More classes

S4

S4 was the second OOP system introduced to R. It is much more formal than S3, which means it can be harder to use but is also more rigorous

Uses special functions to explicitly define classes (`setClass()`), generics (`setGeneric()`), and methods (`setMethod()`).

One way to identify if an object you are looking at is an S4 object is to look for "slots" (accessed using the `@` operator, much like we use `$` in base R)

S4

The group that has most embraced S4 is the Bioconductor community, who have been using almost exclusively S4 since at least 2004.

Bioconductor is analogous to CRAN, and hosts packages related to bioinformatics. Bioinformatics data is much more complicated than the typical "tidy" data we have been thinking about, so it benefits from the added structure of S4.

lubridate

Let's start by looking at a simple example, the Period class in the lubridate package

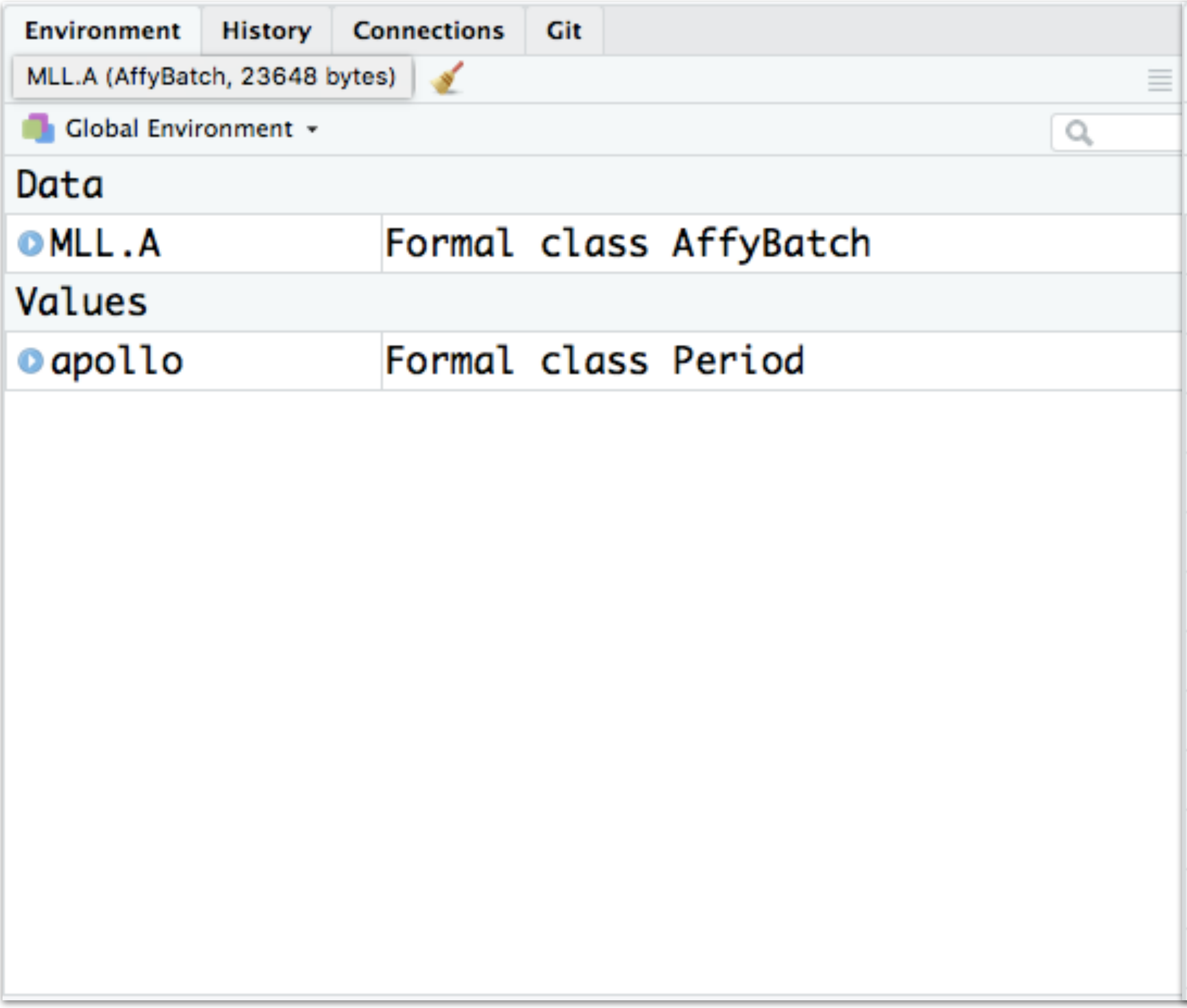
We can use it to define time periods between dates and times. For example, the time since the Apollo launch

```
apollo <-  days(today()-mdy("07-16-1969"))
```

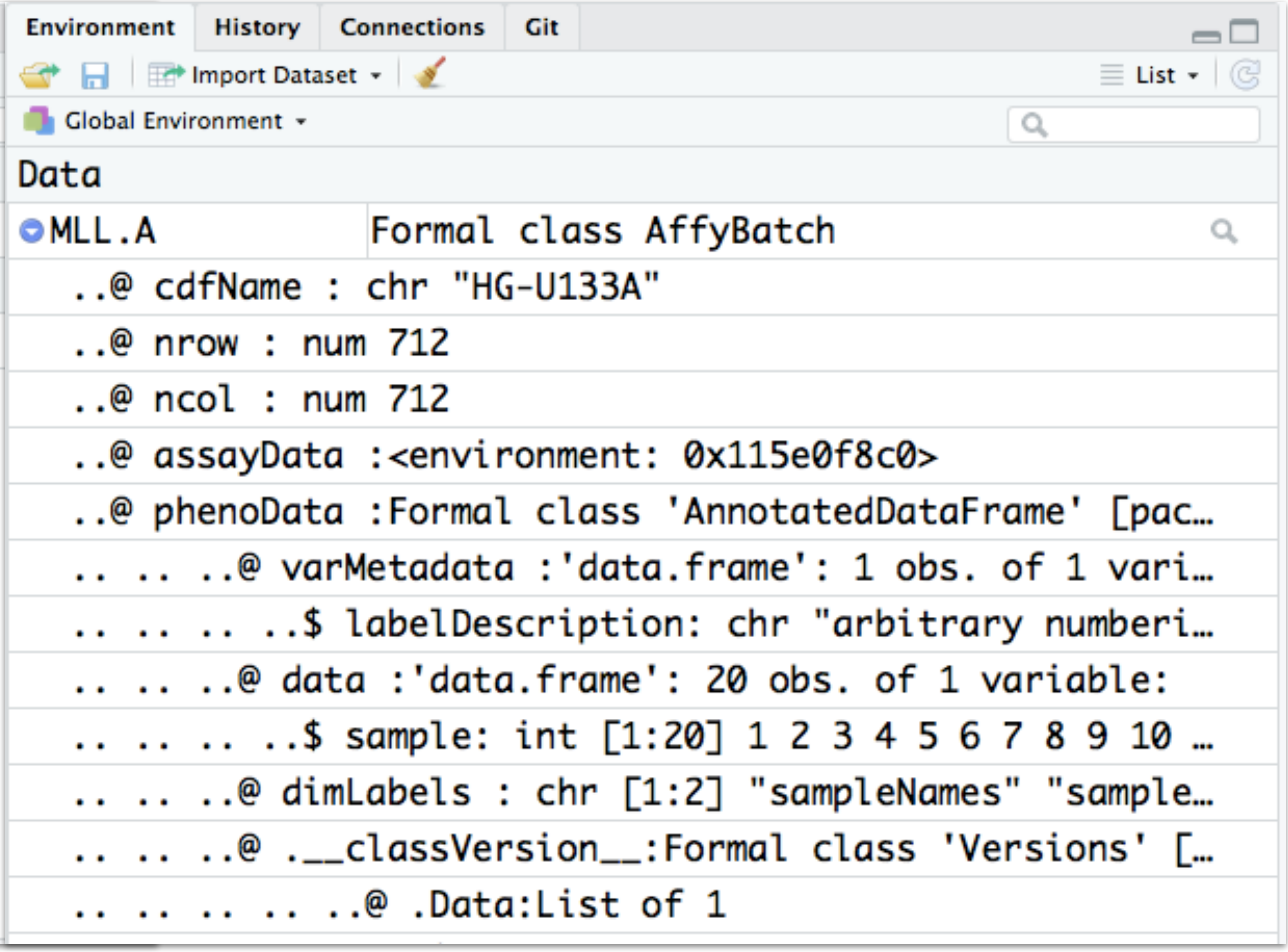
Your Turn

Make an object of class Period and examine it in RStudio.
What slots does the object have? Which are being used?

ALLMLL package data

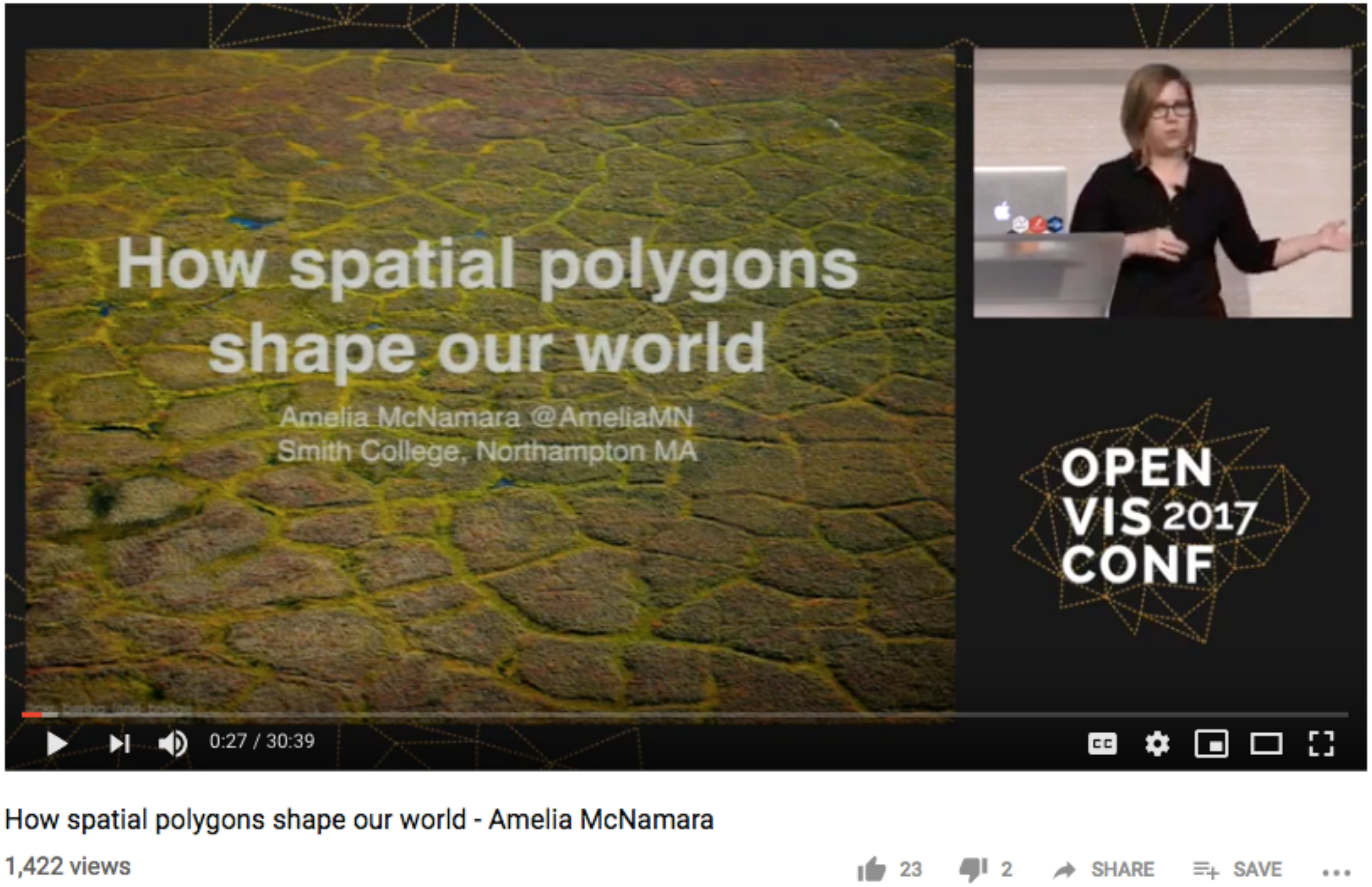


```
Data
MLL.A          Formal class AffyBatch
Values
apollo         Formal class Period
```



```
Data
MLL.A          Formal class AffyBatch
  ..@ cdfName : chr "HG-U133A"
  ..@ nrow : num 712
  ..@ ncol : num 712
  ..@ assayData :<environment: 0x115e0f8c0>
  ..@ phenoData :Formal class 'AnnotatedDataFrame' [pac…
  .. .. ..@ varMetadata :'data.frame': 1 obs. of 1 vari…
  .. .. .. ..$ labelDescription: chr "arbitrary numberi…
  .. .. ..@ data :'data.frame': 20 obs. of 1 variable:
  .. .. .. ..$ sample: int [1:20] 1 2 3 4 5 6 7 8 9 10 …
  .. .. ..@ dimLabels : chr [1:2] "sampleNames" "sample…
  .. .. ..@ .__classVersion__:Formal class 'Versions' […
  .. .. .. .. ..@ .Data:List of 1
```

