POPULATION STUDY:

AN ANALYSIS ON WORLD'S FERTILITY RATE AND THE EXPECTED FUTURE TREND

Group 2

Tseng Chia Yi | Kenneth Low Yan Wei | Peace Tay Jiunn Ching | Wang Kunrui | Yap Pin Yaw

WHICH COUNTRY HAS A HIGHER FERTILITY RATE?

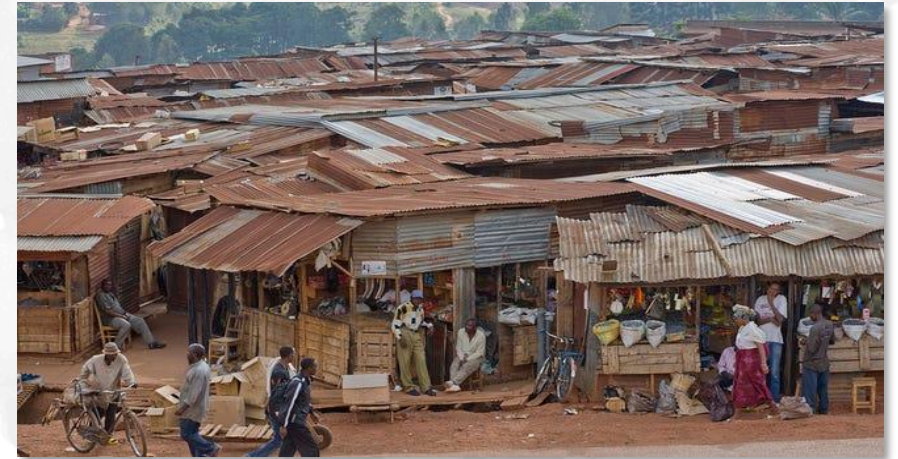**Wealthy and Affluent**

1

High GDP per capita

Educated Population

2

Skilled and competent workforce

Healthy Citizens

3

Advanced medical care and technology

Poor and Poverty Stricken

1

Low GDP per capita

Less Educated Population

2

High proportion of illiterate people

Poor Heathcare

3

Poor medical facilities,
lack of medical technology

CONTENT

OBJECTIVE

Identify global fertility rate trends

Identify factors that affects fertility rate

Determine impact of factors, and project future population and fertility rates

WHY IS THIS IMPORTANT TO YOU?

Economic factors
Higher spending on healthcare and supporting the elderly

Social factors
Heavy burden on the working population

Impact of aging population

A low fertility rate of a country will result in an aging population. There will be increased government spending on healthcare and medical facilities to support the elderly. This means higher taxes on the working population.

Families will have added responsibilities to take care of the elderly, with 1 working adult taking care of 2 elderly parents. This will have ripple effects on societal structure in the long run.

DEVELOPMENTAL INDICATORS

Life Expectancy

Indicates the level of advancement in medical technology and healthcare

Gross % of Secondary School Enrollment

Indicates how educated the population is

GDP Per Capita

Indicates the level of wealth and standard of living

DATA PREPARATION

- ## Data Sources
  - ➢ https://ourworldindata.org/
  - ➢ https://worldbank.org/
  - ➢ https://www.un.org/
  - ➢ https://www.imf.org

- ## Data Issues
  - • Missing data
    - ➢ Exclude missing fertility in 24 countries
    - ➢ Exclude missing GDP in 5 countries
    - ➢ Exclude countries with missing data for certain years on graph visualization

| Variable	Description	How it will be used
Fertility rate	Total number of children that would be born to each woman	To define a country's fertility rate
Gross Domestic Product (GDP per capita)	National GDP divide by its population (The Investopedia Team, 2022)	To define a country's economic level
Infant Mortality rate	Number of deaths of children under one year of age, expressed per 1000 live births (OECD, n.d.)	As a proxy of a country's healthcare accessibility and technological advancement
Secondary School Enrolment %	Percentage of defined age group enrolment in Secondary School education	As a proxy of the population's education level
Countries	Name of Countries	To categorise advanced / emerging and developing countries
Country Standard ID and Codes	List of countries, their associated numeric code, and alpha code	To allow for more seamless merging of data frames

CONTENT

WORLD MAP INDICATORS
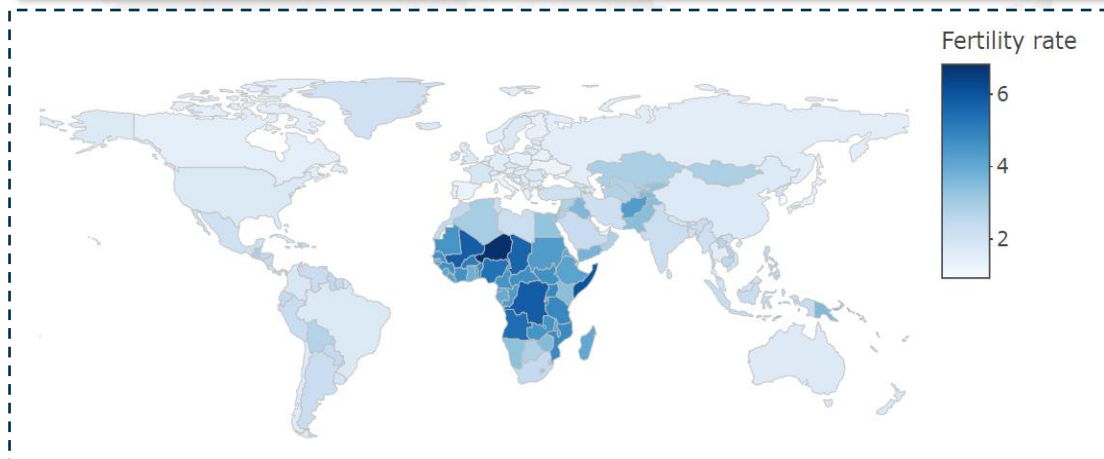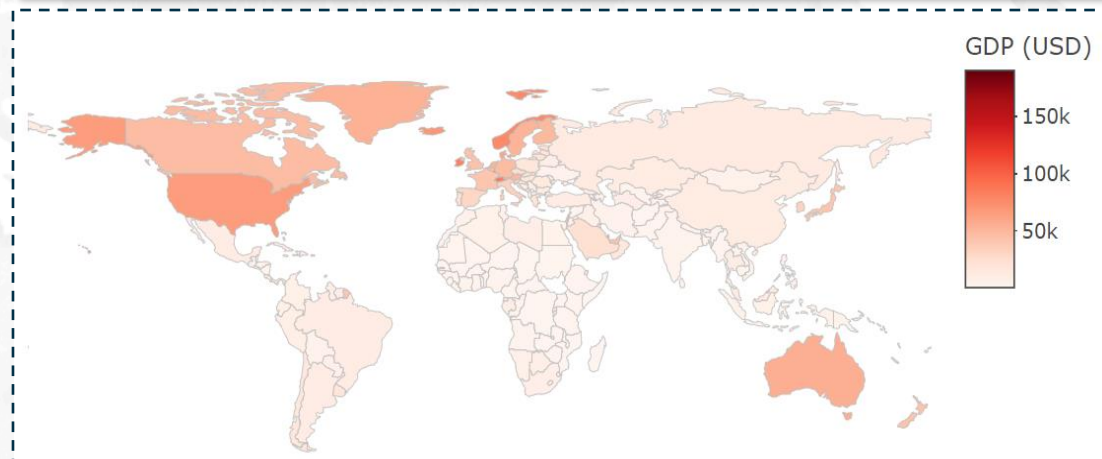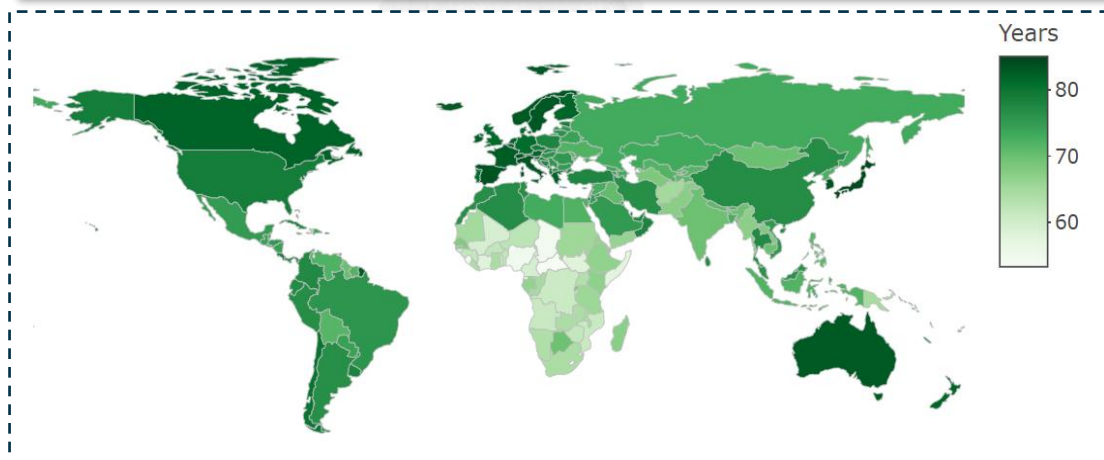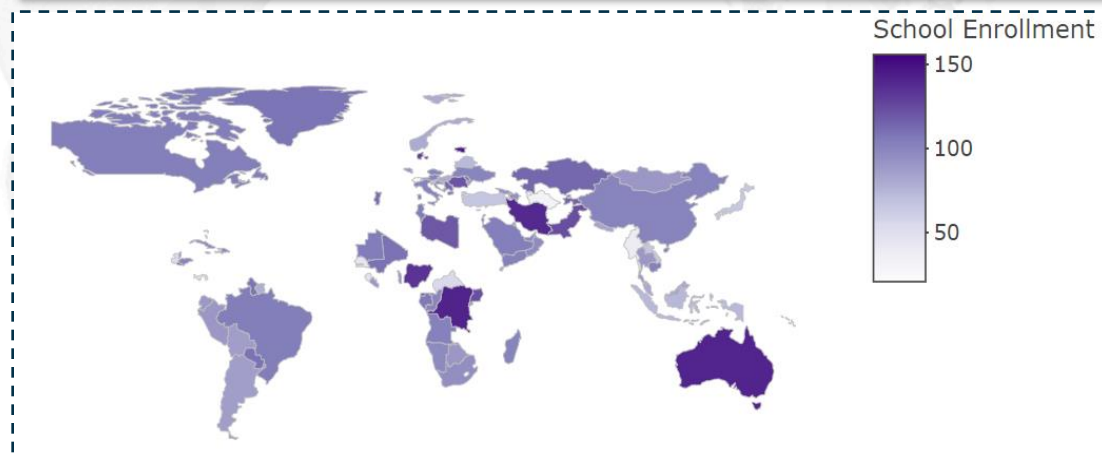
*missing school enrollment data for 2019

