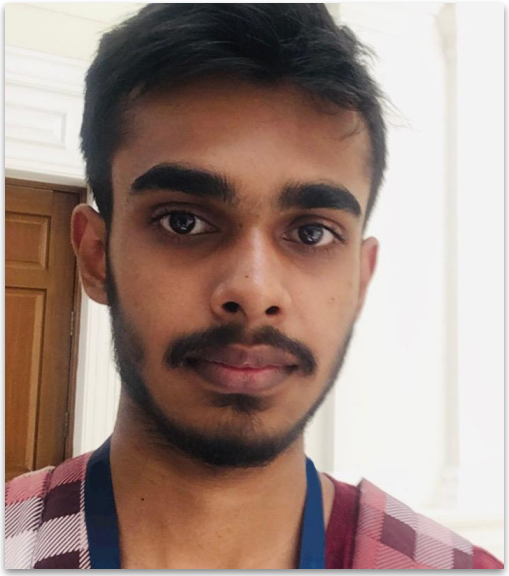
Analytics Vidhya

The DataHour on

Visualising Geospatial Data in Python using Geopandas

by Pareekshith Katti, Senior Data Scientist at ambee

Brief about me..



- My name is Pareekshith Katti
- Senior Data Scientist @ Ambee
- ~4 years of work experience
- Worked with Geospatial Time Series Environmental Data for almost 4 years
- Have a patent on Virtual Air Monitoring System
- Built India’s first and global pollen data product being used by multiple big clients around the world
- Hobbies - Music Production, Books, Comics, Manga and Anime

Table of contents

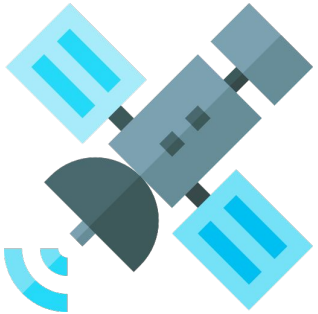
- Introduction
  - What is Geospatial Data?
  - Challenges with handling Geospatial Data
  - Quick overview of formats
  - Libraries for handling geospatial data
- Geospatial Data Visualization with Geopandas
  - Geometries, Centroids, Representative Points
  - Overview of plots - Plotting shapes, Choropleth Maps, Bubble plots, Cartograms, KDE Plots, Heatmaps
- Jupyter Notebook Walkthrough
- Q & A

Analytics Vidhya

Introduction

What is Geospatial Data?

- Any data which has a geographic location associated with it.
- Examples for geospatial data include satellite imagery and remote sensing data, maps, census data, gdp data, weather data etc.
- Geospatial Data is everywhere




Challenges in handling Geospatial Data

- Data formats for geospatial data is not mainstream
- Geospatial Data Science and Machine Learning is a relatively new field compared to Computer Vision or NLP
- Some geospatial data can be huge in size
- Not enough mainstream libraries
- Rapidly developing field that directly ties into very important issues such as climate change, air pollution, weather etc

Quick Overview of Geospatial Data Formats

- Common Data Formats
  - csv, parquet, feather, json, images
- Vector Data Formats
  - Shapefile (shp), GeoJSON, KML, WKT
- Satellite and Raster Data Formats
  - netCDF(nc/nc4), HDF(hdf/hdf5), GRIB, GeoTIFF, ASCII

Libraries

General Data Science Libraries

Geospatial Data Science Libraries

Rasterio

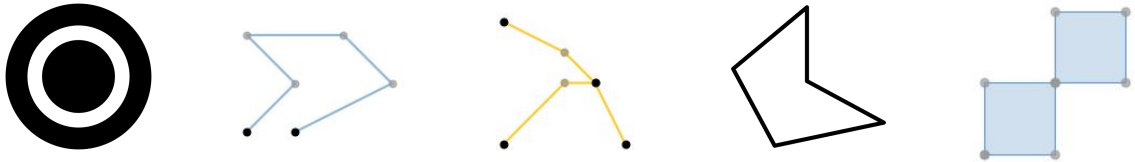
Shapely

Analytics Vidhya

Geospatial Data Visualization with Geopandas

Geometries

- Point, MultiPoint (latitude,longitude)
- LineString (Road/River)
- MultiLineString (River with branches)
- Polygon (Country with simple boundaries)
- MultiPolygon (Country with complex or disconnected boundaries)

