Airbnb Pricing Patterns in Denver

2025-12-14

Abstract Introduction

Airbnb prices vary a lot across Denver. In this project, we ask whether the southeast area has higher nightly prices across Denver because of location or whether it looks expensive just because the area has larger properties on average. We used a public dataset from InsideAirbnb and fitted two GAM models. The first model predicts prices based on location only, and the second one adds some additional features such as number of bedrooms, number of guests the listing can host, room type and review scores. The first GAM model makes the southeast the most expensive area in Denver. However, when we add the other features, the Downtown area becomes the most expensive part of Denver. This shows that Downtown Denver has the strongest location value, and whenever we want to make decisions, we should not rely on basic average prices alone.

Goals of the Analysis

The goal of this project is to understand how location influences Airbnb prices in Denver. specifically,we want to know whether the southeastern neighborhoods have higher prices on average because of location or whether the area is expensive because it has larger houses/apartments on average. To answer this, we fitted two GAM models, one that predicts prices on location only and another one that has some additional features. Comparing both models helps us separate the location effects from other listing features. This will give us a better understanding of which area in Denver has the strongest location premium. In this project, we used GAM models because prices are changing unevenly across locations. We chose the GAM models specifically because they are good at capturing curved, uneven and non-linear patterns.This helps us understand how prices change across the city.

Data Description

The dataset for this project comes from InsideAirbnb, which is an open-source database that provides public Airbnb listing data for many cities, including Denver. Each row from the dataset represents an Airbnb listing with other information, including price, location and other listing features. The dataset includes all Denver's Airbnb listings available when the dataset was downloaded from InsideAirbnb (not just a sample).The key variables of this data set are:

- price_clean: the price of the listing per night (cleaned and converted to numerical values).

- longitude and latitude: the geographic location of each listing.

- bedrooms: number of bedrooms in the listing.

- accommodates: the maximum number of guests the listing can host.

- room_type: whether the listing is an entire home/apt, private room, hotel room or shared room.

- review_scores_rating: average review score given by guests.

To prepare the data, we cleaned the price column by removing the dollar sign and converting it to a numerical value. Also, we dropped the rows that had missing values in the key variables. Moreover, we created simple average summaries, like average price per neighborhood to understand the data before fitting the GAM models.

Distribution of Airbnb Listings in Denver

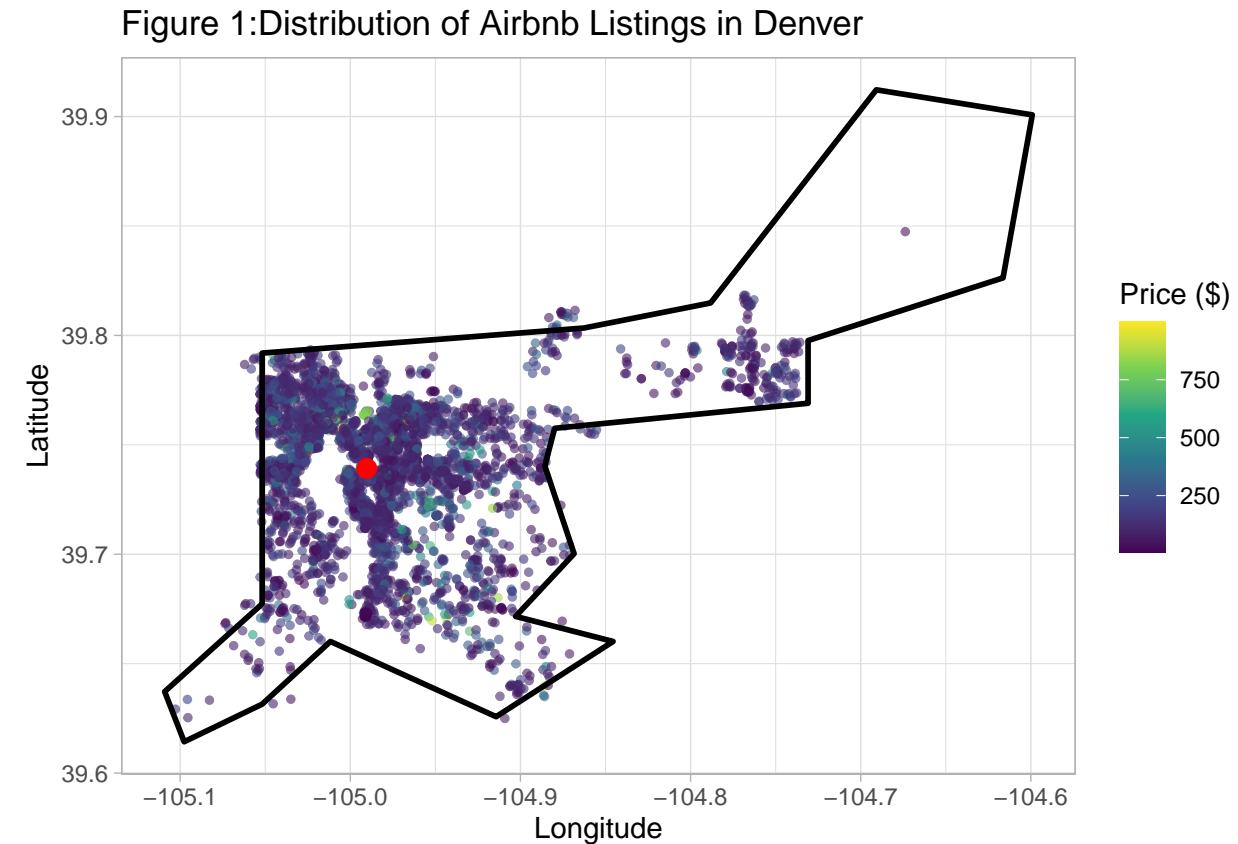Figure 1:Distribution of Airbnb Listings in Denver

Figure 1 above shows the distribution of Airbnb listings across Denver. We can see that most of the listings are concentrated around Downtown Denver, which is marked by the red point. Downtown Denver has the highest density of listings but there is also a cluster of listings in the northwest, and whenever we move farther from these areas, the number of listings becomes smaller. The color scale explains the differences in prices, and both cheap and expensive listings appear to be spread across the city. Overall, this map shows a basic distribution of Airbnb listings across Denver.

