Dsouza_Clinton_Assignment#4

Clinton Dsouza

10/1/2019

Install all the libraries required for this assignment

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
library(forcats)
library(readxl)
```

Loading the farmers_market dataset

```
farmermkt <- read_csv("farmers_market.csv.csv",
                      col_types = cols(Season4Date = col_character(), Season4Time
                                       = col_character()))
head(farmermkt,75)
```

```
## # A tibble: 75 x 59
##      FMID MarketName Website Facebook Twitter Youtube OtherMedia street
##     <dbl> <chr>      <chr>   <chr>    <chr>   <chr>   <chr>      <chr>
##  1 1.02e6 Caledonia~ https:~ https:/~ <NA>    <NA>    <NA>       <NA>
##  2 1.02e6 Stearns H~ http:/~ Stearns~ <NA>    <NA>    <NA>       6975 ~
##  3 1.01e6 106 S. Ma~ http:/~ <NA>     <NA>    <NA>    <NA>       106 S~
##  4 1.01e6 10th Stee~ <NA>    <NA>     <NA>    <NA>    http://ag~ 10th ~
##  5 1.00e6 112st Mad~ <NA>    <NA>     <NA>    <NA>    <NA>       112th~
##  6 1.01e6 12 South ~ http:/~ 12_Sout~ @12sou~ <NA>    @12southf~ 3000 ~
##  7 1.01e6 125th Str~ http:/~ https:/~ https:~ <NA>    Instagram~ 163 W~
##  8 1.01e6 12th & Br~ <NA>    https:/~ <NA>    <NA>    https://w~ 12th ~
##  9 1.01e6 14&U Farm~ <NA>    https:/~ https:~ <NA>    <NA>       1400 ~
## 10 1.01e6 14th & Ke~ <NA>    https:/~ 14KenFM <NA>    instagram~ 5500 ~
## # ... with 65 more rows, and 51 more variables: city <chr>, County <chr>,
## #   State <chr>, zip <chr>, Season1Date <chr>, Season1Time <chr>,
## #   Season2Date <chr>, Season2Time <chr>, Season3Date <chr>,
## #   Season3Time <chr>, Season4Date <chr>, Season4Time <chr>, x <dbl>,
## #   y <dbl>, Location <chr>, Credit <chr>, WIC <chr>, WICcash <chr>,
## #   SFMNP <chr>, SNAP <chr>, Organic <chr>, Bakedgoods <chr>,
## #   Cheese <chr>, Crafts <chr>, Flowers <chr>, Eggs <chr>, Seafood <chr>,
## #   Herbs <chr>, Vegetables <chr>, Honey <chr>, Jams <chr>, Maple <chr>,
## #   Meat <chr>, Nursery <chr>, Nuts <chr>, Plants <chr>, Poultry <chr>,
## #   Prepared <chr>, Soap <chr>, Trees <chr>, Wine <chr>, Coffee <chr>,
## #   Beans <chr>, Fruits <chr>, Grains <chr>, Juices <chr>,
## #   Mushrooms <chr>, PetFood <chr>, Tofu <chr>, WildHarvested <chr>,
## #   updateTime <chr>
```

WarmUp

```r
farmermktcs <- paste(farmermkt$city, farmermkt$State, sep = ", ")
head(farmermktcs, 75)
```

```
##  [1] "Danville, Vermont"                "Parma, Ohio"
##  [3] "Six Mile, South Carolina"         "Lamar, Missouri"
##  [5] "New York, New York"               "Nashville, Tennessee"
##  [7] "New York, New York"               "Wilmington, Delaware"
##  [9] "Washington, District of Columbia" "Washington, District of Columbia"
## [11] "Portland, Oregon"                 "Bronx, New York"
## [13] "New York, New York"               "Minneapolis, Minnesota"
## [15] "Richmond, Virginia"               "Philadelphia, Pennsylvania"
## [17] "Scottsbluff, Nebraska"            "Charleston, Illinois"
## [19] "Chiefland, Florida"               "Woodinville, Washington"
## [21] "Topeka, Kansas"                   "Philadelphia, Pennsylvania"
## [23] "Highlands, New Jersey"            "North Logan, Utah"
## [25] "Philadelphia, Pennsylvania"       "Philadelphia, Pennsylvania"
## [27] "Amherst, Virginia"                "Dayton, Ohio"
## [29] "Morris, Illinois"                 "Rosemary Beach, Florida"
## [31] "Ewing, New Jersey"                "Baltimore, Maryland"
## [33] "Philadelphia, Pennsylvania"       "Indianapolis, Indiana"
## [35] "Sparks, Nevada"                   "Rochester, New York"
## [37] "Philadelphia, Pennsylvania"       "Larimer, Colorado"
## [39] "Indianapolis, Indiana"            "Philadelphia, Pennsylvania"
## [41] "New York, New York"               "Philadelphia, Pennsylvania"
## [43] "Chicago, Illinois"                "New York, New York"
## [45] "New York, New York"               "Dothan, Alabama"
## [47] "Cedar Rapids, Iowa"               "New York, New York"
## [49] "New York, New York"               "Salina, Kansas"
## [51] "SALT LAKE CITY, Utah"             "Boynton Beach, Florida"
## [53] "Tallahassee, Florida"             "Annapolis, Maryland"
## [55] "Abbeville, South Carolina"        "Abbeville, Alabama"
## [57] "Abbotsford, Wisconsin"            "Minneapolis, Minnesota"
## [59] "Aberdeen, South Dakota"           "Aberdeen, Washington"
## [61] "Abilene, Kansas"                  "Abingdon, Virginia"
## [63] "New York, New York"               "Abington, Massachusetts"
## [65] "Clarks Summit, Pennsylvania"      "Abita Springs, Louisiana"
## [67] "Albququerque, New Mexico"         "Mount Bethe, Pennsylvania"
## [69] "Town Hill, Maine"                 "Opelousas, Louisiana"
## [71] "Rome, New York"                   "Birch Tree, Missouri"
## [73] "Loxahatchee, Florida"             "Acton, Massachusetts"
## [75] "Acushnet, Massachusetts"
```

Q1. Clean Facebook and Twitter Column a. Cleaning the Facebook column to contain only the facebook username

