The Moore Institute Open Data Portal:
GIS Standardization of Microplastic Pollution Research

Masters of Science in Geographic Information Science (MSGISci)
Department of Geography, California State University, Long Beach

Holden Ford
Haig Minasian
Anja Oca

Introduction

Microplastics have invaded freshwater ecosystems all over the world. In America, there is a lack of coherence between datasets from rivers in different states. Many of the open datasets only include pictures or descriptions of the location instead of actual geocoded points of the sample site.

Why GIS?

Microplastics are not a localized issue. It is global, and understanding how microplastics move and accumulate downstream requires data visualization. A map is one of the best visualization tools for this because it represents the real world in a two dimensional way.

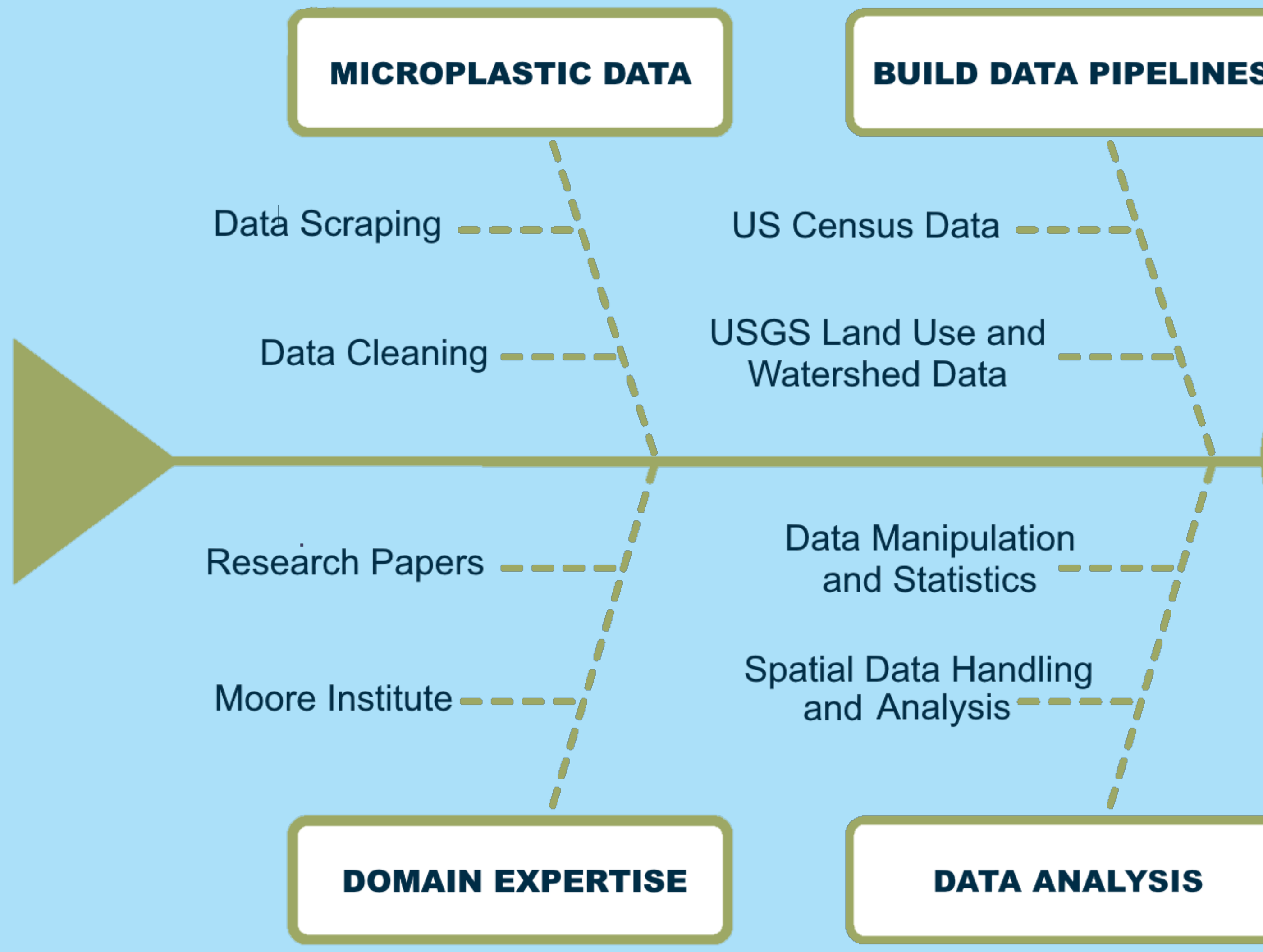
Why Open Source?

