# Main Research Question

Are there discernable demographic, geographic, socio-economic, political or cultural parameters that impact the successful development of artificial intelligence in a country?
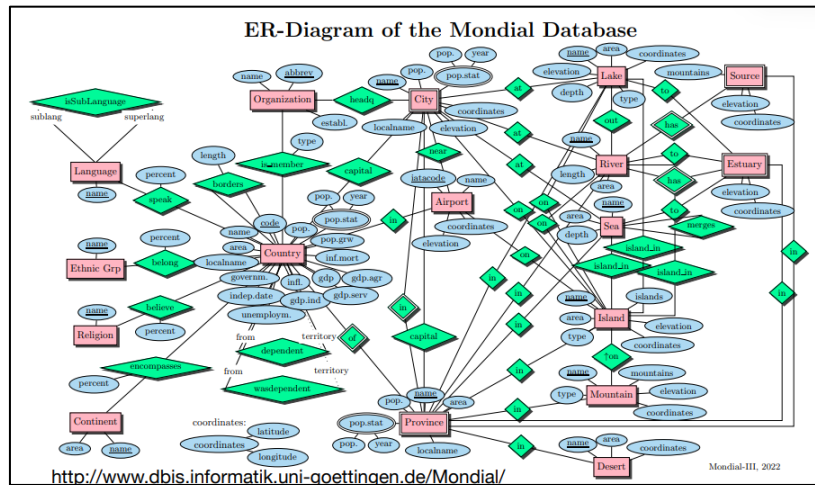
Research Strategy

- Develop an adapted MySQL database that is constructed out of the Mondial database

- Import data from external sources surrounding global developments in AI

- Measuring Success? Find Benchmark data

- Add relevant primary and foreign keys to the imported tables

- Analyze data with the goal of finding relevant correlations
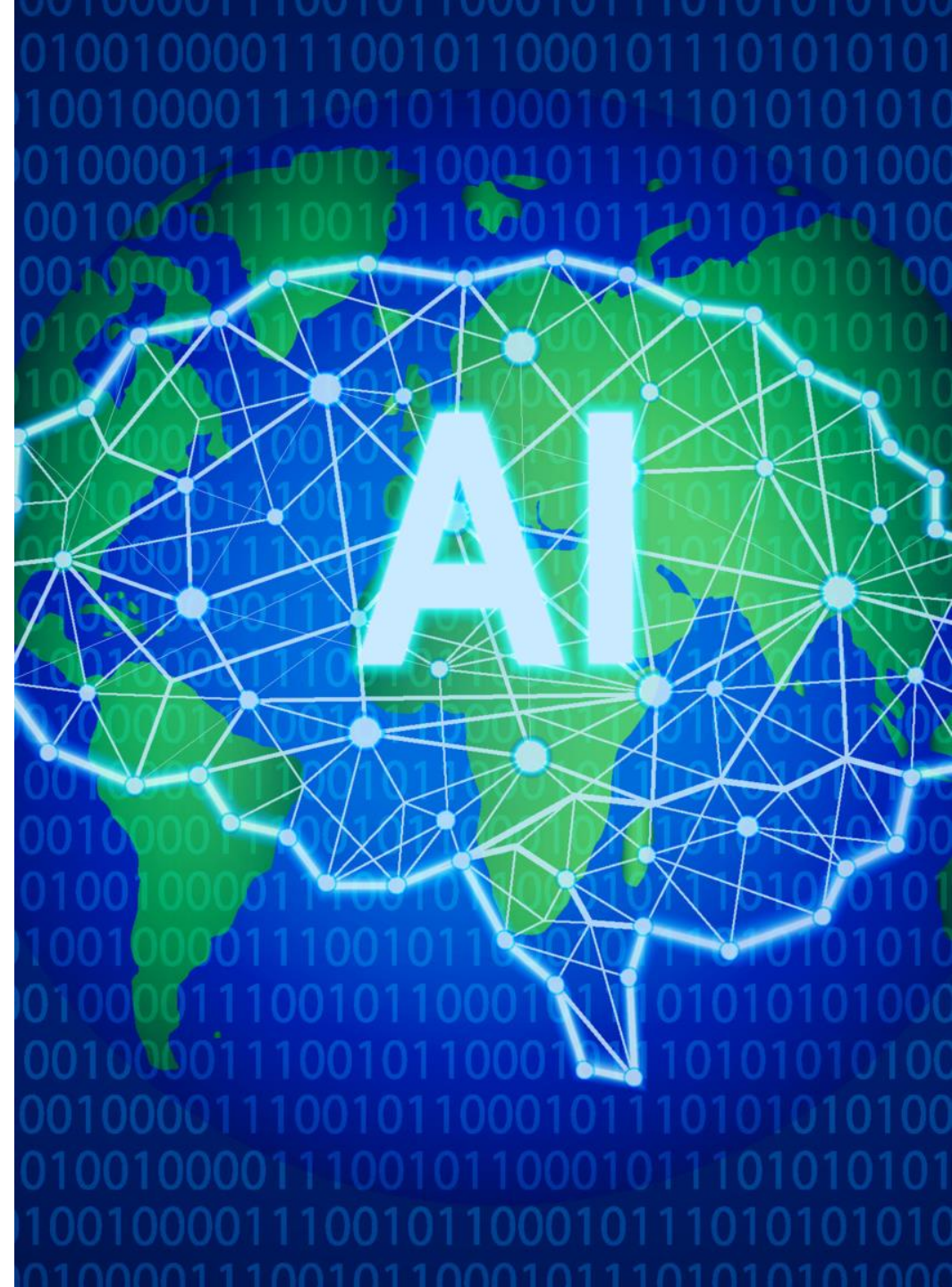
Gathering Data

