adv5

November 10, 2024

[2]:

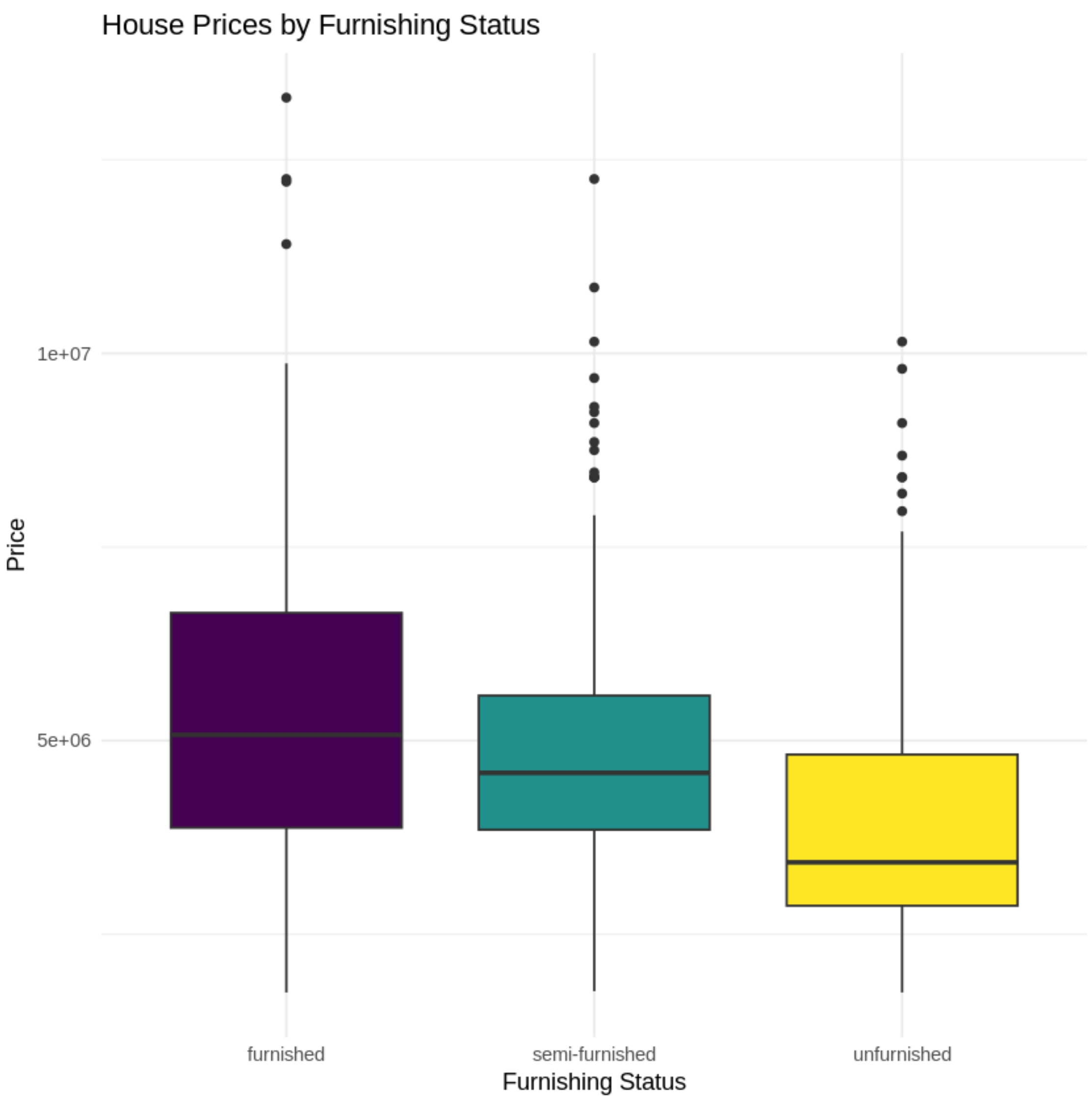
```
# First, install required packages
install.packages(c("ggplot2", "dplyr", "plotly", "viridis", "wordcloud", "tm",␣
 ↪"reshape2", "GGally"))
library(ggplot2)
library(dplyr)
library(plotly)
library(viridis)
library(wordcloud)
library(tm)
library(reshape2)
library(GGally)

# Read the dataset
housing_data <- read.csv("Housing.csv")

# 1. Word Cloud of Categorical Variables
# Combine categorical variables
categorical_text <- paste(
  housing_data$mainroad,
  housing_data$guestroom,
  housing_data$basement,
  housing_data$hotwaterheating,
  housing_data$airconditioning,
  housing_data$prefarea,
  housing_data$furnishingstatus
)

# Create word cloud
wordcloud(words = unlist(strsplit(categorical_text, " ")),
