# W209 Spring 2022 Section 01

# Final Project Exploratory Visualization

Irene Shaffer

Heather Rancic

Austin Jin

Luis Delgado

Tony Angell

## Initial exploration:

Our real estate [dataset](https://www.kaggle.com/deepcontractor/miami-housing-dataset) is from 13,932 single family home sales in Miami throughout the year in 2016. We examined the FRED economic [data](https://fred.stlouisfed.org/series/ATNHPIUS12086A) to have a rudimentary understanding on the economic status of real estate sales in 2016. Were they affected by any larger U.S. economic factors (i.e. the home sales crash of 2007-2010) affecting real estate sales? We noted that 2016 was right in the heart of the last decade’s linear increase in real estate prices. This could have a reasonable impact in sales at the beginning of the year versus those in the end, but as is universally understood, home sales have cyclical impacts throughout the year that help stabilize that impact (i.e. home sale prices peak between April and July and then flatten out in the winter and fall months).

As we examined our data without visualizations, we also noted the key elements that affect home sale prices. It is not simply a division of price by square footage. These variables and our consideration for them, if applicable, are:

1. **SALE\_PRC**, home sale value
2. **TOTLVGAREA**, floor area in square feet: Used in calculating price per square foot
3. **HWY\_DIST & RAIL\_DIST**, distance to highways/railways in feet: Used to calculate the impact of noise and pollution
4. **OCEAN\_DIST**, distance to the ocean: Used to calculate possible impact of sea level rise
5. **WATER\_DIST**, distance to body of water: This is critical since the major body of water in Miami Dade outside of the ocean is the Biscayne Bay, which is directly impacted by sea level rise from the Atlantic.
6. **LATITUDE & LONGITUDE**: Exact location of the home accurate to a 1 square foot. This is very helpful to examine the actual location of homes on Google Maps.
7. **STRUCTURE\_QUALITY**: A 1-5 scale where a 1 is typically a trailer home and a 5 is a very nice home in a highly desired community. We did have to map several coordinates for homes to understand what this variable represented.
8. **MONTH\_SOLD**: Month of the year (2016) the property was sold. This helps us understand seasonal impacts on sales.
9. **PRICE PER SQUARE FOOT**: An internally generated attribute representing the home sale divided by the square footage. This is a good tool to compare home prices.
10. **MONTH OF SALE**: An internally generated attribute translating MONTH\_SOLD numbers to names of the month (i.e. 1 = Jan).

One of the highlights from this initial exploration would be the clustering of high end homes in bay or ocean front locations. This is fairly obvious given the similarities in LAT/LONG coordinates and noting the low values of WATER\_DIST.

As we explore the threat of sea level rise, we relied on a series of studies. The primary one being a 2017 [study](https://www.nytimes.com/2022/02/15/climate/us-rising-sea-levels.html?d) reviewed by the New York Times. This study described the very serious threat of a rise in sea level by one foot by 2050 and two feet by 2100. It is important to note that this study was released one year after our dataset’s sales occurred.

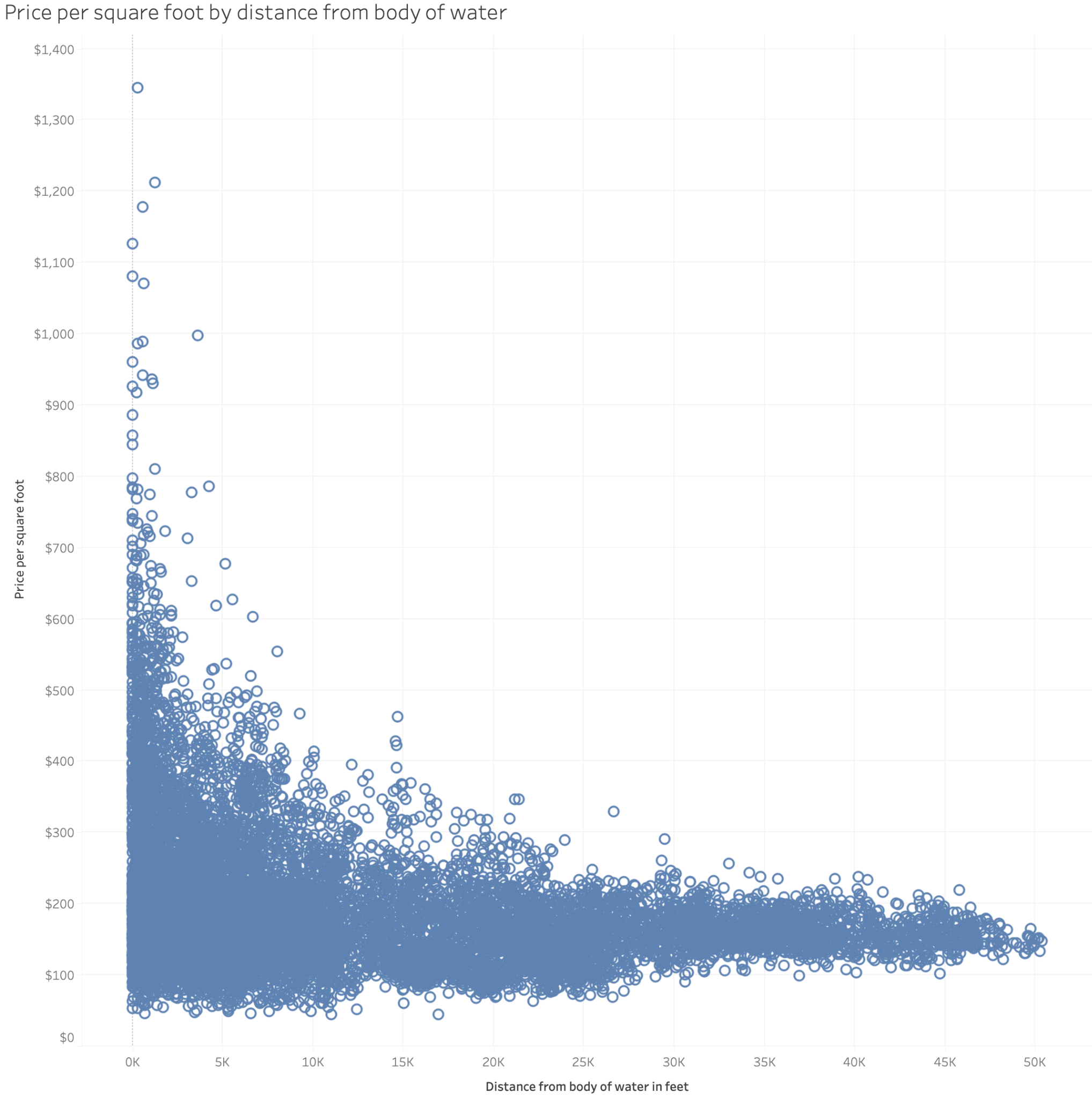
## Resulting Hypotheses:

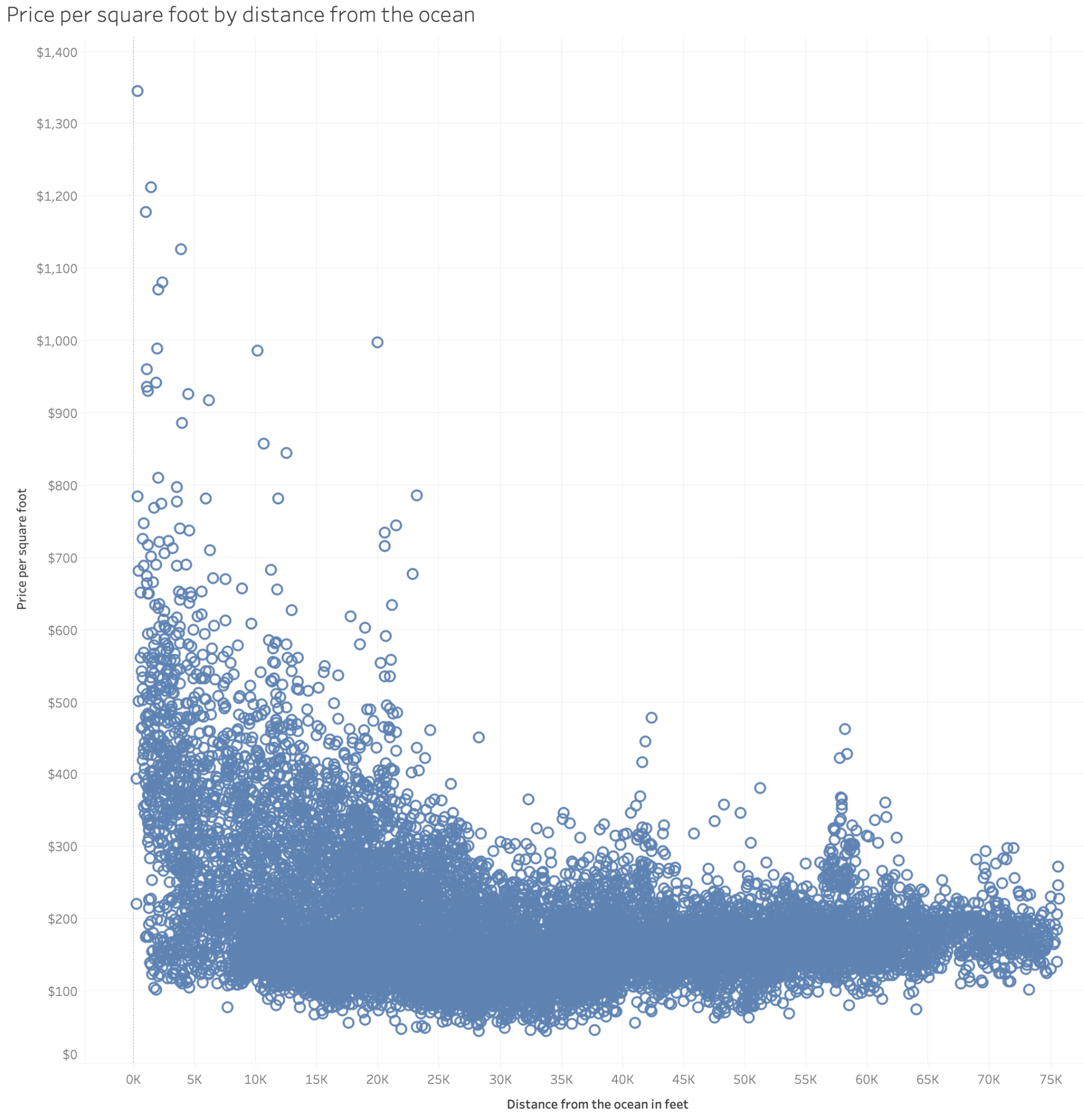
Our initial exploration resulted in the following three hypotheses:

1. The threat of Sea Level Rise (SLR) has a negligible impact on home sale prices.
2. Proximity to noise has a significant effect on comparable home sales.
3. Home prices are more expensive near the ocean with larger homes being more affordable further away from the ocean.

