Data Wrangling Exercise 1

Ankita Hanmiah

November 15, 2018

## Load Data in RStudio and into Data Frame

library("readr")  
library("dplyr")  
library("tidyr")  
refine\_original <- read\_csv("refine\_original.csv")  
df\_refine <- data.frame(refine\_original)

## Clean Up Brand Names

df\_refine$company <- tolower(df\_refine$company)  
df\_refine <- arrange(df\_refine, company)  
df\_refine$company <- sub("ak.\*", "akzo", df\_refine$company)  
df\_refine$company <- sub(".\*ps", "philips", df\_refine$company)  
df\_refine$company <- sub("un.\*", "unilever", df\_refine$company)

## Separate Product Code and Number

df\_refine <- separate(df\_refine, Product.code...number, c("product\_code", "product\_number"), sep = "-")

## Add Product Categories

refine\_p <- df\_refine %>% filter(product\_code == "p") %>% mutate(product\_category = "Smartphone")  
refine\_v <- df\_refine %>% filter(product\_code == "v") %>% mutate(product\_category = "TV")  
refine\_x <- df\_refine %>% filter(product\_code == "x") %>% mutate(product\_category = "Laptop")  
refine\_q <- df\_refine %>% filter(product\_code == "q") %>% mutate(product\_category = "Tablet")  
df\_refine <- bind\_rows(refine\_p, refine\_v, refine\_x, refine\_q)

## Add Full Address For Geocoding

df\_refine <- unite(df\_refine, "full address", address, city, country, sep = ",")

## Create Dummy Variables

df\_refine <- df\_refine %>% mutate(company\_akzo = (ifelse(company %in% "akzo", yes =1, no = 0)))  
df\_refine <- df\_refine %>% mutate(company\_philips = (ifelse(company %in% "philips", yes = 1, no = 0)))   
df\_refine <- df\_refine %>% mutate(company\_unilever = (ifelse(company %in% "unilever", yes = 1, no = 0)))  
df\_refine <- df\_refine %>% mutate(company\_van\_houten = (ifelse(company %in% "van houten", yes = 1, no = 0)))  
df\_refine <- df\_refine %>% mutate(product\_smartphone = (ifelse(product\_category %in% "Laptop", yes = 1, no = 0)))  
df\_refine <- df\_refine %>% mutate(product\_tv = (ifelse(product\_category %in% "TV", yes = 1, no = 0)))  
df\_refine <- df\_refine %>% mutate(product\_laptop = (ifelse(product\_category %in% "Laptop", yes = 1, no = 0)))  
df\_refine <- df\_refine %>% mutate(product\_tablet = (ifelse(product\_category %in% "Tablet", yes = 1, no = 0)))

## Save Cleaned Data

``` refine\_clean <- df\_refine write.csv(refine\_clean, file = “refine\_clean.csv”)