

# Data Analytics

Lecture Series: Part 1

# Welcome!



# Welcome!

In the course, we will:



# Welcome!

In the course, we will:

- Apply data analytics to real estate using R



# Welcome!

In the course, we will:

- Apply data analytics to real estate using R
- Create reproducible working environments



# Welcome!

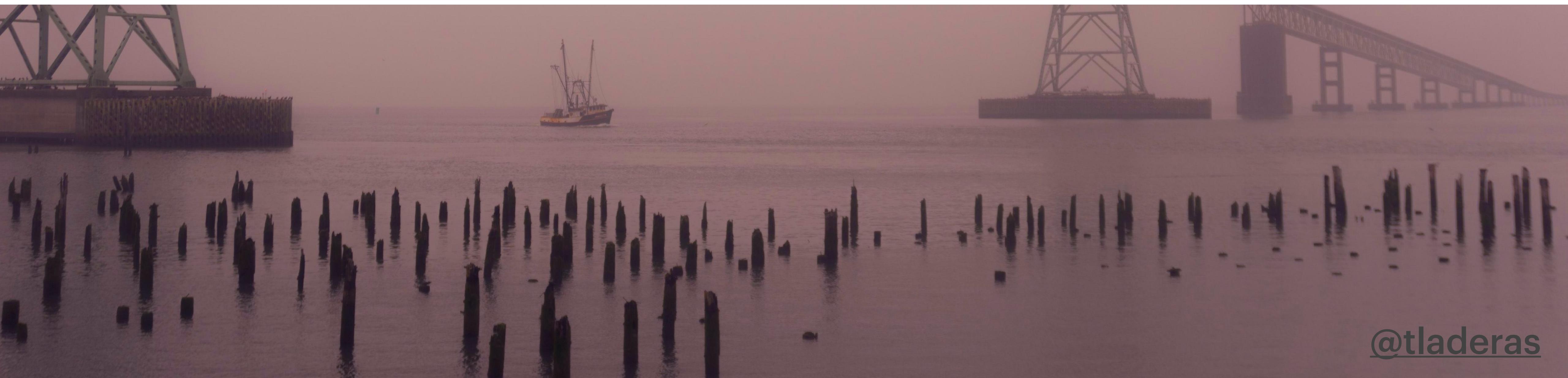
In the course, we will:

- Apply data analytics to real estate using R
- Create reproducible working environments
- Visualize insight and information

