R Data Science: Spatial Data Science 1

steppe

2/14/2022

Contents

1 Example Data **3**
1.1 Libraries . . . 3
1.2 Example Data Methods . . . 3

2 What is a Geographic Information System? **3**
2.1 What is a GIS . . . 3
2.2 A Brief History . . . 4
2.3 What is Spatial Data Science? . . . 5
2.4 Why Use R as a GIS and for Spatial Data Science ? . . . 5

3 Geodesy **5**
3.1 Earth is not a **perfect** sphere . . . 6
3.1.1 The Earth can be represented as an Ellipsoid . . . 6
3.1.2 However, the Earth's surface is not smooth - Geoid . . . 7
3.2 Geodetic Datums . . . 7
3.2.1 Geodetic Coordinates . . . 8
3.2.2 Coordinate Notation . . . 9
3.3 Coordinate Reference System . . . 10
3.4 Problems with working on flat surfaces . . . 11
3.4.1 Geographic Coordinate System . . . 11
3.4.2 Projected Coordinate System . . . 11
3.4.3 Major Map Projections . . . 11
3.5 Geodesy Takeaways . . . 12

4 An Introduction to Geographic Data Models **12**

5 R Data Types Reviewed **13**

6 Vector Data in R **14**
6.1 Simple Features - Standards . . . 14
6.1.1 Simple Features 1 - Attributes . . . 14
6.1.2 Simple Features 2 - Coordinates form a Point . . . 14
6.1.3 Simple Features 3 - Points are/form a SF Geometry (sfg's) . . . 15
6.1.4 Simple Features 4 - A Geometry can be combined to form Geometries (sfg's) . . . 16
6.1.5 Simple Features 5 - Geometries and Spatial Information form an SF Collection (sfc) . . . 17
6.1.6 Simple Features 6 - Recapped. . . . 17
6.2 sp . . . 18
6.2.1 sp - History . . . 18
6.2.2 sp - Spatial Classes for Topology . . . 18
6.2.3 sp - Structures of The Spatial* Class - Topology Only . . . 18

7 sp - Attributes - DataFrame **18**

8 sp Overview of a Spatial*DataFrame **19**
8.0.1 sp - Recapped. . . . 20

9 Raster data in R **20**
9.1 Raster Components . . . 21
9.1.1 Example Raster 1 Create Frame . . . 21
9.1.2 Example Raster 2 Set Values . . . 23
9.1.3 Example Raster 3 . . . 24
9.2 Raster Package - in R! . . . 24
9.2.1 Attributes . . . 24
9.2.2 Raster Layer . . . 24
9.2.3 Raster Stack . . . 24
9.2.4 Raster Brick . . . 25
9.3 Terra . . . 26

10 Cartography **26**

11 Assignments for the Duration of the Spatial Data Science Module: **34**

12 Works Cited **35**
12.1 Packages Cited: . . . 35

List of Figures

1 A Visualization of a GIS, by Anne Sexton . . . 4
2 The Broad Street Pump, A Cholera Outbreak in London, By: John Snow and digitized by National Geographic . . . 4
3 Blue Marble 2012, by Suomi NPP . . . 6
4 Ellipsoids . . . 6
5 Left: Geoid Cross Section. Right: IGCM Geoid . . . 7
6 Number one City Datum. by: Cosmo1976 . . . 8
7 Control Points Number one City Datum . . . 8
8 Angles on Ellipsoid and Geodetic Coordinates. by: Peter Mercator . . . 9
9 UTM Zones of the Conterminous USA. data by ESRI . . . 10
10 Geographic Coordinate Reference System. by: Anna Krystalli . . . 11
11 Orange Peel. by: Nathan Belz . . . 11
12 Map Projections. by: Daniel Strebe . . . 12
13 Vector and Raster Data Models . . . 13
14 Western Plant Predictors. Note each cell is the extent of a single raster tile, each tile contains cells. . . . 21
15 Suprise Valley by: John Glen . . . 22

List of Tables

4 Example Matrix showing values underlaying a raster layer. . . . 23

Assigned Reading: Geocomputation with R Chapter 2.

1 Example Data

1.1 Libraries

1.2 Example Data Methods

I used the United States Geological Surveys 'Earth Explorer' to view images taken from the Sentinel 2 Satellite over the last year at a location on the border of California and Nevada near Reno. I downloaded several images which both A) covered the area of analysis well, and B) did not have much cloud cover. I used the Open Source QGIS, which has a graphical user interface (GUI), to manually geo-reference the images to locations on the earths surface. I then manually marked the edges of bodies of water, and created polygons which cover them. These data were saved as polygon vector files, in a format know as a 'shapefile' and imported them to R.

These steps can all be done programmatically, but I only wanted a few good images, so went through the process by hand.

Here we import our vector data, immediately create a 'Simple Feature' object, and add some attributes regarding the data

We import the Sentinel 2 images here. We will use them as a template to create a raster.

We will very quickly reclassify the input images to populate our new example rasters with data.

We will create empty example rasters.

2 What is a Geographic Information System?

2.1 What is a GIS

- **G**eographic **I**nformation **S**ystem is a system for producing, managing, displaying, and analyzing geographic information.
- Spatial Data Science is an emergent field which utilizes data science approaches in a **GIS**, and is a natural extension of a **GIS**

When we think about it, nearly all data has a spatial dimension, it just tends to be ignored in much of science. Historically this is sensible as performing these analysis is computationally expensive, and requires considerable expertise.

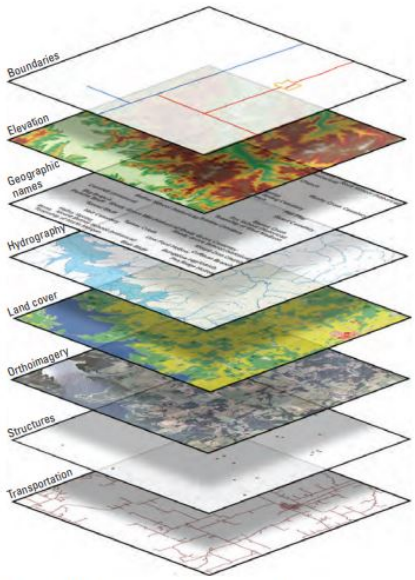


Figure 1: A Visualization of a GIS, by Anne Sexton

This is I think really one of the best and simplest way's of showcasing at heart what a GIS is. A GIS is a system wherein we can explore and study the relationships of different attributes on a process in a spatial context.

2.2 A Brief History

- Dr. John Snow suspected Cholera was spread by water
- 1854 outbreak of Cholera in the Soho neighborhood of London kills 616 persons
- Snow used both *mapping* and *statistics* to identify the water source and stop the outbreak.
- Essentially founded both Epidemiology and GIS, and stopped the outbreak in one stroke

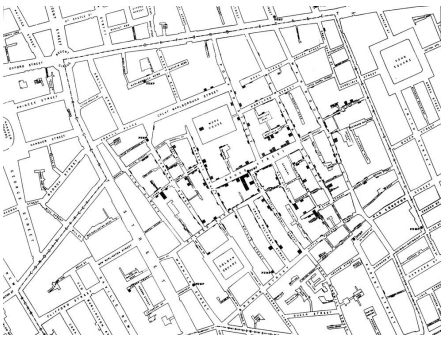


Figure 2: The Broad Street Pump, A Cholera Outbreak in London, By: John Snow and digitized by National Geographic

A good article with a quote by Tufte, anXKCD, and meaningful insight by a journalist at the Guardien is below:

https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map

- Since then GIS has continually made more use of computers and come to be it's own discipline as computers have become more available.

2.3 What is Spatial Data Science?

- The application of Geographic insights, and geospatial analyses to big data sets
- Including spatial terms in statistical models
- Bringing Geographic information to data science questions
- Developing your own spatial products and pipelines

2.4 Why Use R as a GIS and for Spatial Data Science ?

- Work Flow Automation
- Rich Ecosystem (packages, functions, code-sharing, etc.)
- Reproducible
- Self Documenting
- Computationally Efficient
- Parallel Processing/HPC interfaces

There is a great range of computer Geographic Information Systems to choose from. Many of you are likely to be familiar with a software program call 'ArcGIS' and to a lesser extent 'ArcMap' produced by the ESRI company. These products are widely used in nearly all branches of the federal and many state governments, as well as at many large companies especially environmental consultants. I think that this is a good system for you to learn about geospatial operations and workflows in, but nearly all of the work that can be accomlished in ESRI products can be accomplished in R. If you go into several environmental fields, you will need to learn to interface with ESRI products. I advise you to familiarize yourself with them.

On the other hand, I encourage you all to use R as the central piece of your geospatial analysis. If you are not developing new spatial products, you can run all of your analyses in R. If you are interested in making maps - cartography, R is now quite capable for making publication quality maps, as well as for reports and projects. It does lack slightly in cartography aspects, but you can create the data you want to style in another software, such as the open source QGIS, here and export it. But as we will see, cartography is not equivalent to GIS, and GIS is not equivalent with cartography. Do not expect to be able to make *National Georgraphic* quality maps no matter which GIS you use.

If you are very interested in geospatial work than you will see that you use a number of software programs to perform your research. In most cases you will still keep R as the centerpiece, but will likely work within a Linux environment and make good use of Unix and Bash scripts, Python, and QGIS to help fill in some parts of work-flows which R cannot address as well. You will find that R, Python, and QGIS all use many of the same software components and libraries such as GDAL/OGR, GEOS, PROJ, GRASS GIS, which have been developed by the Open Source Geospatial Foundation which was actually founded in Chicago around 20 years ago.

We will draw heavily from the Open Source Geospatial Foundation in these lectures.

3 Geodesy

- 'Geodesy is the science of accurately measuring and understanding the Earth's geometric shape, orientation in space, and gravity field.' - NOAA

This is a very highly specialized field, of which I have no formal training in when it comes to the theory of it. I do believe we have a couple specialists in Earth and Planetary Sciences, but I am far from them. Here I will cover the most basic facets of this field as they pertain to geospatial analysis among non-specialists. A more apt title for this section could be *'Geodesy taught by a dummy'*

3.1 Earth is not a perfect sphere

- Circumference at equator: 40,075 km (24,901 mi)
- Circumference along meridians: 40,009 km (24,860 mi)

Earth is not a sphere; it is ever so slightly longer than wide. The earth is actually around 43 km/27miles wider at the equator. Note the average radius of the earth is around 6371 km/3959 mi so this is quite small! However, trying to represent the Earth as a perfect sphere in geospatial models will lead to inaccurate representation of the location of objects on it.

