ACADEMY :

Education statistics exploration to determine markets abroad.

december 2022 – Julien Le Boucher

Goals

- Determine a list of countries with a lot of potential clients for Academy.
- Assess evolution in those countries.

- Priority : where to operate first ?

Plan

1) Data contents and quality.

2) Data selection for score computing.

3) Markets ranking and comparisons.

4) Projections.

5) Conclusions.

The material

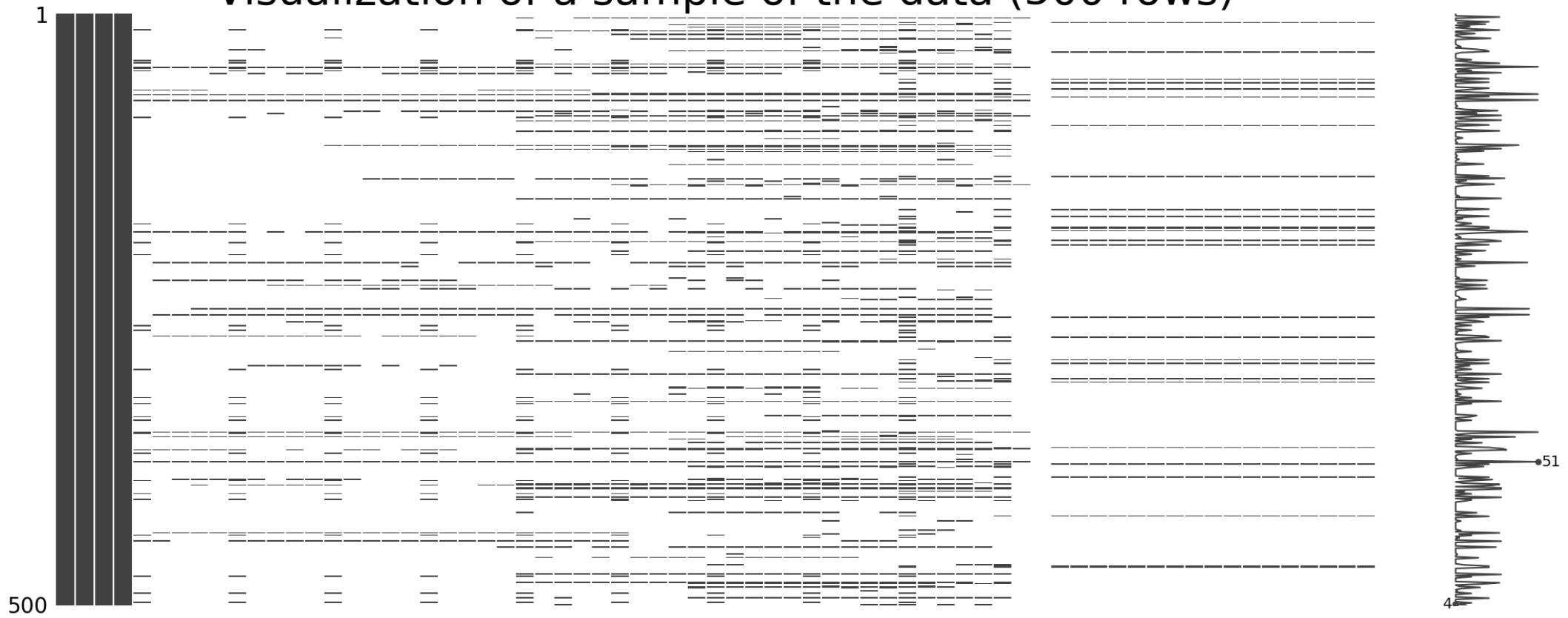
5 CSV files from the world bank open data.

- Global information about countries ;
- Information about observed indicators ;
- Description/source of series ;
- footnotes ;
- The main data.

