R Notebook for SAPE 2019

Bernhard Scheliga

2020-09-04

Data set created by
Author: Bernhard Scheliga

```
## [1] "Dataset version: 0.1"
```

```
## [1] "Date: 2020-10-29"
```

```
## [1] "R version 4.0.3 (2020-10-10)"
```

1. Summary:

Initially, the plan was to included this data directly in the Scottish Vulnerability Resource (SVR). However, the Small Area Population Estimate (SAPE) 2019 data set for Scotland is a bit too large in itself to be included directly. The idea now is, that it is an additional resource to the SVR and we will provide a script for user to combine the SVR and the SVR_SAPE2019 dataset.

The SVR_SAPE2019 resource is postcode(PC) searchable and enables the user to swiftly retrieve data for datazones based on a postcode search. *!Data is not broken down to postcode level!* The spatial resolution of the resource are datazone. For more details see our in depth ***LINK TO THE RIGHT FILE*** documentation on github.com/AbdnCHDS/Scotland_Vulnerability_Resource

Here we describe how the SVR_SAPE2019 was created in *R*, which code and information was used. This is to enable the interested user to follow our process and get a better understanding of the decisions we during the creation process of the SVR_SAPE2019.

The SVR_SAPE2019 is build around the Small Area Population Estimate (SAPE) 2019 data set for Scotland. The current version of the SVR_SAPE 2019 include Female, Male and Both population estimates, the postcodes (PC) of the respective SIMD2020v2 data zones, their data zone names adn the NHS Health board regions.

Keywords: *R script, SAPE 2019, SIMD2020v2, postcode searchable, datazone names, reproducible, open access*

2. Creating the data set

```r
setwd("~/Scotland_Vulnerability_Resource/SVR-data/")
dir()
```

2.1 Loading source data

```
## [1] "2019_Small_Area_Population_Estimates_FeMale_SVR.csv"
## [2] "Scotland-Vulnerability-Resource_v0.1.csv"
## [3] "Scotland-Vulnerability-Resource_v0.2.csv"
```

```
df_SVR <- read.csv("Scotland-Vulnerability-Resource_v0.2.csv")# we only want the first few columns


## The SAPE 2019 Data
setwd("~/Scotland_Vulnerability_Resource/Input-data/")
dir()
```

```
## [1] "Datazone_areas_sizes.csv"
## [2] "Input-data_documentation"
## [3] "NHS_Health_Board_regions.csv"
## [4] "sape-2019-females_Table 1c Females (2019).csv"
## [5] "sape-2019-males_Table 1b Males (2019).csv"
## [6] "SIMD2020v2datazones.csv"
## [7] "SIMD2020v2indicators.csv"
## [8] "SIMD2020v2indicators_desc.csv"
## [9] "SIMD2020v2postcodes.csv"
```

```
df_SAPE2019.Female <-read.csv("sape-2019-females_Table 1c Females (2019).csv", skip = 3)
df_SAPE2019.Male <- read.csv("sape-2019-males_Table 1b Males (2019).csv", skip = 3)
#df_SAPE2019.Person <-read.csv("sape-2019-persons_Table 1a Persons (2019).csv", skip = 3)

library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

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

2.2 Cleaning source data

