Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Working with Geospatial Data in R

Workshop website: https://jornada-im.github.io/2024-06-24-jornada-desc-geo/
Guide to the full DC workshop: https://datacarpentry.org/geospatial-workshop/
Instructors: Greg Maurer, Renée Brown, Darren James
Helpers: Geovany Ramirez

Attendees

Name, institutional affiliation, GeoName (http://www.geonames.org) that is special to you (birthplace, where you live, favorite camping spot)
Rissa Garcia-Prudencio, NMSU, Ysleta del Sur Pueblo
Renée Brown, University of New Mexico, Albuquerque
Jesus Manuel Ochoa-Rivero, UTEP, Chihuahua, MX
Greg Maurer, New Mexico State University, Albuquerque
Kaustuv Neupane, New Mexico Water Resources Research Institute, Nepal
Dylan Stover, UTEP, Carrizo Plain National Monument
Aitor Alameda Martín, PhD student at University of Almeria. Móstoles, Madrid (Spain).
Elizabeth Brock, NMSU, Isleta Pueblo, Graduate Student
Enrique D Garcia, UTEP, Juarez
Lalmani Wagle, New Mexico State University, Nepal
Matthew Becker, New Mexico State University, Organ Mountains Desert Peaks National Monument
Emily Myers, USDA, Raleigh
Sander Denham, USDA, Asheville, NC
Michael Kusibu Mba, PhD Student, University of Nevada, Las Vegas
Mary Salibi, Undergraduate at University of Nevada, Las Vegas, NMSU REU
Lydia Mauroni, Dickinson College/UTEP, Pittsburgh
Maximiliano Dominguez, UTEP, El Paso
Darren James, NMSU/Jornada
Valentina Sánchez Castaño, PhD Student, University of Texas at El Paso.

Bryan Yu, UTEP, El Paso, Cloudcroft

Introduction to R and RStudio

https://datacarpentry.org/r-intro-geospatial/01-rstudio-intro.html

Rstudio is an IDE

Project Management With RStudio

https://datacarpentry.org/r-intro-geospatial/02-project-intro.html

Open and Plot Vector Layers

https://datacarpentry.org/r-raster-vector-geospatial/06-vector-open-shapefile-in-r.html

Challenge 1

Using the steps above, import the HARV_roads and HARVtower_UTM18N vector layers into R. Call the HARV_roads object lines_HARV and the HARVtower_UTM18N point_HARV.

Answer the following questions:

1. What type of R spatial object is created when you import each layer?

Point_harv (POINT) /Lines_Harv (MULTILINESTRING)
lines_HARV(multistring) and point
Multilinestring and points
multiline string and point
multlinestring and point
Point/Multi Line String
multiline string, point
multiline string and point
lines_HARV is multilinestring, point_HARV is point
point_HARV(point)/lines_HARV(multilinestring)

1. What is the CRS and extent for each object?

Point_harv WGS 84 / UTM zone 18N ~ Lines_Harv WGS 84 / UTM zone 18N
lines_HARV: WGS 84 / UTM zone 18N; Points: WGS 84 / UTM zone 18N
WGS 84 / UTM zone 18N
WGS84/UTM Zone 18N
WGS 84 / UTM zone 18N
WFS 84/UTM zone 18N
point_HARV(UMT 18N, WGS84)/lines_HARV(UTM 18N, WGS84)
WGS 84 / UTM zone 18N; WGS 84 / UTM zone 18N
point_HARV and lines_HARV: WGS 84 / UTM zone 18N

1. Do the files contain points, lines, or polygons?

Points and lines
lines and points
Points and Lines
points and lines
points and lines
Lines and points
point and lines

1. How many spatial objects are in each file?

Lines: 13 and Points: 1
lines_HARV: 13, point_HARV: 1
Lines: 13 point: 1
lines=13, point=1
lines: 13, point: 1
lines: 13, point: 1
point=1, lines=13
lines_HARV is 13, point_HARV is 1

Explore and Plot by Vector Layer Attributes

https://datacarpentry.org/r-raster-vector-geospatial/07-vector-shapefile-attributes-in-r.html

