

Cleaning-Bad-Data-in-R

2025-03-19

1.2 Missing Fields

After loading the heating data I need to transform and format some values. When I apply mutate to transform homes heating `%>% mutate(homes = as.numeric(homes))` from Class: character, R show me a **Caused by warning: ! NAs introducidos por coerción**. Then I ask for the data using a filter and found characters values (“,”, “Z”). Finally mutate the values to zero after check with other sources.

```
# Load the tidyverse
library(tidyverse)

# Load the data file
heating <- read_csv("./exercise_files/1_2/heating.csv")

# Tidy the data
heating <- heating %>% gather(key="age", value="homes", -Source)

knitr::kable(
  summary(heating),
  digits=1, align=rep('c', 5))
```

Source	age	homes
Length:112	Length:112	Length:112
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

```
knitr::kable(
  head(heating %>% mutate(homes = as.numeric(homes)), 10),
  digits=1, align=rep('c', 5))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `homes = as.numeric(homes)`.
## Caused by warning:
## ! NAs introducidos por coerción
```

Source	age	homes
Warm-air furnace	Under 25 years old	2546
Steam or hot water system	Under 25 years old	326
Electric heat pump	Under 25 years old	529
Built-in electric units	Under 25 years old	280
Floor, wall, or other built-in hot-air units without ducts	Under 25 years old	267
Room heaters with flue	Under 25 years old	15
Room heaters without flue	Under 25 years old	18
Portable electric heaters	Under 25 years old	122

Source	age	homes
Stoves	Under 25 years old	23
Fireplaces with inserts	Under 25 years old	11

```
knitr::kable(
  heating %>% filter(is.na(as.numeric(homes))),
  digits=1, align=rep('c', 5))
```

```
## Warning: There was 1 warning in `filter()`.
## i In argument: `is.na(as.numeric(homes))`.
## Caused by warning:
## ! NAs introducidos por coerción
```

Source	age	homes
Cooking stove	Under 25 years old	.
Cooking stove	25 to 29 years old	Z
Cooking stove	30 to 34 years old	.

```
heating <-
  heating %>%
  mutate(homes=ifelse(homes=='.', 0, homes)) %>%
  mutate(homes=ifelse(homes=='Z', 0, homes)) %>%
  mutate(homes = as.numeric(homes))

summary(heating)
```

```
##      Source      age      homes
## Length:112      Length:112      Min.   :  0.00
## Class :character Class :character 1st Qu.:  37.25
## Mode  :character Mode  :character Median  : 153.50
##                                     Mean   : 1056.12
##                                     3rd Qu.:  674.00
##                                     Max.   :15348.00
```