Figure 3: Blue Marble 2012, by Suomi NPP

3.1.1 The Earth can be represented as an Ellipsoid

- A simple 3d geometric shape
- An ellipsoid is a slightly to greatly ovaliform shape
- Modeling the Earth as an ellipsoid increases the accuracy of point locations

Because the earth is slightly wider than long, the earth is technically and can be modeled practically as an ellipsoid. This model of representing the earth assumes no terms of gravity, winds, or tides. I.e. it is a perfect geometric shape with bilateral symmetry.

You can imagine that there are certain limitations to the accuracy of our model of the earth without including gravity.

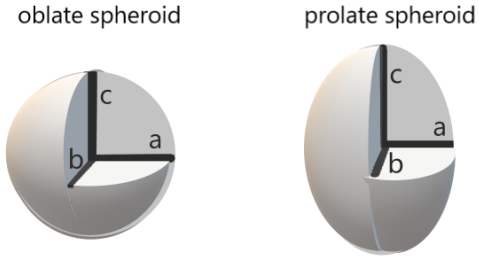


Figure 4: Ellipsoids

3.1.2 However, the Earth's surface is not smooth - Geoid

- Types of models to represent planet Earth
- Include gravity, excluding winds and tides
- Highly accurate since application of GPS technology

The Earth can also be represented as a geoid, which is a model of earth still lacking the effects of wind and tides on the earth – but which includes the major force - gravity. Because the effect of gravity is retained, the earths overall shape is both elliptical however the surface is irregular in regard to distance from the center. The surface of the Geoid is oftentimes roughly equivalent to what we could call a global mean sea level.

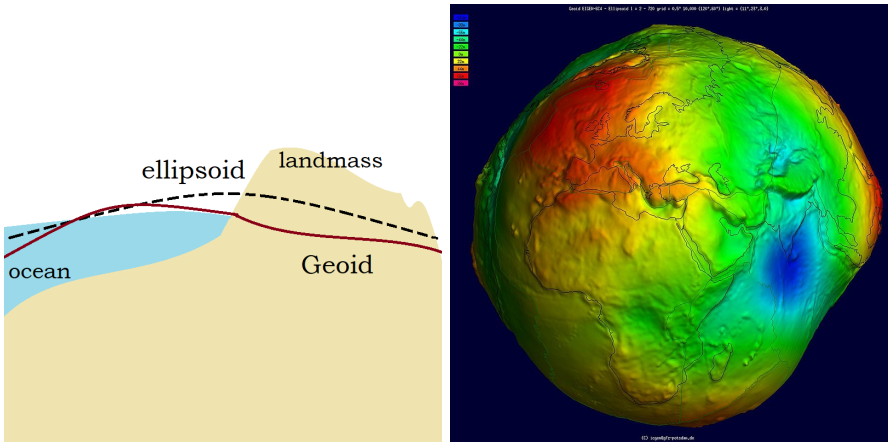


Figure 5: Left: Geoid Cross Section. Right: IGCM Geoid

Oftentimes now, very finely resolved geoids are combined with satellite measurements, and these observations are then fit with ellipsoid models, which we then use as our models of the earths surface.

By International Centre for Global Earth Models (ICGEM) - http://icgem.gfz-potsdam.de/vis3d/longtime / Ince, E. S., Barthelmes, F., Reißland, S., Elger, K., Förste, C., Flechtner, F., Schuh, H. (2019): ICGEM – 15 years of successful collection and distribution of global gravitational models, associated services and future plans. - Earth System Science Data, 11, pp. 647-674,DOI: http://doi.org/10.5194/essd-11-647-2019., CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=81462823

3.2 Geodetic Datums

- Reference frame established to represent locations within the frame.
- Historically these were locally focused and based on Geoids.
- A datum can serve either horizontal (X & Y) or vertical (Z) features.
- Components:
  - reference ellipsoid or geoid
  - origin point (from which measurements run)
  - control points very strictly measured from the origin,

Other locations are then measured in relation to the control points.

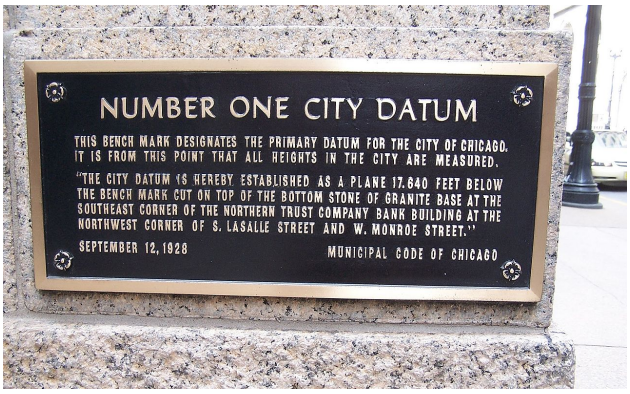


Figure 6: Number one City Datum. by: Cosmo1976

When we think of the Origins of datum points, the Equator and Prime Meridian are the most famous. However, there have been a great many datums in history.

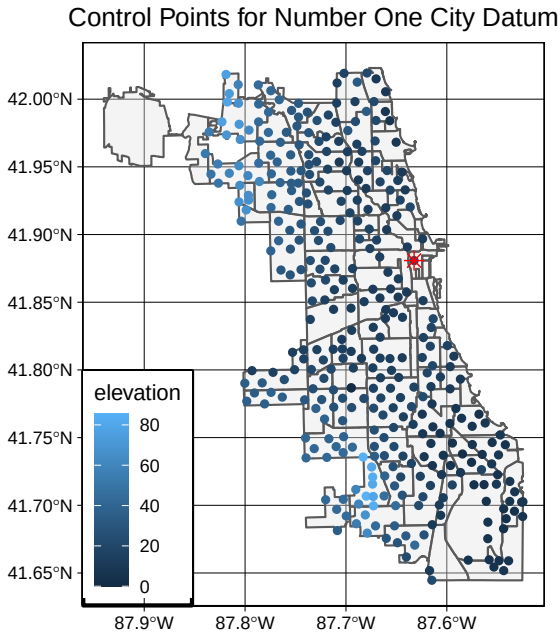


Figure 7: Control Points Number one City Datum

In fact the city of Chicago established it's own elevation datum in the late 1920's. If we look at this quick map, we can see where the Origin point (also pictured) and control points are located.

Also do you really think Chicago is only 80 feet at most above sea level? No of course not! The lowest elvations in the city are roughly 580 feet above sea level, but in the 1920's, it was much easier to define elevation from an adjacent location. So all the values in the map should be X + ~580 ! This is exactly why so many datums are out there.

3.2.1 Geodetic Coordinates

Since our area is 3-dimensional, we use a system related to Cartesian coordinates, however given the curve of the earths surface curvilinear rather than linear coordinates are used.

"The geodetic (or geographic) latitude is the angle between the equatorial plane and the normal (vertical) to the ellipsoid surface at the considered point." - Proj

- phi = geodetic latitude (north/south)
- lambda =longitude (east/west)
- h = ellipsoidal height
- N = Normal (A plane at a right angle to the surface of the ellipsoid)

In the sphere image we have a latitude of 40*, but we do not know to which hemisphere it is associated with.

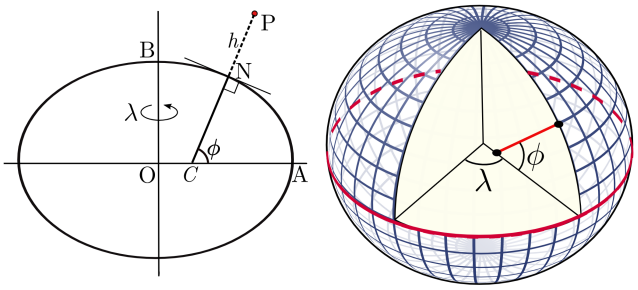


Figure 8: Angles on Ellipsoid and Geodetic Coordinates. by: Peter Mercator

By Peter Mercator - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=17717979

3.2.2 Coordinate Notation

- Trigonometric
  - Degrees Minutes Seconds (DMS), e.g. 42°03'27.7" N
    - Sexagesimal/base 60 (think: minutes in an hour)
  - Decimal Degrees (DD), e.g. 42.05759 N
    - Decimal Fractions of a Degree (portion of 360/2)
- DMS; seldom used in digital formats
- DD; almost exclusively used

Recorded in many forms DMS, DD, (trigonometry) UTM.

As a circle has 360 degrees, so does the earth. Given the nature of datums we split the world into a positive and negative sign for both latitude and longitude; we refer to the 0 lines as '0 meridians'. We have 180 positive and 180 negative degrees, and as these are of a course resolution we retain the decimals of degrees to better locate objects. As you all know, the Western Hemisphere is the negative component of a 180 degree half.

$$\text{Decimal Degrees} = \text{Degrees} + \frac{\text{Minutes}}{60} + \frac{\text{Seconds}}{3600}$$

$$\text{Latitude of Tech} = 42\check{r}03'27.7"N$$

$$42 + (\frac{03'}{60}) + (\frac{27.7"}{3600})$$

$$42 + 0.05 + 0.007694 = 42.05759 \text{ Decimal Degrees}$$

- Universal Transverse Mercator (UTM)
  - Divides the world into 60 zones
  - Flattens each zone
  - Measures distances in Meters

UTM - Common for field work, planar measurements of meters.

