

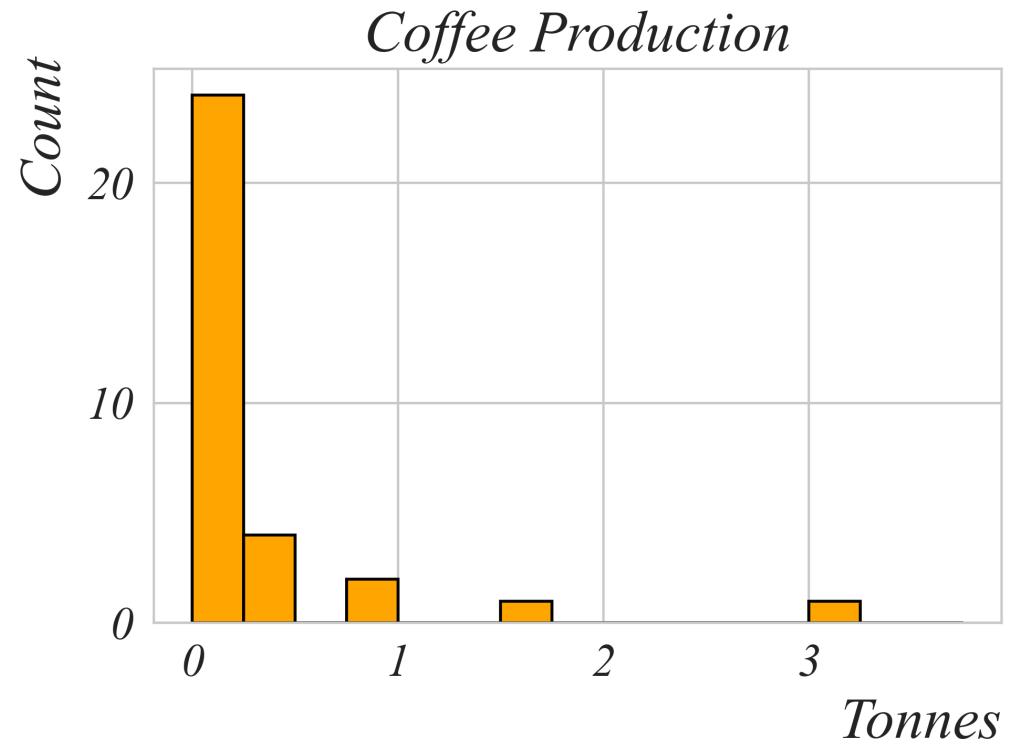
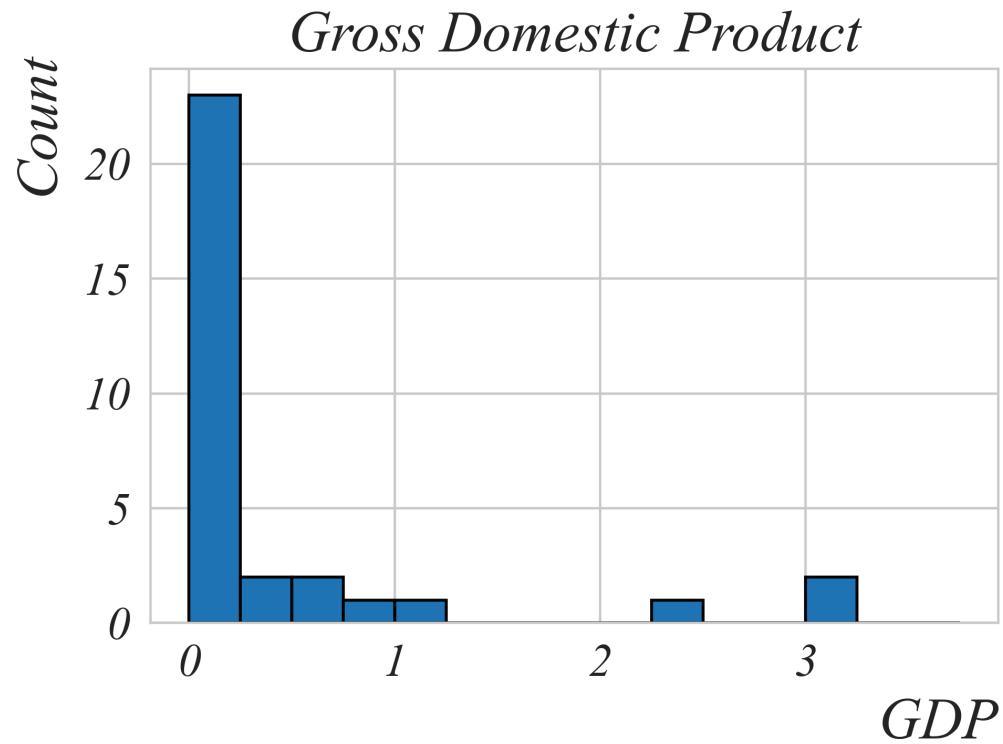
ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 2.1 | Relationships Between Variables

