

Data Analytics

Lecture Series: Part 2

Overview



Overview

In this section, we will:



Overview

In this section, we will:

- Navigate R Studio and R Markdown



Overview

In this section, we will:

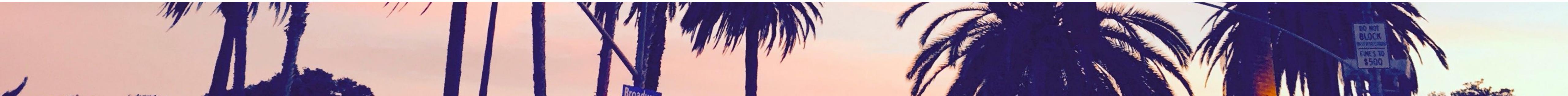
- Navigate R Studio and R Markdown
- Source and wrangle real estate as well as population data







HORST'19



Setup



Setup



Directory



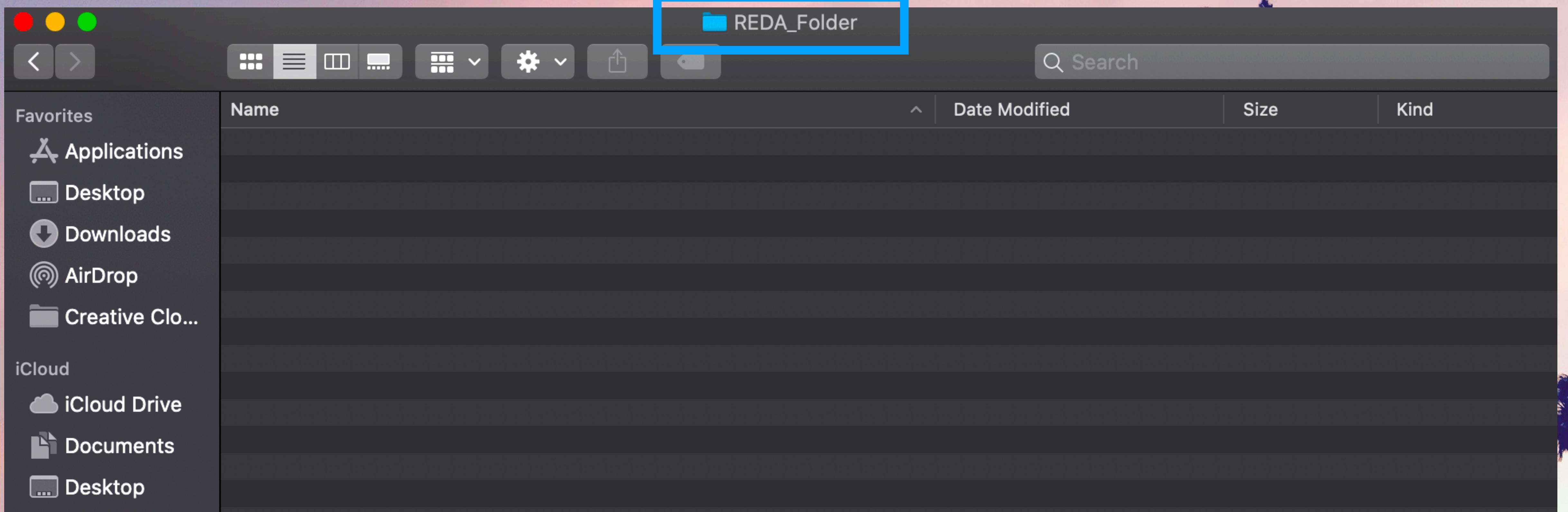
Setup



Setup

Directory :

- Create a desktop folder named **“REDA_Folder”** ... or whatever you like



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

- New File
- New Project...
- Open File... ⌘O
- Reopen with Encoding...
- Recent Files ▶
- Open Project...
- Open Project in New Session...
- Recent Projects ▶
- Import Dataset
- Save ⌘S
- Save As...
- Save with Encoding...
- Save All ⌘S
- Knit Document ⌘K
- Publish...
- Print...