```r
FbClean <- gsub("(.*).com\\/", "", farmermkt$Facebook)
FbClean <- gsub("^(pages\\/)", "", FbClean)
FbClean <- gsub("(\\/)$", "", FbClean)
FbClean <- gsub("(\\?)\\w.*", "", FbClean)
FbClean <- gsub("(\\/)\\w.*", "", FbClean)
FbClean <- gsub("(\\-)\\w.*", "", FbClean)
head(FbClean,75)
```

```
##  [1] "Danville.VT.Farmers.Market"
##  [2] "StearnsHomesteadFarmersMarket"
##  [3] NA
##  [4] NA
##  [5] NA
##  [6] "12_South_Farmers_Market"
##  [7] "125thStreetFarmersMarket"
##  [8] "12th"
##  [9] "14UFarmersMarket"
## [10] "14KennnedyFarmersMarket"
## [11] NA
## [12] "CommunityFoodAction"
## [13] "ManhattanGreenmarkets"
## [14] NA
## [15] "17thStreetFarmersMarket"
## [16] NA
## [17] "ScottsbluffFarmersMarket"
## [18] "18th Street Farmers Market"
## [19] "1927"
## [20] "21Acres"
## [21] NA
## [22] NA
## [23] "Highlands"
## [24] "25th Street Market - North Logan at the Library"
## [25] NA
## [26] NA
## [27] "second stage of AMherst"
## [28] "2ndStreetMarket"
## [29] "3"
## [30] "30aFarmersMarket"
## [31] "31mainfarmersmarket"
## [32] "Baltimores"
## [33] NA
## [34] NA
## [35] "39 North"
## [36] NA
## [37] NA
## [38] NA
## [39] NA
## [40] NA
## [41] "ManhattanGreenmarkets"
## [42] NA
## [43] "61market"
## [44] "ManhattanGreenmarkets"
## [45] "ManhattanGreenmarkets"
## [46] NA
## [47] NA
## [48] "ManhattanGreenmarkets"
## [49] "ManhattanGreenmarkets"
## [50] NA
## [51] "9thwestfarmersmarket"
## [52] "OrganicProduceDelivery"
## [53] NA
## [54] NA
## [55] NA
## [56] "Abbeville"
## [57] NA
## [58] NA
## [59] NA
## [60] "AberdeenSundayMarket"
## [61] NA
## [62] "abingdonfarmersmarket"
## [63] "ManhattanGreenmarkets"
## [64] NA
## [65] NA
## [66] "abitasprings.farmersmarket"
## [67] NA
## [68] "Apple"
## [69] "AcadiaFarmersMarket"
## [70] NA
## [71] NA
## [72] NA
## [73] "AcreageGreenmarket"
## [74] NA
## [75] "Acushnet"
```

b. Cleaning the Twitter column to contain only the Twitter username

```
TwClean <- gsub("(?i)(.*).com\\/", "", farmermkt$Twitter)
TwClean <- gsub("@", "", TwClean)
head(TwClean,75)
```

```
##  [1] NA                 NA                 NA
##  [4] NA                 NA                 "12southfrmsmkt"
##  [7] "FarmMarket125th"  NA                 "14UFarmersMkt"
## [10] "14KenFM"          NA                 "GoodEatsBX"
## [13] NA                 NA                 NA
## [16] NA                 NA                 NA
## [19] NA                 "21acres"          NA
## [22] NA                 NA                 NA
## [25] NA                 NA                 NA
## [28] NA                 NA                 NA
## [31] "31mainfarmmarkt"  NA                 NA
## [34] NA                 "39North Downtown" NA
## [37] NA                 NA                 NA
## [40] NA                 NA                 NA
## [43] "61market"         NA                 NA
## [46] NA                 NA                 NA
## [49] NA                  NA                  "peoplesmarket"
## [52] "OrganicGrownDr"    NA                  NA
## [55] NA                  NA                  NA
## [58] NA                  NA                  NA
## [61] NA                  NA                  NA
## [64] NA                  NA                  NA
## [67] NA                  NA                  NA
## [70] NA                  NA                  NA
## [73] NA                  NA                  NA
```

Q2. Cleaning the city column

