

# Data Science II

## - Introduction to Data Visualization -

*Visualizing Part-to-Whole Relationships and Relationships*



Prof. Dr. Eduard Kromer  
Summer Semester 2024  
University of Applied Sciences Landshut

# **Visualizing Part-to-Whole Relationships**

# Part-to-Whole

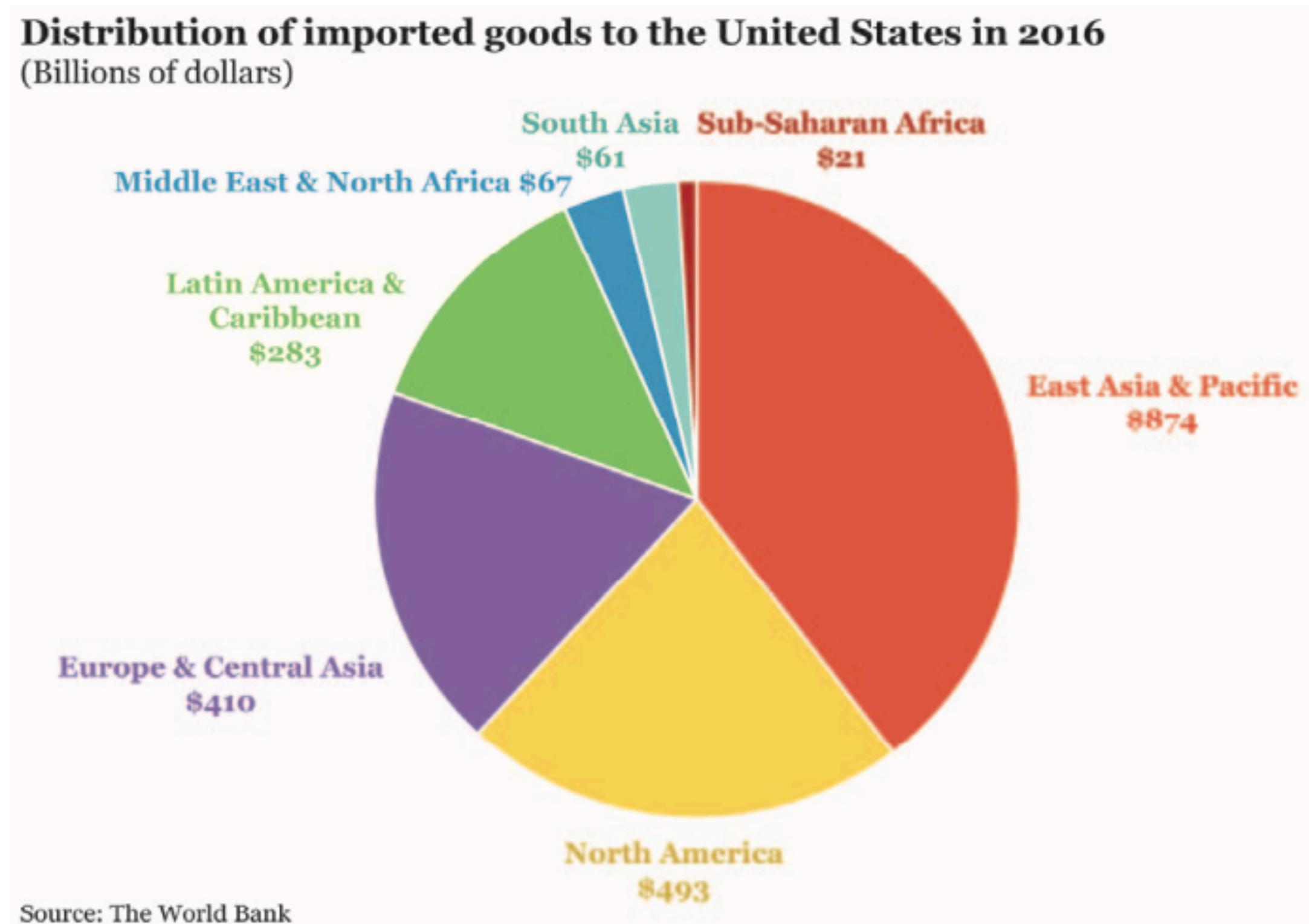
- this class of charts shows how the shares of some amount relate to the total
- most popular graph in this class is the [pie chart](#)
  - ▶ alternatives: [treemap](#) and [sunburst](#)
- graphs in this class can also be used to visualize hierarchical data

# Pie Chart



- the pie chart is very familiar to many people and familiarity can be useful
- **most important rule for pie charts:** the slices must sum to 100 percent
  - ▶ you cannot leave segments out or include segments that sum to more than 100 percent

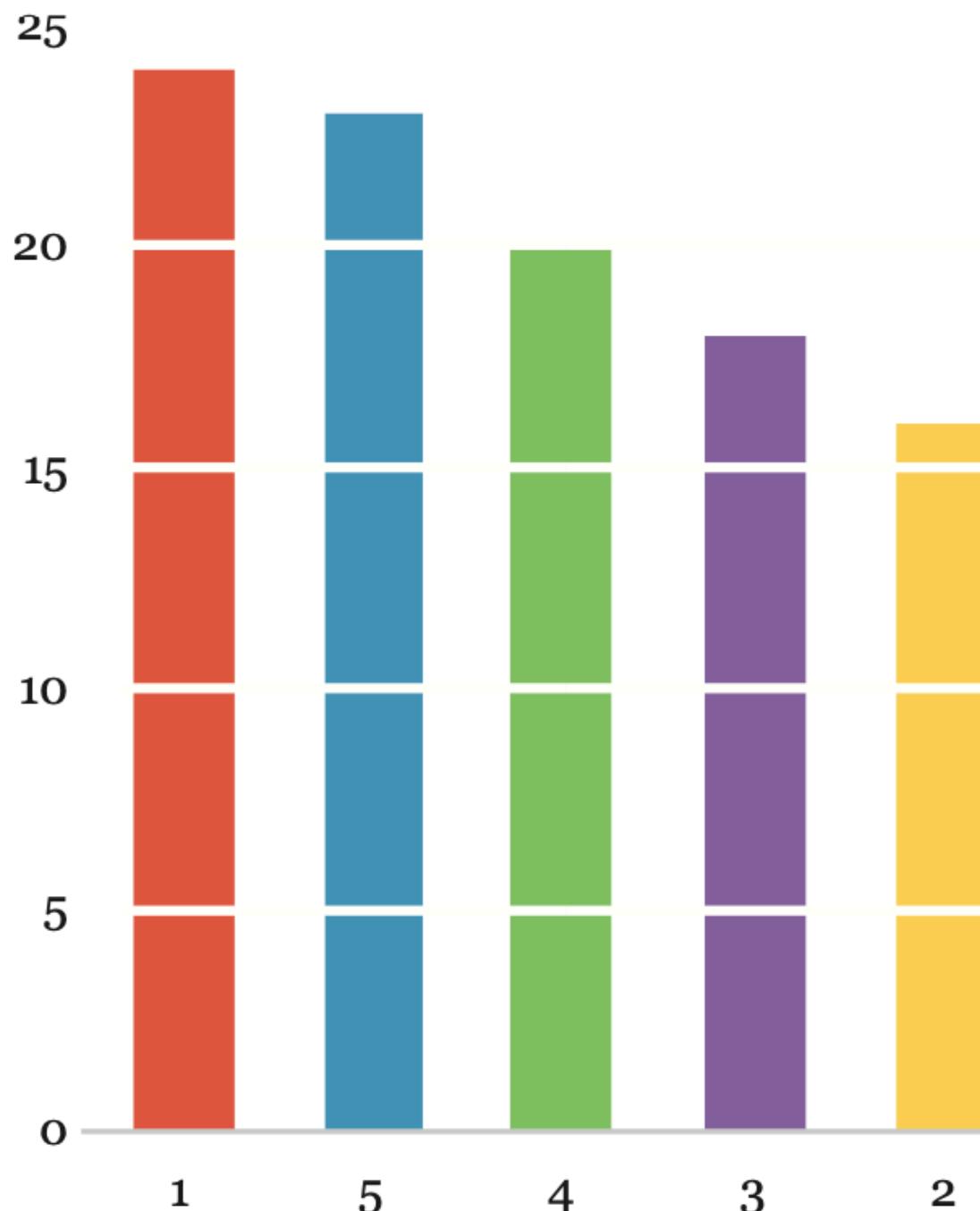
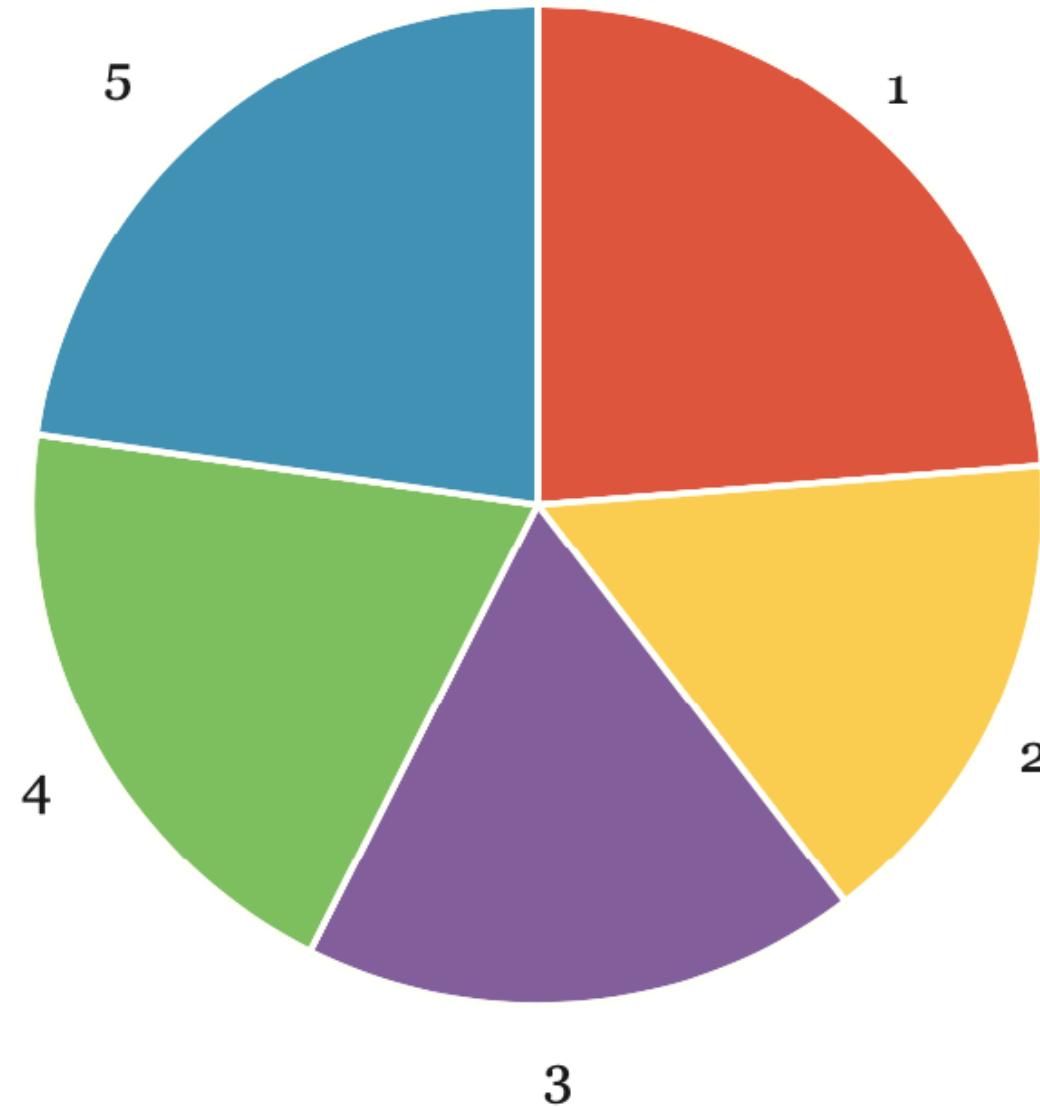
# Pie Chart



arrangement of slices: order from largest to smallest beginning at the 12-o'clock position

[sometimes it is better to order by category and not by value]

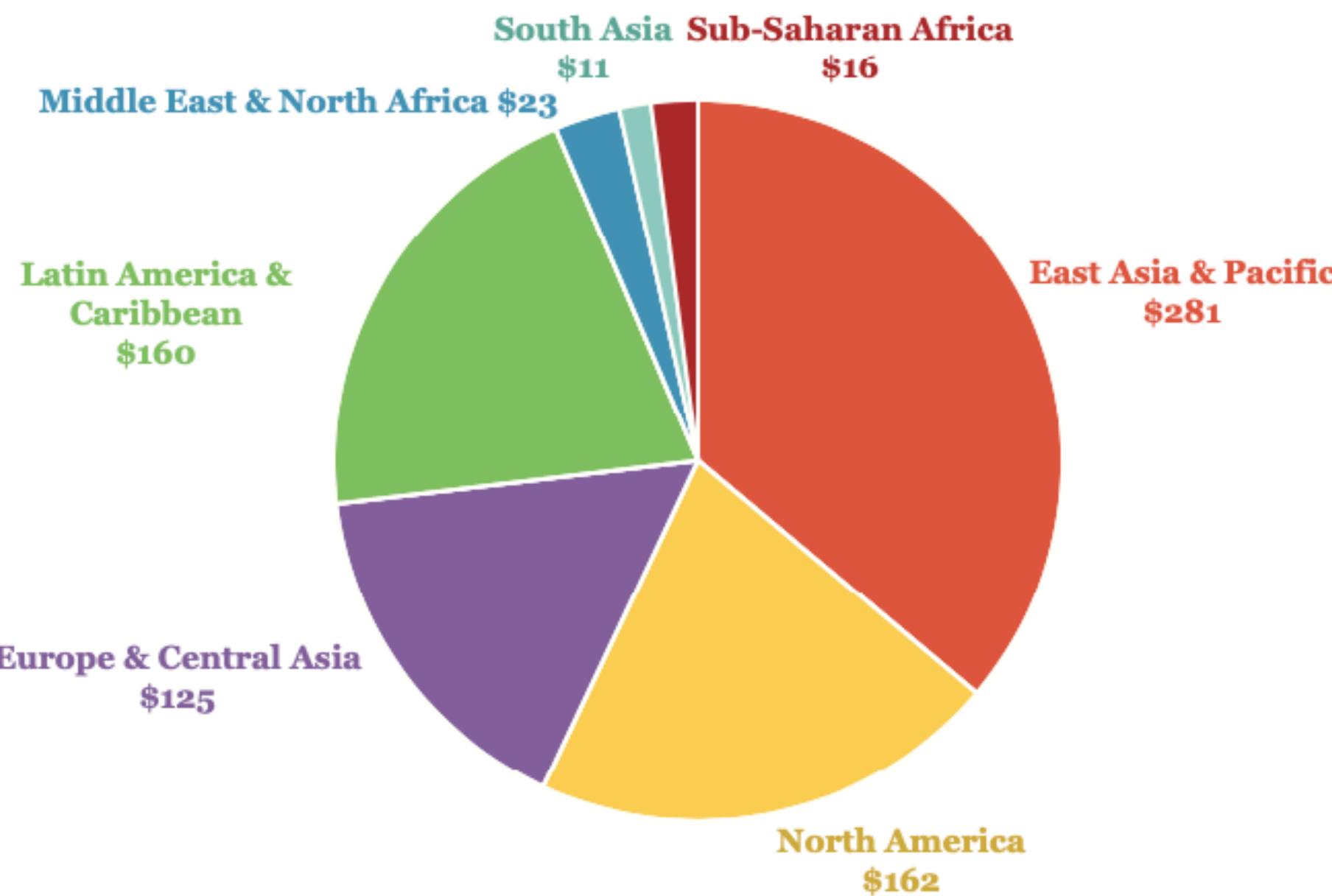
# Problems with Pie Charts



- humans cannot easily compare differently sized slices
- if the goal is to help the reader make clear and accurate determinations about the data, the pie chart is not the best choice
- avoid using pie charts with more than 5 slices (more becomes incomprehensible)

