Welcome to EAES 494!

Data Science & Statistics

Gavin McNicol


UIC

Dr. Gavin McNicol

McNicol Lab @ UIC

Data science

- Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge.
- We're going to learn to do this in a tidy way -- more on that later!
- This is a course on introduction to data science, with an emphasis on statistical thinking.

Q - What data science background does this course assume?
A - None.

Q - Is this an intro stat course?
A - While statistics $\neq$ data science, they are very closely related and have tremendous of overlap. Hence, this course is a great way to get started with statistics. However this course is *not* your typical high school statistics course.

Q - Will we be doing computing?
A - Yes.

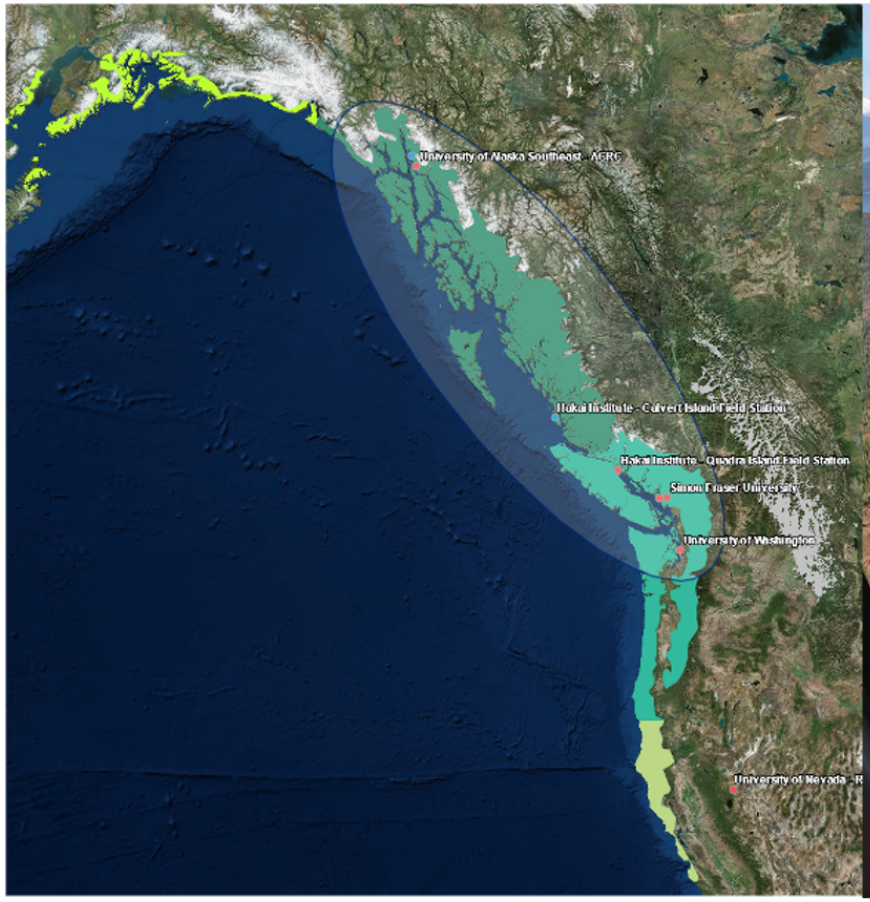
Q - Will we be working with EaESy Data?
A - *Yes!*

Q - Is this an intro CS course?
A - No, but many themes are shared.

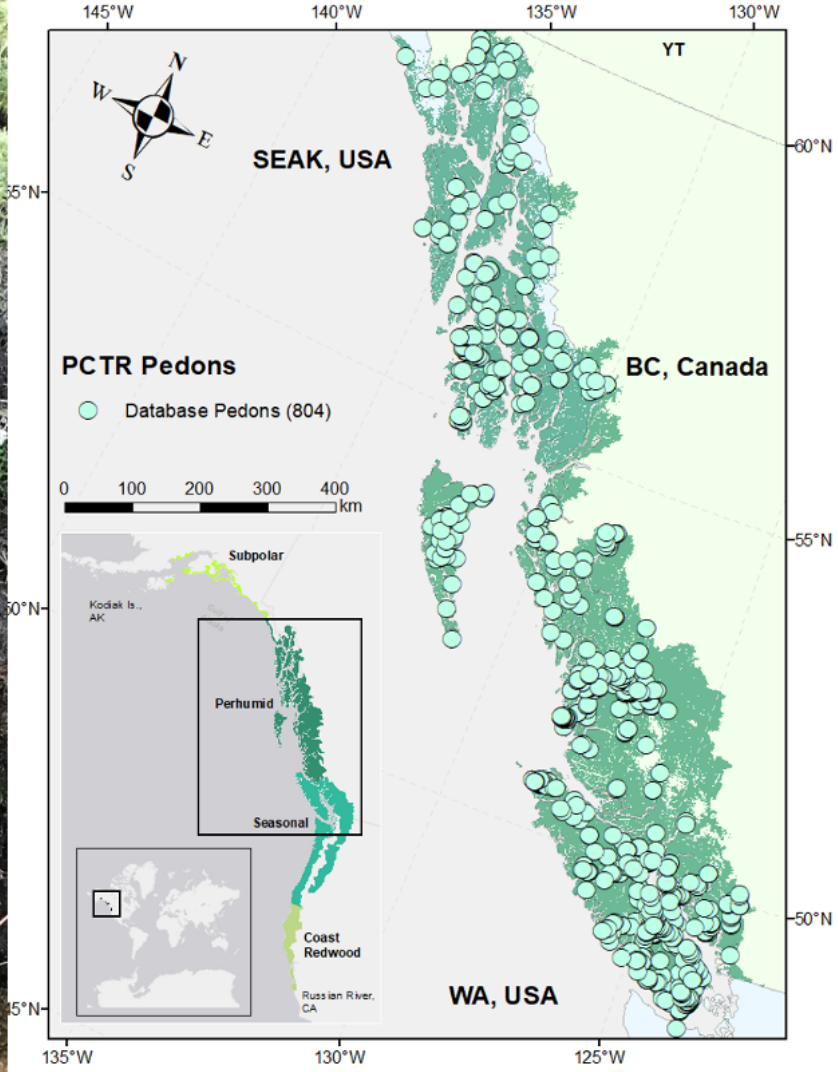
Q - What computing language will we learn?
A - R.

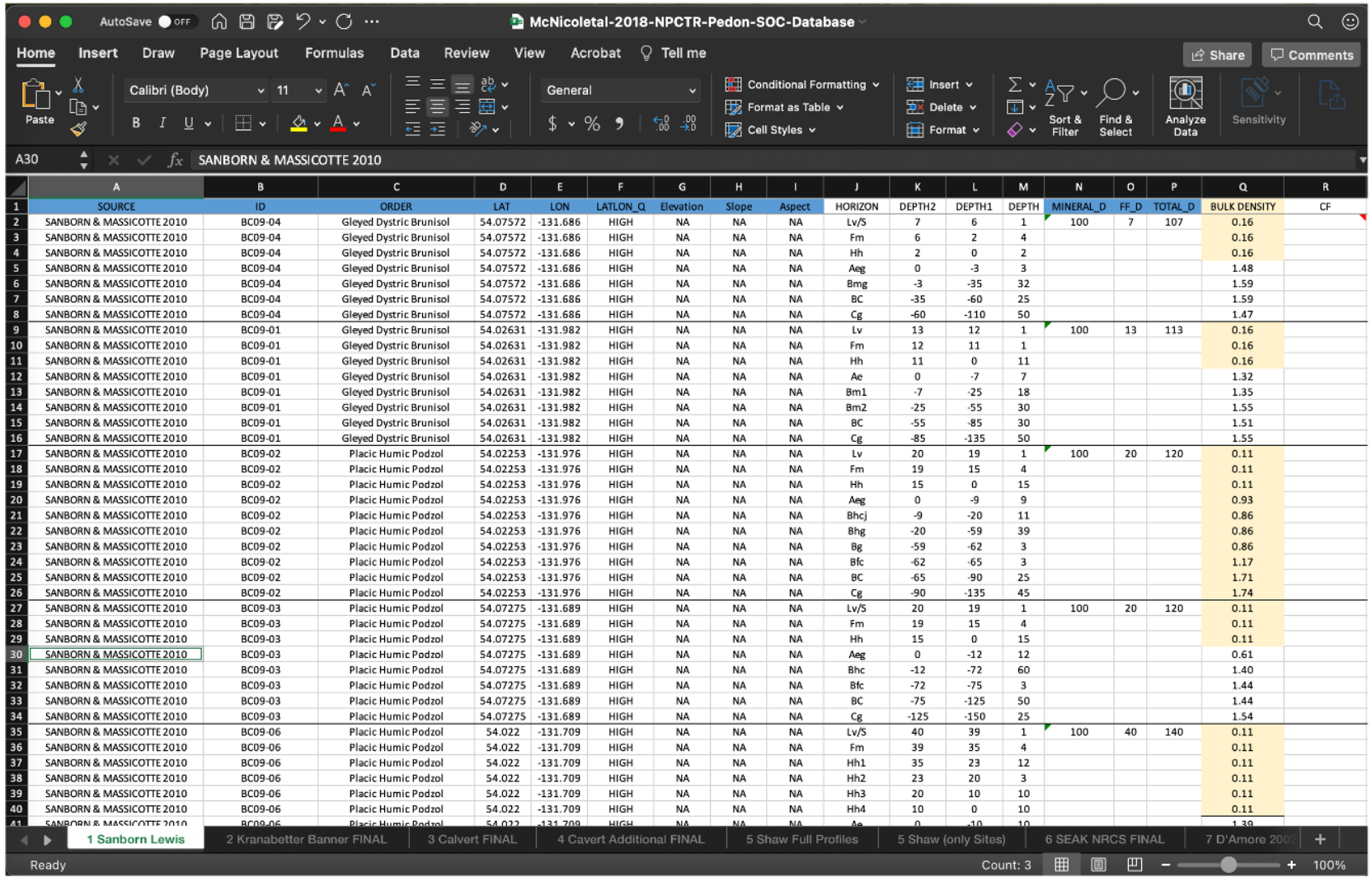
Q: Why not language X?
A: We can discuss that over ☕ .

Software


University of Alaska Southeast - AERC
Hakai Institute - Calvert Island Field Station
Hakai Institute - Quadra Island Field Station
Simon Fraser University
University of Washington
University of Nevada -R


0
20
40
60
80
1
20
40


145°W
140°W
135°W
130°W
N
W
E
S
SEAK, USA
YT
60°N
55°N
50°N
PCTR Pedons
Database Pedons (804)
BC, Canada
0 100 200 300 400
km
Subpolar
Kodiak Is., AK
Perhumid
Seasonal
Coast Redwood
Russian River, CA
WA, USA
135°W
130°W
125°W



McNicoletal-2018-NPCTR-Pedon-SOC-Database

Home Insert Draw Page Layout Formulas Data Review View Acrobat Tell me

Share Comments

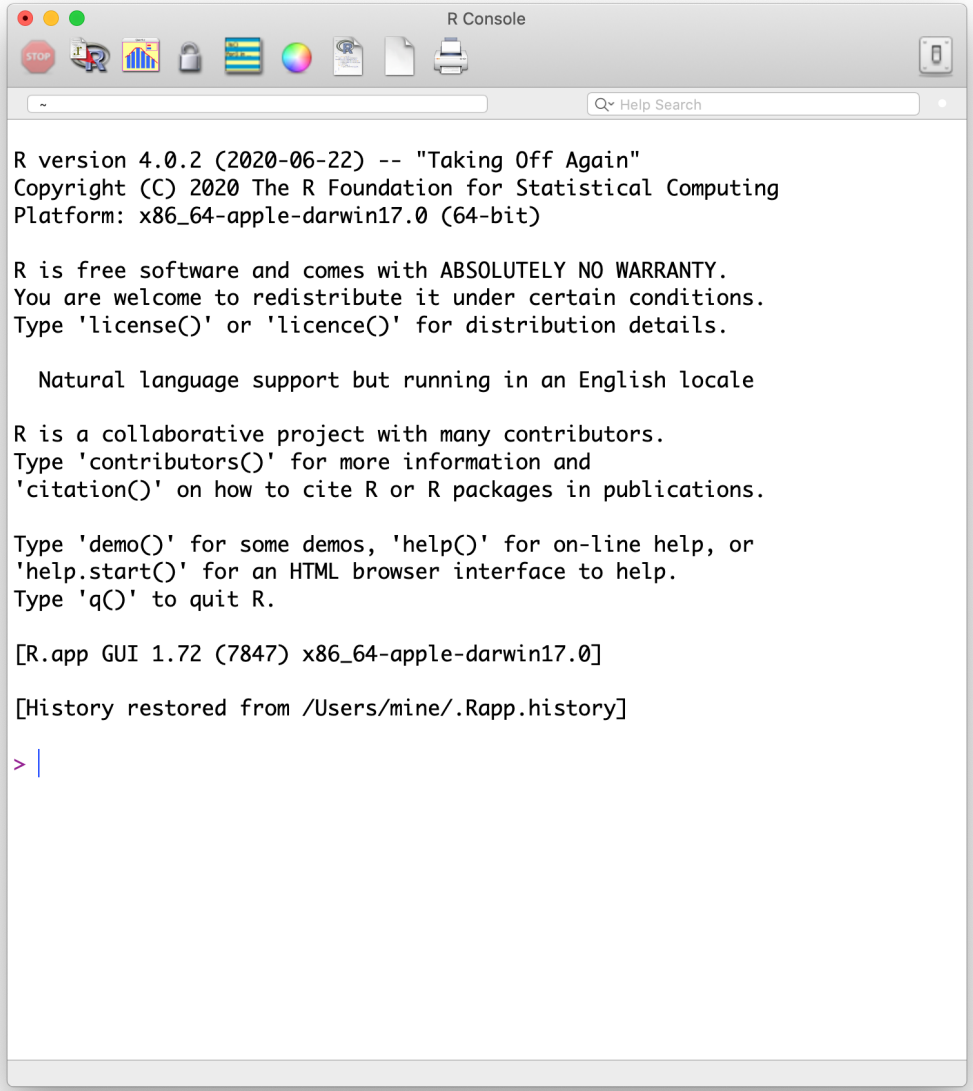
Paste Calibri (Body) 11 General Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select Analyze Data Sensitivity