```
clcity <- str_to_lower(farmermkt$city, locale = "en")
Clcity <- gsub(",.*", "", farmermkt$city)
head(Clcity,75)
```

```
##  [1] "Danville"       "Parma"          "Six Mile"       "Lamar"
##  [5] "New York"       "Nashville"      "New York"       "Wilmington"
##  [9] "Washington"     "Washington"     "Portland"       "Bronx"
## [13] "New York"       "Minneapolis"    "Richmond"       "Philadelphia"
## [17] "Scottsbluff"    "Charleston"     "Chiefland"      "Woodinville"
## [21] "Topeka"         "Philadelphia"   "Highlands"      "North Logan"
## [25] "Philadelphia"   "Philadelphia"   "Amherst"        "Dayton"
## [29] "Morris"         "Rosemary Beach" "Ewing"          "Baltimore"
## [33] "Philadelphia"   "Indianapolis"   "Sparks"         "Rochester"
## [37] "Philadelphia"   "Larimer"        "Indianapolis"   "Philadelphia"
## [41] "New York"       "Philadelphia"   "Chicago"        "New York"
## [45] "New York"       "Dothan"         "Cedar Rapids"   "New York"
## [49] "New York"       "Salina"         "SALT LAKE CITY" "Boynton Beach"
## [53] "Tallahassee"    "Annapolis"      "Abbeville"      "Abbeville"
## [57] "Abbotsford"     "Minneapolis"    "Aberdeen"       "Aberdeen"
## [61] "Abilene"        "Abingdon"       "New York"       "Abington"
## [65] "Clarks Summit"  "Abita Springs"  "Albququerque"   "Mount Bethe"
## [69] "Town Hill"      "Opelousas"      "Rome"           "Birch Tree"
## [73] "Loxahatchee"    "Acton"          "Acushnet"
```

Cleaning the street column

```
clstreet <- farmermkt$street
clstreet <- str_replace_all(clstreet, c("Street" = "St", "Streets" ="St", "St."="St", "street"="St"))
clstreet <- gsub("\\s[a|A]nd", "&", clstreet)
clstreet <- gsub("\\s[A|a]venue", "Ave\\.", clstreet)
clstreet <- gsub("\\s[B|b]roadway", "Bdwy\\.", clstreet)
clstreet <- gsub("\\s[R|r]oad", "Rd\\.", clstreet)
head(clstreet,75)
```

```
##  [1] NA
##  [2] "6975 RidgeRd."
##  [3] "106 S. Main St"
##  [4] "10th Stand Poplar"
##  [5] "112th MadisonAve."
##  [6] "3000 Granny White Pike"
##  [7] "163 West 125th Stand Adam Clayton Powell, Jr. Blvd."
##  [8] "12th & Brandywine St"
##  [9] "1400 U StNW"
## [10] "5500 ColoradoAve., NW"
## [11] "NE 16th Ave &Bdwy."
## [12] "NE Corner of 170th St & TownsendAve."
## [13] "175th Stbetween Wadsworth &Bdwy."
## [14] "1622 6th StNE"
## [15] "100 North 17th St"
## [16] "18th Stand Christian St"
## [17] "18th&Bdwy."
## [18] "825 18th St"
## [19] "NE 7th Ave"
## [20] "13701 NE 171st St"
## [21] "SW 21st& Oakley"
## [22] "22nd& Tasker St"
## [23] "71 WaterwitchAve."
## [24] "475 East 2500 North"
## [25] "26th Stand W AlleghenyAve."
## [26] "29th& Wharton St"
## [27] "194 second St"
## [28] "600 E. 2nd St"
## [29] "123 W. Illinois ave."
## [30] "Rosmary Beach Town Center"
## [31] "1928 PenningtonRd."
## [32] "E. 32nd & Barclay St"
## [33] "N 33rd& Diamond St"
## [34] "3808 N Meridian St"
## [35] "Downtown Sparks Victorian Ave"
## [36] "441 ParsellsAve."
## [37] "N 4th Stand W. LehighAve."
## [38] "315 East 4th St"
## [39] "5200 N. ShadelandAve."
## [40] "N 52nd Stand HaverfordAve."
## [41] "W 57 St & 9 Ave"
## [42] "58th Stand ChesterAve."
## [43] "6100 S. Blackstone Ave"
## [44] "Columbus - W 78 & 81 St."
## [45] "E 82nd St - 1st & York Ave"
## [46] NA
## [47] "8th Ave & 2nd StSE"
## [48] "1st Ave - E 92nd & 93 St."
## [49] "W 97 St & Columbus"
## [50] "304 West GrandAve."
## [51] "1060 South 900 West"
## [52] "Lee Rd. Farm"
## [53] "229 Lake Ella Drive"
## [54] "2001 Medical Parkway, Sajak Pavilion"
## [55] "118 Trinity Stat Livery Stble"
## [56] "Kirkland St"
## [57] "1011 East Spruce St"
## [58] "800 E 28th St"
## [59] "2nd Ave., S.E. & S. Lincoln St"
## [60] "Broadway between Heron & Stte St"
## [61] "East 1st & Buckeye St"
## [62] "Corner of Remsburg Drive & Cummings St"
## [63] "W12 St & 8th Ave"
## [64] "362 Plymouth St"
## [65] "12055 Rose Drove"
## [66] "22049 Main St"
## [67] "NE parking lot of ABQ Uptown shopping center"
## [68] "690 AlleghenyRd."
## [69] NA
## [70] "801 Foreman Drive"
## [71] "115 Black River Blvd"
## [72] "RR1 Box 146"
## [73] "6701 140th Ave. North"
## [74] "1 Pearl St"
## [75] "186 Leonard St"
```

Q3. Creating a tibble that contains the % of farmers in their respective states who have a facebook or twitter account