Denver's Most Expensive Neighborhoods

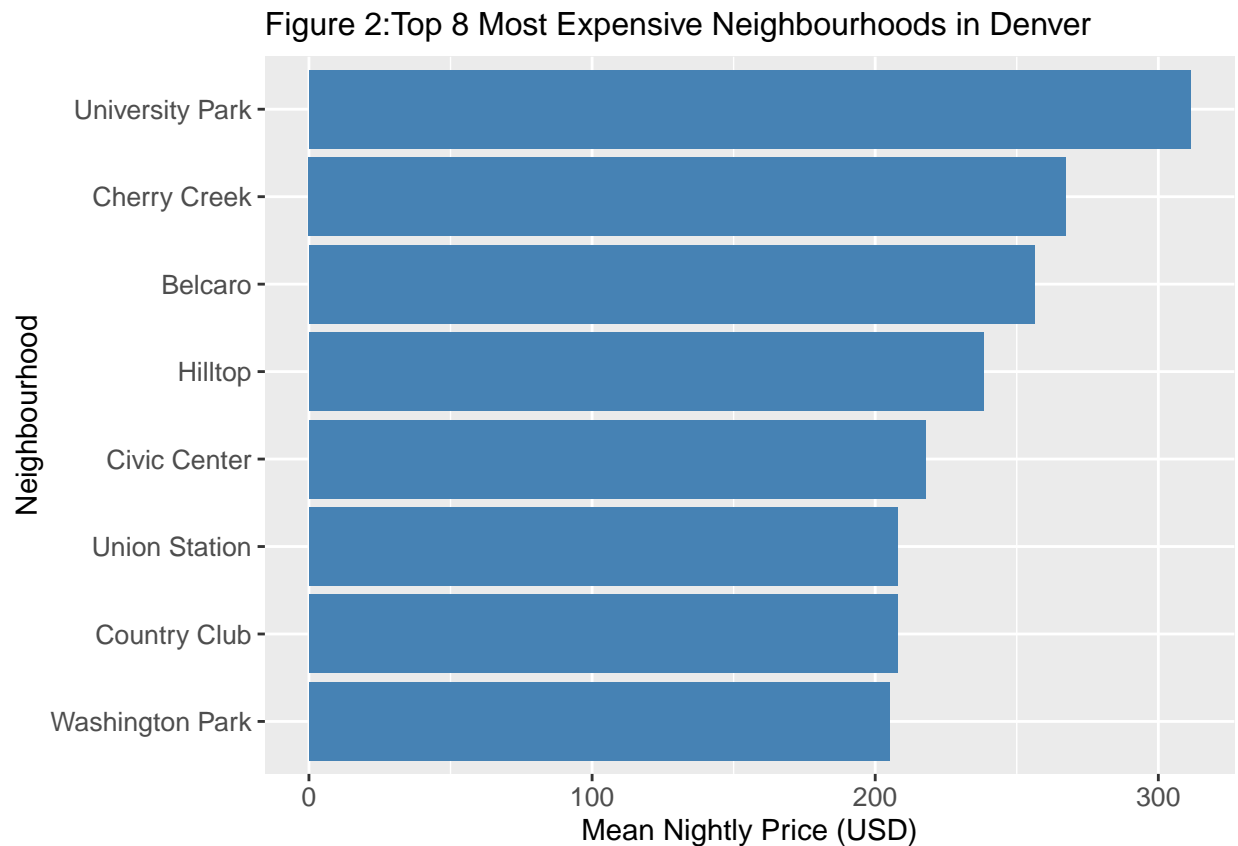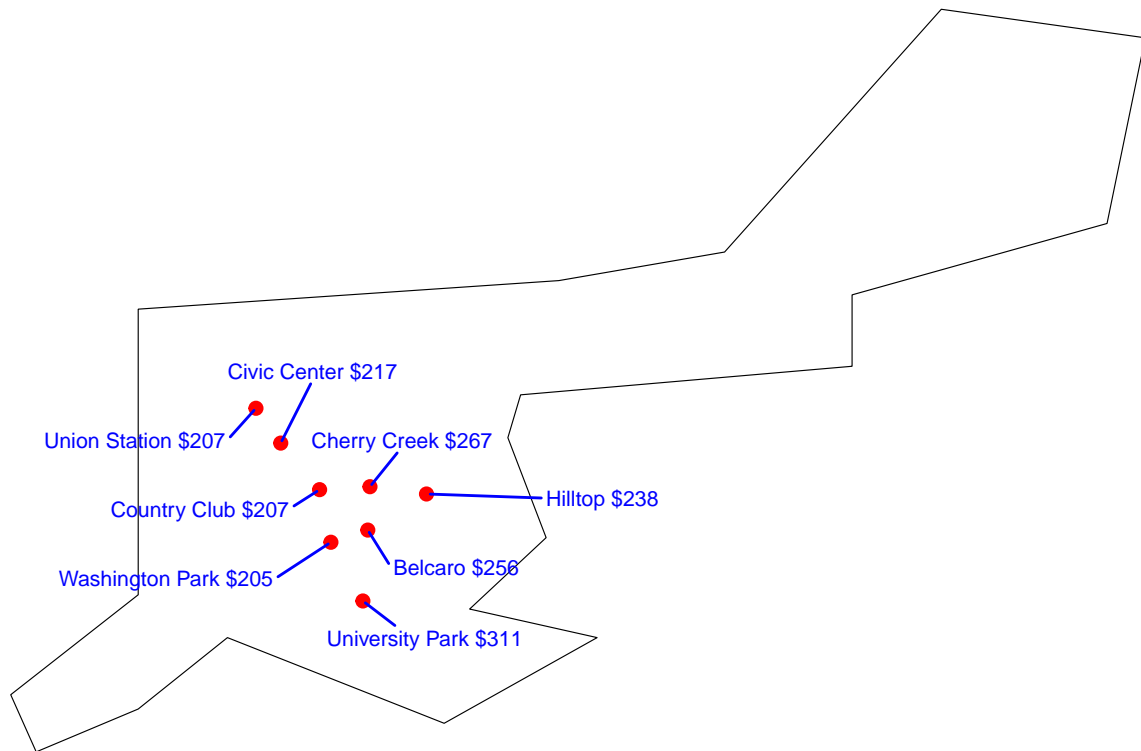Figure 2:Top 8 Most Expensive Neighbourhoods in Denver

