

# Data Analytics

Lecture Series: Part 2

Sourcing Data

Peter J. Mattingly

pjm407@nyu.edu

# Overview



# Overview

In this section, we will:



# Overview

In this section, we will:

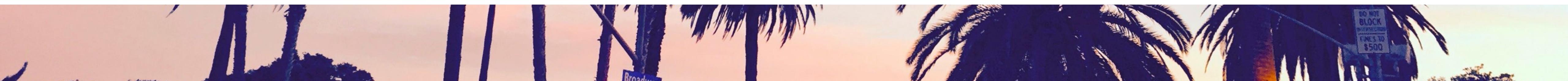
- Source real estate and population data





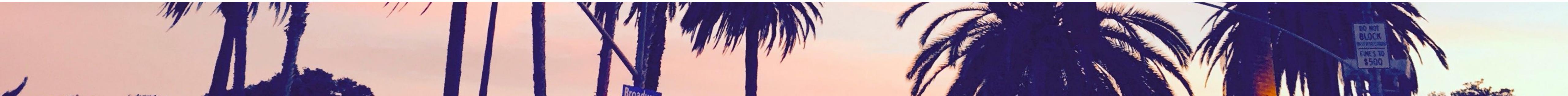
# Rmarkdown

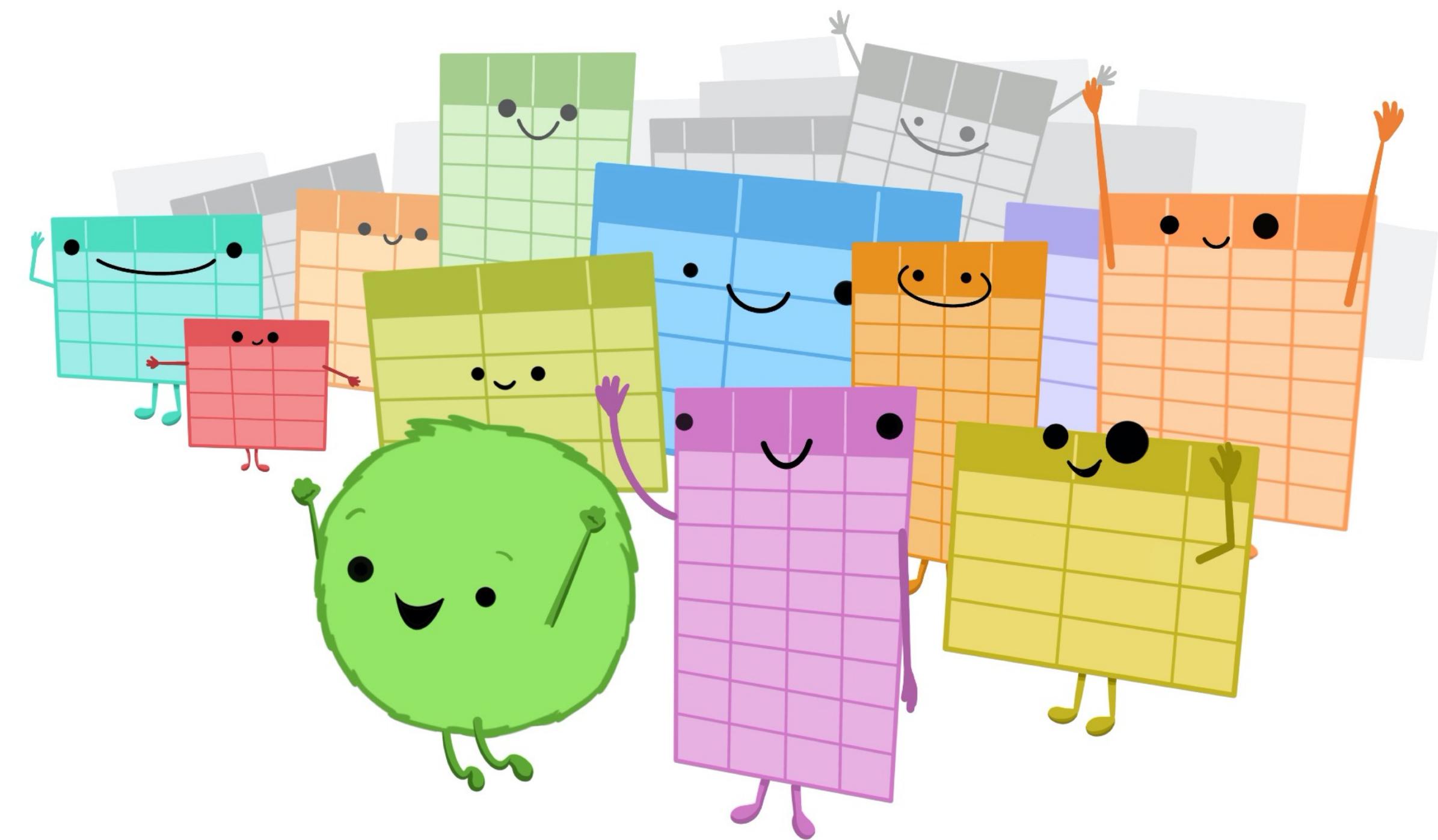
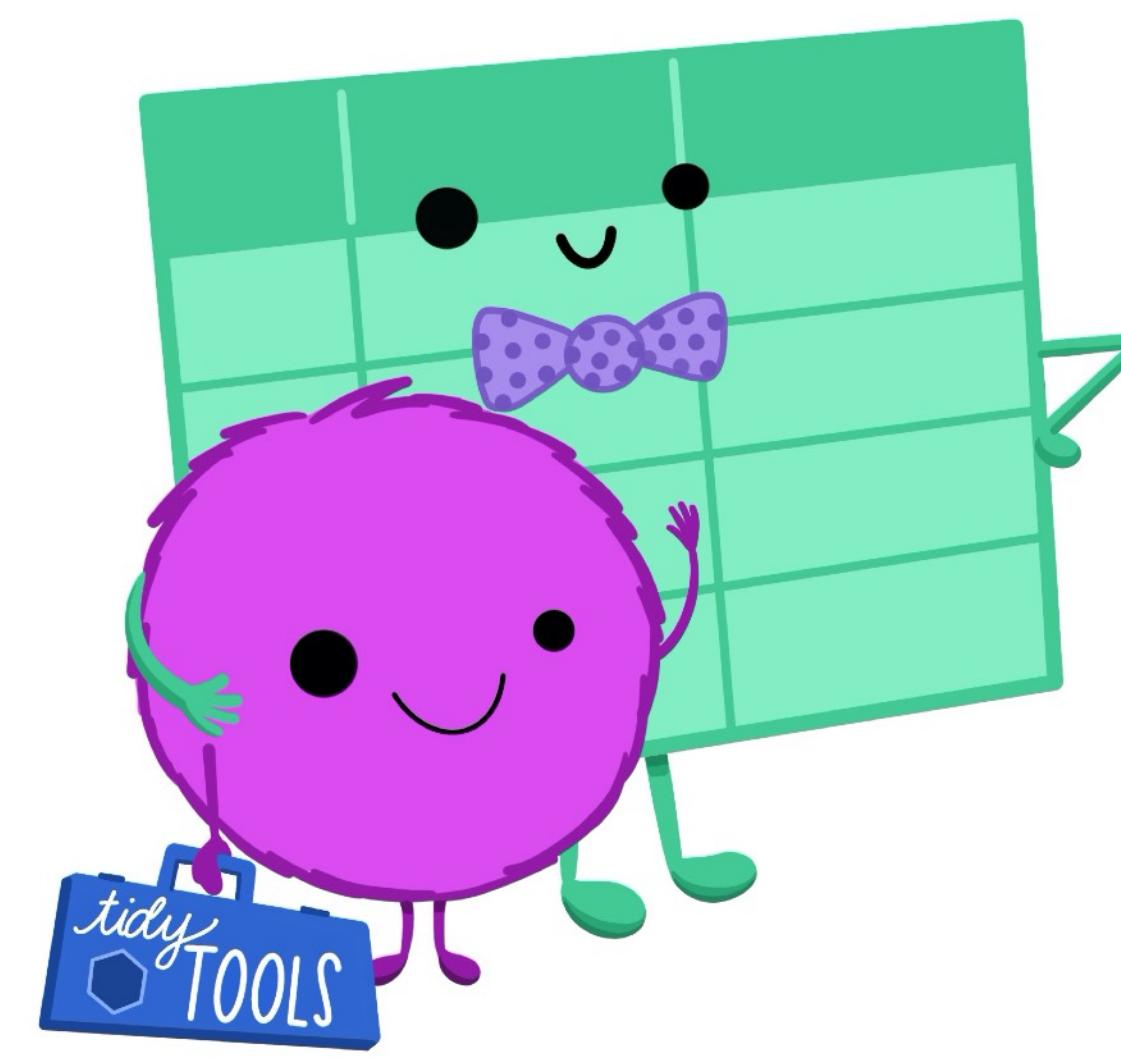
TEXT. CODE. OUTPUT.  
(GET IT TOGETHER, PEOPLE.)





