Housing Data of Boston Area

BC2406 Analytics

Group 1 as State Housing Agency

Who is this data for?

IMMIGRANTS

For those to find desirable living areas based on personal preferences

ESTATE DEVELOPERS

To scout suitable areas for housing developments and population targets

Our target analyses?

Relation between certain housing factors

Top housing factors influencing PRICE

Current segregation of homes in Boston

Data Source

1978 BOSTON HOUSING DATASET

- Collected by the U.S. government
- 506 observations and 13 attributes
- Variable examples: *capita crime rate, average number of rooms per dwelling, proportion of non-retail business acres per town*
- 2 outliers were removed during the preliminary linear regression

Housing Recommendation

- Problem: It's hard for new-comers in Boston to pick the right neighborhood that suit their needs from countless towns and tracts
- Solution: A housing recommendation function

```
> # Example of customer input and recommendation
> # For a customer who has a budget of 30 and cares most about number of rooms, lower status population and crime rate:
> recommendations <- recommend_neighborhoods(
+   budget = 30,
+   number_of_rooms = 5.5,
+   max_Lower_Status_Population = 5,
+   max_Non_retail_Business_Acres = 25,
+   max_Weighted_Distance_to_Employment = 13,
+   max_Proportion_of_Old_Units = 90,
+   max_Nitric_Oxides_Concentration = 0.7,
+   max_Property_Tax_Rate = 700,
+   max_Pupil_Teacher_Ratio = 30,
+   max_Per_Capita_Crime_Rate = 0.1,
+   min_Residential_Land_Proportion = 0,
+   min_Index_of_Highway_Accessibility = 0
+ )
>
> print(recommendations)
            town tract cmedv    crim zn indus chas   nox    rm  age    dis rad tax ptratio      b lstat
299       Sharon  4141  22.5 0.06466 70  2.24    0 0.400 6.345 20.1 7.8278   5 358    14.8 368.24  4.97
66  North Reading  3301  23.5 0.03584 80  3.37    0 0.398 6.290 17.8 6.6115   4 337    16.1 396.90  4.67
1         Nahant  2011  24.0 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
293     Westwood  4123  27.9 0.03615 80  4.95    0 0.411 6.630 23.4 5.1167   4 245    19.2 396.90  4.70
291     Westwood  4121  28.5 0.03502 80  4.95    0 0.411 6.861 27.9 5.1167   4 245    19.2 396.90  3.33
300       Sharon  4142  29.0 0.05561 70  2.24    0 0.400 7.041 10.0 7.8278   5 358    14.8 371.58  4.74
195      Bedford  3592  29.1 0.01439 60  2.93    0 0.401 6.604 18.8 6.2196   1 265    15.6 376.70  4.38
253   Framingham  3839  29.6 0.08221 22  5.86    0 0.431 6.957  6.8 8.9067   7 330    19.1 386.09  3.53
```

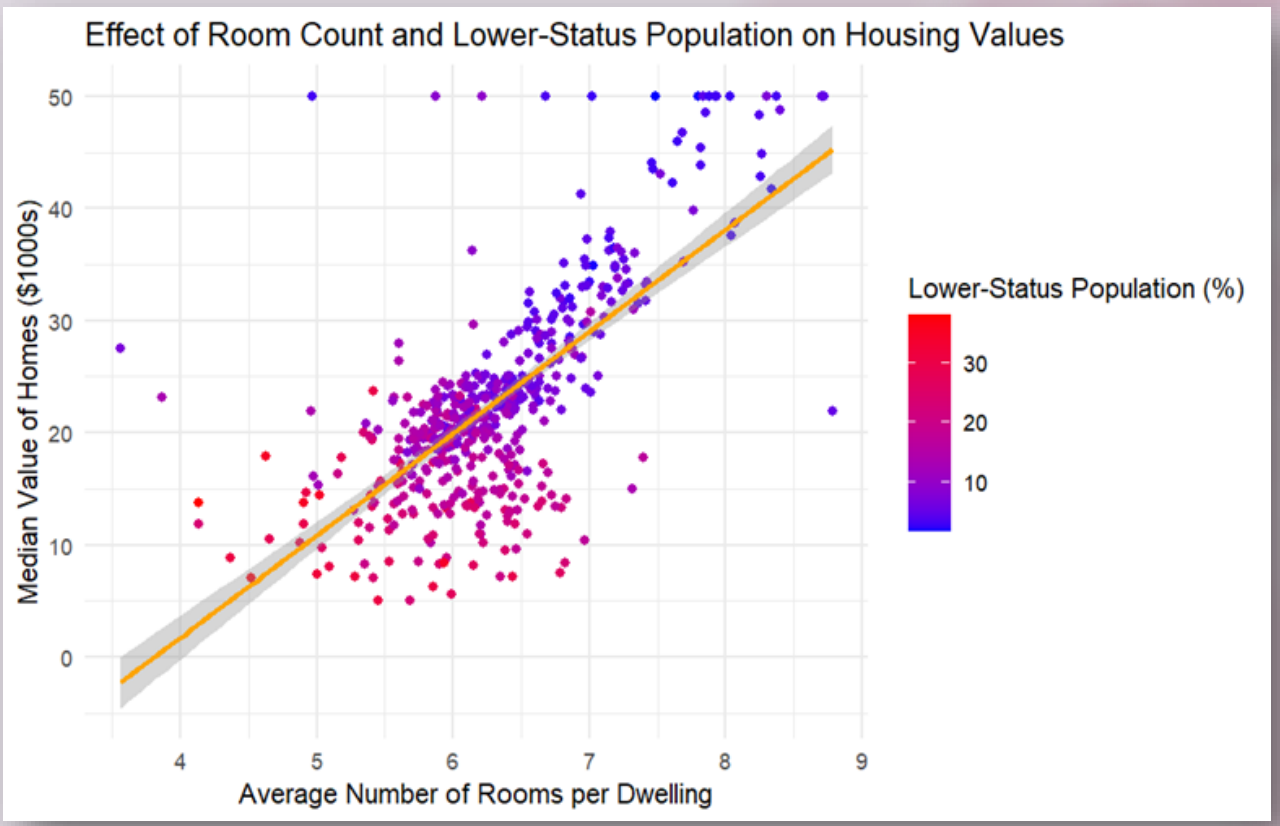
But what about houses built in newly developed neighborhoods?