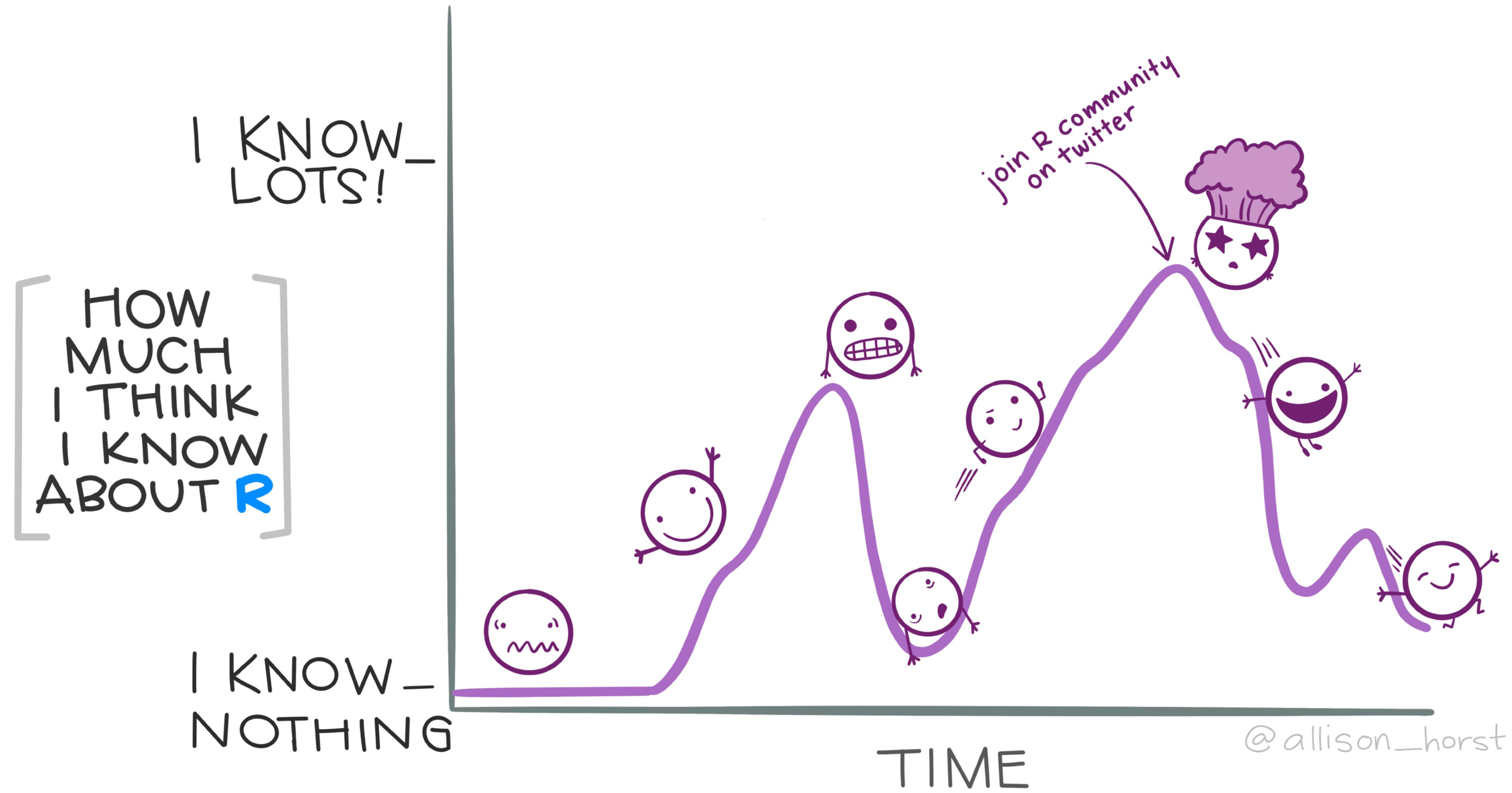




@tladeras



@tladeras



@tladeras

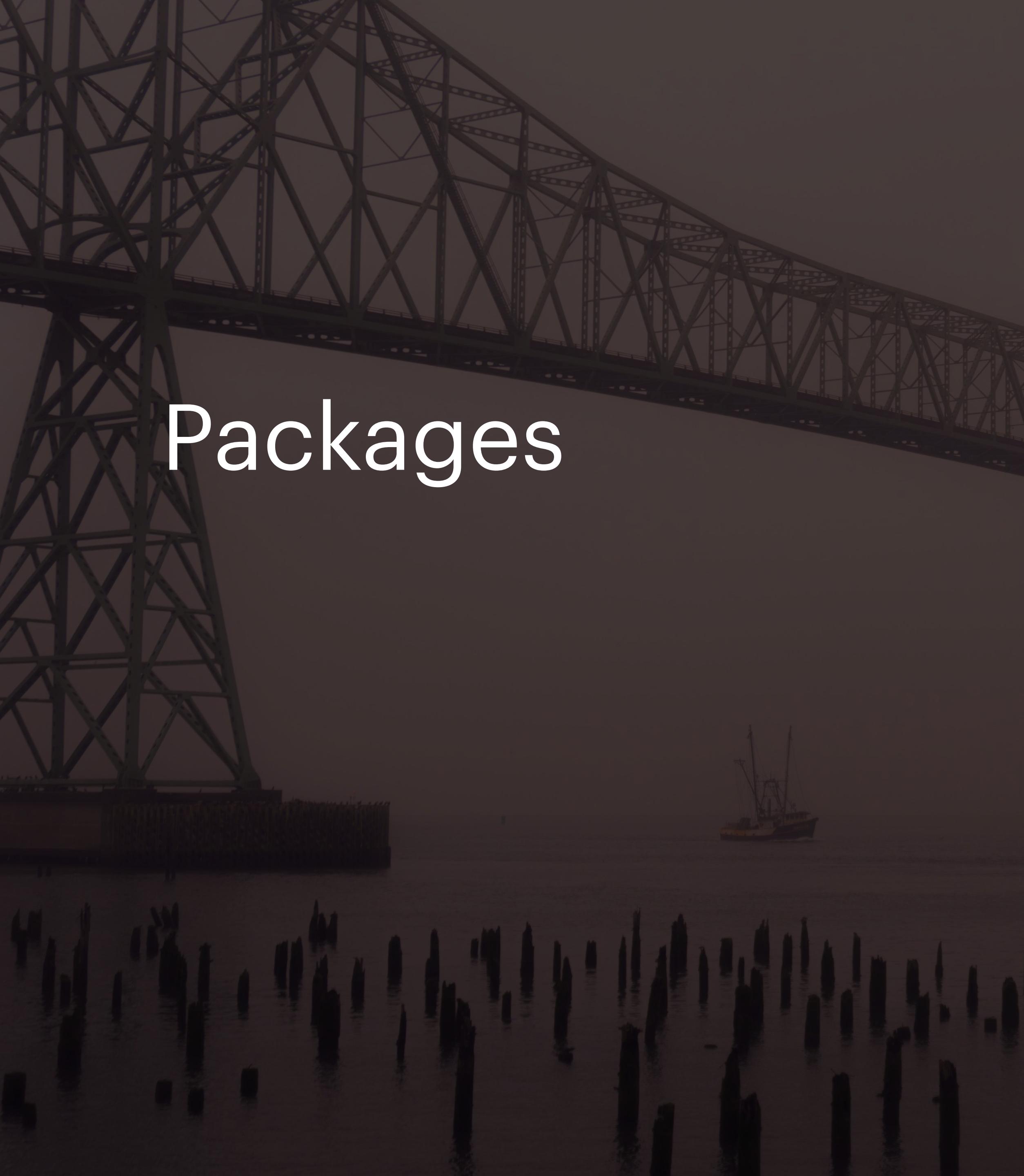


@tladeras

# Terms

# Terms

@tladeras

A large bridge structure, possibly a suspension bridge, spans