- R Script ⌘N
- R Notebook
- R Markdown... ▶
- Shiny Web App...
- Text File
- C++ File
- R Sweave
- R HTML
- R Presentation
- R Documentation

editor.Rmd* outlook_housing

Insert | Run |

	Environment	History	Connections
	Global Environment	Import Dataset	
■ Name			
■ dailyavg_table			
■ dailyavg_wtmeans			
■ data1990			
■ data1990_2018_race_total			
■ data1990_hisp			
■ data1990_main			
■ data1999_2000			
■ data1999_2000_total			
■ data1999_2018_race_total			
■ data1999_2018_total			
■ f1			
■ geo_northern			
■ geospatial			
■ il			
■ labTheme			
■ logo			
■ model1			
■ monthlyavg_countries			
■ name_region			
■ numbers			

Setup

Directory :

- Connect R to folder by setting the working directory

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

New Session

Interrupt R
Terminate R...

Restart R ⌘⌘F10
Restart R and Clear Output
Restart R and Run All Chunks

Set Working Directory ►

To Source File Location
To Files Pane Location

Load Workspace...
Save Workspace As...

Clear Workspace...

Choose Directory... ⌘⌃H

Quit Session...

GreatRecession.Rmd x MEC_0412.Rmd x floodzone_censu

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 'i'
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')

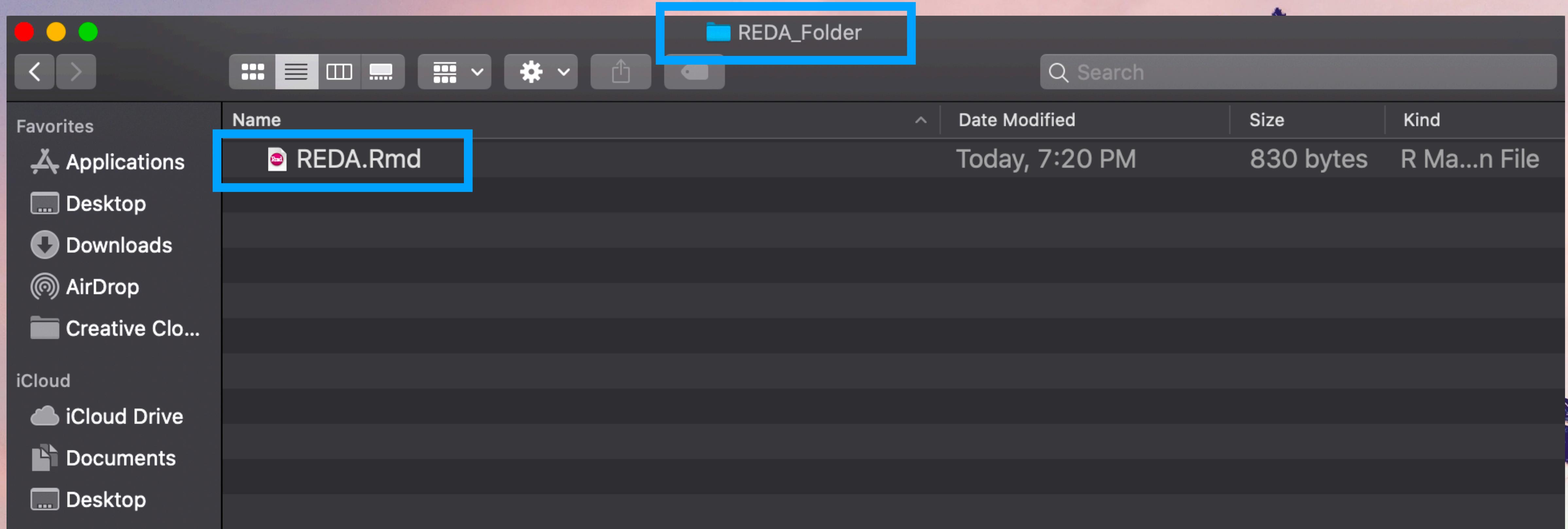
Environment History Connect
Import Dataset
Global Environment
Name
dailyavg_table
dailyavg_wtmeans
data1990
1990_2018_race_total
1990_hisp
1990_main
999_2000
999_2000_total
999_2018_race_total
data1999_2018_total
f1
geo_northern
geospatial
il
labTheme
logo
modell
monthlyavg_countries
name_region
numbers

Files Plots Packages Help
Zoom Export

Setup

Directory :

- Load and save all files and output to this folder



Directory :

- Global options

Setup



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

Install Packages...
Check for Package Updates...