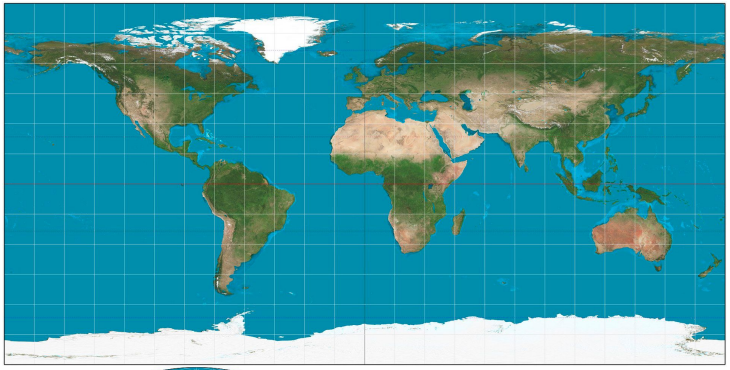


Coordinate Reference System (CRS)

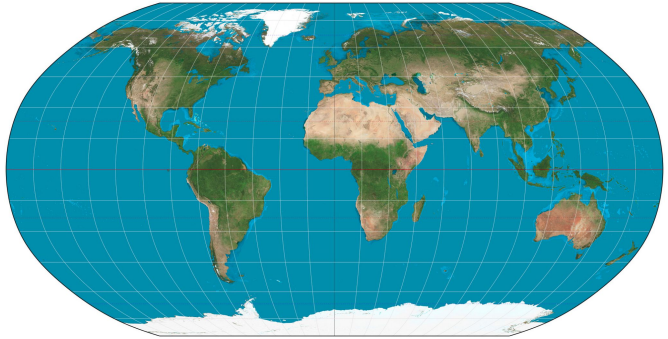
- Earth is spherical/geoid
- In order to make a 2d map, we need to project 3d globe to 2d
- In any projection, there is a distortion in either distance, direction, shape or area
- A coordinate reference system is a system of points/coordinates used to reference a location or geospatial entity in a projection
- The most used CRS is WGS84 (Also known as ESPG:4326) whose coordinates are referenced using latitude and longitude
- Another popular CRS is Web Mercator (ESPG:3857) which is used in web and mapping applications such as Google Maps, OSM, Mapbox etc.