Relationships Between Variables

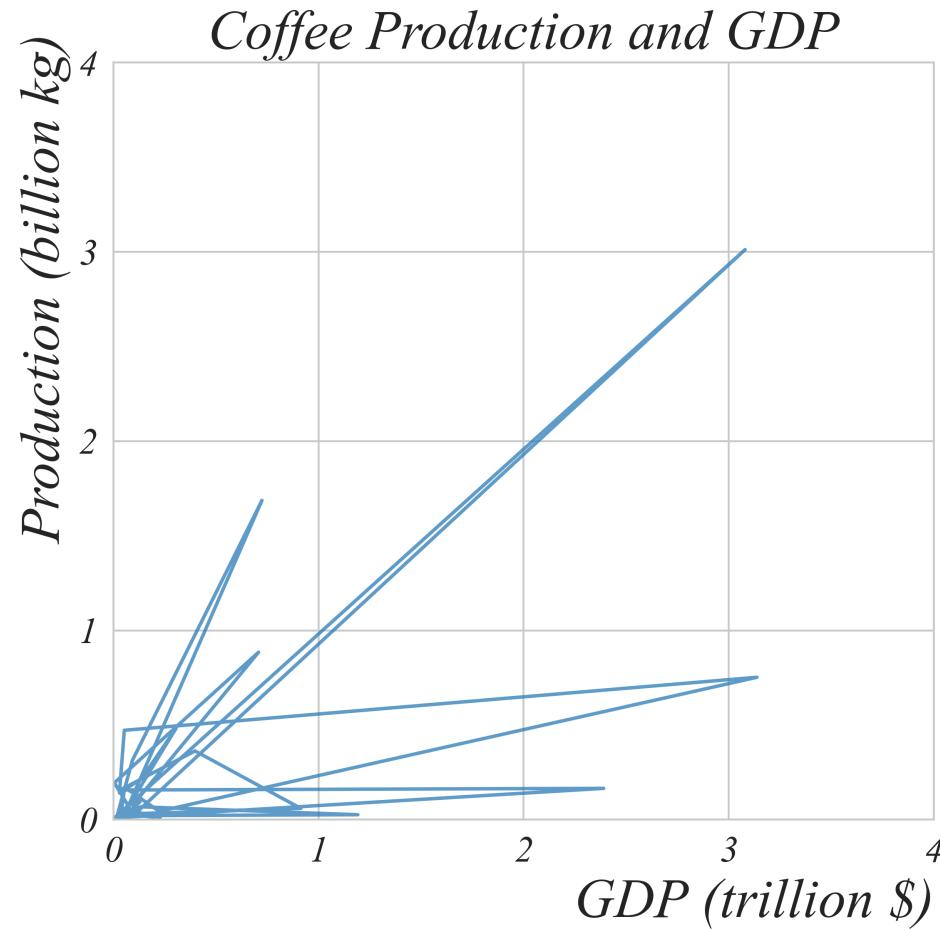
Q. Is there a relationship between GDP and coffee production?



