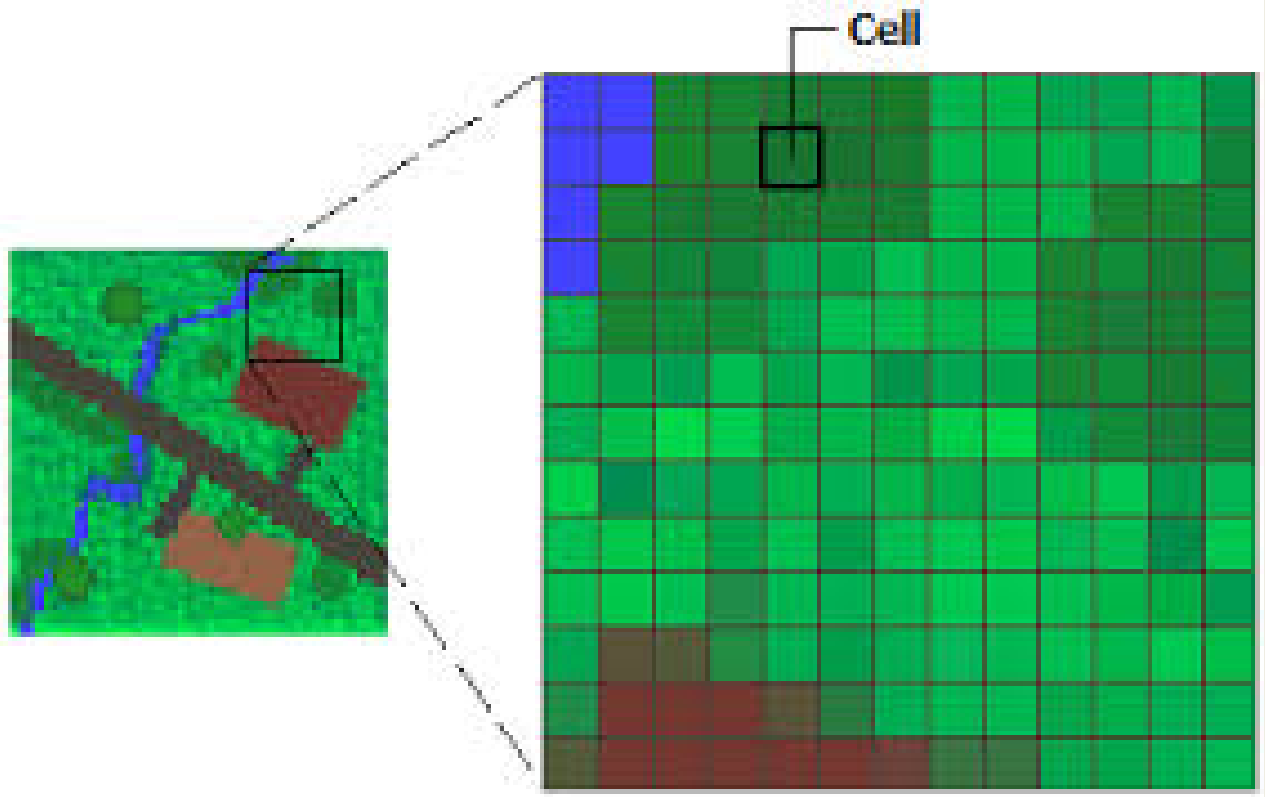
USING SPATIAL DATA WITH CVFS

Adrienne Epstein, PhD candidate
UCSF Department of Epidemiology and Biostatistics

