R Notebook for creation Scotland Vulnerability Resource

Rscript created by
Author: Bernhard Scheliga
Date: 2020-06-11

Data set created by
Author: Bernhard Scheliga

```
## [1] "Dataset version: 0.2"

## [1] "Date: 2020-06-18"

## [1] "R version 4.0.0 (2020-04-24)"
```

1. Summary:

The data set name (Grampian data) is misleading at the moment as it has all of Scotland's SIMD data. Sorry

This script adds additional information to the Scottish Index of Multiple Deprivation indicators (SIMD2020v2) data set. ***Consider adding more details e.g. best on openly accessible information bla and when information access. However, not 100% here would be the right place*** The current version of the script adds the postcodes (PC) of the respective SIMD2020v2 data zones, their data zone names adn the NHS Health board regions.

2. Creating the data set

2.1 Loading source data

```
setwd("~/Scotland_Vulnerability_Resource/Raw_data/")
dir()

## [1] "NHS_Health_Board_regions.csv"  "SIMD2020v2datazones.csv"
## [3] "SIMD2020v2indicators.csv"      "SIMD2020v2indicators_desc.csv"
## [5] "SIMD2020v2postcodes.csv"

df_SIMD2020.indi <- read.csv("SIMD2020v2indicators.csv")
df_SIMD2020.dz <- read.csv("SIMD2020v2datazones.csv")
df_SIMD2020.pc <- read.csv("SIMD2020v2postcodes.csv")
df_NHS_regions <- read.csv("NHS_Health_Board_regions.csv")
```

2.2 Cleaning source data

2.2.1 Removing excess data from source data

```
# df_SIMD.dz we only need column 1,3:13,16,17. See Issues #19 here for
details https://github.com/AbdnCHDS/grampian_data/issues/19
df_SIMD2020.dz <- df_SIMD2020.dz[,c(1,3:13,16,17)]
```

```
# From df_SIMD2020.pc we only need the first two columns "Postcode" & "DZ"
df_SIMD2020.pc <- df_SIMD2020.pc[,c(1,2)]
```

2.2.2 Checking source data for duplicates

```
sapply(df_SIMD2020.indi, function(x) sum(duplicated(x)))
```

```
##              Data_Zone      Intermediate_Zone           Council_area
##                      0                   5726                   6944
##       Total_population Working_age_population            Income_rate
##                   6016                   6201                   6921
##           Income_count        Employment_rate       Employment_count
##                   6605                   6930                   6766
##                    CIF                ALCOHOL                   DRUG
##                   6909                   1109                   2222
##                    SMR                DEPRESS                   LBWT
##                   1577                   6936                   6944
##                  EMERG             Attendance             Attainment
##                   1685                   6920                   6684
##      no_qualifications      not_participating             University
##                   1119                   6941                   6926
##           drive_petrol              drive_GP             drive_post
##                   5931                   5996                   6203
##          drive_primary           drive_retail        drive_secondary
##                   6332                   5682                   5604
##                 PT_GP              PT_post             PT_retail
##                   4972                   5278                   4622
##              Broadband           crime_count             crime_rate
##                   6876                   5636                    927
##      overcrowded_count    nocentralheat_count       overcrowded_rate
##                   6628                   6856                   6922
##     nocentralheat_rate
##                   6955
```

```
# "Data_Zone" should be 0, as they are our primary key here

sapply(df_SIMD2020.dz, function(x) sum(duplicated(x)))
```

```
##                               DZ                 SIMD2020v2_Rank
##                                0                               0
##            SIMD2020v2_Vigintile              SIMD2020v2_Decile
##                             6956                            6966
##             SIMD2020v2_Quintile   SIMD2020v2_Income_Domain_Rank
##                             6971                             471
## SIMD2020_Employment_Domain_Rank  SIMD2020_Education_Domain_Rank
##                             420                               0
##     SIMD2020_Health_Domain_Rank     SIMD2020_Access_Domain_Rank
##                               0                               0
##      SIMD2020_Crime_Domain_Rank    SIMD2020_Housing_Domain_Rank
##                             174                             296
```

```
##                          URclass                           URname
##                             6970                             6970

# "DZ" (Data zone) should be 0, as they are our primary key here

sapply(df_SIMD2020.pc, function(x) sum(duplicated(x)))

## Postcode       DZ
##        0   150858

# "Postcode" should be 0, as they are our primary key here
```