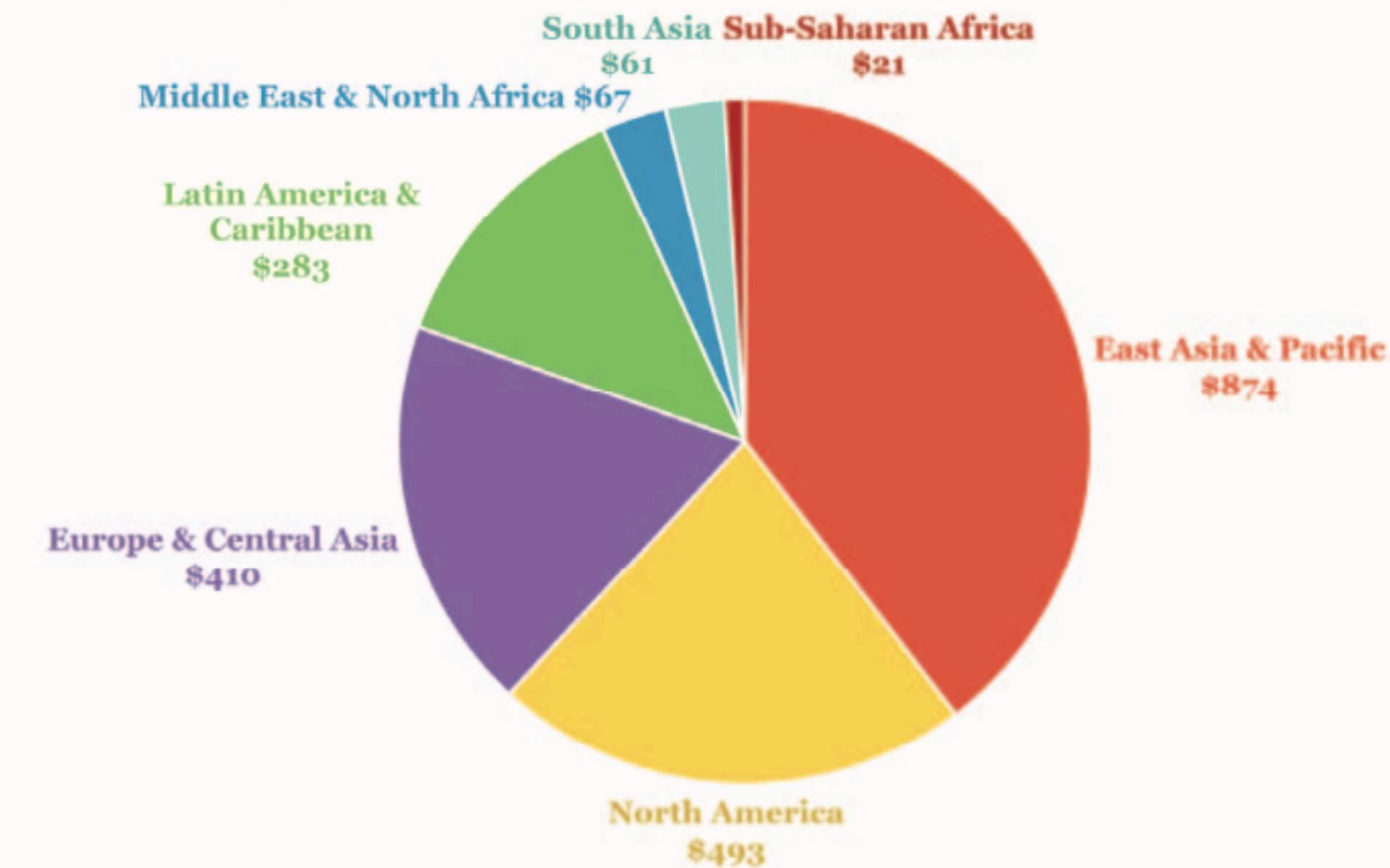
# Problems with Pie Charts

**Distribution of imported goods to the United States in 1996**  
(Billions of dollars)



Source: The World Bank

**Distribution of imported goods to the United States in 2016**  
(Billions of dollars)

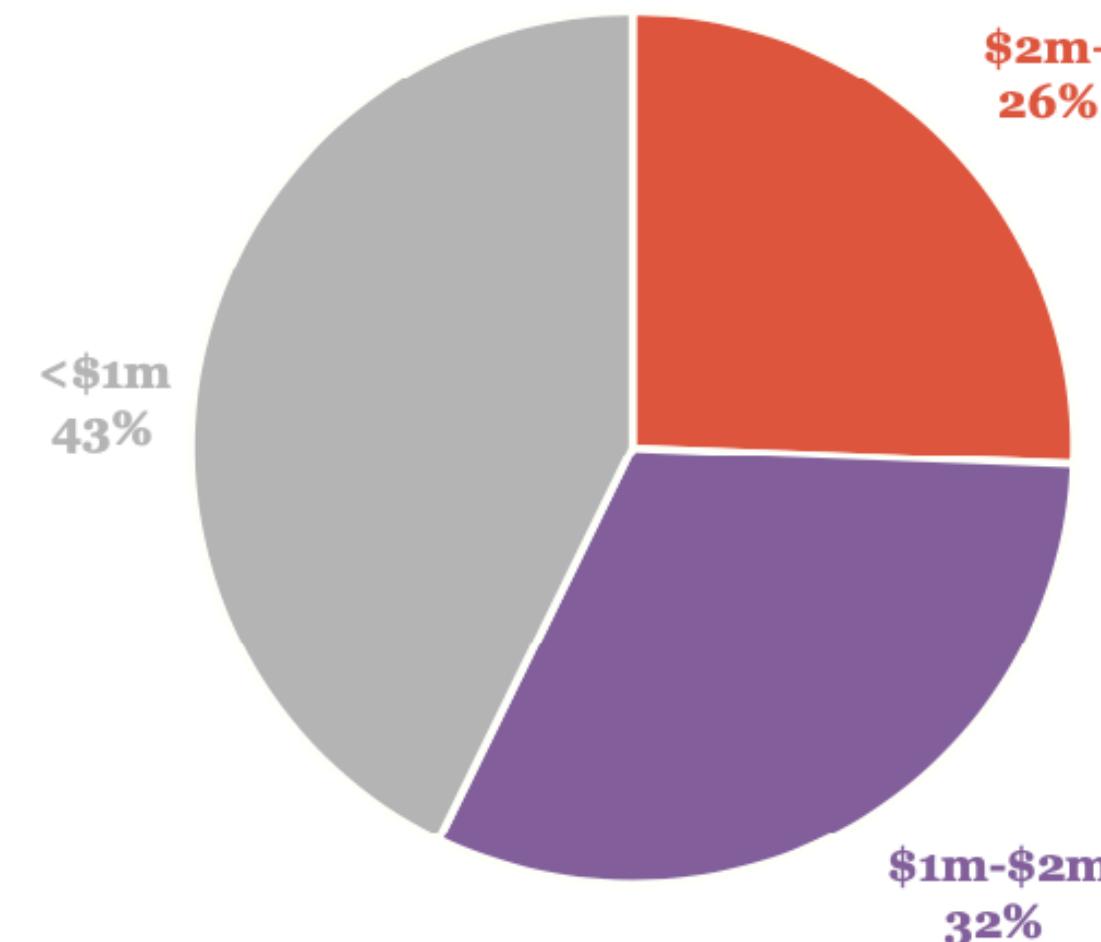


Source: The World Bank

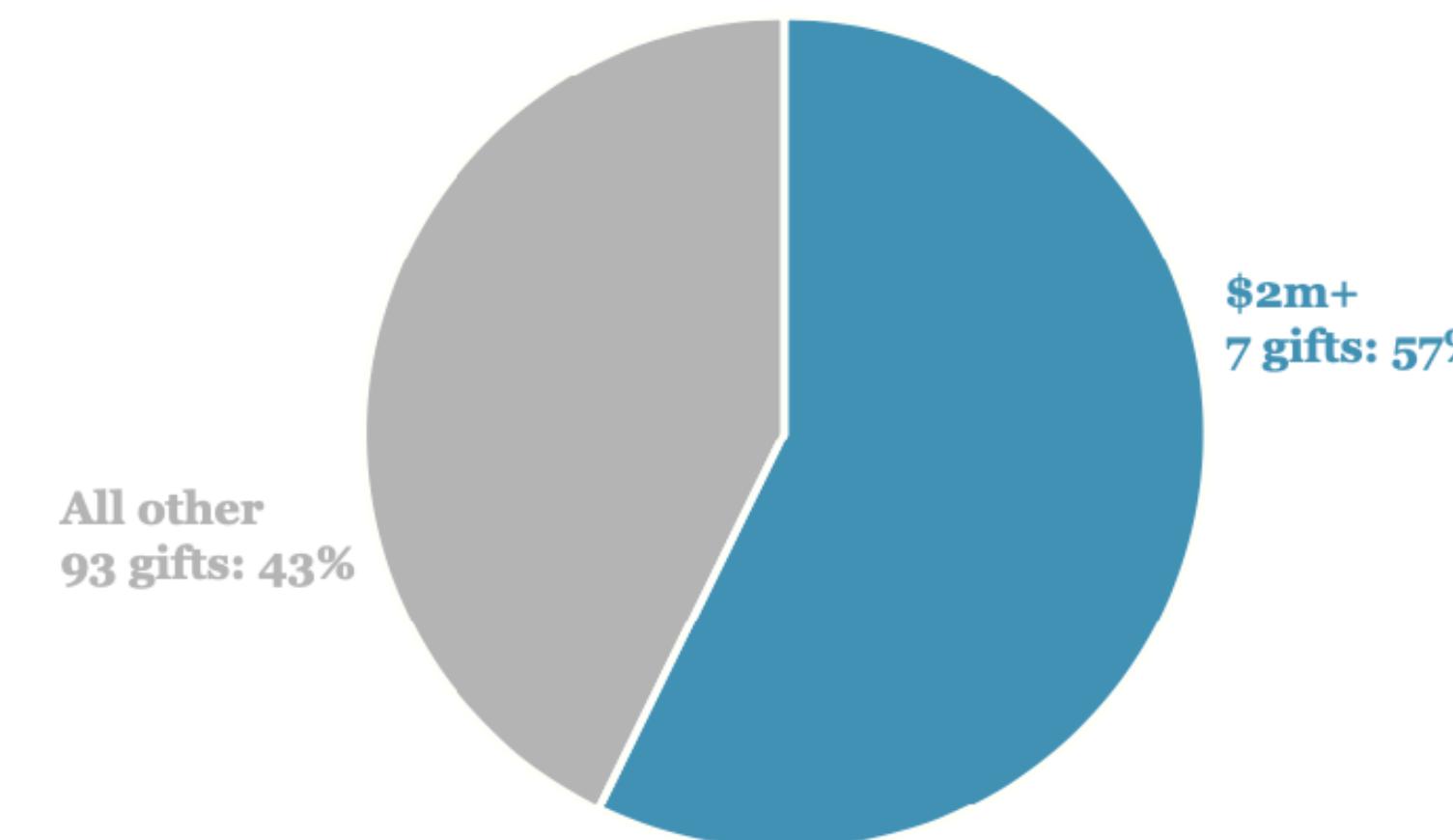
Avoid using pairs of pie charts to show changes over time.

# Advantages of Pie Charts

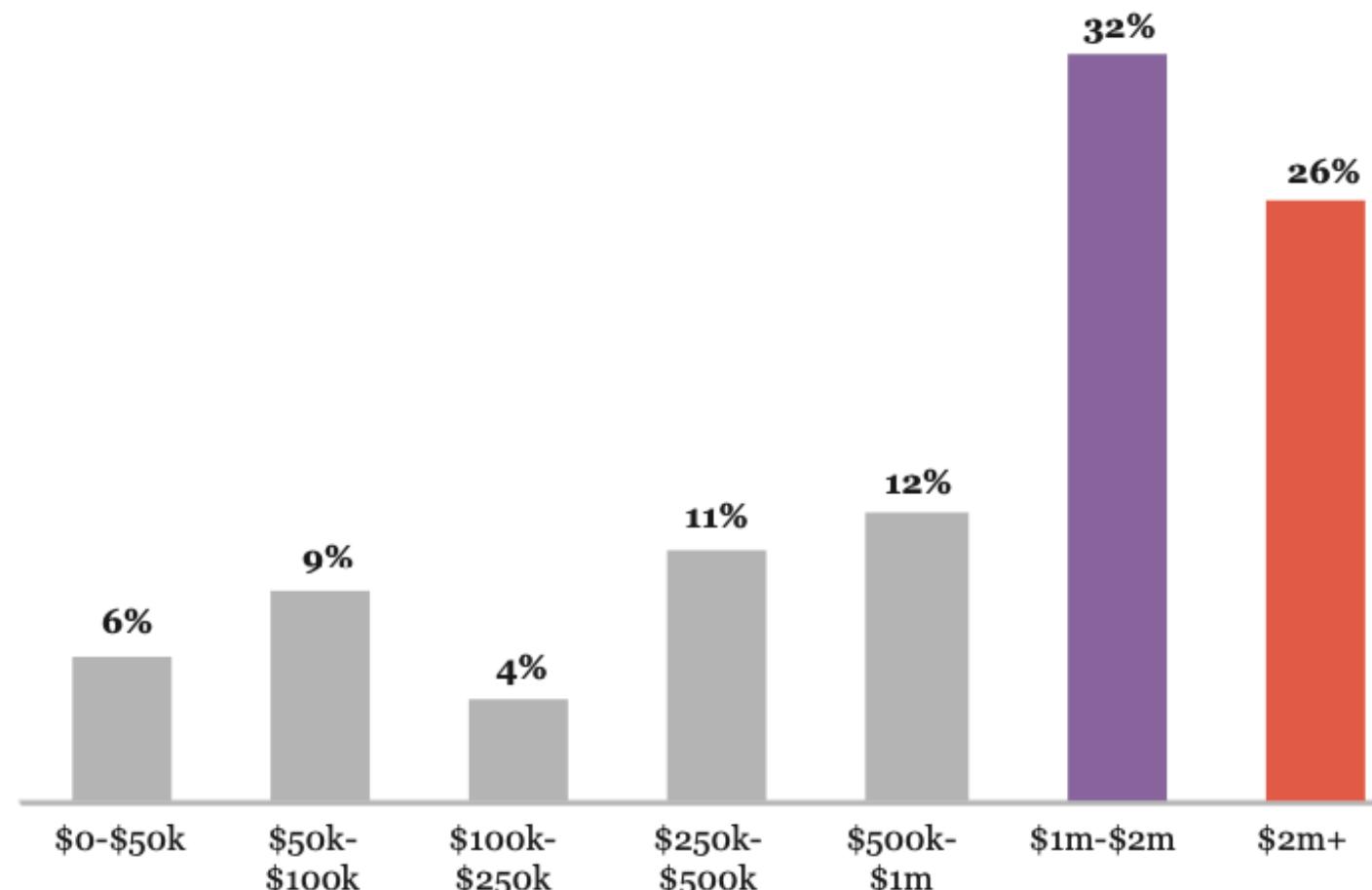
Most donations come from a small number of large gifts



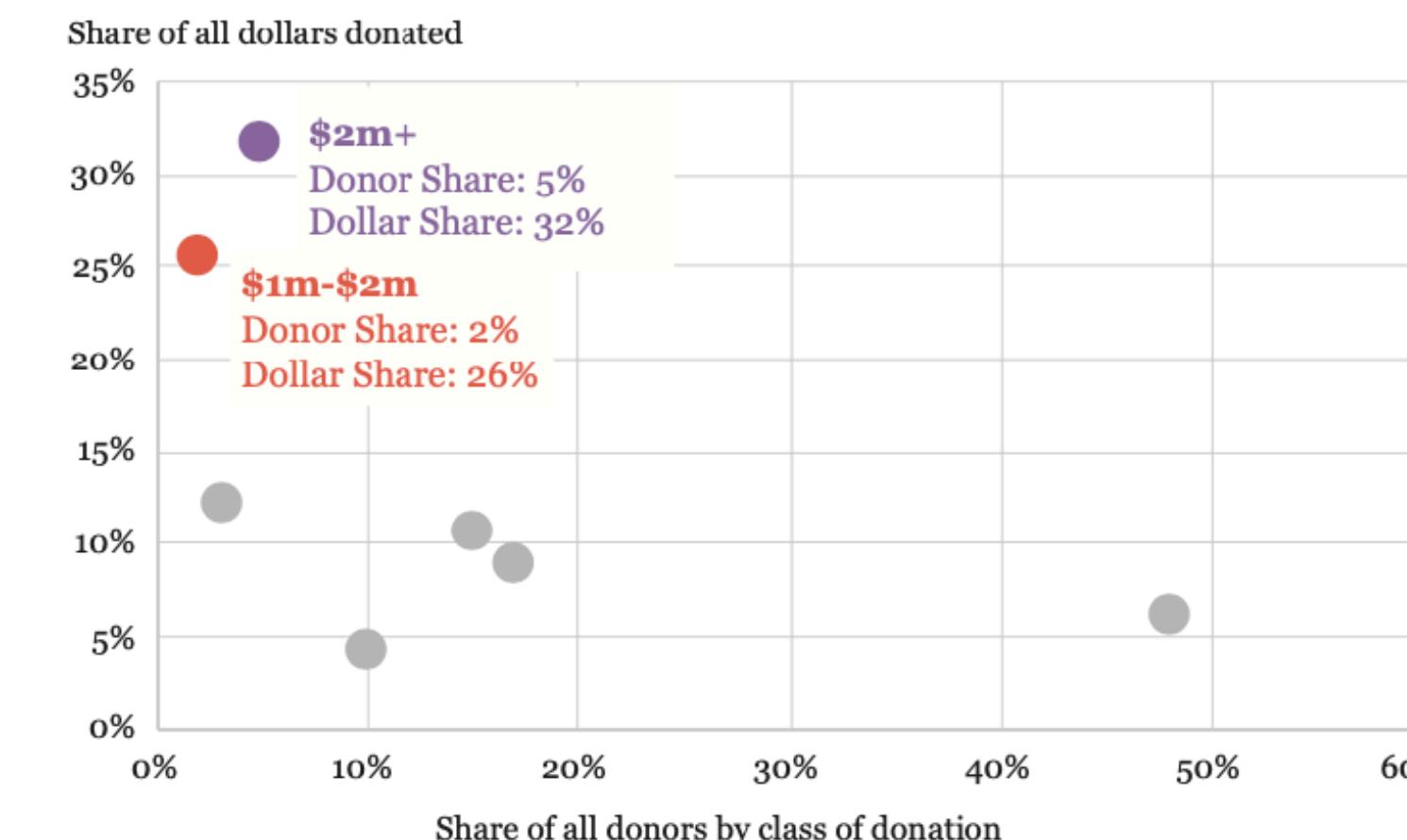
Most donations come from a small number of large gifts