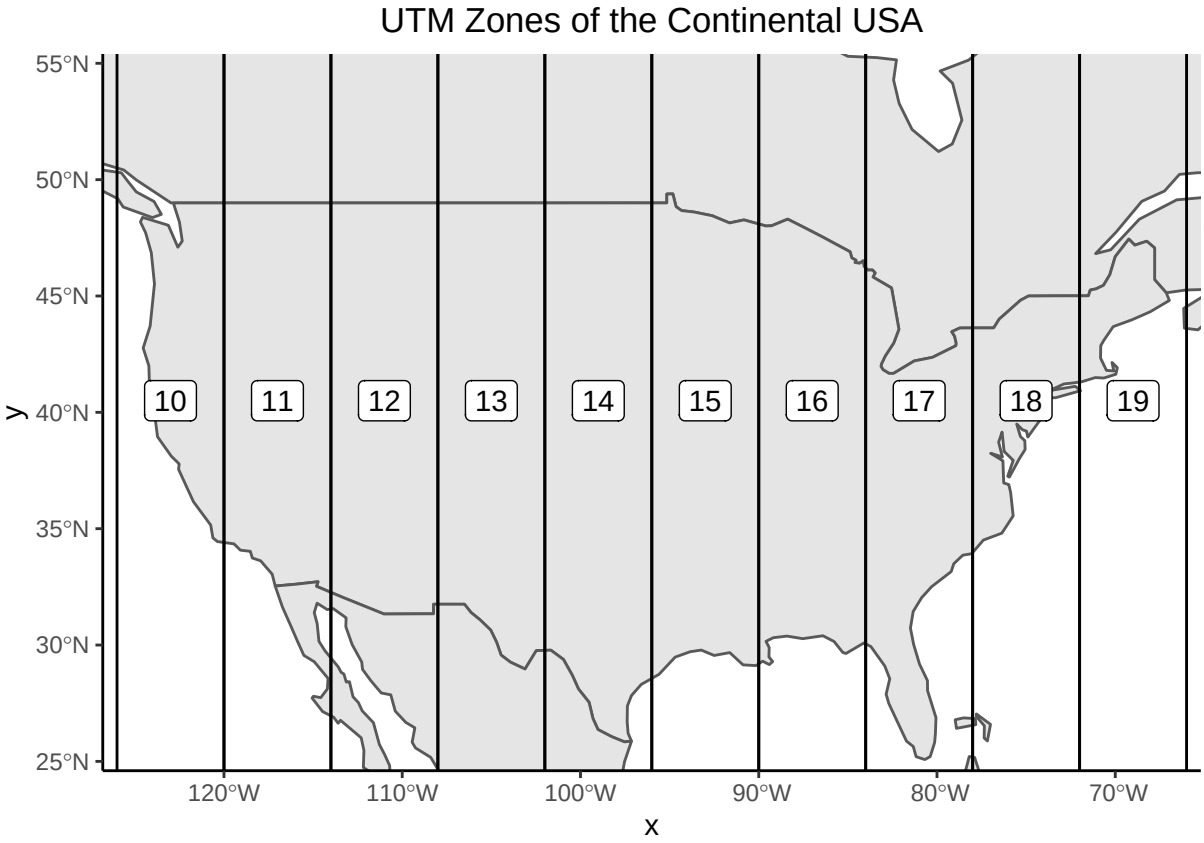


Figure 9: UTM Zones of the Conterminous USA. data by ESRI

3.3 Coordinate Reference System

- System for specifying a location on Earth's surface
- Composed of:
  - a model of earths shape (Geoid or Ellipsoid)
  - geodetic datum
  - generally also a projection

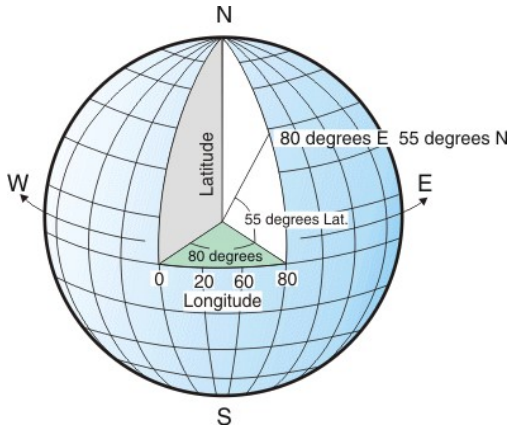


Figure 10: Geographic Coordinate Reference System. by: Anna Krystalli

3.4 Problems with working on flat surfaces

3D to 2d does not work well. But we have worked around this. **sorta**

Figure 11: Orange Peel. by: Nathan Belz

3.4.1 Geographic Coordinate System

- Location on a three-dimensional model of Earths Surface

Where the location is on located earth. We are able to pinpoint locations, quite easily using one of the Global Navigation Satellite Systems, for example the 'Global Positioning System' or GPS. We have a coordinate system for this known as the World Geographic Datum. As we are merely recording the location of an object on a 3d object this is now essentially a trivial process.

3.4.2 Projected Coordinate System

- Location on a two-dimensional model of Earths Surface

Most of us work on flat desks and flat computer screens. We need to represent locations in space on these surfaces. We project coordinates from their geographic locations on earth to locations on flat representations of earth.

3.4.3 Major Map Projections

Many different ways to create two dimensional maps – thousands of projections. None are perfect. There are four main types of projections, each with their own strengths and examples. One of the most common examples of each of these is included in the table below.

- Equal Area: Lambert Cylindrical equal-area
  - Pro: No distortion of area near equator (here)
  - Con: Distorts area at the Polar regions
- Equal Distance: Equi-rectangular (plate carrée projection)
  - Pro: Looks good in mapping applications
  - Con: Distorts both shape and directions
- Conformal:
  - Pro: Boundaries are accurate
  - Con: Distorts Polar area
- Compromise:
  - Pro: Sensitive to both area and direction
  - Con: Sensitive to both area and direction

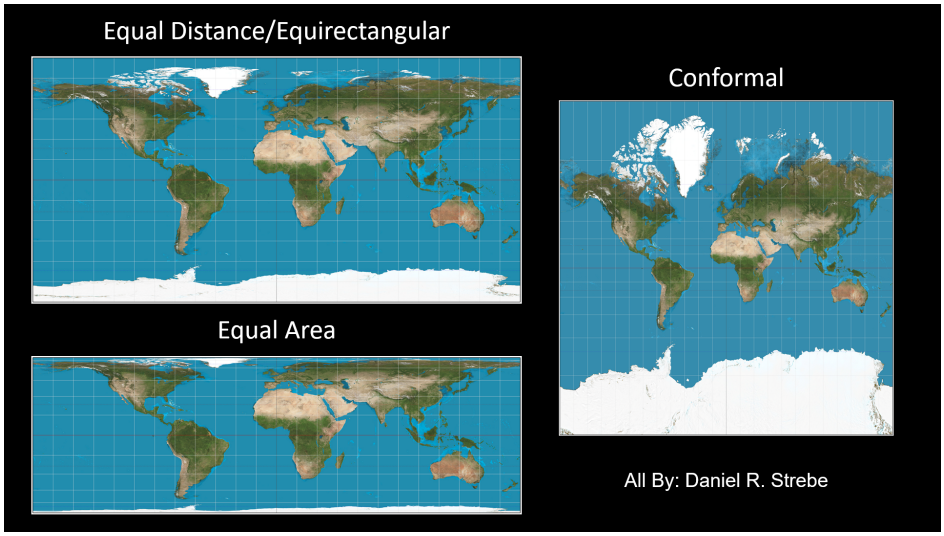


Figure 12: Map Projections. by: Daniel Strebe

3.5 Geodesy Takeaways

Unless you really focus on GIS, hardly anything I have said this lecture will matter to you.

- There are various models (geoids & ellipsoids) to represent the shape of the earth
- Different datums are used to represent different parts of the earth. This is in part due to legacy effects.
- You will almost always use WGS 84 (based on a ellipsoid – which is fit through a special model of earths gravitational fields a geoid) NAD83 (based on a ellipsoid) for a geographic coordinate system. These +/-1 m from each other across much of North America. More useful than a geoid.
- You will usually want to use a UTM grid based on WGS for or State Planes based on NAD83 for projections.
- Different coordinates notation systems are used, focus on Decimal degrees.

4 An Introduction to Geographic Data Models

- **Vector Data Model** represents discrete features on the planet using geometries such as: points, lines, and polygons.
- **Raster Data Model** represents (usually) continuous features on the plant using continuous surfaces, like a tile.

Vector data tends to only include features of interest, e.g. bodies of water; whereas a Raster will include, **explicitly**, the absence of features (e.g. both water and terrestrial areas).

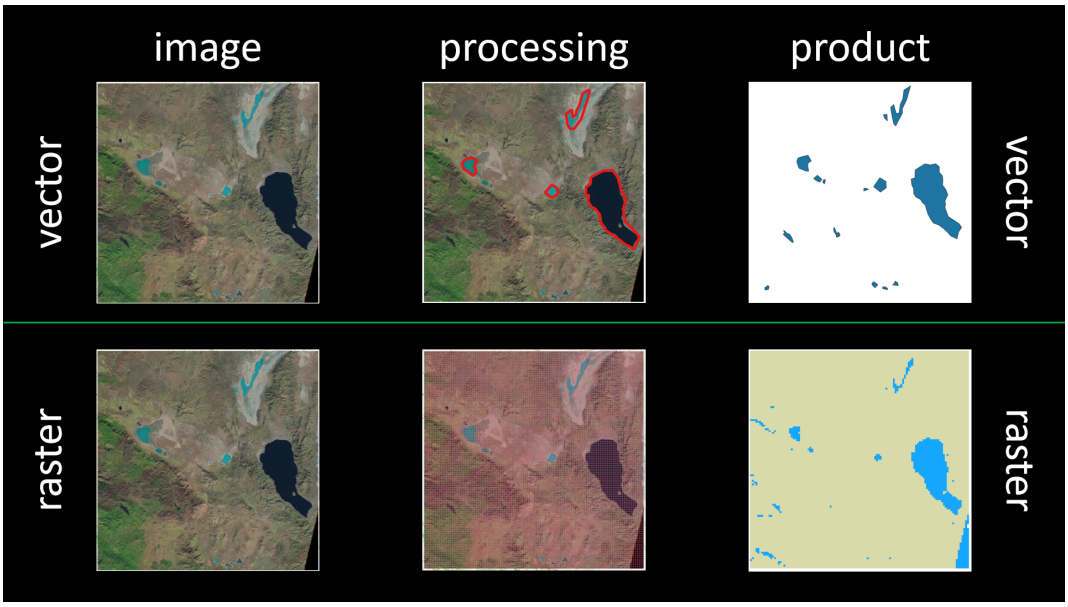


Figure 13: Vector and Raster Data Models

Talking:

In our field, we generally take values from rasters to serve as predictors to our sample unit, many of us store our sample as spatial points. We also tend to predict our models back onto raster surfaces.

5 R Data Types Reviewed

Remember that different data types take up different amounts of memory in pretty much all software systems.

```
When the values in raster cells are stored as integers as opposed to floats (with decimal points) they
```

```
When the values in raster cells are stored as integers as opposed to words they take up only 0.5 as many
```

Please note that our Raster in this scenario is *very* small, small enough we can wrap our brains around it with only minimal inspection.

