MID-OHIO REGIONAL
MORPC
PLANNING COMMISSION

111 Liberty Street, Suite 100
Columbus, Ohio 43215
morpc.org

T. 614. 228.2663
TTY. 1.800.750.0750
info@morpc.org

NOTICE OF A MEETING
REGIONAL INFORMATION & DATA GROUP
MID-OHIO REGIONAL PLANNING COMMISSION
111 LIBERTY STREET, SUITE 100, COLUMBUS, OHIO 43215
SCIOTO CONFERENCE ROOM

January 29, 2020, 3:00 pm – 5:00 pm

AGENDA

1. **Welcome & Introductions**
2. **Data and Mapping Gallery for Census 2020 – *Shoreh Elhami, City of Columbus***
3. **Small Group Breakout Session**
4. **Break**
5. **Census Simplified – *Natalie Hurst, MORPC***
6. **Small Group Data Questions Session**
7. **Closing Remarks**
8. **Adjourn**

Please notify Lynn Kaufman at 614-233-4189 or LKaufman@morpc.org to confirm your attendance for this meeting or if you require special assistance.

The Next Meeting of the Regional Information & Data Group will be in April 2020.
111 Liberty Street, Suite 100, Columbus, Ohio 43215

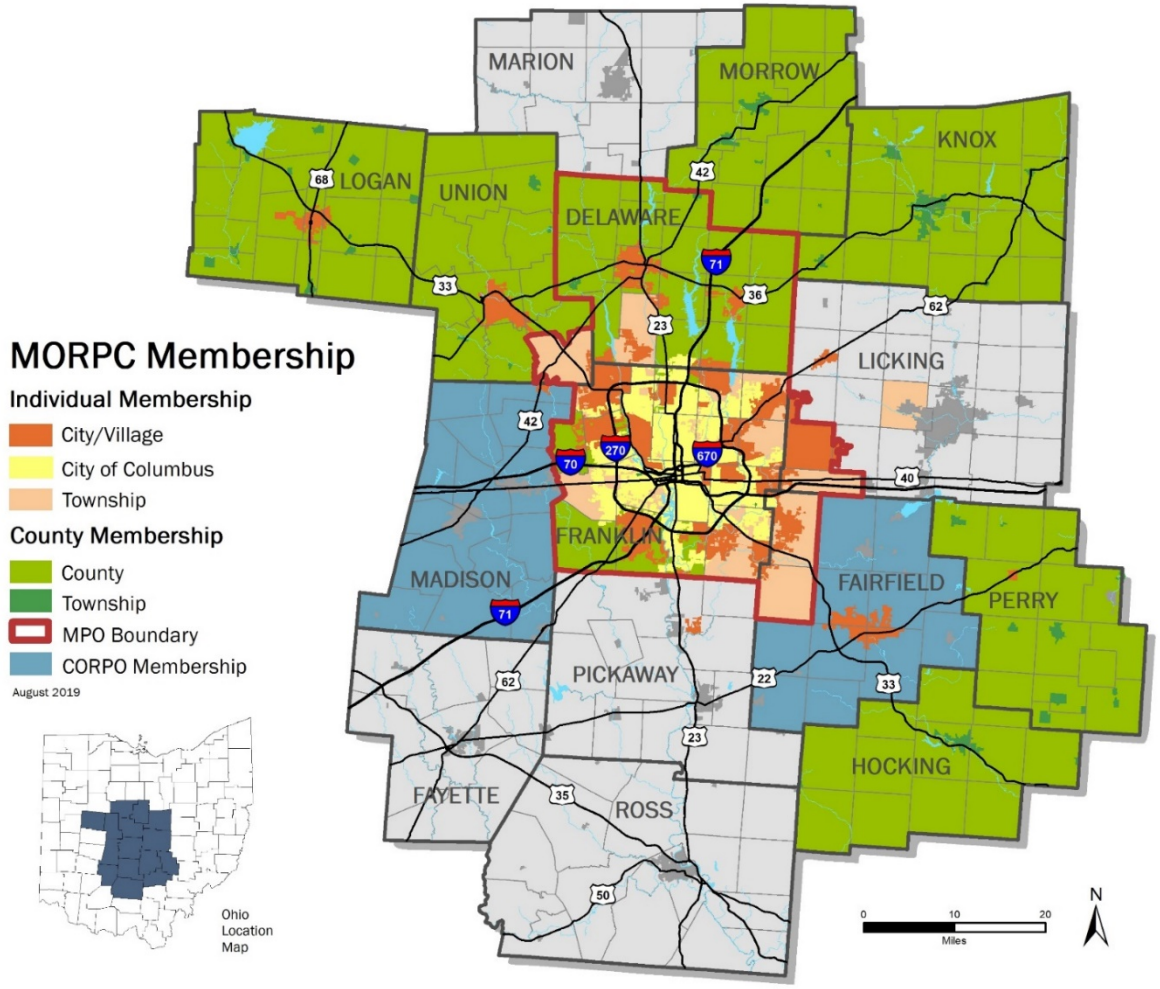
PARKING AND TRANSIT: When parking in MORPC's parking lot, please park in a "MORPC Visitor" space or in a space marked with a yellow "M". Handicapped parking is available at the side of MORPC's building. Indoor bike parking is available for MORPC guests. MORPC is accessible by CBUS. The nearest bus stop is S. Front Street & W. Blenkner St. Buses that accommodate this stop are the Number 61 - Grove City, the Number 5 - West 5th Ave. /Refugee, and the Number 8 - Karl/S. High/Parsons.

Regional Information & Data Group

January 29, 2020

MID-OHIO REGIONAL
MORPC
PLANNING COMMISSION

MORPC Membership & Programming



- Transportation & Infrastructure Development
- Planning & Sustainability
- Data & Mapping
- Residential Services
- Government Affairs
- Communications & Engagement

Regional Data Advisory Committee

"The Regional Data Advisory Committee is a standing committee to **evaluate and advise** the COMMISSION on matters of regional importance pertaining to data analysis. It will **set standards** to enable data interoperability and data credibility. It will **explore public/private partnerships** and other collaboration opportunities around using data."

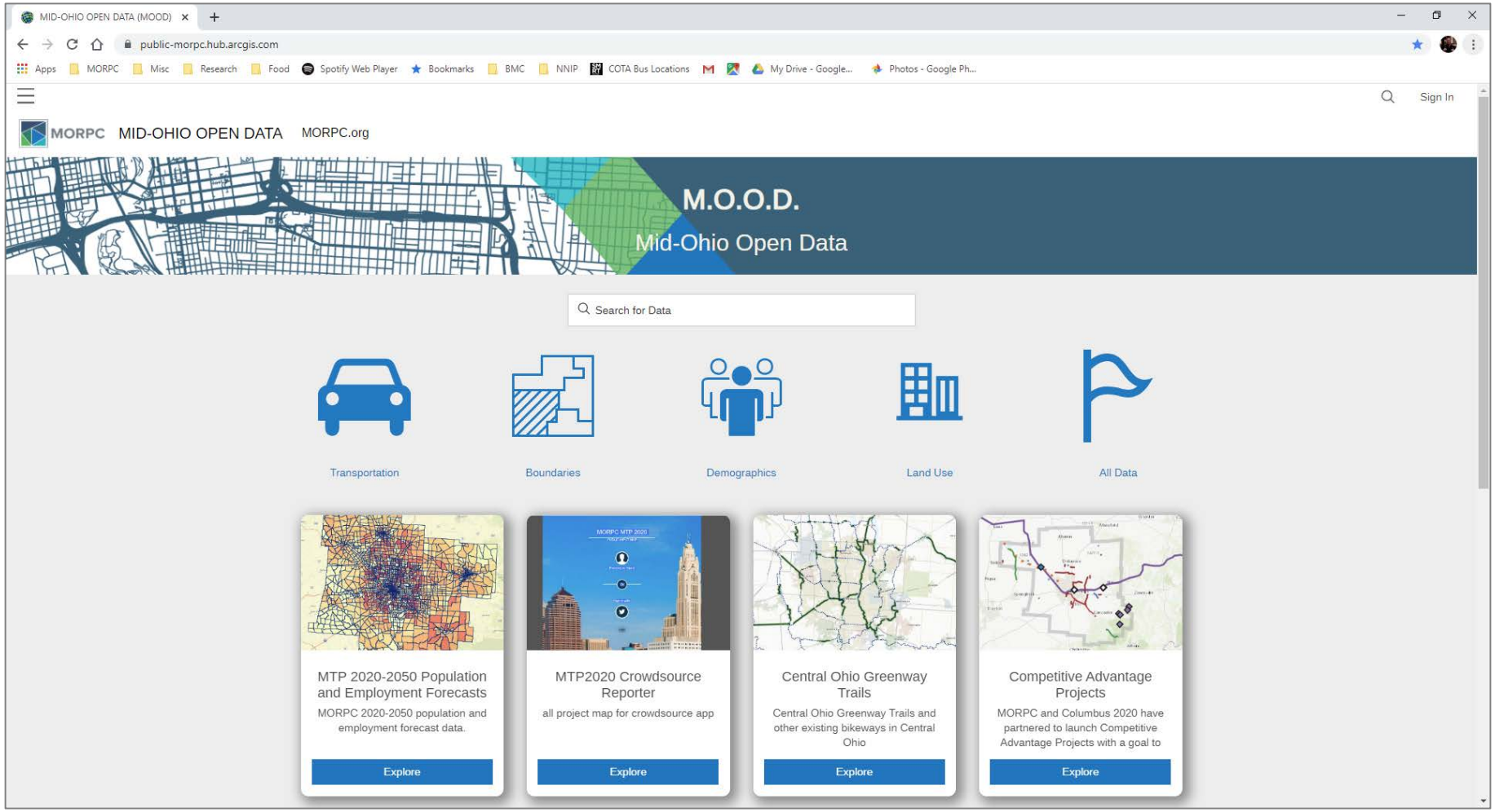
WORKING GROUPS

- **Regional Data Users Group**
- **Regional Sustainability Dashboard**
- **Regional Municipal Fiber Strategy**
- **Data Policy Needs Survey & Toolkit**

Regional Data Agenda

- **Foster Collaboration**
- **Capacity Building & Education**
- **Data Governance & Best Practices**
- **Data Procurement & Development.**
- **Data Access, Inclusivity, & Equity**

Mid-Ohio Open Data (M.O.O.D.)