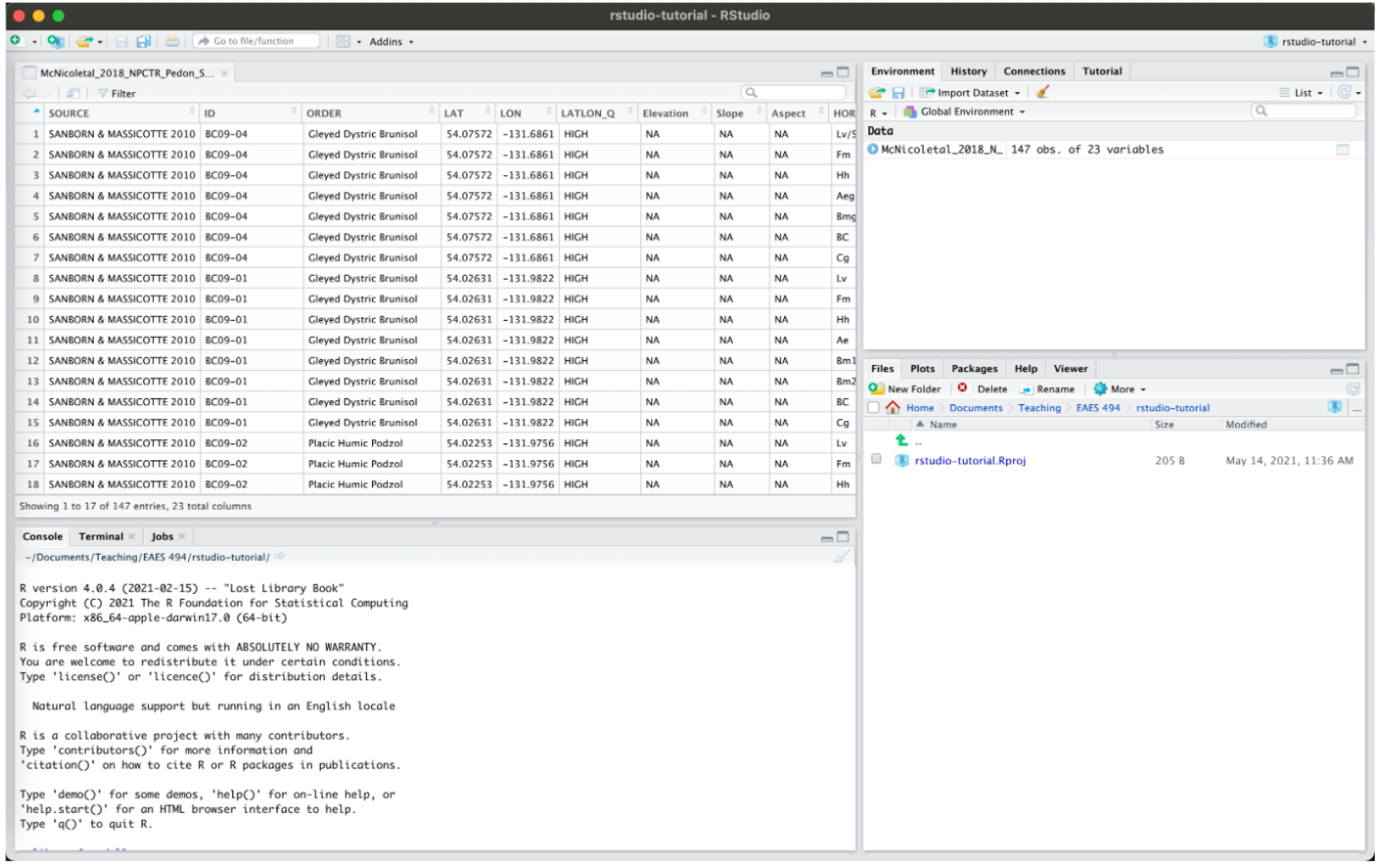
A30 SANBORN & MASSICOTTE 2010

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	SOURCE	ID	ORDER	LAT	LON	LATLON_Q	Elevation	Slope	Aspect	HORIZON	DEPTH2	DEPTH1	DEPTH	MINERAL_D	FF_D	TOTAL_D	BULK DENSITY	CF
2	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Lv/S	7	6	1	100	7	107	0.16	
3	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Fm	6	2	4				0.16	
4	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Hh	2	0	2				0.16	
5	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Aeg	0	-3	3				1.48	
6	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Bmg	-3	-35	32				1.59	
7	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	BC	-35	-60	25				1.59	
8	SANBORN & MASSICOTTE 2010	BC09-04	Gleyed Dystric Brunisol	54.07572	-131.686	HIGH	NA	NA	NA	Cg	-60	-110	50				1.47	
9	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Lv	13	12	1	100	13	113	0.16	
10	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Fm	12	11	1				0.16	
11	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Hh	11	0	11				0.16	
12	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Ae	0	-7	7				1.32	
13	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Bm1	-7	-25	18				1.35	
14	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Bm2	-25	-55	30				1.55	
15	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	BC	-55	-85	30				1.51	
16	SANBORN & MASSICOTTE 2010	BC09-01	Gleyed Dystric Brunisol	54.02631	-131.982	HIGH	NA	NA	NA	Cg	-85	-135	50				1.55	
17	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Lv	20	19	1	100	20	120	0.11	
18	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Fm	19	15	4				0.11	
19	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Hh	15	0	15				0.11	
20	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Aeg	0	-9	9				0.93	
21	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Bhcj	-9	-20	11				0.86	
22	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Bhg	-20	-59	39				0.86	
23	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Bg	-59	-62	3				0.86	
24	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Bfc	-62	-65	3				1.17	
25	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	BC	-65	-90	25				1.71	
26	SANBORN & MASSICOTTE 2010	BC09-02	Placic Humic Podzol	54.02253	-131.976	HIGH	NA	NA	NA	Cg	-90	-135	45				1.74	
27	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Lv/S	20	19	1	100	20	120	0.11	
28	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Fm	19	15	4				0.11	
29	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Hh	15	0	15				0.11	
30	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Aeg	0	-12	12				0.61	
31	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Bhc	-12	-72	60				1.40	
32	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Bfc	-72	-75	3				1.44	
33	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	BC	-75	-125	50				1.44	
34	SANBORN & MASSICOTTE 2010	BC09-03	Placic Humic Podzol	54.07275	-131.689	HIGH	NA	NA	NA	Cg	-125	-150	25				1.54	
35	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Lv/S	40	39	1	100	40	140	0.11	
36	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Fm	39	35	4				0.11	
37	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Hh1	35	23	12				0.11	
38	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Hh2	23	20	3				0.11	
39	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Hh3	20	10	10				0.11	
40	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Hh4	10	0	10				0.11	
41	SANBORN & MASSICOTTE 2010	BC09-06	Placic Humic Podzol	54.022	-131.709	HIGH	NA	NA	NA	Ae	0	-10	10				1.39	

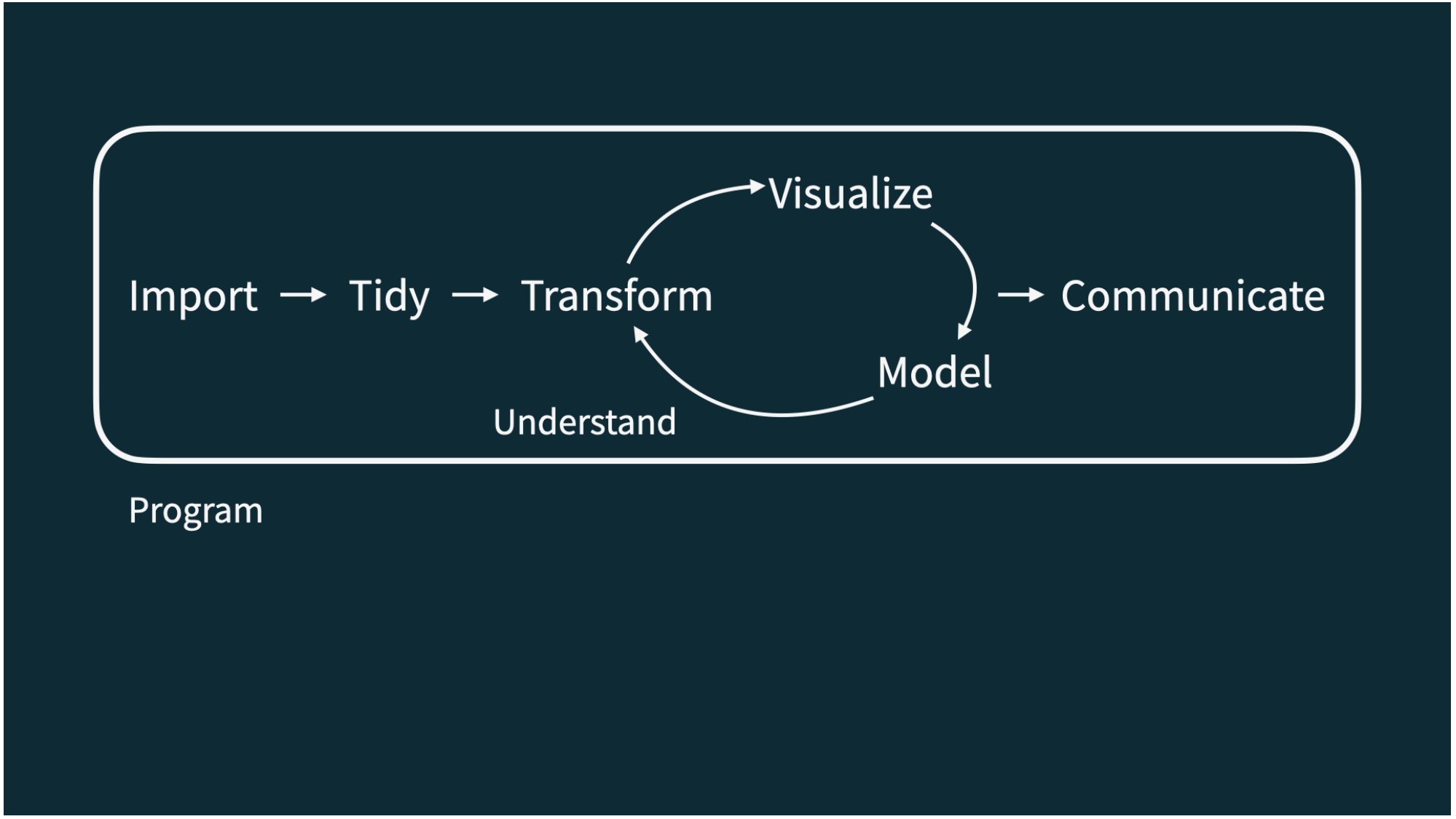
1 Sanborn Lewis | 2 Kranabetter Banner FINAL | 3 Calvert FINAL | 4 Cavert Additional FINAL | 5 Shaw Full Profiles | 5 Shaw (only Sites) | 6 SEAK NRCS FINAL | 7 D'Amore 200

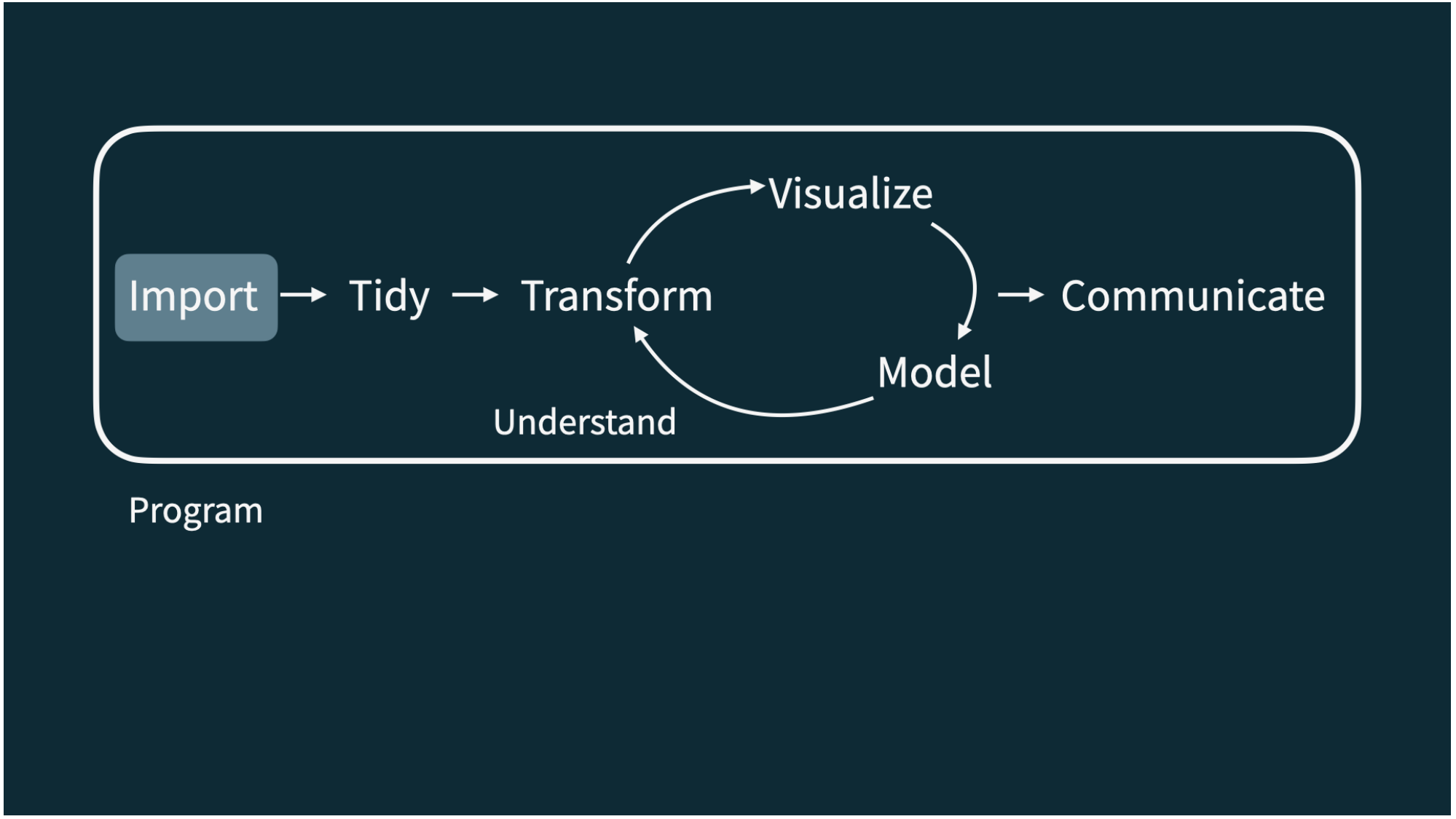
Ready Count: 3 100%

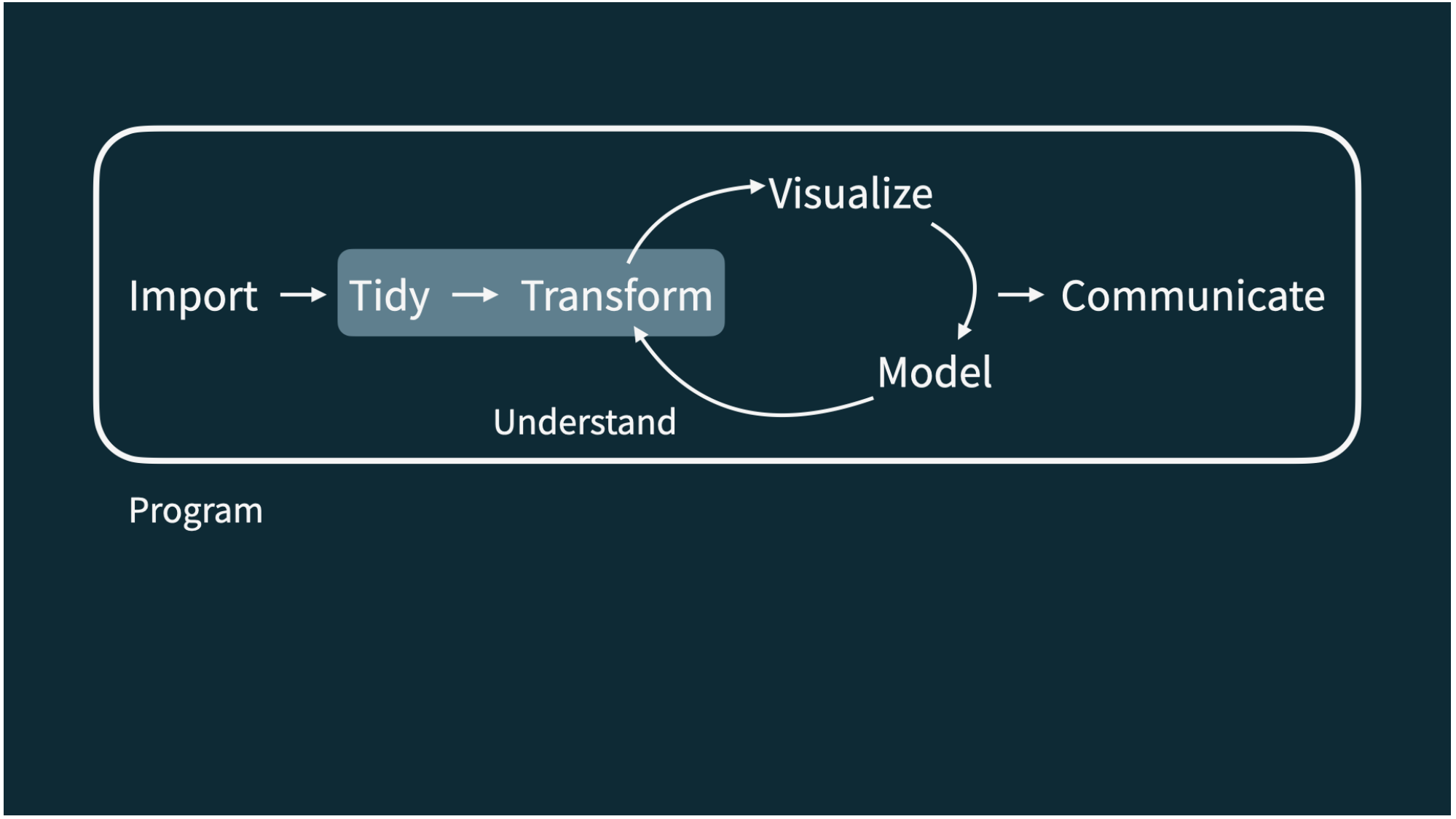

R Console
~
Help Search
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[R.app GUI 1.72 (7847) x86_64-apple-darwin17.0]
[History restored from /Users/mine/.Rapp.history]
>