across a body of water. In the distance, a small sailboat is visible on the water. The sky is overcast.

Packages

# Terms

# Terms

Packages :

- Imported and reusable

# Terms

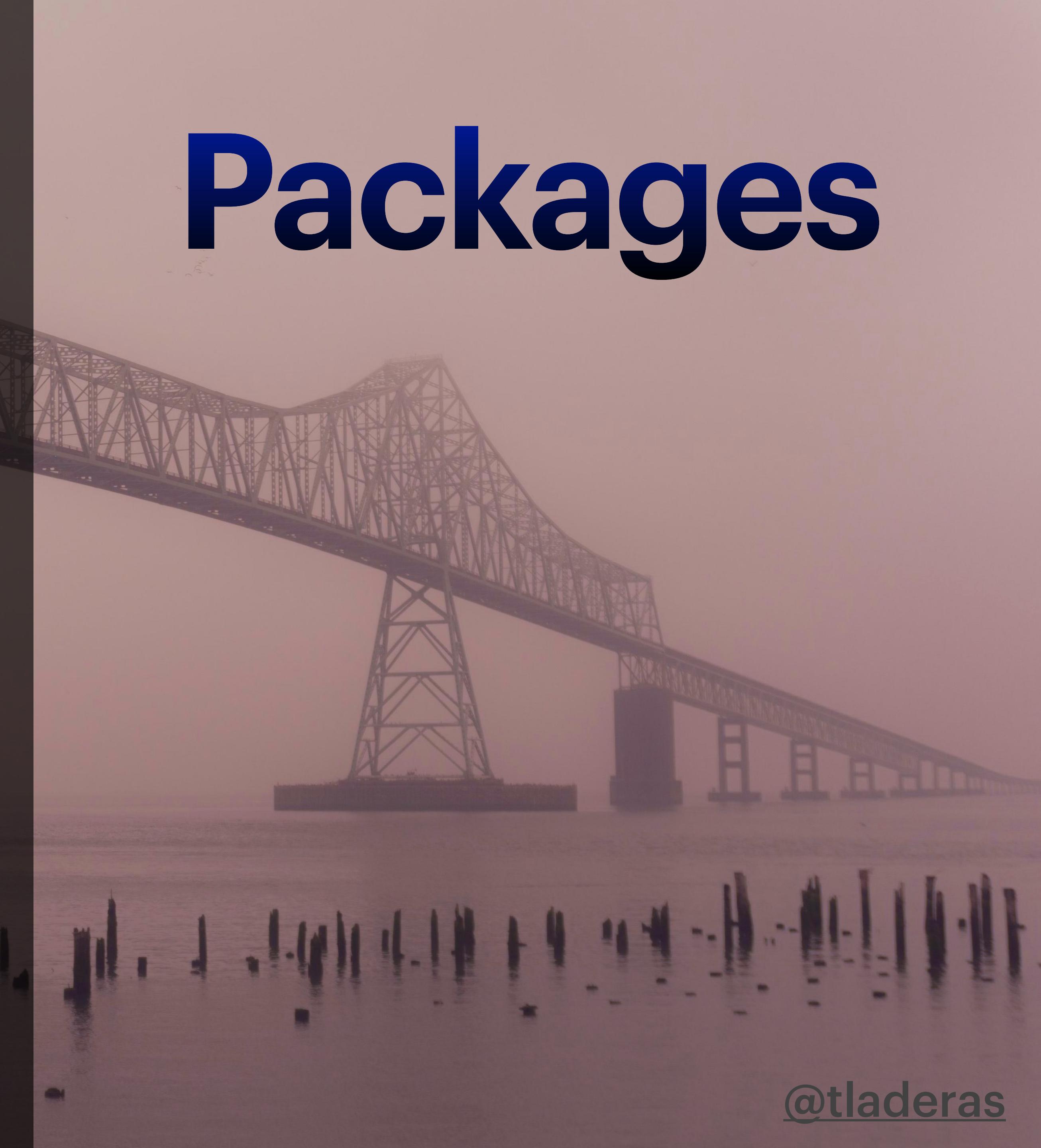
Packages :