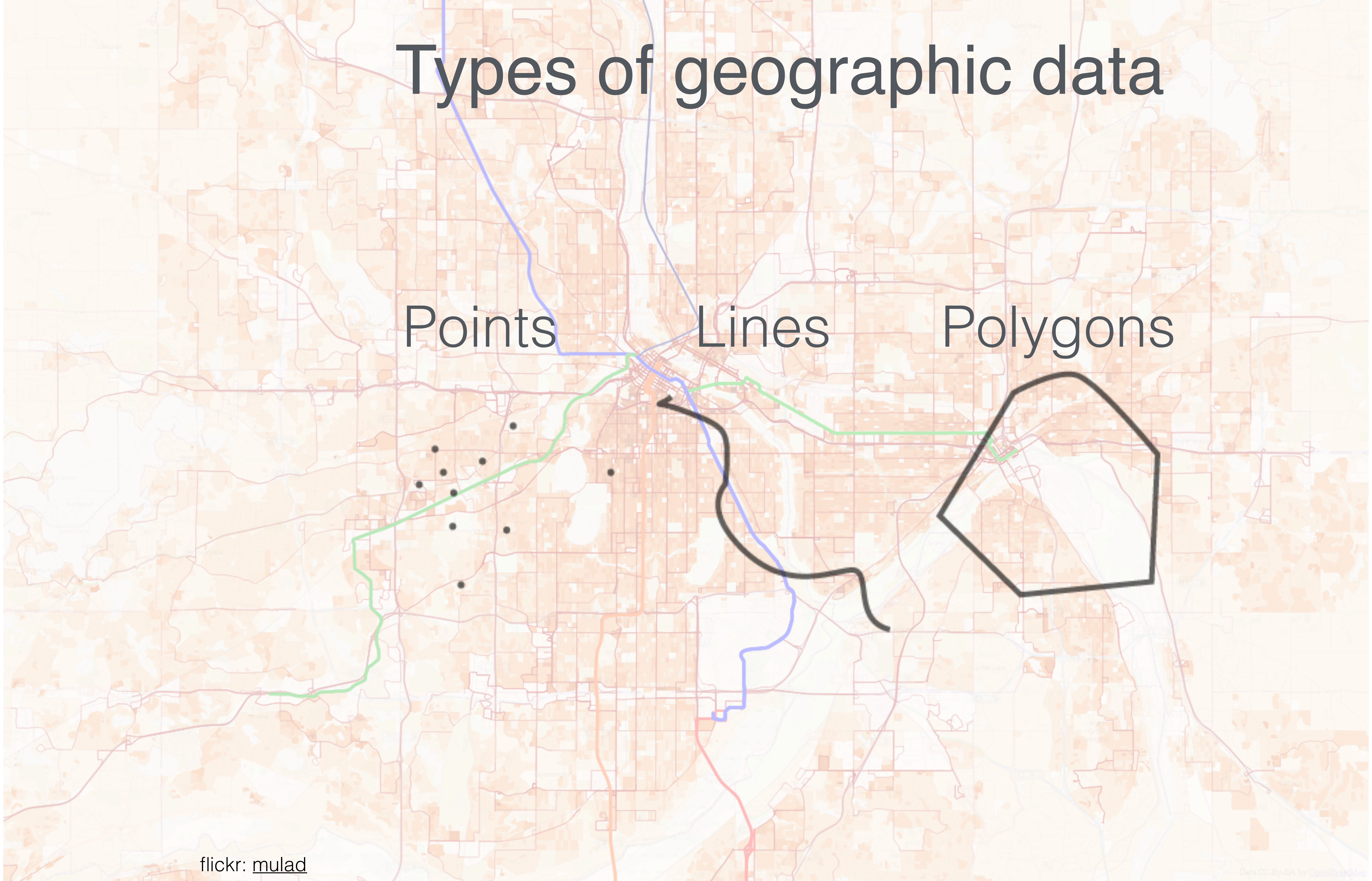
Spatial data


How spatial polygons shape our world
Amelia McNamara @AmeliaMN
Smith College, Northampton MA
OPEN VIS 2017 CONF
0:27 / 30:39
23
2
SHARE
SAVE

How spatial polygons shape our world - Amelia McNamara

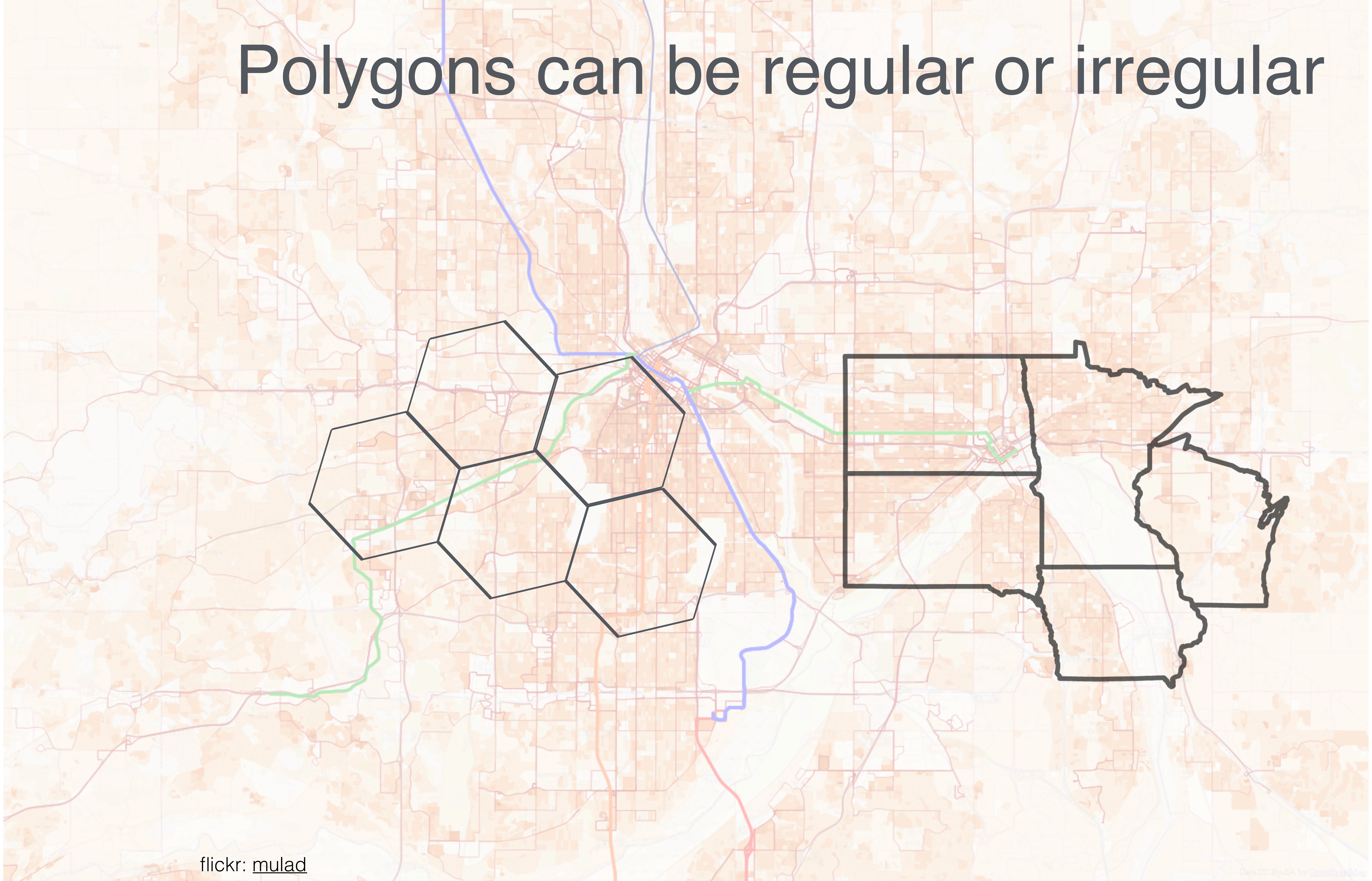
1,422 views

Types of geographic data

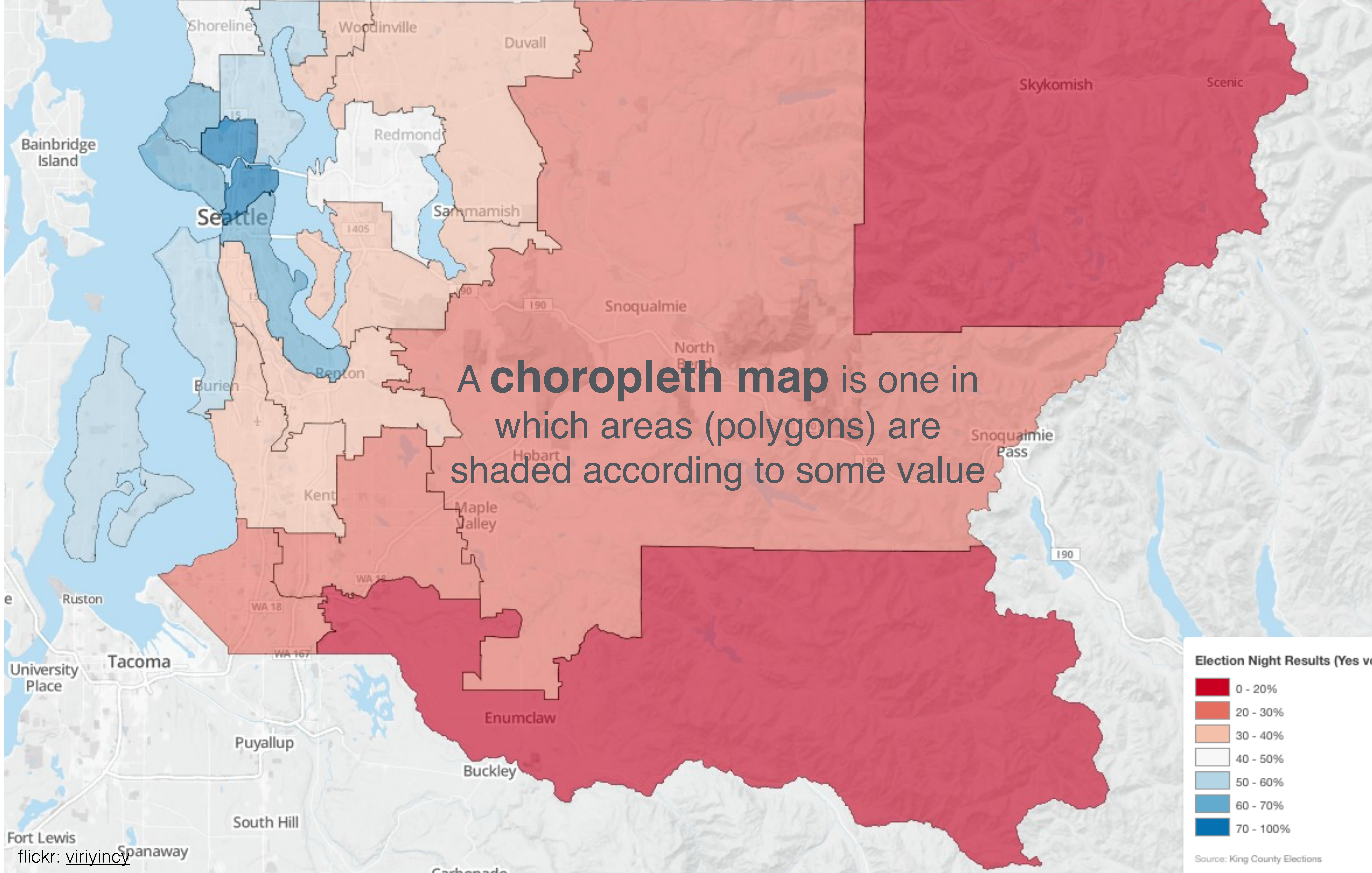


flickr: mulad

Polygons can be regular or irregular

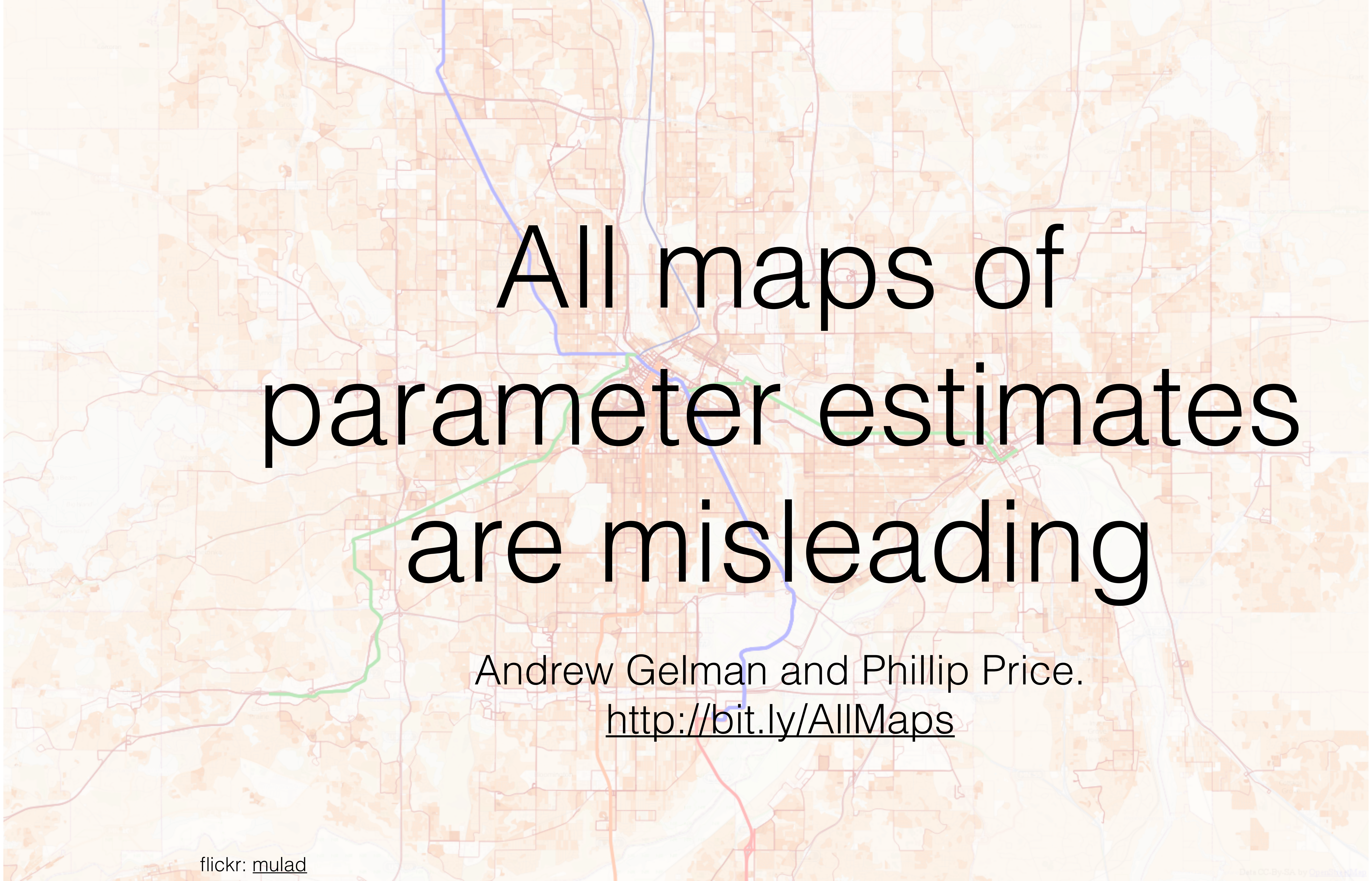


flickr: mulad



A **choropleth map** is one in which areas (polygons) are shaded according to some value