- > maybe, but it's hard to see
- > lets use a two dimensional graph

Relationships Between Variables

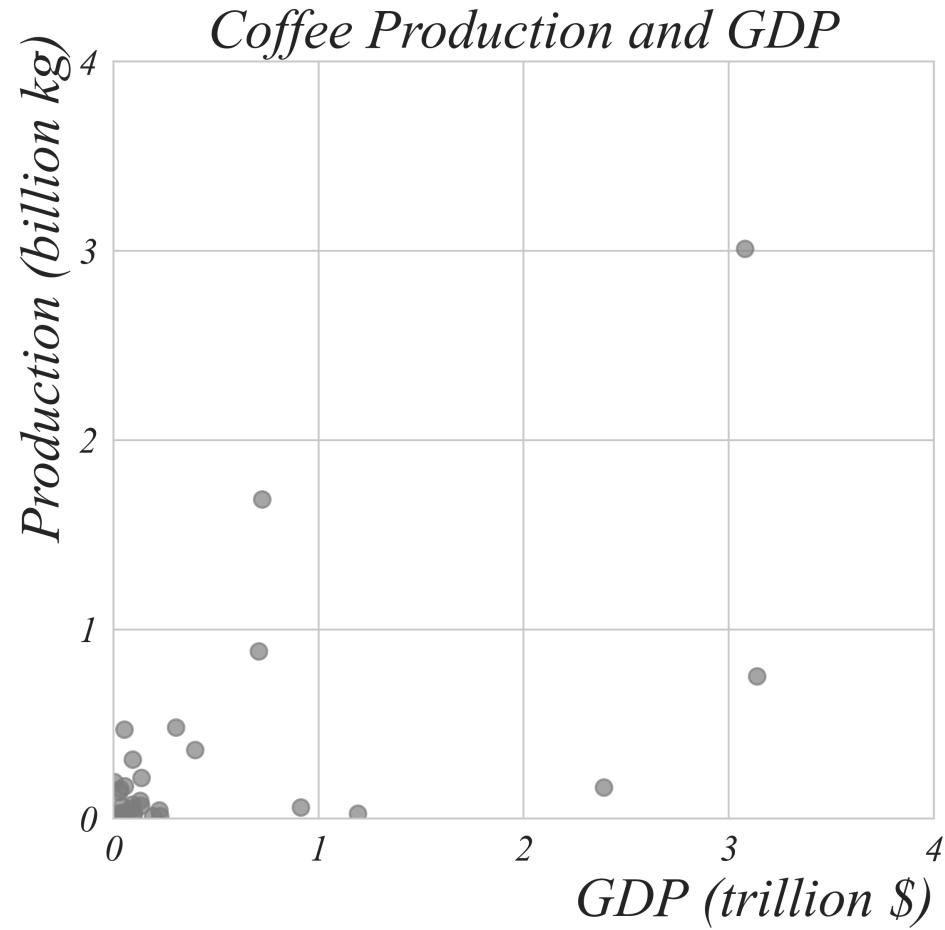
Q. Is there a relationship between GDP and coffee production?