HORST'19





# Data

VARIABLE 1	VARIABLE 2	VARIABLE 3	VARIABLE 4	VARIABLE 5
Medium gray				
Dark gray				
Black	Black	Black	Black	Black

“TIDY DATA” is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

## In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

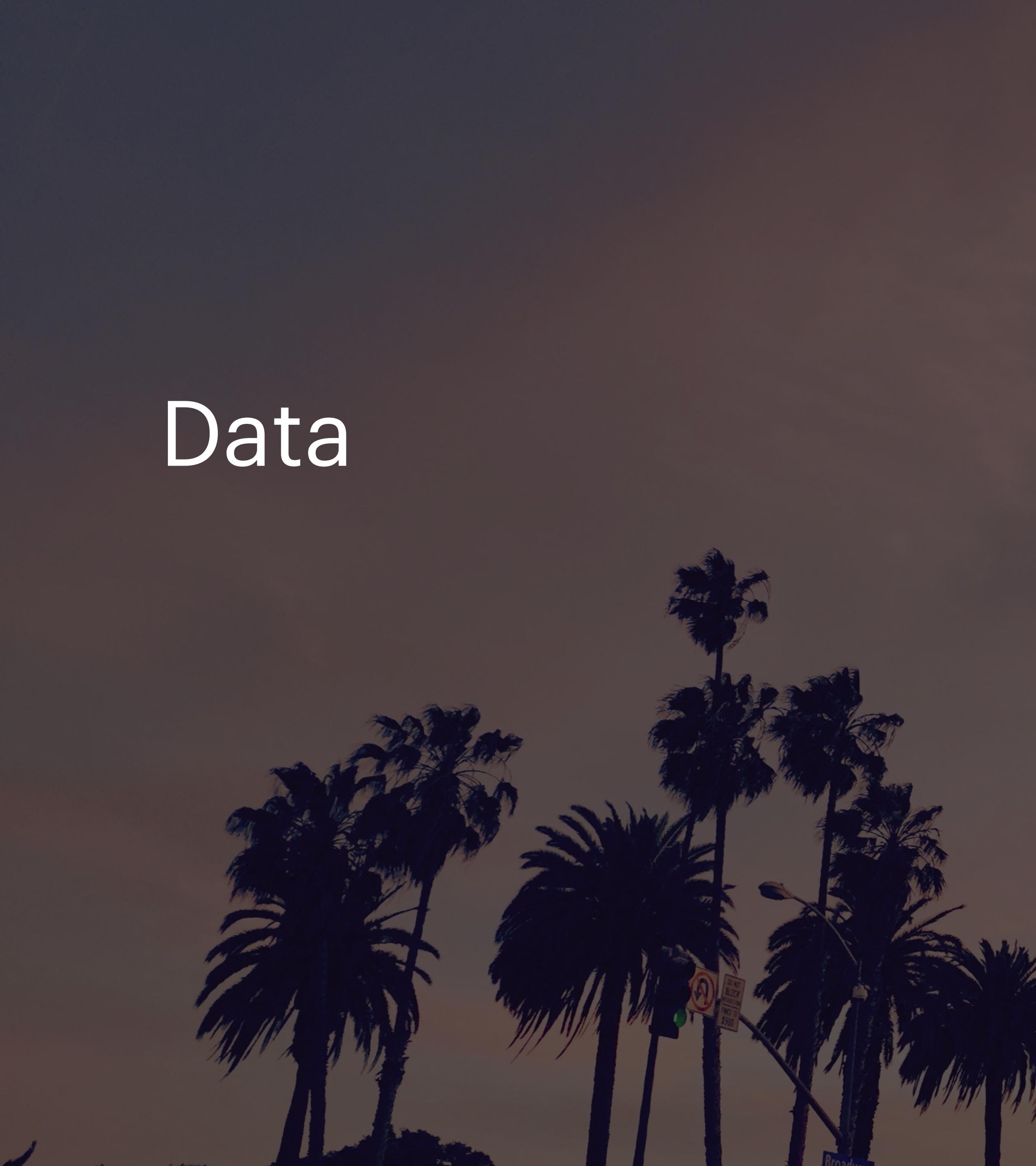
each row an observation

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

# Sourcing





Data



Sourcing

Data :

- Analysis and visualization require quality, formatted data

# Sourcing



Data :

- Federal Reserve

Economic Data  
(FRED)

# Sourcing





# ECONOMIC RESEARCH

FEDERAL RESERVE BANK OF ST. LOUIS

MY ACCOUNT ▾

Search FRED

FRED® Economic Data Information Services Publications Working Papers Economists About