Figure 2 shows the top eight neighborhoods in Denver with the highest average Airbnb prices. We can see that the University Park is the most expensive, with average prices exceeding 300$ per night. Cherry Creek and Belcaro follow next, as both of these neighbourhoods offer high quality housing and strong amenity access. Most of the remaining neighbourhoods, like Civic Center and Union Station, have average housing prices between 200$ and 250$. From this figure, we can see that some neighborhoods have a higher average nightly price likely due to factors such as location, high-quality housing, stronger amenities and property size.

Figure 3: Top 8 Most Expensive Neighborhoods in Denver (Average Price)

Civic Center $217

Union Station $207   Cherry Creek $267

Country Club $207   Hilltop $238

Washington Park $205   Belcaro $256

University Park $311

The map above (Figure 3) shows where the top 8 most expensive neighborhoods are located within Denver. Each red point represents each neighborhood's approximate location. Most of the expensive neighborhoods are located in the southeast, which includes areas like University park, Washington Park and Cherry Creek. Other expensive neighborhoods are located in the downtown area, such as Civic Center and Union Station.Denver boundaries show that these neighborhoods are located within the core of Denver rather than the edges.This pattern shows that location has a strong effect on average nightly prices because neighborhoods that are located in the center usually have better access to amenities and transportation. This shows that location can influence Airbnb prices. However, we need more analysis beyond the average nightly prices to confirm how location affects Airbnb prices.

Interpretation of the Location-Only GAM Model

GAM Model 1: Predicted Airbnb Prices in Denver Using Location Only
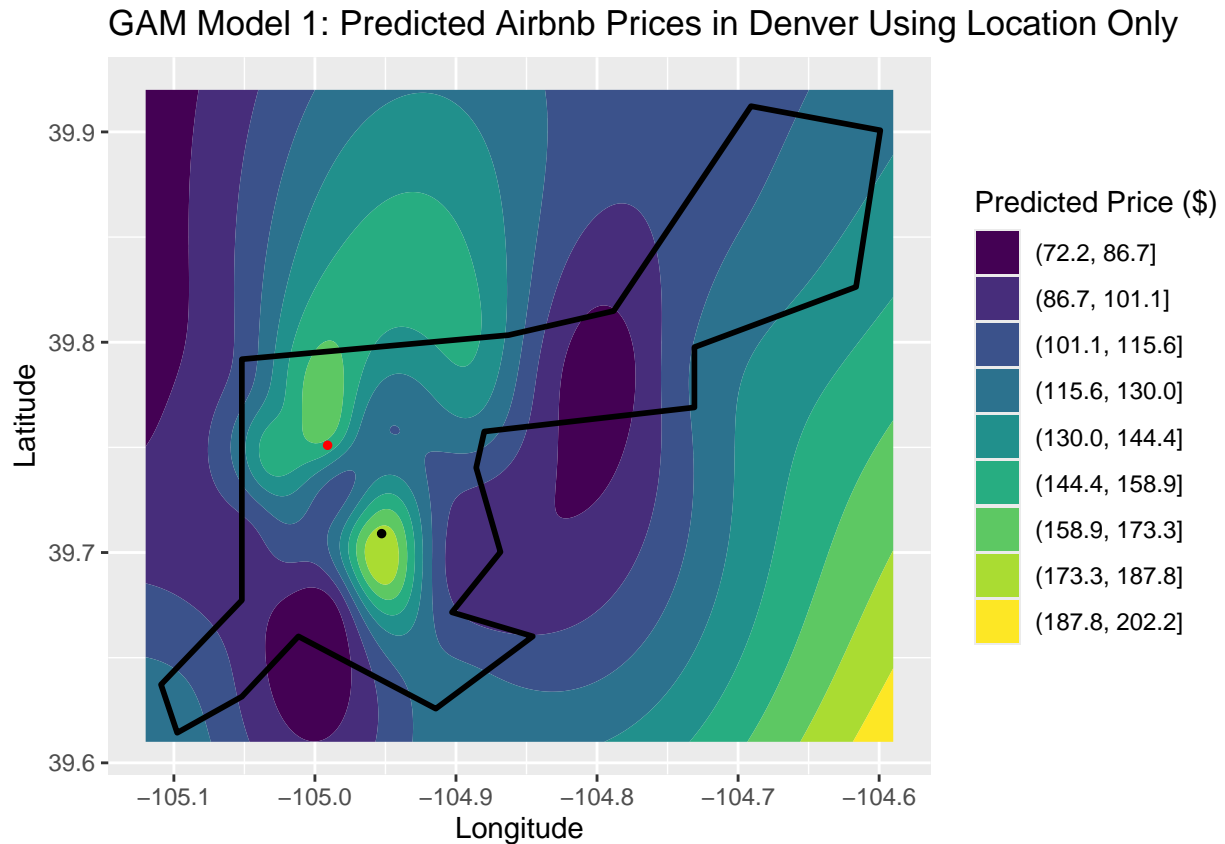
The plot (GAM Model 1) shows how the GAM Model predicts Airbnb prices using location only. The colors represent different price levels. The red point marks the Downtown area, and the black point marks the southeastern neighborhoods which include neighborhoods like Washington Park, University Park and Cherry Creek. The model shows that Downtown Denver near the red point is expensive, but it's not Denver's most expensive area. The model shows that the highest predicted prices are located in the southeastern neighborhoods near the black point, and this is likely because this area is:

1. close to the University of Denver (parents visiting their students)

2. quiet, safe and residential

3. close to Cherry Creek (luxurious commercial area of Denver)

