Assume you have a data frame (df) for patients taking a specific drug. The data consists of a factor variable (Drug) and a numeric variable (N_patients).

Drugs	N_patients
Drug 1	50
Drug 2	40
Drug 3	23
Drug 4	92
Drug 5	70

Later on you filter the data frame for specific levels in the factor variable and saved it in a new data frame called df1.

```
df1 <- df %>% filter(Drugs %in% c("Drug 1", "Drug 2")) %>% print
## Drugs N_patients
## 1 Drug 1 50
## 2 Drug 2 40
```

Although in df1 we have only two observations, the factor variable keeps all of its original levels, even if they do not actually exist as observations.

If we look at the structure of df1:

```
str(df1)
## 'data.frame': 2 obs. of 2 variables:
## $ Drugs : Factor w/ 5 levels "Drug 1","Drug 2",..: 1 2
## $ N_patients: num 50 40
```

Notice that df1 does consists of 2 observations and 2 variables, however looking closely at "Drugs" variable, we notice that it consists of 5 levels.

To see those levels, we can use the levels function. Note that we can use it only on a factor variable.

```
levels(df1$Drugs)
## [1] "Drug 1" "Drug 2" "Drug 3" "Drug 4" "Drug 5"
```

As you can see we do have 5 levels ("Drug 1", "Drug 2", "Drug 3", "Drug 4", "Drug 5") even though only 2 levels ("Drug 1", "Drug 2") are present in df1. For this reason we should drop the levels that are not found in the data frame otherwise it might cause some problems later on when using functions that require factor levels.

There are 2 ways to exclude these levels:

1. Use droplevels function on the variable we want to remove the levels that are not present.

In this case we want to remove the levels ("Drug 3", "Drug 4", "Drug 5") from "Drugs" variable.

```
# If you are only fimiliar with Base R
# df1$Drugs <- droplevels(df1$Drugs)
# If you are fimiliar with dplyr package
```

```
df1 <- df1 %>% mutate(Drugs=droplevels(Drugs))
```

lets check again the levels of Drugs variable:

```
levels(df1$Drugs)
## [1] "Drug 1" "Drug 2"
```

As you can see this is a direct way where we can implement the droplevel function.

2. Indirect way would be as follows:

We can change the vector to a character one then back again to a factor vector:

```
df1 <- df1 %>% mutate(Drugs=as.character(df1$Drugs))
```

Now we have the "Drugs" variable as a character vector. To check the levels, we have to transform it again to a factor one.

```
df1 <- df1 %>% mutate(Drugs=as.factor(df1$Drugs))
levels(df1$Drugs)
## [1] "Drug 1" "Drug 2"
```

It doesn't matter which way we choose as long as we have removed the levels that are not present in the data frame.