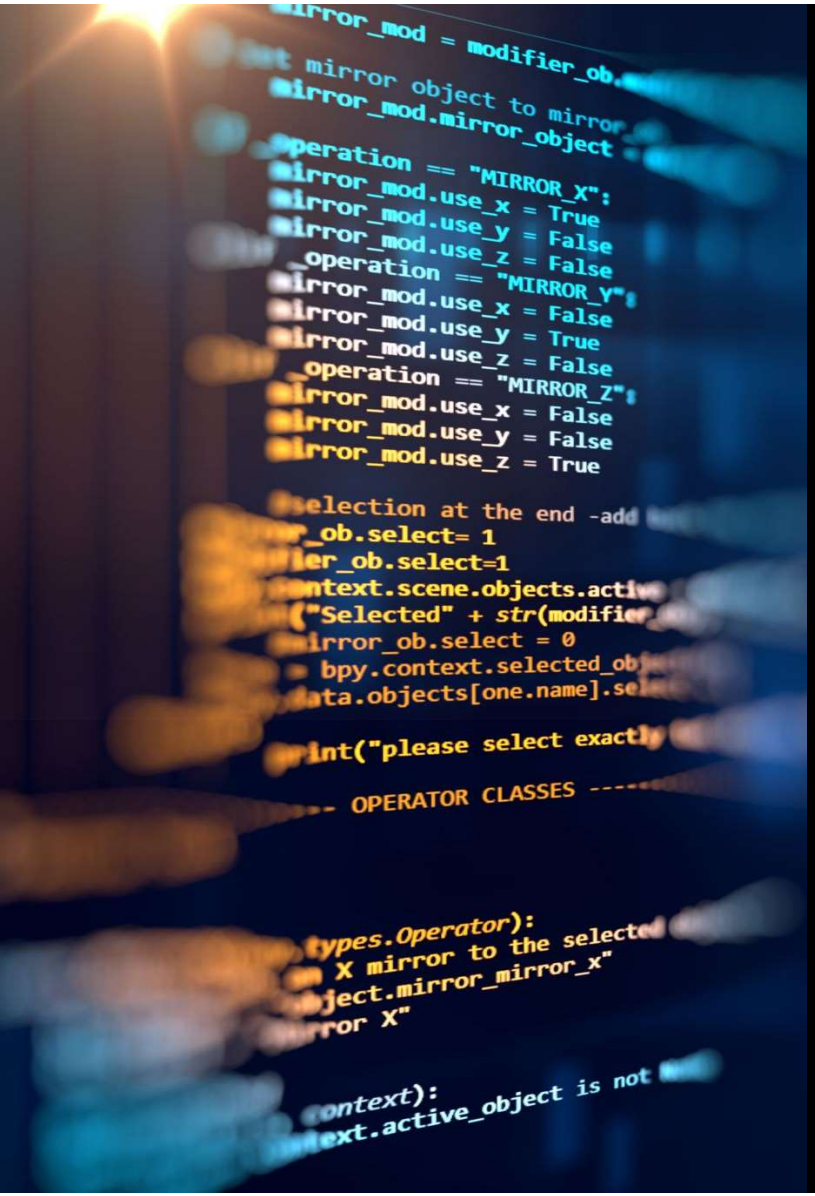
Data science at your fingertips

Ahillan Kumar

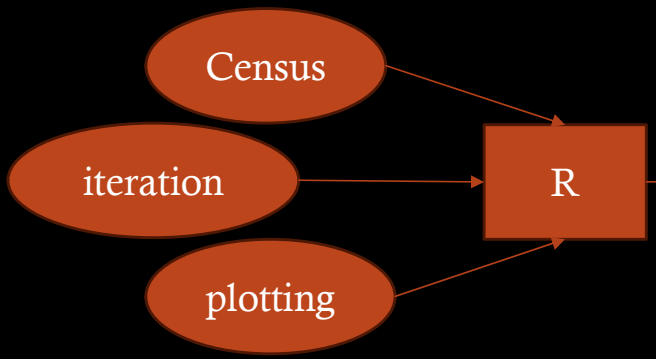


What?

