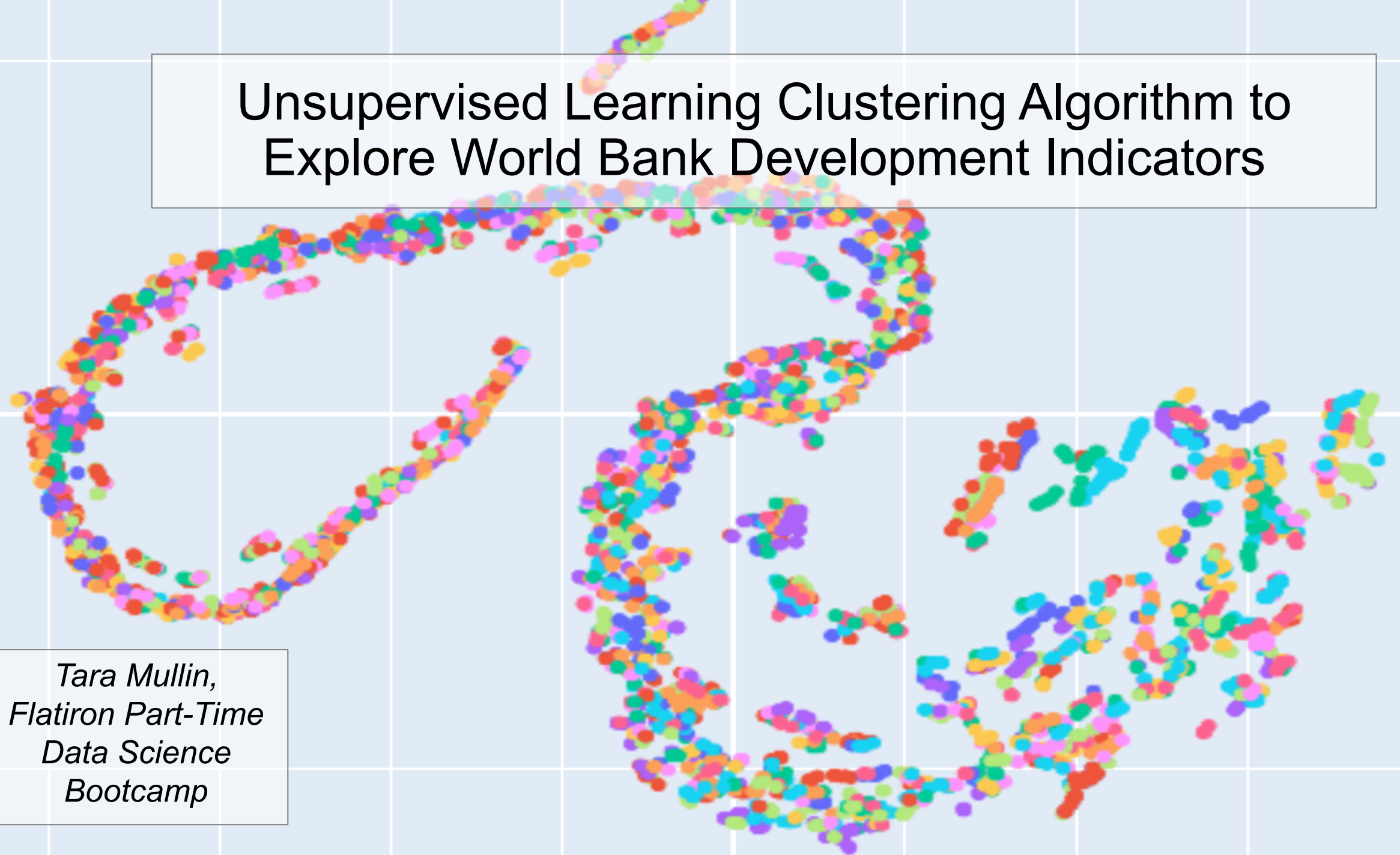# Unsupervised Learning Clustering Algorithm to Explore World Bank Development Indicators