Data, Map & Technology Subcommittee

City of Columbus & Franklin County
Complete Count Committee (CCC)
Census 2020

January 29, 2020

THE CITY OF
COLUMBUS
ANDREW J. GINTHER, MAYOR

Columbus Counts!
CENSUS
2020

THE CITY OF COLUMBUS ANDREW J. GINTHER, MAYOR

Franklin County BOARD OF COMMISSIONERS

Overview

- Why a Data, Map & Technology Subcommittee?
- Who Are the Committee Members?
- What We Have Done So Far?

Data, Map & Technology (DMT) Subcommittee

- The DMT subcommittee was formed to serve the data and mapping needs of other subcommittees of 2020 Complete Count Committee (CCC); 30 in total

List of other CCC Subcommittees

Business and Workforce
Communications and Media
Youth Serving Organizations
Children Ages 0-5
Mental Health Services Organizations
Human Services Organizations
Data, Maps and Technology
Education (Preschool-Grade 12)
Education (Post-Secondary)
Justice-Involved
Faith-Based
Government
Group Quarters
Adult Homelessness
Youth Homelessness
Immigrant
Refugee and Asylum Seekers
African Immigrants
Latino/Hispanic
Asian-American
African-American
LGBTQ
Senior Citizens
Non-Homeowners
Veterans
Philanthropic
Individuals with Disabilities
Organized Labor
Transportation
Elected Officials

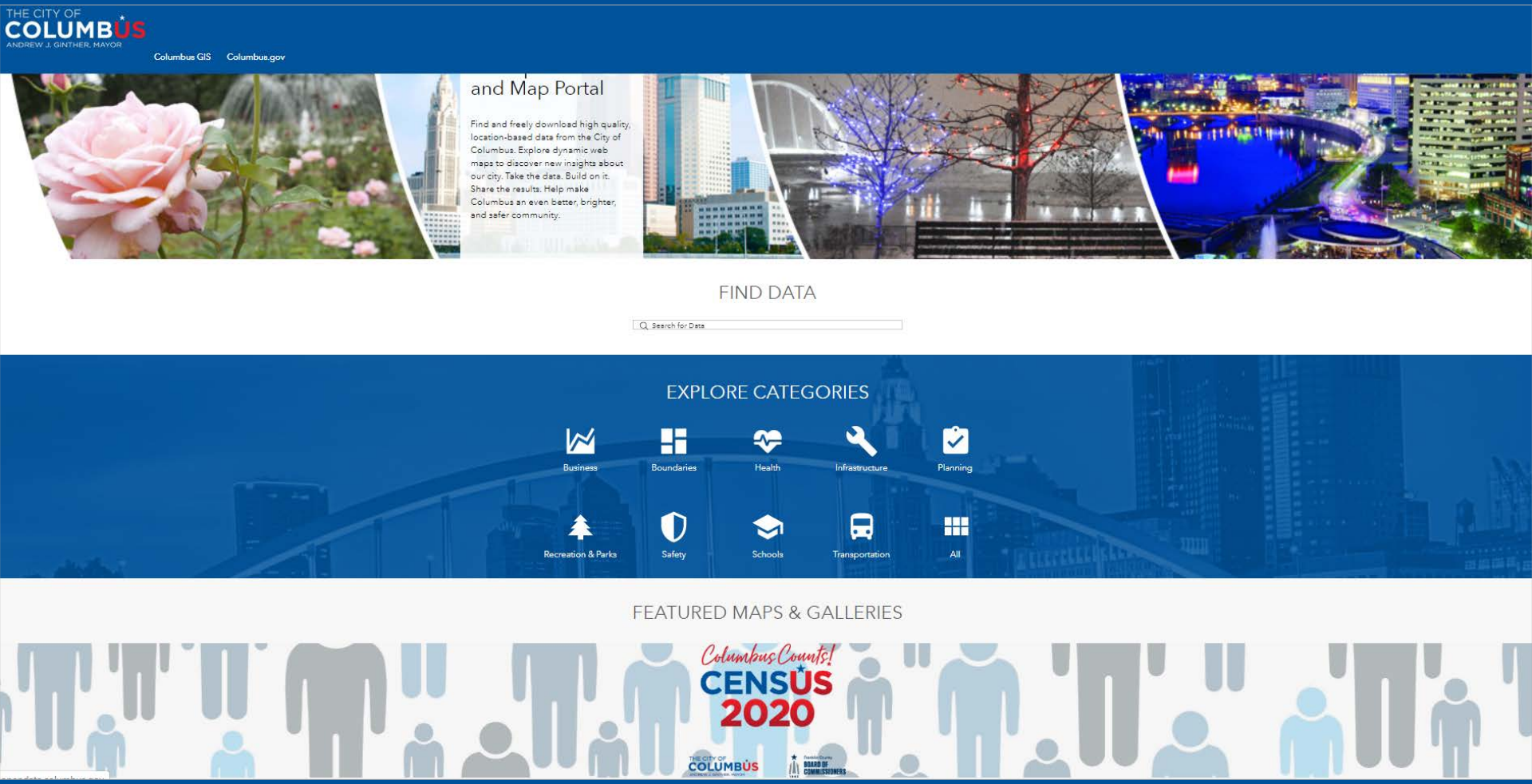
Who Are the Committee Members?

- Columbus
- Franklin County Auditor’s Office
- Morpc
- Hilliard
- Worthington
- Gahanna
- Westerville
- Upper Arlington
- Minerva Park

What We Have Done So Far - Strategy

- Established the DMT Subcommittee – June 2019
- Established a Point of Contact (POC) at the Census Bureau for technical support
- Established contacts (POC) with other 29 subcommittees and sought their data and analytics needs via an online survey – Oct 2019
- Built and added artifacts to the a Data/Map Gallery – December 2019

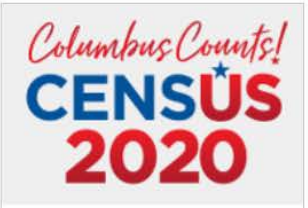
Columbus Open Data Portal:



City of Columbus Census 2020 Map Gallery

African American Employment ACS 2013-2017

African American Occupation ACS 2013-2017

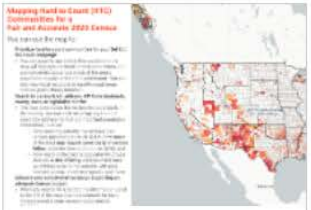


Census 2020 Demographic Resources

Community Resources

Crime Index 2019

Group Quarters in Central Ohio



Hard To Count 2020

Homeless Shelters

Income and Diversity

Internet Connectivity ACS 2013-2017

Language Spoken At Home ACS 2013-2017

People with Disabilities ACS 2013-2017

Response Outreach Area Mapper (ROAM)

USA Tapestry Segmentation

Veterans ACS 2013-2017

City of Columbus Census 2020 Map Gallery

Census 2020 Demographic Resources (a multi-tabbed application)
- Language Spoken At Home
- Veterans
- People with Disabilities
- African American Employment
- African American Occupation
- Internet Connectivity
- Community Resources

Group Quarters in Central Ohio (Morpc)

Hard To Count 2020

Homeless Shelters (Franklin County Auditor's office)

Income and Diversity

Response Outreach Area Mapper (ROAM)

USA Tapestry Segmentation

Crime Index 2019

Note: Data used is ACS 2013-2017




THE CITY OF
COLUMBUS
ANDREW J. GINTHER, MAYOR

Questions

Small group breakout session I

- Introductions:
  - **Name, role, & place of work**

- Small group questions (*please write answers on the large sheet of paper at your tables*):
  - **What project are you spending most of your time on right now?**
  - **What is your favorite dataset?**

Individual Question

- *Please write your answer on one of the notecards at your table before the break:*
  - **What do you most hope to gain from attending this group?**