> two dimensions is nice, but the points have no meaningful relationships

Relationships Between Variables

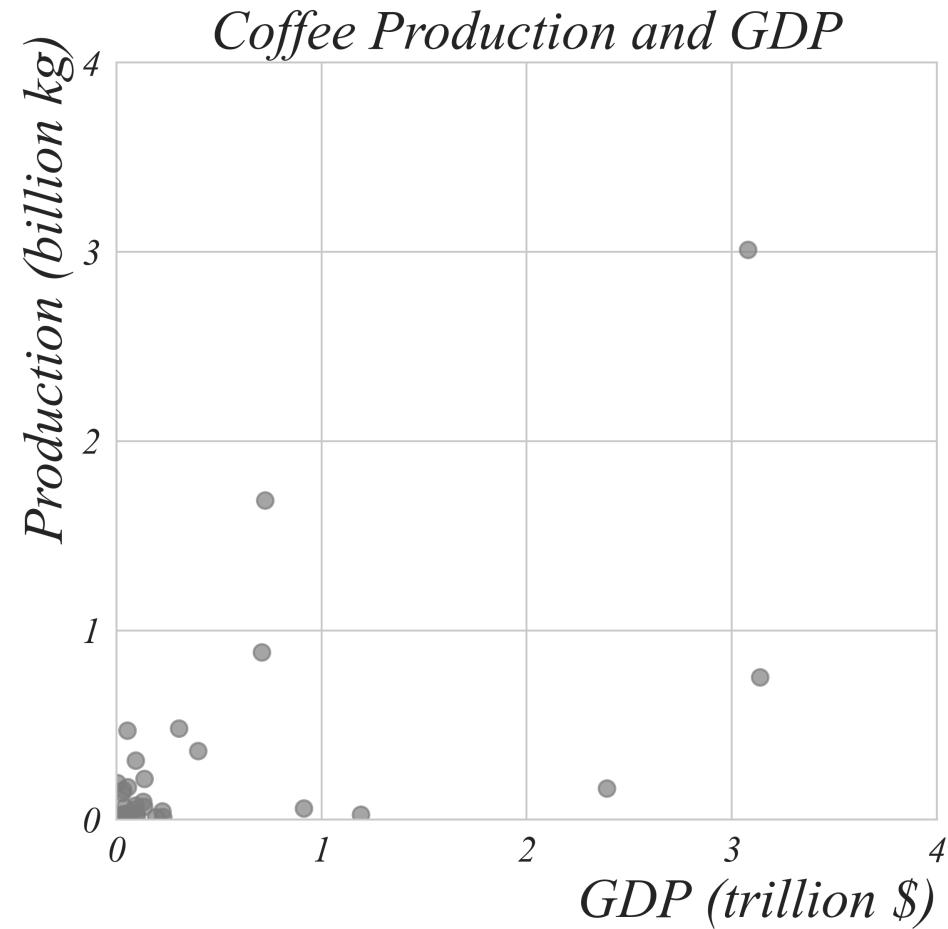
Q. Is there a relationship between GDP and coffee production?



