Assignment 9: Spatial Analysis in R

Halina Malinowski

OVERVIEW

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

Directions

1. Change “Student Name” on line 3 (above) with your name.
2. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. 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.
5. When you have completed the assignment, **Knit** the text and code into a single HTML file.
6. After Knitting, please submit the completed exercise (PDF file) in Sakai. Please add your last name into the file name (e.g., “Fay_A10_SpatialAnalysis.pdf”) prior to submission.

DATA WRANGLING

Set up your session

1. Check your working directory
2. Import libraries: tidyverse, sf, leaflet, and mapview

```
#1.
getwd()
```

```
## [1] "C:/Users/Dell Laptop/Documents/GitHub/EDA/Assignments"
```

```
#2.
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(sf)
```

```
## Warning: package 'sf' was built under R version 4.1.3
```

```
## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1; sf_use_s2() is TRUE
```

```
library(leaflet)
```

```
## Warning: package 'leaflet' was built under R version 4.1.3
```

```
library(mapview)
```

```
## Warning: package 'mapview' was built under R version 4.1.3
```

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
counties_sf_NE<- st_read('/Users/Dell Laptop/Documents/GitHub/EDA/Data/Spatial/cb_2018_us_county
_20m.shp') %>%
  filter(STATEFP == 31)
```

```
## Reading layer `cb_2018_us_county_20m' from data source
##   `C:\Users\Dell Laptop\Documents\GitHub\EDA\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
```

```
class(counties_sf_NE)
```

```
## [1] "sf"         "data.frame"
```

```
names(counties_sf_NE)
```

```
##  [1] "STATEFP"  "COUNTYFP" "COUNTYNS" "AFFGEOID" "GEOID"    "NAME"
##  [7] "LSAD"     "ALAND"    "AWATER"   "geometry"
```

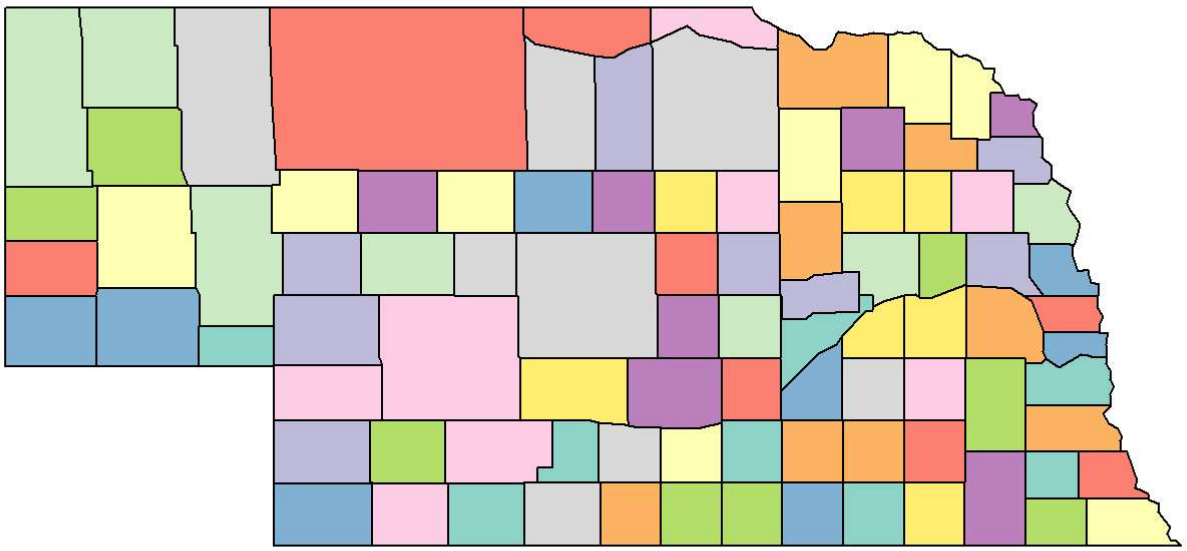
```
#4. Reveal the CRS of the counties features
st_crs(counties_sf_NE)
```