Country Data

- Mondial Database

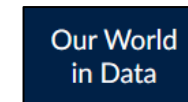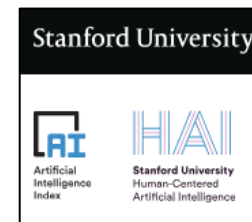- Gapminder Database

AI Data

- World Bank Data

- OECD Data

- Tortoise Media

- Our World In Data

- AI Index

Benchmark Data: Tortoise Media

	Overall	Talent	Infrastructure	Operating Environment	Research	Development	Government Strategy	Commercial
United States	1	1	1	28	1	1	8	1
China	2	20	2	3	2	2	3	2
Singapore	3	4	3	22	3	5	16	4
United Kingdom	4	5	24	40	5	8	10	5
Canada	5	6	23	8	7	11	5	7
South Korea	6	12	7	11	12	3	6	18
Israel	7	7	28	23	11	7	47	3
Germany	8	3	12	13	8	9	2	11
Switzerland	9	9	13	30	4	4	56	9
Finland	10	13	8	4	9	14	15	12
Netherlands	11	8	16	15	10	13	28	20
Japan	12	11	5	10	20	6	18	23
France	13	10	11	25	15	18	13	10
India	14	2	59	12	30	21	38	13
Australia	15	14	44	62	6	16	14	22
Denmark	16	19	15	1	18	19	21	17
Sweden	17	15	21	2	13	17	44	16
Luxembourg	18	31	6	14	19	22	31	14
Ireland	19	17	26	19	27	10	29	15
Austria	20	25	34	5	16	23	33	27

- An extensive research project on global AI development
- Overall Rankings for top 62 countries
- 3 *Ranking Groups/Pillars* used for determining "Success"
  - Implementation (Orange)
    - Talent
    - Infrastructure
    - Operating Environment
  - Innovation (Blue)
    - Research
    - Development
  - Investment (Yellow)
    - Government Strategy
    - Commercial

Analyzing the Data
Query 6, 7, 11: Creating Views

- "View_Name" (basic_country_data) → view that allows us to execute different queries based on country names or codes and simultaneously reveal basic demographic, economic and political data
  - Key Questions:
    - What countries should be considered? (Query 1)
    - Case of Hong Kong and Taiwan? (Query 2-4)
      - Not listed in ai_gapminder_country table, but is listed in benchmark tables

- "View_Name_Economic" (main_economic_rankings) → view that highlights key calculations and visualizations so we can more easily compare key economic data with benchmark data

