Background

I was looking at some breast cancer data recently, and was analyzing
the ER (estrogen receptor) status variable. It turned out that there were three possible outcomes in the data: Positive, Negative and Indeterminate. I had imported this data as a factor, and wanted to convert the Indeterminate level to a missing value, i.e. NA.

My usual method for numeric variables created a rather singular result:

```
x <- as.factor(c('Positive','Negative','Indeterminate'))
x1 <- ifelse(x=='Indeterminate', NA, x)
str(x1)
##  int [1:3] 3 2 NA
```

This process converted it to an integer!! Not the end of the world, but not ideal by any means.

Further investigation revealed two other `tidyverse` strategies.

`dplyr::na_if`

This method changes the values to NA, but keeps the original level in the factor's levels

```
x2 <- dplyr::na_if(x, 'Indeterminate')
str(x2)
##  Factor w/ 3 levels "Indeterminate",..: 3 2 NA
x2
## [1] Positive Negative
## Levels: Indeterminate Negative Positive
```

`dplyr::recode`

This method drops the level that I'm deeming to be missing from the factor

```
x3 <- dplyr::recode(x, Indeterminate = NA_character_)
str(x3)
##  Factor w/ 2 levels "Negative","Positive": 2 1 NA
x3
## [1] Positive Negative
## Levels: Negative Positive
```

This method can also work more generally to change all values not listed to missing values.

```
x4 <- dplyr::recode(x, Positive='Positive', Negative='Negative',
                    .default=NA_character_)
x4
## [1] Positive Negative
## Levels: Negative Positive
```

Other strategies are welcome in the comments.