Data familiarization

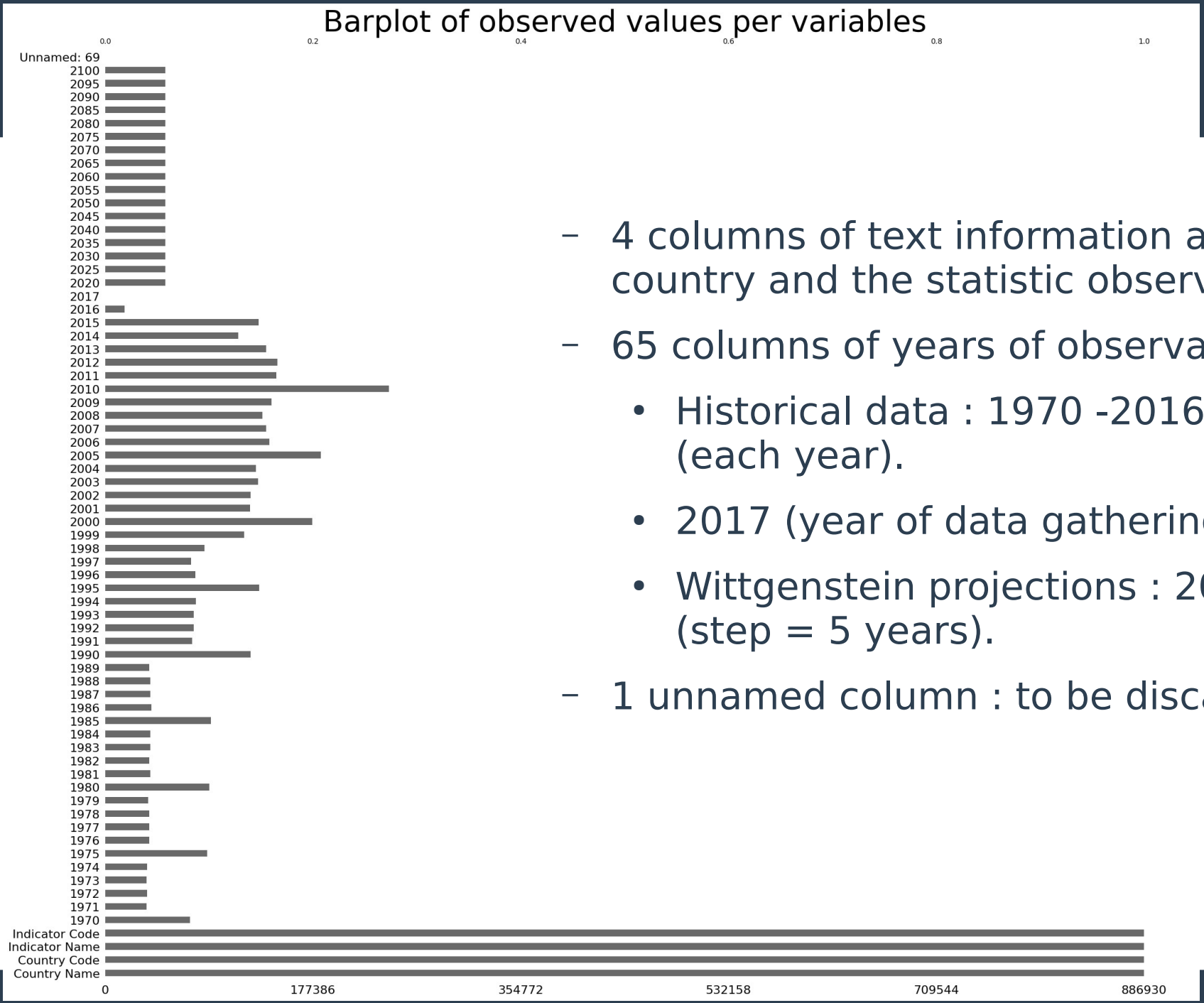
- Data matrix shape : 886 930 rows x 70 columns.

Visualization of a sample of the data (500 rows)



- Percentage of missing data : 86 %.

Observed values and columns meaning



- 4 columns of text information about the country and the statistic observed.
- 65 columns of years of observation :
  - Historical data : 1970 -2016 (each year).
  - 2017 (year of data gathering).
  - Wittgenstein projections : 2020-2100 (step = 5 years).
- 1 unnamed column : to be discarded.

Data set quality

- No duplicates found filtering on (country / indicator name) in text columns.
- Some duplicates where found in time-series

  → seemed reasonable.

  examples :

  - « Theoretical duration of pre-primary education (years) » can be the same in many countries.
  - An observation on the global population was also assigned to the female population.

Data set quality conclusion

- Well-sourced data.
- High percentage of emptiness.
- No obvious duplicates.
- Last historical data observed : 2016.

→ Select **well-provided indicators** (at least 5 values observed in the 2010-2016) and **relevant to Academy's business.**

Relevant for Academy ?
3 indicators to build.

- **Potential clients number :**
  - People in secondary or tertiary education.
  - Graduated people.
- **Access to technology :**
  - With a knowledge of computer use.
  - With access to the internet.
- **Financial wealth :**
  - With a strong purchasing power.

