ML Assignment 1 2025

Lona Uprety

2025-09-08

```
library(readxl)
AHD <- read_excel("C:/Users/lona2/OneDrive - Kent State University/American_Housing_Data_20231209.csv.x
View(AHD)
```

```
#Quantitative descriptive stats
summary(AHD$`Zip Code`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10013   40215   74136   64833   85730   98199
```

```
summary(AHD$Price)
```

```
##    Length     Class      Mode
##     39981 character character
```

```
summary(AHD$Beds)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   3.000   3.000   3.172   4.000  54.000
```

```
#Qualitative descriptive stats
summary(AHD$City)
```

```
##    Length     Class      Mode
##     39981 character character
```

```
summary(AHD$State)
```

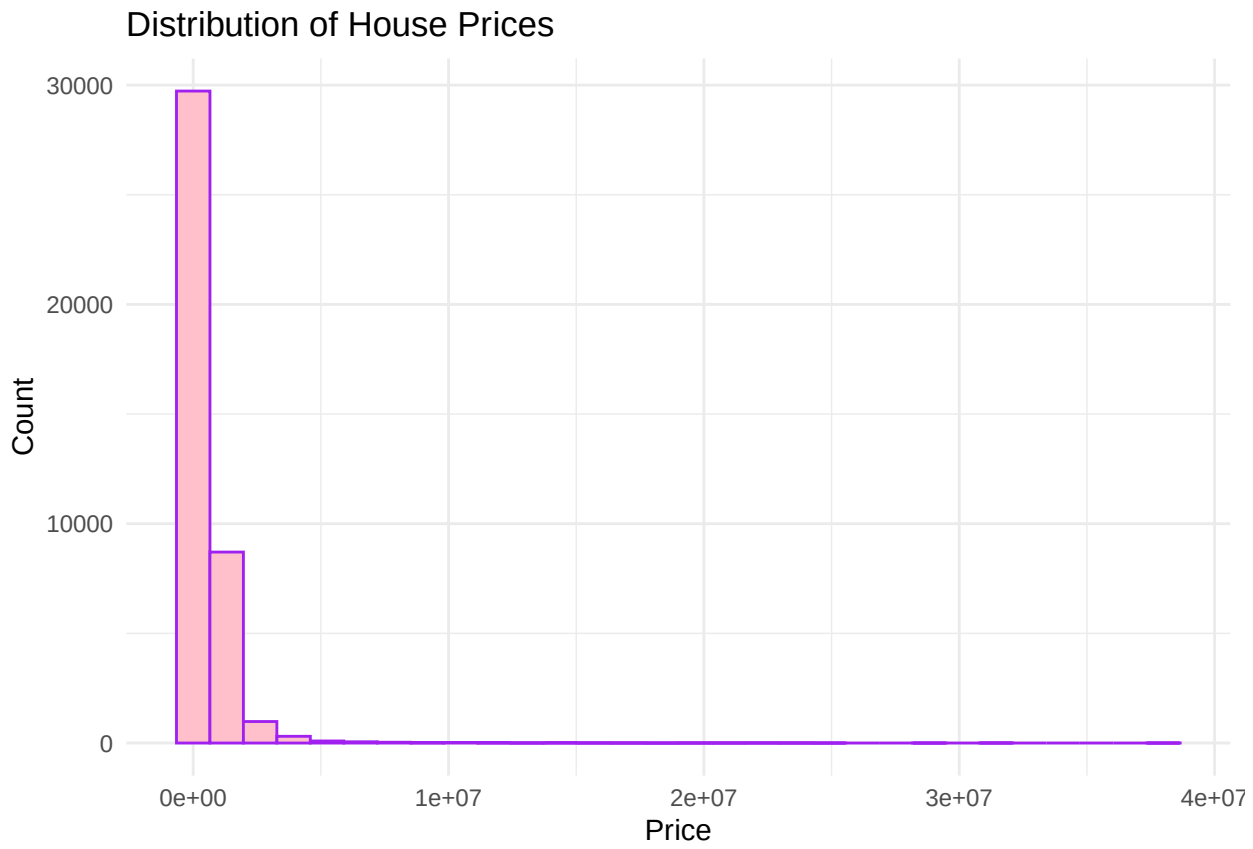
```
##    Length     Class      Mode
##     39981 character character
```

```
#Transforming a variable
AHD$Price <- as.numeric(AHD$Price)
```

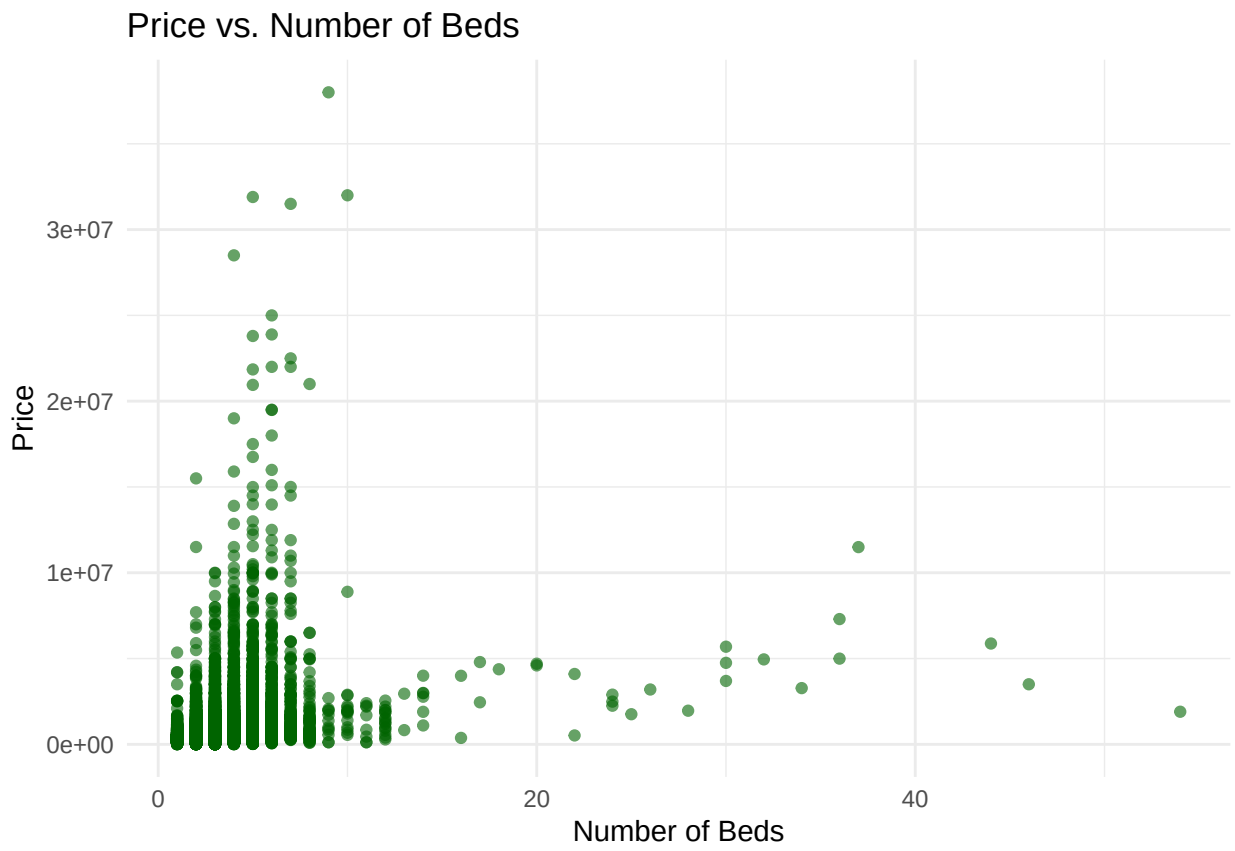
```
#Plot a quantitative variable
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```r
ggplot(AHD, aes(x = Price)) +
  geom_histogram(bins = 30, fill = "pink", color = "purple") +
  labs(title = "Distribution of House Prices", x = "Price", y = "Count") +
  theme_minimal()
```




```r
#Plot a scatterplot
library(ggplot2)
ggplot(AHD, aes(x = Beds, y = Price)) +
  geom_point(color = "darkgreen", alpha = 0.6) +
  labs(title = "Price vs. Number of Beds", x = "Number of Beds", y = "Price") +
  theme_minimal()
```


Price vs. Number of Beds
Price
3e+07
2e+07
1e+07
0e+00
0
20
40
Number of Beds