GIS for Real Estate

I. Overview

In this project, we use the Geographical Information System (GIS) platform for real estate listing and price estimation. The GIS have given researchers the ability to search electronically for property records, many via the web. In the meantime, spatial analysis gives the realtor, and ultimately the seller, the ability to more easily showcase the subject property's characteristics and advantages with respect to its location.

For price estimation, we build a model based on the geographically weighted regression using the estate's attributes and location as input.

The project also involves dealing with big data, screening 200+ features, processing geographic information processing, extracting real estate point coordinates using Google Map API and web application development.

II. GIS

A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographic data.

III. GIS Software

GIS software are normally programs with a graphical user interface that can be manipulated using the mouse and keyboard. The software provides menus near to the top of the window (File, Edit etc.) which, when clicked using the mouse, show a panel of actions. These actions provide a way for you to tell the GIS software what you want to do. There are various GIS software for use and we use ArcGIS in this project.

A. How data are stored in ArcGIS

Geographic data models serve as the foundation on which all geographic information systems are built. A geographic data model is a structure for organizing geospatial data so that it can be easily stored and retrieved. At its most basic level, an ArcGIS geodatabase is a collection of geographic datasets of various types held in a common file system folder, a

Microsoft Access database, or a multi-user relational DBMS (such as Oracle, Microsoft SQL Server, PostgreSQL, Informix, or IBM DB2).

Geodatabases come in many sizes, have varying numbers of users and can scale from small, single-user databases built on files up to larger workgroup, department, and enterprise geodatabases accessed by many users.

A key geodatabase concept is the dataset. It is the primary mechanism used to organize and use geographic information in ArcGIS. The geodatabase contains three primary dataset types: Feature classes, Raster datasets and Tables (Fig. 1) .

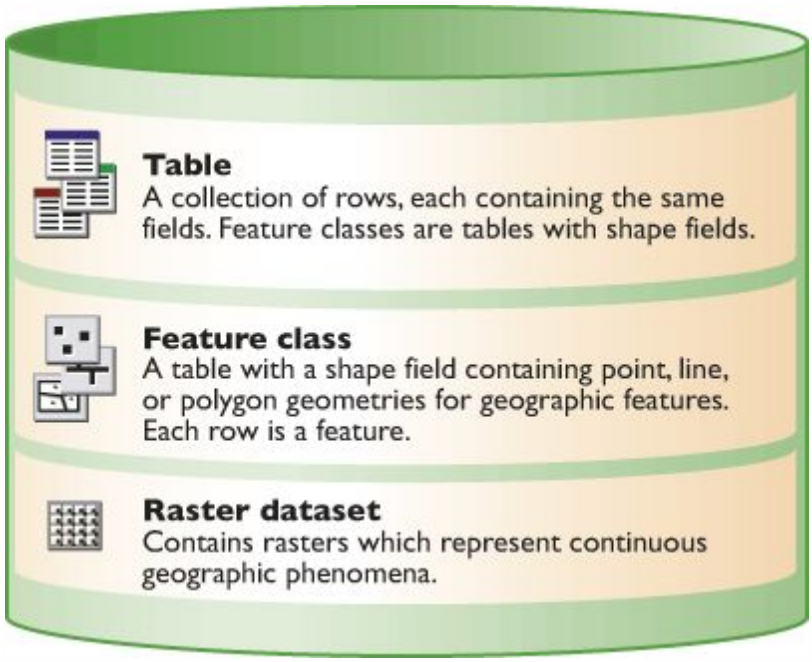

Figure 1: ArcGIS geodatabase

B. Components of ArcGIS

The main components of ArcGIS include ArcMap, ArcCatalog, and ArcToolbox.
ArcMap (Fig. 2) is the main interface for conducting analyses and creating maps. Here, feature classes and shapefiles can be populated, data can be edited, calculations can be performed, and finally, maps can be created for displaying the results of the GIS analysis.