Most donations come from a small number of large gifts



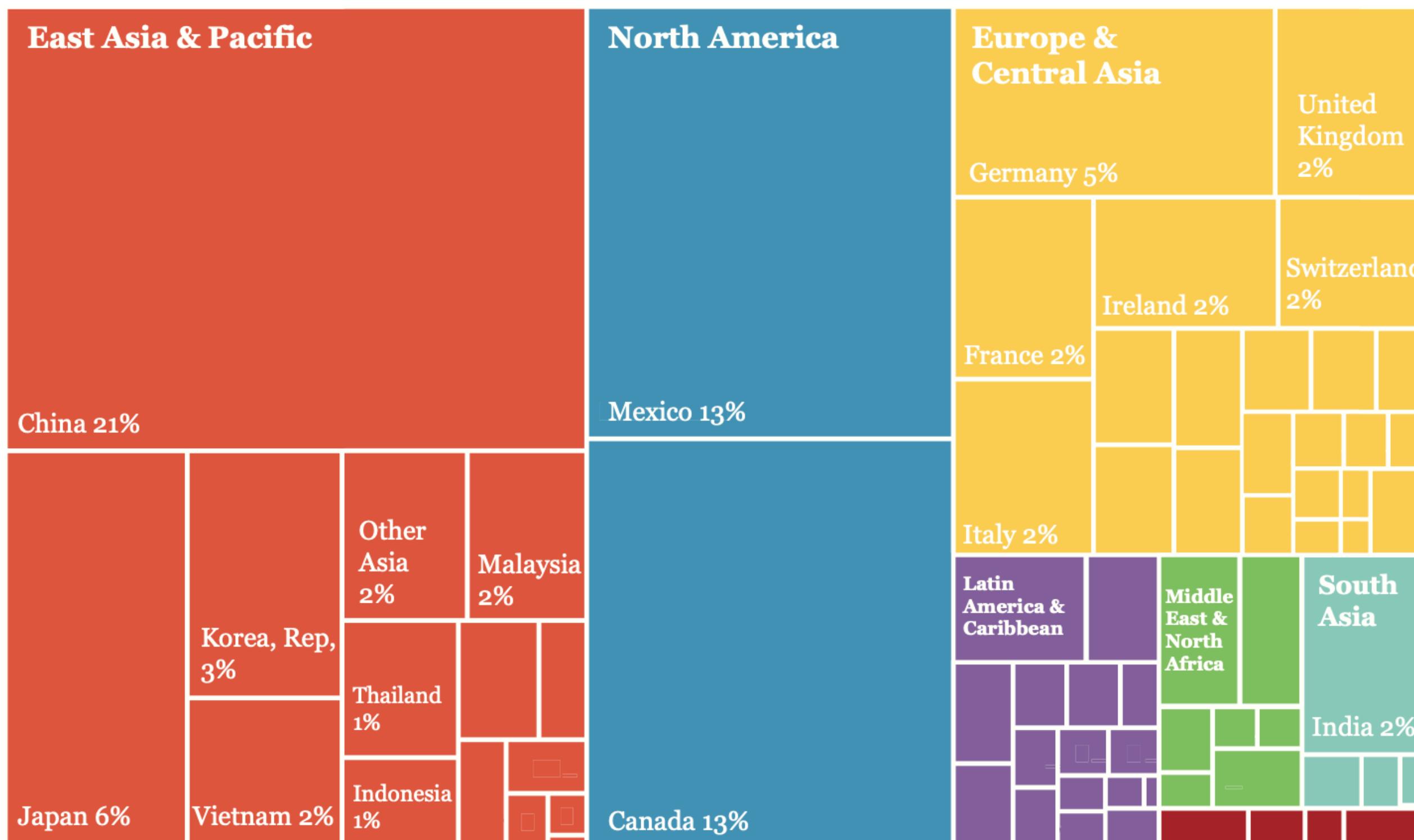
Most donations come from a small number of large gifts



You can read and understand a pie chart almost immediately.

# Treemap

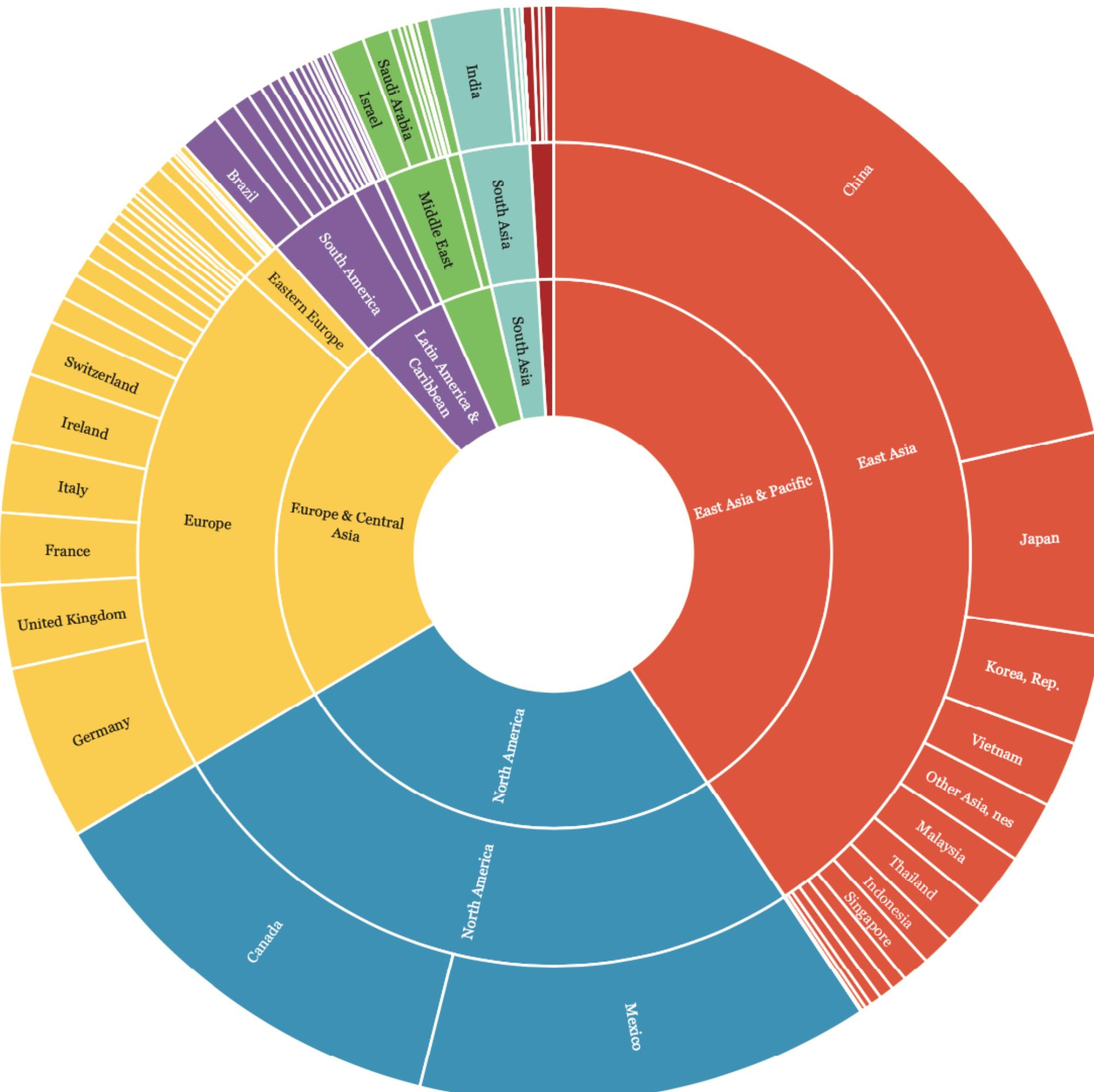
**Distribution of imported goods to the United States in 2016**



Source: The World Bank

The treemap divides sections of a square or rectangles into groups to illustrate a hierarchy or part-to-whole relationship.

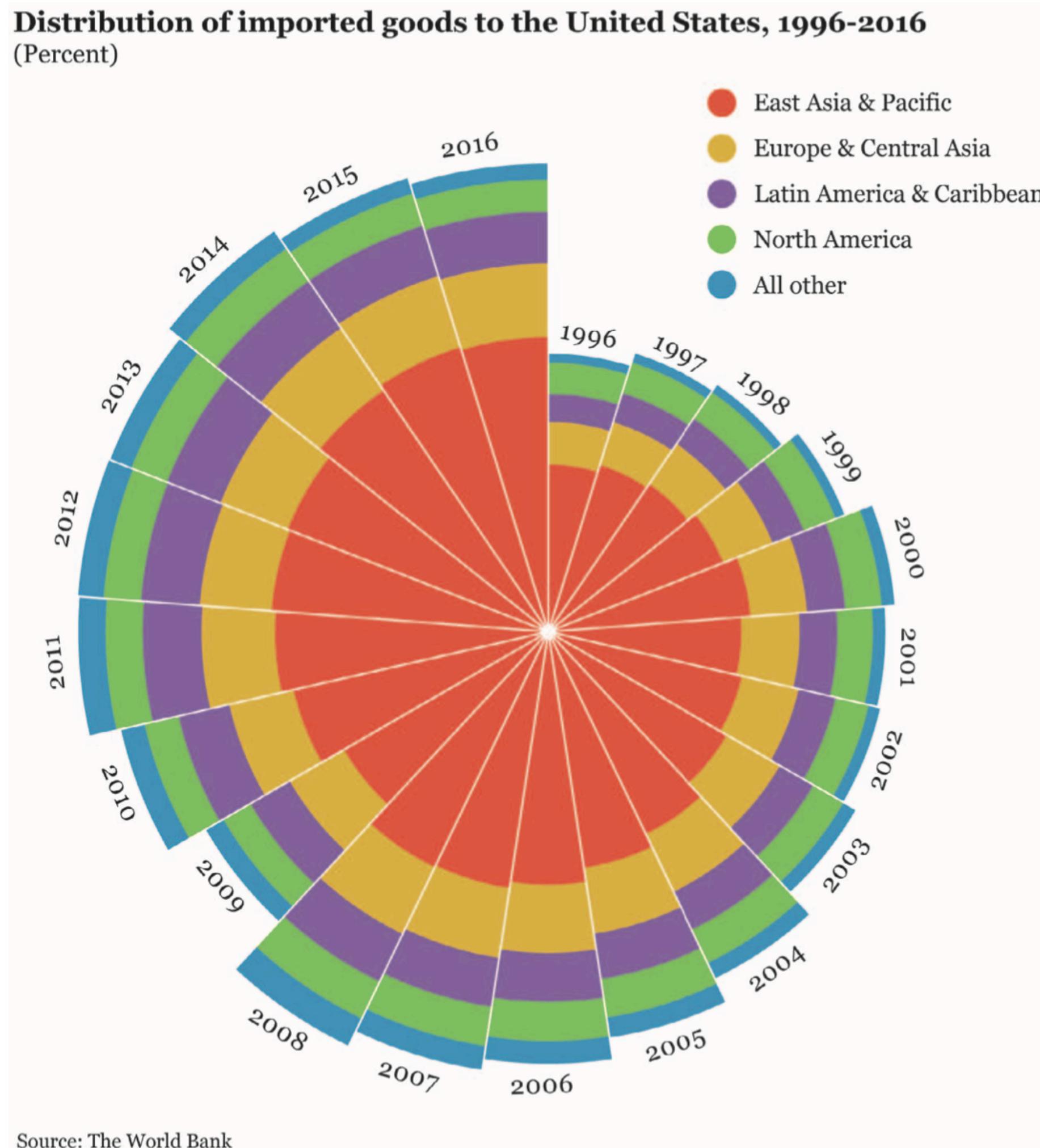
# Sunburst Diagram



The sunburst diagram shows the proportions of parts to a whole at several levels in a hierarchy.

Each ring in a sunburst diagram corresponds to a different level in the hierarchy and the slices of each ring refer to the different subgroups.

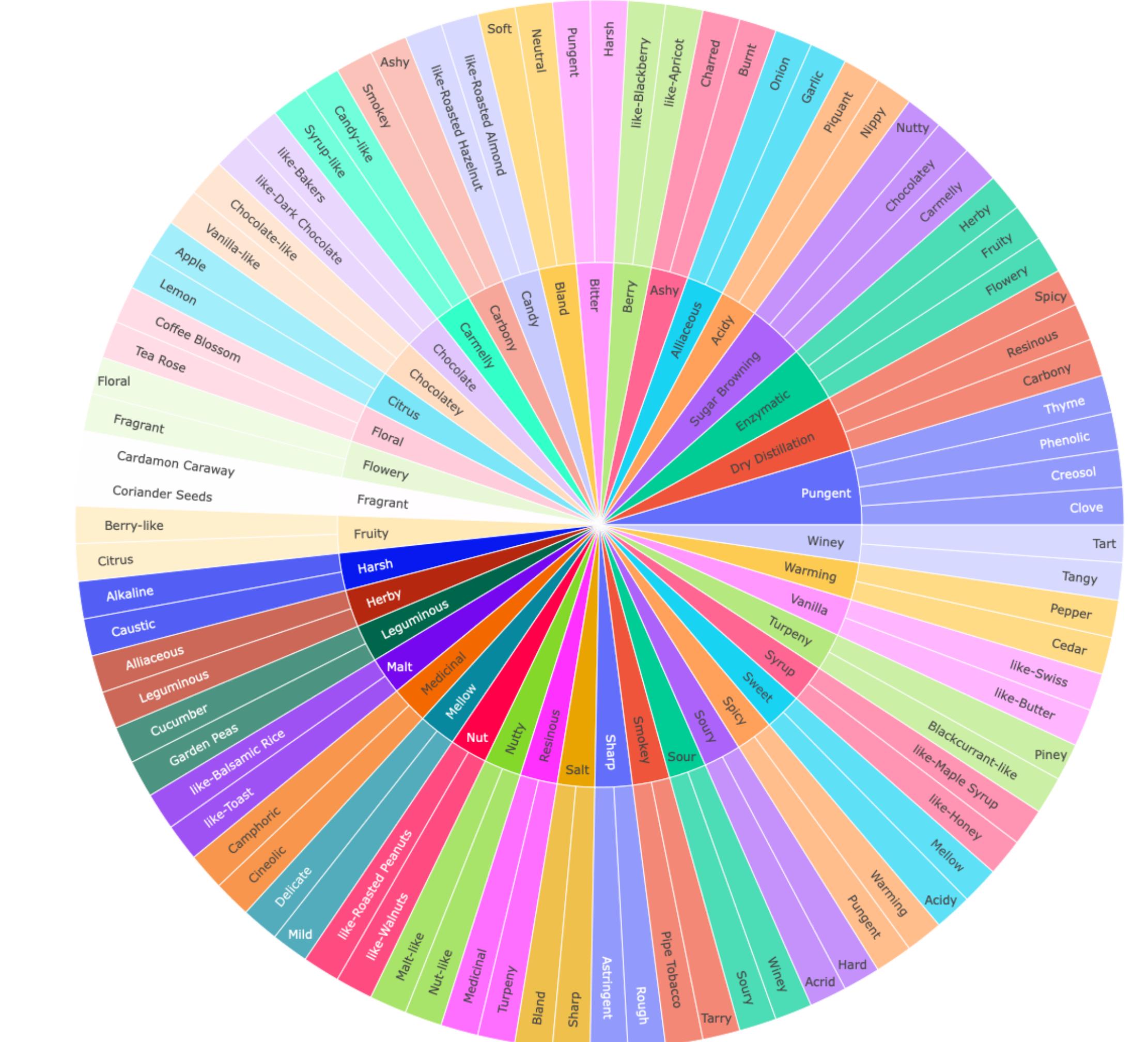
# Nightingale Chart



- the **Nightingale chart** (created by Florence Nightingale) is sometimes called the **coxcomb** or **rose diagram**
- in this chart type the slices have been expanded in different directions
  - ▶ the area of each slice represents its value relative to the whole
  - ▶ the slices are arranged along the time dimension
- this chart shows **both** changes over time and part-to-whole relationship

# Visualizing Sunburst Charts

- read the [plotly tutorial](#) for sunburst charts
- use the data in [coffee.csv](#) to generate a sunburst chart with plotly