> a scatterplot effectively visualizes cross sectional data with two dimensions

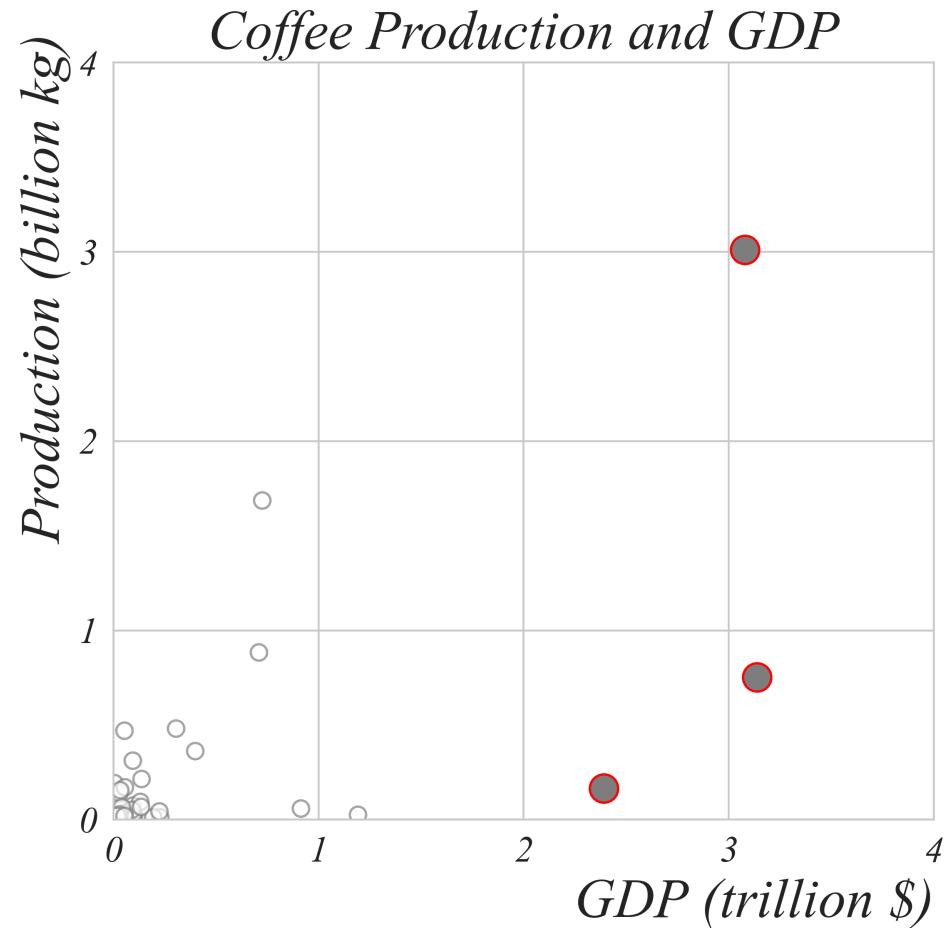
Relationships Between Variables

Which countries have a GDP above \$2 trillion?



Relationships Between Variables

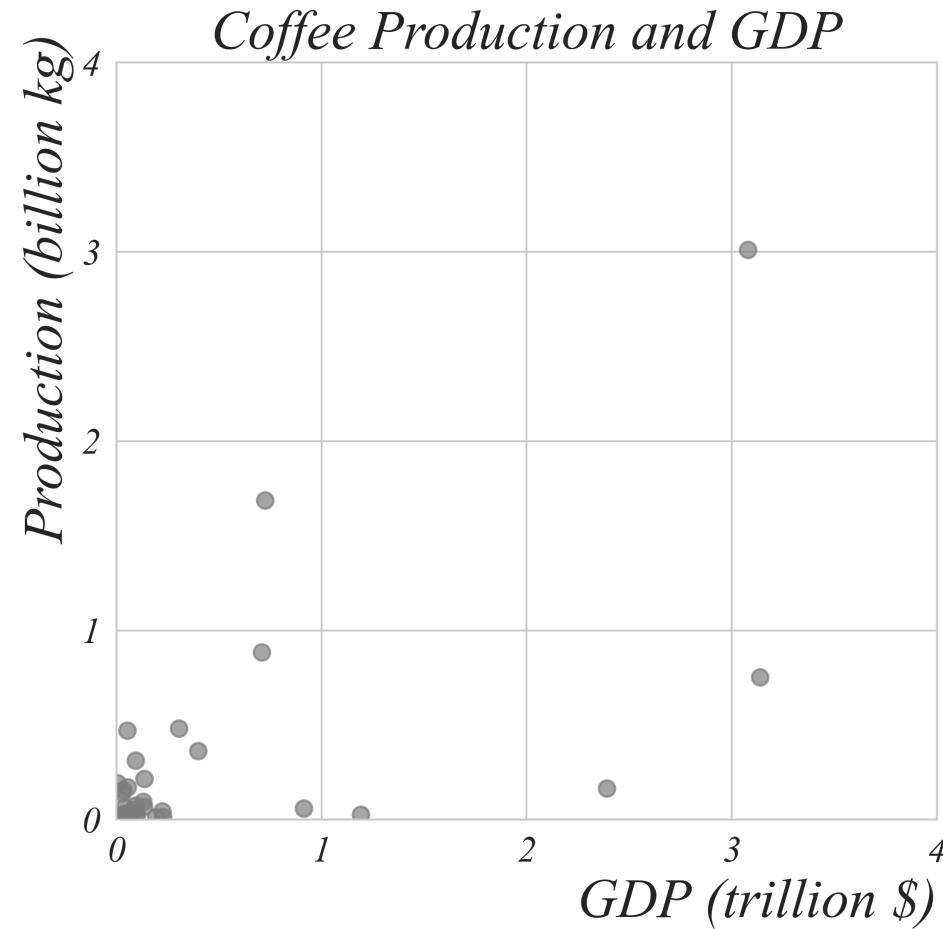
Which countries have a GDP above \$2 trillion?



