

MIL2-Project2

Catherine Young

2024-10-14

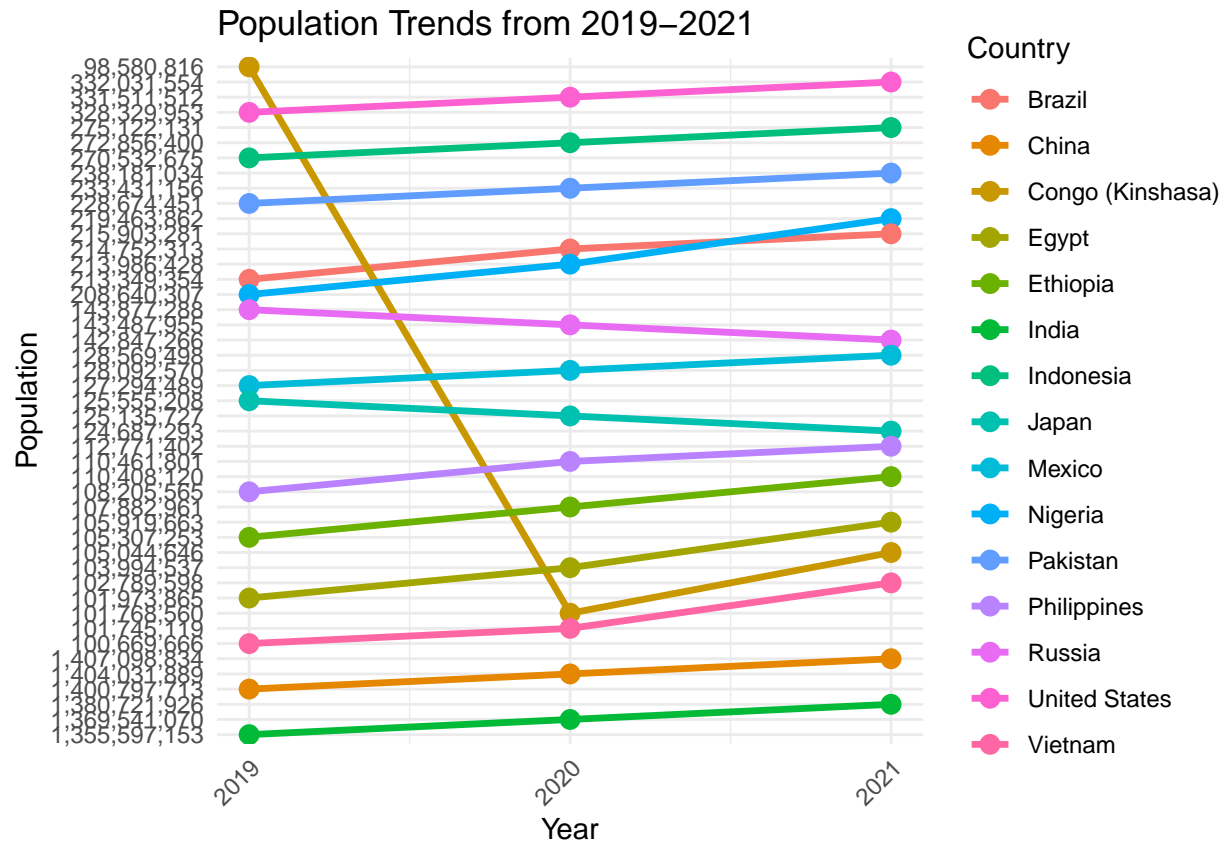
```
#load in original dataset
dataset<-read.csv("C:/Users/young/OneDrive/Desktop/DS_4002/IDB_10-07-2024 (1).csv")
head(data)
```

```
##
## 1 function (... , list = character(), package = NULL, lib.loc = NULL,
## 2     verbose = getOption("verbose"), envir = .GlobalEnv, overwrite = TRUE)
## 3 {
## 4     fileExt <- function(x) {
## 5         db <- grepl("\\\\.([^.]+\\\\\\. (gz|bz2|xz)$", x)
## 6         ans <- sub(".*\\\\\\. ", "", x)
```

```
#load needed package  
library(ggplot2)  