          min.freq = 1,
          scale = c(3, 0.5),
          colors = brewer.pal(8, "Dark2"))

# 2. Box and Whisker Plot
ggplot(housing_data, aes(y = price, x = furnishingstatus, fill =␣
 ↪furnishingstatus)) +
  geom_boxplot() +
```

```r
  theme_minimal() +
  labs(title = "House Prices by Furnishing Status",
       y = "Price",
       x = "Furnishing Status") +
  scale_fill_viridis(discrete = TRUE) +
  theme(legend.position = "none")

# 3. Violin Plot
ggplot(housing_data, aes(x = as.factor(bedrooms), y = price, fill = as.
 ↪factor(bedrooms))) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, fill = "white") +
  theme_minimal() +
  labs(title = "Distribution of Prices by Number of Bedrooms",
       x = "Number of Bedrooms",
       y = "Price") +
  scale_fill_viridis(discrete = TRUE, name = "Bedrooms")

# 4. Regression Plots
# Linear Regression
ggplot(housing_data, aes(x = area, y = price)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = TRUE) +
  theme_minimal() +
  labs(title = "Linear Regression: Price vs Area",
       x = "Area",
       y = "Price")

# Non-linear Regression (using loess)
ggplot(housing_data, aes(x = area, y = price)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = TRUE) +
  theme_minimal() +
  labs(title = "Non-linear Regression: Price vs Area",
       x = "Area",
       y = "Price")

# 5. 3D Chart using plotly
plot_ly(housing_data,
        x = ~area,
        y = ~bedrooms,
        z = ~price,
        type = "scatter3d",
        mode = "markers",
        marker = list(size = 5,
                     color = ~price,