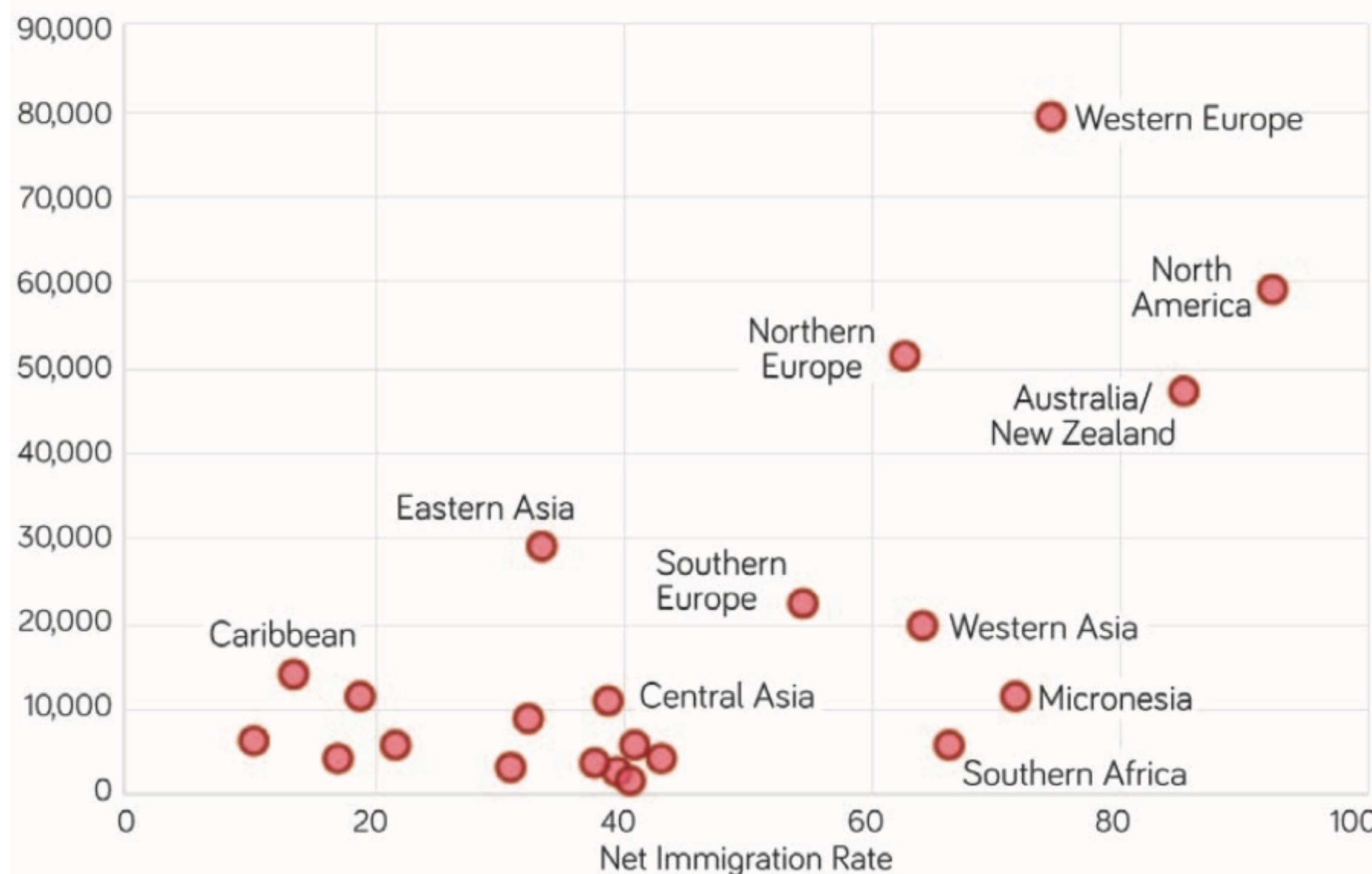


# Visualizing Relationships

# Scatterplot

Positive relationship between the net immigration rate and per capita GDP

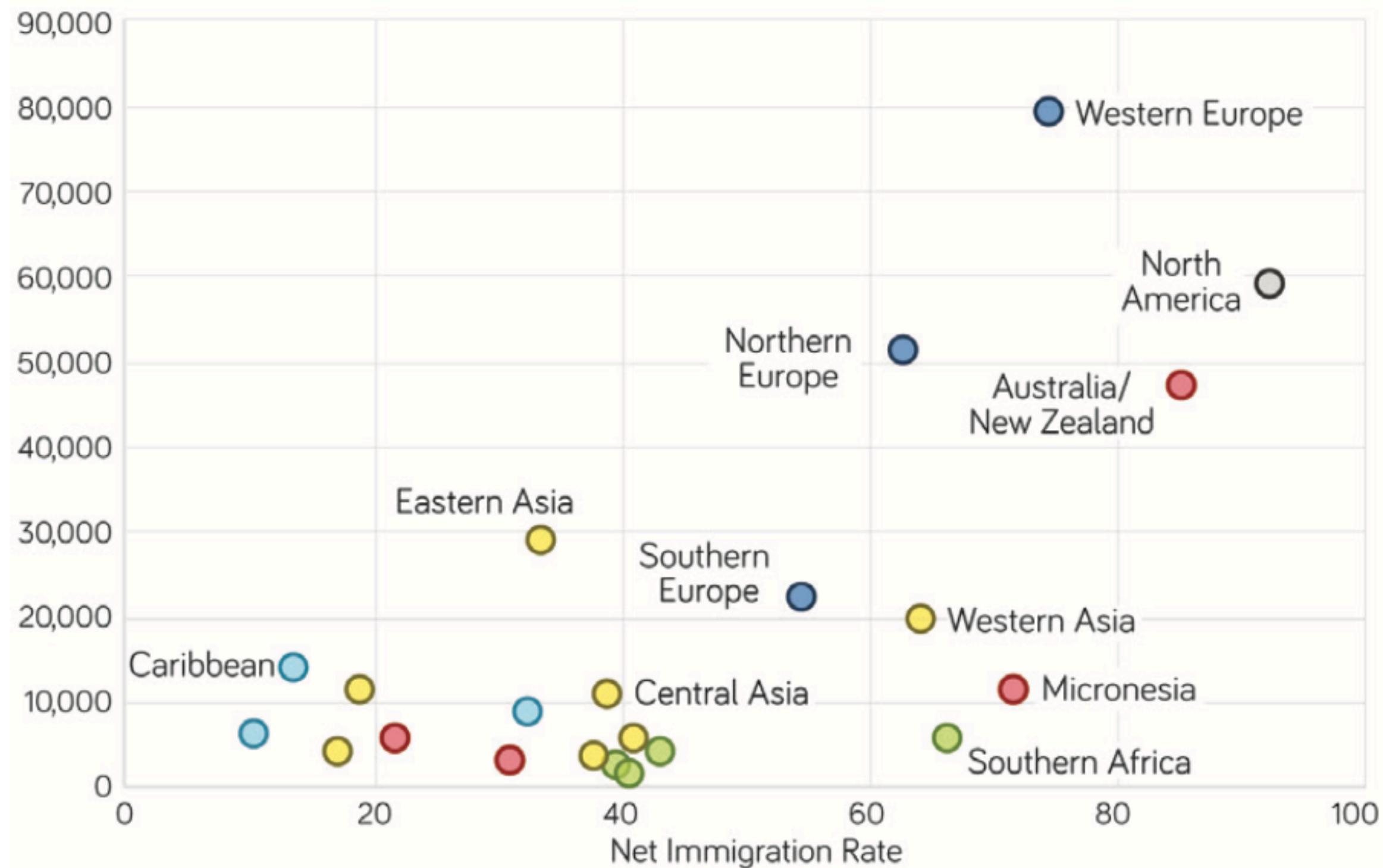
(Per capita GDP)



- the most common visualization to illustrate correlations between two variables
  - ▶ one variable is plotted along a horizontal axis and the other one along a vertical axis
- the axes do not need to start at zero

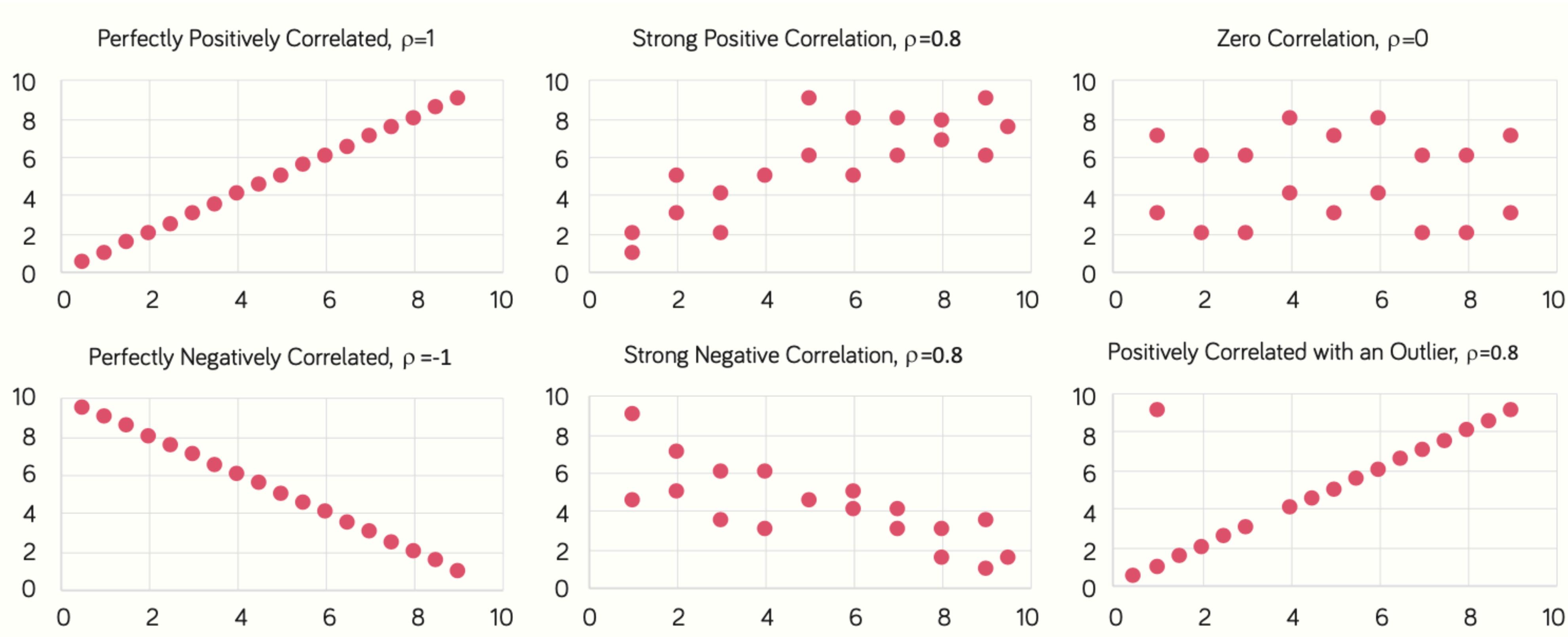
# Scatterplot

Positive relationship between the net immigration rate and per capita GDP  
(Per capita GDP)

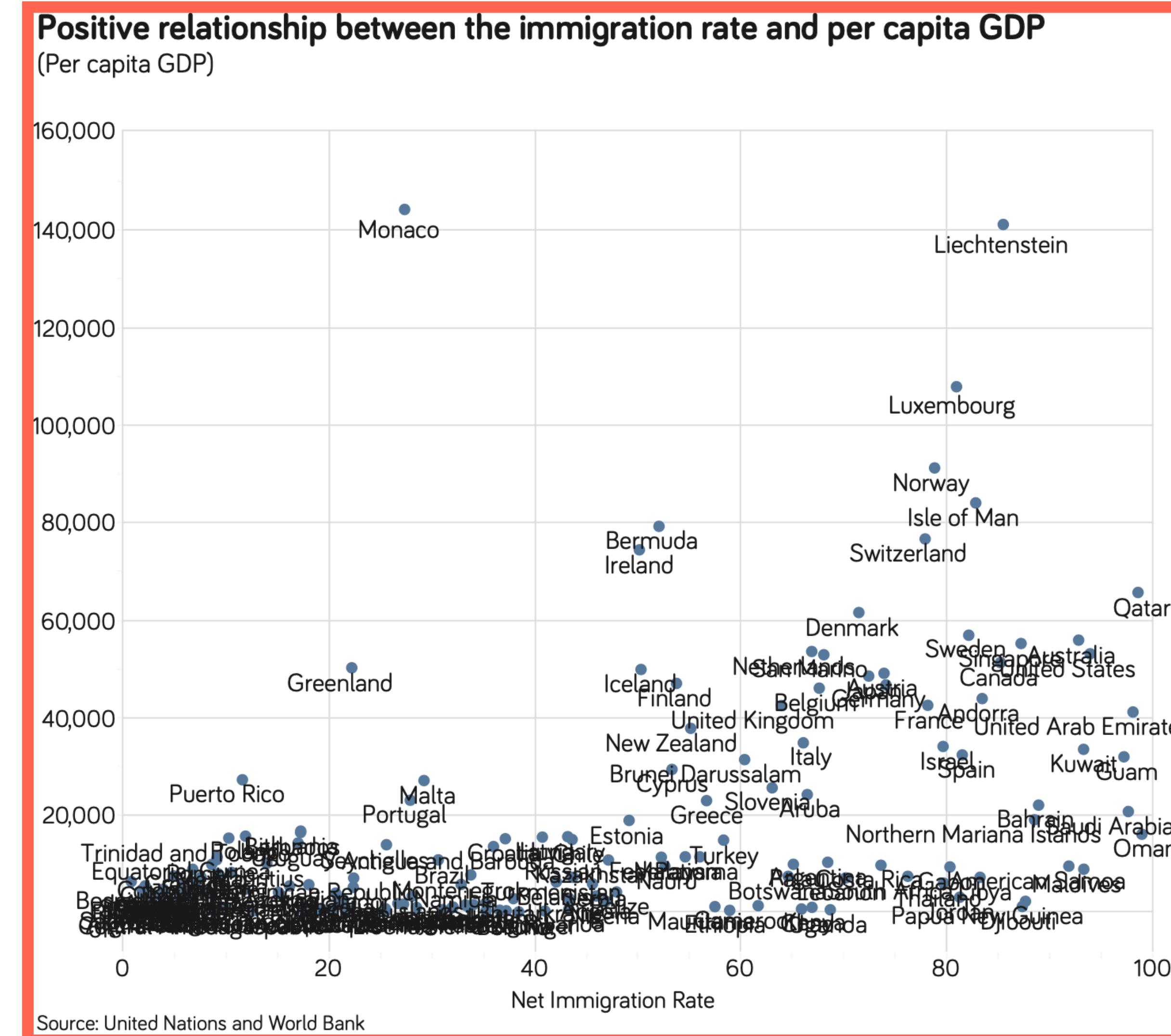


- a scatterplot can help the reader see whether two variables are associated with one another
- two variables can be *positively* or *negatively* or *not correlated*

# Scatterplot



# Scatter Plot

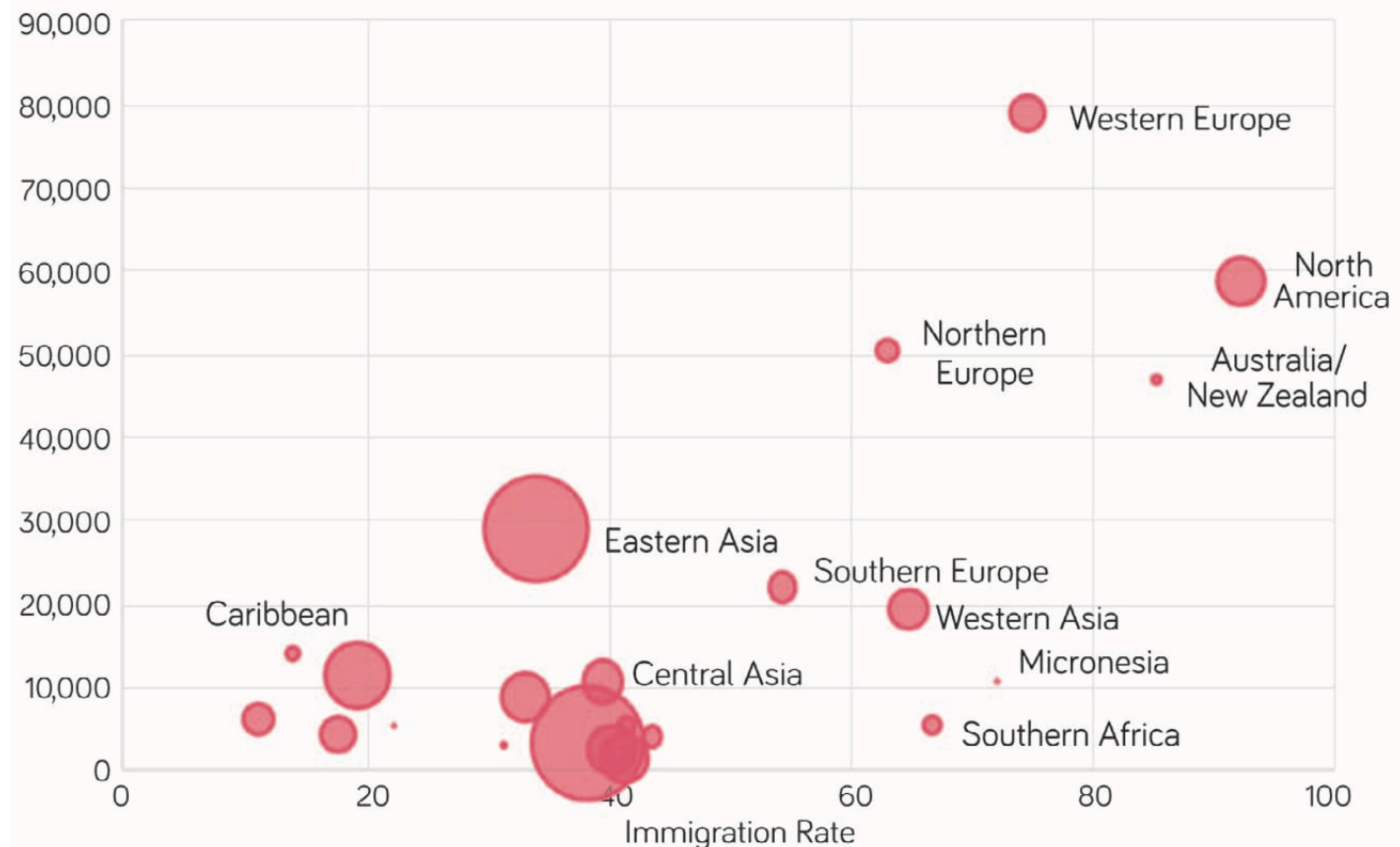


**BAD — Don't do that!**

# Bubble Plot

**Positive relationship between the immigration rate and per capita GDP**

(Per capita GDP; Size of bubble denotes population)

