

## Assignment 2

Developing interactive data visualizations in React and D3js

Fabio Rafael PEREIRA ALVES

020107571C

### I. Introduction

For this assignment we are provided with a dataset that is related to the housing market. As we know, the housing market in general has many variables which can determine the value of a property. This assignment's goal is exactly that, to provide us a better visualization of the housing market and determine their value depending on different variables.

Bedrooms, bathrooms and stories, are only a few of the variables given by the dataset. Our goal is to find a visualization method that allows us to display, and understand the correlation between the price and the area, for example.

### II. Visual Design & Data Properties

#### 1. Scatter plot

Since in the scatter plot is already given, there is not much for us to discuss about the implementation. However, can discuss about their limitation.

As we know a scatter plot can only take two different values, in our case area and price. This is, however only one of the possible combinations with our current dataset. We have in total 36 possible combinations, however, it does not mean that some of those combinations make sense. For example, price with price, or area with area. I will not go into details with other combinations as we will be able to see them more clearly with our chosen solution. Since we have that many possible combinations, it means that we would need to create 36 different scatter plots in order to have a more or less understanding of the housing market, which is why I chose a heatmap.

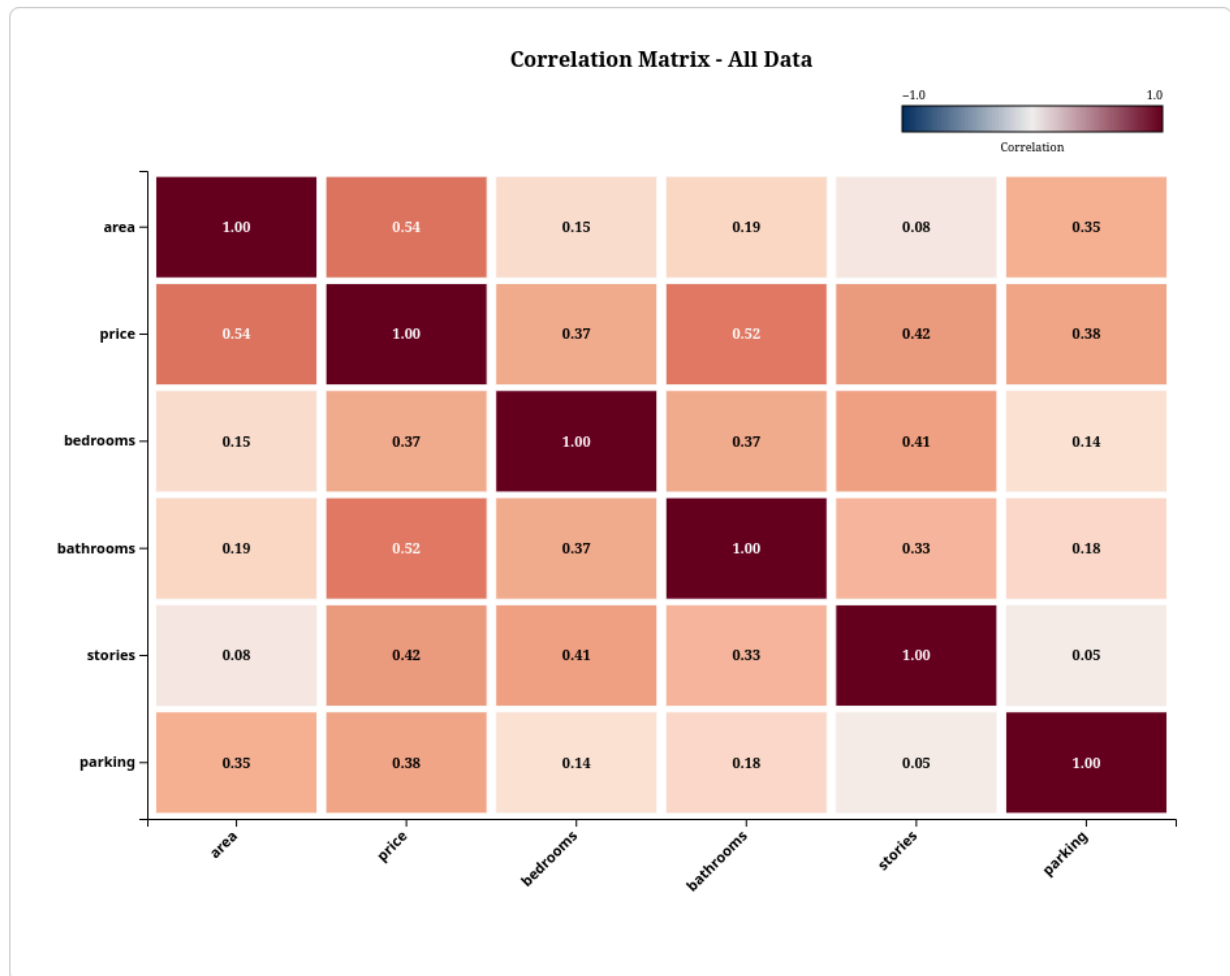
#### 2. Heatmap

Heatmap is one the many data visualization that can represent many variables. The reason as to why I choose this one and not others is quite simple, heat maps look cool and I like them. On a serious note, heat map for me makes it clear to understand the correlation between variables. Not only does it provide a visual presentation, color, but also a numeric representation. Heatmaps are quite versatile for this reason and therefore my choice of visualization.

