Exploring Spatial Data with sf

• • •

Sheila Saia
Thursday, July 19, 2018
Duke Library, The Edge Workshop Room

Lightning Talks & Announcements

Share something interesting and related to R or data science!

Sign-in Link: https://goo.gl/forms/kVY7hL2tjDXfgPwV2

Consider organizing our next R-Ladies RTP Meetup!

Some ideas we've discussed previously...
- Regular expressions

- Low-key chat (coffee shop or bar or both?)

- Unfinished talks

- Shiny (interactive applications and plotting)

- < insert your idea here >

Workshop Materials

You can find the workshop materials on GitHub at
https://github.com/rladies/meetup-presentations_rtp/tree/master/2018-07-19-sf

You will also need to download the data for the workshop. See GitHub download > workshop_data > README.md for the link.

Workshop Learning Outcomes

By the end of this workshop you will be able to:

1. **Describe** what the sf package is and how to use it

2. **Explain** some different sf spatial operations

3. **Apply** sf operations to a real-world geospatial dataset

What is (geo)spatial data?

Major Types of Spatial Data

- Vector data

- Raster data

Can you think of examples of each?
Can you think of example file extensions for each?

What is sf?

- sf stands for simple features

- Package was first published this past May (very new!)

- Was developed to replace the sp package, which didn't meet all ISO simple features standards

- Use for spatial operations & plotting (also works well with the tidyverse! - dplyr, ggplot, etc.)

- More info here: https://github.com/r-spatial/sf

USGS GAGES-II Dataset

"…Geospatial Attributes of Gages for Evaluating Streamflow, version II, provides geospatial data and classifications for 9,322 stream gages maintained by the U.S. Geological Survey (USGS)"

Available online at:
https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml

*We will also use some TIGER data…see README in the GitHub repository for more info.

USGS GAGES-II Dataset (continued)

Includes:
- Watershed bounds by region (reference and non-reference)

- Watershed characteristics (tabular data)

- USGS gage locations

Activity 1: Getting Started with sf

Open up sf_workshop_script.R

Install sf and load libraries

Change the directory paths based on where you saved the GitHub repository

Load in the spatial data using st_read( )

Use class( ) and glimpse () to view one of the new files you loaded in

What might you want to do with your vector data?

Think about spatial operations that might be helpful to use.

Make a list with your group of 2-3 people (5 min)

Discuss with everyone (5 min)

Activity 2: Common Operations in sf

Open up sf_workshop_script.R

Let's try some of the operations we just talked about.
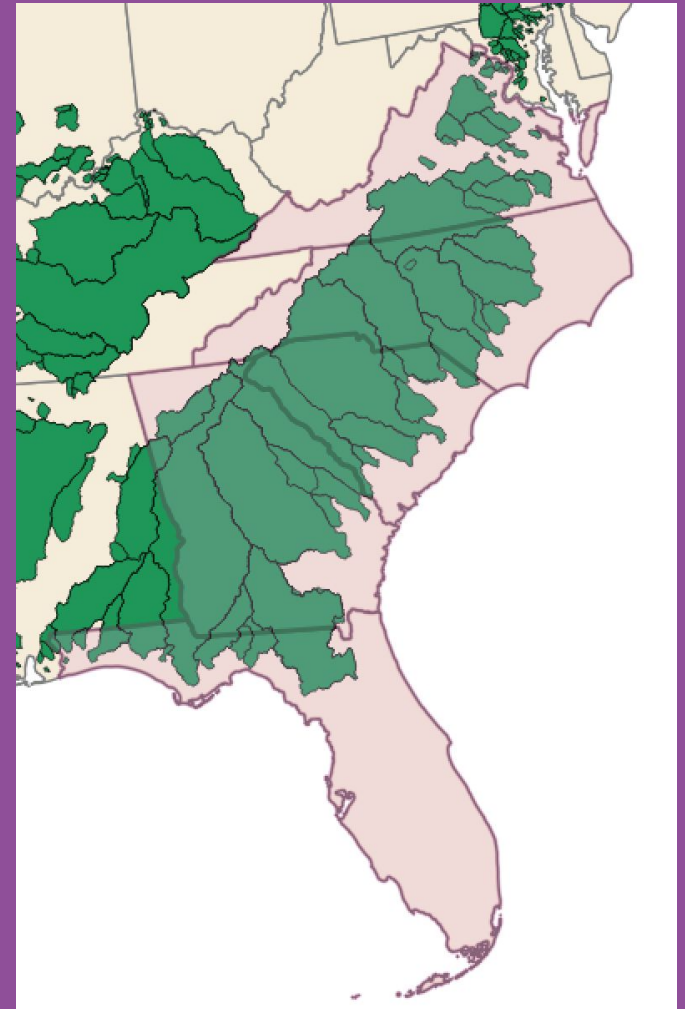
*(For later) Look here for more inspiration:
https://cengel.github.io/rspatial/3_spDataOps.nb.html