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



a	b	c
4	3	1
8	4	0

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

tutorial.R

Source on Save | Run | Source

```
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

Console Terminal Background Jobs

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
>
```

Environment History Connections Tutorial

Import Dataset | 416 MiB

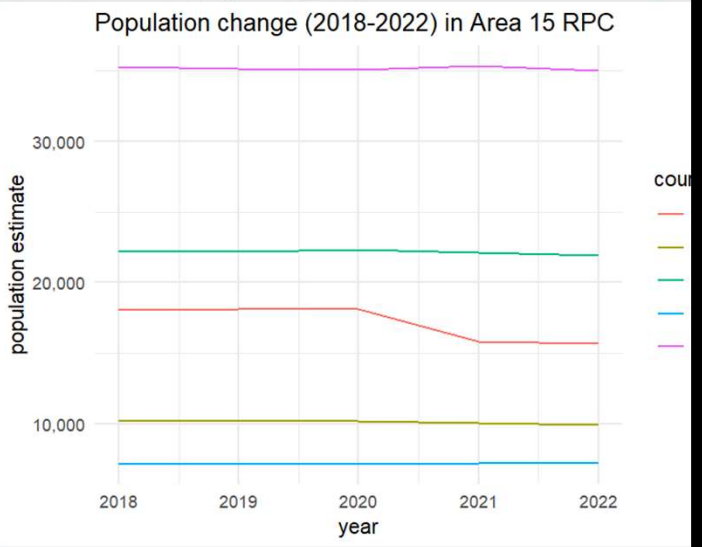
R | Global Environment

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



What??

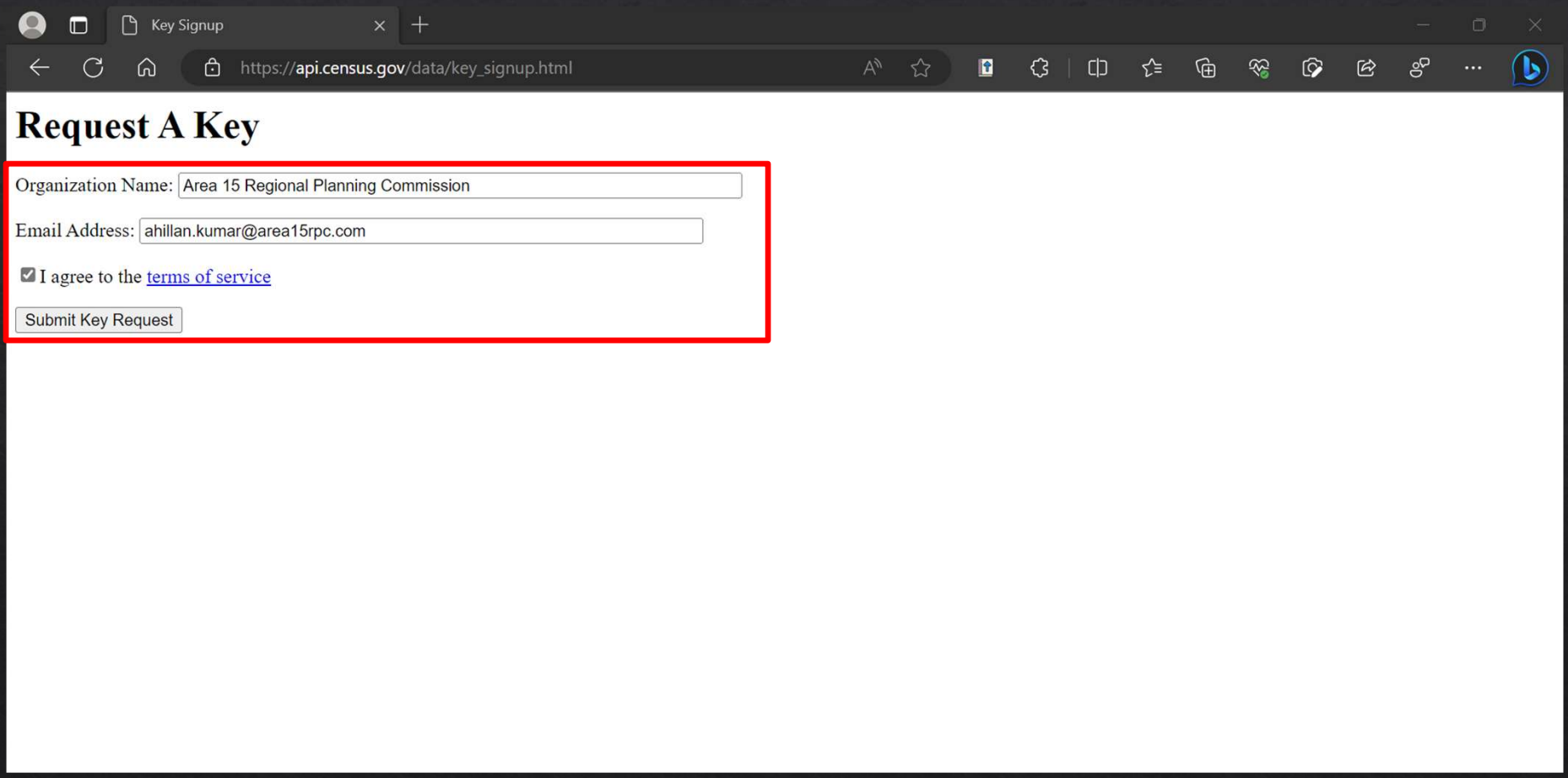
It works because the data is connected to application programming interfaces, APIs, that allow computers to communicate to one another through the internet