Assessment of the number of potential clients per country

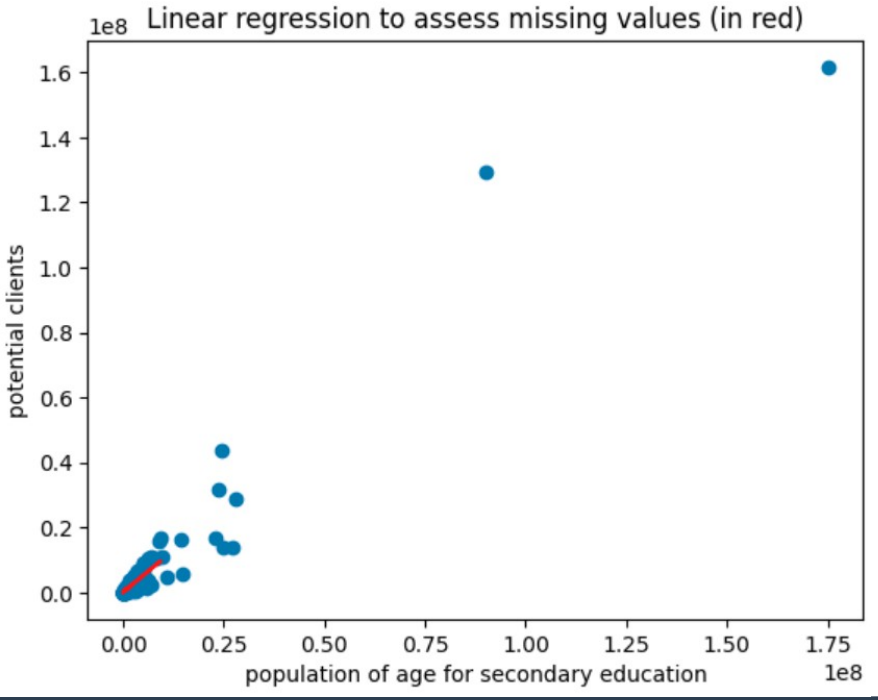
When available for the country, Sum of two indicators :

“Enrolment in secondary education, both sexes (number)”
“Enrolment in tertiary education, all programmes, both sexes (number)”
(177 countries)

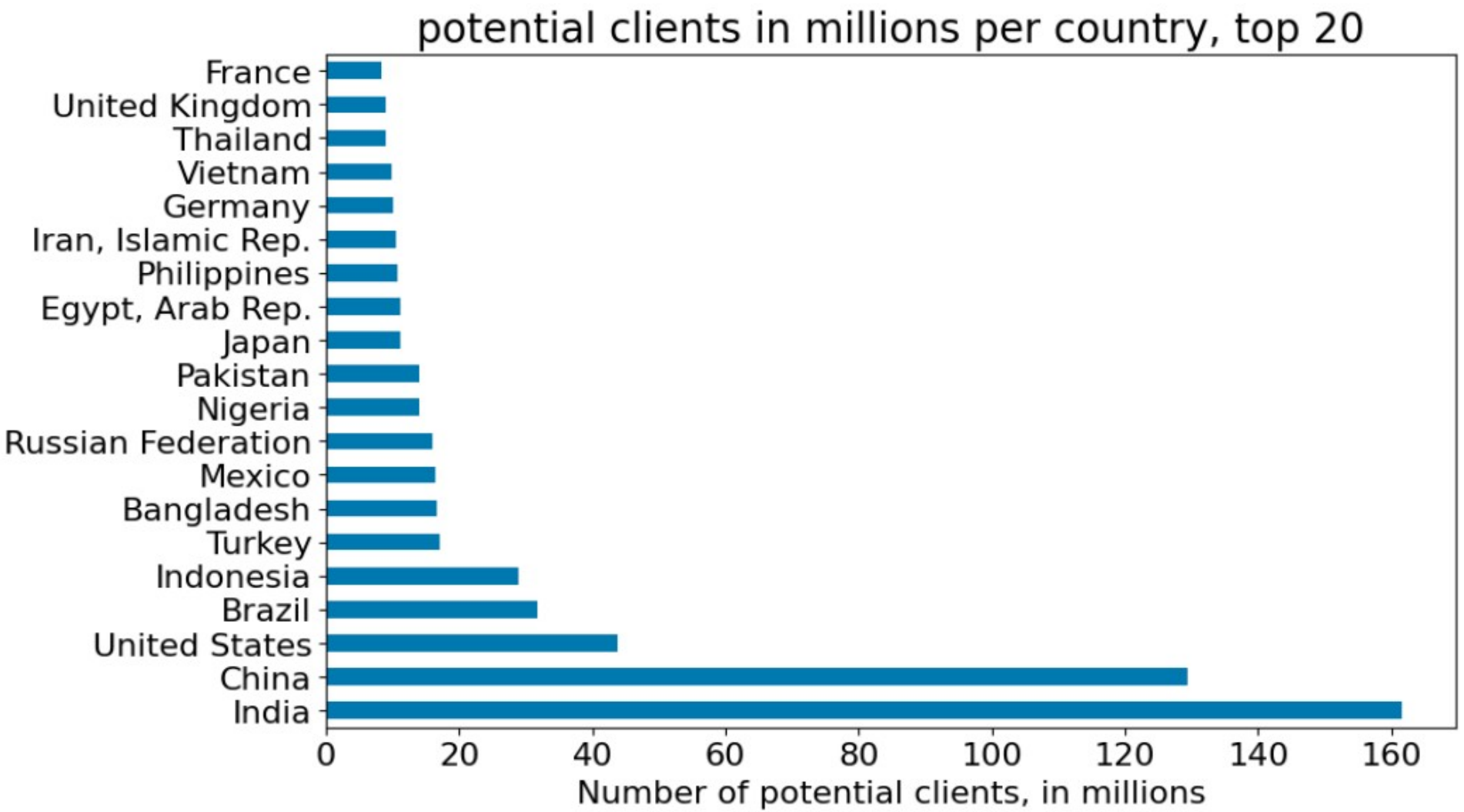
If not available, linear regression model :