##### 2.1 Heatmap Understanding

As I mentioned before, we have a total of 36 combinations, which we can see it now with the heatmap. This will allow us to determine correlations properly.

## Correlation Heatmap



Here you can see my result for the heatmap. The current display is for all the values in the dataset. We already can easily find some correlations compared to the scatter plot, which were quite difficult to find.

in this case, every time we have the same variable comparison, we have the value of **1.00** this means that they have a strong correlation. Which make sense since we are comparing the value with itself and therefore we have a strong correlation. These are comparisons that don't make sense on paper, but are important to understand.

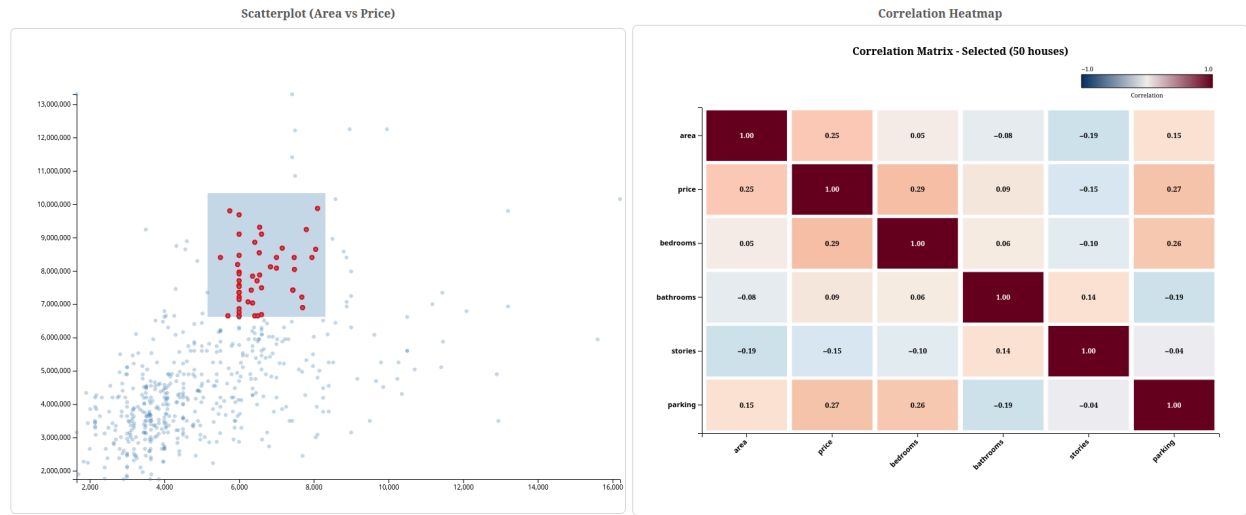
Now, we can take a look a values that are close to zero such as, stories and area. These two show a value of **0.08** . This means that number of stories and the area will not affect on the price. Same goes for stories and parking.

Lastly, positive and negative correlation. For the positive correlation ,  $0.5 <$ , we can safely say that those variables have a correlation, such as price and area. This makes sense since in the house market the more livable surface area we have the more expensive the house is. Contrarily, negative correlation,  $-0.5 >$ , would mean the price would be lower.

### III. Combination of Scatterplot and Heatmap

We had to implement a brush that would allow us to select a highlighted area. That area can be seen on the scatterplot.

## Housing Data Multivariate Analysis



As you can see here, we can successfully highlight a selected number of points, and the heatmap will change accordingly. Currently, the heatmap shows the correlation for all the selected points, not all the points as seen on the previous image. By moving the area around we can see a live change of the heatmap.

### IV. Design Choices: Pros and Cons

For the pros, heatmap does an amazing job portraying the correlation in a proper manner. It allows pattern recognition through color as well as numeric understanding, it allows to handle multiple variables in limited space and reveals symmetric relationships.

For the cons, heatmap will lose individual data points, the color can be quite tricky depending on the user, for example color blind.

### V. Conclusion

It was a nice assignment to combine different types of data visualization tools, heatmap and scatterplot. I think this assignment show how differently data can be displayed if we us different tools and that some data can actually be "missing". For example, heatmap will lack a lot of precision compared to the scatterplot.

I don't think there is a right or wrong choice of visualization tool, but I rather think there are some weaknesses and strengths, which we learn in this assignment.

### Final Note

I used AI to generate and help me code this assignment. This was a fun project and something that I would do again without AI for sure. I used AI mainly because javascript is for me one of the most difficult languages that I have ever learned and when I can avoid it I will. This, however, does not mean that I didn't understand what I was doing nor what it was required from me.

This whole document was written by me and not edited by AI or what so ever.

The AI used was Deepseek.