TYPES OF SPATIAL DATA

Raster

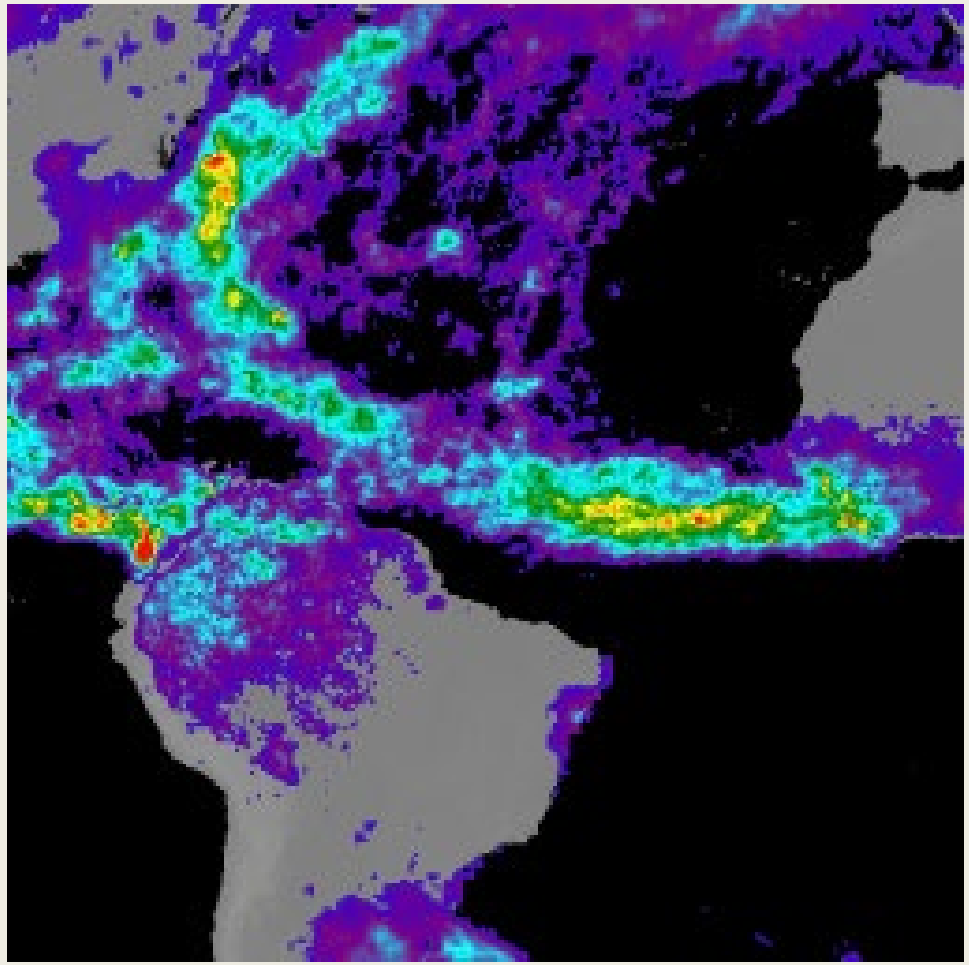
- Raster data is recorded in pixel (cell) units
- Each pixel is associated with 1 unit of information
- Often the format of satellite or photographic information