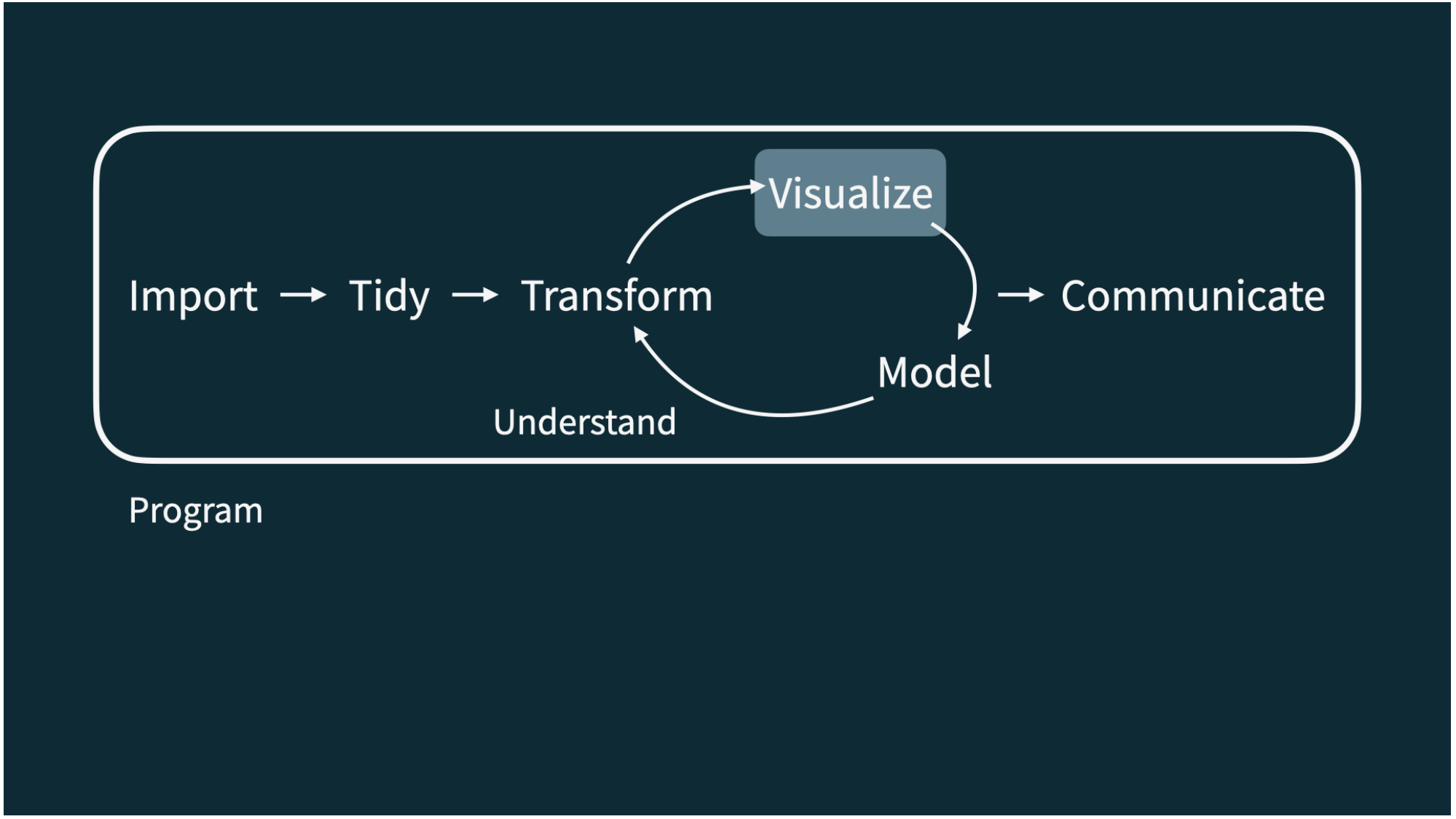

rstudio-tutorial - RStudio
Addins
rstudio-tutorial
McNicoletal_2018_NPCTR_Pedon_S...
Filter
SOURCE ID ORDER LAT LON LATLON_Q Elevation Slope Aspect HOR
1 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Lv/S
2 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Fm
3 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Hh
4 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Aeg
5 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Bmg
6 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA BC
7 SANBORN & MASSICOTTE 2010 BC09-04 Gleyed Dystric Brunisol 54.07572 -131.6861 HIGH NA NA NA Cg
8 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Lv
9 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Fm
10 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Hh
11 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Ae
12 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Bm1
13 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Bm2
14 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA BC
15 SANBORN & MASSICOTTE 2010 BC09-01 Gleyed Dystric Brunisol 54.02631 -131.9822 HIGH NA NA NA Cg
16 SANBORN & MASSICOTTE 2010 BC09-02 Placic Humic Podzol 54.02253 -131.9756 HIGH NA NA NA Lv
17 SANBORN & MASSICOTTE 2010 BC09-02 Placic Humic Podzol 54.02253 -131.9756 HIGH NA NA NA Fm
18 SANBORN & MASSICOTTE 2010 BC09-02 Placic Humic Podzol 54.02253 -131.9756 HIGH NA NA NA Hh
Showing 1 to 17 of 147 entries, 23 total columns
Console Terminal Jobs
~/Documents/Teaching/EAES 494/rstudio-tutorial/
R version 4.0.4 (2021-02-15) -- "Lost Library Book"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment History Connections Tutorial
Import Dataset
List
R
Global Environment
Data
McNicoletal_2018_N... 147 obs. of 23 variables
Files Plots Packages Help Viewer
New Folder
Delete
Rename
More
Home Documents Teaching EAES 494 rstudio-tutorial
Name Size Modified
..
rstudio-tutorial.Rproj 205 B May 14, 2021, 11:36 AM