4. Surrounded by beautiful mountains.

5. Apartments and houses are much larger on average, which pulls the average up.

Interpretation of the Second GAM Model (With Added Features)

Predicted Airbnb Prices After Controlling for Additional Features
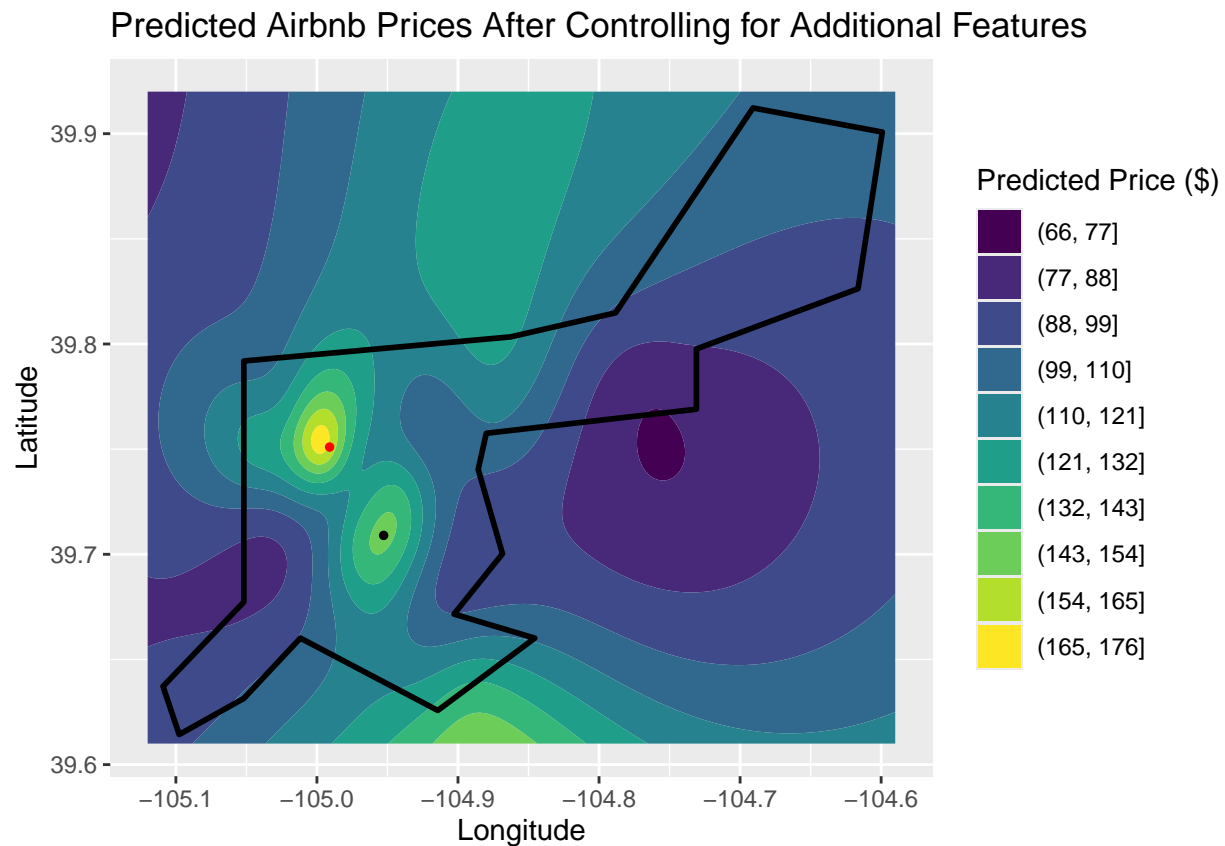
The plot above shows the predicted Airbnb prices after adding more variables to the GAM model, including the number of bedrooms,number of guests allowed,room type and review score. The red point marks the downtown, and the black point marks the southeastern area.

In the first model (GAM Model 1),the southeast area showed the most expensive part of Denver. This is because the model was using the location only. The model noticed that the southeastern area has larger listings on average. Since larger listings are usually more expensive on Airbnb, the model assumed that the entire southeastern area is expensive.

However,once we add other features like number of bedrooms and room type, the model predicts differently. The GAM Model 2 compares the prices of listings with the same type of features. With this fair comparison, the model shows:

- Downtown Denver becomes the most expensive location for a typical Airbnb listing.

- The red point that marks the Downtown area is inside the brightest yellow area showing the highest predicted price.

- This means a median-sized apartment (with the same number of bedrooms, the same accommodations and the same room type) is worth more in downtown than in Washington Park or University Park.This is because the Downtown Area has:

1. High demand from tourists and business travelers

2. strong walkability and better transportation options

3. More things to do within walking distance

In summary, these results answer our main question: the southeastern area is not expensive because the area has the most valuable location premium. The southeast appeared more expensive in the location-only model because the area has larger listings on average. So once we compare the the same listings, we can see clearly that Downtown Denver has the most valuable location.

Interpretation of the Difference Map (Full GAM - Location-Only GAM)

