

# Assignment 9: Spatial Analysis in R

Emma Kaufman

## OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics (ENV872L) on spatial analysis.

## Directions

1. Rename this file `<FirstLast>_A09_SpatialAnalysis.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change “Student Name” on line 3 (above) with your name.
3. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
4. Work through the steps, **creating code and output** that fulfill each instruction.
5. Be sure to **answer the questions** in this assignment document. Space for your answers is provided in this document and is indicated by the “>” character. If you need a second paragraph be sure to start the first line with “>”. You should notice that the answer is highlighted in green by RStudio.
6. When you have completed the assignment, **Knit** the text and code into a single HTML file.

## DATA WRANGLING

### Set up your session

1. Import libraries: tidyverse, sf, leaflet, here, and mapview
2. Execute the `here()` command to display the current project directory

```
#1.  
#Import the tidyverse library  
library(tidyverse, quietly = TRUE)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.3      v readr      2.1.4  
## v forcats    1.0.0      v stringr   1.5.0  
## v ggplot2    3.4.3      v tibble    3.2.1  
## v lubridate  1.9.2      v tidyr     1.3.0  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate); library(here)
```

```
## here() starts at /home/guest/EDE_Fall2023
```

```
#install.packages('sf')
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.1, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
#install.packages('leaflet')
library(leaflet)
#install.packages('mapview')
library(mapview); mapviewOptions(fgb = FALSE)
```

```
## The legacy packages mapproj, rgdal, and rgeos, underpinning the sp package,
## which was just loaded, will retire in October 2023.
## Please refer to R-spatial evolution reports for details, especially
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
## The sp package is now running under evolution status 2
##     (status 2 uses the sf package in place of rgdal)
```

```
library(ggplot2)
library(ggdist)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
#Disable on-the-fly projections
sf::sf_use_s2(FALSE)
```

```
## Spherical geometry (s2) switched off
```

```
#2.
#display current project directory
here()
```

```
## [1] "/home/guest/EDE_Fall2023"
```

### Read (and filter) county features into an sf dataframe and plot

In this exercise, we will be exploring stream gage height data in Nebraska corresponding to floods occurring there in 2019. First, we will import from the US Counties shapefile we've used in lab lessons, filtering it this time for just Nebraska counties. Nebraska's state FIPS code is 31 (as North Carolina's was 37).

3. Read the `cb_2018_us_county_20m.shp` shapefile into an sf dataframe, filtering records for Nebraska counties (State FIPS = 31)

4. Reveal the dataset's coordinate reference system
5. Plot the records as a map (using mapview or ggplot)

```
#3. Read in Counties shapefile into an sf dataframe, filtering for just NE counties
countiesNE_sf_<- st_read(here('Data/Spatial/cb_2018_us_county_20m.shp')) %>%
  filter(STATEFP == 31) #Filter for just NE Counties
```