Challenge 2

Subset out all boardwalk from the lines layer and plot it.

```
boardwalk_HARV <- lines_HARV %>%
  filter(TYPE == "boardwalk")

ggplot() +
  geom_sf(data = boardwalk_HARV, color = "blue", size = 1.5) +
  ggtitle("NEON HARV Forest Field Site", subtitle = "Boardwalks") +
  coord_sf()

boardwalk_HARV <- lines_HARV %>%
  filter(TYPE== "boardwalk")

ggplot()+
  geom_sf(data= boardwalk_HARV, aes(color= factor(OBJECTID)), size = 1.5) +
  labs(color = "Boardwalk ID") +
  ggtitle("NEON HARV Forest Field Site", subtitle = 'boardwalk') +
  coord_sf()

## Subset boardwalks
boardwalk_HARV <- lines_HARV %>%
  filter(TYPE == "boardwalk")

# Plot the lines dataset
```

```
ggplot() +
  geom_sf(data = boardwalk_HARV) +
  ggtitle("NEON HARV Forest Field Site", subtitle = "Boardwalks") +
  coord_sf()

#Use Symbology (colors)
ggplot() +
  geom_sf(data = boardwalk_HARV, aes(color = factor(OBJECTID)), size = 1.5) +
  labs(color = "Boardwalk ID") +
  ggtitle("NEON HARV Forest Field Site", subtitle = "Boardwalks") +
  coord_sf()


 ## Challenge: lotting Boardwalks
  boardwalks_HARV <- lines_HARV %>% filter(TYPE == "boardwalk")

    ## Plots
    ggplot() +
      geom_sf(data = boardwalks_HARV) +
      ggtitle("NEON HARV Forest Field Site", subtitle = "Boardwalks") +
      coord_sf()

ggplot() +
  geom_sf(data = filter(lines_HARV, TYPE=="boardwalk"), size = 1.5) +
  ggtitle("NEON HARV Forest Field Site", subtitle = 'Boardwalks') +
  coord_sf()
```

Challenge 3

Create a plot that emphasizes only roads where bicycles and horses are allowed. To emphasize this, make the lines where bicycles are not allowed A DIFFERENT COLOR than the roads where bicycles are allowed. NOTE: this attribute information is located in the lines_HARV$BicyclesHo attribute.

Be sure to add a title and legend to your map. You might consider a color palette that has all bike/horse-friendly roads displayed in a bright color. All other lines can be black.

```
ggplot()+
  geom_sf(data=lines_HARV,aes(color=BicyclesHo))+scale_color_manual(values=c("green",
"black","black"))+
  theme(legend.text=element_text(size=18),
        legend.box.background=element_rect(linewidth=1))+
  ggtitle("NEON HARV Field Site", subtitle="Bike and Horse Access")

lines_HARV$BicyclesHo

new_colors <- c("green","cyan","yellow")

ggplot() +
  geom_sf(data = lines_HARV, aes(color = BicyclesHo)) +
```

```
  scale_color_manual(values = new_colors) +
  labs(color = 'Allowance') +
  ggtitle("NEON HARV Field Site", subtitle = "Bicycles and Horse Access")

bicycle_colors <- c("green", "red")

lines_HARV_edit <- lines_HARV %>%
  mutate(bikehorse_factor = factor(BIKEHORSE, levels = c("Y","N")))

ggplot() +
  geom_sf(data = lines_HARV_edit, aes(color = bikehorse_factor)) +
  scale_color_manual(values = bicycle_colors, na.value='black') +
  labs(color = 'Bicycles and Horses Allowed?') +
  ggtitle("NEON Harvard Forest Field Site", subtitle = "Bicycle and Horse Access") +
  coord_sf()
```

Plot Multiple Vector Layers

https://datacarpentry.org/r-raster-vector-geospatial/08-vector-plot-shapefiles-custom-legend.html