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

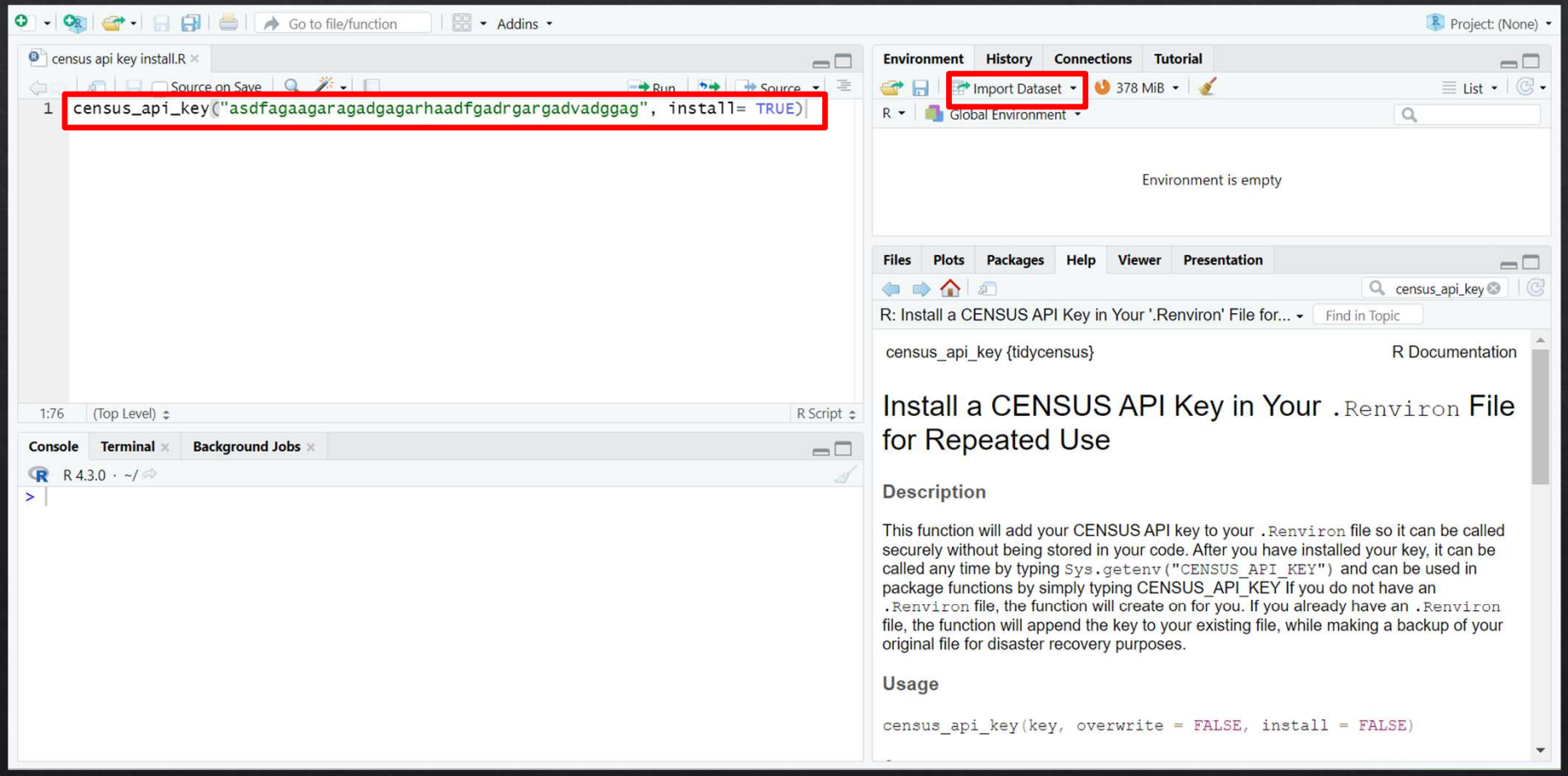
Example

- 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



Connect R Studio to the Census with the key



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)
```

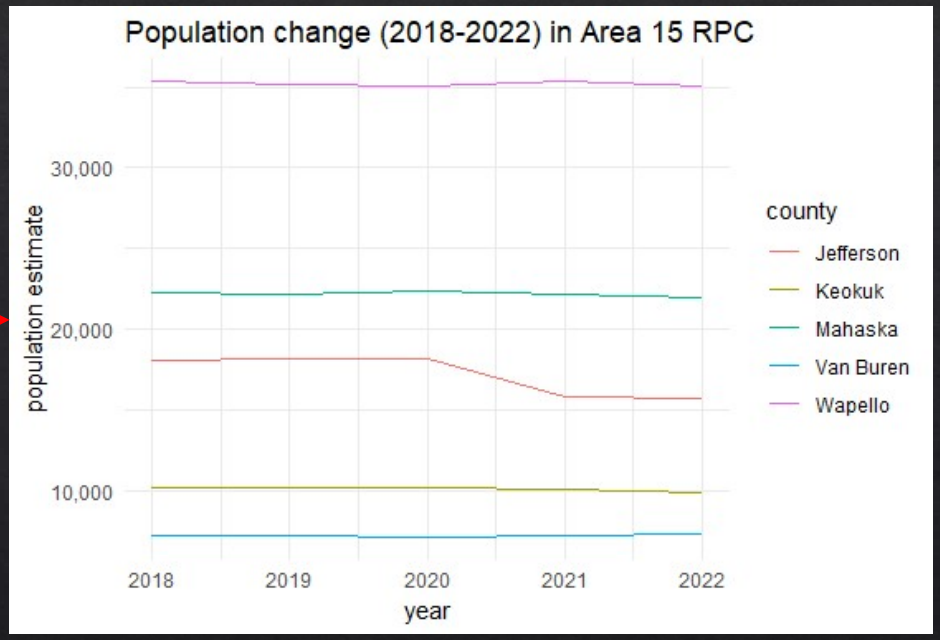
	GEOID	NAME	variable	estimate	moe	year
1	19101	Jefferson	B01003_001	18077	*NA*	2018
2	19107	Keokuk	B01003_001	10200	*NA*	2018