St. Louis Fed Home

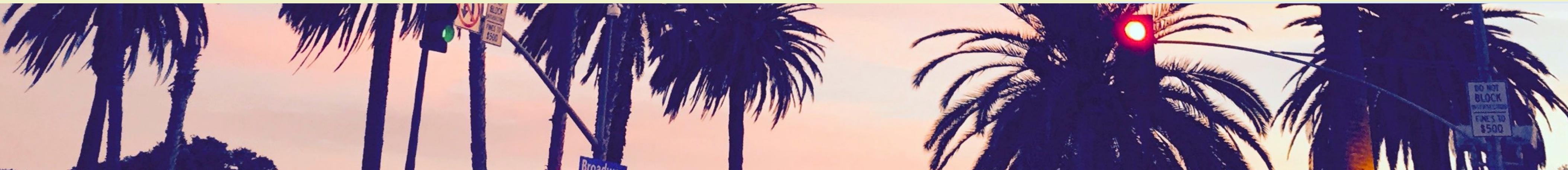
Your trusted data source since 1991.

Download, graph, and track **766,000** US and international time series from **101** sources.

Search FRED data e.g., gdp, inflation, unemployment



Browse data by **Tag**, **Category**, **Release**, **Source**, **Release Calendar** or **Get Help**





Categories &gt; Money, Banking, &amp; Finance &gt; Interest Rates &gt; Treasury Constant Maturity

## ★ 10-Year Treasury Constant Maturity Rate (DGS10)

DOWNLOAD

**Observation:**  
2020-10-08: **0.78** (+ more)  
Updated: Oct 9, 2020

**Units:**  
Percent,  
Not Seasonally Adjusted

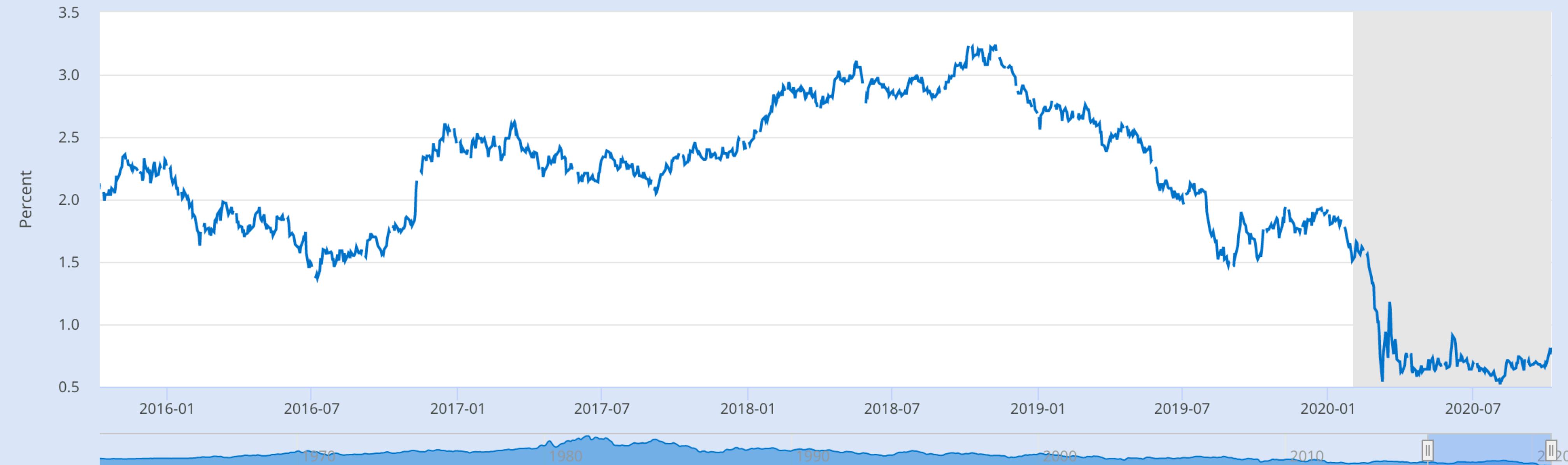
**Frequency:**  
Daily

1Y | 5Y | 10Y | Max

2015-10-08 to 2020-10-08

EDIT GRAPH

FRED — 10-Year Treasury Constant Maturity Rate



Shading indicates U.S. recessions; the most recent one is ongoing.