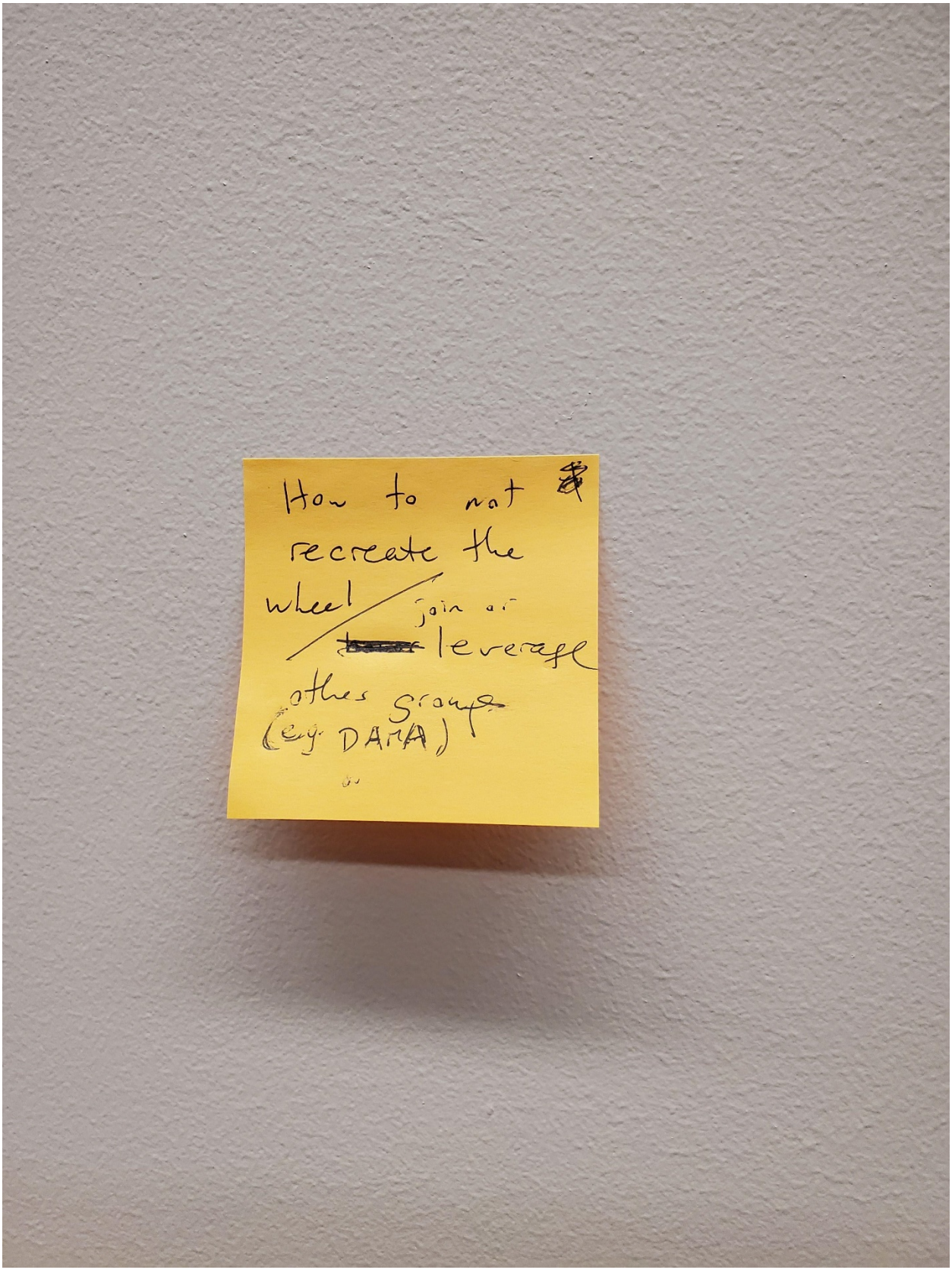
Small group breakout session II

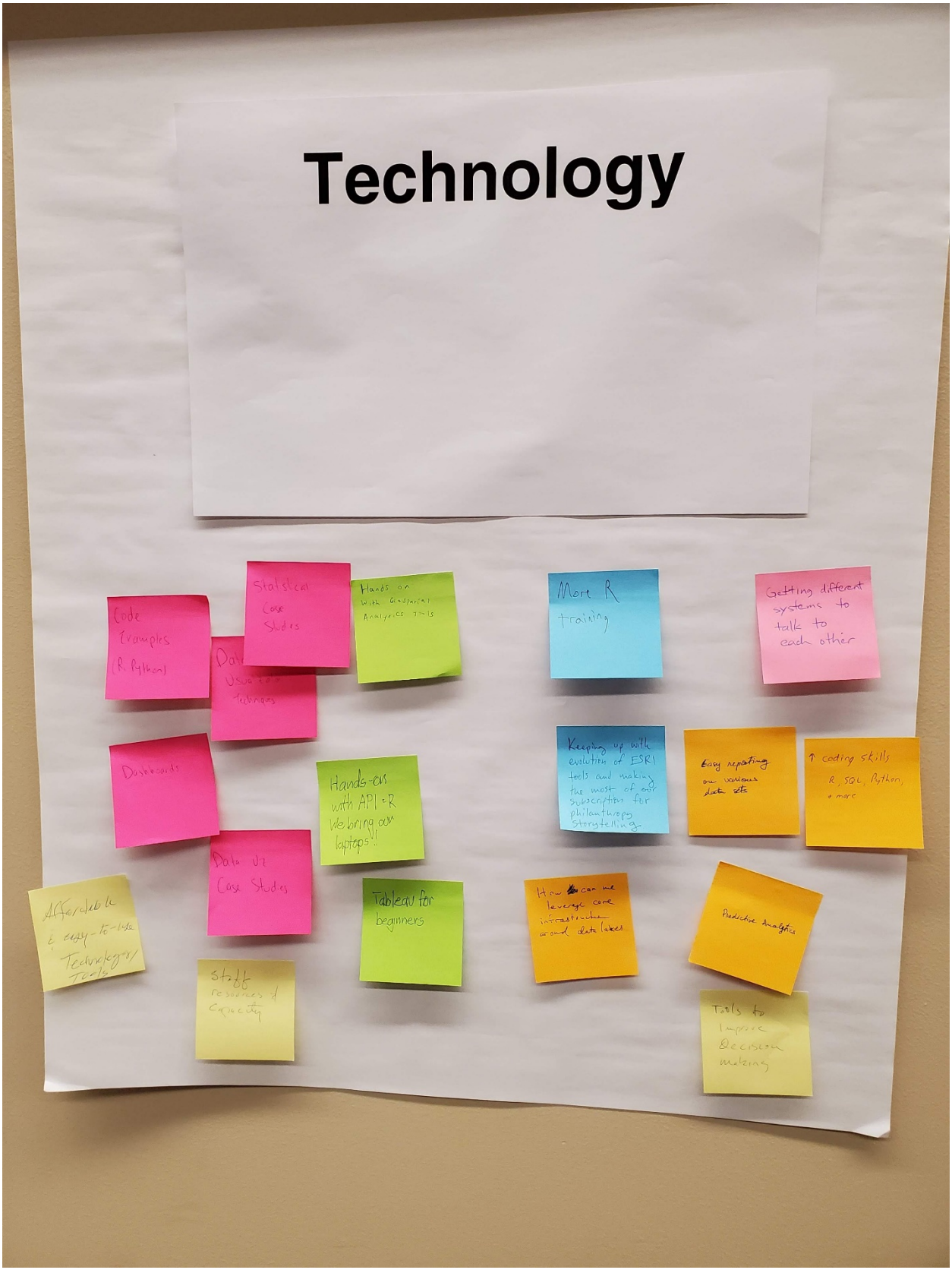
- Write your ideas on sticky notes:

  - **What are your biggest challenges working with data?**

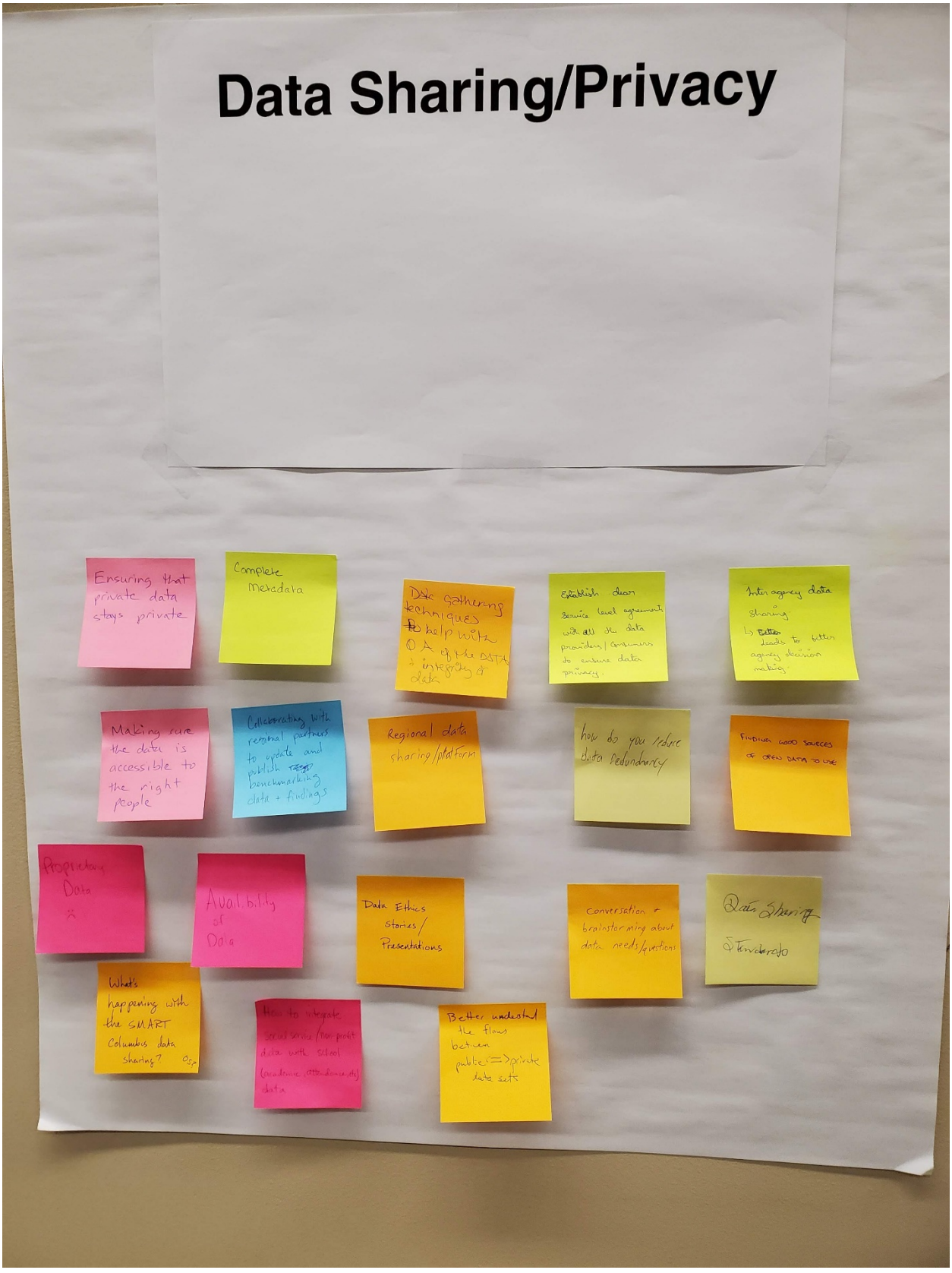
  - **What tools, skills, or other resources would make your job easier or your work more effective?**