> we can also easily use the scale on the axis

Relationships Between Variables

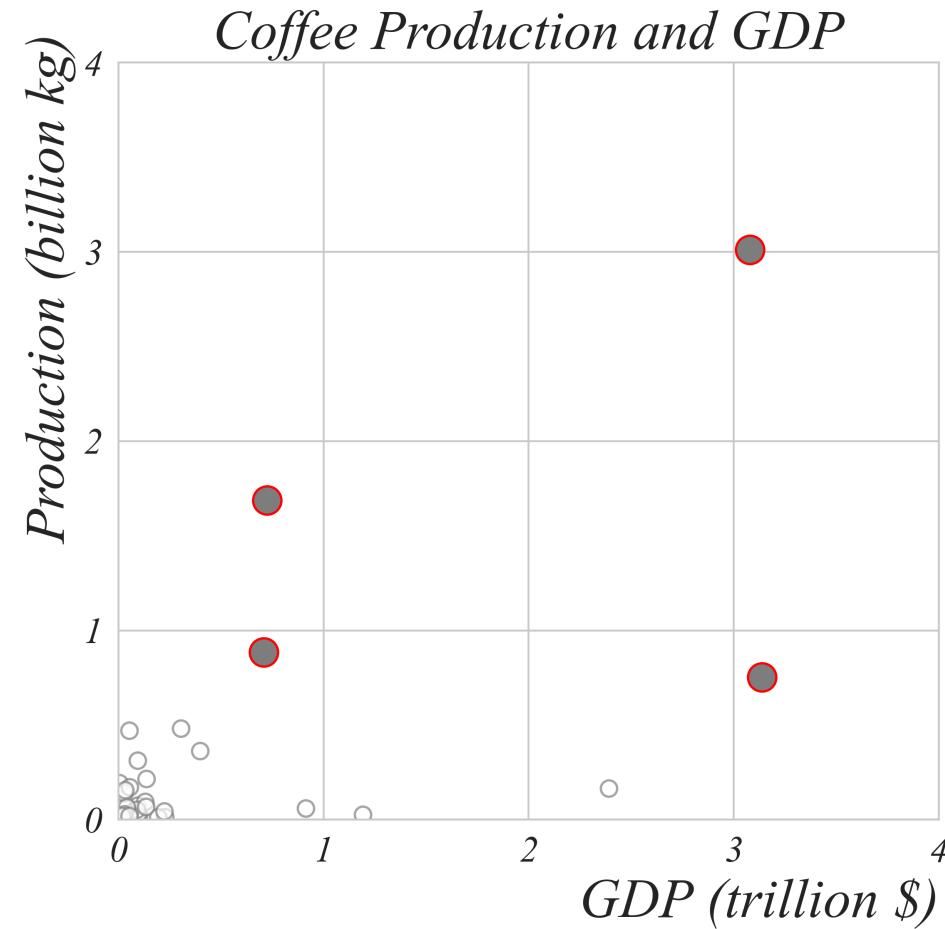
Which countries have a production above $\frac{1}{2}$ billion kg?