Open source is the nurturer of collaboration. An open source program means anyone can take what has been made and make it better. The work can continue after we are no longer there.

Shelly and Charlie Moore at the lab

Image Credit: mooreplasticresearch on Instagram.com

Methods



Data Pipelines
- Analysis required the use of outside data
- Land Use data came from the USGS websites

Data Analysis
- Zonal statistics created the land use rasters
- Pie charts for land use percentage created using R

Data Scraping:
- Journals about plastics were scraped
- Papers about plastic in river systems
- Many categories were scraped at first

Data Cleaning:
- Macroplastics and data outside the US were removed
- Differences in formats between scrapers standardized
- Superfluous and incomplete categories were removed

Domain Expertise:
- As papers were being scraped, we understood the problem of microplastics better
- Material science of plastic was important to understand how it breaks down

Data

Data for 74 waterways in 14 US States was taken from 10 different manuscripts. Twenty-five papers were initially scraped. The decision was made to cut 15 of them due to either being incompatible or located outside the US. Across the remaining 10 manuscripts, there were 996 samples.

A study was deemed incompatible if the concentration was not measured in a particles per volume unit or if it included macroplastics. Columns were removed if they were incomplete in many studies or if information was not pertinent. Finally, remaining data was converted to particles per cubic meter if not already and averaged to represent each river only once per state.

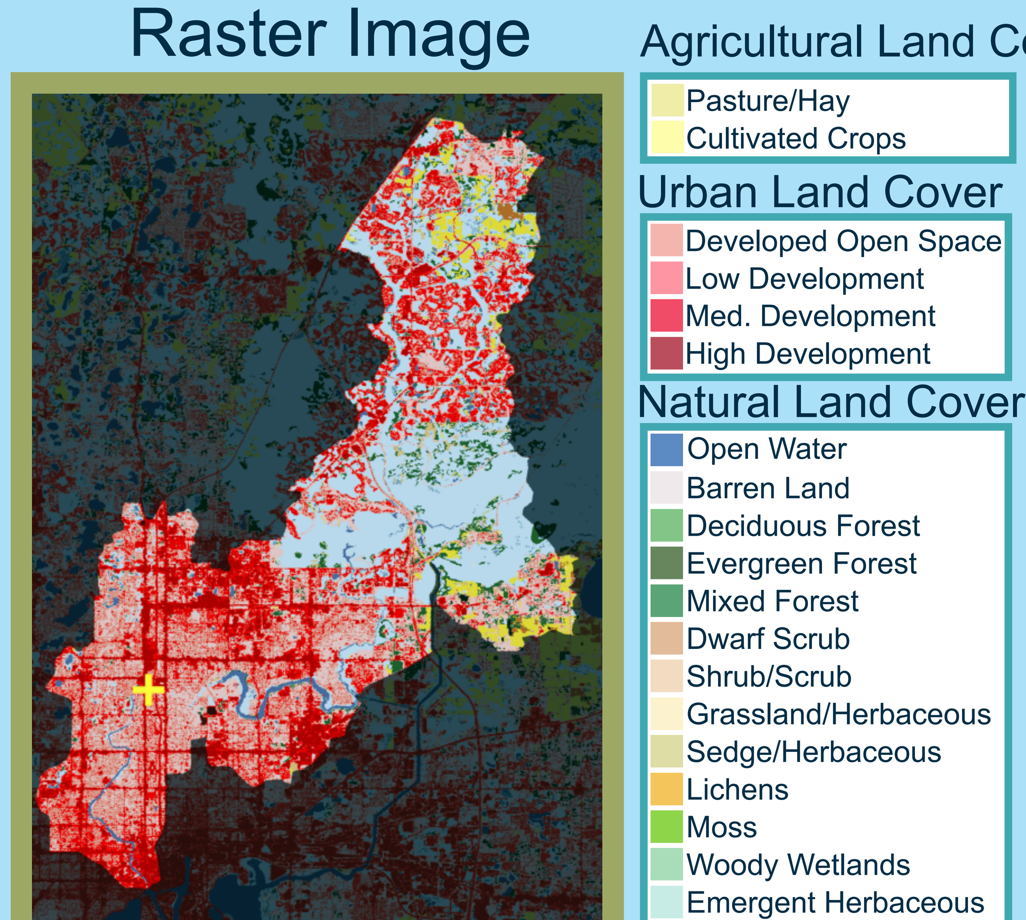
Stats about studies scraped:
- How many had coordinates: 1
- How many had maps: 7
- Max with same concentrations unit: 7

