Biodiversity data analysis workshop – Day 2

Dominic Henry & Lizanne Roxburgh

Conservation Science Unit

dominich@ewt.org.za

lizanner@ewt.org.za

ENDANGERED WILDLIFE TRUST
Protecting forever, together.

science & technology
Department: Science and Technology
REPUBLIC OF SOUTH AFRICA

NRF National Research Foundation

SANBI Biodiversity for Life
South African National Biodiversity Institute

31st October 2019

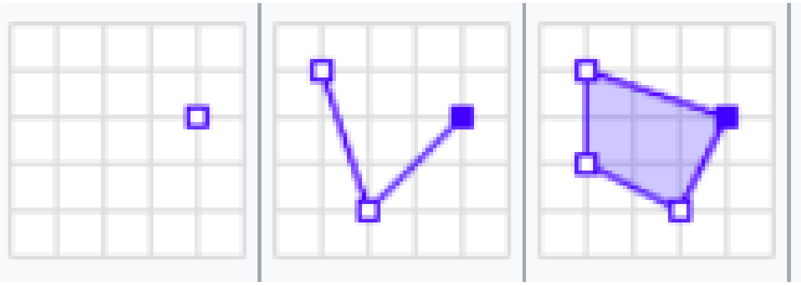
Programme – Day 2

08h30 – 10h00	Session 1 – Introduction to R, RStudio, basics of programming
10h00 – 10h30	Tea break
10h30 – 12h15	Session 2 – Data wrangling with the tidyverse
12h15 – 13h30	Lunch
13h30 – 15h00	Session 3 – Data visualisation using ggplot2
15h00 – 15h30	Tea break
15h30 – 17h00	Session 4 – Handling spatial data in R

Spatial data types

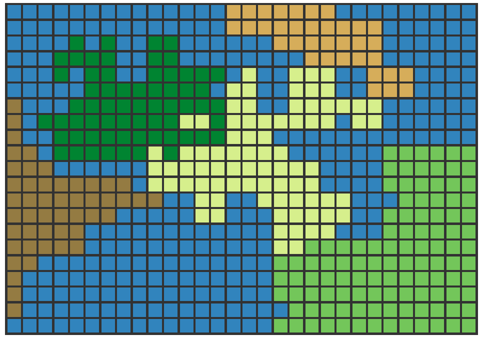
Vector

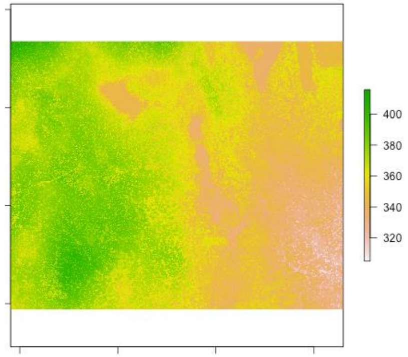
- Points, vertices and paths, polygons



Raster

- A grid with uniformly sized pixels
- Data are discrete or continuous





Vector data

Points

- X and Y, easting and northing, or longitude and latitude
- Within a spatial reference frame
- e.g., species occurrence observations

Lines

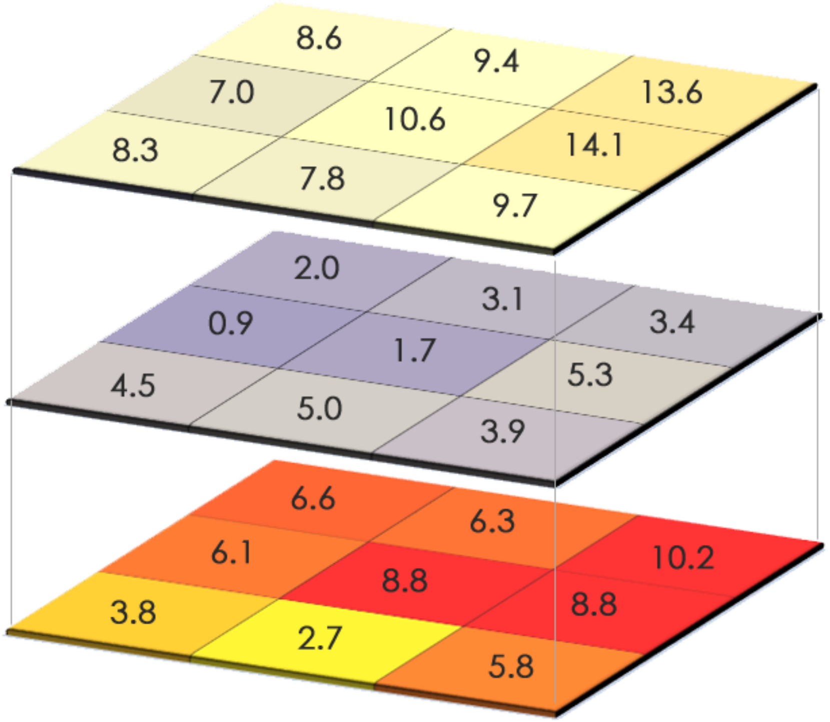
- Connect each vertex (point where two lines meet) with paths
- e.g., rivers, roads, railway lines

