

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")
nrow(who_suicide_statistics_df)
```

```
## [1] 43776
```

```
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")
nrow(filtered_suicide_df)
```

```
## [1] 41520
```

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

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

$$\frac{100 \text{ countries}}{7 \text{ continents}} \approx 14 \text{ to } 15 \text{ countries per continent}$$

Filter countries by continent:

```
install.packages("countrycode", dependencies=TRUE, repos='http://cran.us.r-project.org')
```

```
## Installing package into '/home/bliutwo/R/x86_64-pc-linux-gnu-library/3.6'  
## (as 'lib' is unspecified)
```

```
library(countrycode)  
df <- data.frame(country = c("Afghanistan",  
                             "Algeria",  
                             "USA",  
                             "France",  
                             "New Zealand",  
                             "Fantasyland"))  
df$continent <- countrycode(sourcevar = df[, "country"],  
                             origin = "country.name",  
                             destination = "continent")
```

```
## Warning in countrycode(sourcevar = df[, "country"], origin = "country.name", : Some values were not mapped  
# Get seven dataframes, filtered by list of countries for each continent.
```

4. Interpretation of the results or discussion