Source: ArcGIS

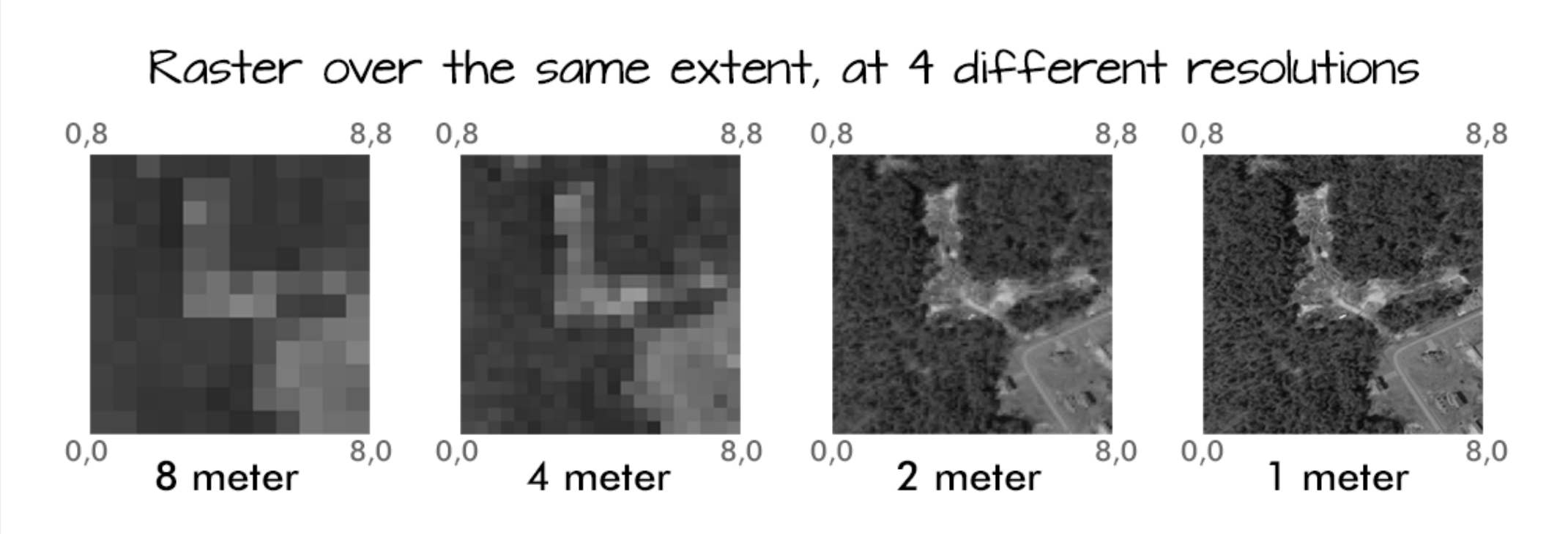
Raster

- Examples of raster files:
  - *Elevation, precipitation, land use, population estimates…*



Raster

Resolution = the pixel size (for example, 5km x 5km resolution)

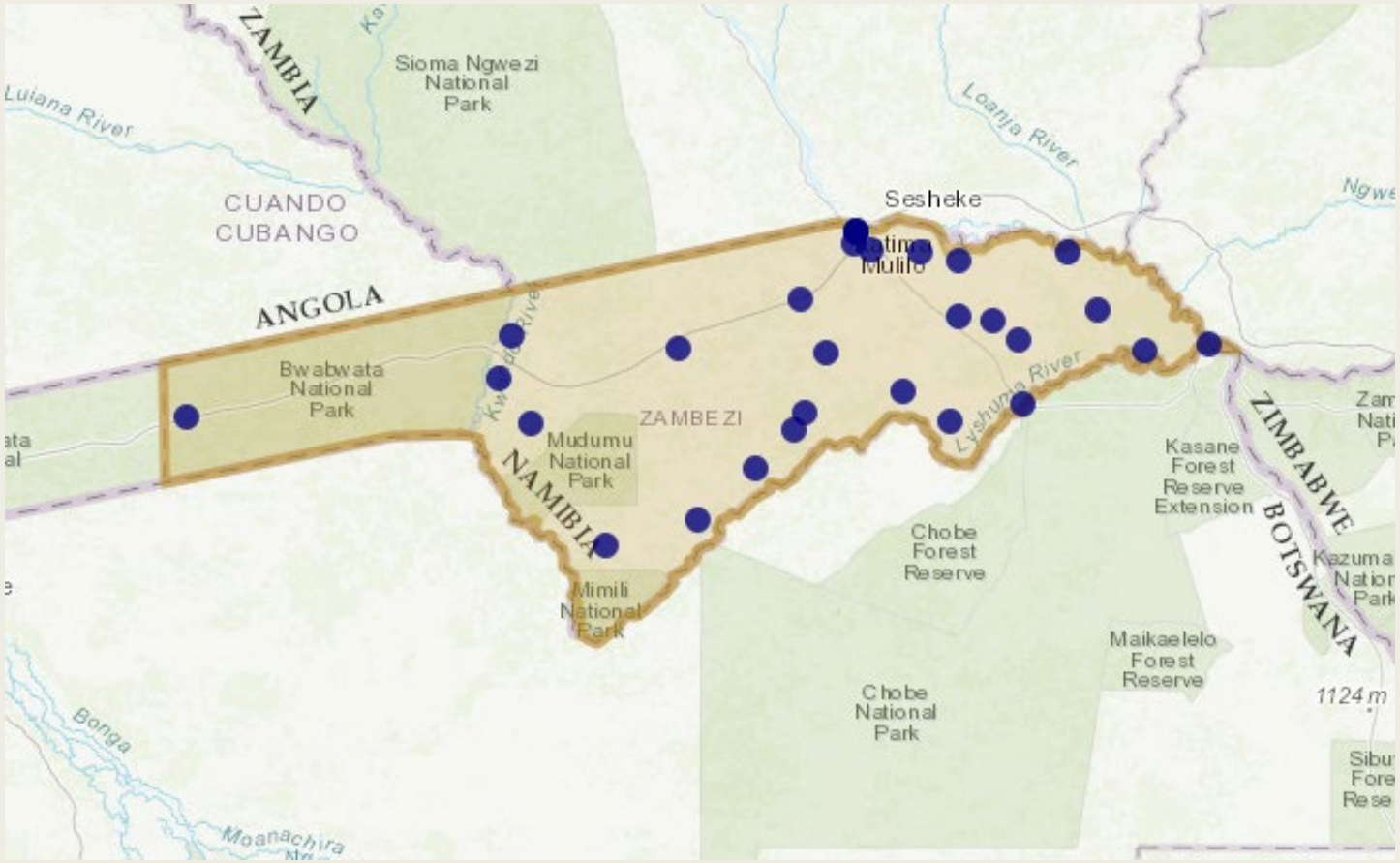


Source: National Ecological Observatory Network (NEON)