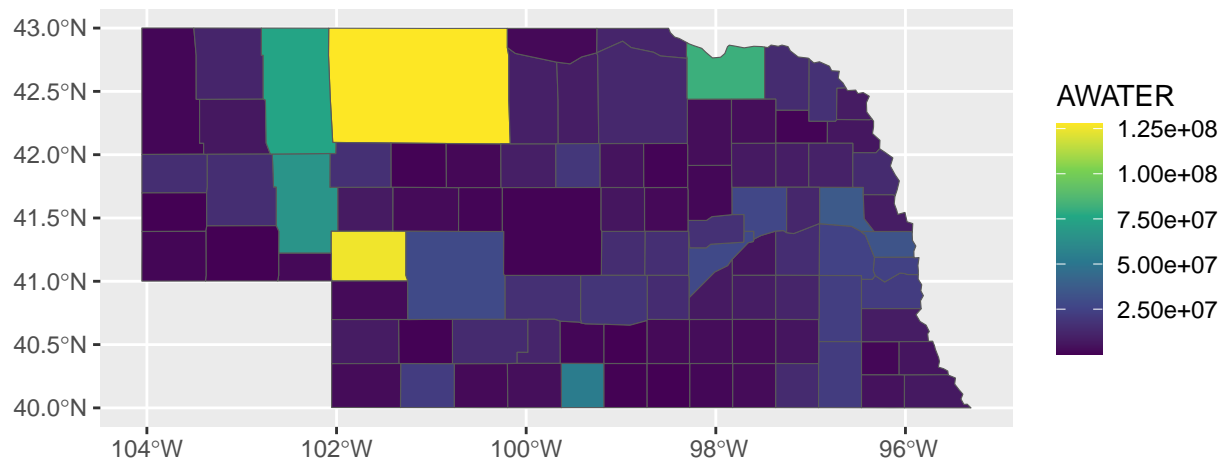
```
## Reading layer 'cb_2018_us_county_20m' from data source
##   '/home/guest/EDE_Fall2023/Data/Spatial/cb_2018_us_county_20m.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 3220 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## Geodetic CRS:   NAD83
```

```
#4. Reveal the CRS of the counties features
st_crs(countiesNE_sf_)
```

```
## Coordinate Reference System:
##   User input: NAD83
##   wkt:
##   GEOGCRS["NAD83",
##     DATUM["North American Datum 1983",
##       ELLIPSOID["GRS 1980",6378137,298.257222101,
##         LENGTHUNIT["metre",1]],
##     PRIMEM["Greenwich",0,
##       ANGLEUNIT["degree",0.0174532925199433]],
##     CS[ellipsoidal,2],
##       AXIS["latitude",north,
##         ORDER[1],
##         ANGLEUNIT["degree",0.0174532925199433]],
##       AXIS["longitude",east,
##         ORDER[2],
##         ANGLEUNIT["degree",0.0174532925199433]],
##     ID["EPSG",4269]]
```

```
#5. Plot the data
#mapView(countiesNE_sf_) #plotting in mapView

#plotting in ggplot with colors based on area of water in each county
ggplot(data=countiesNE_sf_) +
  geom_sf(aes(fill= AWATER)) +
  scale_fill_viridis_c()
```



6. What is the EPSG code of the Counties dataset? Is this a geographic or a projected coordinate reference system? (Or, does this CRS use angular or planar coordinate units?) To what datum is this CRS associated? (Tip: look for the EPSG code on <https://spatialreference.org> and examine the link for “Well Known Text as HTML” or “Human-Readable OGC WKT”...)

ANSWER: 4269. this is a geographic coordinate reference system, NAD83, that uses angular coordinate units. It is associated with the North American Datum 1983.

Read in gage locations csv as a dataframe, then display the column names it contains

Next we'll read in some USGS/NWIS gage location data added to the `Data/Raw` folder. These are in the `NWIS_SiteInfo_NE_RAW.csv` file. (See `NWIS_SiteInfo_NE_RAW.README.txt` for more info on this dataset.)

- Read the `NWIS_SiteInfo_NE_RAW.csv` file into a standard dataframe, being sure to set the `site_no` field as well as other character columns as a factor.
- Display the column names of this dataset.

```
#7. Read in gage locations csv as a dataframe
NEGAGE <- read.csv(here("Data/Raw/NWIS_SiteInfo_NE_RAW.csv"), stringsAsFactors = TRUE, colClasses = c(''))

#8. Reveal the names of the columns
colnames(NEGAGE)
```

```
## [1] "site_no"          "station_nm"        "site_tp_cd"
## [4] "dec_lat_va"       "dec_long_va"       "coord_acy_cd"
## [7] "dec_coord_datum_cd"
```

9. What columns in the dataset contain the x and y coordinate values, respectively?  
 > ANSWER: The x values are in dec\_long\_va, the y values are in dec\_lat\_va. >

### Convert the dataframe to a spatial features (“sf”) dataframe

10. Convert the dataframe to an sf dataframe.

- Note: These data use the same coordinate reference system as the counties dataset

11. Display the column names of the resulting sf dataframe

```
#10. Convert to an sf object
NEgage_sf <- NEGAGE %>%
  st_as_sf(coords = c('dec_long_va', 'dec_lat_va'), crs=4269)

#11. Re-examine the column names
colnames(NEgage_sf)
```

```
## [1] "site_no"          "station_nm"        "site_tp_cd"
## [4] "coord_acy_cd"     "dec_coord_datum_cd" "geometry"
```

12. What new field(s) appear in the sf dataframe created? What field(s), if any, disappeared?

ANSWER: The latitude and longitude columns disappeared, and a geometry field appeared.