Examples of CRS

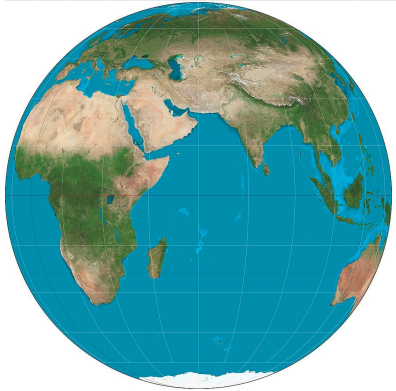
Web Mercator



WGS84



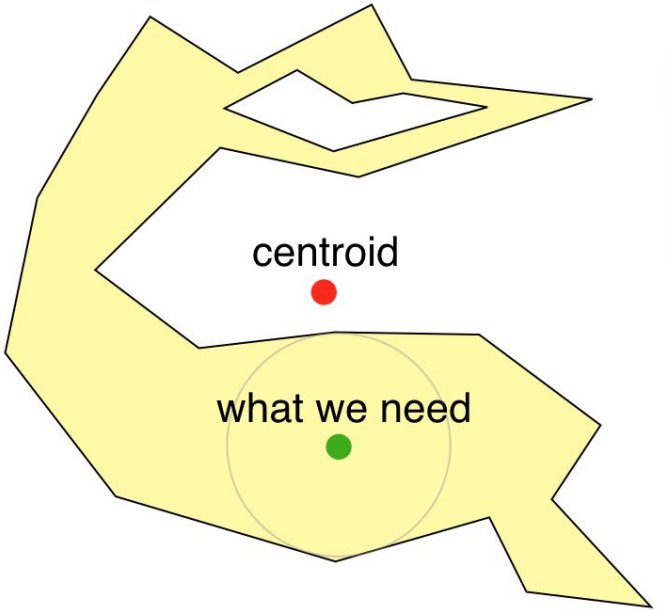
Robinson



Orthographic

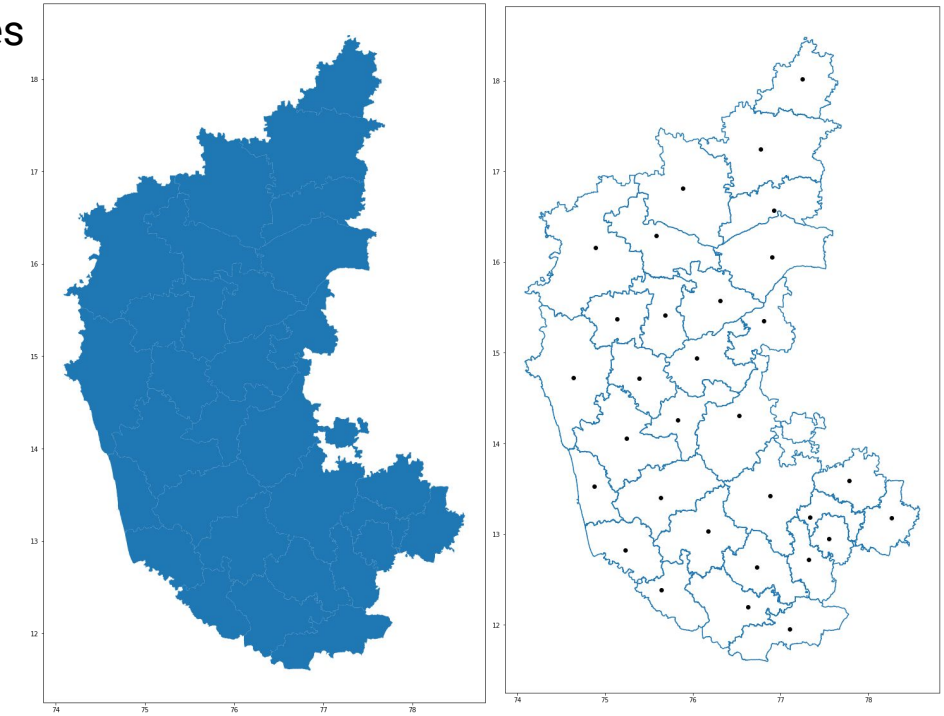
Centroid and Representative Point

- Centroid - Geometric Centre of a polygon/shape or mean of all points inside the shape.
- Depending on the shape, centroid can sometimes be outside the polygon
- Representative Point - A point representing the polygon that is guaranteed to be inside the polygon



Plotting Geometries

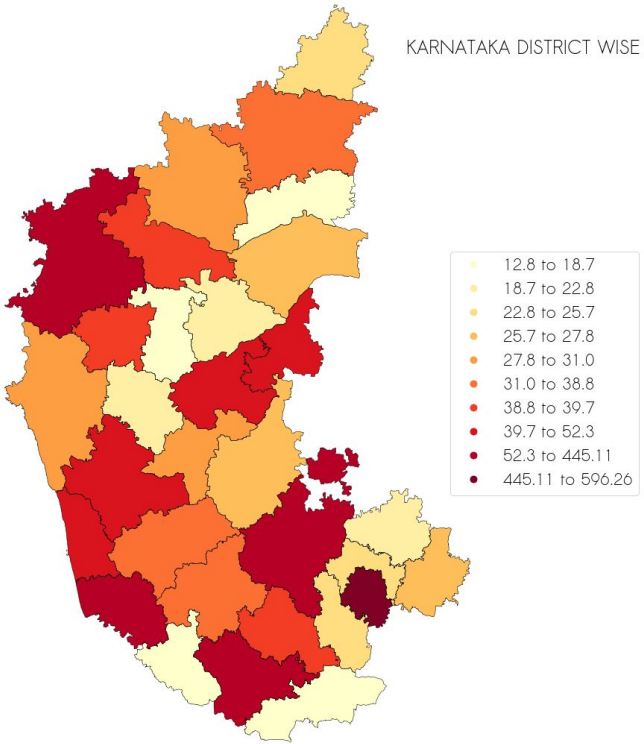
- Plot method of GeoDataFrame/GeoSeries
- Plotting Polygons - Poly Plot
- Plotting Points - Point Plot
- Plotting Lines - Geometric Line Plot
- Thematic Maps