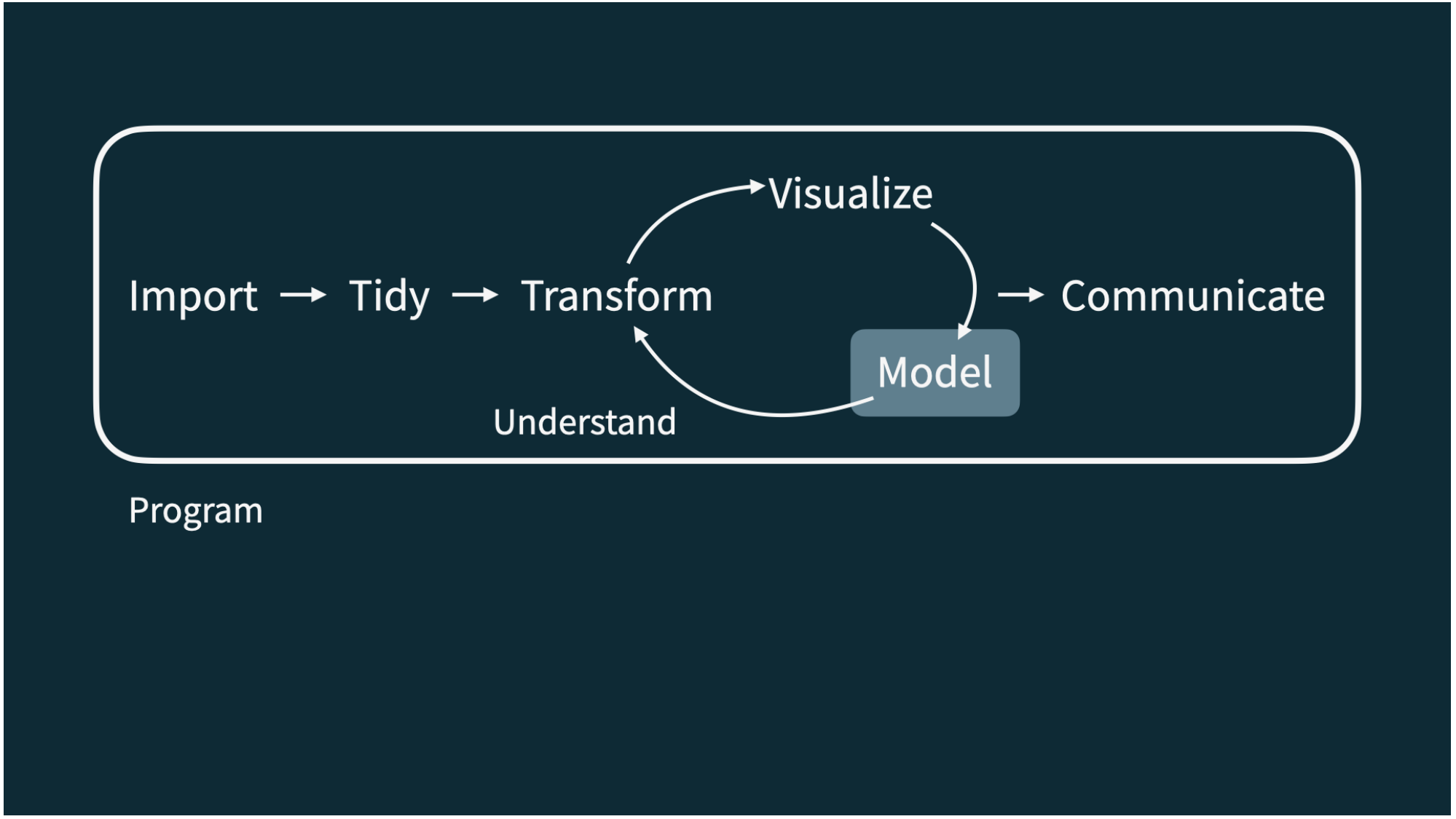
Data science life cycle

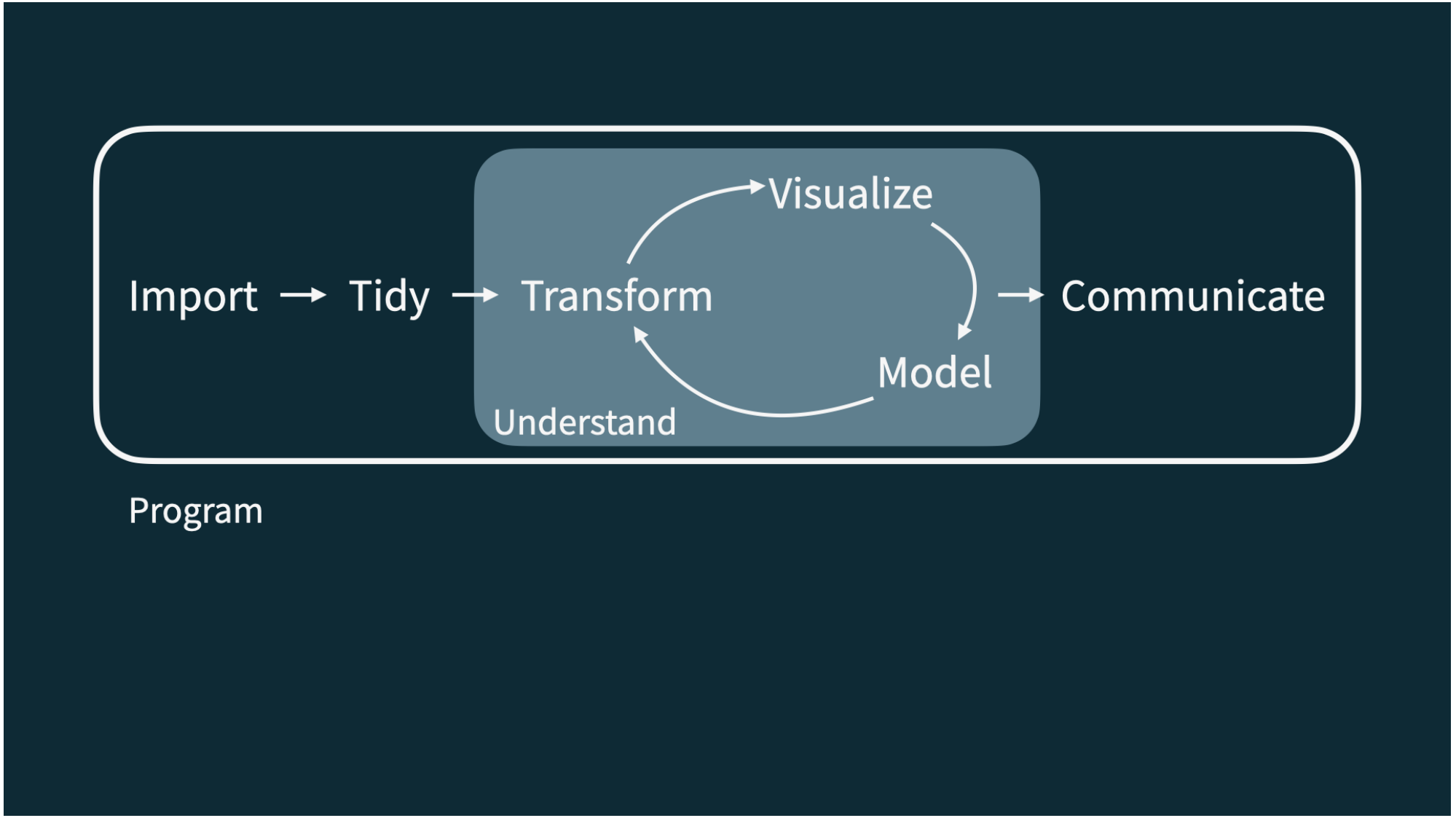

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

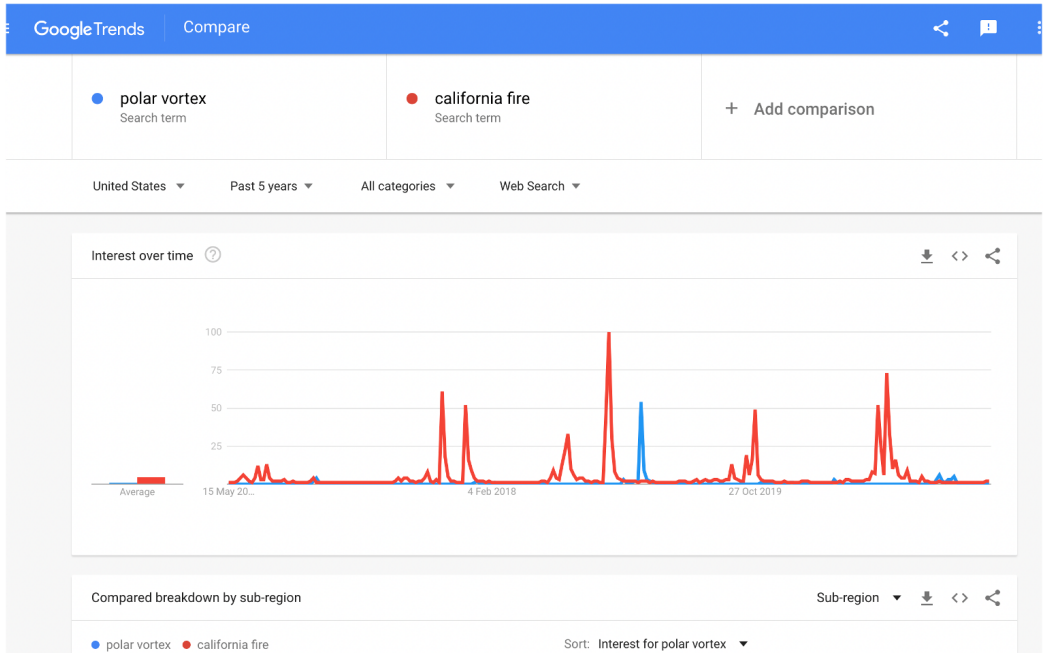

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

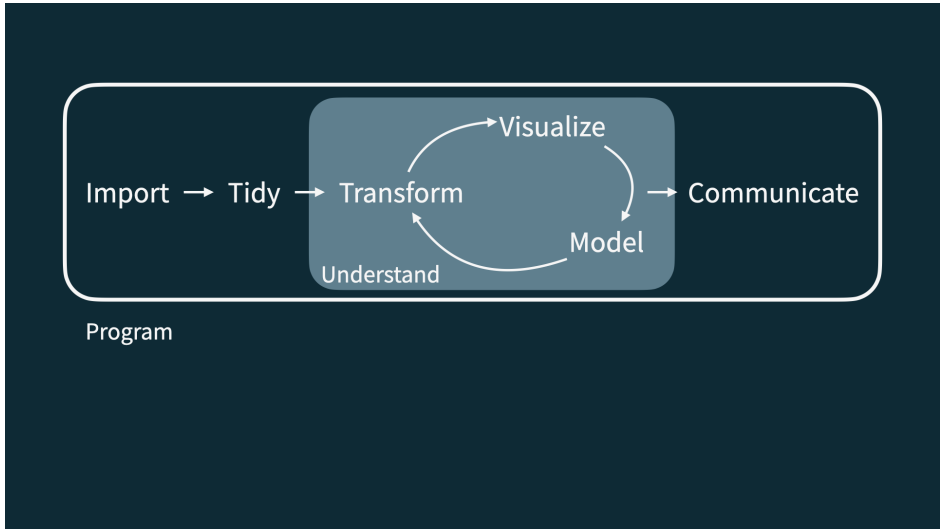

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

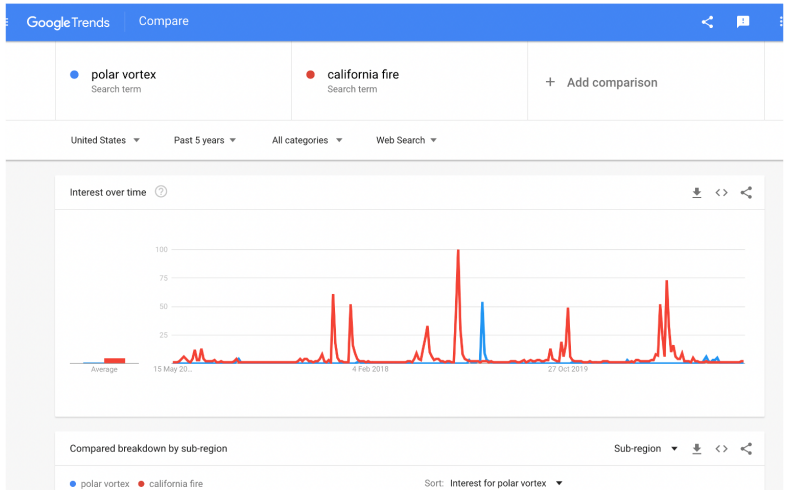

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

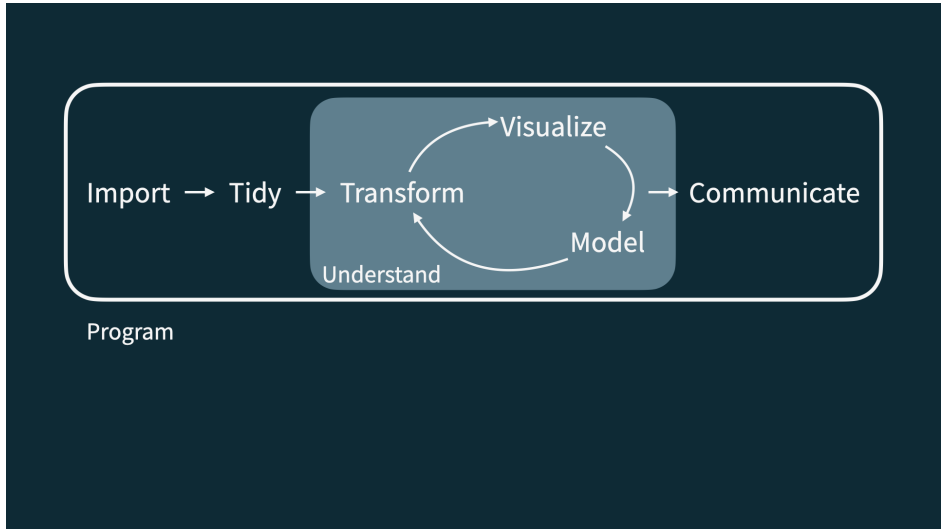

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program


Import
Tidy
Transform
Visualize
Model
Understand
Communicate
Program

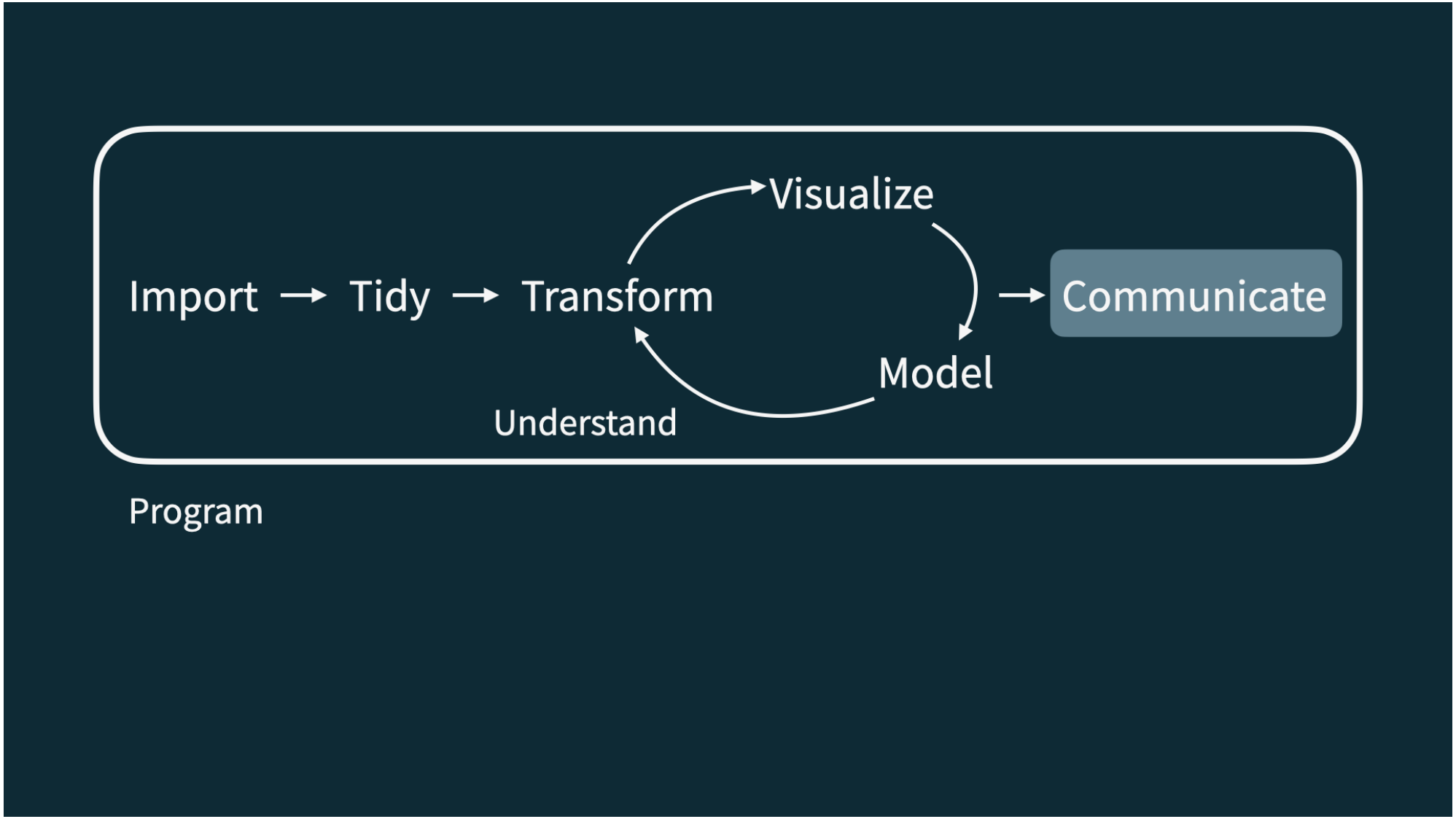

Google Trends
Compare
polar vortex
Search term
california fire
Search term
Add comparison
United States
Past 5 years
All categories
Web Search
Interest over time
100
75
50
25
Average
15 May 20...
4 Feb 2018
27 Oct 2019
Compared breakdown by sub-region
Sub-region
polar vortex
california fire
Sort:
Interest for polar vortex

