

# HDI Analysis: 20 Countries (1990 – 2022)

Sinan KAYMAZ

Ayşe GÜLTEPE

Ahu YILMAZ



Big Data  
Unites  
Sciences  
and Arts



Co-funded by the  
Erasmus+ Programme  
of the European Union



The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## TÜRKİYE-2025

Co-funded by the  
Erasmus+ Programme  
of the European Union



# Table of Contents

- ▶ Dataset & Preprocessing
- ▶ Summary Statistics
- ▶ Missing Data Analysis
- ▶ Correlation Analysis
- ▶ Key Correlated Features
- ▶ Regression Modeling
- ▶ Top 5 HDI Performers (2022)
- ▶ HDI Trends
- ▶ Prediction & Forecast
- ▶ Economic vs HDI Analysis
- ▶ HDI Distribution
- ▶ Boxplot and Whisker
- ▶ Average HDI Over Time
- ▶ Life Expectancy vs HDI Scatter
- ▶ Multidimensional Poverty
- ▶ Policy Recommendations
- ▶ Conclusion & Next Steps

# Dataset & Preprocessing

- ▶ Source: UNDP Human Development Reports ARFF file (1990–2022)
- ▶ 20 countries, 660 annual observations, 44 attributes
- ▶ Decoded categorical fields; handled missing values
- ▶ 2022 subset: dropped nulls in GNI, Life Expectancy, Mean Schooling, HDI
- ▶ Tools: MS Excel, Notepad++, Python (pandas, sklearn, matplotlib), ChatGPT, Tableau Public, WEKA

# Preferred Countries

Preferred Countries

2022

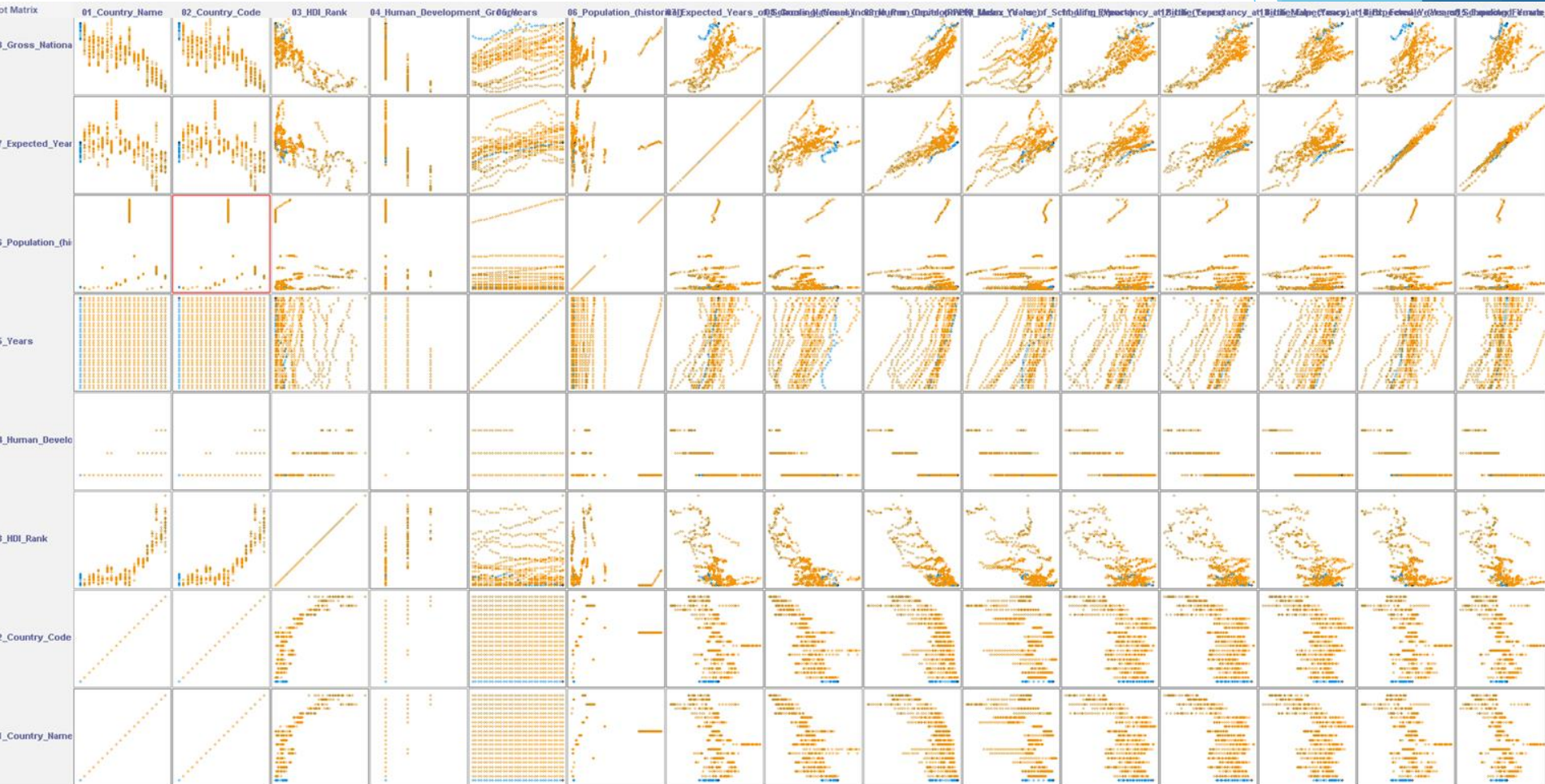


© 2025 Mapbox © OpenStreetMap

Australia Bulgaria Canada Croatia Denmark Estonia Germany  
Japan Iceland Ireland Italy Singapore Spain Sweden  
Netherlands Türkiye Switzerland United States  
Norway Ukraine Austria



# All Attributes Visualize



# Summary Statistics (1990–2022)

Statistic	03_HDI_Rank_(2022)	05_Years	06_Population_(historical)	07_Expected_Years_of_Schooling_(Years)	08_Gross_National_Income_Per_Capita_(PPPS)	09_Human_Development_Index_(Value)	10_Mean_Years_of_Schooling_(Years)	11_Life_Expectancy_at_Birth_(Years)	12_Life_Expectancy_at_Birth_Male_(years)	13_Life_Expectancy_at_Birth_Female_(Years)	14_Expected_Years_of_Schooling_Female_(Years)
count	660	660	660	660	655	655	660	660	660	660	660
mean	21,93	2006	42499719	15,97	40706,08	0,87	11,19	78,32	75,38	81,22	16,27
std	22,41		67858576	2,24	16708,93	0,07	2,02	3,96	4,74	3,23	2,50
min	1,00	1990	254844	8,92	6837,22	0,60	4,46	66,83	61,18	71,41	7,77
0,25	6,00	1998	4833457	14,94	30163,75	0,83	10,39	76,20	73,10	79,34	14,92
0,50	12,00	2006	12759109	15,95	41403,03	0,88	11,55	78,94	76,24	81,60	16,40
0,75	30,00	2014	53266498	17,03	51737,20	0,92	12,64	81,46	79,14	83,62	17,41
max	100,00	2022	341534041	23,25	88761,15	0,97	14,26	84,82	83,17	87,83	23,58

# Summary Statistics (1990–2022)

Statistic	17_Gross_National_Income_Per_Capita_Female	18_Gross_National_Income_Per_Capita_Male	19_HDI_Female_(Value)	20_HDI_Male_(Value)	21_Mean_Years_of_Schooling_Female_(Years)	22_Mean_Years_of_Schooling_Male_(Years)	23_Adolescent_Birth_Rate_(Births_Per_1000_Women_Ages_15_19)	24_Gender_Inequality_Index_(value)	25_Labour_Force_Participation_Rate_Female_(Percentage_Ages_15_and_Older)	26_Labour_Force_Participation_Rate_Male_(Percentage_Ages_15_and_Older)	27_Maternal_Mortality_Ratio_(Deaths_Per_100000_Live_Births)
count	650	650	650	650	660	660	660	655	655	655	660
mean	30615,91	51484	1	0,88	10,98	11,42	16,40	0,16	53,77	70,21	10,56
std	14153,28	19556	0	0,07	2,22	1,85	14,96	0,11	10,91	6,67	8,48
min	5447,36	8312	1	0,64	3,28	5,64	1,81	0,01	23,29	43,10	1,66
0,25	21388,87	39471	1	0,84	10,00	10,61	6,08	0,08	47,80	66,38	5,44
0,50	30047,06	53623	1	0,90	11,45	11,80	10,18	0,13	55,65	70,68	7,63
0,75	38984,05	64897	1	0,93	12,62	12,73	20,56	0,21	59,81	74,20	11,67
max	82518,17	108424	1	0,98	14,14	14,54	69,92	0,60	79,77	88,45	52,40