3	19123	Mahaska	B01003_001	22208	*NA*	2018
4	19177	Van Buren	B01003_001	7223	*NA*	2018
5	19179	Wapello	B01003_001	35315	*NA*	2018

	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

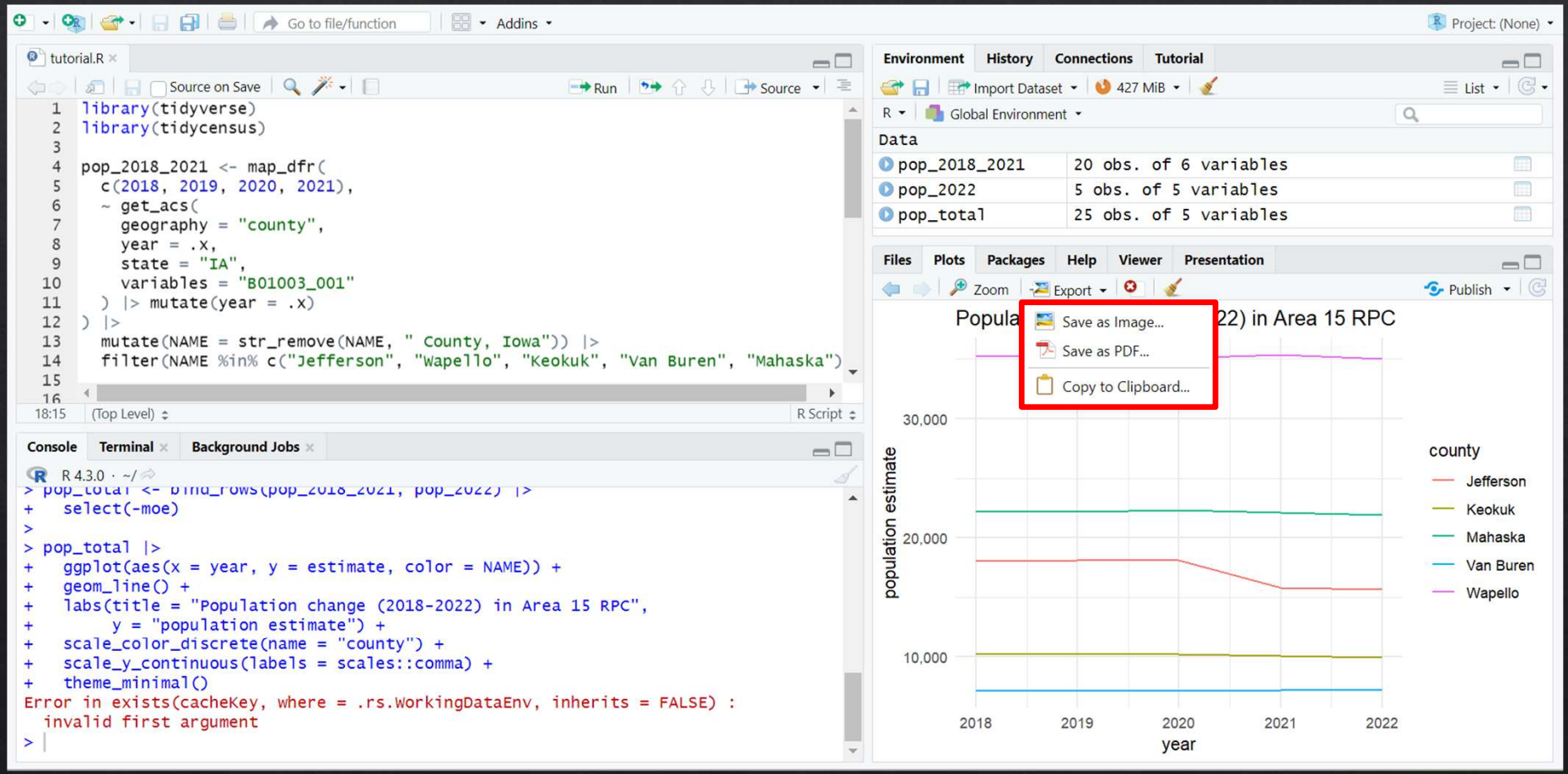
	GEOID	NAME	variable	estimate	year
1	19101	Jefferson	B01003_001	18077	2018
2	19107	Keokuk	B01003_001	10200	2018
3	19123	Mahaska	B01003_001	22208	2018
4	19177	Van Buren	B01003_001	7223	2018
5	19179	Wapello	B01003_001	35315	2018

Visualize the data