Without knowing much about a raster (yet!) We can simulate some of these comparisons to give a clue as to why it contains values like they do.

Rasters generally come in 8-bit signed (unsigned are not uncommon) integers. Pictures are huge amounts of data, these values help reduce them.

```
The size of the our empty raster is 0.01416 MB
```

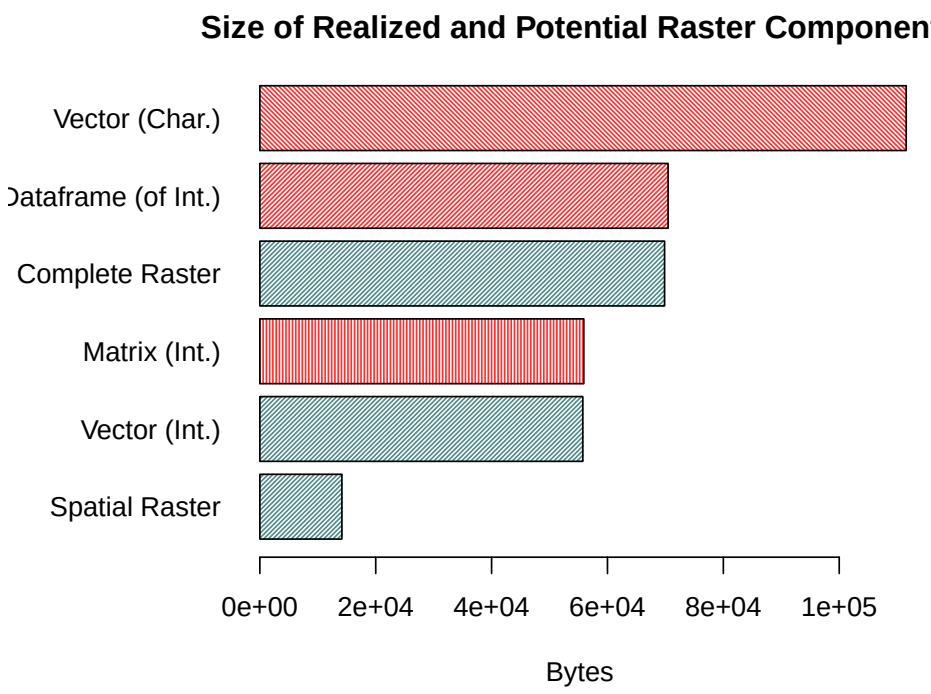
```
The size of the data for our raster as a character vector is 0.11156 MB
```

```
The size of the data for our raster as a numeric (integer !) vector is 0.055744 MB
```

```
The size of the data for our raster in matrix form is 0.055912 MB
```

```
The size of the data for our raster in dataframe form is 0.055912 MB
```

```
The size of our complete raster is 0.069856 MB
```



6 Vector Data in R

- **Vector Data Model** represents discrete features on the planet using geometries such as: points, lines, and polygons.

6.1 Simple Features - Standards

- Open Geospatial Consortium ISO 19125-1:2004: Currently adhered to in ESRI, used in GDAL.
- Features have *geometries* describing their locality on Earth, and properties described by *attributes.*
- If the geometry of a feature is a polygon, it is composed of points, connected by straight lines.
- Lines composing polygons cannot intersect

6.1.1 Simple Features 1 - Attributes

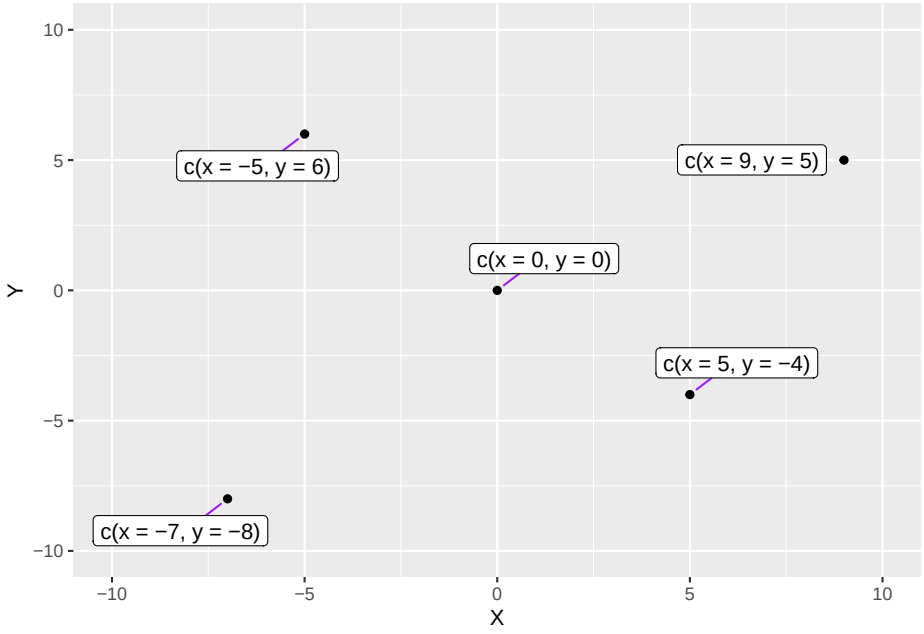
- Attributes of the feature, theoretically devoid of spatial context.
- Described in text, numbers, stored in a data frame type object.

TAXON	DBH	HEIGHT
Robinia pseudoacacia	40	24
Quercus alba	32	21

6.1.2 Simple Features 2 - Coordinates form a Point

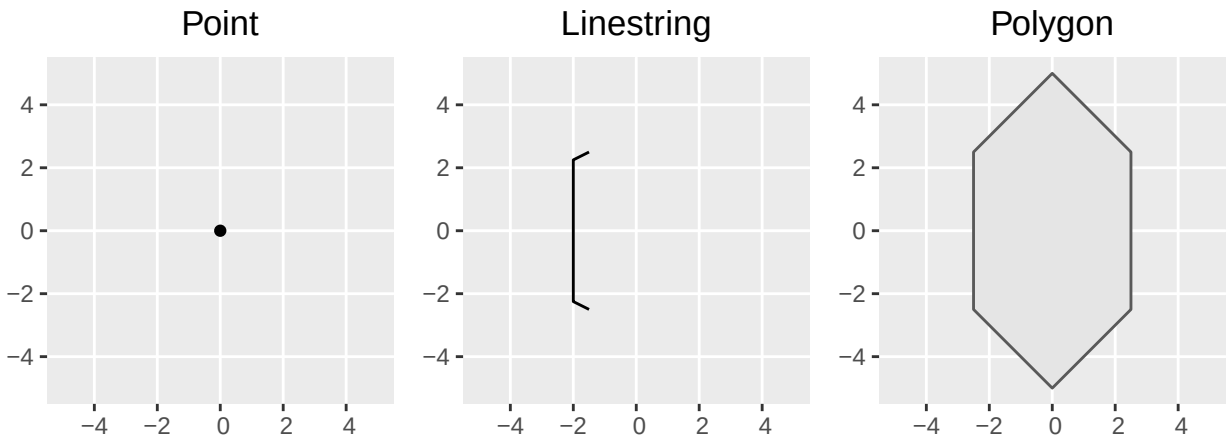
- all geometries are composed of points.
- points only require two coordinates, X & Y.
- Y = Latitude (necessary), X = Longitude (necessary)
- Z = Elevation (somewhat uncommon)

- M = Time or Uncertainty of Measurement (rather uncommon)



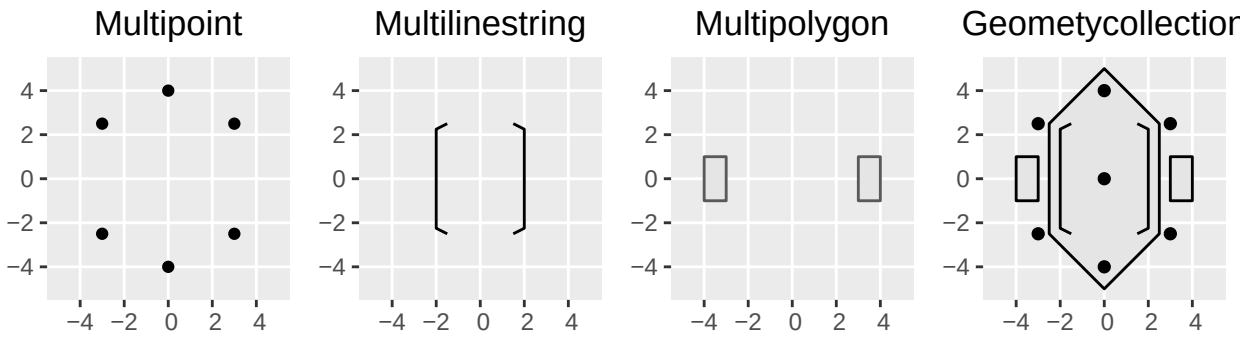
6.1.3 Simple Features 3 - Points are/form a SF Geometry (sfg's)

- SFG is the spatial topology associated with a feature
- POINT - A true dimensionless 1 dimensional point
- LINESTRING - Sequence of points connected by strings.
- POLYGON - Sequence of points connected by strings, which close upon themselves.
- i.e. the first and last point of the polygon are the same.



6.1.4 Simple Features 4 - A Geometry can be combined to form Geometries (sfg's)

- Two or more Points, and sets of Linestring, and polygons can be the geometries of a single feature
- Multi(POINT), multi(LINESTRING), multi(POLYGON) - Forming collections of geometries
- GEOMETRYCOLLECTION - A mixed set of these three or more geometries in the same geometry



6.1.5 Simple Features 5 - Geometries and Spatial Information form an SF Collection (sfc)

- A list column of S3 type containing certain parameters regarding the Geometry/Geometries
- Coordinate Reference System ('crs')
- Precision ('precision')
- Bounding box ('bbox')
- Number Empty ('n_empty')

Wherein precision refers to the precision of the coordinates forming the geometries. the bbox is a rectangle which includes all pairs of coordinates in the geometry. N_empty, denotes whether the geometry is missing coordinates in a position.

