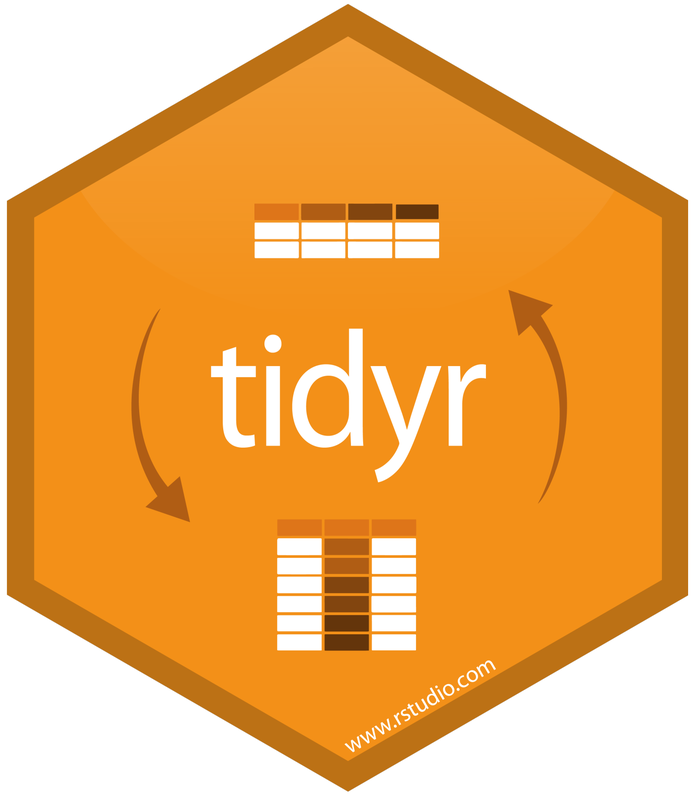
Tidy Data with



In R4DS

Tidy Data

Recovering from Git(Hub) failure

Scenario: You pull and get a merge conflict.

What's the problem?

GitHub can't figure out how to reconcile diffs.

Resolve the conflicts.

Or abort ... and come back later.

Let's create this situation.

Make sure local Git pane is clear.
Make sure local and remote are synced (push, pull).

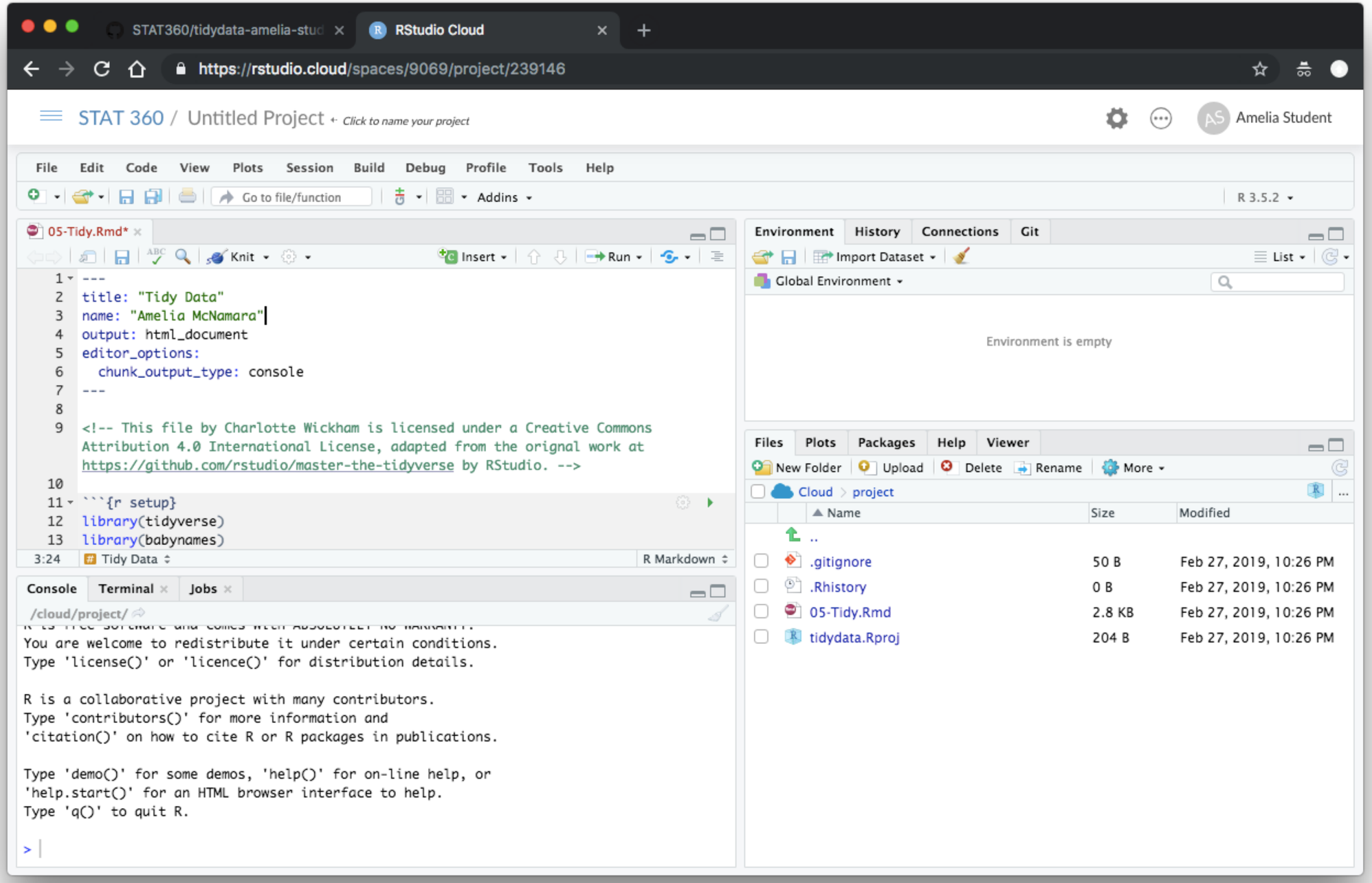
Edit 05-Tidy.Rmd in your RStudio (change my name to your name).

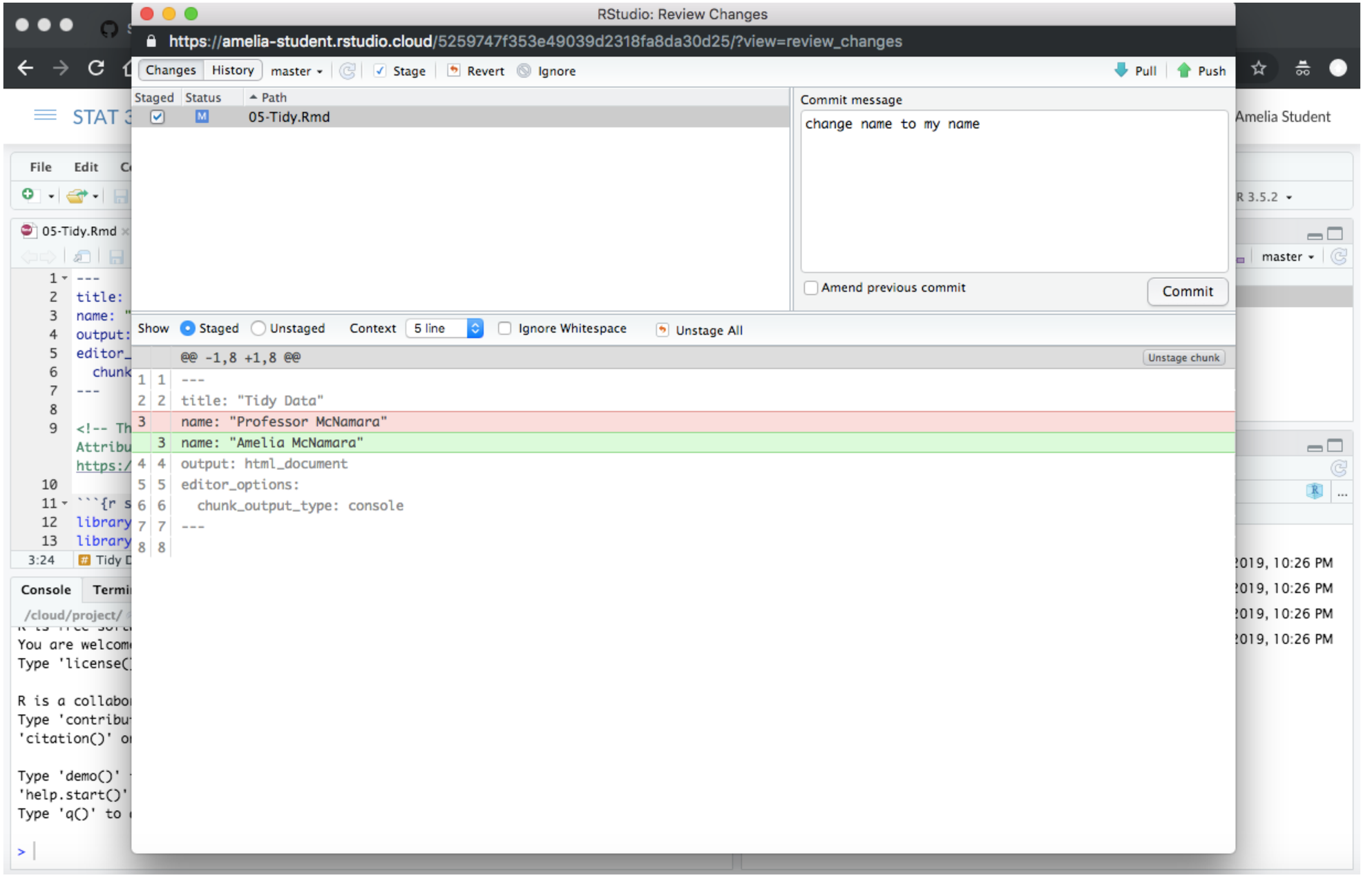
Save your change and commit it .