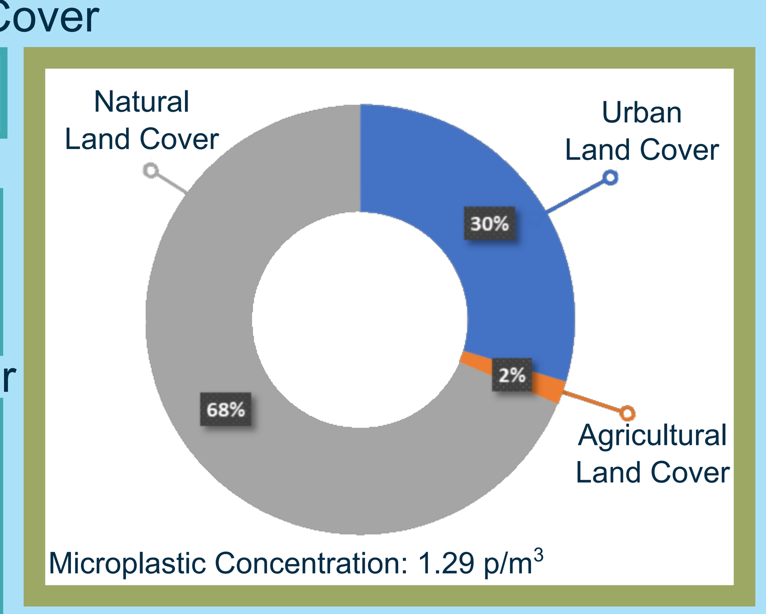
Sample of the Finished Data Table

Sample ID	DOI	Sample River	STATE	No. of Samples	Total Mean Conc	Units	Latitude
MPGISID1	doi.org/10.1021/acs.est.6b02917	Ashtabula River	OH	3	10.3	p/m3	41.904399
MPGISID2	doi.org/10.1016/j.watres.2018.10.013	Bacon Rind Tributary	MT	2	400	p/m3	44.929961
MPGISID3	doi.org/10.1016/j.watres.2018.10.013	Bear Creek Tributary	MT	2	900	p/m3	45.615627
MPGISID4	doi.org/10.1016/j.watres.2018.10.013	Big Sky Tributary	MT	4	550	p/m3	45.267266
MPGISID5	doi.org/10.1016/j.watres.2018.10.013	Black Butte Tributary	MT	6	467	p/m3	45.054072
MPGISID6	doi.org/10.1021/acs.est.6b02917	Black River	OH	3	1.2	p/m3	41.413951
MPGISID7	doi.org/10.3389/frans.2022.857694	Buena Vista Pond	TX	4	1225	p/m3	31.570041
MPGISID8	doi.org/10.1016/j.watres.2018.10.013	Buffalo Horn Creek	MT	6	617	p/m3	45.126678
MPGISID9	doi.org/10.1021/acs.est.6b02917	Buffalo River	NY	3	4.1	p/m3	42.875192
MPGISID10	doi.org/10.1021/acs.est.6b02917	Burns Ditch	IN	3	0.3	p/m3	41.642202
MPGISID11	doi.org/10.1021/acs.est.6b02917	Clinton River	MI	4	12.2	p/m3	42.596659
MPGISID12	doi.org/10.1021/acs.est.6b02917	Cuyahoga River	OH	3	2.6	p/m3	41.385661
MPGISID13	doi.org/10.1016/j.watres.2018.10.013	Deer Creek	MT	6	1067	p/m3	46.840344
MPGISID14	doi.org/10.1002/ecs2.1556	DuPage River	IL	8	8.1	p/m3	41.697565
MPGISID15	doi.org/10.1002/ecs2.1556	E Br DuPage River	IL	8	6	p/m3	41.742995
MPGISID16	doi.org/10.1016/j.scitotenv.2019.02.028	Fall Creek	NY	6	15833.33	p/m3	42.454679