- Bundled functions  
for cleaning,  
wrangling, and  
visualizing data



tidyverse

# Packages

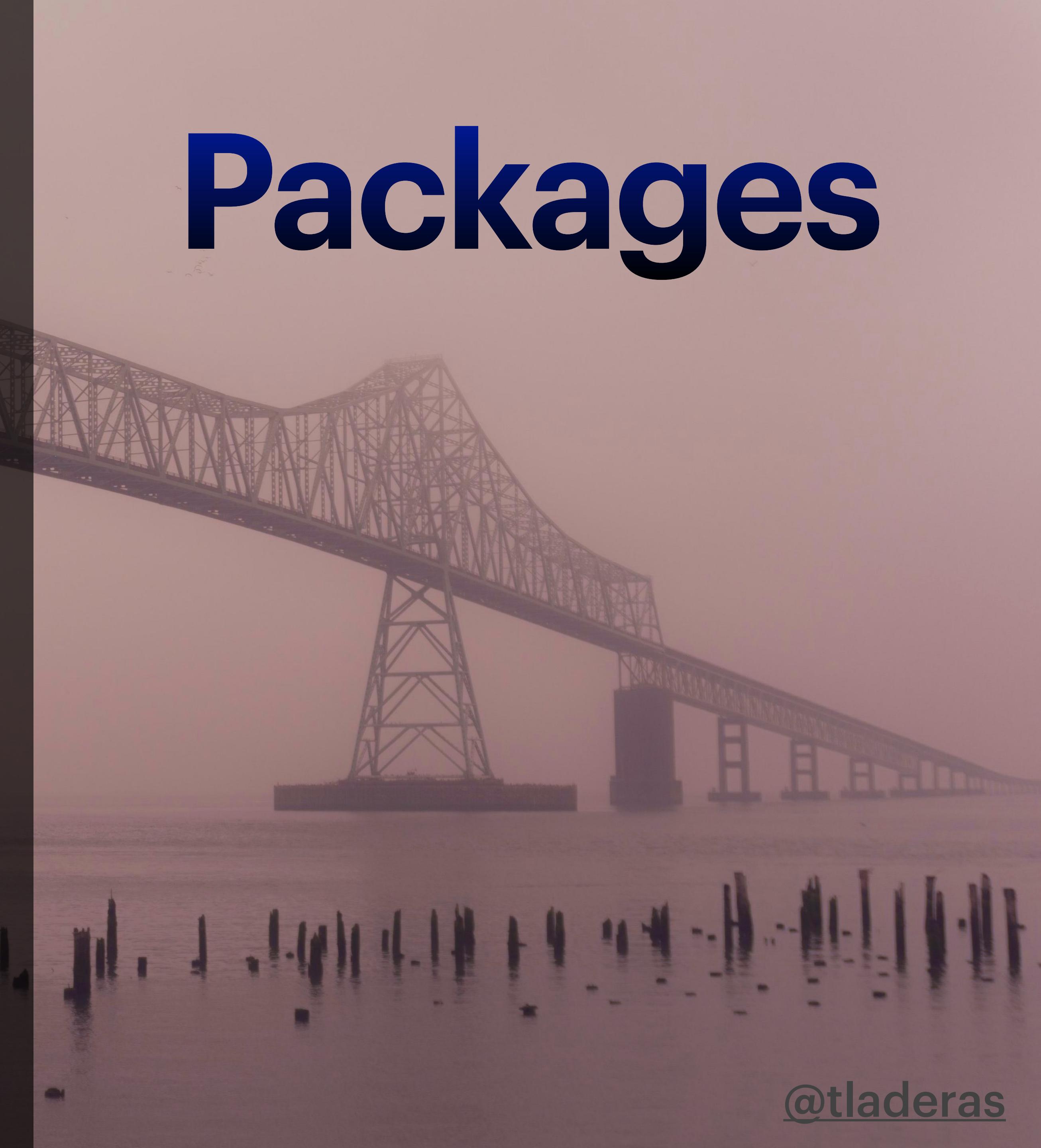


@tladeras



tidyverse  
devtools

# Packages



@tladeras

# Packages

tidyverse

devtools

fredr

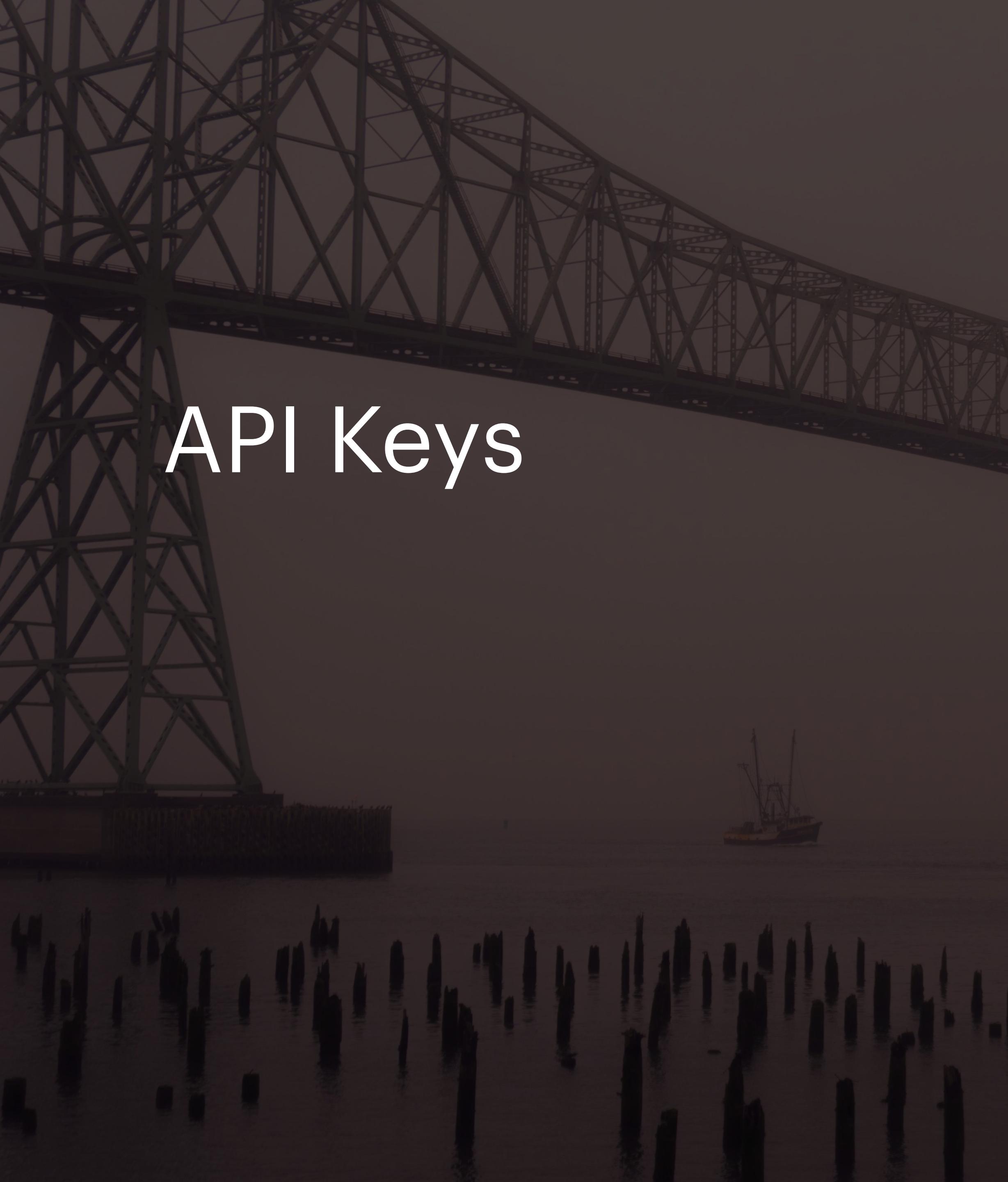
# Packages

tidyverse

devtools

fredr

tidycensus

A large, dark steel truss bridge spans across a body of water. In the background, a small sailboat is visible on the horizon. The foreground shows the dark silhouettes of numerous wooden pilings or posts protruding from the water.

API Keys

# Terms

# Terms

API Keys :

- fredr



Register User Account - St. L... X +

← → ⌂

research.stlouisfed.org/useraccount/login/secure/

Apps

ECONOMIC RESEARCH  
FEDERAL RESERVE BANK *of* ST. LOUIS

Economists ▾ Research and Publications ▾ The Research Division ▾

Already have an account?

[Sign In](#)

[Forgot your password?](#)

Want to create a new account?

[Register](#)

### Why Register?

- Subscribe to email updates for economic data series.
- Create personalized lists of economic data series.
- Save customized graphs and maps for later use.
- Build and share personalized dashboards with series that interest you.
- Access the FRED API to integrate data with your favorite software packages.
- Play FREDcast™.

[Learn more about user accounts](#)

The screenshot shows a web browser window with the title "API Key - St. Louis Fed". The URL "research.stlouisfed.org/useraccount/apikey" is highlighted with an orange box and an arrow pointing to it. The page content includes the St. Louis Fed logo and navigation links for Economists, Research and Publications, and The Research Division. A green box contains the text "Your registered API key is:" followed by a placeholder "YOUR API KEY HERE" which is also highlighted with an orange box and arrow. Below this, a green box contains the text "Documentation is available on the [St. Louis Fed web services website](#)".

API Key

Your registered API key is:

YOUR API KEY HERE

Documentation is available on the [St. Louis Fed web services website](#).

## API Key

**Describe the application or program you intend to write:**

Utilizing the resources and information through FRED and import into R for analysis and visualization.

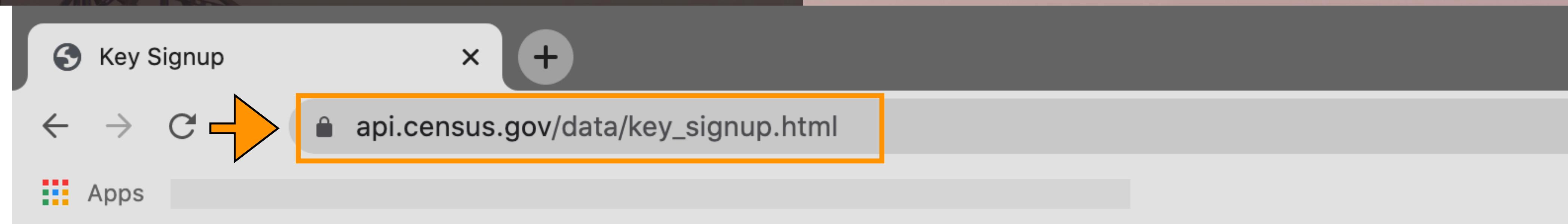
I have read and agree to the St. Louis Fed's [Terms of Use](#), [Privacy Notice & Policy](#), and [Legal Notices](#),

[Request API Key](#)

# Terms

API Keys :

- `tidycensus`



# Request A Key

Organization Name:

Email Address:

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

**Submit Key Request**



R Studio

# Terms

Go to file/funcr Addins Insert Run Knit Project: (None)

```

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)

```

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 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 Publish

