Merge with internal spreadsheet for Census Regional Office (RO).

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | BLOCKID | PLACEFP | ST | RO | EntityID | ENTITY_NAME | FIPS |

Shape
your future
START HERE >

United States®
Census
2020

# CQR Block Count List

BlockID from Crosswalk

AA<span style="color:blue">BBB</span><span style="color:green">CCCCCC</span><span style="color:red">DDDD</span>

State
<span style="color:blue">County</span>
<span style="color:green">Tract</span>
<span style="color:red">Block</span>

**Complete Block Count List**

Help GUs identify blocks in their jurisdiction to populate their CQR Block Count List with one click.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | ENTITYID | STATEFP | COUNTYFP | TRACT | BLOCK | CENSUSHU | CENSUSGQ | CQRHU | CQRGQ |
| | | AA | BBB | CCCCCC | DDDD | 6 | 0 | 9 | 0 |

CENSUS ENTITYID
(unique for each GU)

Shape
your future
START HERE >

United States®
Census
2020

# Customer Service

## Weekly Case Status Updates on 2020 CQR Website

- Enable GU to see status of their case.
- Status based on presence or absence of dates in production control system (PCS).
- PCS data tables accessible by Oracle TOAD software.
- Steps
  - (1) loading list of cases from TOAD
  - (2) writing data to pre-formatted spreadsheet
  - (3) converting it to PDF
  - (4) uploading to website
- All four steps reduced to one-click.
- Script enables Python communication with TOAD and Excel.

**2020 Census Count Question Resolution Operation (CQR)** — As of 9/12/2022
**Cases Received**

**Case status categories:**
**Received:** We have received the case submission, and it is under review. In-scope cases are moved on to processing.
**Closed Due to Incomplete Documentation:** We have informed the submitting governmental unit that the materials submitted for the case are incomplete. They can resubmit the case by June 30, 2023.
**Withdrawn:** We withdrew the case from review at the request of the governmental unit.
**Completed:** We have shared case results with all impacted governmental units.

**Additional information about specific cases:**
\* Case requests review of seven localities at this time.
\*\* CQR accepts boundary or housing count cases. Since the submission requirements and resolution for each case type are different, governmental units that have submitted each case type are listed here separately.

| Case Received Date | Submitting Governmental Unit | Case Status |
|---|---|---|
| 01/10/2022 | White County, GA | Received |
| 01/13/2022 | Chester, GA | Completed |
| 01/13/2022 | Glennville, GA | Completed |
| 01/25/2022 | Fontana-on-Geneva Lake, WI | Completed |
| 02/03/2022 | Wichita, KS | Received |

Shape your future START HERE >

United States Census 2020

# Internal Analytics

- Management-driven Queries (Ad hoc)
  - General curiosity
  - Public Information Office
  - Operational Statute
  - News/Media Inquiries
  - Requires integrating publicly-available data with internal data
- Assessment Tables (Regular)
  - Post-operational assessment to tell story of operation