Relationships to House Prices

Factors : Lower-Status Population, Average Number of Rooms Per Dwelling

FOCUS : Socio-Economic Factors and Property Values

The results indicate that in the dataset, more valuable homes are typically found in neighbourhoods with fewer lower-status residents (lstat).
Additionally, these more valuable homes tend to have more rooms (nr). This reinforces the idea that home size and socio-economic conditions are factors influencing property values.



Relationships to House Prices

Factors : Building Age, Distance to employment Centers

FOCUS : Housing Building Condition and Location

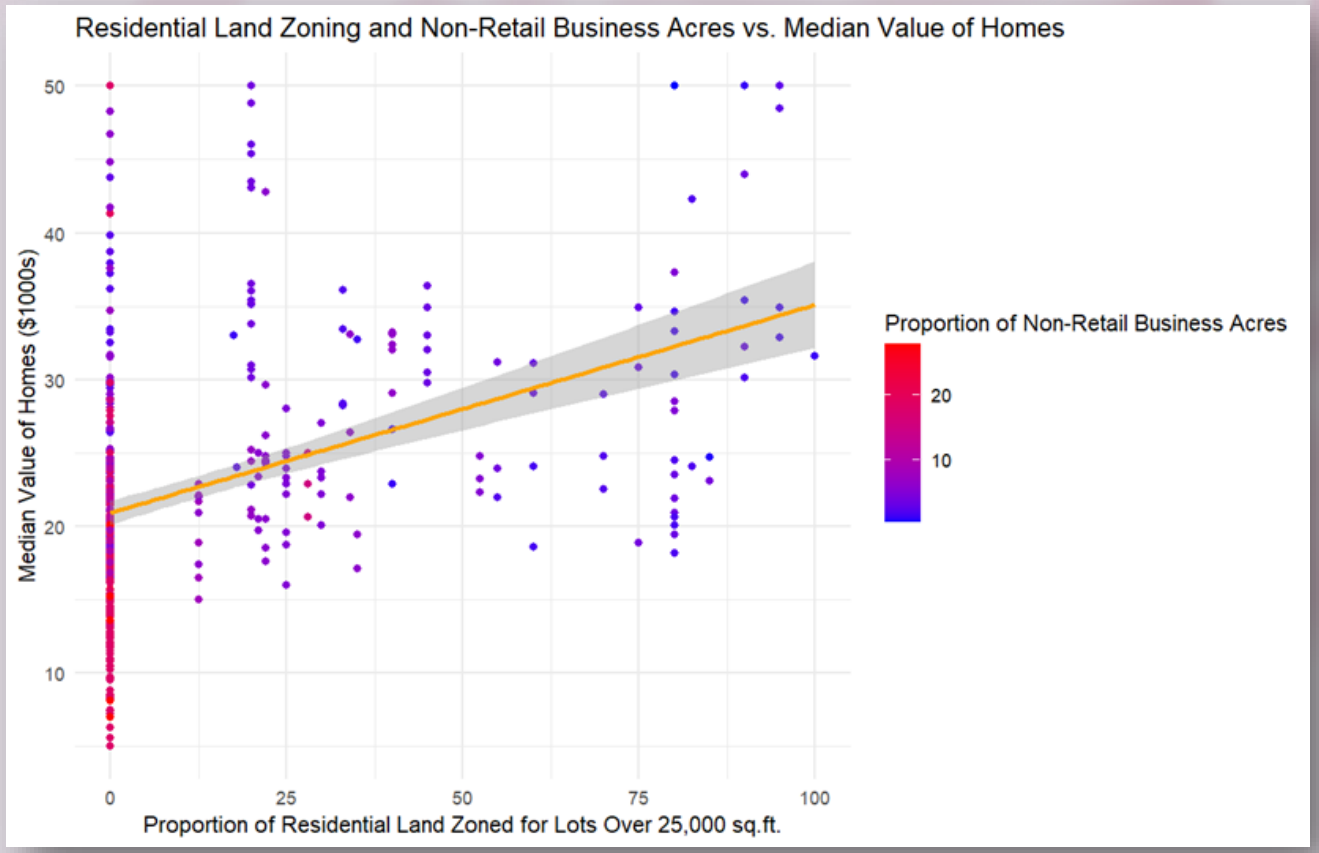
Consider a neighborhood with a large number of older homes and proximity to a major employment hub. In this area, the older homes may be less attractive to buyers due to their age and potentially outdated features, leading to lower property values.
Furthermore, the close proximity to the busy employment center might contribute to a less desirable living environment, further reducing home values.

Relationships to House Prices

Factors : Residential Land Zoning, Non-retail Business Acres

FOCUS : Land Proportionality of Housing and Business

Imagine two neighborhoods: one with a high proportion of land zoned for large residential lots and few non-retail business areas, and another with smaller lot sizes and significant commercial or industrial activity. Typically, the former neighborhood is likely to have higher home values due to the spacious and quieter residential environment, making it more attractive to homebuyers looking for a desirable living area.