```
## 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]]
```

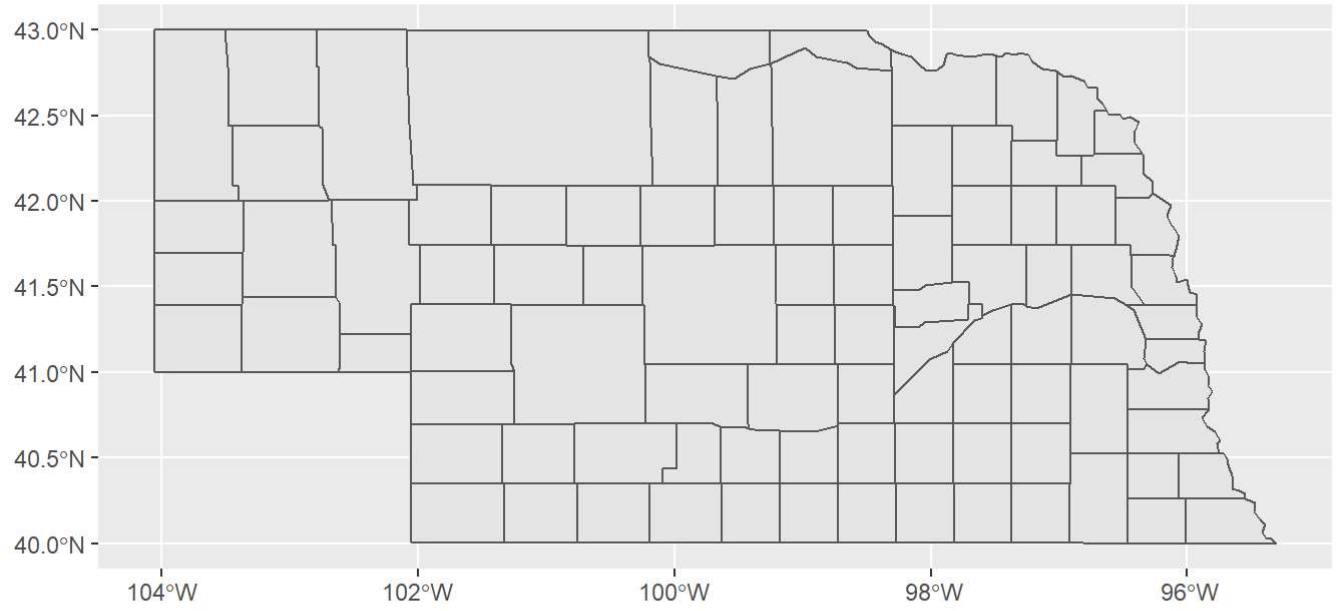
```
st_crs(counties_sf_NE)$epsg
```

```
## [1] 4269
```

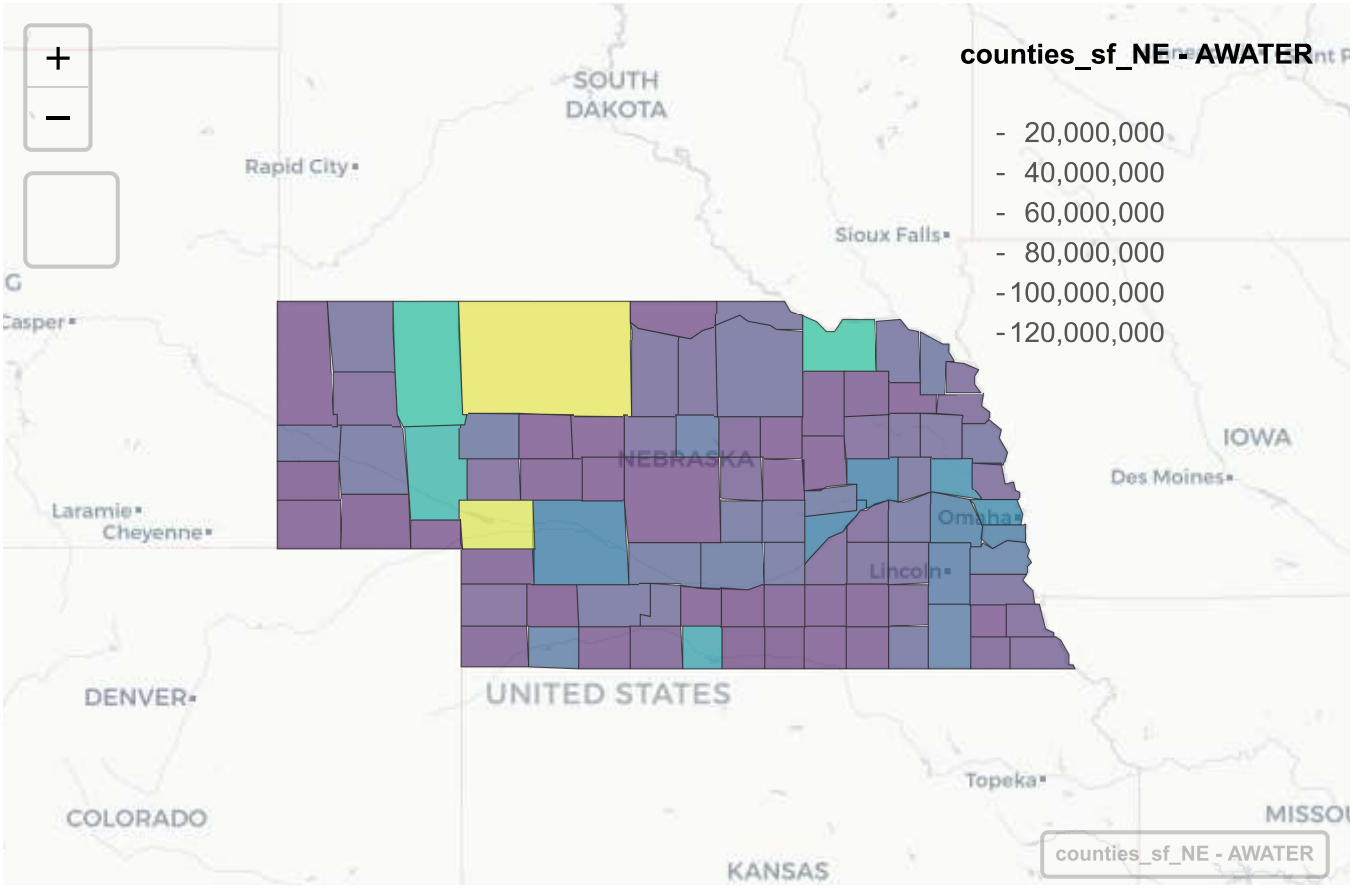
```
#5. Plot the data
plot(counties_sf_NE['COUNTYFP'],pch=16)
```




