Follow along! https://tinyurl.com/odsc-mapping

#ODSC

BOSTON
APR 30 – MAY 3

Mapping Geographic Data in R

Joy Payton
Supervisor, Data Education,
Children's Hospital of Philadelphia

https://tinyurl.com/odsc-mapping

Quick! Before we go any further

You should have R and RStudio installed (with packages at right) OR have a (free!) RStudio Cloud account (https://rstudio.cloud/)

Consider getting a Census API key (free): https://api.census.gov/data/key_signup.html

```
install.packages(c("sp", "spdep", "rgdal", "dplyr",
"tidyr", "leaflet", "leaflet.extras", "broom",
"scales", "xml2", "htmltidy", "jsonlite",
"kableExtra", "tidycensus", "printr", "maptools",
"htmltools", "mapproj"), dependencies = TRUE)
```

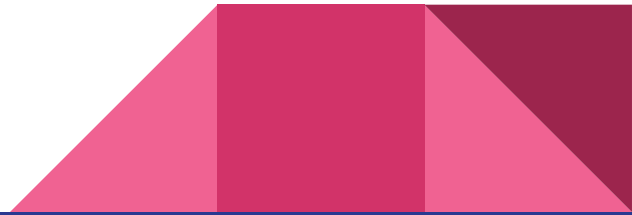
- sp
- spdep
- rgdal
- dplyr
- tidyr
- leaflet
- leaflet.extras
- broom
- scales
- xml2
- htmltidy
- jsonlite
- kableExtra
- tidycensus
- printr
- maptools
- mapproj
- htmltools

About Today (10 m)

About me

Joy Payton

- Not a GIS expert
- A generalist who has to work with lots of use cases
- I teach computational, reproducible research skills (R, Python, git…) to biomedical researchers at the Children's Hospital of Philadelphia



https://tinyurl.com/odsc-mapping

About You

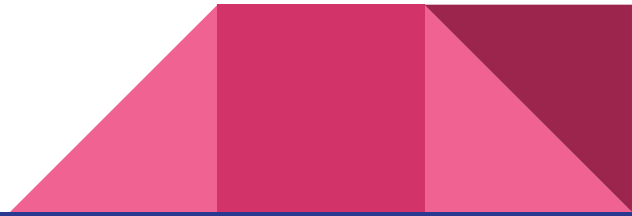
New to R ←→ R Expert ?

New to maps ← → Geo guru?

Care about US Census ← → Nah, just my data

Non-profit ← → Industry

Want to stop early ← → Want to saturate the brain

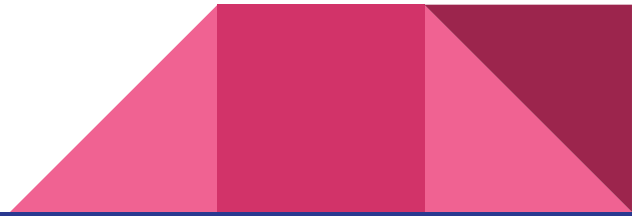


https://tinyurl.com/odsc-mapping

Our Itinerary

- Map files -- what even are they?
- Combining your data with map files
- Obtaining public data to enrich your data and map
- Visualization options

More code than time, we will titrate to the aggregate needs of the audience!