FERTILITY RATE PROJECTION

Singapore, Linear Projection

Ghana, Linear Projection

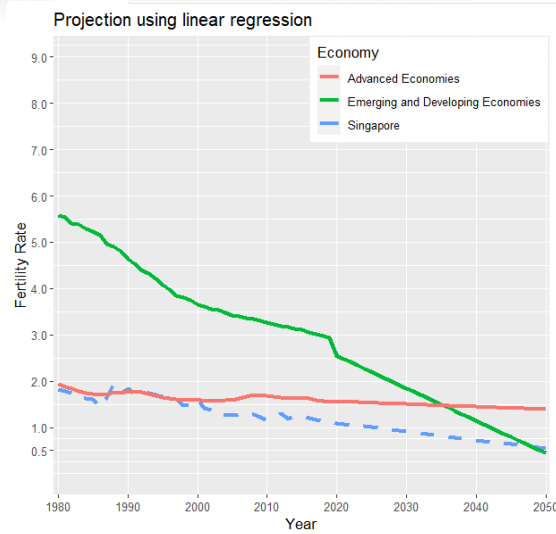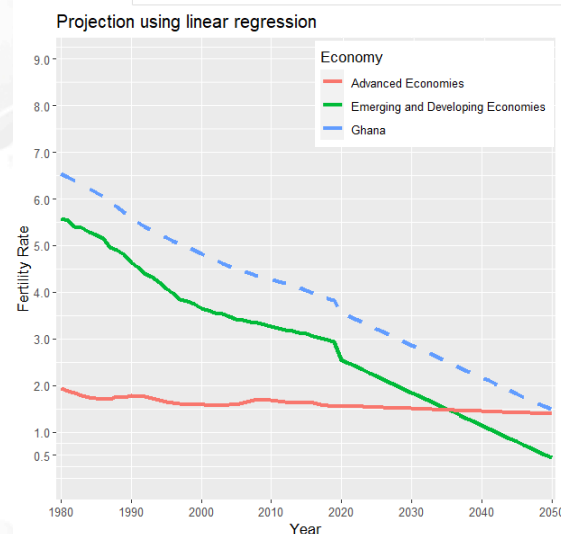### Singapore, Exponential Projection
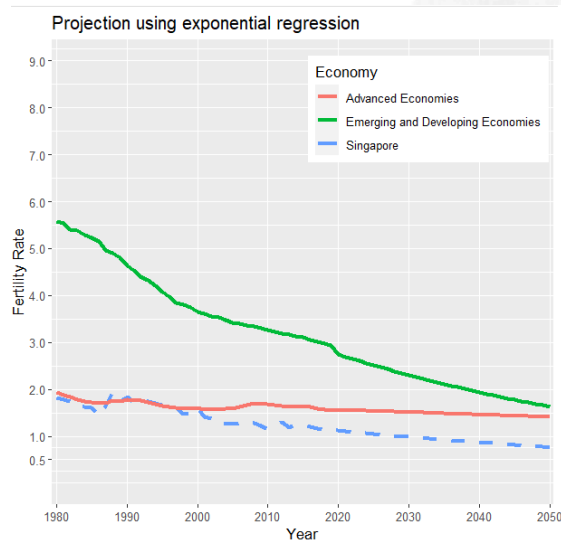
Ghana, Exponential Projection

🔍 Observation

- Advanced economies' countries projection is approximately similar while Emerging & Developing economies' countries are experiencing a high decreasing rate.
- In year 2050, Emerging & Developing economies' countries are having lower fertility rate than advanced economies.
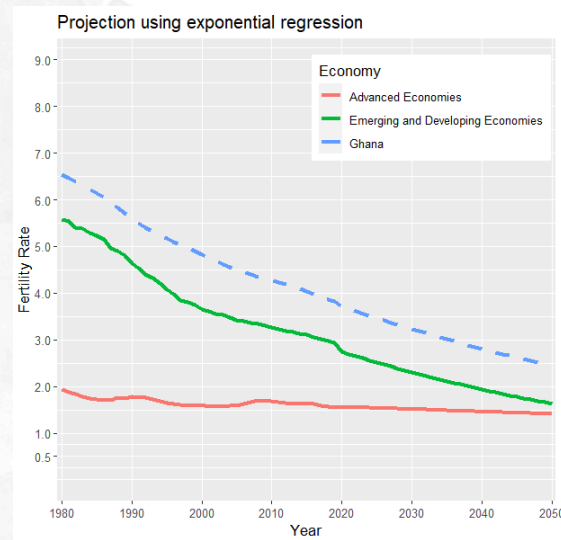
🔍 Observation

- Countries moving towards a plateau state as they continue to be projected over time towards 2050.
- Emerging & Developing economies' countries transforming into advanced economies' countries and exhibit their trend of having low fertility rate.