Version Control ►

Shell...
Terminal
Addins ►

Keyboard Shortcuts Help ⌘↑K
Modify Keyboard Shortcuts...

Project Options... ⌘⌘,

Global Options... (highlighted)

GreatRecession.Rmd x MEC_0412.Rmd x floodzone_censustract.Rmd x R Tutorial.Rmd x REDA.Rmd x

ABC Knit Insert Run

```
1 ---  
2 title: "REDA"  
3 author: "Mattingly"  
4 date: "10/11/2020"  
5 output: pdf_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10 ```  
11  
12 ## R Markdown  
13  
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.  
15  
16 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
17  
18 ```{r cars}  
19 summary(cars)  
20 ```  
21  
22 ## Including Plots  
23
```

Type

- tbl_df
- grouped_df
- tbl_df
- data.frame
- tbl_df
- tbl_df
- grouped_df
- data.frame
- matrix
- data.frame
- function
- data.table
- sf
- function
- rastergrob
- lm
- grouped_df
- data.table
- integer

Options

R General

Code

Appearance

Pane Layout

Packages

R Markdown

Sweave

Spelling

Git/SVN

Publishing

Terminal

Default working directory (when not in a project):

~

Browse...

- Re-use idle sessions for project links
- Restore most recently opened project at startup
- Restore previously open source documents at startup
- Restore .RData into workspace at startup

Save workspace to .RData on exit: Never

- Always save history (even when not saving .RData)
- Remove duplicate entries in history
- Show .Last.value in environment listing
- Use debug error handler only when my code contains errors
- Automatically expand tracebacks in error inspector
- Wrap around when navigating to previous/next tab
- Automatically notify me of updates to RStudio

OK

Cancel

Apply

R Markdown :

- Code chunks

Setup



The image shows the RStudio IDE running on a Mac OS X desktop. The background features a scenic sunset with palm trees silhouetted against a pink and orange sky. The RStudio interface includes a menu bar with File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, and Help. A toolbar above the editor has icons for file operations like Open, Save, Print, and Go to file/function. Below the toolbar, five R Markdown files are listed: R Tutorial.Rmd (selected), R Tutorial Part 2.Rmd, REDA.Rmd, and GreatRecession.Rmd. The main code editor window displays the following R code:

```
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 ``{r}
22 devtools::install_github("sboysel/fredr")
23 library(fredr)
24 ``

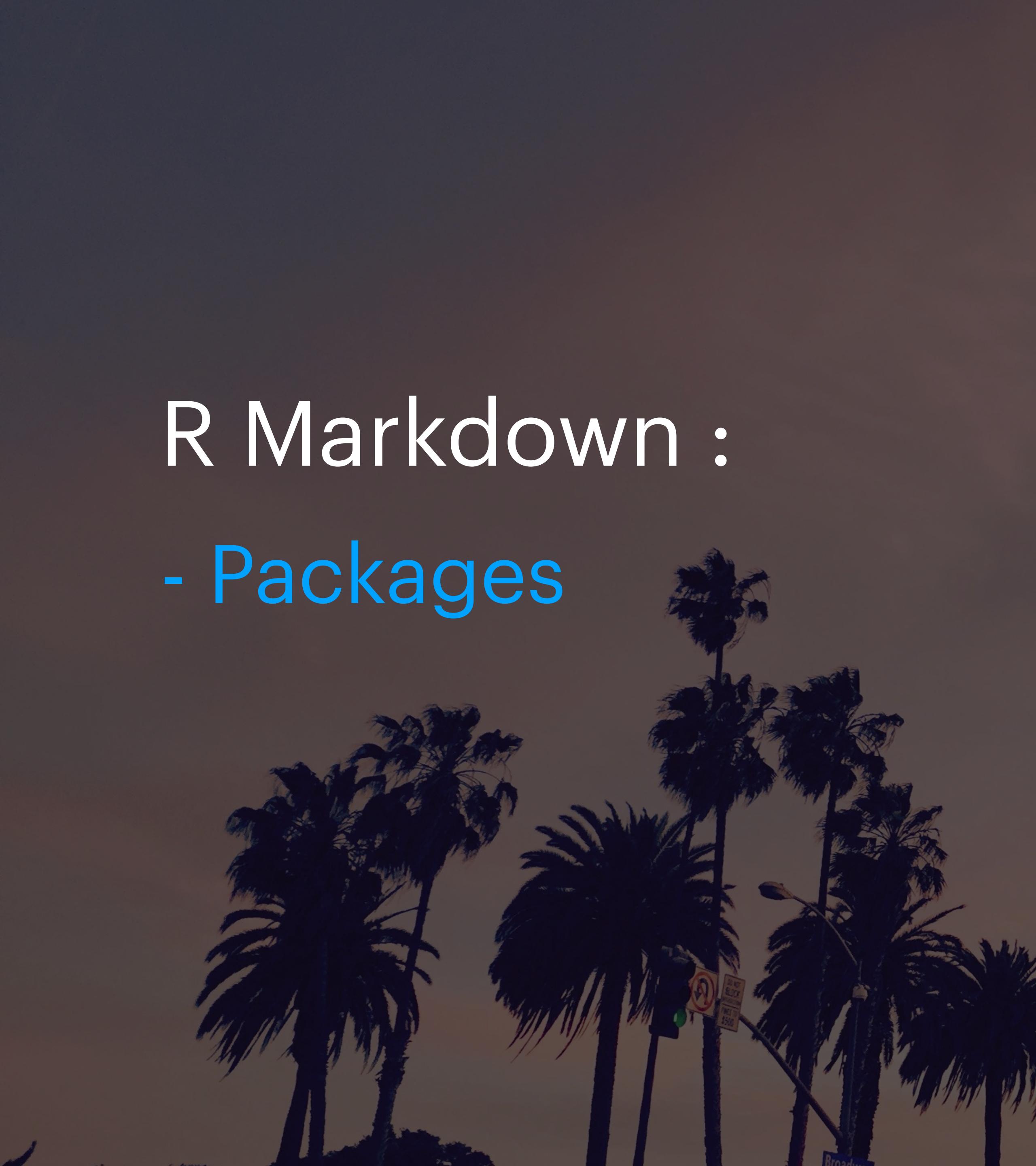
25
26
27
```

The status bar at the bottom indicates "25:1" and "Chunk 34". The R Markdown tab is selected in the bottom right. The Console tab is active, showing the following output:

```
~/
> 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)
>
```

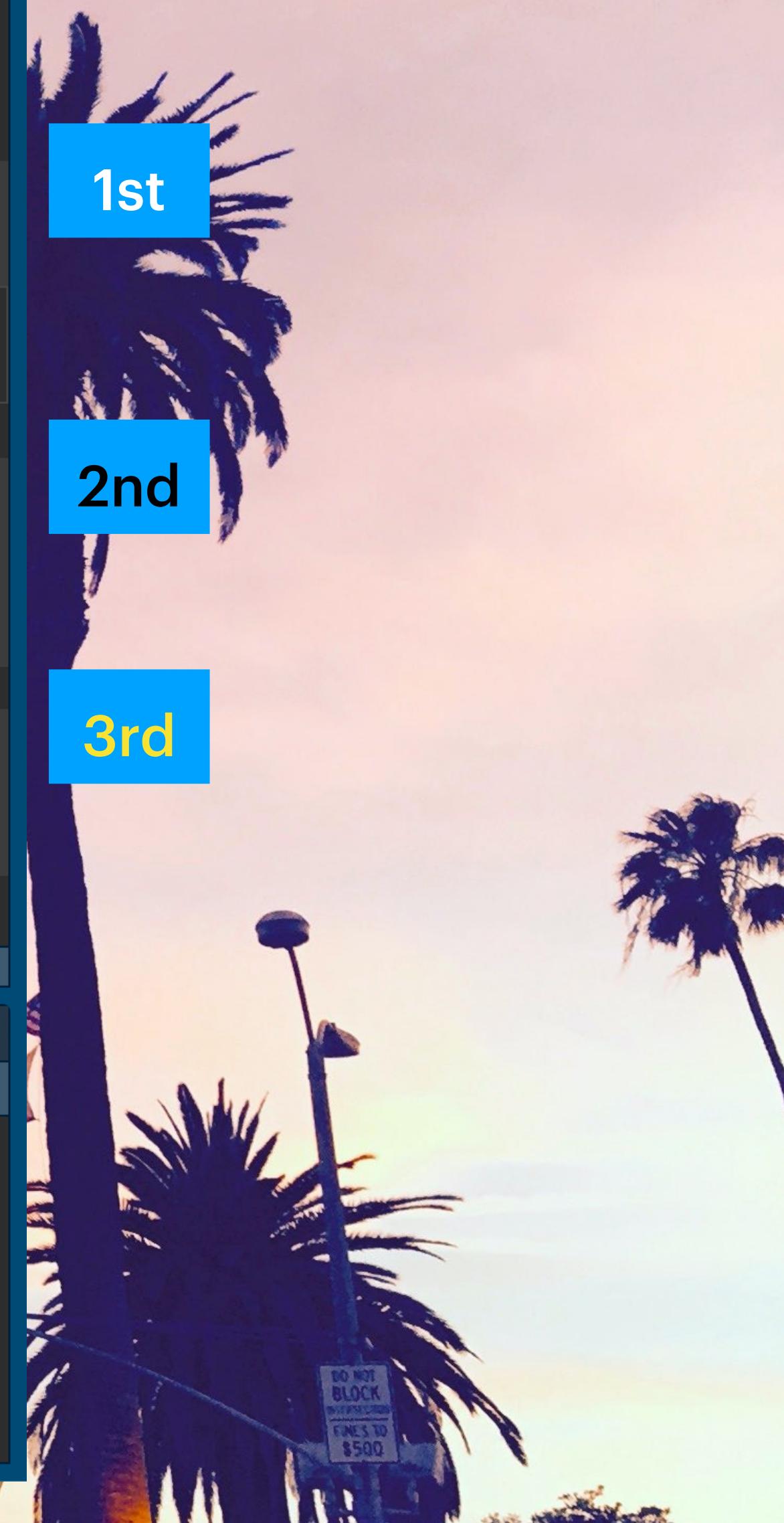
R Markdown :

- Packages



Setup



The code editor has three execution panes below it, each with a blue border and a play button icon. The first pane contains the error message 'Error in install.packages : Updating loaded packages'. The second and third panes contain the library loading code. The status bar at the bottom indicates '25:1' and 'Chunk 34'. The bottom panel shows the 'Console' tab active, displaying the command history and the message 'Skipping install of 'fredr' from a github remote, the SHA1 (97b244ed) has not changed since last install. Use `force = TRUE` to force installation'. The 'Terminal' and 'R Markdown' tabs are also visible.

RStudio

Function |

REDA.Rmd*

Updating Loaded Packages

One or more of the packages that will be updated by this installation are currently loaded. Restarting R prior to updating these packages is strongly recommended.

RStudio can restart R and then automatically continue the installation after restarting (all work and data will be preserved during the restart).

Do you want to restart R prior to installing?

then 'i'
'i'