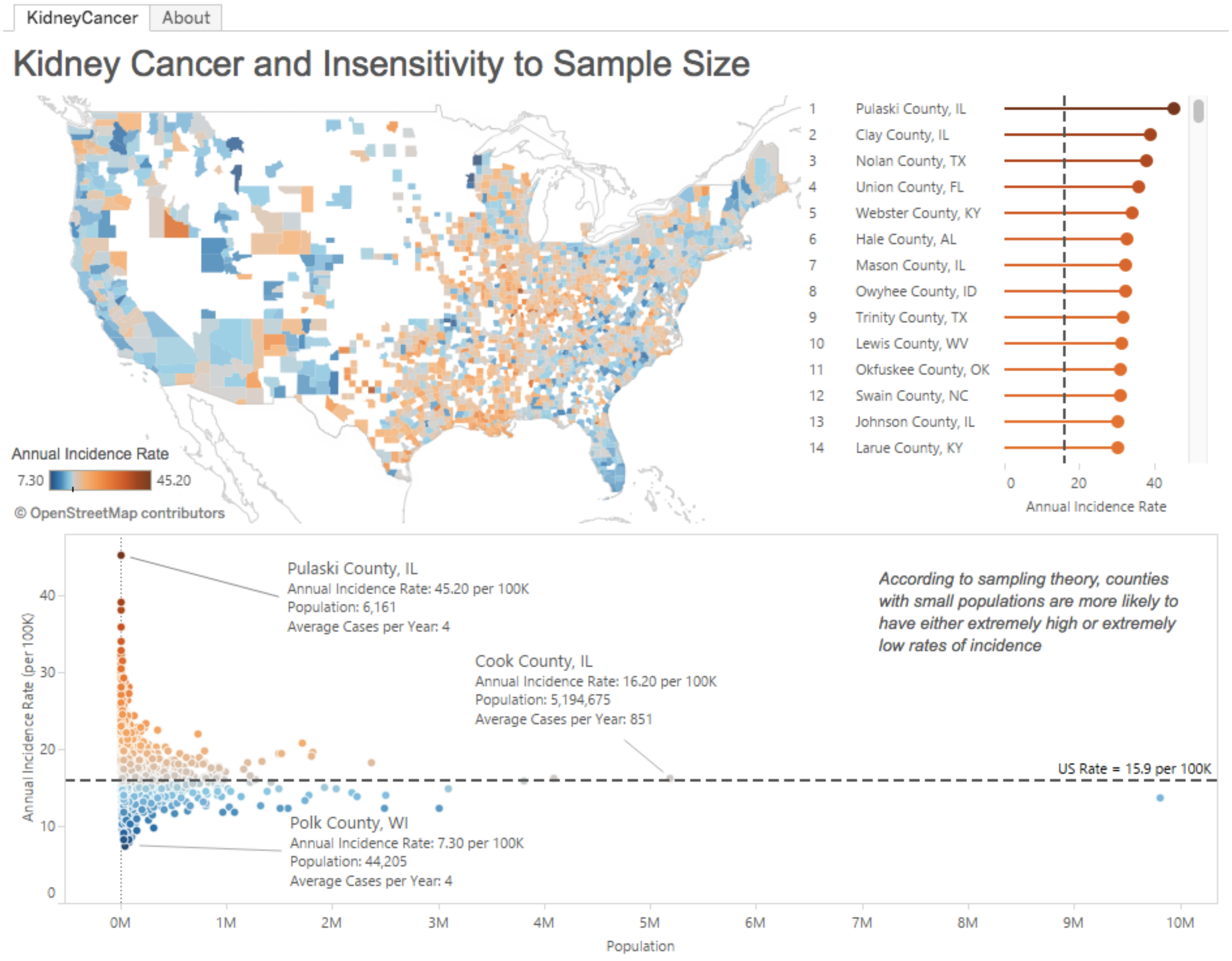
flickr: viriyincy



All maps of parameter estimates are misleading

Andrew Gelman and Phillip Price.
http://bit.ly/AllMaps

flickr: mulad



Kidney Cancer and Insensitivity to Sample Size

1	Pulaski County, IL	
2	Clay County, IL	
3	Nolan County, TX	
4	Union County, FL	
5	Webster County, KY	
6	Hale County, AL	
7	Mason County, IL	
8	Owyhee County, ID	
9	Trinity County, TX	
10	Lewis County, WV	
11	Okfuskee County, OK	
12	Swain County, NC	
13	Johnson County, IL	
14	Larue County, KY	

Pulaski County, IL
Annual Incidence Rate: 45.20 per 100K
Population: 6,161
Average Cases per Year: 4

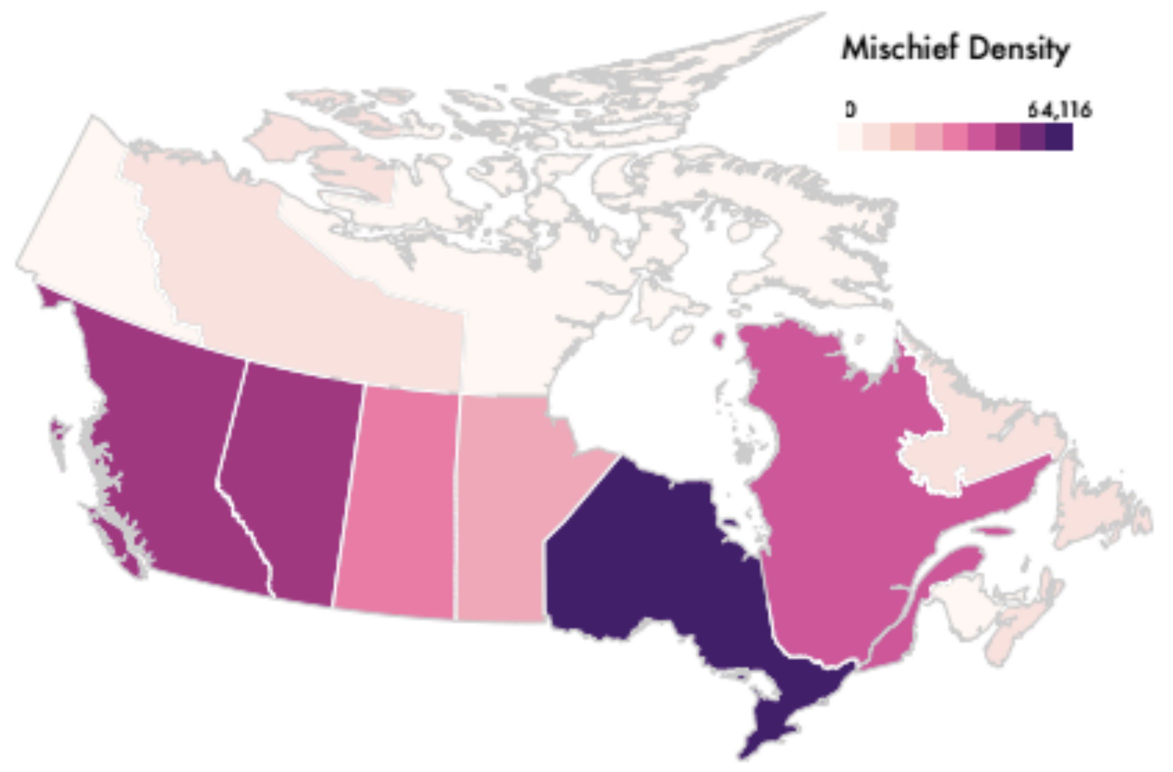
Cook County, IL
Annual Incidence Rate: 16.20 per 100K
Population: 5,194,675
Average Cases per Year: 851

Polk County, WI
Annual Incidence Rate: 7.30 per 100K
Population: 44,205
Average Cases per Year: 4

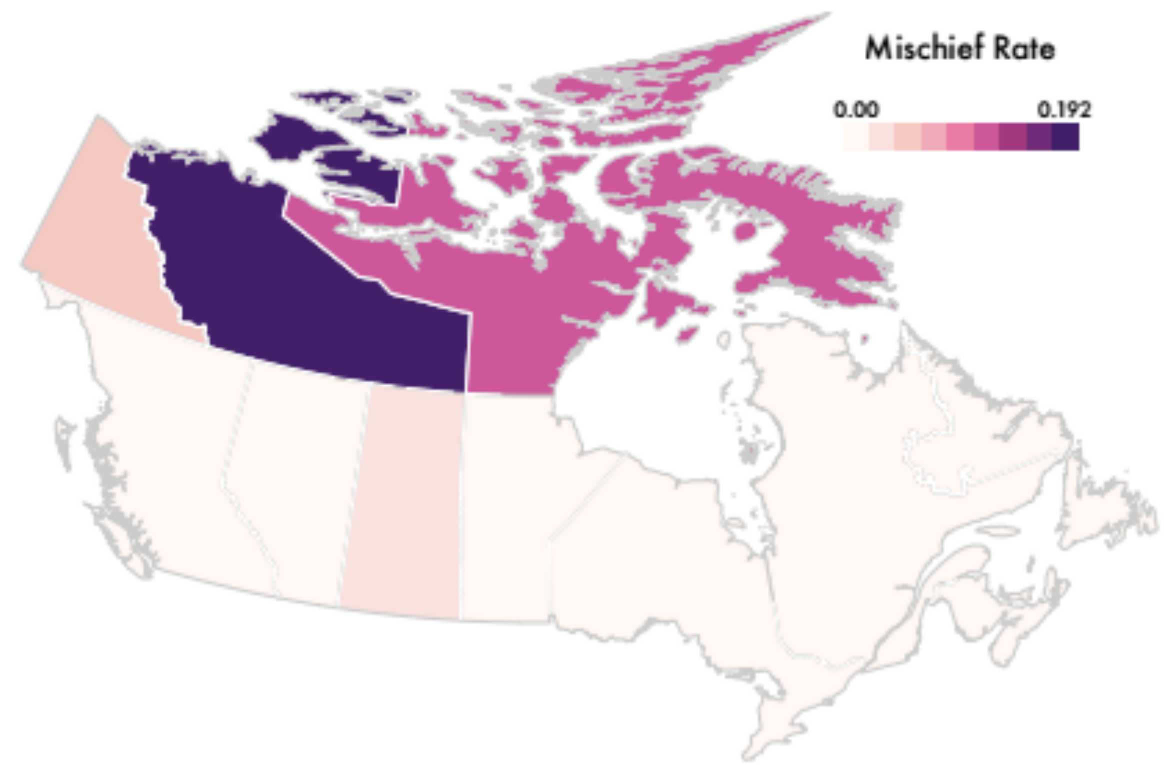
According to sampling theory, counties with small populations are more likely to have either extremely high or extremely low rates of incidence

http://dataremixed.com/2015/01/avoiding-data-pitfalls-part-2/

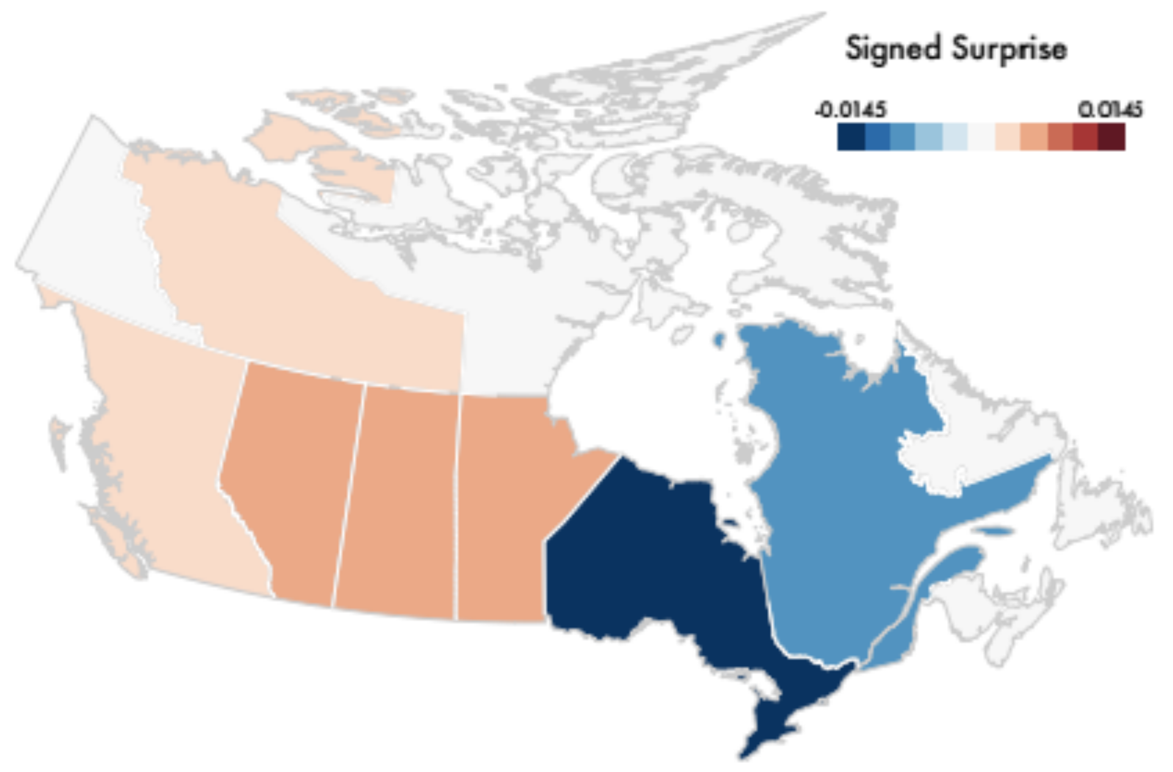
Surprise! Bayesian Weighting for De-Biasing Thematic Maps.