Polygons

- Closed path of connected vertices
- Provides a measure of area
- e.g., habitat patch, lake

Raster data

- Made up of pixels (grid cells)
- Each pixel has its own value
- Continuous surface
  - Elevation
  - Rainfall
  - NDVI
  - Remotely sensed (satellite) data
- Discrete (categorical) surface
  - Land cover
  - Vegetation type
- Ability to stack raster surfaces and perform map algebra



8.6	9.4	13.6
7.0	10.6	14.1
8.3	7.8	9.7

2.0	3.1	3.4
0.9	1.7	5.3
4.5	5.0	3.9

6.6	6.3	10.2
6.1	8.8	8.8
3.8	2.7	5.8

Coordinate reference systems (CRS)

ENDANGERED WILDLIFE TRUST

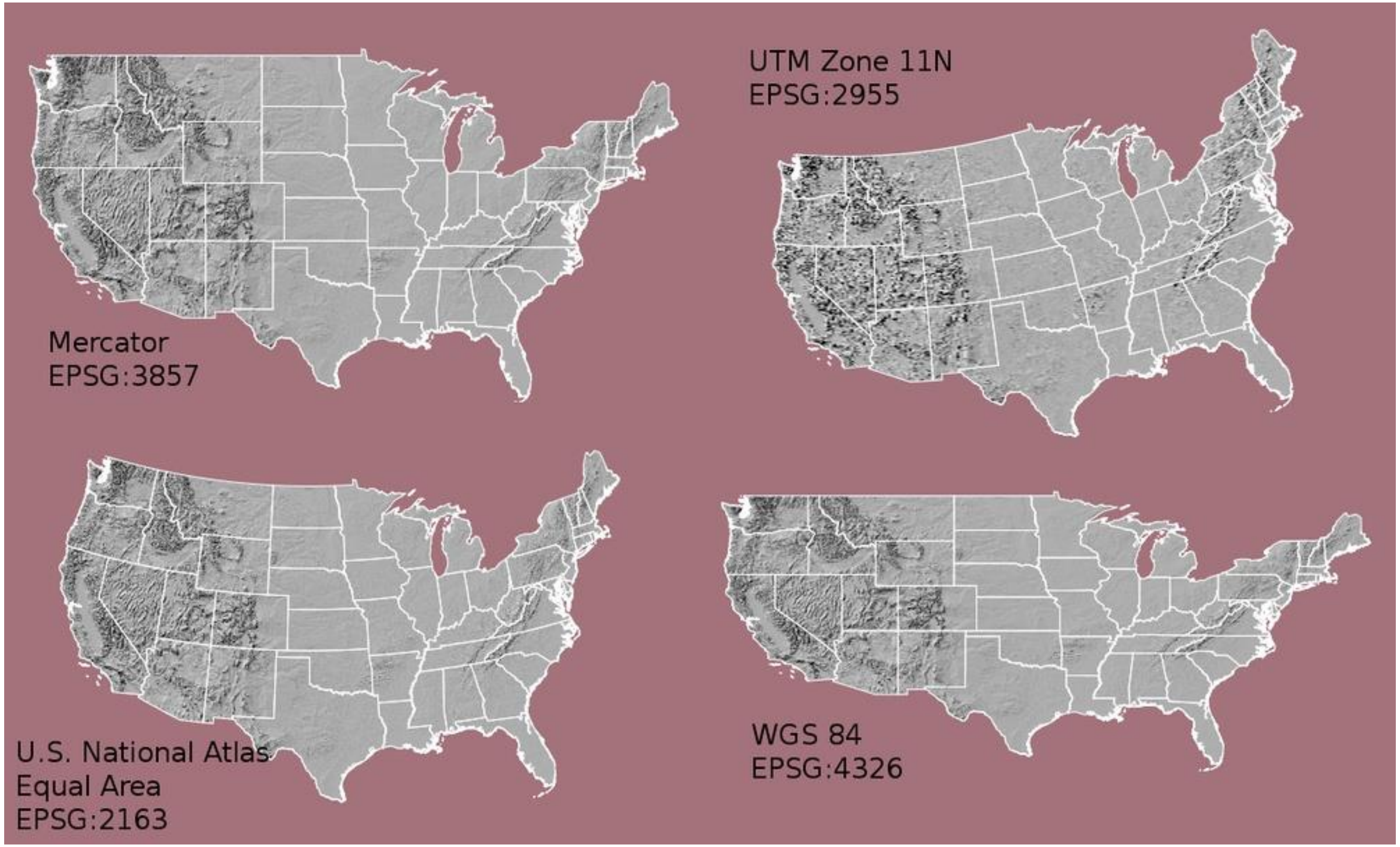
Map projections

- Portray the earth's surface (or part thereof) onto a flat plane
- Spherical shape (3D) to a planar shape (2D)

