


ENVIRONMENTAL DATA ANALYTICS: M8 – SPATIAL ANALYSIS

Fall 2022

Nicholas School of the Environment - Duke University

Catch up

- Making use of office hours

- Remaining schedule
- Class project – presentation & report

- Questions on Time Series Analysis

Making use of office hours

- Friday afternoons are the busiest! Don't wait…
- Check slack; help others
- Debugging strategies:
  - Where does your code break?
  - How is your code different from class examples?
  - Restart R & clear variables
  - Package conflicts (plyr vs dplyr)

Remaining schedule

Date	Class topic
Nov 7/9	Spatial Analysis (assignment due 11/11)
Nov 14/16	Data Scraping (assignment due 11/18)
Nov 21/23	*--Thanksgiving--*
Nov 28/30	Class presentations
Dec 5/7	*--Reading week--*
Dec 12/14	Class Project (due 12/14)

Course Projects

- What is it?
  - Generate a hypothesis and test it using your data skills!
  - Apply the data analytic workflow to a question/dataset of your choosing…

- What is expected?
  - Group project w/2-3 people
  - See rubric on website:
  - Report as knitted Rmd file
  - Use of Git/GitHub

Course Projects: Presentations

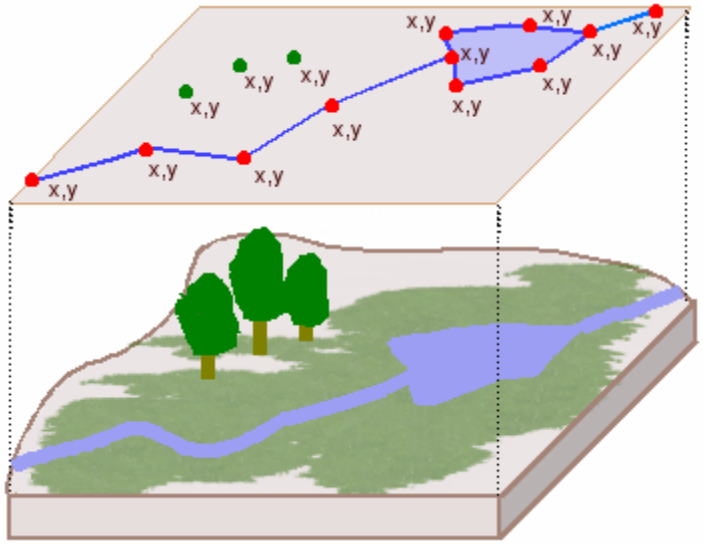
- Not expected to have completed your project!

- 5 minute presentations on your progress…
  - Title slide
  - Central question/hypothesis
  - Data used & relevant details (source, date, challenges)
  - Data exploration and wrangling
  - Data analysis and results
  - Summary and conclusions

M8.1 – Spatial Data

- Importance of Spatial Analysis in EDA
- Representing Spatial Data in R
- Simple Features & Spatial Dataframes
- Creating geometries
- Coordinate Reference Systems, Transformations, EPSG codes
- GeoJSON, Shapefiles, …

