Data Wrangling Exercise 1: Basic Data Manipulation

Suresh Date- 2018-01-26

Load required packages

```
library(dplyr) # DW pkg
library(tidyr) # DW pkg
library(data.table) # Base pkg for DescTools
library(DescTools) # For LIKE and ANY function
library(readr) # write faster csv files
```

Read input file *"refine_original.csv"*

```
orginal <- read.csv("refine_original.csv")
glimpse(orginal)
```

```
## Observations: 25
## Variables: 6
## $ company               <fctr> Phillips, phillips, philips, phllips, p...
## $ Product.code...number <fctr> p-5, p-43, x-3, x-34, x-12, p-23, v-43,...
## $ address               <fctr> Groningensingel 147, Groningensingel 14...
## $ city                  <fctr> arnhem, arnhem, arnhem, arnhem, arnhem,...
## $ country               <fctr> the netherlands, the netherlands, the n...
## $ name                  <fctr> dhr p. jansen, dhr p. hansen, dhr j. Ga...
```

```
orginal <- data.frame(orginal)
```

Clean up brand names

```
comp1 <- orginal$company

for (i in seq(length(comp1)))
{
  tolower(comp1[i])
  if (comp1[i] %like any% c("%li%", "p%", "f%", "%ps")) {
    comp1[i] <- "phillips"
  } else if (comp1[i] %like% "ak%") {
    comp1[i] <- "akzo"
  } else if (comp1[i] %like% "van%") {
    comp1[i] <- "van houten"
  } else if (comp1[i] %like% "un%") {
    comp1[i] <- "unilever"
  }
}
orginal$company <- comp1
```

Function to name product_catetory

```
fetch <- function(x) {
  ifelse(x == "p", "Smartphone",
    ifelse(x == "v", "TV",
      ifelse(x == "x", "Laptop",
```

```
        ifelse(x == "q", "Tablet", NA)
      )
    )
  )
}
```

Edit rest of columns

```
clean <- orginal %>%
  separate(Product.code...number, c("product_code", "product_number"), "-") %>%
  mutate(company = tolower(company)) %>%
  mutate(product_category = fetch(product_code)) %>%
  unite(full_address, address, city, country, sep = ", ") %>%
  mutate(company_philips = ifelse(company == "phillips", 1, 0)) %>%
  mutate(company_akzo = ifelse(company == "akzo", 1, 0)) %>%
  mutate(company_van_houten = ifelse(company == "van houten", 1, 0)) %>%
  mutate(company_unilever = ifelse(company == "unilever", 1, 0)) %>%
  mutate(product_smartphone = ifelse(product_code == "p", 1, 0)) %>%
  mutate(product_tv = ifelse(product_code == "v", 1, 0)) %>%
  mutate(product_laptop = ifelse(product_code == "x", 1, 0)) %>%
  mutate(product_tablet = ifelse(product_code == "q", 1, 0))
glimpse(clean)
```

```
## Observations: 25
## Variables: 14
## $ company            <chr> "phillips", "phillips", "phillips", "philli...
## $ product_code       <chr> "p", "p", "x", "x", "x", "p", "v", "v", "x"...
## $ product_number     <chr> "5", "43", "3", "34", "12", "23", "43", "12...
## $ full_address       <chr> "Groningensingel 147, arnhem, the netherlan...
## $ name               <fctr> dhr p. jansen, dhr p. hansen, dhr j. Ganse...
## $ product_category   <chr> "Smartphone", "Smartphone", "Laptop", "Lapt...
## $ company_philips    <dbl> 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1...
## $ company_akzo       <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0...
## $ company_van_houten <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ company_unilever   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ product_smartphone <dbl> 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0...
## $ product_tv         <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1...
## $ product_laptop     <dbl> 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0...
## $ product_tablet     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0...
```

Write output to final file

```
clean %>% write_csv("refine_clean.csv")
```