

pennsylvania
DEPARTMENT OF ENVIRONMENTAL PROTECTION

BUREAU OF CLEAN WATER

Geospatial Operations in R

Learn how to use R as GIS and incorporate geospatial operations into your R workflows

Matt Shank

Feb. 9, 2024

PA AFS Workshop - Meadville, PA

Outline

- Introduction
  - R background and capabilities
  - Importance of projects
  - Keyboard shortcuts
  - Piping and tidycoding principles
  - Overall utility of geospatial workflows in R
- Background on the example – acidification in PA
- Module 1: Importing/accessing geospatial data in R
- Module 2: Geospatial Operations
- Module 3: Exporting spatial objects
- Module 4: Map creation

Introduction

- What is R?
  - R is a language and environment for statistical computing and graphics.
- The R environment
  - R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:
    - an effective data handling and storage facility,
    - a suite of operators for calculations on arrays, in particular matrices,
    - a large, coherent, integrated collection of intermediate tools for data analysis,
    - graphical facilities for data analysis and display either on-screen or on hardcopy, and
    - a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.
- TLDR: R can handle your database querying, data wrangling, cleaning, statistical, reporting, and **geospatial** problems!

Source: https://www.r-project.org/about.html

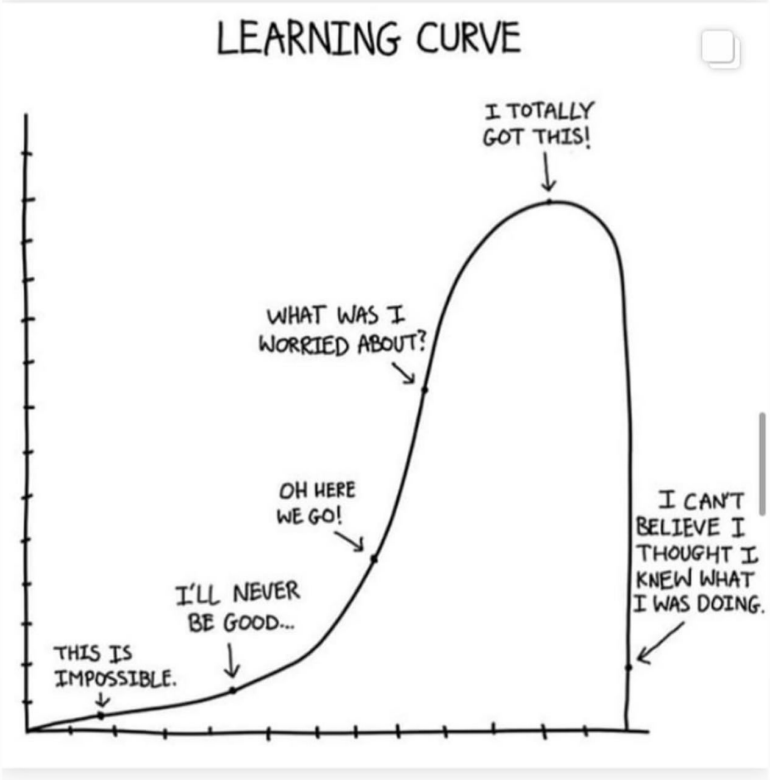
Introduction

***First**, teaching R is often confounded with teaching statistics. Pick one (preferably R first), and then move on to the other.*

- I'm teaching both geospatial techniques and R in four hours today (sorry)

***Second**, start by only teaching the most essential and important functions first. Don't overwhelm your students with all the functions they might ever need to know. And if you know two ways to do the same thing? Just pick one.*

- tried to strike a balance here, focusing on only the most commonly used sf functions, wrapping it into a tangible example workflow



Source: https://www.r-bloggers.com/2021/07/the-myth-of-the-r-learning-curve/

Introduction

- What is an R?

R is the engine

R Studio is the dashboard, plus bells, whistles, etc…

Introduction

- First some ground rules
  - Create a project

> *If the first line of your R script is*
>
> ```
> setwd("C:\Users\jenny\path\that\only\I\have")
> ```
>
> *I will come into your office and SET YOUR COMPUTER ON FIRE 🔥.*