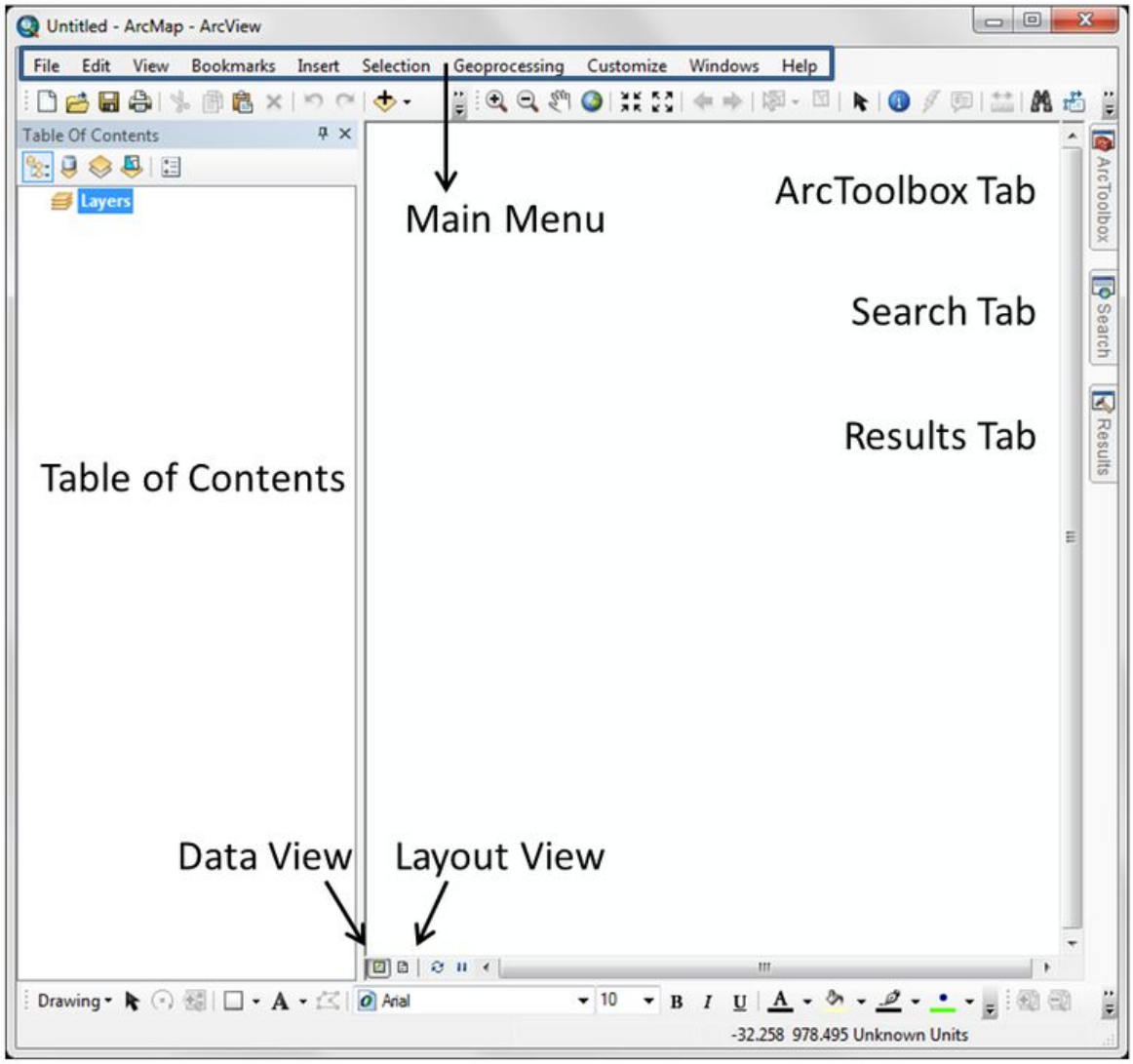

Figure 2: An overview of ArcMap interface

ArcCatalog (Fig. 3) is where GIS data can be searched for, previewed, and managed.