Shape
your future
START HERE >

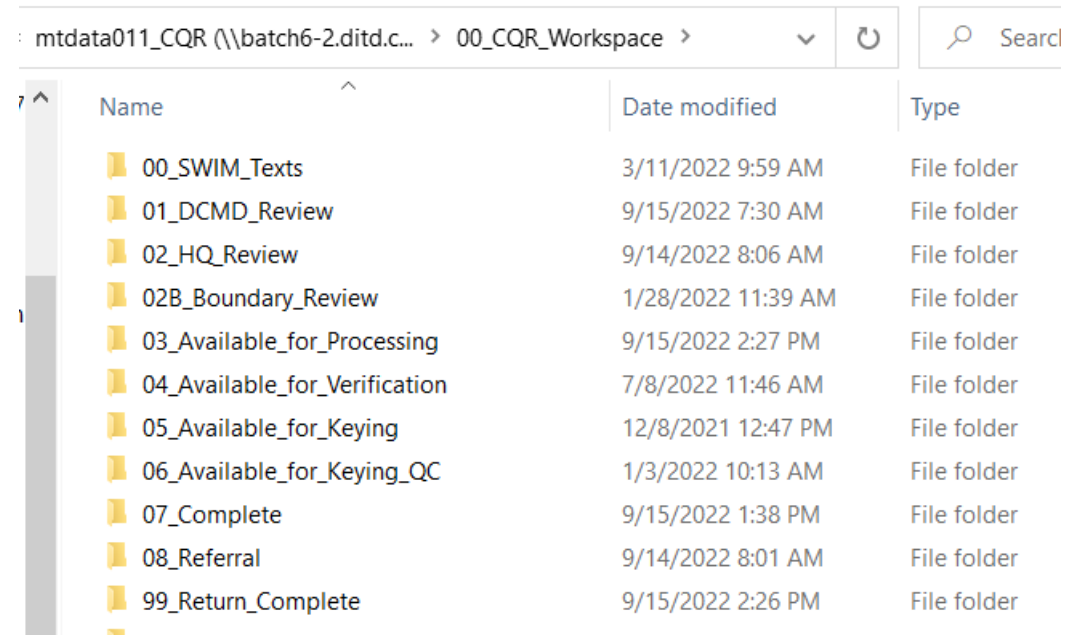United States®
Census
2020

# Assessment Tables

- Info from production system and master file queried TOAD. Field names may not be consistent from different databases.
- Row and Column totals for multiple tables; functions handy.

| Region | GU Type | Boundary | Housing Count |
|---|---|---|---|
| New York | Tribe | | |
| | State | | |
| | County | | |
| | MCD | | |
| | Place | | |
| | Total | | |

Shape
your future
START HERE >

United States®
Census
2020

# Tracking Workflow Script

- No customized software for tracking pre-processing workflow.

- Common to other certain geographic operations that have pre-processing.

- Working directory folder structure. Numbered folders for processing steps, personal folders to "check out" case.



| Name | Date modified | Type |
|------|---------------|------|
| 00_SWIM_Texts | 3/11/2022 9:59 AM | File folder |
| 01_DCMD_Review | 9/15/2022 7:30 AM | File folder |
| 02_HQ_Review | 9/14/2022 8:06 AM | File folder |
| 02B_Boundary_Review | 1/28/2022 11:39 AM | File folder |
| 03_Available_for_Processing | 9/15/2022 2:27 PM | File folder |
| 04_Available_for_Verification | 7/8/2022 11:46 AM | File folder |
| 05_Available_for_Keying | 12/8/2021 12:47 PM | File folder |
| 06_Available_for_Keying_QC | 1/3/2022 10:13 AM | File folder |
| 07_Complete | 9/15/2022 1:38 PM | File folder |
| 08_Referral | 9/14/2022 8:01 AM | File folder |
| 99_Return_Complete | 9/15/2022 2:26 PM | File folder |

mtdata011_CQR (\\batch6-2.ditd.c... > 00_CQR_Workspace >

Shape
your future
START HERE >

United States®
Census
2020

# Tracking Workflow Script

- Cases are recorded on .txt files and sent from one folder to another via cut-and-paste.

- One analyst works on case, leaves notes, moves it to numbered folder.