between previous potential clients number and a highly correlated indicator : “Population of the official age for secondary education, both sexes (number)”

(+20 countries)



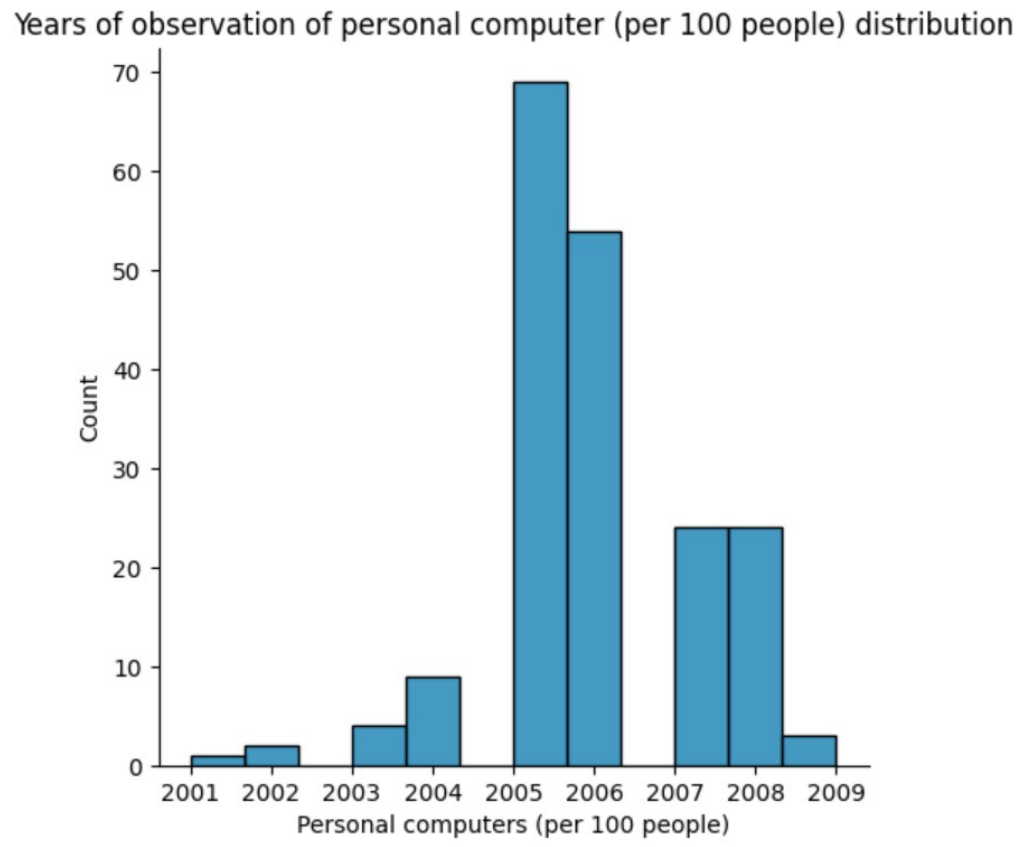
Potential clients : Top 20



Technology skills per country

Unfortunately, the data set does not allow the assessment of technological skills in a country. Some indicators are available, but they were observed before 2010, or only on very few countries :

- “Personal computers (per 100 people)”
- “PIAAC: Adults by proficiency level in problem solving in technology-rich environments (%).”



