

18/12/2020 – DAFT – Final project

Life expectancy across the globe

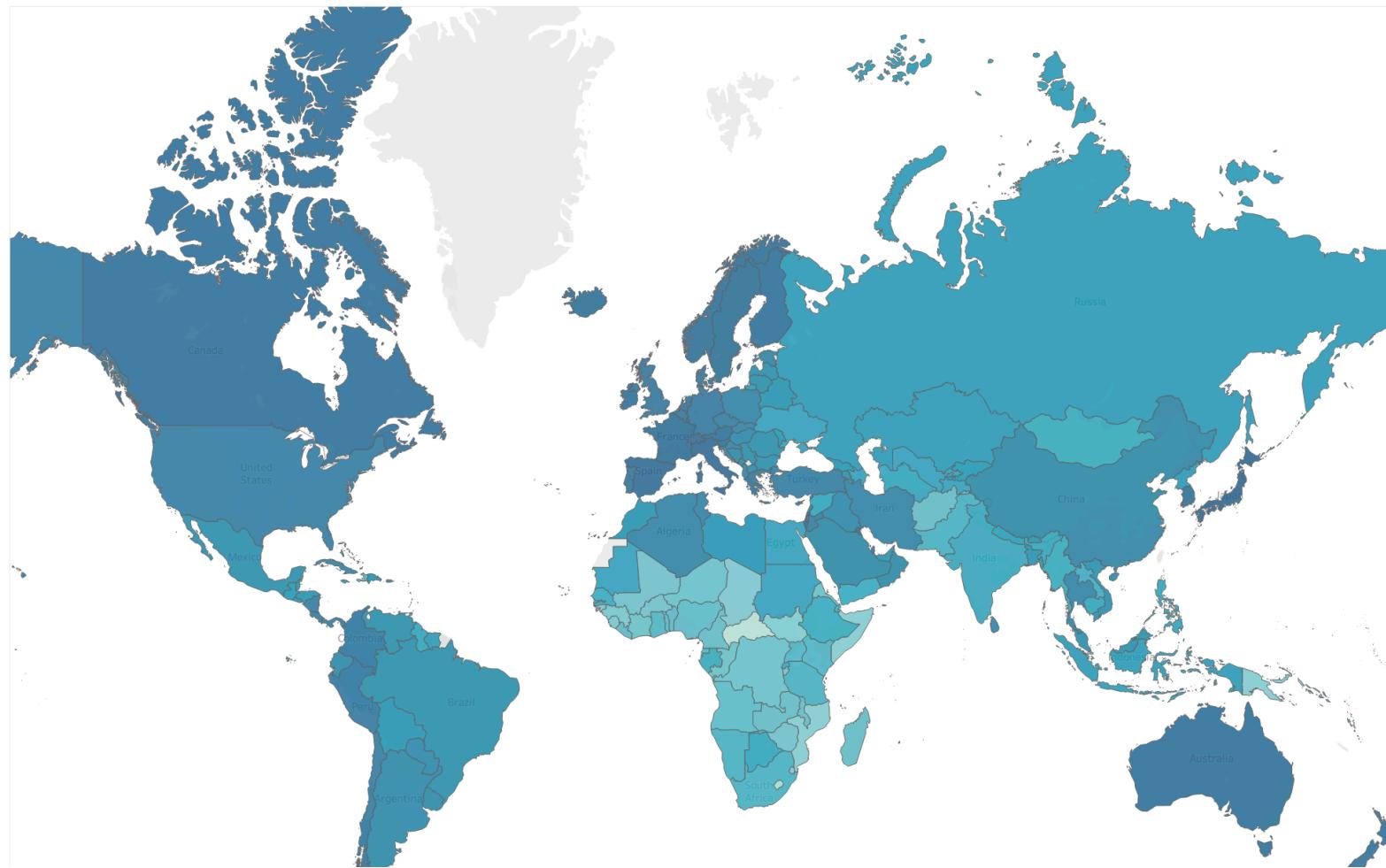
Jamil Akili



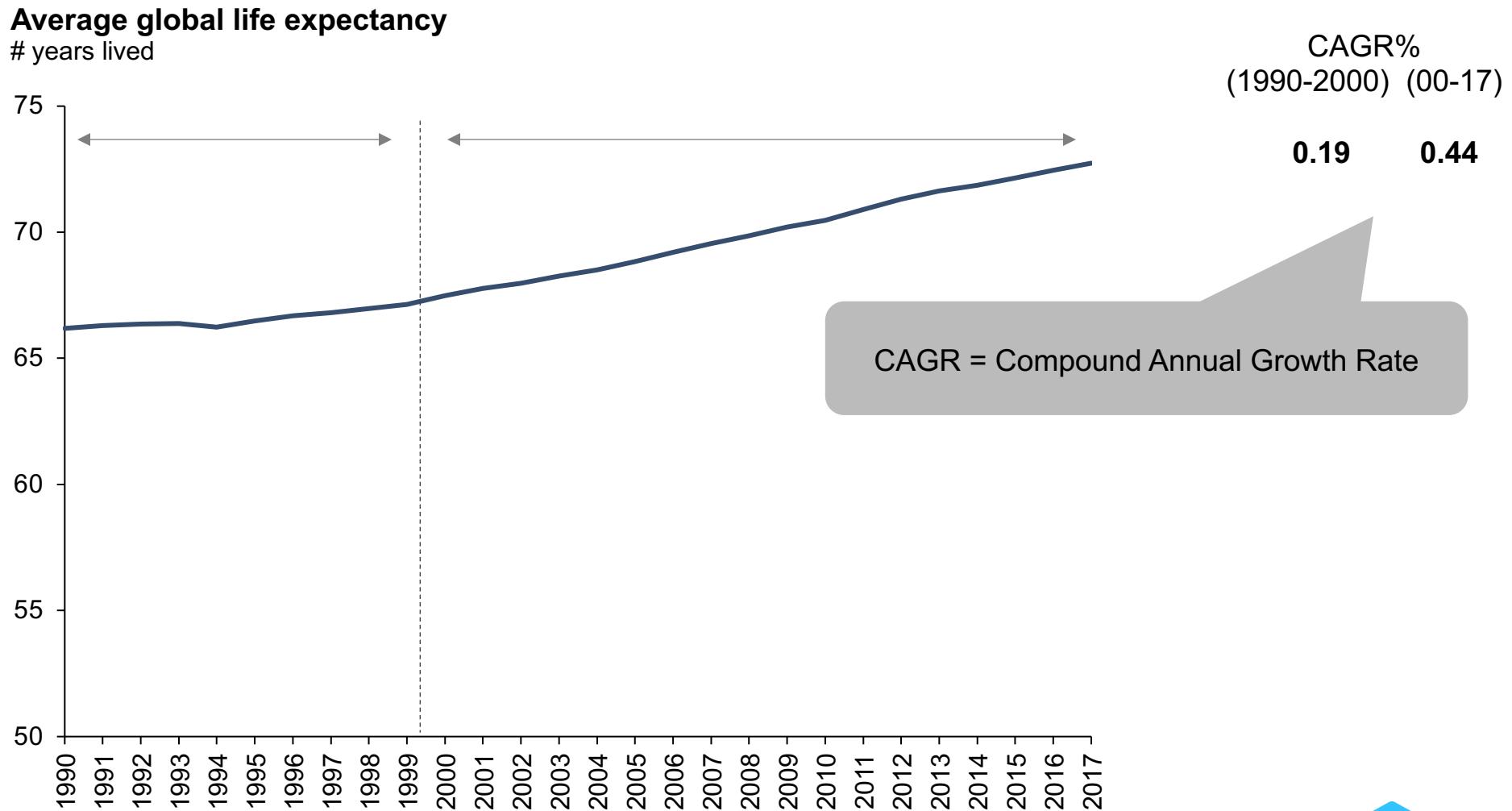
Global life expectancy values today mostly reflect the gap between developed and developing countries

Average life expectancy (2017)