Go to Github and make a conflicting edit to 05-Tidy.Rmd (delete the line with the name).

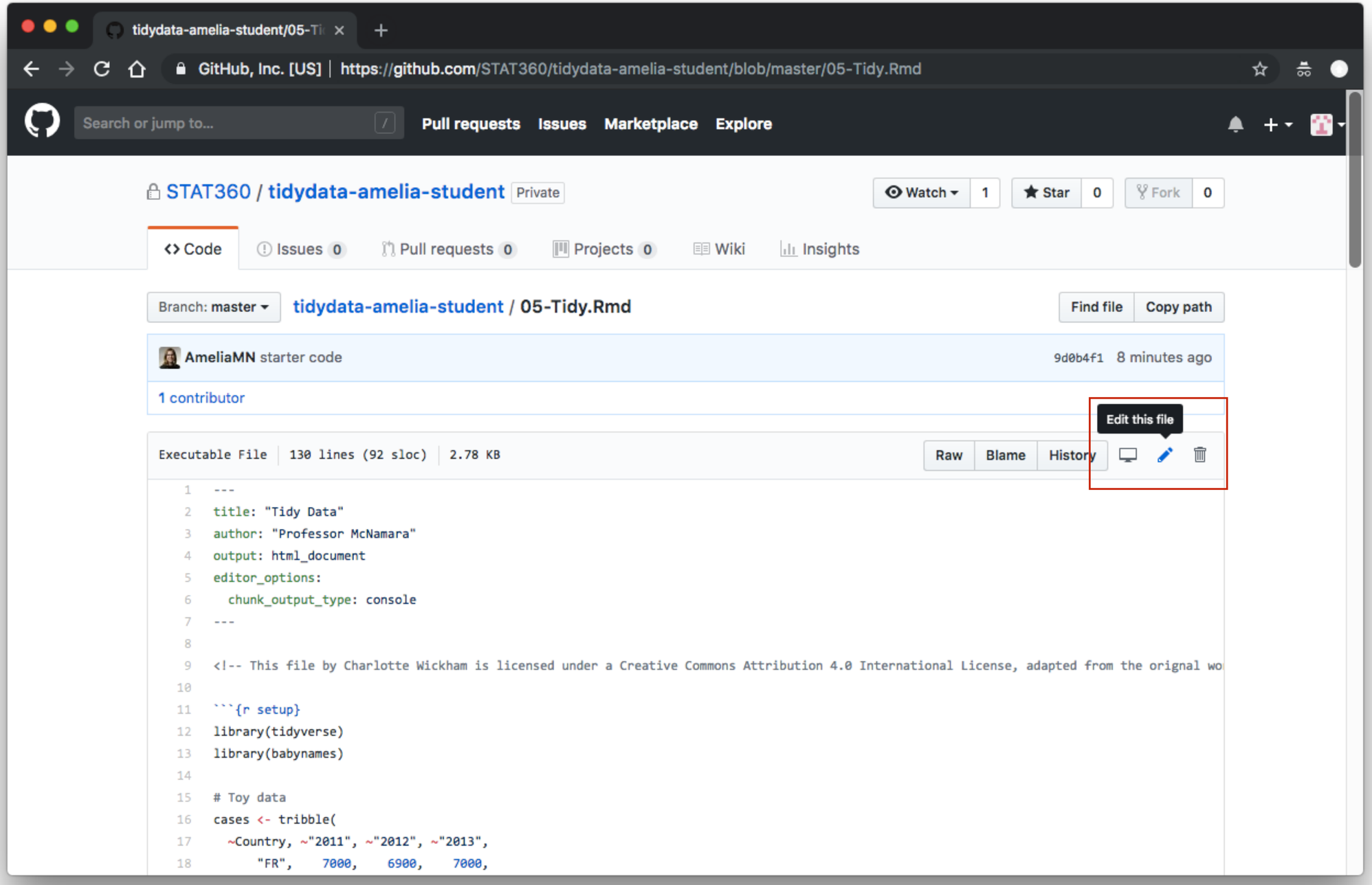
Save your change and commit on Gitub.

Try to push. You will fail. Try to pull. You will fail. All is fail.

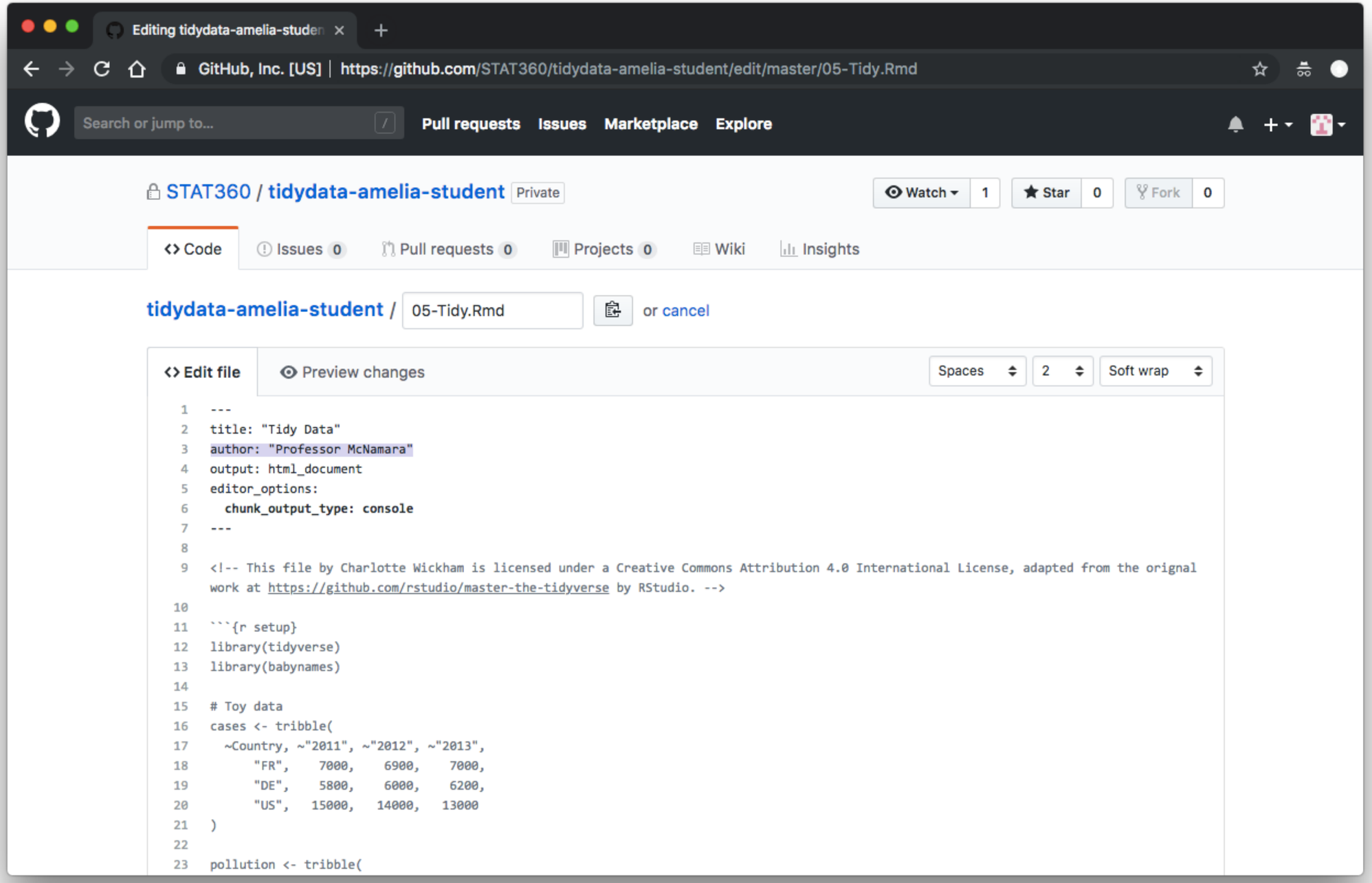

STAT360/tidydata-amelia-stud
RStudio Cloud
https://rstudio.cloud/spaces/9069/project/239146
STAT 360 / Untitled Project Click to name your project
Amelia Student
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 3.5.2
05-Tidy.Rmd*
Knit
Insert
Run
1 ---
2 title: "Tidy Data"
3 name: "Amelia McNamara"
4 output: html_document
5 editor_options:
6 chunk_output_type: console
7 ---
8
9 <!-- This file by Charlotte Wickham is licensed under a Creative Commons
Attribution 4.0 International License, adapted from the orignal work at
https://github.com/rstudio/master-the-tidyverse by RStudio. -->
10
11 ```{r setup}
12 library(tidyverse)
13 library(babynames)
3:24
Tidy Data
R Markdown
Console
Terminal
Jobs
/cloud/project/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
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
Git
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.gitignore 50 B Feb 27, 2019, 10:26 PM
.Rhistory 0 B Feb 27, 2019, 10:26 PM
05-Tidy.Rmd 2.8 KB Feb 27, 2019, 10:26 PM
tidydata.Rproj 204 B Feb 27, 2019, 10:26 PM