```


df_SAPE2019.Female %>% head()
```

2.2.1 Removing excess data from source data

```
##                  X              X.1                X.2       X.3 X.4   AGE0
## 1                          SCOTLAND                    2,800,297  NA 24,634
## 2 DataZone2011Code DataZone2011Name CouncilArea209Name             NA       
## 3        S01006506      Culter - 01      Aberdeen City       456  NA      1
## 4        S01006507      Culter - 02      Aberdeen City       417  NA      1
## 5        S01006508      Culter - 03      Aberdeen City       292  NA      5
## 6        S01006509      Culter - 04      Aberdeen City       281  NA      4
##     AGE1   AGE2   AGE3   AGE4   AGE5   AGE6   AGE7   AGE8   AGE9  AGE10  AGE11
## 1 25,675 26,221 27,272 27,931 28,087 28,829 29,392 30,495 29,216 30,000 29,891
## 2
## 3      2      5      0      2      7      5      5     12     11      6      6
## 4      4      4      4      4      3      3      6      5      3      4     11
## 5      3      3      0      2      1      0      3      1      0      2      1
## 6      1      2      2      3      3      2      7      2      2      6      2
##    AGE12  AGE13  AGE14  AGE15  AGE16  AGE17  AGE18  AGE19  AGE20  AGE21  AGE22
## 1 28,588 28,135 27,573 27,014 26,428 26,256 27,931 30,157 32,426 33,432 35,056
## 2
## 3      1      7      8      5      5      5      1      4      0      2      1
## 4      4      8      0      1      1      2      3      0      2      2      6
## 5      0      3      0      3      1      1      1      1      1      3      4
## 6      5      6      5      5      3      3      2      1      0      0      1
##    AGE23  AGE24  AGE25  AGE26  AGE27  AGE28  AGE29  AGE30  AGE31  AGE32  AGE33
## 1 35,144 35,332 36,292 37,629 39,778 39,398 38,013 37,992 38,188 37,272 36,601
## 2
## 3      6      2      5      6     11      6     12      1      6      6      1
## 4      5      6      4      8      9      7      6      3      3      3      3
## 5      8      5      6     10      7      4     10      7     11      2      3
## 6      1      0      4      1      0      5      5      8      4      2      1
##    AGE34  AGE35  AGE36  AGE37  AGE38  AGE39  AGE40  AGE41  AGE42  AGE43  AGE44
## 1 36,775 35,630 36,112 36,673 36,109 35,374 33,774 31,920 30,822 32,847 33,279
## 2
## 3     15      7      3      9      7      9      9      5      3      4      5
## 4      5      1      4      5      4     11      6      4     10      6      5
## 5      7      5      5      2      3      5      2      4      5      2      6
## 6      2      5      5      3      1      5      9      2      6      4      3
##    AGE45  AGE46  AGE47  AGE48  AGE49  AGE50  AGE51  AGE52  AGE53  AGE54  AGE55
## 1 33,508 35,671 38,220 39,969 39,278 40,785 41,473 41,221 41,241 42,430 42,245
## 2
## 3      7     10     10      4      7      9      5      3      2      8      8
## 4      2      5      1      6      2      3      7      5      5      9      5
## 5      2      5     10      2      5      1      7      4      7      6      2
## 6      5      4      6      3      2      4      2      1      2      2      4
##    AGE56  AGE57  AGE58  AGE59  AGE60  AGE61  AGE62  AGE63  AGE64  AGE65  AGE66
## 1 41,909 40,514 39,564 38,116 37,971 36,854 35,598 34,439 32,979 32,241 31,538
## 2
## 3      8      5      9      5      7      7      6      3      3      5      3
## 4      6     11      4      4     14      4      4      2      7      6      4
## 5      8      7      5      5      6      1      3      1      5      3      5
## 6      5      5      0      3      4      2      4      5      4      0      6
##    AGE67  AGE68  AGE69  AGE70  AGE71  AGE72  AGE73  AGE74  AGE75  AGE76  AGE77
## 1 30,335 30,482 30,388 30,897 31,711 34,215 25,481 24,213 24,270 23,327 21,360
## 2
## 3      5      5      0      8      7      9      2      4      9      4      3
## 4      6     10      9      5     10      9      8      1      6      0      2
## 5      5      1      3      2      2      2      3      1      0      0      3
## 6      8      2      9      3      4      5      0      1      1      2      3
##    AGE78  AGE79  AGE80  AGE81  AGE82  AGE83  AGE84  AGE85  AGE86  AGE87 AGE88
## 1 19,566 19,542 18,728 17,788 16,665 15,555 14,223 12,865 11,405 10,617 9,213
## 2
## 3      0      3      2      1      3      2      0      1      5      2     1
## 4      4      6      4      3      6      0      3      3      2      1     3
## 5      0      0      1      1      0      0      0      1      4      0     0
## 6      3      2      3      3      1      2      1      3      1      2     1
##   AGE89 AGE90. X.5 X.6
## 1 7,847 30,247  NA  NA
## 2              NA  NA
## 3     2      5  NA  NA
## 4     0      1  NA  NA
## 5     1      4  NA  NA
## 6     1      4  NA  NA
```

This is what all SAPE 2019 data set look like. THe first two row and last two rows we do not need. Row 1 has the total poptulation in Scotland for the respective gender and age group. Row 2 has only the header for for the first 3 columns. We will remove column X.1, X.2, X.3, X.4, X.5 & X.6

