Converting categorical variables into numerical

Using the `caret` package in R is a straightforward task that converts every categorical variable into a **flag one**, also known as a *dummy* variable.

If the original categorical variable has thirty possible values, then it will result in 30 new columns holding the value `0` or `1`, where `1` represents the presence of that category in the row.

If we use the caret package from R, then this conversion only takes two lines of code:

```
library(caret) # contains dummyVars function
library(dplyr) # data munging library
library(funModeling) # df_status function
```

```
# Checking categorical variables
status=df_status(heart_disease, print_results = F)
filter(status,  type %in% c("factor", "character")) %>%
select(variable)
```

```
# It converts all categorical variables (factor and
character) into numerical variables
# It skips the original variable, so no need to remove it
after the conversion, the data is ready to use.

dmy = dummyVars(" ~ .", data = heart_disease)
heart_disease_2 = data.frame(predict(dmy, newdata =
heart_disease))
```

```r
# Checking the new numerical data set:
colnames(heart_disease_2)
```

Original data `heart_disease` has been converted into `heart_disease_2` with no categorical variables, only numerical and dummy. Note that every new variable has a *dot* followed by the *value*.

If we check the before and after for the 7th patient (row) in variable `chest_pain` which can take the values `1`, `2`, `3` or `4`, then

```r
# before
as.numeric(heart_disease[7, "chest_pain"])
## [1] 4

# after
heart_disease_2[7, c("chest_pain.1", "chest_pain.2",
"chest_pain.3", "chest_pain.4")]
##   chest_pain.1 chest_pain.2 chest_pain.3 chest_pain.4
## 7            0            0            0            1
```

Having kept and transformed only numeric variables while excluding the nominal ones, the data `heart_disease_2` are ready to be used.