Real Estate in Boston

Nicole Victory, Curtis Fisher, Anushka Wani

ABSTRACT

This paper presents a visualization aimed at assisting property buyers in identifying areas with opportunities for purchasing properties in Martha's Vineyard. The visualization employs a heat map created using spatial autocorrelation and data obtained from a regional multi-listing service (MLS) to display areas with properties for sale based on their average price. Additionally, users can explore individual listings that meet their criteria through tooltips containing relevant information, including contact details of the seller. Listings will be presented based on expressed area of interest. The visualization is intended to streamline the property buying process by reducing research time and facilitating interactions with unbiased agents and brokers. By providing an efficient means of locating properties, the visualization offers buyers a practical solution to the daunting task of purchasing property.

1 INTRODUCTION

This visualization supports property buyers looking to purchase property in certain areas of Martha's Vineyard. The first high level task that the visualization supports is identifying areas of interest using a heat map. The heat map allows users to discover, explore, and compare the average price of different areas of Martha's Vineyard. The heat map uses spatial autocorrelation, a form of data analysis increasingly used by real estate companies, to condense data from different areas of Martha's Vineyard and present the summarized data in an easily digestible format that can help property buyers easily decide if they want to pursue looking for properties in that specific area. Each area of the heat map will display the average price of a property in that area upon user interaction, but

additional information may be conveyed such as average size and qualitative details about the area. Learning about a new geographical market is a strenuous process, and a single buyer trying to discover and condense all the data available or unavailable about a new geographic area can be so difficult that the buyer gives up or loses precious time trying to learn and discover more. This visualization would expedite that process by collecting and summarizing the data so that the end user does not have to.

This visualization also supports buyers hunting for individual properties and listings through by allowing buyers to discover, browse, and identify individual properties that fit buyer criteria based on what area those properties are in. After using the visualization to discover possible areas to buy properties in within Martha's Vineyard, users can delve further into discovering individual listings for sale. The visualization also provides the functionality of easily displaying information of each individual listing, including useful data such as size, number of bedrooms, number of bathrooms, asking price, and any other useful information that may be readily available. This heavily reduces the cognitive load involved in scanning individual properties and trying to find the available information about potential properties to be bought.

The final high-level task supported by this visualization is providing the functionality for users to contact agents who are selling properties that are potential fits for the buyer annotate, lookup, and identify the contact information of properties that fit buyer criteria using a tooltip. The next step in the property buying process after exploring areas and individual listings is contacting the external parties that have private

information and can provide in-person tour experiences of a property as well as begin the negotiation process if desired by the buyer. This visualization significantly improves the process of searching for properties to purchase by condensing and summarizing information about general areas and specific listings that are available for purchase, as well as easily providing the contact information for the agents that are needed to continue the property buying process.

2 RELATED WORK

This paper [1] reviews the literature on spatial autocorrelation in real estate studies, focusing on the use of hedonic regression for modeling house prices. Spatial autocorrelation is a phenomenon where values of a variable located within a certain geographic area show a similar pattern. The paper highlights the importance of considering spatial autocorrelation in real estate studies, and outlines ways of dealing with this phenomenon.

The paper discusses two main approaches of modeling spatial autocorrelation: the spatial weight matrix and the geo-statistical approaches. The former is preferred in previous real estate studies that involve economic analysis. The paper concludes by emphasizing the importance of considering spatial autocorrelation when cross-sectional data are used.

The paper notes that while spatial autocorrelation has gained attention in real estate studies, evidence has mainly come from the USA. It suggests that evidence from other countries, such as Malaysia, would enrich the literature on spatial autocorrelation consideration in real estate studies. Overall, this paper provides a useful overview of the importance of considering spatial autocorrelation in real estate studies and suggests ways of addressing this phenomenon in hedonic regression models.

How this informs our project: The paper mainly focuses on the importance of considering spatial autocorrelation in real estate studies and outlines

ways of dealing with this phenomenon in hedonic regression models. While the paper does not provide specific visualization techniques or tools, it highlights the importance of understanding the spatial relationships among the data in real estate studies.

In the context of real estate data visualization, this paper emphasizes the need to account for spatial relationships when creating visualizations of real estate data. For example, it suggests that using spatial weight matrices may be a useful approach for modeling spatial autocorrelation in real estate studies, which could potentially inform the creation of spatially aware visualizations of real estate data.

Overall, this paper provides a useful theoretical foundation for understanding the importance of spatial relationships in real estate data and could inform the development of spatially aware visualization techniques for real estate data.