FERTILITY RATE ANALYSIS

Linear versus Exponential Regression

Ghana, Linear Projection

Projection using linear regression

Ghana, Exponential Projection

Projection using exponential regression
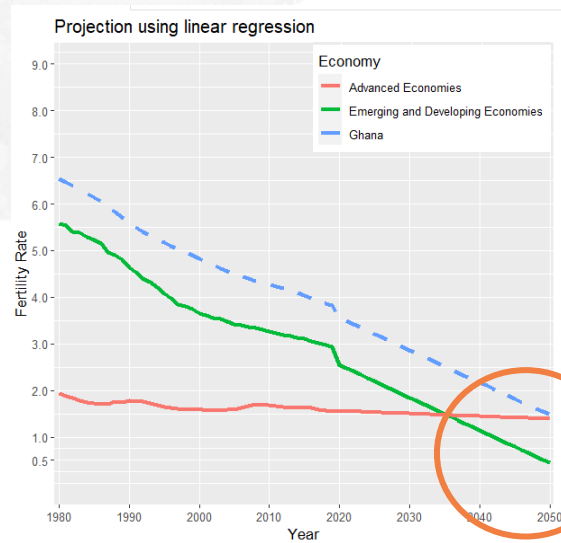
Years → ∞

Fertility Rate → -ve

Fertility rate cannot be negative!
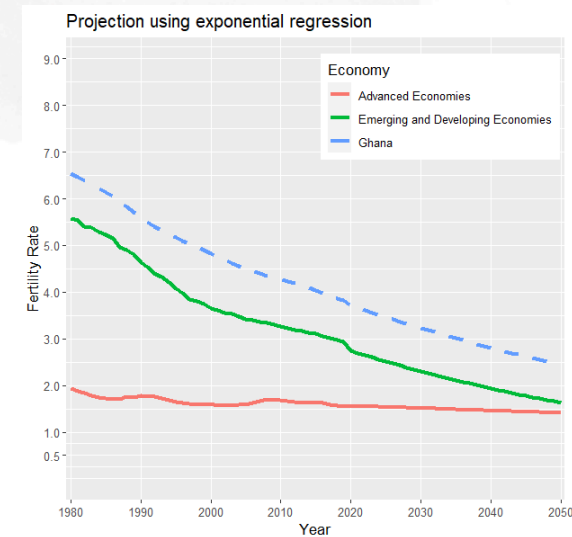
- As a country advances, the fertility rate tends to hit a plateau (as shown in the advanced economies' countries projection)

- Exponential regression proved to be a much effective and reliable projection of fertility rate on time-series as compared to linear regression.

POPULATION DEMOGRAPHICS

- For the country of Ghana
- For year 1980

1980

Population Size of Males | Females('000)
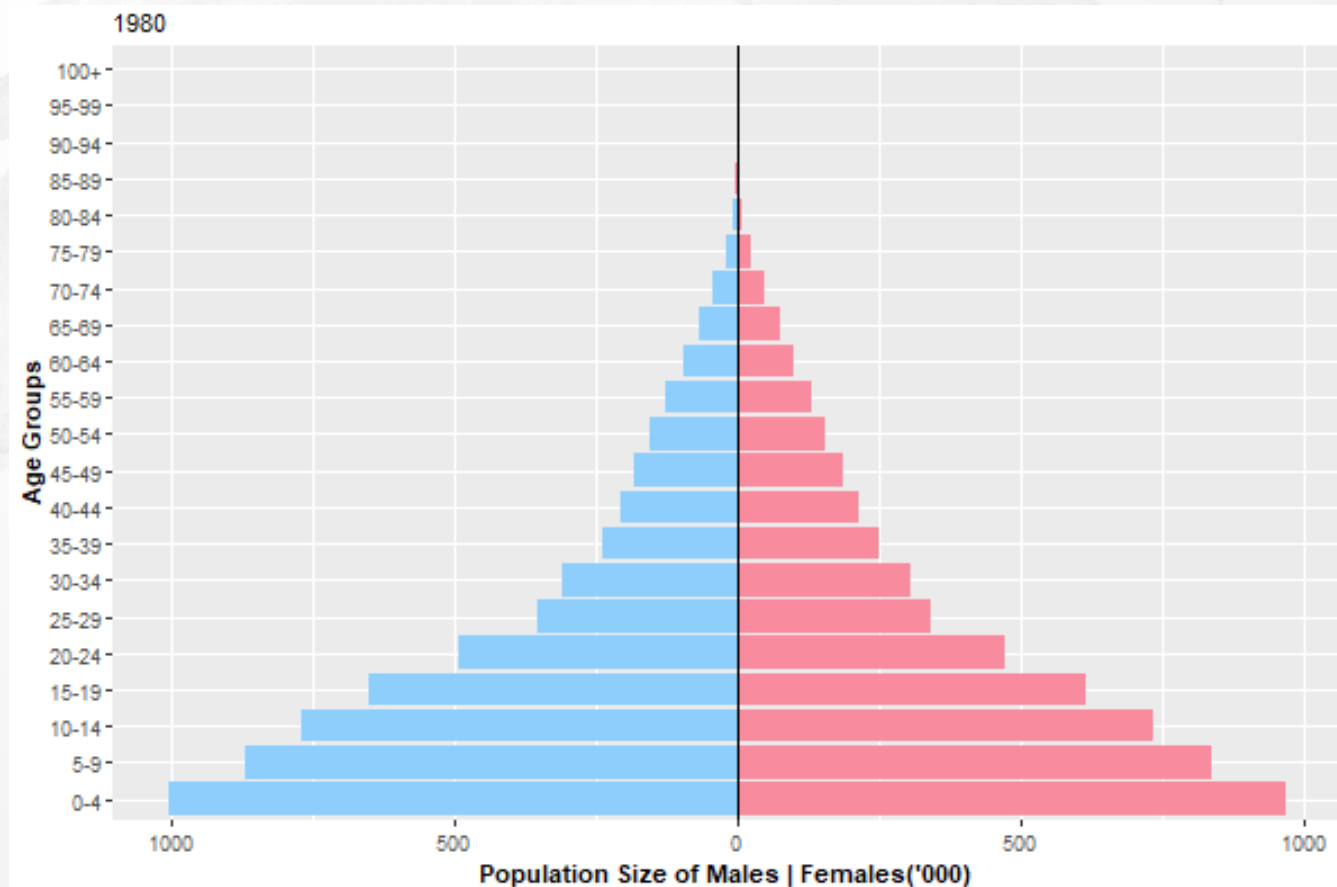
POPULATION DEMOGRAPHICS

Ghana, 1980
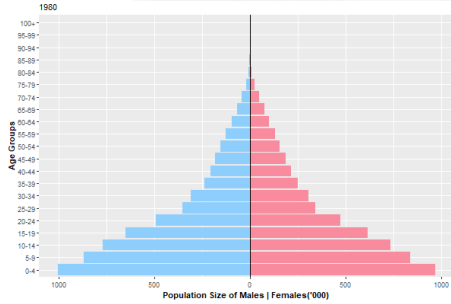
Singapore, 1980
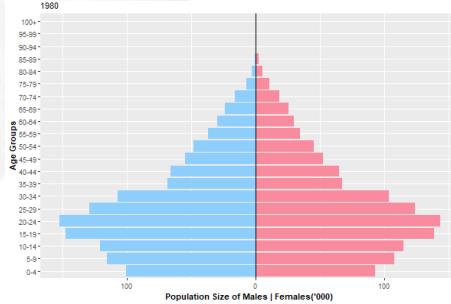
Ghana, 2020
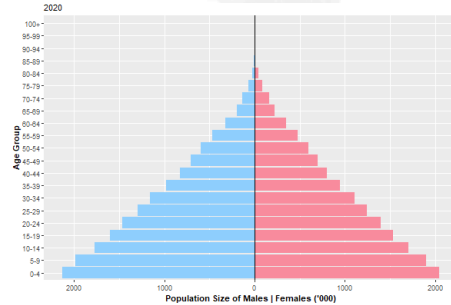
Singapore, 2020

Ghana, 2050
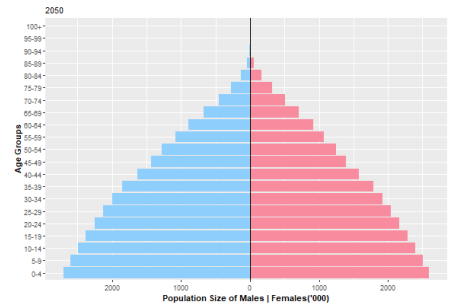
Singapore, 2050
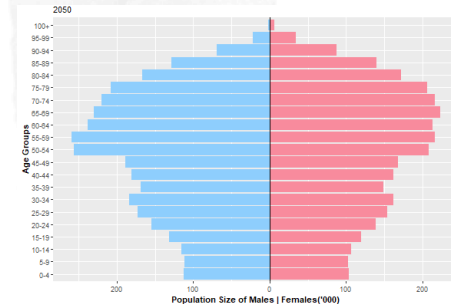
According to WHO, ideal ratio is 105 males : 100 females

Pre-Reproductive:	Reproductive:	Post-Reproductive:
0Y – 14Y	15Y – 44Y	45Y - EOL

Singapore, 1980	Singapore, 2020	Singapore, 2050
Male : Female	Male : Female	Male : Female
104 : 100	113 : 100	113 : 100