Data Governance
Establishing a dedicated data management practice (not a hobby) to create a "single version of the truth" and address quality issues
Setting up an enterprise Data Governance practice
Establishing Master Data Mgmt rules, hierarchies, matching, golden record, practitioners etc to serve the whole company
How to uncover use cases where data and analytics can help, among depts who have not thought about DnA
How do you monitor + measure data quality?
Buy-in for data management policy (org wide)
MOST DATA IS HORRIBLE AND REQUIRES EXTENSIVE CLEANING
How to govern public vs. private data
Best Practices for Improving Organizational Data Literacy?
Leveraging best/emergent practices for civic data
Thinking through data architecture for efficiency
How do maintain datasets in an Enterprise geodatabase vs. web data
Metadata standards
How do you sell the benefits of Data Analysis to decision makers?
master data management/ organization strategies + recommendations
Lack of standardization
Data Quality & Improvement

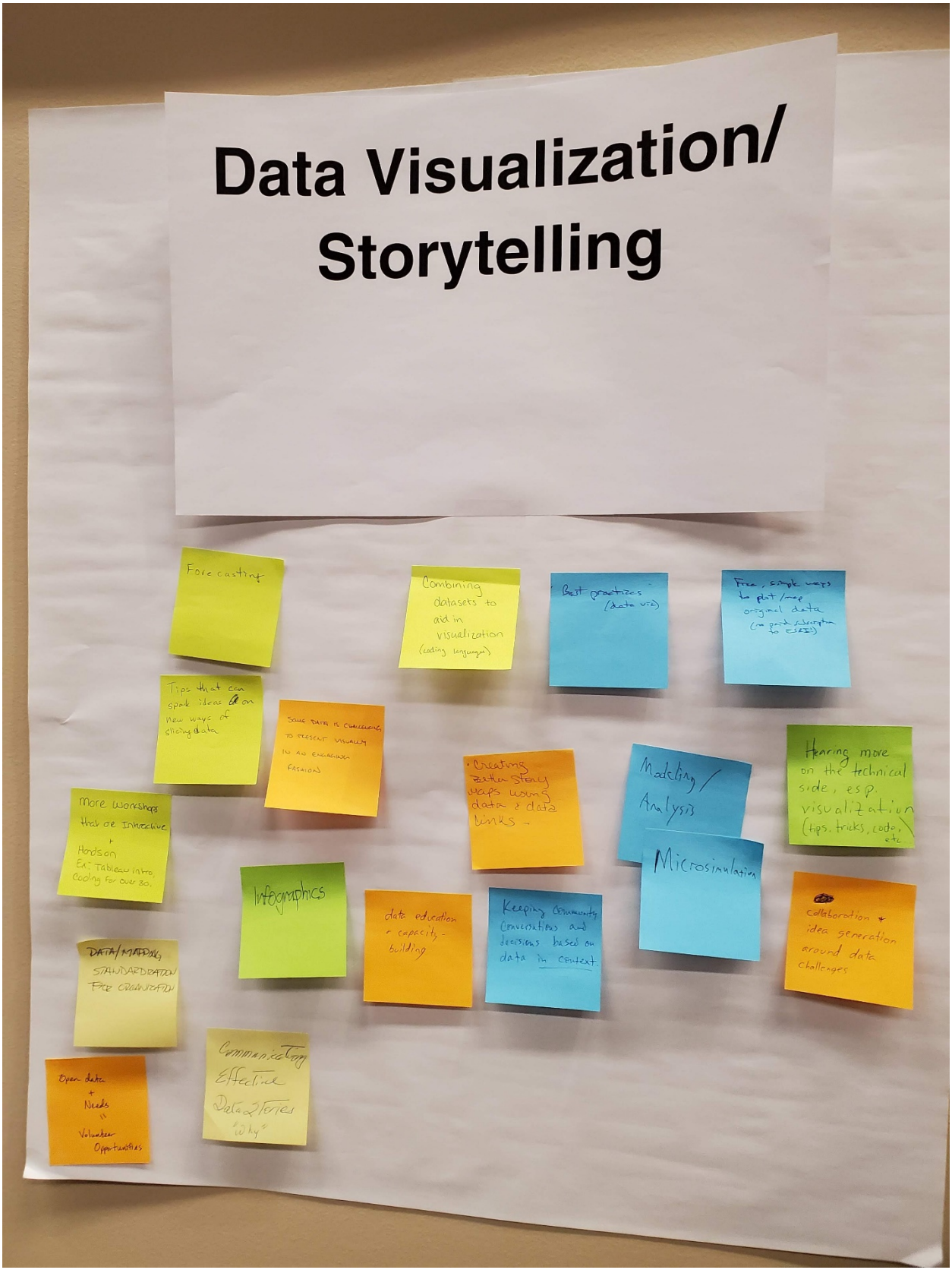

How to not recreate the wheel
join or
leverage
other groups
(eg. DAMA)


Technology
Code Examples (R, Python)
Statistical Case Studies
Hands on with Geospatial Analytics Tools
Dashboards
Hands-on with API + R
We bring our laptops!
Data Viz Case Studies
Tableau for beginners
Staff resources & capacity
Mort R training
Keeping up with evolution of ESRI tools and making the most of our subscription for philanthropy storytelling
How can we leverage core infrastructure around data lakes
Getting different systems to talk to each other
Easy reporting on various data sets
↑ coding skills R, SQL, Python, + more
Predictive Analytics
Tools to improve decision making

Data Sharing/Privacy




Data Sources
Consistency in data collection practices
Data standardization
Establish standardized data entries. This ensures clean & quality data.
Keeping track of data sources
Tutorials on data extraction from variety of sources
- help on understanding margin of error, etc.
- CTPP workshop is great example (Ed Christopher is Penelope)
Best practices for keeping data fresh
Availability of Data
Minimize open text fields. Be aware by PII detection in open text field is a challenge
INCENTIVES DATA
ACTIVITY CENTERS
CDCs
EZs
JEDDs
JEDDs
Move on local or regional data sources
Data quality
LOCAL UTILITY DATA
WATER/SEWER
FIRE
EMS
Housing Data Related to School Districts
When do you ignore the margin of error
SERVICES - VS - STAND ALONE DATASETS
Inconsistent data points across programs
Inability to determine Data lineage
Connect to data resources
- sources of data
- training
Benchmarking
- the numbers
- the sources
- data quality
Develop an inventory of data gaps & solve


Data Visualization/ Storytelling
Fore casting
Combining datasets to aid in visualization (coding language)
Best practices (data viz)
Free, simple ways to plot/map original data (no paid subscription to ESRI)
Tips that can spark ideas on new ways of slicing data
Some data is challenging to present visually in an engaging fashion
Creating better story maps using data & data links
Modeling/ Analysis
Hearing more on the technical side, esp. visualization (tips, tricks, code, etc)
More workshops that are interactive + Hands on Ex: Tableau intro. Coding for over 30.
Microsimulation
Infographics
data education - capacity - building
Keeping Community Conversations and decisions based on data in context
Collaboration + idea generation around data challenges
DATA/MAPPING STANDARDIZATION FOR ORGANIZATION
Open data + Needs = Volunteer Opportunities
Communicating Effective Data 2 Stories "Why"

CENSUS SIMPLIFIED

Natalie Sioux Hurst, MORPC

MID-OHIO REGIONAL
MORPC
PLANNING COMMISSION

PRESENTATION OVERVIEW

- Potential uses of Census data

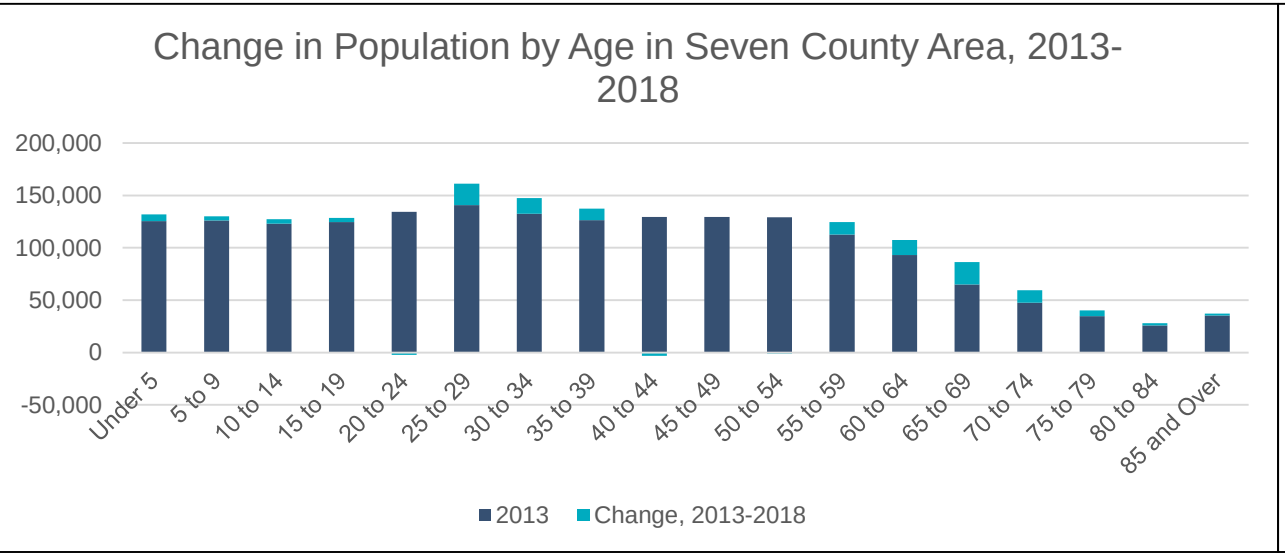
- Ways to acquire Census data (at varying technical skill levels) and resource guides
  - Data.Census.Gov
    - In web browser
  - Census API
    - In web browser and R
  - Census FTP
    - In web browser and R