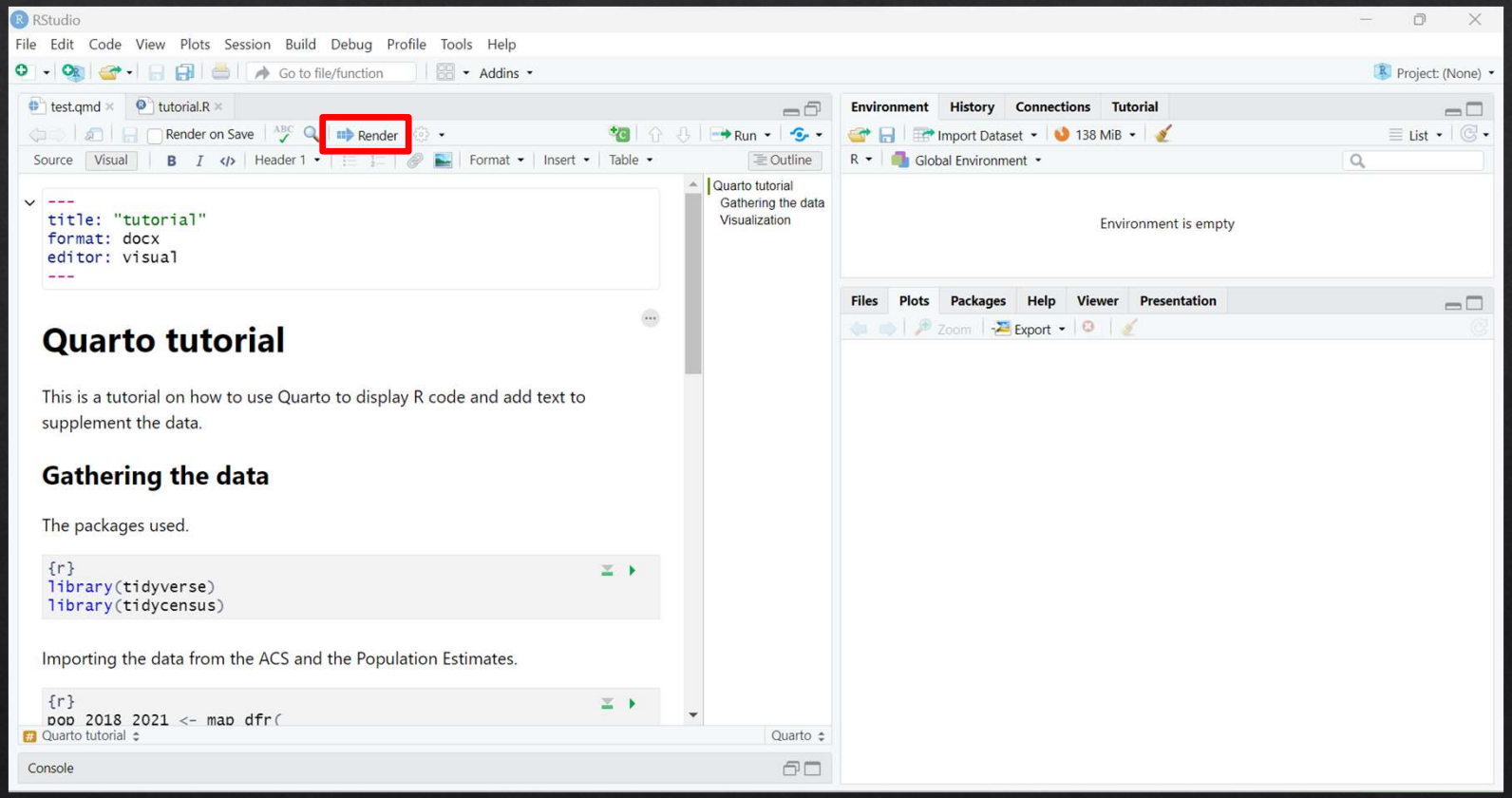
```
pop_total |>
  ggplot(aes(x = year, y = estimate, color = NAME)) +
  geom_line() +
  labs(title = "Population change (2018-2022) in Area 15 RPC",
       y = "population estimate") +
  scale_color_discrete(name = "county") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal()
```



Export the visualization



Reports


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
test.qmd
tutorial.R
Render on Save