#filter the dataset for the years 2019-2021 using the 'Region' column  
filtered_data <- dataset[dataset$Year %in% c(2019, 2020, 2021), ]
```

```
#time series plot on pop
ggplot(filtered_data, aes(x=Year, y=Population, color=Name, group=Name)) +
  geom_line(size=1.2) +
  geom_point(size=3) +
  labs(title="Population Trends from 2019-2021",
       x="Year",
       y="Population",
       color="Country") +
  scale_x_continuous(breaks=c(2019, 2020, 2021)) +
  theme_minimal() +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

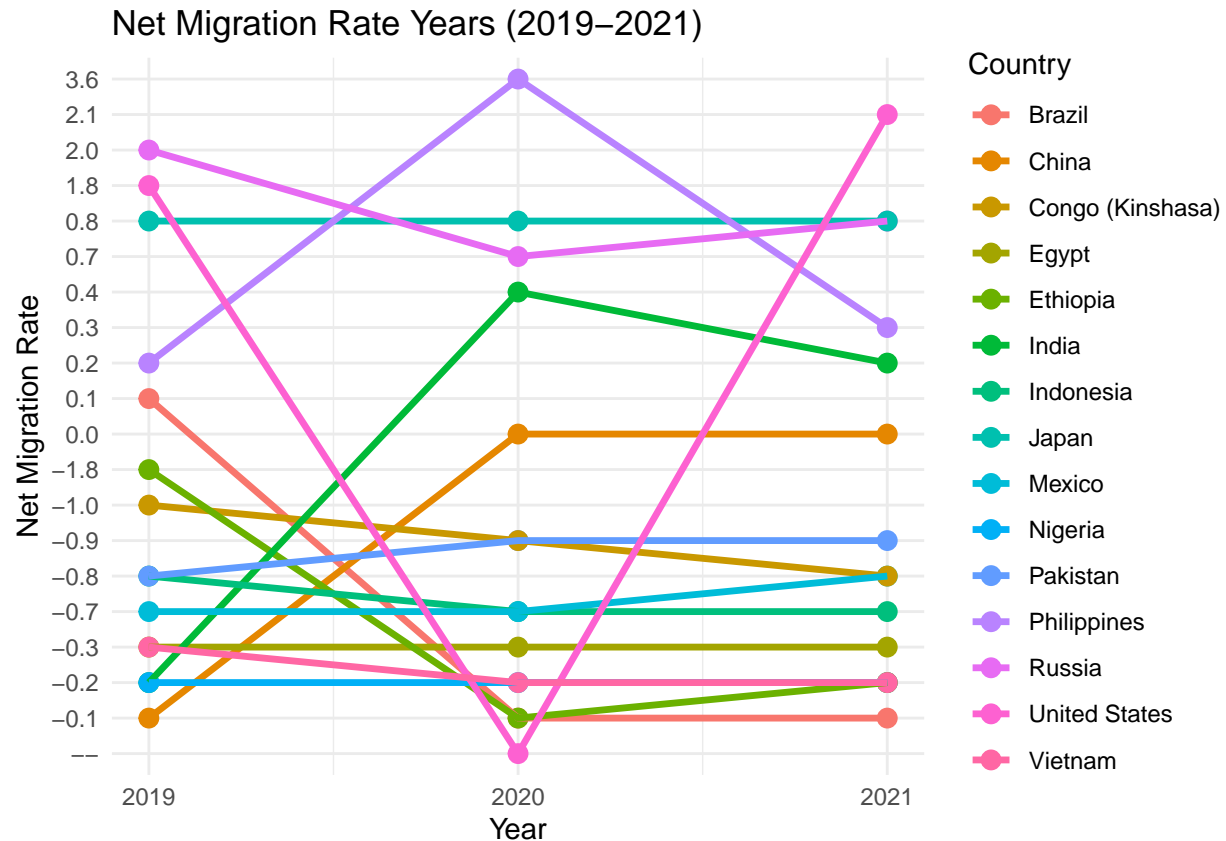


#save the graph as image