# Summary Statistics (1990–2022)

Statistic	31_Population_With_at_Least_Some_Secondary_Education_Male_(Percentage_Ages_25_and_Older)	32_Carbon_dioxide_Emissions_Per_Capita_(Production)_Tonnes	33_Difference_From_HDI_Value_(Percentage)	34_Material_Footprint_Per_Capita_(Tonnes)	35_Planetary_Pressures_Adjusted_Human_Development_Index_(Value)	36_Age_standardized_deaths_from_malignant_neoplasms_in_both_sexes_per_100000_people	37_Age_standardized_deaths_from_cardiovascular_diseases_in_both_sexes_per_100000_people	38_Age_standardized_deaths_from_respiratory_infections_in_both_sexes_per_100000_people	39_Age_standardized_deaths_from_infectious_and_parasitic_diseases_in_both_sexes_per_100000_people	40_Age_standardized_deaths_from_digestive_diseases_in_both_sexes_per_100000_people	41_Age_standardized_deaths_from_unintentional_injuries_in_both_sexes_per_100000_people
count	660	660	651	654	651	613	613	613	613	613	613
mean	84,07	10	15	24,75	0,73	13,83	25,90	23,26	13,36	24,76	22,81
std	15,28	4	6	10,33	0,06	24,27	27,52	17,37	8,67	13,52	9,08
min	24,92	3	5	3,48	0,54	2,06	7,26	2,68	2,56	1,25	6,07
0,25	77,29	6	11	16,86	0,69	5,78	13,99	13,59	9,00	15,24	16,03
0,50	89,79	9	15	24,49	0,73	7,32	18,95	18,68	11,53	23,16	20,53
0,75	95,05	11	18	32,06	0,78	9,12	26,94	25,56	14,71	34,02	29,18
max	100,00	24	36	78,24	0,84	147,67	212,37	145,98	88,55	78,85	51,88

# Missing Data Analysis

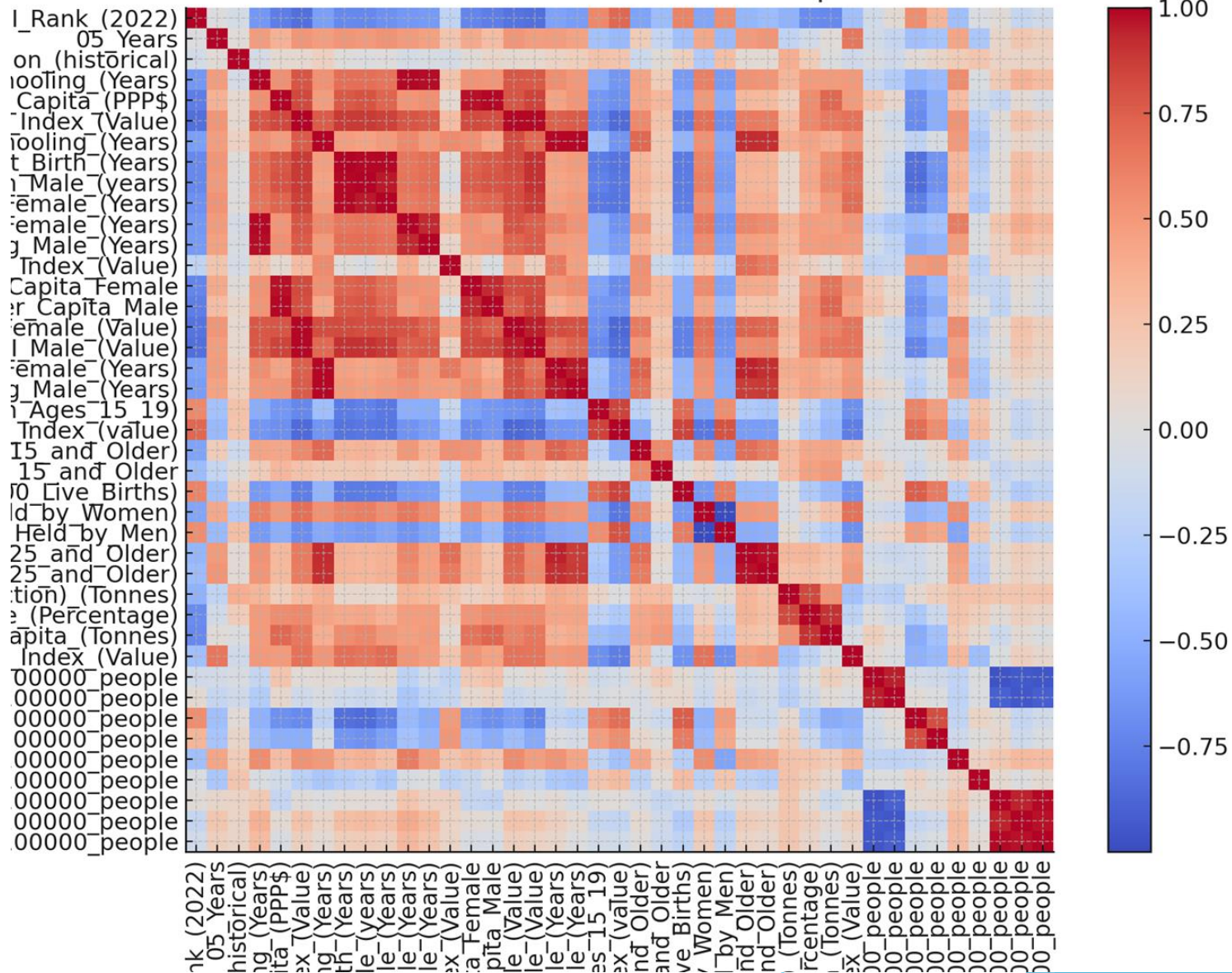
## Top 6 Attributes with Most Missing Values:

- ▶ 36 Age-standardized deaths from malignant neoplasms in both sexes per 100.000 people : 47 missing
- ▶ 37 Age-standardized deaths from cardiovascular diseases in both sexes per 100.000 people : 47 missing
- ▶ 38 Age-standardized deaths from respiratory infections in both sexes per 100.000 people : 47 missing
- ▶ 39 Age-standardized deaths from infectious and parasitic diseases in both sexes per 100.000 people : 47 missing
- ▶ 40 Age-standardized deaths from digestive diseases in both sexes per 100.000 people : 47 missing
- ▶ 41 Age-standardized deaths from unintentional injuries in both sexes per 100.000 people: 47 missing
- ▶ **Total missing values: %7.12**

