# Lab 2 - James - Correlations B/W Price and Other Predictors

2024-02-25

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
bmw <- read.csv("C:/Users/Temporality/Documents/College & University/Boston University/Spring 2024/MA 5
summary(bmw)
```

```
##   maker_key          model_key            mileage          engine_power
## Length:4843        Length:4843        Min.   :     -64   Min.   :  0
## Class :character   Class :character   1st Qu.: 102914   1st Qu.:100
## Mode  :character   Mode  :character   Median : 141080   Median :120
##                                       Mean   : 140963   Mean   :129
##                                       3rd Qu.: 175196   3rd Qu.:135
##                                       Max.   :1000376   Max.   :423
## registration_date     fuel          paint_color         car_type
## Length:4843        Length:4843        Length:4843        Length:4843
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## feature_1          feature_2          feature_3          feature_4
## Mode :logical      Mode :logical      Mode :logical      Mode :logical
## FALSE:2181         FALSE:1004         FALSE:3865         FALSE:3881
## TRUE :2662         TRUE :3839         TRUE :978          TRUE :962
##
##
##
## feature_5          feature_6          feature_7          feature_8
## Mode :logical      Mode :logical      Mode :logical      Mode :logical
## FALSE:2613         FALSE:3674         FALSE:329          FALSE:2223
## TRUE :2230         TRUE :1169         TRUE :4514         TRUE :2620
##
##
##
##     price            sold_at            obs_type
## Min.   :   100   Length:4843        Length:4843
## 1st Qu.: 10800   Class :character   Class :character
## Median : 14200   Mode  :character   Mode  :character
## Mean   : 15828
## 3rd Qu.: 18600
## Max.   :178500
```

```r
sold_at_split <- strsplit(bmw$sold_at, "/")

registration_split <- strsplit(bmw$registration_date, "/")

# assign month only; all sold in 2018
bmw$month_sold <- sapply(sold_at_split, function(x) as.integer(x[1]))

bmw$year_sold <- sapply(sold_at_split, function(x) as.integer(x[3]))

bmw$month_registered <- sapply(registration_split, function(x) as.integer(x[1]))

bmw$year_registered <- sapply(registration_split, function(x) as.integer(x[3]))

bmw <- subset(bmw, select = -c(maker_key, registration_date, obs_type, year_sold))

response <- "price"
predictors <- c("mileage", "fuel", "paint_color", "car_type", "feature_1", "feature_2", "feature_3", "f
bmw_subset <- subset(bmw, select = c(response, predictors))

for (predictor in predictors) {
  # Subset the dataframe for the current predictor variable
  plot_data <- subset(bmw_subset, select = c(response, predictor))

  # Create a scatterplot/correlation graph
  g <- ggpairs(plot_data,
        title = paste("Scatterplot and correlation for price and", predictor),
        upper = list(continuous = wrap("points", alpha=0.3, size=0.1)),
        lower = list(continuous = "cor", method = "spearman")) + theme(axis.text = element_text(size = 

  # Print the graph
  print(g)
}
```
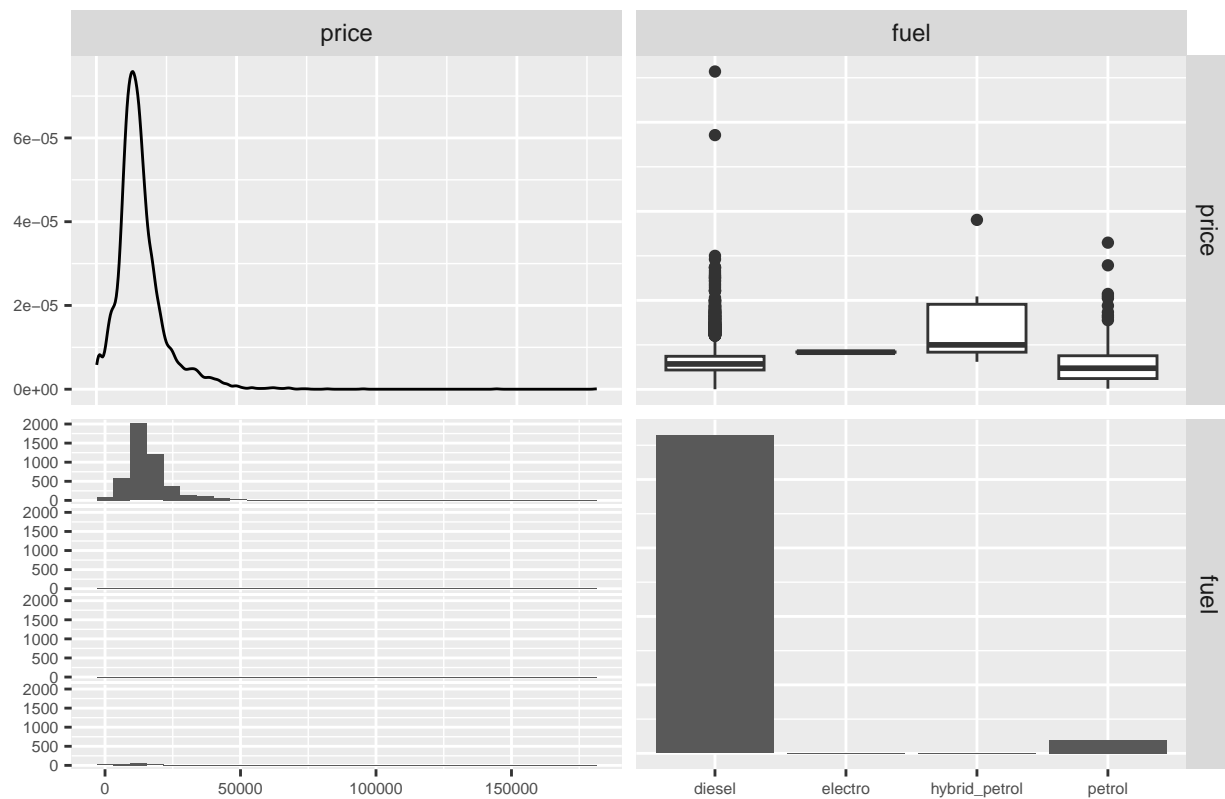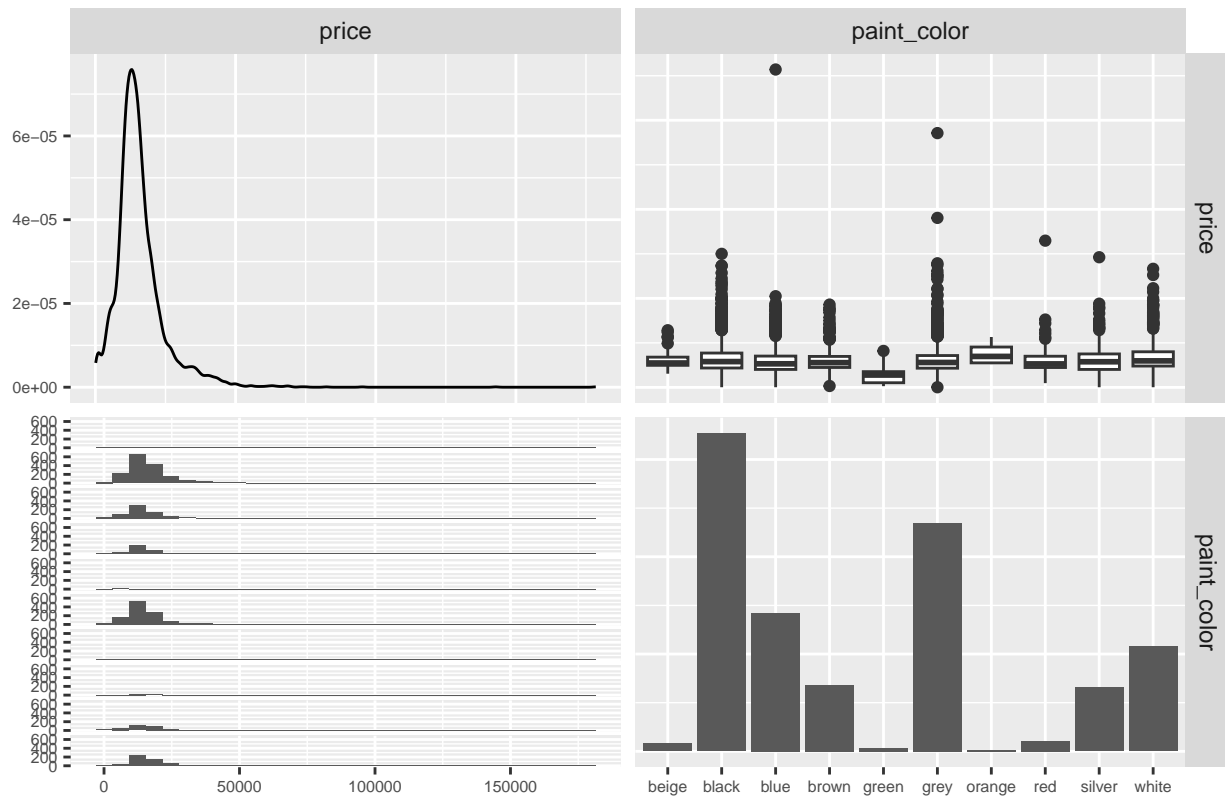
## Scatterplot and correlation for price and mileage
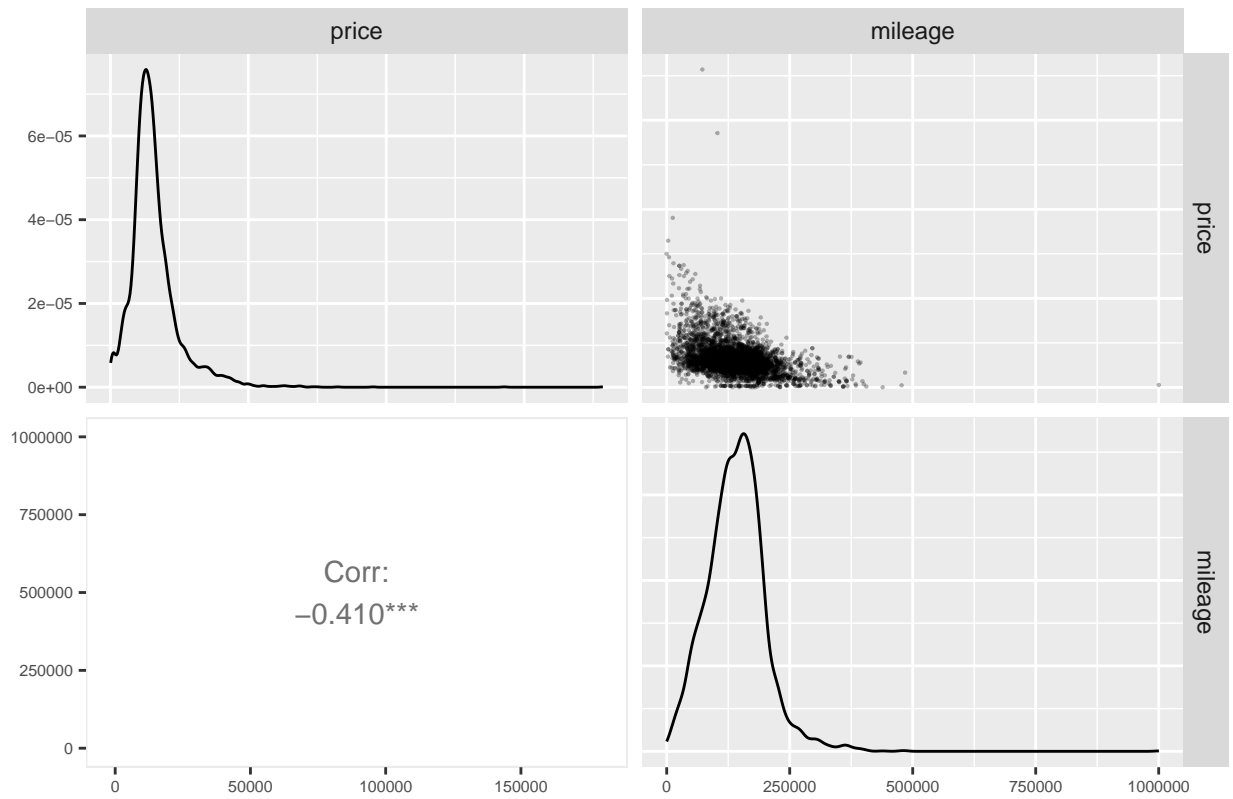


```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

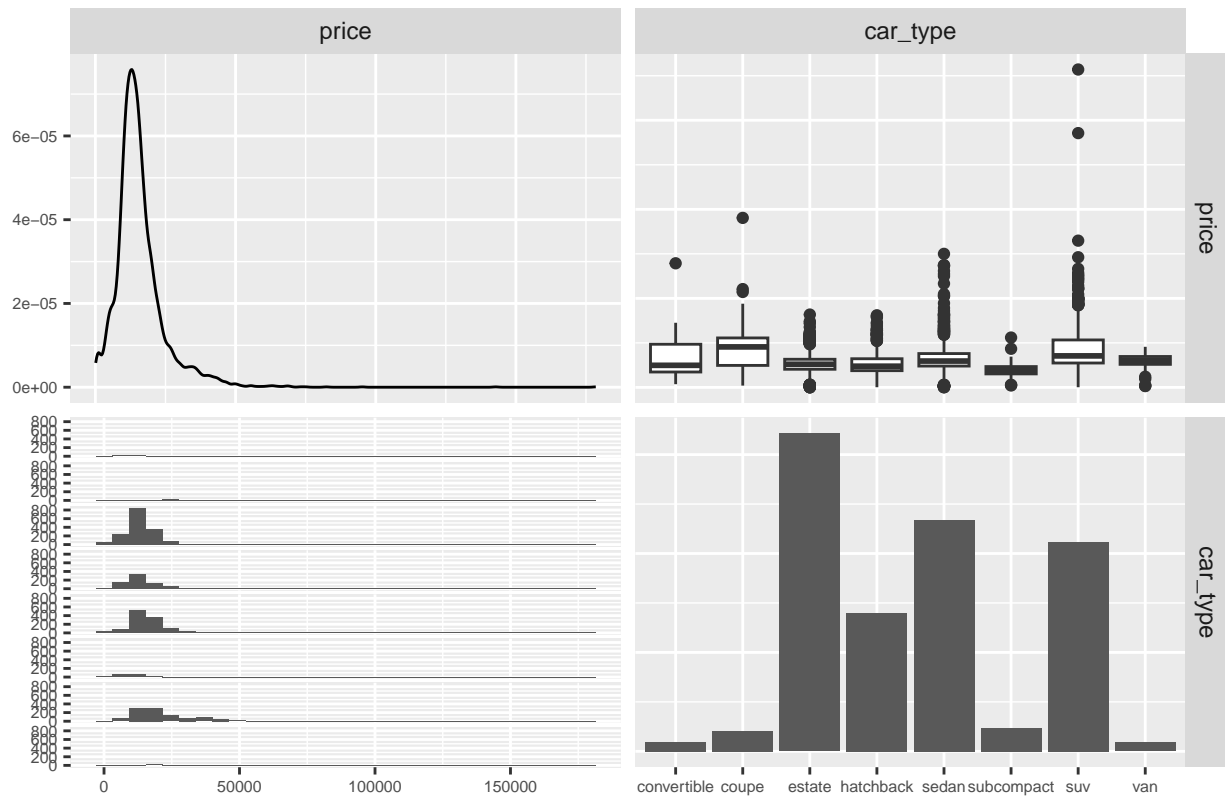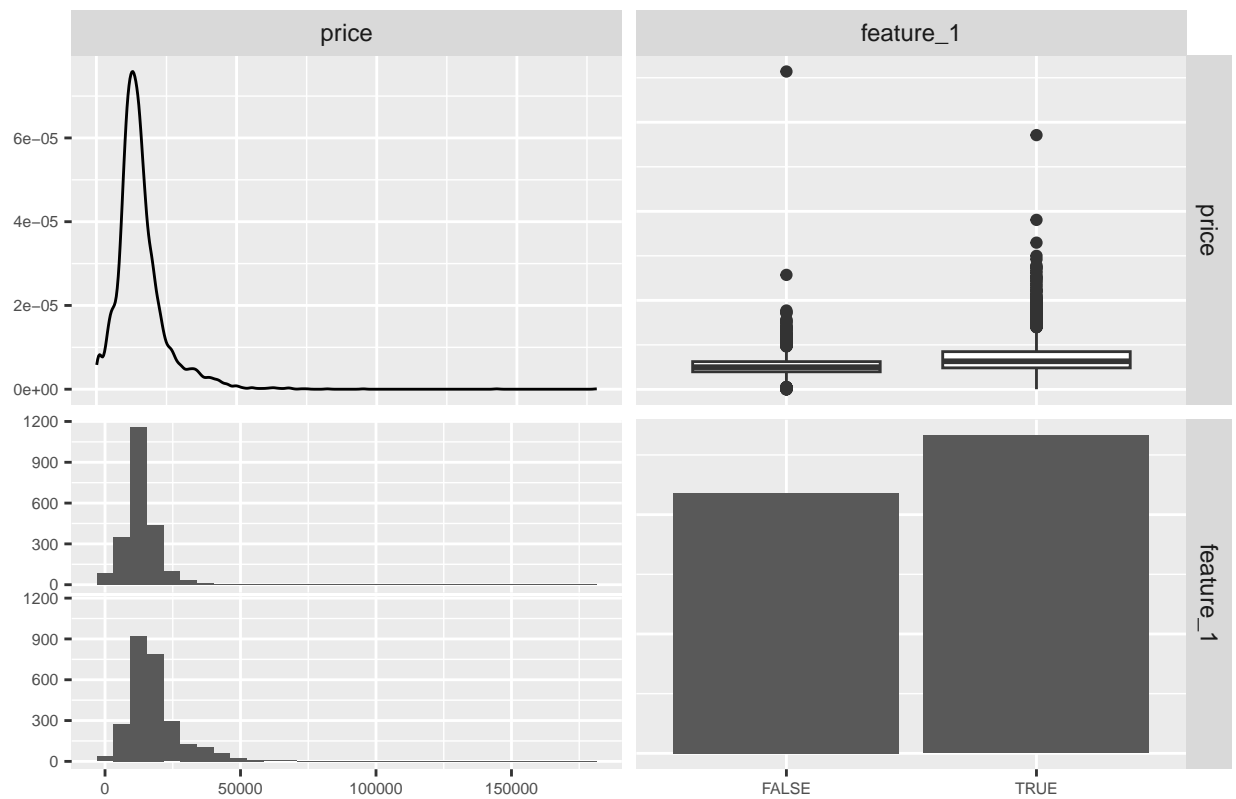# Scatterplot and correlation for price and fuel



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

# Scatterplot and correlation for price and paint_color



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
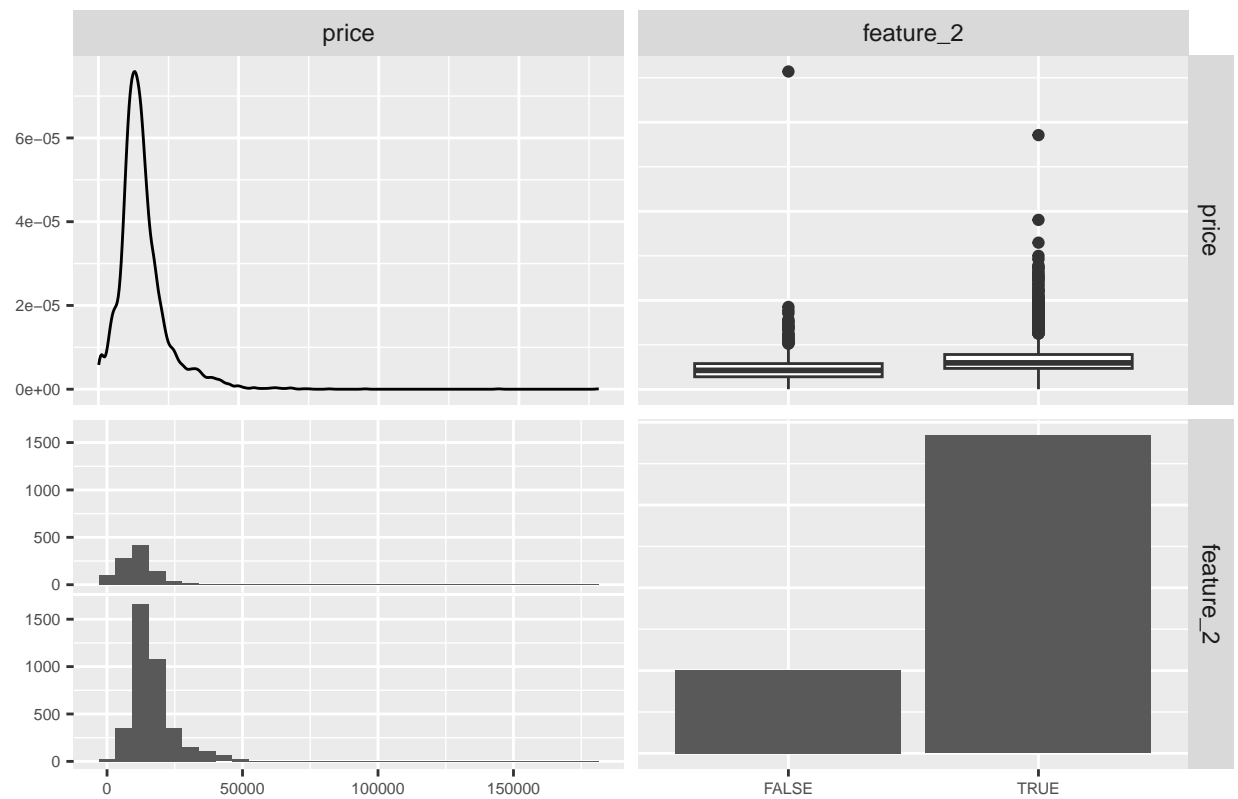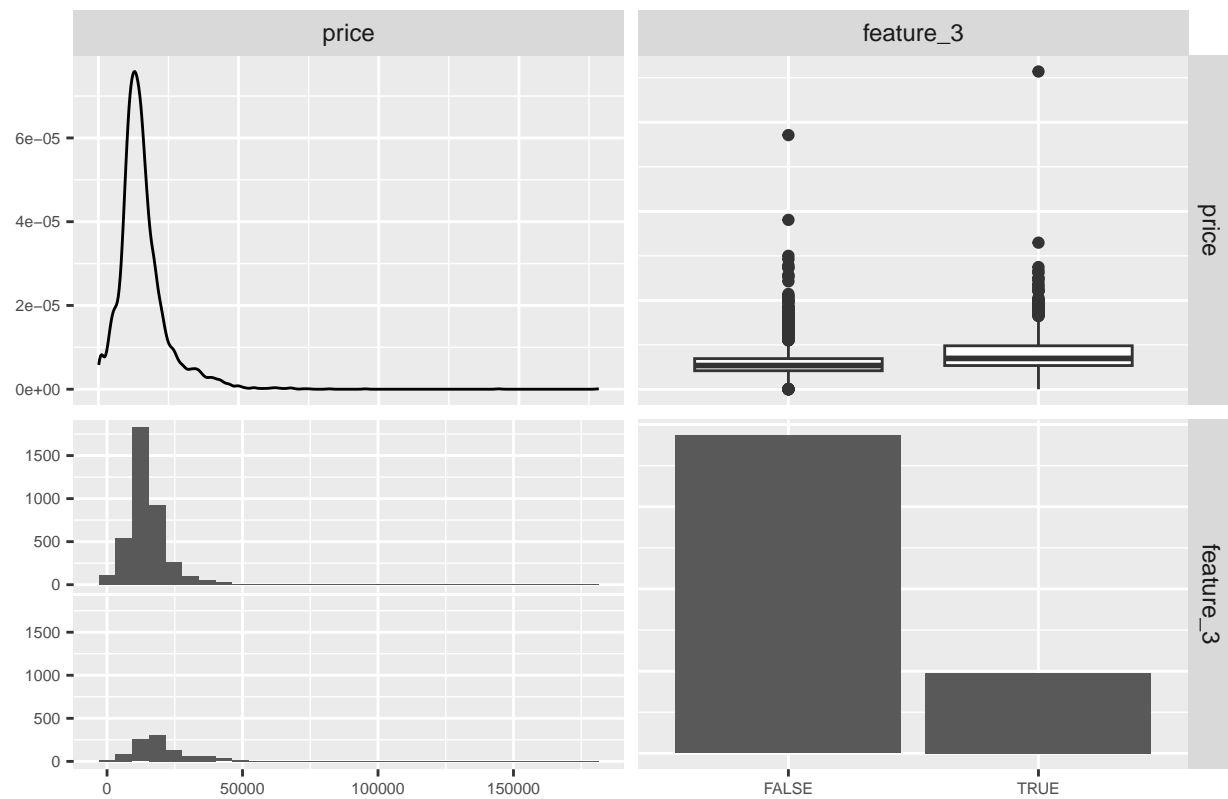
## Scatterplot and correlation for price and car_type
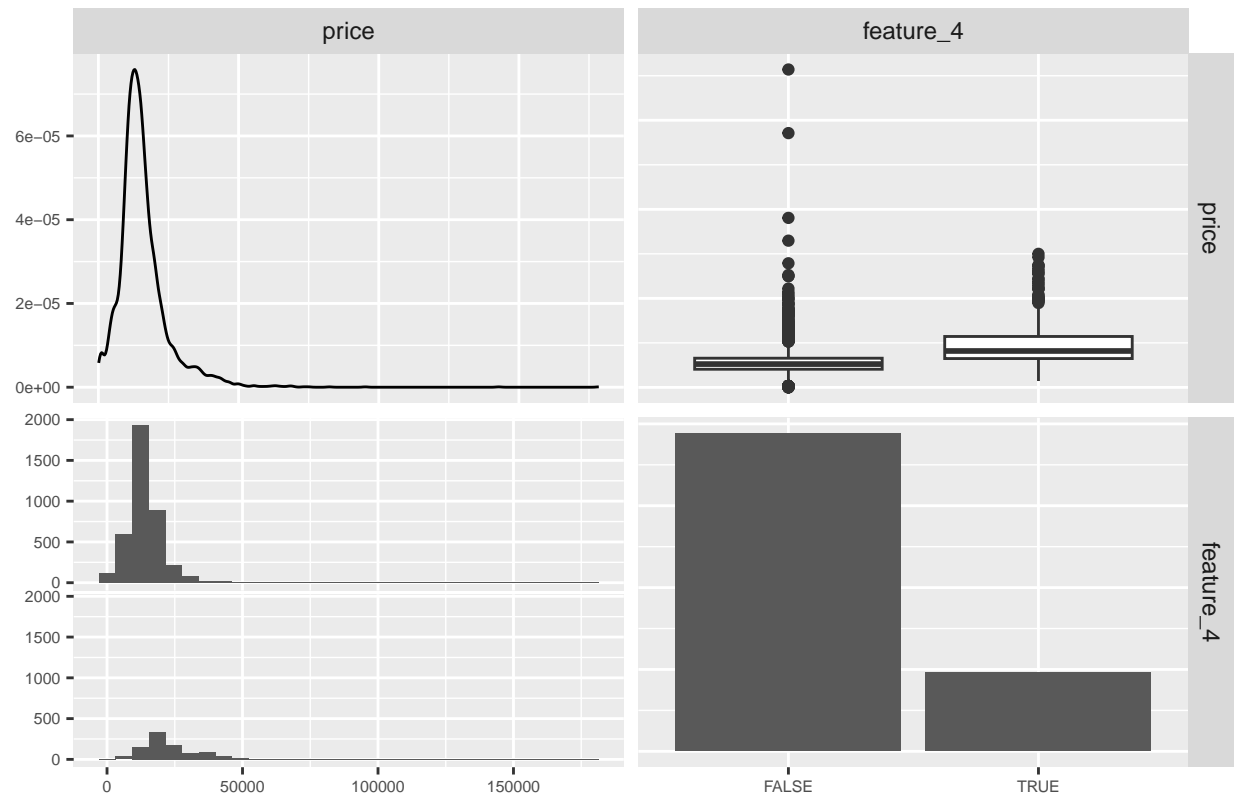


```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

## Scatterplot and correlation for price and feature_1



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

## Scatterplot and correlation for price and feature_2



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

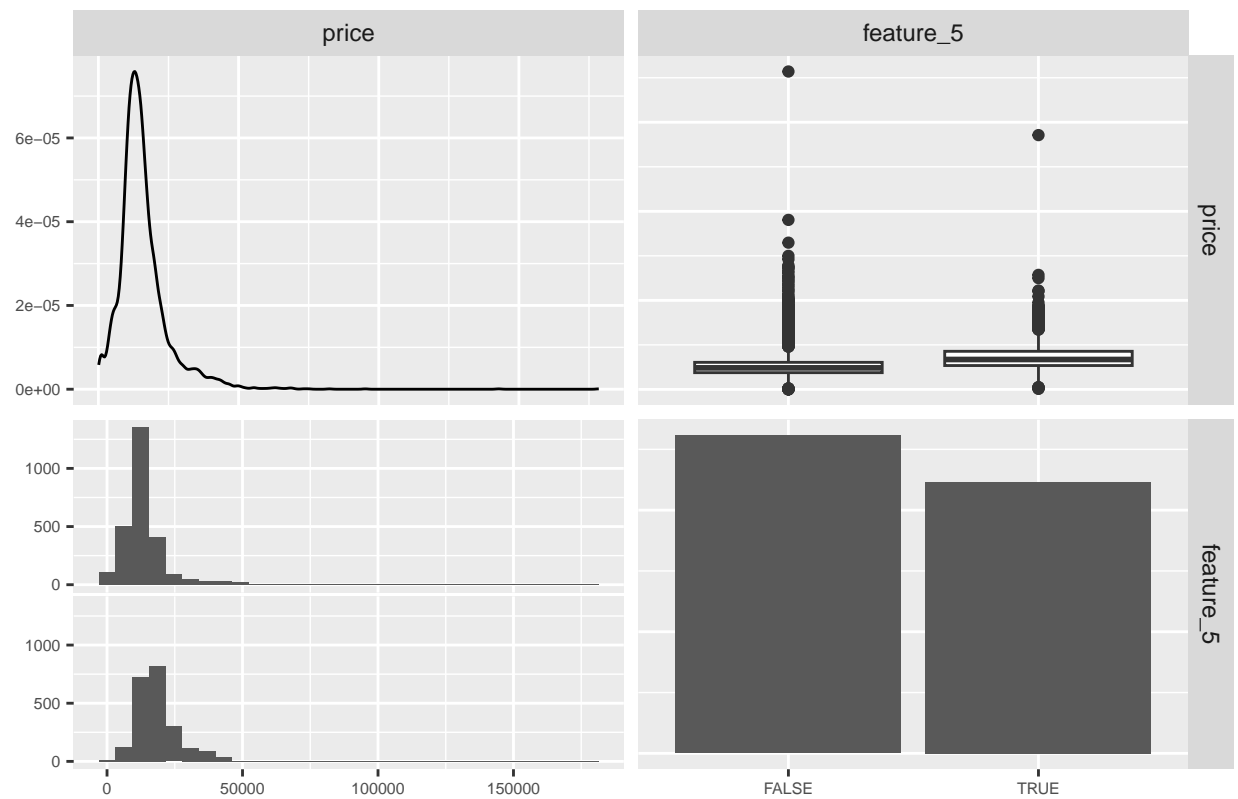## Scatterplot and correlation for price and feature_3



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
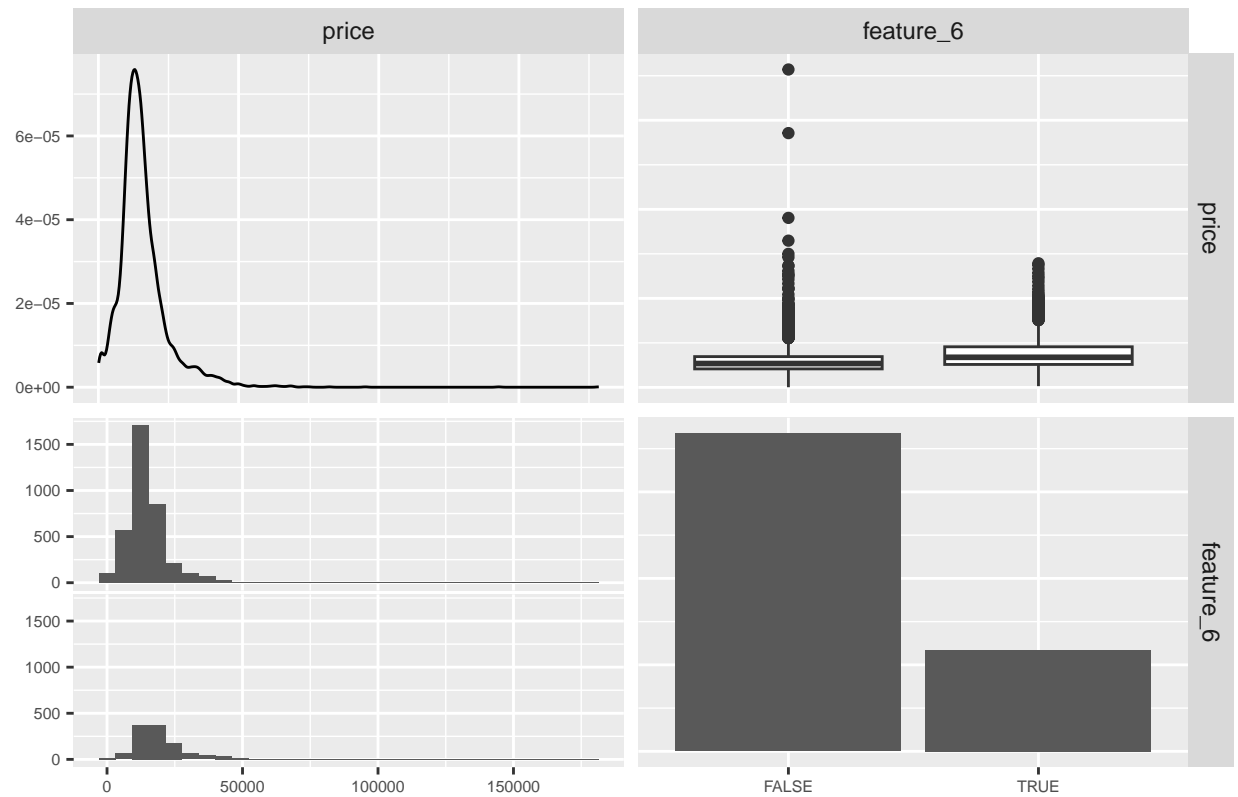
## Scatterplot and correlation for price and feature_4



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

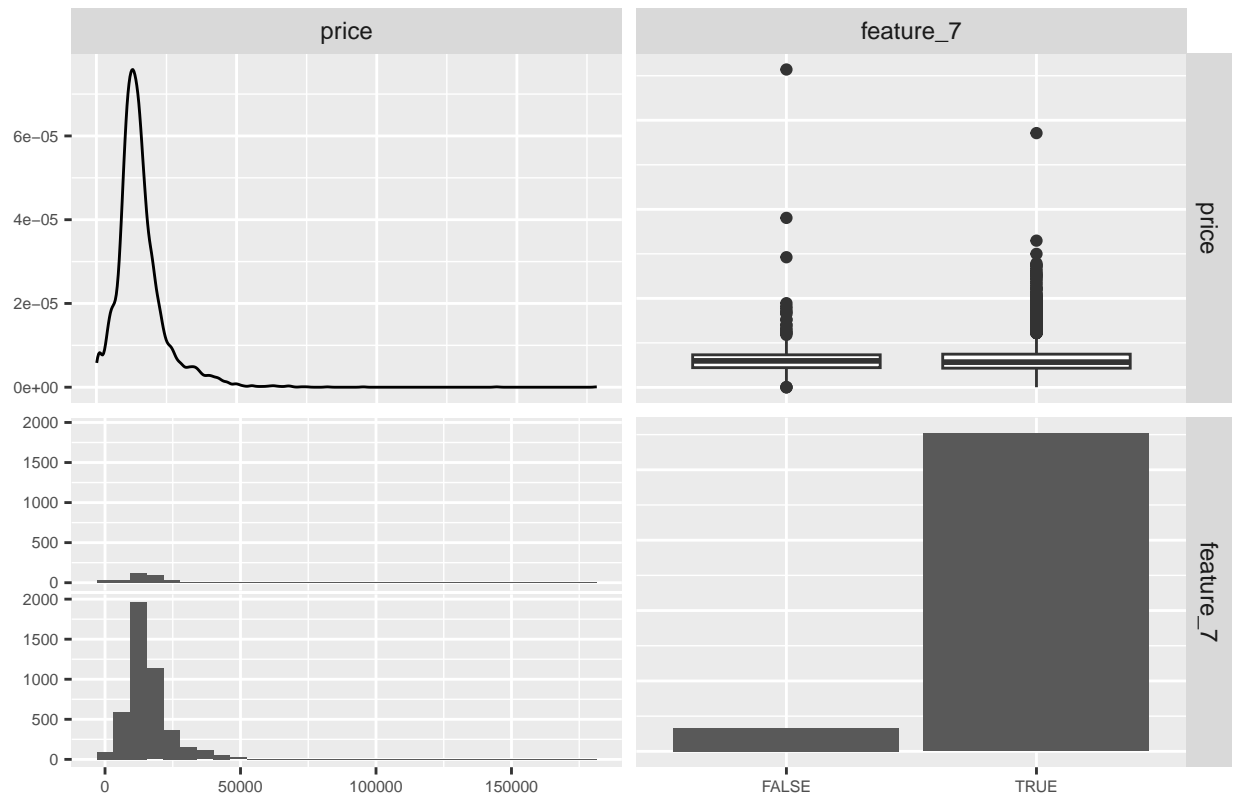## Scatterplot and correlation for price and feature_5



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

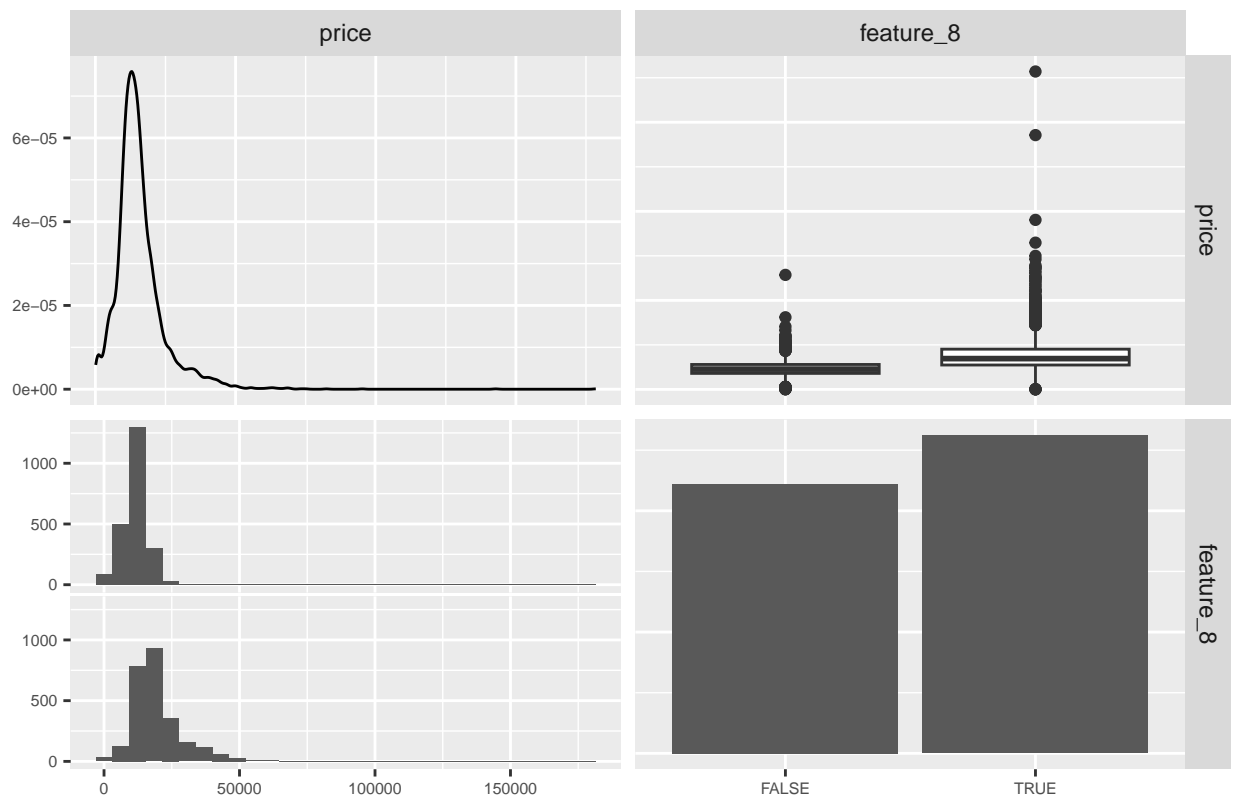## Scatterplot and correlation for price and feature_6



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

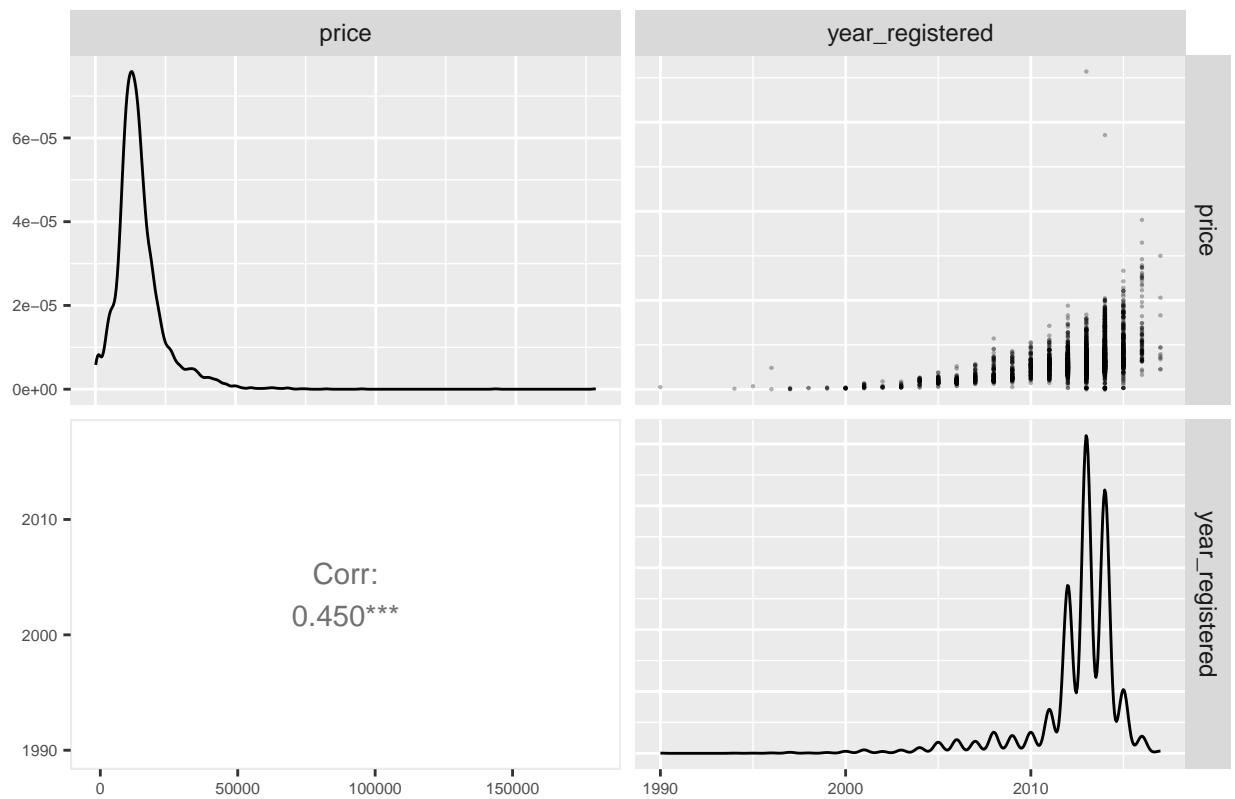## Scatterplot and correlation for price and feature_7



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

# Scatterplot and correlation for price and feature_8

# Scatterplot and correlation for price and year_registered

# Scatterplot and correlation for price and month_registered