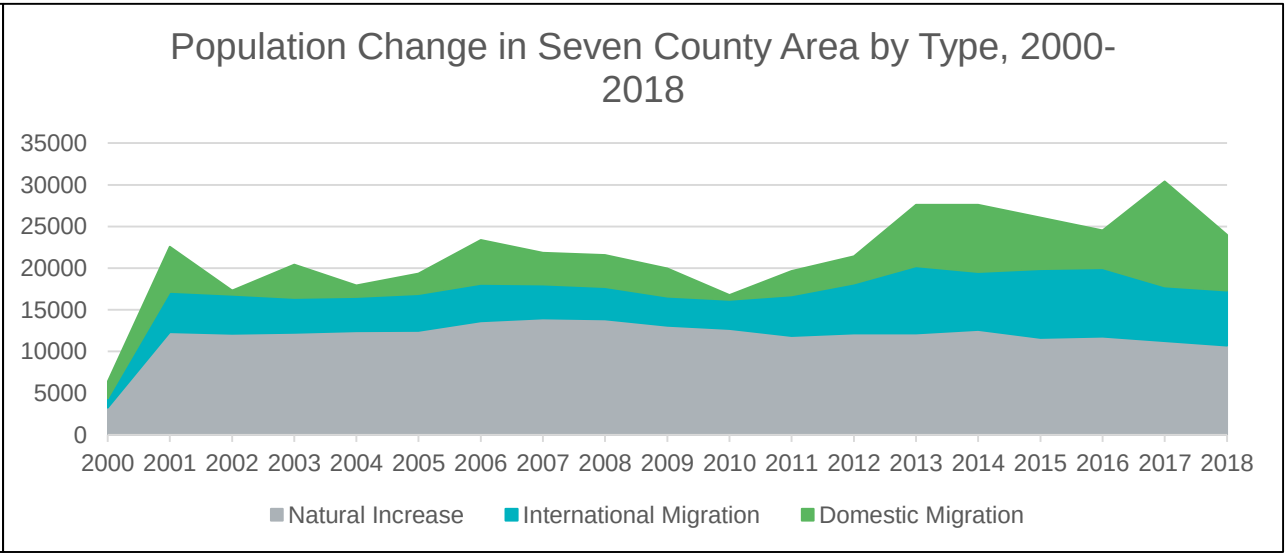
MORPC PROJECTS WITH CENSUS DATA

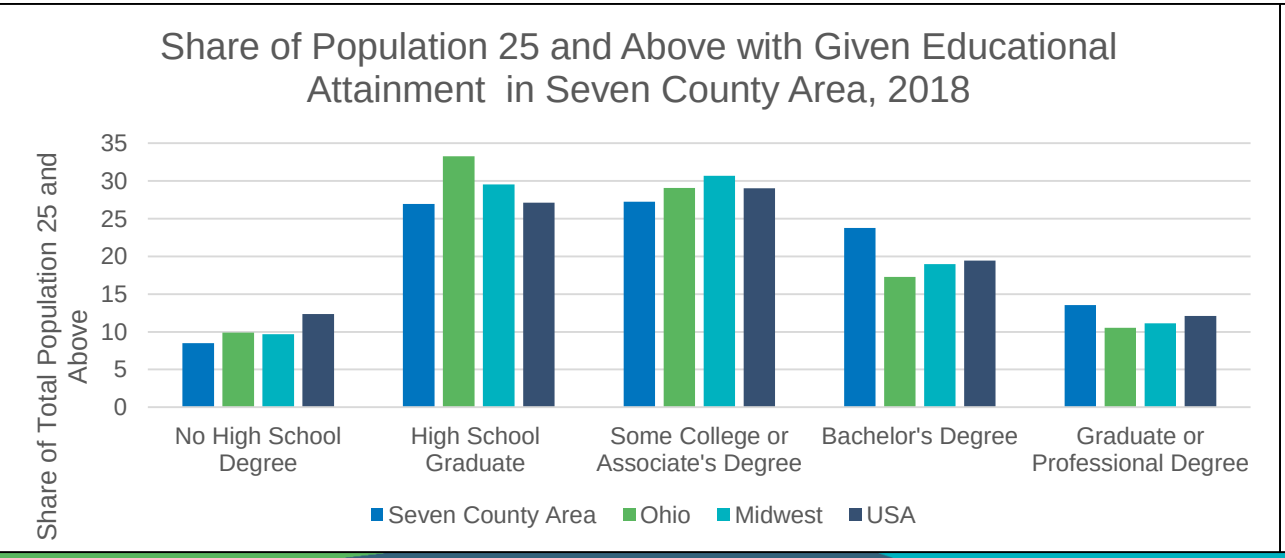
- Almost every MORPC project uses Census data
  - Metropolitan Transportation Plan
  - Income and Population Forecasting
  - Regional Housing Strategy
  - Member Requests

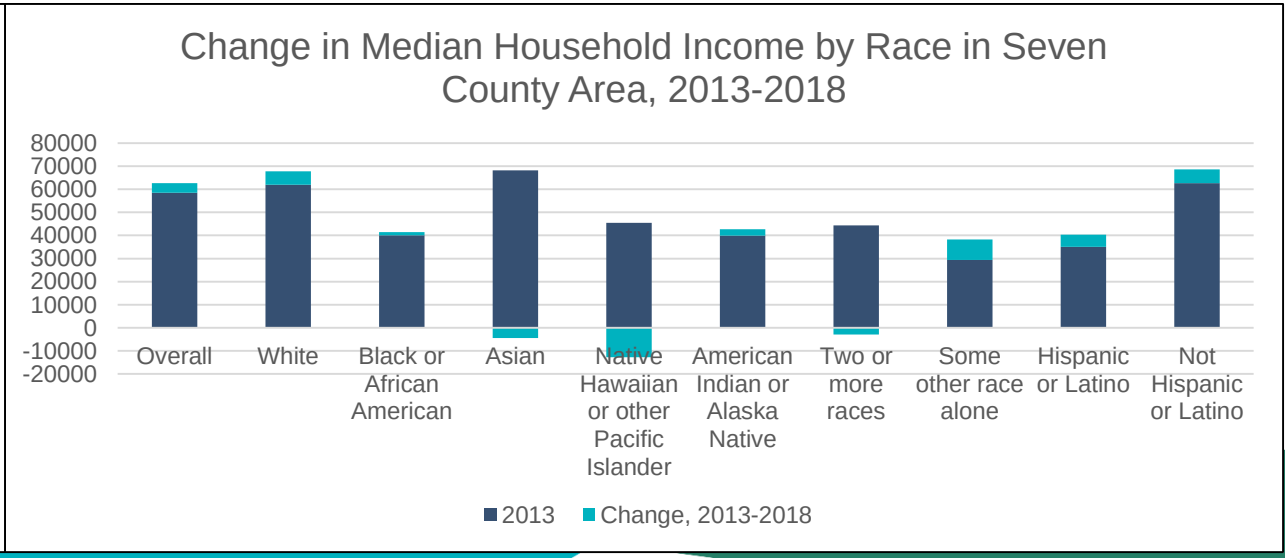
VISUALIZATIONS CREATED WITH CENSUS DATA

MORPC









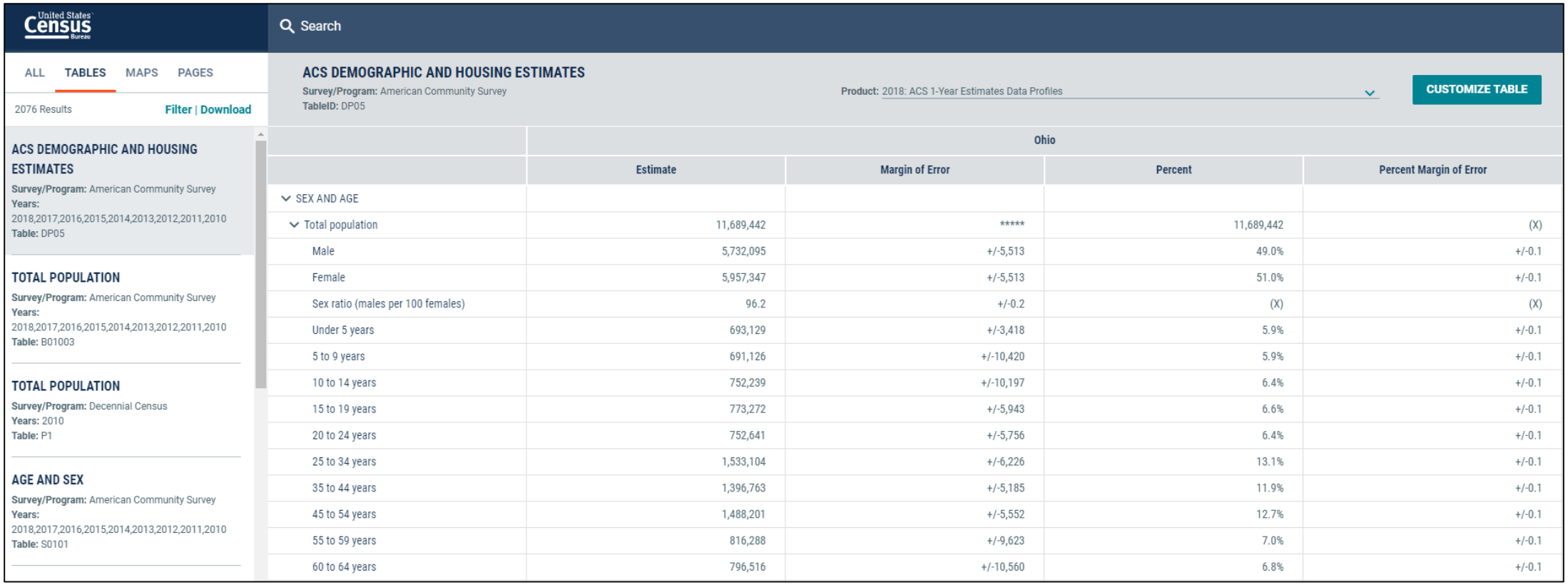
HOW DO WE ACCESS CENSUS DATA?