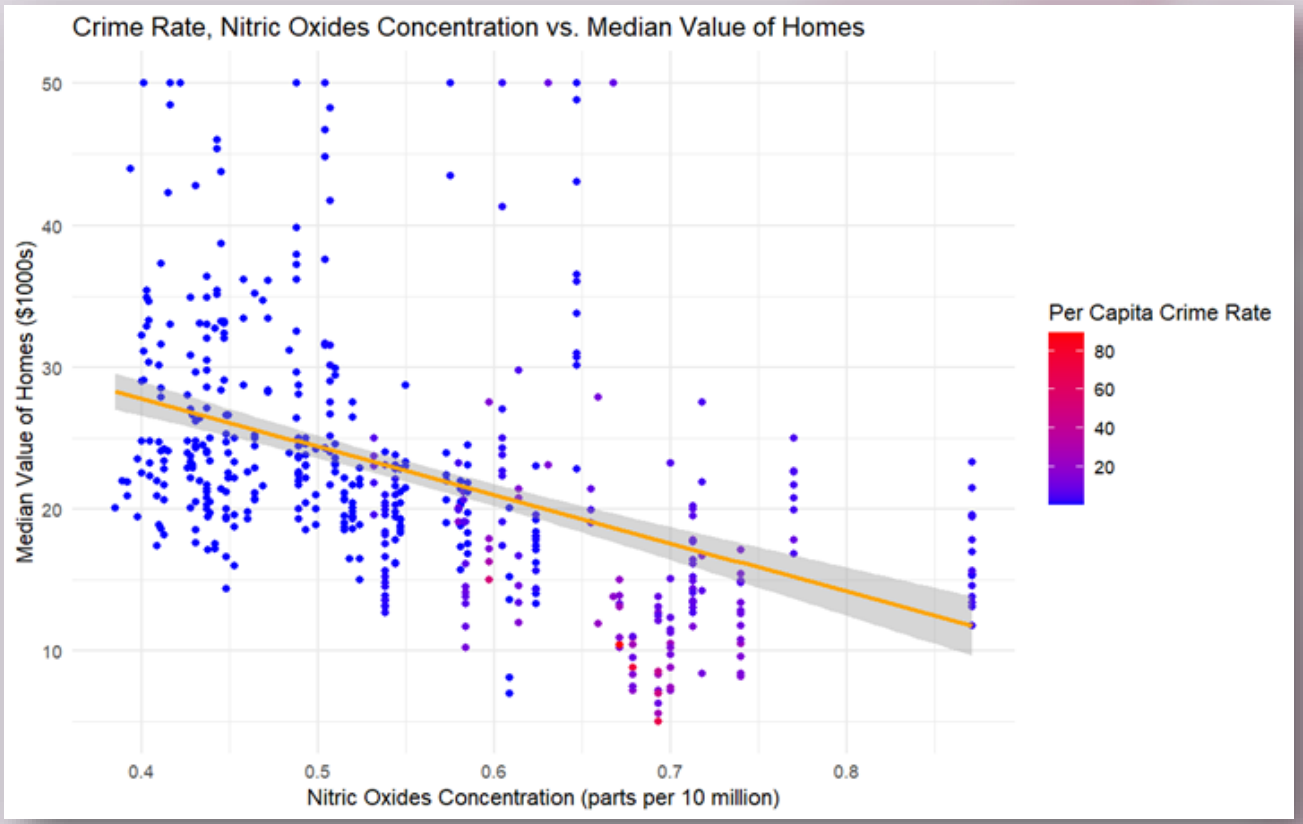


Relationships to House Prices

Factors : Crime Rate, Nitric Oxide Concentration

FOCUS : Physical and Mental Environment (crime rates and air pollution)

The analysis shows a positive relationship between environmental quality and safety with housing values. Specifically, neighborhoods with lower nitric oxides concentration and lower crime rates are associated with higher median home values. This indicates that homebuyers place a high value on both clean air and safety, and are willing to pay more for properties in such areas.



EDA - Testing Hypothesized Correlations (Linear Regression)

- Median Value Housing Price =

-0.5068(lower status population)

+ 4.239(avg no. of rooms)

-1.045(parent-teacher ratio)

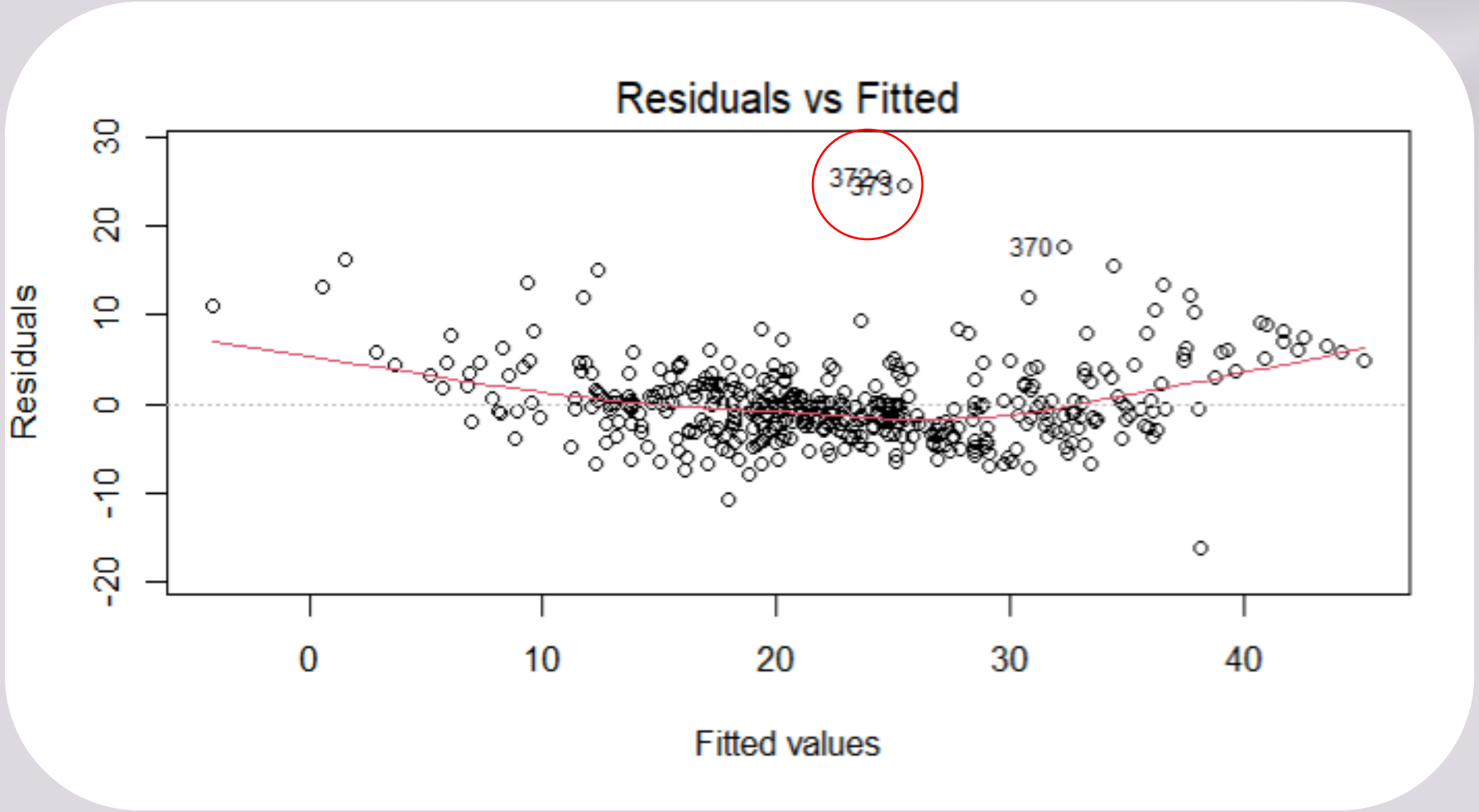
- 1.607(distance to employment centers)

+ ...

