**Spatial Data Summary**

Spatial data consists of positional information. Two spatial data models are vector and raster data model.

**Vector Data**-

The vector data types are points, lines and polygons. Geometry of these data structures consist of set of coordinate pairs (x, y). Points are simplest case with one coordinate pair and many associated variables. Lines are represented as ordered set of coordinates representing river, tributaries for example in spatial analysis. Polygon is a set of closed polylines. Geometry of polygon is similar to lines but the last coordinate pair coincides with first pair.

**Raster Data**-

Raster data represents spatially continuous phenomena like elevation. Raster divides world into a grid of equally sized cells. Geometry is stored as spatial resolution which is the size of each cell in grid.

In R important packages that define the spatial data structures are **sp** and **raster.**

**sp** package defines set of classes to represent spatial data. The basic types of vector data in sp package are SpatialPoints, SpatialLines, SpatialPolygons. These classes only represent geometry. To store attributes with geometry classes available are SpatialPointsDataFrame, SpatialLinesDataFrame, SpatialPolygonsDataframe.

sp package classes for raster data are SpatialGridDaataFrame and SpatialPixelDataFrame. Raster package classes for raster data are RasterLayer, RasterBrick, RasterStack.

**Coordinate Reference System**

Coordinate Reference System(CRS) of a location is longitude/latitude which is angular system associated with an ellipse model shape of earth. These angles are estimated using models similar to shape of earth called datums. Most commonly used datum is WGS84(World Geodesic System). A planar CRS is defined by a projection, datum, and a set of parameters.

In R, two main ways to describe CRS are an epsg code or a proj4string definition.

GRASS is a major open source for GIS developed as the Geographic Resources Analysis Support System provided in CRAN package rgrass7 or spgrass6.

**Reading and Writing of spatial data**

Shapefile is most commonly used file format to read and write vector data.

readOGR() and writeOGR() from rgdal library can be used to read and write vector data.

**Synthetic data**

synthACS package can be used to enable spatial microsimulation at any US geographic level taking advantage of reliable statistics on small area population from the ACS. Spatial Microsimulation is a technique that helps to understand individual characteristic based on aggregate statistics. synthACS is targeted for users interested in modelling and analyzing United States geographic data.

I was able to access 5 years bachelors data for California county from ACS macro data for specific USA geographic location.

Generation of synthetic micro data from ACS macro is memory intensive process.

**Point Pattern Analysis**

Point process is a random process to observe locations of some events of interest within bounded region.

Spatial, spatstat, splancs packages provide implementations and other methods for the analysis of different types of point process.

Basic test performed during point process analysis is Complete Spatial Randomness (CSR), which means that events are distributed independently at random and uniformly over the study area.

To measure the degree of accomplishment of CSR, several functions like G function (distance to the nearest point), F function (Distance from a point to nearest point) can be computed on data.

Currently trying to understand more about Statistical analysis of point process using Poisson process.

**Confusion and Doubts** –

1. For synthetic data generation how to plan for a scenario for what if analysis.
2. Fetching synthetic micro data from ACS macro data did not work on my system as it is memory intense.
3. ACS macro data gives information about population of specific US geographic area. If we still plan to consider spatial analysis for checking strength of the cell signal I am not sure how ACS data will help.