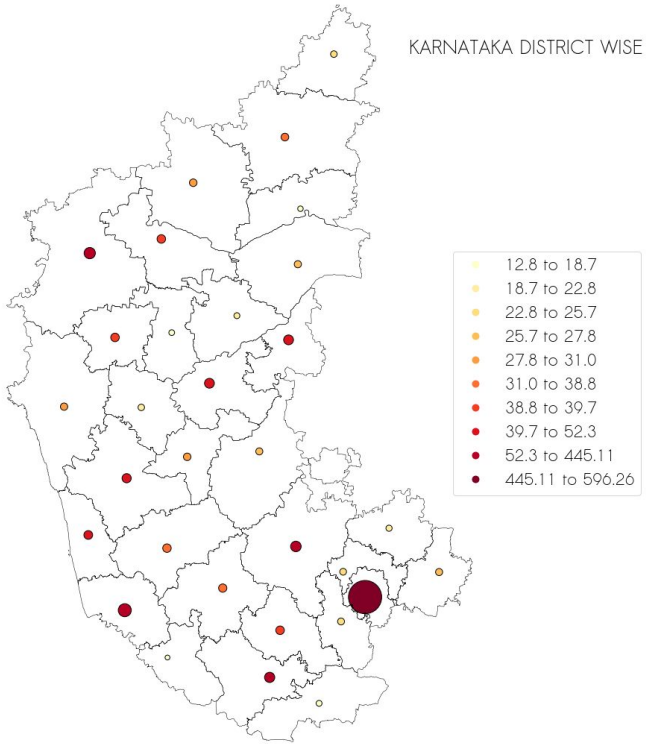


Representing Data on a Map - Thematic Maps, Density Maps

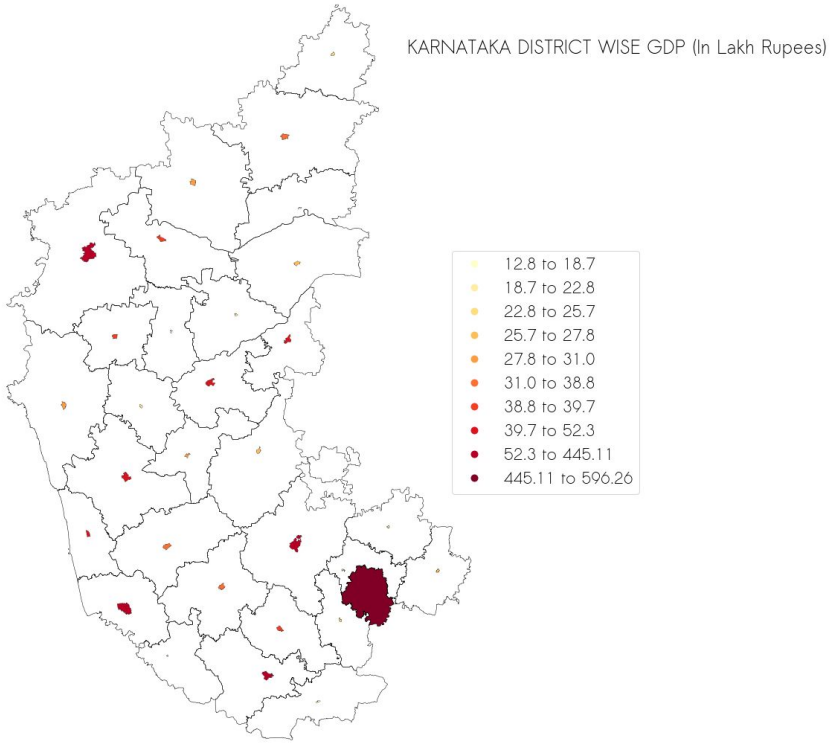
- **Choropleth Map** - Used to show magnitude of a statistic/numeric data across different geographies. Data is binned and a color is assigned for each bin. Example: Country Wise GDP
- **Bubble Plots** - Shows the difference in size along with the categorical bins
- **Cartogram** - Shows the difference in size by scaling geographies
- **Heatmap** - Used to show magnitude of a statistic/numeric data while ignoring geographical boundaries. Works well for datasets having large number of data points.
- **Density Map** - Used to know geographical distribution of data (KDE Plot)

