GEOG 498G

Matthew Tralka

Scientia potential est

From smartphones to agriculture, data surrounds every aspect of modern life on Earth. When identified and analyzed, big data has the applicational capacity to impact every facet of human existence. However, pulling actionable intelligence from the overwhelming swarm of data generated every day is a job often restricted to information science professionals. Data is all around us, often hiding the knowledge we need. The goal of this project is to open the door for anyone to use and analyze social media big data by creating an open source friendly all-inclusive web interface to process, gather, and distribute geospatial data.

Data Source

As per the objective of this project, the proposed program will not produce new geospatial intelligence. Rather, it will wrap, scrape, and repackage traditional API sources for dissemination to the general public. The initial evolutions of this program will aim to implement the API for Flickr, a popular social media photo sharing site. The Flickr API is RESTful and open for non-commercial development; therefore, it is an ideal pivot point for initial project operations. Following preliminary release, tertiary objectives are in place to integrate other open API sources such as Twitter or Facebook.

Data distribution to end-users will tentatively encompass the comma-separated values (CSV) and Geographic JavaScript Object Notation (GeoJSON) file formats. It is assumed that end-users with varying degrees of technological aptitude can easily identify and process at least one of the above file formats. Further review is underway to assess other feasible file distribution formats.

Technical Implementation

I will implement the technical side of the proposed project into two main segments - the ‘frontend’ and the ‘backend’. The frontend, the outwardly facing publicly accessible portion of the project, will be constructed in semantic and accessible HTML, CSS, and JavaScript. Additional frontend libraries will be imported with discretion and on an ‘as needed’ basis to reduce overall complexity and dependencies. The web interface will be constructed from the ground up with no large CSS helper-frameworks while abiding by modern web design standards in accessibility and responsiveness. I aim to implement the server-side ‘backend’ largely in Python using Flask, UWSGI (deployment), and other libraries as needed. Using Python, known for easy readability and low bar of access, will ensure the critical parts of program code are easily humanly readable. Python will abide by PEP8 conventions where appropriate. Database construction will be in SQLite with an SQLAlchemy ORM wrapper for easy interaction. Nginx is to handle reverse proxy and static file serving. Final deployment is designed for the financially-sound five-dollar Digital Ocean Virtual Private Server (VPS). Code will be made available on GitHub under the GNU GPL-v3 license modified for purpose.

Challenges

Challenges expected for this project lie with API interactions and web service scaling. The project is fully reliant on open APIs. At any point, companies such as Flickr may change terms of use or revoke API keys. Crafting the project in a legal manor that abides by such usage terms is essential. I have

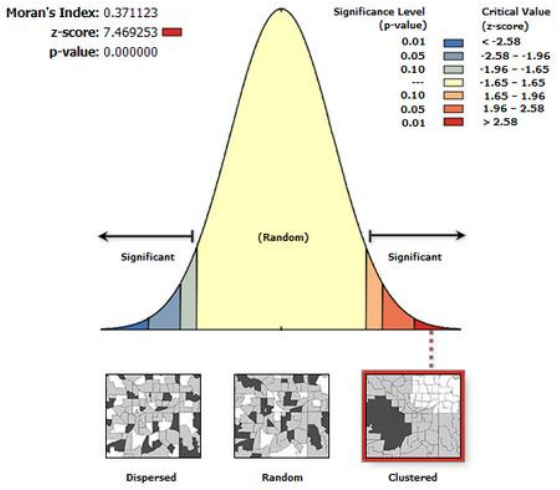
limited experience with such legalities and as such expect this portion of the project to be challenging. Additionally, building a local API wrapper is entirely different to deploying a multi-user thread-safe web application. I have some experience in this field; however, I have no formal training and always expect the worst. I aim to mitigate deployment challenges by building the app completely from the ground up for web deployment and by extensive case testing. Further issues not explicitly mentioned are expected and will be discussed upon discovery.

Implications

I believe the overall implications of this project are enormous. The limitation of big data within the upper echelon of the technological literate is inexcusably limiting. This project will allow anyone with a computer and a base level of internet competency the ability to gather geospatial data from all integrated API providers in one place. Overcoming this gap of equitable access has the potential to spark new research, foster engagement into the field of data science, and raise awareness for internet privacy and data security.

Exploratory Data Analysis (ESDA)

A program produced ESDA is outside the scope of this proposed project. However, I aim to demonstrate the capabilities of this program through featuring different data analysis methods and project ideas on the main web page. Below, I have included two figures representing a dataset acquired though a preliminary version of the above described Flickr API wrapper. Statistics used, Global Moran's I and Local Moran's I, display spatial clustering and outlier analysis on a Flickr photo dataset identical to the one the above proposed project will disseminate.



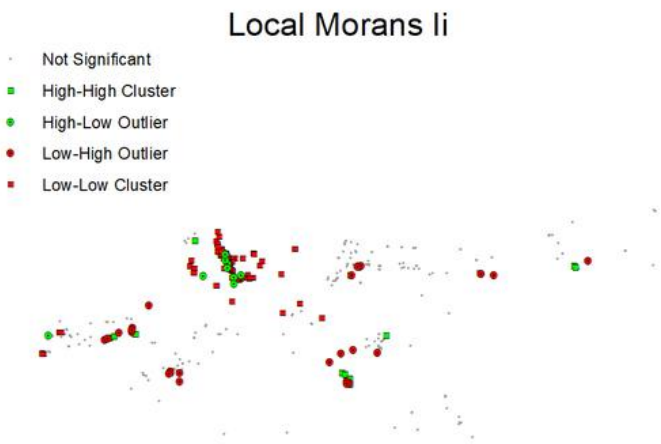


Figure 1: Global Moran's I **Figure 2:** Local Moran's I

Spatial clustering of Flickr photos taken above 1500m ASL and between 2015 - 2020 containing plants as general subjects in the Northern Ticino Alps, Switzerland. Data acquired through Flickr GeoSearch, classified through ClassyFi by Dr. B. Hale and M. Tralka