STATS 205: Final Project Write-Up

Brian Liu

6/14/2019

1. Background of the data and why it is interesting or important

The data we are using is the data from WHO suicide statistics from Kaggle. This gives population-based statistics on suicide rate...

2. Explanation of the method studied and its properties

3. Data analysis or simulation study

We will use the crude rate of suicide per 100,000 people.

This analysis provides information on age-standardized rates...

```
who_suicide_statistics_df <- read.csv("who_suicide_statistics.csv")
head(who_suicide_statistics_df)
```

```
##   country year    sex         age suicides_no population
## 1 Albania 1985 female 15-24 years          NA     277900
## 2 Albania 1985 female 25-34 years          NA     246800
## 3 Albania 1985 female 35-54 years          NA     267500
## 4 Albania 1985 female  5-14 years          NA     298300
## 5 Albania 1985 female 55-74 years          NA     138700
## 6 Albania 1985 female   75+ years          NA      34200
```

```
colnames(who_suicide_statistics_df)
```

```
## [1] "country"     "year"        "sex"         "age"         "suicides_no"
## [6] "population"
```

Filter and save countries with missing suicide rate.

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

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

```
## v ggplot2 3.1.1     v purrr   0.3.2
## v tibble  2.1.1     v dplyr   0.8.1
## 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()
```

```
filtered_suicide_df <- drop_na(who_suicide_statistics_df, "suicides_no")
head(filtered_suicide_df)
```

```
##    country year    sex         age suicides_no population
## 25 Albania 1987 female 15-24 years          14     289700
## 26 Albania 1987 female 25-34 years           4     257200
## 27 Albania 1987 female 35-54 years           6     278800
## 28 Albania 1987 female  5-14 years           0     311000
## 29 Albania 1987 female 55-74 years           0     144600
## 30 Albania 1987 female   75+ years           1      35600
```

After filtering countries with missing suicide rate, take a random sample of 100 countries and make sure each continent has approximately equal countries.

Filter countries by continent:

```
library(countrycode)
filtered_suicide_df$continent <- countrycode(sourcevar = filtered_suicide_df[, "country"],
                            origin = "country.name",
                            destination = "continent")
```

```
## Warning in countrycode(sourcevar = filtered_suicide_df[, "country"], origin = "country.name", : Some
```

```
## Warning in countrycode(sourcevar = filtered_suicide_df[, "country"], origin = "country.name", : Some
```

```
head(filtered_suicide_df)
```

```
##    country year    sex         age suicides_no population continent
## 25 Albania 1987 female 15-24 years          14     289700    Europe
## 26 Albania 1987 female 25-34 years           4     257200    Europe
## 27 Albania 1987 female 35-54 years           6     278800    Europe
## 28 Albania 1987 female  5-14 years           0     311000    Europe
## 29 Albania 1987 female 55-74 years           0     144600    Europe
## 30 Albania 1987 female   75+ years           1      35600    Europe
```

```
write.csv(filtered_suicide_df, 'filtered_suicide.csv')
```

Let us find out which continents are counted:

```
# Get list of continents
list_of_continents <- unique(filtered_suicide_df$continent); list_of_continents
```

```
## [1] "Europe"   "Americas" "Asia"     "Oceania"  "Africa"   NA
```

Therefore,

$$\frac{100 \text{ countries}}{6 \text{ continents}} \approx 16 \text{ to } 17 \text{ countries per continent}$$

we should randomly sample 17 countries from each continent.

Notably, there are countries that are not on any of the listed continents. Let us see which ones those are:

```
not_in_a_continent = filtered_suicide_df[is.na(filtered_suicide_df$continent),]
write.csv(not_in_a_continent, 'not_in_a_continent.csv')
head(not_in_a_continent)
```

```
##         country year    sex         age suicides_no population continent
## 32317 Rodrigues 2001 female 15-24 years           0         NA      <NA>
## 32318 Rodrigues 2001 female 25-34 years           0         NA      <NA>
```

```
## 32319 Rodrigues 2001 female 35-54 years           0         NA      <NA>
## 32320 Rodrigues 2001 female  5-14 years           0         NA      <NA>
## 32321 Rodrigues 2001 female 55-74 years           0         NA      <NA>
## 32322 Rodrigues 2001 female   75+ years           0         NA      <NA>
```

```
unique(not_in_a_continent$country)
```

```
## [1] Rodrigues            Virgin Islands (USA)
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
```

Let us make the choice not to include these countries in the analysis, since there are only two countries.

```
# Take off `NA` from list of continents
list_of_continents <- list_of_continents[-length(list_of_continents)]
list_of_continents
```

```
## [1] "Europe"   "Americas" "Asia"     "Oceania"  "Africa"
```