This paper [2] discusses the automation of spatial data analysis using ArcGIS, a popular software for working with geographic data. The paper focuses on two ways of automating tasks in ArcGIS - using Python as a programming language and using ModelBuilder, a visual programming language within ArcGIS. The paper presents a model created with ModelBuilder, which automates operations related to visualizing the spatial distribution of land prices in the North-Western part of the city of Lublin, Poland. The model takes in input data such as property location, unit price, and location of the city center, and uses concentric circles around the city center to determine the average property price dependence on distance from the city center. The advantage of using ModelBuilder is that it allows for the creation of geoprocessing models and scripts without requiring knowledge of programming language and can be exported to Python for advanced users to improve upon.

How this informs our project: This paper provides a useful approach for automating operations

related to visualization of spatial data in ArcGIS using ModelBuilder. The model created in the paper can be used to analyze the spatial distribution of land prices in a specific area.

The concentric circle visualizations presented in the paper provide a clear and effective way to illustrate the relationship between real estate prices and important locations such as the city center. By using concentric circles to represent distance from the city center, it becomes easy to see how property values change as one moves further away from the center. This can help to identify real estate hotspots, or areas where property values are particularly high due to their proximity to important locations. By examining these hotspots, it is possible to derive the valuation drivers behind them, such as access to transportation, amenities, or desirable features like waterfront views. This information can then be used to inform real estate investment decisions, marketing strategies, and urban planning initiatives. Overall, the visualizations presented in the paper offer a useful framework for analyzing real estate prices in relation to important locations and can help to reveal valuable insights into the factors that drive property values.

The paper also highlights the importance of automating repetitive processes in order to reduce the risk of human error and speed up work. This can be particularly useful in real estate data visualization where large amounts of data are often analyzed. The model created in this paper could be adapted or extended to analyze other aspects of real estate data such as the spatial distribution of property types or amenities.

3 USE CASE

Potential buyers of homes often look for homes or residences with certain price points. However, when moving to a new area, it can be difficult, time consuming, and tedious to learn about an area and discover if the properties available fit a buyer's budget preferences. This visualization will help potential buyers looking to move to Martha's Vineyard discover the different neighborhoods of Martha's Vineyard in a manner that will help them choose which potential areas they may choose to buy properties in. The visualization will use a heat map of Martha's Vineyard that colors areas of Martha's Vineyard based on the price points of those areas. Users can click into different areas of the heat map to zoom in on those areas and can then see the individual properties available for sale within those areas. Users may click on individual properties that are for sale in order to see data such as size, asking price, number of beds/baths, and contact information for the property owner to make further inquiries.

4 DATA

This real estate dataset was collected by LINKMV, a Cape and Islands based real estate multi-listing service (MLS) provider. The purpose of a multi-listing service is to work directly with real estate brokerages to aggregate real estate sales into a single unified data service. This dataset, therefore, contains information about real estate sales in the Cape and Islands.

One important point about this dataset is that it is relatively unbiased because LINKMV is a neutral third-party to the real estate market. However, it's worth noting that bias may be introduced in the form of the "Description" attribute, which is provided by the brokerages themselves. These brokerages have a profit incentive to sell more properties and may therefore provide descriptions that are overly positive or even misleading.

In terms of ethical considerations, this dataset provides granular information about people's homes, including the sale price, location, and features of the property. While this information is publicly available, a certain degree of inspection must be carried out by the brokerages, the MLS provider, and the data team to ensure that no sensitive information makes its way into the dataset. Sensitive information could include things like the personal contact information of the property owners, or details about the property

that could be used to identify the owner. It's important to ensure that this information is protected and not made public in the dataset.

Another important consideration when using a real estate dataset is the potential for spatial autocorrelation, which occurs when nearby observations are more similar to each other than they are to distant observations. In the context of real estate data, this means that the sale prices of nearby properties may be more similar to each other than they are to sale prices of properties located further away. This can be an issue when using statistical models that assume independent observations, as it can lead to biased estimates of coefficients and standard errors. To account for spatial autocorrelation, techniques such as spatial regression or spatial filtering may be used.
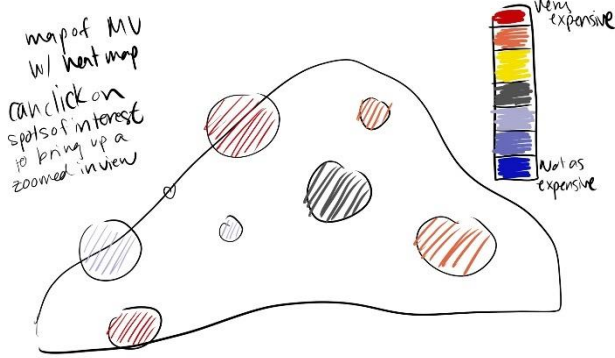
Another consideration when working with real estate data is the potential for missing or incomplete data. Incomplete data can arise from various sources, such as data entry errors, data suppression due to privacy concerns, or missing information on certain attributes of a property. Missing data can have an impact on the accuracy of statistical models, as it can lead to biased estimates of coefficients and standard errors. Therefore, it's important to carefully handle missing data by imputing missing values or using statistical techniques that can handle missing data.