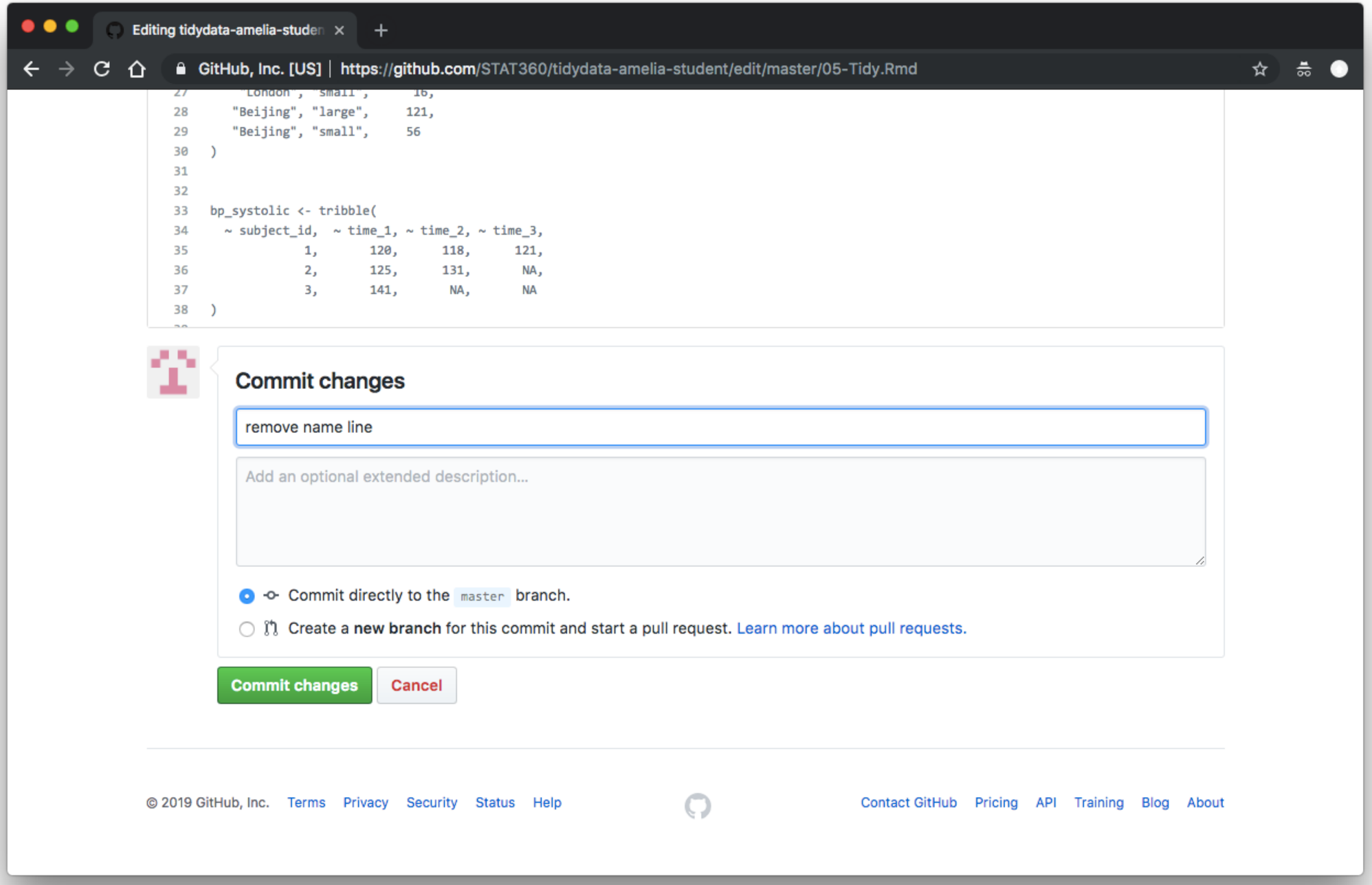

RStudio: Review Changes
https://amelia-student.rstudio.cloud/5259747f353e49039d2318fa8da30d25/?view=review_changes
Changes
History
master
Stage
Revert
Ignore
Pull
Push
Staged
Status
Path
05-Tidy.Rmd
Commit message
change name to my name
Amend previous commit
Commit
Show
Staged
Unstaged
Context
5 line
Ignore Whitespace
Unstage All
@@ -1,8 +1,8 @@
Unstage chunk
---
title: "Tidy Data"
name: "Professor McNamara"
name: "Amelia McNamara"
output: html_document
editor_options:
chunk_output_type: console

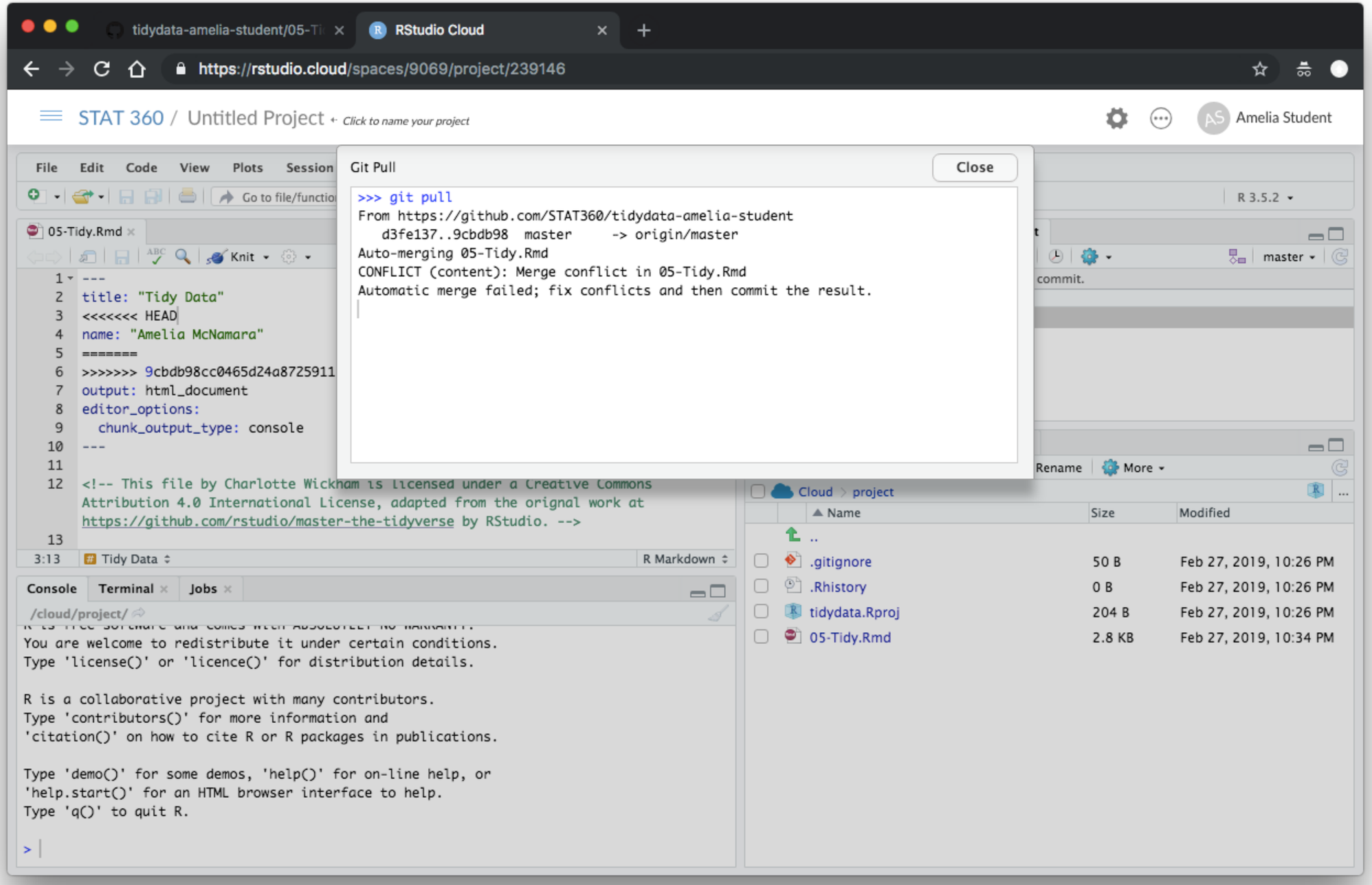
Amelia Student
File
Edit
05-Tidy.Rmd
Console
Terminal
/cloud/project/
You are welcome
Type 'license()'
R is a collabo
Type 'contribu
'citation()' o
Type 'demo()'
'help.start()'
Type 'q()' to
2019, 10:26 PM
2019, 10:26 PM
2019, 10:26 PM
2019, 10:26 PM

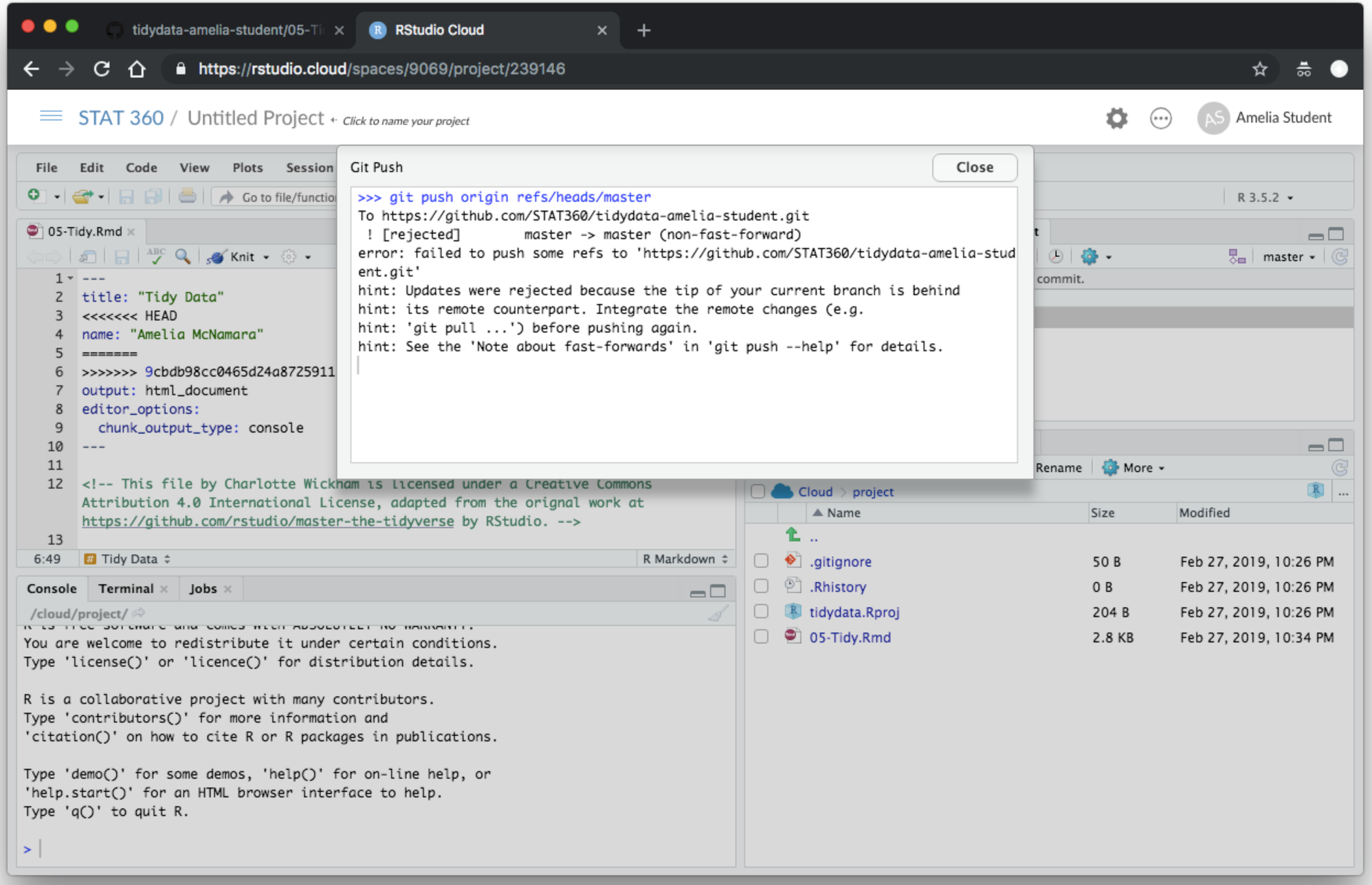

tidydata-amelia-student/05-Ti
GitHub, Inc. [US] | https://github.com/STAT360/tidydata-amelia-student/blob/master/05-Tidy.Rmd
Search or jump to...
Pull requests
Issues
Marketplace
Explore
STAT360 / tidydata-amelia-student
Private
Watch 1
Star 0
Fork 0
Code
Issues 0
Pull requests 0
Projects 0
Wiki
Insights
Branch: master
tidydata-amelia-student / 05-Tidy.Rmd
Find file
Copy path
AmeliaMN starter code
9d0b4f1 8 minutes ago
1 contributor
Edit this file
Executable File | 130 lines (92 sloc) | 2.78 KB
Raw
Blame
History
---
title: "Tidy Data"
author: "Professor McNamara"
output: html_document
editor_options:
chunk_output_type: console

<!-- This file by Charlotte Wickham is licensed under a Creative Commons Attribution 4.0 International License, adapted from the orignal wo
```{r setup}
library(tidyverse)
library(babynames)
# Toy data
cases <- tribble(
~Country, ~"2011", ~"2012", ~"2013",
"FR", 7000, 6900, 7000,
```

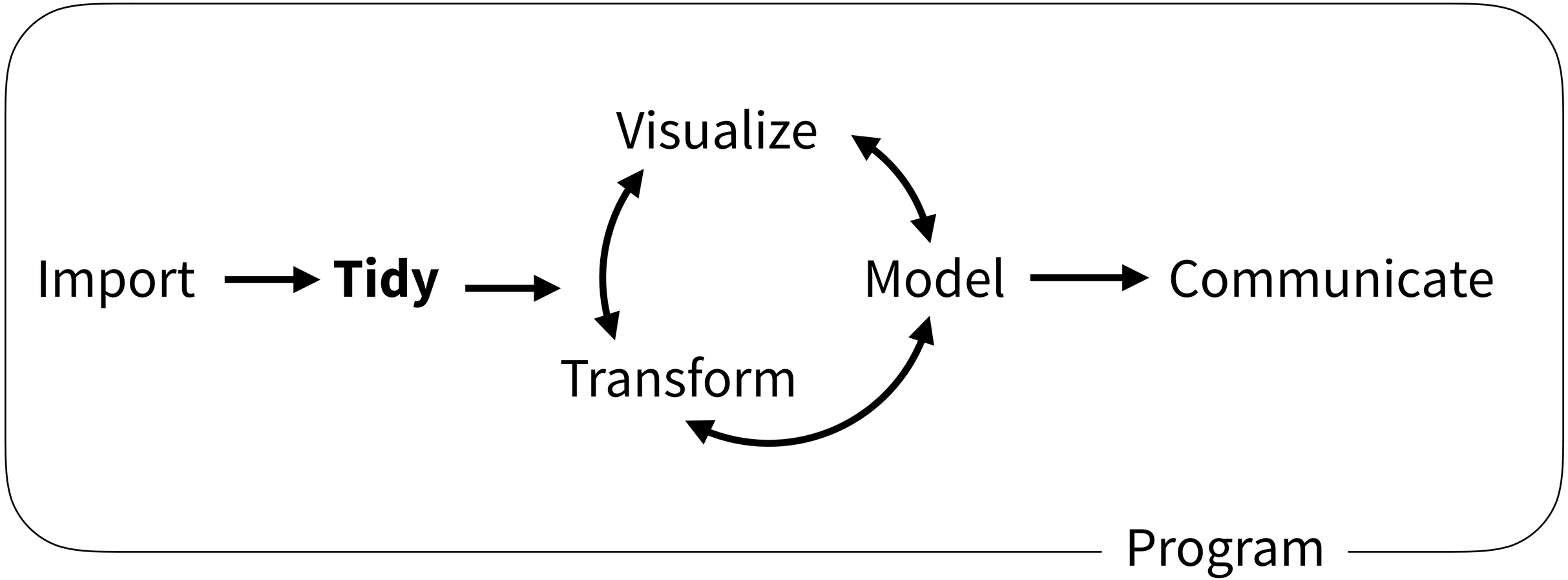

Editing tidydata-amelia-studen
GitHub, Inc. [US] | https://github.com/STAT360/tidydata-amelia-student/edit/master/05-Tidy.Rmd
Search or jump to...
Pull requests
Issues
Marketplace
Explore
STAT360 / tidydata-amelia-student
Private
Watch 1
Star 0
Fork 0
Code
Issues 0
Pull requests 0
Projects 0
Wiki
Insights
tidydata-amelia-student / 05-Tidy.Rmd
or cancel
Edit file
Preview changes
Spaces
2
Soft wrap
1 ---
2 title: "Tidy Data"
3 author: "Professor McNamara"
4 output: html_document
5 editor_options:
6 chunk_output_type: console
7 ---
8
9 <!-- This file by Charlotte Wickham is licensed under a Creative Commons Attribution 4.0 International License, adapted from the orignal work at https://github.com/rstudio/master-the-tidyverse by RStudio. -->
10
11 ```{r setup}
12 library(tidyverse)
13 library(babynames)
14
15 # Toy data
16 cases <- tribble(
17 ~Country, ~"2011", ~"2012", ~"2013",
18 "FR", 7000, 6900, 7000,
19 "DE", 5800, 6000, 6200,
20 "US", 15000, 14000, 13000
21)
22
23 pollution <- tribble(


Editing tidydata-amelia-studen
GitHub, Inc. [US] | https://github.com/STAT360/tidydata-amelia-student/edit/master/05-Tidy.Rmd
"London", "small", 16,
"Beijing", "large", 121,
"Beijing", "small", 56
)
bp_systolic <- tribble(
~ subject_id, ~ time_1, ~ time_2, ~ time_3,
1, 120, 118, 121,
2, 125, 131, NA,
3, 141, NA, NA
)
Commit changes
remove name line
Add an optional extended description...
Commit directly to the master branch.
Create a new branch for this commit and start a pull request. Learn more about pull requests.
Commit changes
Cancel
© 2019 GitHub, Inc. Terms Privacy Security Status Help
Contact GitHub Pricing API Training Blog About


tidydata-amelia-student/05-Ti
RStudio Cloud
https://rstudio.cloud/spaces/9069/project/239146
STAT 360 / Untitled Project Click to name your project
Amelia Student
File Edit Code View Plots Session
Git Pull
Close
>>> git pull
From https://github.com/STAT360/tidydata-amelia-student
d3fe137..9cbdb98 master -> origin/master
Auto-merging 05-Tidy.Rmd
CONFLICT (content): Merge conflict in 05-Tidy.Rmd
Automatic merge failed; fix conflicts and then commit the result.
R 3.5.2
05-Tidy.Rmd
Knit
1 ---
2 title: "Tidy Data"
3 <<<<<<< HEAD
4 name: "Amelia McNamara"
5 =======
6 >>>>>>> 9cbdb98cc0465d24a8725911
7 output: html_document
8 editor_options:
9 chunk_output_type: console
10 ---
11
12 <!-- This file by Charlotte Wickham is licensed under a Creative Commons
Attribution 4.0 International License, adapted from the orignal work at
https://github.com/rstudio/master-the-tidyverse by RStudio. -->
13
3:13 Tidy Data R Markdown
master
commit.
Rename More
Cloud project
Name Size Modified
.. 50 B Feb 27, 2019, 10:26 PM
.gitignore 0 B Feb 27, 2019, 10:26 PM
.Rhistory 204 B Feb 27, 2019, 10:26 PM
tidydata.Rproj 2.8 KB Feb 27, 2019, 10:34 PM
05-Tidy.Rmd
Console Terminal Jobs
/cloud/project/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
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.
>


tidydata-amelia-student/05-Ti
RStudio Cloud
https://rstudio.cloud/spaces/9069/project/239146
STAT 360 / Untitled Project Click to name your project
Amelia Student
File Edit Code View Plots Session
Git Push
Close
>>> git push origin refs/heads/master
To https://github.com/STAT360/tidydata-amelia-student.git
! [rejected] master -> master (non-fast-forward)
error: failed to push some refs to 'https://github.com/STAT360/tidydata-amelia-student.git'
hint: Updates were rejected because the tip of your current branch is behind
hint: its remote counterpart. Integrate the remote changes (e.g.
hint: 'git pull ...') before pushing again.
hint: See the 'Note about fast-forwards' in 'git push --help' for details.
R 3.5.2
05-Tidy.Rmd
Knit
1 ---
2 title: "Tidy Data"
3 <<<<<<< HEAD
4 name: "Amelia McNamara"
5 =======
6 >>>>>>> 9cbdb98cc0465d24a8725911
7 output: html_document
8 editor_options:
9 chunk_output_type: console
10 ---
11
12 <!-- This file by Charlotte Wickham is licensed under a Creative Commons Attribution 4.0 International License, adapted from the orignal work at https://github.com/rstudio/master-the-tidyverse by RStudio. -->
13
6:49 Tidy Data R Markdown
master
commit.
Rename More
Cloud project
Name Size Modified
..
.gitignore 50 B Feb 27, 2019, 10:26 PM
.Rhistory 0 B Feb 27, 2019, 10:26 PM
tidydata.Rproj 204 B Feb 27, 2019, 10:26 PM
05-Tidy.Rmd 2.8 KB Feb 27, 2019, 10:34 PM
Console Terminal Jobs
/cloud/project/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
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.
>

(Applied) **Data Science**



Tidy tools

Tidy tools

Functions are easiest to use when they are:

1. **Simple** - They do one thing, and they do it well
2. **Composable** - They can be combined with other functions for multi-step operations
3. **Smart** - They can use R objects as input.

Tidy functions do these things in a specific way.

Tidy tools

Functions are easiest to use when they are:

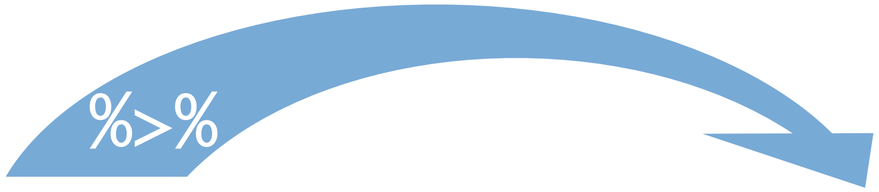
1. **Simple** - They do one thing, and they do it well
2. **Composable** - They can be combined with other functions for multi-step operations
3. **Smart** - They can use R objects as input.

Tidy functions do these things in a specific way.

1. Simple - They do one thing, and they do it well

filter() - extract **cases**
arrange() - reorder **cases**
group_by() - group **cases**
select() - extract **variables**
mutate() - create new **variables**
summarise() - summarise **variables** / create **cases**

2. **Composable** - They can be combined with other functions for multi-step operations




```
babynames          mutate(_____, percent = prop * 100)
```

Each dplyr function takes a data frame as its first argument and returns a data frame. As a result, you can directly pipe the output of one function into the next.

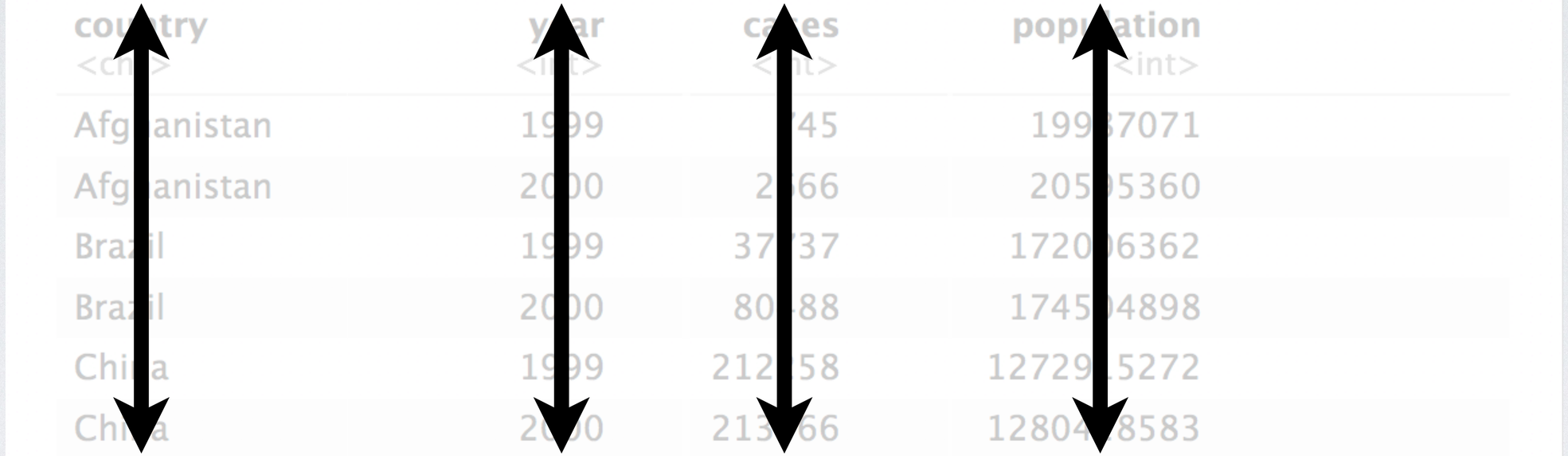
"Data are not just numbers,
they are numbers with a context."

- George Cobb and David Moore (1997)

Consider

What are the variables in this data set?

```
table1
```



country	year	cases	population
<chr>	<int>	<int>	<int>
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

6 rows

Consider

What are the variables in this data set?

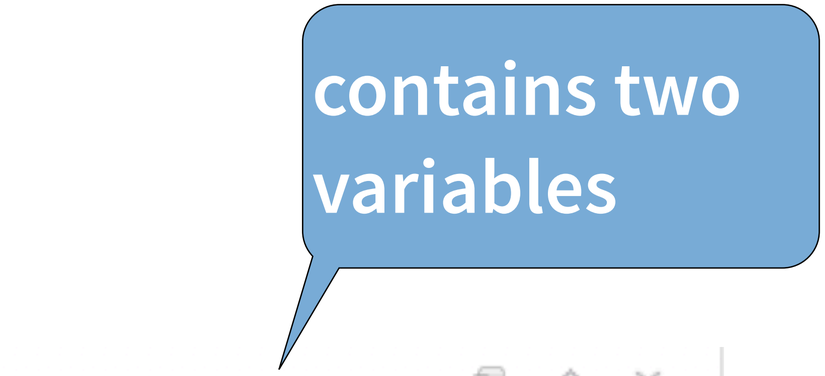
```
table2
```

country	year	type	count
<chr>	<int>	<chr>	<int>
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272

1–10 of 12 rows

Previous 1 2 Next

`table2` isn't tidy



country <chr>	year <int>	type <chr>	count <int>
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272

1–10 of 12 rows

Previous 1 2 Next

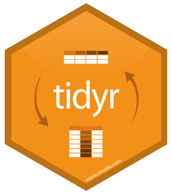
"Data comes in many formats, but R prefers just one: tidy data. "

- Garrett Grolemund

Tidy data

A data set is **tidy** iff:

1. Each **variable** is in its own **column**
2. Each **case** is in its own **row**
3. Each **value** is in its own **cell**



Your Turn 1

Is bp_systolic tidy? What are the variables?

Your Turn 1

Is bp_systolic tidy? What are the variables?

subject_id <dbl>	time_1 <dbl>	time_2 <dbl>	time_3 <dbl>
1	120	118	121
2	125	131	*NA*
3	141	*NA*	*NA*

3 rows

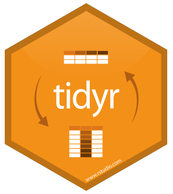
Variables:

- subject
- time
- systolic blood pressure

bp_systolic2 is tidy

subject_id <dbl>	time <dbl>	systolic <dbl>
1	1	120
1	2	118
1	3	121
2	1	125
2	2	131
3	1	141

6 rows



Your Turn 2

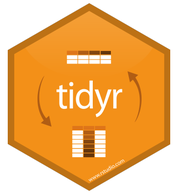
Using `bp_systolic2` with `group_by()` and `summarise()`:

- Find the average systolic blood pressure for each subject

- Find the last time each subject was measured

```
bp_systolic2 %>%
  group_by(subject_id) %>%
  summarise(avg_sys = mean(systolic),
    last_measurement = max(time))
```

```
# A tibble: 3 x 3
  subject_id avg_sys last_measurement
       <dbl>   <dbl>            <dbl>
1          1    120.                3
2          2    128                 2
3          3    141                 1
```

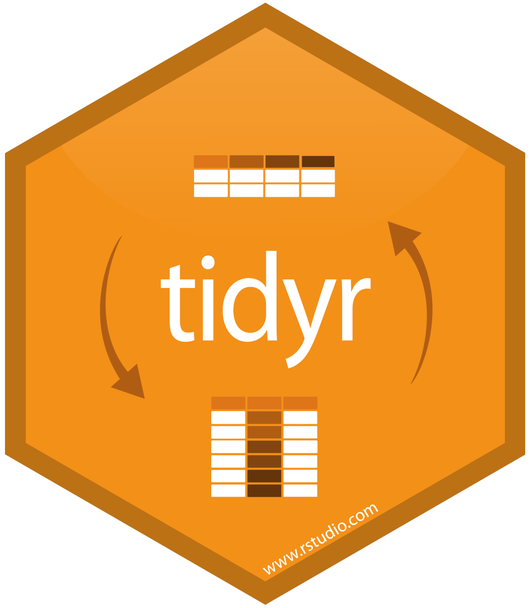


"Tidy data sets are all alike; but every messy data set is messy in its own way."

- Hadley Wickham

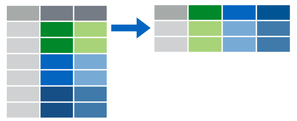
tidyr

tidyr

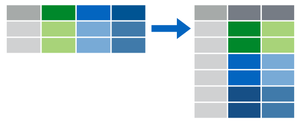


A package that reshapes the layout of tabular data.

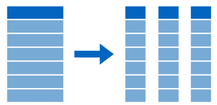
tidyr verbs



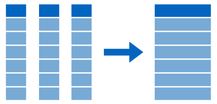
Move values into column names with **spread()**



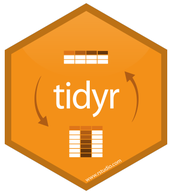
Move column names into values with **gather()**



Split a column with **separate()** or **separate_rows()**

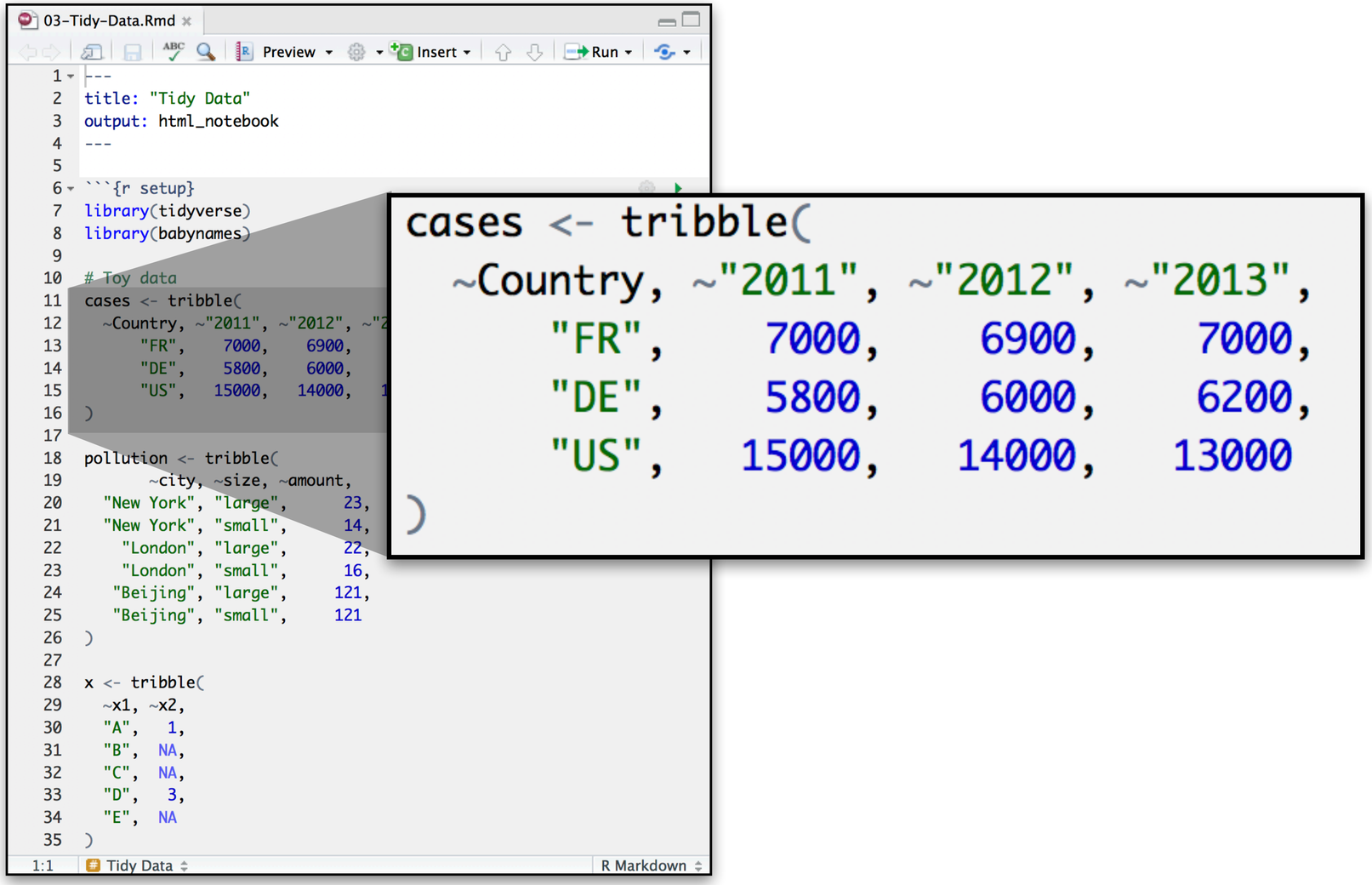


Unite columns with **unite()**



gather()

Toy data



Consider

What are the variables in cases?

Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Consider

What are the variables in cases?

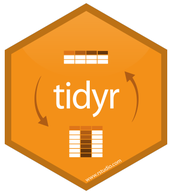
- Country
- Year
- Count

Your Turn 3

On a sheet of paper, draw how the cases data set would look if it had the same values grouped into three columns: *country*, *year*, *n*

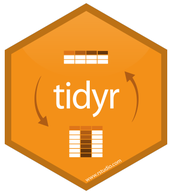
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000



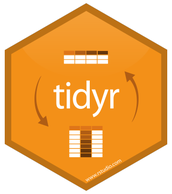
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n



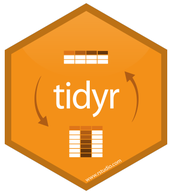
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000



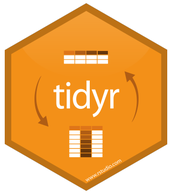
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800



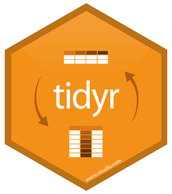
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000



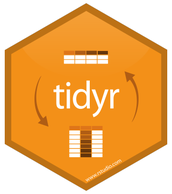
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900



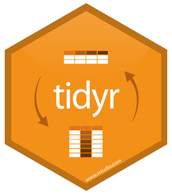
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000



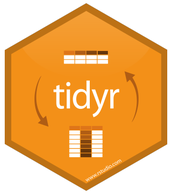
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000



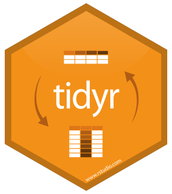
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000



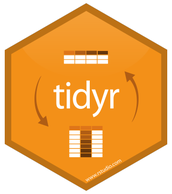
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200

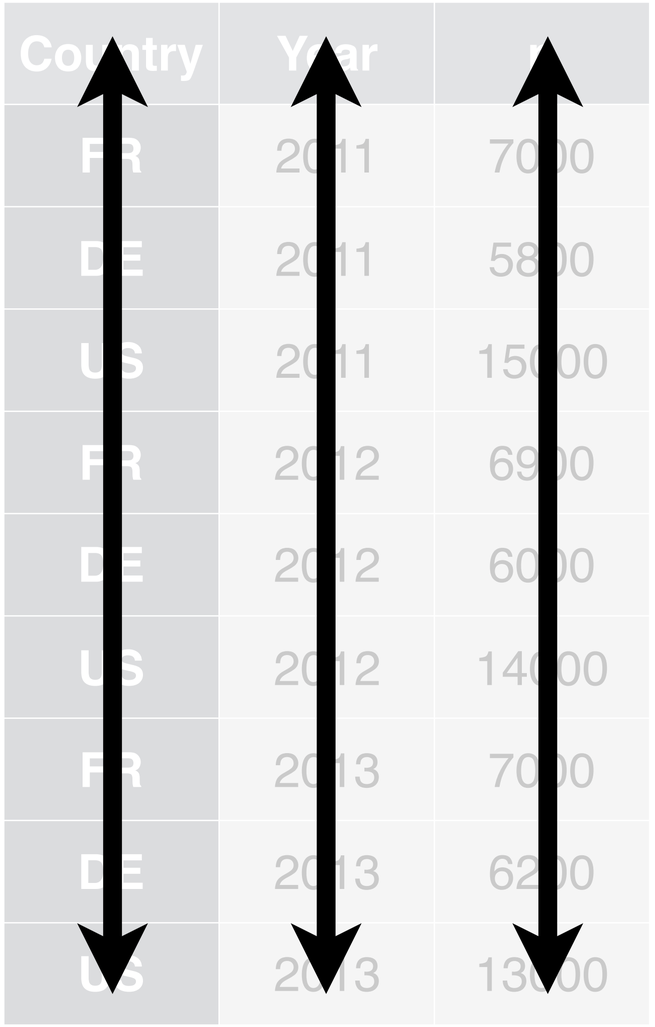


Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

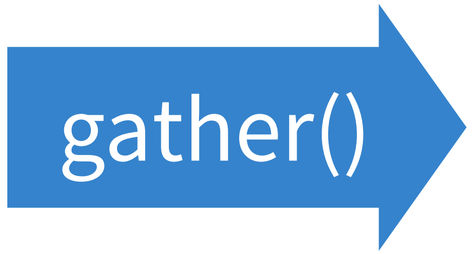
Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000



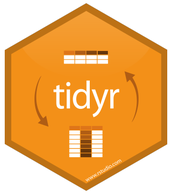
Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000



Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000



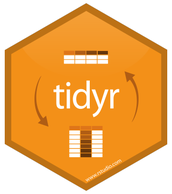
Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000



Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

1 2

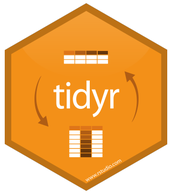
Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000



Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

key (former column names)

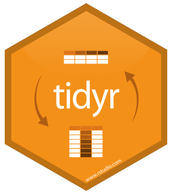
Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000



key **value** (former cells)

Country	2011	2012	2013
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

Country	Year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000



gather()

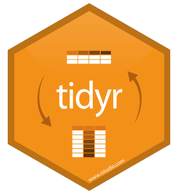
```
cases %>% gather(key = "year", value = "n", 2:4)
```

data frame to reshape

name of the new key column (a character string)

name of the new value column (a character string)

numeric indexes of columns to collapse (or names)

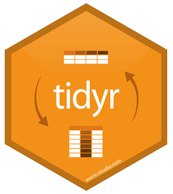


gather()

```
cases %>% gather("year","n", 2:4)
```

numeric indexes

	2	3	4
Country <chr>	**2011** <dbl>	**2012** <dbl>	**2013** <dbl>
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

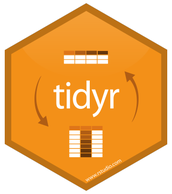


gather()

```
cases %>% gather("year","n", "2011", "2012", "2013")
```

names

	2011	2012	2013
Country <chr>	**2011** <dbl>	**2012** <dbl>	**2013** <dbl>
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000

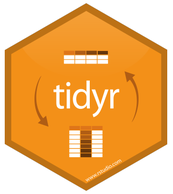


gather()

```
cases %>% gather("year","n", -Country)
```

Everything except…

	Not Country	Not Country	Not Country
Country <chr>	**2011** <dbl>	**2012** <dbl>	**2013** <dbl>
FR	7000	6900	7000
DE	5800	6000	6200
US	15000	14000	13000



Your Turn 4

Use **gather()** to reorganize **table4a** into three columns: *country*, *year*, and *cases*.

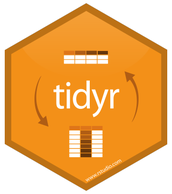
	country <chr>	1999 <int>	2000 <int>
1	Afghanistan	745	2666
2	Brazil	37737	80488
3	China	212258	213766

3 rows

```
table4a %>%
  gather(key = "year", value = "n", 2:3)
```

country <chr>	year <chr>	n <int>
Afghanistan	1999	745
Brazil	1999	37737
China	1999	212258
Afghanistan	2000	2666
Brazil	2000	80488
China	2000	213766

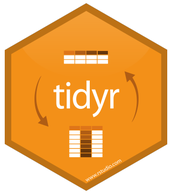
6 rows



```
table4a %>%
  gather(key = "year", value = "n", 2:3, convert = TRUE)
```

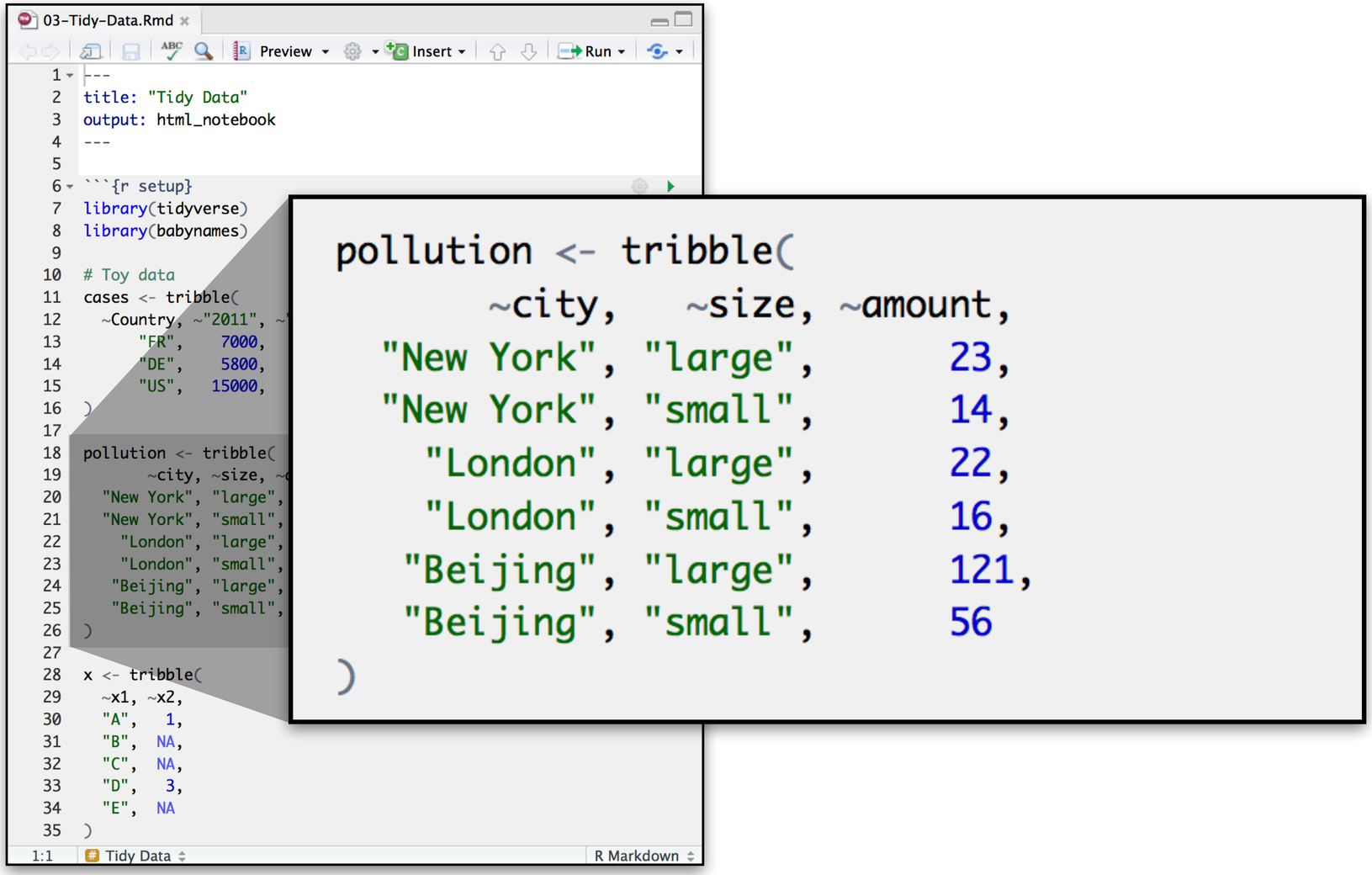
country <chr>	year <int>	n <int>
Afghanistan	1999	745
Brazil	1999	37737
China	1999	212258
Afghanistan	2000	2666
Brazil	2000	80488
China	2000	213766

6 rows



spread()

Toy data



Consider

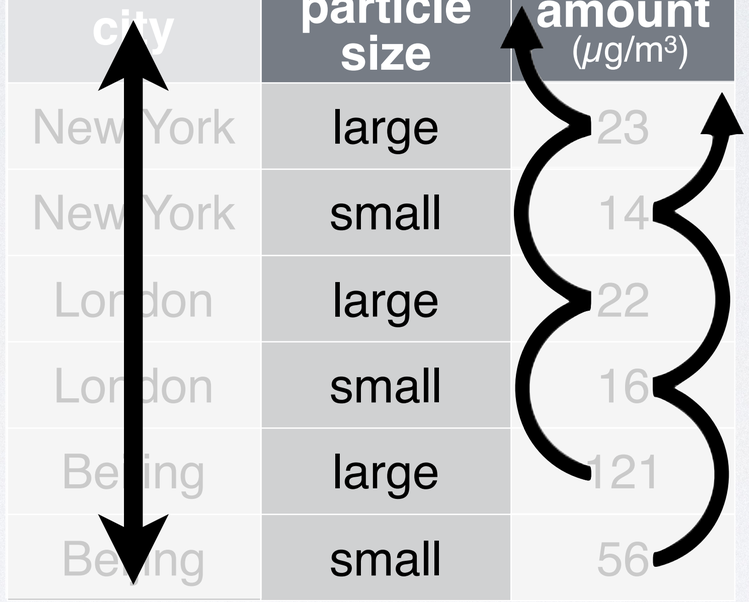
What are the variables in pollution?

city	particle size	amount ($\mu g/m^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

Consider

What are the variables in pollution?

city	particle size	amount ($\mu g/m^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56



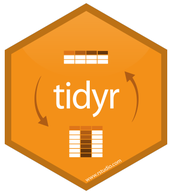
- City
- Amount of large particulate
- Amount of small particulate

Your Turn 5

On a sheet of paper, draw how this data set would look if it had the same values grouped into three columns: *city*, *large*, *small*

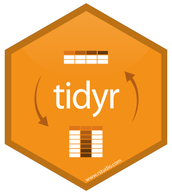
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56



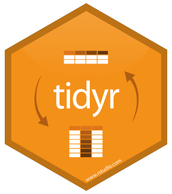
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small



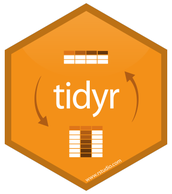
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	



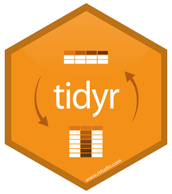
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14



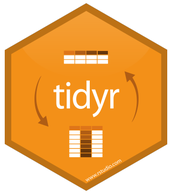
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14
London	22	



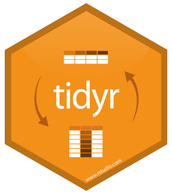
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14
London	22	16



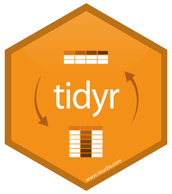
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14
London	22	16
Beijing	121	



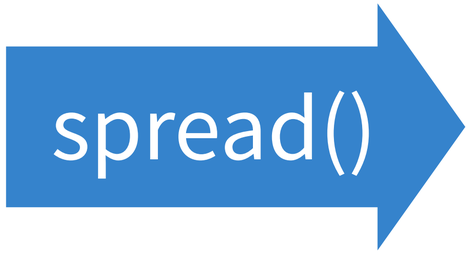
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14
London	22	16
Beijing	121	56

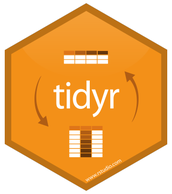


city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

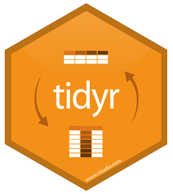


city	large	small
New York	23	14
London	22	16
Beijing	121	56



	1	2
city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

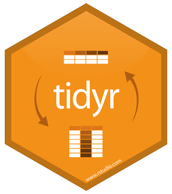
city	large	small
New York	23	14
London	22	16
Beijing	121	56



key (new column names)

city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

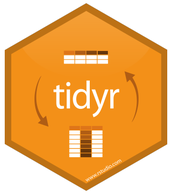
city	large	small
New York	23	14
London	22	16
Beijing	121	56



key **value** (new cells)

city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	large	small
New York	23	14
London	22	16
Beijing	121	56



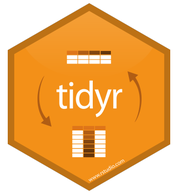
spread()

```
pollution %>% spread(key = size, value = amount)
```

data frame to reshape

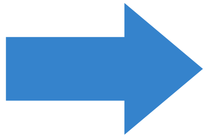
column to use for keys (becomes new column names)

column to use for values (becomes new column cells)

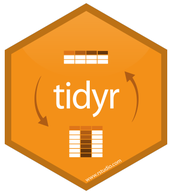


```
pollution %>% spread(size, amount)
```

	city	size	amount
1	New York	large	23
2	New York	small	14
3	London	large	22
4	London	small	16
5	Beijing	large	121
6	Beijing	small	56



	city	large	small
1	Beijing	121	56
2	London	22	16
3	New York	23	14



Your Turn 6

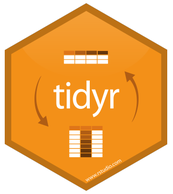
Use **spread()** to reorganize **table2** into four columns: *country*, *year*, *cases*, and *population*.

country <chr>	year <int>	type <chr>	count <int>
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362

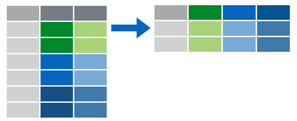
```
table2 %>%
  spread(key = type, value = count)
```

	country <chr>	year <int>	cases <int>	population <int>
1	Afghanistan	1999	745	19987071
2	Afghanistan	2000	2666	20595360
3	Brazil	1999	37737	172006362
4	Brazil	2000	80488	174504898
5	China	1999	212258	1272915272
6	China	2000	213766	1280428583

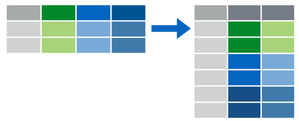
6 rows



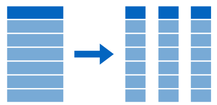
tidyr verbs



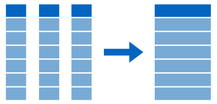
Move values into column names with **spread()**



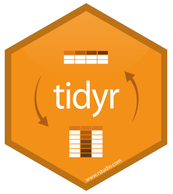
Move column names into values with **gather()**



Split a column with **separate()** or **separate_rows()**



Unite columns with **unite()**



Tidy Data with