2.2.3 Checking source data for NA-values

```
sapply(df_SIMD2020.indi, function(x) sum(is.na(x)))

##              Data_Zone      Intermediate_Zone           Council_area
##                      0                      0                      0
##       Total_population Working_age_population            Income_rate
##                      0                      0                      0
##           Income_count        Employment_rate       Employment_count
##                      0                      0                      0
##                    CIF               ALCOHOL                   DRUG
##                      0                      0                      0
##                    SMR               DEPRESS                   LBWT
##                      0                      0                      0
##                  EMERG            Attendance             Attainment
##                      0                      0                      0
##      no_qualifications      not_participating             University
##                      0                      0                      0
##           drive_petrol              drive_GP             drive_post
##                      0                      0                      0
##          drive_primary          drive_retail        drive_secondary
##                      0                      0                      0
##                  PT_GP               PT_post              PT_retail
##                      0                      0                      0
##              Broadband           crime_count             crime_rate
##                      0                      0                      0
##      overcrowded_count    nocentralheat_count       overcrowded_rate
##                      0                      0                      0
##     nocentralheat_rate
##                      0

# Data_Zone should be 0. However, currently in the SIM2020v2 source data set
missing values and suppressed values are denoted by "*"

# Lets check how many missing values and suppressed values are denoted by "*"
in the source data
sapply(df_SIMD2020.indi, function(x) sum(x=="*"))

##              Data_Zone      Intermediate_Zone           Council_area
##                      0                      0                      0
```

```
##       Total_population Working_age_population            Income_rate
##                      0                      0                      3
##           Income_count        Employment_rate       Employment_count
##                      0                      3                      0
##                    CIF               ALCOHOL                   DRUG
##                      3                      2                      2
##                    SMR               DEPRESS                   LBWT
##                      2                      1                      1
##                  EMERG            Attendance             Attainment
##                      2                    567                    189
##      no_qualifications      not_participating             University
##                      0                      3                      2
##           drive_petrol              drive_GP             drive_post
##                      0                      0                      0
##          drive_primary          drive_retail        drive_secondary
##                      0                      0                      0
##                 PT_GP               PT_post              PT_retail
##                      0                      0                      0
##              Broadband           crime_count             crime_rate
##                      2                    500                    501
##      overcrowded_count    nocentralheat_count       overcrowded_rate
##                      0                      0                      0
##     nocentralheat_rate
##                      0
```

```
# storing for a quick comparision
df_Star <-sapply(df_SIMD2020.indi, function(x) sum(x=="*"))

# Here we replace the "*" denotation with NA in the SIMD data set
df_SIMD2020.indi[df_SIMD2020.indi=="*"] <- NA
# Comparing, if number of "*" denotes is the same as the number of NA now for
the   respective columns
df_Star == sapply(df_SIMD2020.indi, function(x) sum(is.na(x)))
```

```
##              Data_Zone      Intermediate_Zone           Council_area
##                   TRUE                   TRUE                   TRUE
##       Total_population Working_age_population            Income_rate
##                   TRUE                   TRUE                   TRUE
##           Income_count        Employment_rate       Employment_count
##                   TRUE                   TRUE                   TRUE
##                    CIF               ALCOHOL                   DRUG
##                   TRUE                   TRUE                   TRUE
##                    SMR               DEPRESS                   LBWT
##                   TRUE                   TRUE                   TRUE
##                  EMERG            Attendance             Attainment
##                   TRUE                   TRUE                   TRUE
##      no_qualifications      not_participating             University
##                   TRUE                   TRUE                   TRUE
##           drive_petrol              drive_GP             drive_post
##                   TRUE                   TRUE                   TRUE
```

```
##          drive_primary           drive_retail        drive_secondary
##                   TRUE                   TRUE                   TRUE
##                  PT_GP                PT_post              PT_retail
##                   TRUE                   TRUE                   TRUE
##              Broadband            crime_count             crime_rate
##                   TRUE                   TRUE                   TRUE
##      overcrowded_count    nocentralheat_count       overcrowded_rate
##                   TRUE                   TRUE                   TRUE
##     nocentralheat_rate
##                   TRUE
```

```
sapply(df_SIMD2020.dz, function(x) sum(is.na(x)))
```

```
##                              DZ                SIMD2020v2_Rank
##                               0                              0
##            SIMD2020v2_Vigintile               SIMD2020v2_Decile
##                               0                              0
##             SIMD2020v2_Quintile   SIMD2020v2_Income_Domain_Rank
##                               0                              0
## SIMD2020_Employment_Domain_Rank  SIMD2020_Education_Domain_Rank
##                               0                              0
##     SIMD2020_Health_Domain_Rank     SIMD2020_Access_Domain_Rank
##                               0                              0
##      SIMD2020_Crime_Domain_Rank    SIMD2020_Housing_Domain_Rank
##                               0                              0
##                         URclass                         URname
##                               0                              0
```

```
# DZ (Data zone) should be 0
sapply(df_SIMD2020.pc, function(x) sum(is.na(x)))
```

```
## Postcode       DZ
##        0        0
```

```
# Postcode should be 0
```