Source: Board of Governors of the Federal Reserve System (US)

fred.stlouisfed.org

FREDR



Sourcing



FREDR :

- Open source

# Sourcing





⚠ Not Secure

sboysel.github.io/fredr/articles/fredr.html

fredr 1.1.0.9000



Get started

Reference

Articles ▾

Changelog

# Getting started with fredr

Sam Boysel

2020-05-19

Source: vignettes/fredr.Rmd

---

## Introduction

`fredr` provides a complete set of `R` bindings to the [Federal Reserve Economic Data \(FRED\)](#) RESTful API, provided by the Federal Reserve Bank of St. Louis. The functions allow the user to search for and fetch time series observations as well as associated metadata within the FRED database. The core functions are

- `fredr_set_key()` - Set the required FRED API key for the session.
- `fredr()` or `fredr_series_observations()` - Fetch a FRED series.
- `fredr_series_search_text()` - Search for a FRED series by text.
- `fredr_request()` - Send a general request to the FRED API.

Objects are returned as `tibbles`. The user is strongly encouraged to read the full [FRED API](#) documentation to leverage the full power of `fredr` and the FRED API.

FREDR :

- Series ID
- Date range
- Frequency
- Units

# Sourcing





fred.stlouisfed.org

ECONOMIC RESEARCH  
FEDERAL RESERVE BANK OF ST. LOUIS

MY ACCOUNT ▾

Search FRED

FRED® Economic Data Information Services Publications Working Papers Economists About

St. Louis Fed Home

Categories &gt; Money, Banking, &amp; Finance &gt; Interest Rates &gt; Treasury Constant Maturity

## ★ 10-Year Treasury Constant Maturity Rate (DGS10)

DOWNLOAD

**Observation:**  
2020-10-08: **0.78** (+ more)  
Updated: Oct 9, 2020

**Units:**  
Percent,  
Not Seasonally Adjusted

**Frequency:**  
Daily

1Y | 5Y | 10Y | Max

2015-10-08 to 2020-10-08

EDIT GRAPH

FRED — 10-Year Treasury Constant Maturity Rate



Shading indicates U.S. recessions; the most recent one is ongoing.

Source: Board of Governors of the Federal Reserve System (US)

fred.stlouisfed.org

Data :

- Census Bureau

# Sourcing



# IPUMS

IPUMS  
LOG IN TO IPUMS USA

## LOG IN TO IPUMS USA

### EMAIL ADDRESS

Enter email

### PASSWORD

Enter password

SIGN IN

IPUMS uses a **common user management system** for several data projects: IPUMS International, IPUMS USA, IPUMS CPS, IPUMS Higher Ed, NHGIS, IPUMS Health Surveys, IPUMS Time Use, and IPUMS Terra.

If you already have an account with any of these projects, sign in [here](#). If you've not yet registered for IPUMS USA, you will be directed to the project registration page after signing in.

If you do not have an account for any IPUMS project, you may create one [here](#).

### HELP

› [Forgot your password?](#)

› [Create an account](#) – *only for users who do not have an IPUMS account*



# Geography in tidycensus

To get decennial Census data or American Community Survey data, tidycensus users supply an argument to the required `geography` parameter. Arguments are formatted as consumed by the Census API, and specified in the table below. Not all geographies are available for all surveys, all years, and all variables. Most Census geographies are supported in tidycensus at the moment; if you require a geography that is missing from the table below, please file an issue at <https://github.com/walkerke/tidycensus/issues>.

If **state** or **county** is in bold face in “Available by”, you are required to supply a state and/or county for the given geography.

Geography	Definition	Available by	Available in
"us"	United States		<code>get_acs()</code> , <code>get_decennial()</code>
"region"	Census region		<code>get_acs()</code> , <code>get_decennial()</code>
"division"	Census division		<code>get_acs()</code> , <code>get_decennial()</code>
"state"	State or equivalent	state	<code>get_acs()</code> , <code>get_decennial()</code>
"county"	County or equivalent	state, county	<code>get_acs()</code> , <code>get_decennial()</code>
"county subdivision"	County subdivision	<b>state</b> , county	<code>get_acs()</code> , <code>get_decennial()</code>
"tract"	Census tract	<b>state</b> , county	<code>get_acs()</code> , <code>get_decennial()</code>
"block group"	Census block group	<b>state</b> , county	<code>get_acs()</code> , <code>get_decennial()</code>

## TIDYCENSUS :

- Geography
- Variables
- Year
- Sample

# Sourcing





# Script

```

1 --
2 title: "R Tutorial"
3 author: "Mattingly"
4 date: "2/10/2020"
5 output: pdf_document
6 ---
7
8 getwd()
9 setwd("/Users/petermattingly/Desktop/")
10
11 ## creating a notebook chunk
12 'control' + 'option', then
13
14 ``{r}
15
16 ``
17
18 ## running individual lines of code
19 # mac: 'command' then 'return'
20 # pc: 'control' then 'enter'
21
22 ## assignment operator <-
23
24
25 ## creating pipe operator %>%
26 'command' 'shift' 'm' =
27
28
29 ## libraries and packages
30
31 ``{r}
32 install.packages('data.table', 'tidyverse')
33 library(data.table)
34 library(tidyverse)