- Fact Finder (?)
  - After all data is migrated to new Census data site, FactFinder will be no more

RIP FACT FINDER

NEW CENSUS DATA WEBSITE: Data.Census.Gov

- Data.Census.Gov



	Ohio			
	Estimate	Margin of Error	Percent	Percent Margin of Error
SEX AND AGE				
Total population	11,689,442	*****	11,689,442	(X)
Male	5,732,095	+/-5,513	49.0%	+/-0.1
Female	5,957,347	+/-5,513	51.0%	+/-0.1
Sex ratio (males per 100 females)	96.2	+/-0.2	(X)	(X)
Under 5 years	693,129	+/-3,418	5.9%	+/-0.1
5 to 9 years	691,126	+/-10,420	5.9%	+/-0.1
10 to 14 years	752,239	+/-10,197	6.4%	+/-0.1
15 to 19 years	773,272	+/-5,943	6.6%	+/-0.1
20 to 24 years	752,641	+/-5,756	6.4%	+/-0.1
25 to 34 years	1,533,104	+/-6,226	13.1%	+/-0.1
35 to 44 years	1,396,763	+/-5,185	11.9%	+/-0.1
45 to 54 years	1,488,201	+/-5,552	12.7%	+/-0.1
55 to 59 years	816,288	+/-9,623	7.0%	+/-0.1
60 to 64 years	796,516	+/-10,560	6.8%	+/-0.1

NEW CENSUS DATA WEBSITE: Data.Census.Gov

MORPC

- Changes
  - Acquired additional hardware to increase capacity
  - Streamlined table customization / search layout
  - QuickFacts
  - Tables can be copied and pasted
  - Right click to export as csv or xlsx
  - Website has mapping capabilities

- Educational Materials
  - How to materials for website on homepage
  - Webinar: https://www.census.gov/data/academy/webinars/2019/new-updates-data-census-site.html

GETTING CENSUS DATA VIA API

MORPC

- What is an API???
  - “Application Programming Interface”
    - Allows you to communicate with features or data of an operating system, application, or other service
      - A secret passage of sorts
  - Rather than downloading data from a page, you can query it directly
- Advantages
  - Get raw data directly from source
  - Customizable queries make large data tasks easier
  - More reproducible
  - If you use a coding system, no importing or exporting / saving files locally
- Disadvantages
  - Learning curve

AVAILABLE APIs

- Over 100 available Census APIs
  - American Community Survey (1-Year, 3-Year, 5-Year, Migration Flows)
  - Economic Indicators Time Series
  - Decennial Census (SF1/SF3) – 1990, 2000, 2010
  - Economic Census – 2002, 2007, 2012
  - County Business Patterns – 1986-2016
  - International Trade – annual 2005-present, monthly 2013-present
  - And more (!)
    - Page with all:
    - API Base URL's and Variables:

USING THE API IN A WEB BROWSER

- API queries can be ran in a web browser
- Query Components:
- Available Data:
- Steps (see screen sharing video on next slide)
  - Run query in web browser
  - Save page as txt file
  - Open txt file in Excel to convert to spreadsheet format
- Text to Excel Tutorial:

USING THE API IN A R

- While the Census API works in a web browser, it also works in many statistical programming languages (in my case, R)
  - R queries allows for better replicability
  - Data manipulation is easier when R
  - Since API can be read directly into R, no wasted storage space

OBTAINING AN API KEY

- Step 1: Request API Key
  - http://api.census.gov/data/key_signup.html

Hello!

Thank you for your interest in the Census Data API. Your API key is [string of letters and numbers]. Please click here to activate your key.

Save this email for future reference.

Have Fun,

The Census Bureau API Team

~Save this e-mail~

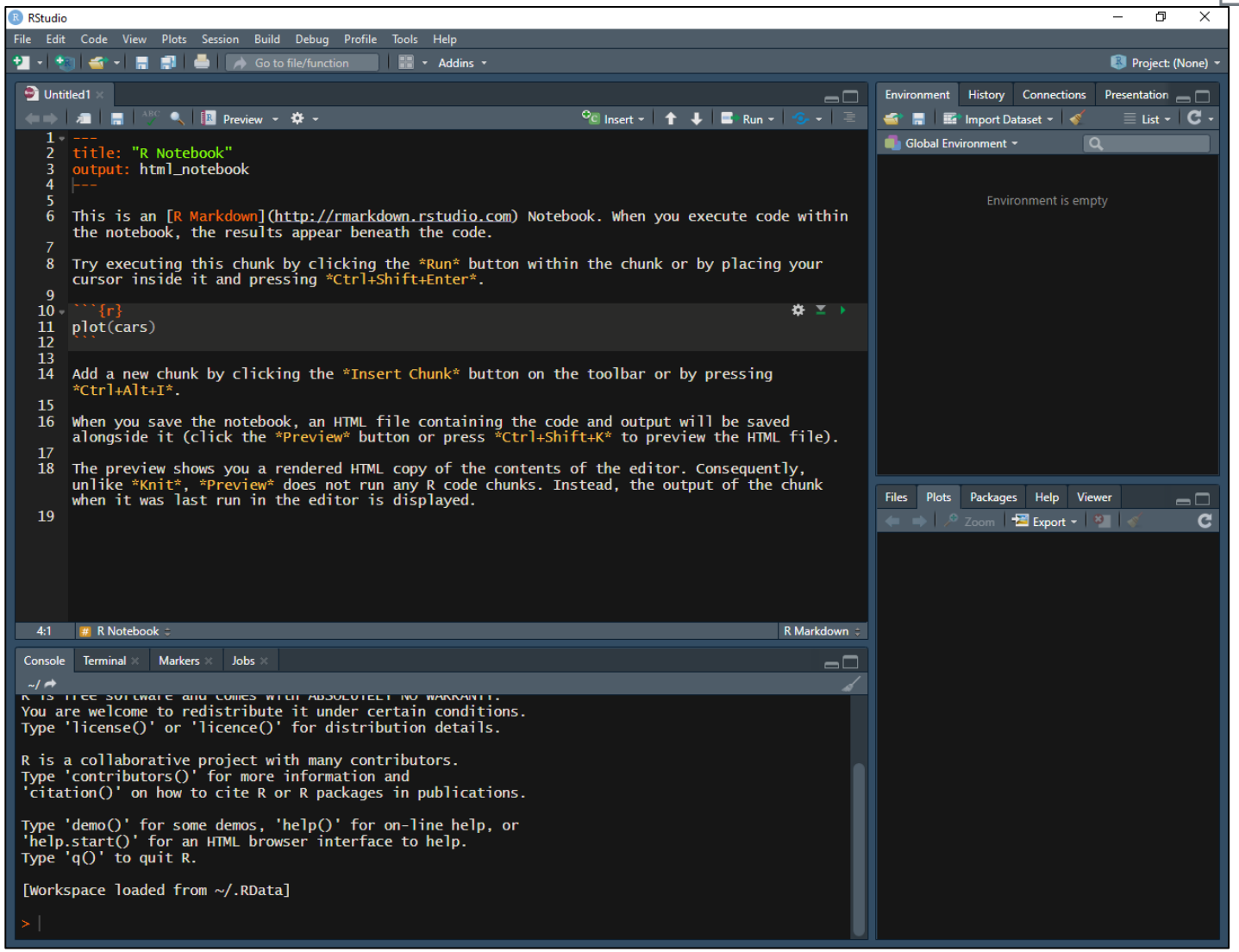
DOWNLOADING R / R STUDIO

- Step 2: Download R and R Studio
  -
    - Use USA Cran Mirror
  - R Studio:

R INTERFACE ANATOMY

MORPC

Notebook: This is where you write your script. You can save it and keep a record of you analysis.



Environment: Any data/objects you import and create will be stored here

Command Line/Console: You will see your code here as well as any results.

Files, plots, packages, help, and viewer: Place to see file these objects

Getting started with R / R Studio Tutorial: https://ourcodingclub.github.io/2016/11/13/intro-to-r.html

R DEFINITIONS

- R Lingo
  - Package: Collection of R functions, compiled code, and sample data
    - Metaphor: A toolbox
  - Function: R tool that completes specific task
    - A tool within the toolbox
  - Arguments: Components of function
    - Components the tool needs
      - Ex: A power drill needs a certain drill bit and a power cord

- More fun R language guides
  - R Core Team:

DOWNLOADING R PACKAGES

MORPC

- There are multiple R packages that can be used to pull Census data
  - “censusapi”
  - “tidycensus”
- To run a function, you must install and load the package that contains that function
  - Two options

```
install.packages("pacman")
library(pacman)

#Install and load Census R packages
pacman::p_load(censusapi, tidycensus, tidyverse, tidyr, dplyr)
```

R FUNCTION USAGE

MORPC

- Within these packages, there are a few different functions
  - getCensus() – Retrieve Census data from any available API
  - get_acs() – Retrieve data and feature geometry for five-year ACS
  - get_decennial() – Obtain data and feature geometry for decennial Census
  - get_estimates() – Get data from US Census Bureau Population Estimates Program

- Arguments will vary based on function
  - To see required arguments, read R package documentation or use args() functions
    - Census API Package:
    - Tidycensus Package:
    - args(function_name) will provide arguments needed

QUERY DECENNIAL DATA

- get_decennial()

```r
Race_County_2010 <- get_decennial(geography = "county", state = 39, county = c(041, 045, 049, 089, 097, 129, 159), variables = c("P003001", "P003002", "P003003", "P003004", "P003005", "P003006", "P003007", "P003008"), year = 2010) #Query data
```

- Result: Data table in long format

GEOID	NAME	variable	value
39041	Delaware County, Ohio	P003001	174214
39045	Fairfield County, Ohio	P003001	146156
39049	Franklin County, Ohio	P003001	1163414
39089	Licking County, Ohio	P003001	166492
39097	Madison County, Ohio	P003001	43435
39129	Pickaway County, Ohio	P003001	55698
39159	Union County, Ohio	P003001	52300
39041	Delaware County, Ohio	P003002	156328
39045	Fairfield County, Ohio	P003002	131830
39049	Franklin County, Ohio	P003002	805617
39089	Licking County, Ohio	P003002	155190
39097	Madison County, Ohio	P003002	39364