Nothing concerning here :)

2.3 Joining the source data set

2.3.1 Merging/Joining data zone names to SIMD2020v2

```
df_SIMD2020.1merge <- merge(df_SIMD2020.indi, df_SIMD2020.dz,
by.x="Data_Zone", by.y="DZ", all = TRUE) #all= TRUE to include potential
missing values. In case something goes wrong with merge().

# Check if we introduced NA values
sapply(df_SIMD2020.1merge, function(x) sum(is.na(x)))
```

```
##                      Data_Zone               Intermediate_Zone
##                               0                              0
##                   Council_area                Total_population
```

```
##                                   0                                   0
##          Working_age_population                         Income_rate
##                                   0                                   3
##                    Income_count                     Employment_rate
##                                   0                                   3
##                Employment_count                                 CIF
##                                   0                                   3
##                         ALCOHOL                                DRUG
##                                   2                                   2
##                             SMR                              DEPRESS
##                                   2                                   1
##                            LBWT                               EMERG
##                                   1                                   2
##                      Attendance                          Attainment
##                                 567                                 189
##               no_qualifications                     not_participating
##                                   0                                   3
##                      University                          drive_petrol
##                                   2                                   0
##                        drive_GP                            drive_post
##                                   0                                   0
##                    drive_primary                          drive_retail
##                                   0                                   0
##                  drive_secondary                                PT_GP
##                                   0                                   0
##                         PT_post                             PT_retail
##                                   0                                   0
##                       Broadband                           crime_count
##                                   2                                 500
##                      crime_rate                     overcrowded_count
##                                 501                                   0
##             nocentralheat_count                      overcrowded_rate
##                                   0                                   0
##              nocentralheat_rate                       SIMD2020v2_Rank
##                                   0                                   0
##            SIMD2020v2_Vigintile                     SIMD2020v2_Decile
##                                   0                                   0
##             SIMD2020v2_Quintile        SIMD2020v2_Income_Domain_Rank
##                                   0                                   0
## SIMD2020_Employment_Domain_Rank       SIMD2020_Education_Domain_Rank
##                                   0                                   0
##     SIMD2020_Health_Domain_Rank          SIMD2020_Access_Domain_Rank
##                                   0                                   0
##      SIMD2020_Crime_Domain_Rank         SIMD2020_Housing_Domain_Rank
##                                   0                                   0
##                         URclass                             URname
##                                   0                                   0
```