Activity 3: Putting sf to Work with USGS GAGES-II Data

Pick your favorite SE state
(Hint: Make a new variable)

Find all the watersheds that overlap

Plot the watersheds in your favorite
state.

Activity 4: Bringing in the tidyverse

Use your favorite state's watersheds (from Activity 3)

Select all the gages in that watershed

Join some of the tabular data and color these gages by your favorite variable (Hint: see the tabular_data file for GAGES-II variable descriptions)

Extra: Download USGS gage data for one (or multiple) gage(s) in your state using the dataRetrieval package
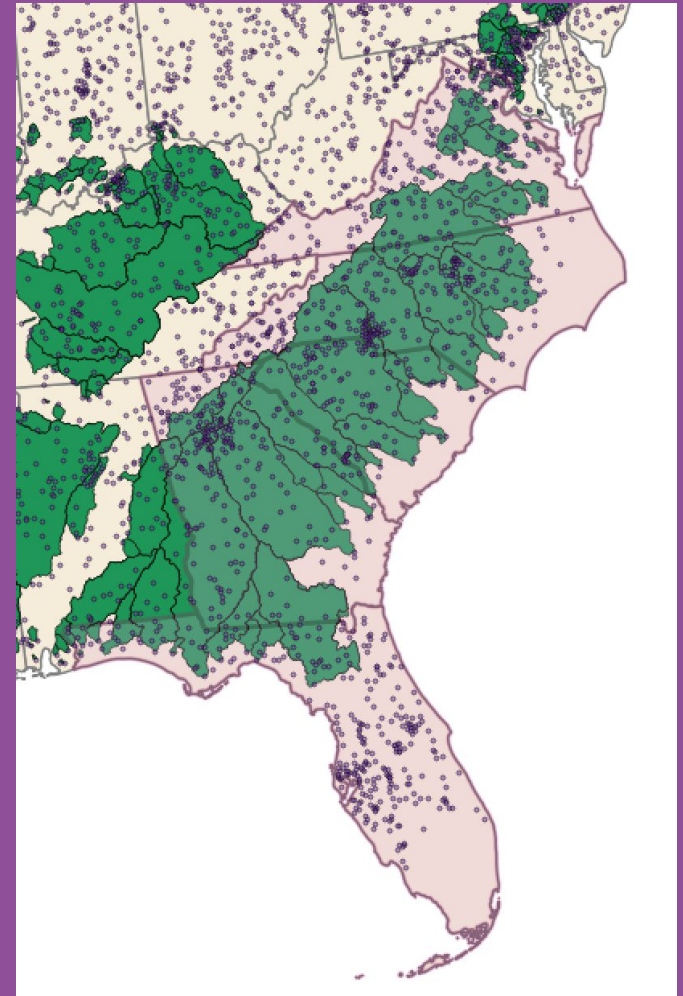
Workshop Learning Outcomes Revisited

Now that it's the end of this workshop, we hope you are able to:

1. **Describe** what the sf package is and how to use it

2. **Explain** some different sf spatial operations

3. **Apply** sf operations to a real-world geospatial dataset

Consider organizing our next R-Ladies RTP Meetup!

Some ideas we've discussed previously...
- Regular expressions

- Low-key chat (coffee shop or bar or both?)

- Unfinished talks

- Shiny (interactive applications and plotting)

- < insert your idea here! >

Sign-in Link: https://goo.gl/forms/kVY7hL2tjDXfgPwV2