2_summarise

Fiona Neilson

23/06/2021

2 SUMMARISE DATA

2.1 Import libraries

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(tidyr)
library(tibble)
library(stringr)
```

2.2 Read in dataframe

```
df = read.csv("out_df.csv", header=TRUE)
```

2.3 Compute weighted counts

2.3.1 Sum of cases

```
sum(df$total)
```

```
## [1] 57825
```

2.3.2 Counts per variable

2.3.2.1 LANP - Language Spoken at Home

This shows the counts of people who speak languages other than English at home, in descending order.

```
df %>% count(LANP, wt = total) %>% arrange(desc(n))
```

```
##                                       LANP     n
## 1                                Vietnamese 10342
## 2                                     Khmer  5275
## 3                                   Punjabi  4071
## 4                                  Mandarin  3690
## 5                                 Cantonese  2730
## 6                                 Sinhalese  2548
## 7                                 Hazaraghi  2114
## 8                                     Tamil  2096
## 9                                     Hindi  1797
## 10                                     Dari  1692
## 11                                    Greek  1496
```

```
## 12                                    Serbian  1375
## 13                                    Italian  1205
## 14                                     Arabic  1062
## 15                                    Burmese   890
## 16                                       Urdu   828
## 17                                    Spanish   802
## 18                                    Bosnian   798
## 19                                    Tagalog   775
## 20                                  Malayalam   678
## 21                                    Turkish   638
## 22                                     French   627
## 23                                   Albanian   625
## 24                                   Filipino   537
## 25                                    Bengali   488
## 26                  Persian (excluding Dari)   476
## 27                                     Polish   471
## 28                                     Telugu   443
## 29                                   Croatian   409
## 30                                   Gujarati   364
## 31                                     Pashto   352
## 32                                     Samoan   352
## 33                                    Min Nan   351
## 34                                 Indonesian   347
## 35                                       Thai   323
## 36                                   Rohingya   303
## 37                                     Korean   272
## 38                                      Hakka   255
## 39                                  Hungarian   240
## 40                                   Romanian   231
## 41                                    Kannada   226
## 42                        Mauritian Creole   212
## 43                                    Russian   212
## 44                     Maori (Cook Island)   188
## 45                                 Macedonian   153
## 46                                      Malay   149
## 47                                     Nepali   145
## 48                                    Maltese   141
## 49                                      Oromo   127
## 50                                     German   123
## 51                                        Lao   120
## 52                                    Marathi   116
## 53                                Chinese nfd   110
## 54                                     Somali   102
## 55                                  Portuguese    81
## 56                                       Dutch    79
## 57                  Southern Asian Languages    78
## 58                         French Creole nfd    77
## 59                                    Swahili    77
## 60  Serbo-Croatian/Yugoslavian so described    65
## 61                                    Japanese    60
```

```
## 62                                  Karen    58
## 63                                Amharic    57
## 64                                   Nuer    54
## 65                                  Shona    42
## 66                                Tigrinya    42
## 67                               Afrikaans    38
## 68                                Armenian    38
## 69                    Maori (New Zealand)    37
## 70                                Konkani    27
## 71                                  Dinka    26
## 72                                   Krio    23
## 73                                 Bisaya    21
## 74                                  Czech    20
## 75                              Ukrainian    20
## 76                                 Tongan    19
## 77                                Tibetan    18
## 78                             Iranic nfd    16
## 79                                  Uygur    16
## 80                                 Yoruba    16
## 81                                Shilluk    15
## 82                  Chaldean Neo-Aramaic    14
## 83                                 Harari    14
## 84                                Slovene    14
## 85                               Bulgarian    11
## 86                     Fijian Hindustani    11
## 87                                  Tetum    11
## 88                                 Acholi    10
## 89                                 Ndebele     9
## 90                                    Tulu     9
## 91                                 Cebuano     8
## 92                                    Igbo     8
## 93                           Indo-Aryan nfd     8
## 94                    African Languages nec     6
## 95                                    Akan     6
## 96                              Creole nfd     6
## 97                                  Fijian     6
## 98                                 Kurdish     6
## 99               Tok Pisin (Neomelanesian)     6
## 100                               Hebrew     5
## 101                               IIokano     5
## 102                              Timorese     5
## 103                              Finnish     4
## 104                   Kinyarwanda (Rwanda)     4
## 105                        Kirundi (Rundi)     4
## 106                                 Oriya     4
## 107                                 Slovak     4
## 108                   African Languages nfd     3
## 109                               Chin Haka     3
## 110                                   Hausa     3
```