2.3.2 Merging/Joining postcodes to SIMD2020v2

```
df_SIMD2020.2merge <- merge(df_SIMD2020.1merge, df_SIMD2020.pc,
by.x="Data_Zone", by.y="DZ", all = TRUE)
# reordering the data frame, placing the postcode column in the first
position
df_SIMD2020.2merge <- df_SIMD2020.2merge[,c(51,1:50)]


# find the column with the NA values
sapply(df_SIMD2020.2merge, function(x) sum(is.na(x)))
```

```
##                          Postcode                         Data_Zone
##                                 2                                 0
##                 Intermediate_Zone                       Council_area
##                                 0                                 0
##                  Total_population           Working_age_population
##                                 0                                 0
##                       Income_rate                     Income_count
##                                 4                                 0
##                    Employment_rate                   Employment_count
##                                 4                                 0
##                               CIF                           ALCOHOL
##                                 4                                 3
##                              DRUG                               SMR
##                                 3                                 3
##                           DEPRESS                              LBWT
##                                 1                                 1
##                             EMERG                         Attendance
##                                 3                             12739
##                        Attainment                 no_qualifications
##                              4722                                 0
##                 not_participating                        University
##                                44                                 3
##                      drive_petrol                          drive_GP
##                                 0                                 0
##                       drive_post                      drive_primary
##                                 0                                 0
##                      drive_retail                    drive_secondary
##                                 0                                 0
##                             PT_GP                           PT_post
##                                 0                                 0
##                        PT_retail                         Broadband
##                                 0                                 2
##                       crime_count                       crime_rate
##                              8976                             8978
##                 overcrowded_count               nocentralheat_count
##                                 0                                 0
##                  overcrowded_rate                nocentralheat_rate
##                                 0                                 0
##                    SIMD2020v2_Rank              SIMD2020v2_Vigintile
```

```
##                                 0                                 0
##               SIMD2020v2_Decile             SIMD2020v2_Quintile
##                                 0                                 0
##   SIMD2020v2_Income_Domain_Rank SIMD2020_Employment_Domain_Rank
##                                 0                                 0
##  SIMD2020_Education_Domain_Rank     SIMD2020_Health_Domain_Rank
##                                 0                                 0
##     SIMD2020_Access_Domain_Rank      SIMD2020_Crime_Domain_Rank
##                                 0                                 0
##    SIMD2020_Housing_Domain_Rank                           URclass
##                                 0                                 0
##                           URname
##                                 0
```

```
# find affected rows
df_SIMD2020.2merge[is.na(df_SIMD2020.2merge$Postcode),]
```

```
##       Postcode Data_Zone Intermediate_Zone Council_area Total_population
## 85669     <NA> S01010206        Petershill Glasgow City                0
## 86063     <NA> S01010226         Sighthill Glasgow City                0
##       Working_age_population Income_rate Income_count Employment_rate
## 85669                      0        <NA>            0            <NA>
## 86063                      0        <NA>            0            <NA>
##       Employment_count  CIF ALCOHOL  DRUG    SMR DEPRESS LBWT EMERG
Attendance
## 85669                0 <NA>    <NA>  <NA>   <NA>    <NA> <NA>  <NA>
<NA>
## 86063                0 <NA>   95.22 57.20 153.32    0.01 0.00 87.37
0.84
##       Attainment no_qualifications not_participating University
drive_petrol
## 85669       <NA>            353.08              <NA>       <NA>
2.64
## 86063       <NA>            202.42              0.00       0.24
2.41
##       drive_GP drive_post drive_primary drive_retail drive_secondary PT_GP
## 85669     4.19       4.17          3.66         5.48            5.22  7.31
## 86063     2.74       2.53          3.00         2.60            2.92  7.93
##       PT_post PT_retail Broadband crime_count crime_rate overcrowded_count
## 85669   12.67     13.54      <NA>        <NA>       <NA>               243
## 86063   10.40      9.51      <NA>        <NA>       <NA>               339
##       nocentralheat_count overcrowded_rate nocentralheat_rate
SIMD2020v2_Rank
## 85669                  21             0.49               0.04
4172
## 86063                  45             0.42               0.06
6058
##       SIMD2020v2_Vigintile SIMD2020v2_Decile SIMD2020v2_Quintile
## 85669                   12                 6                   3
## 86063                   18                 9                   5
```

```
##       SIMD2020v2_Income_Domain_Rank SIMD2020_Employment_Domain_Rank
## 85669                          6969                            6974
## 86063                          6969                            6974
##       SIMD2020_Education_Domain_Rank SIMD2020_Health_Domain_Rank
## 85669                            811                        3436
## 86063                           3517                        4559
##       SIMD2020_Access_Domain_Rank SIMD2020_Crime_Domain_Rank
## 85669                        1682                       6928
## 86063                        4634                       6928
##       SIMD2020_Housing_Domain_Rank URclass            URname
## 85669                           18       1 Large Urban Areas
## 86063                           50       1 Large Urban Areas
```