```
Farmer_Online_account  <- farmermkt %>%
                      select(State, Facebook, Twitter) %>%
                      group_by(State) %>%
                      summarise(TotalMarket = n(), Fbcount =
                      sum(!is.na(Facebook)),      percent_FB=(Fbcount/TotalMarket)*100,
                      TWcount = sum(!is.na(Twitter)), percent_TW = (TWcount/TotalMarket)*100, TWFb =sum

Farmer_Online_account
```

```
## # A tibble: 53 x 8
##    State TotalMarket Fbcount percent_FB TWcount percent_TW  TWFb
##    <chr>       <int>   <int>      <dbl>   <int>      <dbl> <int>
##  1 Alab~         140      37       26.4       9       6.43    46
##  2 Alas~          37      17       45.9       4      10.8     21
##  3 Ariz~          93      54       58.1      25      26.9     79
##  4 Arka~         111      58       52.3       5       4.50    63
##  5 Cali~         759     316       41.6     110      14.5    426
##  6 Colo~         160      70       43.8      16      10       86
##  7 Conn~         158      55       34.8      18      11.4     73
##  8 Dela~          36      22       61.1       4      11.1     26
##  9 Dist~          58      30       51.7      25      43.1     55
## 10 Flor~         264     115       43.6      23       8.71   138
## # ... with 43 more rows, and 1 more variable: percent_TWFb <dbl>
```

```
summary_online_account <- tibble(state= Farmer_Online_account$State, percent_FB= Farmer_Online_account$

