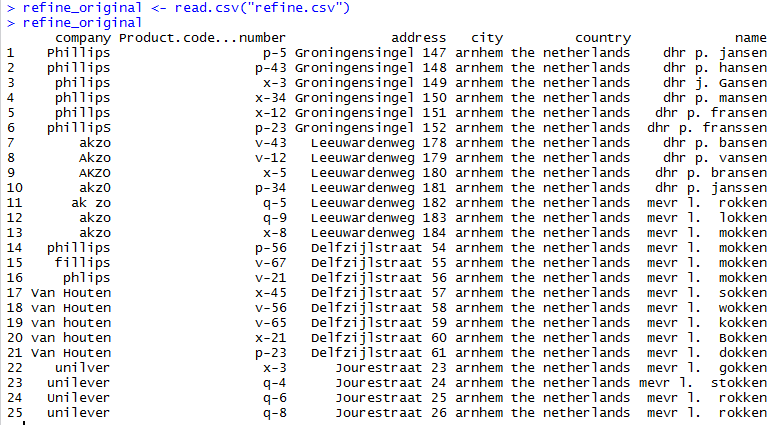
**Data Science Unit 3: Exercise 1 Stephen Golden**

**2016 08 30**

Using R, clean [the provided excel spreadsheet] to make it easier to visualize and analyze.

Save the spreadsheet to *refine\_original.csv*. Then, I copied the file to *refine.csv*.

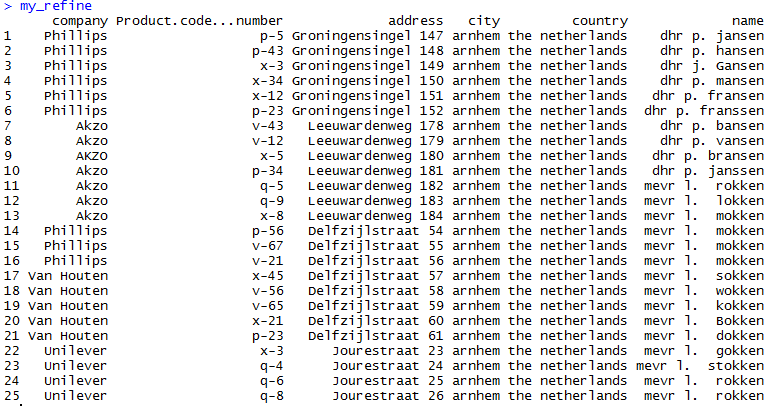
Read the CSV into a data frame



Clean it up:

Sample clean-up command: my\_refine$company[my\_refine$company=="unilver"] <- "Unilever"

Data frame after name clean up:



**Place “Product.code…number” Column data into a Data Frame**

Put column data into a vector (Take the second column and read it into a vector)

colvector <- my\_refine[,2]

**Substring the code-nums values into individual vectors**

codevector <- substr(colvector, 1, 1)

numvector <- substr(colvector, 3, 4)

**Create new column names**

my\_refine["prod\_code"] <- NA

my\_refine["prod\_num"] <- NA

**Add vectors to columns in data frame**

my\_refine$prod\_code <- codevector

my\_refine$prod\_num <- numvector

**Replace product codes with their product names**

my\_refine$prod\_code[my\_refine$prod\_code=="p"] <- "Smartphone"

my\_refine$prod\_code[my\_refine$prod\_code=="v"] <- "TV"

my\_refine$prod\_code[my\_refine$prod\_code=="x"] <- "Laptop"

my\_refine$prod\_code[my\_refine$prod\_code=="q"] <- "Tablet"

**Create new address column (This will also reset an existing column)**

my\_refine["full\_address"] <- NA

**Concatenate address fields into a single field**

my\_refine$full\_address <- paste(my\_refine $address, my\_refine $city, my\_refine $country, sep=", ")

**Create binary variable columns (like the Info/Man Master Bit List) for Company**

my\_refine$company\_phillips <- ifelse(my\_refine$company=="Phillips", 1, 0)

my\_refine$company\_akzo <- ifelse(my\_refine$company=="Akzo", 1, 0)

my\_refine$company\_van\_houten <- ifelse(my\_refine$company=="Van Houten", 1, 0)

my\_refine$company\_unilever <- ifelse(my\_refine$company=="Unilever", 1, 0)

**Create binary variable columns (like the Info/Man Master Bit List) for Product**

my\_refine$product\_smartphone <- ifelse(my\_refine$prod\_code=="Smartphone", 1, 0)

my\_refine$product\_tv <- ifelse(my\_refine$prod\_code =="TV", 1, 0)

my\_refine$product\_laptop <- ifelse(my\_refine$prod\_code =="Laptop", 1, 0)

my\_refine$product\_tablet <- ifelse(my\_refine$prod\_code =="Tablet", 1, 0)

**Write cleaned up data to csv**

write.csv(my\_refine, file = "refine\_clean.csv")