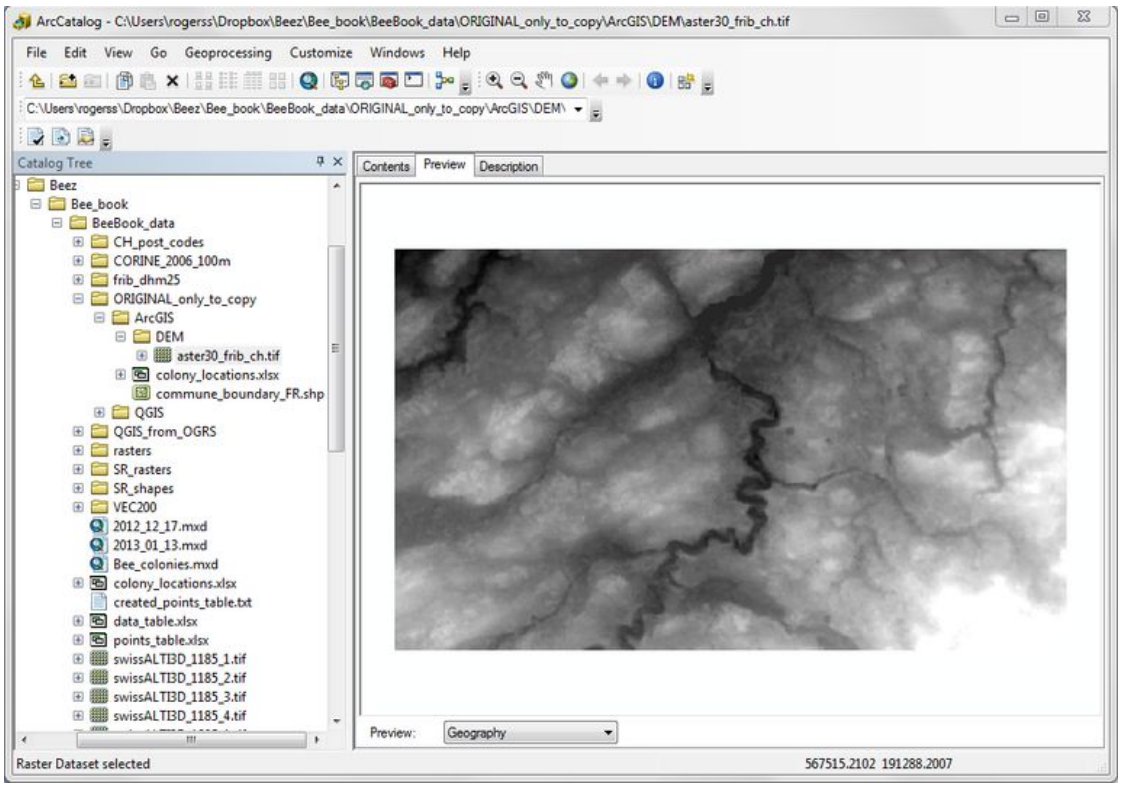

Figure 3: An overview of ArcCatelog interface

ArcToolbox (Fig. 4) is where all of the geoprocessing and spatial analysis tools are located. There are hundreds of tools available and they range from basic to advanced.

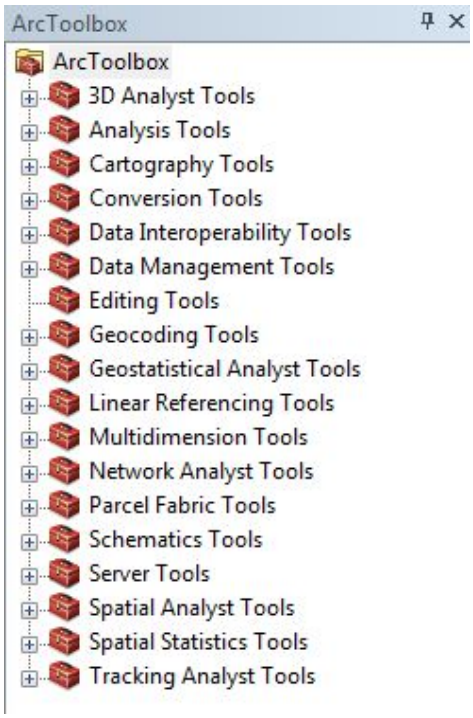

Figure 4: An overview of ArcCatelog interface

C. GIS on the Web

Web GIS is a type of distributed information system, comprising at least a server and a client, where the server is a GIS server and the client is a web browser, desktop application, or mobile application. In its simplest form, web GIS can be defined as any GIS that uses web technology to communicate between a server and a client. Web GIS has many advantages such as a global reach, a large number of users, better cross-platform capability, low cost as averaged by the number of users, easy to use, unified updates, diverse applications and etc. We can take advantage of web GIS through ArcGIS Server and ArcGIS Online.

D. Extensions of ArcGIS

ArcGIS has a suite of extensions that provide extended capabilities to the core product such as spatial analyst, geostatistical analyst, 3D analyst and etc. Those extensions are solutions for specific needs and markets.