Shape
your future
START HERE >

United States®
Census
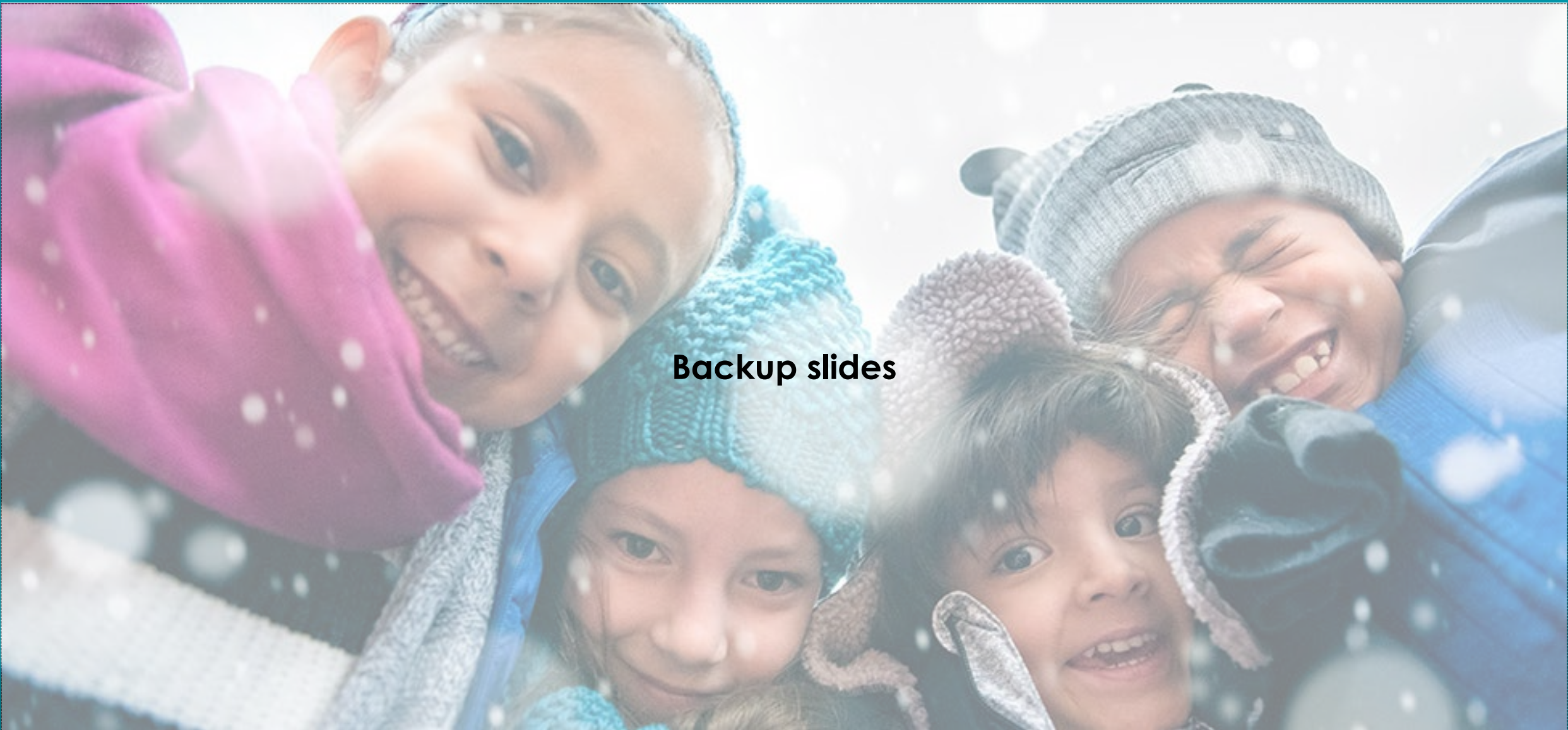2020

# Locating Cases Files

- File names: Case ID + extension.
- Extract Case IDs from Oracle.
- Use Python to determine whether file exist under a certain directory.
- Scan for duplicates.
- Avoid need to look through every folder and subfolder.

Shape
your future
START HERE >

United States®
Census
2020

# Challenges

1. IT limitations.

2. Economies of Scale not always reached.

Shape
your future
START HERE >

United States®
Census
2020

# Questions?

# Thank you!

Shape
your future
START HERE >

United States®
Census
2020

Backup slides

Shape
your future
START HERE >

United States®
Census
2020

# Key Python Packages

- Communicating with Oracle TOAD: cx_Oracle

- Manipulating Excel: openpyxl, xlsxwriter

- Scraping and uploading data to website: selenium

- Scan for file: pathlib

Shape
your future
START HERE >

United States®
Census
2020