Difference Map: Full GAM vs Location–Only GAM
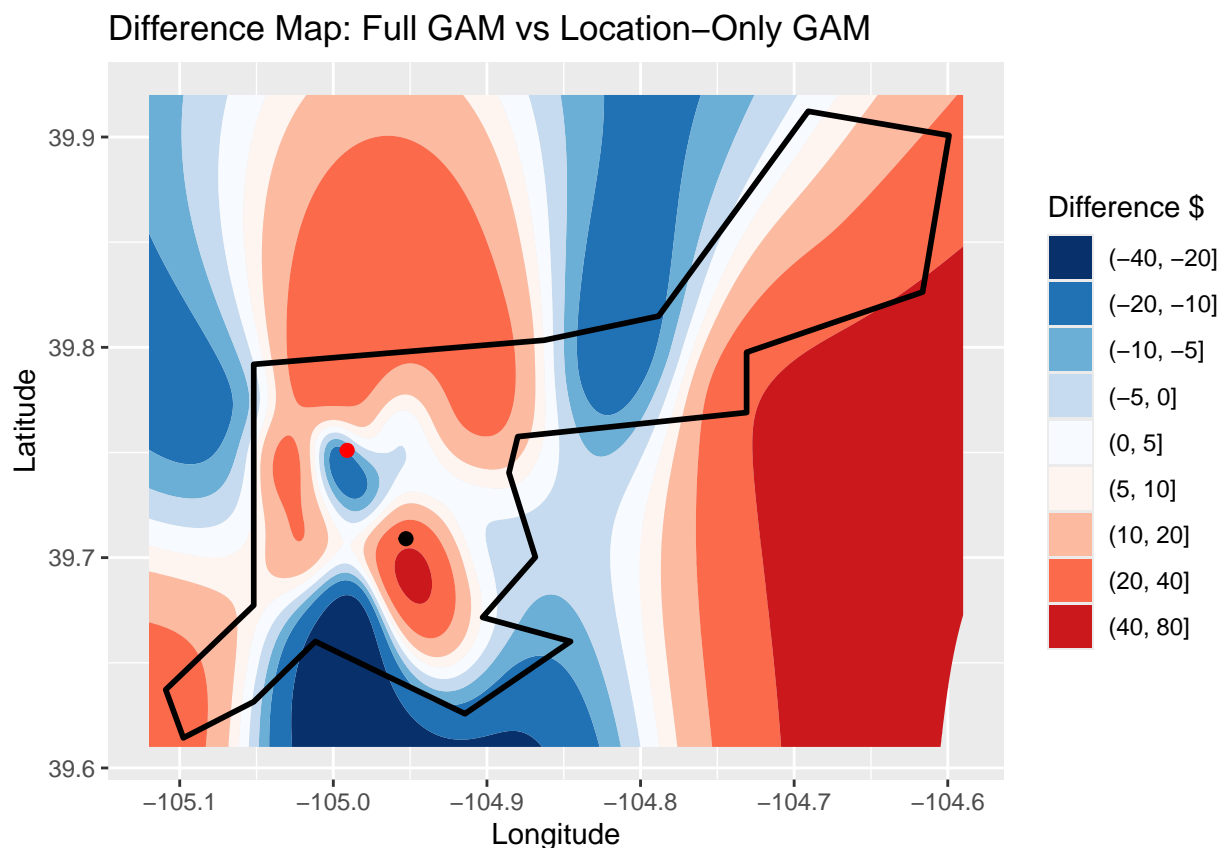
The map plot above shows the difference between the full GAM model (GAM Model 2) and the location-only GAM model (GAM Model 1). The colors tell us:

- Blue color: The full GAM model predicts higher prices than the location-only model.

- Red color: The location-only GAM model predicts higher prices than the full GAM model.

- White color: both full and location-only models predict similar prices.

The red point marks the Downtown area, and the black point marks the southeast area, including Washington Park, University Park and Cherry Creek.From the map we can see:

- The Downtown area is mostly blue, which means that the model predicts higher prices once we control for features like the number of bedrooms, room type, etc. However, when we compare identical listings, the Downtown area becomes the most expensive because of its strong location premium.

- The southeastern area is mostly red, which means the location-only model predicts higher prices than the full GAM and this is because the location-only model sees that the area has bigger houses on average, which pushes the average prices up. However, when we add the features to the full GAM model, the model is no longer assuming that the area is expensive.

In other words, red areas show where the location-only model overestimated prices because it confuses big houses with expensive locations. On the other hand, blue areas show where the full GAM model adjusts prices after comparing identical listings.Overall, These results are useful in real life. If we only look at the average nightly prices, we might think that the southeast is the most expensive area in Denver. However, our model shows that the southeast has the highest average nightly prices only because the area has larger houses on average and the downtown area has the most valuable location in Denver. For Airbnb hosts, this means that small or medium listings are worth more in Downtown Denver than in the southeast area. Overall, the results show us that we should separate the effects of location and home features before making decisions based on simple averages.

Note: The model worked much better when we added more features. The full GAM had an R-squared of 0.58, which means the model explains about 58% of the differences in prices. However, the location-only model had only 0.07, which means that it explains only 8% of the variation (see Appendix for details).
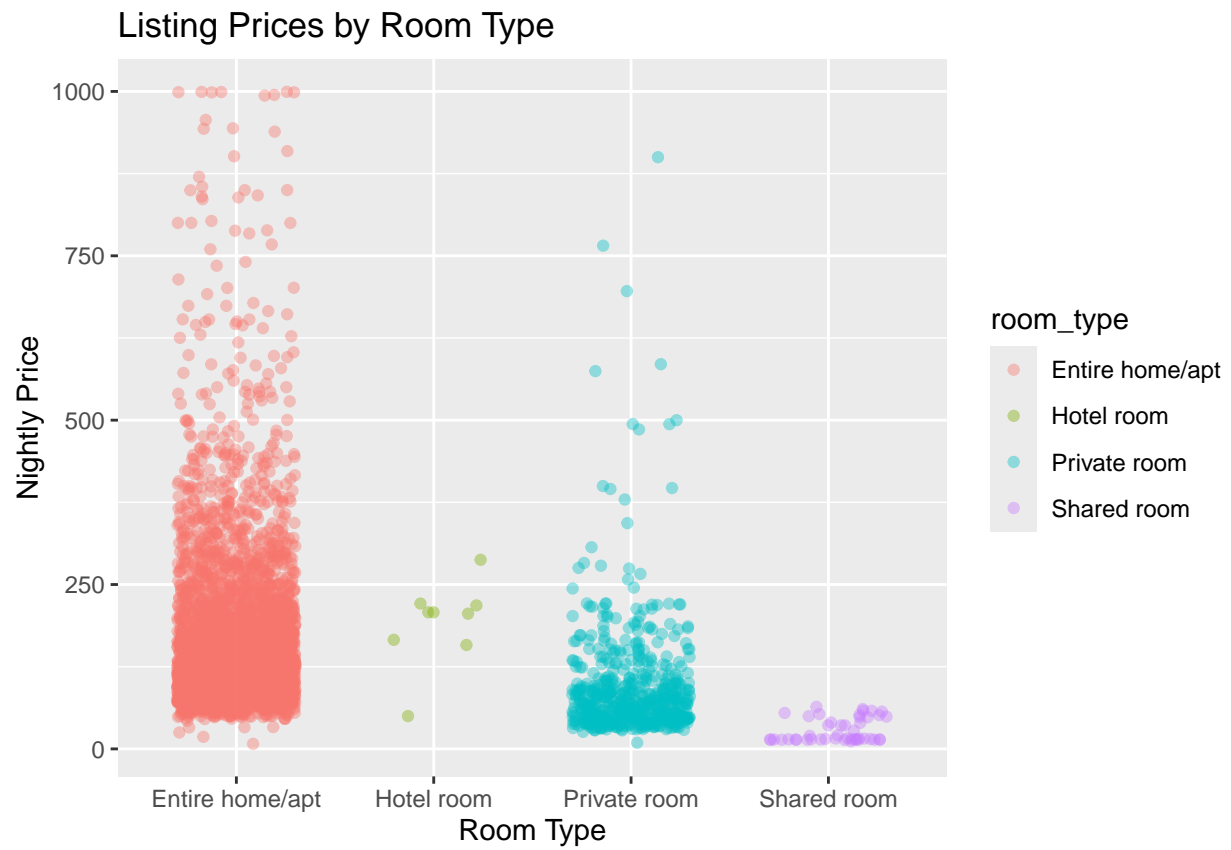
Conclusions