Finally, when using a real estate dataset, it's important to consider the potential impact on communities and individuals. Real estate data can reveal patterns of segregation, gentrification, and discrimination in housing markets, which can have a profound impact on people's lives. Therefore, it's important to use real estate data in an ethical and responsible manner, taking into account the potential social and economic implications of the findings. This may involve working with community organizations or advocacy groups to ensure that the findings are used in a way that promotes fairness and justice in housing markets.
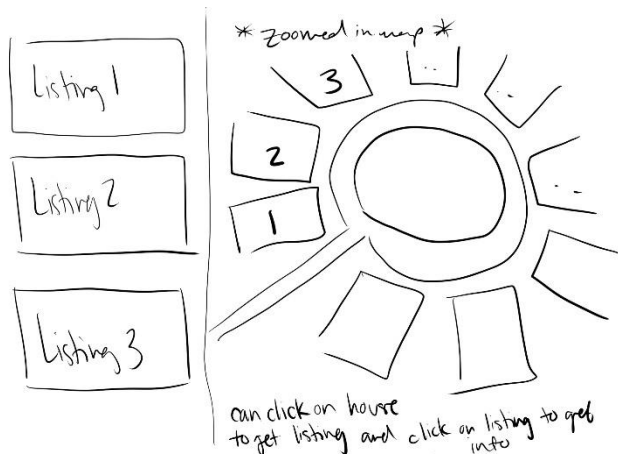
To clean the data, we decided to remove the photos columns and other columns that had an excess of missing data. We also decided to remove columns that included links or urls. We combined the exterior features columns to create a singular column with all the exterior features, formatted with commas separating the values. We also re-formatted missing values from NaN's to empty strings, so as to have ease in concatenation, without resulting in a loss of information. Missing values were not filled in with averages or other such methods, as the data missing involved attributes that would be displayed as information about a listing, and weren't major attributes like price or location. Missing values such as "last renovated date" cannot be found through mathematical derivation, and therefore needed to be treated as information that needs to be left out of the additional information portion of the listing. In terms of derived columns, creating a column describing price category allows us to organize the listings into 6 different groups based on the listing's price ranges. This provides the ability to create the heat-map of the different areas of Martha's Vineyard with certain price ranges to complete part of our specified domain tasks. Other simple reformatting of columns was performed for consistency, and reordering of columns was needed so that the information being presented felt organized when looked at in its tabulated format.

5 DESIGN PROCESS

The first sketch shows the preliminary heat map, all just sample values with no actual backing of data. The actual positions for the area circles will be determined through the implementation of the visualization. This rough sketch was a partial sketch of what the final map view would look like. The final version is mostly the same, just with added detail on functionality and an updated divergent color scale.

map of MV w/ heat map

can click on spots of interest to bring up a zoomed in view
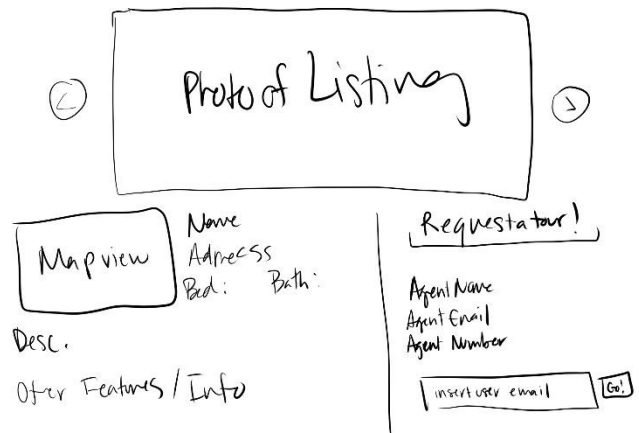
Very expensive

Not as expensive

The second sketch shows a partial view of the listing view, with the map on the right side. This is an extremely zoomed in version, showing a specific neighborhood. One could click on the actual shape of the house to get it's listing. In the final design that was changed to include simple points to represent each listing, with limited zoom functionality so that there isn't a mass of buttons in the middle of the screen that is impossible to click on, but not zoomed so far as to need to render the house shapes. This functionality may be expanded to include rendering of the house shape based on further exploration in the implementation phase.