6.1.6 Simple Features 6 - Recapped.

TAXON	DBH	HEIGHT	LONG	LAT	geometry
Robinia pseudoacacia	40	24	-87.67607	42.05663	POINT (-87.67607 42.05663)
Quercus alba	32	21	-87.67399	42.05715	POINT (-87.67399 42.05715)

```
Geometry set for 1 feature
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: -87.67607 ymin: 42.05663 xmax: -87.67607 ymax: 42.05663
Geodetic CRS:  WGS 84
```

```
Precision:     2
```

```
sfc_POINT of length 1; first list element:  'XY' num [1:2] -87.7 42.1
```

- We now finally have *both* attributes and coordinates which are forming a point geometry.
- This is a Simple Feature.

6.2 sp

- Predecessor to the 'SF' R package
- Many spatial statistics programs still only take sp objects as input
- A good introduction to S4 objects; popular in new spatial packages too!

6.2.1 sp - History

- Released in 2005
- First package to hold all major vector types of geometries
- Unified Spatial class allowed for support in many spatial statistics packages
- Improved mapping of spatial objects
- Formed the centerpiece of spatial statistics in R for over a dozen years.

6.2.2 sp - Spatial Classes for Topology

- SF holds different geometries in list columns, SP has *many* different types of objects/classes for holding geometries.
- SpatialPoints
- SpatialLines
- SpatialPolygons
- SpatialGrids

Similar in theory to the simple feature geometries we learned about...

6.2.3 sp - Structures of The Spatial* Class - Topology Only

```
## Formal class 'SpatialPoints' [package "sp"] with 3 slots
##   ..@ coords     : num [1:15, 1:2] 0.9 0.62 0.32 0.39 0.62 0.4 0.41 0.24 0.39 0.44 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : NULL
##   .. .. ..$ : chr [1:2] "xc" "yc"
##   ..@ bbox       : num [1:2, 1:2] 0.03 0 0.96 1.67
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:2] "xc" "yc"
##   .. .. ..$ : chr [1:2] "min" "max"
##   ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
##   .. .. ..@ projargs: chr NA
```

7 sp - Attributes - DataFrame

- Attachment of attributes, in a data frame, to a Spatial topology can create:
- SpatialPointsDataFrame
- SpatialLinesDataFrame

- SpatialPolgonsDataFrame
- SpatialGridsDataFrame
- Each S*DF is accessed the same way:
  - SPDF@data[[‘col_name’]]
  - SPDF$‘col_name’

```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
##   ..@ data       :'data.frame':  15 obs. of  2 variables:
##   .. ..$ temperature: num [1:15] 3.95 6.39 4.32 4.74 5.23 4.95 5.05 5.85 3.94 5.24 ...
##   .. ..$ aspect     : int [1:15] 29 1 40 19 43 36 20 37 5 45 ...
##   ..@ coords.nrs : num(0)
##   ..@ coords     : num [1:15, 1:2] 0.9 0.62 0.32 0.39 0.62 0.4 0.41 0.24 0.39 0.44 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : NULL
##   .. .. ..$ : chr [1:2] "xc" "yc"
##   ..@ bbox       : num [1:2, 1:2] 0.03 0 0.96 1.67
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:2] "xc" "yc"
##   .. .. ..$ : chr [1:2] "min" "max"
##   ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
##   .. .. ..@ projargs: chr NA
```

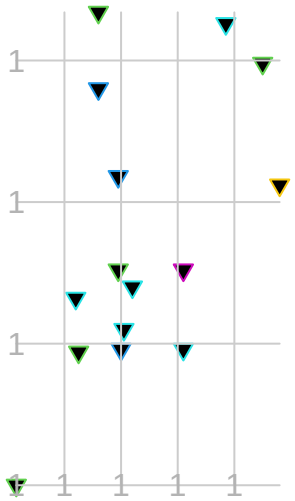
temperature	aspect
3.95	29
6.39	1
4.32	40
4.74	19
5.23	43
4.95	36

We see here that SF contains all of the spatial information in a Simple Feature Collection which can be tidily tucked away in the geometry list column, in SP the analogue to the SFC is the object itself.

SP is mostly geometry with a data slot, SF is mostly data frame with a geometry column.

8 sp Overview of a Spatial*DataFrame

- Data frame is held in a different slot from the geometry and topology
- Data frame columns may be subset
- Data frame columns accessed via object@data[[‘colname’]] indexing



8.0.1 sp - Recapped.

- If you need to use spatial statistics, you will come across these objects.
- Don't sweat them **too** much; you can always convert from and to sf to run certain analyses

9 Raster data in R

- Reminder: the Raster data Model is a way to represent (continuous) features on the planet using grids
- Excellent format for storing and sharing data
- Divides space into continuous grids
- Raster data sets are often distributed in *tiles*

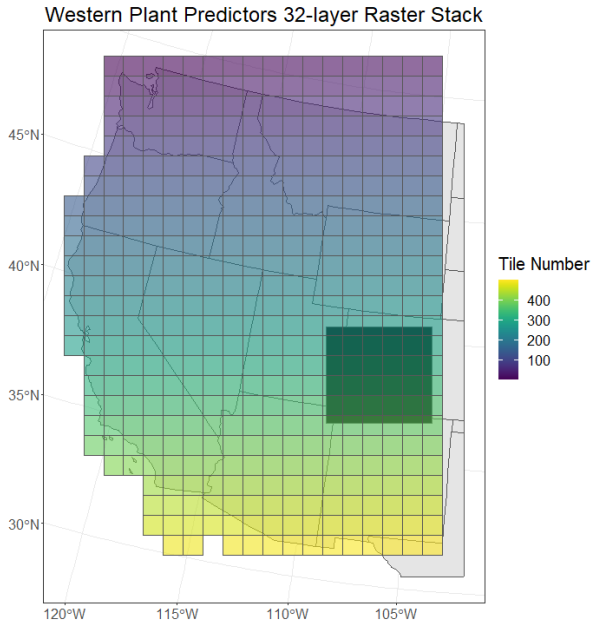


Figure 14: Western Plant Predictors. Note each cell is the extent of a single raster tile, each tile contains cells.

9.1 Raster Components

- Bounding Coordinate(s)
- Cell Resolution (Size of cell)
- Dimensions (No of cells in rows and columns)
- Coordinate Reference System (CRS)

Note that both *cell sizes*, and the *resolution of the values* in cells can, within reason, be converted to finer and coarser resolution. We will discuss these types of calculations next class.

9.1.1 Example Raster 1 Create Frame

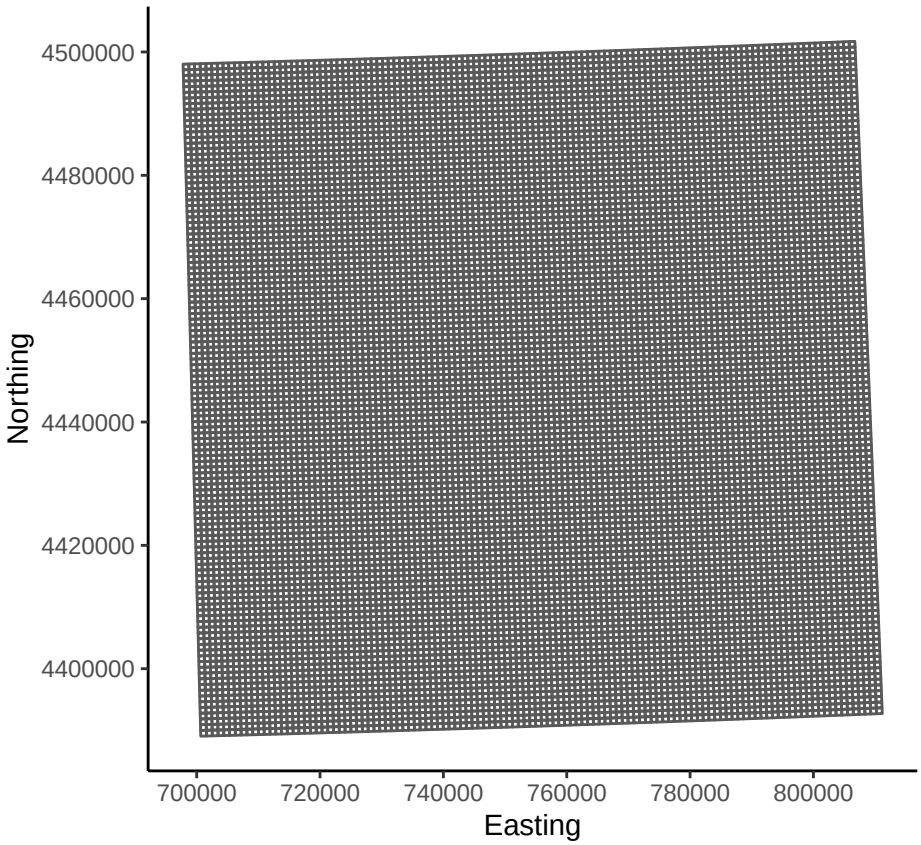
Rasters tend to confuse people. We are going to create our own example here before we talk about them much more.

Fairy shrimp live in bodies of water which dry out in late Spring, and refill in early Fall. We seek to determine where suitable habitat for fairy shrimp are in the Great Basin. In order to do so, we will use satellite imagery to detect valleys which fill with water in Fall, and dry out in late Spring.

Figure 15: Suprise Valley by: John Glen

```
empty_raster <- raster(

  # rasters have 4 bounding edges
  # Here we define each 'corner' of the raster'
    xmn = 697129.7,
    xmx = 811775.7,
    ymn = 4388466,
    ymx = 4502382,

  # Here we set the number of cells 118*118
    nrows = 118,
    ncols = 118,

    # we do NOT manually specify the cell size here.

    crs = "+proj=utm +zone=10 +datum=WGS84",
    # set the rasters Coordinate Reference System
)
```

In the code above, we are going to define all four of the essential components of a raster. Clearly, we are defining three explicitly (Bounding Coordinates, Dimensions, CRS), and the remaining one implicitly (Cell resolution).