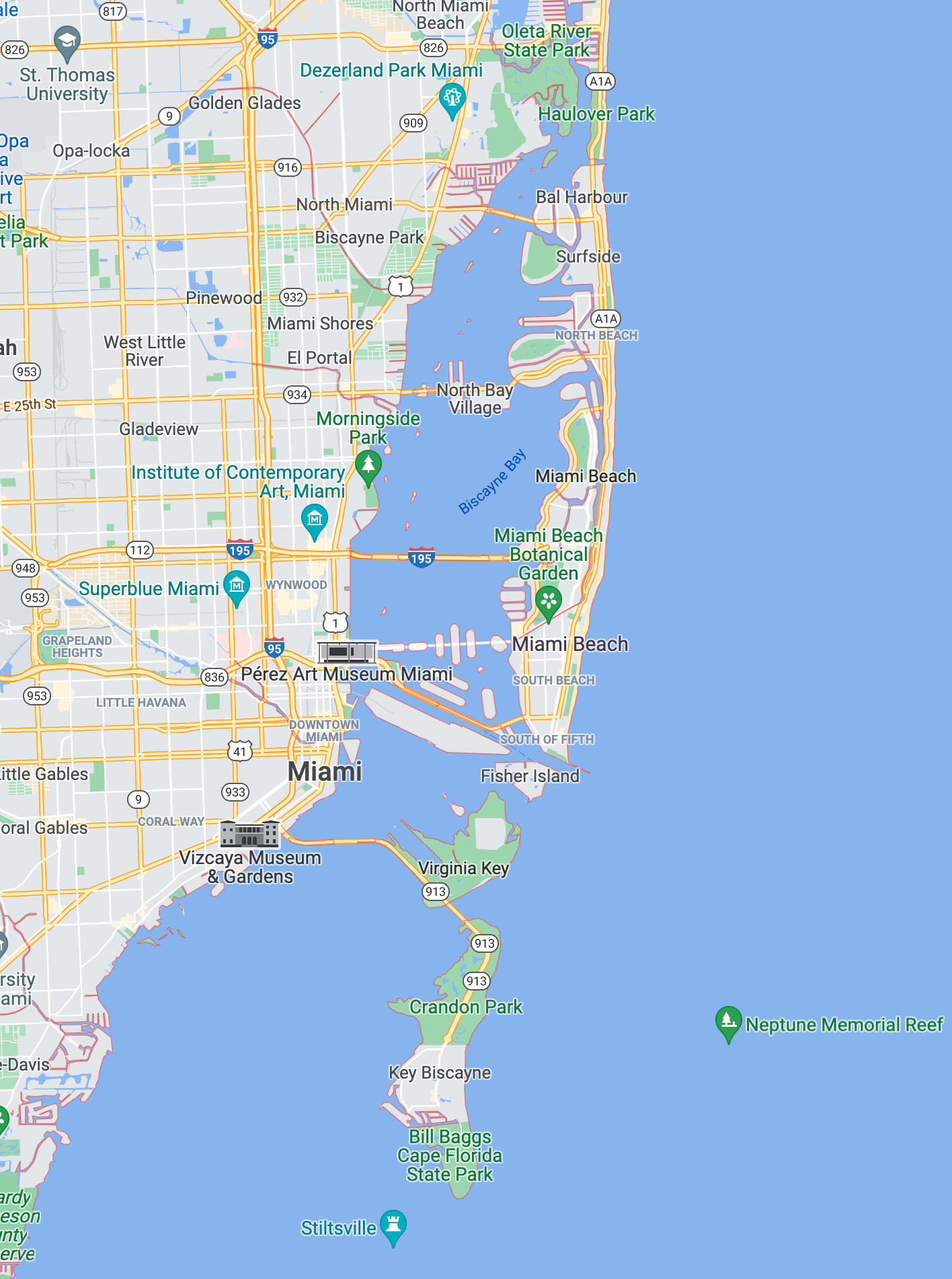
## Hypothesis 1: The threat of Sea Level Rise has a negligible impact on home sale prices.

Our first visualization displays the price per square foot against the distance in feet from a body of water. The second one has the same metric except it is the distance from the ocean.



