

# Exercise 1 – Explore the Hitters data



```
library(tree)
library(ISLR)
library(ggplot2)
```

```
##### Regression trees #####
```

```
|
# load and inspect the Hitters data set
data("Hitters")
?Hitters # The dataset records and salaries for baseball players depending on their number of hits etc.
head(Hitters)

# remove incomplete observations
sum(is.na(Hitters))
Hitters <- na.omit(Hitters)

# Salary is our target variable
# We log-transform it so that its distribution has more of a typical bell-shape.
# We need to remember later on that Salary is logarithmized!
hist(Hitters$Salary) # Salary is measured in thousands of dollars
Hitters$Salary <- log(Hitters$Salary)
hist(Hitters$Salary)

attach(Hitters)
```

See R-code on Moodle!