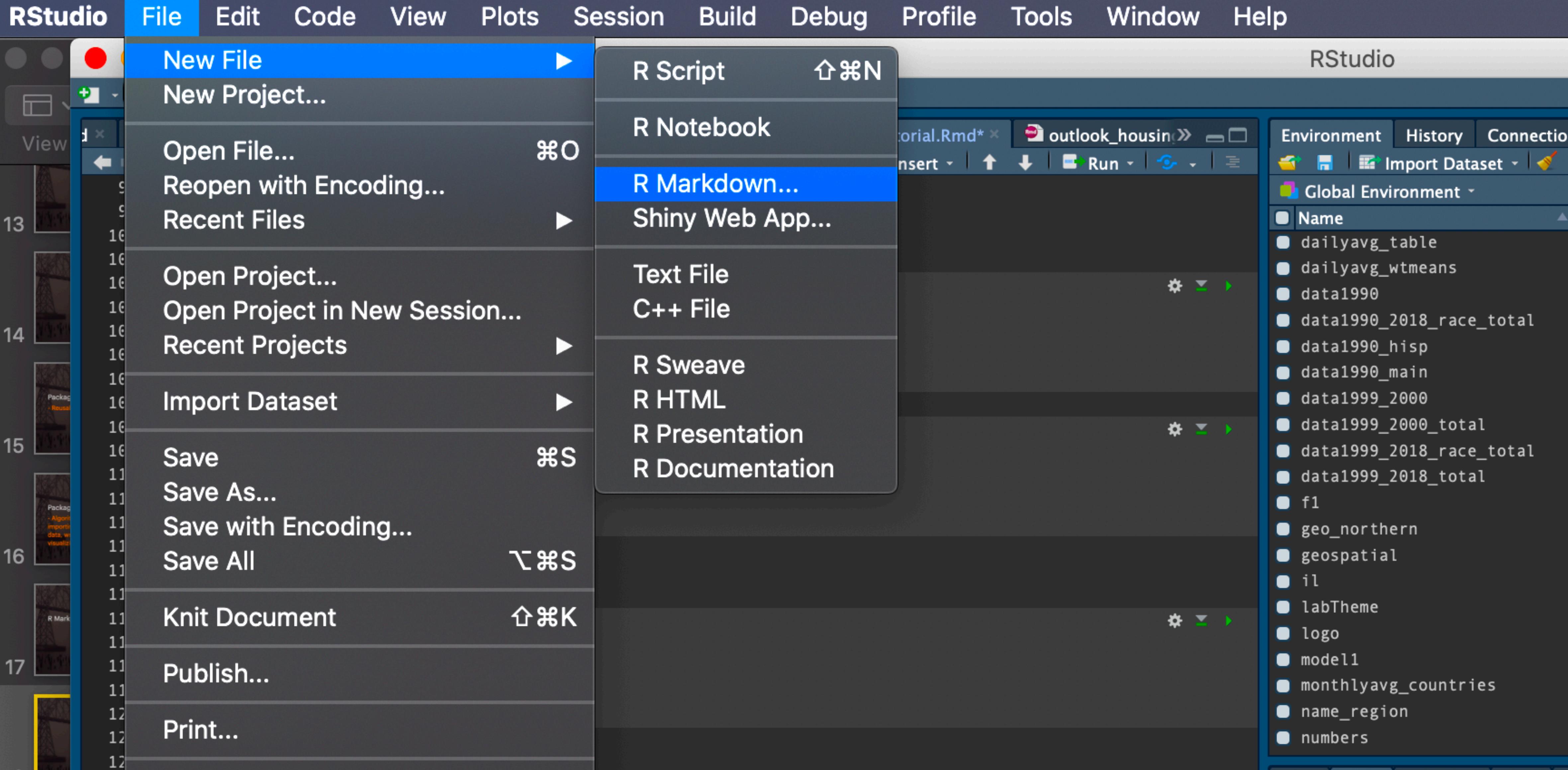
### Daily Interest Rates Since 2000

The scatter plot displays the relationship between '3 Month Yields' (X-axis) and '10 Year Yields' (Y-axis) from 2000 onwards. The X-axis ranges from approximately 0.5 to 6.5, and the Y-axis ranges from 1 to 7. A strong positive linear trend is visible, represented by a red regression line.

# Terms

R Studio :

- R Markdown



# Terms

R Studio :

- 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  
1999\_2000  
1999\_2000\_total  
data1999\_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

# Terms

R Markdown :  
- Code chunks

```
98  ## subsetting
99
100 #### subsetting by value
101
102 ``{r}
103 ### base r
104 setosa <- iris[iris$Species == "setosa",]
105 glimpse(setosa)
106 ```
107
108 ``{r}
109 ### dplyr
110 setosa_tidy <- iris %>% filter(Species = "setosa")
111 glimpse(setosa_tidy)
112 ```
113
114 #### subsetting by columns
115
116 ``{r}
117 ### base r
118 iris_length <- iris[, c(1,3,5,9)]
119 glimpse(iris_length)
120 ```
121
122
123 ``{r}
124 ### dplyr
125 iris_length_dplyr <- iris %>% dplyr::select(matches("(Length|Species)"))
126 glimpse(iris_length_dplyr)
127 ```
128
```