(a) The **Event Density** of “mischief” in Canada.



(b) The per-capita **Event Rate** of mischief.



(c) The **Surprise Map** of mischief.

Michael Correll and Jeffrey Heer
http://bit.ly/SurpriseMaps

Point data

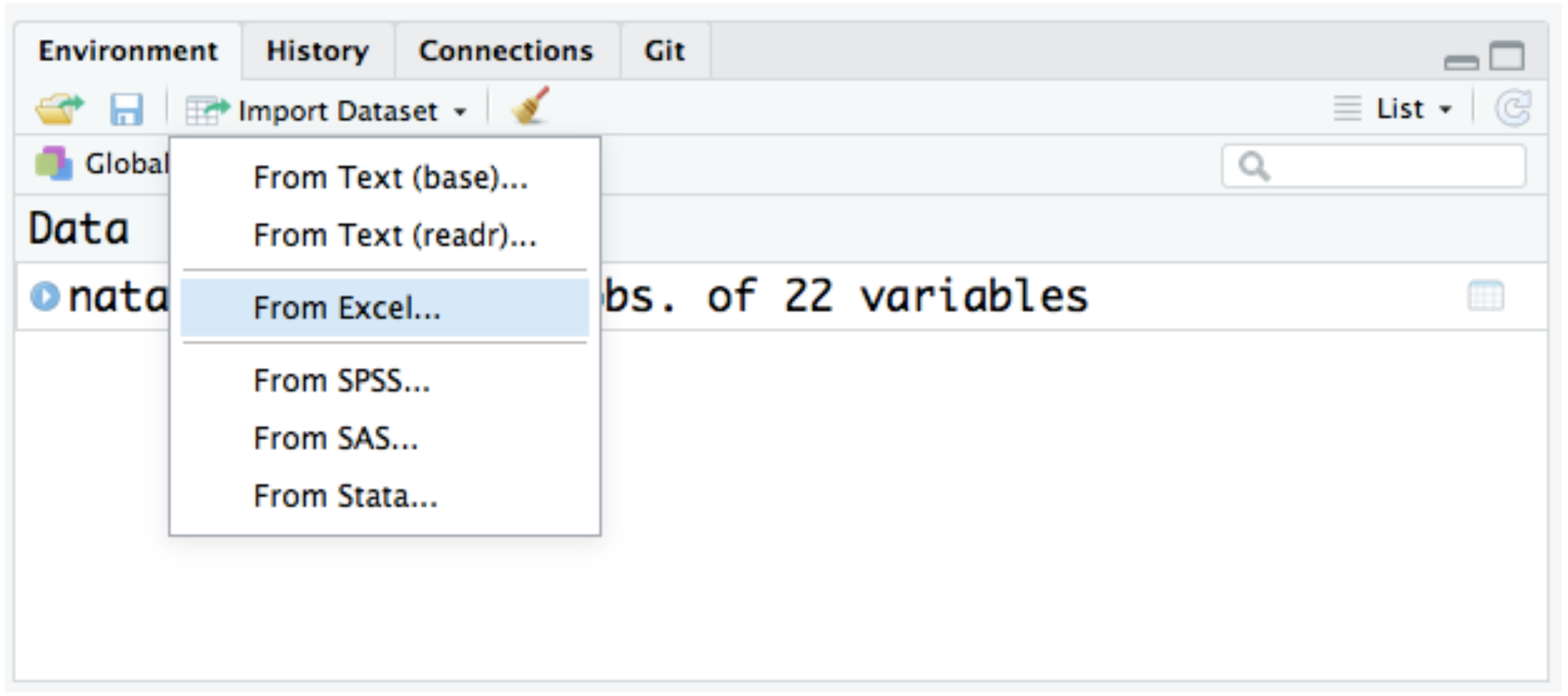
S3 class

"flat file"

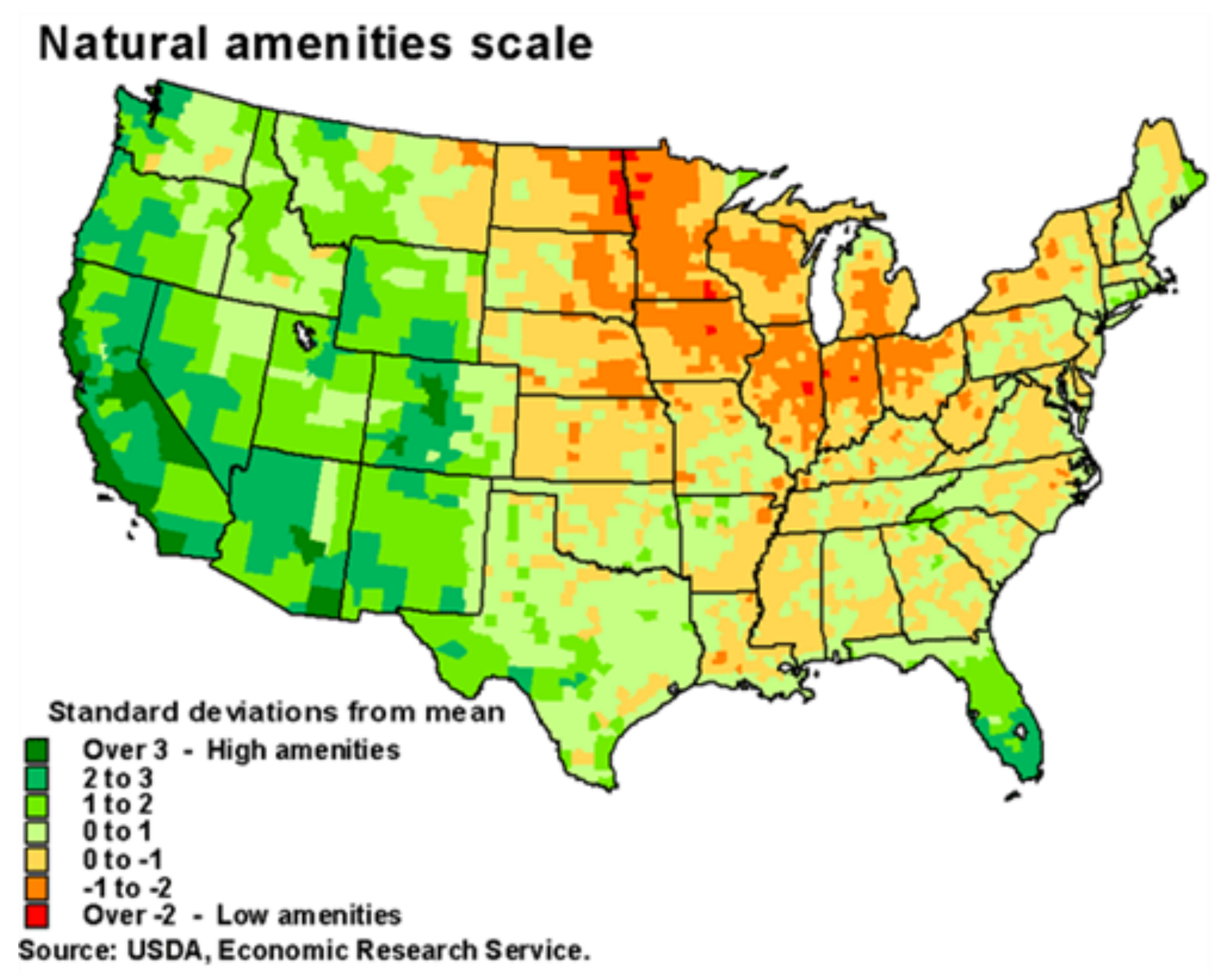
read in with `readr::read_csv()`, `readxl::read_excel()` or RStudio Import button

Natural amenities score

https://www.ers.usda.gov/data-products/natural-amenities-scale.aspx



```
Data
nata      bs. of 22 variables
```



Your Turn

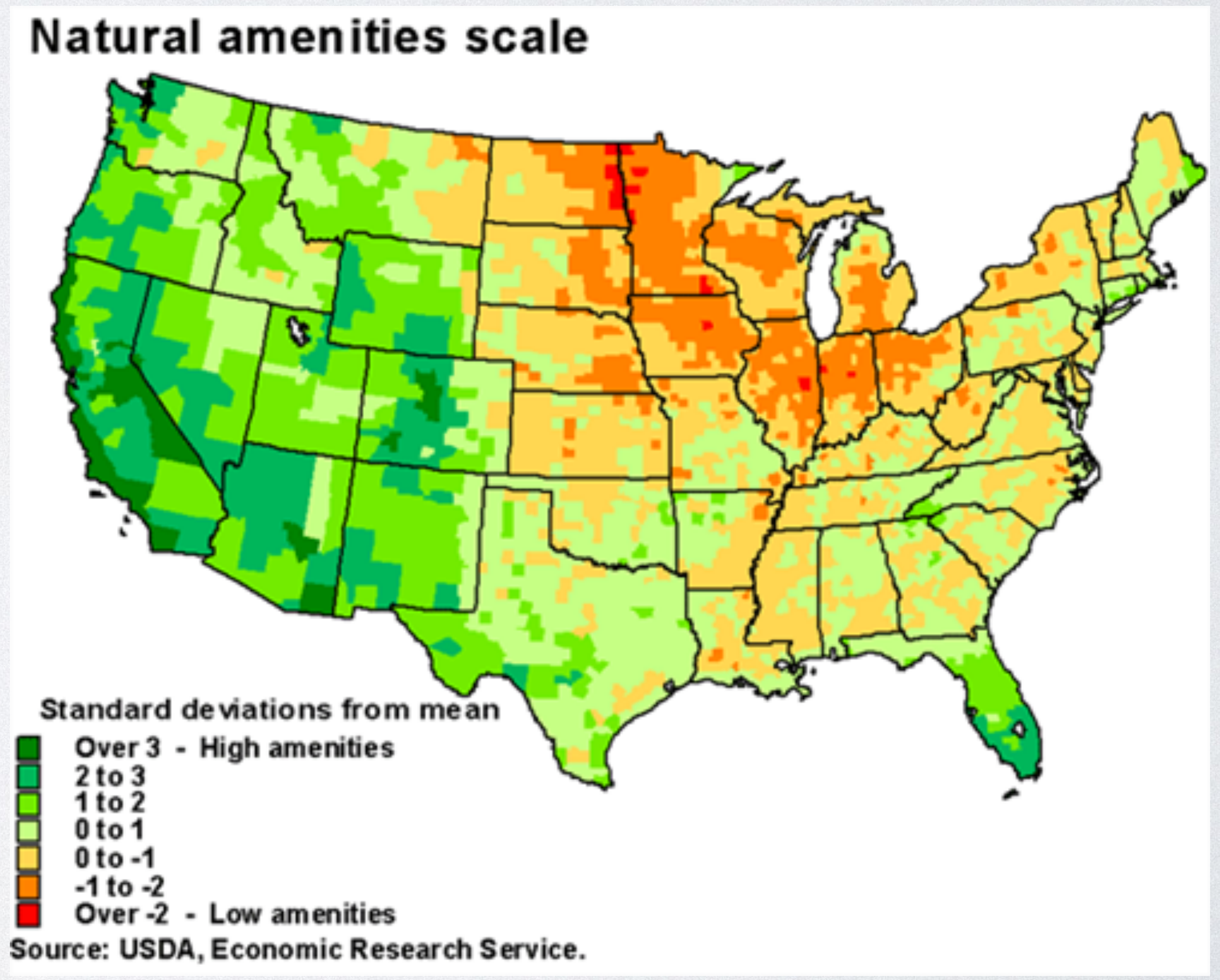
Download the natural amenities data from

https://www.ers.usda.gov/data-products/natural-amenities-scale.aspx

Upload it to RStudio Cloud

Load it in to R (hint: skip 104 rows)

Put your load-in code into your Rmd

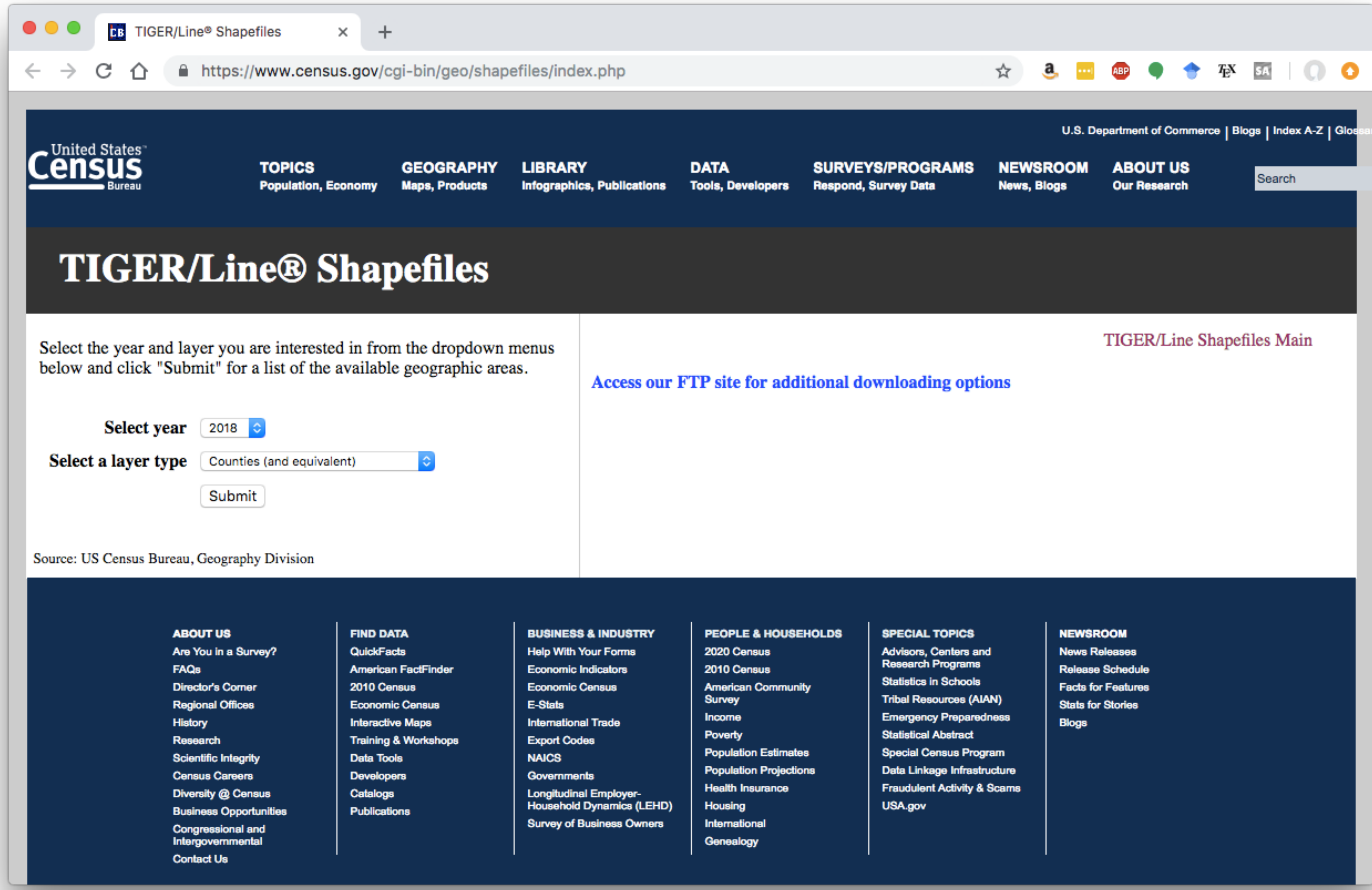


Polygon data

"Shapefiles" (proprietary format from ESRI, but readable by R)