```
ggplot(data=counties_sf_NE) + geom_sf()
```




```
mapView(counties_sf_NE, zcol = "AWATER")
```



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 the EPSG code on https://spatialreference.org (https://spatialreference.org))

> ANSWER:

> The EPSG code for the Counties dataset is 4269. This code is for NAD83 and is a geographic reference system that uses angular coordinate units measured in degrees. The datum this CRS is associated with is the North American Datum 1983 or NAD83.

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.)

7. Read the NWIS_SiteInfo_NE_RAW.csv file into a standard dataframe.

8. Display the column names of this dataset.

```
#7. Read in gage locations csv as a dataframe
gage_locations <- read.csv('/Users/Dell Laptop/Documents/GitHub/EDA/Data/Raw/NWIS_SiteInfo_NE_RAW.csv',
                           stringsAsFactors = TRUE)
class(gage_locations)
```

```
## [1] "data.frame"
```

```
#8. Reveal the names of the columns
(names(gage_locations))
```

```
## [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:

There are two columns in the dataset that contain the x and y coordinates. The x-coordinate or the longitude is contained in “dec_long_va” column, while the y-coordinate or the latitude is contained in the “dec_lat_va” column.

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

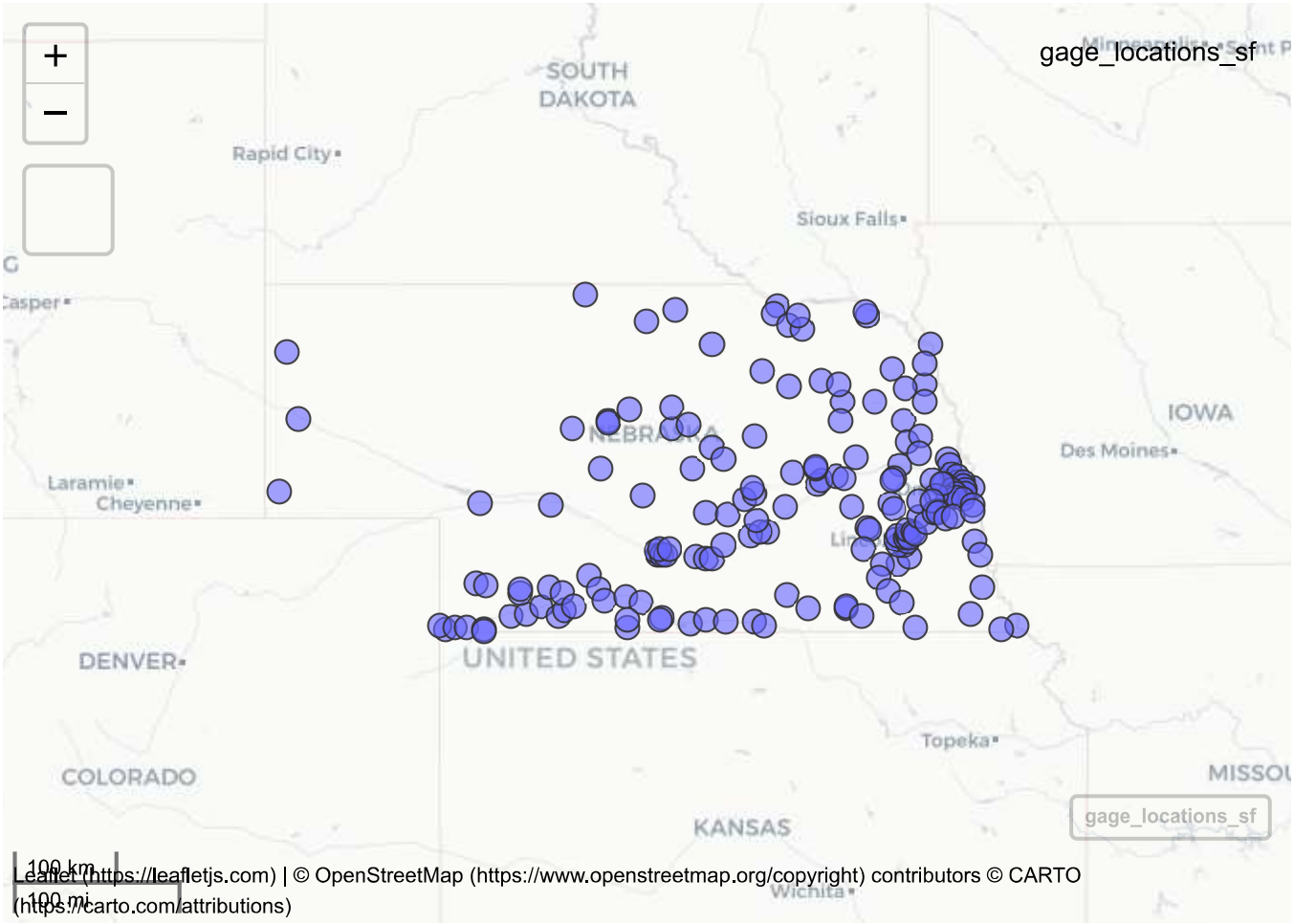
10. Convert the dataframe to an sf dataframe.
    - Note: These data use the same coordinate reference system as the counties dataset
11. Display the column names of the resulting sf dataframe

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

class(gage_locations_sf)
```

```
## [1] "sf"         "data.frame"
```

```
mapview(gage_locations_sf)
```



```
#11. Re-examine the column names
(names(gage_locations_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:

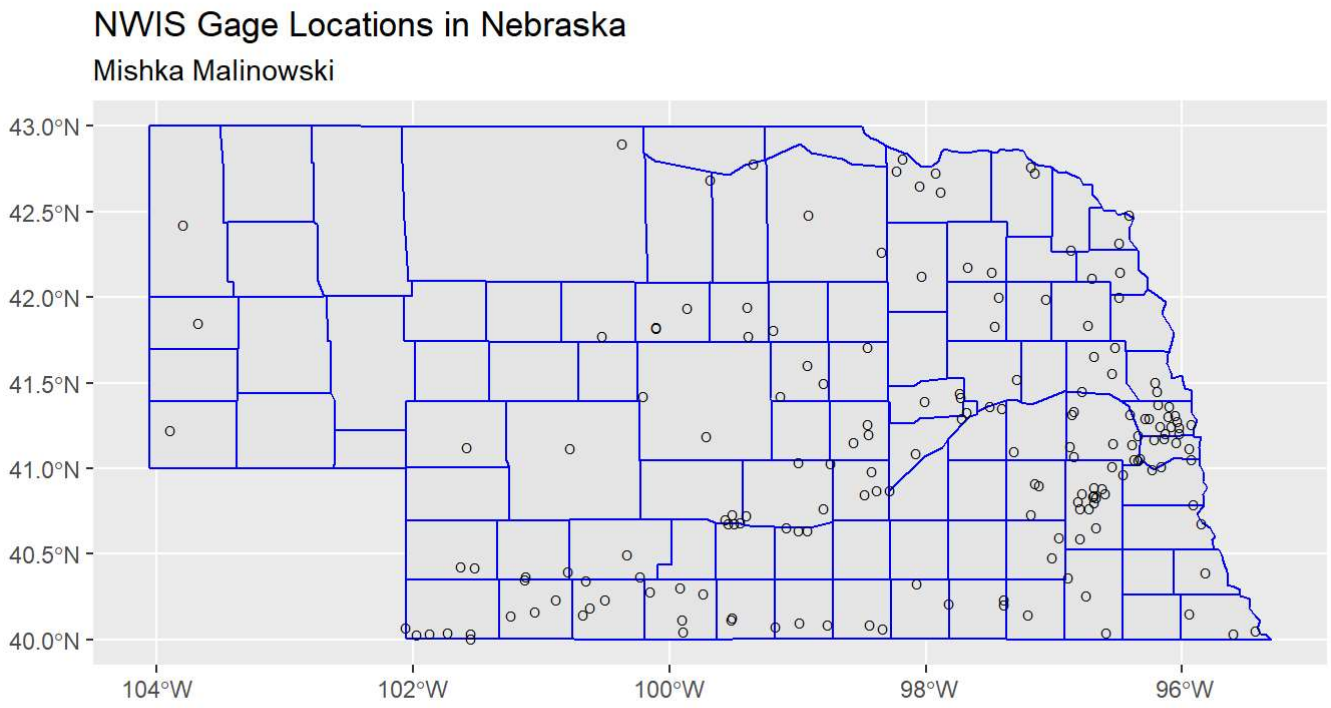
One new field appears in the sf dataframe named “geometry” which has the combined coordinates. The “dec_long_va” and “dec_lat_va” columns have disappeared.