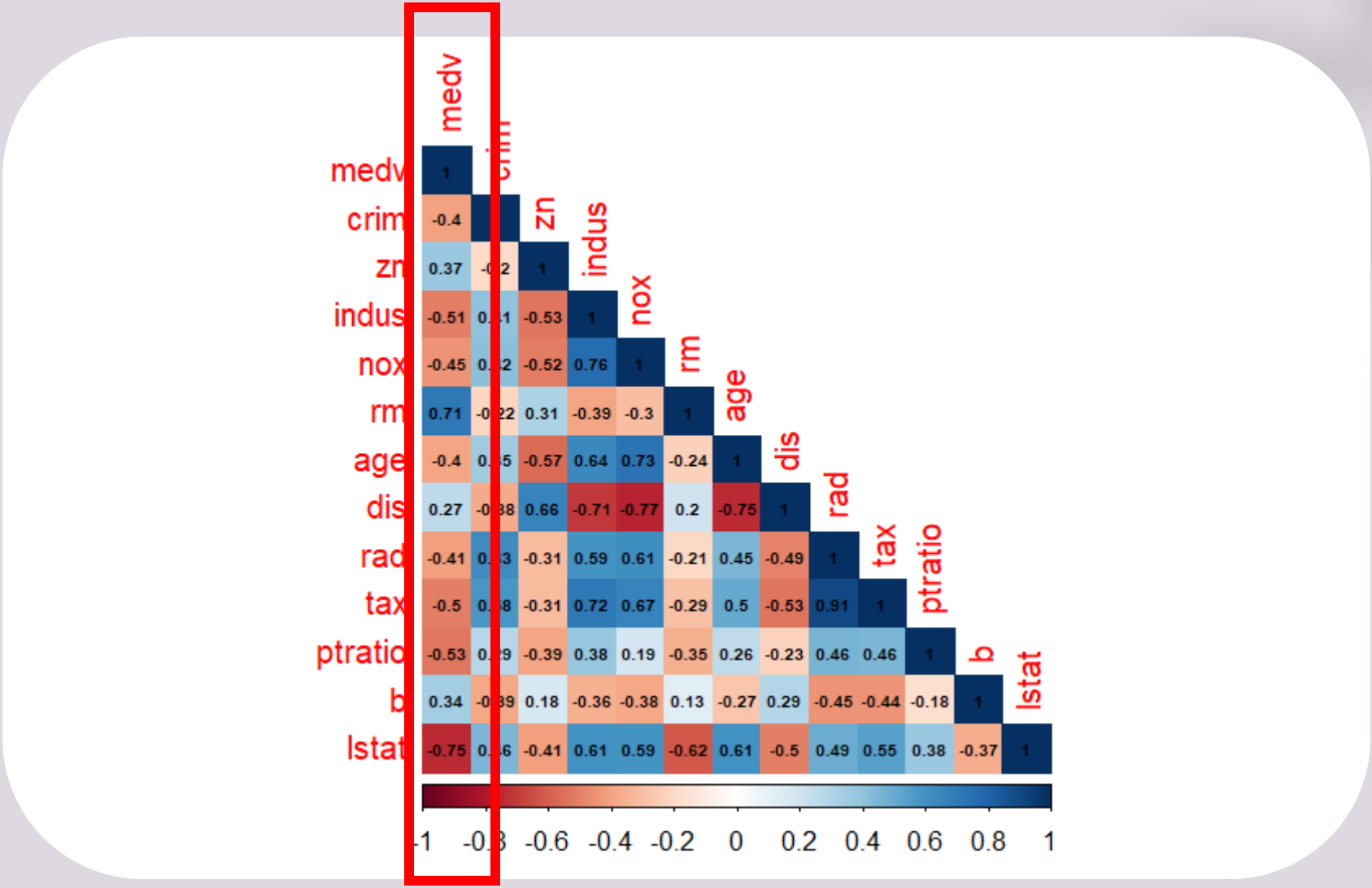
This model is able to explain 75.7% of the cases

Root-mean-squared-error: 4.9

EDA - Testing Hypothesized Correlations (Outliers)



EDA - Correlations Table



EDA - Testing Hypothesized Correlations (Improved Linear Regression)

EDA - Testing Hypothesized Correlations (Improved Linear Regression)

- Median Value Housing Price =

-0.426(lower status population)

+ 4.92(avg no. of rooms)

-0.892(parent-teacher ratio)

+ ...