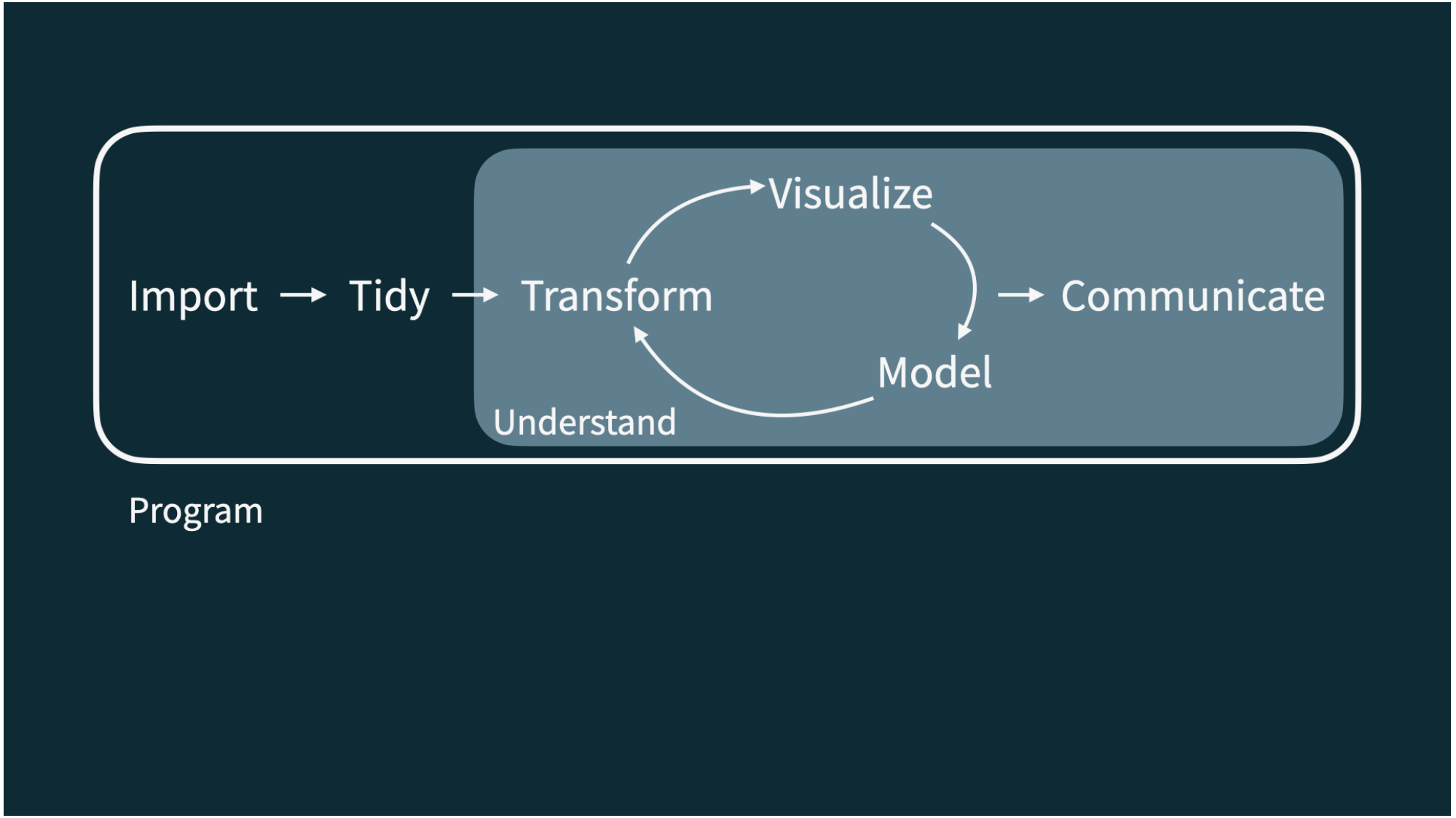

Import
Tidy
Transform
Visualize
Model
Understand
Communicate
Program

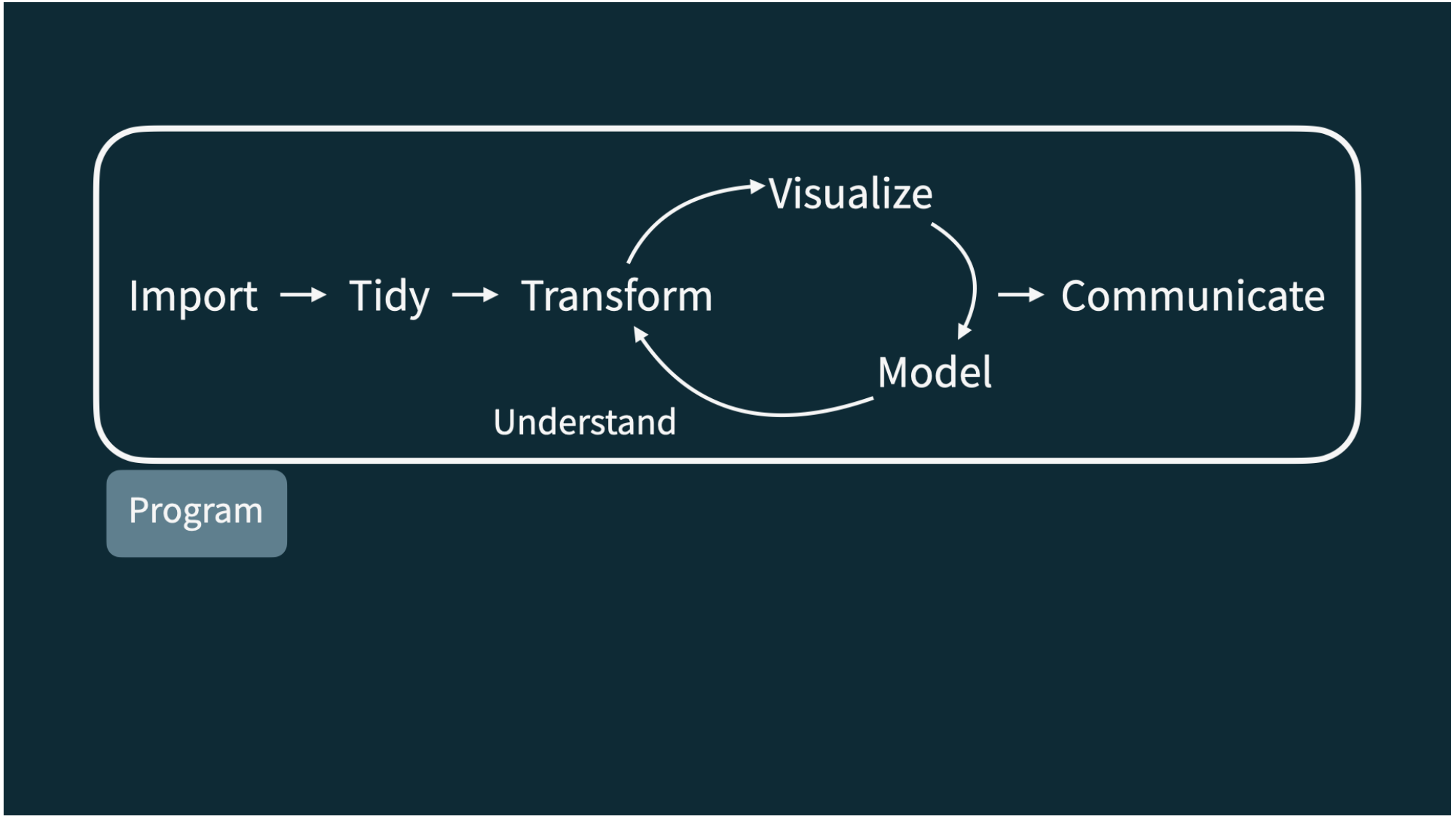


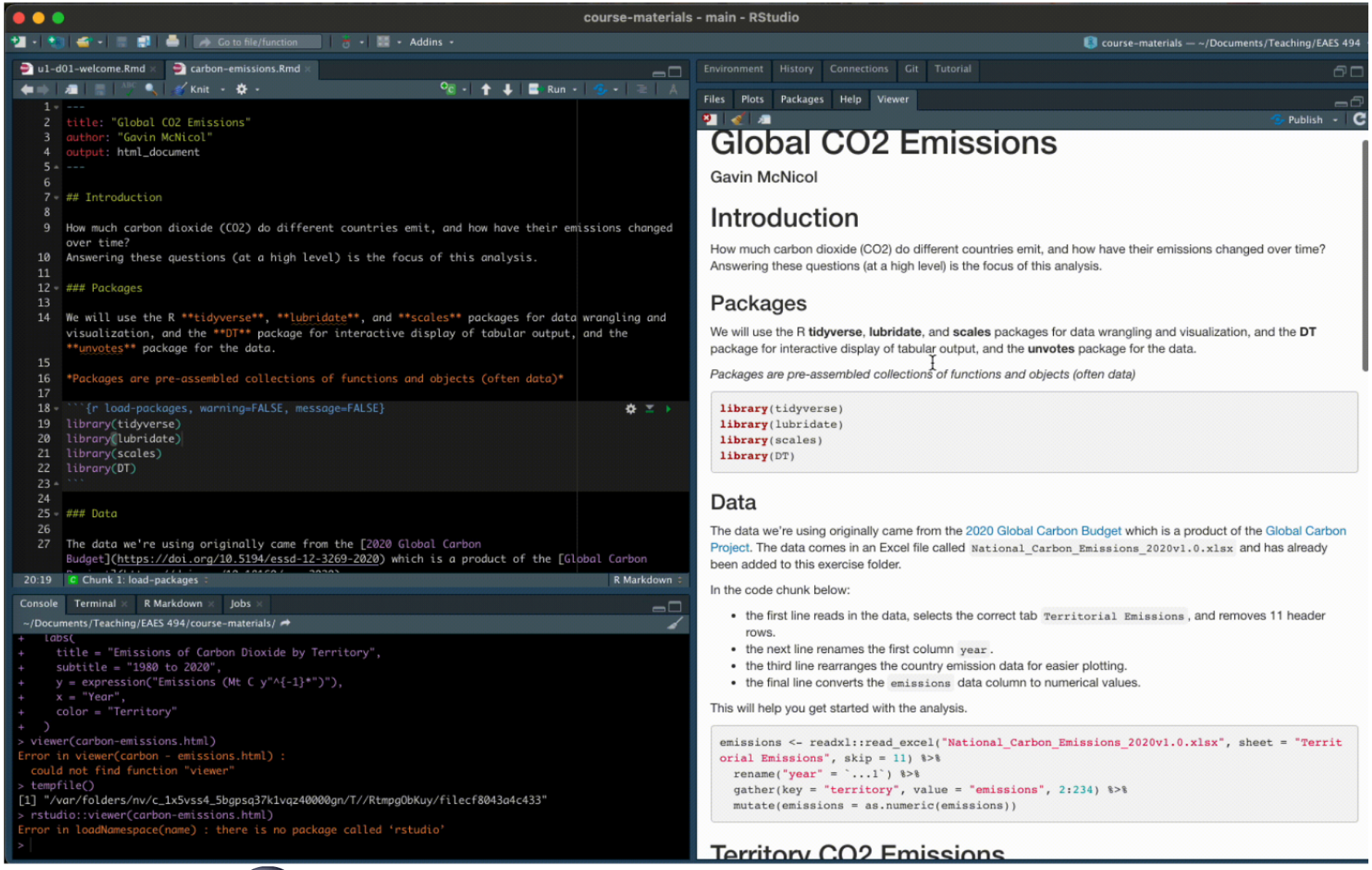



```
## # A tibble: 5 x 2
##   date             season
##   <chr>            <chr>
## 1 23 January 2017  winter
## 2 4 March 2017     spring
## 3 14 June 2017     summer
## 4 1 September 2017 fall
## 5 ...              ...
```

