R_Recap_02

Felix Mueller

2/27/2017

Prepare your R session

```
setwd("~/Google_Drive/github/R/R_Recap/session_02")
library(dplyr)
library(ggplot2)
```

Iris dataset

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class: – Iris Setosa – Iris Versicolour – Iris Virginica

Source: `https://archive.ics.uci.edu/ml/datasets/Iris`

Loading the Iris dataset

```
glimpse(iris)
```

```
## Observations: 150
## Variables: 5
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6,
## $ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4,
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4,
## $ Petal.Width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3,
## $ Species      <fctr> setosa, setosa, setosa, setosa, se
```

Boston Housing data

1. CRIM per capita crime rate by town
2. ZN proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS proportion of non-retail business acres per town
4. CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
5. NOX nitric oxides concentration (parts per 10 million)
6. RM average number of rooms per dwelling
7. AGE proportion of owner-occupied units built prior to 1940
8. DIS weighted distances to five Boston employment centres
9. RAD index of accessibility to radial highways
10. TAX full-value property-tax rate per $10,000
11. PTRATIO pupil-teacher ratio by town
12. B 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13. LSTAT % lower status of the population
14. MEDV Median value of owner-occupied homes in $1000's

Source: `http://archive.ics.uci.edu/ml/datasets/Housing`

Loading the Boston dataset

```
library(MASS)
glimpse(Boston)
```

```
## Observations: 506
## Variables: 14
## $ crim    <dbl> 0.00632, 0.02731, 0.02729, 0.03237, 0.06
## $ zn      <dbl> 18.0, 0.0, 0.0, 0.0, 0.0, 0.0, 12.5, 12.
## $ indus   <dbl> 2.31, 7.07, 7.07, 2.18, 2.18, 2.18, 7.87
## $ chas    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
## $ nox     <dbl> 0.538, 0.469, 0.469, 0.458, 0.458, 0.458
## $ rm      <dbl> 6.575, 6.421, 7.185, 6.998, 7.147, 6.430
## $ age     <dbl> 65.2, 78.9, 61.1, 45.8, 54.2, 58.7, 66.6
## $ dis     <dbl> 4.0900, 4.9671, 4.9671, 6.0622, 6.0622,
## $ rad     <int> 1, 2, 2, 3, 3, 3, 5, 5, 5, 5, 5, 5, 5, 4
## $ tax     <dbl> 296, 242, 242, 222, 222, 222, 311, 311,
## $ ptratio <dbl> 15.3, 17.8, 17.8, 18.7, 18.7, 18.7, 15.2
## $ black   <dbl> 396.90, 396.90, 392.83, 394.63, 396.90,
## $ lstat   <dbl> 4.98, 9.14, 4.03, 2.94, 5.33, 5.21, 12.4
```