### Plot the gage locations on top of the counties

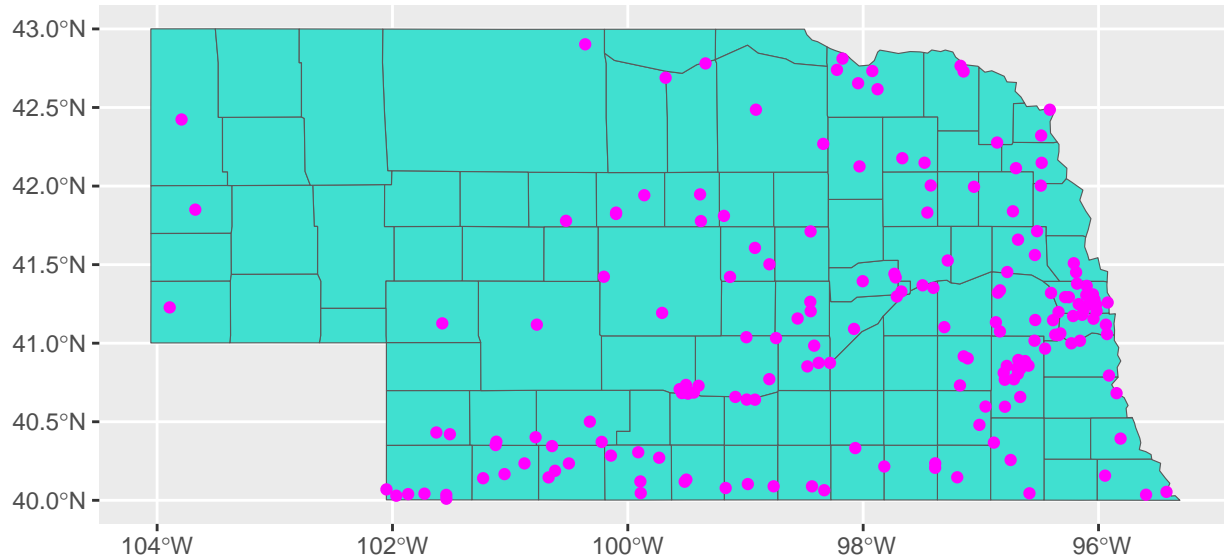
13. Use ggplot to plot the county and gage location datasets.

- Be sure the datasets are displayed in different colors
- Title your plot “NWIS Gage Locations in Nebraska”
- Subtitle your plot with your name

```
#13. Plot the gage locations atop the county features
ggplot(data=countiesNE_sf_) +
  geom_sf(fill= 'turquoise') +
  geom_sf(data= NEgage_sf$geometry, color= 'magenta')+
  labs(title='NWIS Gage Locations in Nebraska',
       subtitle = 'Emma Kaufman')
```

## NWIS Gage Locations in Nebraska

Emma Kaufman



**Read in the gage height data and join the site location data to it.**

Lastly, we want to attach some gage height data to our site locations. I've constructed a csv file listing many of the Nebraska gage sites, by station name and site number along with stream gage heights (in meters) recorded during the recent flood event. This file is titled `NWIS_SiteFlowData_NE_RAW.csv` and is found in the `Data/Raw` folder.

14. Read the `NWIS_SiteFlowData_NE_RAW.csv` dataset in as a dataframe

- Pay attention to which fields should be imported as factors!

