ADS_group

WXL

2024-04-11

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
substance_use <- read.csv("/Users/wangxiaolei/Desktop/ADS Group ICA/substance_use.csv")
anyNA(substance_use)
```

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

```
head(is.na(substance_use))
```

```
##      measure location   sex   age cause metric  year   val upper lower
## [1,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [2,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [3,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [4,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [5,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [6,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
```

```
tail(substance_use)
```

```
##          measure                   location    sex      age
## 15115 Prevalence Europe & Central Asia - WB   Male 55 to 59
## 15116 Prevalence Europe & Central Asia - WB Female 55 to 59
## 15117 Prevalence Europe & Central Asia - WB   Male 60 to 64
## 15118 Prevalence Europe & Central Asia - WB Female 60 to 64
## 15119 Prevalence Europe & Central Asia - WB   Male 65 to 69
## 15120 Prevalence Europe & Central Asia - WB Female 65 to 69
```

```
##                       cause  metric year         val       upper        lower
## 15115 Opioid use disorders Percent 2019 0.002193432 0.002957747 0.0015733604
## 15116 Opioid use disorders Percent 2019 0.001373674 0.001884760 0.0009665065
## 15117 Opioid use disorders Percent 2019 0.001753564 0.002433040 0.0011996883
## 15118 Opioid use disorders Percent 2019 0.001179834 0.001708881 0.0007723861
## 15119 Opioid use disorders Percent 2019 0.001327149 0.001827547 0.0009152947
## 15120 Opioid use disorders Percent 2019 0.001099951 0.001512615 0.0007387951
```

```
tail(is.na(substance_use))
```

```
##          measure location   sex   age cause metric  year   val upper lower
## [15115,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [15116,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [15117,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [15118,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [15119,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
## [15120,]   FALSE    FALSE FALSE FALSE FALSE  FALSE FALSE FALSE FALSE FALSE
```

```
apply(is.na(substance_use), 2, which)
```

```
## integer(0)
```

```
dim(substance_use)
```

```
## [1] 15120    10
```

```
data.noNA = substance_use[complete.cases(substance_use), ]
dim(data.noNA)
```

```
## [1] 15120    10
```

```
frw.idx = which(duplicated(data.noNA))
rvs.idx = which(duplicated(data.noNA, fromLast = TRUE))
data.noNA[c(frw.idx, rvs.idx), ]
```

```
##  [1] measure  location sex      age      cause    metric   year     val
##  [9] upper    lower
## <0 rows> (or 0-length row.names)
```

```
dim(data.noNA)
```

```
## [1] 15120    10
```

```
data.noNA.noDup = data.noNA[!duplicated(data.noNA),]
dim(data.noNA.noDup)
```

```
## [1] 15120    10
```

```
head(data.noNA.noDup)
```

```
##   measure                 location    sex      age                 cause
## 1  Deaths East Asia & Pacific - WB   Male 25 to 29 Alcohol use disorders
## 2  Deaths East Asia & Pacific - WB Female 25 to 29 Alcohol use disorders
## 3  Deaths East Asia & Pacific - WB   Male 30 to 34 Alcohol use disorders
## 4  Deaths East Asia & Pacific - WB Female 30 to 34 Alcohol use disorders
## 5  Deaths East Asia & Pacific - WB   Male 35 to 39 Alcohol use disorders
## 6  Deaths East Asia & Pacific - WB Female 35 to 39 Alcohol use disorders
##    metric year         val       upper       lower
## 1 Percent 1990 0.004355489 0.005574785 0.003579575
## 2 Percent 1990 0.002316023 0.002622133 0.002052042
## 3 Percent 1990 0.006539015 0.007974114 0.005392593
## 4 Percent 1990 0.002667792 0.002950154 0.002417720
## 5 Percent 1990 0.007597508 0.010585770 0.006359210
## 6 Percent 1990 0.002744876 0.003049935 0.002468063
```

```
plot(x = data.noNA.noDup$year, y = data.noNA.noDup$val, pch = 20, col = "darkgoldenrod4",
     las = 1, xlab = "year", ylab = "val",
     main = "year~val", bty = "l")
text(data.noNA.noDup$year, data.noNA.noDup$cal, labels = data.noNA.noDup$X, col = "dimgray", cex = 0.7,
```