Data Wrangling for Life Expectancy

Patrick Kelly

2/14/2019

Life Expectancy at Birth: World Bank Data

Add this to the TB Incidence data

Click Here"Life Exspectancy"

```
suppressMessages(library(data.table))
suppressMessages(library(dplyr))
```

Load data for years 2000-2016

```
data <- fread("Life_Exp.csv", skip = 4,
        select=c(1,2,45:61),header = FALSE)
data<- data[2:265,]
colnames(data) <- c("Country","alpha_3","2000","2001",
  "2002","2003","2004","2005","2006","2007",
  "2008","2009","2010","2011","2012","2013",
  "2014","2015","2016")
cc<- complete.cases(data)
data <- data[cc]
data = melt(data,
      measure.vars = c("2000", "2001", "2002",
      "2003","2004","2005","2006","2007","2008",
      "2009","2010","2011","2012","2013","2014",
      "2015","2016"),
      variable.name = "year", value.name = "Life_Exp")
data <- arrange(data,Country)
data$Life_Exp <- round(data$Life_Exp,1)
rm(cc)
```

Load the TB Incidence data

```
tb_regions <- fread("tb_regions.csv")
all_tb <- data.frame(alpha_3=unique(tb_regions$alpha_3))
all_tb$alpha_3 <- as.character(all_tb$alpha_3)
all_tb <- all_tb[-c(44), ]
all_tb <- data.frame(alpha_3=all_tb)
all_tb$alpha_3 <- as.character(all_tb$alpha_3)
```

```
data2 <- left_join(all_tb,data, by = "alpha_3")
data2 <- data2[complete.cases(data2), ]
data2 <- select(data2,year,alpha_3,Life_Exp)
class(data2$year)
```

```
## [1] "factor"
```

```
data2$year <- as.numeric(as.character(data2$year))
tb_Life_exp <- left_join(data2,tb_regions, by = c("alpha_3","year"))
```

```
tb_Life_exp  <- tb_Life_exp[complete.cases(tb_Life_exp), ]
anyNA(tb_Life_exp)
```

```
## [1] FALSE
```

```
tb_Life_exp <- select(tb_Life_exp,Country,alpha_3,alpha_2,year,
    population,incidence,Life_Exp,sub_region,region_code)
dim(tb_Life_exp)
```

```
## [1] 3190    9
```

```
head(tb_Life_exp)
```

```
##       Country alpha_3 alpha_2 year population incidence Life_Exp
## 1 Afghanistan     AFG      AF 2000      20.09     189.1     55.5
## 2 Afghanistan     AFG      AF 2001      20.97     190.7     56.0
## 3 Afghanistan     AFG      AF 2002      21.98     191.1     56.6
## 4 Afghanistan     AFG      AF 2003      23.06     190.8     57.2
## 5 Afghanistan     AFG      AF 2004      24.12     190.7     57.9
## 6 Afghanistan     AFG      AF 2005      25.07     187.5     58.5
##   sub_region   region_code
## 1       Asia Southern Asia
## 2       Asia Southern Asia
## 3       Asia Southern Asia
## 4       Asia Southern Asia
## 5       Asia Southern Asia
## 6       Asia Southern Asia
```

```
tail(tb_Life_exp)
```

```
##       Country alpha_3 alpha_2 year population incidence Life_Exp
## 3225 Zimbabwe     ZWE      ZW 2011      14.39     382.2     54.8
## 3226 Zimbabwe     ZWE      ZW 2012      14.71     353.5     56.5
## 3227 Zimbabwe     ZWE      ZW 2013      15.05     305.6     58.1
## 3228 Zimbabwe     ZWE      ZW 2014      15.41     279.0     59.4
## 3229 Zimbabwe     ZWE      ZW 2015      15.78     240.8     60.4
## 3230 Zimbabwe     ZWE      ZW 2016      16.15     210.5     61.2
##      sub_region    region_code
## 3225     Africa Eastern Africa
## 3226     Africa Eastern Africa
## 3227     Africa Eastern Africa
## 3228     Africa Eastern Africa
## 3229     Africa Eastern Africa
## 3230     Africa Eastern Africa
```

```
# write.csv(tb_Life_exp,"tb_Life_exp.csv", row.names = FALSE)
```