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() +
  geom_sf(data = counties_sf_NE ,color='blue') +
  geom_sf(data = gage_locations_sf,color='black',shape='O') +
  ggtitle(label = "NWIS Gage Locations in Nebraska",
          subtitle = "Mishka Malinowski")
```



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.
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
site_flow <- read.csv('/Users/Dell Laptop/Documents/GitHub/EDA/Data/Raw/NWIS_SiteFlowData_NE_RA
W.csv',
                      stringsAsFactors = TRUE)
class(site_flow)
```

```
## [1] "data.frame"
```

```
#15. Show the column names
(names(site_flow))
```

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

```
(names(gage_locations_sf))
```

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

```
#16. Join location data to it
siteflow_gagelocationssf_join = merge(x = gage_locations_sf,
                      y = site_flow,
                      by.x = 'site_no',
                      by.y = 'site_no',
                      na.rm = T)

summary(siteflow_gagelocationssf_join)#checking for NAs
```

```
##     site_no                                                  station_nm.x
##  Min.   :6453600   Big Blue River at Barneston, Nebr.                :  2
##  1st Qu.:6610795   Big Papillion Creek at Fort Street at Omaha, Nebr.:  2
##  Median :6799050   Elkhorn River at West Point, Nebr.                :  2
##  Mean   :6746717   Little Blue River at County Line nr Deshler, Nebr.:  2
##  3rd Qu.:6813750   Little Blue River near Deweese, Nebr.             :  2
##  Max.   :6884000   Little Papillion Creek at Irvington, Nebr.        :  2
##                    (Other)                                          :124
##  site_tp_cd coord_acy_cd dec_coord_datum_cd
##  ST:136     1:21         NAD83:136
##             5: 3
##             F:12
##             H: 7
##             M: 1
##             S:92
##
##                                             station_nm.y
##  Big Blue River at Barneston, Nebr.                :  2
##  Big Papillion Creek at Fort Street at Omaha, Nebr.:  2
##  Elkhorn River at West Point, Nebr.                :  2
##  Little Blue River at County Line nr Deshler, Nebr.:  2
##  Little Blue River near Deweese, Nebr.             :  2
##  Little Papillion Creek at Irvington, Nebr.        :  2
##  (Other)                                           :124
##                   date       gage_ht                geometry
##  2019-03-20 14:45:00:43   Min.   : 0.480   POINT        :136
##  2019-03-20 14:15:00:25   1st Qu.: 3.715   epsg:4269    :  0
##  2019-03-20 14:30:00:15   Median : 5.890   +proj=long...:  0
##  2019-03-20 14:00:00:14   Mean   : 7.585
##  2019-03-20 13:30:00: 7   3rd Qu.: 9.370
##  2019-03-20 14:50:00: 7   Max.   :32.970
##  (Other)            :25
```

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

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

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

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

```
nrow(siteflow_gagelocationssf_join)
```

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

```
ncol(siteflow_gagelocationssf_join)
```