The project shows that the southeast of Denver looks expensive because the listings in that area are larger. When we fitted a GAM model based on location only, the southeast appeared to be the most expensive area in Denver. However, when we add other features like bedrooms, accommodations, room type, and reviews, the results change. After adding the features to the second GAM model, the results show that downtown is the most expensive area in Denver. This means that the exact same listing is worth more downtown than in the southeast. The southeast is expensive only because the area has larger listings on average, not because it has the best location value. Overall, we should not rely on simple averages to decide whether a certain area is more expensive than the other.
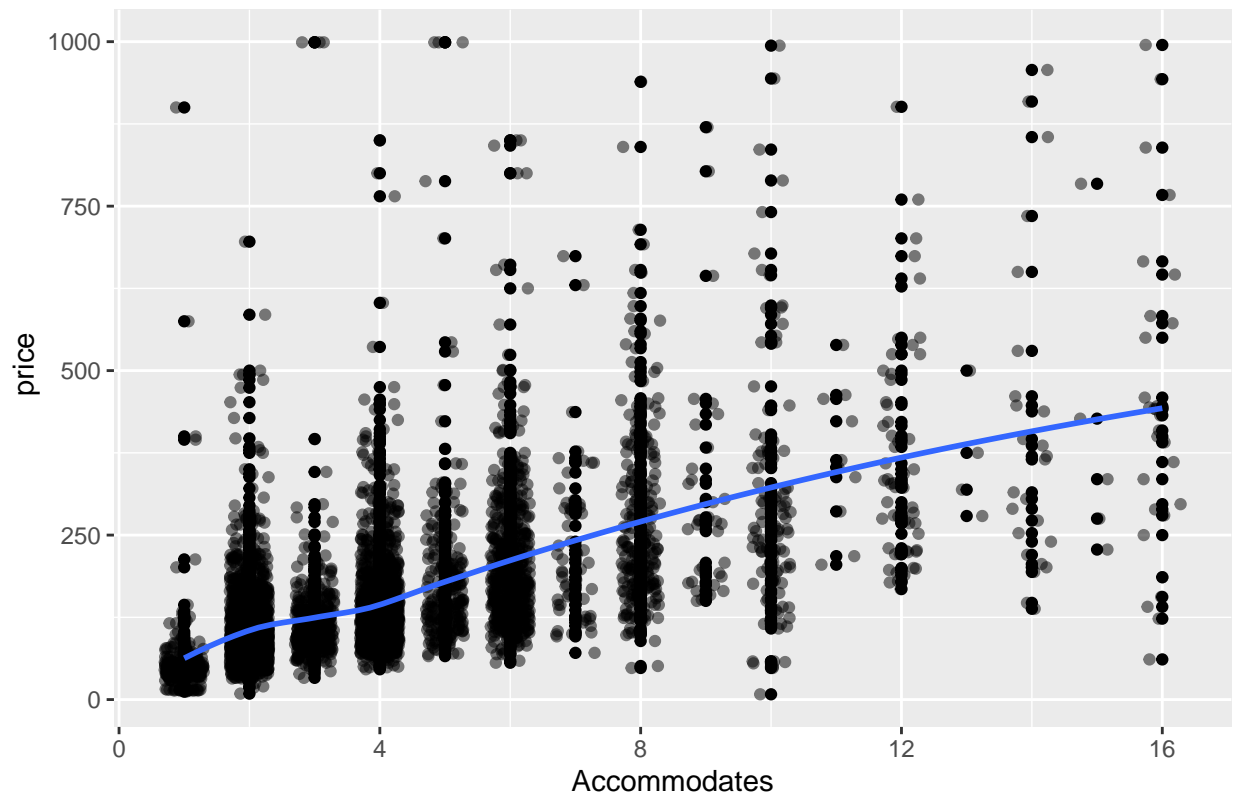
Limitations and Future Work

The project has some limitations; when we fitted the models, we only used features like number of bedrooms, number of guests the listing can host, room type and review scores. There are some important factors that are not included in the models, such as the area of the property, the year the property was built and how much the property is worth in the market. Also, we didn't account for how prices change during the seasons. Airbnb prices usually are more expensive during peak periods like holidays or during summer.