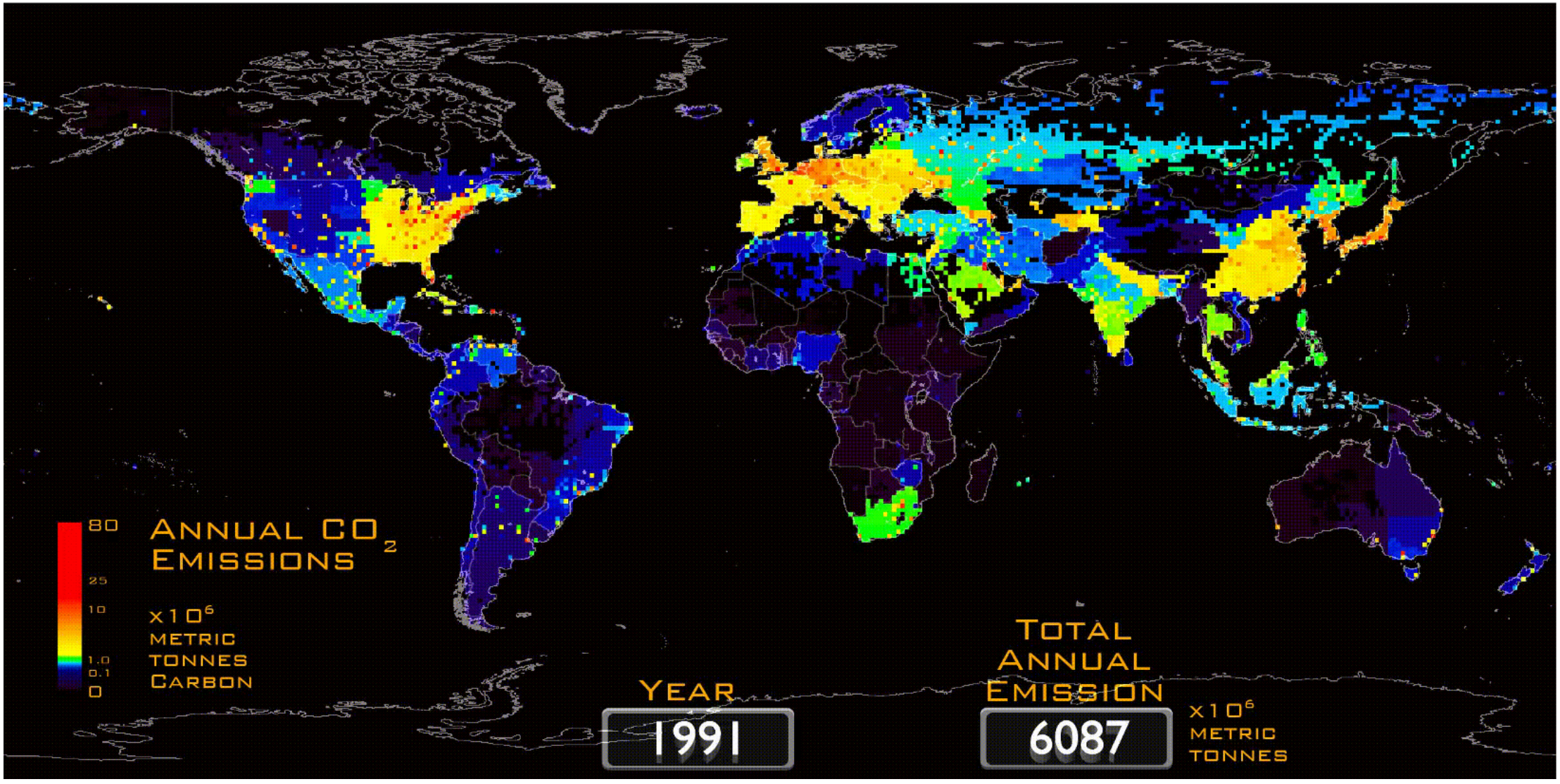

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

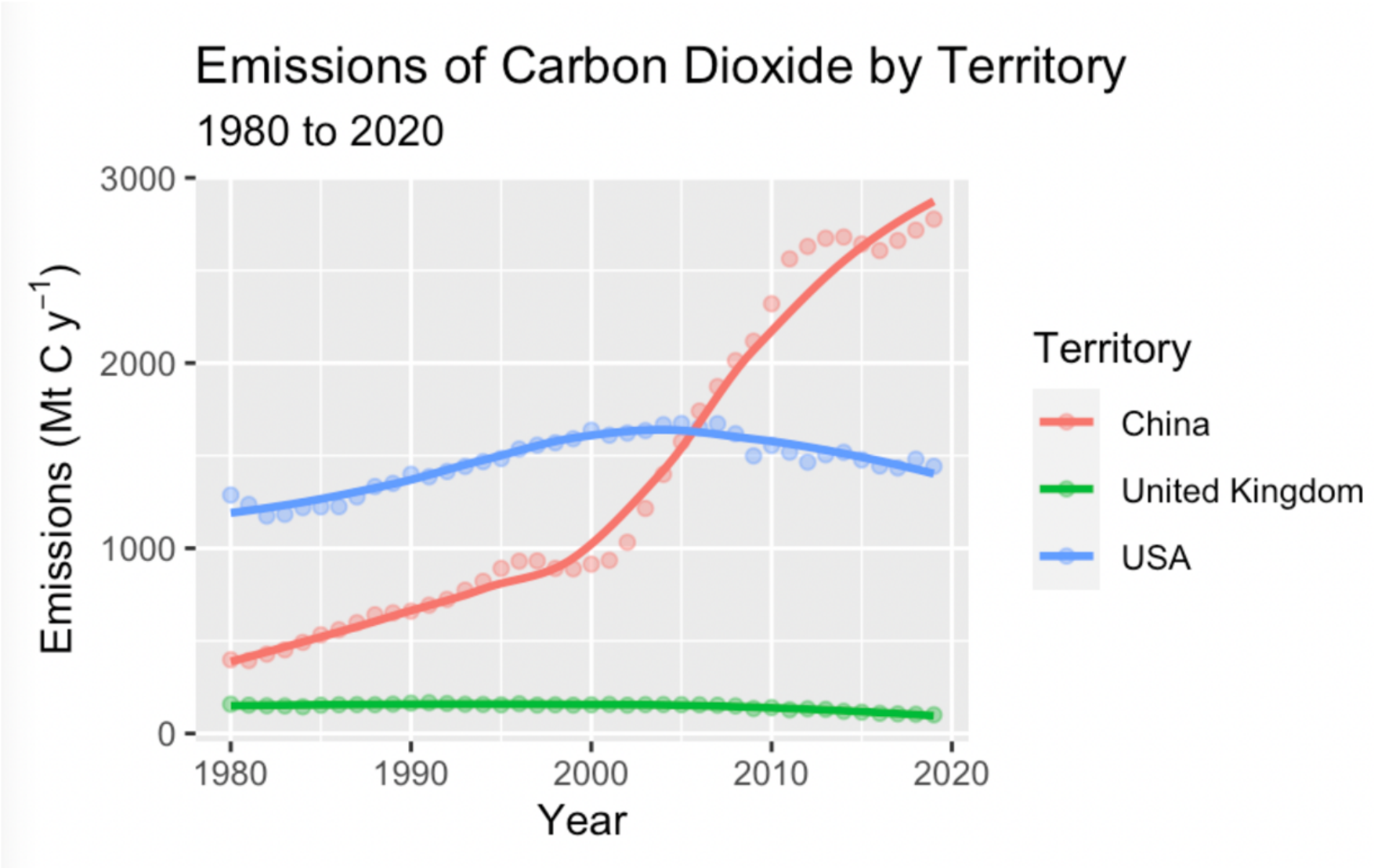

Import
Tidy
Transform
Visualize
Model
Understand
Communicate
Program

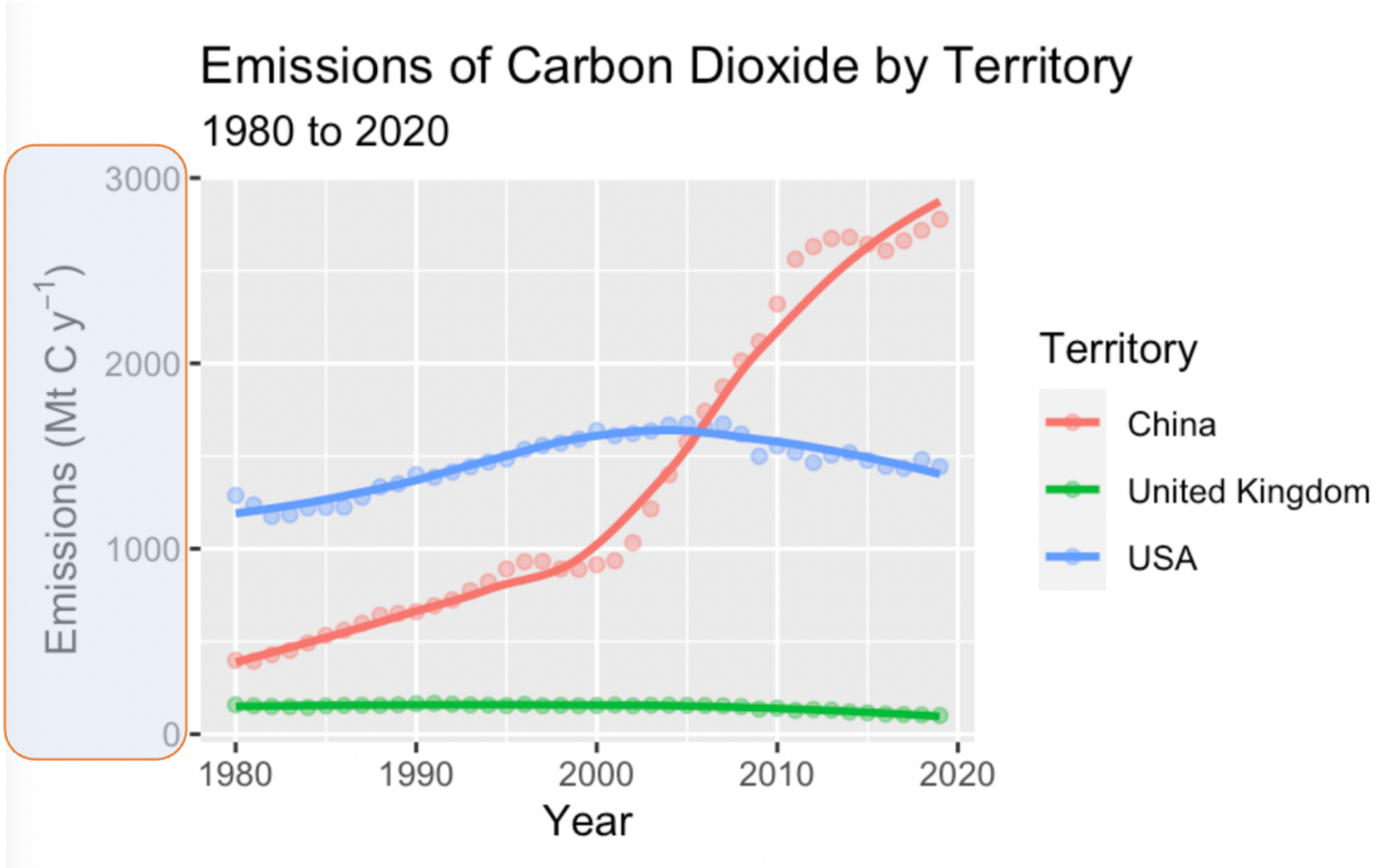

Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

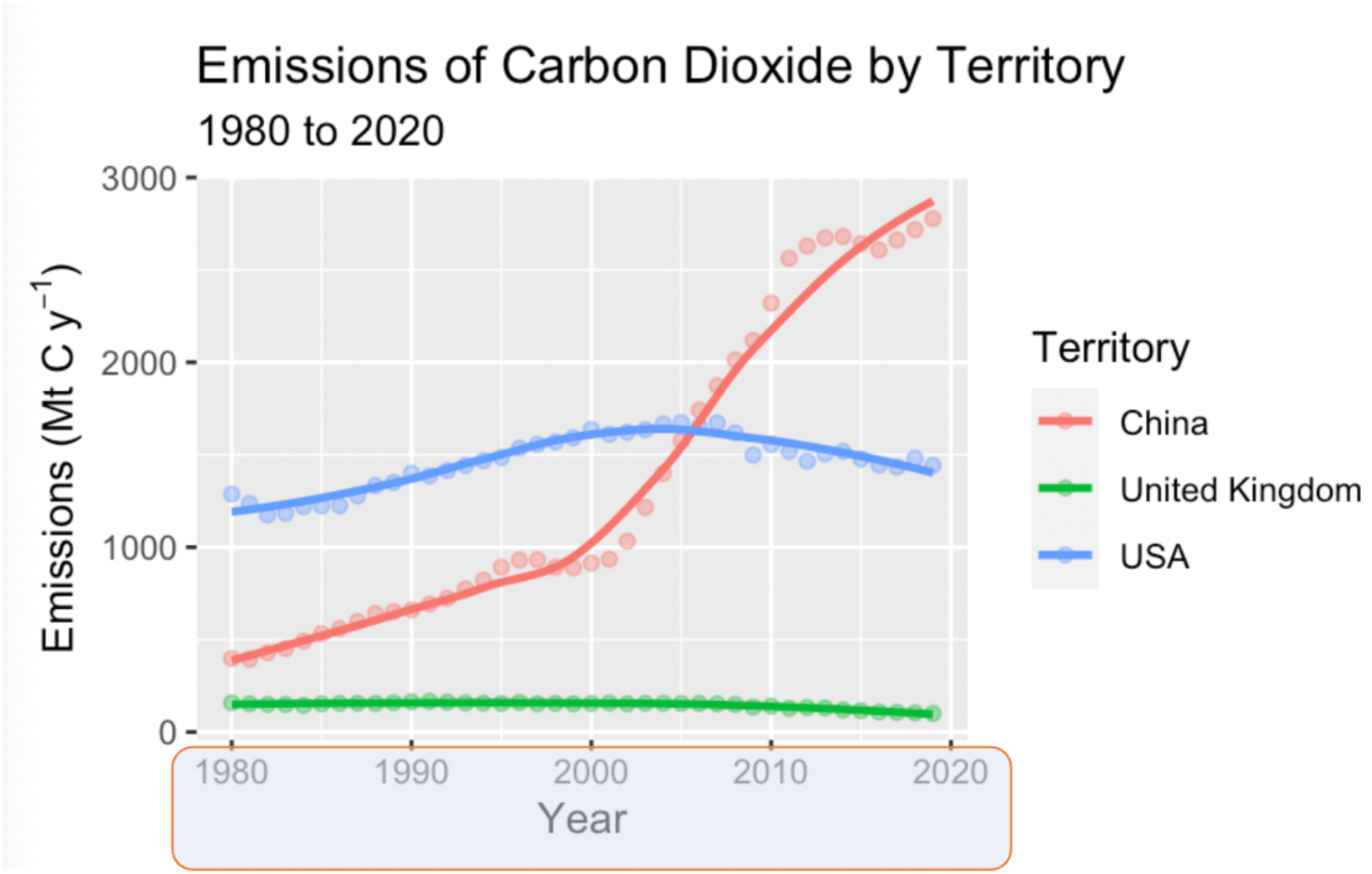

course-materials - main - RStudio
u1-d01-welcome.Rmd
carbon-emissions.Rmd
Knit
Run
1 ---
2 title: "Global CO2 Emissions"
3 author: "Gavin McNicol"
4 output: html_document
5 ---
6
7 ## Introduction
8
9 How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time?
10 Answering these questions (at a high level) is the focus of this analysis.
11
12 ### Packages
13
14 We will use the R **tidyverse**, **lubridate**, and **scales** packages for data wrangling and visualization, and the **DT** package for interactive display of tabular output, and the **unvotes** package for the data.
15
16 *Packages are pre-assembled collections of functions and objects (often data)*
17
18 ```{r load-packages, warning=FALSE, message=FALSE}
19 library(tidyverse)
20 library(lubridate)
21 library(scales)
22 library(DT)
23 ```
24
25 ### Data
26
27 The data we're using originally came from the [2020 Global Carbon Budget](https://doi.org/10.5194/essd-12-3269-2020) which is a product of the [Global Carbon
20:19
Chunk 1: load-packages
R Markdown
Console
Terminal
R Markdown
Jobs
~/Documents/Teaching/EAES 494/course-materials/
+ labs(
+ title = "Emissions of Carbon Dioxide by Territory",
+ subtitle = "1980 to 2020",
+ y = expression("Emissions (Mt C y"^{-1}*")"),
+ x = "Year",
+ color = "Territory"
+)
> viewer(carbon-emissions.html)
Error in viewer(carbon - emissions.html) :
could not find function "viewer"
> tempfile()
[1] "/var/folders/nv/c_1x5vss4_5bgpsq37k1vqz40000gn/T//RtmpgObKuy/filecf8043a4c433"
> rstudio::viewer(carbon-emissions.html)
Error in loadNamespace(name) : there is no package called 'rstudio'
course-materials — ~/Documents/Teaching/EAES 494
Environment
History
Connections
Git
Tutorial
Files
Plots
Packages
Help
Viewer
Publish
Global CO2 Emissions
Gavin McNicol
Introduction
How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time? Answering these questions (at a high level) is the focus of this analysis.
Packages
We will use the R tidyverse, lubridate, and scales packages for data wrangling and visualization, and the DT package for interactive display of tabular output, and the unvotes package for the data.
Packages are pre-assembled collections of functions and objects (often data)
library(tidyverse)
library(lubridate)
library(scales)
library(DT)
Data
The data we're using originally came from the 2020 Global Carbon Budget which is a product of the Global Carbon Project. The data comes in an Excel file called National_Carbon_Emissions_2020v1.0.xlsx and has already been added to this exercise folder.
In the code chunk below:
the first line reads in the data, selects the correct tab Territorial Emissions, and removes 11 header rows.
the next line renames the first column year.
the third line rearranges the country emission data for easier plotting.
the final line converts the emissions data column to numerical values.
This will help you get started with the analysis.
emissions <- readxl::read_excel("National_Carbon_Emissions_2020v1.0.xlsx", sheet = "Territorial Emissions", skip = 11) %>%
rename("year" = `...1`) %>%
gather(key = "territory", value = "emissions", 2:234) %>%
mutate(emissions = as.numeric(emissions))
Territory CO2 Emissions

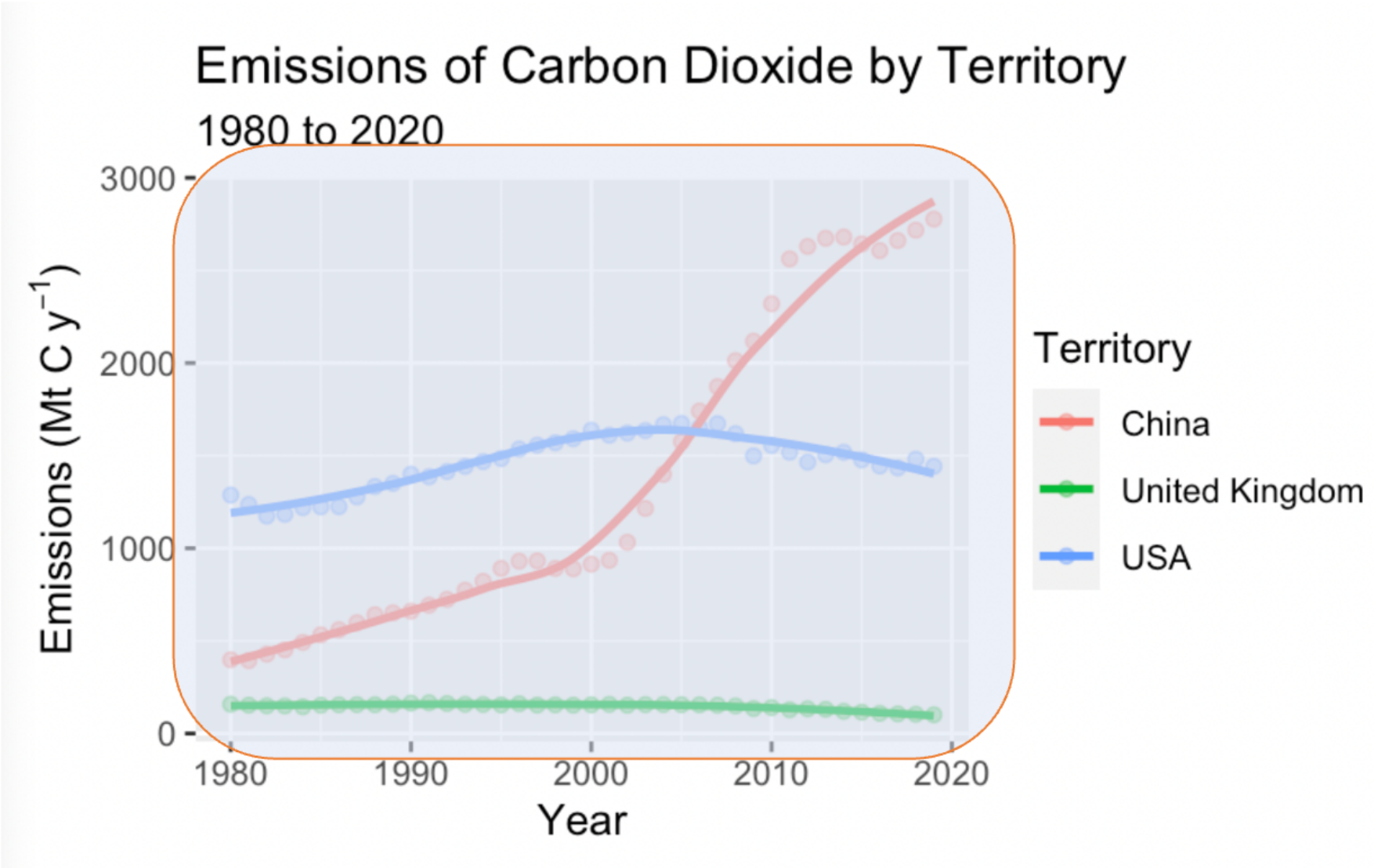
Let's dive in!


80
25
10
1.0
0.1
0
ANNUAL CO2 EMISSIONS
x10^6 METRIC TONNES CARBON
YEAR
1991
TOTAL ANNUAL EMISSION
6087
x10^6 METRIC TONNES


Emissions of Carbon Dioxide by Territory
1980 to 2020
Emissions (Mt C y^{-1})
3000
2000
1000
0
1980
1990
2000
2010
2020
Year
Territory
China
United Kingdom
USA


Emissions of Carbon Dioxide by Territory
1980 to 2020
Emissions (Mt C y^{-1})
3000
2000
1000
0
1980
1990
2000
2010
2020
Year
Territory
China
United Kingdom
USA


Emissions of Carbon Dioxide by Territory
1980 to 2020
Emissions (Mt C y^{-1})
3000
2000
1000
0
1980
1990
2000
2010
2020
Year
Territory
China
United Kingdom
USA


Emissions of Carbon Dioxide by Territory
1980 to 2020
Emissions (Mt C y^{-1})
3000
2000
1000
0
1980
1990
2000
2010
2020
Year
Territory
China
United Kingdom
USA

emissions

Filter

	year	territory	emissions
1	1959	Afghanistan	0.1049594
2	1960	Afghanistan	0.1129598
3	1961	Afghanistan	0.1339513
4	1962	Afghanistan	0.1879351
5	1963	Afghanistan	0.1928865
6	1964	Afghanistan	0.2288621
7	1965	Afghanistan	0.2748135
8	1966	Afghanistan	0.2978054
9	1967	Afghanistan	0.3498540
10	1968	Afghanistan	0.3338949
11	1969	Afghanistan	0.2568865
12	1970	Afghanistan	0.4558943
13	1971	Afghanistan	0.5168072
14	1972	Afghanistan	0.4176876
15	1973	Afghanistan	0.4463936
16	1974	Afghanistan	0.5221996
17	1975	Afghanistan	0.5790436
18	1976	Afghanistan	0.5407152
19	1977	Afghanistan	0.6507874
20	1978	Afghanistan	0.5877777
21	1979	Afghanistan	0.6094850
22	1980	Afghanistan	0.4793848
23	1981	Afghanistan	0.5400434
24	1982	Afghanistan	0.5717566
25	1983	Afghanistan	0.6877769
26	1984	Afghanistan	0.7702047