                     colorscale = "Viridis",
                      showscale = TRUE))

# 6. Jitter Plot
ggplot(housing_data, aes(x = as.factor(stories), y = price, color = as.
  ↪factor(stories))) +
  geom_jitter(width = 0.2) +
  theme_minimal() +
  labs(title = "Price Distribution by Number of Stories",
       x = "Number of Stories",
       y = "Price") +
  scale_color_viridis(discrete = TRUE, name = "Stories")

# Save plots
ggsave("boxplot.png", width = 10, height = 6)
ggsave("violin_plot.png", width = 10, height = 6)
ggsave("regression_linear.png", width = 10, height = 6)
ggsave("regression_nonlinear.png", width = 10, height = 6)
ggsave("jitter_plot.png", width = 10, height = 6)
```

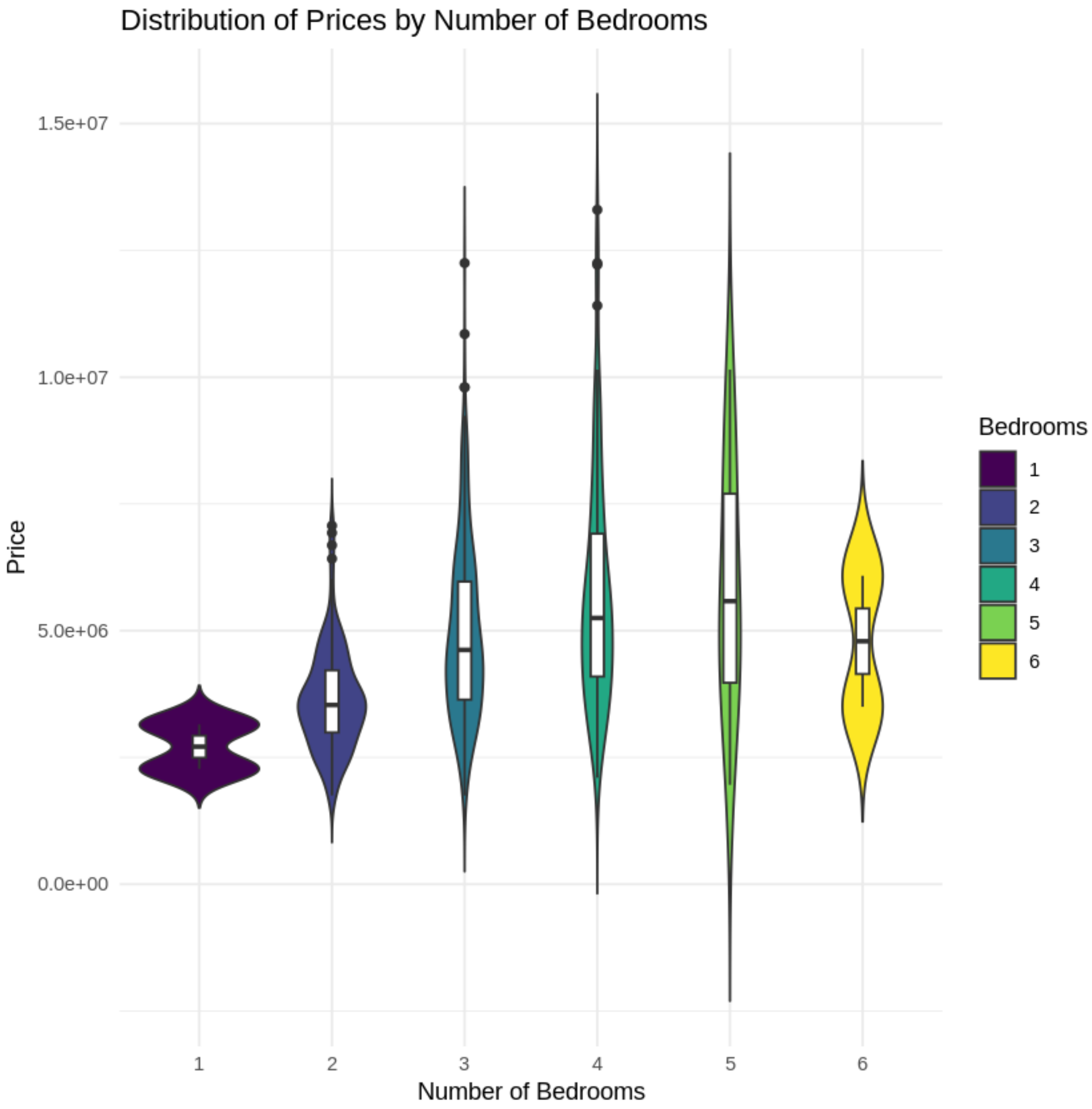
```