Listing 1

Listing 2

Listing 3

* zoomed in map *

3

2

1

can click on house to get listing and click on listing to get info

The third sketch shows the individual listing view with the formatting of the different information involved with each listing. The photos are displayed at the top, whilst the information is on

the bottom left and the request a tour functionality with agent information is on the right. This was improved upon to have a more structured format in the final design, and with added screens to either apply for the listing, or to request a tour, with each button taking one to a different form that would get sent to the agent email provided if submitted.

Photo of Listing

Map view

Name
Address
Bed: Bath:

Desc.

Other Features / Info

Request a tour!

Agent Name
Agent Email
Agent Number

insert user email   Go!

The final sketch includes 4 views. The first being the heat-map, that uses divergent color scaling to indicate price range. The marks are the different heat zones, and the channels are position, size, and color. The color indicates the price range, and the size of the heat zone indicates the number of listings in that specific area. The placement of the marks also indicates position on the map and shows the locations of the different price zones in Martha's Vineyard. As explained before, one can click on a zone on the heat map to get a zoomed in view on that area and to be able to click on the listings, now being represented with marks that are points, through the position channel to indicate location on the map portion. This view also offers the list view of the listings that can also be clicked on for more information about the listing. Clicking on a point or a listing card will bring you to the individual listing view, that has the photos, information, area to get back to the map view, and a section to request more information from an agent. Clicking on either button, request a tour or apply, will bring the user to a different form. One

form will request to tour the listing, while the other will start the application form.
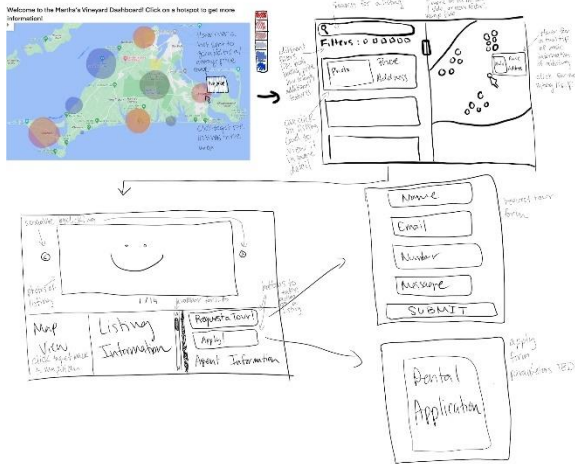
Discover possible areas to buy properties in Martha's Vineyard	Discover, explore, and compare the average price of different areas of Martha's Vineyard using a heat map
Discover individual properties for sale in Martha's Vineyard	Discover, browse, and identify individual properties that fit buyer criteria
Contact agents who are selling properties that are potential fits for the buyer	Annotate, lookup, and identify the contact information of properties that fit buyer criteria using a tooltip.

6 FINAL DESIGN

7 DISCUSSION

8 CONCLUSION

REFERENCES

[1] S. Ismail, "Spatial Autocorrelation and Real Estate Studies: A Literature Review."

[2] P. Pochwatka, U. Litwin, T. Teterycz and A. Bitner, "Cartographic Visualization in The Real Estate Market Investigation with the Use of GIS Tools," 2017 Baltic Geodetic Congress (BGC Geomatics), Gdansk, Poland, 2017, pp. 105-109, doi: 10.1109/BGC.Geomatics.2017.53.

APPENDIX

Data Abstraction

id	name	description	address	area	beds	baths	price	Price category
Categorical	Categorical	Categorical	Categorical	quantitative	quantitative	quantitative	ordinal	ordinal

town	squareFeet	lotSize	type	yearBuilt	lastRenovated	dateListed	status	Appliances
categorical	quantitative	quantitative	categorical	ordinal	ordinal	ordinal	categorical	categorical

Exterior Features	furnished	basement	fireplaceFeatures	heating	laundryFeatures	parkingFeatures	sewer	waterSource
categorical	categorical	categorical	categorical	categorical	categorical	categorical	categorical	categorical

listingAgent/name	listingAgent/phoneNumber	listingAgent/email	listingAgent/linkMlsId	listingAgency/name	listingAgency/phoneNumber	listingAgency/faxNumber	listingAgency/email	location/latit...
categorical	categorical	categorical	categorical	categorical	categorical	categorical	categorical	location

Each row in the dataset, ie each item, represents a different listing, with the different attribute types identified above.

Task Abstraction

Domain Task	Task Abstraction