CRS

- Defines how real places on earth relate to 2D projected map (using a mathematical model)
- Type depends on regional extent and global position of data
- Necessary to calculate distances and areas accurately
- Tells mapping software where the data are and what method should be used to project the data in geographic space

Coordinate reference systems (CRS)



Components of CRS

Datum

- Model of the shape of the earth (angular units where starting point [0,0] is defined so that angles can reference meaningful spot)
- Most common WGS84

Projection

- Mathematical transformation of angular measurements
- Units are usually meters or feet

Additional parameters

- Definition of the center of the map and others

Components of CRS

Datum

Projection

Describing CRS

Several formats

- EPSG
- PROJ.4 (widely used in R, but *sf* is moving towards using EPSG)

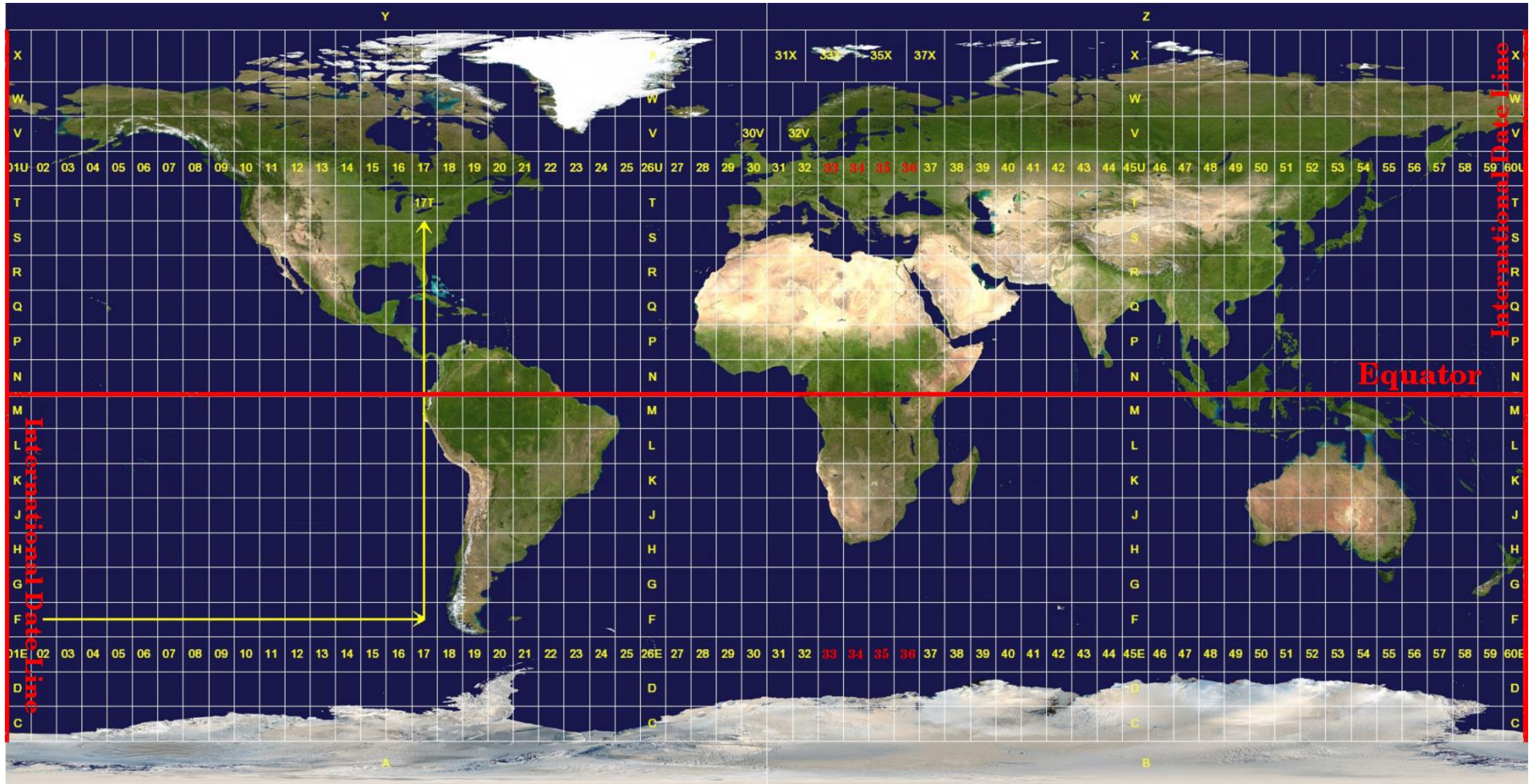
A PROJ4 string includes the following information:

- **proj=**: the projection of the data
- **zone=**: the zone of the data (unique UTM projection)
- **datum**=: the datum use
- **units**=: the units for the coordinates of the data
- **ellps**=: the ellipsoid (how the earth's roundness is calculated) for the data