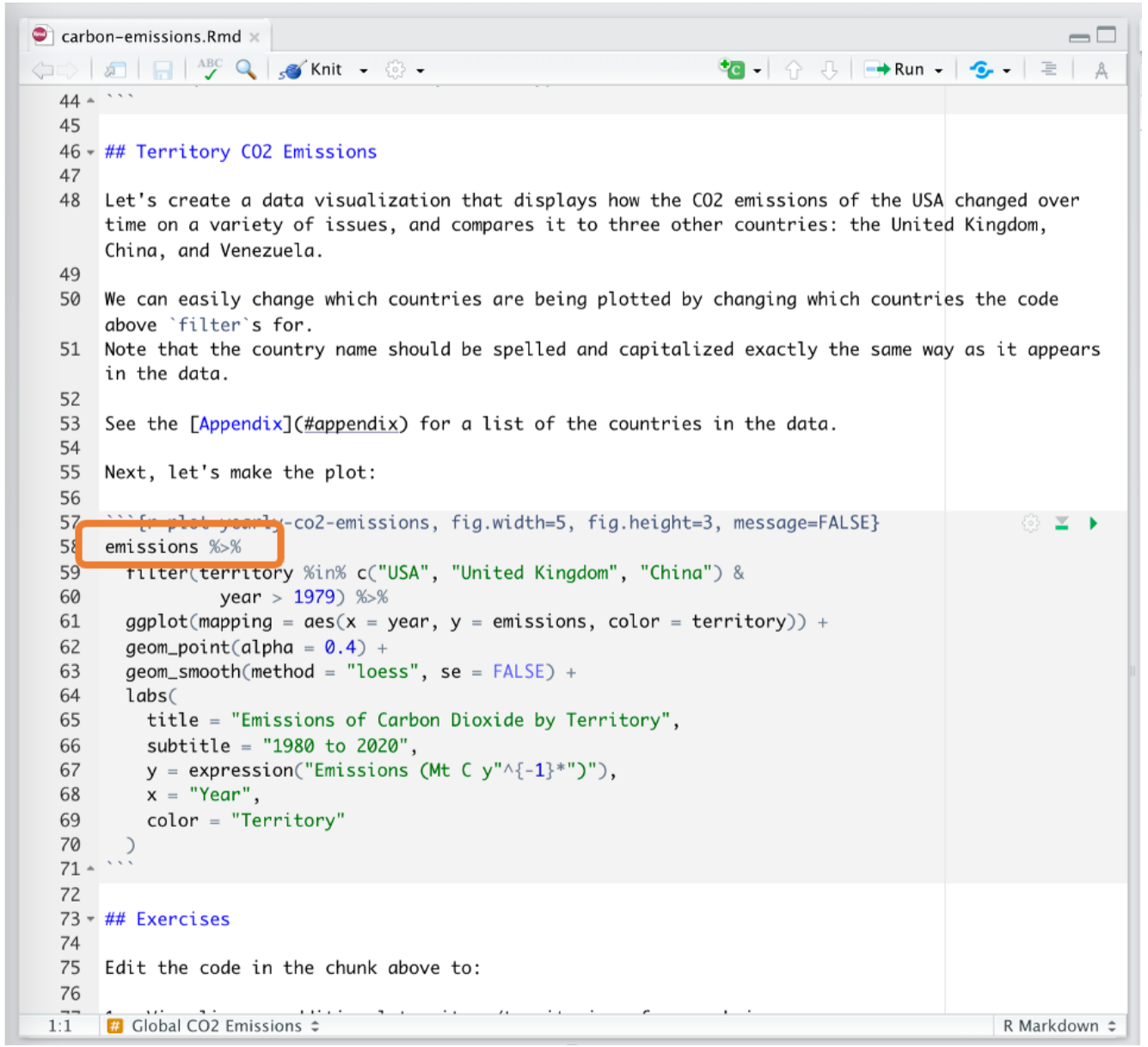
Showing 1 to 26 of 14,213 entries, 3 total columns

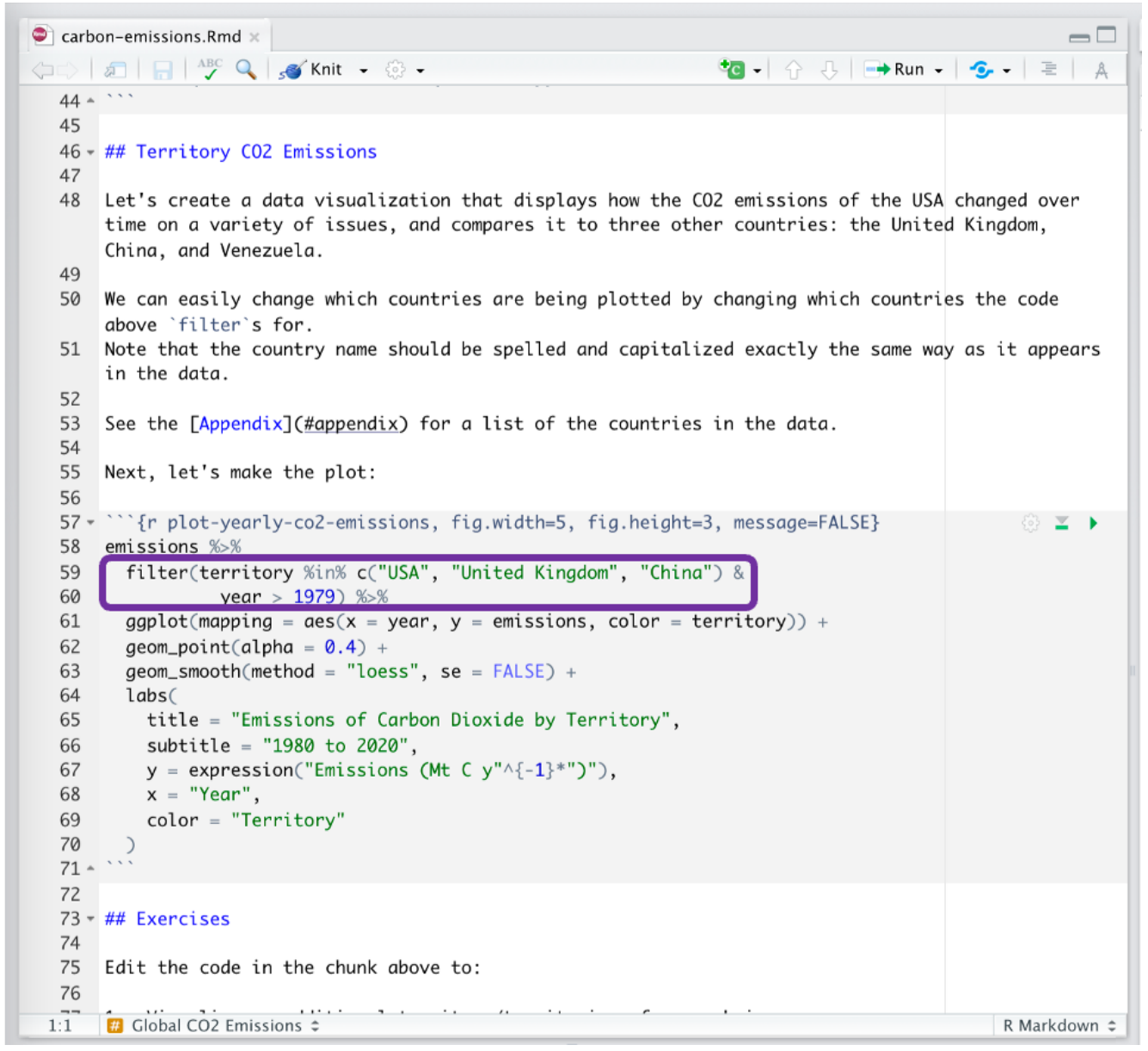
carbon-emissions.Rmd

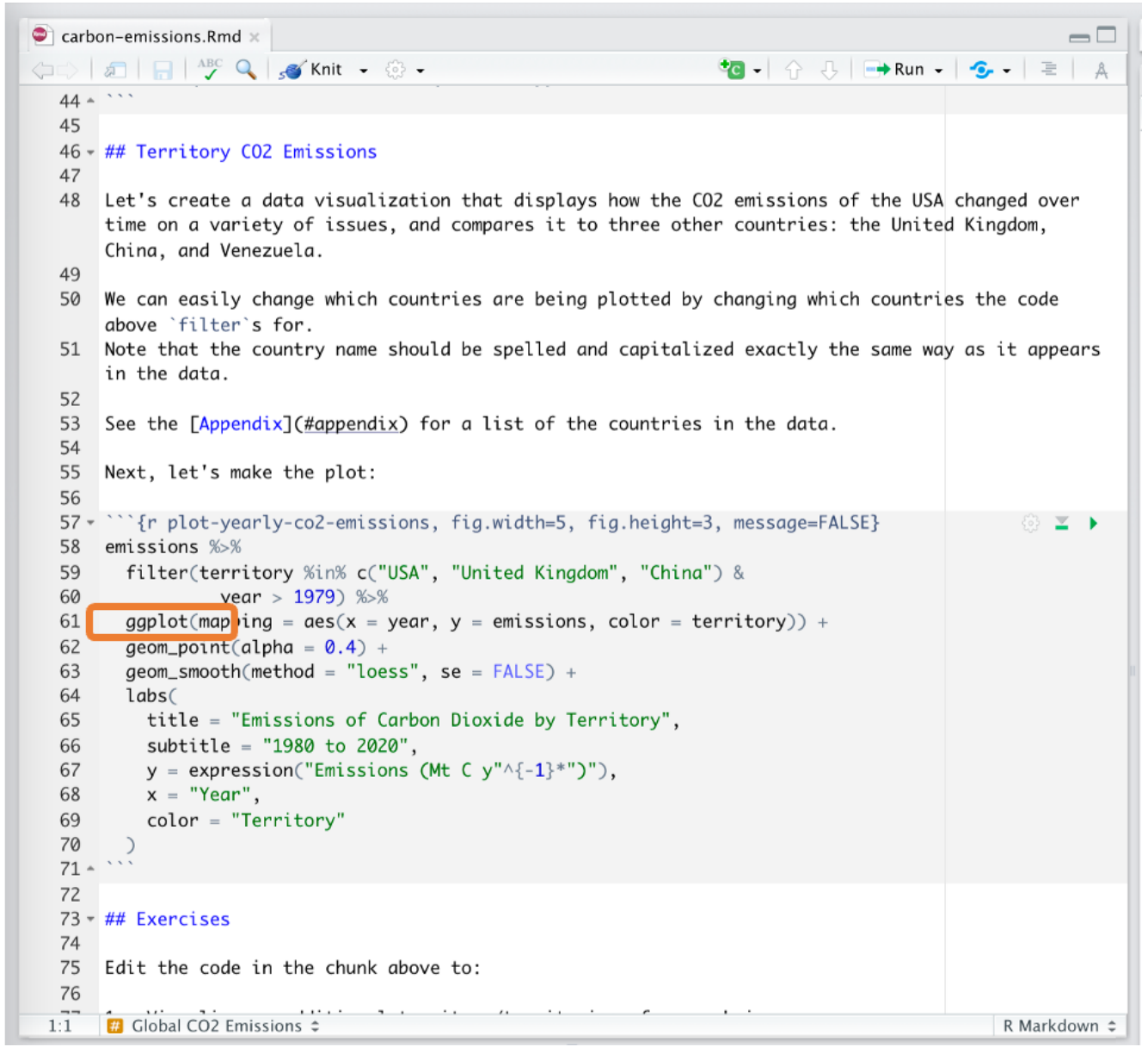
Knit | Run

```
44 ```
45
46 ## Territory CO2 Emissions
47
48 Let's create a data visualization that displays how the CO2 emissions of the USA changed over
   time on a variety of issues, and compares it to three other countries: the United Kingdom,
   China, and Venezuela.
49
50 We can easily change which countries are being plotted by changing which countries the code
   above `filter`s for.
51 Note that the country name should be spelled and capitalized exactly the same way as it appears
   in the data.
52
53 See the [Appendix](#appendix) for a list of the countries in the data.
54
55 Next, let's make the plot:
56
57 ```{r plot-yearly-co2-emissions, fig.width=5, fig.height=3, message=FALSE}
58 emissions %>%
59   filter(territory %in% c("USA", "United Kingdom", "China") &
60            year > 1979) %>%
61   ggplot(mapping = aes(x = year, y = emissions, color = territory)) +
62   geom_point(alpha = 0.4) +
63   geom_smooth(method = "loess", se = FALSE) +
64   labs(
65     title = "Emissions of Carbon Dioxide by Territory",
66     subtitle = "1980 to 2020",
67     y = expression("Emissions (Mt C y"^{-1}*")"),
68     x = "Year",
69     color = "Territory"
70   )
71 ```
72
73 ## Exercises
74
75 Edit the code in the chunk above to:
76
```

1:1 | Global CO2 Emissions | R Markdown


carbon-emissions.Rmd
Knit
Run
Territory CO2 Emissions
Let's create a data visualization that displays how the CO2 emissions of the USA changed over time on a variety of issues, and compares it to three other countries: the United Kingdom, China, and Venezuela.
We can easily change which countries are being plotted by changing which countries the code above `filter`s for.
Note that the country name should be spelled and capitalized exactly the same way as it appears in the data.
See the [Appendix](#appendix) for a list of the countries in the data.
Next, let's make the plot:
```{r plot-yearly-co2-emissions, fig.width=5, fig.height=3, message=FALSE}
emissions %>%
filter(territory %in% c("USA", "United Kingdom", "China") &
year > 1979) %>%
ggplot(mapping = aes(x = year, y = emissions, color = territory)) +
geom_point(alpha = 0.4) +
geom_smooth(method = "loess", se = FALSE) +
labs(
title = "Emissions of Carbon Dioxide by Territory",
subtitle = "1980 to 2020",
y = expression("Emissions (Mt C y"^{-1}*")"),
x = "Year",
color = "Territory"
)
```
Exercises
Edit the code in the chunk above to:
1:1
Global CO2 Emissions
R Markdown


carbon-emissions.Rmd
Knit
Run
44 ```
45
46 ## Territory CO2 Emissions
47
48 Let's create a data visualization that displays how the CO2 emissions of the USA changed over time on a variety of issues, and compares it to three other countries: the United Kingdom, China, and Venezuela.
49
50 We can easily change which countries are being plotted by changing which countries the code above `filter`s for.
51 Note that the country name should be spelled and capitalized exactly the same way as it appears in the data.
52
53 See the [Appendix](#appendix) for a list of the countries in the data.
54
55 Next, let's make the plot:
56
57 ```{r plot-yearly-co2-emissions, fig.width=5, fig.height=3, message=FALSE}
58 emissions %>%
59   filter(territory %in% c("USA", "United Kingdom", "China") &
60   year > 1979) %>%
61   ggplot(mapping = aes(x = year, y = emissions, color = territory)) +
62   geom_point(alpha = 0.4) +
63   geom_smooth(method = "loess", se = FALSE) +
64   labs(
65   title = "Emissions of Carbon Dioxide by Territory",
66   subtitle = "1980 to 2020",
67   y = expression("Emissions (Mt C y"^{-1}*")"),
68   x = "Year",
69   color = "Territory"
70   )
71 ```
72
73 ## Exercises
74
75 Edit the code in the chunk above to:
76
1:1
Global CO2 Emissions
R Markdown