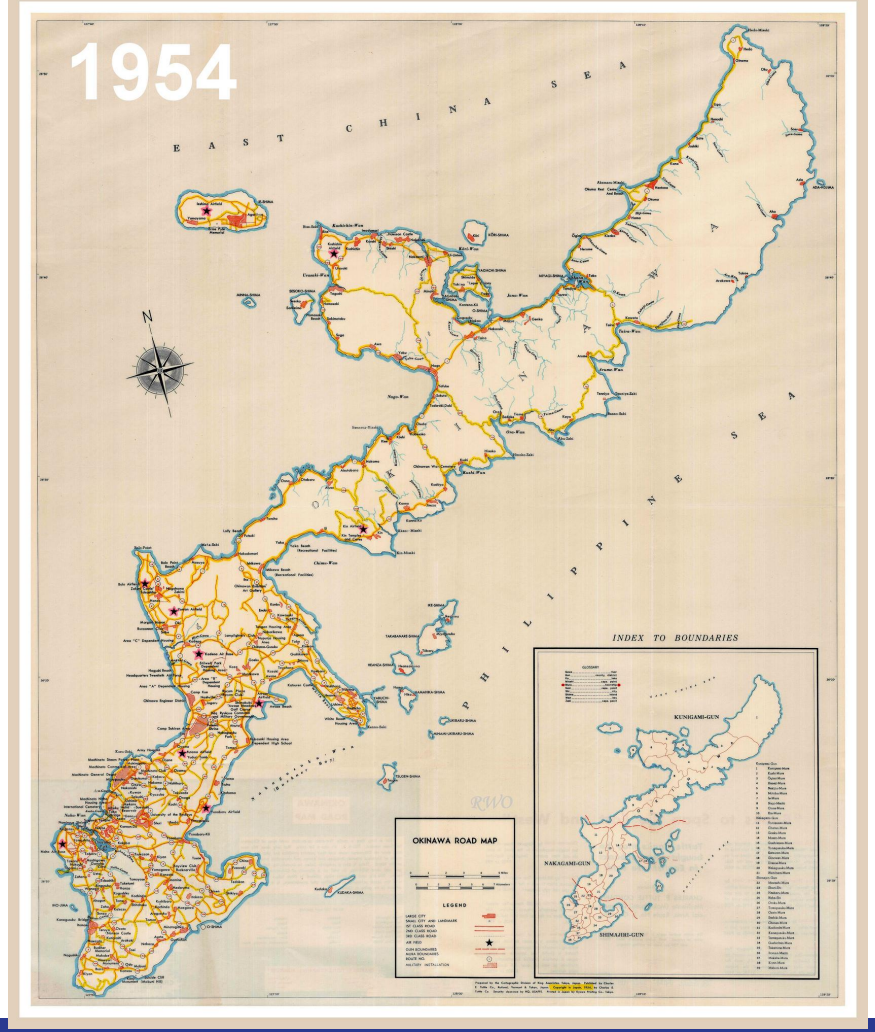


https://tinyurl.com/odsc-mapping

Why geospatial?

- We have some kind of data that we want to understand better in a geographic sense:
  - Customers (how far are they from my store?)
  - Patients (do they live in a high-crime area?)
  - Sensors (what are the air quality patterns near my child's school?)
  - Stores (proximity to gas stations = high sales?)
- We want to use maps to present data (why?)

https://tinyurl.com/odsc-mapping



https://tinyurl.com/odsc-mapping

Same idioms, ~ 10 centuries later!

Map of Okinawa

https://tinyurl.com/odsc-mapping

What's in a map?

Shapes:	Colors:	Sizes:	Words:
• Points • Lines • Polygons	• Hue (water is blue) • Intensity (e.g. water depth)	• Thick / thin lines • Large / small points • Solid / broken lines	• Scales • Numbers • Words

https://tinyurl.com/odsc-mapping

Why a map?

- Our data are influenced by geography
- Lots of public data we can combine w individual data
- Maps highlight things quickly for stakeholders
- Maps show patterns we might not see otherwise

Map Files (25 m)

https://tinyurl.com/odsc-mapping

Map files

- GeoJSON (IETF Standard) https://tools.ietf.org/html/rfc7946 (usually one file)
- Shapefile (ESRI standard) https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf (usually multiple files in a .zip)

Along with many folks (see, e.g. http://switchfromshapefile.org/) , I believe that geoJSON is a better format than Shapefile, but this is mostly due to the fact that JSON itself is so well-understood and easy to work with, so it's a simpler jump for me. Your needs may be different!

There are others with smaller market share:

- OGC GeoPackage
- OGC GML
- SpatiaLite
- OGC KML

Let’s open some map files!

Let’s just start working with some map files to see what they look like under the hood.

All of my files can be found at https://github.com/pm0kjp/mapping-geographic-data-in-r or https://rstudio.cloud/project/334226 (YMMV)

https://tinyurl.com/odsc-mapping

Takeaways:

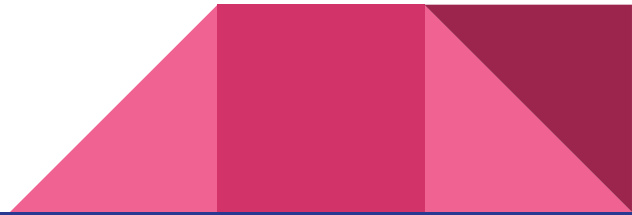
Main map files in the wild:

- Shapefiles (expect a .zip)
- geoJSON (expect .json)

Both can be read into R using the rgdal library and will have:

- A data frame with information about geographic features
- Some geography that gives the shape of those features

Leaflet is one quick way to draw a map



Incorporating Your Data (20 m)

Your data

You have some sort of geographic data that intersects with your map:

- Lat / long points of shale oil wells in Pennsylvania, and a map of congressional districts in PA
- Census tracts of patients in Long Island and a map of the census tracts in Long Island
- Air Quality index measurements in with lat / long and a listing of locations of cattle farms in Texas
- Sales records by country and a map of country boundaries in EMEA
- Philadelphia violent crime points (in lat/long) and a map of city council districts in Philadelphia

Some major assumptions:

Your data is tabular (is a data frame or can be coerced to become a data frame)

There's some sort of geographic locator in your data

… but I don't have your data!

And I work in a hospital, so I can't use mine...

The methods of data combining are the same, whether we're talking about your proprietary data or public data, so we'll use public data here for the most part (but we'll fabricate some data, too!)

All of my files can be found at https://github.com/pm0kjp/mapping-geographic-data-in-r or https://rstudio.cloud/project/334226 (YMMV)

Census Data (25m)

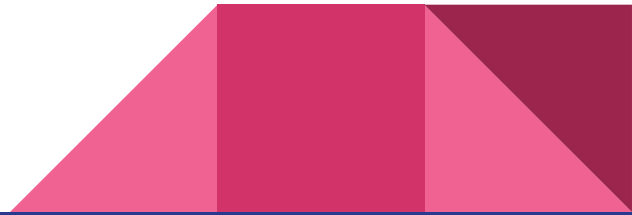
https://tinyurl.com/odsc-mapping

The power of public data: aggregation

Your patients, customers, employees might not give you information about the poverty, crime rate, or walkability of their neighborhood.

But with their geographic locations you can get aggregate information that can help you make some (broad, sweeping, therefore suspect!) generalizations.

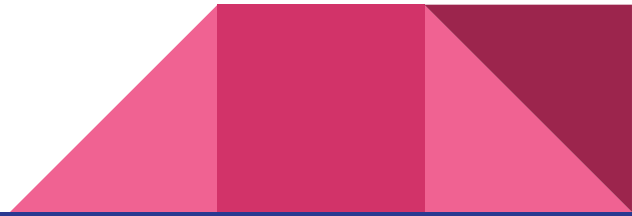
Sources include public data portals, US Census Bureau, World Bank...



https://tinyurl.com/odsc-mapping

United States Census Bureau

- Full decennial (10 year) census -- attempts to get full population
- American Community Survey (1- or 5- year)
- Economic, Government Censuses
- Free API (can be tricky, tidycensus can help!)



https://tinyurl.com/odsc-mapping

Let's look at combining public data!

Once again:

All of my files can be found at
https://github.com/pm0kjp/mapping-geographic-data-in-r or
https://rstudio.cloud/project/334226 (YMMV)

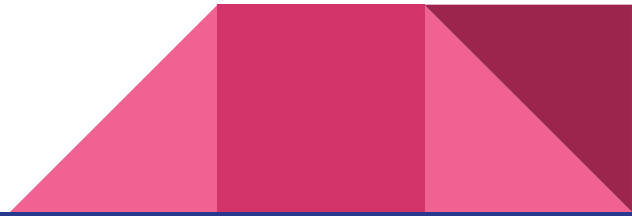
Variations on Mapping -- ggplot (if time remains, 10m)

We're not limited to leaflet...

You can map geospatial data using various R packages, like:

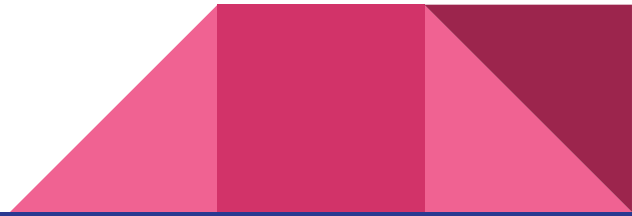
- Plain old "plot"
- plotly
- bokeh
- highchartr (maybe?)
- ggplot

Here we're just going to concentrate on ggplot2



ggplot2

- If you're not familiar, this is part of the so-called "Hadleyverse", now the "tidyverse"



Questions?

Feel free to tweet at me @KJoyPayton
Liked the session? Compliment me on Twitter!

Please also consider giving me all feedback! Your time is valuable -- did I give you value?