Ghana, 1980	Ghana, 2020	Ghana, 2050
Male : Female	Male : Female	Male : Female
100 : 100	104 : 100	104 : 100

GDP PER CAPITA VS FERTILITY RATE, 2019

GDP per capita

Region
- East Asia & Pacific
- Europe & Central Asia
- Latin America & Caribbean
- Middle East & North Africa
- North America
- South Asia
- Sub-Saharan Africa

IMF
- Advanced Economies
- Emerging and Developing Economies
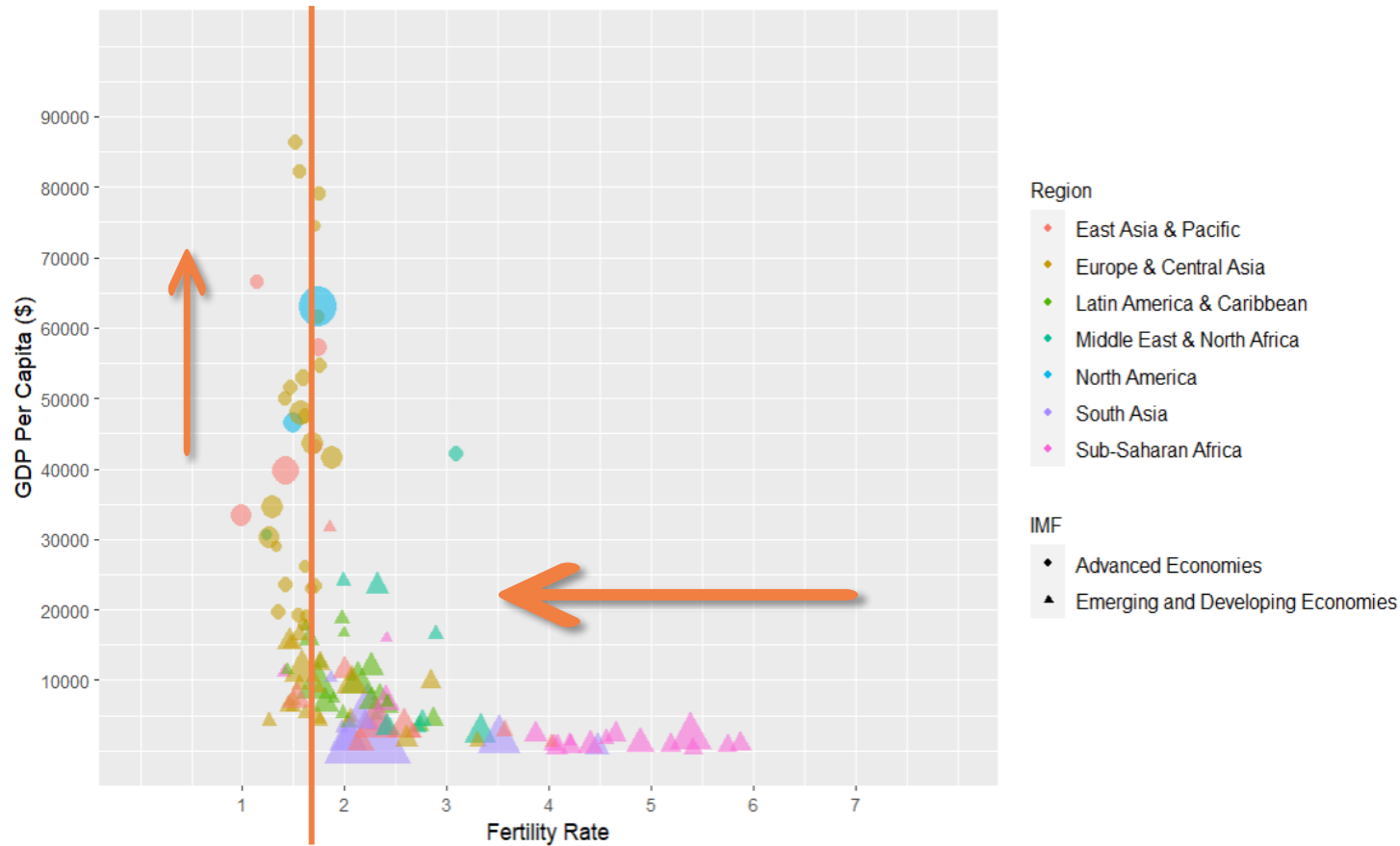
🔍 Observation

- Across the years, the elements are clearly trending towards the left for both Advanced and Developing countries

- All Advanced Country's fertility rate falls to the range of +- 1.5 and stagnates there, while GDP continues to increase

- African countries have one of the highest Fertility Rates and lowest GDP per capita

💡 Analysis

- This is consistent with trends around the world where citizens of Advanced Countries are getting married at a later age, and families prefer to have only 1 or 2 children*

- African country's population are still heavily reliant on agriculture and mining. Therefore, GDP is low while families prefer to have more children to help with manual labor

OTHER DEVELOPMENTAL INDICATORS VS FERTILITY RATE, 2019

Secondary School Enrolment
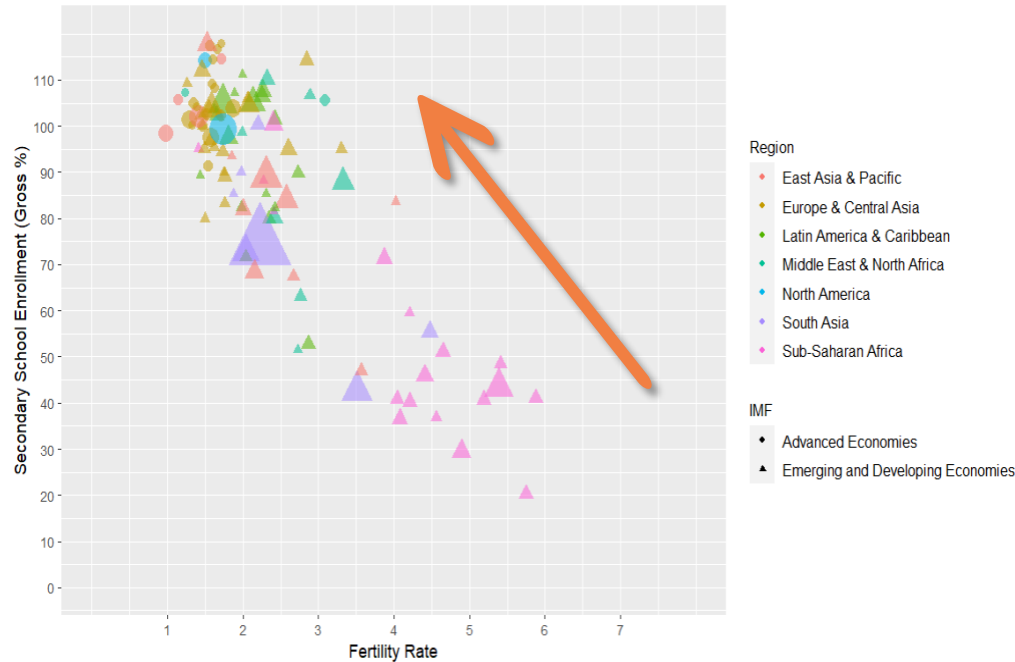
Life expectancy
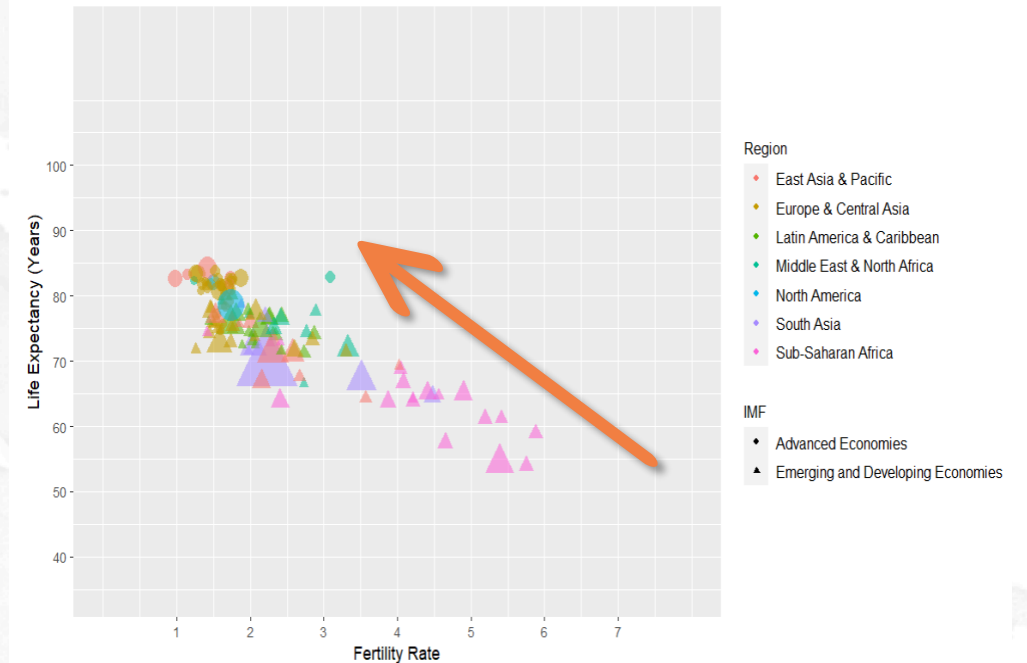
🔍 Observation