```

11:30 # creating a notebook chunk

```

Console Terminal R Markdown
~/
+   xlab=TeX("3 Month Yields"), ylab=TeX("10 Year Yields"),
+   main="Daily Interest Rates Since 2000", pch=16, col='blue')
Error in (function (formula, data = NULL, subset = NULL, na.action = na.fail, :
  invalid type (list) for variable 'strptime(threemonth$value, "%Y-%m-%d")'
> plot(strptime(threemonth$value, "%Y-%m-%d"), strptime(tenyear$value, "%Y-%m-%d"),
+   xlab=TeX("3 Month Yields"), ylab=TeX("10 Year Yields"),
+   main="Daily Interest Rates Since 2000", pch=16, col='blue')
Error in plot.window(...) : need finite 'xlim' values
In addition: Warning messages:
1: In min(x) : no non-missing arguments to min; returning Inf
2: In max(x) : no non-missing arguments to max; returning -Inf
3: In min(x) : no non-missing arguments to min; returning Inf
4: In max(x) : no non-missing arguments to max; returning -Inf
> plot(threemonth$value, tenyear$value,
+   xlab=TeX("3 Month Yields"), ylab=TeX("10 Year Yields"),
+   main="Daily Interest Rates Since 2000", pch=16, col='blue')
> cor(tenyear$value ~ threemonth$value)
Error in cor(tenyear$value ~ threemonth$value) :
  supply both 'x' and 'y' or a matrix-like 'x'
> cor(tenyear$value, threemonth$value)
[1] 0.7608
> threemonth = drop_na(fredr(series_id = "DGS3M0", observation_start = as.Date("2000-01-01")))
> tenyear = drop_na(fredr(series_id = "DGS10", observation_start = as.Date("2000-01-01")))
> plot(threemonth$value, tenyear$value,
+   xlab=TeX("3 Month Yields"), ylab=TeX("10 Year Yields"),
+   main="Daily Interest Rates Since 2000", pch=16, col='blue')

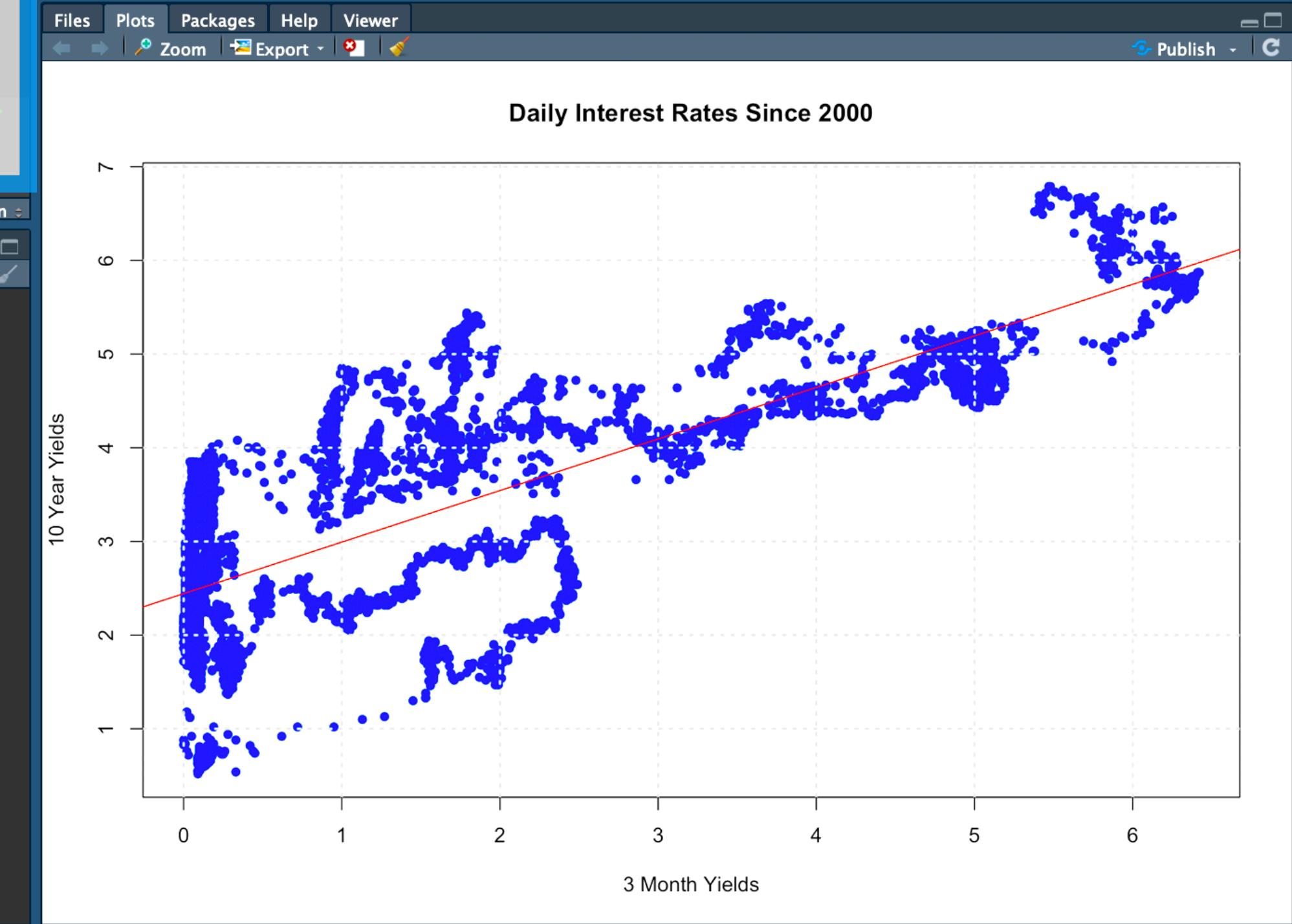
```