Access to technology per country

Indicator used :

“Internet users (per 100 people)”

Definition :
individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.

Financial Wealth per country

Indicator used :

“GDP per capita, PPP (current international $)”

Definition :
gross domestic product per capita, based on purchasing power parity.
It uses the prices of specific goods to compare the absolute purchasing power and readjust the standard GDP.

→ A Boost for emerging countries in the ranking.

synthesis

Country Name	potential_clients	access_to_technology	financial_wealth
India	1.616495e+08	29.547163	1709.591808
China	1.294946e+08	53.200000	8123.180873
United States	4.376150e+07	76.176737	57638.159088
Brazil	3.178726e+07	59.682747	8649.948492
Indonesia	2.886432e+07	25.366301	3570.294888
Turkey	1.703206e+07	58.347734	10862.600399
Bangladesh	1.663513e+07	18.246938	1358.779029

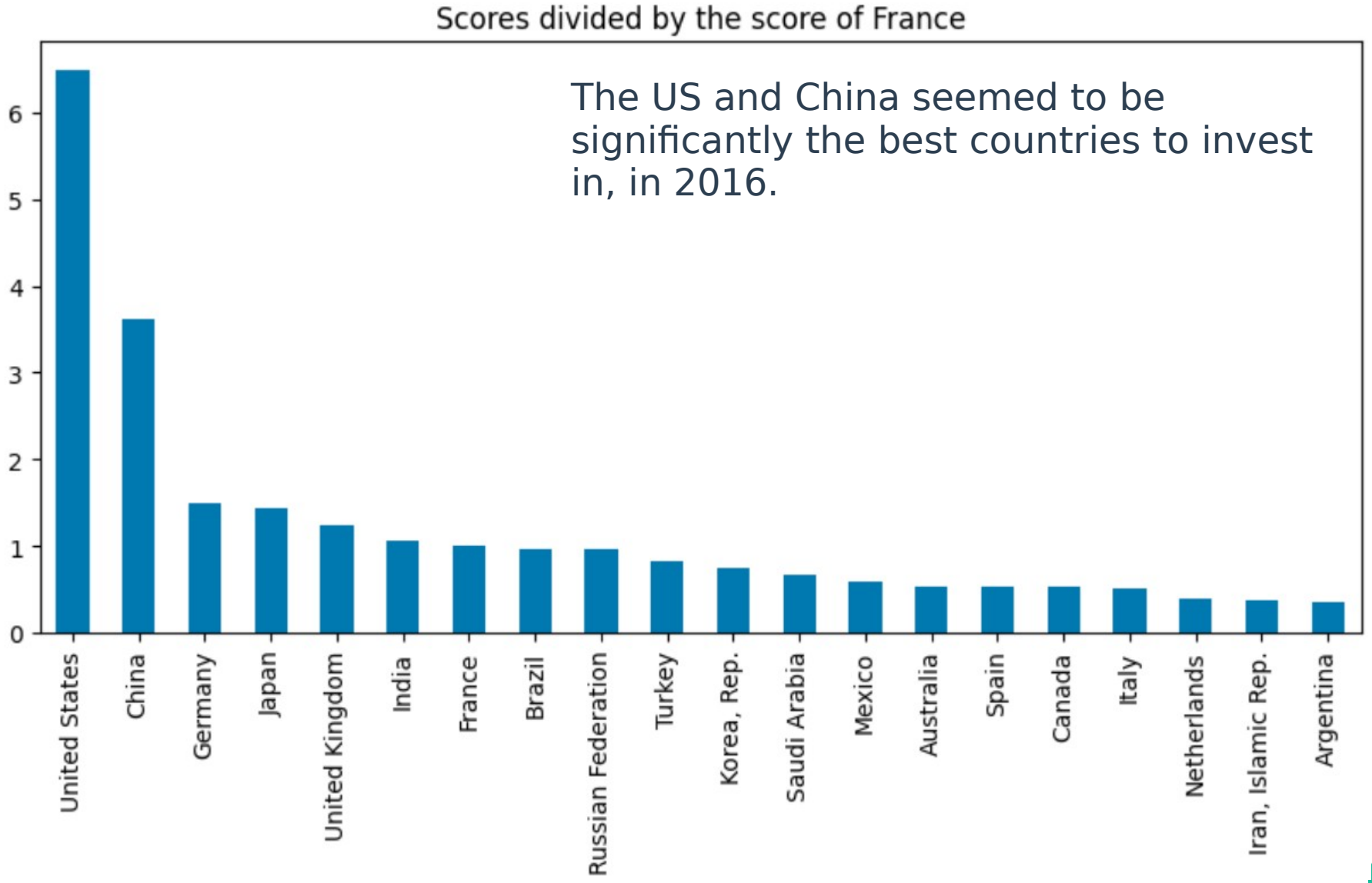
Values provided for 194 countries.