Usage

```
get_decennial(
  geography,
  variables = NULL,
  table = NULL,
  cache_table = FALSE,
  year = 2010,
  sumfile = "sf1",
  state = NULL,
  county = NULL,
  geometry = FALSE,
  output = "tidy",
  keep_geo_vars = FALSE,
  shift_geo = FALSE,
  summary_var = NULL,
  key = NULL,
  show_call = FALSE,
  ...
)
```

INITIATE CENSUS API IN R

MORPC

- Step 3: Create new R notebook; Install packages and get to downloading

```
#This section of the notebook demonstrates how to use the Census API to retrieve Census Data

install.packages("pacman")
library(pacman)

#Install and load Census R packages
pacman::p_load(censusapi, tidycensus, tidyverse, tidyr, dplyr)

census_api_key("YOUR_API_KEY", overwrite = TRUE)

Sys.setenv(CENSUS_KEY = "YOUR_API_KEY")
```

MANIPULATE CENSUS DATA IN R

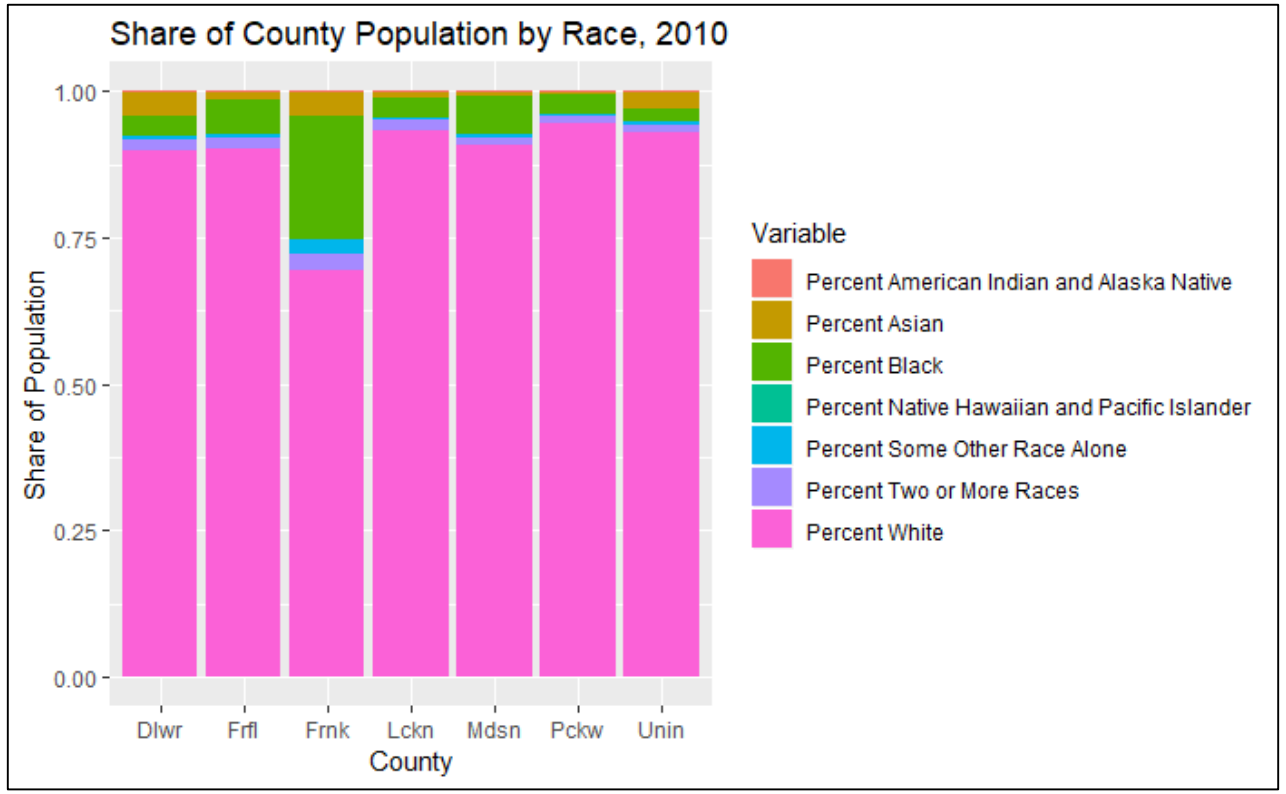
MORPC

- Change shape of data and calculate race percentages

```
Race_County_2010 <- spread(Race_County_2010, variable, value) #change shape of data from wide to long

#Calculate portion of population belonging to each racial category

Race_County_2010 <- mutate(Race_County_2010, "Percent White" = P003002 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent Black" = P003003 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent American Indian and Alaska Native" = P003004 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent Asian" = P003005 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent Native Hawaiian and Pacific Islander" = P003006 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent Some Other Race Alone" = P003007 / P003001)

Race_County_2010 <- mutate(Race_County_2010, "Percent Two or More Races" = P003008 / P003001)

#Reduce to percents only
Race_County_2010 <- Race_County_2010[c("GEOID", "NAME", "Percent White", "Percent Black", "Percent American Indian and Alaska Native",
"Percent Asian", "Percent Native Hawaiian and Pacific Islander", "Percent Some Other Race Alone", "Percent Two or More Races")]

Race_County_2010 <- mutate(Race_County_2010, NAME_Short = str_sub(Race_County_2010$NAME, start = 1, end = -14)) #create variable with
shortened county name for visualization

Race_County_2010_long <- gather(Race_County_2010, Variable, Estimate, "Percent White":"Percent Two or More Races") #make data long agair
```

PLOT DECENNIAL DATA

```
pacman::p_load(ggplot2)

#Add graph
ggplot(Race_County_2010_long, aes(x=NAME_Short, y=Estimate, fill=Variable))+
  geom_bar(stat='identity') + ylab("Share of Population") + xlab("County") + ggtitle("Share of County Population by Race,
2010") + scale_x_discrete(labels = abbreviate)
```



MORE IN CODE

- Example 1
  - Retrieve decennial and American-Community Survey data at county-level for 2010 and 2018
    - Calculate percentage of each race
    - Plot share of population by race for each county (2010, 2018, and both together)
- Example 2
  - Retrieve decennial and American-Community Survey data at tract-level for 2010 and 2018
    - Calculate percentage of each race
    - Map share of population by race at tract level

RESOURCES

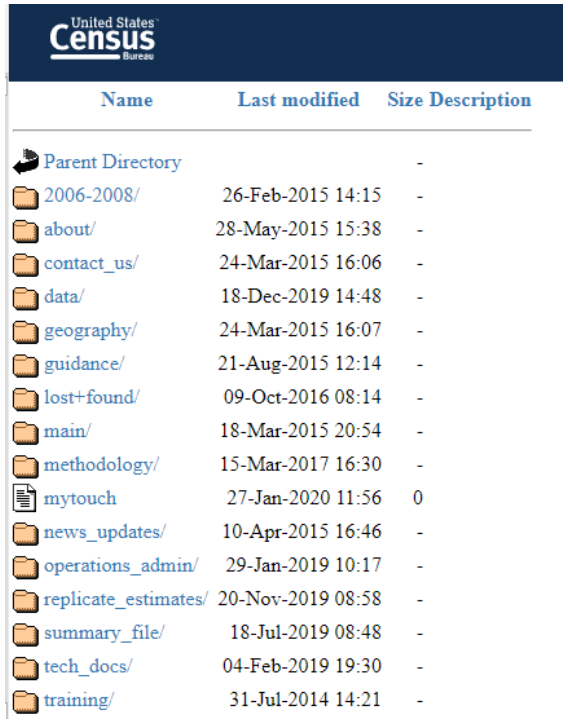
- Census Data API User Guide:

- TidyCensus R Package Documentation:
- CensusAPI R Package Documentation:
- Census API Tutorial:
- Tigris Package Tutorial:

GETTING CENSUS DATA VIA FTP

MORPC

- FTP
  - "File Transfer Protocol"
    - URL that allows for transferring files over the Internet
  - Rather than downloading data from a page, you can query it directly
- Advantages
  - Good for weird geographies (ie part of / remainder)
  - Customizable queries make large data tasks easier
  - More reproducible than online downloads
- Disadvantages
  - Data website and code are a little more complex than API
  - Although variables can be downloaded directly, some tables must be saved to network



GETTING CENSUS DATA VIA FTP

- A handful of Census data products are available via FTP
  - American Community Survey (1-Year, 3-Year, 5-Year, Migration Flows)
  - Decennial Census
  - Public Use Microdata Sample Files
  - And more (!)
    - Info Page:

USING THE FTP IN A WEB BROWSER

- Data can be downloaded directly from FTP website
- Available Programs and Surveys:
- Steps
  - Use ACS variable lookup table to find variables
  - Identify sequence of the variables (column F)
  - Download the sequence zip file from desired survey year
    - programs-surveys/acs/data/5_year_seq_by_state/Ohio/All_Geographies_Not_Tracts_Block_Groups
  - Download needed sequence file and download
    - Sequence 3 = 20185oh0003000.zip
  - Save page as txt file
  - Open txt file in Excel to convert to spreadsheet

USING THE FTP IN R

- Step 1: Download R and R Studio
- Step 2: Download necessary files and save to network
  - Variable Lookup Table: 
    Microsoft Excel Worksheet
  - Geographic Identifiers:
  - Table Headings:
- Step 3: Use R to query data
  - Requires many of the same steps as downloading directly from FTP, but R does more of the work