# Correlation Analysis

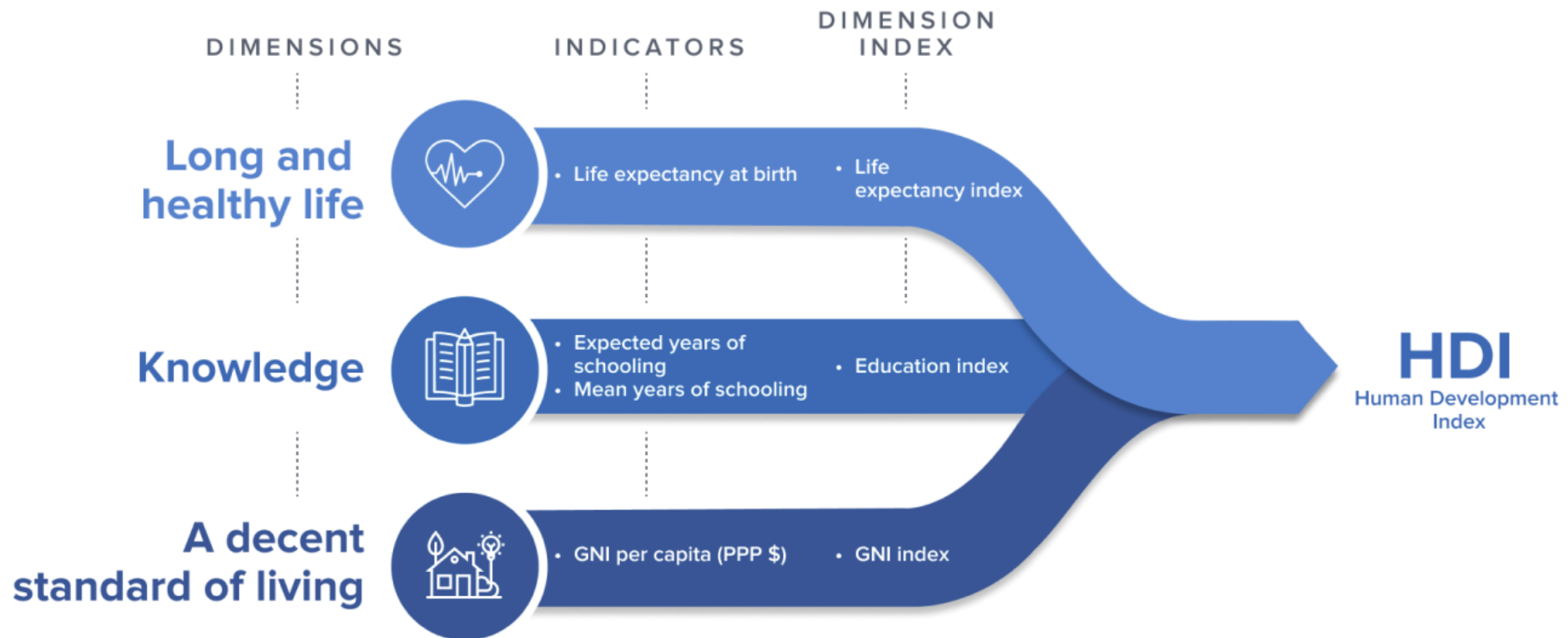


Correlation Matrix Heatmap



# HDI - Human Development Index

## HDI Dimensions and Indicators



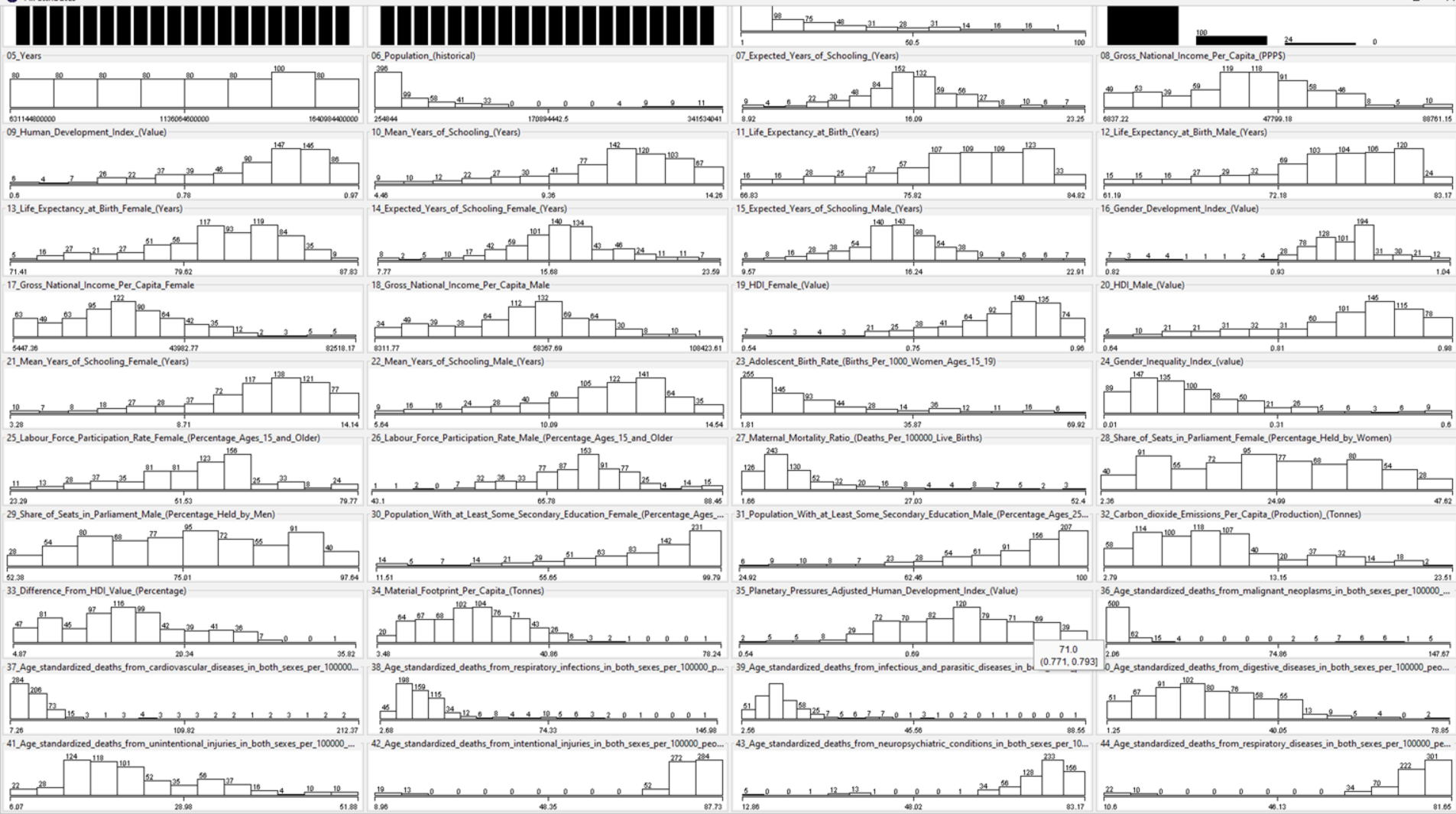
# Top 10 Features Correlated with HDI

	<b> r  value</b>	<b>Attributes</b>
1	0.9724	20_HDI_Male_(Value)
2	0.9691	19_HDI_Female_(Value)
3	0.8811	11_Life_Expectancy_at_Birth_(Years)
4	0.8768	12_Life_Expectancy_at_Birth_Male_(Years)
5	0.8593	13_Life_Expectancy_at_Birth_Female_(Years)
6	0.8349	08_Gross_National_Income_Per_Capita_(PPP\$)
7	0.8137	17_Gross_National_Income_Per_Capita_Female
8	0.7966	18_Gross_National_Income_Per_Capita_Male
9	0.789	07_Expected_Years_of_Schooling_(Years)
10	0.788	04_Human_Development_Groups

Weka: Attribute Evaluator (supervised, Class (numeric): 09\_Human\_Development\_Index\_(Value)): Correlation Ranking Filter

# All Attributes (44)

All attributes



# Regression Modeling (2022)

Linear Regression predicting HDI from key indicators

Cross-validation			
0.9999	R <sup>2</sup>	Correlation coefficient	higher value indicates more relationship
0.0006	MAE	Mean absolute error	The lower the better.
0.0009	RMSE	Root mean squared error	The lower the better.
1.0354 %	RAE	Relative absolute error	The lower the better.
1.1618 %	RRSE	Root relative squared error	The lower the better.
655	Total Number of Instances		
5	Ignored Class Unknown Instances		

Weka: Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4

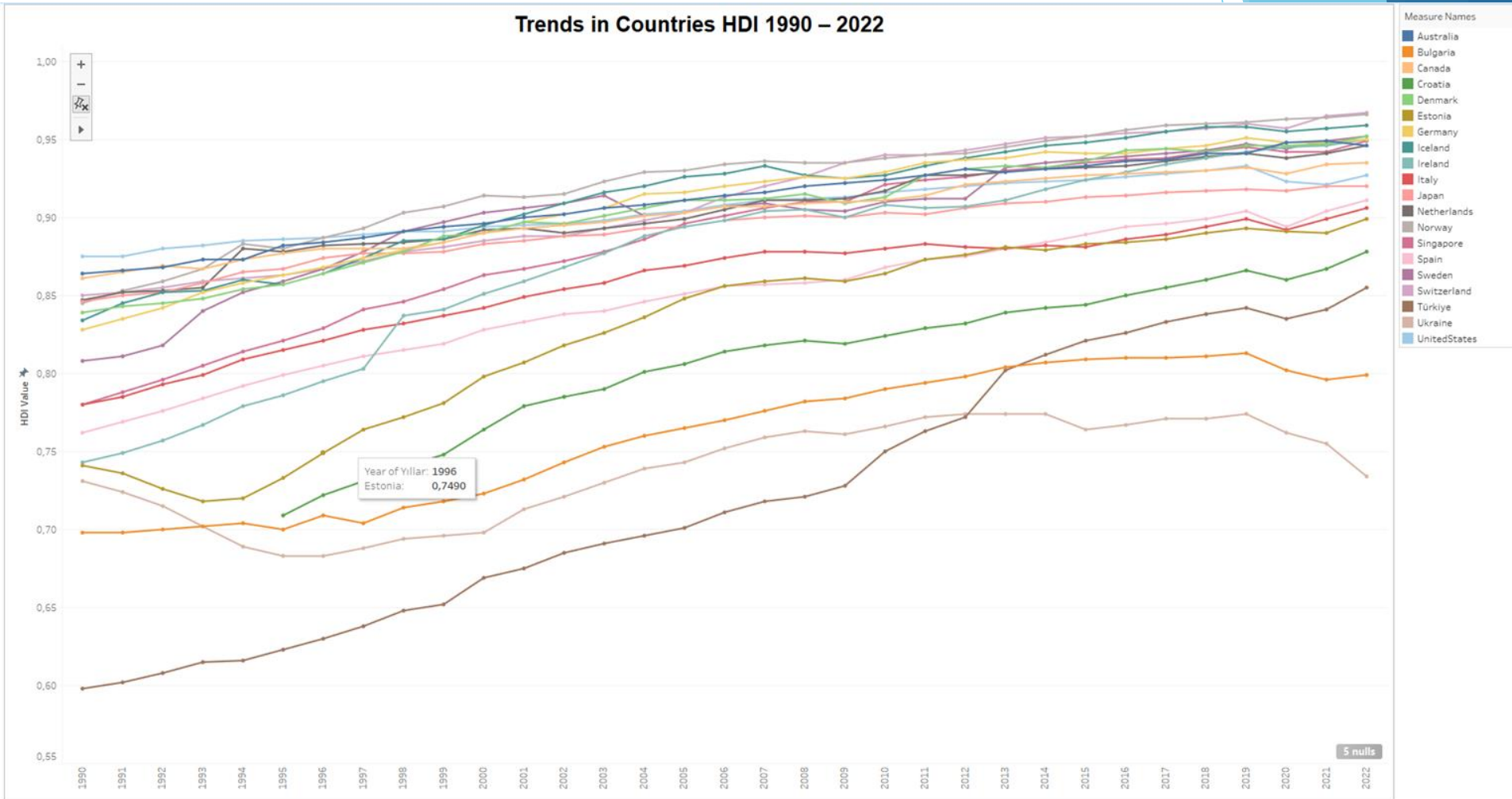
09\_Human\_Development\_Index\_(Value)