Comprehensive library - https://spatialreference.org and https://proj.org/operations/projections/index.html

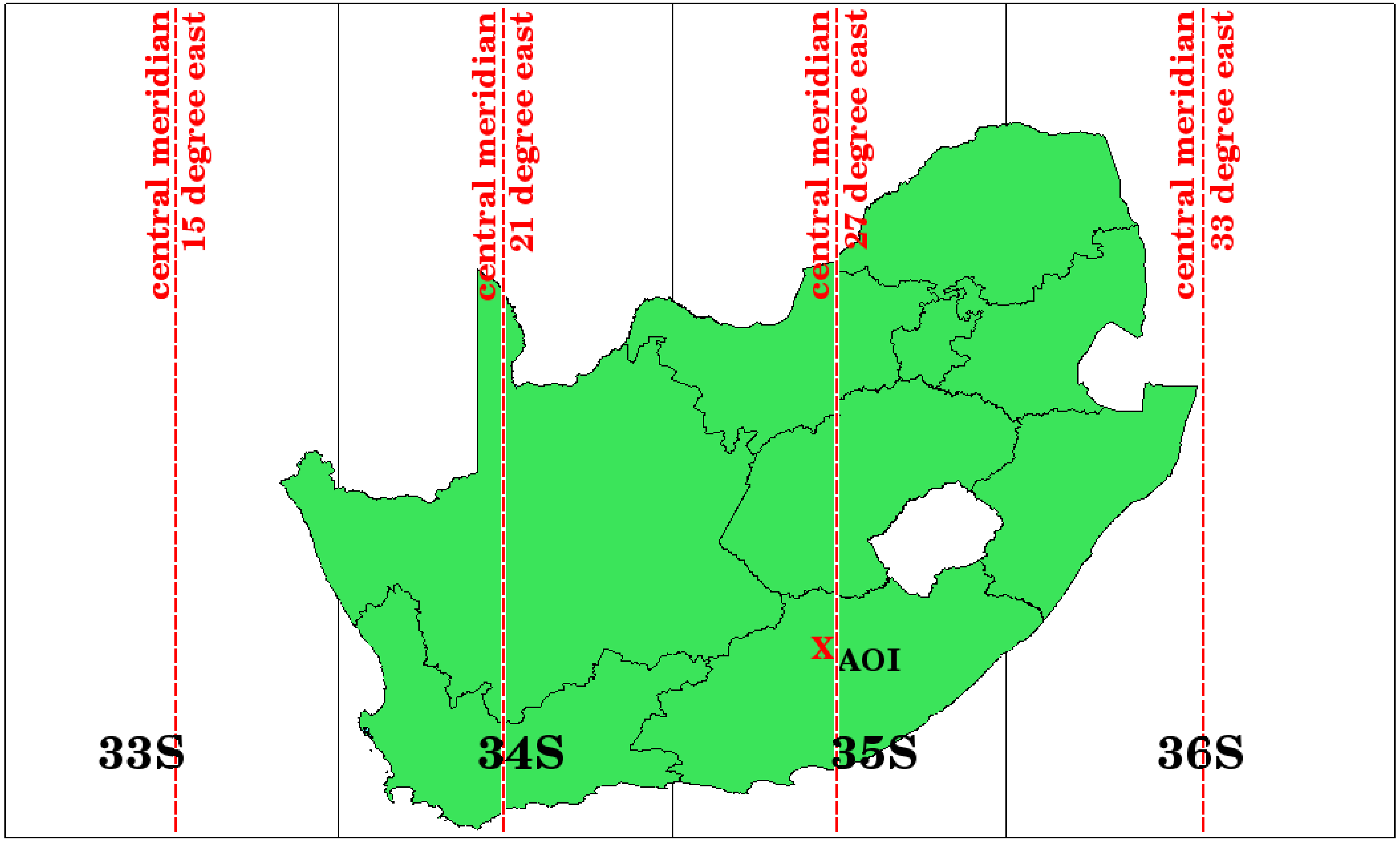
Universal Transverse Mercator (UTM)

UTM is a global map projection and to avoid distortion the world is divided into 60 equal zones (6 degrees wide)


Y
Z
X
31X
33X
35X
37X
W
V
30V
32V
01U 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26U 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45U 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60U
17T
T
S
R
Q
P
N
Equator
M
L
K
J
H
G
F
01E 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26E 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45E 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60E
D
C
A
B
International Date Line
International Date Line

Universal Transverse Mercator (UTM)

ENDANGERED WILDLIFE TRUST



Proj4 example

Proj4 demo link - LSM

Vector data with the *sf* package

Simple features

- Geometry describing where they are
- May also have additional attributes which describe properties of the geometries

