Advanced Geography Statistical Methods project proposal

Robert Chastain

24 February, 2021

What are you planning on modeling?

For my project I plan to model home values. Specifically, I plan to create a heat-map which indicates home values; the value of a specific house will be determined by where the house falls on the heat map given it's latitude and longitude.

What is the extent and resolution of your proposed study area?

Because house values are generally very localized I propose to study only values in the city limits of Reno. Initial research indicates that a resolution of 10 square meters is appropriate.

Where will you get your locality data?

I plan to use three categories of data:

- Home value data I plan to source from Clear Capital's proprietary data set.
- Elevation data I plan to source from the U.S. Geological Survey.
- Zoning and flood data I plan to source directly from the City of Reno.

In addition to these data sets, I'd like to include crime data, noise pollution, and a 'walkability' score. Unfortunately I haven't been able to find a source raster source for this data, but I will keep searching.

What environmental data will you likely be using, and why are you choosing these layers?

Most of the data I've chosen directly relates to the desirability of the neighborhood in which a house is located, and as the realtor cliche goes it's all about location, location, location.

The home value data is required as it serves as the target variable for training my model.

Home exist in a three dimensional world, the elevation of the house provides the z axis location of the house (where latitude and longitude are the x and y). Additionally elevation serves as a proxy for improved views and high priced homes in the Reno area.

Zoning and flood data both contribute to the desirability of a house. Homes in flood plains and homes that are near commercial zones are generally less desirable.

What obstacles might you face in getting this project put together?

I anticipate two significant data related obstacles with this project. First and foremost, I'm still waiting to get final approval to use Clear Capital home value data. The data I plan to use is older data and is mostly publicly available, but there is still a non-zero risk that I won't have access to it for these purposes.

Additionally, home values are influenced by the individual characteristics as much as by the homes location (characteristics such as lot size, gross living area, bedroom count, etc.). Not having these data points available

to the model will likely hurt overall performance. Unfortunately I don't see a reasonable way to include these data points in the model.

Aside from this most data sets appear to be rather complete and should be straight forward to work with.