Vector

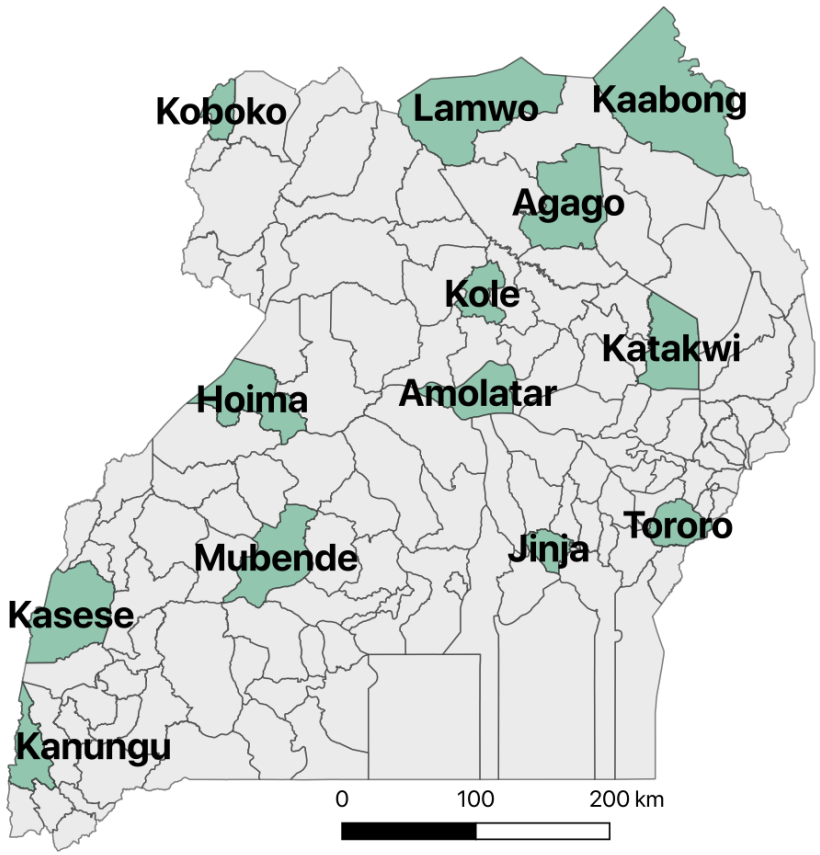
- Scalable data geometrically represented points, lines, or polygons
- For depicting boundaries, roads, regional areas, bodies of water, etc.

Vector: points



Vector: lines

Vector: polygons



How are vector files stored?