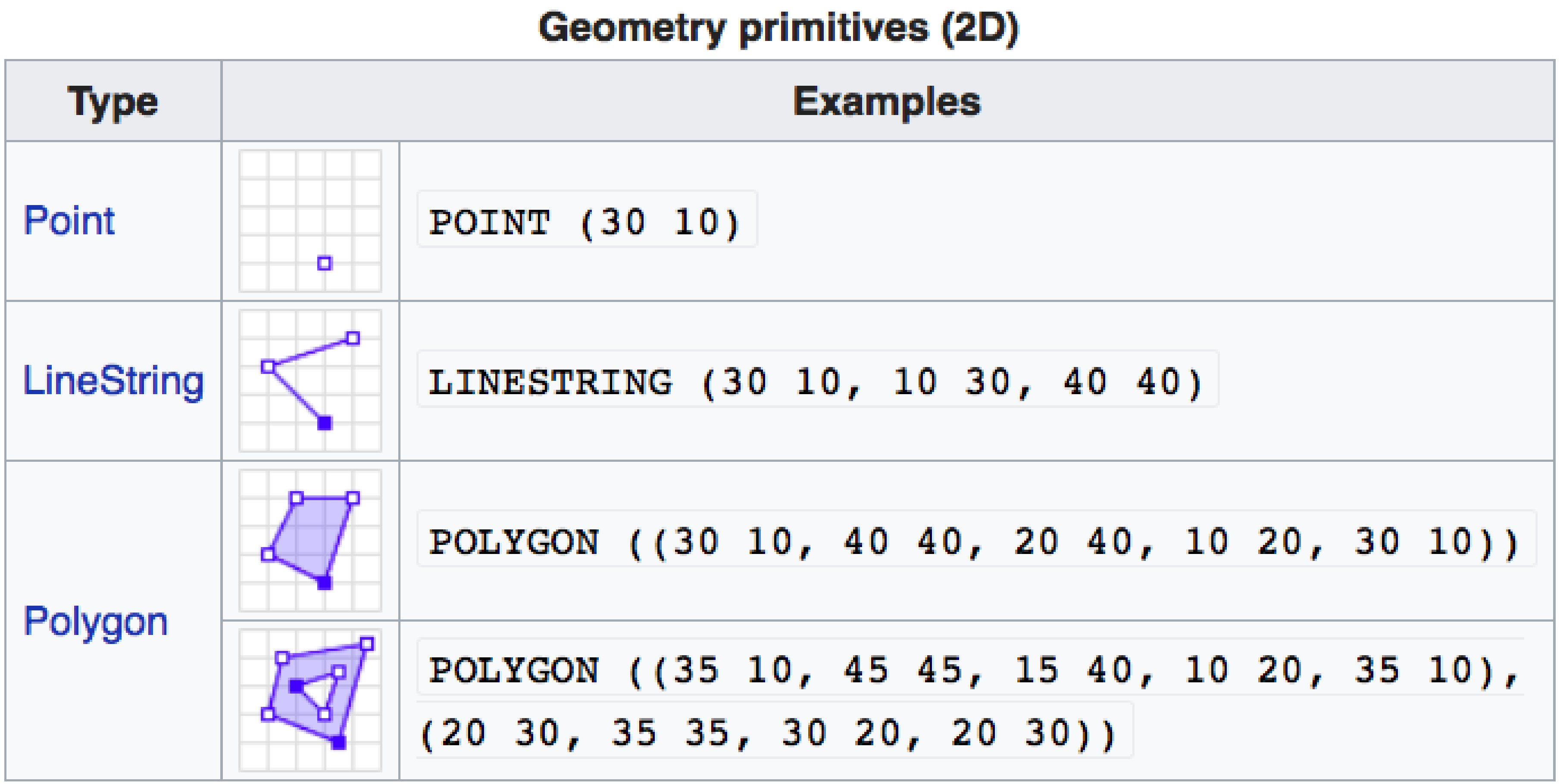
Simple point feature

- Geometry: XY coordinate
- Attributes: species name, date, collector

Simple feature geometry types

Type	Description
POINT	zero-dimensional geometry containing a single point
LINESTRING	sequence of points connected by straight, non-self intersecting line pieces; one-dimensional geometry
POLYGON	geometry with a positive area (two-dimensional); sequence of points form a closed, non-self intersecting ring; the first ring denotes the exterior ring, zero or more subsequent rings denote holes in this exterior ring
MULTIPOINT	set of points; a MULTIPOINT is simple if no two Points in the MULTIPOINT are equal
MULTILINESTRING	set of linestrings
MULTIPOLYGON	set of polygons
GEOMETRYCOLLECTION	set of geometries of any type except GEOMETRYCOLLECTION