Environment History Connections

Import Dataset Grid C

Global Environment

Name	Type	Length	Size	Value
dailyavg_table	tbl_df	7	2 KB	3 obs. of 7 variables
dailyavg_wtmeans	grouped_df	4	66.4 KB	1095 obs. of 4 variables
data1990	tbl_df	6	22 KB	373 obs. of 6 variables
data1990_2018_race_total	data.frame	5	8.7 KB	174 obs. of 5 variables
data1990_hisp	tbl_df	6	7 KB	62 obs. of 6 variables
data1990_main	tbl_df	6	19.1 KB	311 obs. of 6 variables
data1999_2000	grouped_df	5	4.4 KB	12 obs. of 5 variables
data1999_2000_total	data.frame	5	4.3 KB	66 obs. of 5 variables
data1999_2018_race_total	matrix	10	7.9 KB	List of 10
data1999_2018_total	data.frame	5	8.6 KB	174 obs. of 5 variables
f1	function	1	10.1 KB	function (x, y, p = 0)
geo_northern	data.table	9	30.6 KB	97 obs. of 9 variables
geospatial	data.table	9	73.7 KB	246 obs. of 9 variables
il	sf	6	1.4 MB	408 obs. of 6 variables
labTheme	function	1	18 KB	function (base_size = 48)
logo	rastergrob	12	1.8 MB	Large rastergrob (12 elements, 1.8 Mb)
model1	lm	12	1.3 MB	Large lm (12 elements, 1.3 Mb)
monthlyavg_countries	grouped_df	7	47 KB	730 obs. of 7 variables
name_region	data.table	5	38.5 KB	246 obs. of 5 variables
numbers	integer	10	96 B	int [1:10] 1 2 3 4 5 6 7 8 9 10
numlist	numeric	10	176 B	num [1:10] 1 2 3 4 5 6 7 8 9 10
open_daily_graph	gg	9	24.7 KB	List of 9



RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

RStudio

R Tutorial.Rmd x R Tutorial Part 2.Rmd x REDA.Rmd x GreatRecession.Rmd x

ABC Knit Insert Run

1  
2  
3  
4  
5  
6  
7  
8 **### creating a notebook chunk**  
9 # on a mac: 'control' + 'option', then 'i'  
10 # on a pc: 'control' + 'alt', then 'i'  
11  
12 **```{r}**  
13 install.packages(c("tidyverse", "devtools", "tidycensus"))  
14 **```**

Error in install.packages : Updating loaded packages

15  
16 **```{r}**  
17 library(tidyverse)  
18 library(devtools)  
19 library(tidycensus)  
20 **```**