In the future, we can add additional features to our models, such as the number of bathrooms, the year of build, the area of the property and transportation accessibility. These features will help us understand how prices vary from place to place in the city. Also, it would be useful if we studied how the prices change over seasons to see whether some neighborhoods become more expensive during a specific season or not.
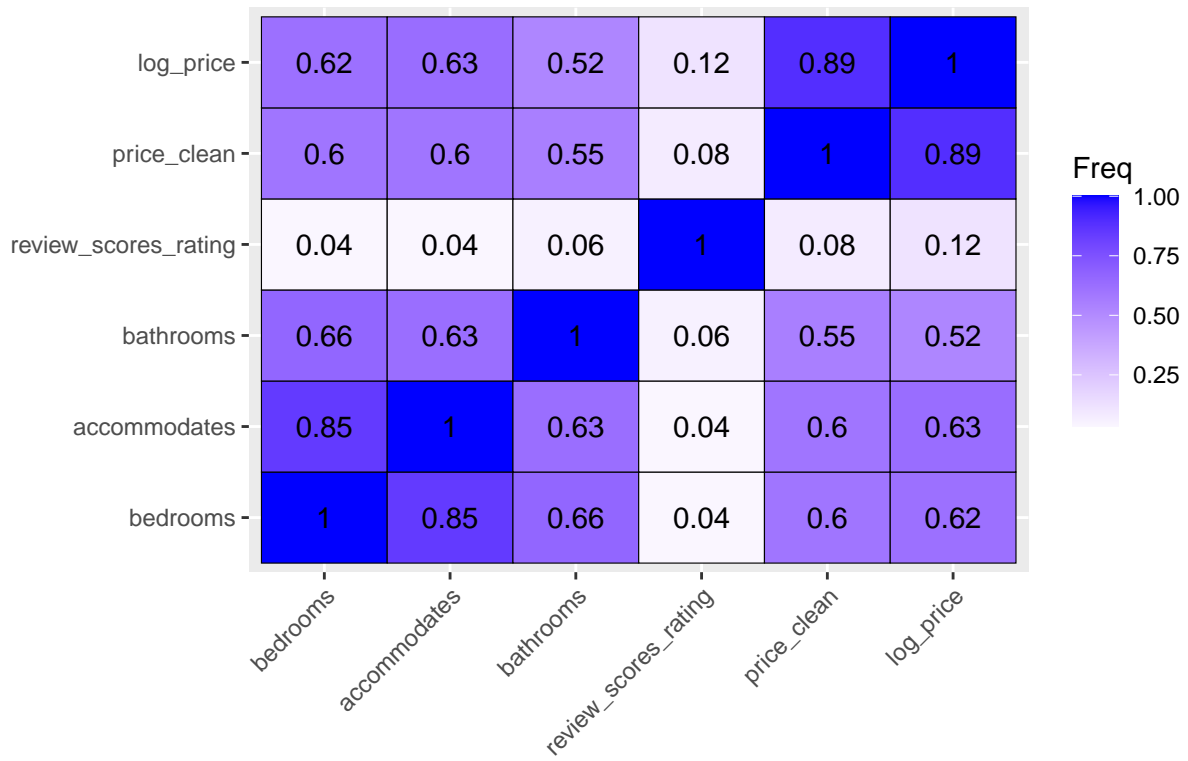
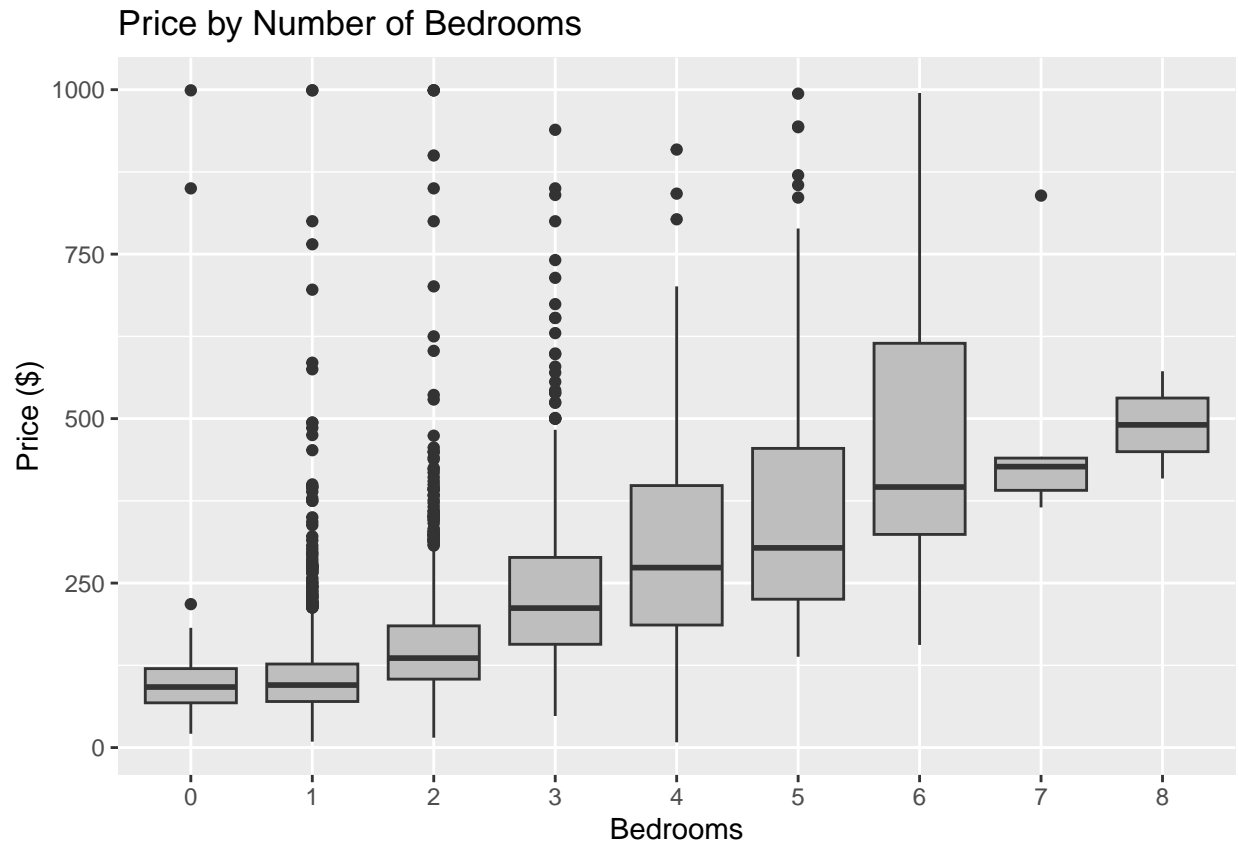**Appendix A: EXPLORATORY DATA ANALYSIS**