Shapefiles

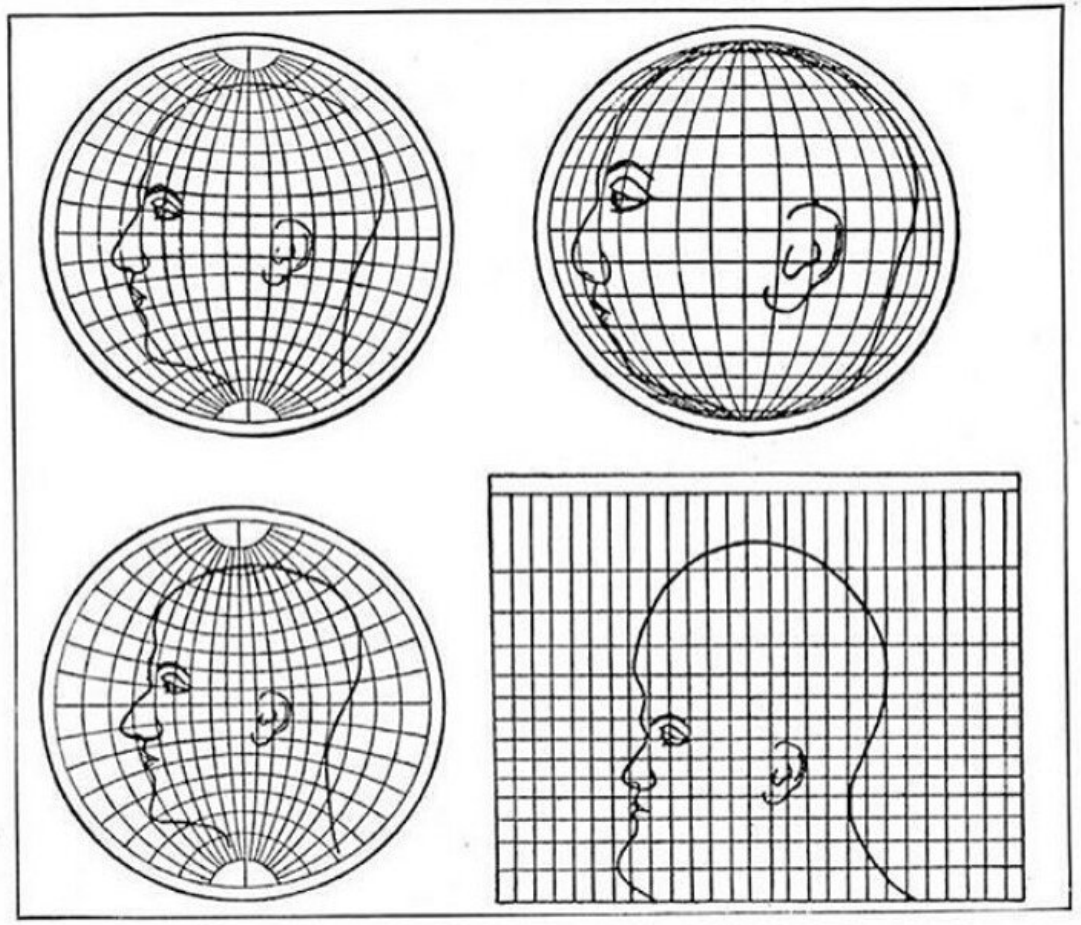
Shapefiles

- Comprised of the following files:
  - *.shp – the main file that stores geometry*
  - *.dbf –stores the attribute information of features*
  - *.prj – stores coordinate system information (more on this later)*
  - *.shx – the index file that stores the index feature of the geometry*

Name	Date Modified	Size	Kind
villages_aduku.dbf	Jul 19, 2019 at 12:58 PM	129 KB	Document
villages_aduku.prj	Jul 19, 2019 at 12:58 PM	143 bytes	Document
villages_aduku.shp	Jul 19, 2019 at 12:58 PM	99 KB	ESRI S...cument
villages_aduku.shx	Jul 19, 2019 at 12:58 PM	1 KB	Document

Coordinate reference systems

- To make a map, you must convert 3-dimensional earth (latitude and longitude) into a 2-dimensional map!
- This will create distortion no matter what



Upper left: Globular. *Upper right:* Orthographic. *Lower left:* Stereographic. *Lower right:* Mercator

What four commonly used projections do, as shown on a human head

Projections

- A very common projection is WGS 84
- WGS 84 is referred to as **EPSG:4326**
- We won't go too far into projections today, but **make sure all of the layers in your map use the same projection!!!**
- This source is useful for more information: https://www.earthdatascience.org/courses/earth-analytics/spatial-data-r/intro-to-coordinate-reference-systems/

SPATIAL DATA IN CVFS

Neighborhoods are stored as **points** (centroids)

nx	ny	neighid
[redacted]	[redacted]	001
[redacted]	[redacted]	002
[redacted]	[redacted]	003
[redacted]	[redacted]	004
[redacted]	[redacted]	005
[redacted]	[redacted]	006
[redacted]	[redacted]	007
[redacted]	[redacted]	008
[redacted]	[redacted]	009