We will now create six dataframes, filtered by list of countries for each continent.

```
# library(rlist)
countries_per_continent <- list()

for (i in seq_along(list_of_continents))
{
    countries_per_continent[[i]] <- filtered_suicide_df[filtered_suicide_df$continent == list_of_contin
}

length(countries_per_continent)
```

```
## [1] 5
```

```
length(countries_per_continent)
```

```
## [1] 5
```

```
for (i in seq_along(countries_per_continent))
{
    print(head(countries_per_continent[[i]]))
    print(length(countries_per_continent[[i]]))
    cat("\n")
}
```

```
##    country year    sex         age suicides_no population continent
## 25 Albania 1987 female 15-24 years          14     289700    Europe
## 26 Albania 1987 female 25-34 years           4     257200    Europe
## 27 Albania 1987 female 35-54 years           6     278800    Europe
## 28 Albania 1987 female  5-14 years           0     311000    Europe
## 29 Albania 1987 female 55-74 years           0     144600    Europe
## 30 Albania 1987 female   75+ years           1      35600    Europe
## [1] 7
##
##      country year    sex         age suicides_no population continent
## 373 Anguilla 1983 female 15-24 years           0         NA  Americas
## 374 Anguilla 1983 female 25-34 years           0         NA  Americas
## 375 Anguilla 1983 female 35-54 years           0         NA  Americas
## 376 Anguilla 1983 female  5-14 years           0         NA  Americas
## 377 Anguilla 1983 female 55-74 years           0         NA  Americas
## 378 Anguilla 1983 female   75+ years           0         NA  Americas
```

```
## [1] 7
##
##      country year    sex         age suicides_no population continent
## 1501 Armenia 1981 female 15-24 years           5     348000      Asia
## 1502 Armenia 1981 female 25-34 years           6     242200      Asia
## 1503 Armenia 1981 female 35-54 years           6     333500      Asia
## 1504 Armenia 1981 female  5-14 years           0     295200      Asia
## 1505 Armenia 1981 female 55-74 years          10     164300      Asia
## 1506 Armenia 1981 female   75+ years           7      43100      Asia
## [1] 7
##
##        country year    sex         age suicides_no population continent
## 2161 Australia 1979 female 15-24 years          71    1236800   Oceania
## 2162 Australia 1979 female 25-34 years          86    1138500   Oceania
## 2163 Australia 1979 female 35-54 years         171    1572100   Oceania
## 2164 Australia 1979 female  5-14 years           1    1246500   Oceania
## 2165 Australia 1979 female 55-74 years         135    1137800   Oceania
## 2166 Australia 1979 female   75+ years          15     309900   Oceania
## [1] 7
##
##         country year    sex         age suicides_no population continent
## 7669 Cabo Verde 2011 female 15-24 years           1      56039    Africa
## 7670 Cabo Verde 2011 female 25-34 years           0      38528    Africa
## 7671 Cabo Verde 2011 female 35-54 years           2      49078    Africa
## 7672 Cabo Verde 2011 female  5-14 years           0      56558    Africa
## 7673 Cabo Verde 2011 female 55-74 years           2      19887    Africa
## 7674 Cabo Verde 2011 female   75+ years           0       7582    Africa
## [1] 7
```

This text links to very important information about why a `for` loop doesn't print anything.[1]

Link to Pandoc Markdown formatting

Randomly sample 17 countries from each continent:

```
list_of_continents
```

```
## [1] "Europe"   "Americas" "Asia"     "Oceania"  "Africa"
```

```
for (i in seq_along(countries_per_continent))
{
    print(list_of_continents[i])
    countries <- unique(countries_per_continent[[i]]$country)
    print(countries)
    print(length(countries))
    cat("\n")
}
```

```
## [1] "Europe"
##  [1] Albania                Austria                Belarus
##  [4] Belgium                Bosnia and Herzegovina Bulgaria
##  [7] Croatia                Czech Republic         Denmark
## [10] Estonia                Finland                France
```

