Data Wrangling Exercise 1

Swati Jani Joshi

June 14, 2016

This is an R Markdown document containing the code for the exercise 1.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
refine_original <- read.csv("~/Documents/DW/DW_1/refine_original.csv")
View(refine_original)
df <- refine_original
company_names <- c("philips", "akzo", "unilever", "van houten")
df1 <- df %>% mutate(company = tolower(company), first = substr(company,1,1))
df <- df1 %>% filter(first == "p" | first == "f'") %>% mutate (company = company_names[1])
company_initials <- c("p","a","u","v")
for (j in 2:4){df2 <- df1 %>% filter(first == company_initials[j]) %>% mutate(company = company_names[j]
df <- rbind(df, df2)}
df$first <- NULL
df <- df %>% separate(Product.code...number, c("product_code", "product_number"), sep = '-')
category <- list(p = 'Smartphone', v = 'TV', x = 'Laptop', q = 'Tablet')
df <- df %>% mutate(product_category = paste0(category[product_code]))
df <- df %>% mutate(full_address = paste(address, city, country, sep = ","))
for (k  in 1:4) {
  comp <- company_names[k]
  dummy_var1 <- paste("company", comp, sep ="_")
  df[[paste0(dummy_var1)]]<- as.numeric(df$company == paste0(comp))
}
for (l in 1:4) {
  product <- category[l]
  dummy_var2 <- paste("product", product, sep = "_")
  df[[paste0(dummy_var2)]]<- as.numeric((df$product_category == paste0(product)))
}
write.csv(df, file = "refine_clean.csv")
```