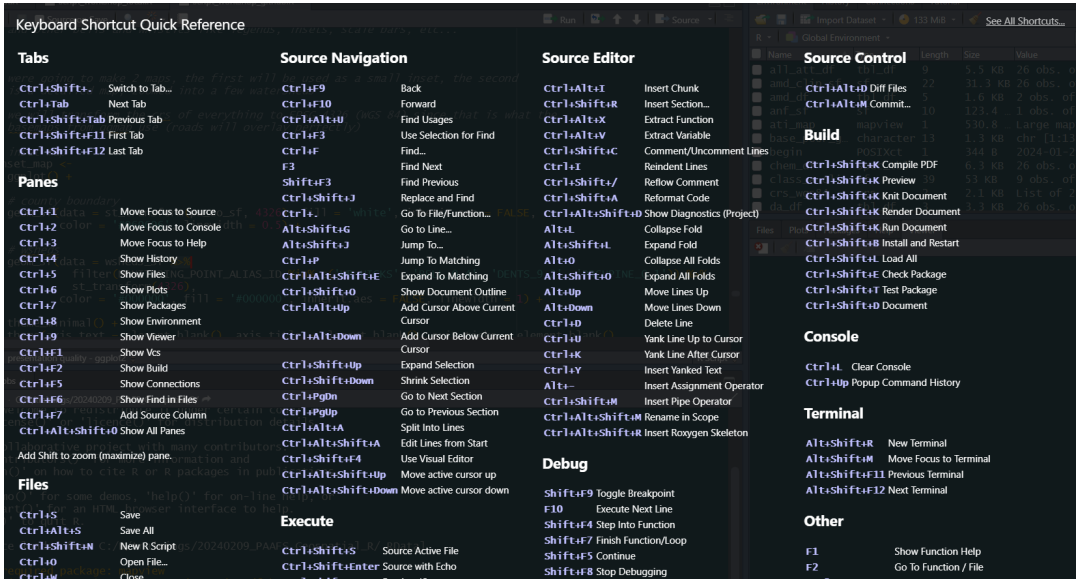
> *If the first line of your R script is*
>
> ```
> rm(list = ls())
> ```
>
> *I will come into your office and SET YOUR COMPUTER ON FIRE 🔥.*

Recommended reading: https://r4ds.had.co.nz/workflow-projects.html ; https://www.tidyverse.org/blog/2017/12/workflow-vs-script/

Introduction

- First some ground rules
  - Learn your keyboards shortcuts.
  - Do you know how annoying it is to type <- or %>%???

  - ***Tools → Keyboard Shortcuts Help***



- Alt + - = <-
- Ctrl + shift + m = %>%
- Ctrl + shift + c = #
- Ctrl + I = indent
- Ctrl + c = COPY
- Ctrl + v = PASTE
- Ctrl + shift + r = new section

Introduction

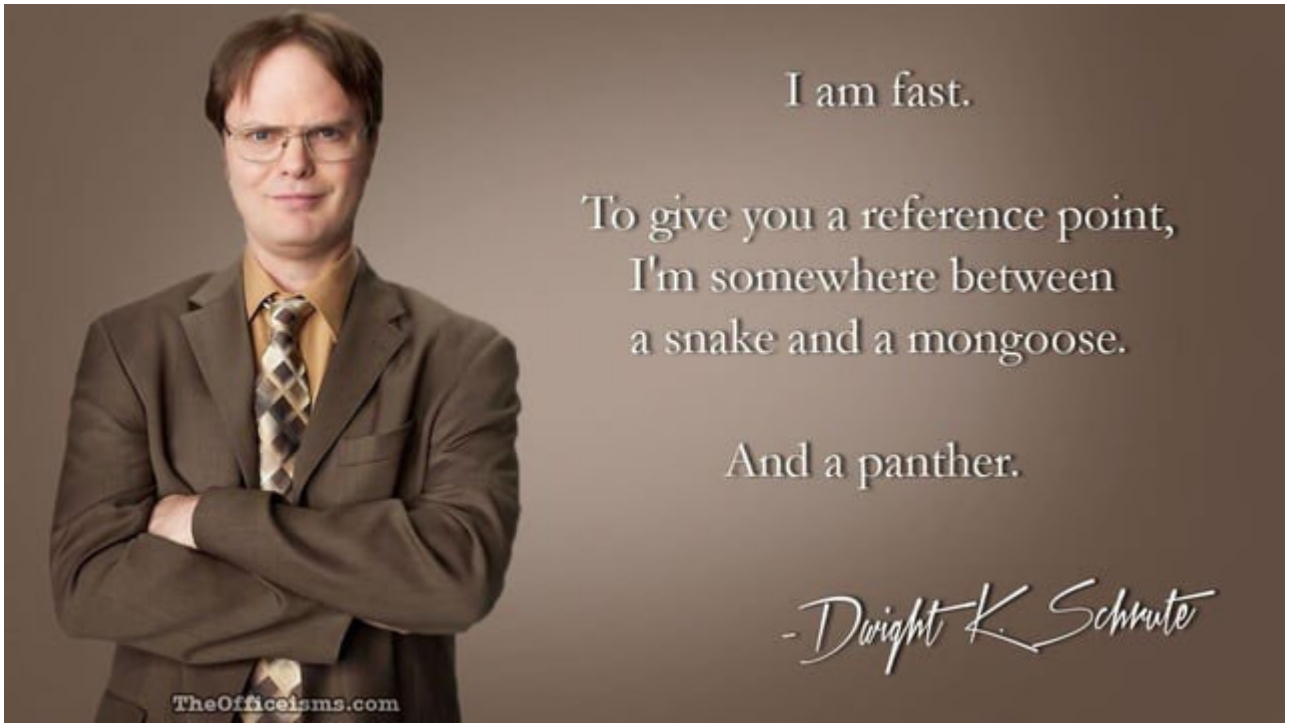
- First some ground rules
  - Today we will follow tidy coding principles
  - Our code will be ***human centered, consistent, composable, and inclusive***

