

# Insurance

Georgios Papadopoulos

Vasileios Papadopoulos

## Load Dataset

Dataset consists of 1338 records and 7 features. The column *charges* is the independent variable, the other 6 will be used to analyze the impact they have.

```
df <- read.csv(file = 'insurance.csv')
nrow(df)
```

```
## [1] 1338
```

```
ncol(df)
```

```
## [1] 7
```

```
head(df)
```

```
##   age    sex    bmi children smoker   region  charges
## 1  19 female 27.900         0    yes southwest 16884.924
## 2  18   male 33.770         1     no southeast  1725.552
## 3  28   male 33.000         3     no southeast  4449.462
## 4  33   male 22.705         0     no northwest 21984.471
## 5  32   male 28.880         0     no northwest  3866.855
## 6  31 female 25.740         0     no southeast  3756.622
```