Used to always be represented as an S4 class, including "slots" for data and polygons

Now, packages in the tidyverse have provided representations in S3, but support for modeling isn't complete



TIGER/Line® Shapefiles

Select the year and layer you are interested in from the dropdown menus below and click "Submit" for a list of the available geographic areas.

Access our FTP site for additional downloading options

Select year 2018

Select a layer type Counties (and equivalent)

Source: US Census Bureau, Geography Division

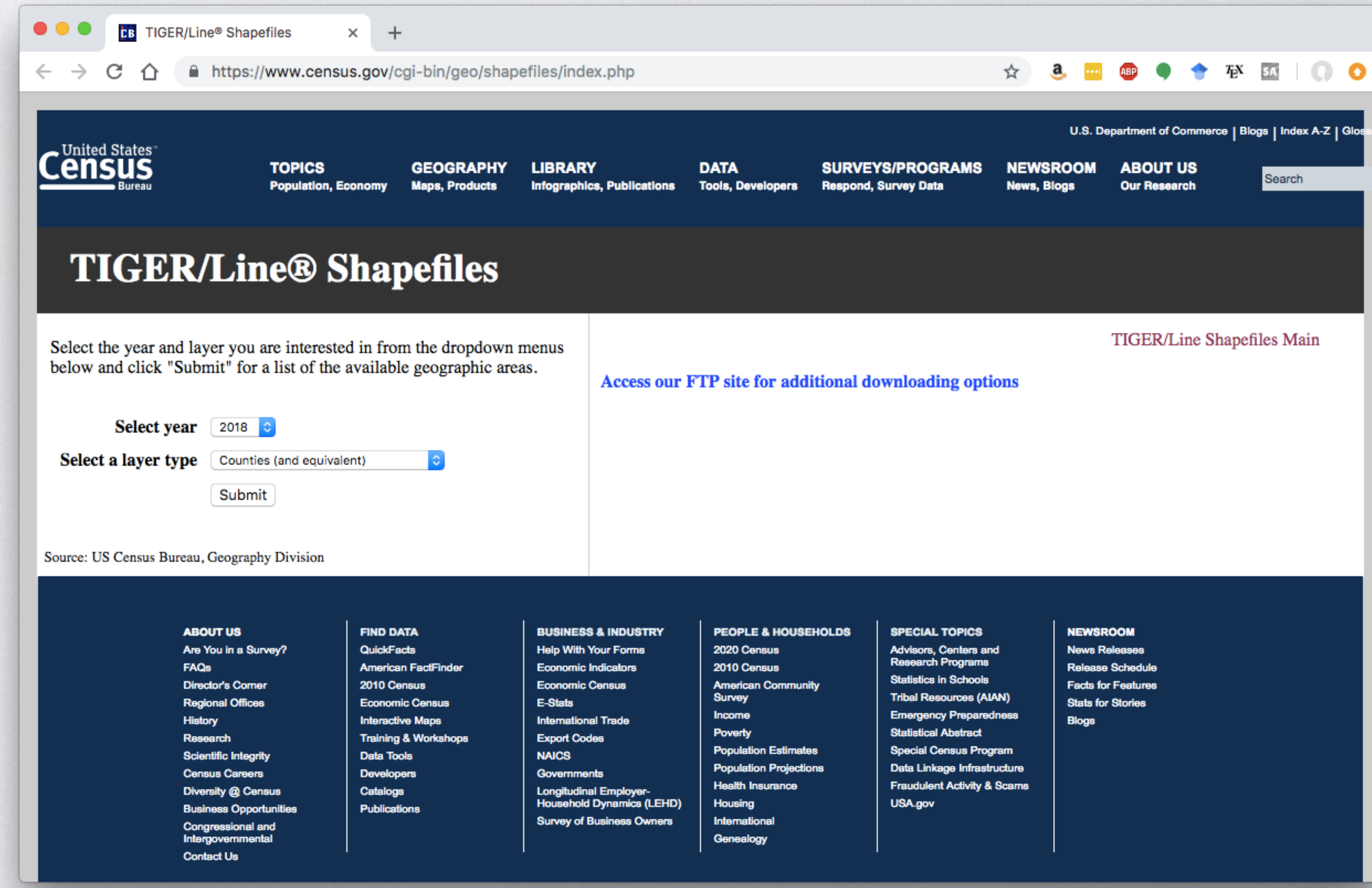
Your Turn

Download county shape files from

https://www.census.gov/cgi-bin/geo/shapefiles/index.php

This will be a folder of files

Upload the **zipped folder** to RStudio Cloud



TIGER/Line® Shapefiles

https://www.census.gov/cgi-bin/geo/shapefiles/index.php

U.S. Department of Commerce | Blogs | Index A-Z | Glos

United States Census Bureau

TOPICS Population, Economy
GEOGRAPHY Maps, Products
LIBRARY Infographics, Publications
DATA Tools, Developers
SURVEYS/PROGRAMS Respond, Survey Data
NEWSROOM News, Blogs
ABOUT US Our Research

Search

TIGER/Line® Shapefiles

Select the year and layer you are interested in from the dropdown menus below and click "Submit" for a list of the available geographic areas.

Select year: 2018

Select a layer type: Counties (and equivalent)

Submit

Source: US Census Bureau, Geography Division

TIGER/Line Shapefiles Main

Access our FTP site for additional downloading options

ABOUT US
Are You in a Survey?
FAQs
Director's Corner
Regional Offices
History
Research
Scientific Integrity
Census Careers
Diversity @ Census
Business Opportunities
Congressional and Intergovernmental
Contact Us

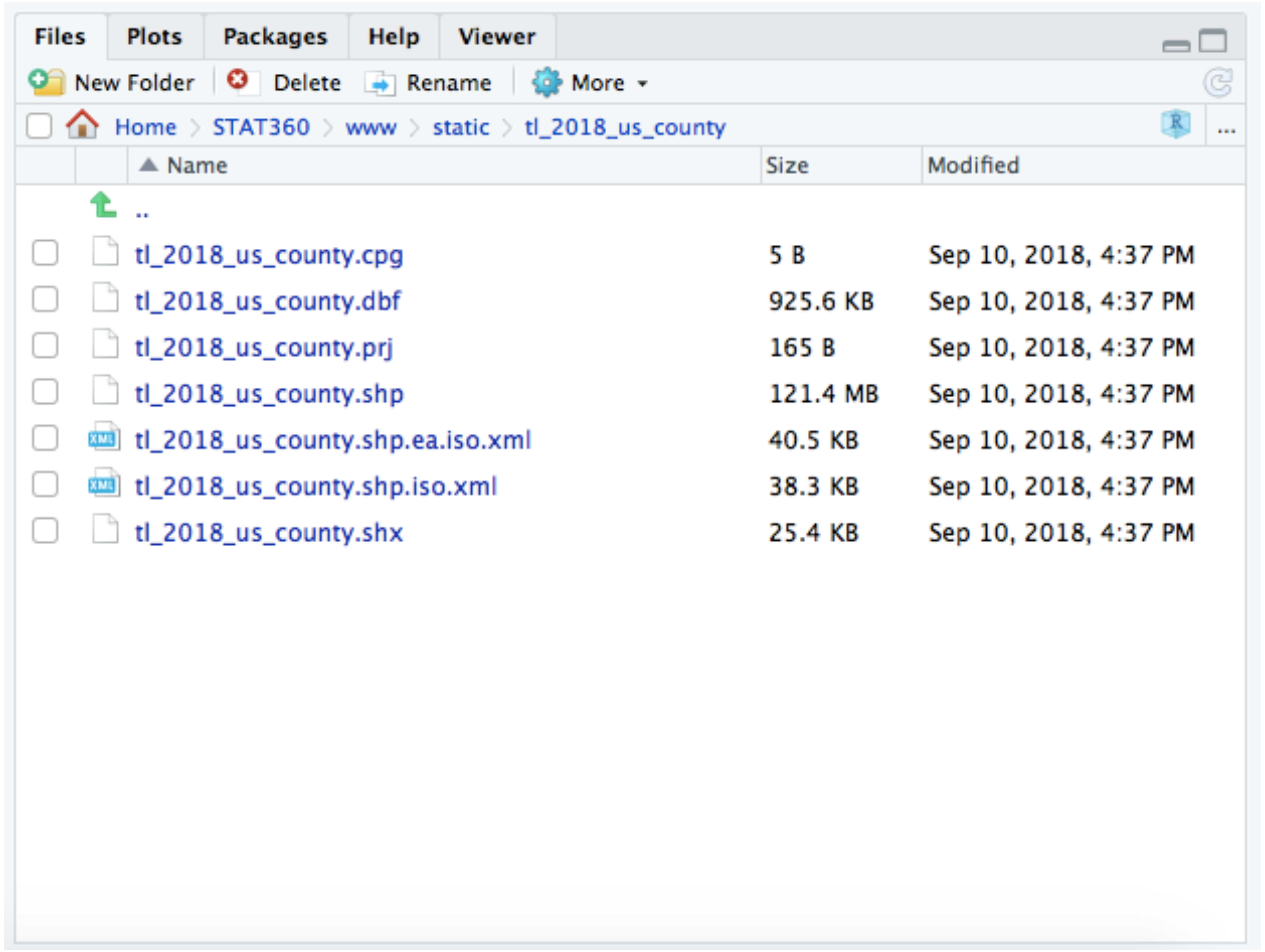
FIND DATA
QuickFacts
American FactFinder
2010 Census
Economic Census
Interactive Maps
Training & Workshops
Data Tools
Developers
Catalogs
Publications

BUSINESS & INDUSTRY
Help With Your Forms
Economic Indicators
Economic Census
E-Stats
International Trade
Export Codes
NAICS
Governments
Longitudinal Employer-Household Dynamics (LEHD)
Survey of Business Owners

PEOPLE & HOUSEHOLDS
2020 Census
2010 Census
American Community Survey
Income
Poverty
Population Estimates
Population Projections
Health Insurance
Housing
International
Genealogy

SPECIAL TOPICS
Advisors, Centers and Research Programs
Statistics in Schools
Tribal Resources (AIAN)
Emergency Preparedness
Statistical Abstract
Special Census Program
Data Linkage Infrastructure
Fraudulent Activity & Scams
USA.gov

NEWSROOM
News Releases
Release Schedule
Facts for Features
Stats for Stories
Blogs



Name	Size	Modified
..		
tl_2018_us_county.cpg	5 B	Sep 10, 2018, 4:37 PM
tl_2018_us_county.dbf	925.6 KB	Sep 10, 2018, 4:37 PM
tl_2018_us_county.prj	165 B	Sep 10, 2018, 4:37 PM
tl_2018_us_county.shp	121.4 MB	Sep 10, 2018, 4:37 PM
tl_2018_us_county.shp.ea.iso.xml	40.5 KB	Sep 10, 2018, 4:37 PM
tl_2018_us_county.shp.iso.xml	38.3 KB	Sep 10, 2018, 4:37 PM
tl_2018_us_county.shx	25.4 KB	Sep 10, 2018, 4:37 PM

Loading shapefile data— the oldschool, S4 way

```
library(rgdal)

counties_rgdal <- readOGR("www/static/tl_2018_us_county/",
layer="tl_2018_us_county")
```

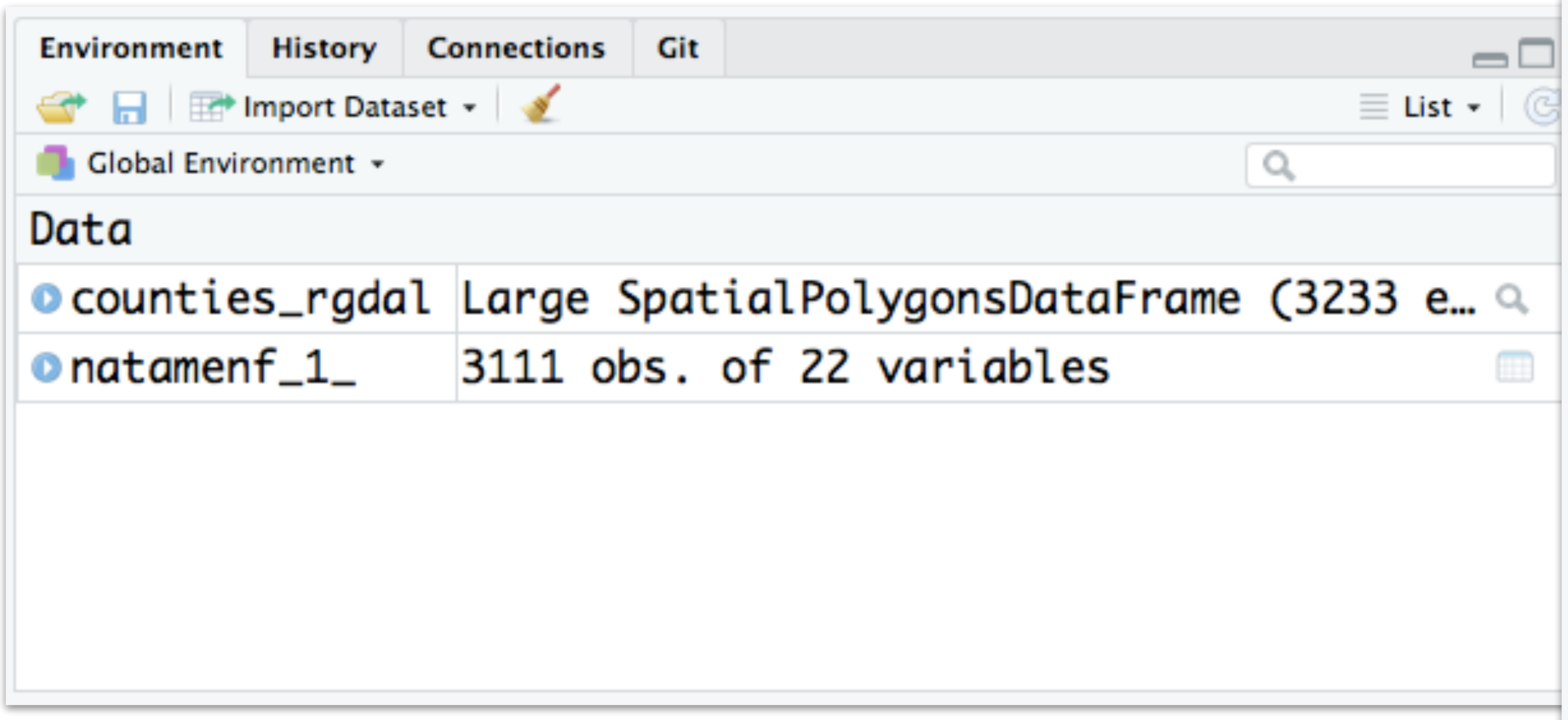
folder name

file names

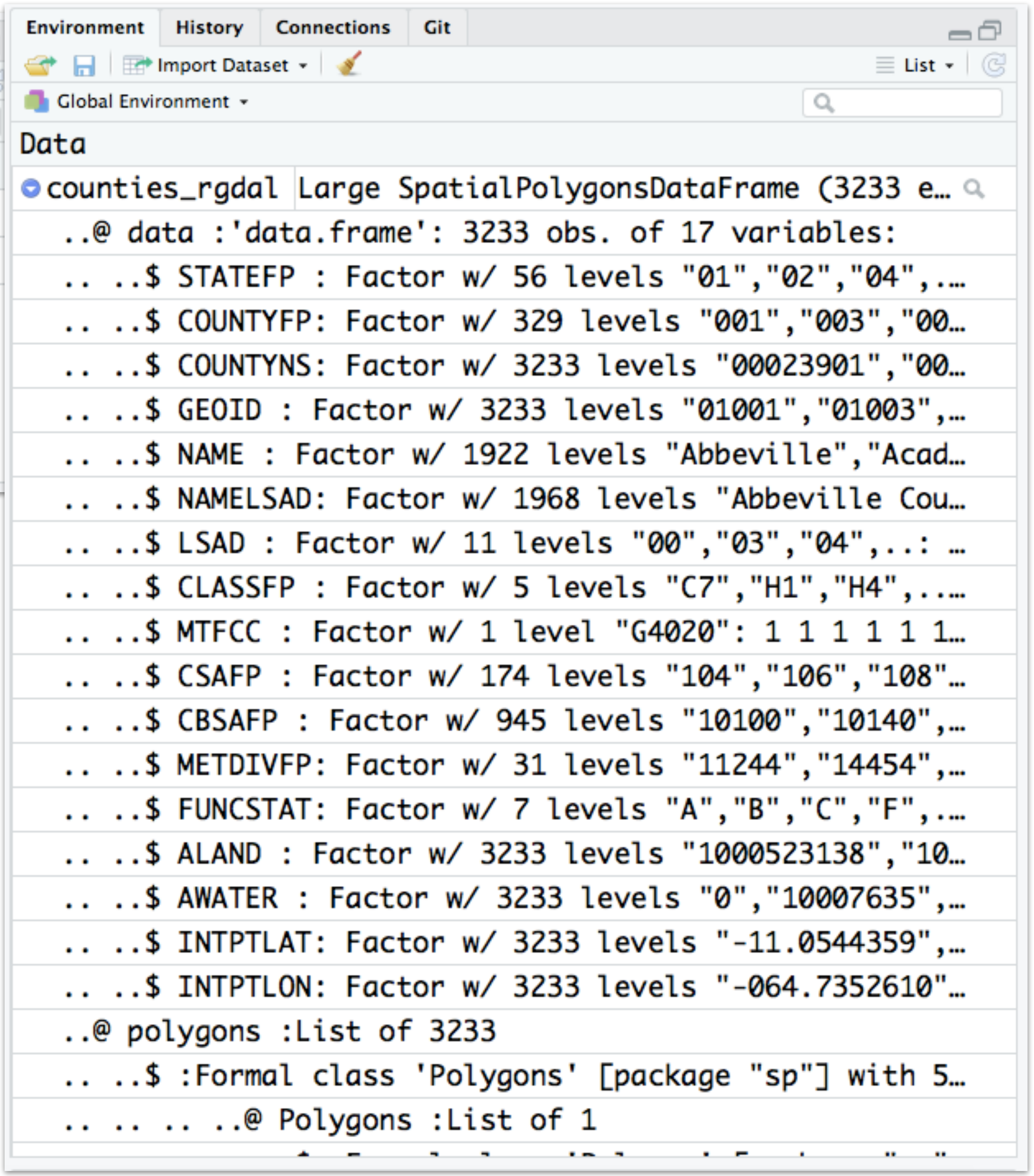
Your Turn

Load the county data in using rgdal

Look into the object. What slots does it have?



```
Data
counties_rgdal  Large SpatialPolygonsDataFrame (3233 e…
natamenf_1_     3111 obs. of 22 variables
```



```
Data
counties_rgdal Large SpatialPolygonsDataFrame (3233 e…
  ..@ data :'data.frame': 3233 obs. of 17 variables:
  .. ..$ STATEFP : Factor w/ 56 levels "01","02","04",..…
  .. ..$ COUNTYFP: Factor w/ 329 levels "001","003","00…
  .. ..$ COUNTYNS: Factor w/ 3233 levels "00023901","00…
  .. ..$ GEOID : Factor w/ 3233 levels "01001","01003",…
  .. ..$ NAME : Factor w/ 1922 levels "Abbeville","Acad…
  .. ..$ NAMELSAD: Factor w/ 1968 levels "Abbeville Cou…
  .. ..$ LSAD : Factor w/ 11 levels "00","03","04",..: …
  .. ..$ CLASSFP : Factor w/ 5 levels "C7","H1","H4",..…
  .. ..$ MTFCC : Factor w/ 1 level "G4020": 1 1 1 1 1 1…
  .. ..$ CSAFP : Factor w/ 174 levels "104","106","108"…
  .. ..$ CBSAFP : Factor w/ 945 levels "10100","10140",…
  .. ..$ METDIVFP: Factor w/ 31 levels "11244","14454",…
  .. ..$ FUNCSTAT: Factor w/ 7 levels "A","B","C","F",..…
  .. ..$ ALAND : Factor w/ 3233 levels "100523138","10…
  .. ..$ AWATER : Factor w/ 3233 levels "0","10007635",…
  .. ..$ INTPTLAT: Factor w/ 3233 levels "-11.0544359",…
  .. ..$ INTPTLON: Factor w/ 3233 levels "-064.7352610"…
  ..@ polygons :List of 3233
  .. ..$ :Formal class 'Polygons' [package "sp"] with 5…
  .. .. .. ..@ Polygons :List of 1
```

```
> slotNames(counties_rgdal)

[1] "data"        "polygons"    "plotOrder"   "bbox"        "proj4string"

> slot(counties_rgdal, "data")

   STATEFP COUNTYFP COUNTYNS GEOID        NAME             NAMELSAD LSAD CLASSFP MTFCC

0       31      039 00835841 31039      Cuming        Cuming County   06      H1 G4020

1       53      069 01513275 53069   Wahkiakum     Wahkiakum County   06      H1 G4020

2       35      011 00933054 35011     De Baca       De Baca County   06      H1 G4020

3       31      109 00835876 31109   Lancaster     Lancaster County   06      H1 G4020

4       31      129 00835886 31129    Nuckolls      Nuckolls County   06      H1 G4020
```

```
> class(counties_rgdal)
[1] "SpatialPolygonsDataFrame"
attr(,"package")
[1] "sp"
> methods(class="SpatialPolygonsDataFrame")
 [1] [              [[             [[<-           [<-            $              $<-
 [7] addAttrToGeom  as.data.frame  bbox           coerce         coerce<-       coordinates
[13] coordinates<-  coordnames     coordnames<-   dim            dimensions     disaggregate
[19] fullgrid       geometry       geometry<-     gridded        is.projected   length
[25] merge          names          names<-        over           plot           polygons
[31] polygons<-     proj4string    proj4string<-  rbind          recenter       row.names
[37] row.names<-    spChFIDs       spChFIDs<-     split          sppanel        spplot
[43] spsample       spTransform    summary
see '?methods' for accessing help and source code
```

Loading shapefile data— the tidyverse, S3 way

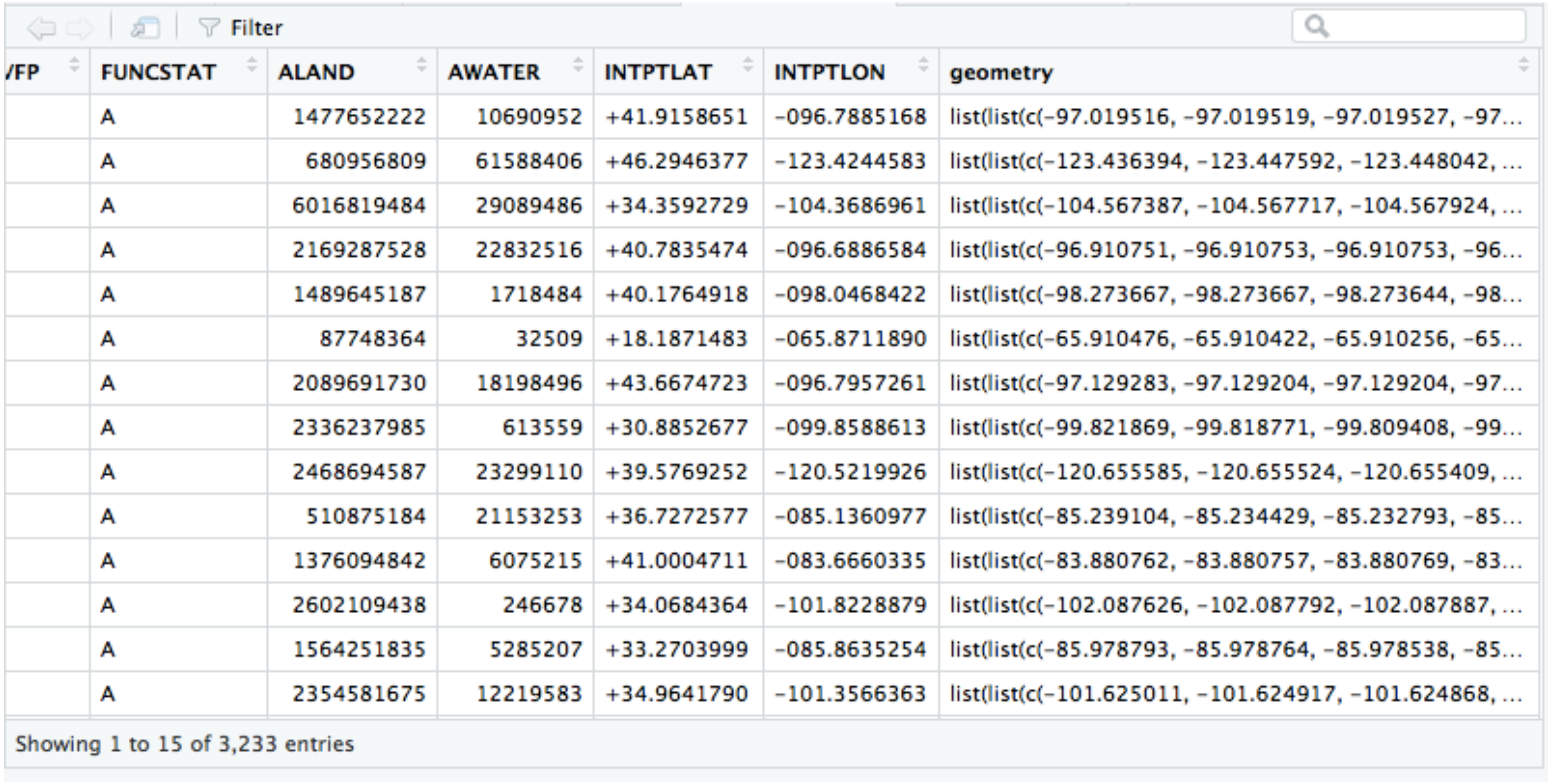
```
library(sf)

counties_sf <- st_read("www/static/tl_2018_us_county/")
```

folder name

Your Turn

Load the county data in using sf

Look into the object. What does it look like?

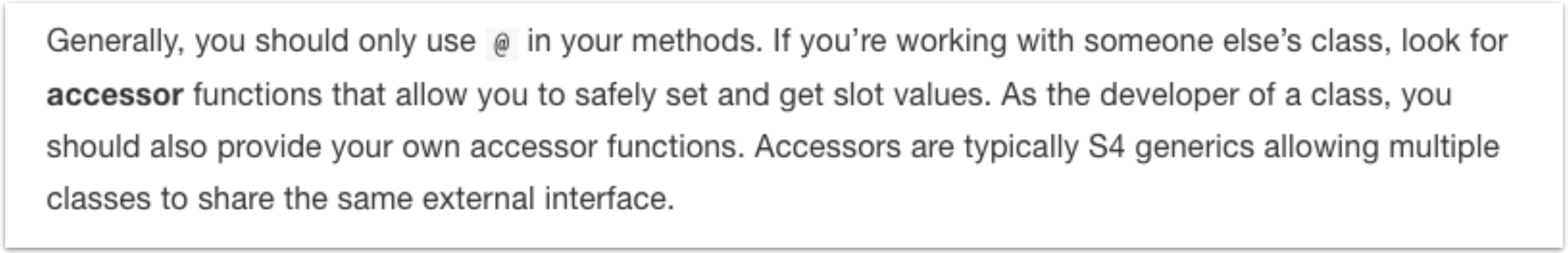


/FP	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPTLON	geometry
	A	1477652222	10690952	+41.9158651	-096.7885168	list(list(c(-97.019516, -97.019519, -97.019527, -97...
	A	680956809	61588406	+46.2946377	-123.4244583	list(list(c(-123.436394, -123.447592, -123.448042, ...
	A	6016819484	29089486	+34.3592729	-104.3686961	list(list(c(-104.567387, -104.567717, -104.567924, ...
	A	2169287528	22832516	+40.7835474	-096.6886584	list(list(c(-96.910751, -96.910753, -96.910753, -96...
	A	1489645187	1718484	+40.1764918	-098.0468422	list(list(c(-98.273667, -98.273667, -98.273644, -98...
	A	87748364	32509	+18.1871483	-065.8711890	list(list(c(-65.910476, -65.910422, -65.910256, -65...
	A	2089691730	18198496	+43.6674723	-096.7957261	list(list(c(-97.129283, -97.129204, -97.129204, -97...
	A	2336237985	613559	+30.8852677	-099.8588613	list(list(c(-99.821869, -99.818771, -99.809408, -99...
	A	2468694587	23299110	+39.5769252	-120.5219926	list(list(c(-120.655585, -120.655524, -120.655409, ...
	A	510875184	21153253	+36.7272577	-085.1360977	list(list(c(-85.239104, -85.234429, -85.232793, -85...
	A	1376094842	6075215	+41.0004711	-083.6660335	list(list(c(-83.880762, -83.880757, -83.880769, -83...
	A	2602109438	246678	+34.0684364	-101.8228879	list(list(c(-102.087626, -102.087792, -102.087887, ...
	A	1564251835	5285207	+33.2703999	-085.8635254	list(list(c(-85.978793, -85.978764, -85.978538, -85...
	A	2354581675	12219583	+34.9641790	-101.3566363	list(list(c(-101.625011, -101.624917, -101.624868, ...

Showing 1 to 15 of 3,233 entries

Joining spatial data— the oldschool, S4 way

```
counties_rgdal@data <-
left_join(counties_rgdal@data, natamenf_1_, by =
c("GEOID" = "FIPS Code"))
```



Generally, you should only use `@` in your methods. If you're working with someone else's class, look for **accessor** functions that allow you to safely set and get slot values. As the developer of a class, you should also provide your own accessor functions. Accessors are typically S4 generics allowing multiple classes to share the same external interface.

¯_(ツ)_/¯

```
attr(,"package")
[1] "sp"
> methods(class="SpatialPolygonsDataFrame")
 [1] [             [[            [[<-          [<-           $             $<-
 [7] addAttrToGeom as.data.frame bbox          coerce        coerce<-      coordinates
[13] coordinates<- coordnames    coordnames<-  dim           dimensions    disaggregate
[19] fullgrid      geometry      geometry<-    gridded       is.projected  length
[25] merge         names         names<-       over          plot          polygons
[31] polygons<-    proj4string   proj4string<- rbind         recenter      row.names
[37] row.names<-   spChFIDs      spChFIDs<-    split         sppanel       spplot
[43] spsample      spTransform   summary
see '?methods' for accessing help and source code
```

Joining spatial data— the tidyverse, S3 way

```
> counties_sf <- counties_sf %>%
left_join(natamenf_1_, by=c("GEOID" = "FIPS Code"))
```

Your Turn

Join the data together, one or both ways

Base plotting of spatial objects

Remember the generic function, plot()? It has methods for both these data types

```
plot(states_rgdal)

plot(states_sf["Yes"])
```

Leaflet

Leaflet is a Javascript library for interactive maps. A bunch of people worked to make an R package that works with leaflet, but you can use leaflet in many more situations (for example, if you do data visualization in d3.js, it's easy to integrate with leaflet).

```
library(leaflet)


pal <- colorNumeric(
  palette = "Greens",
  domain = counties_rgdal$Yes
)


m <- leaflet(data=counties_rgdal) %>%
  addProviderTiles("Stamen.Watercolor") %>%
  setView(lng = -98.35, lat = 39.8, zoom = 03) %>%
  addPolygons(stroke = FALSE, fillOpacity = 0.5, smoothFactor = 0.5, color =~pal(Scale)
  ) %>%
  addLegend("bottomright", pal = pal, values = ~Scale,
            title = "Natual ammeniries score",
            opacity = 1
  )
```

Leaflet options

Check out the leaflet options on the RStudio documentation page

- Basemaps: `?addProviderTiles` for different base maps

- Colors: colors from RColorBrewer are based on ColorBrewer. You can see all the available palettes by using

```
library(RColorBrewer)

display.brewer.all(type="seq")
```

- Legends: check out `?addLegend` to see options. In particular, you might want to adjust the `bins`

Your Turn

Customize your map! Change at least two things (the variable you're plotting, the colors, the bin breaks, the legend text, etc., etc.)

Knit your document!

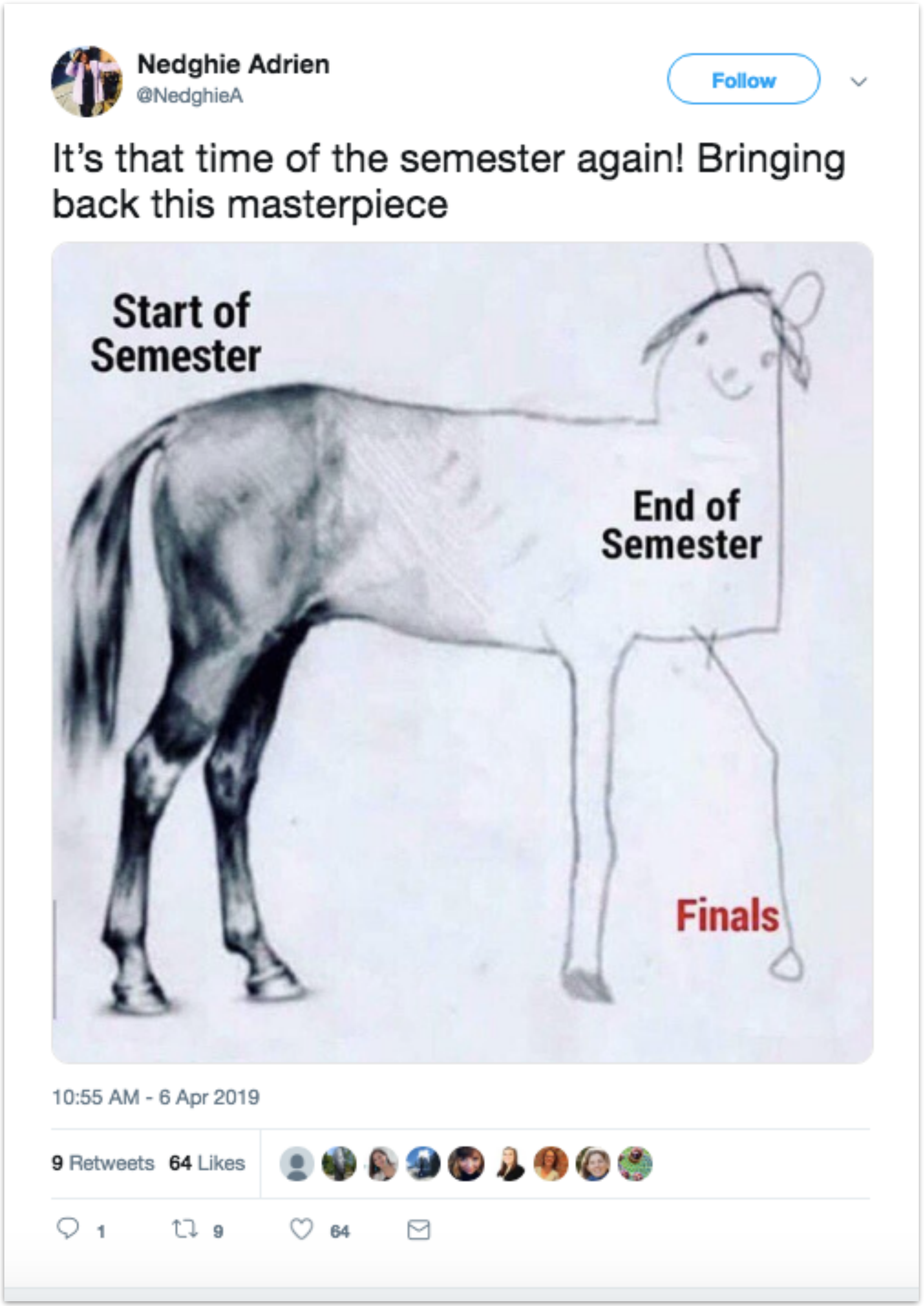
Hint: DO NOT COMMIT SHAPEFILES

They are large, large files and Github won't accept them

You may want to edit your .gitignore file to ignore them

One way to save yourself is with

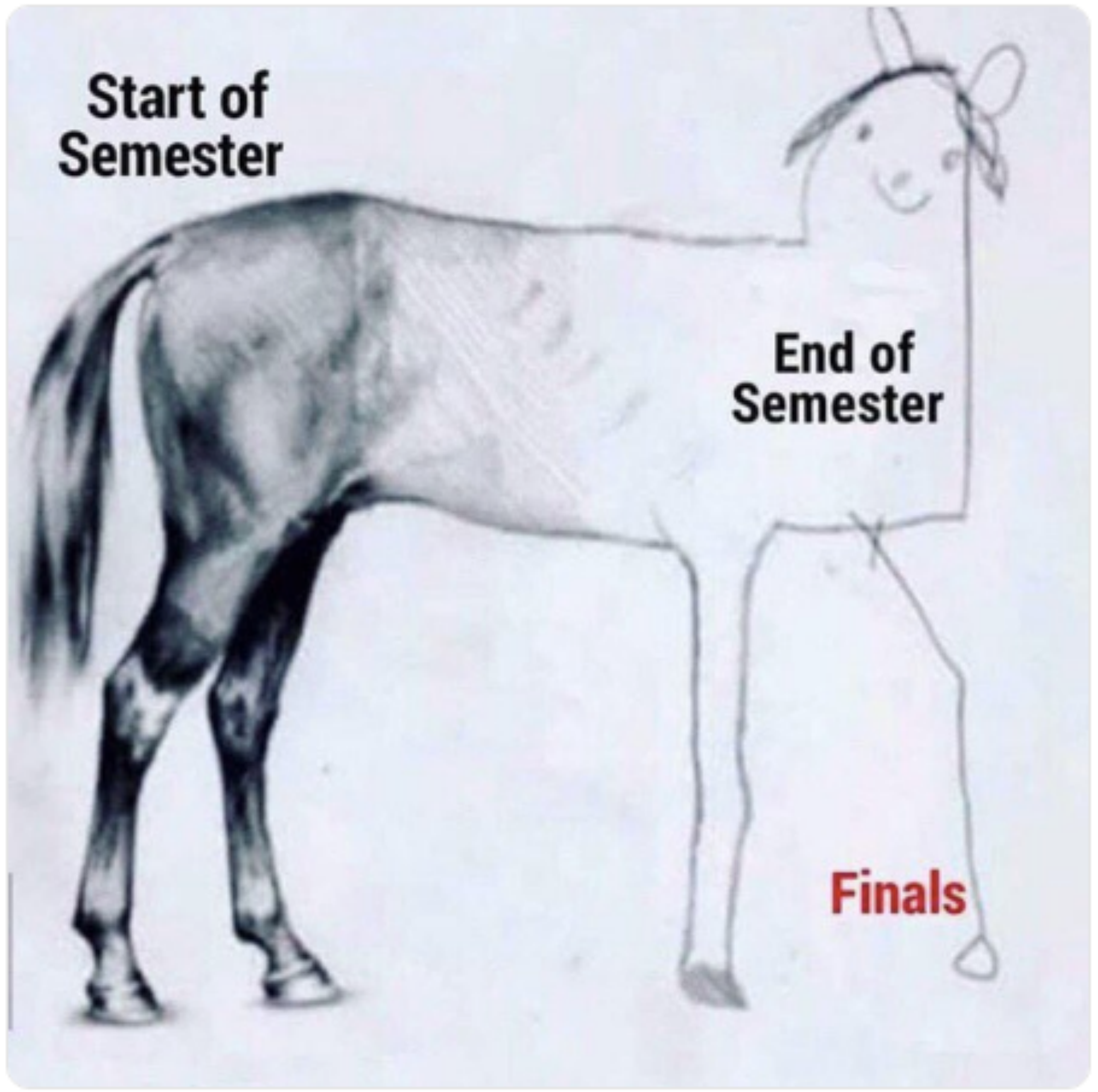
```
git rm --cached giant file

git commit --amend -CHEAD
```



Nedghie Adrien
@NedghieA

Follow

It's that time of the semester again! Bringing back this masterpiece

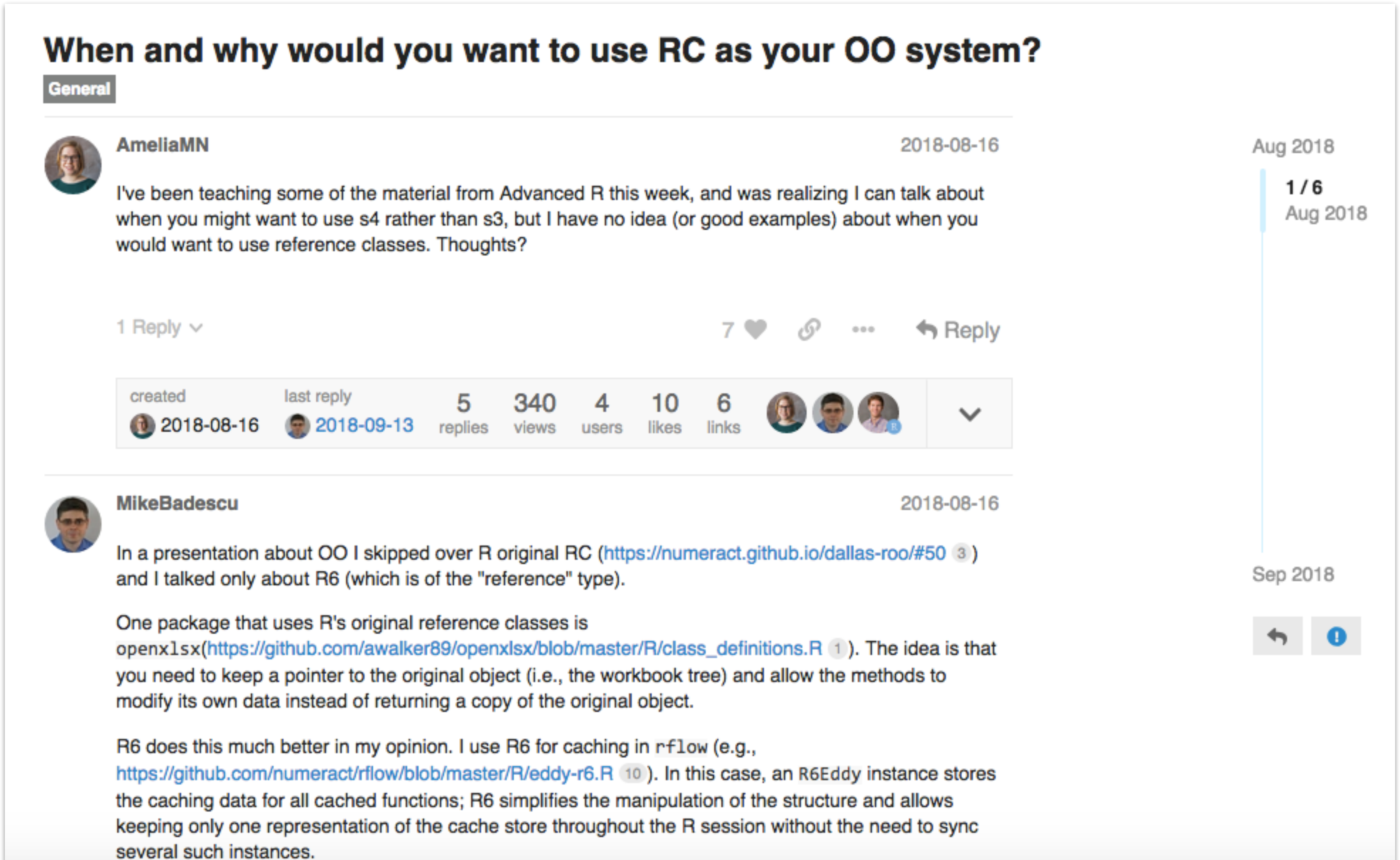


10:55 AM - 6 Apr 2019

9 Retweets 64 Likes

1 9 64

RC and R6?



When and why would you want to use RC as your OO system?

General

AmeliaMN 2018-08-16

I've been teaching some of the material from Advanced R this week, and was realizing I can talk about when you might want to use s4 rather than s3, but I have no idea (or good examples) about when you would want to use reference classes. Thoughts?

1 Reply 7 Reply

created	last reply	5 replies	340 views	4 users	10 likes	6 links
2018-08-16	2018-09-13					

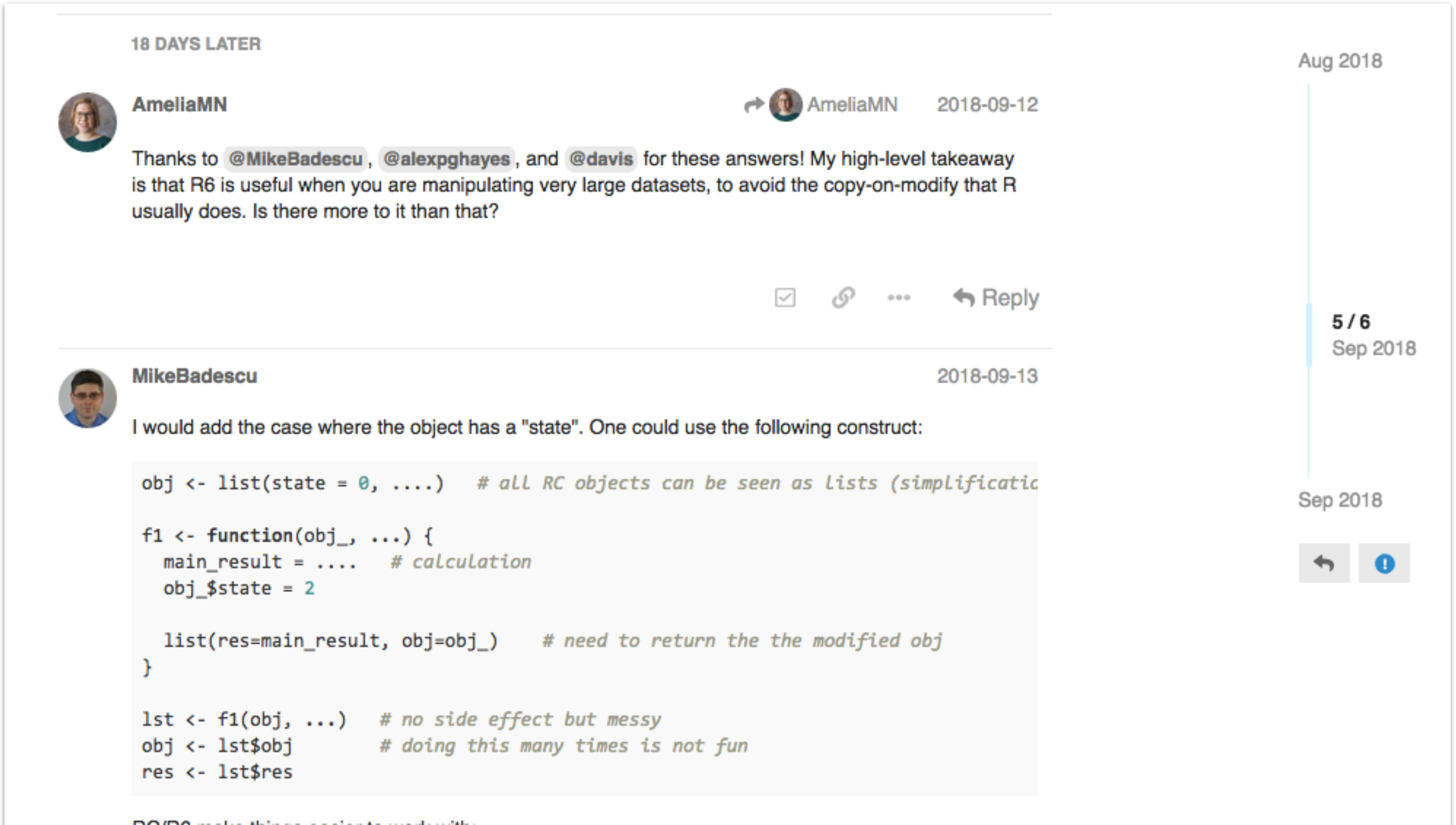
MikeBadescu 2018-08-16

In a presentation about OO I skipped over R original RC (https://numeract.github.io/dallas-roo/#50 3) and I talked only about R6 (which is of the "reference" type).

One package that uses R's original reference classes is `openxlsx`(https://github.com/awalker89/openxlsx/blob/master/R/class_definitions.R 1). The idea is that you need to keep a pointer to the original object (i.e., the workbook tree) and allow the methods to modify its own data instead of returning a copy of the original object.

R6 does this much better in my opinion. I use R6 for caching in `rflow` (e.g., https://github.com/numeract/rflow/blob/master/R/eddy-r6.R 10). In this case, an `R6Eddy` instance stores the caching data for all cached functions; R6 simplifies the manipulation of the structure and allows keeping only one representation of the cache store throughout the R session without the need to sync several such instances.

RC and R6?



18 DAYS LATER

AmeliaMN — AmeliaMN 2018-09-12

Thanks to @MikeBadescu, @alexpghayes, and @davis for these answers! My high-level takeaway is that R6 is useful when you are manipulating very large datasets, to avoid the copy-on-modify that R usually does. Is there more to it than that?

MikeBadescu 2018-09-13

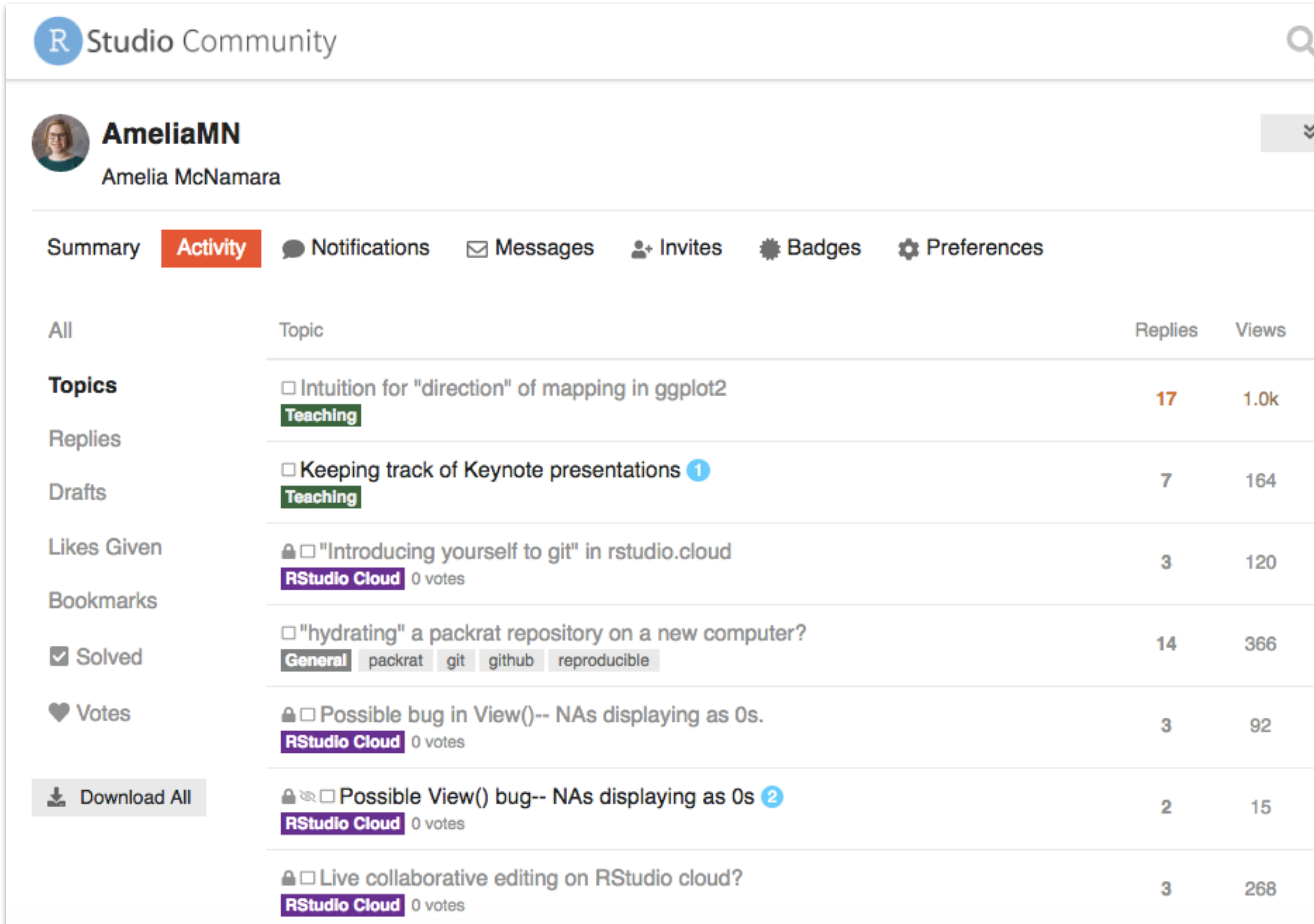
I would add the case where the object has a "state". One could use the following construct:

```
obj <- list(state = 0, ....)   # all RC objects can be seen as lists (simplificatio

f1 <- function(obj_, ...) {
  main_result = ....    # calculation
  obj_$state = 2

  list(res=main_result, obj=obj_)    # need to return the the modified obj
}

lst <- f1(obj, ...)   # no side effect but messy
obj <- lst$obj        # doing this many times is not fun
res <- lst$res
```

RC/R6 make things easier to work with:

An aside: RStudio Community

A great place to ask "dumb" questions that might get negative responses on, for example, Stack Overflow



RStudio Community

AmeliaMN
Amelia McNamara

Summary | Activity | Notifications | Messages | Invites | Badges | Preferences

All
Topics
Replies
Drafts
Likes Given
Bookmarks
Solved
Votes

Download All

Topic	Replies	Views
Intuition for "direction" of mapping in ggplot2 Teaching	17	1.0k
Keeping track of Keynote presentations 1 Teaching	7	164
"Introducing yourself to git" in rstudio.cloud RStudio Cloud 0 votes	3	120
"hydrating" a packrat repository on a new computer? General packrat git github reproducible	14	366
Possible bug in View()-- NAs displaying as 0s. RStudio Cloud 0 votes	3	92
Possible View() bug-- NAs displaying as 0s 2 RStudio Cloud 0 votes	2	15
Live collaborative editing on RStudio cloud? RStudio Cloud 0 votes	3	268