

Tidy

Isaac Verminck

2018-04-30

```
listings <- read_raw_listings()
```

Tidying consists of the following steps: * remove irrelevant data (e.g. rent always mentions it's in euro) * map some values to english (e.g. state of building is coded in Dutch but English is more consistent) * give all variables their correct data type * make clearly incorrect values missing (e.g. rent of 1 euro) * remove cases offering no information (all variables contain missing values)

```
listings <- listings %>%  
  extract_digit_variables %>%  
  extract_postcode %>%  
  state_to_english %>%  
  correct_data_types %>%  
  make_incorrect_values_missing %>%  
  remove_empty_cases
```

The basic data is now in the shape we want, but we're still missing some variables like the name of the municipality (easy to add based on the postal code) and the total rental cost.

```
listings <- listings %>%  
  add_municipality %>%  
  add_total_cost %>%  
  add_price_per_square_m
```

```
save_tidy_listings(listings)
```