Simple feature geometry types

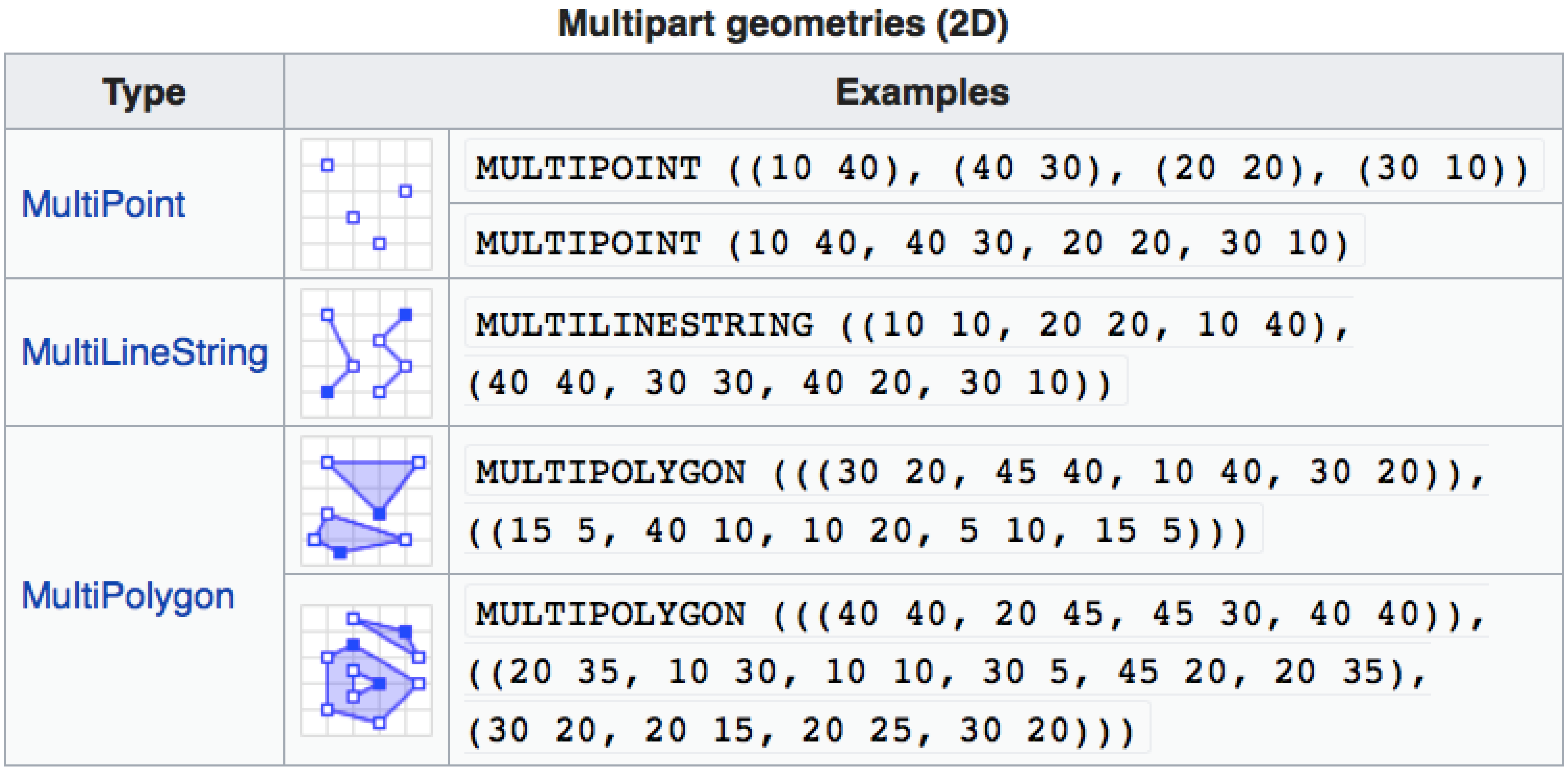


Geometry primitives (2D)

Type	Examples	
Point		`POINT (30 10)`
LineString		`LINESTRING (30 10, 10 30, 40 40)`
Polygon		`POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))`

		`POLYGON ((35 10, 45 45, 15 40, 10 20, 35 10),`
		`(20 30, 35 35, 30 20, 20 30))`

Simple feature geometry types



Multipart geometries (2D)

Type	Examples	
MultiPoint		MULTIPOINT ((10 40), (40 30), (20 20), (30 10))
		MULTIPOINT (10 40, 40 30, 20 20, 30 10)
MultiLineString		MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))
MultiPolygon		MULTIPOLYGON (((30 20, 45 40, 10 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5)))
		MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 10 30, 10 10, 30 5, 45 20, 20 35), (30 20, 20 15, 20 25, 30 20)))

sf package

- sf represents simple features as R objects

- All functions and methods in sf that operate on spatial data are prefixed by **`st_`**, which refers to *spatial and temporal*; this makes them easily findable by command-line completion (TAB key)

- Typical use involves reading, manipulating and writing of sets of features, with attributes and geometries.