Query 6

```python
view_name = 'basic_country_data'

query = f"""CREATE VIEW {view_name} AS
SELECT t1.code, t1.name, t1.area, t1.population,
t2.iso3166_1_alpha2 AS ISO2, t2.iso3166_1_alpha3 AS ISO3, t2.world_4region, t2.g77_and_oecd_countries AS club, t2.income_grou
t3.GDP, SUBSTRING(t4.government,1,50) AS Government,
t5.overall,
t3.GDP/t1.population*1000000 AS GDPPC
FROM country AS t1
LEFT JOIN ai_gapminder_country AS t2 ON t1.name = t2.name
LEFT JOIN economy AS t3 ON t1.code = t3.country
LEFT JOIN politics AS t4 ON t1.code = t4.country
LEFT JOIN ai_tortoise_indicators AS t5 ON t1.name = t5.country
WHERE t5.overall IS NOT NULL OR t2.name IS NOT NULL;
"""

try:
    with engine.connect() as conn:
        drop_view_query = f"DROP VIEW IF EXISTS {view_name}"
        conn.execute(text(drop_view_query))
        conn.execute(text(query))
except Exception as e:
    print(e)
```

Query 7

```python
query = f"SELECT * FROM {view_name}"

try:
    # Execute the query and store the result in a DataFrame
    df = pd.read_sql_query(query, engine)

    # Apply formatting for thousands and decimal points
    formatted_df = df.style \
        .format({
            'area': '{:,.0f}',
            'population': '{:,.0f}',
            'GDP': '{:,.0f}',
            'GDPPC': '{:,.2f}',
            'overall': '{:,.2f}'
        }, thousands=".", decimal=",")
    # Display the formatted DataFrame
    display(formatted_df)
except Exception as e:
    print(e)
```

	code	name	area	population	ISO2	ISO3	world_4region	club	income_groups	GDP	Government	overall	GDPPC
0	A	Austria	83.850	8.023.244	AT	AUT	europe	oecd	high_income	152.000	federal republic	27,70	18.944,96
1	AFG	Afghanistan	647.500	22.664.136	AF	AFG	asia	g77	low_income	12.800	transitional government	nan	564,77
2	AG	Antigua and Barbuda	442	65.647	AG	ATG	americas	g77	high_income	425	parliamentary democracy	nan	6.474,02
3	AL	Albania	28.750	3.249.136	AL	ALB	europe	others	upper_middle_income	4.100	emerging democracy	nan	1.261,87

Query 11

```python
view_name_economic = 'main_economic_rankings'

query = f"""CREATE VIEW {view_name_economic} AS
SELECT Name, Population, GDPPC, GDP,
RANK() OVER ( ORDER BY Population DESC) population_rank,
RANK() OVER ( ORDER BY GDPPC DESC) GDPPC_rank,
ROW_NUMBER() OVER (ORDER BY GDP DESC) GDP_rank,
RANK() OVER ( ORDER BY overall DESC) overall_rank
from {view_name} ORDER BY GDP DESC;
"""

try:
    with engine.connect() as conn:
        drop_view_query = f"DROP VIEW IF EXISTS `{view_name_economic}`"   # Using backticks in case of special characters
        conn.execute(text(drop_view_query))
        conn.execute(text(query))
except Exception as e:
    print(e)
```

Analyzing the Data (countries)
Query 9, 12, 13: Comparing country GDP data with benchmark ratings

- Query 9 & 10 → How does each country's GDP compare with benchmark rankings?
  - Correlation of 0,67 for GDP per capita when only countries with a ranking are taken into account
- Query 12 → Are there any countries with top 50 GDP_rank but outside benchmark data ?
  - Venzuela seems to be somewhat of an outlier. Why?
- Query 13, 14 & 15 → Are there any countries with GDP_rank outside top 50 but within benchmark data?
  - Estonia seems to be an interesting country to research further primarily because it has a very impressive investment performance indicator

Query 9

```
# RANK() OVER (ORDER BY ...) returns the ranking within the dataframe following the "ORDER BY" clause between brackets
query = f"""
SELECT Name, Population, GDPPC, GDP,
RANK() OVER ( ORDER BY Population DESC) population_rank,
RANK() OVER ( ORDER BY GDPPC DESC) GDPPC_rank,
RANK() OVER (ORDER BY GDP DESC) GDP_rank,
RANK() OVER ( ORDER BY overall DESC) overall_rank
FROM {view_name}
ORDER BY GDP DESC
"""
try:
    # Apply formatting and coloring using the modified function
    df = pd.read_sql_query(query, engine)
    formatted_df = df.style\
        .format(precision=0, thousands=".", decimal=",")
    # Display the formatted DataFrame
    display(formatted_df)
except Exception as e:
    print(e)
```

	name	population	GDPPC	GDP	population_rank	GDPPC_rank	GDP_rank	overall_rank
0	United States	266.476.278	27.198	7.247.700	3	2	1	1
1	China	1.210.004.956	2.893	3.500.000	1	101	2	2
2	Japan	125.449.703	21.357	2.679.200	8	11	3	12
3	Germany	83.536.115	17.384	1.452.200	12	26	4	8
4	India	952.107.694	1.480	1.408.700	2	127	5	14
5	France	58.317.450	20.114	1.173.000	20	13	6	13
6	United Kingdom	58.489.975	19.463	1.138.400	19	16	7	4
7	Italy	57.460.274	18.945	1.088.600	21	20	8	23
8	Brazil	162.661.214	6.005	976.800	5	64	9	34

Query 12

```
query = f"SELECT * FROM {view_name_economic} WHERE overall_rank = 63 AND GDP_rank < 51 AND GDPPC_rank < 51 ORDER BY GDP DESC"
try:
    # Apply formatting using the modified function
    df = pd.read_sql_query(query, engine)
    formatted_df = df.style\
        .format(precision=0, thousands=".", decimal=",")
    # Display the formatted DataFrame
    display(formatted_df)
except Exception as e:
    print(e)
```

	Name	Population	GDPPC	GDP	population_rank	GDPPC_rank	GDP_rank	overall_rank
0	Venezuela	21.983.188	8.893	195.500	42	49	28	63

Query 13

```
query = f"SELECT * FROM {view_name_economic} WHERE overall_rank < 63 AND GDP_rank > 50 AND GDPPC_rank > 50 ORDER BY overall_r
try:
    # Apply formatting using the modified function
    df = pd.read_sql_query(query, engine)
    formatted_df = df.style\
        .format(precision=0, thousands=".", decimal=",")
    # Display the formatted DataFrame
    display(formatted_df)
except Exception as e:
    print(e)
```

	Name	Population	GDPPC	GDP	population_rank	GDPPC_rank	GDP_rank	overall_rank
0	Estonia	1.459.428	8.428	12.300	139	51	100	25
1	Hungary	10.002.541	7.248	72.500	70	57	54	38
2	Lithuania	3.646.041	3.648	13.300	114	87	98	43
3	Slovakia	5.374.362	7.257	39.000	100	56	67	50
4	Uruguay	3.238.952	7.533	24.400	123	54	78	53
5	Armenia	3.463.574	2.627	9.100	117	106	112	54
6	Tunisia	9.019.687	4.113	37.100	78	83	68	56
7	Morocco	29.779.156	2.935	87.400	32	100	51	57
8	Sri Lanka	18.553.074	3.536	65.600	49	89	57	60
9	Kenya	28.176.686	1.306	36.800	36	135	70	62

Analyzing the Data (a wider net)
Queries 16, 17, 18: From Country to Continent

- Geography (Query 16)
  - AI_Performance is spread – apart from Africa – relatively evenly
  - "None" as reference to Taiwan and Hong Kong

- Income(Query 17,18)
  - High_income status = higher ai_performance
  - Asia (mainly because of China) is an exception

	continent	ai_performance
0	americas	31,387500
1	asia	27,394118
2	europe	26,562069
3	None	23,950000
4	africa	12,650000

	continent	income_groups	ai_performance
0	None	None	23,95
1	americas	high_income	44,20
2	asia	high_income	29,69
3	asia	upper_middle_income	30,37
4	europe	high_income	27,37

	continent	income_groups	ai_performance
0	africa	lower_middle_income	12,36
1	africa	upper_middle_income	14,10
2	americas	upper_middle_income	18,58
3	asia	lower_middle_income	19,42
4	europe	upper_middle_income	19,60

Query 16   Query 17   Query 18

Analyzing the Data (organizations)
Query 22: On the importance of organizational membership

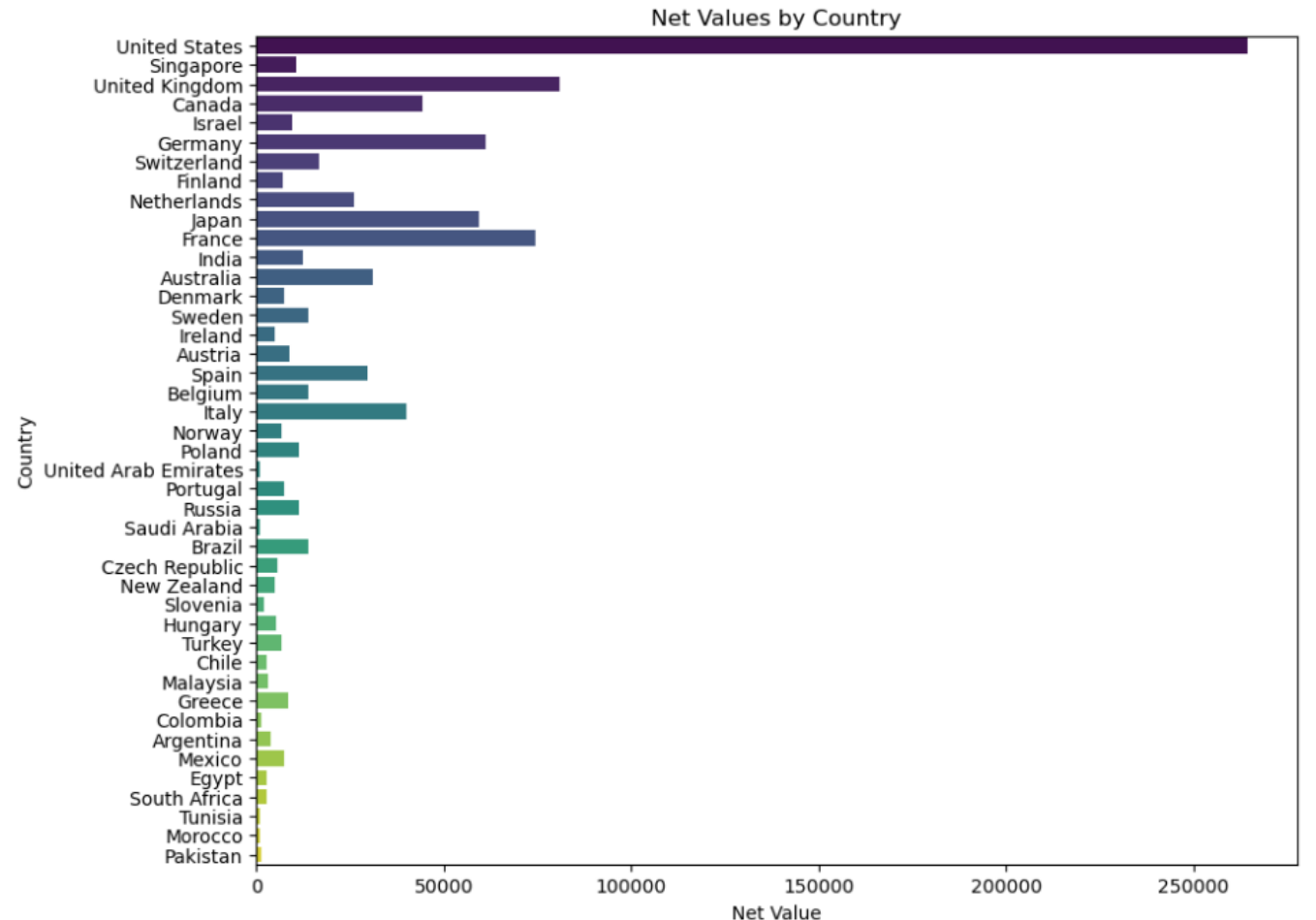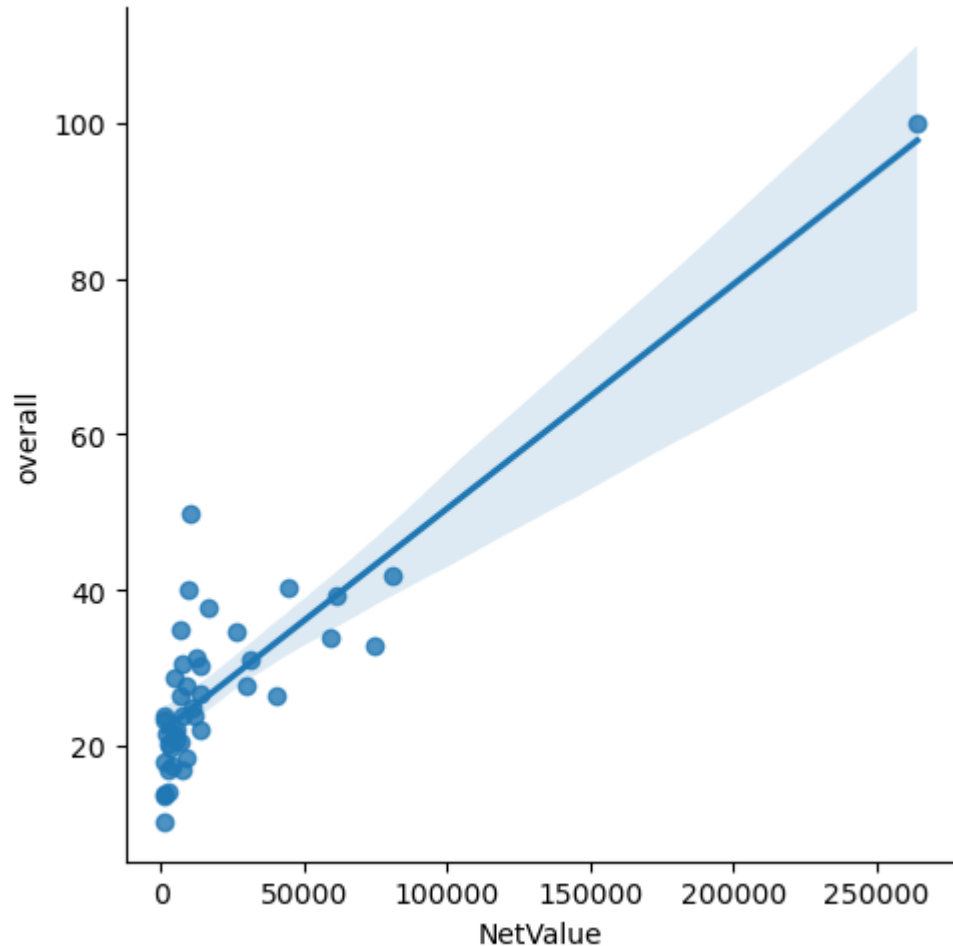|	club	code	name	overall	overall_rank	HQ_Count
0	oecd	USA	United States	100	1	14
1	oecd	GB	United Kingdom	42	4	4
2	oecd	CDN	Canada	40	5	1
3	oecd	CH	Switzerland	38	9	17
4	oecd	SF	Finland	35	10	1
5	oecd	F	France	33	13	10
6	oecd	AUS	Australia	31	15	1
7	oecd	S	Sweden	30	17	1
8	oecd	E	Spain	28	20	1
9	oecd	A	Austria	28	20	4
10	oecd	B	Belgium	27	22	8
11	oecd	I	Italy	26	23	2
12	g77	UAE	United Arab Emirates	24	28	1
13	g77	SA	Saudi Arabia	23	31	3
14	oecd	CZ	Czech Republic	22	34	1
15	oecd	H	Hungary	21	38	1
16	g77	RCH	Chile	20	41	1
17	g77	RI	Indonesia	18	46	2
18	g77	RA	Argentina	18	49	1
19	oecd	MEX	Mexico	17	51	1
20	g77	ROU	Uruguay	16	53	1
21	g77	RSA	South Africa	14	55	1
22	g77	MA	Morocco	14	57	1
23	g77	CL	Sri Lanka	10	60	1

Based on organizational membership we get a correlation coefficient of **0,64** and interesting exceptions on the upper and lower side of the rankings.
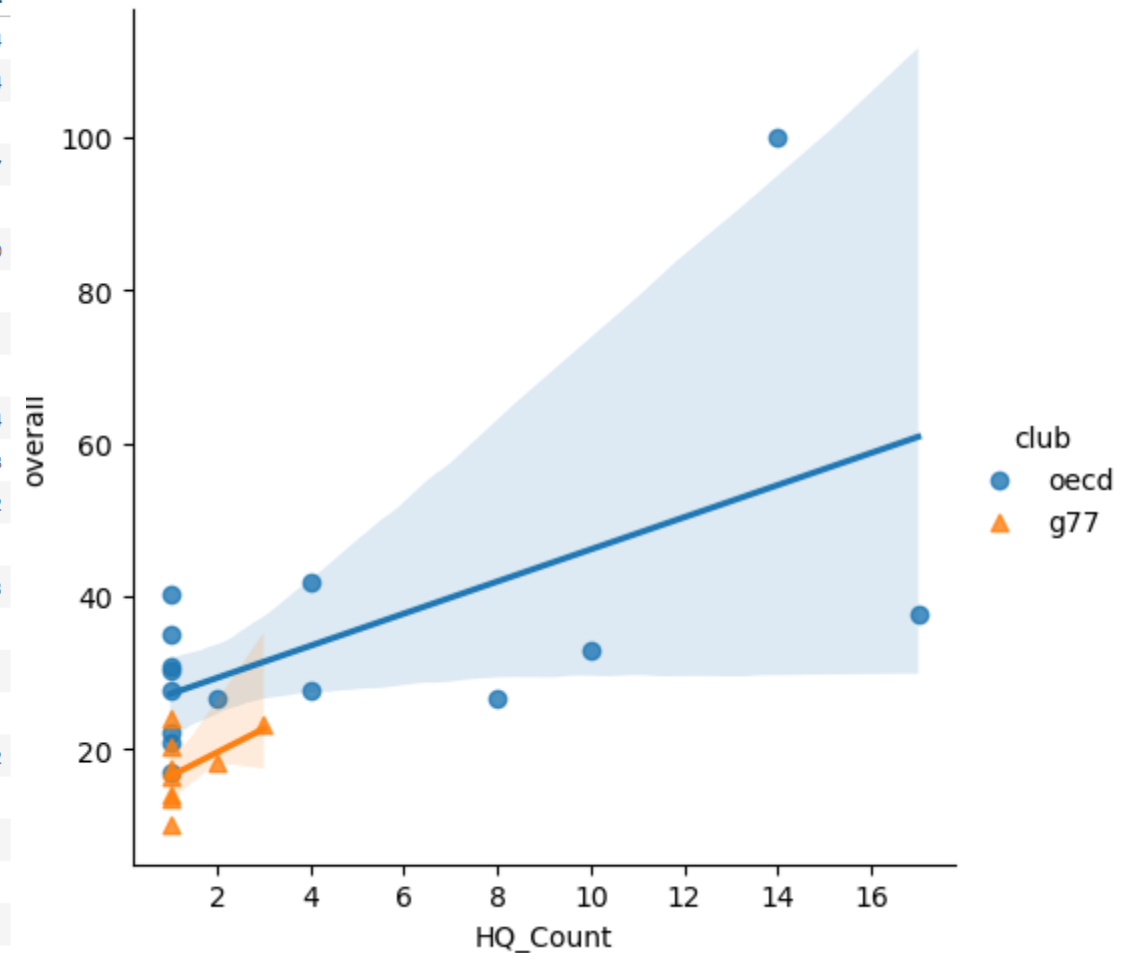
Analyzing the Data (networks)
Query 23: Research and Prominence on the AI Stage (Final Query)

Conclusion

- Based on the data we used to investigated the prominence of countries on the AI scene, being basic economic and demographic data, complemented with prominence in organizational and research networks, we can conclude that there are clear indications that **economic as well as network integration are prominent factors in the performance of a country on the AI stage**.

- There are many **more possibilities to investigate** why certain countries have prominence in the ongoing ai revolution, however within the confines of this project we will limit ourselves to some of the **hints for further research** we referred to earlier.

- Many of the tables we collected to support such an investigation where not used, but could be interesting to look at. It is also very important to **update the data** from the Mondial database to make accurate conclusions as to which factors do count for AI prominence as a country. As said, all this would take much **more research** but **it would definitely be worthwhile**.