```
## 111                                        Mon     3
## 112                                  Pidgin nfd     3
```

2.3.2.2 HEAP - Level of Highest Educational Attainment

```
df %>% count(HEAP, wt = total) %>% arrange(match(HEAP, c("Higher", "Secondary
", "Non-Secondary")))
```

```
##            HEAP     n
## 1        Higher 18020
## 2     Secondary 25022
## 3 Non_Secondary 14783
```

2.3.2.3 EETP - Engagement in Employment, Education and Training

```
df %>% count(EETP, wt = total) %>% arrange(match(EETP, c("Fully", "Partial",
"Not_Engaged")))
```

```
##          EETP     n
## 1       Fully 27093
## 2     Partial  9128
## 3 Not_Engaged 21604
```

2.3.2.4 NEDD - Dwelling Internet Connection

```
df %>% count(NEDD, wt = total)
```

```
##          NEDD     n
## 1    Internet 52415
## 2 No_Internet  5410
```

2.3.2.5 ENGP - Proficiency in Spoken English

```
df %>% count(ENGP, wt = total)
```

```
##         ENGP     n
## 1   Adequate 41182
## 2 Inadequate 16643
```

2.3.2.6 BPLP - Country of Birth

```
df %>% count(BPLP, wt = total) %>% arrange(desc(n)) %>% head(10)
```

```
##                                BPLP     n
## 1                           Vietnam 11562
## 2                             India  8126
## 3                          Cambodia  5690
## 4                         Sri Lanka  3991
## 5  China (excludes SARs and Taiwan)  3789
## 6                       Afghanistan  3618
## 7                       Philippines  1388
## 8                            Greece  1379
## 9                          Pakistan  1282
## 10          Bosnia and Herzegovina  1229
```

2.3.2.7 YARRP - Year of Arrival in Australia

```
df %>% count(YARRP, wt = total) %>% arrange(desc(n))
```

```
##      YARRP     n
## 1 Pre_2006 32045
## 2  2006-15 24666
## 3     2016  1114
```

2.3.2.8 SEXP - Sex

```
df %>% count(SEXP, wt = total)
```

```
##     SEXP     n
## 1 Female 28666
## 2   Male 29159
```

2.4 Exploration

2.4.1 Combine variables and summarise

2.4.2.1 For example, group Engagement in Employment, Education and Training with Internet Access. Show counts.

```
df %>% select(EETP, NEDD, total) %>% head(10)
```

```
##           EETP        NEDD total
## 1      Partial    Internet     5
## 2      Partial    Internet     6
## 3      Partial    Internet    70
## 4      Partial    Internet    33
## 5  Not_Engaged    Internet    12
## 6  Not_Engaged    Internet    18
## 7  Not_Engaged    Internet   111
## 8  Not_Engaged    Internet    21
## 9  Not_Engaged No_Internet     9
## 10       Fully    Internet    19
```

2.4.2.2 Another example: group by Language Spoken at Home and Engagement in Employment, Education and Training. Show counts.