head(summary_online_account,75)
```

```
## # A tibble: 53 x 4
##    state                percent_FB percent_TW percent_FBorTW
##    <chr>                     <dbl>      <dbl>          <dbl>
##  1 Alabama                    26.4       6.43           32.9
##  2 Alaska                     45.9      10.8            56.8
##  3 Arizona                     58.1      26.9            84.9
##  4 Arkansas                    52.3       4.50           56.8
##  5 California                  41.6      14.5            56.1
##  6 Colorado                    43.8      10              53.8
##  7 Connecticut                 34.8      11.4            46.2
##  8 Delaware                    61.1      11.1            72.2
##  9 District of Columbia        51.7      43.1            94.8
## 10 Florida                     43.6       8.71           52.3
## # ... with 43 more rows
```

Q4. forcats::fct_recode() The farmer market names are quite long in this data set. Every observation in the "MarketName" has a unique observation. By using fct_recode() function, we will not be able to change/rename every observation in the variable "Market Name". The purpose of fct_recode() is to set distinct categories in the variable such that the entire variable is grouped with fixed categorical observations. This will result in better visualization and further analysis. fct_recode() function will take a fixed string and match the exact string to replace it as given in the argument. In this situation, since the column MarketName has lots of variation in the names, we will have to use regex to rename the column which is a better solution.

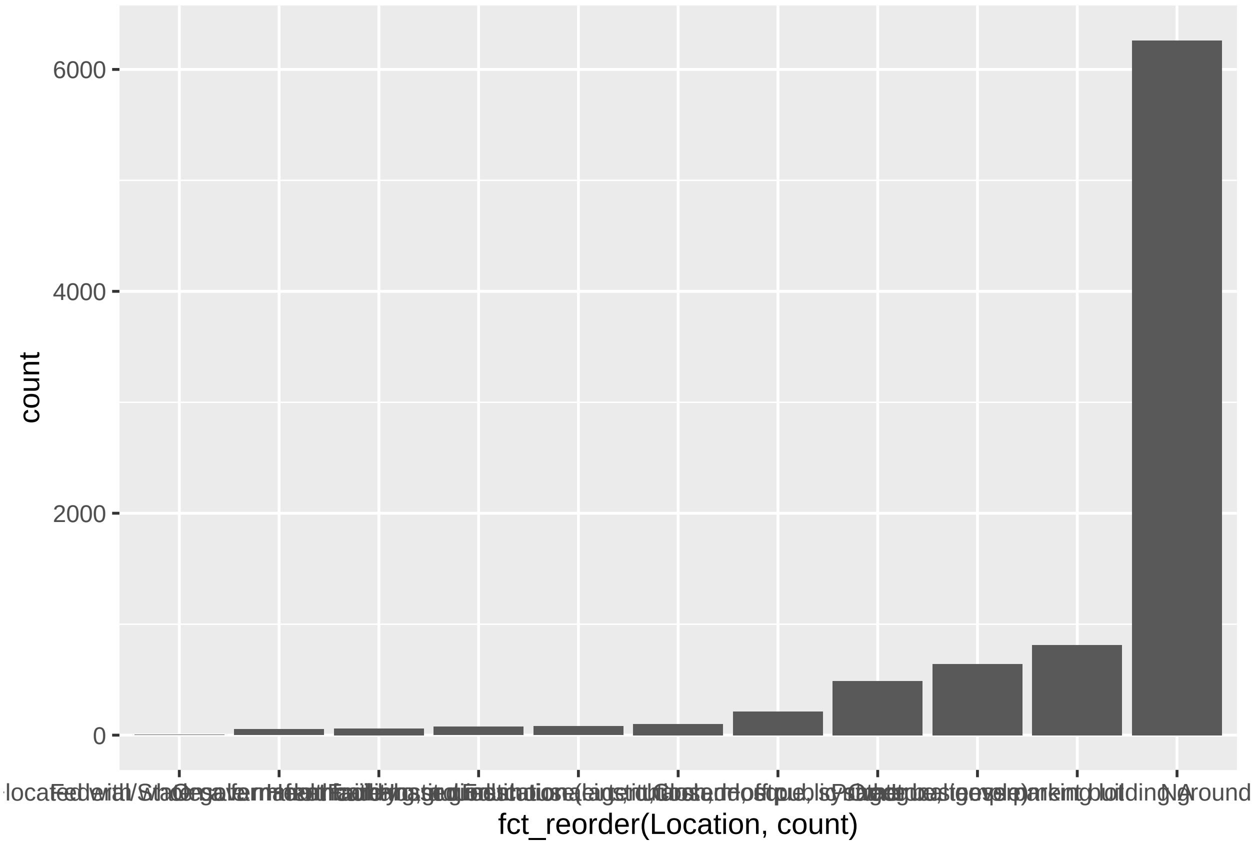
Creating a tibble using dplyr which has the details of location type