(years) 51.90 84.80



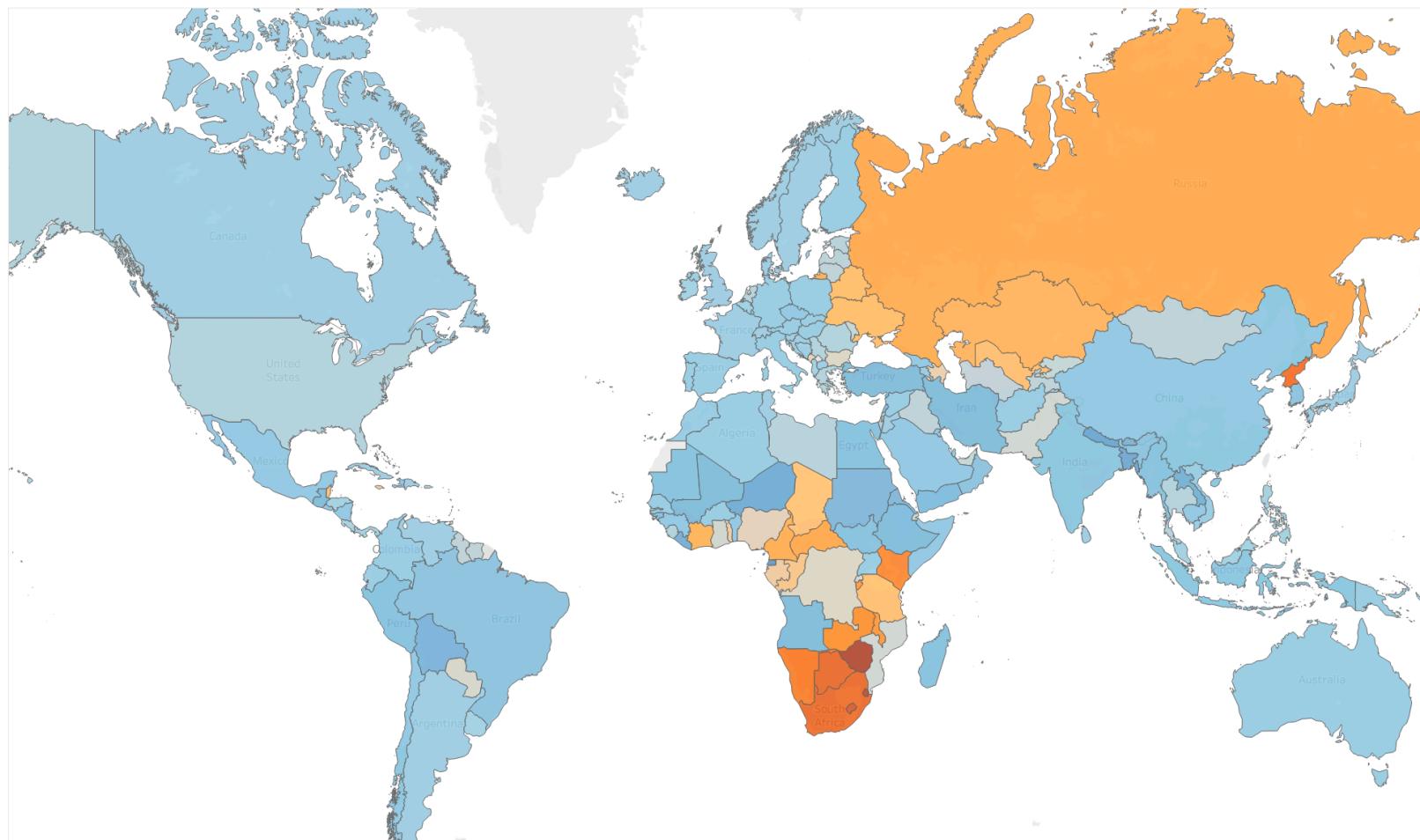
Global average life expectancy has been growing at c.0.5% p.a. since 2000 after almost stagnating in the 90-ies



During the 1990s, average life expectancy was still decreasing in many countries...