- Similarly, all developmental indicators improve as fertility rate falls and approaches 1.5
- African countries continue to lack behind

💡 Analysis

- Developing Countries follows the trend of Advanced Countries, as society advances and improves, the fertility rate drops

CLUSTER PLOT [1]

GDP per capita vs Fertility Rate
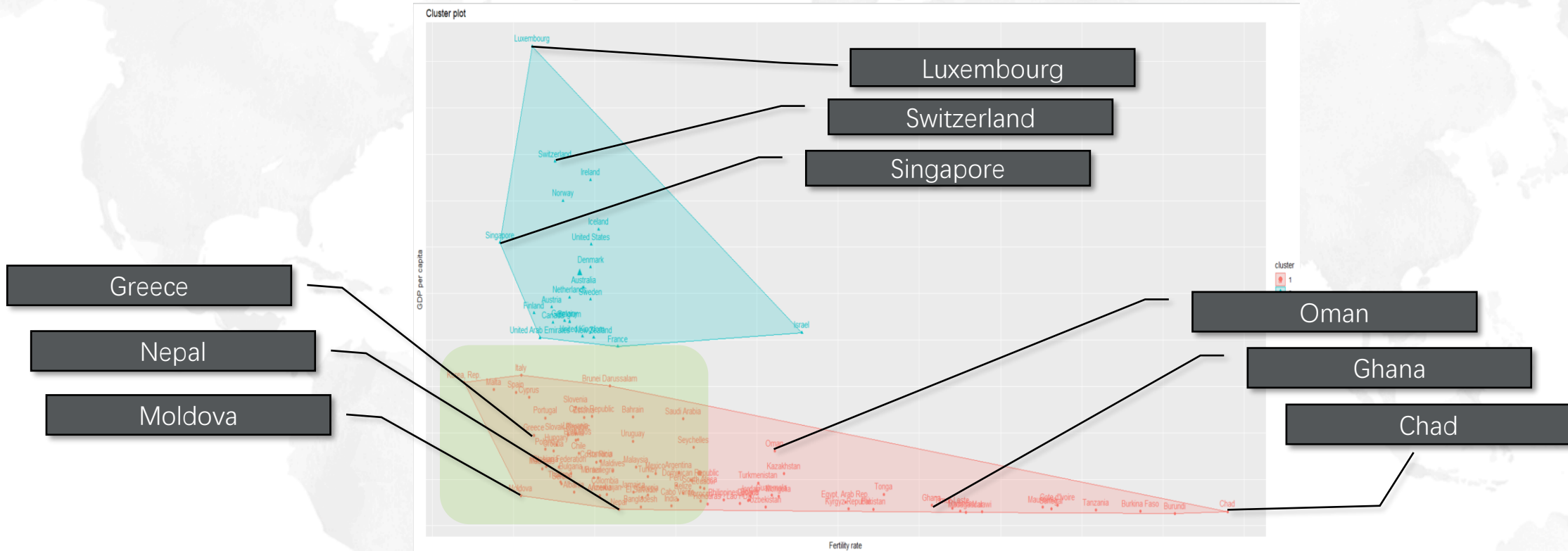
🔍 **Observation**

Highly affluent countries (Luxembourg, Switzerland, Singapore etc.) are grouped in the blue box, while less developed countries (Oman, Ghana etc.) are grouped in the red box. *The country Chad, has one of the lowest GDP per capita and highest fertility rates in the world*

💡 **Analysis**

Countries in the green box have relatively low GDP per capita and Fertility rate. These countries experiences a lower standard of living and aging population. Foreign investors will shun such countries as a result.

Reference: https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/

CLUSTER PLOT [2]

Life Expectancy vs Fertility Rate
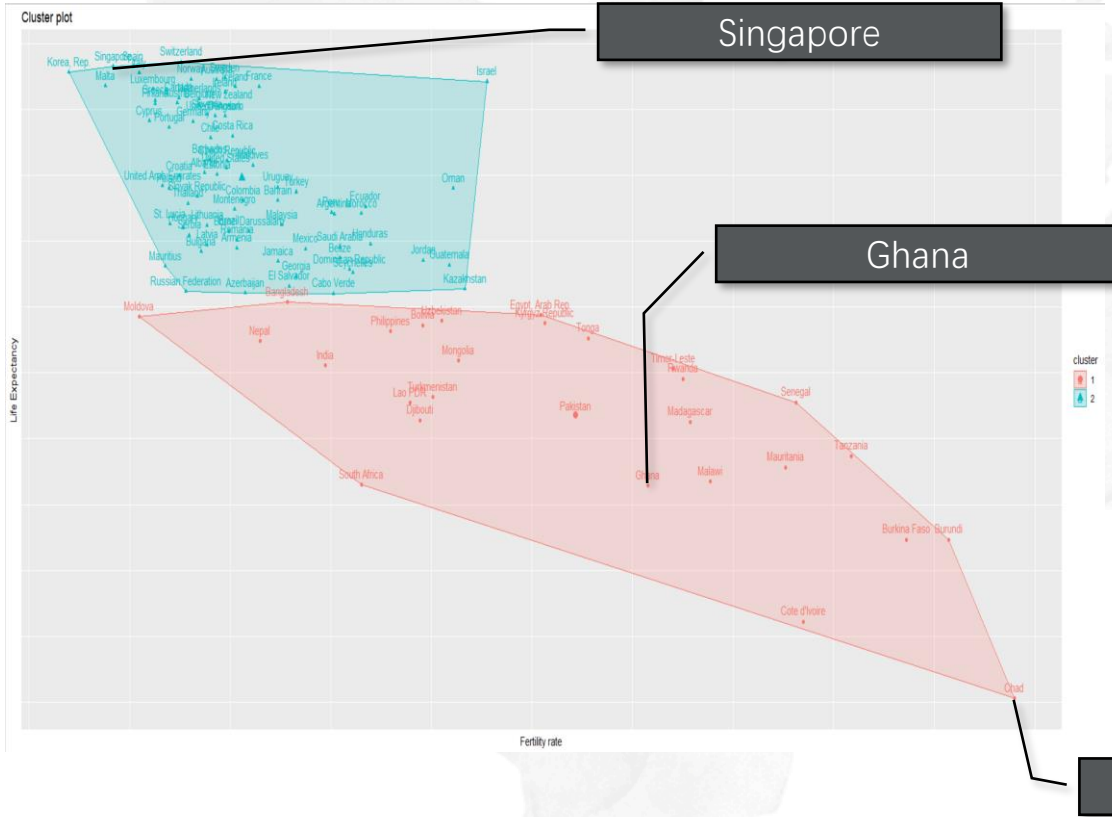
Secondary School Enrollment vs Fertility Rate
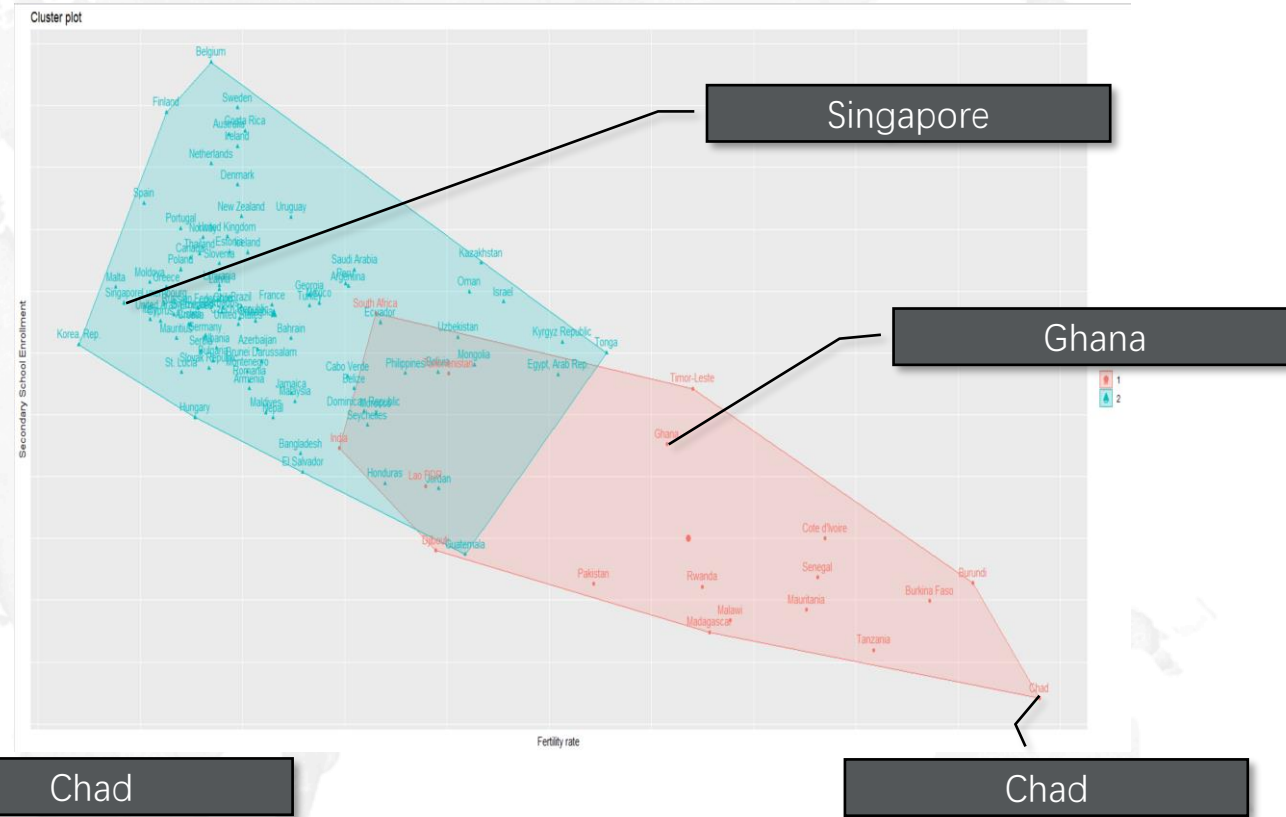
Analysis