```
ggsave("population_trends_2019-2021.png", width = 10, height = 6)
```

#create the scatter plot for Net Migration Rate on Year

```
ggplot(filtered_data, aes(x=Year, y=`Net.Migration.Rate`, color=Name, group=Name)) +
  geom_point(size=3) +
  geom_line(size=1.2) +
  labs(title = "Net Migration Rate Years (2019-2021)",
       x="Year",
       y="Net Migration Rate",
       color="Country") +
  scale_x_continuous(breaks=c(2019, 2020, 2021)) +
  theme_minimal()
```



#save the graph as image

```
ggsave("net_migration_rate_2019-2021.png", width = 10, height = 6)
```

#clean the data. Read in the dataset with values entered for nas

```
clean_data<-read.csv("C:/Users/young/OneDrive/Desktop/DS 4002/IDB_10-07-2024_clean.csv")
filtered_data_clean<- clean_data[clean_data$Year %in% c(2019, 2020, 2021), ]
head(filtered_data_clean)
```

##	Name	Region	GENC	Year	Population
## 34	China	2019,China	CN	2019	1,400,797,713
## 35	India	2019,India	IN	2019	1,355,597,153
## 36	United States	2019,United States	US	2019	328,329,953
## 37	Indonesia	2019,Indonesia	ID	2019	270,532,675
## 38	Pakistan	2019,Pakistan	PK	2019	228,674,451
## 39	Brazil	2019,Brazil	BR	2019	213,349,354
##	Total.Fertility.Rate	Crude.Birth.Rate	Life.Expectancy.at.Birth..Both.Sexes		
## 34	1.45	10.6	77.7		
## 35	2.13	17.1	71.5		
## 36	2.00	11.4	70.0		
## 37	2.11	16.2	72.3		
## 38	3.68	27.9	68.8		
## 39	1.77	14.1	75.9		
##	Crude.Death.Rate	Net.Migration.Rate	Net.international.migrants..both.sexes		
## 34	7.2	-0.1	-180,000		
## 35	6.6	-0.2	-267,593		

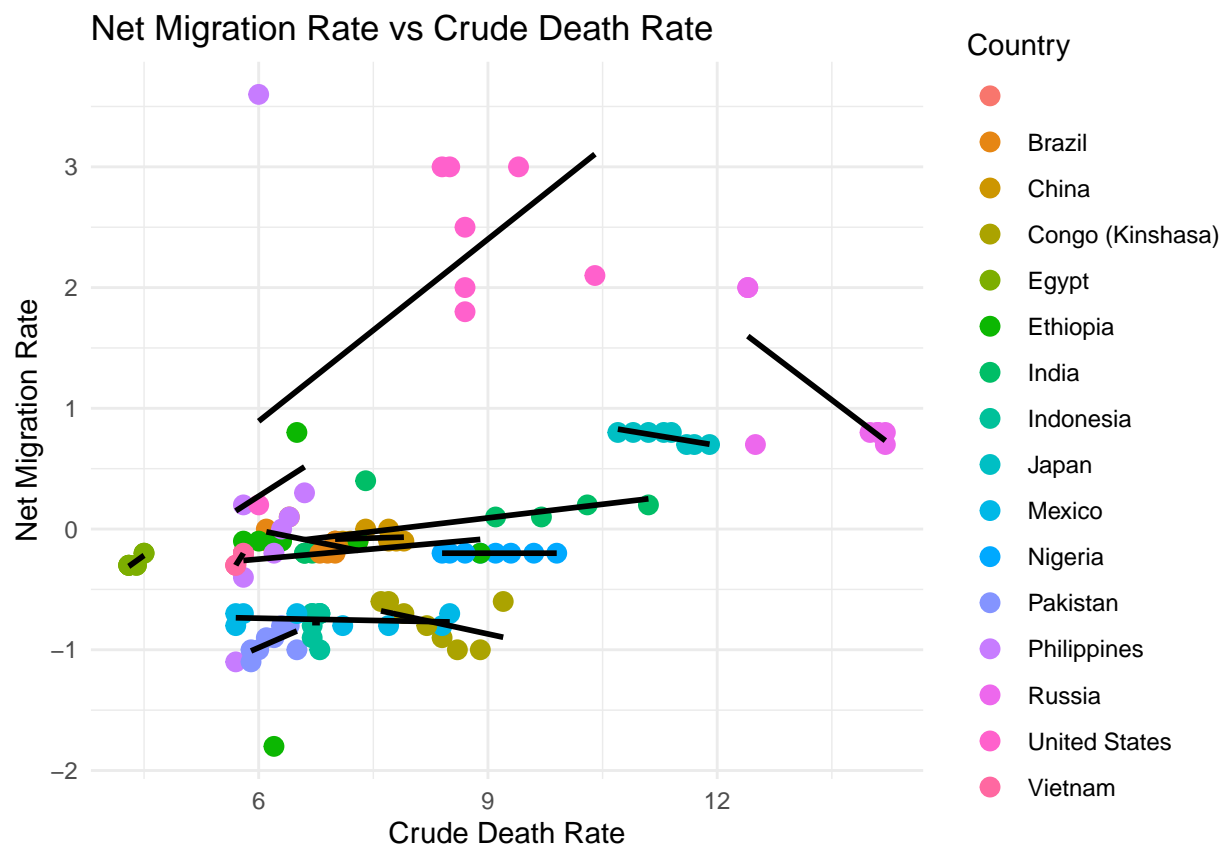
```
## 36          8.7          1.8          594,330
## 37          6.7         -0.8         -217,436
## 38          6.3         -0.8         -187,193
## 39          6.4          0.1          29,646
```