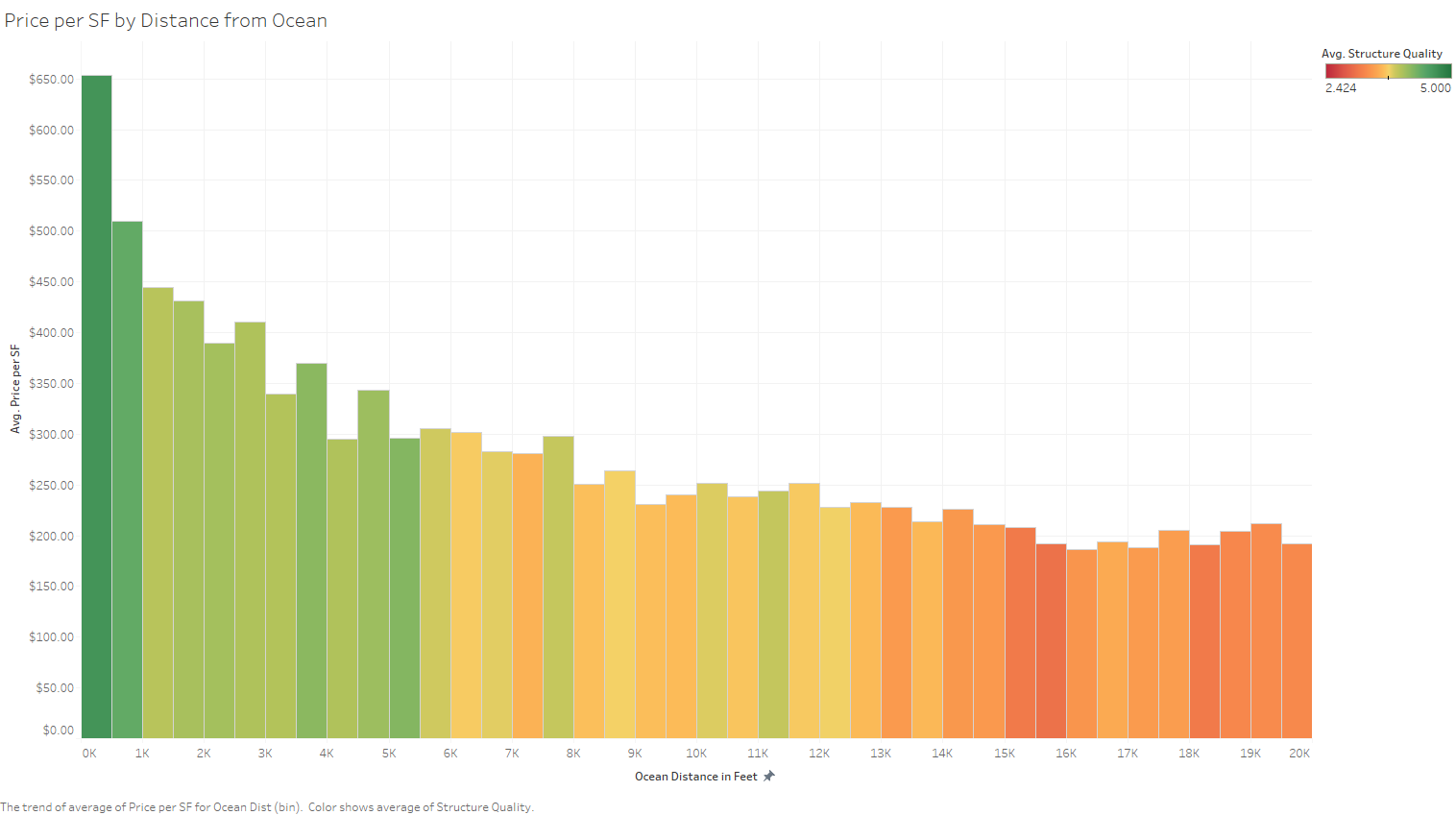


It is critical to have an understanding of the geography of the Miami-Dade metropolitan area. As you can see below:



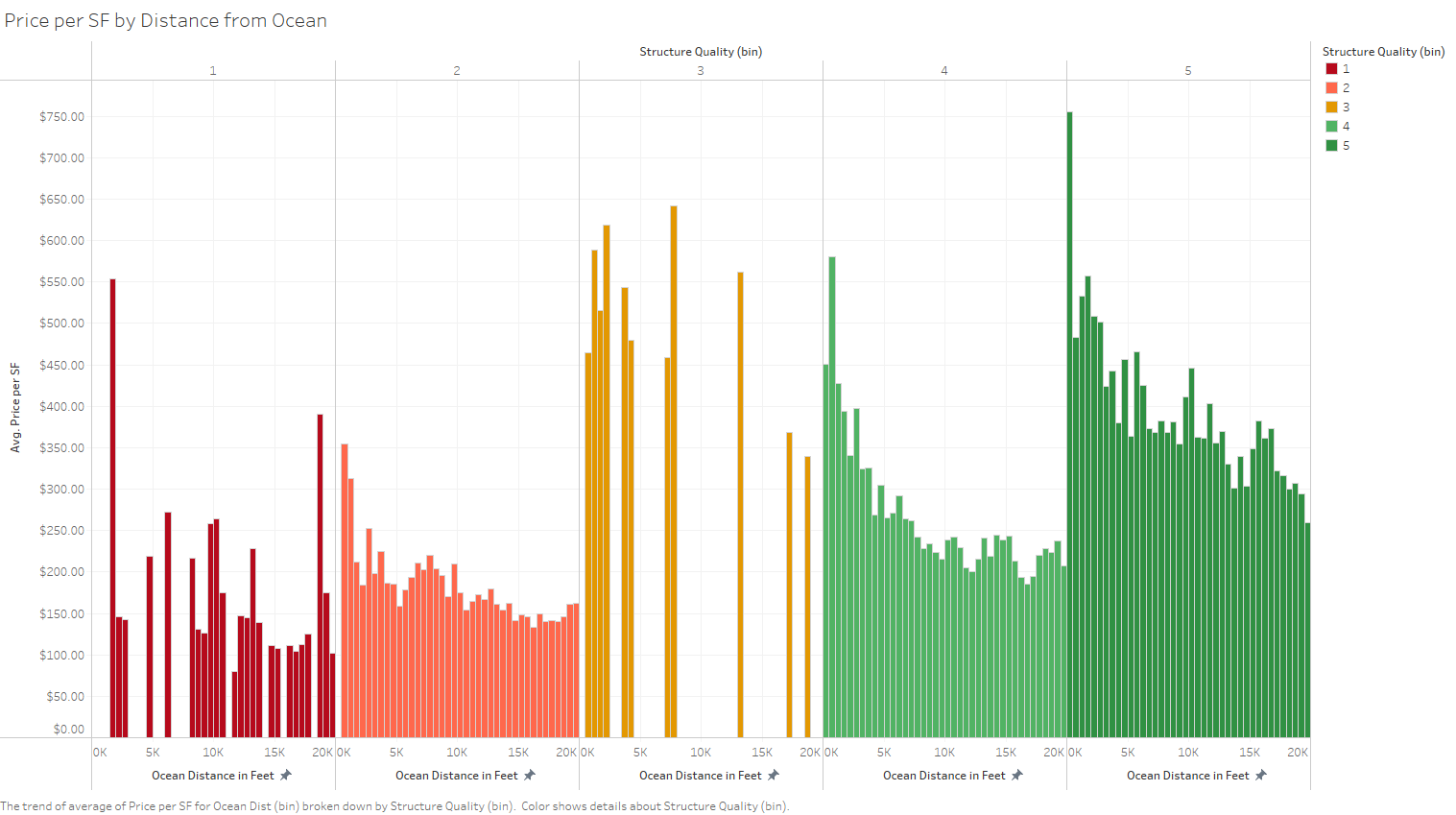
The two major bodies of water are the Atlantic Ocean and Biscayne Bay. Biscayne Bay is directly fed and affected by any changes in the Atlantic Ocean.

We further refine our visualization to better understand the price per square foot and distance to the ocean:



From: <https://public.tableau.com/views/EDA_16453930877820/Sheet1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

We also separate the distance into groupings to check for clusters of distances to the ocean.



From: <https://public.tableau.com/views/EDA_16453930877820/Sheet12?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

**Evidence for or against hypotheses:** The results from the chart show evidence against the hypotheses since the distance from the ocean does in fact have an impact on the price per square foot.