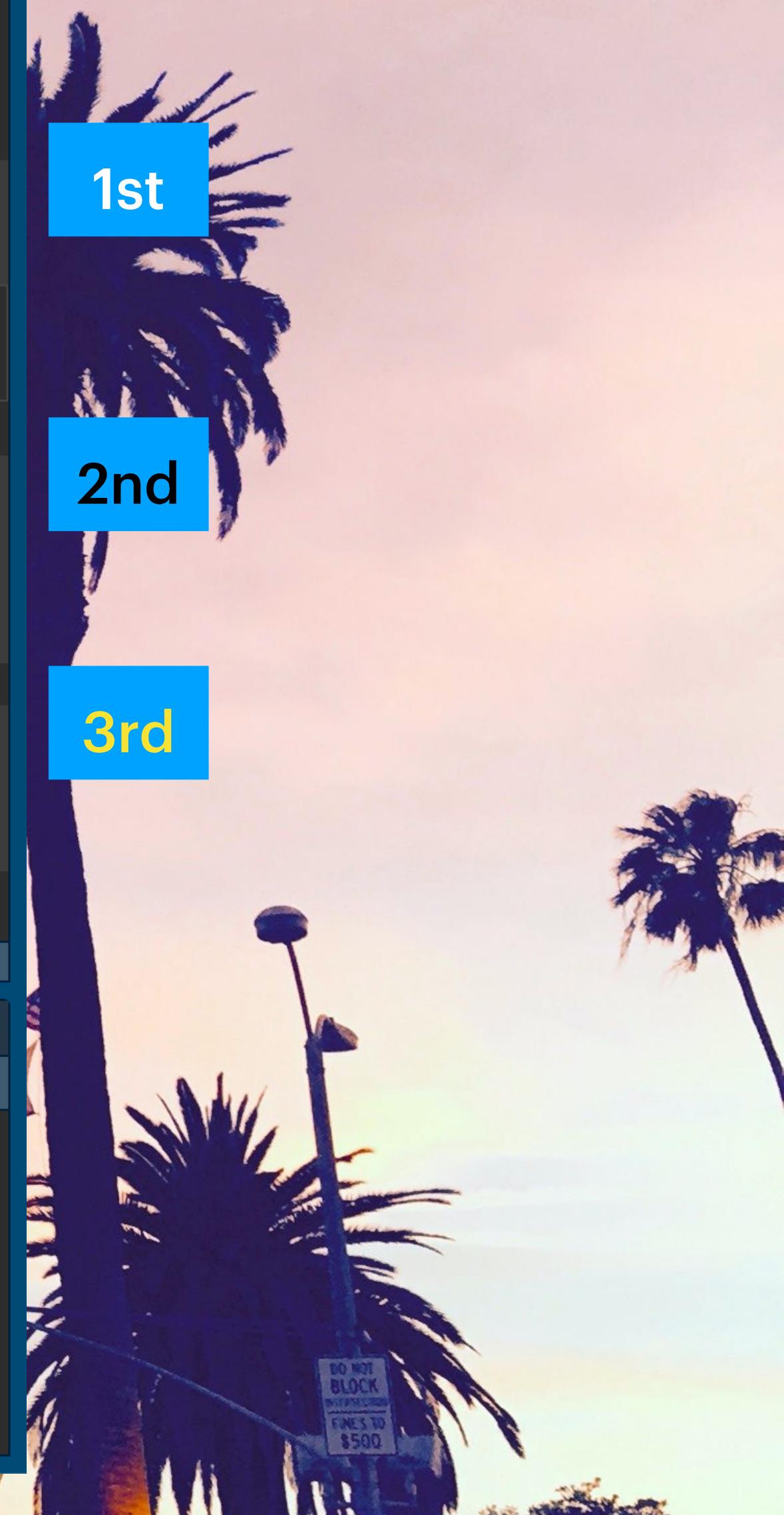
21  
22 **```{r}**  
23 devtools::install\_github("sboysel/fredr")  
24 library(fredr)  
25 **```**

26  
27

25:1 C Chunk 34 R Markdown

Console Terminal x R Markdown x

```
~/>  
> library(tidyverse)  
> library(devtools)  
> library(tidycensus)  
> devtools::install_github("sboysel/fredr")  
Skipping install of 'fredr' from a github remote, the SHA1 (97b244ed) has not changed since last install.  
Use `force = TRUE` to force installation  
> library(fredr)  
>
```



RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

RStudio

+ | Go to file/function | Addins

R Tutorial.Rmd R Tutorial Part 2.Rmd REDA.Rmd GreatRecession.Rmd

ABC Knit Insert Run

```
18 library(devtools)
19 library(tidyCensus)
20 ``
21 ``
22 ``{r}
23 devtools::install_github("sboysel/fredR")
24 library(fredR)
25 ``
26 
27 **FRED API KEY**
28 
29 ``{r}
30 fredr_set_key('YOUR API KEY HERE')
31 ``
32 
33 **CENSUS API KEY**
34 
35 ``{r}
36 census_api_key('YOUR API KEY HERE')
37 ````
```

1:1 # REDA R Markdown