[1] Basically, `for` loops are functions themselves. R prints out the result of a command automatically, but functions are not inherently a command, and will therefore not print anything unless explicitly run. The solution is to have `print(command())` to get output. You will never again spend hours trying to find out why a `for` loop doesn't print anything because you're no longer an R newbie.

```
## [13] Germany                Greece                 Hungary
## [16] Iceland                Ireland                Italy
## [19] Latvia                 Lithuania              Luxembourg
## [22] Malta                  Monaco                 Montenegro
## [25] Netherlands            Norway                 Poland
## [28] Portugal               Republic of Moldova    <NA>
## [31] Romania                Russian Federation     San Marino
## [34] Serbia                 Slovakia               Slovenia
## [37] Spain                  Sweden                 Switzerland
## [40] TFYR Macedonia         Ukraine                United Kingdom
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 42
##
## [1] "Americas"
##  [1] Anguilla                          Antigua and Barbuda
##  [3] Argentina                         Aruba
##  [5] Bahamas                           Barbados
##  [7] Belize                            Bermuda
##  [9] Bolivia                           Brazil
## [11] British Virgin Islands            Canada
## [13] Cayman Islands                    Chile
## [15] Colombia                          Costa Rica
## [17] Cuba                              Dominica
## [19] Dominican Republic                Ecuador
## [21] El Salvador                       Falkland Islands (Malvinas)
## [23] French Guiana                     Grenada
## [25] Guadeloupe                        Guatemala
## [27] Guyana                            Haiti
## [29] Honduras                          Jamaica
## [31] Martinique                        Mexico
## [33] Montserrat                        Netherlands Antilles
## [35] Nicaragua                         Panama
## [37] Paraguay                          Peru
## [39] Puerto Rico                       <NA>
## [41] Saint Kitts and Nevis             Saint Lucia
## [43] Saint Pierre and Miquelon         Saint Vincent and Grenadines
## [45] Suriname                          Trinidad and Tobago
## [47] Turks and Caicos Islands          United States of America
## [49] Uruguay                           Venezuela (Bolivarian Republic of)
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 50
##
## [1] "Asia"
##  [1] Armenia                     Azerbaijan
##  [3] Bahrain                     Brunei Darussalam
##  [5] Cyprus                      Georgia
##  [7] Hong Kong SAR               Iran (Islamic Rep of)
##  [9] Iraq                        Israel
## [11] Japan                       Jordan
## [13] Kazakhstan                  Kuwait
## [15] Kyrgyzstan                  Macau
## [17] Malaysia                    Maldives
## [19] Mongolia                    Occupied Palestinian Territory
## [21] Oman                        Philippines
```

```
## [23] Qatar                          Republic of Korea
## [25] <NA>                           Saudi Arabia
## [27] Singapore                      Sri Lanka
## [29] Syrian Arab Republic           Tajikistan
## [31] Thailand                       Turkey
## [33] Turkmenistan                   United Arab Emirates
## [35] Uzbekistan
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 35
##
## [1] "Oceania"
## [1] Australia   Fiji        Kiribati    New Zealand <NA>
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 5
##
## [1] "Africa"
##  [1] Cabo Verde            Egypt                 Mauritius
##  [4] Mayotte               Morocco               Reunion
##  [7] <NA>                  Sao Tome and Principe Seychelles
## [10] South Africa          Tunisia               Zimbabwe
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 12
```

Since there are only 5 countries in Oceania and 12 countries in Africa, we will use all 5 countries of Oceania and all 12 countries of Africa.

```
samples_of_countries <- list()
num_samples <- 17
for (i in seq_along(countries_per_continent))
{
    countries <- unique(countries_per_continent[[i]]$country)
    current_sample <- list()
    if (length(countries) >= num_samples)
    {
        current_sample <- sample(countries, 17)
    } else {
        current_sample <- sample(countries, length(countries))
    }
    samples_of_countries[[i]] <- current_sample
}
```

Let's see the countries that we will be sampling:

```
for (i in seq_along(samples_of_countries))
{
    print(list_of_continents[i])
    print(samples_of_countries[[i]])
    print(length(samples_of_countries[[i]]))
    cat("\n")
}
```

```
## [1] "Europe"
##  [1] Belgium               Montenegro            Slovakia
##  [4] Sweden                Malta                 Monaco
##  [7] TFYR Macedonia        Bosnia and Herzegovina Russian Federation
## [10] Switzerland           Portugal              Luxembourg
```

```
## [13] Norway                 Spain                  Croatia
## [16] Finland                Greece
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 17
##
## [1] "Americas"
##  [1] Cuba                    Guadeloupe
##  [3] El Salvador             French Guiana
##  [5] Costa Rica              Aruba
##  [7] Ecuador                 Turks and Caicos Islands
##  [9] Jamaica                 Nicaragua
## [11] Bolivia                 Guyana
## [13] Honduras                Mexico
## [15] Bahamas                 Peru
## [17] Uruguay
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 17
##
## [1] "Asia"
##  [1] Kyrgyzstan          Jordan              Azerbaijan
##  [4] Qatar               Hong Kong SAR       Thailand
##  [7] Israel              Tajikistan          Brunei Darussalam
## [10] Oman                Republic of Korea   Mongolia
## [13] Turkey              Macau               Syrian Arab Republic
## [16] Philippines         Kuwait
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 17
##
## [1] "Oceania"
## [1] <NA>        Fiji        New Zealand Kiribati    Australia
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 5
##
## [1] "Africa"
##  [1] South Africa          <NA>                  Reunion
##  [4] Sao Tome and Principe Seychelles            Mauritius
##  [7] Zimbabwe              Morocco               Tunisia
## [10] Mayotte               Egypt                 Cabo Verde
## 141 Levels: Albania Anguilla Antigua and Barbuda Argentina ... Zimbabwe
## [1] 12
```

Let's filter the original dataframe only to include countries that we have sampled:

```
# Make singular list of countries
countries_to_test <- list()
library(rlist)
for (i in seq_along(samples_of_countries))
{
    current_samples <- samples_of_countries[[i]]
    list.append(countries_to_test, current_samples)
}

length(countries_to_test)
```

```
## [1] 0
```

```
for (i in seq_along(countries_to_test))
{
    print(countries_to_test[[i]])
}
```

4. Interpretation of the results or discussion