```
LocationType <- farmermkt %>%
  select(Location) %>%
  group_by(Location) %>%
  summarise(count = n()) %>%
  arrange(desc(count))
head(LocationType,75)
```

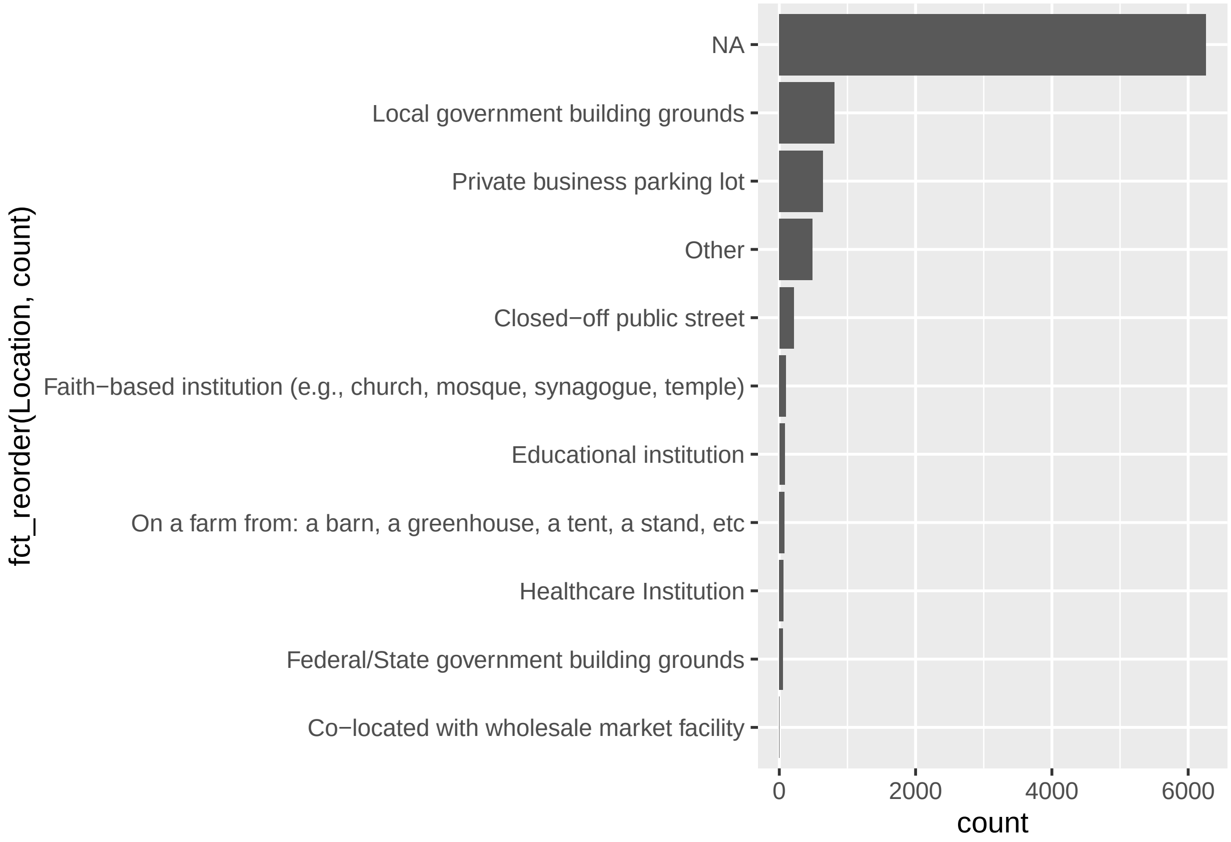
```
## # A tibble: 11 x 2
##    Location                                                       count
##    <chr>                                                          <int>
##  1 <NA>                                                            6262
##  2 Local government building grounds                                812
##  3 Private business parking lot                                     642
##  4 Other                                                            488
##  5 Closed-off public street                                         212
##  6 Faith-based institution (e.g., church, mosque, synagogue, temple)   101
##  7 Educational institution                                           80
##  8 On a farm from: a barn, a greenhouse, a tent, a stand, etc         75
##  9 Healthcare Institution                                            60
## 10 Federal/State government building grounds                         53
## 11 Co-located with wholesale market facility                          3
```

Plotting the graph of number of farmer markets as per location type

```
LTGraph <- ggplot(LocationType) + geom_bar(aes(fct_reorder(Location, count),count), stat = "identity")
LTGraph
```



```
LTGraph + coord_flip()
```



Q5. Perform Sanity check on the kyfprojects dataset

a. Reading the data set

```
kyf <- read_excel("kyfprojects.xls.xls")
head(kyf,50)
```

```
## # A tibble: 50 x 18
##    `Project Title` `Program Name` `Program Abbrev~  Year State Town  Zip
##    <chr>           <chr>          <chr>            <dbl> <chr> <chr> <chr>
##  1 "\"Buy Illinoi~ Specialty Cro~ SCBG              2009 IL    Spri~ 62702
##  2 "\"Growing Far~ Farmers Marke~ FMPP              2009 MN    Onam~ 56359
##  3 "\"Growing the~ Farmers Marke~ FMPP              2009 NM    Sant~ 87501
##  4 "\"Health Food~ Farmers Marke~ FMPP              2009 LA    Bato~ 70803
##  5 2009 RBEG-Ajo ~ Rural Busines~ RBEG              2009 AZ    Ajo   85321
##  6 2011 Internati~ Specialty Cro~ SCBG              2009 WV    Char~ 25305
##  7 21st Century Y~ Specialty Cro~ SCBG              2009 AL    Mont~ 36107
##  8 "A \"Field to ~ Specialty Cro~ SCBG              2009 DC    Wash~ 20010
##  9 A Demand Drive~ Federal-State~ FSMIP             2009 NJ    Camd~ 08901
## 10 A Garlic Commu~ Sustainable A~ SARE              2009 ND    Full~ <NA>
## # ... with 40 more rows, and 11 more variables: `USDA Agency` <chr>, `USDA
## #   Mission Area` <chr>, Recipient <chr>, `Recipient Type` <chr>, `Funding
## #   Amount ($)` <dbl>, `Funding Type` <chr>, Description <chr>,
## #   Topic_A <chr>, Topic_B <chr>, Topic_C <chr>, `More Information` <chr>
```

b. Check if Program Abbreviation has the same match across all the Program Names i.e. Program abbreviation is the same overall for every distinct Program names