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

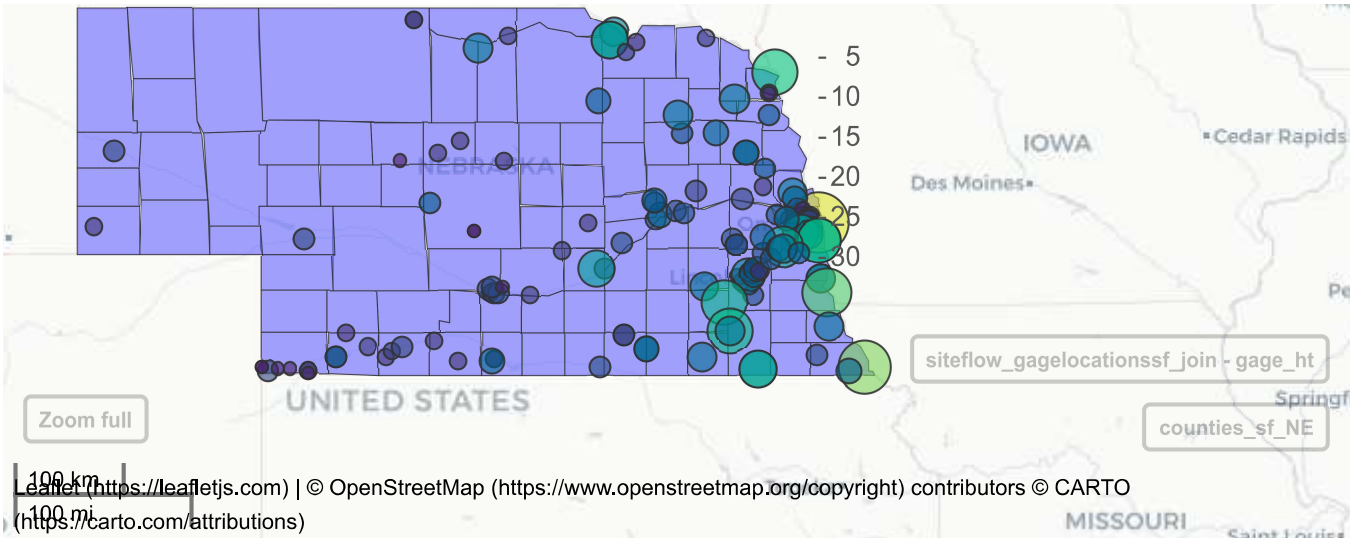
```
str(siteflow_gagelocationssf_join)
```

```
## Classes 'sf' and 'data.frame':   136 obs. of  9 variables:
##  $ site_no           : int  6453600 6461500 6461500 6463500 6463720 6465500 6465500 6465700 6
466500 6478522 ...
##  $ station_nm.x      : Factor w/ 175 levels " BIG SANDY CR AT ALEXANDRIA, NE",..: 117 87 87 7
0 86 88 88 162 8 23 ...
##  $ site_tp_cd        : Factor w/ 1 level "ST": 1 1 1 1 1 1 1 1 1 ...
##  $ coord_acy_cd      : Factor w/ 6 levels "1","5","F","H",..: 6 6 6 6 6 6 6 6 1 ...
##  $ dec_coord_datum_cd: Factor w/ 1 level "NAD83": 1 1 1 1 1 1 1 1 1 ...
##  $ station_nm.y      : Factor w/ 133 levels "Antelope Creek at 27th Street at Lincoln, Neb
r.",..: 89 65 65 44 63 66 66 123 3 15 ...
##  $ date              : Factor w/ 20 levels "2018-12-11 01:00:00",..: 13 13 13 3 19 4 4 8 10 1
6 ...
##  $ gage_ht           : num  12.58 3.63 3.61 11.38 3.75 ...
##  $ geometry          :sfc_POINT of length 136; first list element:  'XY' num  -98.2 42.8
##  - attr(*, "sf_column")= chr "geometry"
##  - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",..: NA NA NA NA NA NA NA NA
##   ..- attr(*, "names")= chr [1:8] "site_no" "station_nm.x" "site_tp_cd" "coord_acy_cd" ...
```

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.

```
#Map the points, sized by gage height
mapview(counties_sf_NE)+
  mapview(siteflow_gagelocationssf_join,
        zcol = 'gage_ht', cex = 'gage_ht')
```



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 (EPGS = 32614).