15. Show the column names .

16. Join our site information (already imported above) to these gage height data

- The `site_no` and `station_nm` can both/either serve as joining attributes
- Construct this join so that the result only includes spatial features where both tables have data

17. Show the column names in this resulting spatial features object

18. Show the dimensions of the resulting joined dataframe

```
#14. Read the site flow data into a data frame
```

```
NEflow <- read.csv(here("Data/Raw/NWIS_SiteFlowData_NE_RAW.csv"), stringsAsFactors = TRUE, colClasses =
```

```
#Set dates to date format
NEflow$date <- ymd_hms(NEflow$date)
```

```
#15. Show the column names
colnames(NEflow)
```

```
## [1] "site_no"      "station_nm" "date"        "gage_ht"
```

```
#16. Join the flow data to our NWIS gage location spatial dataframe
NE_gage_flow <- NEgage_sf %>%
  inner_join(NEflow, by = "station_nm")
```

```
#17. Show the column names of the joined dataset
colnames(NE_gage_flow)
```

```
## [1] "site_no.x"      "station_nm"      "site_tp_cd"
## [4] "coord_acy_cd"   "dec_coord_datum_cd" "site_no.y"
## [7] "date"           "gage_ht"         "geometry"
```

```
#18. Show the dimensions of this joined dataset
dim(NE_gage_flow)
```

```
## [1] 136    9
```

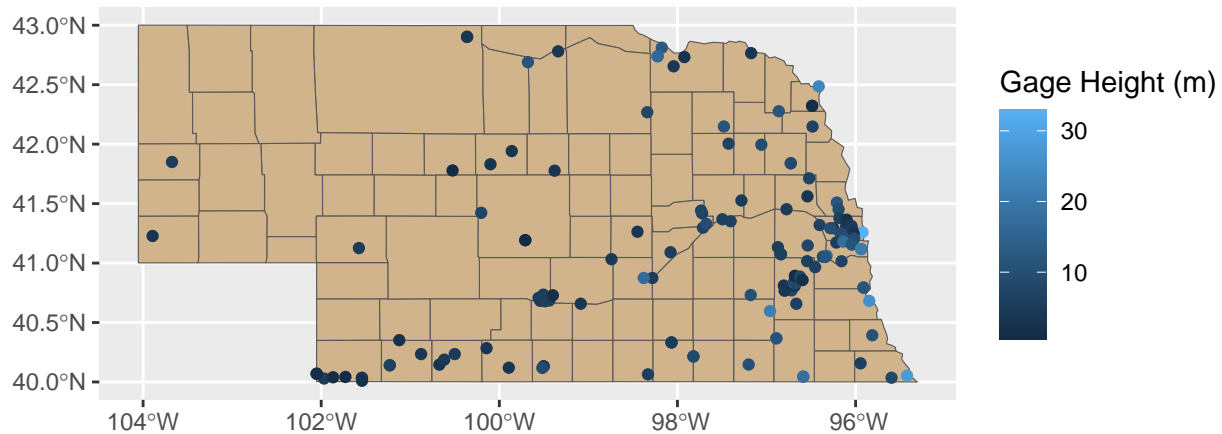
## Map the pattern of gage height data

Now we can examine where the flooding appears most acute by visualizing gage heights spatially. 19. Plot the gage sites on top of counties (using `mapview`, `ggplot`, or `leaflet`) \* Show the magnitude of gage height by color, shape, other visualization technique.

```
#19. Map the points, sized by gage height
ggplot(data=countiesNE_sf_) +
  geom_sf(fill= 'tan') +
  geom_sf(data= NE_gage_flow, aes(color= gage_ht))+
  labs(title='NWIS Gage Locations in Nebraska mapped by Gage Height',
        subtitle = 'Emma Kaufman',
        color = "Gage Height (m)")
```

## NWIS Gage Locations in Nebraska mapped by Gage Height

Emma Kaufman



---

## SPATIAL ANALYSIS

Up next we will do some spatial analysis with our data. To prepare for this, we should transform our data into a projected coordinate system. We'll choose UTM Zone 14N (EPSG = 32614).