USING THE FTP IN R

MORPC

- Code to read in Census data via FTP

```
pacman::p_load(readxl, dplyr, tidyverse)

temp <- tempfile()
download.file("https://www2.census.gov/programs-surveys/acs/summary_file/2018/data/5_year_seq_by_state/Ohio/All_Geographies_Not_Tracts_Block_Groups/20185oh0014000.zip", temp) #In the example below the '103' part of the file name represents the sequence ID. Replace with desired sequence ID from lookup table. You can also change the year in the file path, or the estimate type (e.g. 5-year), or change to pull in tracts and block groups. Go to the ftp site here https://www2.census.gov/programs-surveys/acs/summary_file to explore path name structure.

data <-read.table(unz(temp, "e20185oh0014000.txt")) # Create a temporary file on the PC to store the zipped file, then extract the data (it comes in as one field with all values separated by commas. This is resolved in a later step)

seq_head <- read.csv("N:/DATA/Data Management/Lookup Tables/Templates/Seq14.csv") # Call in the variable names (when pulling in the template, replace 'Seq103' with 'Seq1' or 'Seq12' or 'Seq102' in the file name, depending on which you are pulling in)

geo_lookup <- read_xlsx("N:/DATA/Data Management/Lookup Tables/oh_geo_crosswalk.xlsx", "oh") # Call in the geography names

data_sf1acs <- data %>%
  separate(V1, c(variable.names(seq_head)), ",") # Split the variables into separate columns then assign the correct variable names

subsetacs <- data_sf1acs %>%
  select(LOGRECNO, B06009_001, B06009_002, B06009_003, B06009_004, B06009_005, B06009_006) %>% # Select the variables
  left_join (geo_lookup, by = c("LOGRECNO" = "Logical Record Number")) # Join to geography names and identifiers. Now the "Geography ID" field can be used to filter out a subset of the geographies required.
```

USING THE FTP IN R

- Output of code on previous slide provides specified variables for all Ohio geographies (besides block groups and tracts – these are in another FTP file)
- Next, you will select reduce the data down by selecting the geography you care about

```
MPO_Edu_Attainment_2018 <- subset(subsetacs, LOGRECNO %in% MPO_GEOID$LOGRECNO) #select only geographies in MPO
(references MPO geography data frame created above)

MPO_Edu_Attainment_2018 <- gather(MPO_Edu_Attainment_2018, Variable, Estimate, B06009_001:B06009_006) #Change data from
wide to long

MPO_Edu_Attainment_2018$Estimate <- as.numeric(MPO_Edu_Attainment_2018$Estimate) #Change estimate column to numeric

MPO_Edu_Attainment_2018 <- na.omit(MPO_Edu_Attainment_2018) #Remove na values from data frame
```

LOGRECNO	State	Geography ID	Geography Name	Variable	Estimate
0000033	OH	05000US39041	Delaware County, Ohio	B06009_001	128544
0000037	OH	05000US39049	Franklin County, Ohio	B06009_001	848603
0016912	OH	15500US3911332045	Fairfield County (part), Canal Winchester city, Ohio	B06009_001	252
0016984	OH	15500US3918000045	Fairfield County (part), Columbus city, Ohio	B06009_001	4134
0017059	OH	15500US3922694159	Union County (part), Dublin city, Ohio	B06009_001	4919
0017354	OH	15500US3944086045	Fairfield County (part), Lithopolis village, Ohio	B06009_001	364
0017643	OH	15500US3961112089	Licking County (part), Pataskala city, Ohio	B06009_001	123
0017659	OH	15500US3962498045	Fairfield County (part), Pickerington city, Ohio	B06009_001	292
0017714	OH	15500US3966390045	Fairfield County (part), Reynoldsburg city, Ohio	B06009_001	195
0017716	OH	15500US3966390089	Licking County (part), Reynoldsburg city, Ohio	B06009_001	853

USING THE FTP IN R

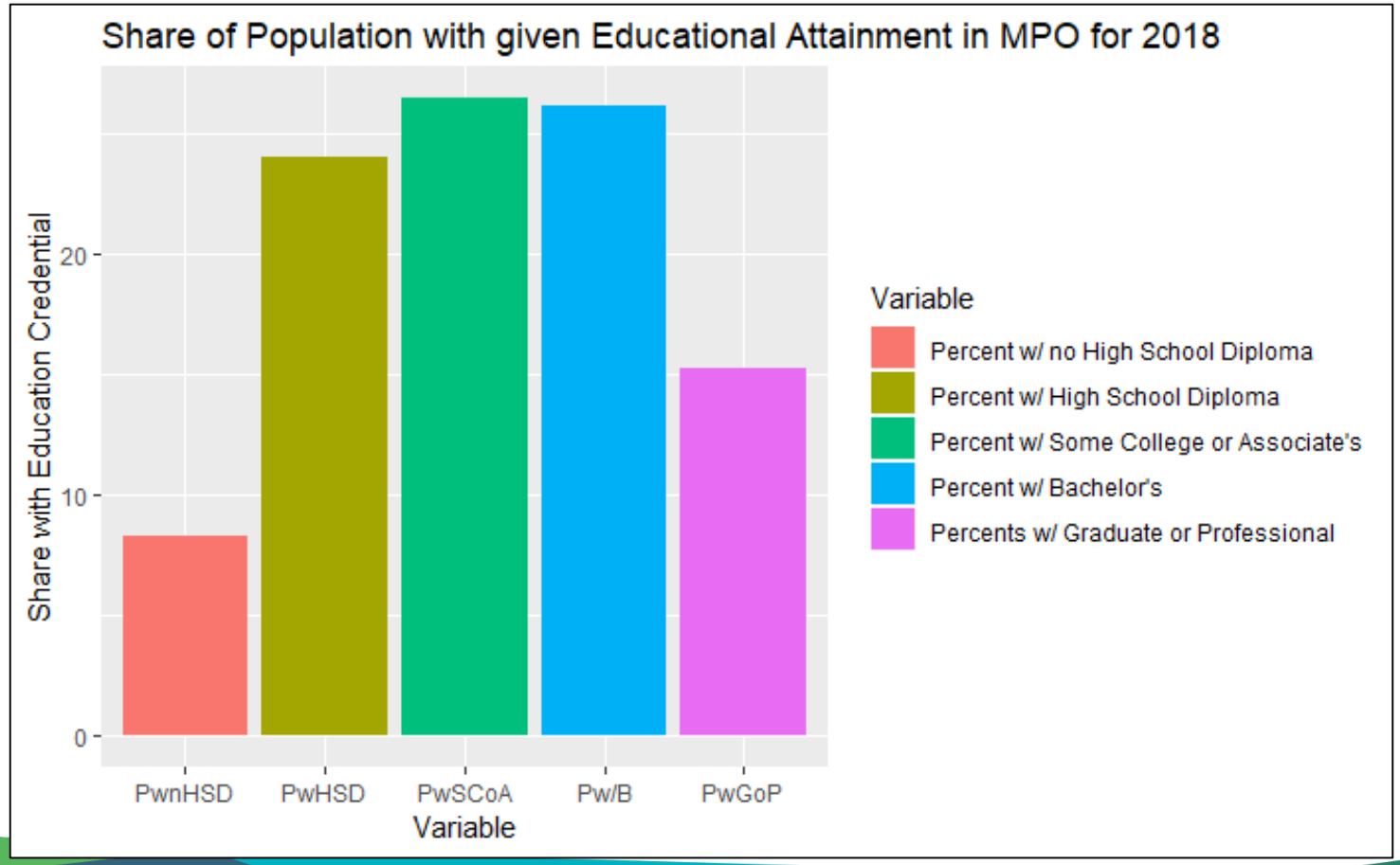
- Sum each variable for entire MPO, make data wide, then calculate share of population with educational attainment

```
#Sum MPO
MPO_Edu_Attainment_2018 <- MPO_Edu_Attainment_2018 %>%
  group_by(Variable) %>%
  summarise(Estimate = sum(Estimate))

#Make data wide
MPO_Edu_Attainment_2018 <- spread(MPO_Edu_Attainment_2018, Variable, Estimate)

#Calculate %
MPO_Edu_Attainment_2018 <- mutate(MPO_Edu_Attainment_2018, "Percent w/ no High School Diploma" = (B06009_002 /
B06009_001) * 100)

MPO_Edu_Attainment_2018 <- mutate(MPO_Edu_Attainment_2018, "Percent w/ High School Diploma" = (B06009_003 / B06009_001) *
100)

MPO_Edu_Attainment_2018 <- mutate(MPO_Edu_Attainment_2018, "Percent w/ Some College or Associate's" = (B06009_004 /
B06009_001) * 100)

MPO_Edu_Attainment_2018 <- mutate(MPO_Edu_Attainment_2018, "Percent w/ Bachelor's" = (B06009_005 / B06009_001) * 100)

MPO_Edu_Attainment_2018 <- mutate(MPO_Edu_Attainment_2018, "Percents w/ Graduate or Professional" = (B06009_006 /
B06009_001) * 100)
```

B06009_001	B06009_002	B06009_003	B06009_004	B06009_005	B06009_006	Percent w/ no High School Diploma	Percent w/ High School Diploma	Percent w/ Some College or Associate's	Percent w/ Bachelor's	Percents w/ Graduate or Professional
988279	81471	237206	261053	258035	150514	8.243725	24.00193	26.41491	26.10953	15.22991

USING THE FTP IN R

- Plot share of population with given educational attainment
  - Plotting in R can be a little weird, so the code features a few steps to make the graph



MORE IN CODE

- Example 1
  - Retrieve American Community Survey data for MPO in 2018
    - Calculate percentage of population with educational attainment
    - Plot share of population with educational attainment in MPO for 2018
- Example 2
  - Retrieve American Community Survey data for MPO and Ohio in 2018
    - Calculate poverty rate for MPO and Ohio
    - Plot poverty rate for 2018

NATALIE SIOUX HURST
Data Analyst
Mid-Ohio Regional Planning Commission

T: 614.233.4164 | **M**: 740.263.9409
nhurst@morpc.org
111 Liberty Street, Suite 100
Columbus, OH 43215

Questions?