```
# grouping and summarising - arranged by population
df %>% group_by(LANP, EETP) %>% summarise(total_pop = sum(total)) %>% arrange
(desc(total_pop))
```

```
## `summarise()` has grouped output by 'LANP'. You can override using the `.g
roups` argument.
```

```
## # A tibble: 268 x 3
## # Groups:   LANP [112]
##    LANP       EETP        total_pop
##    <chr>      <chr>           <int>
##  1 Vietnamese Fully            4874
##  2 Vietnamese Not_Engaged      3765
##  3 Punjabi    Fully            2566
```

```
##  4 Khmer      Fully            2404
##  5 Khmer      Not_Engaged      2120
##  6 Mandarin   Fully            1773
##  7 Vietnamese Partial          1703
##  8 Sinhalese  Fully            1466
##  9 Mandarin   Not_Engaged      1249
## 10 Hazaraghi  Fully            1231
## # … with 258 more rows
```

2.4.2.3 Another example: filter by language group and explore English proficiency. Show as percentage.

```
# English proficiency by gender for a language group (Urdu)
df %>% filter(LANP == "Urdu") %>% group_by(ENGP, SEXP) %>% summarise(Total =
sum(total)) %>% mutate(Grand_Total = sum(Total)) %>% mutate(Perc = Total / Gr
and_Total * 100)
```

```
## # A tibble: 4 x 5
## # Groups:   ENGP [2]
##   ENGP       SEXP   Total Grand_Total  Perc
##   <chr>      <chr>  <int>       <int> <dbl>
## 1 Adequate   Female   311         803  38.7
## 2 Adequate   Male     492         803  61.3
## 3 Inadequate Female    22          25  88
## 4 Inadequate Male       3          25  12
```

2.4.2.4 Another example: explore the relationship between Proficiency in Spoken English and those not engaged in Engagement in Employment, Education or Training. Show counts.

```
# what relationship does English (ENGP) have with the portion of the Non_Enga
ged (EETP)?
# 1 - select these groups
EETP_group <- df %>%
  group_by(
    EETP, ENGP
      )
# 2 - summarise all levels
EETP_group %>%
  summarise(total = sum(total))
```

```
## # A tibble: 6 x 3
## # Groups:   EETP [3]
##   EETP        ENGP       total
##   <chr>       <chr>      <int>
## 1 Fully       Adequate   22858
## 2 Fully       Inadequate  4235
## 3 Not_Engaged Adequate   11305
## 4 Not_Engaged Inadequate 10299
## 5 Partial     Adequate    7019
## 6 Partial     Inadequate  2109
```

```
# 3 filter to one level
NotEngaged <- EETP_group %>%
  filter(
    EETP == "Not_Engaged"
  ) %>%
  summarise(total = sum(total))
NotEngaged
```

```
## # A tibble: 2 x 3
## # Groups:   EETP [1]
##   EETP        ENGP       total
##   <chr>       <chr>      <int>
## 1 Not_Engaged Adequate   11305
## 2 Not_Engaged Inadequate 10299
```

2.4.2.5 Another example: explore the relationship between Proficiency in Spoken English and Language Spoken at Home

```
# what relationship does English Proficiency have with language spoken at hom
e (LANP)?
# 1 - select these groups
English <- df %>%
  group_by(
    LANP, ENGP
      ) %>%
  summarise(
    total = sum(total)
    )
```

```
## `summarise()` has grouped output by 'LANP'. You can override using the `.g
roups` argument.
```

```
English
```

```
## # A tibble: 176 x 3
## # Groups:   LANP [112]
##    LANP                  ENGP       total
##    <chr>                 <chr>      <int>
##  1 Acholi                Adequate      10
##  2 African Languages nec Adequate       6
##  3 African Languages nfd Adequate       3
##  4 Afrikaans             Adequate      38
##  5 Akan                  Adequate       6
##  6 Albanian              Adequate     430
##  7 Albanian              Inadequate   195
##  8 Amharic               Adequate      57
##  9 Arabic                Adequate     903
## 10 Arabic                Inadequate   159
## # … with 166 more rows
```

2.5 Export file

```
write.csv(df, file="out_2_df.csv")
```