Scoring function

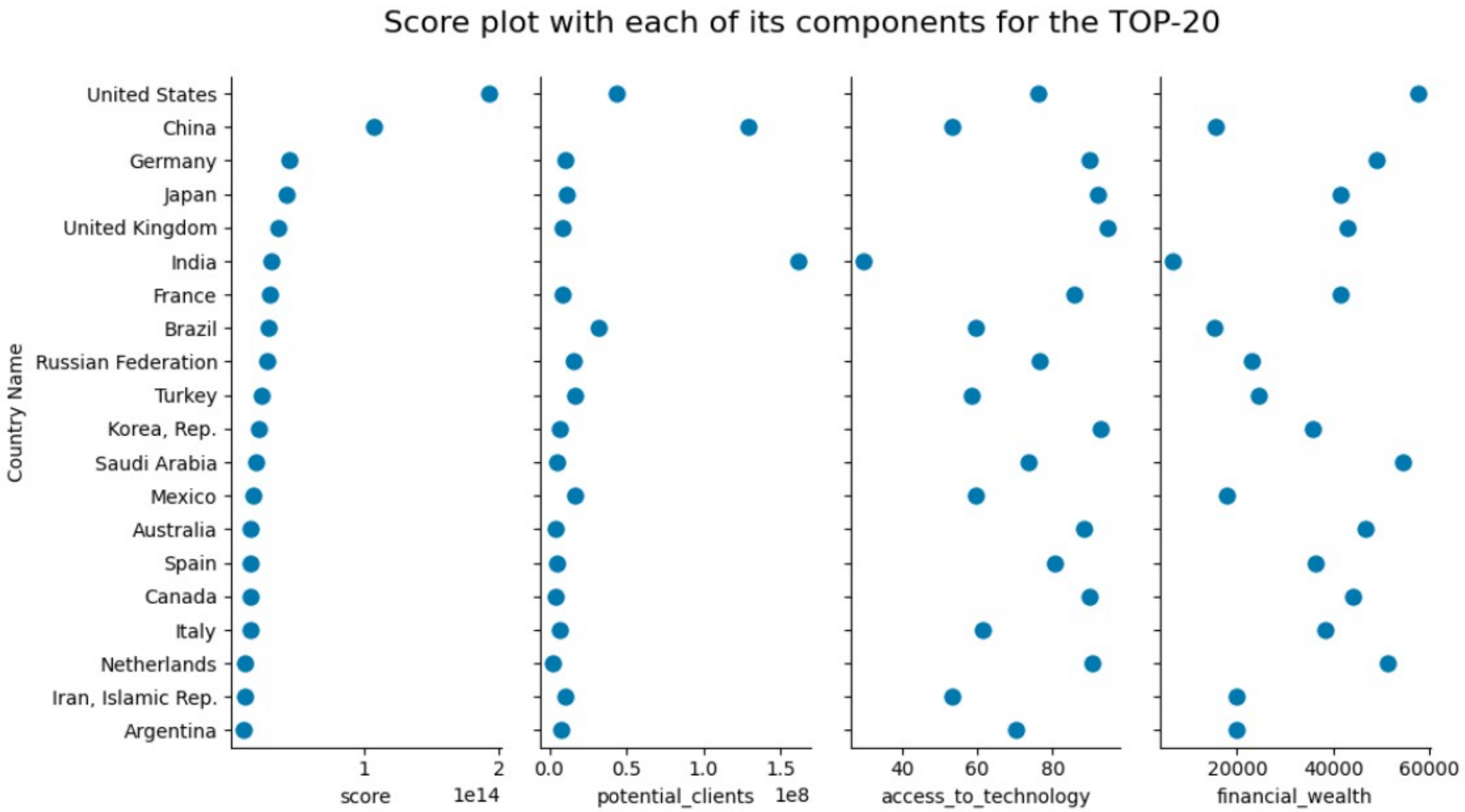
Score = potential_clients * access_to_technology * financial_wealth

Score and $ are homogeneous.

