Title: Analyzing Stormwater Best Management Plan Data

Data encased in lists, wrapped in vectors, swaddled in JSON

Erik H. Beck

USEPA R1

17 Oct 2023

Outline

Visualization

Figure: (https://images.freeimages.com/images/large-previews/6a4/russian-nesting-doll-1187383.jpg)

Nature of the problem:

- Data encased in lists, wrapped in vectors, swaddled in JSON

Speedy Talk: Quick Overview

- Stormwater::CWA 319::BMPs::GRTS
- Q: Why the GRTS System?
- A: Strings attached to Govt. money!
- Oracle Business Intelligence: Clunky and not very capable
- Basic questions to answer:
  - What is the mean level of funding for Nonpoint Source projects in New England (and the confidence interval)?
  - How is that funding scattered across appropriation years? What does that look like graphically?
  - Can't answer these basic questions using OBI

Stormwater, CWA 319, Nonpoint Source Pollution, BMPs, and GRTS

- EPA's GRTS system a required part of getting EPA funding under CWA 319 (NPS)
- States have to enter the data on projects using EPA funding
- Entered data is broad info on work done under EPA grants
- Data goes into Grants Results and Tracking System (GRTS)
- OW/OWOW builds and maintains the system
  - Data entered by states, tribes, regions
  - System can export some data 'natively' to PDF and CSV, but pretty limited

Getting the Data to R

- OW has developed one Application Programming Interface (API) to GRTS
  - Uses HUC12 to find data
  - One to many relationships of data
  - One HUC may have zero to many BMP projects associated with it across years, grants, etc
  - More in development
- Using the API returns a single JSON structure for a HUC12
- JSON is deeply nested; challenge is picking apart the material and massaging it for analysis

Code Fragment


```
HUCmetaData <- function (DataRequestRaw) {
  HUCdataBlobRaw <- DataRequestRaw$content
  HUCdataBlob <- rawToChar(HUCdataBlobRaw)
  HUCdataBlob %>% spread_all -> HUCmetaDataBall

  document.id <- HUCmetaDataBall$document.id
  count <- HUCmetaDataBall$count
  # The Payload is in 'HUCmetaDataBall$..JSON'.
  ## Extract elsewhere.
  HUCMetaData <- data.frame (document.id,count)

  return (HUCMetaData)
 }
```

Basic Analysis

After the GRTS data is tamed, we can start answering the basic questions

Question One

What's the mean and 'data spread' of GRTS funding in New England?

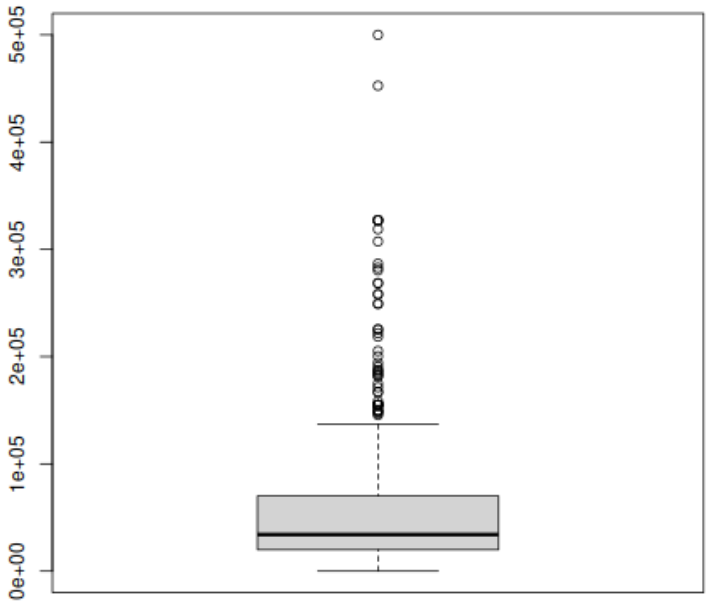

Figure: USD ($) for CWA 319 Stormwater BMP Funding in New England

Question Two

What does the data look like when grouped by appropriation years?

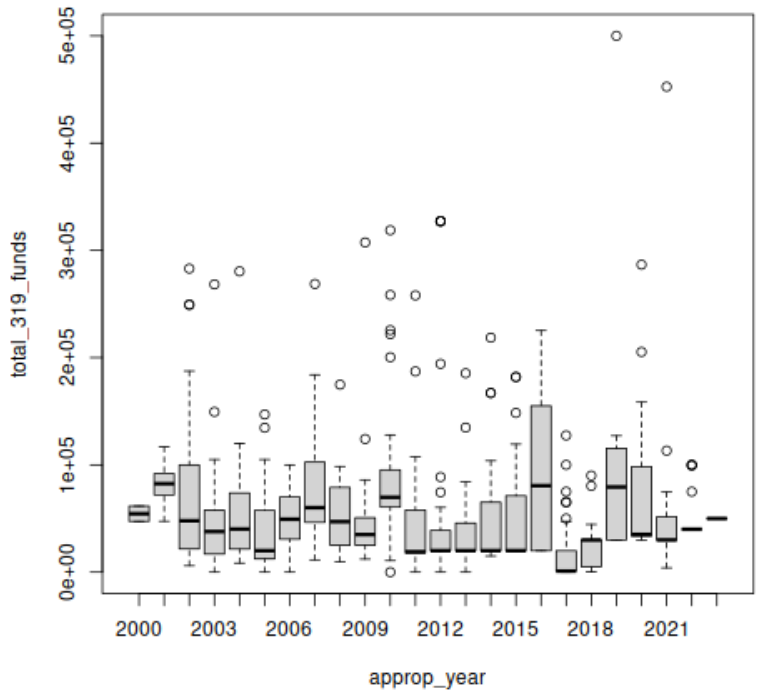

Figure: CWA 319 Stormwater Funding by Federal Fiscal Year in New England

Conclusion

- Using R to analyze GRTS data is very useful
- APIs need some work to provide more parts of the dataset
- Other APIs in development need to mature and move to the production system
- Working with the deeply nested JSON data is worth it; trying to coerce this info from OBI would be very difficult or impossible.

Questions?

I can take a few questions.