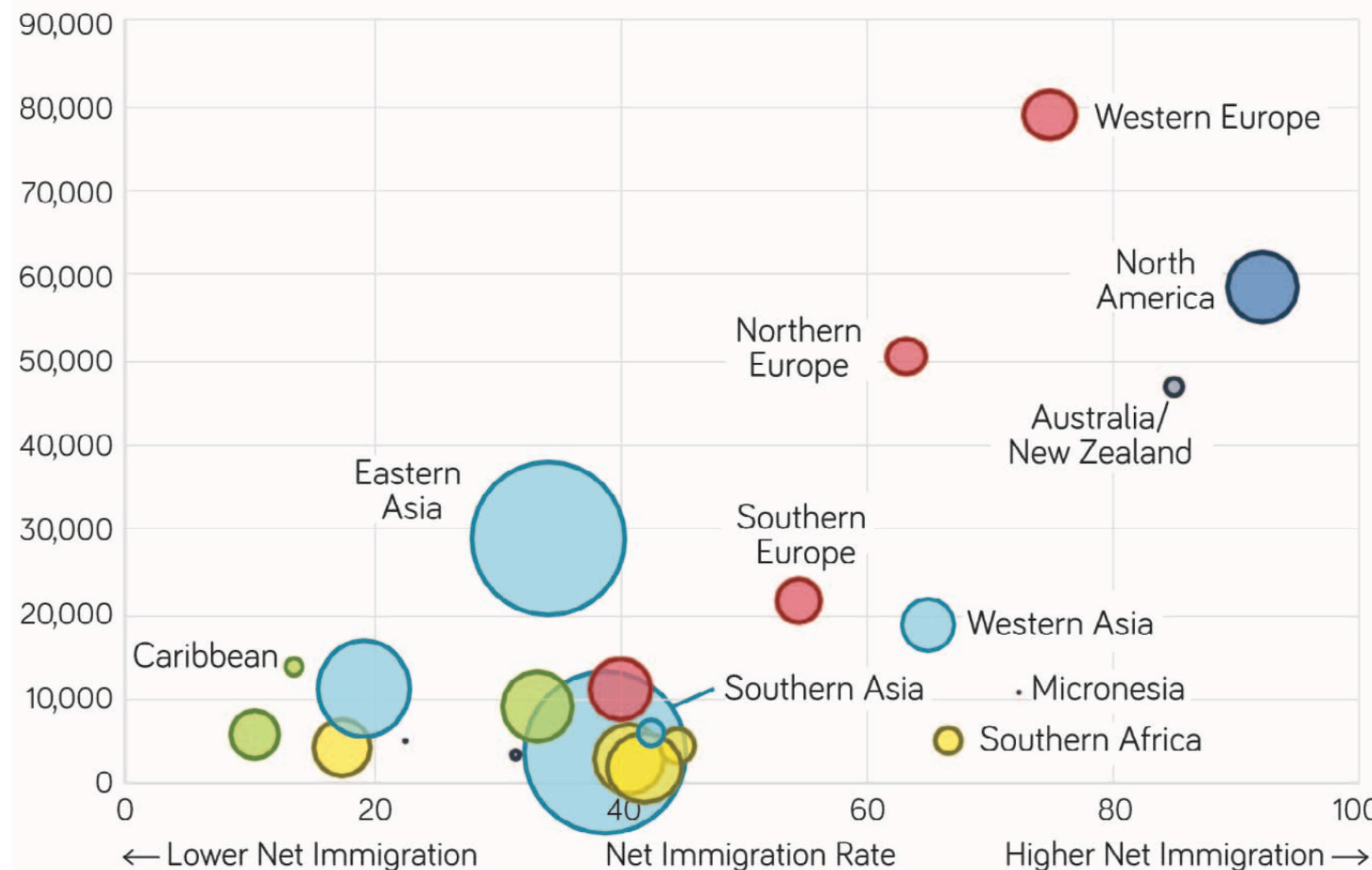


- the scatterplot can be transformed into a bubble plot by varying the sizes of the circles to a third variable
- the data points can be circles or any other shape that does not distort our perception of the data

# Bubble Plot

**Positive relationship between the net immigration rate and per capita GDP**

(Per capita GDP; Size of bubble denotes population)



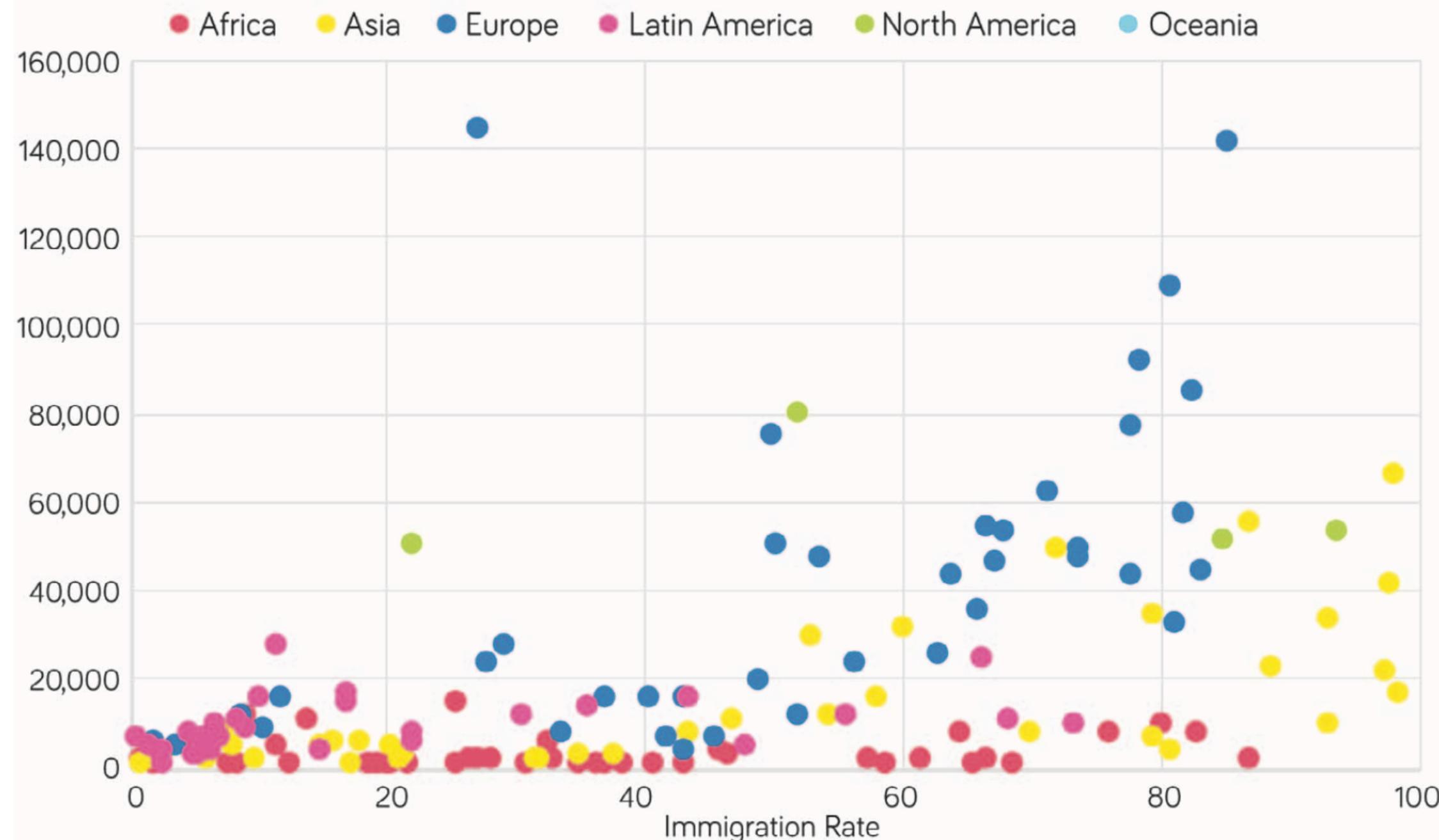
Source: United Nations and World Bank

- more colors can be added to denote another variable, such as the region of the world

# Choosing the Right Scale

**Positive relationship between the immigration rate and per capita GDP**

(Per capita GDP)

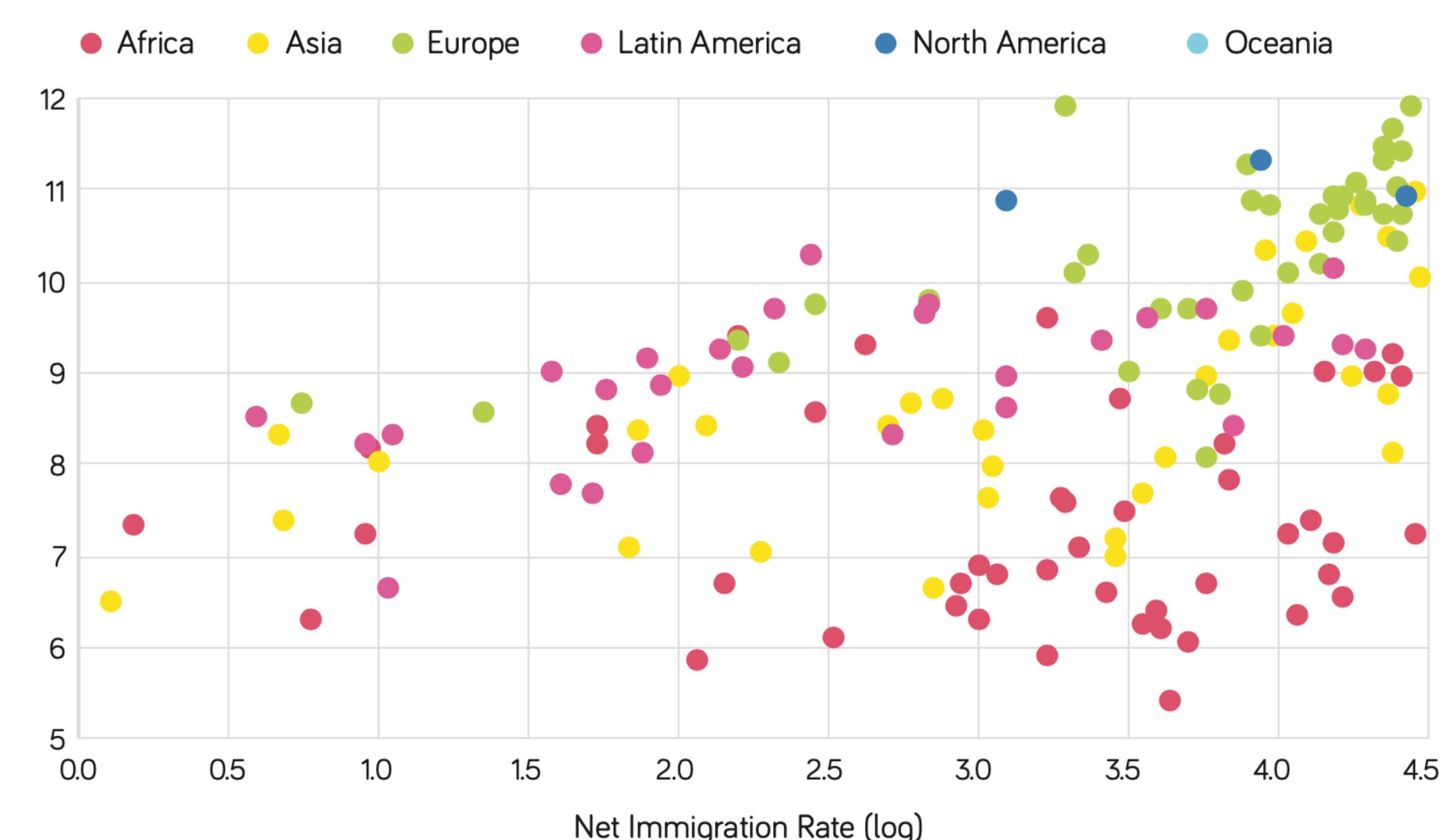


Source: United Nations and World Bank

**linear scale**

**Positive relationship between the immigration rate and per capita GDP**

(Per capita GDP, log)



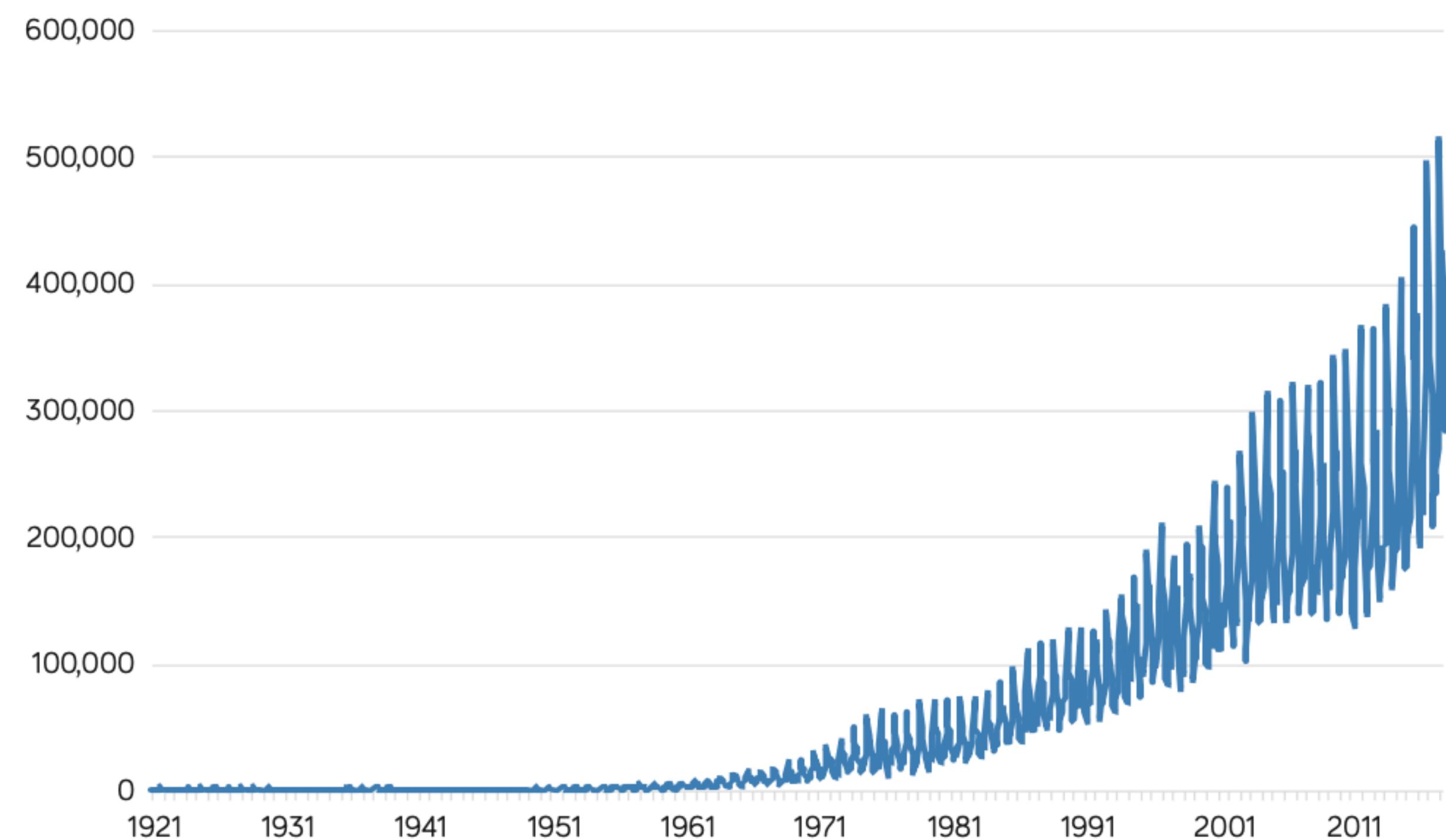
Source: United Nations and World Bank

**log scale**

# Choosing the Right Scale

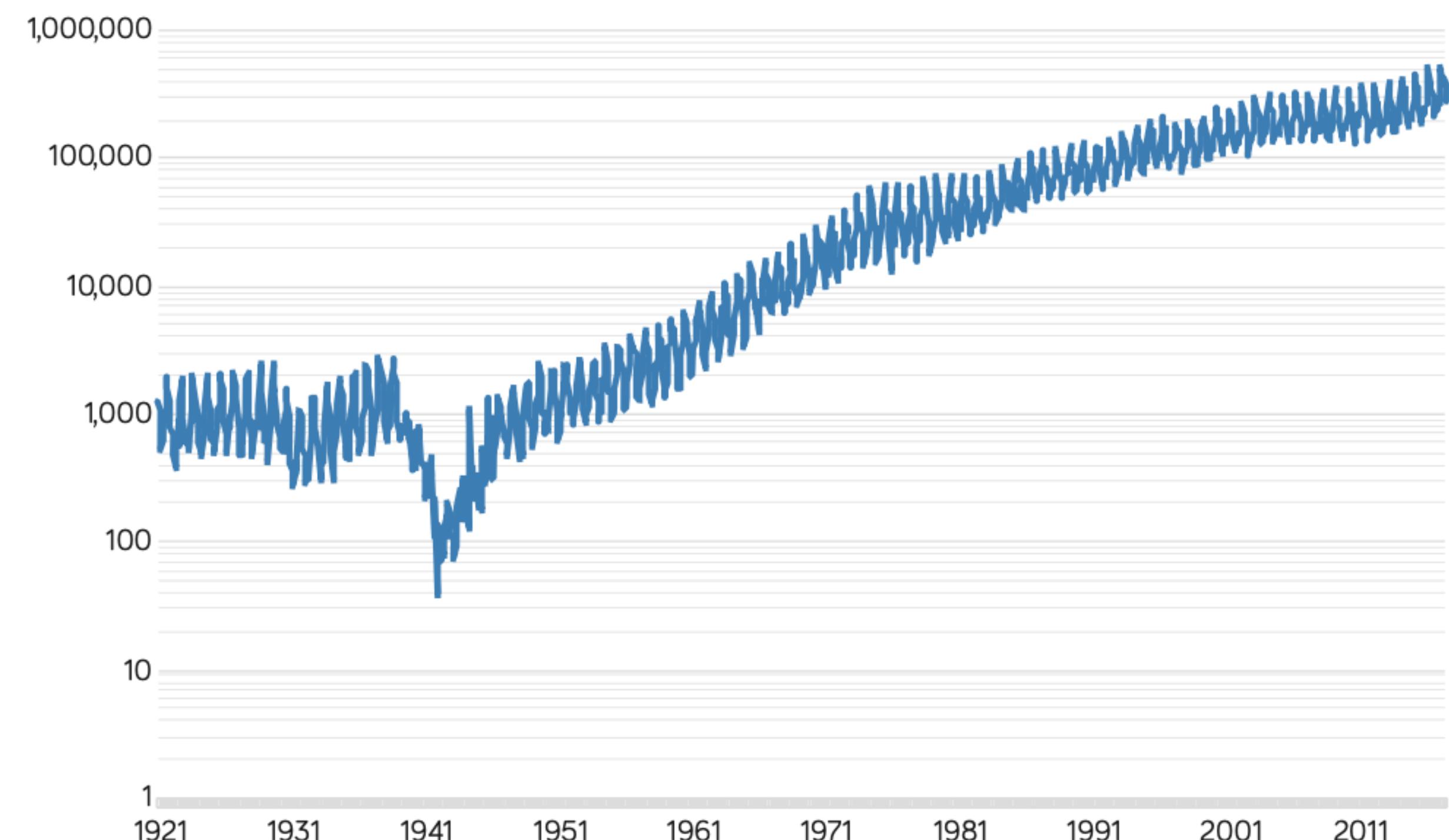
**New Zealand Tourists**

Number of overseas visitors whose intended length of stay is less than 12 months, per month, 1921-2018