```
#scatterplot on: Net Migration Rate vs Crude Death Rate
ggplot(clean_data, aes(x=`Crude.Death.Rate`, y=`Net.Migration.Rate`, color=Name)) +
  geom_point(size=3) +
  geom_smooth(method="lm", se=FALSE, color="black", aes(group=Name)) + #adds line of best fit
labs(title= "Net Migration Rate vs Crude Death Rate",
      x= "Crude Death Rate",
      y= "Net Migration Rate",
      color= "Country") +
theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 8 rows containing missing values ('geom_point()').
```



```
#save the graph as image
ggsave("net_migration_vs_crude_death_rate.png", width= 10, height = 6)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

```
## Removed 8 rows containing missing values ('geom_point()').
```

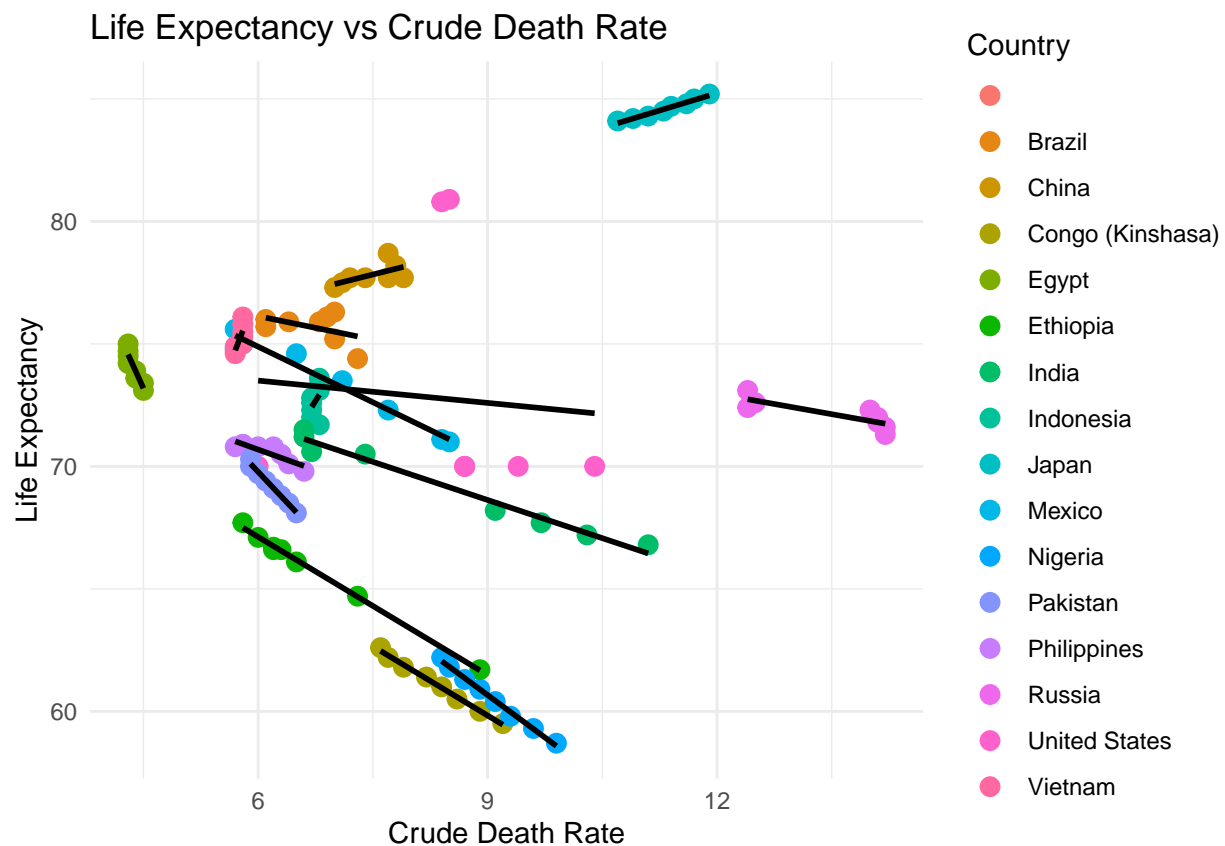
```
#scatterplot on: Life Expectancy vs Crude Death Rate
```

```
ggplot(clean_data, aes(x=`Crude.Death.Rate`, y= `Life.Expectancy.at.Birth..Both.Sexes`,color=Name)) +  
  geom_point(size = 3) +  
  geom_smooth(method="lm", se=FALSE, color="black",aes(group=Name)) + #adds line of best fit  
  labs(title = "Life Expectancy vs Crude Death Rate",  
        x= "Crude Death Rate",  
        y= "Life Expectancy",  