Across all developmental metrics, more advanced countries have better standards of living and developmental indicators, with a lower fertility rate. Conversely, developing countries are grouped together with a higher fertility rate and lower standard of living. Once again, country Chad has the highest fertility rate and lowest standard of living

WHERE IS GHANA?

Africa

Republic of Ghana

Ghana is a country in West Africa. Ghana has experienced multiple and long-lasting civil and drug wars, and is rife governmental corruption*

RADAR

A RADAR PLOT SHOWS THE CORRELATION BETWEEN KEY VARIABLES AND **FERTILITY RATE**

IN GHANA…

🔍 Observation

A strong effect is observed between GDP and Fertility (-0.75), and Education and Fertility (-0.74).
A very strong effect (-0.96) is observed between Life Expectancy and Fertility.
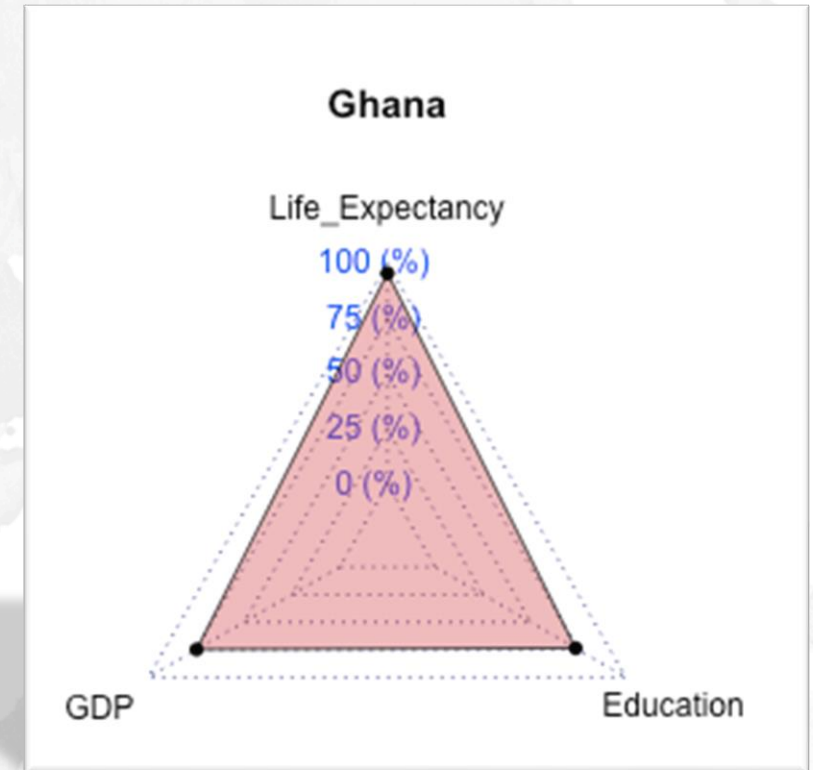
Ghana

	Life Expectancy	GDP	Education
Correlation	-0.96	-0.75	-0.74

RADAR ANALYSIS

Analysis

- Life Expectancy shares the strongest correlation with Fertility

- Ghana has the world's lowest life expectancy at 64.2 years and one the world's highest fertility rate

- Might be missing out on the demographic transition phenomenon

Ghana

	Life Expectancy	GDP	Education
Correlation	-0.96	-0.75	-0.74

RADAR ANALYSIS

IN Singapore…

DEMOGRAPHIC TRANSITION

Phenomenon which refers to the historical shift from

High birth rates + High death rates

to

Low birth rates + Low death rates

in societies with advancement in technology, education and economic development

Singapore
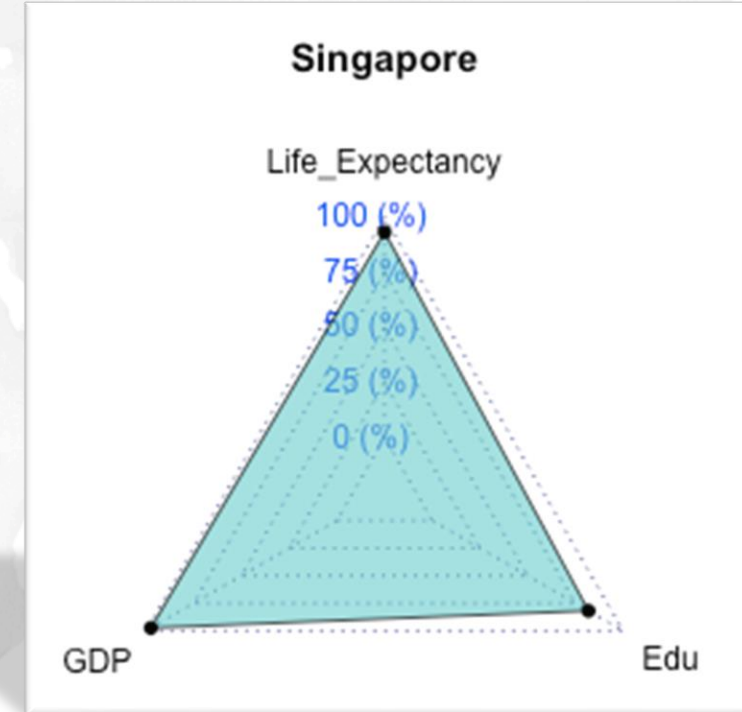
	Life Expectancy	GDP	Education
Correlation	-0.93	-0.97	-0.81

LINEAR REGRESSION

Constant/Intercept   Independent Variable

$$Y_i = \beta_0 + \beta_1 X_i$$

Dependent Variable   Slope/Coefficient

GOAL:

To find out if there is significant linear regression between each Independent Variable (GDP, Life Expectancy, Education) with Dependent Variable (Fertility).