Simple Features…



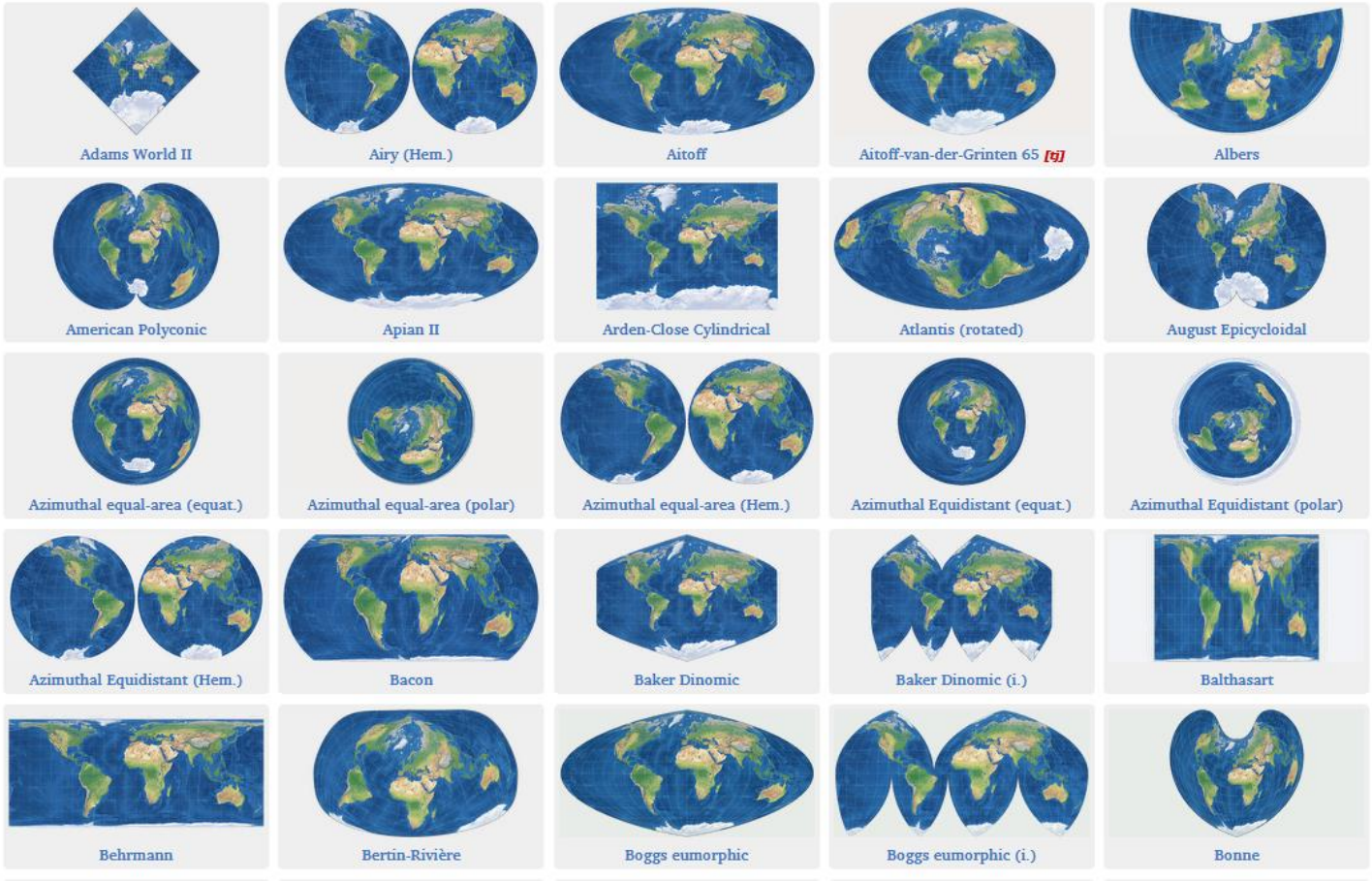

```
## Simple feature collection with 100 features and 6 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -84.32385 ymin: 33.88199 xmax: -75.45698 ymax: 36.58965
## epsg (SRID):    4267
## proj4string:    +proj=longlat +datum=NAD27 +no_defs
## precision:      double (default; no precision model)
## First 3 features:
##   BIR74 SID74 NWBIR74 BIR79 SID79 NWBIR79                              geom
## 1  1091     1       10  1364     0       19 MULTIPOLYGON(((-81.47275543...
## 2   487     0       10   542     3       12 MULTIPOLYGON(((-81.23989105...
## 3  3188     5      208  3616     6      260 MULTIPOLYGON(((-80.45634460...
```

Simple feature

Simple feature geometry list–colum (sfc)

Simple feature geometry (sfg)

Coordinate Reference Systems



https://www.nceas.ucsb.edu/sites/default/files/2020-04/OverviewCoordinateReferenceSystems.pdf

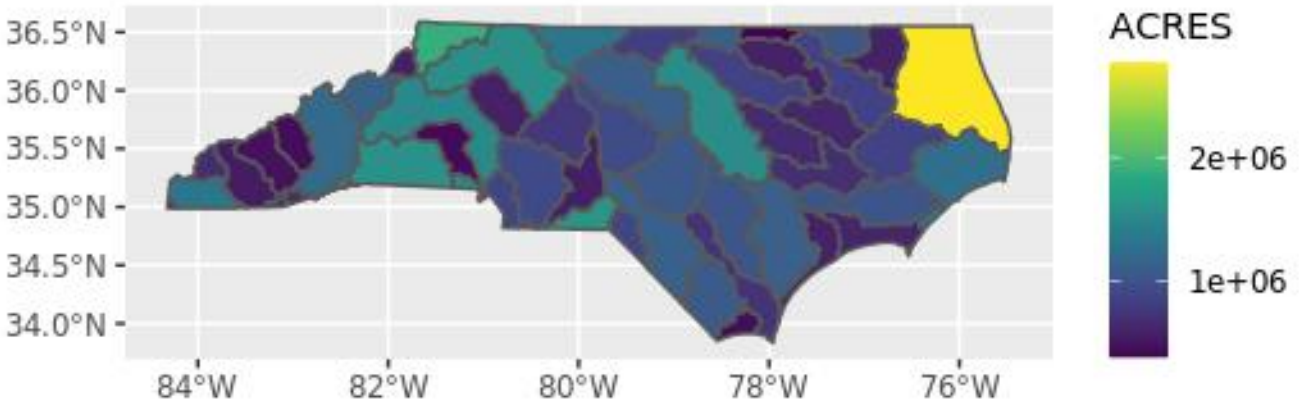
M8.2 – Spatial Analysis

- Reading spatial data into R
- Attribute joins
- Spatial aggregation
- Coordinate system transformations
- Intersecting data
- Clipping data
- Spatial Selection

Exercise 2.2.1

- Read in the NC HUC-8 shapefile & filter for : `./Data/Spatial/NCHUC8.shp`

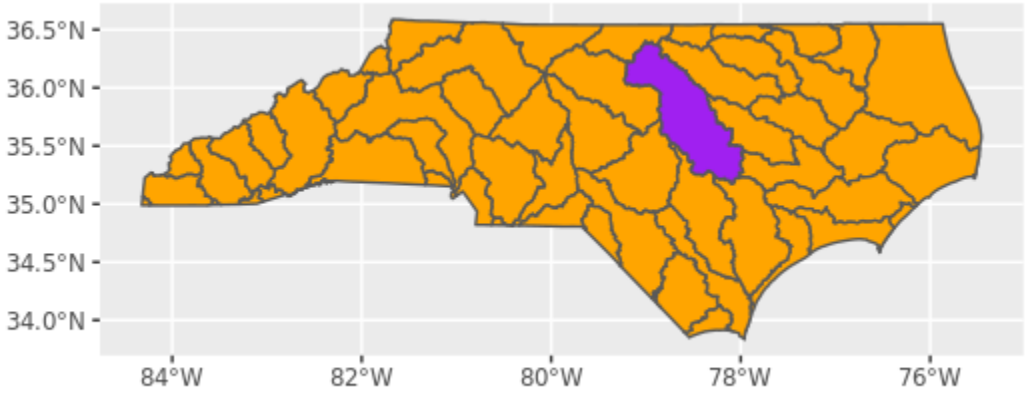
- View features, colored by ACRES



Exercise 2.2.2 - Challenge

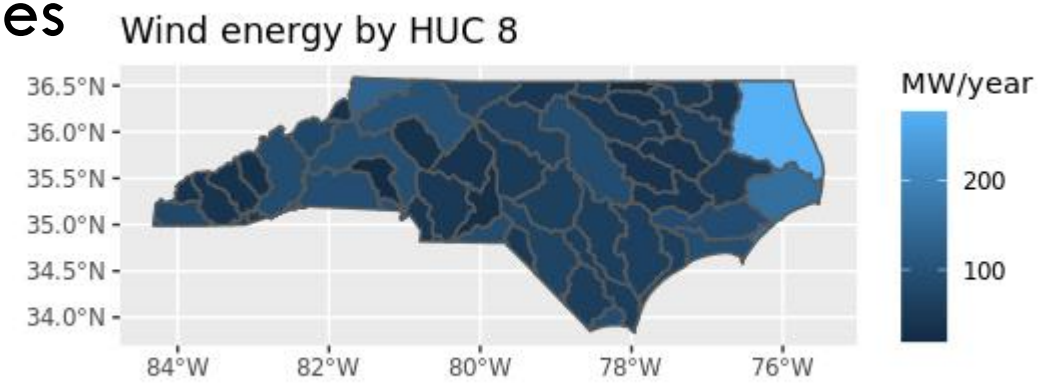
- Read in the NC HUC-8 shapefile & *filter for SUBBASIN is "Upper Neuse"*

- View all HUCs in orange, Upper Neuse in purple



Exercise 3.1.1

- Read an online CSV file into a dataframe\
  https://raw.githubusercontent.com/ENV859/EnviroAtlasData/main/Wind_Energy.csv
  - Set `HUC12` column to be a factor (*colClasses*)
  - Compute `HUC8` from `HUC12` (*substr*)
  - Group on `HUC8`
  - Compute sum of AvgWindEnergy for each HUC8
- Join to HUC8 features



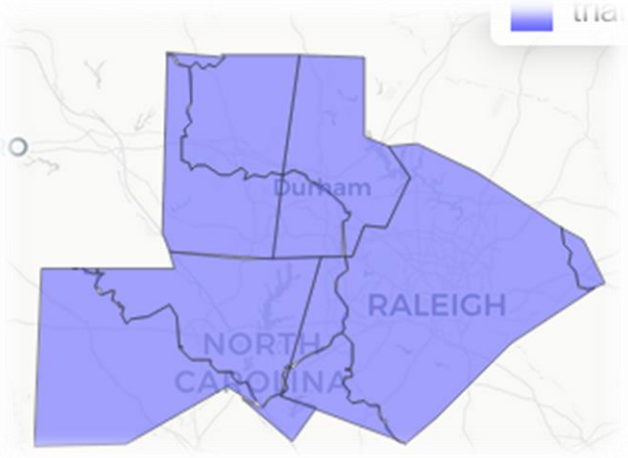
Exercise 3.2.1

- Transform all data to UTM Zone 17

```
```{r Transform the datasets to other coordinate reference systems}
#Convert all to UTM Zone 17 (crs = 26917)
epa_sf_utm <-       st_transform(epa_pm25_sites_sf, crs = 26917)
counties_sf_utm <-
state_sf_utm <-
huc8_sf_utm <-
huc2_utm <-
```
```

Exercise 3.3.1

- Select Triangle counties from all counties: *Chatham*, *Durham*, *Orange*, and *Wake*
- Select HUC8s that intersect the Triangle counties
- Intersect (clip) the HUC8 areas falling w/in Triangle



M8.3 – Spatial Data Visualization

- ggplot() + geomsf()
- mapview
- leaflet