We can imagine that the raster we are currently creating looks like this. A frame without content. We can see where the bounding coordinates are,

9.1.2 Example Raster 2 Set Values

```
raster_matrix <- matrix(rast_vals_num, # # fill matrix with values,
                        nrow = empty_raster@nrows, # create matrix same dimensions.
                        ncol = empty_raster@ncols, # create matrix same dimensions
#ensure filling of matrix goes from upper left to lower right.
                        byrow = T)

raster_matrix[90:118,112:118] <- 1 # fix the clipping image at edge.
```

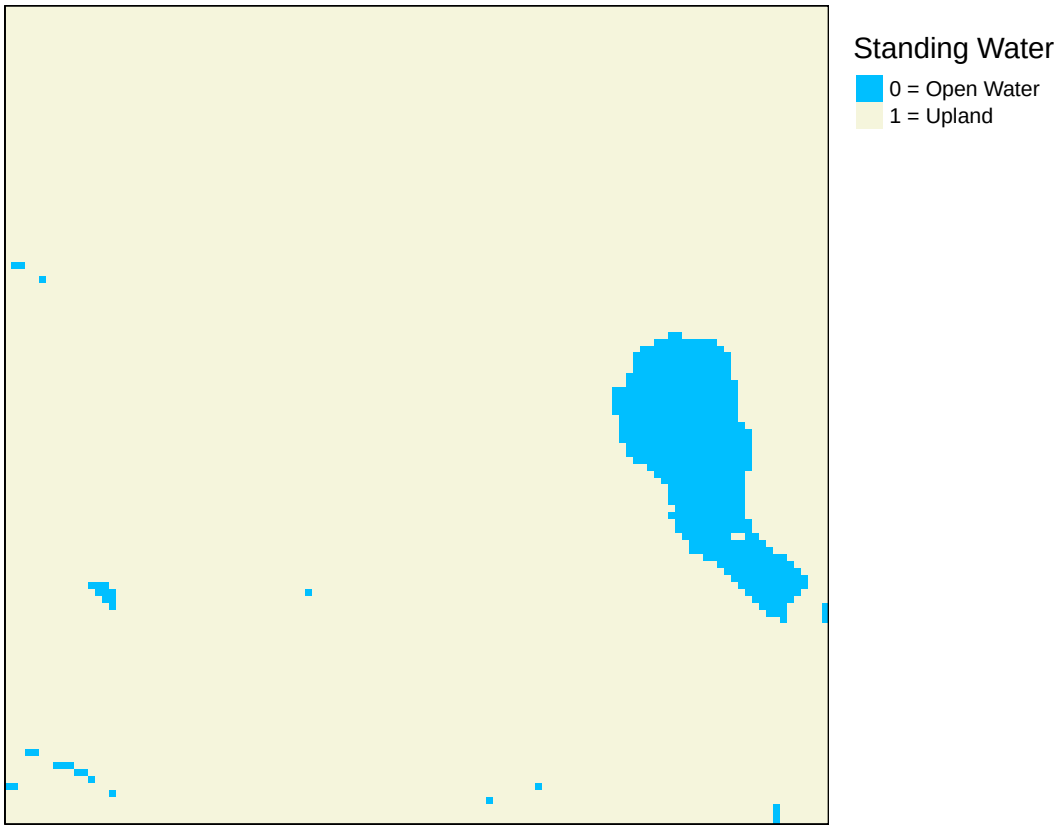
Table 4: Example Matrix showing values underlaying a raster layer.

0	1	1	0	1	0	0	1
1	0	1	0	0	1	1	0
1	0	0	1	1	0	1	0
0	1	1	0	1	0	0	1
1	0	1	0	0	1	1	0
1	0	0	1	1	0	1	0

I think it is easiest to imagine that the values within a raster are in the shape of a matrix. Hence, using the code above we could take a vector of values, and bend them, so that each position within the vector matches up with the beginning of a new row.

```
example_raster_dec <- setValues(empty_raster, raster_matrix)
```

9.1.3 Example Raster 3




```
The width of each raster cell is: 971.57627 meters
The height of each raster cell is: 965.38983 meters
```

```
This example_raster_dec contains: 13924 elements
```

9.2 Raster Package - in R!

- Does not need to load all files into active memory at once
- Many functions based on functions in 'base' R.

9.2.1 Attributes

- Cells
- Values

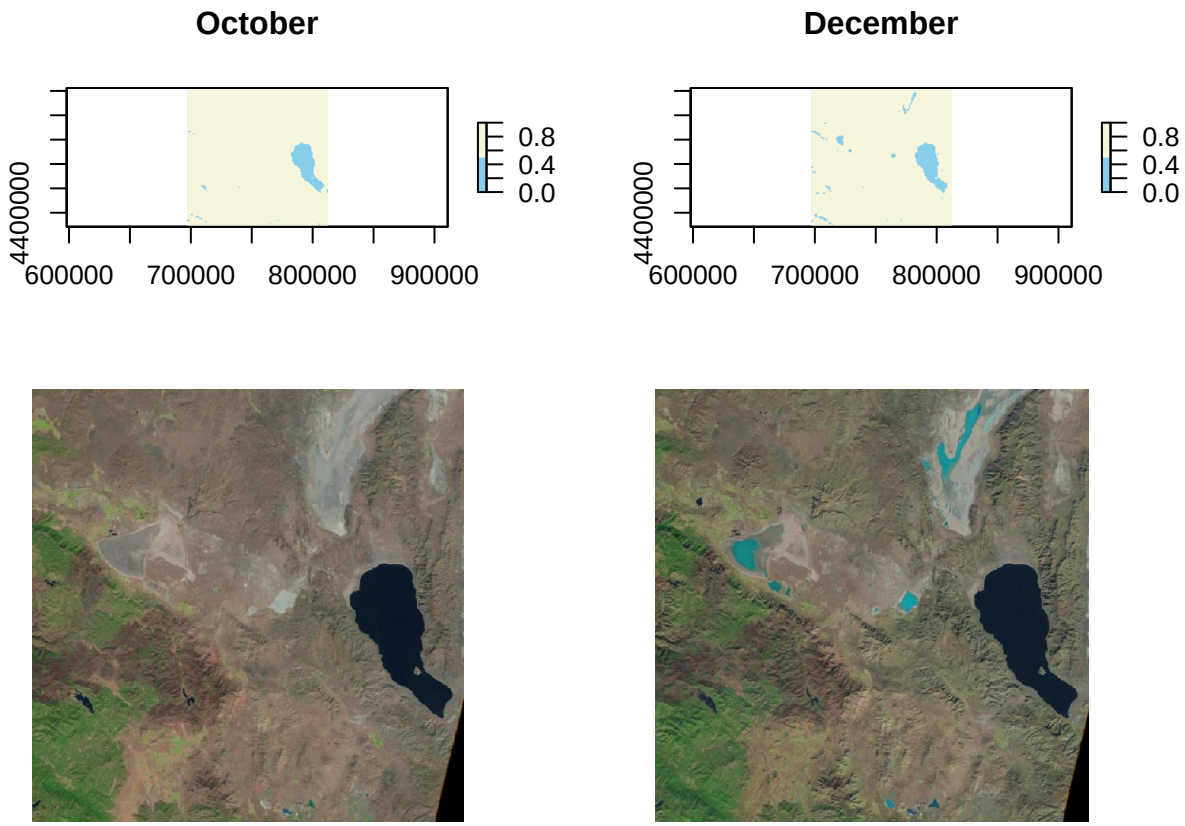
9.2.2 Raster Layer

A single raster layer.

9.2.3 Raster Stack

In general these are rasters which have been classified and we want to extract values from or run calculations with. Now what is great about layers, is that we can do one big thing bricks cannot do, we can load in many

layers from many files and create a stack of attributes we are interested in studying on the fly.



The main use of Raster Stacks is to hold layers of similar themes.

Two usage examples:

1) Each layer is the same variable from a different time.

- e.g. mean monthly temperature from January -> December (12 layers per stack)

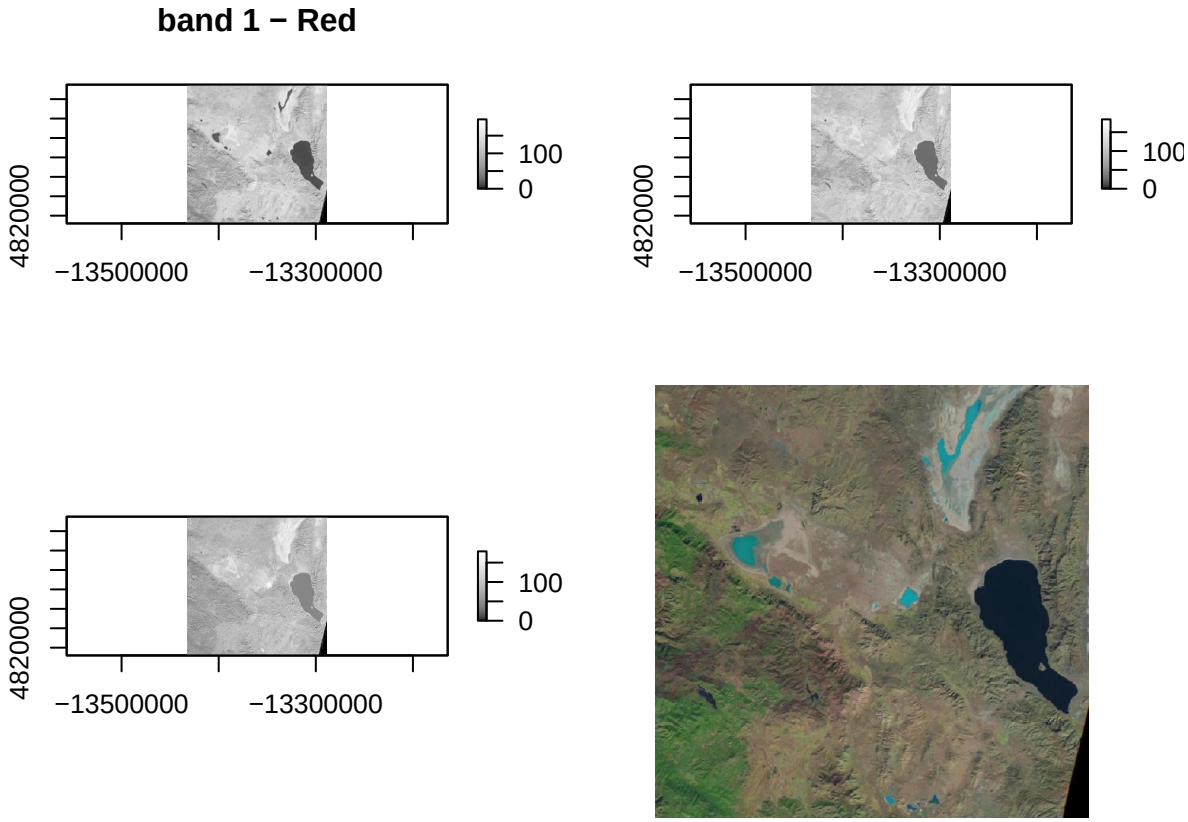
2) Each layer is a different variable (theme) in an analysis.