# Top 5 HDI Performers (2022)

Country	HDI	GNI per Capita (\$)	Life Expectancy (Years)	Mean Schooling (Years)
Switzerland	0.967	69432.79	84.3	13.9
Norway	0.966	69189.76	83.4	13.1
Iceland	0.959	54688.38	82.8	13.8
Denmark	0.952	62018.96	81.9	13.0
Sweden	0.952	56995.85	83.5	12.7

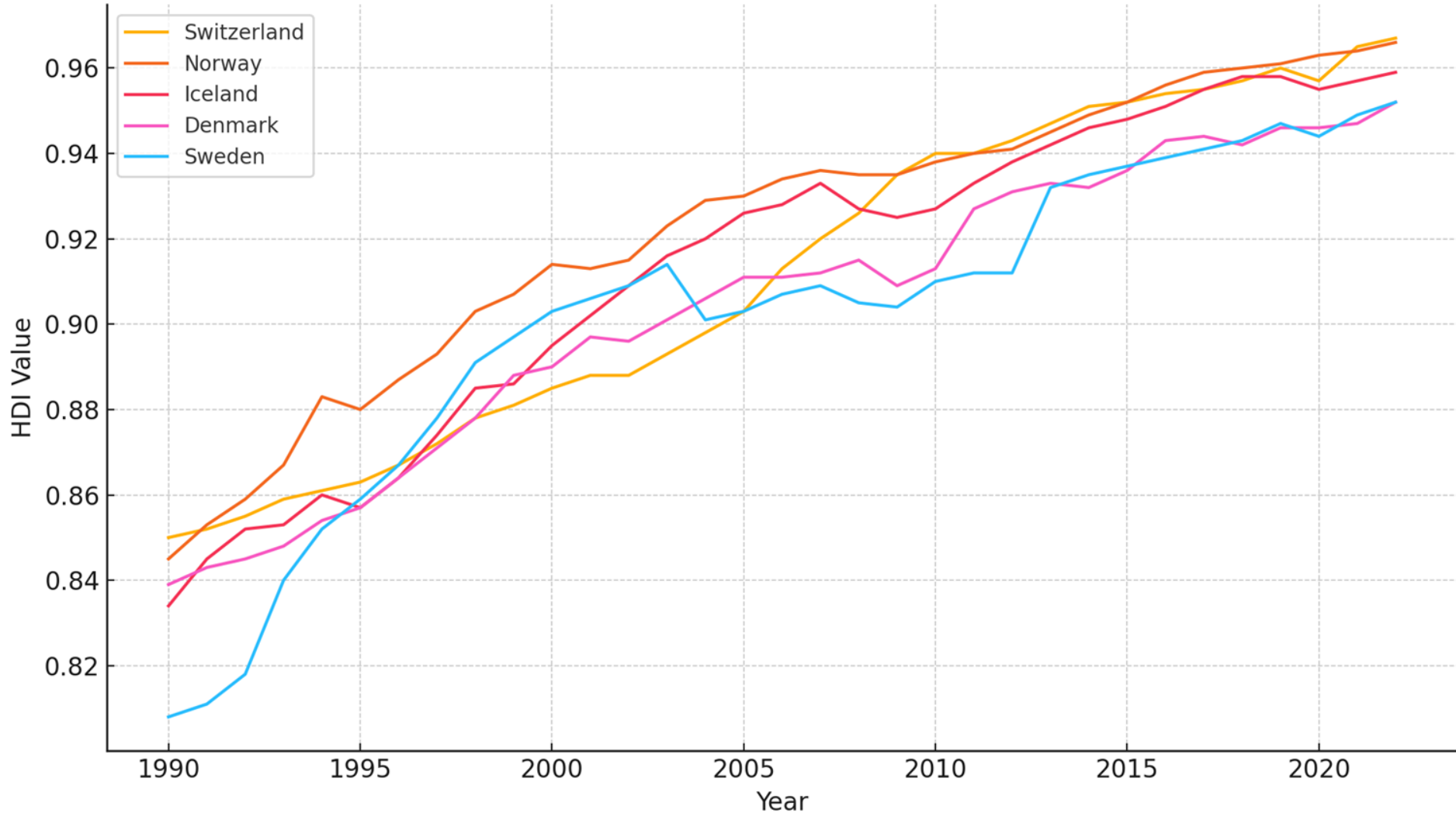
# HDI Trends Over Time (20)