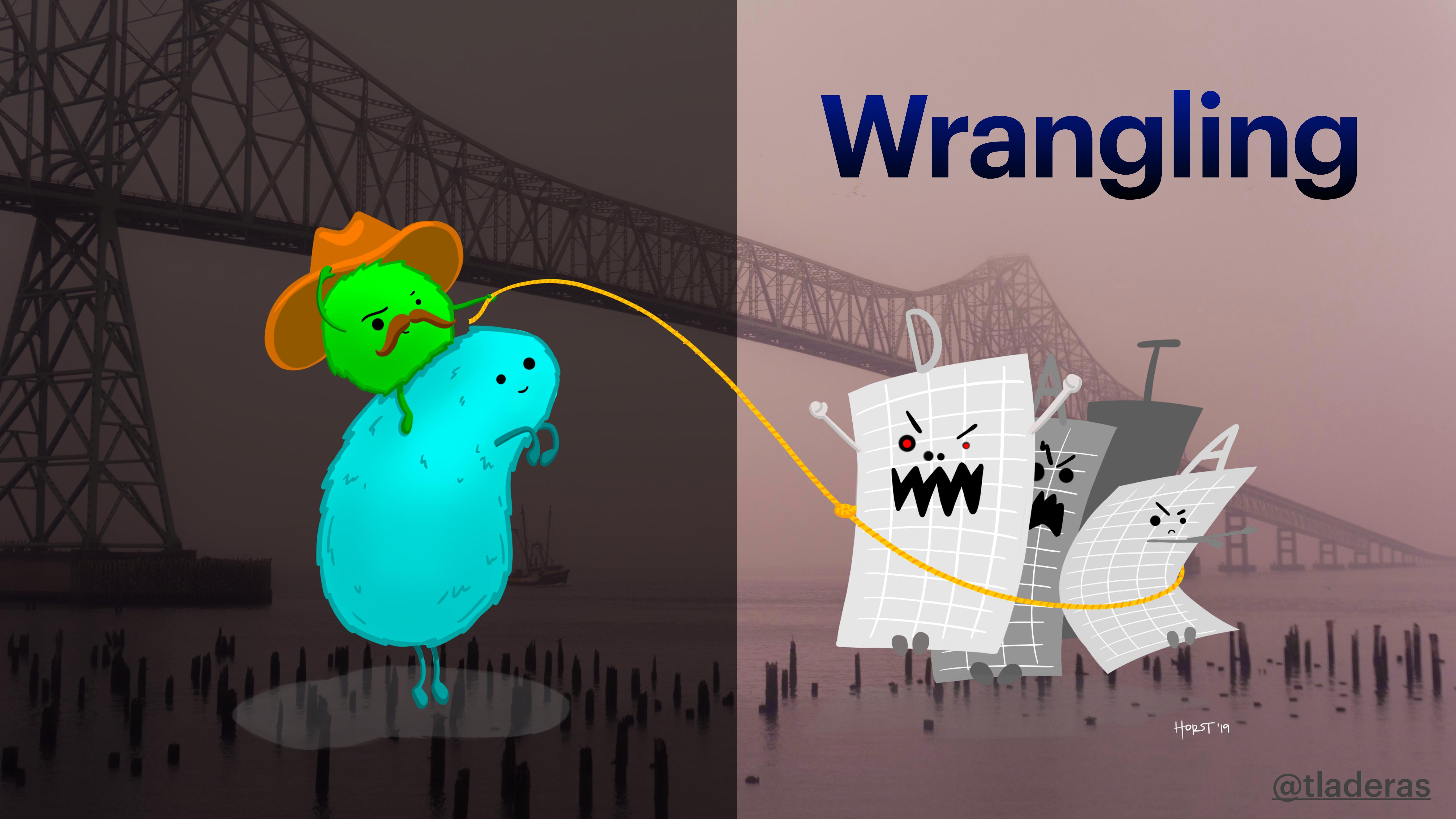


@tladeras



@tladeras

# Wrangling



@tladeras

A large bridge structure, likely a suspension bridge, spans across a body of water. In the foreground, the silhouettes of many wooden pilings are reflected in the water. A small boat with two masts is visible in the distance under a hazy sky.

Data wrangling

# Terms

# Terms

Data wrangling :

- Reshaping by lengthening or widening data

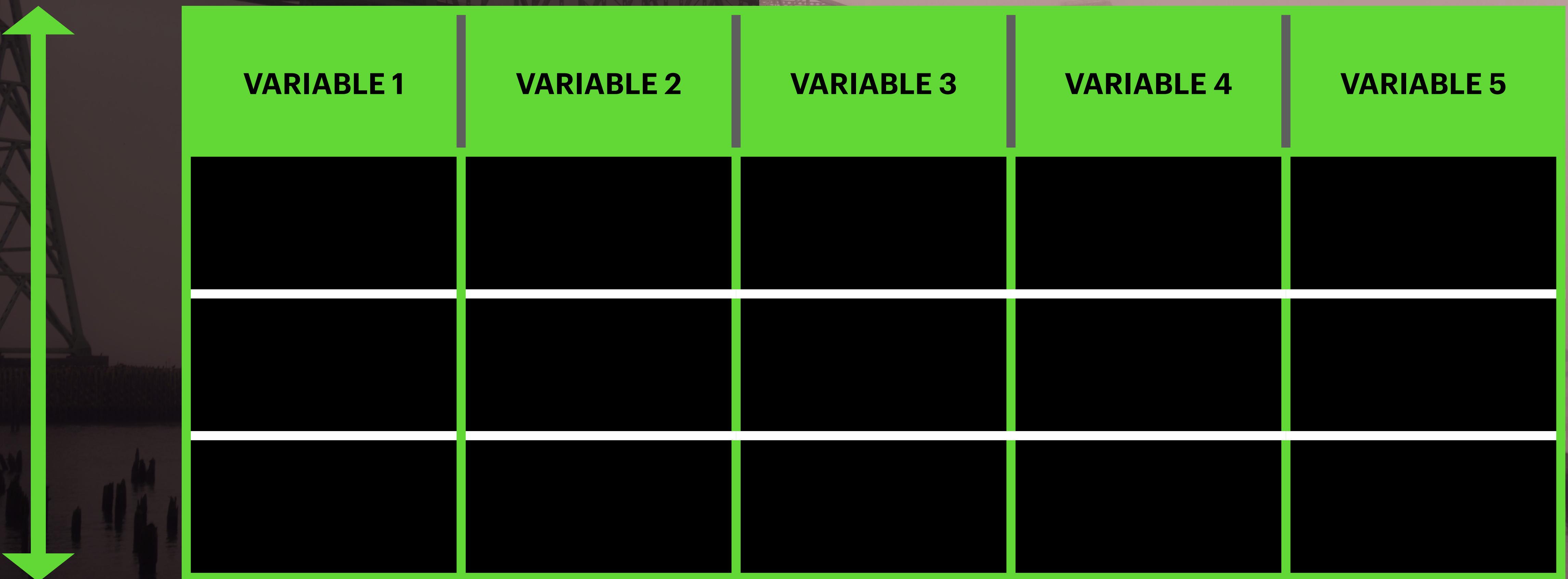
# Data

@tladeras

# Data

VARIABLE 1	VARIABLE 2	VARIABLE 3	VARIABLE 4	VARIABLE 5

# Columns



# ROWS

	VARIABLE 1	VARIABLE 2	VARIABLE 3	VARIABLE 4	VARIABLE 5
VARIABLE 1					
VARIABLE 2					
VARIABLE 3					
VARIABLE 4					
VARIABLE 5					

# Terms

Data wrangling :

- “Gathering” or lengthening with more observations/rows

# Original

Var 1	Var 2	Date 1	Date 2	Date 3

@tladeras

# Original

Var 1	Var 2	Date 1	Date 2	Date 3



# Gathering

Var 1	Var 2	Date	Value
		1	
		2	
		3	

# Terms

Data wrangling :

- “Spreading” or widening with more columns/variables

# Original

Var 1	Var 2	Date	Value
		1	Blue
		2	Cyan
		3	Green

# Original

Var 1	Var 2	Date	Value
		1	Blue
		2	Cyan
		3	Green

# Spreading

Var 1	Var 2	Date 1	Date 2	Date 3
		Blue	Cyan	Green

# Terms

Data wrangling :

- Variable creation or “mutation”
- Descriptive statistics
- Formulas



# Terms

Data wrangling :

- Working with variables like dates

- Dates in R:

“YYYY-MM-DD”

LUBRIDATE: wrangle  
times + dates!