        color= "Country") +  
  theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 8 rows containing missing values ('geom_point()').
```



```
#save the graph as image
```

```
ggsave("life_expectancy_vs_crude_death_rate.png", width=10, height= 6)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

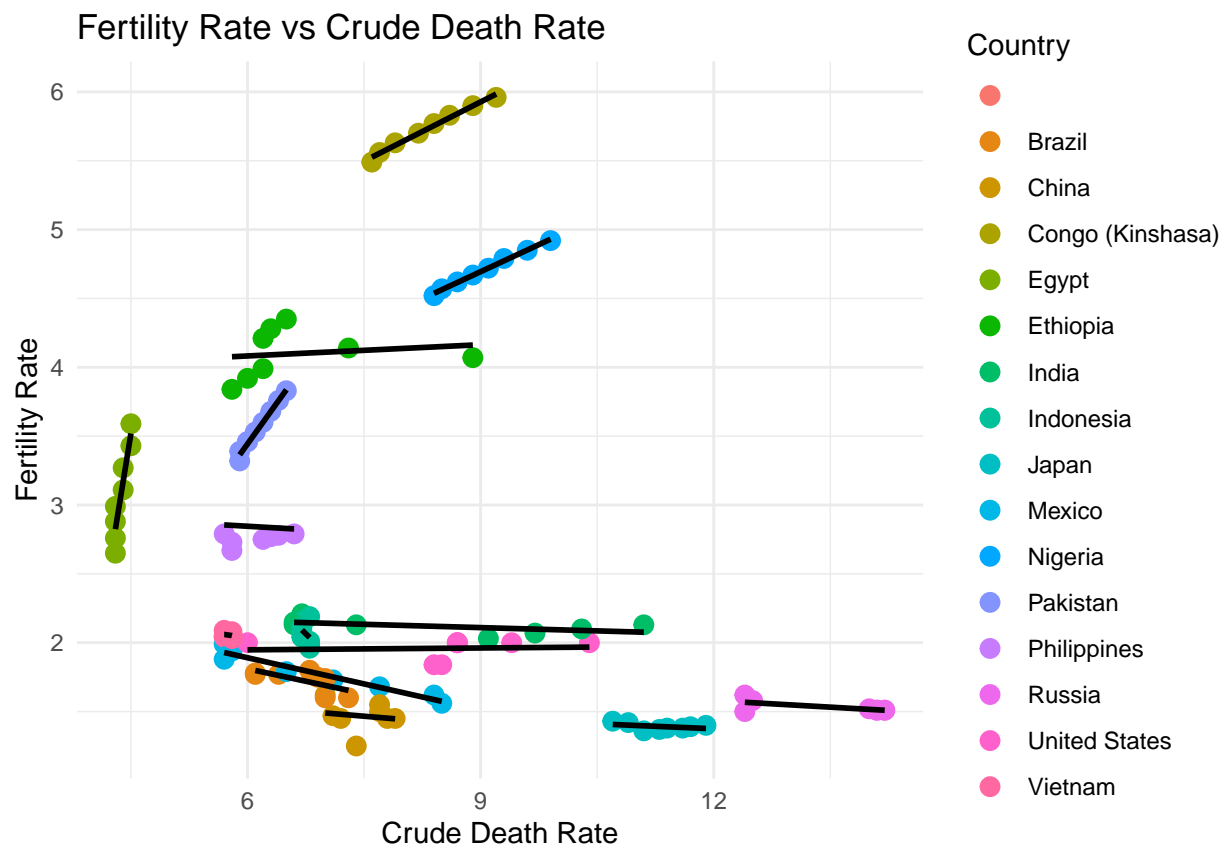
```
## Removed 8 rows containing missing values ('geom_point()').
```

```
#scatterplot on: Total Fertility Rate vs Crude Death Rate
ggplot(clean_data, aes(x=`Crude.Death.Rate`, y=`Total.Fertility.Rate`, color=Name)) +
  geom_point(size=3) +
  geom_smooth(method="lm", se=FALSE, color="black", aes(group=Name)) + #adds line of best fit
labs(title = "Fertility Rate vs Crude Death Rate",
     x= "Crude Death Rate",
     y= "Fertility Rate",
     color= "Country") +
theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 8 rows containing missing values ('geom_point()').
```



```
#save the graph as image
ggsave("fertility_rate_vs_crude_death_rate.png", width=10, height=6)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 8 rows containing non-finite values ('stat_smooth()').
```

```
## Removed 8 rows containing missing values ('geom_point()').
```