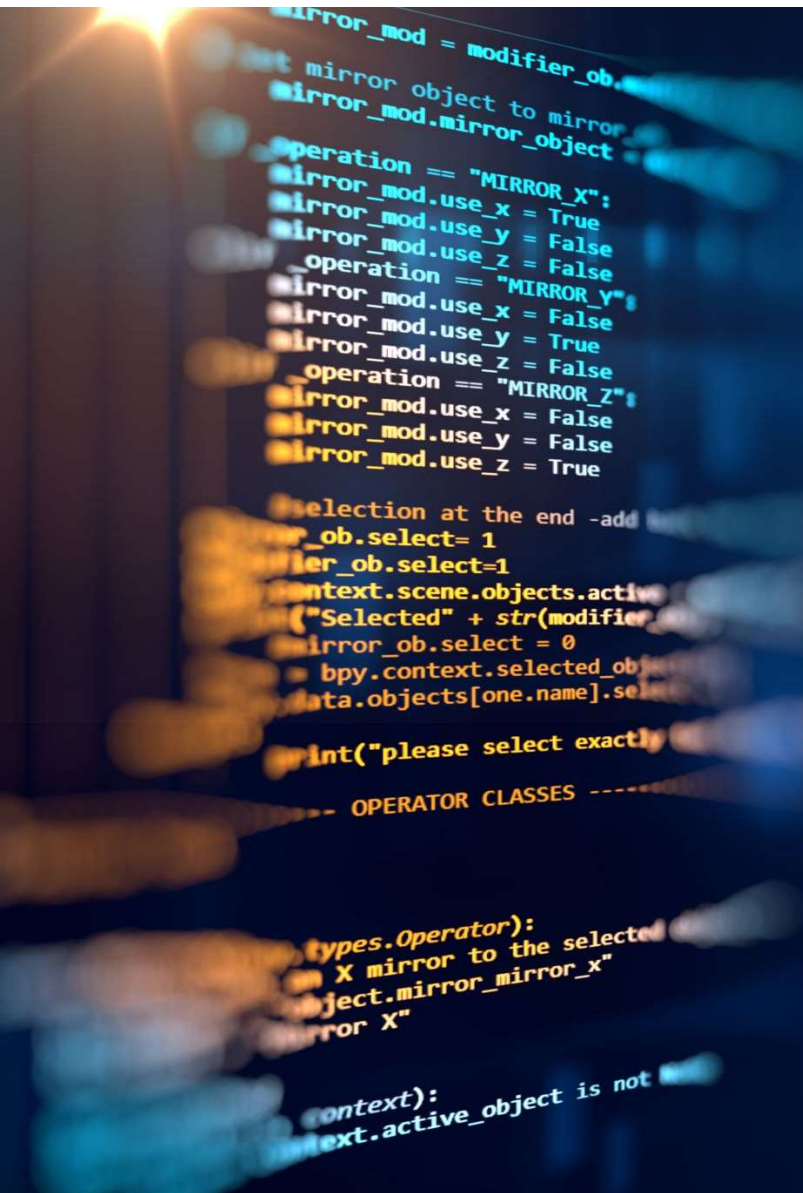


Data science at your fingertips

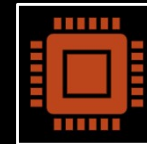
Ahillan Kumar



What?



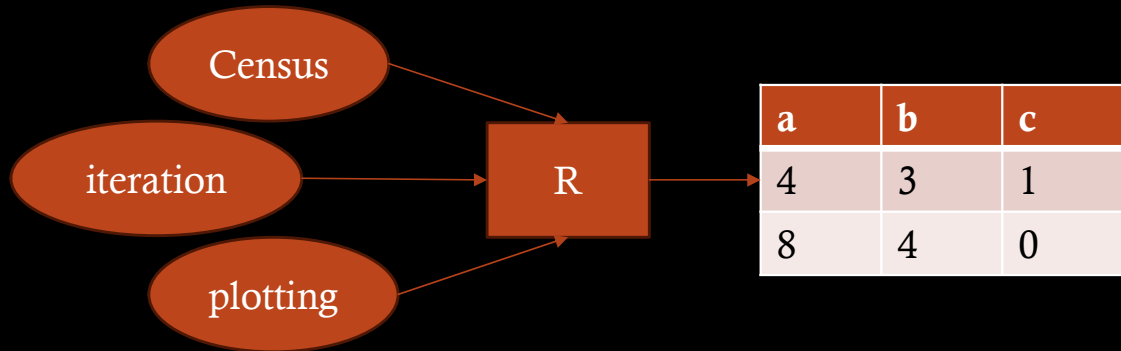
R Studio is a free program created and maintained by Posit, that uses R, a computer programming language