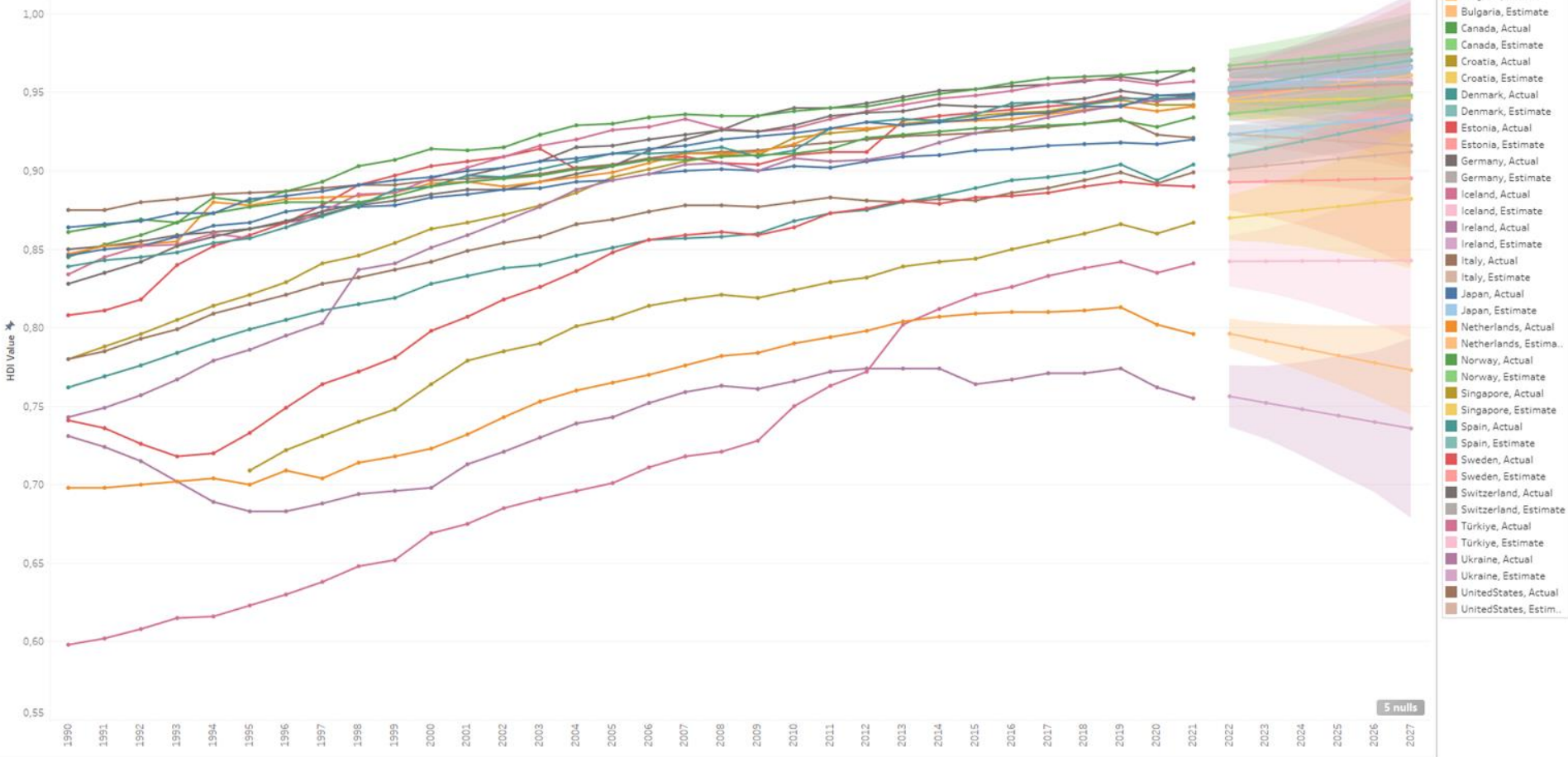
Trends in Countries HDI 1990 – 2022



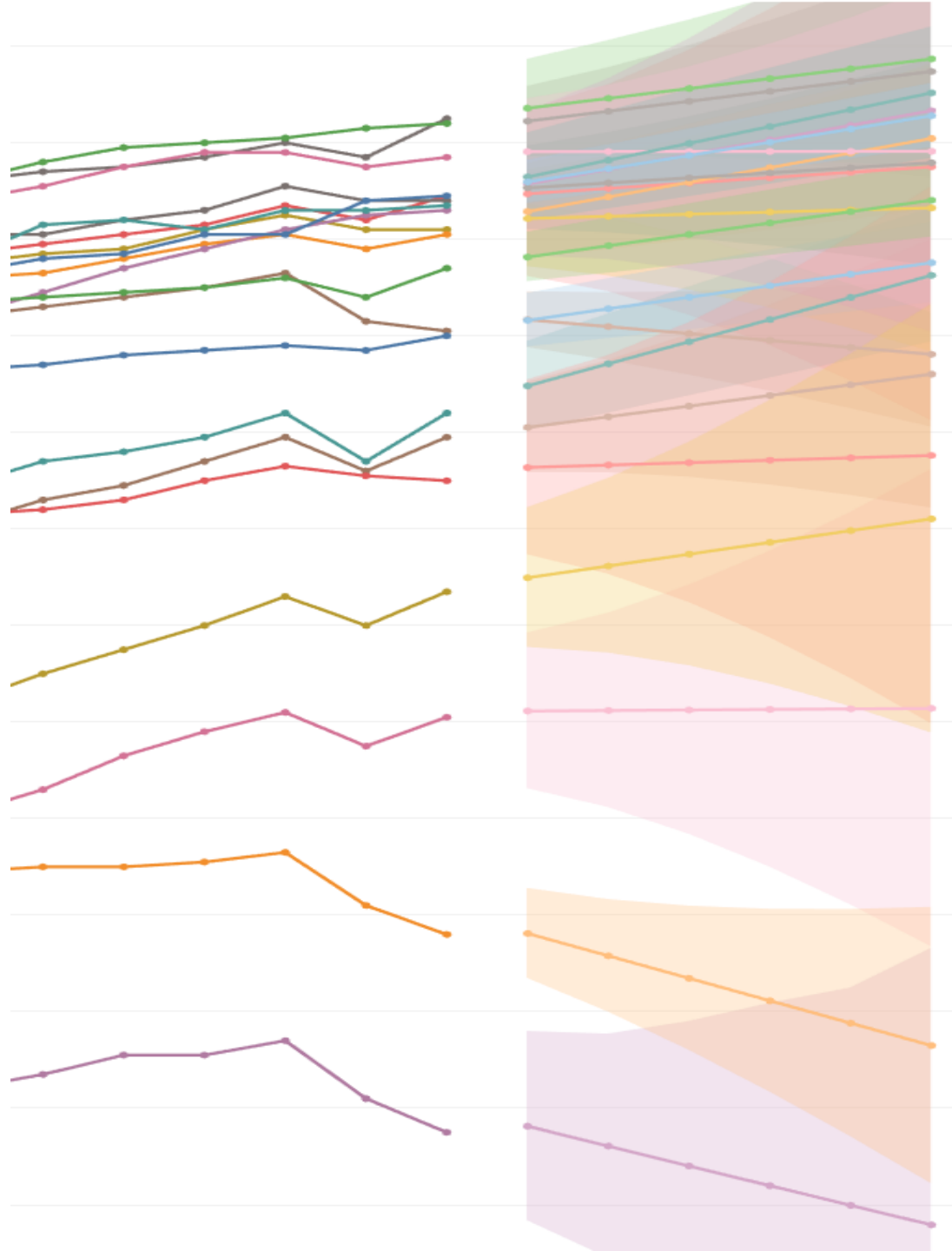
# HDI Trends Over Time (Top 5)

HDI Trends (1990-2022) for Top 5 Countries

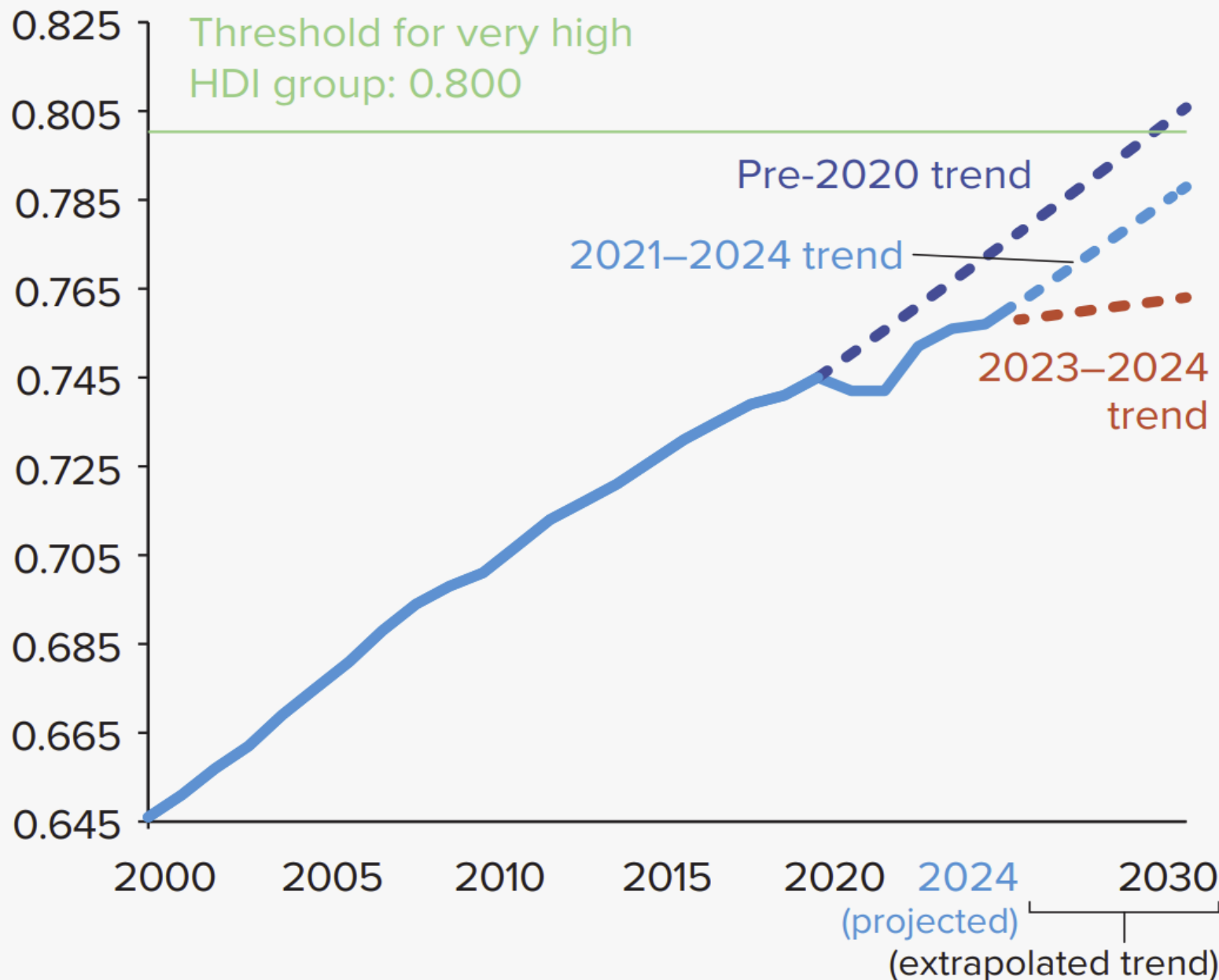


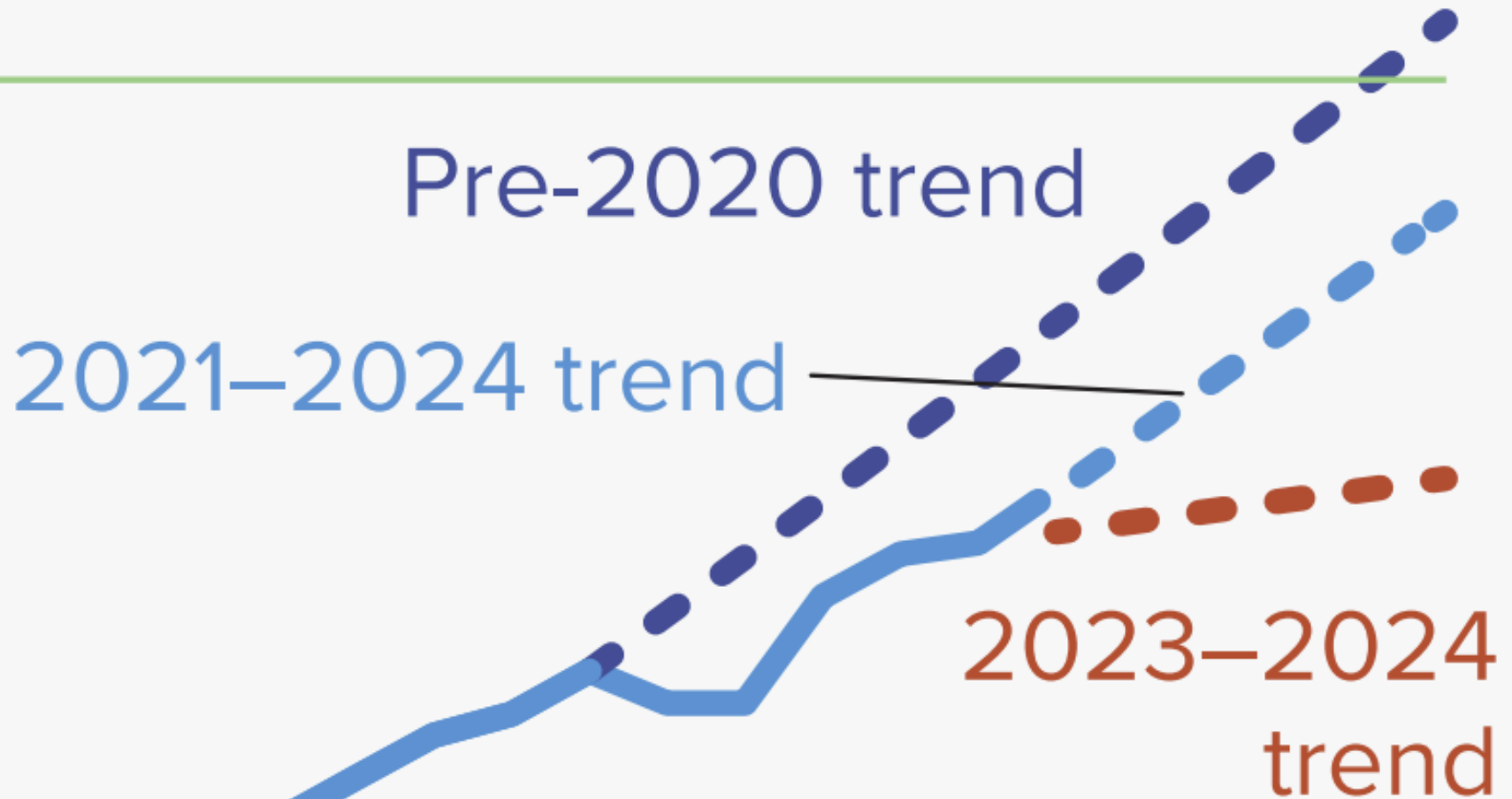


# Prediction & Forecast



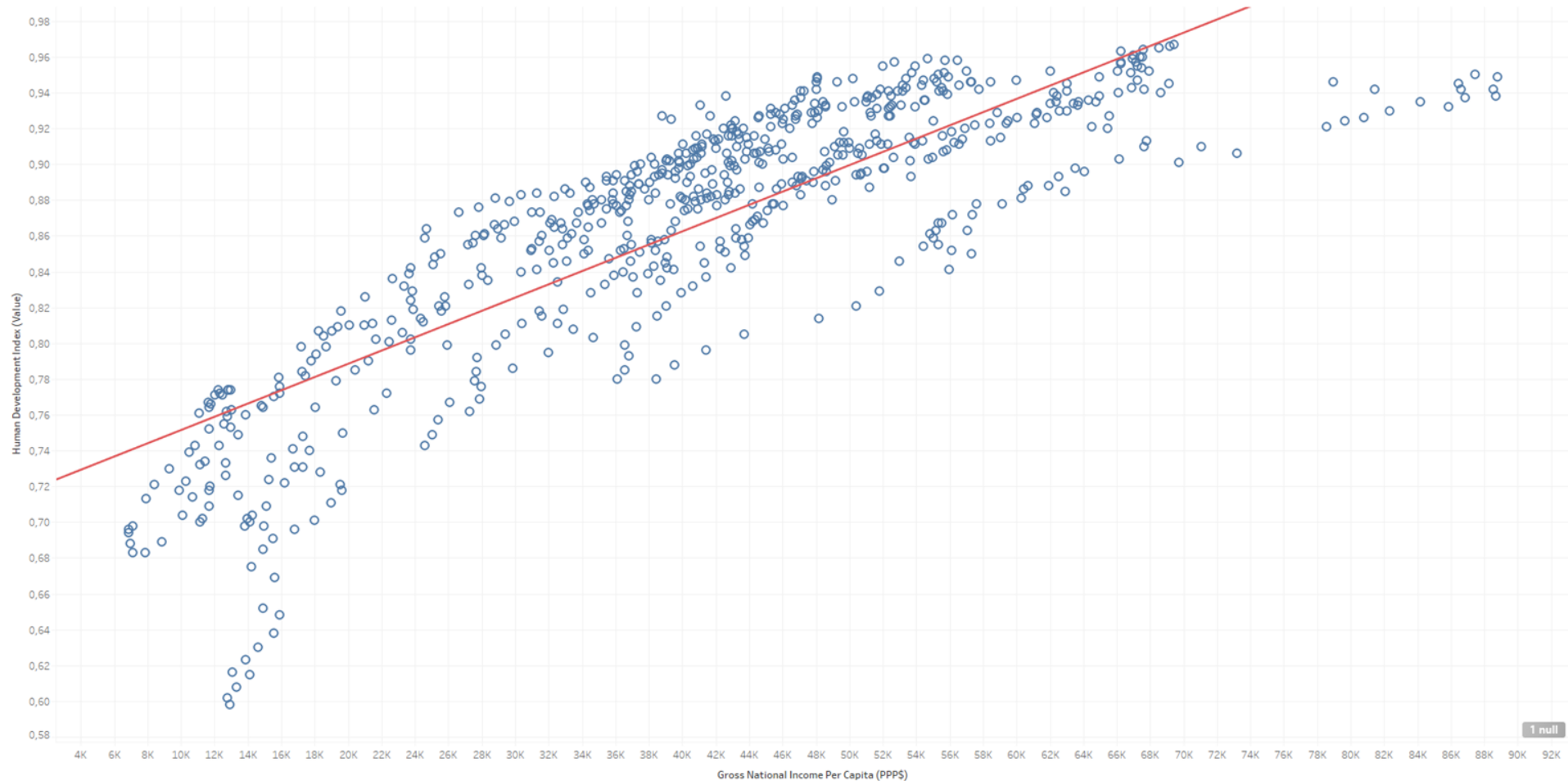
# Global Human Development Index (HDI) value