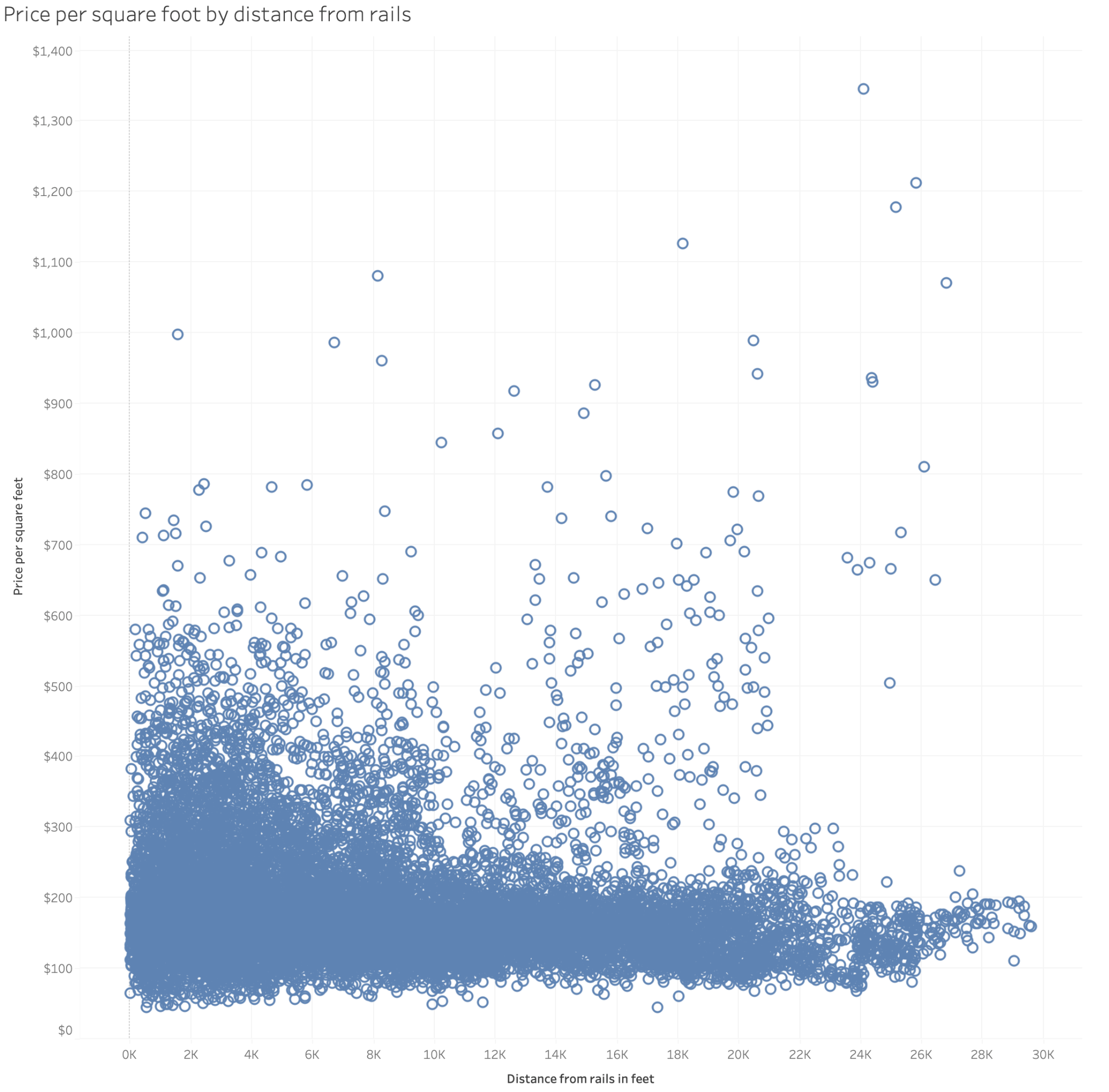
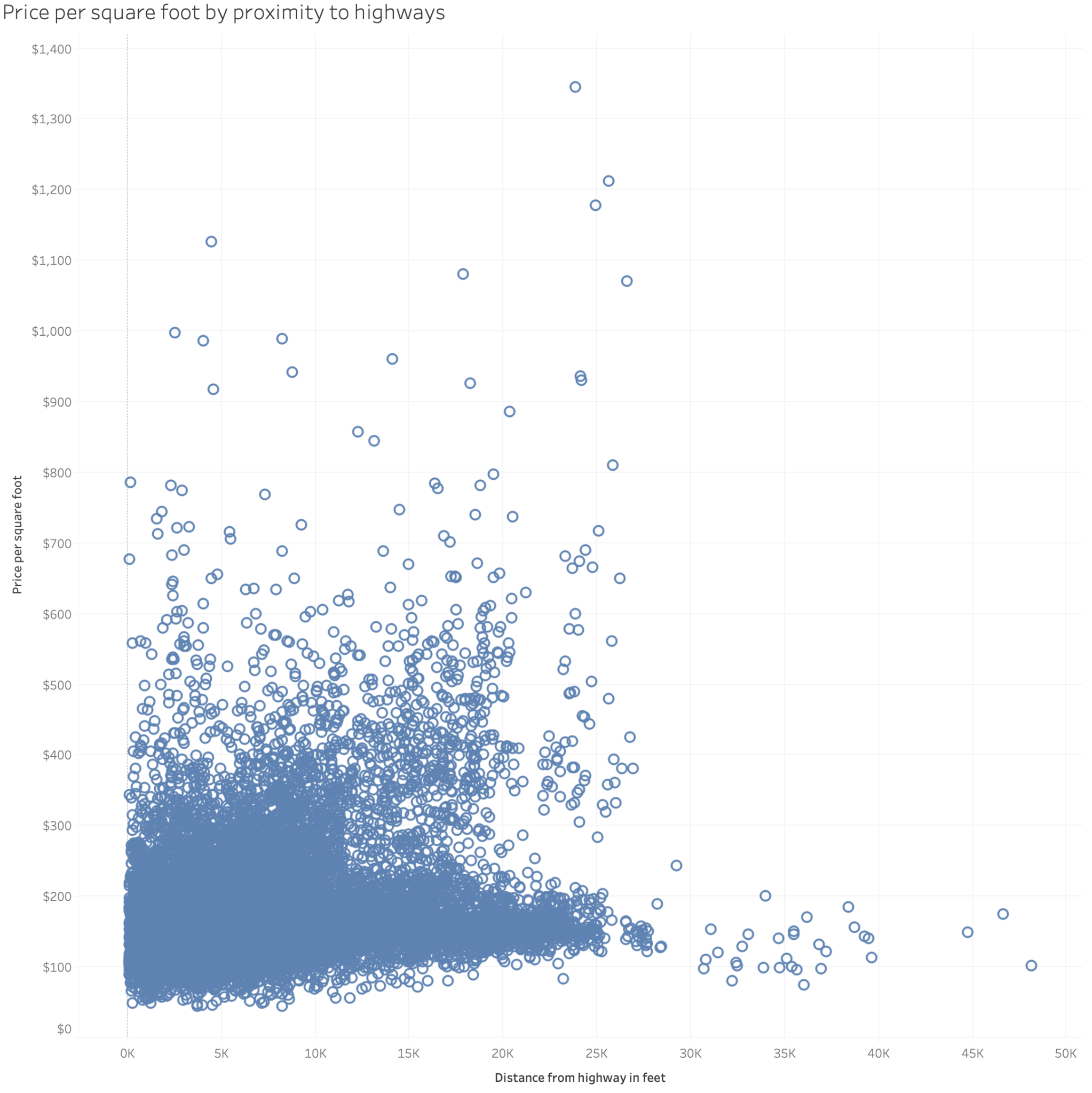
**Clusters, Outliers, and Correlations:** There seems to be a few outliers when the house is very close to the ocean (e.g. highest price per square foot in between $1,300 and $1,400 when the distance from the ocean in feet is close to 0K). There are also clusters of houses within the range of around $100 to $250 per square foot and distanced 67.5k to 75k feet away from the ocean. Even though the data points are fairly scattered throughout the chart, there is an explicit negative correlation between the price per square foot decreasing as the distance from the ocean in feet increases.

