

# 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[-length(list_of_continents)]
```

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

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

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_continents[i]]
}

length(countries_per_continent)
```

```
## [1] 6
```

```
length(countries_per_continent)
```

```
## [1] 6
```

```
for (i in seq_along(countries_per_continent))
{
  print(head(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
##
##   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
```

```
##
##      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
##
##      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
##
##      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
##
##      country year    sex    age suicides_no population continent
## NA      <NA>    NA <NA> <NA>          NA          NA      <NA>
## NA.1    <NA>    NA <NA> <NA>          NA          NA      <NA>
## NA.2    <NA>    NA <NA> <NA>          NA          NA      <NA>
## NA.3    <NA>    NA <NA> <NA>          NA          NA      <NA>
## NA.4    <NA>    NA <NA> <NA>          NA          NA      <NA>
## NA.5    <NA>    NA <NA> <NA>          NA          NA      <NA>
```

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

Basically, `for` loops are functions themselves. Since R prints out the result of a command automatically, but functions are not inherently a command (unless explicitly run), you need to have `print(command())` in order to get output.

## 4. Interpretation of the results or discussion