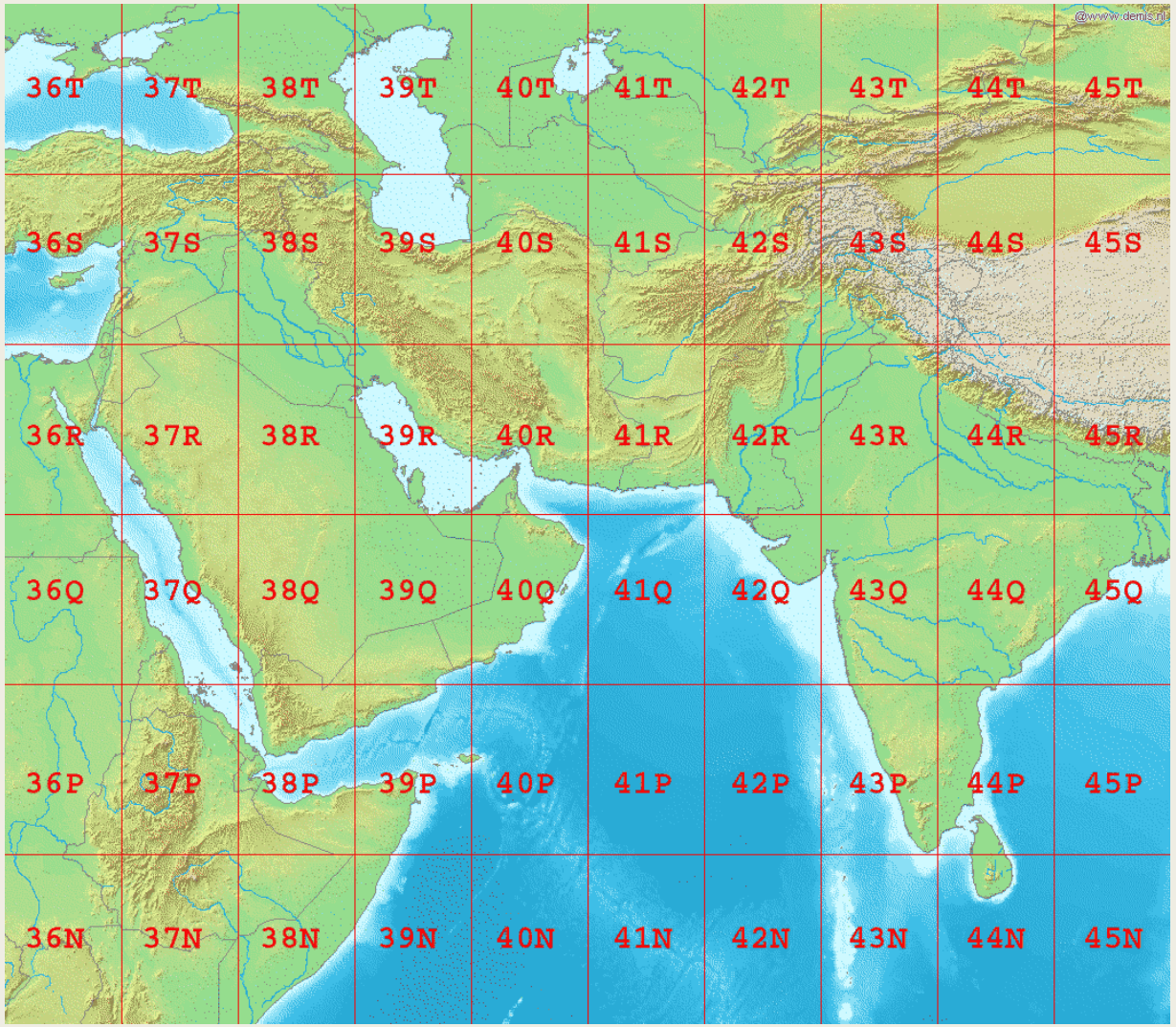
nx = latitude (UTM)

ny = longitude (UTM)

Note: We are not showing exact GIS coordinates of sampled CVFS neighborhoods to protect respondent privacy and confidentiality.

Neighborhoods are stored as **points** (centroids)

- Stored as Universal Transverse Mercator (UTM) projection
- Zone 45



Extract the value of a raster in a neighborhood

- Example: modeled raster surface of prevalence of houses with access to improved water source (from 2016 Demographic and Health Survey)