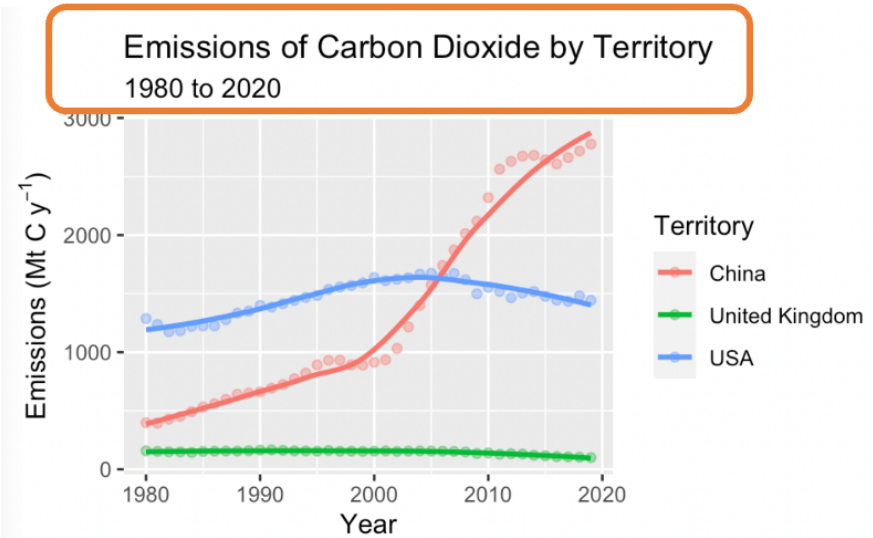

carbon-emissions.Rmd
Knit
Run
Territory CO2 Emissions
Let's create a data visualization that displays how the CO2 emissions of the USA changed over time on a variety of issues, and compares it to three other countries: the United Kingdom, China, and Venezuela.
We can easily change which countries are being plotted by changing which countries the code above `filter`s for.
Note that the country name should be spelled and capitalized exactly the same way as it appears in the data.
See the [Appendix](#appendix) for a list of the countries in the data.
Next, let's make the plot:
```{r plot-yearly-co2-emissions, fig.width=5, fig.height=3, message=FALSE}
emissions %>%
filter(territory %in% c("USA", "United Kingdom", "China") &
year > 1979) %>%
ggplot(mapping = aes(x = year, y = emissions, color = territory)) +
geom_point(alpha = 0.4) +
geom_smooth(method = "loess", se = FALSE) +
labs(
title = "Emissions of Carbon Dioxide by Territory",
subtitle = "1980 to 2020",
y = expression("Emissions (Mt C y"^{-1}*")"),
x = "Year",
color = "Territory"
)
```
Exercises
Edit the code in the chunk above to:
1:1
Global CO2 Emissions
R Markdown

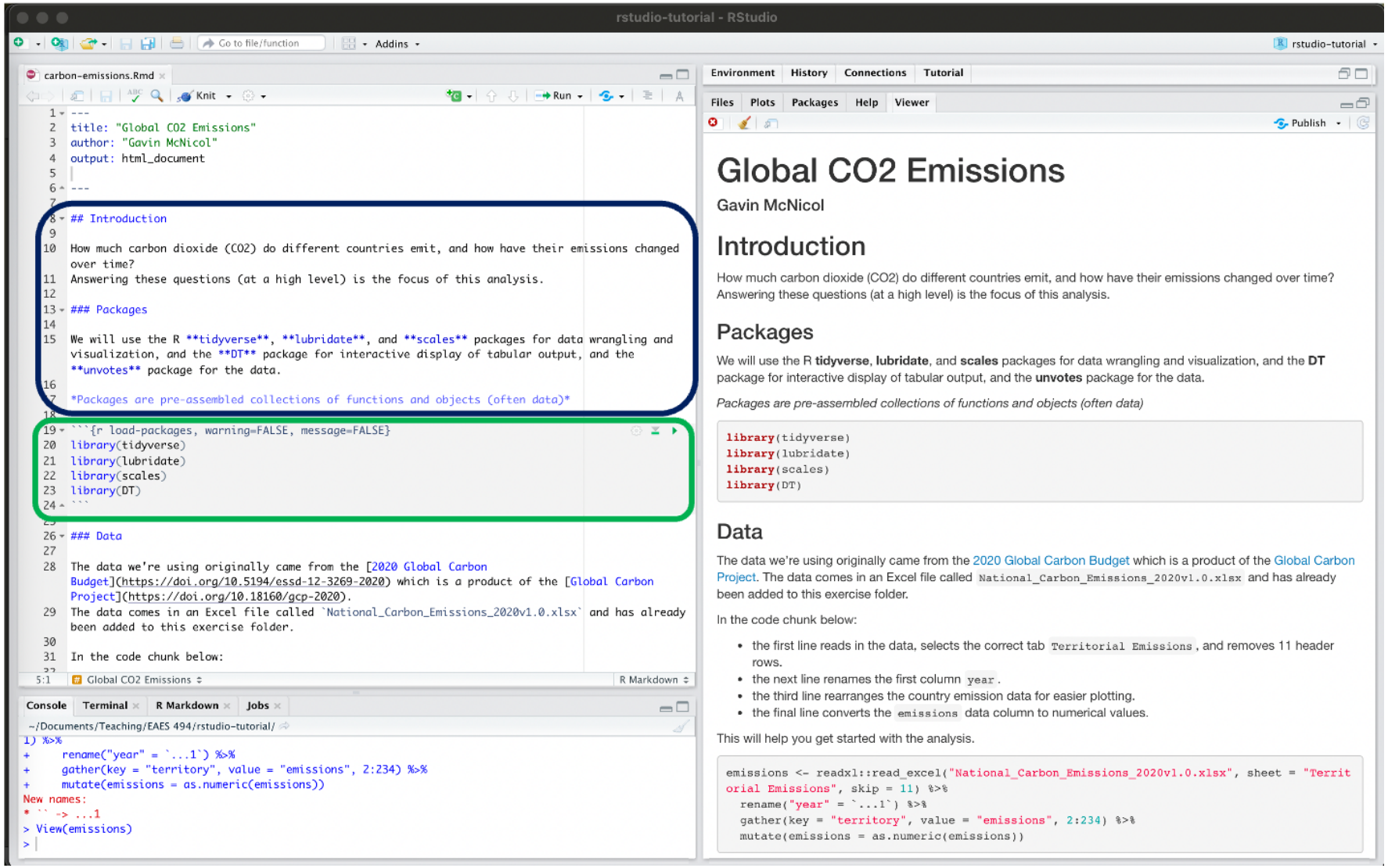
carbon-emissions.Rmd

Knit | Run

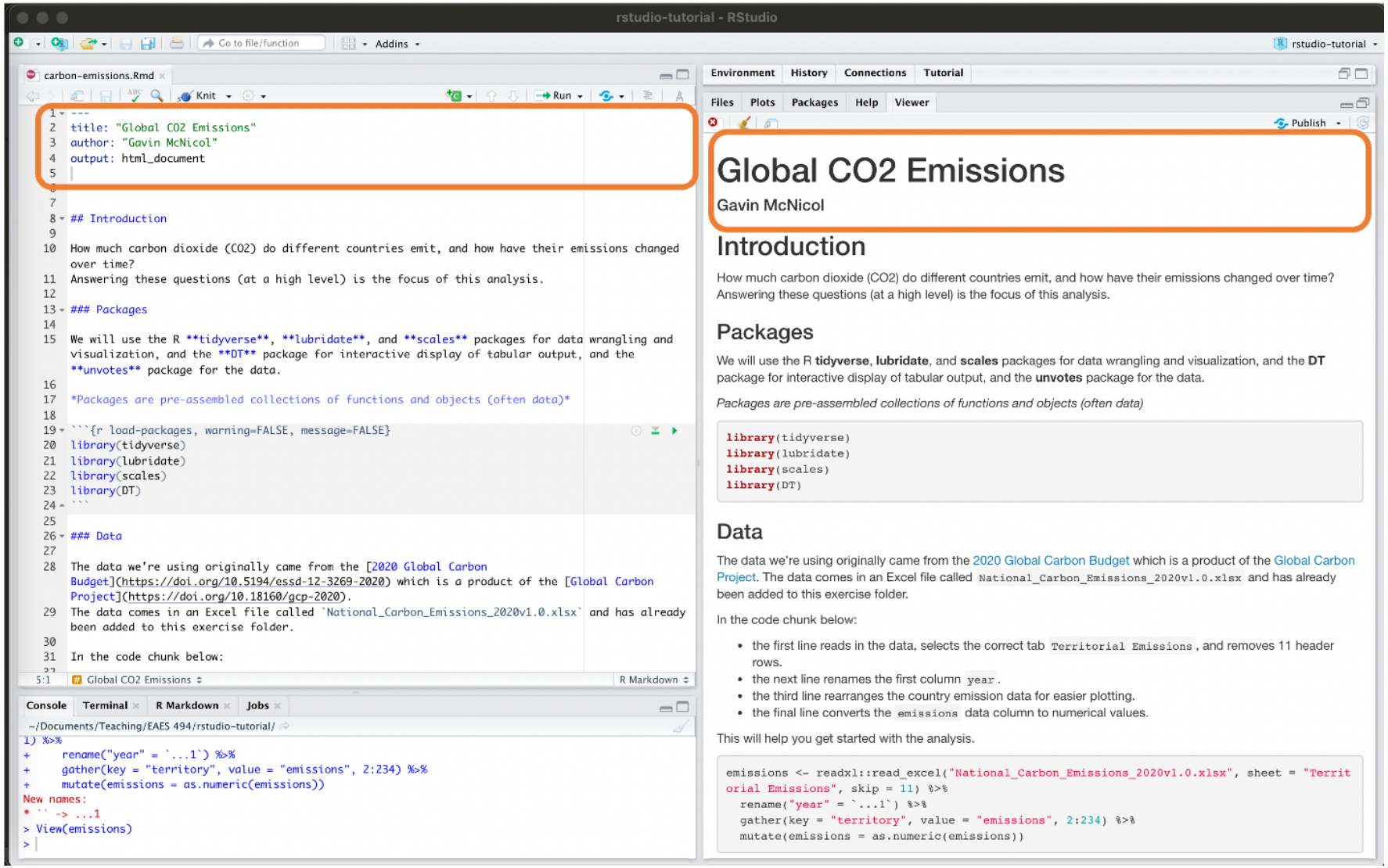
```
44 ```

45

46 ## Territory CO2 Emissions

47

48 Let's create a data visualization that displays how the CO2 emissions of the USA changed over
   time on a variety of issues, and compares it to three other countries: the United Kingdom,
   China, and Venezuela.

49

50 We can easily change which countries are being plotted by changing which countries the code
   above `filter`s for.
51 Note that the country name should be spelled and capitalized exactly the same way as it appears
   in the data.

52

53 See the [Appendix](#appendix) for a list of the countries in the data.

54

55 Next, let's make the plot:

56

57 ```{r plot-yearly-co2-emissions, fig.width=5, fig.height=3, message=FALSE}
58 emissions %>%
59   filter(territory %in% c("USA", "United Kingdom", "China") &
60            year > 1979) %>%
61   ggplot(mapping = aes(x = year, y = emissions, color = territory)) +
62   geom_point(alpha = 0.4) +
63   geom_smooth(method = "loess", se = FALSE) +
64   labs(
65     title = "Emissions of Carbon Dioxide by Territory",
66     subtitle = "1980 to 2020",
67     y = expression("Emissions (Mt C y"^{-1}*")"),
68     x = "Year",
69     color = "Territory"
70   )
71 ```

72

73 ## Exercises

74

75 Edit the code in the chunk above to:

76
```

1:1 | Global CO2 Emissions | R Markdown




rstudio-tutorial - RStudio
carbon-emissions.Rmd
---
title: "Global CO2 Emissions"
author: "Gavin McNicol"
output: html_document

Introduction
How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time?
Answering these questions (at a high level) is the focus of this analysis.
Packages
We will use the R **tidyverse**, **lubridate**, and **scales** packages for data wrangling and visualization, and the **DT** package for interactive display of tabular output, and the **unvotes** package for the data.
Packages are pre-assembled collections of functions and objects (often data)
```{r load-packages, warning=FALSE, message=FALSE}
library(tidyverse)
library(lubridate)
library(scales)
library(DT)
```
Data
The data we're using originally came from the [2020 Global Carbon Budget](https://doi.org/10.5194/essd-12-3269-2020) which is a product of the [Global Carbon Project](https://doi.org/10.18160/gcp-2020).
The data comes in an Excel file called `National_Carbon_Emissions_2020v1.0.xlsx` and has already been added to this exercise folder.
In the code chunk below:
Global CO2 Emissions
R Markdown
Console
Terminal
R Markdown
Jobs
~/Documents/Teaching/EAES 494/rstudio-tutorial/
1) %>%
+ rename("year" = `...1`) %>%
+ gather(key = "territory", value = "emissions", 2:234) %>%
+ mutate(emissions = as.numeric(emissions))
New names:
* `` -> ...1
> View(emissions)
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Publish
Global CO2 Emissions
Gavin McNicol
Introduction
How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time? Answering these questions (at a high level) is the focus of this analysis.
Packages
We will use the R tidyverse, lubridate, and scales packages for data wrangling and visualization, and the DT package for interactive display of tabular output, and the unvotes package for the data.
Packages are pre-assembled collections of functions and objects (often data)
library(tidyverse)
library(lubridate)
library(scales)
library(DT)
Data
The data we're using originally came from the 2020 Global Carbon Budget which is a product of the Global Carbon Project. The data comes in an Excel file called National_Carbon_Emissions_2020v1.0.xlsx and has already been added to this exercise folder.
In the code chunk below:
the first line reads in the data, selects the correct tab Territorial Emissions, and removes 11 header rows.
the next line renames the first column year.
the third line rearranges the country emission data for easier plotting.
the final line converts the emissions data column to numerical values.
This will help you get started with the analysis.
emissions <- readxl::read_excel("National_Carbon_Emissions_2020v1.0.xlsx", sheet = "Territorial Emissions", skip = 11) %>%
rename("year" = `...1`) %>%
gather(key = "territory", value = "emissions", 2:234) %>%
mutate(emissions = as.numeric(emissions))


rstudio-tutorial - RStudio
carbon-emissions.Rmd
---
title: "Global CO2 Emissions"
author: "Gavin McNicol"
output: html_document
Introduction
How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time?
Answering these questions (at a high level) is the focus of this analysis.
Packages
We will use the R **tidyverse**, **lubridate**, and **scales** packages for data wrangling and visualization, and the **DT** package for interactive display of tabular output, and the **unvotes** package for the data.
Packages are pre-assembled collections of functions and objects (often data)
```{r load-packages, warning=FALSE, message=FALSE}
library(tidyverse)
library(lubridate)
library(scales)
library(DT)
```
Data
The data we're using originally came from the [2020 Global Carbon Budget](https://doi.org/10.5194/essd-12-3269-2020) which is a product of the [Global Carbon Project](https://doi.org/10.18160/gcp-2020).
The data comes in an Excel file called `National_Carbon_Emissions_2020v1.0.xlsx` and has already been added to this exercise folder.
In the code chunk below:
Global CO2 Emissions
R Markdown
Console
Terminal
R Markdown
Jobs
~/Documents/Teaching/EAES 494/rstudio-tutorial/
1) %>%
+ rename("year" = `...1`) %>%
+ gather(key = "territory", value = "emissions", 2:234) %>%
+ mutate(emissions = as.numeric(emissions))
New names:
* `` -> ...1
> View(emissions)
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Publish
Global CO2 Emissions
Gavin McNicol
Introduction
How much carbon dioxide (CO2) do different countries emit, and how have their emissions changed over time? Answering these questions (at a high level) is the focus of this analysis.
Packages
We will use the R tidyverse, lubridate, and scales packages for data wrangling and visualization, and the DT package for interactive display of tabular output, and the unvotes package for the data.
Packages are pre-assembled collections of functions and objects (often data)
library(tidyverse)
library(lubridate)
library(scales)
library(DT)
Data
The data we're using originally came from the 2020 Global Carbon Budget which is a product of the Global Carbon Project. The data comes in an Excel file called National_Carbon_Emissions_2020v1.0.xlsx and has already been added to this exercise folder.
In the code chunk below:
the first line reads in the data, selects the correct tab Territorial Emissions, and removes 11 header rows.
the next line renames the first column year.
the third line rearranges the country emission data for easier plotting.
the final line converts the emissions data column to numerical values.
This will help you get started with the analysis.
emissions <- readxl::read_excel("National_Carbon_Emissions_2020v1.0.xlsx", sheet = "Territorial Emissions", skip = 11) %>%
rename("year" = `...1`) %>%
gather(key = "territory", value = "emissions", 2:234) %>%
mutate(emissions = as.numeric(emissions))

Now, your turn!

Open the **exercise-01 carbon emissions** project in RStudio Cloud!