This model is able to explain 79.2% of the cases

Root-mean-squared-error: 4.7

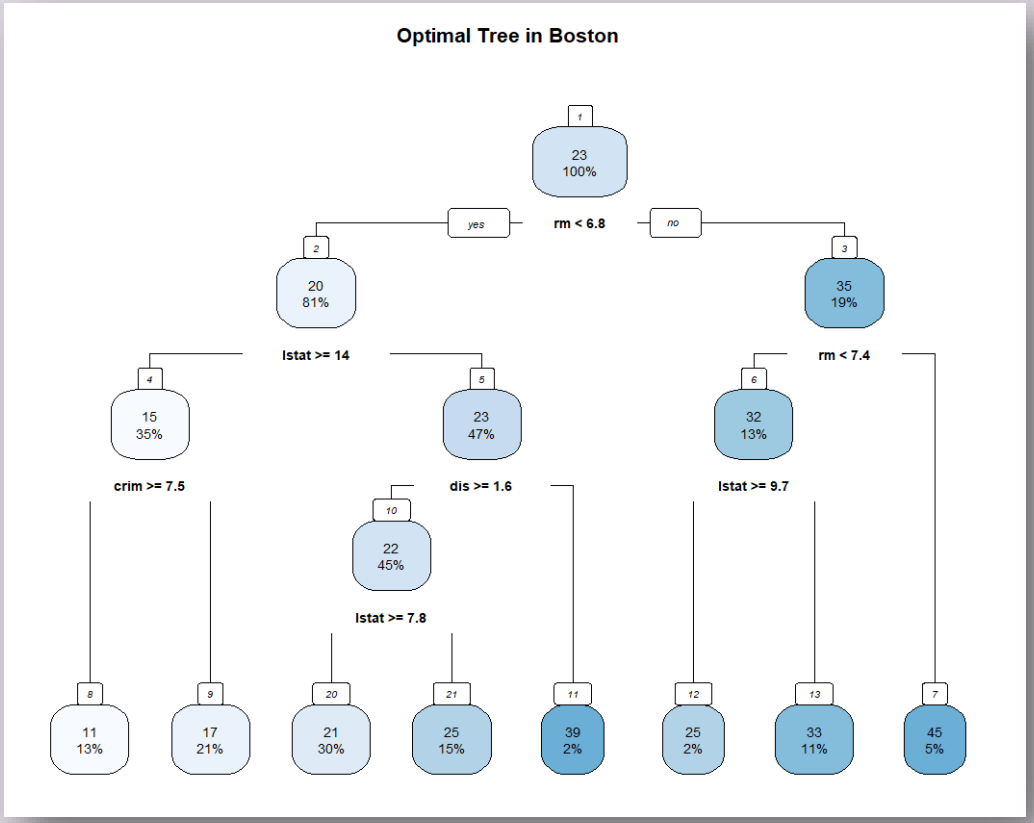
Most important variables in descending order:

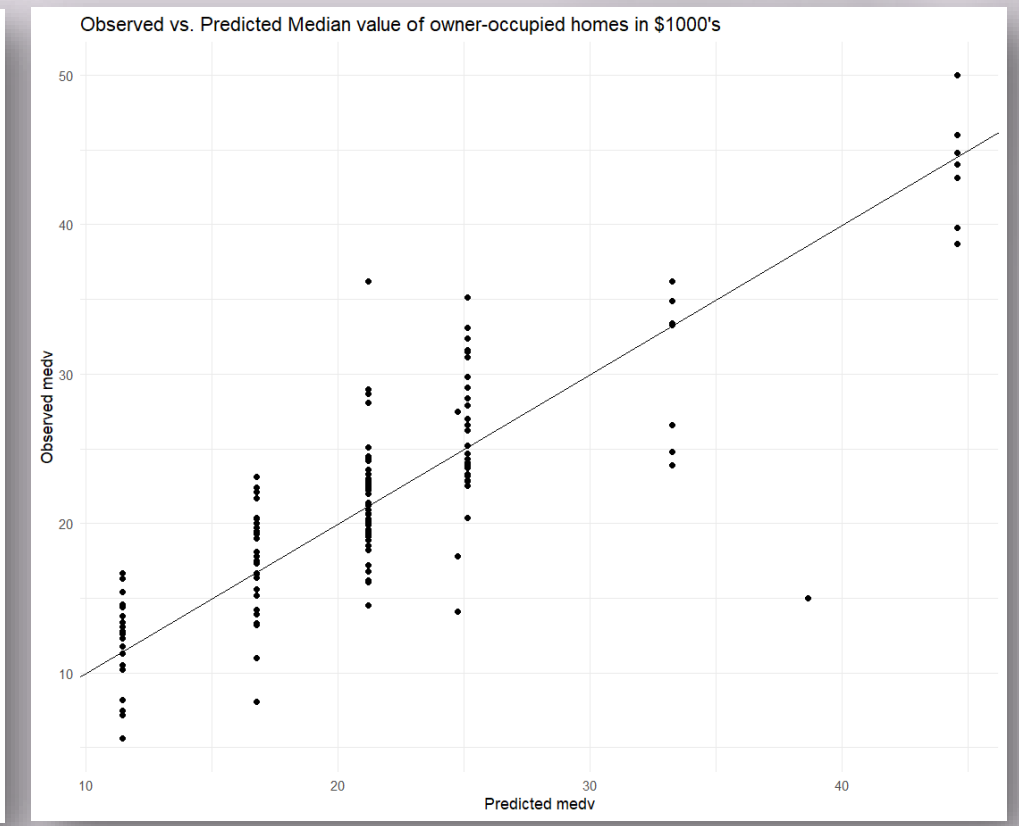
1. Lower Status Population
2. Average Number of Rooms per Dwelling
3. Parent – Teacher Ratio

Housing Factors Analysis (CART)

City expansion and new houses

- Problem: New houses are being built in newly developed neighborhoods as Boston expands, and people don't know whether these houses can fulfill their preferences or if they are affordable.
- Solution: A CART model to predict housing prices