```
df_SAPE2019.Female <- df_SAPE2019.Female %>% select(-c(X.1, X.2, X.3, X.4, X.5, X.6)) %>% slice(-c(1,2,
df_SAPE2019.Male <- df_SAPE2019.Male %>% select(-c(X.1, X.2, X.3, X.4, X.5, X.6)) %>% slice(-c(1,2,6979
#df_SAPE2019.Person <- df_SAPE2019.Person %>% select(-c(X.1, X.2, X.3, X.4, X.5, X.6)) %>% slice(-c(1,2
```

We need to change two column names, see below

```
colnames(df_SAPE2019.Female)[c(1,92)]
```

```
## [1] "X"      "AGE90."
```

```
colnames(df_SAPE2019.Female)[c(1,92)] <- c("Data_Zone","AGE90PLUS")
colnames(df_SAPE2019.Male)[c(1,92)] <- c("Data_Zone","AGE90PLUS")
#colnames(df_SAPE2019.Person)[c(1,92)] <- c("Data_Zone","AGE90PLUS")

colnames(df_SAPE2019.Female)[c(1,92)]
```

```
## [1] "Data_Zone" "AGE90PLUS"
```

Better.

Adding a gender column This is done to distinglish the data later once it is merged

```
df_SAPE2019.Female$Gender <- "Female"
df_SAPE2019.Male$Gender <- "Male"
#df_SAPE2019.Person$Gender <- "Person"
```

3. Joining the Datasets We actually just need to use bind_rows() the datasets. ***Note!***: We only joined the female & male data since the all three datasets together would be to big for GitHub.

```
df_SAPE2019 <- bind_rows(df_SAPE2019.Female,df_SAPE2019.Male)
```

Now, we have all three SAPE 2019 data set in one object. The next step is to take the *Postcode* -, *Data_zone* -, *Intermediate_Zone* -,*Council_area* - *& NHS_Health_Board_Region* - columns from the SVR dataset and join them to the SAPE 2019 (*df_SAPE2019*).

```
df_SAPE2019.SVRfront <- df_SVR %>% select(Postcode, Data_Zone, Intermediate_Zone,Council_area, NHS_Heal
# quick reordering of the columns
df_SAPE2019.SVRfront <- df_SAPE2019.SVRfront[,c(1:5,97,6:96)]