solve complex problems by breaking them down into small pieces, supporting a rapid cycle of exploratory iteration to find the best solution.

```
first _df %>% # start with a data frame

 left_join(other_stuff_df, by = “shared_col") %>% # join other stuff

 select(-dumb_col) %>%  # drop dumb column

 mutate(new_col = number_col * 18.265) # multiply existing column by a constant
```

Introduction

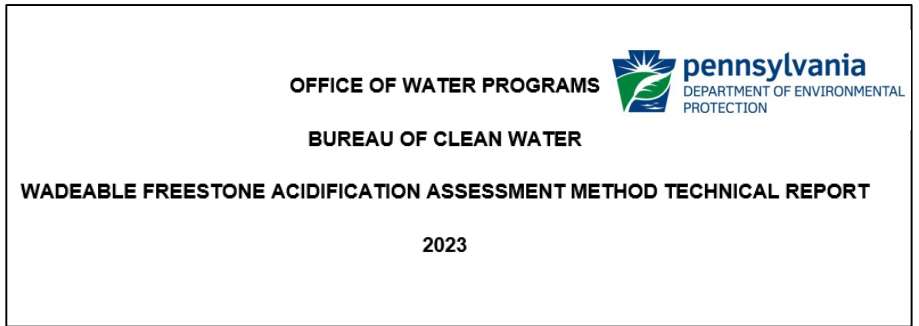
- Overall utility of geospatial workflows in R
  - Unite data wrangling and statistical analysis capabilities with powerful geospatial functions



Acidification Example

- PADEP just developed a draft acidification assessment method
- This method documented relationships between surface water acidification (low pH) and geology, roads, and AMD
- The method also created a new macroinvertebrate metric, the Acid Tolerance Index (ATI), which is responsive to stream pH

- Core concepts
  - 'Base-poor' geology formations – detailed in Greg Moyers presentation yesterday
    - The higher % base-poor geology in a watershed, the lower the stream pH
    - The lower the stream pH, the higher the macroinvertebrate ATI score
  - High road density can increase pH
  - AMD can lower pH



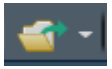
Acidification Example

- Today we will bring in 26 watersheds from the acidification study in the four counties surrounding the Allegheny National Forest
  - Access data from github, REST services (PADEP, PASDA)
  - Make tabular data with lat & long spatial object (geometry column)
  - Clip geology, streams, roads, AMD discharges to watersheds
  - Buffer streams, dissolve geometry, clip roads again
  - Calculate statistics (length, area, density, ratios)
  - Plot tabular geospatial output against pH and ATI scores
  - Create a presentation quality static map, and a stand-alone interactive map
  - And some other stuff….