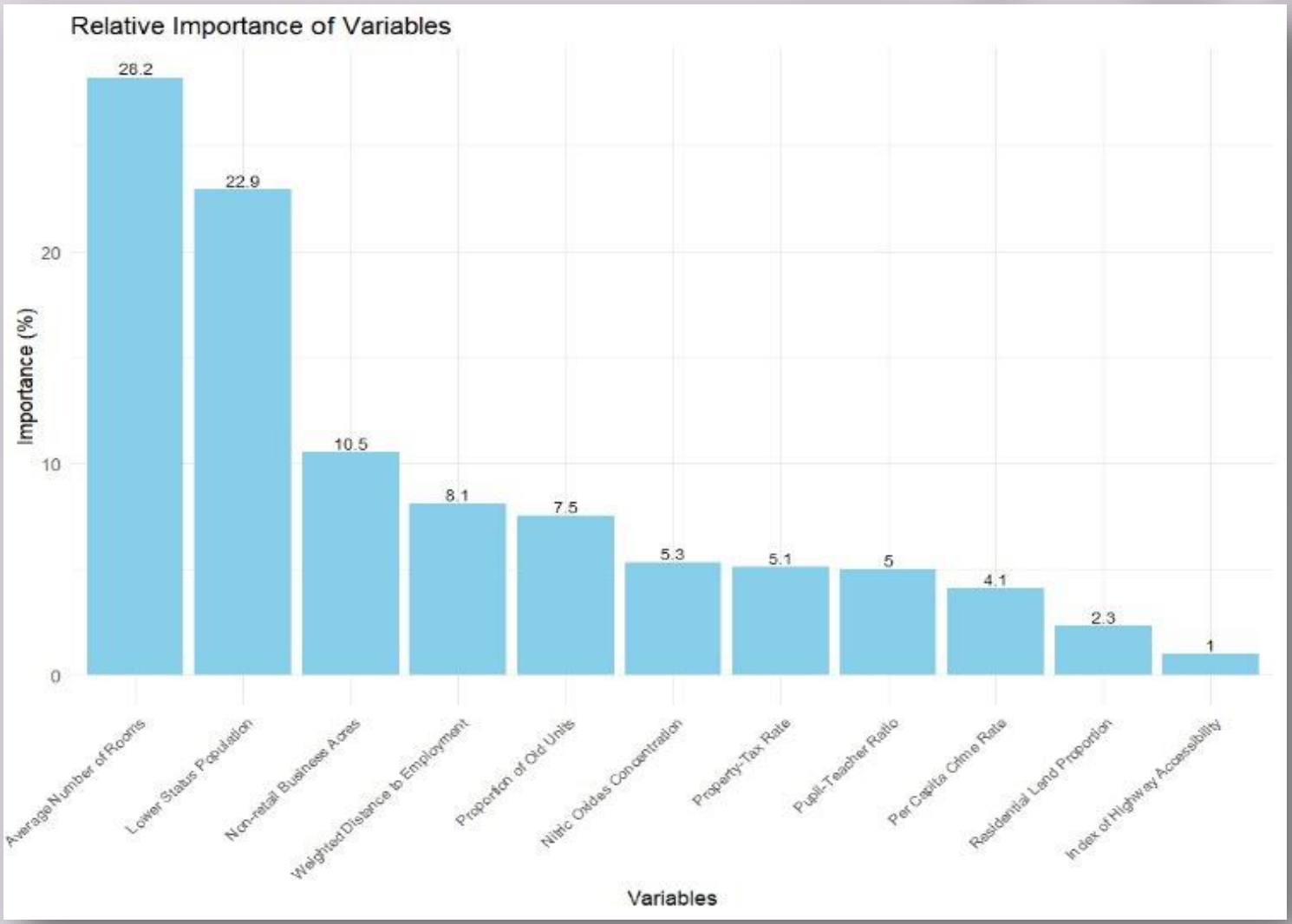


Housing Factors Analysis (CART)

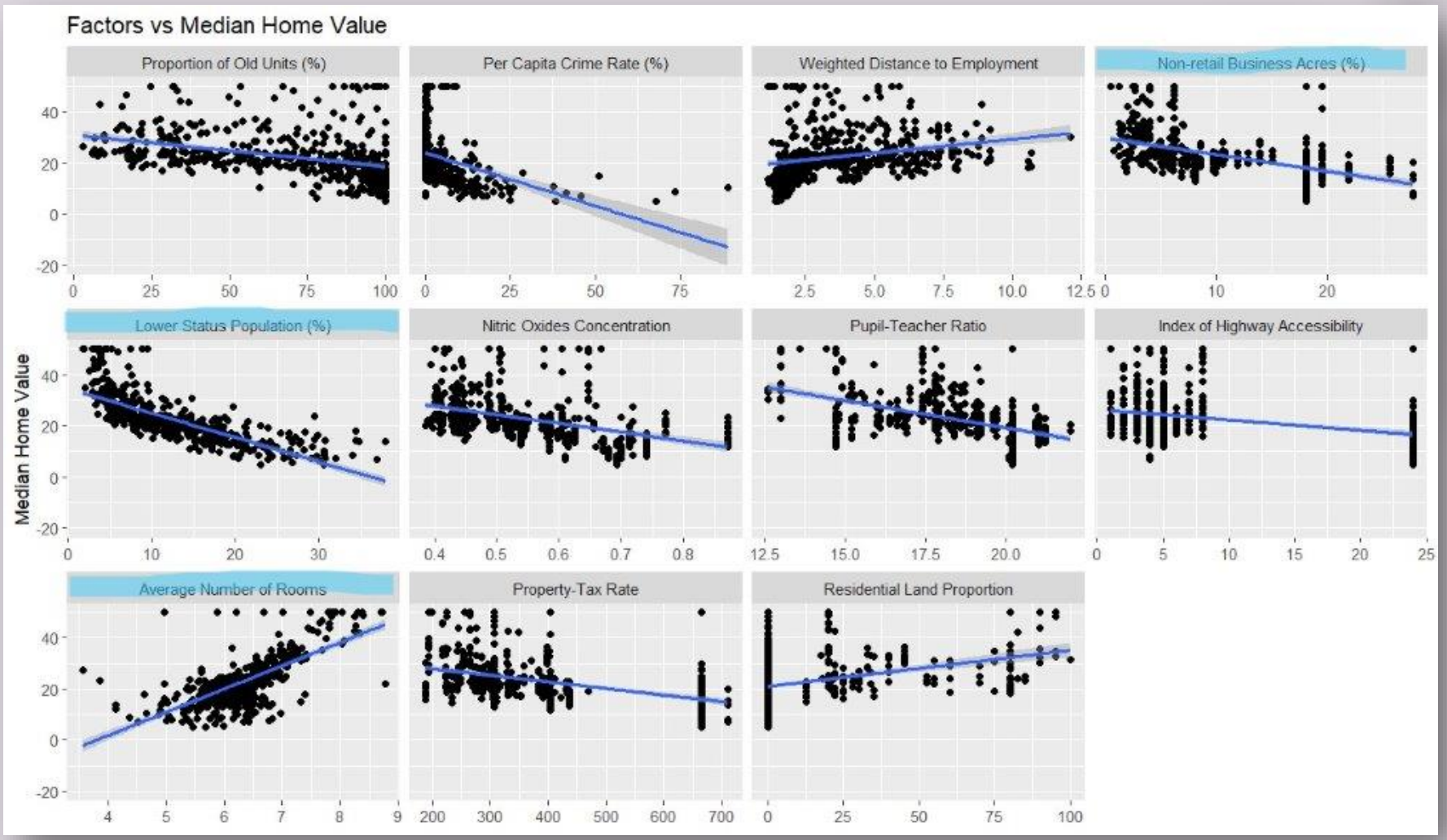
Our CART Model takes all variables and ranks **their *importance* in predicting median price values of Boston houses**.