Installing packages into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘patchwork’, ‘lazyeval’, ‘crosstalk’,
‘gridExtra’, ‘NLP’, ‘slam’, ‘BH’, ‘plyr’, ‘ggstats’




Attaching package: ‘dplyr’


The following objects are masked from ‘package:stats’:

    filter, lag


The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union




Attaching package: ‘plotly’


The following object is masked from ‘package:ggplot2’:

    last_plot
```

```
The following object is masked from ‘package:stats’:

    filter


The following object is masked from ‘package:graphics’:

    layout


Loading required package: viridisLite

Loading required package: RColorBrewer

Loading required package: NLP


Attaching package: ‘NLP’


The following object is masked from ‘package:ggplot2’:

    annotate


Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2

Warning message in tm_map.SimpleCorpus(corpus, tm::removePunctuation):
"transformation drops documents"
Warning message in tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x,
tm::stopwords())):
"transformation drops documents"
```

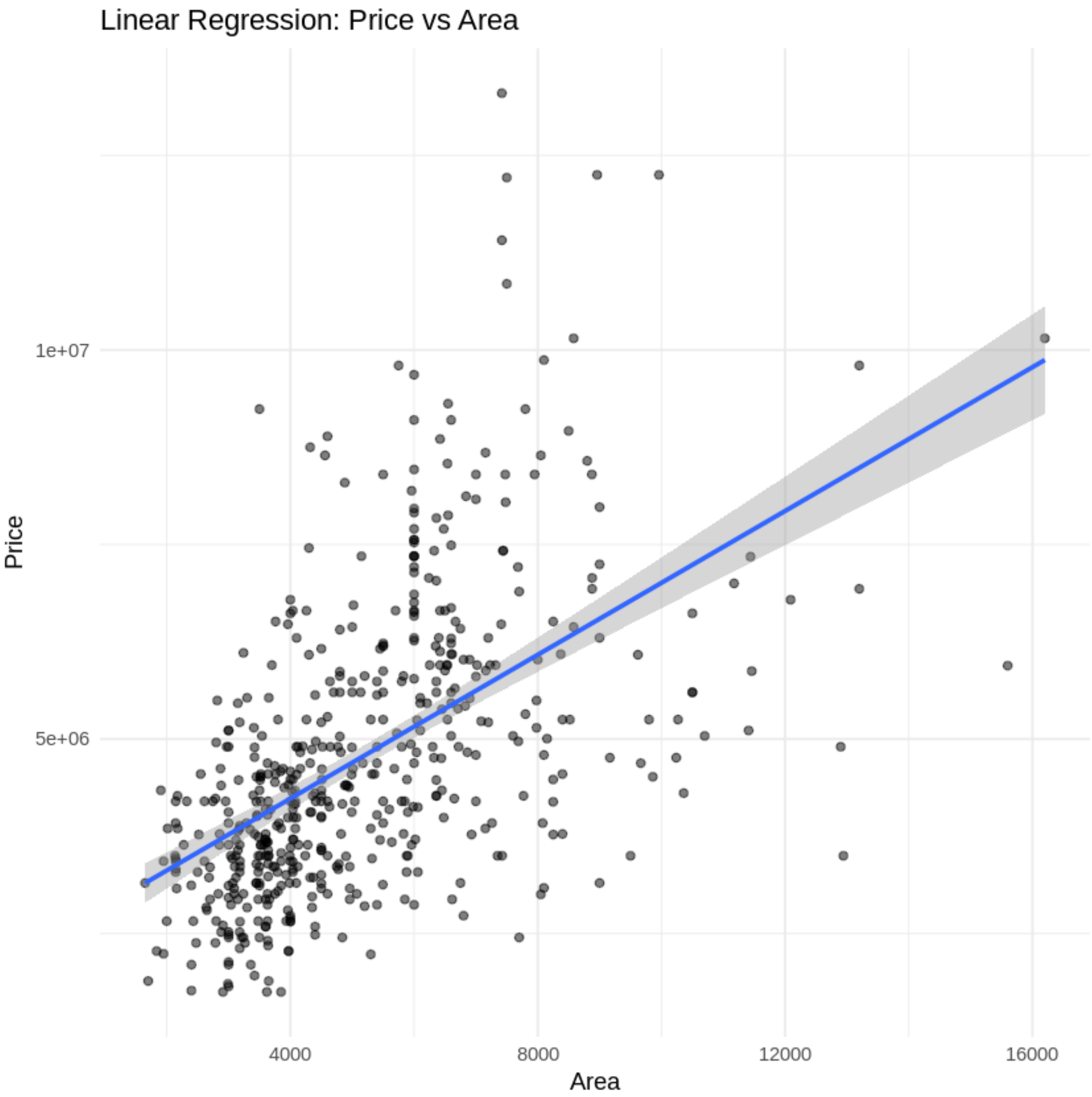
furnished
semifurnished
unfurnished
yes



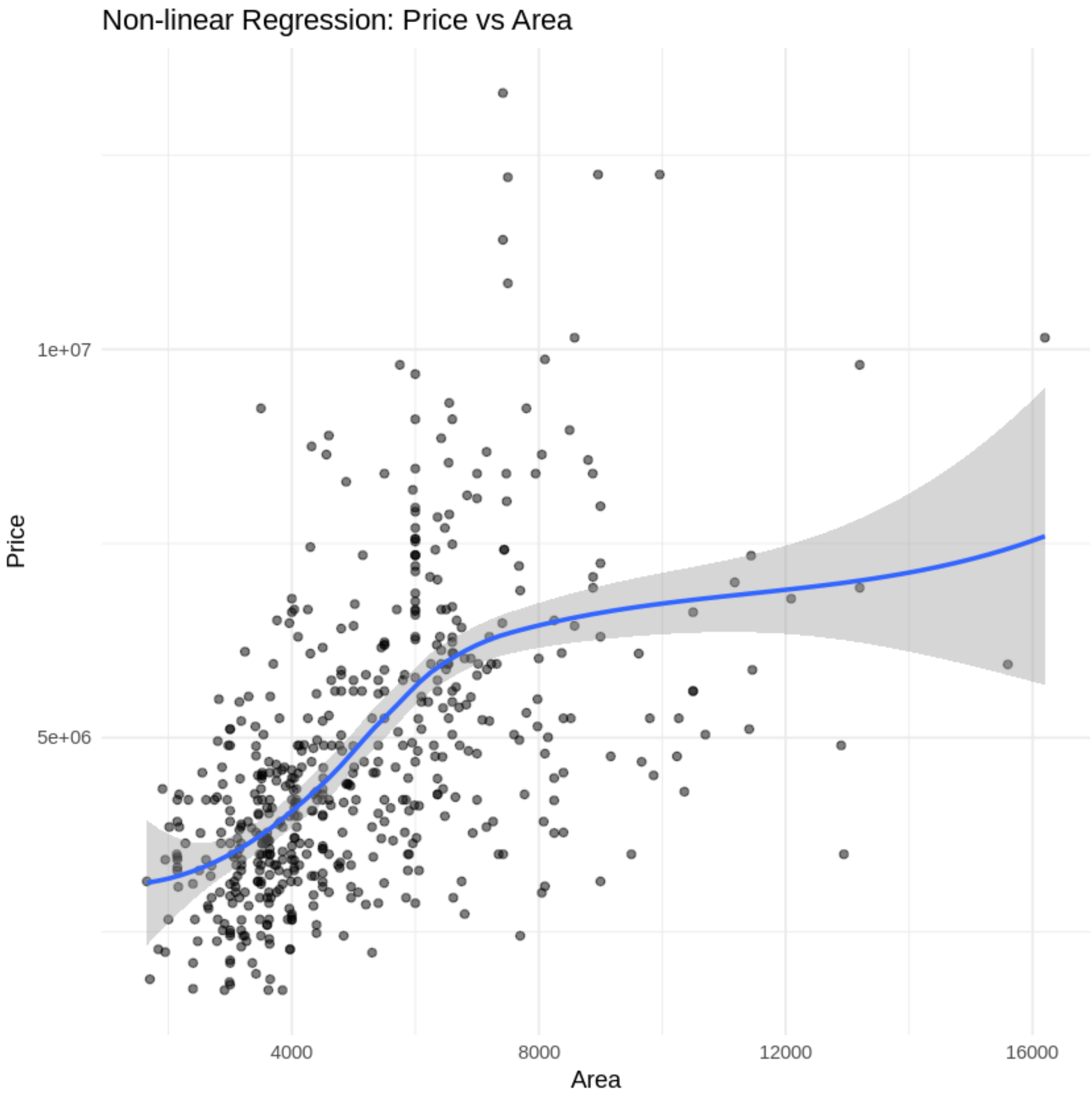
```
`geom_smooth()` using formula = 'y ~ x'
```

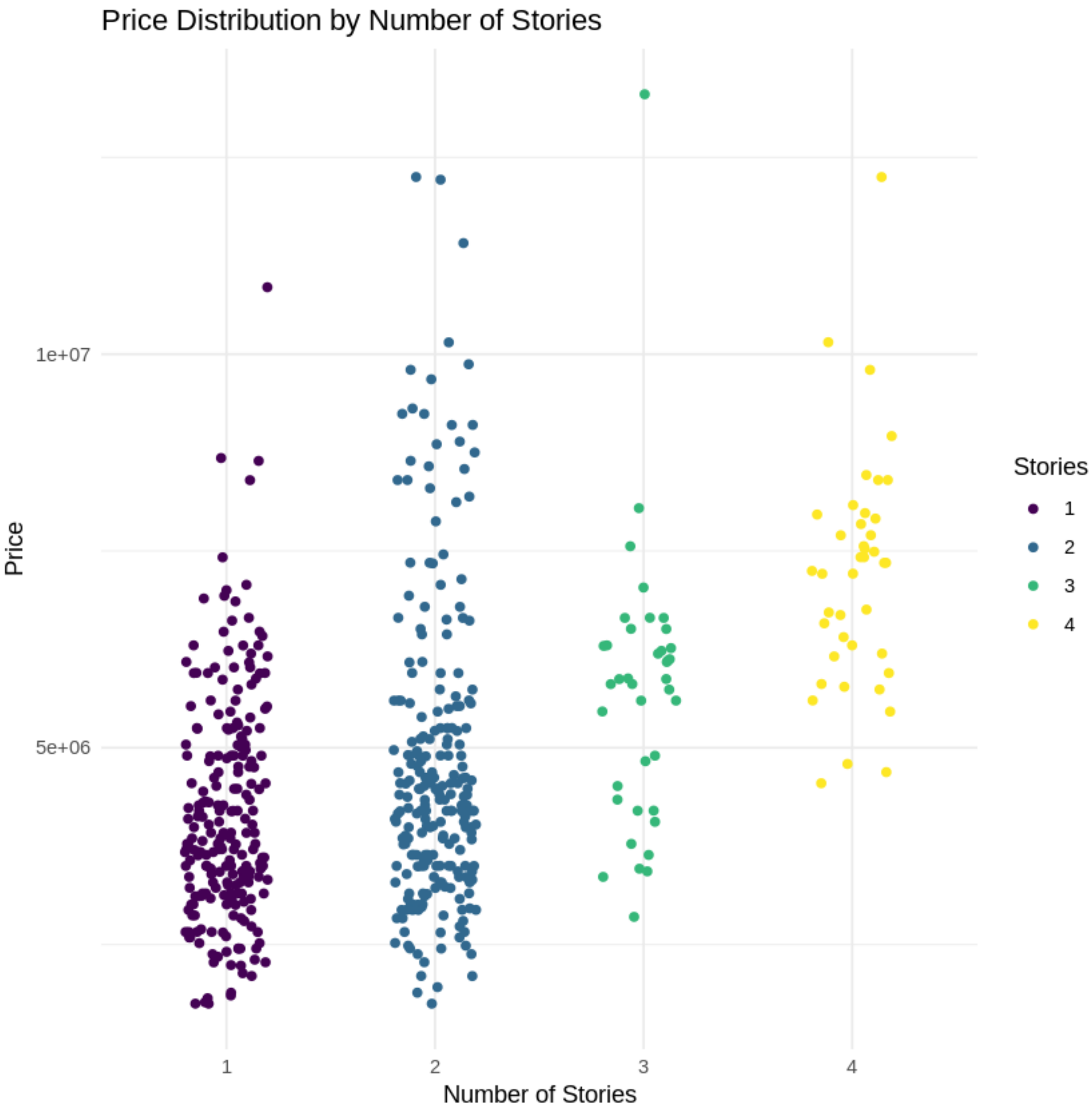


```
`geom_smooth()` using formula = 'y ~ x'
```



```
HTML widgets cannot be represented in plain text (need html)
```


Non-linear Regression: Price vs Area
Price
1e+07
5e+06
4000
8000
12000
16000
Area


Price Distribution by Number of Stories
Price
1e+07
5e+06
1
2
3
4
Number of Stories
Stories
● 1
● 2
● 3
● 4