Horst '18

@tladeras

# Terms

Data wrangling :

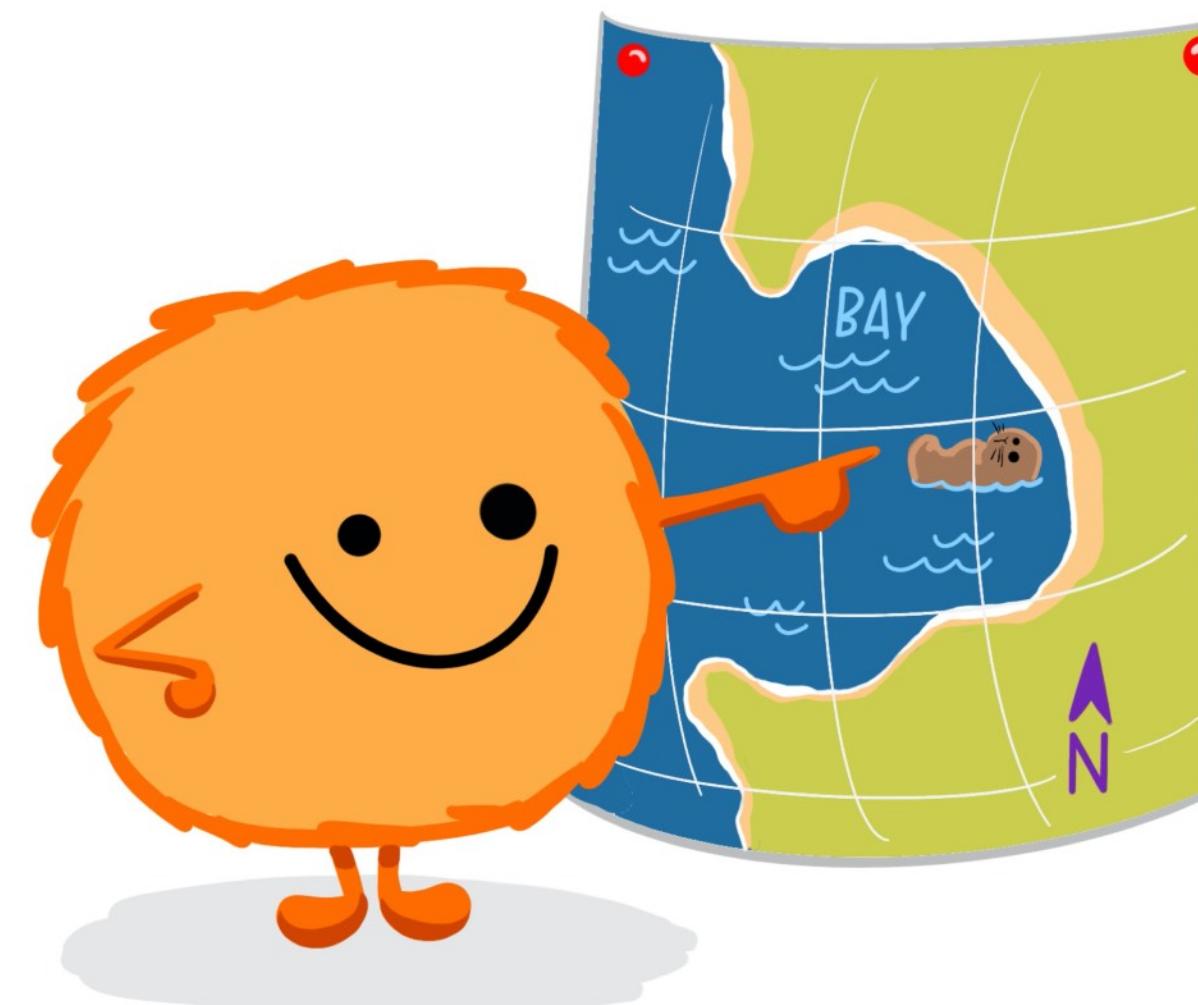
- Subsetting or filtering

# dplyr::filter()

KEEP ROWS THAT  
s.a.t.i.s.f.y  
*your CONDITIONS*

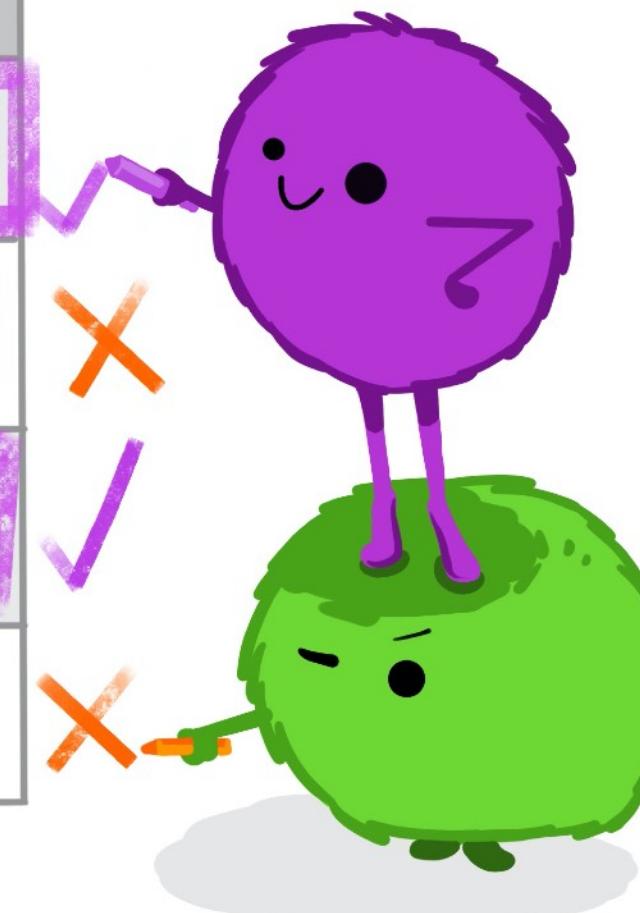
keep rows from... this data... ONLY IF... type is "otter"  
AND site is "bay"

```
filter(df, type == "otter" & site == "bay")
```



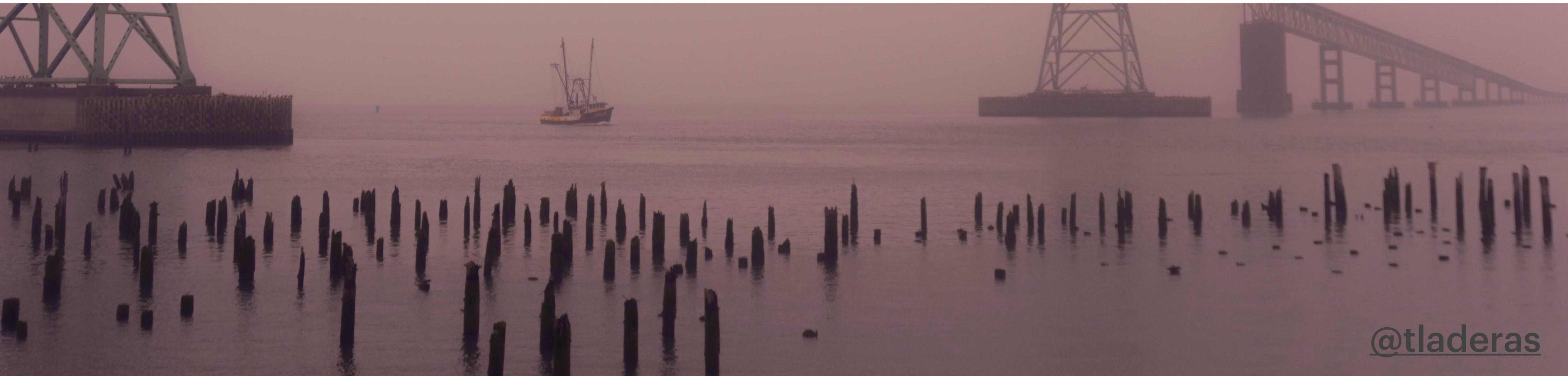
type	food	site
otter	urchin	bay
Shark	seal	channel
otter	abalone	bay
otter	crab	wharf

@allison\_horst





@tladeras

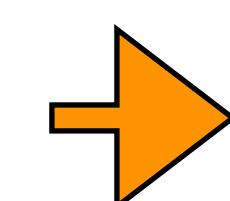
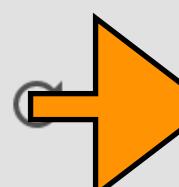


@tladeras

Download RStudio - RStudio

rstudio.com/products/rstudio/download/

Apps Data Experience Miscellaneous News PhD Work



### RStudio Desktop

Open Source License

**Free**

**DOWNLOAD**

[Learn more](#)

### RStudio Desktop Pro

Commercial License

**\$995**

/year

**BUY**

[Learn more](#)

### RStudio Server

Open Source License

**Free**

**DOWNLOAD**

[Learn more](#)

### RStudio Server Pro

Commercial License

**\$4,975**

/year

(5 Named Users)

**BUY**

[Evaluation | Learn more](#)

Integrated Tools for R



Priority Support



Access via Web Browser



RStudio Professional Drivers



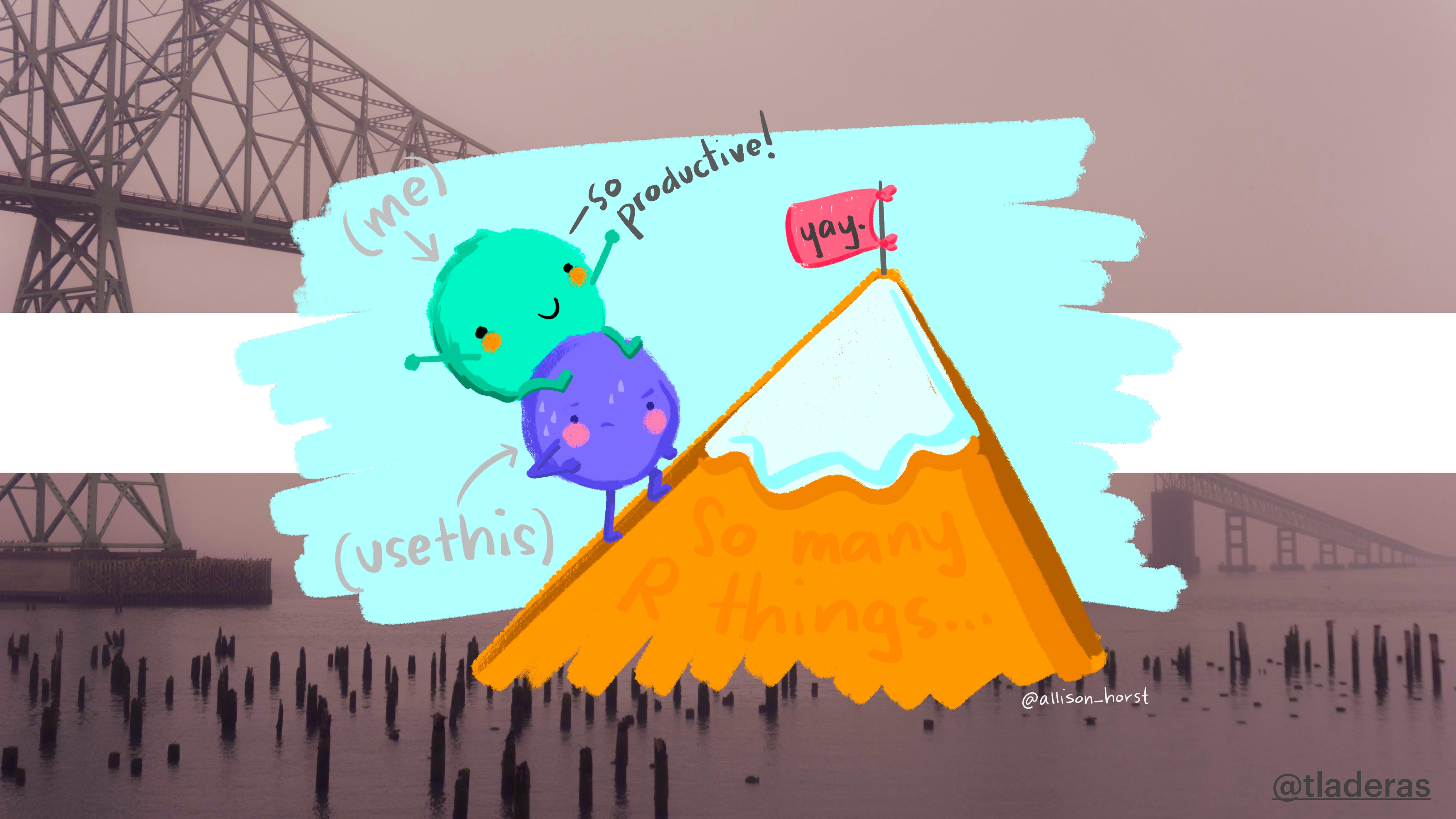
Connect to RStudio Server



Pro remotely



@tladeras



@allison\_horst

@tladeras