**The trend in world population development**

Initial questions:

* Will it keep on growing or will it stagnate?
* Will it decrease?
* What are the differences between countries and continents?
* Do you have some own theory of reasons behind differences, and can you justify it based on data?
* How will the age distribution change?

Our hypothesis:

* The overall trend in the world will be increasing in a diminishing tempo
* In Europe the trend will be downward sloping, in Asia upward sloping and in Africa also increasing
* China will remain the most populated country in the world
* The general life expectancy will rise

Work partition:

* Each member will analyse data for a certain region of the world: western Asia, western Europe, western Africa, Sahara
* data for population between 2000 and 2022

Open points:

* Do we have enough observations for net migration, or do we exclude it? –
  + Net migration was only observed in 2002, 2007, 2012 and 2017
* Net migration does not affect population growth (but which countries we have?)
* Should British and French oversee territories like Cayman Island, Curacao, French Polynesia, Sint Maarten,
* Virgin Islands and similar be included? Are they included in the data for England/France?

Observations during the process:

* Noticed that we are missing data for birth rate and life expectancies for 2020 – haven’t figured out why
* There are few countries in which there was a sudden increase in population growth between 2004 and 2010 (would be nice to find which ones and why it happened)
* Birth rate and population growth are positively correlated
* Life expectancy does not seem to influence population growth (if it does than it is slightly negative)
* Life expectancy is increasing over time everywhere
* With increasing life expectancy the birth rate declines
* Birth rate is declining in all of the regions
* Sub-Saharan Africa has by far the highest birth rate, but it is decreasing rapidly. At the same time it has the lowest life expectancy which increased by more than 10 years in the last 20 years
* The immigrants who leave their region (not only the country) are going to Europe and Central Asia, North America and the rest of the regions have negative net migration (South Asia the most)

TO-DO-LIST

* Start with a final presentation (ppt)
* Start with a final project report (pdf)
* Analyse regions
* Compare regions (and countries but we have to decide which ones)
* Figure out how to present the relationship between GDP and population growth

**Executive Summary**

* **Data Acquisition - Timelines and Metrics**
  + **Data Source:** ‘The World Bank’ data base with information on ‘Health Nutrition and Population statistics’
  + **Timelines:** Considered from 2000 to 2021
  + **Countries:** All Countries in the world
  + **Key Metrics:** Birth Rate, Life expectancy, Gender, Population Growth, Net Migration, GDP
* **Data Cleaning** 
  + Removing unnecessary rows and columns
  + Two datasets combined based on ‘Country and Year’
  + Two subsets of data are created - regions and countries separately
* **Data Transformation**
  + Datatype conversion
  + Panel Data Frame
  + Individual metric standardization
* **Data Analysis**
  + Noticed that we are missing data for birth rate and life expectancies for 2020 –haven’t figured out why
  + There are few countries in which there was a sudden increase in population growth between 2004 and 2010 (would be nice to find which ones and why it happened)
  + Birth rate and population growth are positively correlated
  + Life expectancy does not seem to influence population growth (if it does than it is slightly negative)
  + Life expectancy is increasing over time
* **Insights**
  + High-level trends on Birth Rate, Life expectancy, Population Growth, Net Migration, and GDP
  + Gender level differentiation on these metrics
  + Based on the high-level insights, deep dive into interested regions and compare with other regions
  + From Regions, deep dive into countries that are significant to the metrics
  + **Ultimately to answer how the metrics evolved from 2000 to 2021 and compare between regions and countries.**