We also see a clear correlation between structure quality and proximity to the ocean (Figure 3). Homes sold near the ocean tend to be higher quality, which may also contribute to the higher price per SF. However, when we plot the different Structure Quality values separately, we continue to see the trend of higher price per SF closer to the Ocean (Figure 4).

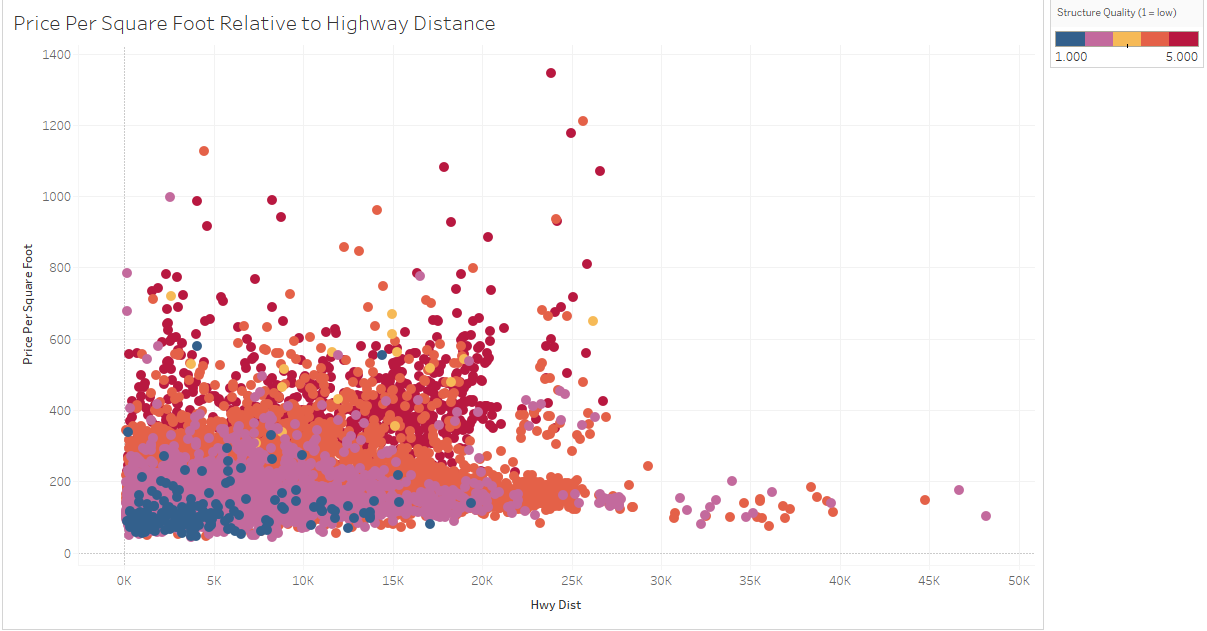
**Conclusion:** As we can see from the chart, the price per square feet decreases as the house is distanced further away from water, and this trend remains even when holding structure quality constant.

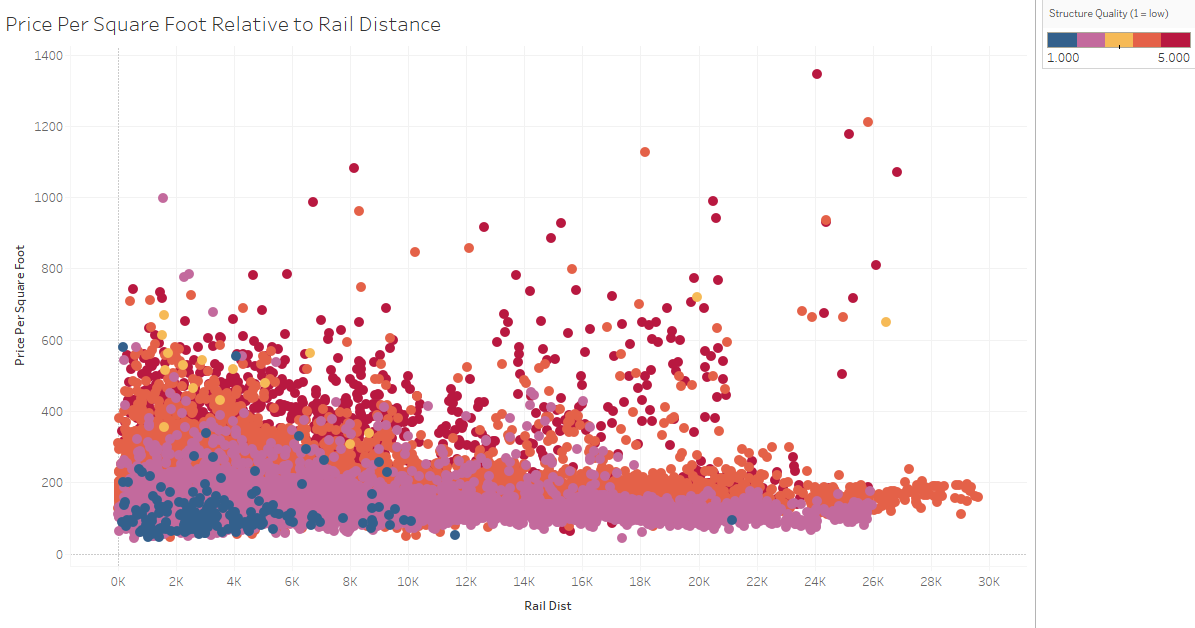
## Hypothesis 2: Proximity to noisy areas has a negative impact on comparable home sales.