```r
kyfPN <- kyf$`Program Name`
kyfPN <- gsub("[a-z \\s \\-]", "", kyfPN)
kyfPN <- gsub("(\\s Grants)$", "", kyfPN)
kyfPN <- gsub("^CFPCG$", "CFP", kyfPN)
kyfPN <- gsub("^RMEOP$", "RMEO", kyfPN)
kyfPN <- gsub("^B&ILG", "B and I", kyfPN)
kyfPN <- gsub("^HFCG", "HFC", kyfPN)

kyfPA <- kyf$`Program Abbreviation`

str_detect(kyfPA, kyfPN)
```

```
##    [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [14] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [27] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [40] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [53] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [66] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [79] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [92] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [105] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [118] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [131] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [144] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [157] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [170] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [183] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [196] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [209] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [222] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [235] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [248] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [261] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [274] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [287] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [300] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [313] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [326] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [339] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [352] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [365] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [378] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [391] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [404] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [417] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [430] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [443] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [456] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [469] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [482] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [495] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [508] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [521] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [534] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [547] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [560] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [573] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [586] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [599] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [612] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [625] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [638] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [651] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [664] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [677] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [690] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [703] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [716] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [729] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [742] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [755] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [768] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [781] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [794] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [807] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [820] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [833] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [846] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [859] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [872] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [885] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [898] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [911] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [924] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [937] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [950] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [963] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [976] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [989] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1002] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1015] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1028] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1041] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1054] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1067] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1080] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1093] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1106] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1119] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1132] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1145] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1158] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1171] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1184] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1197] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1210] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1223] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1236] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1249] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1262] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1275] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1288] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1314] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1327] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1340] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1353] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1366] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1379] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1392] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1405] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1418] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1431] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1444] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1457] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1470] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1483] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1496] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1509] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1522] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1535] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1548] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1561] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1574] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1587] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1600] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1613] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1626] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1639] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1652] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1665] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1678] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1691] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1704] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1717] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1730] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1743] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1756] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1769] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1782] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1795] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1808] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1821] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1834] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1847] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1860] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1873] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1886] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1899] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1912] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1925] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1938] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1951] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1964] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1977] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1990] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2003] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2016] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2029] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2042] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2055] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2068] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2081] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2094] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2107] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2120] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2133] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2146] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2159] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2172] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2185] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2198] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2211] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2224] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2237] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2250] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2263] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2276] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2289] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2302] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2315] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2328] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2341] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2354] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2367] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Since everything returns a TRUE value that means the Program Name matches the Program Abbreviation across all rows and is distinct for a particular set

c. Cleaning all categorical variables and creating a distinct set by replacing similar entries

c.1. Clean the column of Funding Type

```r
kyf %>%
  group_by(`Funding Type`) %>%
  count(`Funding Type`)
```

```
## # A tibble: 2 x 2
## # Groups:   Funding Type [2]
##   `Funding Type`     n
##   <chr>          <int>
## 1 Grant           2308
## 2 Loan              71
```

This set returns a clean distinct set

c.2. Clean the column of USDA Agency

```
kyf$`USDA Agency` <- gsub("^[N].*[e]$","NIFA",kyf$`USDA Agency`)
kyf$`USDA Agency` <- gsub("^[F].*[e]$","FNS",kyf$`USDA Agency`)
kyf$`USDA Agency` <- gsub("^[(Ag)|(ag)].*[e]$","AMS",kyf$`USDA Agency`)
kyf$`USDA Agency` <- gsub("^[R].*[t]$","RBS",kyf$`USDA Agency`)
kyf$`USDA Agency` <- gsub("^[R].[S]$","RBS",kyf$`USDA Agency`)
kyf %>%
  group_by(`USDA Agency`) %>%
  count(`USDA Agency`)
```

```
## # A tibble: 5 x 2
## # Groups:   USDA Agency [5]
##   `USDA Agency`     n
##   <chr>         <int>
## 1 AMS            1392
## 2 FNS              16
## 3 NIFA            287
## 4 RBS             657
## 5 RMA              27
```

This returns a new clean distinct variable for USDA Agency

c.3. Clean the column of USDA Mission Area

```
kyf$`USDA Mission Area` <- gsub("^(Fo).*[s]$","Food, Nutrition and Consumer Services", kyf$`USDA Missio
kyf %>%
  group_by(`USDA Mission Area`) %>%
  count(`USDA Mission Area`)
```

```
## # A tibble: 5 x 2
## # Groups:   USDA Mission Area [5]
##   `USDA Mission Area`                      n
##   <chr>                                <int>
## 1 Farm and Foreign Agricultural Services    27
## 2 Food, Nutrition and Consumer Services     16
## 3 Marketing and Regulatory Programs       1392
## 4 Research, Education and Economics        287
## 5 Rural Development                        657
```

This returns a new clean variable for USDA MIssion Area

c.4. Recipient variable has a lot of different observations and they cannot be categorized into particular sets

c.5. Clean the column of Recipient Type

```
kyf$`Recipient Type` <- gsub("[N].*", "Nonprofit", kyf$`Recipient Type`)
kyf %>%
  group_by(`Recipient Type`) %>%
  count(`Recipient Type`)
```

```
## # A tibble: 7 x 2
## # Groups:   Recipient Type [7]
##   `Recipient Type`     n
##   <chr>            <int>
## 1 Academic           256
## 2 Business           108
## 3 Businesses           1
## 4 Government         430
## 5 Nonprofit         1320
## 6 Other                6
## 7 Producer           258
```