> and we can use either axis

Relationships Between Variables

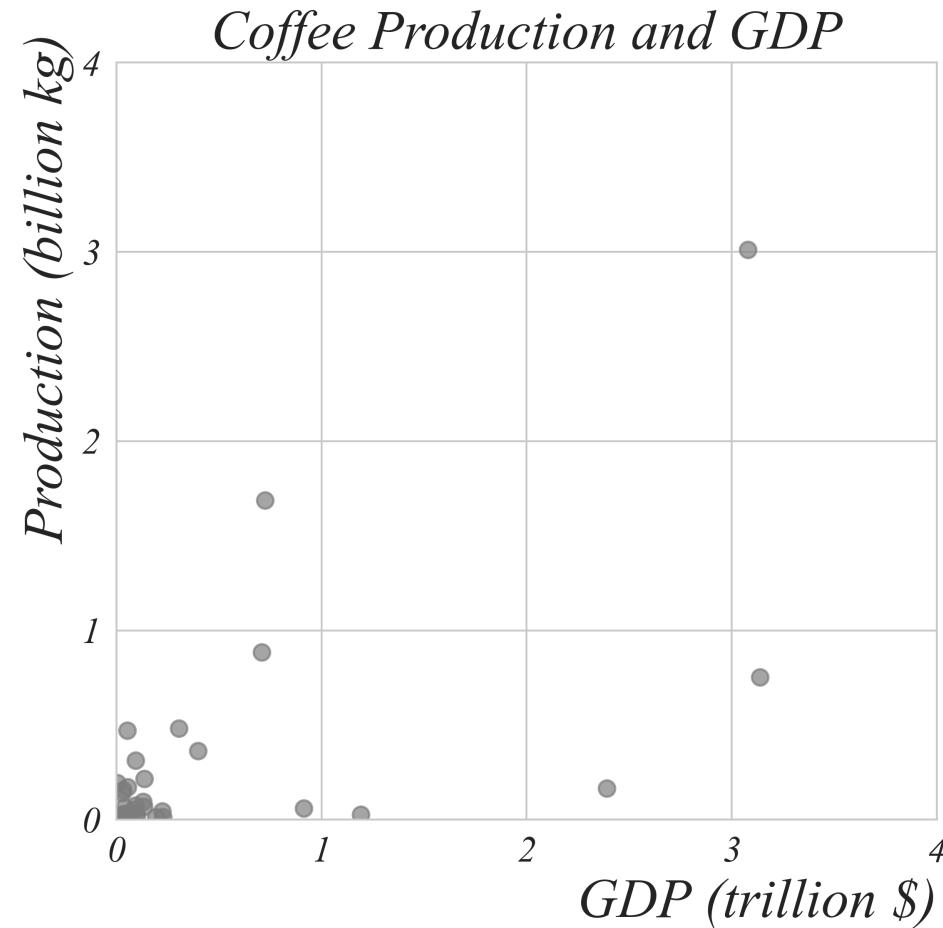
Which countries have a production above $\frac{1}{2}$ billion kg?



> and we can use either axis

Relationships Between Variables

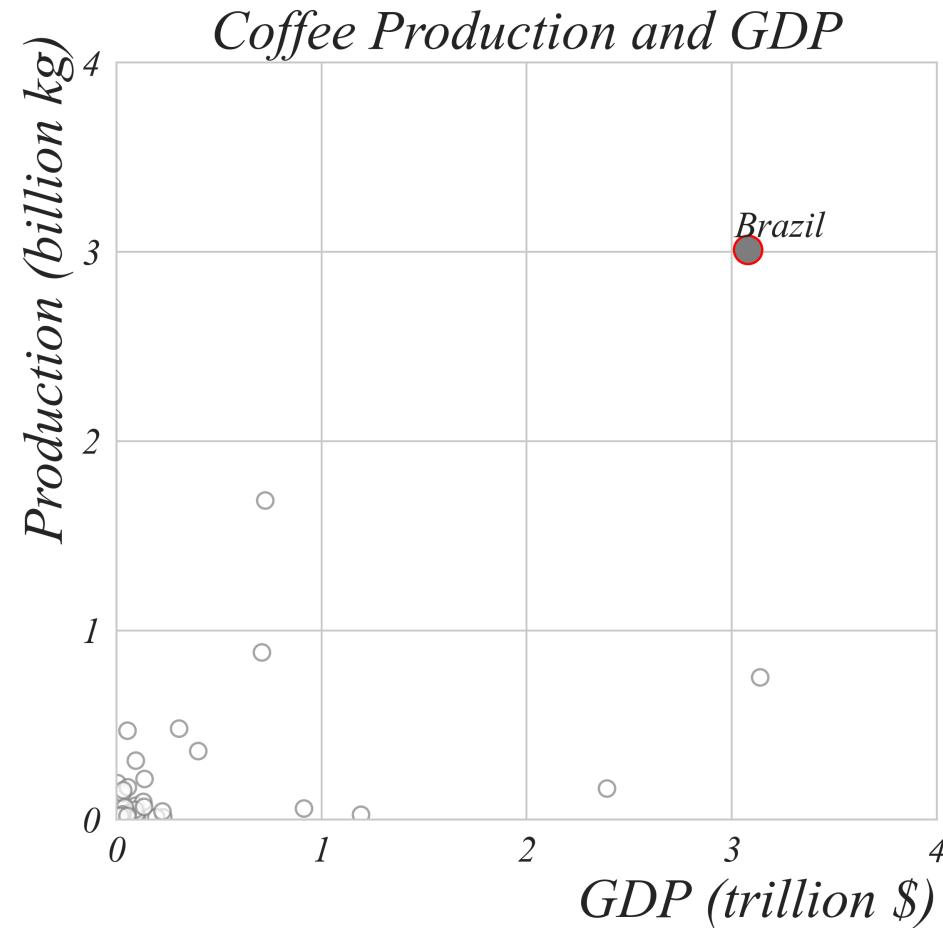
Which countries produce less coffee per dollar than Brazil?