## Check for NA
sapply(df_SAPE2019.SVRfront, function(x) sum(is.na(x)))
```

```
##                Postcode               Data_Zone       Intermediate_Zone
##                       4                       0                       0
##            Council_area NHS_Health_Board_Region                  Gender
##                       0                       0                       0
##                    AGE0                    AGE1                    AGE2
##                       0                       0                       0
##                    AGE3                    AGE4                    AGE5
##                       0                       0                       0
##                    AGE6                    AGE7                    AGE8
##                       0                       0                       0
##                    AGE9                   AGE10                   AGE11
##                       0                       0                       0
##                   AGE12                   AGE13                   AGE14
##                       0                       0                       0
##                   AGE15                   AGE16                   AGE17
##                       0                       0                       0
##                   AGE18                   AGE19                   AGE20
##                       0                       0                       0
##                   AGE21                   AGE22                   AGE23
##                       0                       0                       0
##                   AGE24                   AGE25                   AGE26
##                       0                       0                       0
##                   AGE27                   AGE28                   AGE29
##                       0                       0                       0
##                   AGE30                   AGE31                   AGE32
##                       0                       0                       0
##                   AGE33                   AGE34                   AGE35
##                       0                       0                       0
##                   AGE36                   AGE37                   AGE38
##                       0                       0                       0
##                   AGE39                   AGE40                   AGE41
##                       0                       0                       0
##                   AGE42                   AGE43                   AGE44
##                       0                       0                       0
##                   AGE45                   AGE46                   AGE47
##                       0                       0                       0
##                   AGE48                   AGE49                   AGE50
##                       0                       0                       0
##                   AGE51                   AGE52                   AGE53
##                       0                       0                       0
##                   AGE54                   AGE55                   AGE56
##                       0                       0                       0
##                   AGE57                   AGE58                   AGE59
##                       0                       0                       0
##                   AGE60                   AGE61                   AGE62
##                       0                       0                       0
##                   AGE63                   AGE64                   AGE65
##                       0                       0                       0
##                   AGE66                   AGE67                   AGE68
##                       0                       0                       0
##                   AGE69                   AGE70                   AGE71
##                       0                       0                       0
##                   AGE72                   AGE73                   AGE74
##                       0                       0                       0
##                   AGE75                   AGE76                   AGE77
##                       0                       0                       0
##                   AGE78                   AGE79                   AGE80
##                       0                       0                       0
##                   AGE81                   AGE82                   AGE83
##                       0                       0                       0
##                   AGE84                   AGE85                   AGE86
##                       0                       0                       0
##                   AGE87                   AGE88                   AGE89
##                       0                       0                       0
##               AGE90PLUS
##                       0
```

6 NA in Postcode. That will be the same NA postcodes as in the SVR data (Petershill & Sighthill) just time three

```
df_SAPE2019.SVRfront[is.na(df_SAPE2019.SVRfront$Postcode),]
```

```
##        Postcode Data_Zone Intermediate_Zone Council_area
## 169411     <NA> S01010206        Petershill Glasgow City
## 169412     <NA> S01010206        Petershill Glasgow City
## 170199     <NA> S01010226         Sighthill Glasgow City
## 170200     <NA> S01010226         Sighthill Glasgow City
##          NHS_Health_Board_Region Gender AGE0 AGE1 AGE2 AGE3 AGE4 AGE5 AGE6 AGE7
## 169411 Greater Glasgow and Clyde Female    0    0    0    0    0    0    0    0
## 169412 Greater Glasgow and Clyde   Male    0    0    0    0    0    0    0    0
## 170199 Greater Glasgow and Clyde Female    0    0    0    0    0    0    0    0
## 170200 Greater Glasgow and Clyde   Male    0    0    0    0    0    0    0    0
##        AGE8 AGE9 AGE10 AGE11 AGE12 AGE13 AGE14 AGE15 AGE16 AGE17 AGE18 AGE19
## 169411    0    0     0     0     0     0     0     0     0     0     0     0
## 169412    0    0     0     0     0     0     0     0     0     0     0     0
## 170199    0    0     0     0     0     0     0     0     0     0     0     0
## 170200    0    0     0     0     0     0     0     0     0     0     0     0
##        AGE20 AGE21 AGE22 AGE23 AGE24 AGE25 AGE26 AGE27 AGE28 AGE29 AGE30 AGE31
## 169411     0     0     0     0     0     0     0     0     0     0     0     0
## 169412     0     0     0     0     0     0     0     0     0     0     0     0
## 170199     0     0     0     0     0     0     0     0     0     0     0     0
## 170200     0     0     0     0     0     0     0     0     0     0     0     0
##        AGE32 AGE33 AGE34 AGE35 AGE36 AGE37 AGE38 AGE39 AGE40 AGE41 AGE42 AGE43
## 169411     0     0     0     0     0     0     0     0     0     0     0     0
## 169412     0     0     0     0     0     0     0     0     0     0     0     0
## 170199     0     0     0     0     0     0     0     0     0     0     0     0
## 170200     0     0     0     0     0     0     0     0     0     0     0     0
##        AGE44 AGE45 AGE46 AGE47 AGE48 AGE49 AGE50 AGE51 AGE52 AGE53 AGE54 AGE55
## 169411     0     0     0     0     0     0     0     0     0     0     0     0
## 169412     0     0     0     0     0     0     0     0     0     0     0     0
## 170199     0     0     0     0     0     0     0     0     0     0     0     0
## 170200     0     0     0     0     0     0     0     0     0     0     0     0
##        AGE56 AGE57 AGE58 AGE59 AGE60 AGE61 AGE62 AGE63 AGE64 AGE65 AGE66 AGE67
## 169411     0     0     0     0     0     0     0     0     0     0     0     0
## 169412     0     0     0     0     0     0     0     0     0     0     0     0
## 170199     0     0     0     0     0     0     0     0     0     0     0     0
## 170200     0     0     0     0     0     0     0     0     0     0     0     0
##        AGE68 AGE69 AGE70 AGE71 AGE72 AGE73 AGE74 AGE75 AGE76 AGE77 AGE78 AGE79
## 169411     0     0     0     0     0     0     0     0     0     0     0     0
## 169412     0     0     0     0     0     0     0     0     0     0     0     0
## 170199     0     0     0     0     0     0     0     0     0     0     0     0
## 170200     0     0     0     0     0     0     0     0     0     0     0     0
##        AGE80 AGE81 AGE82 AGE83 AGE84 AGE85 AGE86 AGE87 AGE88 AGE89 AGE90PLUS
## 169411     0     0     0     0     0     0     0     0     0     0         0
## 169412     0     0     0     0     0     0     0     0     0     0         0
## 170199     0     0     0     0     0     0     0     0     0     0         0
## 170200     0     0     0     0     0     0     0     0     0     0         0
```

We want to check now, if there are only *numeric* value in the *AGE*-columns. By changing the value type from *character* to *numeric* in those columns non-numeric values should turn into NA-values.