Also, if there is any difference between Advanced and Emerging Economies

LINEAR REGRESSION TESTING_GDP

H0: *Linear regression is not significant*
H1: *Linear regression is significant*

Advanced countries

```
Call:
lm(formula = Advanced_FR$Fertility ~ Advanced_GDP$GDP)

Residuals:
     Min       1Q    Median       3Q       Max
-0.164905 -0.048104  0.008657  0.048640  0.221825

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.787e+00  3.167e-02   56.418  < 2e-16 ***
Advanced_GDP$GDP -4.097e-06  1.079e-06   -3.797 0.000513 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08407 on 38 degrees of freedom
Multiple R-squared:  0.275,     Adjusted R-squared:  0.2559
F-statistic: 14.42 on 1 and 38 DF,  p-value: 0.0005135
```

H_0 is rejected

The linear regression for fertility rate with GDP is significant in advanced countries.

Developing countries

```
Call:
lm(formula = Emerging_FR$Fertility ~ Emerging_GDP$GDP)

Residuals:
     Min       1Q    Median       3Q       Max
-0.74311 -0.44216 -0.04787  0.29229  1.36098

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       5.280e+00  2.023e-01   26.099  < 2e-16 ***
Emerging_GDP$GDP -3.546e-04  4.489e-05   -7.899 1.54e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5575 on 38 degrees of freedom
Multiple R-squared:  0.6215,     Adjusted R-squared:  0.6116
F-statistic:  62.4 on 1 and 38 DF,  p-value: 1.545e-09
```

H_0 is rejected

The linear regression for fertility rate with GDP is significant in emerging and developing countries.

LINEAR REGRESSION TESTING_LIFE EXPECTANCY

H0: *Linear regression is not significant*
H1: *Linear regression is significant*

Advanced countries

Developing countries

```
Call:
lm(formula = Advanced_FR$Fertility ~ Advanced_LifeExp$LifeExp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.16908 -0.05211  0.01494  0.05115  0.19205

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.684809   0.434739   8.476 2.74e-10 ***
Advanced_LifeExp$LifeExp -0.025881   0.005603  -4.619 4.33e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07902 on 38 degrees of freedom
Multiple R-squared:  0.3596,    Adjusted R-squared:  0.3428
F-statistic: 21.34 on 1 and 38 DF,  p-value: 4.327e-05
```

```
Call:
lm(formula = Emerging_FR$Fertility ~ Emerging_LifeExp$LifeExp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.46563 -0.25710  0.04619  0.23269  0.43166

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              18.91025    0.77476   24.41   <2e-16 ***
Emerging_LifeExp$LifeExp -0.23123    0.01187  -19.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2734 on 38 degrees of freedom
Multiple R-squared:  0.909,     Adjusted R-squared:  0.9066
F-statistic: 379.5 on 1 and 38 DF,  p-value: < 2.2e-16
```

H_0 is rejected

The linear regression for fertility rate with life expectancy is significant in advanced countries.

H_0 is rejected

The linear regression for fertility rate with life expectancy is significant in emerging and developing countries.

LINEAR REGRESSION TESTING_EDUCATION LEVEL

H0: *Linear regression is not significant*
H1: *Linear regression is significant*

Advanced countries

```
Call:
lm(formula = Advanced_FR$Fertility ~ Advanced_Edu$Edu)

Residuals:
     Min        1Q    Median        3Q       Max
-0.141935 -0.037517  0.001695  0.038362  0.120049

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.582265   0.103330   24.99  < 2e-16 ***
Advanced_Edu$Edu  -0.008851   0.001007   -8.79 1.08e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05669 on 38 degrees of freedom
Multiple R-squared:  0.6703,   Adjusted R-squared:  0.6616
F-statistic: 77.26 on 1 and 38 DF,  p-value: 1.083e-10
```

H₀ is rejected

The linear regression for fertility rate with education level is significant in advanced countries.

Developing countries

```
Call:
lm(formula = Emerging_FR$Fertility ~ Emerging_Edu$Edu)

Residuals:
    Min       1Q   Median       3Q      Max
-0.48411 -0.14970  0.00193  0.20402  0.31391

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        7.401789   0.155368   47.64   <2e-16 ***
Emerging_Edu$Edu  -0.059826   0.002539  -23.57   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2293 on 38 degrees of freedom
Multiple R-squared:  0.936,    Adjusted R-squared:  0.9343
F-statistic: 555.3 on 1 and 38 DF,  p-value: < 2.2e-16
```

H₀ is rejected

The linear regression for fertility rate with education level is significant in emerging and developing countries.

LINEAR REGRESSION TEST FOR SIGNIFICANCE

	Advanced Countries	Developing Countries
GDP vs Fertility	Linear regression is significant	**Reject!** Linear regression is significant
Life Expectancy vs Fertility	**Reject!** Linear regression is significant	**Reject!** Linear regression is significant
Education vs Fertility	**Reject!** Linear regression is significant	**Reject!** Linear regression is significant

MULTI LINEAR REGRESSION TESTING

```
Call:
lm(formula = allcountries$Fertility ~ allcountries$GDP + allcountries$LifeExp +
    allcountries$Edu)

Residuals:
     Min        1Q    Median        3Q       Max
-0.242513 -0.080870 -0.007791  0.093342  0.257812

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             4.307e+00  3.435e+00   1.254   0.2180
allcountries$GDP        4.916e-05  1.943e-05   2.530   0.0159 *
allcountries$LifeExp    6.638e-02  6.440e-02   1.031   0.3095
allcountries$Edu       -8.556e-02  1.260e-02  -6.792 6.16e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1327 on 36 degrees of freedom
   (1 observation deleted due to missingness)
Multiple R-squared:  0.9703,    Adjusted R-squared:  0.9678
F-statistic: 391.9 on 3 and 36 DF,  p-value: < 2.2e-16
```

Observation

- P values calculated are large, making this multi-linear regression not statistically significant.

- StepAIC to be carried out to determine which independent variable contributes more to the variation of the dependent variable.

MULTI LINEAR REGRESSION TESTING

Additional step of using stepAIC to determine the best representation of the independent variables.

```
Start:  AIC=-157.82
allcountries$Fertility ~ allcountries$GDP + allcountries$LifeExp +
    allcountries$Edu

                        Df Sum of Sq    RSS      AIC
- allcountries$LifeExp   1    0.01870 0.65218  -158.65
<none>                                0.63348  -157.82
- allcountries$GDP       1    0.11264 0.74613  -153.27
- allcountries$Edu       1    0.81181 1.44529  -126.82

Step:   AIC=-158.65
allcountries$Fertility ~ allcountries$GDP + allcountries$Edu

                        Df Sum of Sq    RSS      AIC
<none>                                0.6522  -158.65
+ allcountries$LifeExp   1    0.0187 0.6335  -157.82
- allcountries$GDP       1    0.4181 1.0703  -140.84
- allcountries$Edu       1    4.6105 5.2626   -77.13
```

```
Call:
lm(formula = allcountries$Fertility ~ allcountries$GDP + allcountries$Edu)

Residuals:
      Min        1Q    Median        3Q       Max
-0.242267 -0.100206 -0.008088  0.095035  0.269085

Coefficients:
                    Estimate Std. Error t value  Pr(>|t|)
(Intercept)        7.842e+00  2.041e-01  38.419  < 2e-16 ***
allcountries$GDP   6.394e-05  1.313e-05   4.871  2.1e-05 ***
allcountries$Edu  -7.345e-02  4.541e-03 -16.173  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1328 on 37 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9694,     Adjusted R-squared:  0.9678
F-statistic: 586.3 on 2 and 37 DF,  p-value: < 2.2e-16
```

Analysis

- StepAIC eliminates Life Expectancy independent variable. Enabling the new multi-linear regression combination has a more stable AIC.
- According to the p-values on the new multi-linear regression, the intercept & slopes (GDP, Education) are all statistically significant.

CHI-SQUARE TESTING

H0: *There is no association.*
H1: *There is association.*

🔍 **Observation**

	Fertility Rate	
GDP	=< 2	>2
<= USD 11K	n = 19	n = 92
> USD 11K	n = 27	n = 15

```
              Pearson's Chi-squared test

data:  Matriz
X-squared = 32.242, df = 1, p-value = 1.361e-08
```

H_0 is rejected

There is association between fertility rate and GDP.