Set up Project; Download script

1. Create a new project!
    1. Navigate to your preferred directory location
    2. Create a new folder (call it 'R_geospatial_workshop')
    3. File -> New Project -> Existing Directory -> Browse to above location. Click Open
2. In this directory, create 2 new folders called:
    1. github_files (we will download objects from github in here)
    2. output (we will save output from R in here)

- Navigate to https://github.com/FWeco/Geospatial_R
- Download script_workshop_github.R 

- Place in your directory – open in R studio 
- Now we are good to go

Module 1 - Importing

- Importing/accessing geospatial data in R
  - Import various formats
    - .rds, .shp, tabular (x,y) - covered
  - Various spatial objects
    - Vector - Points, lines, polygons - covered
    - Raster - not covered
  - From various sources
    - Online REST services (e.g. PASDA, PADEP)
    - Local or network directories/GIS geodatabases
    - R-ArcGIS Bridge – not covered (but amazing)
      - Set up tutorial
  - Checking and transforming coordinate reference system (crs)

Module 1 - Importing

- R-ArcGIS Bridge (not covered)
  - A couple notes:
  - Allows you to access ESRI GIS resources as long as you have a valid ESRI license
  - Magical
  - Install the arc bridge
  - Useful vignette

Get started:

```
install.packages("arcgisbinding", repos="https://r.esri.com", type="win.binary")
library(arcgisbinding)
arc.check_product()
```

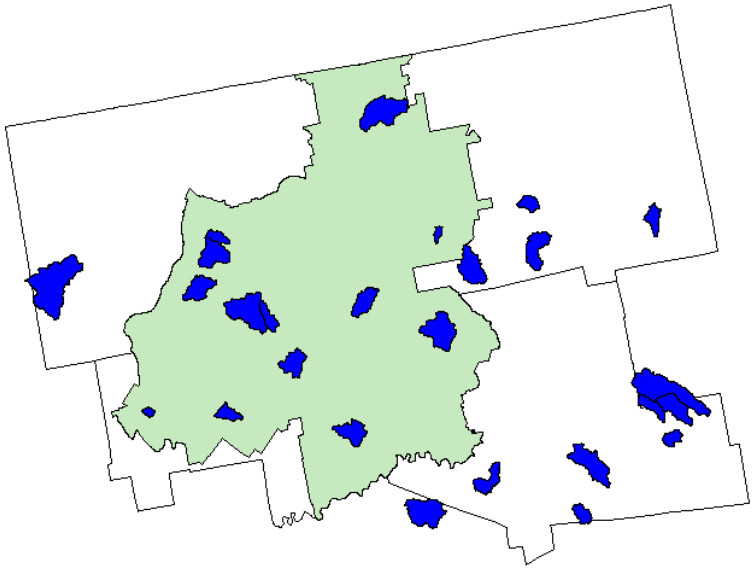
* Not covered today, but highly recommended

Module 1 - Importing

- Online REST services (covered)
  - Representational state transfer
  - API (application programming interface) that helps you communicate with the information source
  - TONS of geospatial resources available
  - PASDA REST services tutorial:
    - https://rpubs.com/mattshank20/pasda_api
- Local/network geospatial data in various formats (covered)[1]
- Local/network tabular data in various formats (covered)[1]
  - [1]Our data will be sourced from github for ease of access

Module 1 - Importing

- Run Module 1 code
- Github repo:
  - https://github.com/FWeco/Geospatial_R
- url needed for R code:
  - https://raw.githubusercontent.com/FWeco/Geospatial_R/master/