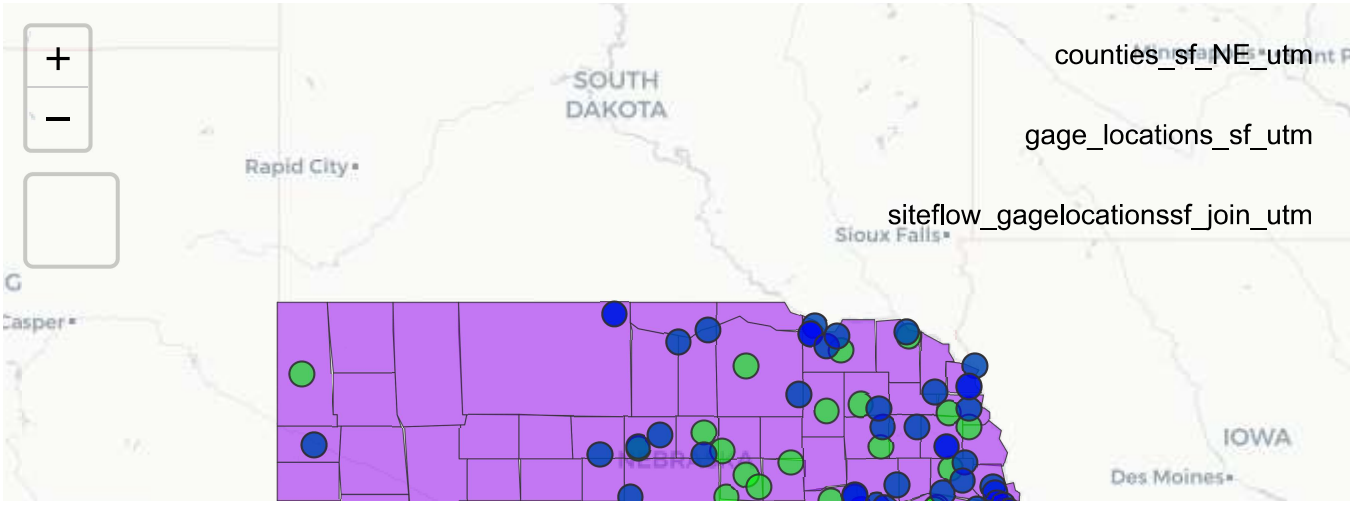
Transform the counties and gage site datasets to UTM Zone 14N

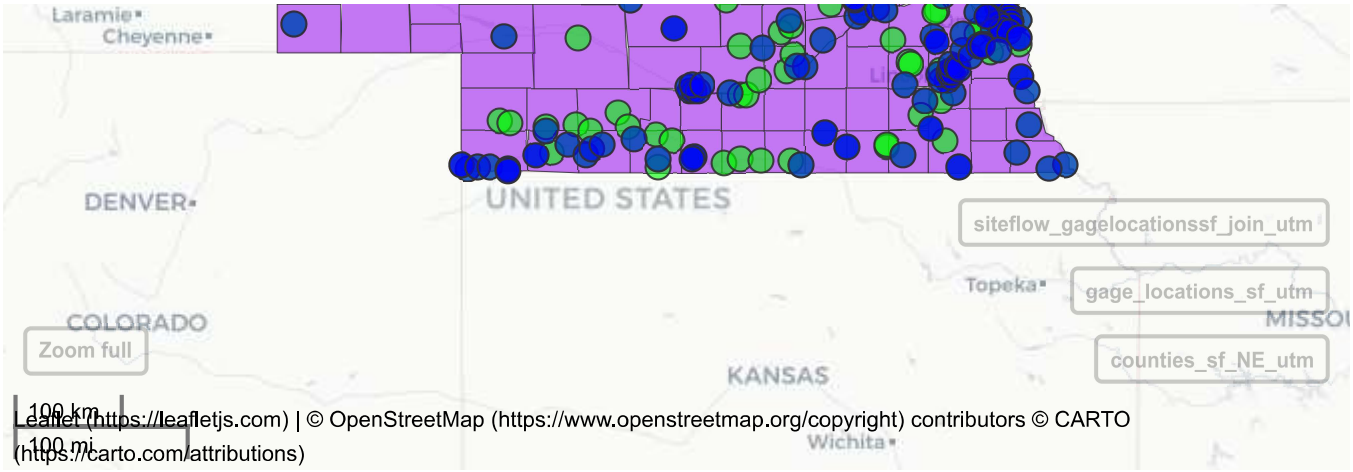
20. Transform the counties and gage sf datasets to UTM Zone 14N (EPGS = 32614).
21. Using `mapview` or `ggplot` , plot the data so that each can be seen as different colors

```
#20 Transform the counties and gage location datasets to UTM Zone 14

counties_sf_NE_utm <- st_transform(counties_sf_NE, crs = 32614)
gage_locations_sf_utm <- st_transform(gage_locations_sf, crs = 32614)
siteflow_gagelocationssf_join_utm <- st_transform(siteflow_gagelocationssf_join,
                                                  crs = 32614)

#21 Plot the data
mapview(counties_sf_NE_utm, col.regions = "purple")+
  mapview(gage_locations_sf_utm, col.regions = "green")+
  mapview(siteflow_gagelocationssf_join_utm, col.regions = "blue")
```