Real estate developers can focus on the more significant variables when choosing locations to build new dwelling areas.

It also benefits house buyers as more houses will be built to cater their needs, instead of based on the judgement of real estate developers, which may not really capture what residents care most about.

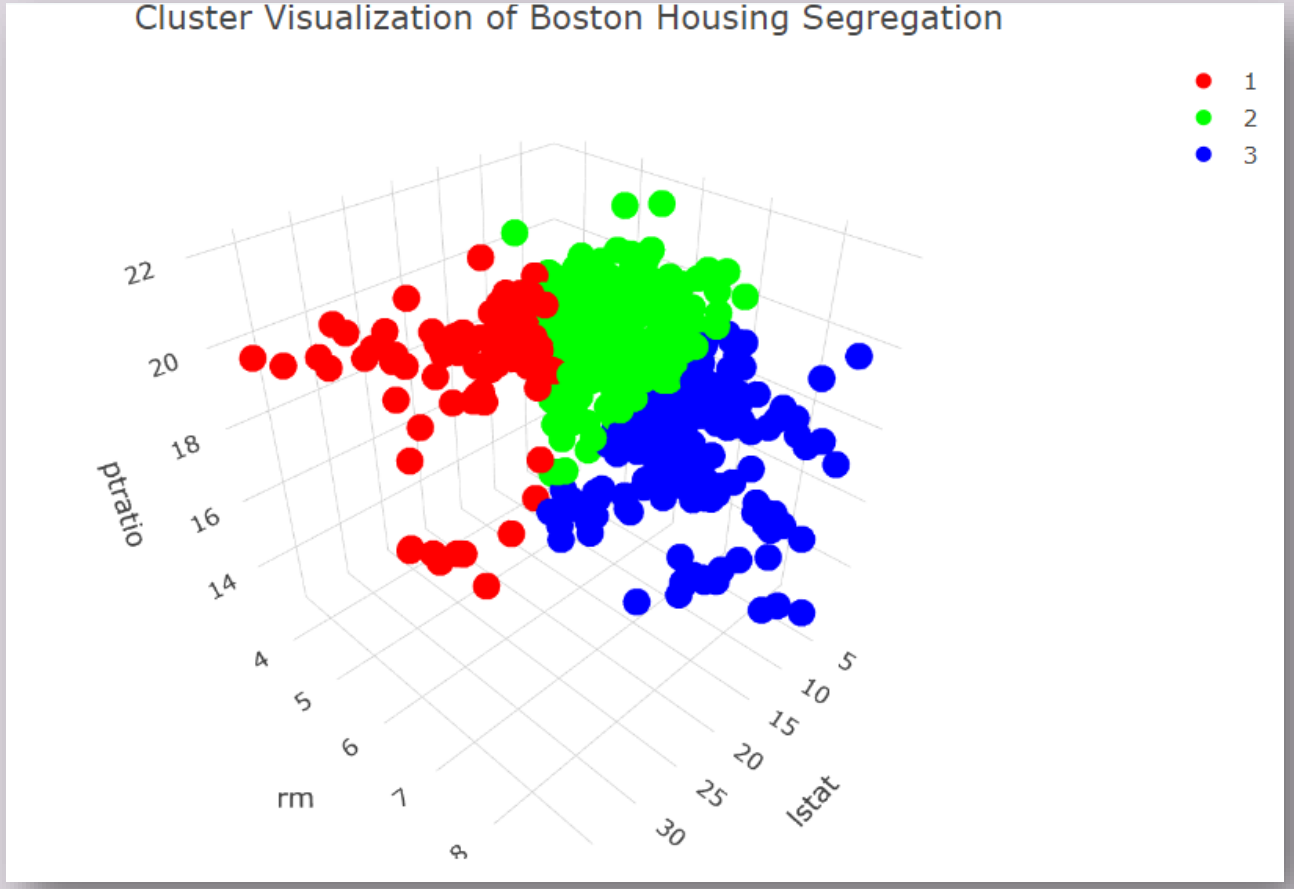


Housing Factors Analysis (CART)



FACET CHART VISUALIZATION

How are houses in Boston *currently* clustered?



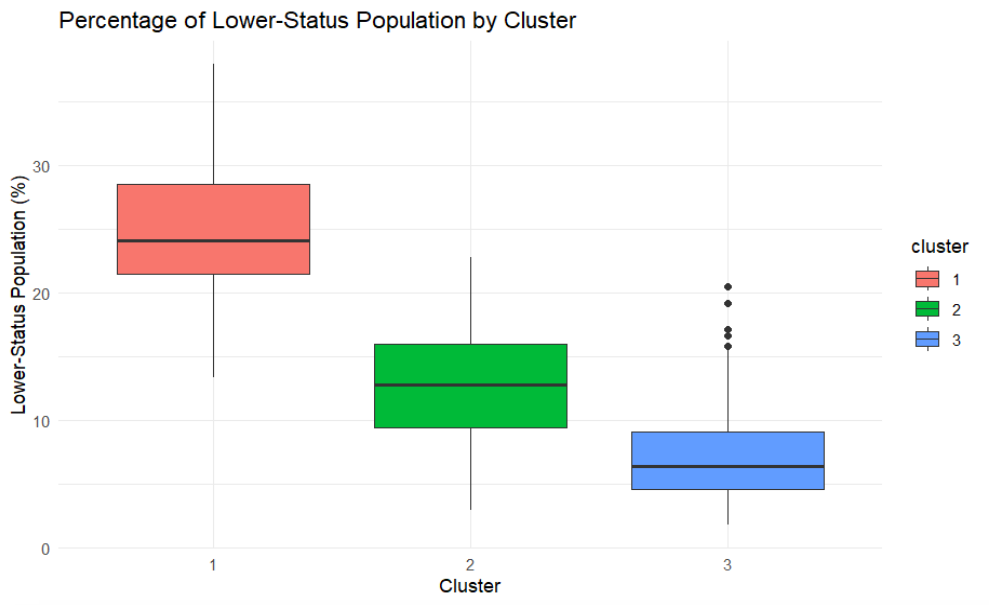
Here's **how Boston housing is segregated into groups**, and the characteristics of each group.