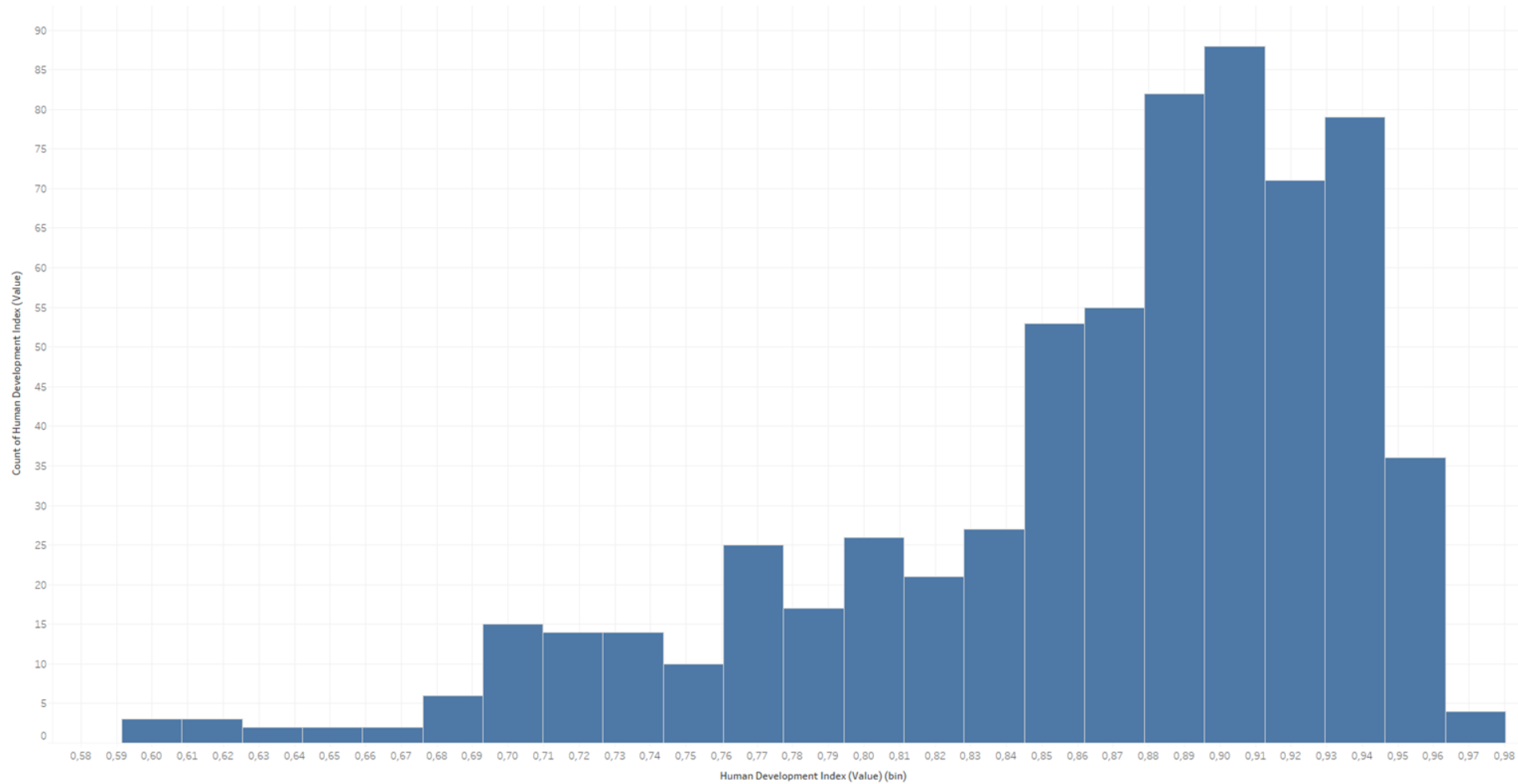
# Economic Growth vs HDI

Economic Growth vs HDI with Trendline

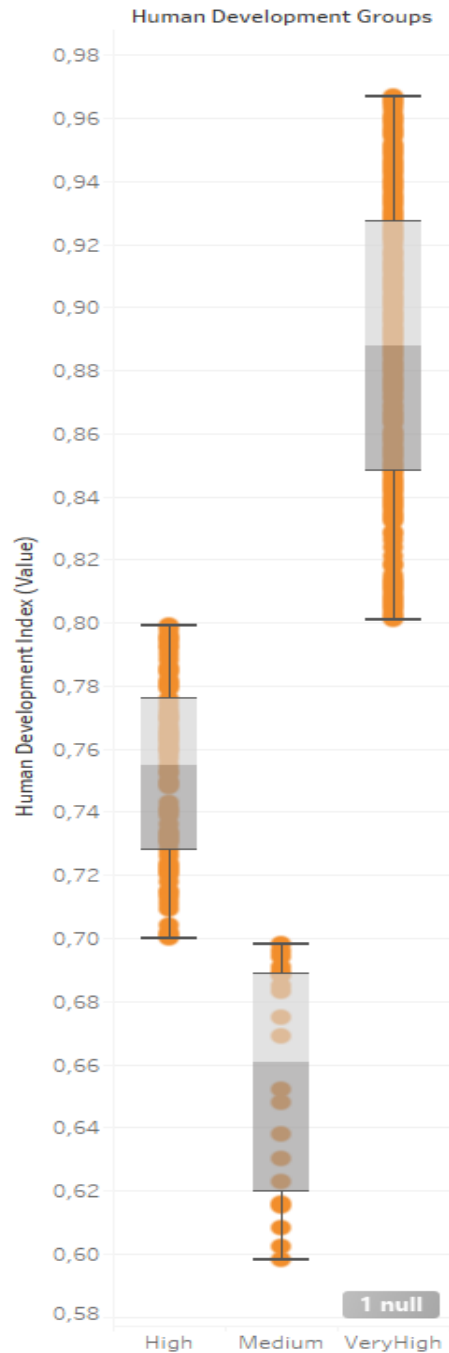


# HDI Distribution

HDI Frequency

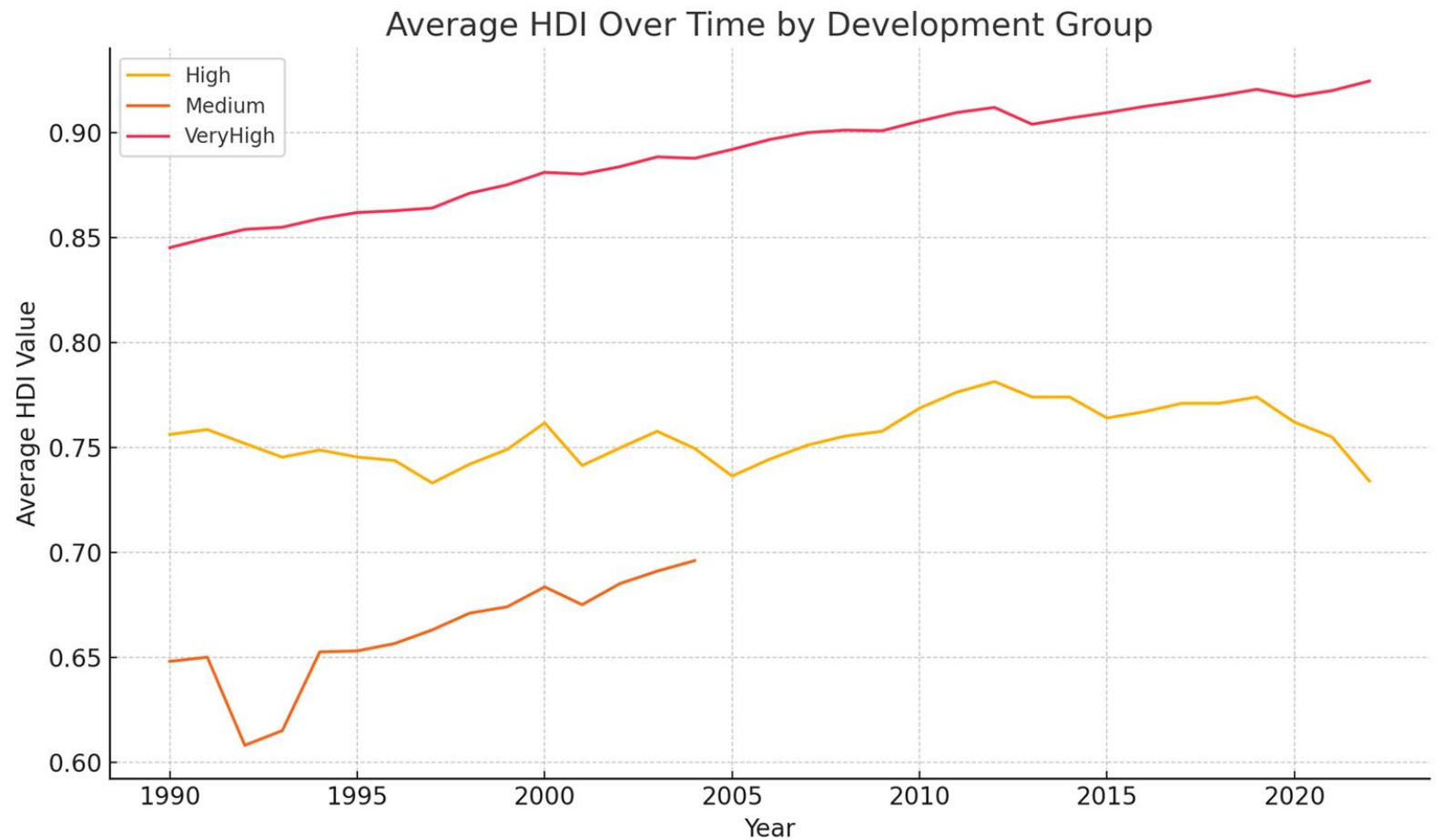


# HDI Box and Whisker Plots



## HDI Boxplot by Development Group

# Average HDI Over Time



What are the common goals of all countries?

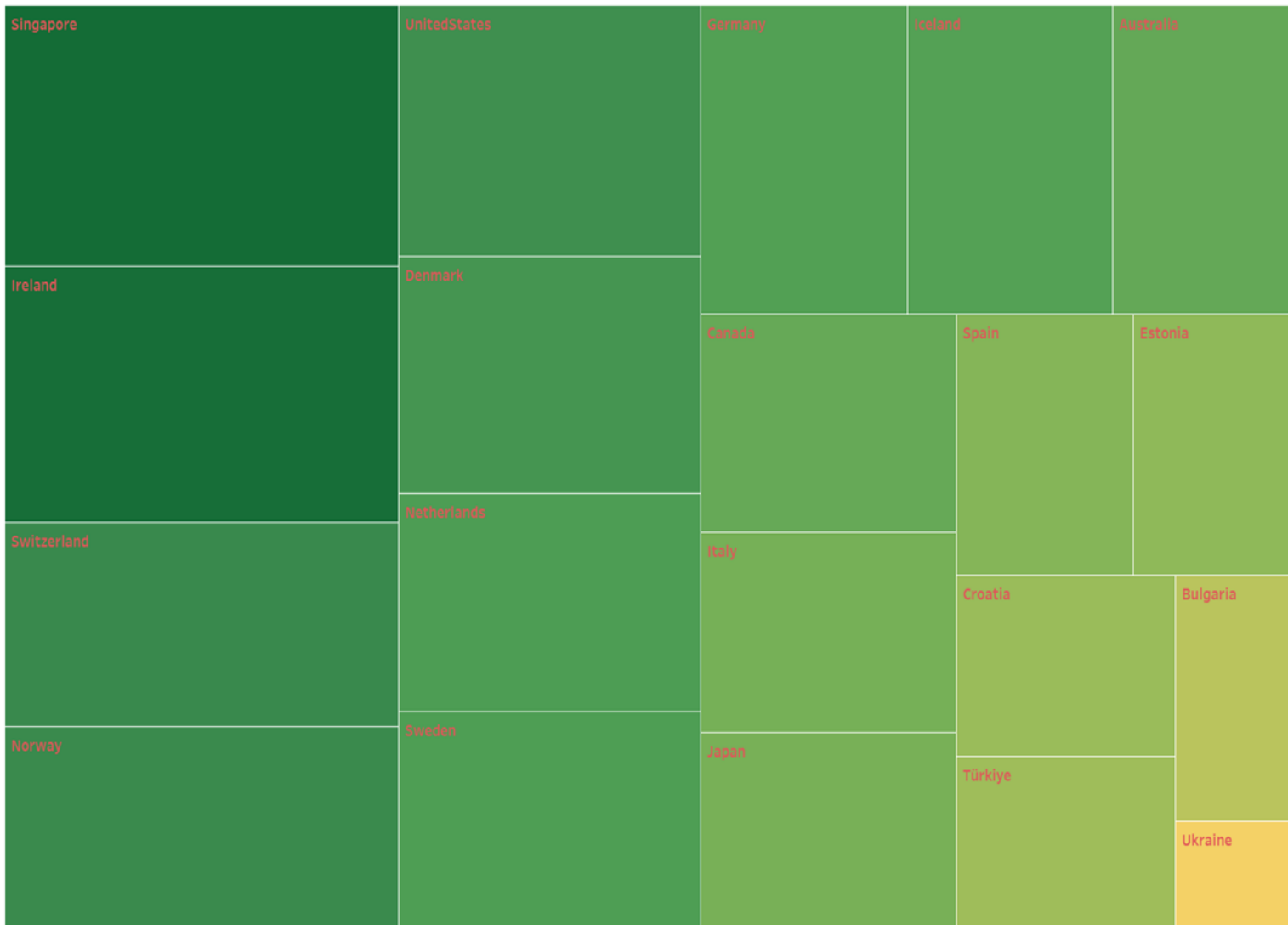
???