Challenge 4

Using the NEON-DS-Site-Layout-Files/HARV/PlotLocations_HARV.shp ESRI shapefile, create a map of study plot locations, with each point colored by the soil type (soilTypeOr). How many different soil types are there at this particular field site? Overlay this layer on top of the lines_HARV layer (the roads). Create a custom legend that applies line symbols to lines and point symbols to the points.

```
  # Challenge 4
PlotLocations_HARV <- st_read("data/NEON-DS-Site-Layout-Files/HARV/PlotLocations_HARV.shp")
#Need to color each point by soil type
PlotLocations_HARV$soilTypeOr
# Plot attempt
ggplot() +
  geom_sf(data = PlotLocations_HARV, aes(color = soilTypeOr), size = 2)+
  ggtitle("NEON HARV Field Soil Types")
#new colors
New_colors = c("blue","red","green","yellow", "orange", "navy")
# Plot over original map
ggplot() +
  geom_sf(data = PlotLocations_HARV, aes(color = soilTypeOr), size = 2)+
  geom_sf(data = lines_HARV, aes(color = TYPE), size = 2) +
  scale_color_manual(values = New_colors, name = "Line Type") +
  ggtitle("NEON HARV Field Soil Types")

 ggplot() +
     geom_sf(data = lines_HARV, aes(color = TYPE), size = 1, show.legend = "line") +
```

geom_sf(data = plot_locations, aes(fill = soilTypeOr), show.legend = "point") +
scale_color_manual(values = road_colors, "Path Type")
scale_fill_manual(values = soil_col, "Soil Type") +
coord_sf()

Intro to Raster Data

https://datacarpentry.org/r-raster-vector-geospatial/01-raster-structure.html

Coordinate Reference Systems (CRS) in R: It is discouraged to use 'PROJ4 strings' any longer to represent a CRS; several string elements for CRSs are now deprecated or unsupported. Use EPSG codes instead. For more information see this link: https://inbo.github.io/tutorials/tutorials/spatial_crs_coding/

In this workshop our method for plotting rasters is to first create a data.frame from the raster using terra::as.data.frame() and then plot the data.frame with ggplot() +geom_raster().
However, it's possible to use ggplot() to plot "terra" rasters directly with the geom_spatraster() function in the tidyterra package: https://dieghernan.github.io/tidyterra/reference/geom_spatraster.html

Plot Raster Data

https://datacarpentry.org/r-raster-vector-geospatial/02-raster-plot.html

Challenge: Create DSM for SJER (and DTM if you feel like it)

Use the files in the data/NEON-DS-Airborne-Remote-Sensing/SJER/ directory to create a Digital Surface Model map of the San Joaquin Experimental Range field site.

Make sure to:

- include hillshade in the maps,
- optional: label axes on the DSM map and exclude them from the DTM map,
- include a title for each map,
- experiment with various alpha values and color palettes to represent the data.

DSM:
```
# import DSM data
DSM_SJER <-
  rast("data/NEON-DS-Airborne-Remote-Sensing/SJER/DSM/SJER_dsmCrop.tif")

# convert to data frame for ggplot
DSM_SJER_df <- as.data.frame(DSM_SJER, xy=TRUE)

DSM_hill_SJER <-
  rast("data/NEON-DS-Airborne-Remote-Sensing/SJER/DSM/SJER_dsmHill.tif")

DSM_hill_SJER_df <- as.data.frame(DSM_hill_SJER, xy=TRUE)
ggplot() +
```

```
geom_raster(data = DSM_SJER_df, aes(x = x, y = y, fill = SJER_dsmCrop)) +
geom_raster(data = DSM_hill_SJER_df, aes(x = x, y = y, alpha = SJER_dsmHill)) +
scale_fill_viridis_c() +
scale_alpha(range = c(0.15, 0.65), guide = "none") +
theme(axis.title.x=element_blank(),
      axis.text.x=element_blank(),
      axis.ticks.x=element_blank()) +
theme(axis.title.y=element_blank(),
      axis.text.y=element_blank(),
      axis.ticks.y=element_blank()) +
ggtitle("Elevation with Hillshade") +
coord_quickmap()
```

Reproject Raster Data

https://datacarpentry.org/r-raster-vector-geospatial/03-raster-reproject-in-r.html

Raster Calculations

https://datacarpentry.org/r-raster-vector-geospatial/04-raster-calculations-in-r.html

It's often a good idea to explore the range of values in a raster dataset just like we might explore a dataset that we collected in the field.

1. What is the min and maximum value for the Harvard Forest Canopy Height Model (CHM_HARV) that we just created?

min = 0, max = 38.16998

1. What are two ways you can check this range of data for CHM_HARV?

minmax() and summary()
min() and max()
range()

1. What is the distribution of all the pixel values in the CHM?

1. Plot a histogram with 6 bins instead of the default and change the color of the histogram.
2. Plot the CHM_HARV raster using breaks that make sense for the data. Include an appropriate color palette for the data, plot title and no axes ticks / labels.