Source: https://spatialdata.dhsprogram.com/modeled-surfaces/#survey=NP|2016|DHS

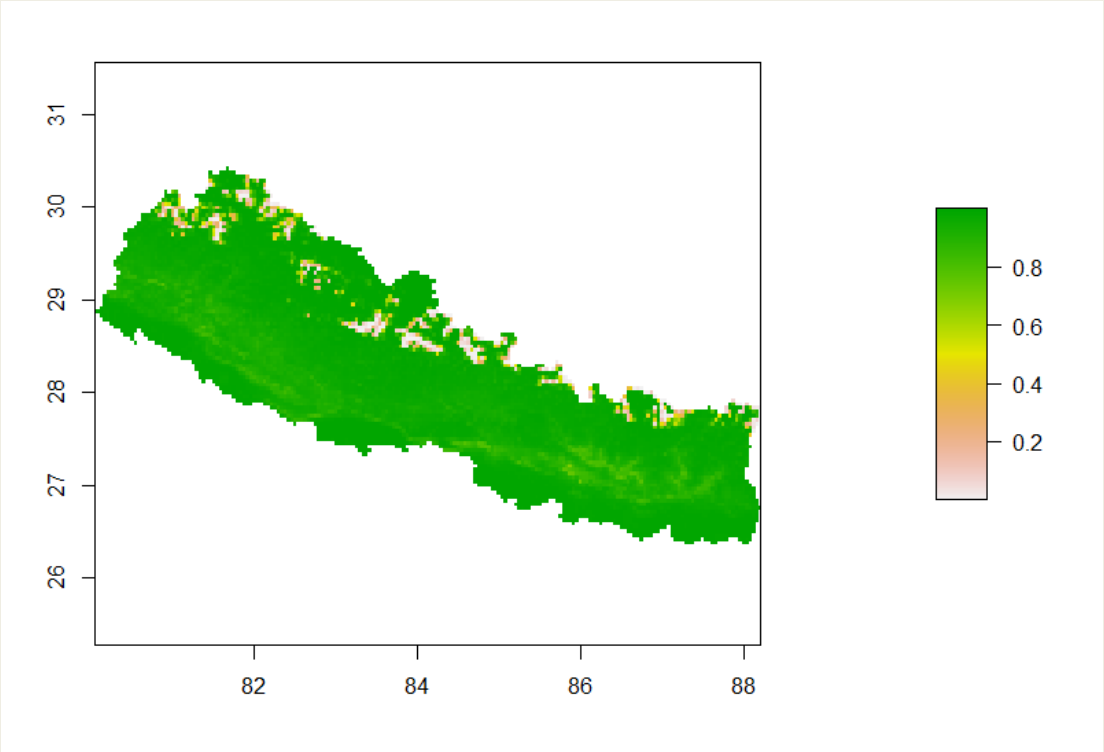
Extract the value of a raster in a neighborhood

```
xy <- read.dta13("U:/SPE-Data-Limited/Drought-Migration/Data/neigh_xycoord.dta")
summary(xy)

#Coordinates are in UTM, need to be converted to lat long
utmcoor <- SpatialPointsDataFrame(cbind(xy$nx,xy$ny), proj4string=CRS("+proj=utm +zone=45"), data = xy)

neighb<-spTransform(utmcoor,CRS("+proj=longlat"))

imprv_water <- raster("U:/SPE-Data-Limited/Drought-Migration/NP2016DHS_WSSRCEPIMP_MS_MEAN_v01.tif")
plot(imprv_water)

crs(imprv_water) ##same as neighb

neighb$imprv_water <- raster::extract(imprv_water, neighb)

summary(neighb$imprv_water)
View(neighb@data)
```

Extract the value of a raster in a neighborhood

```
xy <- read.dta13("U:/SPE-Data-Limited/Drought-Migration/Data/neigh_xycoord.dta")
summary(xy)

#Coordinates are in UTM, need to be converted to lat long
utmcoor <- SpatialPointsDataFrame(cbind(xy$nx,xy$ny), proj4string=CRS("+proj=utm +zone=45"), data = xy)

neighb<-spTransform(utmcoor,CRS("+proj=longlat"))

imprv_water <- raster("U:/SPE-Data-Limited/Drought-Migration/NP2016DHS_WSSRCEPIMP_MS_MEAN_v01.tif")
plot(imprv_water)

crs(imprv_water) ##same as neighb

neighb$imprv_water <- raster::extract(imprv_water, neighb)

summary(neighb$imprv_water)
View(neighb@data)
```

Extract the value of a raster in a neighborhood



Extract the value of a raster in a neighborhood

```
xy <- read.dta13("U:/SPE-Data-Limited/Drought-Migration/Data/neigh_xycoord.dta")
summary(xy)

#Coordinates are in UTM, need to be converted to lat long
utmcoor <- SpatialPointsDataFrame(cbind(xy$nx,xy$ny), proj4string=CRS("+proj=utm +zone=45"), data = xy)

neighb<-spTransform(utmcoor,CRS("+proj=longlat"))

imprv_water <- raster("U:/SPE-Data-Limited/Drought-Migration/NP2016DHS_WSSRCEPIMP_MS_MEAN_v01.tif")
plot(imprv_water)

crs(imprv_water) ##same as neighb

neighb$imprv_water <- raster::extract(imprv_water, neighb)

summary(neighb$imprv_water)
View(neighb@data)
```

