

Exercise

Using R, clean this data set to make it easier to visualize and analyze. Specifically, these are the tasks you need to do:

0: Load the data in RStudio

Save the data set as a CSV file called *refine_original.csv* and load it in RStudio into a data frame.

1: Clean up brand names

Clean up the 'company' column, so all of the misspellings of the brand names are standardized. For example, you can transform the values in the column to be: *philips*, *akzo*, *van houten* and *unilever* (all lowercase).

2: Separate product code and number

Separate the product code and product number into separate columns i.e. add two new columns called *product_code* and *product_number*, containing the product code and number respectively

3: Add product categories

You learn that the product codes actually represent the following product categories:

- p = Smartphone
- v = TV
- x = Laptop
- q = Tablet

In order to make the data more readable, add a column with the product category for each record.

4: Add full address for geocoding

You'd like to view the customer information on a map. In order to do that, the addresses need to be in a form that can be easily geocoded. Create a new column *full_address* that concatenates the three address fields (*address*, *city*, *country*), separated by commas.

5: Create dummy variables for company and product category

Both the company name and product category are categorical variables i.e. they take only a fixed set of values. In order to use them in further analysis you need to create dummy variables. Create dummy binary variables for each of them with the prefix *company_* and *product_* i.e.

1. Add four binary (1 or 0) columns for company: *company_philips*, *company_akzo*, *company_van_houten* and *company_unilever*
2. Add four binary (1 or 0) columns for product category: *product_smartphone*, *product_tv*, *product_laptop* and *product_tablet*

6: Submit the project on Github

Include your code, the original data as a CSV file *refine_original.csv*, and the cleaned up data as a CSV file called *refine_clean.csv*.