Results

Raster Image





Middle Hillsborough River Watershed

A raster of the proportions of land use was made using National Land Cover Database (NLCD) from 2019 made by the USGS. The pixels in the raster were counted up to create a pie chart. This pie chart can be used in a microplastics vs land use study.

The NLCD package has many land uses, some only making up a few pixels of the raster. The land uses were combined into 3 groups for the pie chart. Agricultural land cover is the smallest group comprised of only pastures and cultivated crops. Urban land cover includes all developed areas, including open space like parks. Natural land cover is comprised of all natural environments.

R Code

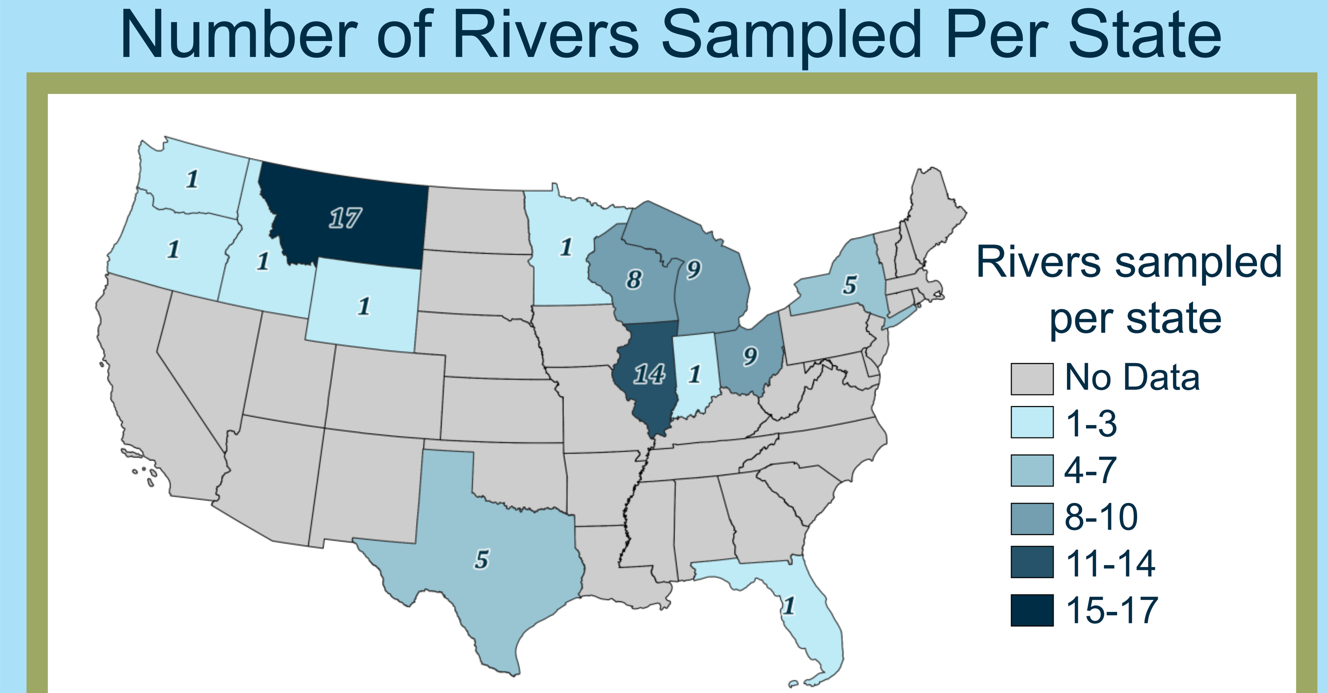
```r
# Step 4: Run Zonal Histogram processing tool to get raster values
zonal_hist_result <- qgis_run_algorithm("native:zonalhistogram",
                                        INPUT_RASTER = clip_output,
                                        RASTER_BAND = 1,
                                        INPUT_VECTOR = watershed_output,
                                        OUTPUT = "TEMPORARY_OUTPUT")
zonal_hist_output <- qgis_extract_output(zonal_hist_result)

# Define a mapping from original categories to new ones
category_mapping <- data.frame(
  Original = c("Developed, Open Space", "Developed, Low Intensity",
               "Developed, Medium Intensity", "Developed, High Intensity",
               "Pasture/Hay", "Cultivated Crops",
               "Barren Land (Rock/Sand/Clay)", "Decidous Forest", "Evergreen Forest",
               "Mixed Forest", "Shrub/Scrub", "Grasslands/Herbaceous", "Woody Wetlands",
               "Emergent Herbaceous Wetlands"),
  New = c("Urban Land Cover", "Urban Land Cover", "Urban Land Cover", "Urban Land Cover",
          "Agricultural Land Cover", "Agricultural Land Cover",
          "Natural Land Cover", "Natural Land Cover", "Natural Land Cover",
          "Natural Land Cover", "Natural Land Cover", "Natural Land Cover",
          "Natural Land Cover", "Natural Land Cover")
)

# Rebin the histogram results into the new categories
zonal_hist_output <- zonal_hist_output %>%