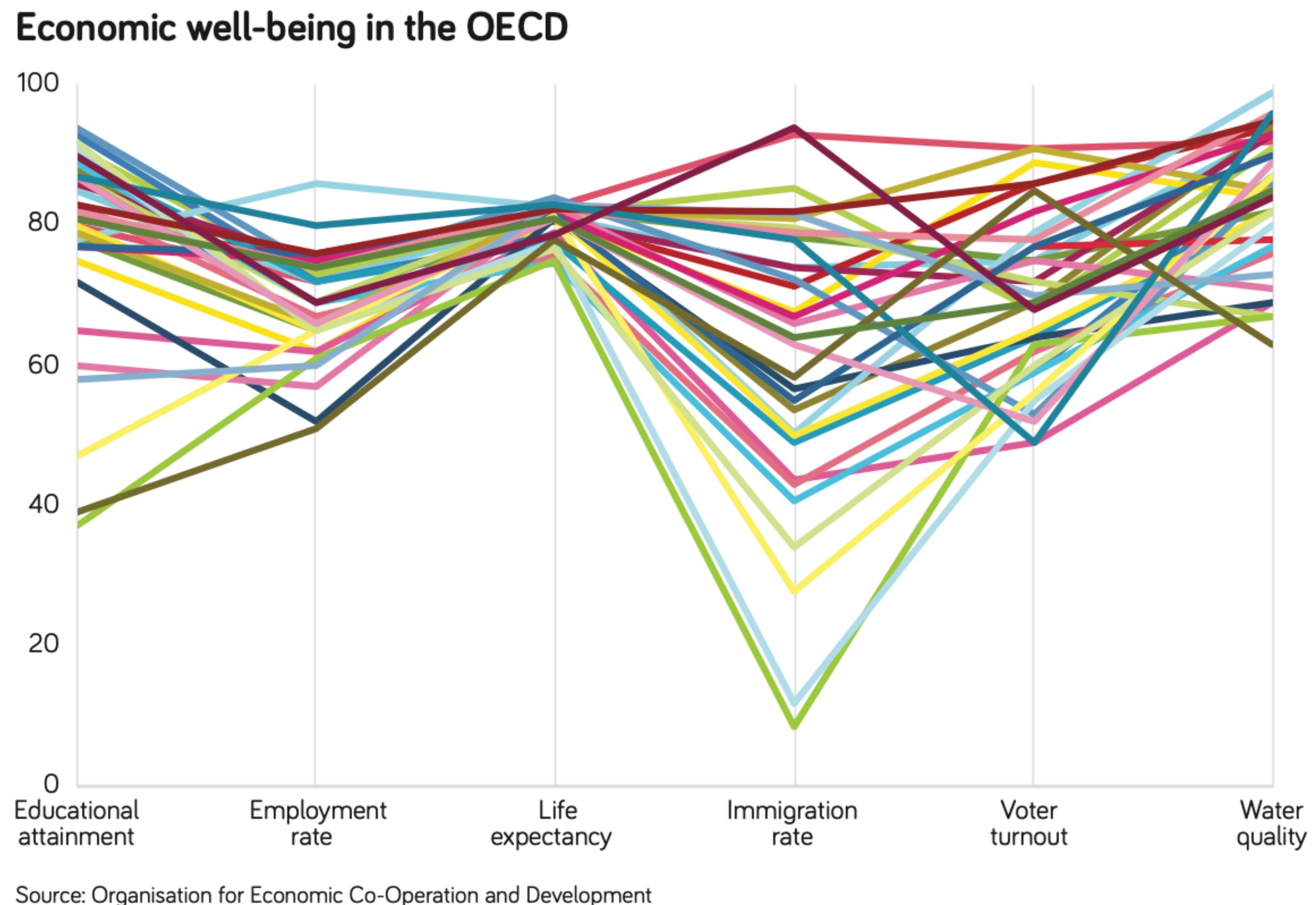


**New Zealand Tourists**

Log of the number of overseas visitors whose intended length of stay is less than 12 months, per month, 1921-2018



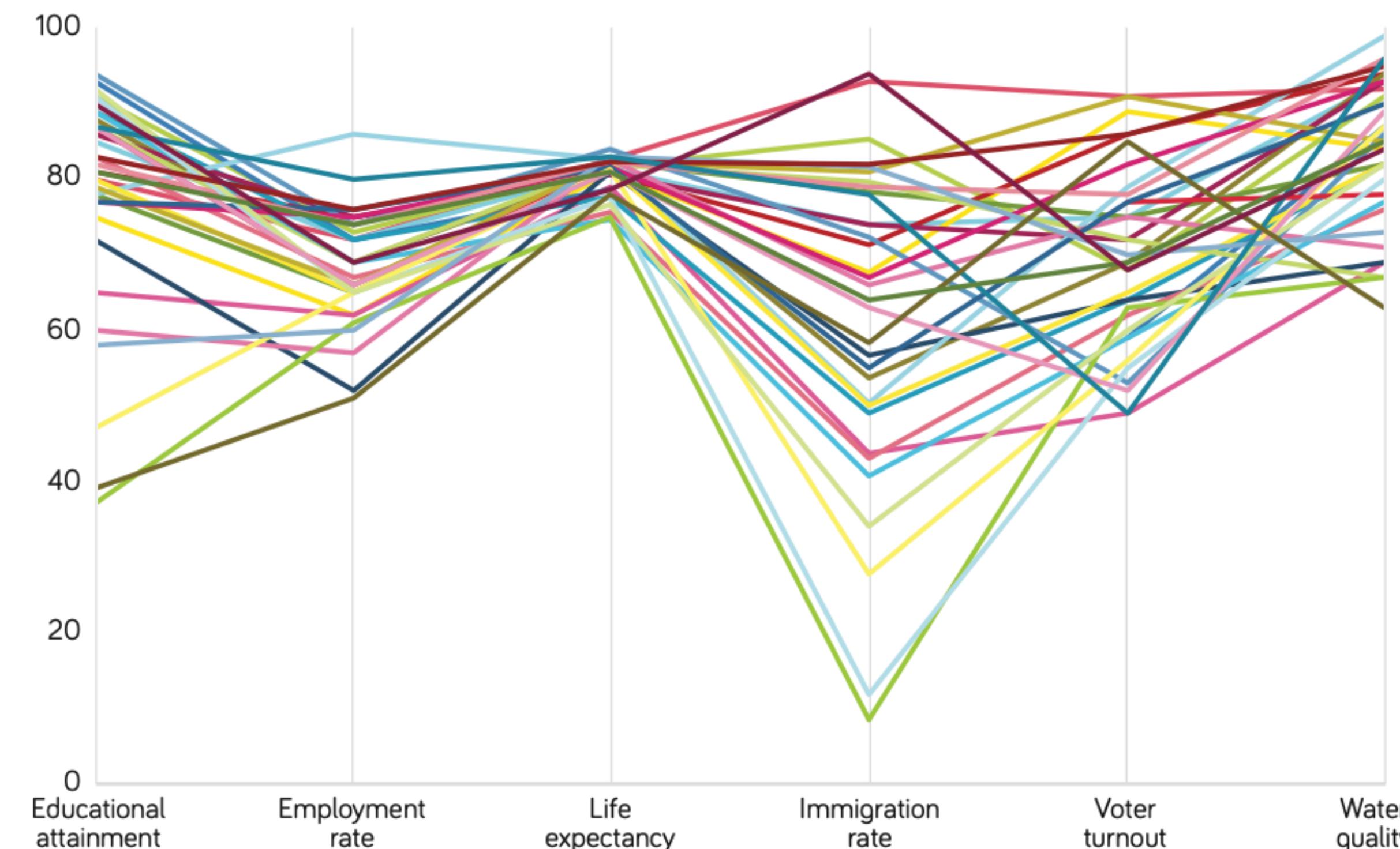
# Parallel Coordinates Plot



- data values are plotted along multiple vertical axes and connected by lines
  - the axes can have different units of measurement (or can be normalized to keep scales uniform)
  - the parallel coordinates plot permits multiple correlations within a single view

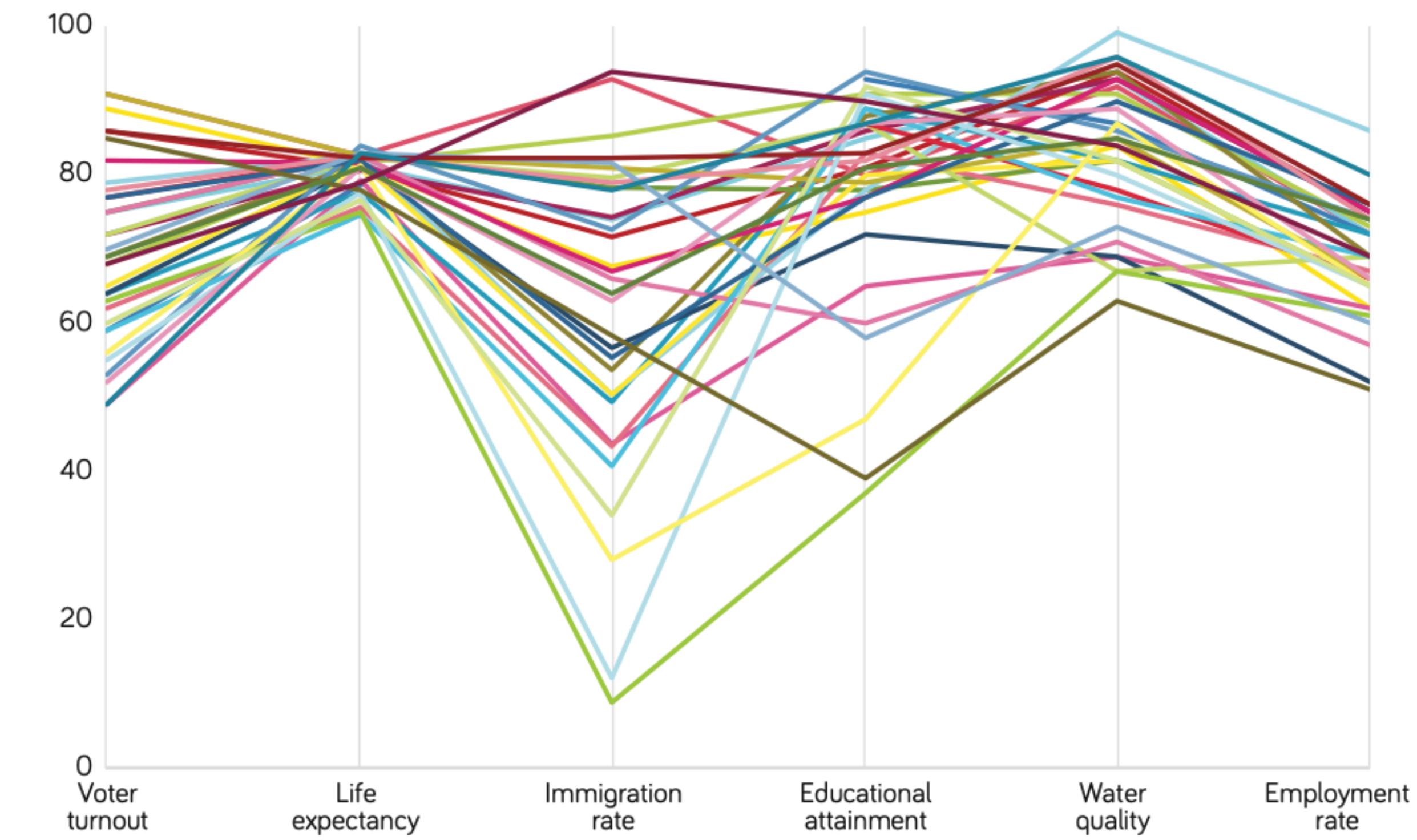
# Parallel Coordinates Plot

Economic well-being in the OECD



Source: Organisation for Economic Co-Operation and Development

Economic well-being in the OECD

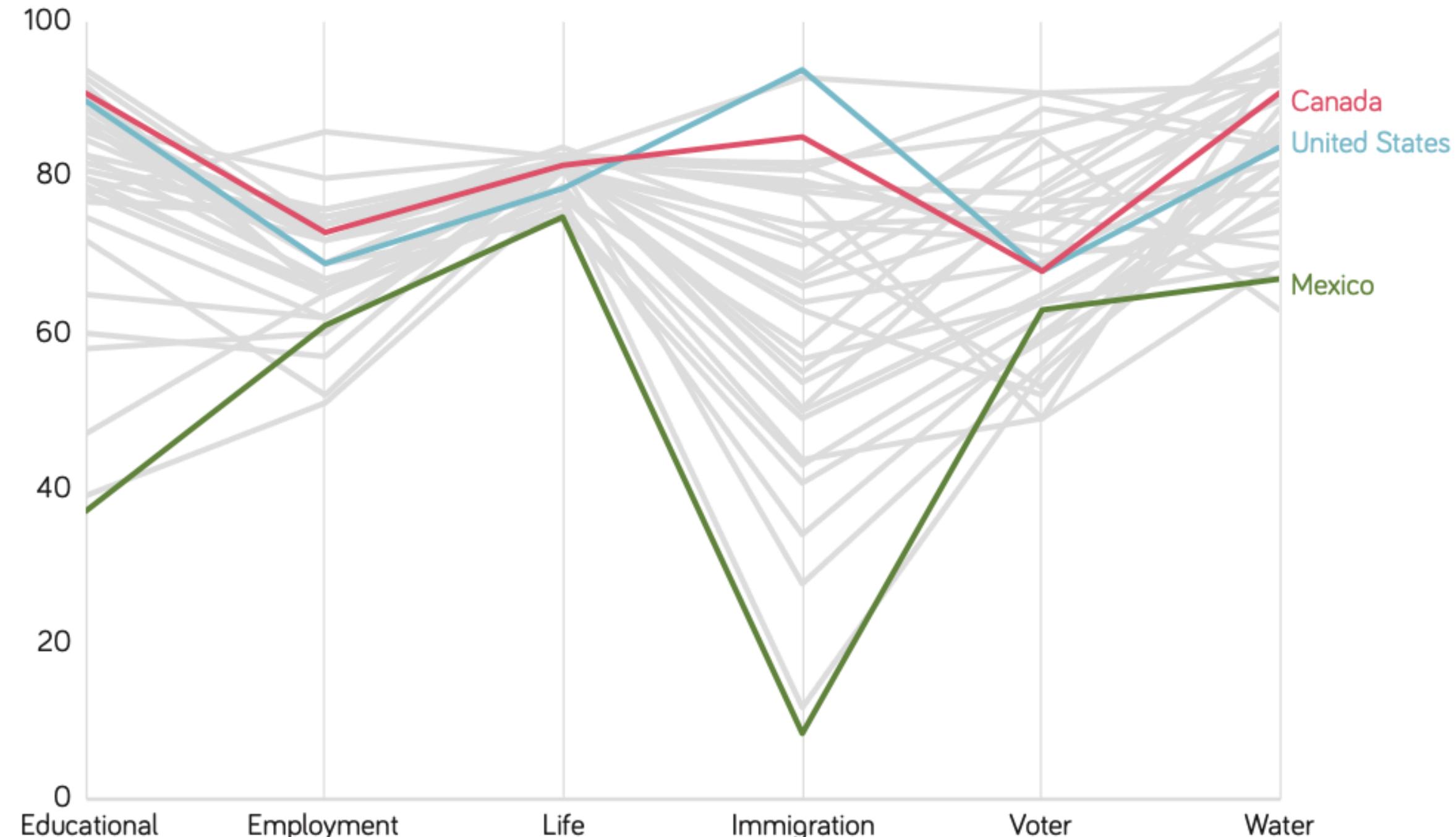


Source: Organisation for Economic Co-Operation and Development

Changing the order of the vertical axes lets us discover different relationships.