Evolution of life expectancy
1990 - 2000

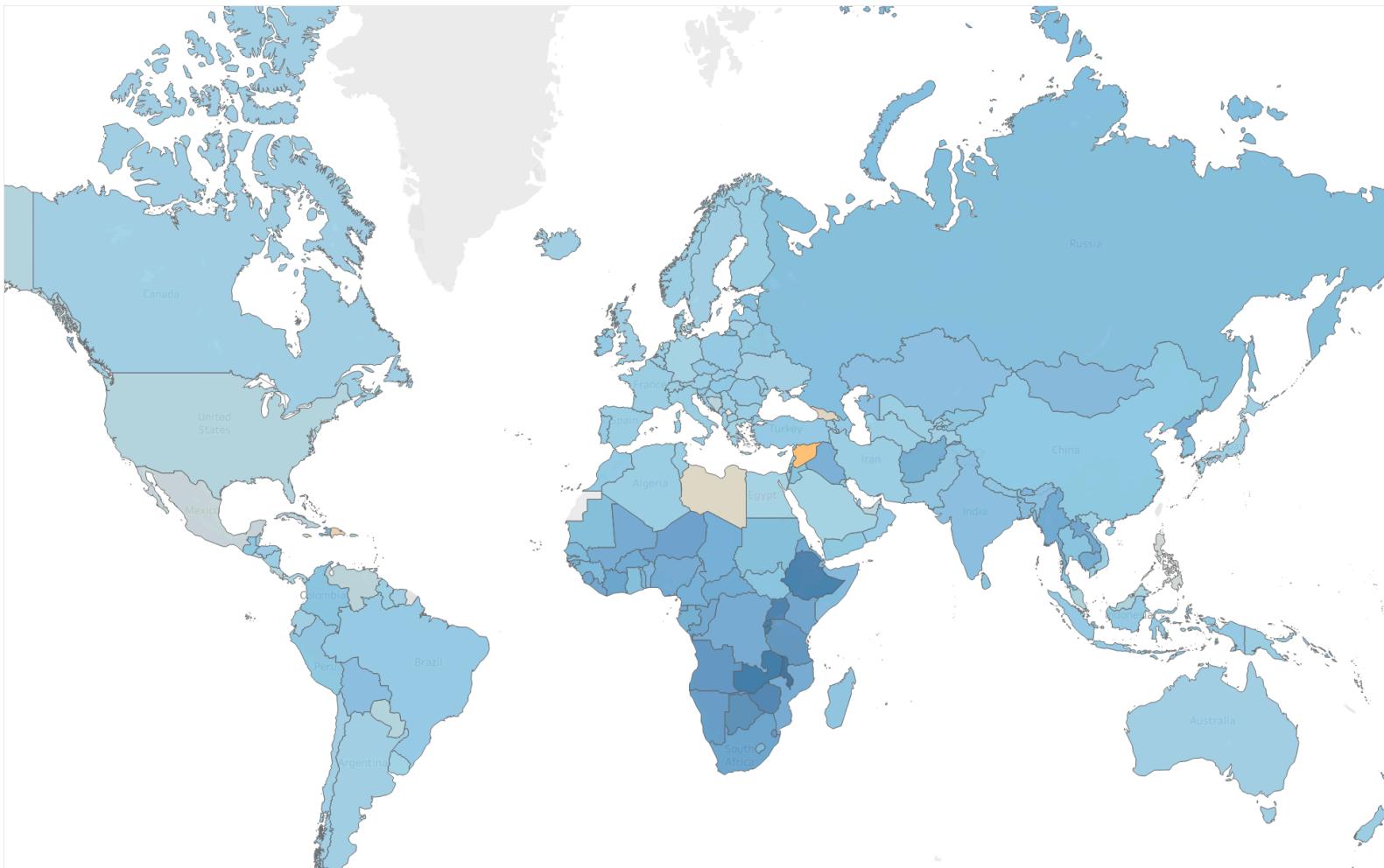
% CAGR
-2.641 1.417



... but the trend has been positive across the globe since 2000

Evolution of life expectancy
2000 - 2017

% CAGR
-0.223 2.072



Project goal

What are the main factors influencing life expectancy?

Is it possible to use machine learning to predict life expectancy in a given country?

Alternative: building a dataset from scratch and running a multiple linear regression on it

- Constructing dataset from different sources, with a focus on 1990-2017
- Merged datasets used include:
 - Life expectancy, total population, GDP per capita
 - Access to basic sanitation, access to water, doctors per 1 000, child mortality
 - HIV, measles, TB



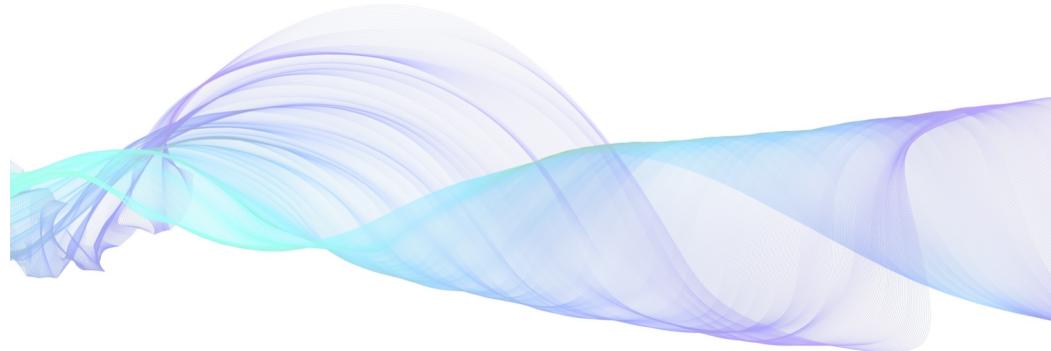
- New dataset (5460 rows × 12 columns)
- Good accuracy for regression models

	model	R2	R2A	MSE	RMSE	MAE
0	linreg	0.90	0.87	8.11	2.85	2.07
1	lasso	0.90	0.87	8.29	2.88	1.99
2	ridge	0.91	0.88	7.69	2.77	2.00
3	elastic	0.90	0.87	8.30	2.88	1.99
4	rfr	0.77	0.70	18.57	4.31	2.41

Key takeaways

- Always pay close attention to data source and collection method (time series)
- Relatively limited influence on life expectancy of indicators related to wealth and economic development (i.e. *GDP per capita, doctors available per 1000 pop*)
- Much higher impact for simpler factors (e.g. *access to basic sanitation, access to clean water*) : UN programs are efficient
- Explore the impact of other factors like *schooling level, poverty/inequality index, or public funding of healthcare* → **next step**
- Finalize database by assigning data to missing values (depending on the number of values already available for a given year & country) → **next step**

Thank you!



References:

- Kaggle : <https://www.kaggle.com/kumarajarshi/life-expectancy-who>
- Gapminder : <https://www.gapminder.org/data/>
- OECD : <https://data.oecd.org/healthstat/life-expectancy-at-birth.htm>
- WHO : <https://www.who.int/data/gho/gho-search>