Cleaning Data 2 assignment

Alec Brooks

2024-10-05

Setup

```
library(readr)
GDP <- read_csv("GDP.csv")

POP <- read_csv("POP.csv")
```

clean GDP

```
GDP <- GDP[c(5:221),-c(2,3,4,6)]
names(GDP) <- c("code","GDP")
GDP$GDP <- as.numeric(gsub(",", "", GDP$GDP))
```

clean POP

```
POP <- POP[c(5:221),-c(2,3,6)]
names(POP) <- c("code","country","POP")
POP$POP <- as.numeric(gsub(",", "", POP$POP))
```

merge, reorder columns, repalce NAs

```
P2G <- merge(GDP, POP, by = "code")
P2G <- P2G[, c(1, 3, 4, 2)]
```

all NAs are in Col 4.

```
which(is.na(P2G), arr.ind = TRUE)

##      row col
## [1,]  59   4
## [2,]  72   4
## [3,] 159   4
## [4,] 178   4
## [5,] 207   4
## [6,] 208   4
## [7,] 214   4

P2G$GDP[is.na(P2G$GDP)] <- round(mean(P2G$GDP, na.rm = TRUE))
```

report First and Last 6

```
head(P2G, 6)
```

```
##   code            country   POP    GDP
## 1  ABW              Aruba   106   3545
## 2  AFG        Afghanistan 41129  14266
## 3  AGO             Angola 35589 106783
## 4  ALB            Albania  2778  18916
## 5  AND            Andorra    80   3352
## 6  ARE United Arab Emirates  9441 507064
```

```
tail(P2G, 6)
```

```
##     code     country   POP    GDP
## 212  WSM       Samoa   222    833
## 213  XKX      Kosovo  1762   9409
## 214  YEM  Yemen, Rep. 33697 476955
## 215  ZAF South Africa 59894 405271
## 216  ZMB      Zambia 20018  29164
## 217  ZWE    Zimbabwe 16321  27367
```

Full Code

```
library(readr)
GDP <- read_csv("GDP.csv")
POP <- read_csv("POP.csv")
GDP <- GDP[c(5:221),-c(2,3,4,6)]
names(GDP) <- c("code","GDP")
GDP$GDP <- as.numeric(gsub(",", "", GDP$GDP))
POP <- POP[c(5:221),-c(2,3,6)]
names(POP) <- c("code","country","POP")
POP$POP <- as.numeric(gsub(",", "", POP$POP))
P2G <- merge(GDP, POP, by = "code")
P2G <- P2G[, c(1, 3, 4, 2)]
which(is.na(P2G), arr.ind = TRUE)
P2G$GDP[is.na(P2G$GDP)] <- round(mean(P2G$GDP, na.rm = TRUE))
```

```
head(P2G, 6)
tail(P2G, 6)
```