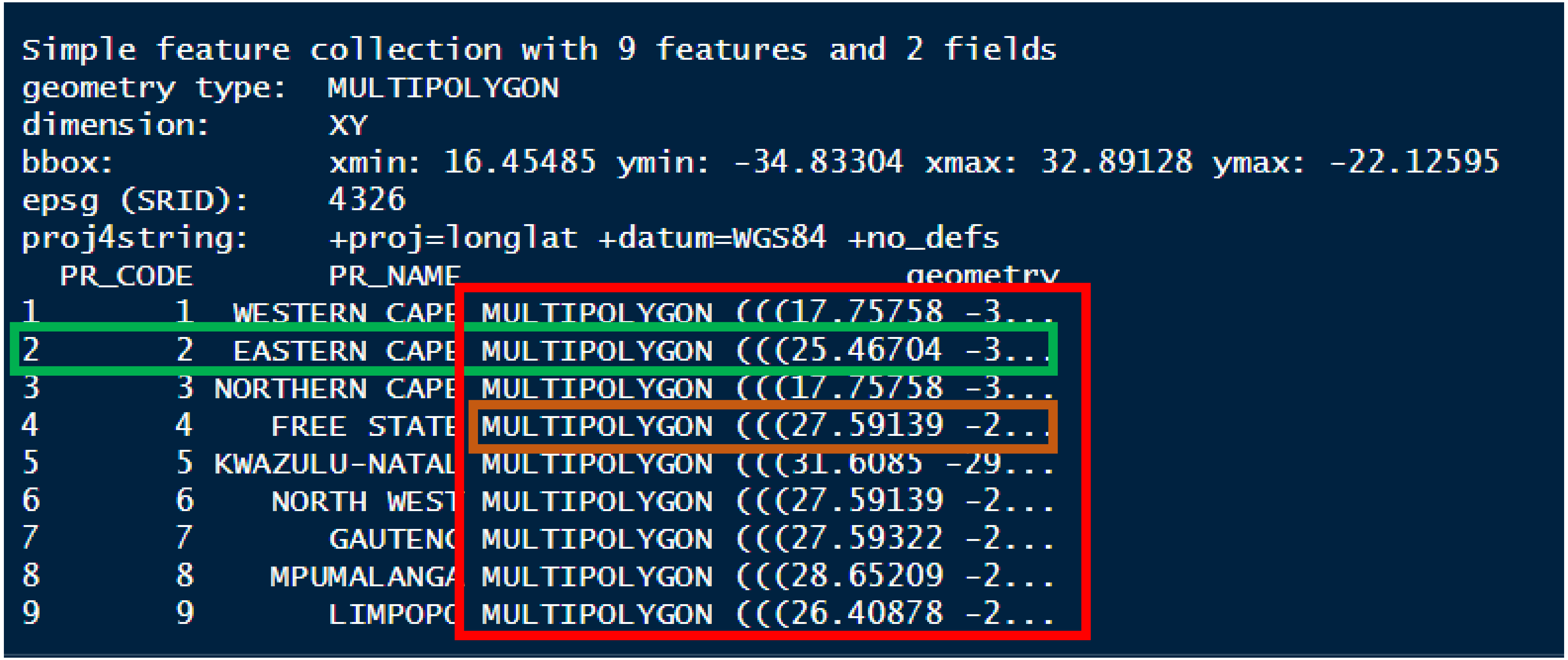
sf package

- Attributes are typically stored in **data.frame or tbl_df** objects
- Geometry feature are also stored in **data.frame** column.
- Geometries are not single-valued and are therefore in a list-column (length equal to number of records and each element holds simple feature geometry of that feature)

The three classes used to represent simple features are:

- **sf**, the table (data.frame) with feature attributes and feature geometries, which contains
- **sfc**, the list-column with the geometries for each feature (record), which is composed of
- **sfg**, the feature geometry of an individual simple feature.

sf classes



```
Simple feature collection with 9 features and 2 fields
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: 16.45485 ymin: -34.83304 xmax: 32.89128 ymax: -22.12595
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
  PR_CODE       PR_NAME                       geometry
1       1  WESTERN CAPE MULTIPOLYGON (((17.75758 -3...
2       2  EASTERN CAPE MULTIPOLYGON (((25.46704 -3...
3       3 NORTHERN CAPE MULTIPOLYGON (((17.75758 -3...
4       4    FREE STATE MULTIPOLYGON (((27.59139 -2...
5       5 KWAZULU-NATAL MULTIPOLYGON (((31.6085 -29...
6       6    NORTH WEST MULTIPOLYGON (((27.59139 -2...
7       7       GAUTENG MULTIPOLYGON (((27.59322 -2...
8       8    MPUMALANGA MULTIPOLYGON (((28.65209 -2...
9       9       LIMPOPO MULTIPOLYGON (((26.40878 -2...
```

Simple feature (sf)

Simple feature geometry list column (sfc)