Tara Mullin, Flatiron Part-Time Data Science Bootcamp

Dataset

- World Bank Development Indicator Data

- 1,599 countries & groups of countries

- 10 indicators

- 1990 - 2015

Factors Used to Define Clusters

1. Population density (people per sq. km of land area)

2. % of labor force that is female

3. Under 5 mortality rate (per 1,000 live births)

4. GDP per capital (current US$)

5. Death rate, crude (per 1,000 people)

6. Birth rate, crude (per 1,000 people)

7. Renewable electricity output (% of total electricity output)

8. Years of primary education

Findings

- 8 Clusters

- 2 very clear, 6 not so clear…

Findings

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Findings – Cluster -1 "Developed Countries"

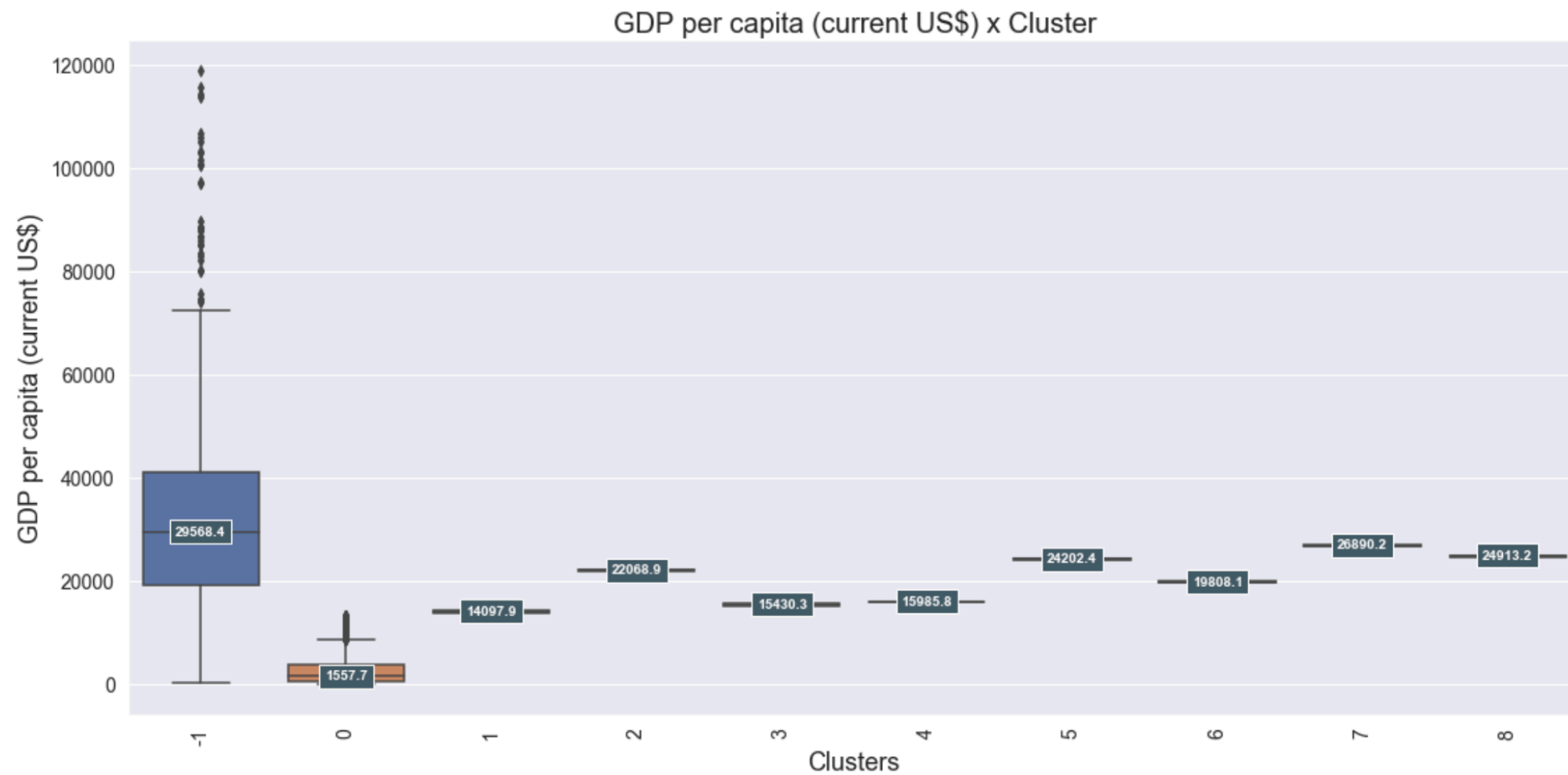|	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Highest GDP per capital
- Highest population density (urbanized)

Examples:
- USA 1992 - 2015
- Australia
- Japan
- Singapore

- High income countries 1997 – 2015
- EU since 1996

Cluster Characteristics - GDP per Capita

GDP per capita (current US$) x Cluster

Findings – Cluster 0 "Developing Countries"

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Lowest GDP per capital
- Low population density (more rural)
- Lowest life expectancy
- Highest under 5 mortality
- Highest birth and growth rate

Examples:
- China (all years)
- India (all years)
- Brazil (all years)
- Central Europe & the Baltics 1990 – 2007
- Russia 1990 – 2010
- Mexico (all years)

Cluster Characteristics – Birth Rate
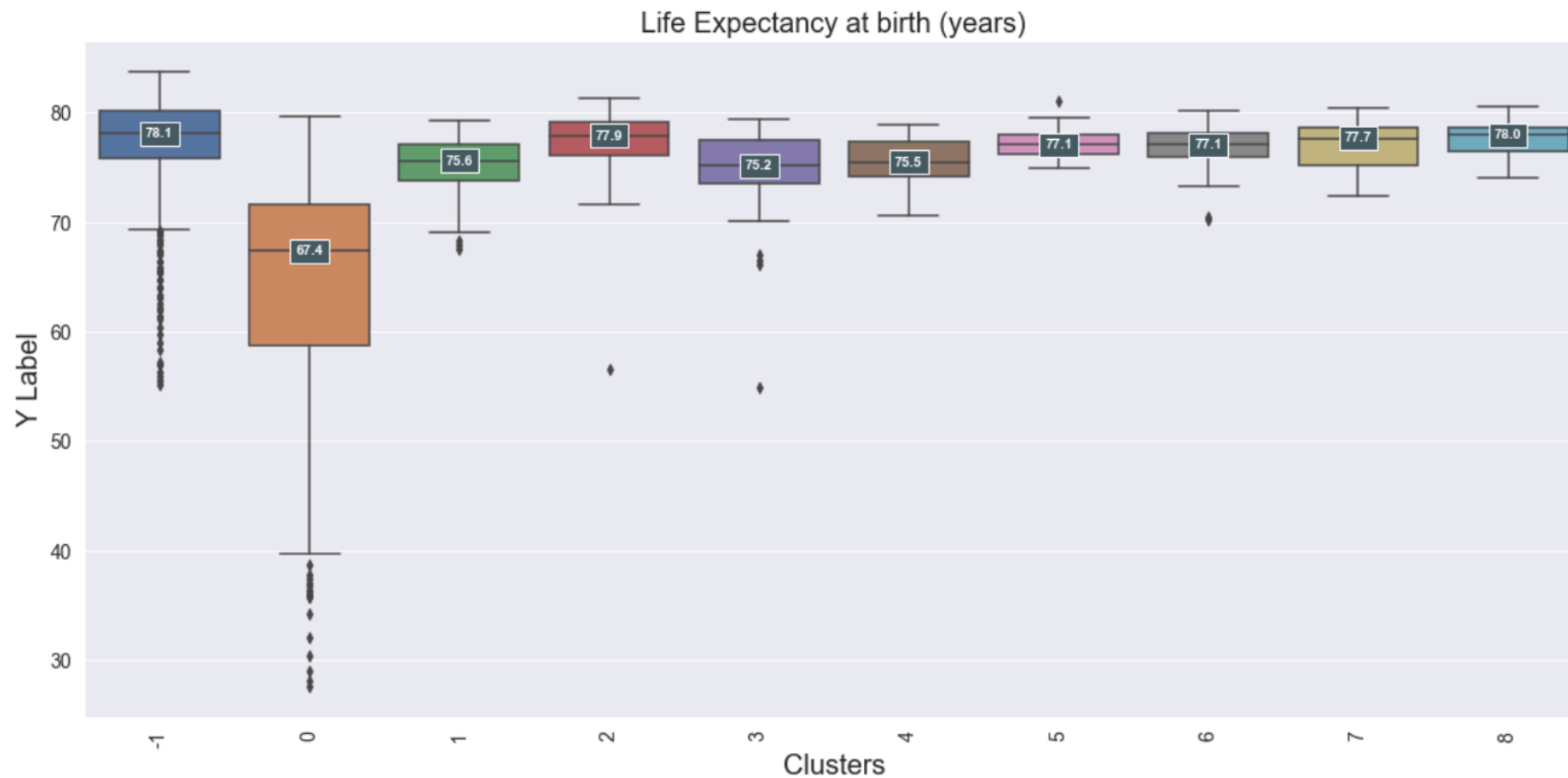

Birth rate, crude (per 1,000 people) x Cluster

Cluster Characteristics – Mortality Rate

Mortality rate, under-5 (per 1,000 live births) x Cluster

Cluster Characteristics – Life Expectancy

Life Expectancy at birth (years)

Findings – Cluster 1

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Second lowest GDP per capital
- Low population density
- Fairly low female participation in labor force
- Low population growth

Examples:
- "Small states" 2012 – '14
- Ireland 1990 – '91
- Israel 1991 – '93
- New Zealand 1990, '98 – '01
- Greece 1996 – 1997 & 2002
- Spain 1990 – '91, '94 , '97, '00
- Russia 2011
- Poland 2008, '11, '13, '14

Findings – Cluster 2

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Mid level population density
- Second highest death rate but low birth rate (aging society?)
- Low population growth

Examples:
- Euro Area 1997 – '99, 2002
- Europe & Central Asia – 2015
- Greece 2004 – '05, '12 – '14
- High Income Countries 1994
- Italy 1991, '97 – '99, 2002
- Portugal 2010, '13, '14

Findings – Cluster 3

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Low per capita GDP
- Low population density
- Second lowest female participation in labor market

Examples:
- Cyprus '95 – '99
- Spain '95, '98 – '99, '01
- Russia '12
- Europe & Central Asia '03

- New Zealand '94 – '99
- Portugal '03

Findings – Cluster 4

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Lowest female labor participation
- Low population density
- Lowest death rate

Examples:
- EU 1990
- Ireland '94
- Israel '94
- Oman '07
- Qatar '95
- Russia '13
- Spain '92, '96

Findings – Cluster 5

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Lowest population density
- High GDP per capital
- Slightly higher death rate

Examples:
- Euro Area '96
- EU '03
- High Income '95 – '96
- North America 1991
- USA '90 – '91
- UK '96

Findings – Cluster 6

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Lower GDP per capita
- Highest death rate
- Lowest growth rate

Examples:
- Canada '93-'94
- Czech Republic '09 – '14
- EU '95, '98-'99, '02
- Italy 2000
- UK '91, '94

Findings – Cluster 7

	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Second highest GDP per capita
- High population density
- Lowest growth rate

Examples:
- Germany '94, '97, ' 99
- Denmark '90 – '91
- North America 1994
- Euro Area 2003