R is a data science tool that uses addons, called packages, that add functionality and usability to the programmer



I used R Studio to make plots, tables, and maps that can be accessed and customized online



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

tutorial.R x

```

1 library(tidyverse)
2 library(tidycensus)
3
4 pop_2018_2021 <- map_dfr(
5   c(2018, 2019, 2020, 2021),
6   ~ get_acs(
7     geography = "county",
8     year = .x,
9     state = "IA",
10    variables = "B01003_001"
11  ) |> mutate(year = .x)
12 ) |>
13 mutate(NAME = str_remove(NAME, " County, Iowa")) |>
14 filter(NAME %in% c("Jefferson", "Wapello", "Keokuk", "Van Buren", "Mahaska"))
15
16
4:26 (Top Level) R Script

```

Environment History Connections Tutorial

Import Dataset 416 MiB

R Global Environment

Data

pop_2018_2021	20 obs. of 6 variables
pop_2022	5 obs. of 5 variables
pop_total	25 obs. of 5 variables

Files Plots Packages Help Viewer Presentation

Zoom Export

Population change (2018-2022) in Area 15 RPC

population estimate

year

Console Terminal x Background Jobs x

```

R 4.3.0 ~-/?
> pop_total <- bind_rows(pop_2018_2021, pop_2022) |>
+   select(-moe)
>
> pop_total |>
+   ggplot(aes(x = year, y = estimate, color = NAME)) +
+   geom_line() +
+   labs(title = "Population change (2018-2022) in Area 15 RPC",
+        y = "population estimate") +
+   scale_color_discrete(name = "county") +
+   scale_y_continuous(labels = scales::comma) +
+   theme_minimal()
Error in exists(cacheKey, where = .rs.WorkingDataEnv, inherits = FALSE) :
  invalid first argument
>

```

What??



It works because the data is connected to application programming interfaces, APIs, that allow computers to communicate to one another through the internet



Once R Studio is connected to the Census API, I can make plots, tables, and maps with any data available through that API

Example



Get a Census API key



Connect R Studio to the Census with the key



Import the data



Tidy the data



Transform the data



Visualize the data



Export the visualization



Get a Census API key

Browser navigation bar showing the URL https://api.census.gov/data/key_signup.html and various browser icons.

Request A Key

Organization Name:

Email Address:

I agree to the [terms of service](#)

Connect R Studio to the Census with the key

The screenshot displays the R Studio interface. In the top-left pane, the R script editor shows the following code line, which is highlighted with a red box:

```
1 census_api_key("asdfagaagaragadgagarhaadfgadrgargadvadggag", install= TRUE)
```

The top-right pane shows the Environment tab with the 'Import Dataset' button highlighted by a red box. Below it, the Environment is empty.

The bottom-right pane shows the Files tab with the search results for 'census_api_key'. The search results include the following text:

R: Install a CENSUS API Key in Your '.Renviron' File for... Find in Topic

census_api_key {tidycensus} R Documentation

Install a CENSUS API Key in Your .Renviron File for Repeated Use

Description

This function will add your CENSUS API key to your .Renviron file so it can be called securely without being stored in your code. After you have installed your key, it can be called any time by typing `Sys.getenv("CENSUS_API_KEY")` and can be used in package functions by simply typing `CENSUS_API_KEY`. If you do not have an .Renviron file, the function will create one for you. If you already have an .Renviron file, the function will append the key to your existing file, while making a backup of your original file for disaster recovery purposes.

Usage

```
census_api_key(key, overwrite = FALSE, install = FALSE)
```



Import the data



Tidy the data



Transform the data

```

library(tidyverse)
library(tidycensus)

pop_2018_2021 <- map_dfr(
  c(2018, 2019, 2020, 2021),
  ~ get_acs(
    geography = "county",
    year = .x,
    state = "IA",
    variables = "B01003_001"
  ) |> mutate(year = .x)
) |>
mutate(NAME = str_remove(NAME, " County, Iowa")) |>
filter(NAME %in% c("Jefferson", "Wapello", "Keokuk", "Van Buren", "Mahaska"))

pop_2022 <- get_estimates(
  geography = "county",
  year = 2022,
  state = "IA",
  variables = "POPESTIMATE"
) |>
rename(estimate = value) |>
mutate(NAME = str_remove(NAME, " County, Iowa")) |>
filter(NAME %in% c("Jefferson", "Wapello", "Keokuk", "Van Buren", "Mahaska"))

pop_total <- bind_rows(pop_2018_2021, pop_2022) |>
select(-moe)

```

	GEOID	NAME	variable	estimate	moe	year
1	19101	Jefferson	B01003_001	18077	NA	2018
2	19107	Keokuk	B01003_001	10200	NA	2018
3	19123	Mahaska	B01003_001	22208	NA	2018
4	19177	Van Buren	B01003_001	7223	NA	2018
5	19179	Wapello	B01003_001	35315	NA	2018

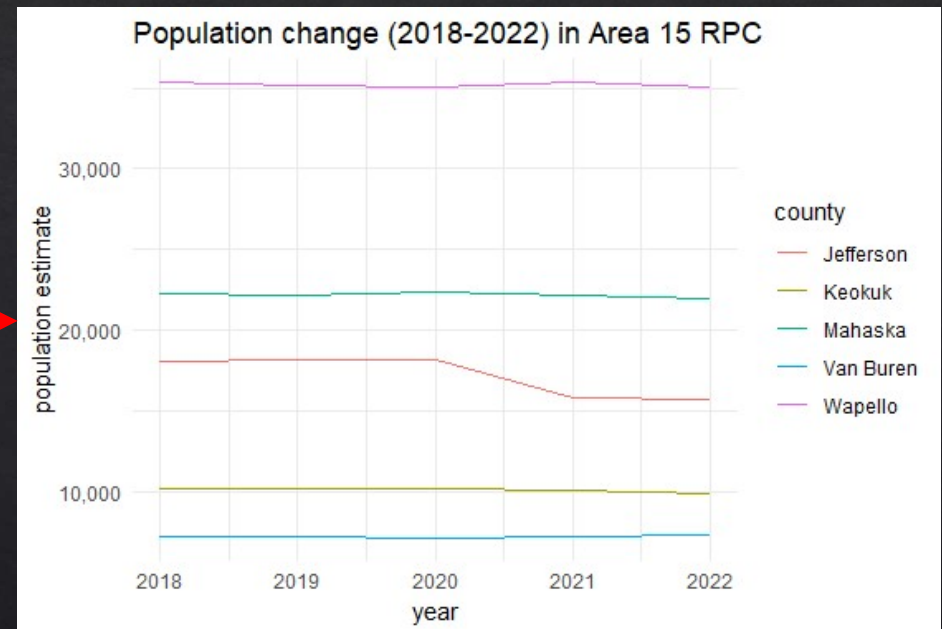
	GEOID	NAME	variable	year	estimate
1	19101	Jefferson	POPESTIMATE	2022	15698
2	19107	Keokuk	POPESTIMATE	2022	9904
3	19123	Mahaska	POPESTIMATE	2022	21946
4	19177	Van Buren	POPESTIMATE	2022	7256
5	19179	Wapello	POPESTIMATE	2022	35043

	GEOID	NAME	variable	estimate	year
1	19101	Jefferson	B01003_001	18077	2018
2	19107	Keokuk	B01003_001	10200	2018
3	19123	Mahaska	B01003_001	22208	2018
4	19177	Van Buren	B01003_001	7223	2018
5	19179	Wapello	B01003_001	35315	2018



Visualize the data

```
pop_total |>
  ggplot(aes(x = year, y = estimate, color = NAME)) +
  geom_line() +
  labs(title = "Population change (2018-2022) in Area 15 RPC",
       y = "population estimate") +
  scale_color_discrete(name = "county") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal()
```



Export the visualization

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation and plotting.
- Environment:** Lists data objects: `pop_2018_2021` (20 obs. of 6 variables), `pop_2022` (5 obs. of 5 variables), and `pop_total` (25 obs. of 5 variables).
- Plots:** Shows a line plot titled "Population change (2018-2022) in Area 15 RPC" with a red box highlighting the "Export" menu options: "Save as Image...", "Save as PDF...", and "Copy to Clipboard...".
- Console:** Shows the execution of the R code and an error message: `Error in exists(cacheKey, where = .rs.WorkingDataEnv, inherits = FALSE) : invalid first argument`.

```
1 library(tidyverse)
2 library(tidycensus)
3
4 pop_2018_2021 <- map_dfr(
5   c(2018, 2019, 2020, 2021),
6   ~ get_acs(
7     geography = "county",
8     year = .x,
9     state = "IA",
10    variables = "B01003_001"
11  ) |> mutate(year = .x)
12 ) |>
13 mutate(NAME = str_remove(NAME, " County, Iowa")) |>
14 filter(NAME %in% c("Jefferson", "Wapello", "Keokuk", "Van Buren", "Mahaska"))
15
16
```

```
> pop_total <- bind_rows(pop_2018_2021, pop_2022) |>
+   select(-moe)
>
> pop_total |>
+   ggplot(aes(x = year, y = estimate, color = NAME)) +
+   geom_line() +
+   labs(title = "Population change (2018-2022) in Area 15 RPC",
+        y = "population estimate") +
+   scale_color_discrete(name = "county") +
+   scale_y_continuous(labels = scales::comma) +
+   theme_minimal()
Error in exists(cacheKey, where = .rs.WorkingDataEnv, inherits = FALSE) :
  invalid first argument
>
```

county	2018	2019	2020	2021	2022
Jefferson	18,000	18,000	18,000	16,000	16,000
Keokuk	10,000	10,000	10,000	10,000	10,000
Mahaska	22,000	22,000	22,000	22,000	22,000
Van Buren	7,000	7,000	7,000	7,000	7,000
Wapello	32,000	32,000	32,000	32,000	32,000

Reports

The screenshot displays the RStudio interface with a Quarto report open in the editor. The report content is visible in the main editor pane, and the 'Render' button in the toolbar is highlighted with a red box. The report content includes a title, a description, and R code for loading packages and importing data.

Render

```
---  
title: "tutorial"  
format: docx  
editor: visual  
---
```

Quarto tutorial

This is a tutorial on how to use Quarto to display R code and add text to supplement the data.

Gathering the data

The packages used.

```
{r}  
library(tidyverse)  
library(tidycensus)
```

Importing the data from the ACS and the Population Estimates.

```
{r}  
p00 2018 2021 <- map_dfr(  
Quarto tutorial
```

Environment: Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export

Console

AutoSave On test - Compatibility M... - Upload Blocked

File Home Insert Draw Design Layout References Mailings Review View Help Acrobat

Calibri (Headings) 18

Clipboard Paste Font Paragraph Styles Editing

SIGN IN REQUIRED We can't upload or download your changes because your cached credentials have expired. Sign In

tutorial

Quarto tutorial

This is a tutorial on how to use Quarto to display R code and add text to supplement the data.

Gathering the data

The packages used.

```
library(tidyverse)
```

Warning: package 'ggplot2' was built under R version 4.3.1

Warning: package 'purrr' was built under R version 4.3.1

Page 1 of 3 344 words Text Predictions: On Accessibility: Unavailable

AutoSave On test - Compatibility M... - Upload Blocked

File Home Insert Draw Design Layout References Mailings Review View Help Acrobat

Calibri (Headings) 18

Clipboard Paste Font Paragraph Styles Editing

SIGN IN REQUIRED We can't upload or download your changes because your cached credentials have expired. Sign In

Population change (2018-2022) in Area 15 RPC

Year	Jefferson	Keokuk	Mahaska	Van Buren	Wapello
2018	18,000	10,000	22,000	6,000	34,000
2019	18,000	10,000	22,000	6,000	34,000
2020	18,000	10,000	22,000	6,000	34,000
2021	16,000	10,000	22,000	6,000	34,000
2022	16,000	10,000	22,000	6,000	34,000

Page 3 of 3 344 words Text Predictions: On Accessibility: Unavailable

Web app

The screenshot shows a web browser window displaying a web application. The browser's address bar shows the URL <https://posit.cloud/content/6244127>. The application header includes the Posit Cloud logo, the text "Your Workspace / disability_map", and a user profile for "Ahillan Kumar".

The main content area is titled "disability map". On the left, there is a sidebar with navigation options: "Spaces" (Your Workspace, L RTP chapter 3 plots, table Ahillan Kumar, New Space), "Learn" (Guide, What's New, Primers, Cheat Sheets), and "Help" (Current System Status, Posit Community).

The central control panel contains the following fields and options:

- start year:** A dropdown menu set to "2017".
- end year:** A text input field containing "2021".
- title:** A text input field containing "Figure 3.10: Area disability".
- legend:** A text input field containing "number of people".
- A "download plot" button with a download icon.

On the right, two choropleth maps are displayed side-by-side, labeled "2017" and "2021". A legend titled "number of people" is positioned to the right of the maps, with a color scale ranging from 0 (dark purple) to 400 (yellow). The maps show a geographic area with varying shades of purple and blue, indicating the number of people in different regions for each year.

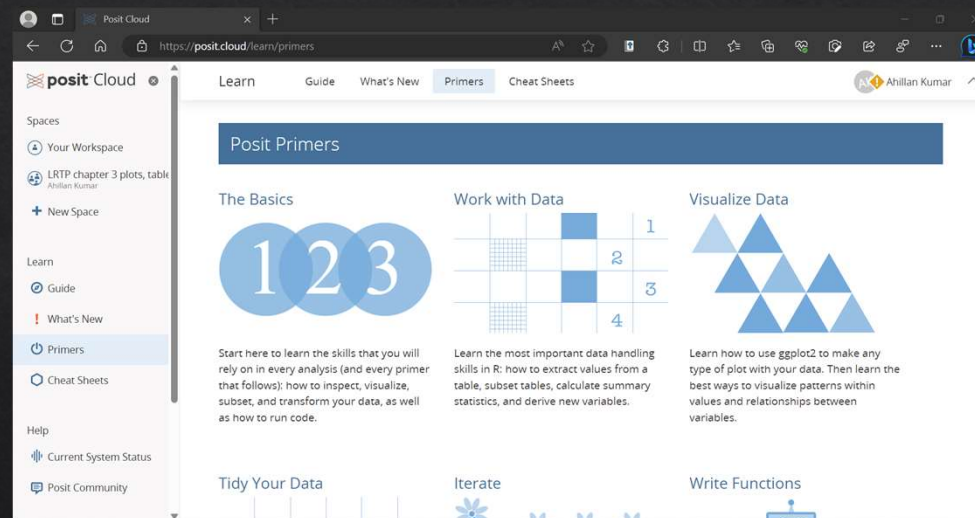
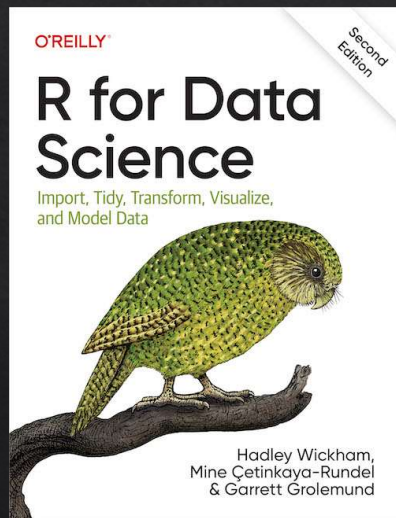
Source: US Census Bureau: 2013-2017 & 2017-2021 American Community Survey

How?

- ◆ Individual programs were written in the R programming language for each plot, table, and map
- ◆ They were then converted into web applications using the Shiny package and uploaded to my Posit Cloud space
- ◆ The apps have customizable labels and data selection by year
- ◆ If an error occurs, I have written a basic set of instructions on how to find the error log and send it to me for fixing
- ◆ The apps aren't without limitations on customization and computational limits as per the free cloud storage policy

How can I?

- ◇ Introduced to and learned basics in Iowa State University class
- ◇ Further documentation



- ◇ Create plots, interactive apps, and reports

SWOT Assessment



Strengths

Quickly and easily customize plots, tables, and maps
Almost instant results
Vast documentation available online



Weaknesses

Basic R coding knowledge required to write and update scripts
Learn package specific functions and know how to use them



Opportunities

Explore data distributions and aesthetics
Automate reports with figures
Create web apps



Threats

Learning curve/time to mastery

Uses



USED IT TO UPDATE
THE LONG-RANGE
TRANSPORTATION
PLAN



CREATE
STANDALONE
TABLES, PLOTS, AND
MAPS



EXPLORE DATA
RELATIONSHIPS



CREATE OR UPDATE
ANY REPORT STYLE
DOCUMENT



CREATE
INTERACTIVE WEB
APPS FOR NON-
PROGRAMMERS

Contact

✉ Ahillan Kumar, ahillank@gmail.com

↓ Download and learn R Studio, <https://posit.co>

🖥 Learn R Studio, <https://r4ds.hadley.nz>

☁ Use R Studio online and save in the cloud, <https://posit.cloud>

🗄 Get a Census API key, https://api.census.gov/data/key_signup.html

✓ Comprehensive list of R packages, https://cran.r-project.org/web/packages/available_packages_by_name.html