Results (2016)



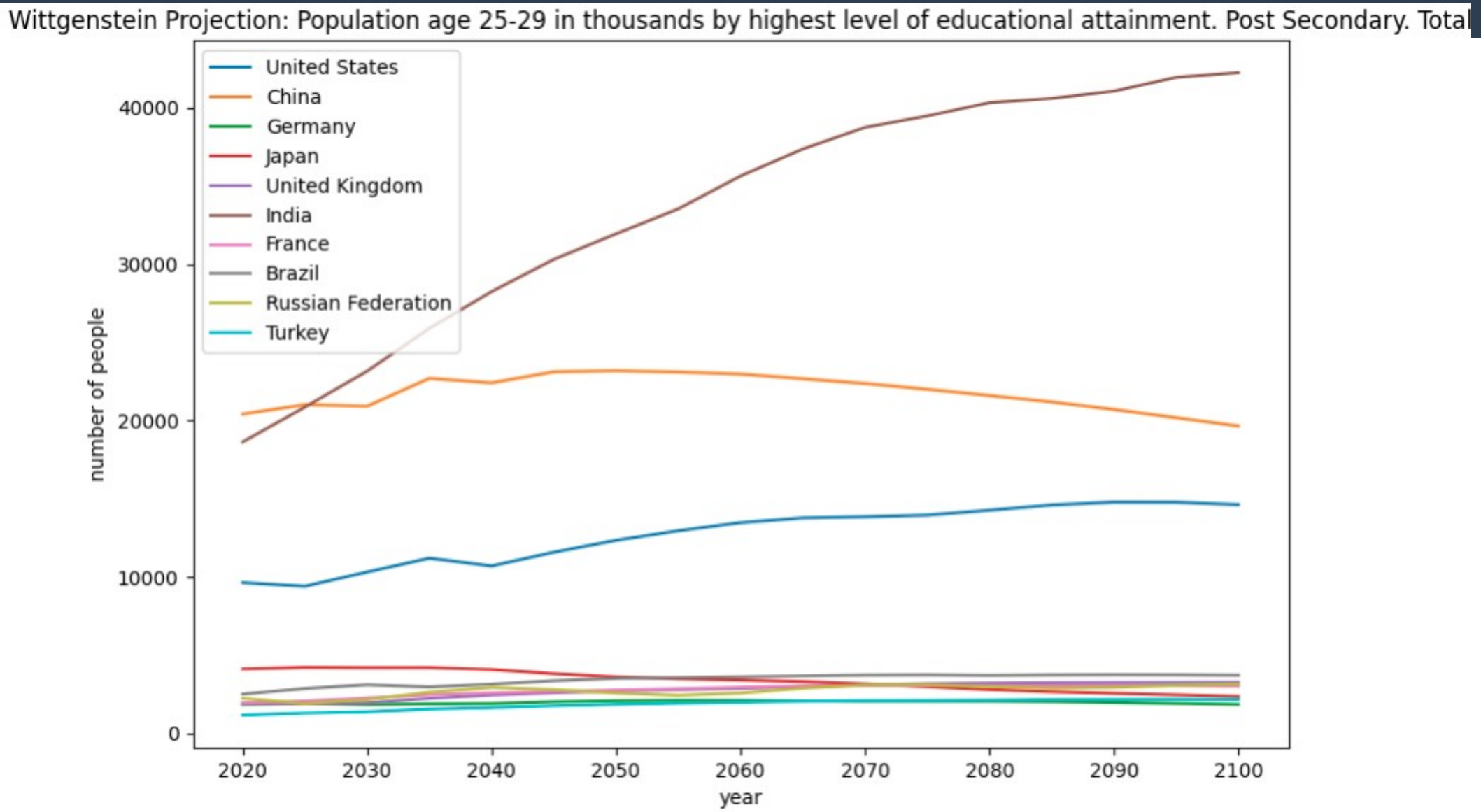
Results (2016)