  left_join(category_mapping, by = c("Label" = "Original")) %>%
  group_by(New) %>%
  summarise(Value = sum(Value))

# Create a pie chart
pie_chart <- plot_ly(zonal_hist_output, labels = ~New, values = ~Value, type = 'pie') %>%
  layout(title = "Zonal Histogram")

return(pie_chart)
}
```

When a new NLCD dataset is released by the MRLC, the map and pie charts can be updated with this code. Eventually, with enough data, this can be used to examine if there is a correlation between microplastic concentration and land use. Code, instructions and recommendations were uploaded to GitHub for future microplastics researchers interested in GIS studies.

Further Research

Studies generally focus on a single watershed. Sometimes this is only one state or one river. That might not seem like much, but together they can cover a larger area. This is why open data is so important. With collaboration, a larger coverage of states will be possible.

Number of Rivers Sampled Per State



Visualizations and geospatial analyses were limited by access to and quality of data in the papers scraped. Open collaboration within the microplastics community can fill in the gaps encountered. Once there is more representation in North America, the focus can be shifted internationally.

The creation of standardized methods for collecting microplastics data as well as an open data portal makes information about microplastics more accessible and transparent to those who need it.

Recomendations:

- Include exact coordinates for all sample sites
- A majority of concentration data is in particles per volume
- A majority of those papers use particles per cubic meter
- Create a unique identifier for all sites
- Only one value per site unless showing change over time

Our main takeaway from this project is that in order to combat this human health crisis we all must work together.

Acknowledgements

We would like to thank the members of the Moore Institute for their help and support especially Dr. Win Cowger and Hannah Hapich

Contact

Holden Ford - rholdenford@gmail.com
Haig Minasian - haigjackminasian@gmail.com
Anja Oca - anjaoca@gmail.com
https://mooreplasticresearch.org/

GitHub