Render
Run
Source
Visual
Header 1
Format
Insert
Table
Outline
---
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}
pop_2018_2021 <- map_dfr(
Quarto tutorial
Gathering the data
Visualization
Quarto tutorial
Quarto
Console
Environment
History
Connections
Tutorial
Import Dataset
138 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export

tutorial

Quarto tutorial

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

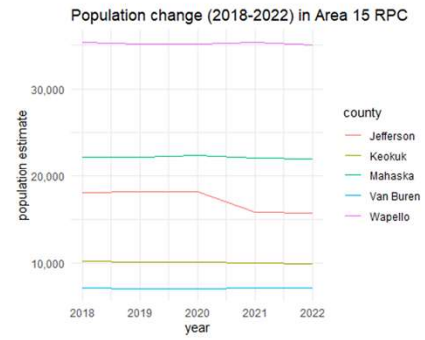
Gathering the data

The packages used.

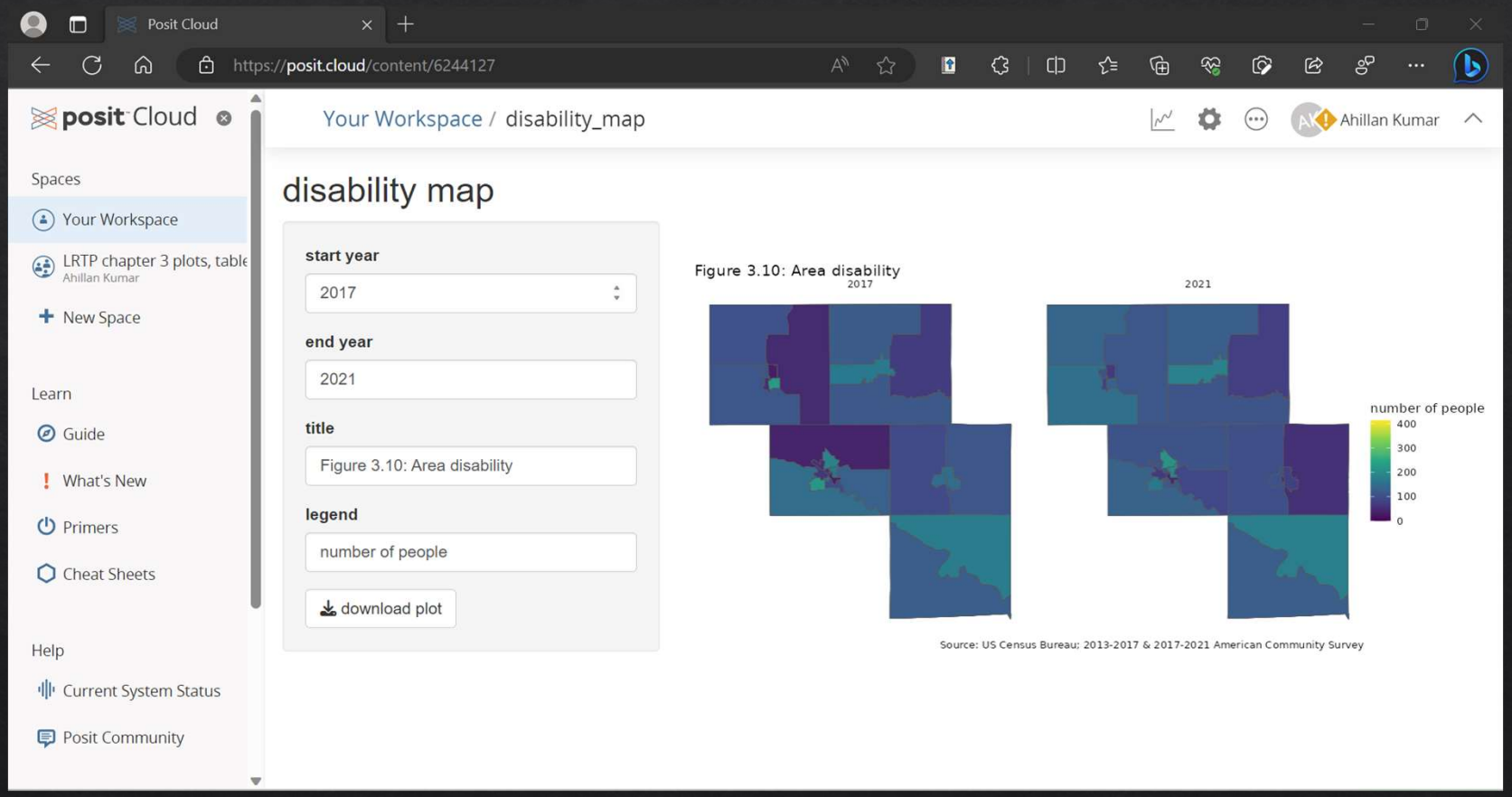
```
library(tidyverse)
```

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

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



Web app

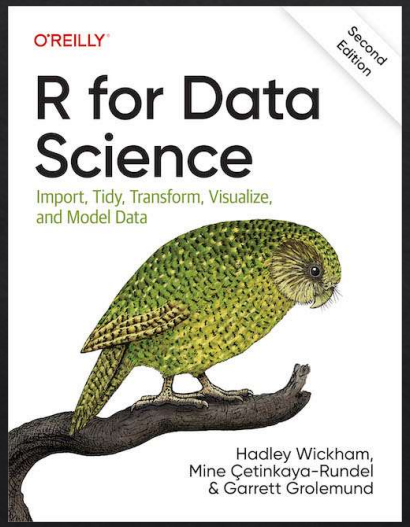

Posit Cloud
https://posit.cloud/content/6244127
posit Cloud
Your Workspace / disability_map
Ahillan Kumar
Spaces
Your Workspace
LRTP chapter 3 plots, table
Ahillan Kumar
New Space