Listing Prices by Room Type

Price by Number of Guests

Correlation Heatmap

	bedrooms	accommodates	bathrooms	review_scores_rating	price_clean	log_price
log_price	0.62	0.63	0.52	0.12	0.89	1
price_clean	0.6	0.6	0.55	0.08	1	0.89
review_scores_rating	0.04	0.04	0.06	1	0.08	0.12
bathrooms	0.66	0.63	1	0.06	0.55	0.52
accommodates	0.85	1	0.63	0.04	0.6	0.63
bedrooms	1	0.85	0.66	0.04	0.6	0.62

Freq
1.00
0.75
0.50
0.25

Price by Number of Bedrooms

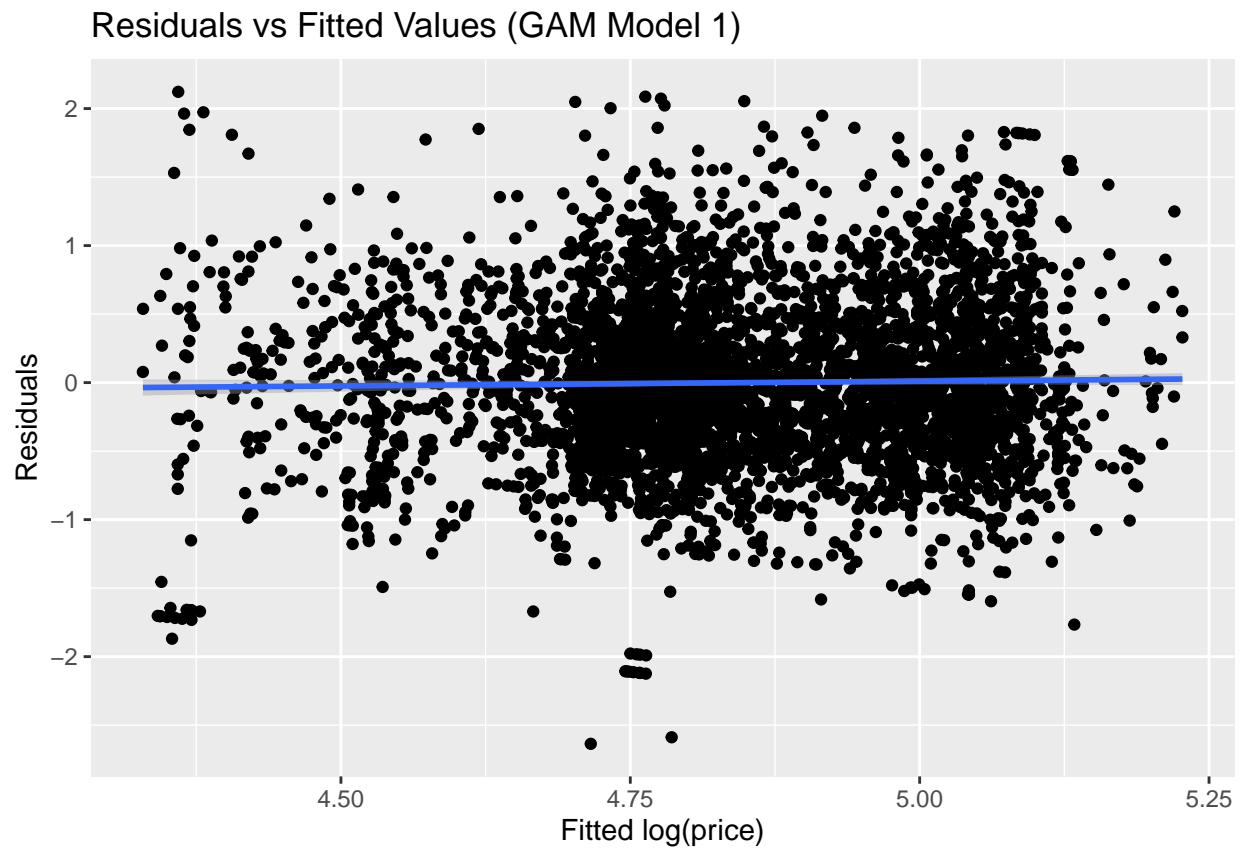## Appendix B: MODEL RESULTS AND DIAGNOSTICS

Summary of Model 1:

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## log(price_clean) ~ s(longitude, latitude)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.852403   0.009213   526.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                       edf Ref.df      F p-value
## s(longitude,latitude) 25.42  28.22 12.67  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0777   Deviance explained = 8.32%
## -REML = 3959.7  Scale est. = 0.36379   n = 4286
```

Summary of Model 2:

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## log(price_clean) ~ log(accommodates) + bedrooms + review_scores_rating +
##     room_type + review_scores_rating + s(longitude, latitude)
##
## Parametric coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.359735   0.116521  28.834  < 2e-16 ***
## log(accommodates)     0.362231   0.019652  18.432  < 2e-16 ***
## bedrooms              0.188670   0.009816  19.221  < 2e-16 ***
## review_scores_rating  0.153870   0.023786   6.469 1.12e-10 ***
## room_typeHotel room   0.357459   0.134229   2.663  0.00778 **
## room_typePrivate room -0.151264   0.022425  -6.745 1.76e-11 ***
## room_typeShared room  -1.172015   0.077562 -15.111  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                       edf Ref.df     F p-value
## s(longitude,latitude) 25.33  28.18 22.35  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.577   Deviance explained = 58.1%
## -REML = 1932.7  Scale est. = 0.15904   n = 3751
```

Residuals Plot of Location-Only GAM

Residuals vs Fitted Values (GAM Model 1)

Residuals Plot of Full GAM

Residuals vs Fitted Values (GAM Model 2)