This returns a clean set for Recipient Type

c.6. Funding Type

```
kyf %>%
  group_by(`Funding Type`) %>%
  count(`Funding Type`)
```

```
## # A tibble: 2 x 2
## # Groups:   Funding Type [2]
##   `Funding Type`     n
##   <chr>          <int>
## 1 Grant           2308
## 2 Loan              71
```

d. Compare USDA Mission Area and USDA Agency

```
USDAAb <- gsub("^(M).*(s)$", "AMS", kyf$`USDA Mission Area`)
USDAAb <- gsub("^(R).*(t)$", "RBS", USDAAb)
USDAAb <- gsub("^(R).*(s)$", "NIFA", USDAAb)
USDAAb <- gsub("^(Fa).*(s)$", "RMA", USDAAb)
USDAAb <- gsub("^(Fo).*(s)$", "FNS", USDAAb)

str_detect(kyf$`USDA Agency`, USDAAb)
```

```
##    [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [14] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [27] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [40] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [53] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [66] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [79] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   [92] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [105] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [118] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [131] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [144] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [157] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [170] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [183] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [196] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [209] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [222] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [235] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [248] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [261] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [274] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [287] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [300] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [313] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [326] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [339] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [352] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [365] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [378] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [391] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [404] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [417] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [430] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [443] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [456] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [469] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [482] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [495] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [508] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [521] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [534] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [547] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [560] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [573] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [586] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [599] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [612] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [625] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [638] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [651] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [664] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [677] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [690] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [703] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [716] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [729] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [742] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [755] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [768] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [781] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [794] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [807] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [820] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [833] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [846] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [859] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [872] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [885] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [898] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [911] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [924] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [937] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [950] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [963] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [976] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##  [989] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1002] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1015] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1028] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1041] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1054] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1067] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1080] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1093] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1106] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1119] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1132] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1145] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1158] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1171] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1184] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1197] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1210] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1223] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1236] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1249] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1262] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1275] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1288] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1314] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1327] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1340] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1353] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1366] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1379] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1392] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1405] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1418] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1431] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1444] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1457] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1470] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1483] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1496] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1509] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1522] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1535] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1548] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1561] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1574] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1587] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1600] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1613] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1626] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1639] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1652] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1665] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1678] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1691] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1704] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1717] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1730] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1743] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1756] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1769] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1782] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1795] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1808] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1821] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1834] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1847] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1860] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1873] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1886] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1899] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1912] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1925] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1938] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1951] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1964] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1977] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [1990] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2003] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2016] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2029] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2042] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2055] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2068] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2081] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2094] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2107] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2120] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2133] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2146] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2159] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2172] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2185] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2198] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2211] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2224] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2237] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2250] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2263] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2276] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2289] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2302] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2315] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2328] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2341] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2354] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2367] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Since this returns all TRUE values, we can be now certain that the variables are matching across all observations. If this task was performed before the cleaning of variables, then we would not get TRUE values across every string.

e. Final view of the clean kyfprojects data

```
head(kyf,100)
```

```
## # A tibble: 100 x 18
##    `Project Title` `Program Name` `Program Abbrev~  Year State Town  Zip
##    <chr>           <chr>          <chr>            <dbl> <chr> <chr> <chr>
##  1 "\"Buy Illinoi~ Specialty Cro~ SCBG              2009 IL    Spri~ 62702
##  2 "\"Growing Far~ Farmers Marke~ FMPP              2009 MN    Onam~ 56359
##  3 "\"Growing the~ Farmers Marke~ FMPP              2009 NM    Sant~ 87501
##  4 "\"Health Food~ Farmers Marke~ FMPP              2009 LA    Bato~ 70803
##  5 2009 RBEG-Ajo ~ Rural Busines~ RBEG              2009 AZ    Ajo   85321
##  6 2011 Internati~ Specialty Cro~ SCBG              2009 WV    Char~ 25305
##  7 21st Century Y~ Specialty Cro~ SCBG              2009 AL    Mont~ 36107
##  8 "A \"Field to ~ Specialty Cro~ SCBG              2009 DC    Wash~ 20010
##  9 A Demand Drive~ Federal-State~ FSMIP             2009 NJ    Camd~ 08901
## 10 A Garlic Commu~ Sustainable A~ SARE              2009 ND    Full~ <NA>
## # ... with 90 more rows, and 11 more variables: `USDA Agency` <chr>, `USDA
## #   Mission Area` <chr>, Recipient <chr>, `Recipient Type` <chr>, `Funding
## #   Amount ($)` <dbl>, `Funding Type` <chr>, Description <chr>,
## #   Topic_A <chr>, Topic_B <chr>, Topic_C <chr>, `More Information` <chr>
```