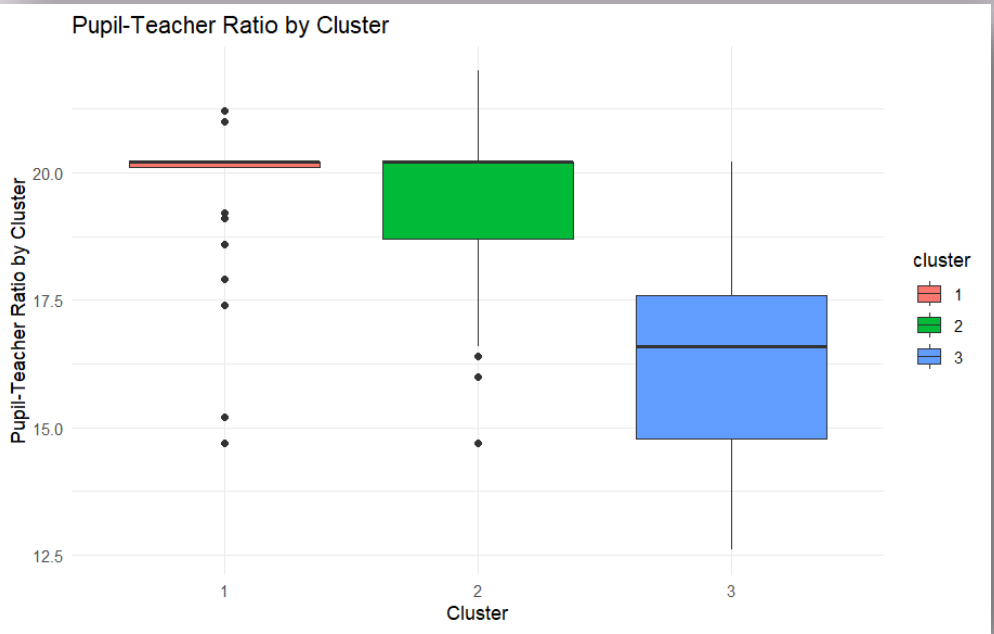
First, how many groups will exist?
THREE.

- K-Means Clustering with k=3 through Elbow Method

What will the groups be analyzed on?
TOP 3 PRIMARY HOUSING FACTORS.

- *AVERAGE ROOMS PER DWELLING (rm)*
- *LOWER STATUS POPULATION (lstat)*
- *PUPIL TEACHER RATIO (ptratio)*





Cluster Variable Analysis

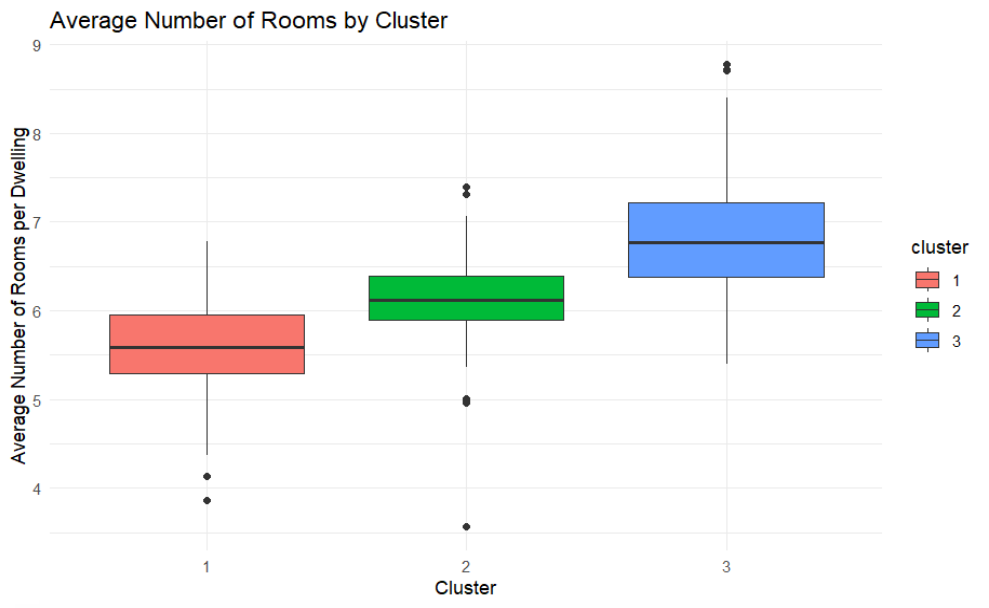
Distinction between each housing clusters based on these factors :

Percentage of lower status population

This factor influences the living environment and community.

Pupil-teacher ratio in schools

Small ratio is usually desirable as it relates to more focused learning per student and thus better education quality.



Cluster Variable Analysis

Distinction between each housing clusters based on these factors :

Average number of rooms per dwelling

More rooms per dwelling often suggests bigger-sized (more accommodating) houses.

Median price of houses

The median price of houses per cluster is also calculated.

The meanings?
1978 Boston Housing clusters are as follows.

CLUSTER 1 Cheap Houses	CLUSTER 2 Moderately Priced	CLUSTER 3 Expensive Houses
Mostly are houses around lower-class communities with less # of schools (or major class sizes) and smaller houses.	Houses located around middle-class communities with decent # of schools (or mid class sizes) and mid-sized houses.	Houses belonging majority to upper-class communities with more exclusive schools (or minor class sizes) and bigger houses.

Conclusions

- Boston's housing data has been analyzed using Linear Regression, Cart and K-means Clustering
- Future buyers and investors are advised to make more **informed choices** based on
- **Primary factors influencing house prices:**

  Average rooms per dwelling

  Percentage of lower-status population

  Pupil-teacher ratios in schools

  Non-retail business in acres of land

Current grouping of Boston housing: