

Boston housing

Tristan-Mihai Radulescu, Hugo Ponthieu, Benoit Planche

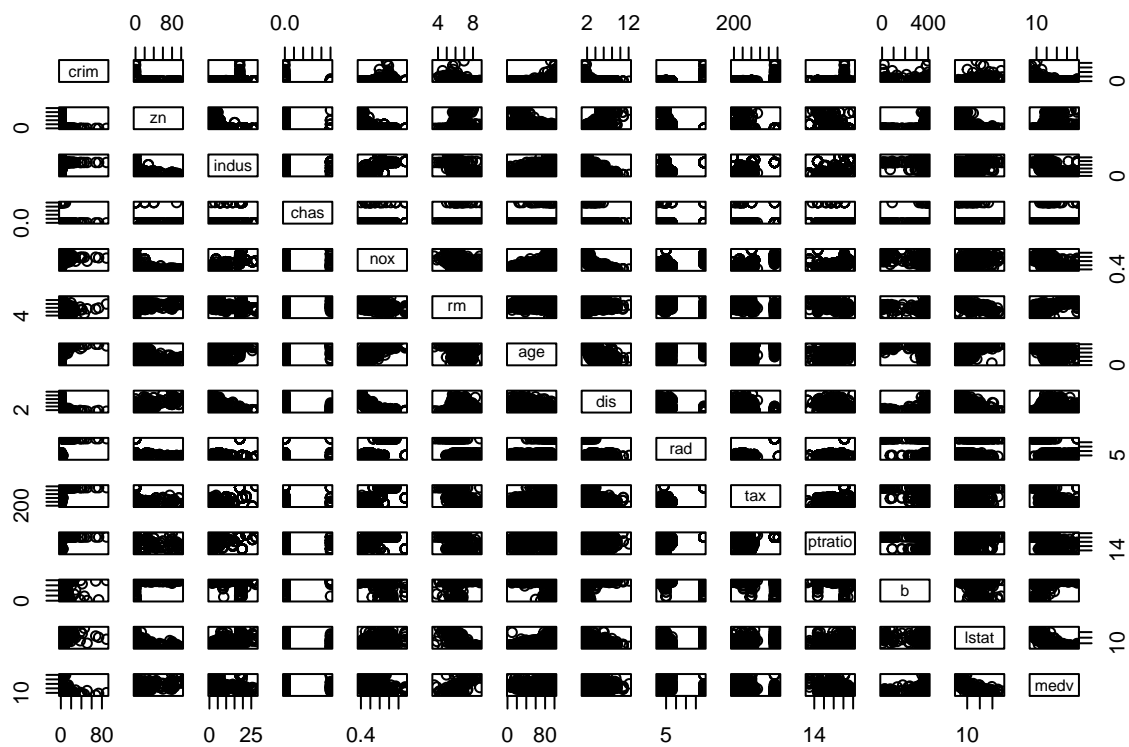
```
BostonHousing <- read.csv("./BostonHousing.csv")
head(BostonHousing)
```

```
##      crim zn indus chas   nox    rm age    dis rad tax ptratio    b lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

```
str(BostonHousing)
```

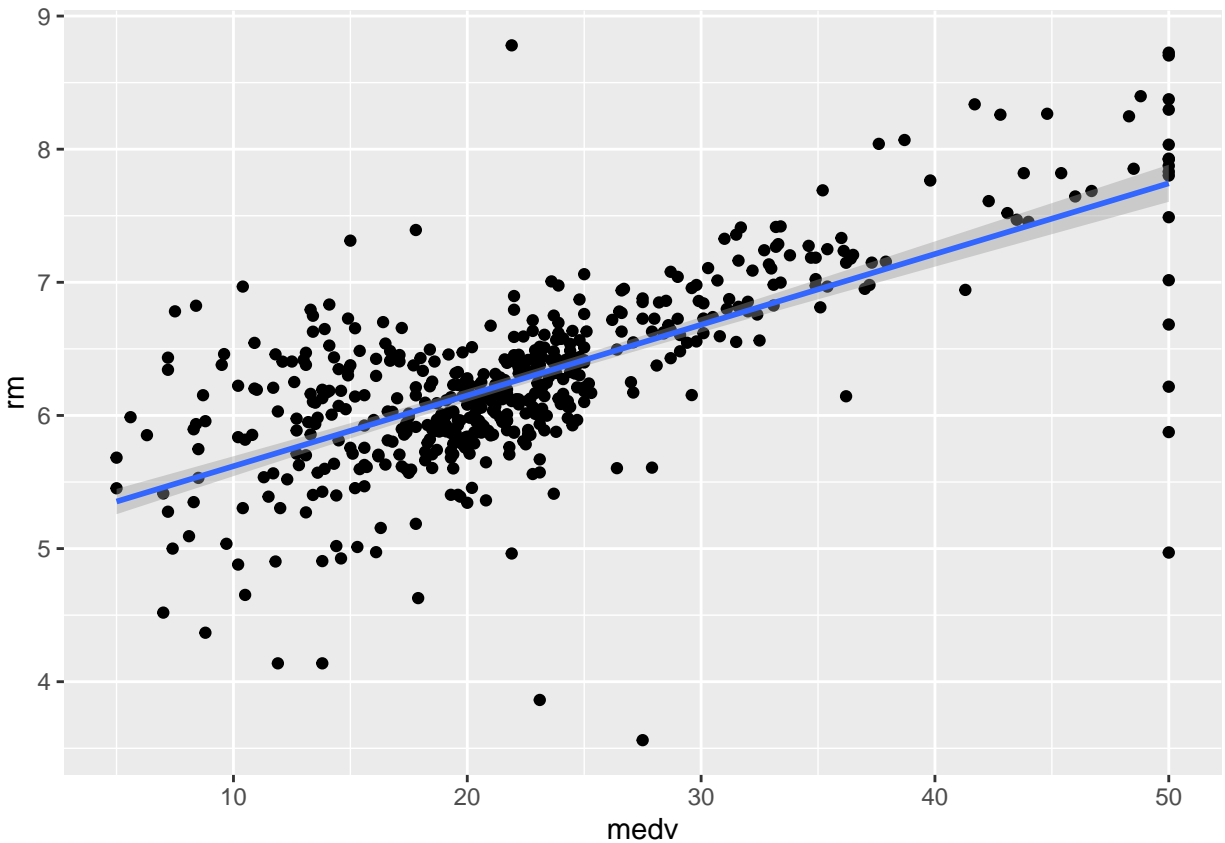
```
## 'data.frame':    506 obs. of  14 variables:
## $ crim   : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
## $ zn     : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
## $ indus  : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
## $ chas   : int   0 0 0 0 0 0 0 0 0 0 ...
## $ nox    : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...
## $ rm     : num  6.58 6.42 7.18 7 7.15 ...
## $ age    : num  65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...
## $ dis    : num  4.09 4.97 4.97 6.06 6.06 ...
## $ rad    : int   1 2 2 3 3 3 5 5 5 5 ...
## $ tax    : int  296 242 242 222 222 222 311 311 311 311 ...
## $ ptratio: num  15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...
## $ b      : num  397 397 393 395 397 ...
## $ lstat  : num  4.98 9.14 4.03 2.94 5.33 ...
## $ medv   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
```

```
pairs(data=BostonHousing,
~ crim + zn + indus + chas + nox + rm + age + dis + rad + tax + ptratio + b + lstat + medv)
```



```
lin_reg <- lm(medv ~ rm, data=BostonHousing)
ggplot(data=BostonHousing, aes(x=medv, y=rm)) + geom_point() + geom_smooth(method="lm")
```

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



Can you make a box plot of the value of the rivers in the BostonHousing dataset?

```
# Load necessary libraries
library(ggplot2)

# Create a boxplot for 'medv' colored by 'chas'
ggplot(BostonHousing, aes(x = factor(chas), y = medv, fill = factor(chas))) +
  geom_boxplot() +
  labs(title = "Boxplot of Median Home Values by River Boundary",
       x = "Bounds River",
       y = "Median Home Value (in $1000s)",
       fill = "Bounds River") +
  scale_fill_manual(values = c("No" = "red", "Yes" = "blue")) +
  theme_minimal()
```

```
## Warning: No shared levels found between 'names(values)' of the manual scale and the
## data's fill values.
## No shared levels found between 'names(values)' of the manual scale and the
## data's fill values.
```