nx	ny	neighid	imprv_water
[illegible]	[illegible]	001	0.9891113
[illegible]	[illegible]	002	0.9891113
[illegible]	[illegible]	003	0.9891113
[illegible]	[illegible]	004	0.9891113
[illegible]	[illegible]	005	0.9919132
[illegible]	[illegible]	006	0.9891113
[illegible]	[illegible]	007	0.9919132
[illegible]	[illegible]	008	0.9940205
[illegible]	[illegible]	009	0.9891113
[illegible]	[illegible]	010	0.9891113
[illegible]	[illegible]	011	0.9891113
[illegible]	[illegible]	012	0.9891113
[illegible]	[illegible]	013	0.9926102
[illegible]	[illegible]	014	0.9926102
[illegible]	[illegible]	015	0.9926102
[illegible]	[illegible]	016	0.9926102
[illegible]	[illegible]	017	0.9910276

Case study: drought and migration

"Drought and migration: An analysis of the effects of climate change on temporary labor migration and return migration from a migrant-sending area." Presented at Population Association of America Annual Meeting, 2021.

Aim: to evaluate the association between drought status at the neighborhood level and out-migration

Steps

1. Download monthly precipitation data from CHIRPS
2. Generate monthly drought variable

   How much did drought in the previous year deviate from rain in the prior 29 years?
3. Extract drought status at each neighborhood for each month of observation
4. Merge monthly drought/neighborhood information with household registry data

1. Download monthly precipitation data from CHIRPS

```r
#Download CHIRPS from FTP
#Monthly raster datasets for globe

setwd("/Users/adrienne/Box Sync/UCSF Research Projects/Climate/Migration/CVFS/CHIRPS data/") #This is where the data will download

library(RCurl)
url <- "ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_monthly/tifs/"
filenames = getURL(url, ftp.use.epsv = FALSE, dirlistonly = TRUE)
filenames <- strsplit(filenames, "\n")
filenames <- unlist(filenames)
filenames

length(filenames) #This is the number of months, where 1 = January 1981
#You can choose the month you want to start with, and you don't have to go all the way
#til this month if you don't want
filenames <- filenames[301:432]

for (filename in filenames) {
  download.file(paste(url, filename, sep = ""), paste(getwd(), "/", filename,
                                                      sep = ""))
}
```

2. Generate monthly drought variable

- Defined as rainfall percentile deviation in past year relative to last 29 years
  - *Value of 0.5 = median level of rainfall*
  - *Drought is defined as < 0.3 (below the 30th percentile)*

- For every month, now have a rasters of **percentiles**

3. Extract drought status at each neighborhood for each month of observation

```
#Load in percentile rasters
file_loc <- "U:/SPE-Data-Limited/Drought-Migration/Monthly rainfall percentile rasters"
pxtile_rasters <- c()
all_percentiles_df <- c()
for(year in 2011:2018){
  for(month in 1:12){
    file <- paste("pxtile_", year, "_", month.abb[[month]], ".tif", sep = "")
    filename <- paste(file_loc, file, sep = "/")
    raster <- raster(filename)
    pxtile_rasters <- append(pxtile_rasters, raster)
    percentiles <- as.data.frame(raster::extract(raster, neighb))
    all_percentiles_df <- rbind(all_percentiles_df, percentiles)
  }
}
```

percentile	neighid	monthyear_date
0.2758621		2011-01-01
0.3103448		2011-02-01
0.3448276		2011-03-01
0.4482759		2011-04-01
0.5172414		2011-05-01
0.4482759		2011-06-01
0.3103448		2011-07-01
0.2758621		2011-08-01
0.2758621		2011-09-01
0.2758621		2011-10-01
0.2758621		2011-11-01
0.2758621		2011-12-01
0.2758621		2012-01-01
0.3103448		2012-02-01
0.2758621		2012-03-01
0.2413793		2012-04-01
0.2413793		2012-05-01
0.2413793		2012-06-01
0.2758621		2012-07-01

4. Merge monthly drought/neighborhood information with household registry data

- Now have on drought status for each individual at each month

- Were people more likely to migrate when their neighborhood experienced drought? Yes!

*aOR for internal drought-out migration association: **1.24**, 95% CI 1.11-1.40*

*aOR for international drought-out migration association: **1.16**, 95% CI 1.05-1.28*

EMAIL WITH QUESTIONS OR IF YOU'D LIKE SAMPLE R CODE

adrienne.epstein@ucsf.edu