Findings – Cluster 8

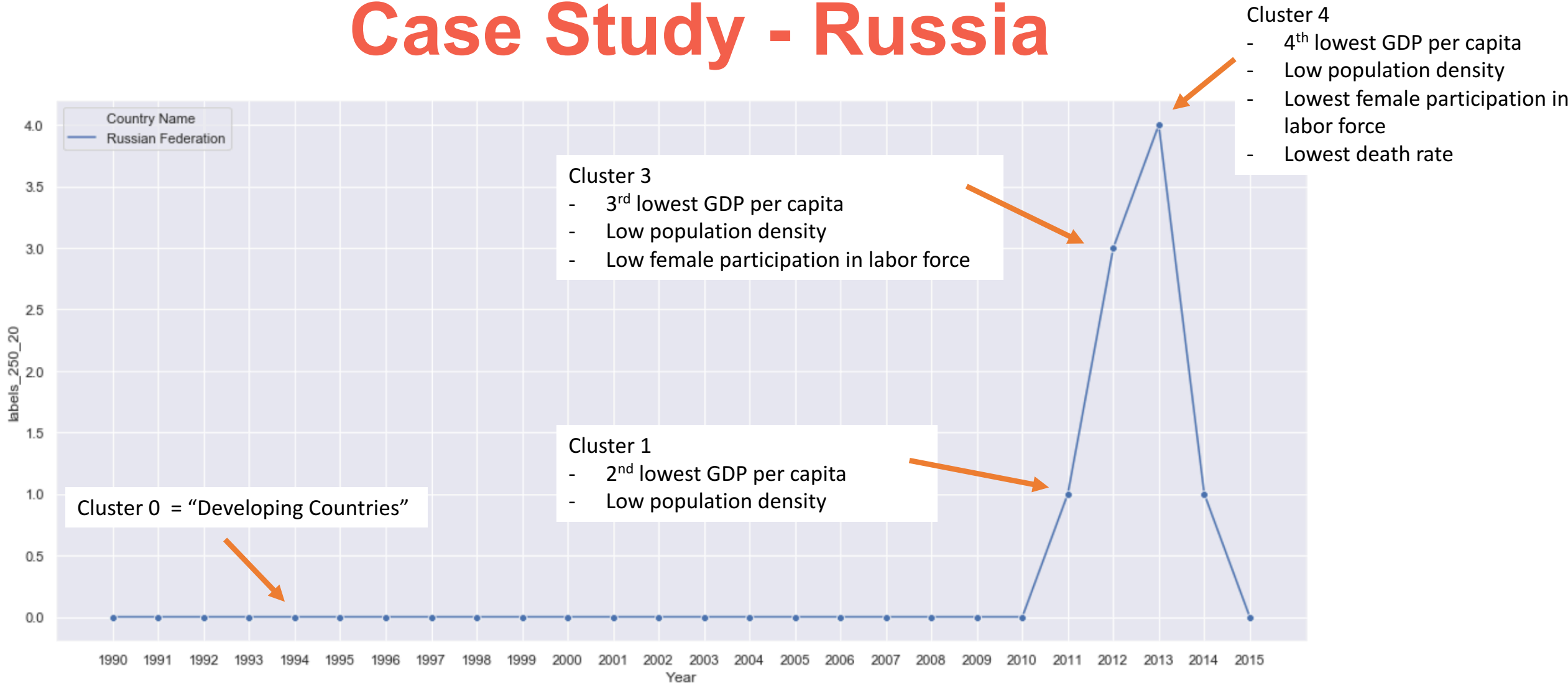|	GDP per Capita	Life Expectancy (years	Population Density (people per sq. km of land)	Labor Force Female (% of total labor force)	Under 5 Mortality (per 1,000 live births)	Death Rate (per 1,000 people)	Birth Rate (per 1,000 people)	Population Growth (annual %)
-1	$ 29,568.40	78	120	44%	7	8	13	0.8%
0	$ 1,557.70	67	50	42%	50	8	26	1.7%
1	$ 14,097.90	76	55	41%	8	9	13	0.5%
2	$ 22,068.90	78	89	43%	7	10	12	0.6%
3	$ 15,430.30	75	42	40%	9	8	14	0.8%
4	$ 15,985.80	76	56	38%	9	7	13	0.9%
5	$ 24,202.40	77	32	44%	7	9	13	0.7%
6	$ 19,808.10	77	118	43%	7	10	11	0.4%
7	$ 26,890.20	78	97	44%	7	9	13	0.4%
8	$ 24,890.20	78	33	44%	6	9	12	0.6%

Characteristics:
- Low population density
- Lowest under 5 mortality
- Slightly higher death rate and slightly lower growth rate (perhaps aging society)

Examples:
- High income countries '99 – '01
- Europe & Central Asia 2012
- North America 1992
- France '98-'99

Case Study - Russia

Cluster 4
- 4th lowest GDP per capita
- Low population density
- Lowest female participation in labor force
- Lowest death rate

Cluster 3
- 3rd lowest GDP per capita
- Low population density
- Low female participation in labor force

Cluster 1
- 2nd lowest GDP per capita
- Low population density

Cluster 0 = "Developing Countries"

- Developing country 1990 – 2010 --> Cluster 1 --> Cluster 3 --> Cluster 5 --> Cluster 1 --> Developing country status

Take-Aways

- Proof of concept using clustering to study development patterns

- Globally, since 1990, we've seen a pattern of development - countries moving into the most developed category

Use Cases

- Study development patterns & predict development paths

- Study which indicators at which levels may suggest a country is on the path to developing

- Compare countries with similar development patterns

- Compare countries with very different development patterns

Lessons Learned & Limitations

Lessons Learned:

• Clustering most useful to explore domain expertise data

• Carefully pre-screen features

• Leave more time for analysis & interpretation

Limitations:

• Missing data

• Data dependent on countries self-reporting

Future Work

Future work:

- Continue exploring findings

- Create a new model that includes:
  - Older data
  - Additional features

- Research techniques to address missing values

Thank you!

Questions?