	Fertility Rate	
Life Expectancy	=< 2	>2
<= 70	n = 38	n = 71
> 70	n = 40	n = 36

```
              Pearson's Chi-squared test

data:  Matriz
X-squared = 5.7979, df = 1, p-value = 0.01605
```

H_0 is rejected

There is association between fertility rate and life expectancy.

	Fertility Rate	
Education Level	=< 2	>2
<= 85	n = 10	n = 10
> 85	n = 50	n = 31

```
              Pearson's Chi-squared test

data:  Matriz
X-squared = 0.91491, df = 1, p-value = 0.3388
```

H_0 is not rejected

There is no association between fertility rate and education level.

CONTENT

CONTENT

Conclusion

Advanced Countries

Linear regression for all factors were found to be significant. As these countries continue to progress, fertility stagnates around ~1.5 in 2050.

Developing Countries

Linear regression for all factors were found to be significant. As these countries advances, fertility rate drops exponentially.

All other Countries

Life expectancy is the least significant for the multi-linear regression model

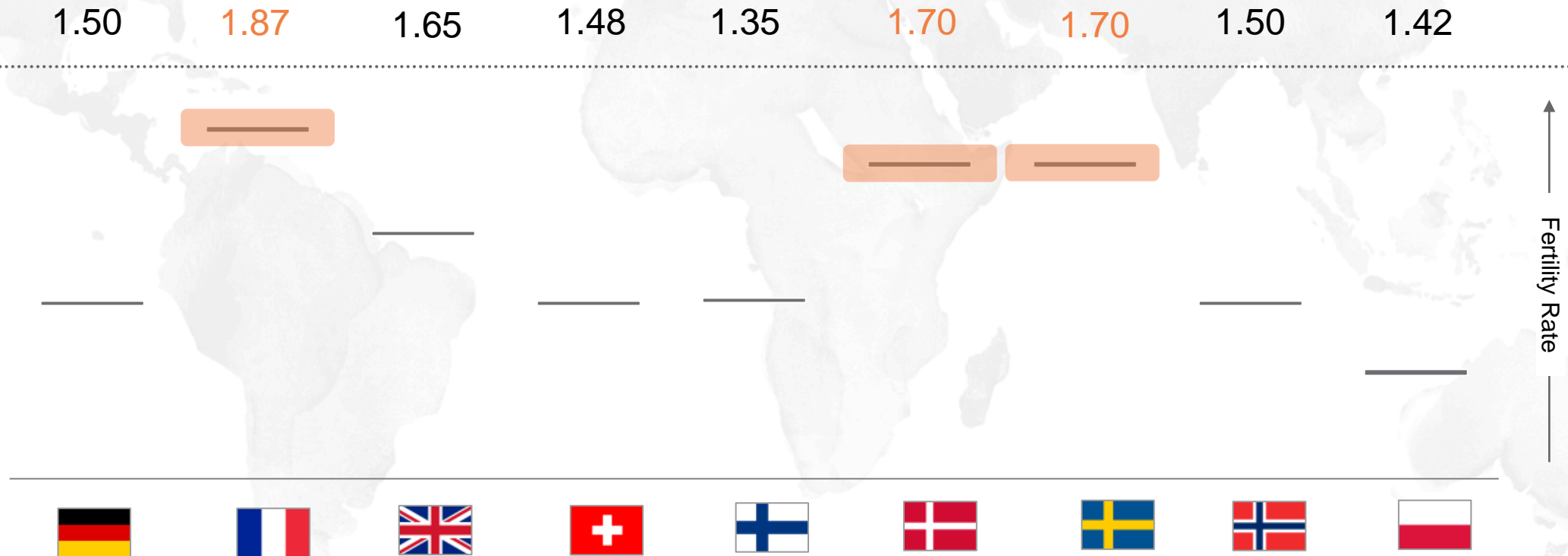GDP per capita and Life Expectancy have association with Fertility Rate

Education has no association with Fertility Rate

FUTURE WORK [1]

Study Policies of 'Model Countries'

To study and implement social policies of **France, Denmark and Sweden** who have relatively high fertility rate.

1.50   1.87   1.65   1.48   1.35   1.70   1.70   1.50   1.42

Fertility Rate

For year 2019

FUTURE WORK [2]

To include other variables for model to predict fertility rates

Fertility rate projection

1 — Other indicators for level of healthcare e.g. Infant mortality rate
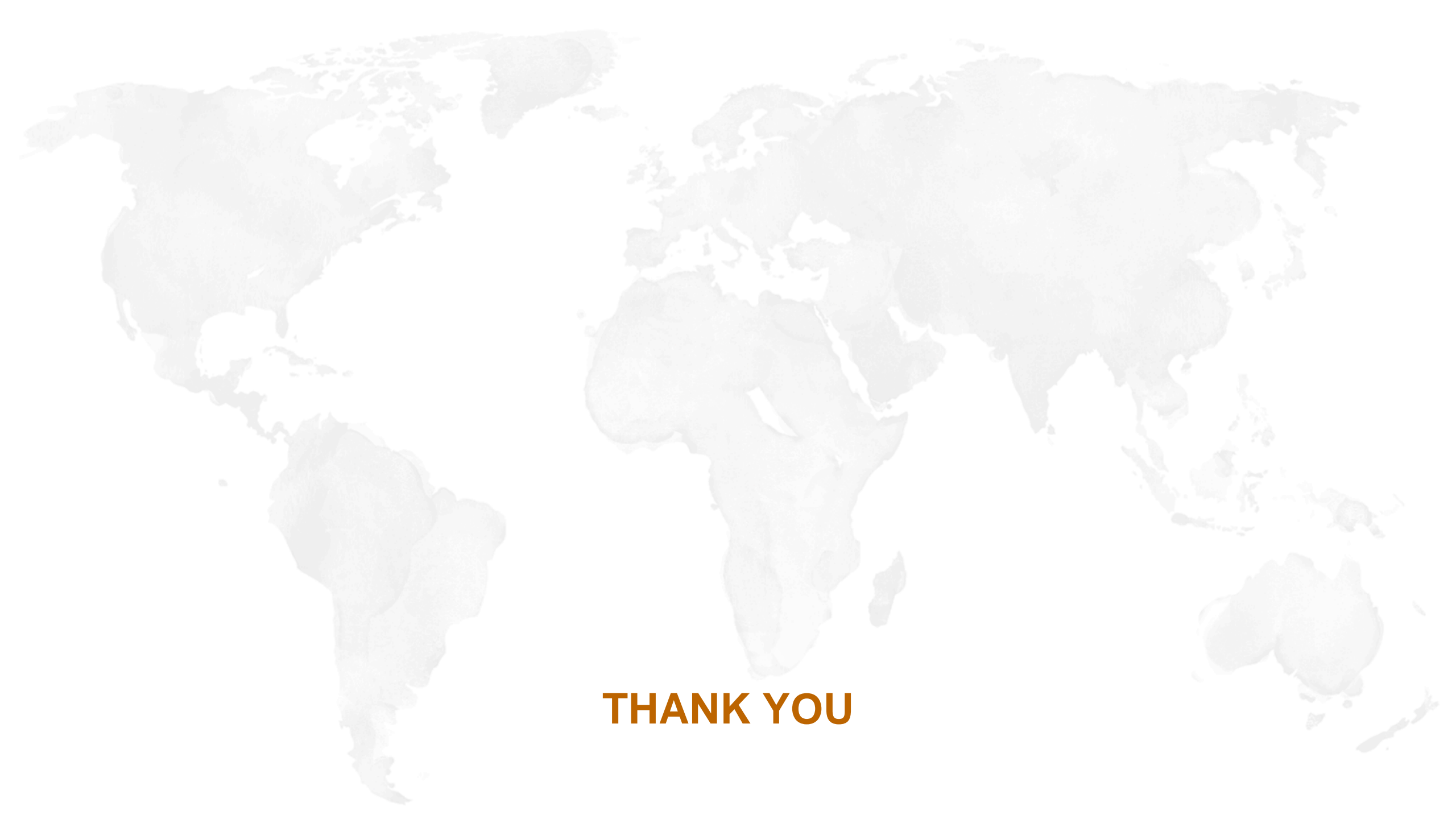
2 — Population Density

3 — Migration trends of countries

4 — Other indicators for standard of living

5 — Better indicators for level of education of workforce

THANK YOU