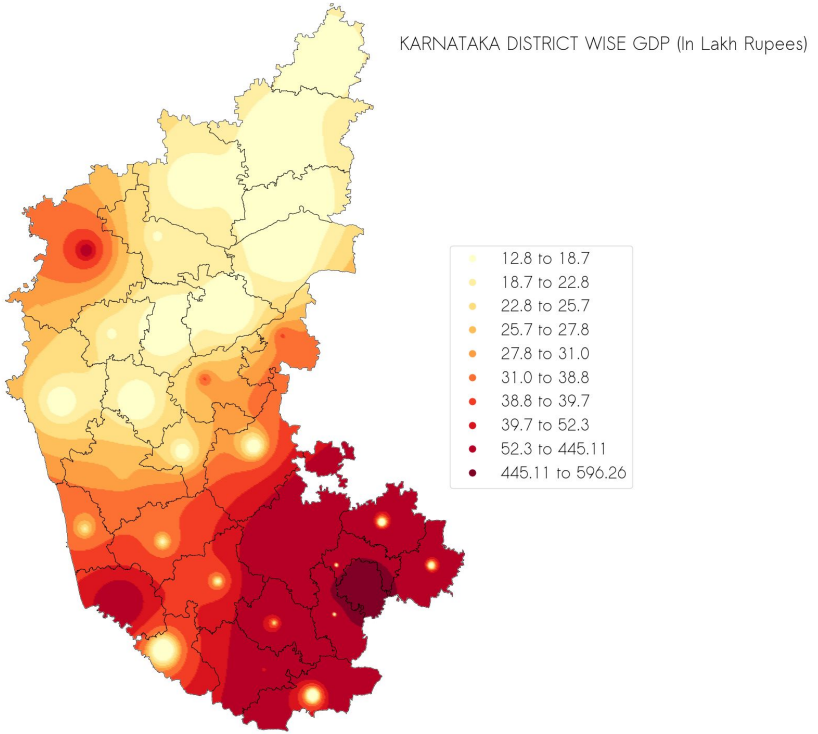
Choropleth and Bubble Maps





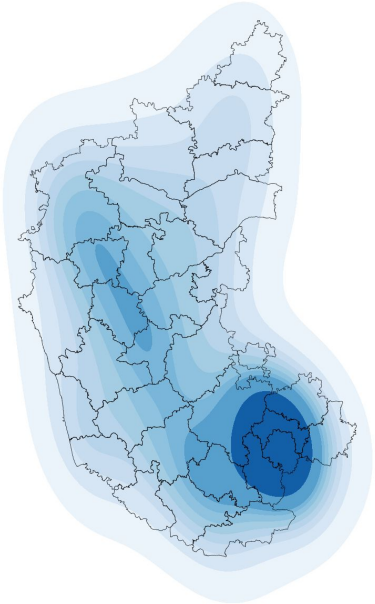
Cartogram and Heatmap



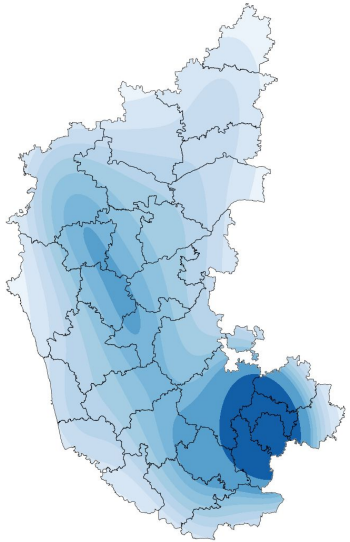


Density Maps

KARNATAKA LOKSABHA ASSEMBLIES



KARNATAKA LOKSABHA ASSEMBLIES



Analytics Vidhya

Jupyter Notebook Walkthrough

gspatial_plot

- Github - https://github.com/ambeelabs/gspatial_plot
- PyPi - https://pypi.org/project/gspatial-plot/
- Readthedocs - https://gspatial-plot.readthedocs.io

Conclusion

- You can find the datasets, presentation and code on my github repo
  - (Link)
- Connect with me on LinkedIn - https://in.linkedin.com/in/pareekshith-katti-ab180988
- Github - https://github.com/Paree24
- Icons used in this presentation are from https://www.flaticon.com

Analytics Vidhya

DataHour : Q&A

Analytics Vidhya

Thank you!