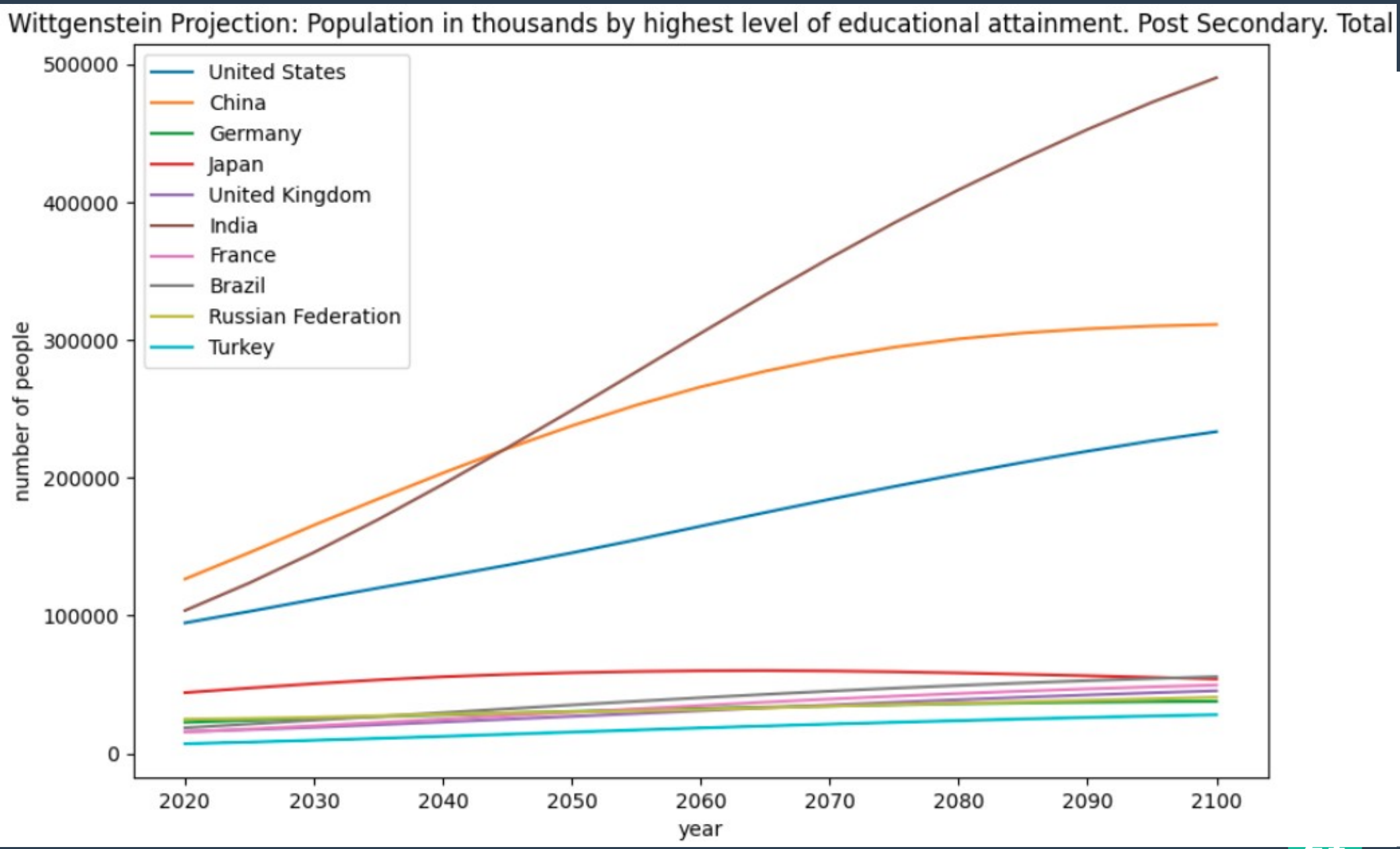
Projections

- Since 2016, evolutions might already have been huge for some countries regarding the access to technology.
- It is probably going to converge toward 100 % for many countries quite fast. Having less impact on the ranking in the future.
- Clients number and financial wealth must be looked at projection-wise.
- The data only provide projections of clients number.

Number of well-educated people in the age range 25-29 per country



Number of well-educated people in the global population per country



Conclusion

- No risk of decrease in the top 10 countries concerning high-graduated people number. **The US (1) and China (2) are solid investments** because the demographic projections confirm important growths and they already are ahead.

- If **India** was 6th in the score ranking in 2016, It **might become the first in 20 years or so.** Indeed, the internet infrastructures are going to evolve, for example with the 5G, and the demographic gap will be an advantage.

- Brazil could also become important at some point.