Select the gages falling within a given county

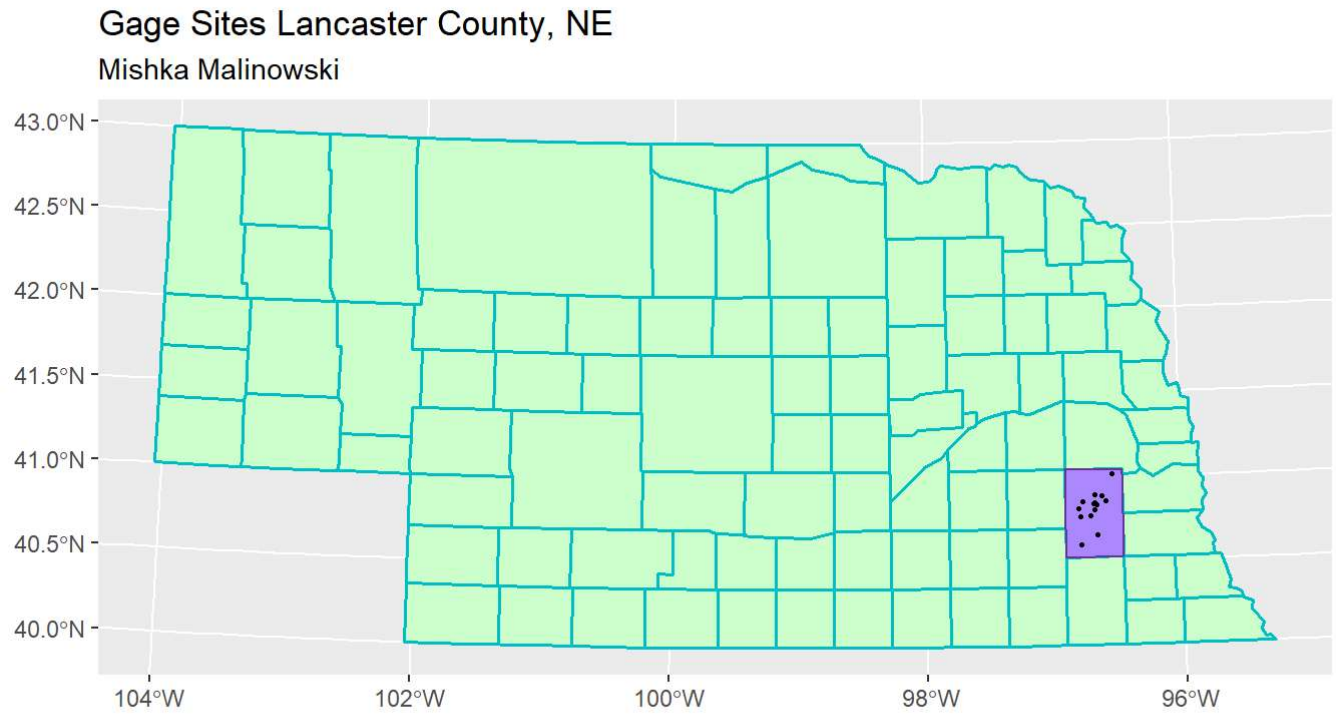
Now let's zoom into a particular county and examine the gages located there. 22. Select Lancaster county from your county sf dataframe 23. Select the gage sites falling `within` that county * Use either matrix subsetting or tidy filtering 24. Create a plot showing: * all Nebraska counties, * the selected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

```
#22 Select the county
Lancaster_county <- counties_sf_NE_utm %>%
  filter(NAME == 'Lancaster')

#23 Select gages within the selected county
Lancaster_county_gages_intersect <- gage_locations_sf_utm[Lancaster_county,]

Lancaster_county_gages_intersect2 <- gage_locations_sf_utm %>%
  filter(st_intersects(x = ., y = Lancaster_county, sparse = FALSE))

#24 Plot
ggplot() +
  geom_sf(data = counties_sf_NE_utm, color='#00BFC4', size=0.6, fill = '#CCFFCC')+
  geom_sf(data = Lancaster_county, color = '#663399', fill = '#AC88FF')+
  geom_sf(data = Lancaster_county_gages_intersect, color='black', size=0.5) +
  ggtitle( label = "Gage Sites Lancaster County, NE", subtitle = "Mishka Malinowski")
```


Gage Sites Lancaster County, NE
Mishka Malinowski
43.0°N
42.5°N
42.0°N
41.5°N
41.0°N
40.5°N
40.0°N
104°W
102°W
100°W
98°W
96°W

```
```