```
df_SAPE2019.SVRfront[,c(7:97)] <- sapply(df_SAPE2019.SVRfront[,c(7:97)], as.numeric) # Couldn't find a

df_SAPE2019.SVRfront%>%select(AGE0:AGE90PLUS)%>%sapply(function(x) sum(is.na(x)))
```

```
##      AGE0      AGE1      AGE2      AGE3      AGE4      AGE5      AGE6      AGE7
##         0         0         0         0         0         0         0         0
##      AGE8      AGE9     AGE10     AGE11     AGE12     AGE13     AGE14     AGE15
##         0         0         0         0         0         0         0         0
##     AGE16     AGE17     AGE18     AGE19     AGE20     AGE21     AGE22     AGE23
##         0         0         0         0         0         0         0         0
##     AGE24     AGE25     AGE26     AGE27     AGE28     AGE29     AGE30     AGE31
##         0         0         0         0         0         0         0         0
##     AGE32     AGE33     AGE34     AGE35     AGE36     AGE37     AGE38     AGE39
##         0         0         0         0         0         0         0         0
##     AGE40     AGE41     AGE42     AGE43     AGE44     AGE45     AGE46     AGE47
##         0         0         0         0         0         0         0         0
##     AGE48     AGE49     AGE50     AGE51     AGE52     AGE53     AGE54     AGE55
##         0         0         0         0         0         0         0         0
##     AGE56     AGE57     AGE58     AGE59     AGE60     AGE61     AGE62     AGE63
##         0         0         0         0         0         0         0         0
##     AGE64     AGE65     AGE66     AGE67     AGE68     AGE69     AGE70     AGE71
##         0         0         0         0         0         0         0         0
##     AGE72     AGE73     AGE74     AGE75     AGE76     AGE77     AGE78     AGE79
##         0         0         0         0         0         0         0         0
##     AGE80     AGE81     AGE82     AGE83     AGE84     AGE85     AGE86     AGE87
##         0         0         0         0         0         0         0         0
##     AGE88     AGE89 AGE90PLUS
##         0         0         0
```

No NA-values, perfect!

```
setwd("~/Scotland_Vulnerability_Resource/SVR-data/")

write.csv(df_SAPE2019.SVRfront, paste("2019_Small_Area_Population_Estimates_FeMale_SVR",".csv", sep = "
```

2.4 Saving the data set