There are no postcodes for those to data zone. Also they don't have any population. See issue #9 on link.

```
## duplicated values
sapply(df_SIMD2020.2merge, function(x) sum(duplicated(x)))
```

```
##                        Postcode                       Data_Zone
##                               1                          150858
##                Intermediate_Zone                     Council_area
##                          156584                          157802
##                 Total_population           Working_age_population
##                          156874                          157059
##                      Income_rate                    Income_count
##                          157779                          157463
##                   Employment_rate                 Employment_count
##                          157788                          157624
##                             CIF                         ALCOHOL
##                          157767                          151967
##                             DRUG                             SMR
##                          153080                          152435
##                         DEPRESS                             LBWT
##                          157794                          157802
##                           EMERG                      Attendance
##                          152543                          157778
##                      Attainment                 no_qualifications
##                          157542                          151977
##                not_participating                      University
##                          157799                          157784
##                     drive_petrol                       drive_GP
##                          156789                          156854
##                    drive_post                     drive_primary
##                          157061                          157190
##                    drive_retail                   drive_secondary
##                          156540                          156462
##                           PT_GP                         PT_post
##                          155830                          156136
##                       PT_retail                       Broadband
```

```
##                            155480                            157734
##                       crime_count                        crime_rate
##                            156494                            151785
##                 overcrowded_count               nocentralheat_count
##                            157486                            157714
##                  overcrowded_rate                nocentralheat_rate
##                            157780                            157813
##                   SIMD2020v2_Rank             SIMD2020v2_Vigintile
##                            150858                            157814
##                 SIMD2020v2_Decile              SIMD2020v2_Quintile
##                            157824                            157829
##    SIMD2020v2_Income_Domain_Rank SIMD2020_Employment_Domain_Rank
##                            151329                            151278
##   SIMD2020_Education_Domain_Rank     SIMD2020_Health_Domain_Rank
##                            150858                            150858
##      SIMD2020_Access_Domain_Rank      SIMD2020_Crime_Domain_Rank
##                            150858                            151032
##     SIMD2020_Housing_Domain_Rank                           URclass
##                            151154                            157828
##                            URname
##                            157828

# However, there is one duplicated postcode. The Postcodes should be 0
# let find it

df_SIMD2020.2merge[duplicated(df_SIMD2020.2merge$Postcode, fromLast=FALSE),]

##       Postcode Data_Zone Intermediate_Zone Council_area Total_population
## 86063     <NA> S01010226         Sighthill Glasgow City                0
##       Working_age_population Income_rate Income_count Employment_rate
## 86063                      0        <NA>            0            <NA>
##       Employment_count  CIF ALCOHOL  DRUG    SMR DEPRESS LBWT EMERG
Attendance
## 86063                0 <NA>   95.22 57.20 153.32    0.01 0.00 87.37
0.84
##       Attainment no_qualifications not_participating University
drive_petrol
## 86063       <NA>            202.42              0.00       0.24
2.41
##       drive_GP drive_post drive_primary drive_retail drive_secondary PT_GP
## 86063     2.74       2.53             3          2.6            2.92  7.93
##       PT_post PT_retail Broadband crime_count crime_rate overcrowded_count
## 86063    10.4      9.51      <NA>        <NA>       <NA>               339
##       nocentralheat_count overcrowded_rate nocentralheat_rate
SIMD2020v2_Rank
## 86063                  45             0.42               0.06
6058
##       SIMD2020v2_Vigintile SIMD2020v2_Decile SIMD2020v2_Quintile
## 86063                   18                9                   5
##       SIMD2020v2_Income_Domain_Rank SIMD2020_Employment_Domain_Rank
```

```
## 86063                            6969                                6974
##       SIMD2020_Education_Domain_Rank SIMD2020_Health_Domain_Rank
## 86063                           3517                          4559
##       SIMD2020_Access_Domain_Rank SIMD2020_Crime_Domain_Rank
## 86063                        4634                        6928
##       SIMD2020_Housing_Domain_Rank URclass            URname
## 86063                           50       1 Large Urban Areas

# duplicated also picked up on the two NA in the postcodes
```