- e.g. yearly mean temperature, mean precipitation, etc.
- Do not need be held in memory

9.2.4 Raster Brick

- Multiple bands of imagery held in the same file
- The sensors of a camera, e.g. Red, Green, and Blue
- Used for performing image classification to produce data raster layers
- Now bricks of over 100 sensor bands exist...

As I have mentioned rasters are often generated from satellite imagery; we will discuss rasters which are not developed this way next lecture. Historically most pictures are imaged via the use of three bands. These having spectral values of Red, Green, and Blue (RGB) associated with them. For example our .tif file, is composed of three bands. Note that a Rasterbrick is most often used for loading in these types of imaging data, which can then be processed to form a more typical 'raster' dataset.



Rasters are often built from satellite imagery. Historically most pictures are split into three copies, each one having spectral values of Red, Green, and Blue (RGB) associated with them. For example our .tif file, is composed of three layers. Note that a Rasterbrick is most often used for loading in these types of data.

Bricks are very important for processing and classifying raw image data. While we have three bands here, nowadays LIDAR equipped with Hyperspectral sensors are likely to have many more bands up to a couple hundred depending on the application.

One important technical point to note with the Rasterbrick is that each band of the brick loads from the same individual file, for example picture. And that bands are not typically combined from different picture sources.

Raster bands also do not need to be held in memory, allowing one to work through large amounts of them.

9.3 Terra

- Developed by the same team as the Raster Package
- Functionally virtually identical to Raster, but calculations run more quickly ! :-)
- Rasterbricks/layers no longer need specification, all objects are Spatstats ?
- Will supercede Raster, but see points 1 & 2.

10 Cartography

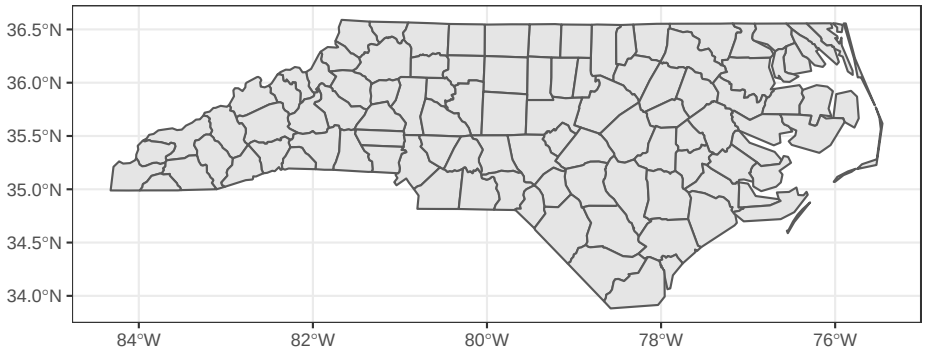
- Focus on sf here
- You have scripts to do many types of mapping in your course resources
- sf objects are ggplot2 compliant

- the order of mapping operations is more important than code
- Import example data sets which come with two packages

```
data("us_states", package = "spData")
us_states <- st_as_sf(us_states)

north_carolina <- read_sf(system.file("shape/nc.shp", package = "sf"))
```

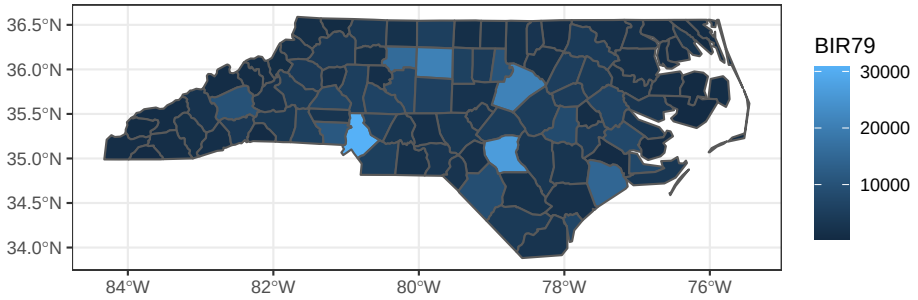
- Map an sf object using their own geom 'geom_sf'

```
ggplot(north_carolina) +
  geom_sf() +
  theme_bw() # we will use this theme for class.
```



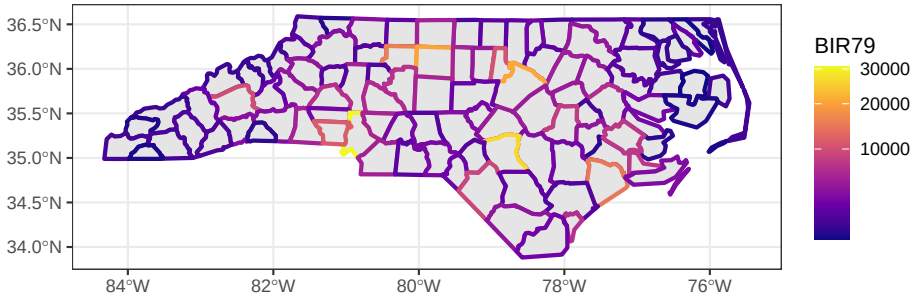
- Fill the interior of polygons by a variable

```
ggplot(north_carolina) +
  geom_sf(aes(fill = BIR79))  + # fill the interior of polygons by a variable
  theme_bw()
```



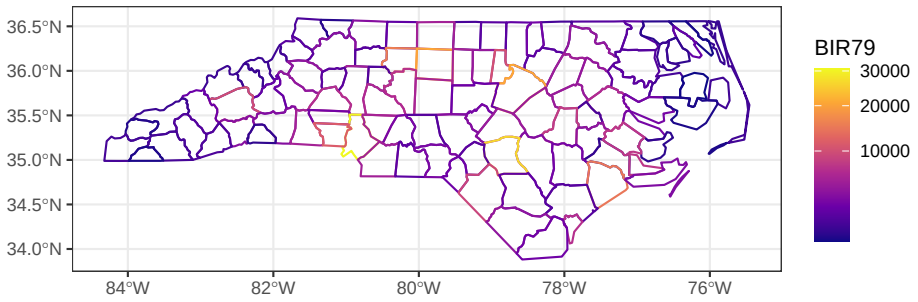
- Color the borders of each polygon in an sf object

```
ggplot(north_carolina) +
  geom_sf(aes(color = BIR79),  # color the borders of polygons by a variable
          lwd = 1 # just making the borders thicker.
          )  +
  scale_color_viridis_c(option = "plasma", trans = "sqrt") +
  # just using a very colourful color scheme so you can see it.
  theme_bw()
```



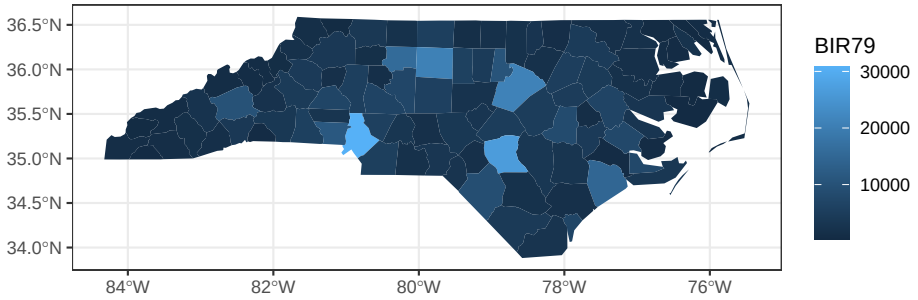
- Color the border and remove any fill from the polygon interior

```
ggplot(north_carolina) +
  geom_sf(aes(color = BIR79),
          fill = NA  # set to 'NA' to 'remove' the interior of polygons
          ) +
  scale_color_viridis_c(option = "plasma", trans = "sqrt") +
  theme_bw()
```



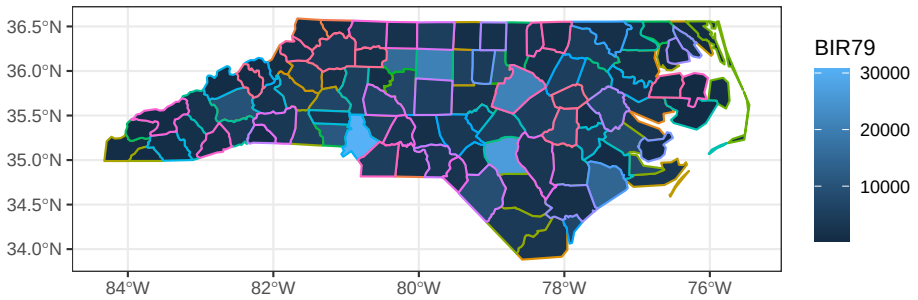
- Color the interior of polygons and remove the border
- Can be useful to declutter maps

```
ggplot(north_carolina) +
  geom_sf(aes(fill = BIR79),
          color = NA  # set to 'NA' to 'remove' the interior of polygons
          ) +
  scale_color_viridis_c(option = "plasma", trans = "sqrt") +
  theme_bw()
```



- Both fill the interior of polygons and color the borders

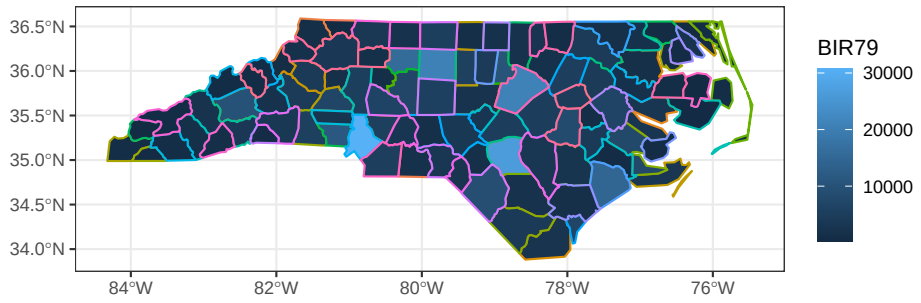
```
ggplot(north_carolina) +
  geom_sf(aes(fill = BIR79,
              color = FIPS)
          ) + # both fill and color
  guides(color = 'none') + # removed the categorical legend (is too big!)
  theme_bw()
```



- all of this could also have been done as:

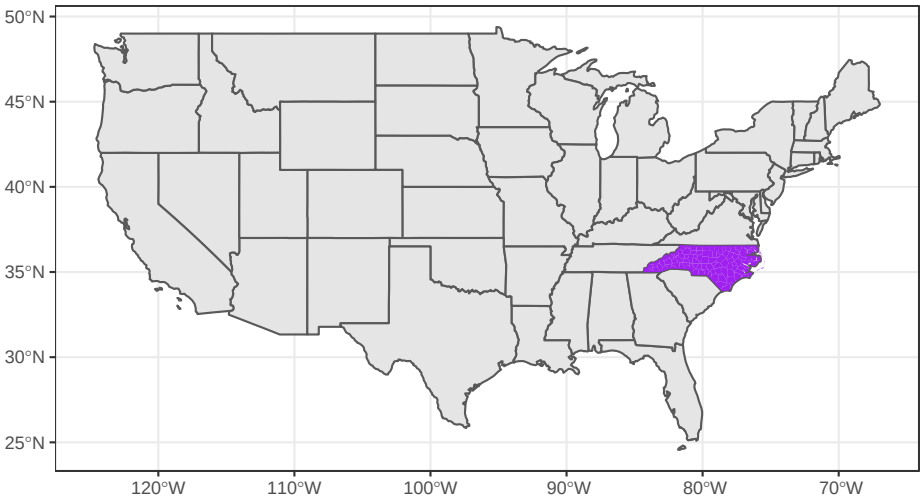
```
ggplot() + # leave empty
  geom_sf(data = north_carolina, # put data here
```

```
          aes(fill = BIR79,
              color = FIPS)
          ) + # both fill and color
  guides(color = 'none') + # removed the categorical legend (is too big!)
  theme_bw()
```



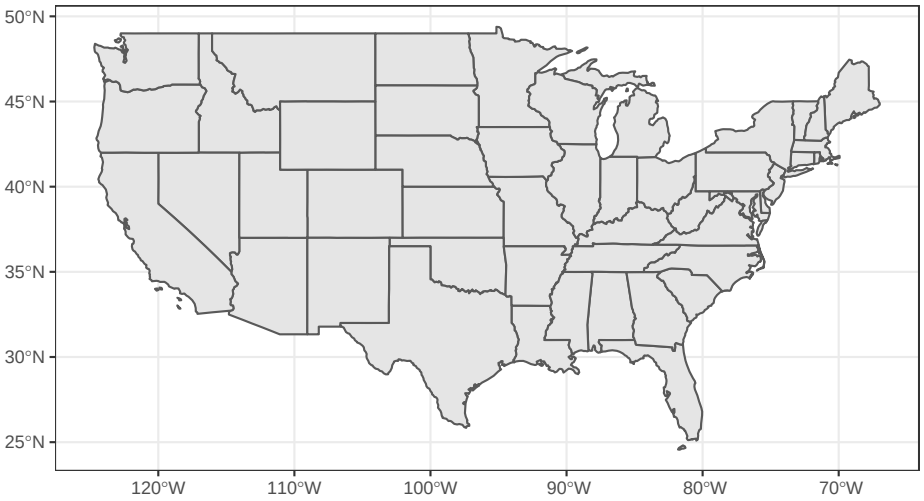
- Use two data sets to create one map

```
ggplot(us_states) + # two data sets.
  geom_sf() +
  geom_sf(data = north_carolina,
          fill = 'purple',
          color = NA # hash this line out in your own time to see the difference
          ) +
  theme_bw()
```



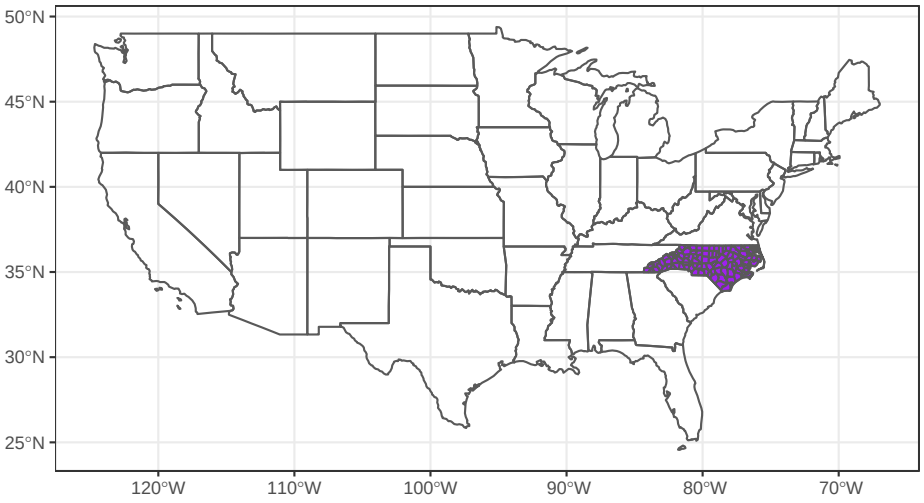
- Be diligent about the order of data sets

```
ggplot(north_carolina) + # oh no we drew over North Carolina!
  geom_sf(fill = 'purple') +
  geom_sf(data = us_states) +
  theme_bw()
```



- remember to build plots from the 'bottom' up

```
ggplot() + # two data sets.
  geom_sf(data = us_states, fill = NA) +
  geom_sf(data = north_carolina, fill = 'purple') +
  theme_bw()
```

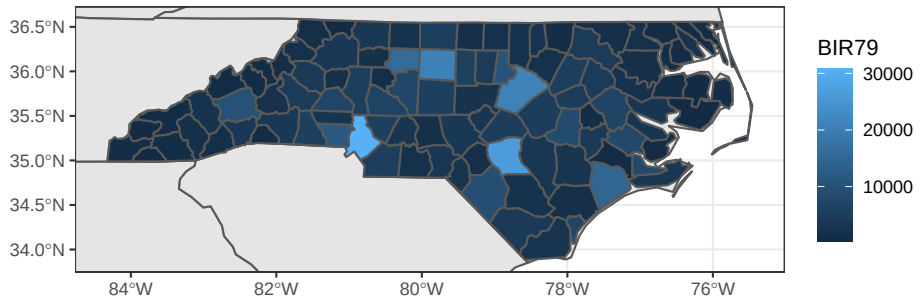


- coord_sf is a helpful modifier to geom_sf
- Can set a CRS for the map, modify extent, and change some rendering styles
- remember: '?coord_sf' for arguments!

```
ggplot() +
  geom_sf(data = us_states) +
  geom_sf(data = north_carolina) +
  coord_sf(crs = 4267, # we can convert each dataset to the same CRS
           datum = NA # we can remove the grid lines/graticules from plot
           )
```

- clip the extent of a map using the bbox of the data set of interest

```
bound <- st_bbox(north_carolina) # retrieve the bbox from the sfc list column

ggplot() +
  geom_sf(data = us_states) +
  geom_sf(data = north_carolina, aes(fill = BIR79)) +
  coord_sf( # use the bbox to 'crop' the extent of the map
           xlim = c(bound[1], bound[3]),
           ylim = c(bound[2], bound[4])
           ) +
  theme_bw()
```



11 Assignments for the Duration of the Spatial Data Science Module:

For next Lab:

Please install these packages (if you have not done so already):

```
install.packages("sf", "raster", "terra", "sp", "tmap", "leaflet", "ggmap")
# optional packages for using parallel processing at a step
# (in our example it won't actually speed up the analyses
# they actually may slow them down!)
install.packages('snow','parallel')
```

Download the labs .R script from the course website.

If you are interested in how Drone/Lidar data are collected please check out the 'RMBL Spatial Data Science Webinar Series':

- https://github.com/ikb-rmbl/SpatialDataScienceWebinars2020
- Collecting UAS Data (Video): https://youtu.be/Pq8btEZRCvM (1.25 hours)

For next Lecture: Assigned Reading: Chapter 3 of Spatial Data Science

Future Bonus SDS Office Hours Wednesday Night at 5:00 - 6:00.

Notes on this Lab My lecture notes are in the R script I used to generate all of the novel figures for this presentation. Likewise this presentation is an .HTML file and can be launched from your computer (it was rendered directly from R using the script).

12 Works Cited

Krystalli, A. https://annakrystalli.me/intro-r-gis/gis.html Accessed 01.20.2022

https://geocompr.robinlovelace.net/spatial-class.html Accessed 01.09.2022

Hijman, R. 05.12.2019 'The raster Package'

https://rspatial.org/raster/RasterPackage.pdf Accessed 01.09.2022

https://www.neonscience.org/resources/learning-hub/tutorials/dc-multiband-rasters-r Accessed 01.19.2022

Pebesma, E. https://r-spatial.github.io/sf/articles/sf1.html Accessed 01.10.2022

https://proj.org/operations/conversions/geoc.html Accessed 01.20.2022.

https://rspatial.org/raster/spatial/8-rastermanip.html Accessed 01.11.2022

https://en.wikipedia.org/wiki/Open_Source_Geospatial_Foundation Accessed 01.09.2022

NOAA. What is geodesy? National Ocean Service website, https://oceanservice.noaa.gov/facts/geodesy.html, 1/25/17.

Geocomputation with R. Lovelace, R., Nowosad, J., Muenchow, J. 2022-01-25.

12.1 Packages Cited:

```
c("raster", "sp", "sf", "tidyverse", "terra") %>%
  map(citation) %>%
  print(style = "text", na.print = '')
```

```
[[1]]
Hijmans R (2022). _raster: Geographic Data Analysis and Modeling_. R
package version 3.5-15, <URL:
https://CRAN.R-project.org/package=raster>.

[[2]]
Pebesma EJ, Bivand RS (2005). "Classes and methods for spatial data in
R." _R News_, *5*(2), 9-13. <URL:
https://CRAN.R-project.org/doc/Rnews/>.

Bivand RS, Pebesma E, Gomez-Rubio V (2013). _Applied spatial data
analysis with R, Second edition_. Springer, NY. <URL:
https://asdar-book.org/>.

[[3]]
Pebesma E (2018). "Simple Features for R: Standardized Support for
Spatial Vector Data." _The R Journal_, *10*(1), 439-446. doi:
10.32614/RJ-2018-009 (URL: https://doi.org/10.32614/RJ-2018-009), <URL:
https://doi.org/10.32614/RJ-2018-009>.

[[4]]
Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,
```

Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi: 10.21105/joss.01686 (URL: https://doi.org/10.21105/joss.01686).

[[5]]
Hijmans R (2022). _terra: Spatial Data Analysis_. R package version 1.5-18, <URL: https://rspatial.org/terra/>.