Our initial visualization showing the price per square foot compared to the distance from highways and rails, respectively:



We further refined these visualizations to better understand the distances from noise sensitive areas grouped by the quality of the structure. As a reminder higher quality structures are very expensive homes with architectural complexity. Lower ones (scores of 1) are usually trailers.



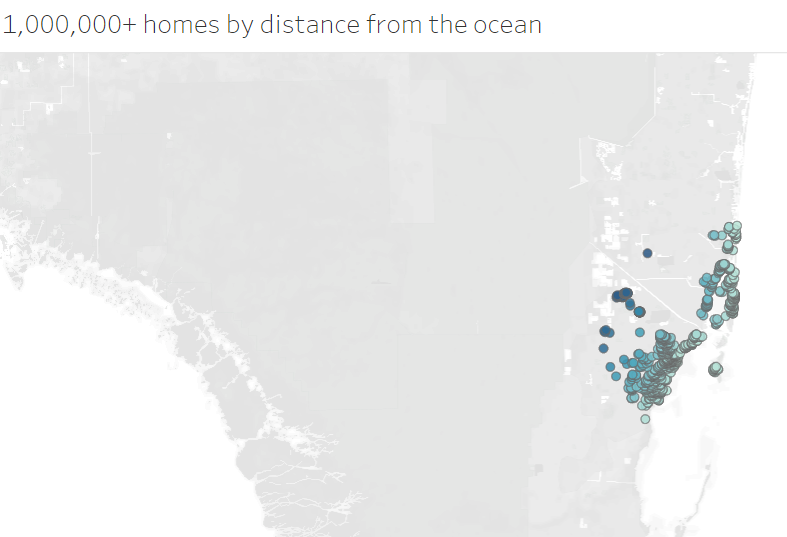


**Evidence for or against hypotheses:** The results from our charts show that, in general, distance from the highway and railways do not have a large impact on price. However, when we add the structural quality where 1 is lowest quality, a trailer home, and 5 is highest quality, a nice home in a nice neighborhood, we see some differences. For the level 5 homes, we do see some difference. In those cases, there are no level 5 homes beyond 25k feet from the highway so the prices only go up once they hit that wall. In terms of rail distance, it doesn’t seem to have too much of an impact.

**Clusters, Outliers, and Correlations:** In terms of highway distance, it does seem like the vast majority of homes are within a particular distance and price range (up to 25k feet from the highway, up to 600k price) and then after that we could consider ‘outliers.’ If we investigated some additional dimensions like do they have any special features (eg. a pool) we may find that these are not outliers and they do have a particular pattern but if we’re restricted to structural quality it seems like they are. Additionally, for rail distance the main takeaway is that price decreases, in general, as distance increases. However, there are many level 4 and 5 homes, in particular, that break that trend. Overall, it seems like homes in the 1-3 level range and 4-5 level range have similar behavior patterns.

**Conclusion:** From this analysis, it looks like rail and highway distance do not strongly indicate an effect from noise on price. It seems more likely that all homes are at least within 25k feet of a highway. Railways are more rare. An alternate explanation may be more relevant here - who needs access to transportation and how are they transported? Most likely if we could measure noise inside of homes we could arrive at a better conclusion since things like insulation and other features can reduce noise and this is a poor approximation of the effect of noise on the home.

## Hypothesis 3: Home prices are more expensive near the ocean with larger homes being more affordable further away from the ocean.



**Evidence for or against hypotheses:** The results from the chart show evidence for the hypotheses since the distance from the ocean does in fact have an impact on the price per square foot. It seems that home prices are more expensive if near the ocean compared to home prices that are further away from the ocean.

**Conclusion:** As we can see from the chart, the price per square feet increases as the house is distanced closer to water.

## Data Sources:

### Miami Housing Dataset:

<https://www.kaggle.com/deepcontractor/miami-housing-dataset>

### Miami-Dade County 3-D Sea Level Rise (SLR) impacts:

<https://mdc.maps.arcgis.com/apps/webappviewer3d/index.html?id=b92a9fa4ff8847bf97f3e628a195a398>

### Miami-Dade County Sea Level Rise (SLR) predictions:

<https://southeastfloridaclimatecompact.org/wp-content/uploads/2020/04/Sea-Level-Rise-Projection-Guidance-Report_FINAL_02212020.pdf>

### Coastal Sea Levels in U.S. predicted to rise by 1 foot by 2050

<https://www.nytimes.com/2022/02/15/climate/us-rising-sea-levels.html?d>

And

<https://oceanservice.noaa.gov/hazards/sealevelrise/sealevelrise-tech-report.html>

### FRED Economic Data on House Price Index for Miami-Dade County, FL

<https://fred.stlouisfed.org/series/ATNHPIUS12086A>