### Transform the counties and gage site datasets to UTM Zone 14N

20. Transform the counties and gage sf datasets to UTM Zone 14N (EPSG = 32614).
21. Using mapview or ggplot, plot the data so that each can be seen as different colors

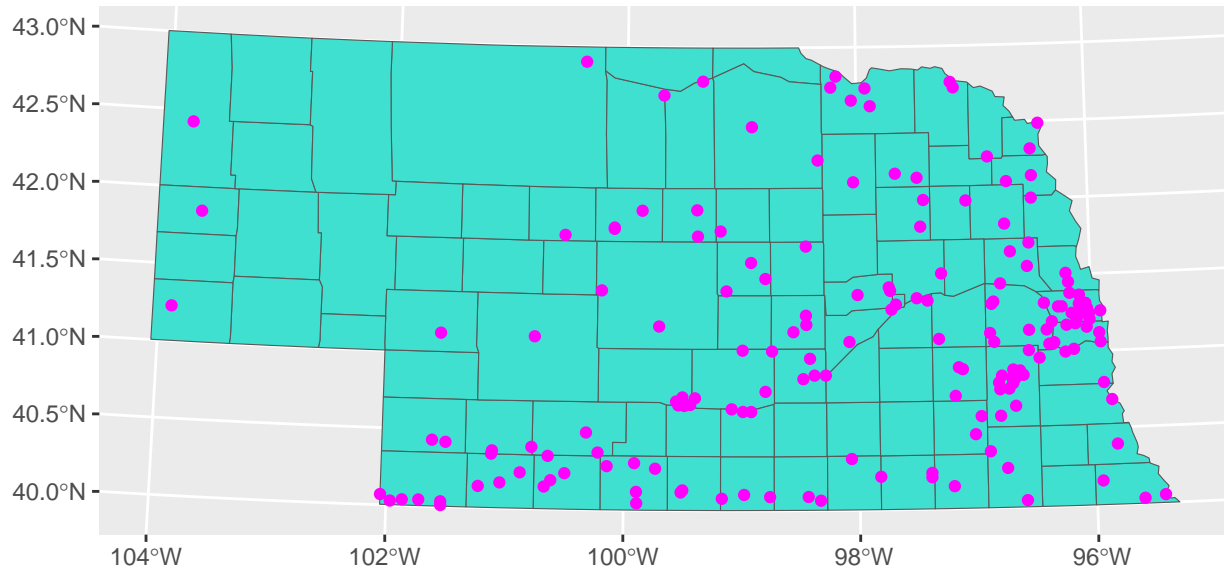
```
#20 Transform the counties and gage location datasets to UTM Zone 14
counties_UTM <- st_transform(countiesNE_sf_, c=32614)
gage_UTM <- st_transform(NEgage_sf, c= 32614)

#21 Plot the data
ggplot(data=counties_UTM ) +
  geom_sf(fill= 'turquoise') +
  geom_sf(data= gage_UTM$geometry, color= 'magenta')+
  labs(title='NWIS Gage Locations in Nebraska (UTM)',
        subtitle = 'Emma Kaufman')
```



## NWIS Gage Locations in Nebraska (UTM)

Emma Kaufman



### Select the gages falling within a given county

Now let's zoom into a particular county and examine the gages located there. 22. Select Dawson county from your projected county sf dataframe 23. Select the gage sites falling **within** that county \* Use either matrix subsetting or tidy filtering 24. Create a plot showing: \* all Nebraska counties, \* the selected county, \* and the gage sites in that county

```
#22 Select the county
Dawson <- counties_UTM %>%
  filter(NAME == "Dawson")

#23 Spatially select gages within the selected county
Gage_intersect_Dawson <- gage_UTM[Dawson,]

GageDawson_intersect_2 <- gage_UTM %>%
  filter(st_intersects(x = ., y = Dawson, sparse = FALSE))
```

```
## Warning: Using one column matrices in 'filter()' was deprecated in dplyr 1.1.0.
## i Please use one dimensional logical vectors instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
#24 Plot
ggplot() +
  geom_sf(data=counties_UTM, fill= 'turquoise') +
  geom_sf(data = Dawson, fill = 'violet') +
  geom_sf(data= Gage_intersect_Dawson, color= 'darkblue')+
  labs(title='NWIS Gage Locations in Dawson County Nebraska (UTM)',
        subtitle = 'Emma Kaufman')
```