# Parallel Coordinates Plot

Economic well-being in the OECD



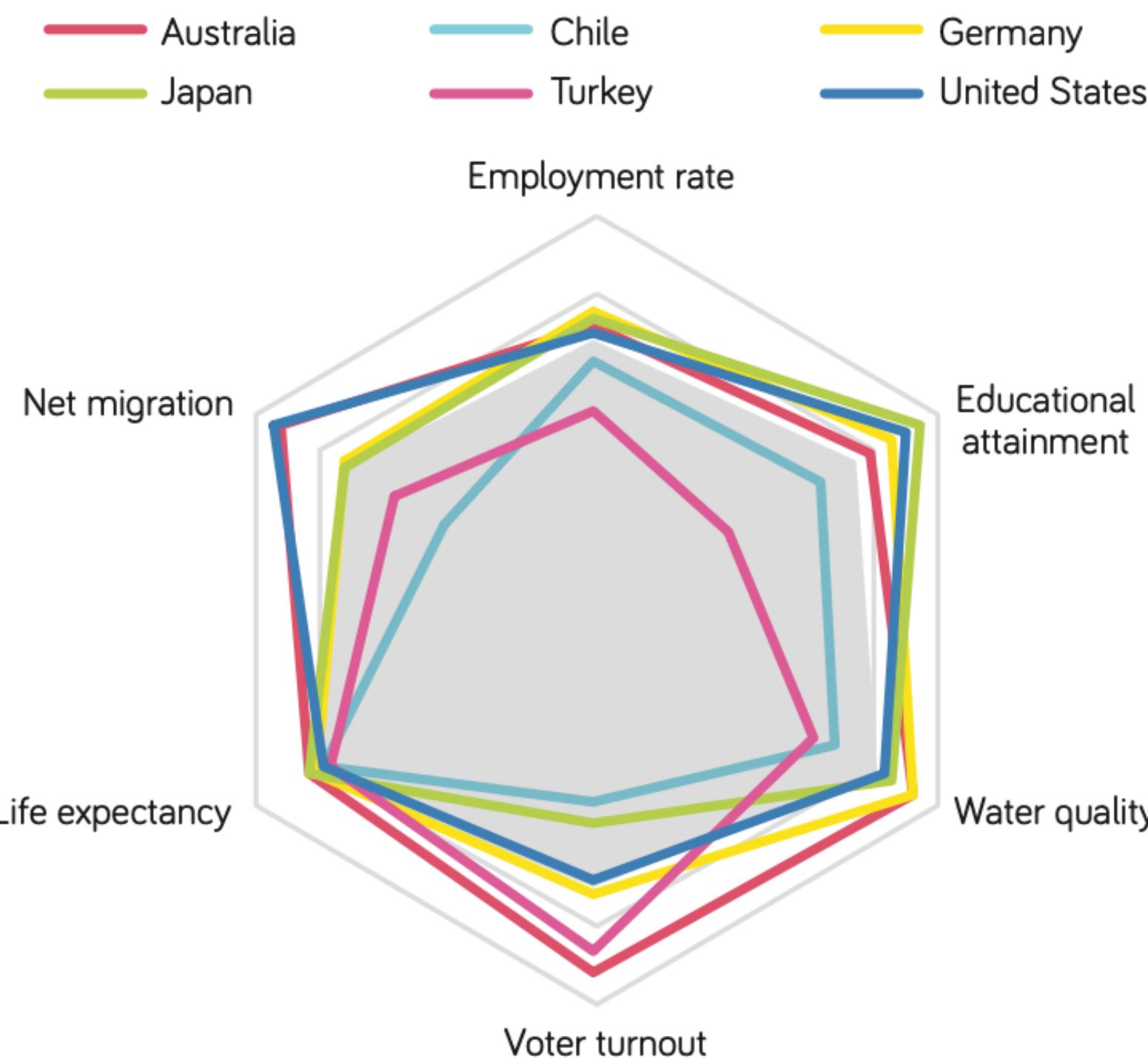
Source: Organisation for Economic Co-Operation and Development

Declutter the parallel coordinates plot with the “start with gray” strategy.

# Radar Charts

## Economic well-being in the OECD

(Gray area denotes overall average)

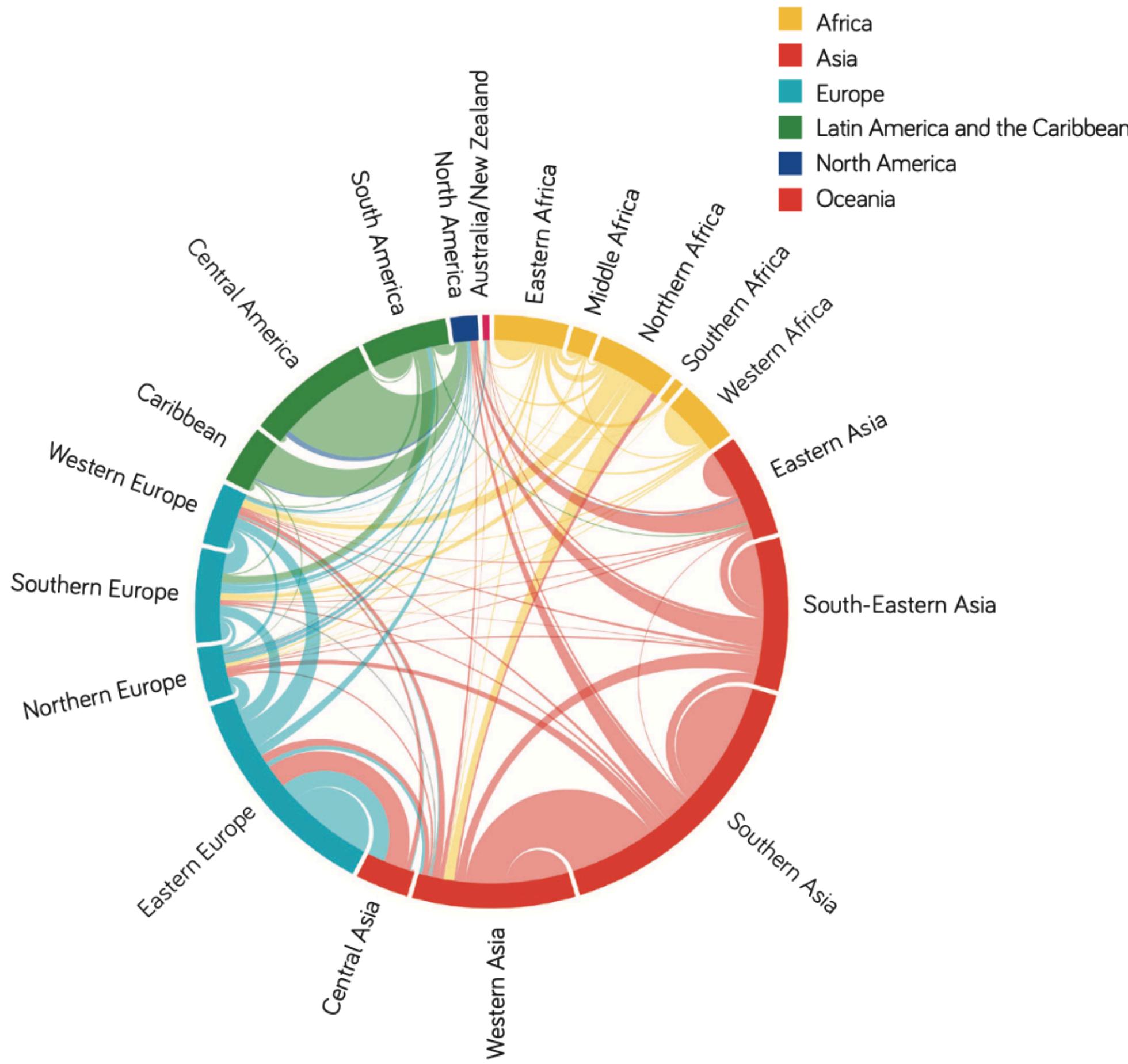


Source: Organisation for Economic Co-Operation and Development

- radar charts are like parallel coordinate plots, but the lines wrap around a circle instead of being arranged parallel to one another
- they are compact and especially good at highlighting outliers [see Turkey]

# Chord Diagram

Migration around the world



Source: Organisation for Economic Co-operation and Development  
Note: Data limited to a minimum of 200,000 immigrants or emigrants

- in chord diagrams, observations [nodes] are located around the circumference of the circle and connected by arcs within the circle to illustrate connections
- the thickness of the arcs represent the degree of the connection between the different groups

# Correlation Matrix

## World Migration

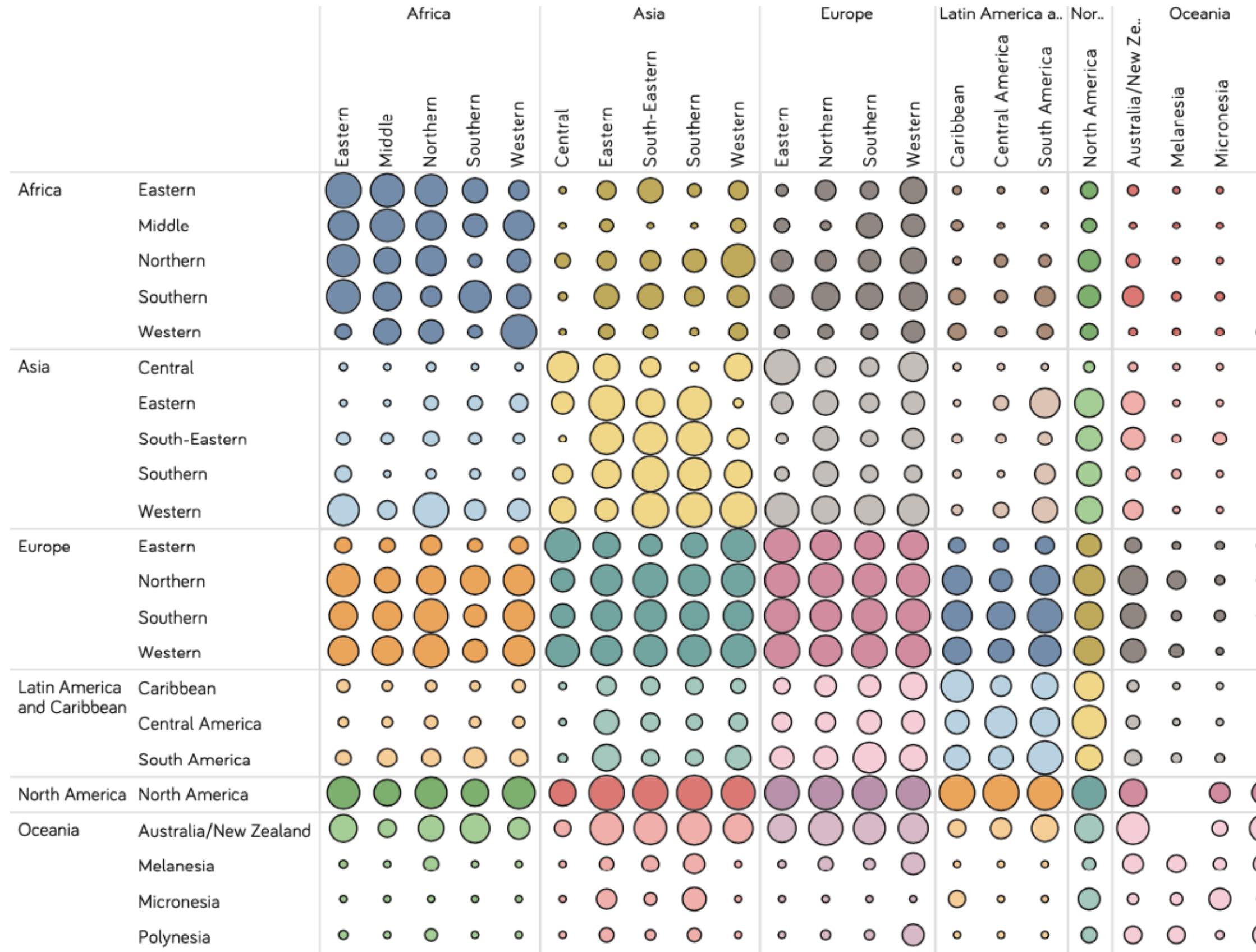
	Africa					Asia					Europe					Latin American and Caribbean			Oceania			
	Eastern	Middle	Northern	Southern	Western	Central	Eastern	South-Eastern	Southern	Western	Eastern	Northern	Southern	Western	Caribbean	Central America	South America	North America	Australia/New Zealand	Melanesia	Micronesia	Polynesia
Africa	Eastern	49.0	10.0	7.1	0.6	0.1	0.0	0.1	0.6	0.0	0.1	0.0	0.1	0.1	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Middle	3.9	17.0	3.7	0.5	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
	Northern	7.2	1.1	3.2	0.0	0.4	0.0	0.1	0.1	0.4	9.0	0.2	0.2	0.3	0.8	0.0	0.0	0.0	0.2	0.0	0.0	0.0
	Southern	14.0	2.0	0.2	7.1	0.5	0.0	0.5	0.8	0.1	0.2	0.4	1.7	1.1	1.8	0.0	0.0	0.1	0.3	0.2	0.0	0.0
	Western	0.0	1.5	0.6	0.0	58.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.1	0.0	0.0	0.1	0.0	0.0	0.0
Asia	Central	0.0	0.0	0.0	0.0	0.0	4.9	1.0	0.1	0.0	1.6	44.0	0.1	0.1	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Eastern	0.0	0.0	0.0	0.0	0.1	0.3	53.0	1.8	12.0	0.0	0.2	0.6	0.1	0.3	0.0	0.0	3.6	2.2	0.4	0.0	0.0
	South-Eastern	0.0	0.0	0.0	0.0	0.0	0.0	9.2	13.0	68.0	0.2	0.0	0.6	0.1	0.2	0.0	0.0	0.0	0.9	0.4	0.0	0.0
	Southern	0.0	0.0	0.0	0.0	0.0	0.1	2.1	110.0	8.6	1.6	0.0	0.6	0.1	0.1	0.0	0.0	0.2	0.5	0.0	0.0	0.0
	Western	6.0	0.1	38.0	0.2	0.3	0.9	0.3	170.0	40.0	130.0	12.0	1.7	2.9	5.4	0.0	0.0	0.6	1.4	0.1	0.0	0.0
Europe	Eastern	0.0	0.0	0.2	0.0	0.1	56.0	1.5	0.4	1.0	21.0	100.0	4.7	2.5	3.9	0.0	0.0	0.1	0.5	0.0	0.0	0.0
	Northern	8.5	0.7	2.0	2.4	4.5	0.4	5.0	23.0	5.9	8.8	26.0	20.0	8.9	8.8	2.4	0.3	2.7	4.2	2.2	0.1	0.0
	Southern	2.0	2.4	15.0	0.3	5.3	0.5	4.1	6.2	2.3	3.2	32.0	5.8	31.0	16.0	3.8	1.7	25.0	1.9	0.6	0.0	0.0
	Western	4.2	4.0	34.0	0.5	6.5	12.0	4.8	8.7	8.5	31.0	60.0	7.7	49.0	29.0	2.0	0.7	7.3	3.6	0.6	0.0	0.0
Latin American and Caribbean	Caribbean	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.2	0.3	1.4	7.1	0.2	1.0	2.6	0.0	0.0	0.0	0.0
	Central America	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.1	0.0	0.1	0.1	0.5	0.4	0.5	6.5	2.1	9.8	0.0	0.0	0.0	0.0
	South America	0.0	0.1	0.1	0.2	0.1	0.0	1.8	0.1	0.1	0.6	0.5	0.5	8.0	1.6	0.8	0.5	42.0	1.3	0.0	0.0	0.0
North America	North America	8.2	1.3	6.7	1.5	8.1	1.2	53.0	47.0	53.0	17.0	23.0	18.0	19.0	15.0	66.0	160.0	34.0	12.0	1.6	0.7	0.2
	Oceania	1.6	0.1	0.8	2.5	0.2	0.0	8.8	8.3	10.0	3.1	1.9	18.0	6.8	3.7	0.1	0.2	1.3	2.1	7.4	1.4	0.0
	Australia/New Zealand	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.2	0.1	0.0	0.1
	Melanesia	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
	Micronesia	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
	Polynesia	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.2

Source: Organisation for Economic Co-Operation and Development  
Note: Data limited to a minimum of 200,000 immigrants or emigrants

- a correlation matrix is a table with the variables listed along the horizontal and vertical axes
- numbers in each cell represent the strength of that relationship (Pearson's correlation coefficient)

# Correlation Matrix

## World Migration

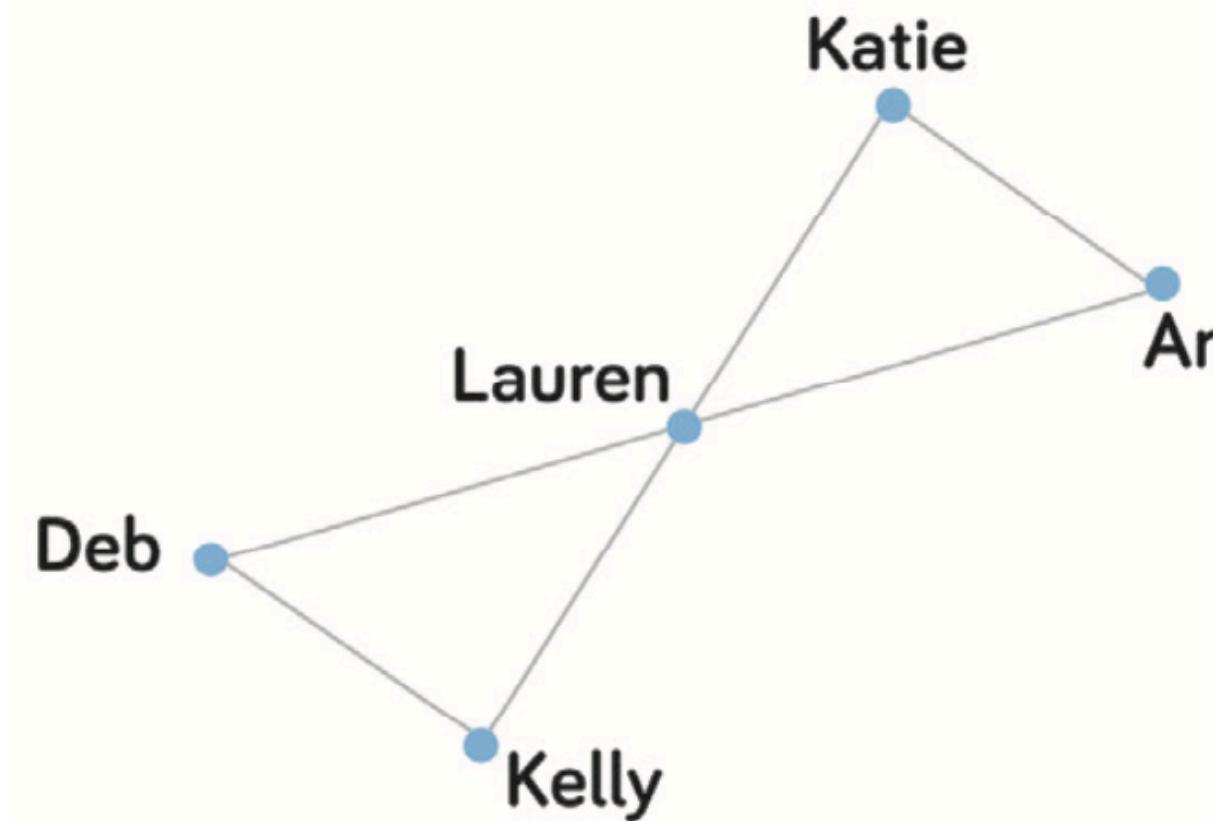


Source: Organisation for Economic Co-Operation and Development

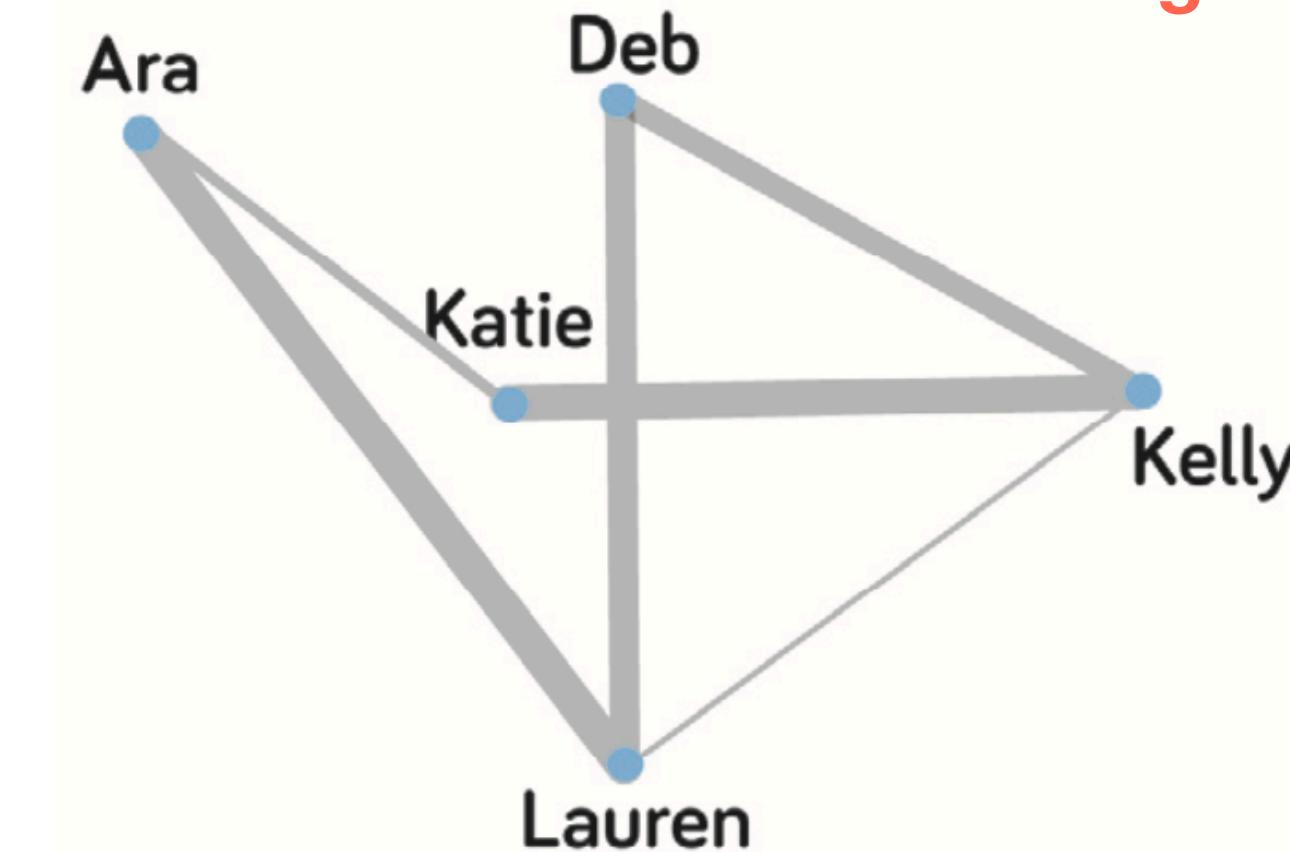
- use circles or other shapes and
- use color to organize the space

# Network Diagrams

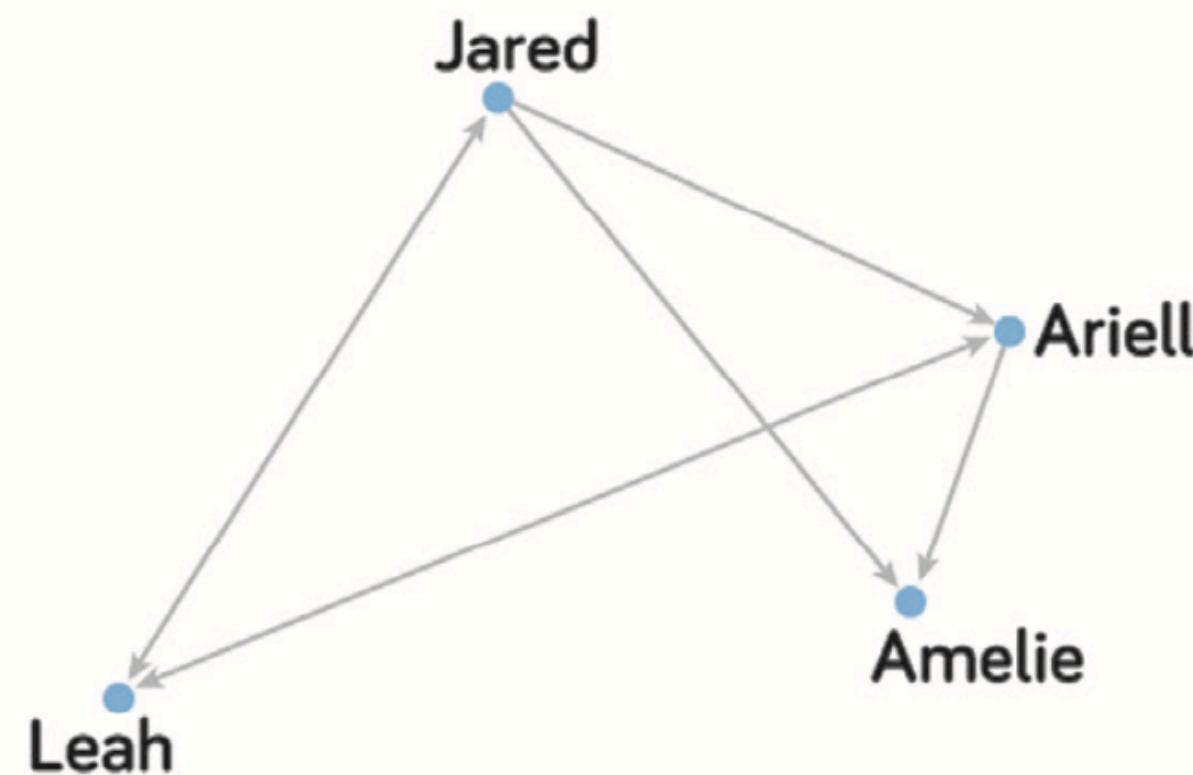
undirected and unweighted



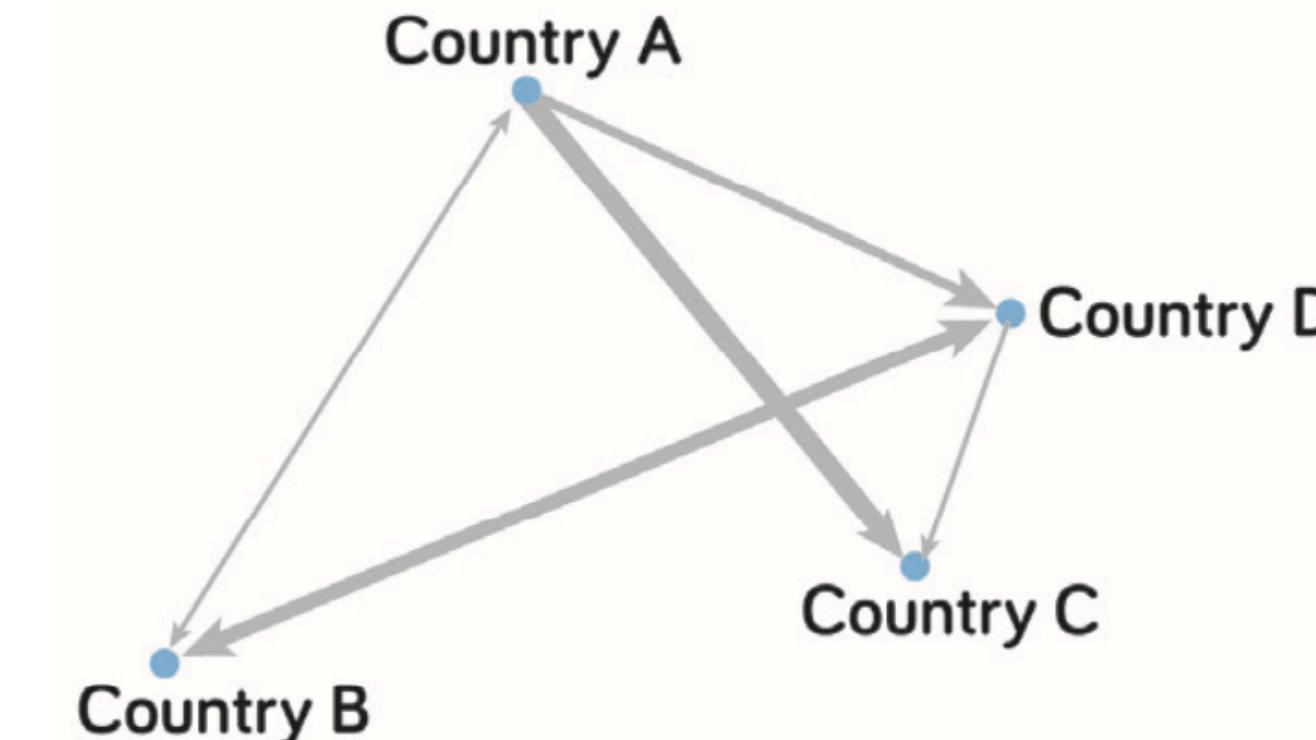
undirected and weighted



directed and unweighted

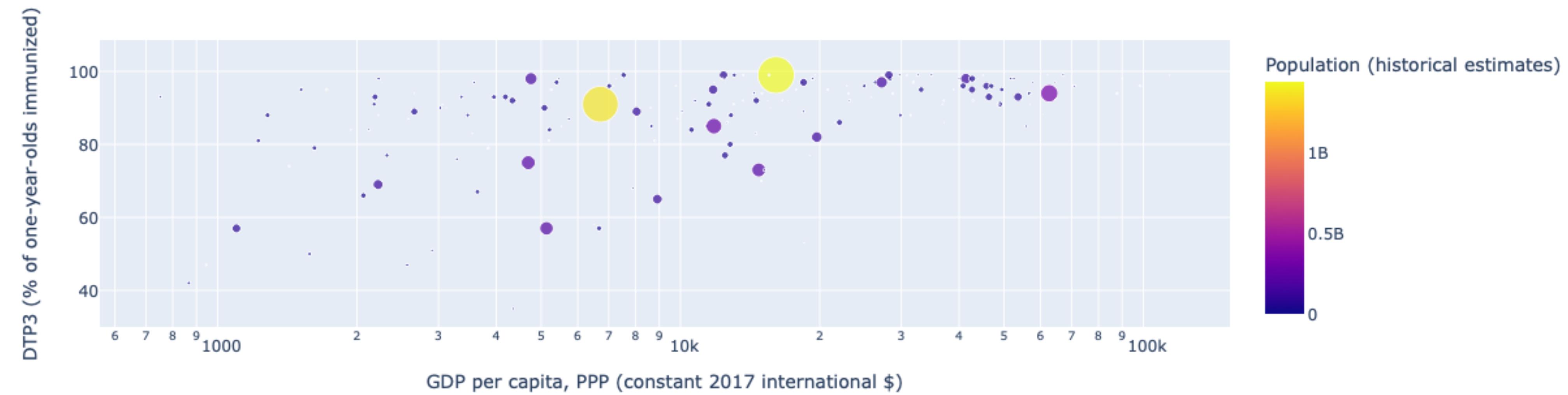


directed and weighted



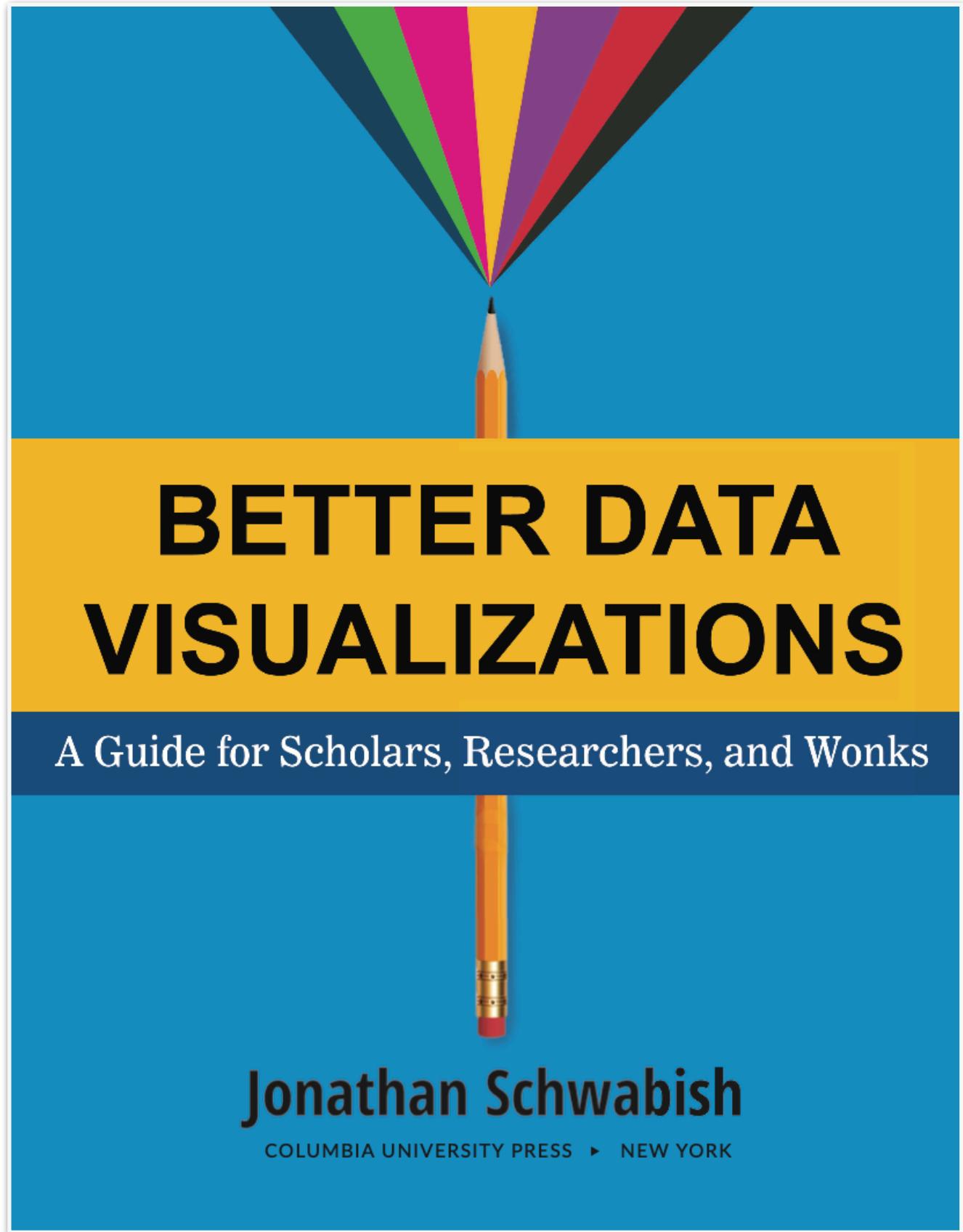
# Visualizing Relationships

- use the data set [dtp3-vs-gdp-per-capita.csv](#) to generate a bubble plot similar to this one using `matplotlib`, `seaborn` or `plotly`



- What can you improve about this chart?

# Literature



# References

- Slide 73-80, 83-98; Image Source: J. Schwabisch - Better Data Visualizations, Columbia University Press