2.3.2 Merging/Joining NHS Health Board regions to SIMD2020v2

```
df_SIMD2020.3merge <- merge(df_SIMD2020.2merge, df_NHS_regions,
by.x="Council_area", by.y="Council_area", all = TRUE)
# reordering the data frame
df_SIMD2020.3merge <- df_SIMD2020.3merge[,c(2:4,1,52,5:51)]
# find the column with the NA values
sapply(df_SIMD2020.3merge, function(x) sum(is.na(x)))
```

```
##                         Postcode                        Data_Zone
##                                2                                0
##                Intermediate_Zone                     Council_area
##                                0                                0
##          NHS_Health_Board_Region                 Total_population
##                                0                                0
##           Working_age_population                      Income_rate
##                                0                                4
##                     Income_count                   Employment_rate
##                                0                                4
##                Employment_count                              CIF
##                                0                                4
##                          ALCOHOL                             DRUG
##                                3                                3
##                              SMR                          DEPRESS
##                                3                                1
##                             LBWT                            EMERG
##                                1                                3
##                       Attendance                       Attainment
##                            12739                             4722
##               no_qualifications                 not_participating
##                                0                               44
##                       University                     drive_petrol
##                                3                                0
##                         drive_GP                       drive_post
##                                0                                0
##                    drive_primary                     drive_retail
##                                0                                0
##                  drive_secondary                            PT_GP
##                                0                                0
##                         PT_post                          PT_retail
##                                0                                0
```

```
##                         Broadband                       crime_count
##                                 2                              8976
##                        crime_rate                overcrowded_count
##                              8978                                 0
##              nocentralheat_count                 overcrowded_rate
##                                 0                                 0
##               nocentralheat_rate                   SIMD2020v2_Rank
##                                 0                                 0
##             SIMD2020v2_Vigintile                SIMD2020v2_Decile
##                                 0                                 0
##              SIMD2020v2_Quintile   SIMD2020v2_Income_Domain_Rank
##                                 0                                 0
## SIMD2020_Employment_Domain_Rank  SIMD2020_Education_Domain_Rank
##                                 0                                 0
##     SIMD2020_Health_Domain_Rank     SIMD2020_Access_Domain_Rank
##                                 0                                 0
##      SIMD2020_Crime_Domain_Rank    SIMD2020_Housing_Domain_Rank
##                                 0                                 0
##                           URclass                           URname
##                                 0                                 0

# find affected rows
df_SIMD2020.3merge[is.na(df_SIMD2020.3merge$Postcode),]

##       Postcode Data_Zone Intermediate_Zone Council_area
## 84706     <NA> S01010206        Petershill Glasgow City
## 85100     <NA> S01010226         Sighthill Glasgow City
##         NHS_Health_Board_Region Total_population Working_age_population
## 84706 Greater Glasgow and Clyde                0                      0
## 85100 Greater Glasgow and Clyde                0                      0
##       Income_rate Income_count Employment_rate Employment_count  CIF
ALCOHOL
## 84706        <NA>            0            <NA>                0 <NA>
<NA>
## 85100        <NA>            0            <NA>                0 <NA>
95.22
##        DRUG    SMR DEPRESS LBWT EMERG Attendance Attainment
no_qualifications
## 84706  <NA>   <NA>    <NA> <NA>  <NA>       <NA>       <NA>
353.08
## 85100 57.20 153.32    0.01 0.00 87.37       0.84       <NA>
202.42
##       not_participating University drive_petrol drive_GP drive_post
## 84706              <NA>       <NA>         2.64     4.19       4.17
## 85100              0.00       0.24         2.41     2.74       2.53
##       drive_primary drive_retail drive_secondary PT_GP PT_post PT_retail
## 84706          3.66         5.48            5.22  7.31   12.67     13.54
## 85100          3.00         2.60            2.92  7.93   10.40      9.51
##       Broadband crime_count crime_rate overcrowded_count
nocentralheat_count
```

```
## 84706      <NA>        <NA>       <NA>               243
21
## 85100      <NA>        <NA>       <NA>               339
45
##       overcrowded_rate nocentralheat_rate SIMD2020v2_Rank
SIMD2020v2_Vigintile
## 84706             0.49               0.04            4172
12
## 85100             0.42               0.06            6058
18
##       SIMD2020v2_Decile SIMD2020v2_Quintile SIMD2020v2_Income_Domain_Rank
## 84706                 6                   3                          6969
## 85100                 9                   5                          6969
##       SIMD2020_Employment_Domain_Rank SIMD2020_Education_Domain_Rank
## 84706                            6974                            811
## 85100                            6974                           3517
##       SIMD2020_Health_Domain_Rank SIMD2020_Access_Domain_Rank
## 84706                        3436                       1682
## 85100                        4559                       4634
##       SIMD2020_Crime_Domain_Rank SIMD2020_Housing_Domain_Rank URclass
## 84706                       6928                           18       1
## 85100                       6928                           50       1
##                  URname
## 84706 Large Urban Areas
## 85100 Large Urban Areas
```

2.4 Saving the data set

```
setwd("~/Scotland_Vulnerability_Resource/Processed_data/")

write.csv(df_SIMD2020.3merge,
paste("Scotland_Vulnerability_Resource_v",Dataset_version,".csv", sep = ""),
row.names=FALSE)
```