`devtools", "tidycensus"))`

History Connections
Import Dataset
Environment
Type Length

Type	Length
tbl_df	6
01...	data.frame 5
isp	tbl_df 6
data1990_main	tbl_df 6
data1999_2000	grouped_df 5
data1999_200...	data.frame 5
data1999_201...	matrix 10
data1999_201...	data.frame 5
f1	function 1
geo_northern	data.table 9

Project: (None) ▾

GreatRecession.Rmd MEC_0412.Rmd floodzone_censubtract.Rmd R Tutorial.Rmd outlook_housing.Rmd R Tutorial Par ▾

Knit Insert Run Grid C

```

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 'i'
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)

```

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

Files Plots Packages Help Viewer

Zoom Export C

Daily Interest Rates Since 2000

10 Year Yields

3 Month Yields

19
14:1 C Chunk 34 R Markdown

Console Terminal R Markdown

~/

```
(as LIO is unspecified)
also installing the dependency 'covr'

There are binary versions available but the source versions are later:
  binary source needs_compilation
covr      3.5.0  3.5.1          TRUE
devtools   2.2.2  2.3.2         FALSE
tidycensus 0.9.6 0.10.2        FALSE

Do you want to install from sources the package which needs compilation? (Yes/no/cancel) Yes
```

API Keys :

- fredr
- tidycensus

Setup



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(tidy census)
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