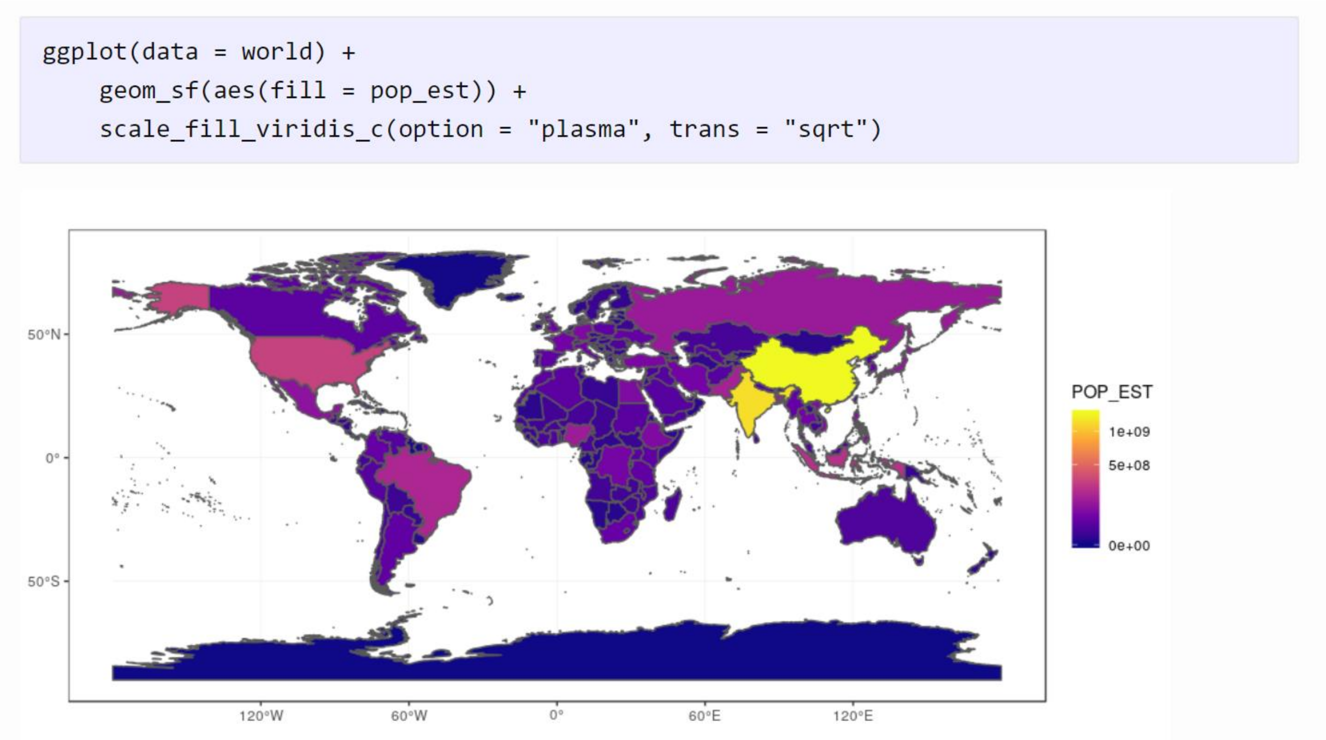
Simple feature geometry (sfg)

Reading and writing

```
read_sf(“input/South_Africa_prov.shp”)
write_sf(RSA, “outputsouth_africa.shp”)
```

Plotting sf objects

- Each attribute can be individually plotted
- Integration with ggplot2()



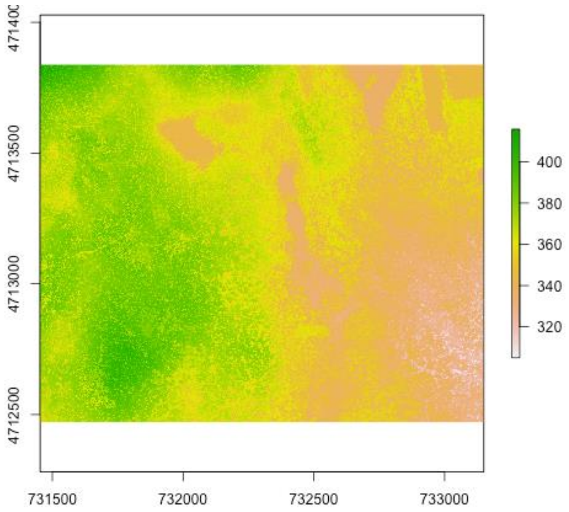
```
ggplot(data = world) +
    geom_sf(aes(fill = pop_est)) +
    scale_fill_viridis_c(option = "plasma", trans = "sqrt")
```

Rmd demo link - LSM

Rasters

Rasters have a regular structure and are defined by:

- coordinates of its origin
- a distance or cell size in each direction
- a dimension or numbers of cells in each direction
- an array of cell values



raster package

- Major extension of spatial data classes to access large rasters and process very large files
- Object classes include **RasterLayer**, **RasterStacks** and **RasterBricks**
- Simple functions for converting among these classes, and operators for computations on the raster data.
- Vector to raster conversions are also possible

Rasters

Rmd demo link – LSM

Rmd demo link – handling spatial data