IV. GWR Model

Geographically weighted regression (GWR) is an exploratory technique mainly intended to indicate where non-stationarity is taking place on the map, that is where locally weighted regression coefficients move away from their global values. Its basis is the concern that the fitted coefficient values of a global model, fitted to all the data, may not represent detailed local variations in the data adequately – in this it follows other local regression implementations. It differs, however, in not looking for local variation in ‘data’ space, but by moving a weighted window over the data, estimating one set of coefficient values at every chosen ‘fit’ point. The fit points are very often the points at which observations were made, but do not have to be. If the local coefficients vary in space, it can be taken as an indication of non-stationarity.

The technique involves first selecting a bandwidth for an isotropic spatial weights kernel, typically a Gaussian kernel with a fixed bandwidth chosen by leave-one-out cross-validation. Choice of the bandwidth can be very demanding, as n regressions must be fitted at each step. Alternative techniques are available, for example for adaptive bandwidths, but they may often be even more compute-intensive.

Figure 5 illustrates the weighting scheme of GWR.

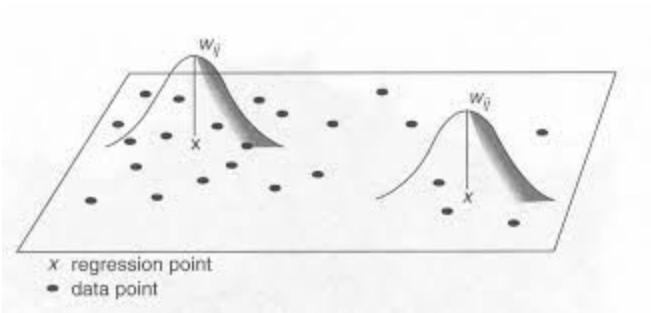

Figure 5: Weighting scheme of GWR

Formula (1) shows how to calculate the coefficients of regression for each data point.

$$\beta_i = (\sum W_{ij} X_j X_j^T)^{-1} (\sum W_{ij} X_j Y_j) \quad \text{.....(1)}$$

where X_j are the explanatory variables for all observations except i, Y_j is a vector heights for all observations except i, W_{ij} is the weight that building j given for i and the summations given by $\sum$ are taken over all buildings j.

V. Web Application Development

In this project, we use online model to conduct price estimation. After customers click on a location and select a circle range, the system collect the data points in the range and feeds them to the model. The model runs based on the inputs and returns the price estimation results. The results returned by the model is shown a box as follows (Fig. 6). The interactive website is developed based on Google Map.

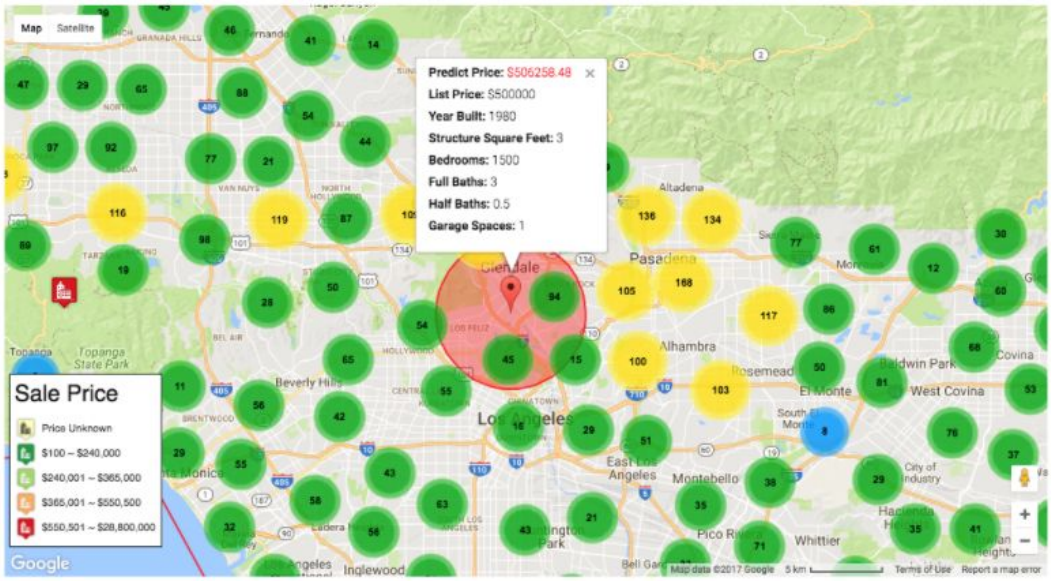

Figure 6: Display of model results