Module 2 – Geospatial operations

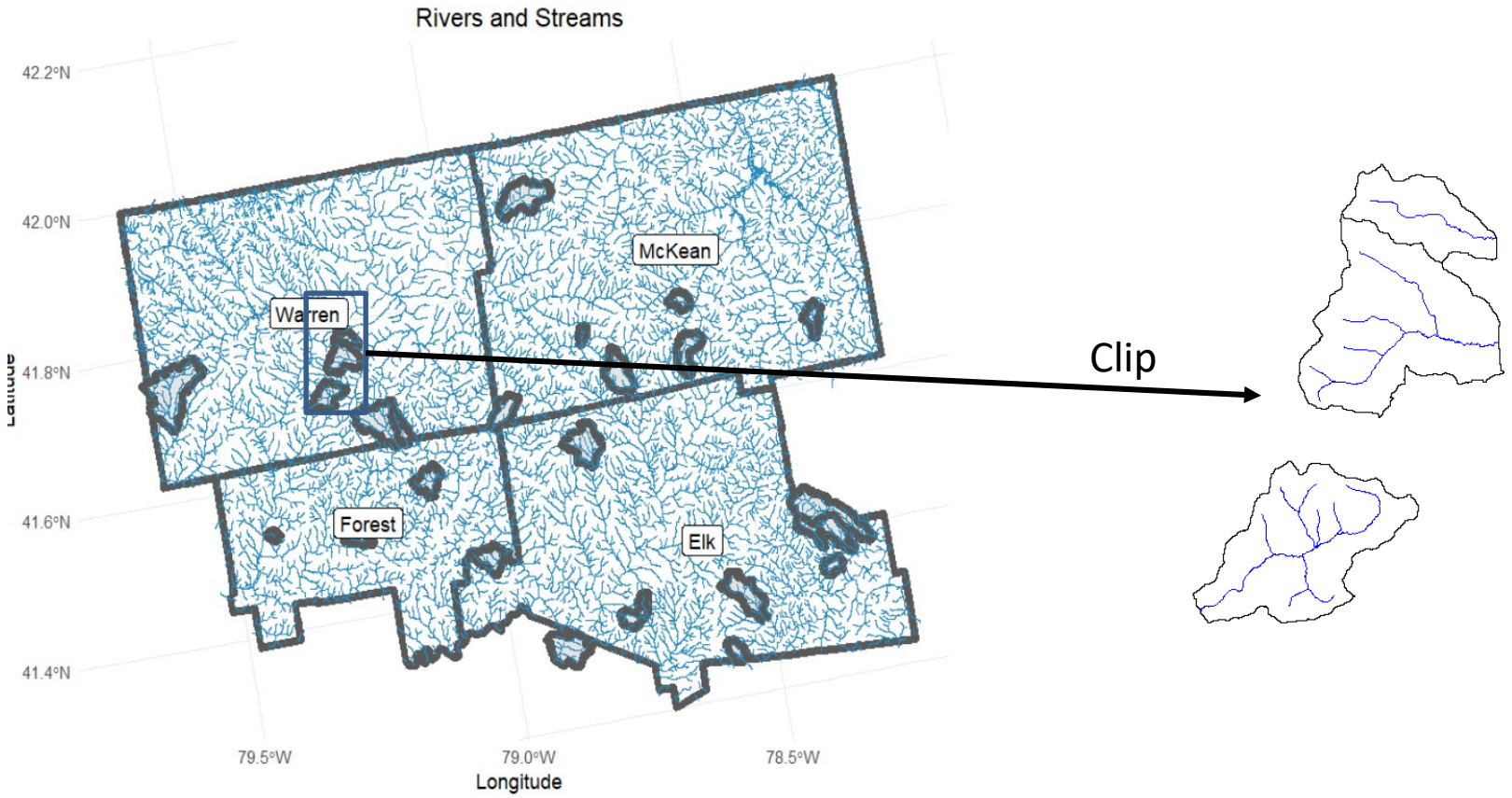
- After we have imported our geospatial and tabular data, we can begin

- Geospatial operations:
  - clip, buffer, dissolve, spatial selection, spatial join (covered)

- Calculate statistics:
  - length, area, density, ratios

Module 2 – Geospatial operations

- Run Module 2 code



Module 3 - Exporting

- Can save to any preferred file format:
  - .rds, .rda, .shp
  - In one step, can project tabular (x,y | lat,long) data and export to .shp!

Module 4 - Map creation

- Static
  - Base plots (what?)
  - ggplots (Okay!)
- Interactive
  - mapview() - html (YEAHHHH!)

For detailed training on interactive maps with **leaflet**, see the 2021 PA AFS workshop
Interactive Mapping in R: Don't get caught up in the static
https://rpubs.com/mattshank20/intMappingR

Wrap Up

- Takeaways
  - Between your agency/school/institution's ESRI resources and REST services from places like PASDA and PADEP, there is a mountain of geospatial data available to you right now
  - Importing / analyzing / exporting / and map creation is simple
  - This removes the need for ‘intermediate steps’ in workflows
    - ~~OLD: pull tabular data from database to excel, export to GIS for geospatial analysis, reimport to R for modeling~~
    - NEW: pull data from database with R (ODBC, DBI, dbplyr), geospatial analysis in R, model in R
  - The ‘ideal world’ of workflows

Wrap Up

- Takeaways
  - Keep st_zm() and st_make_valid() at the ready if you run into errors with invalid geometry
  - after you do something, plot it to give it a quick look
  - Run your code a chunk at a time to evaluate each step
  - Clean up your friggin environment. If you don't need something again, DECLUTTER

Wrap Up

- Takeaways
  - Interactive maps are one of my favorite things in the world. You can communicate very complex dynamics very clearly by sharing a well made map
  - Static maps also have their place. And with a few rounds of trial and error (and attention to detail), ggplot makes some damn fine maps

Acknowledgments

- R code troubleshooting
  - Dustin Shull, Ty Wagner, Becky Whiteash, Mark Hoger

Send me your cool R code:
mattheshan@pa.gov

Miners Run – a beautiful acidified stream