Learn
Guide
What's New
Primers
Cheat Sheets
Help
Current System Status
Posit Community
disability map
start year
2017
end year
2021
title
Figure 3.10: Area disability
legend
number of people
download plot
Figure 3.10: Area disability
2017
2021
number of people
400
300
200
100
0
Source: US Census Bureau; 2013-2017 & 2017-2021 American Community Survey

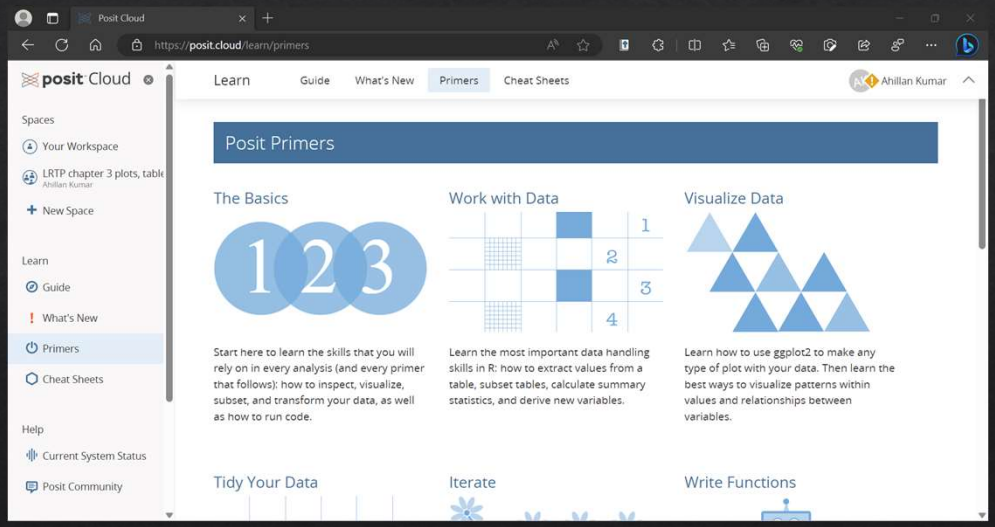
How?

- 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