- \* High GNI
- \* Good healthcare system
- \* Quality education

# Gross National Income Per Capita (\$)



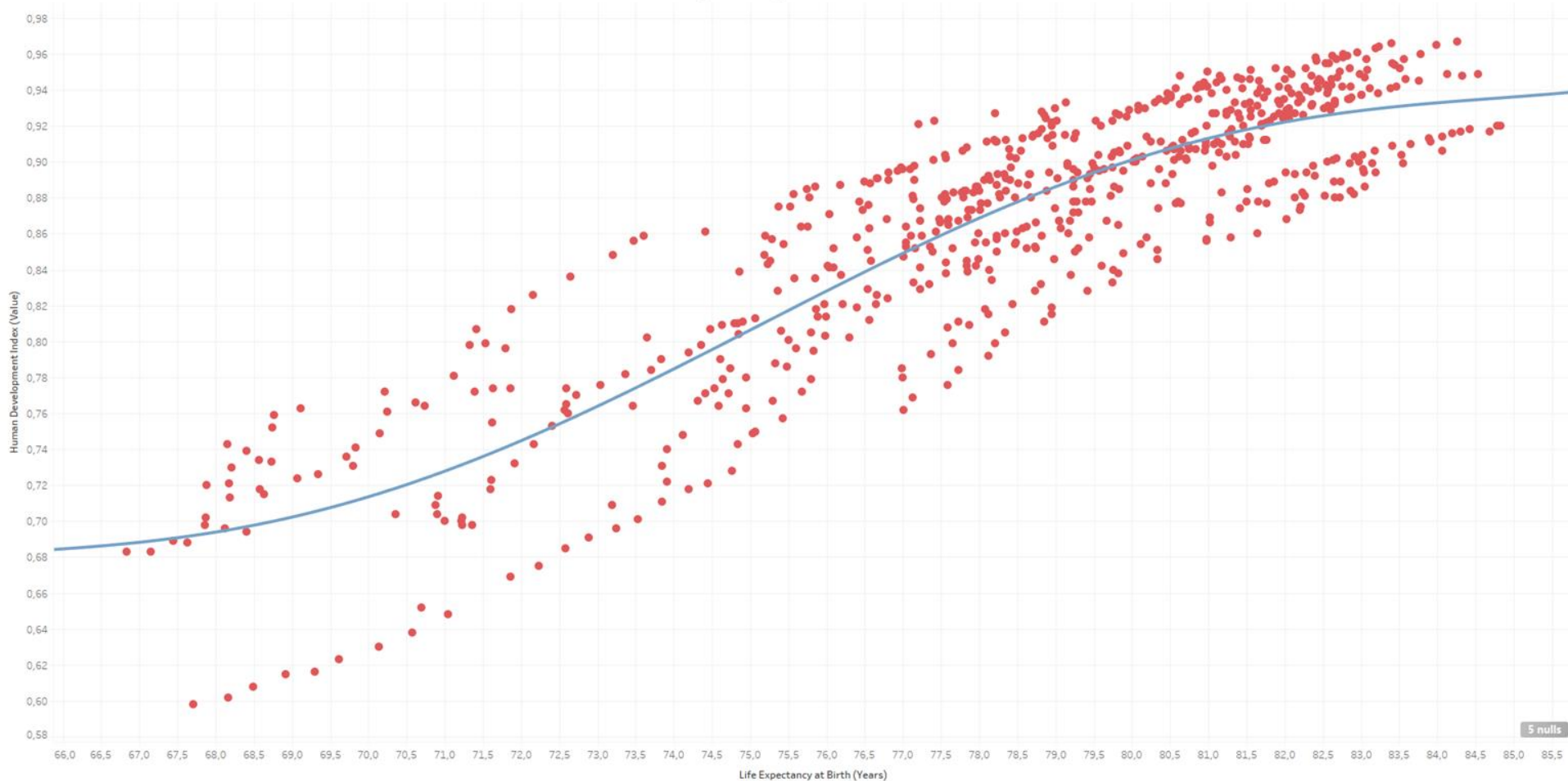
Gross National Income Per Capita (\$)



SUM(Gross National In.  
11.416 88.761

# Life Expectancy vs HDI

Life Expectancy at Birth vs HDI



Trend Line: Polynominal

Degree: 5

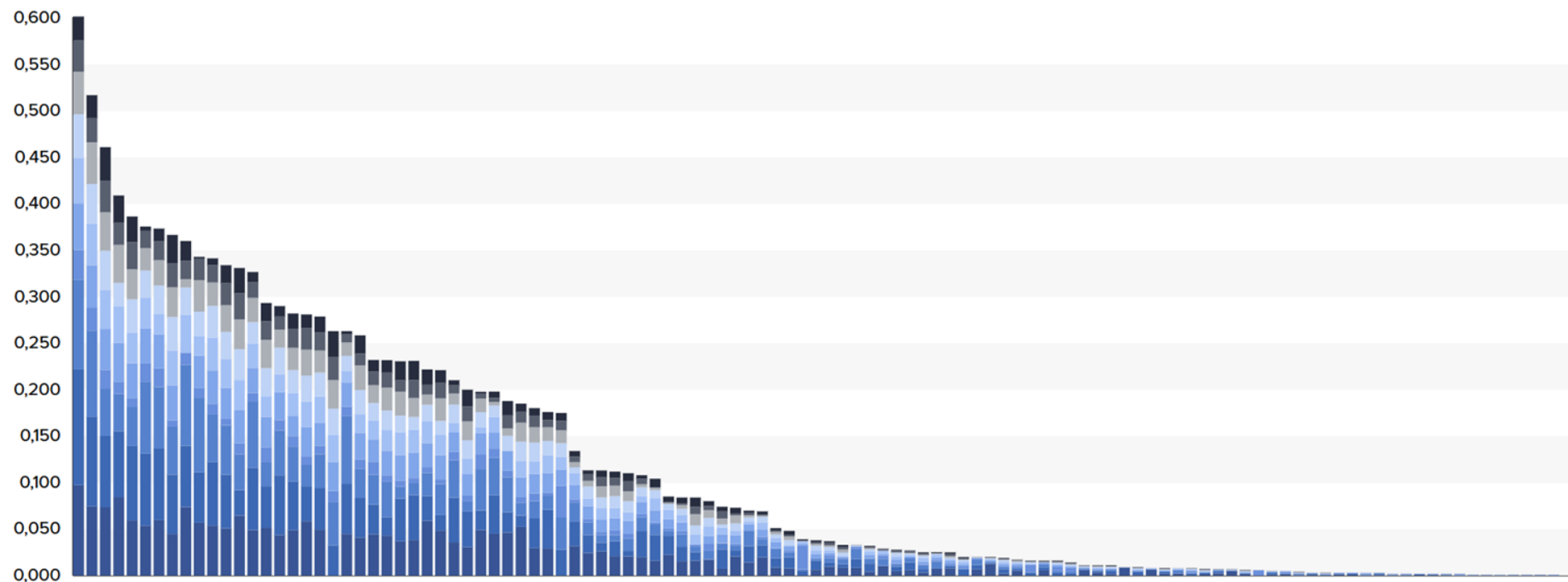
What is the other side of the  
coin?

# Multidimensional Poverty

Contribution of deprivation in indicator to overall multidimensional poverty



**MPI** Note: the lower MPI values represent a better performance regarding multidimensional poverty.



# Policy Recommendations

- ▶ Invest in education: Expand access to early childhood and secondary schooling
- ▶ Enhance healthcare systems: Strengthen primary care and reduce child mortality
- ▶ Promote inclusive economic growth: Targeted social programs to raise GNI
- ▶ Encourage gender equality: Bridge gaps in education and workforce participation
- ▶ Integrate environmental sustainability into development planning

# Conclusion & Next Steps

- ▶ HDI strongly linked to GNI, health, and education indicators.
- ▶ Regression model explains ~95.6% of variance in HDI.
- ▶ Top performers maintain holistic development strategies.

## Next Steps

- ▶ Expand model with more variables (e.g., gender inequality, environment, schooling).
- ▶ Consider time series forecasting for future HDI prediction.



# Thank You!