> we can also compare *BETWEEN* data points

Relationships Between Variables

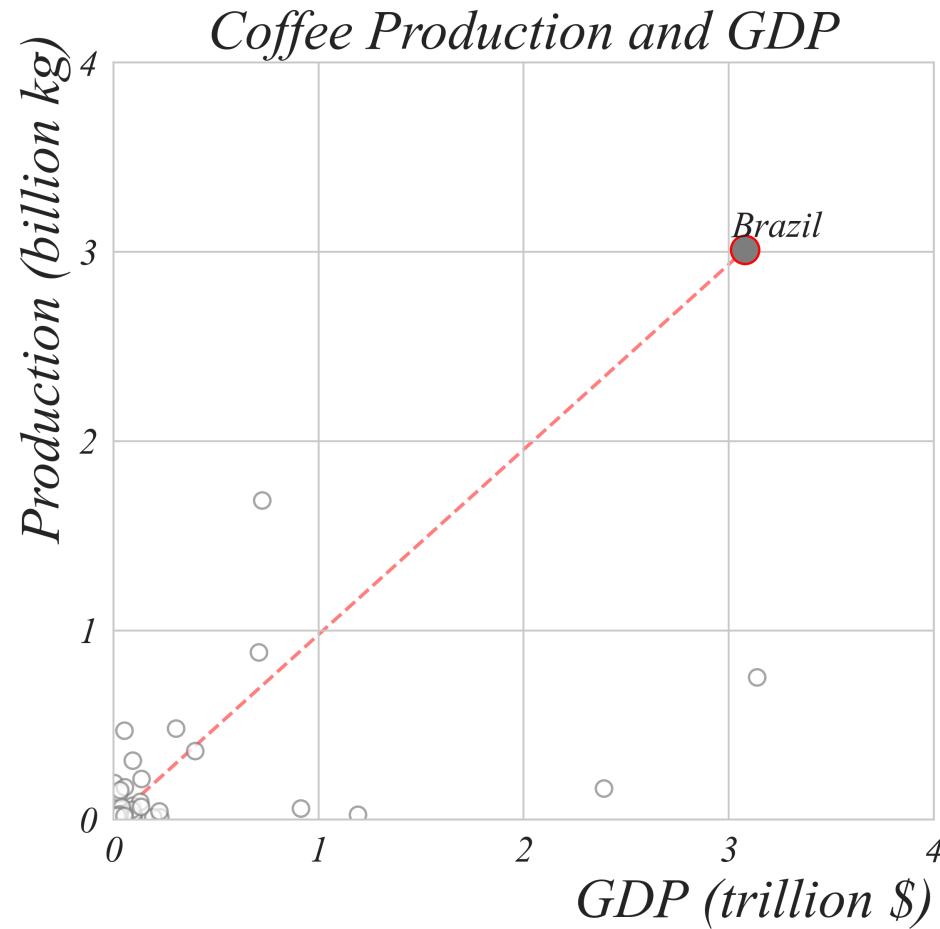
Which countries produce less coffee per dollar than Brazil?



> we can also compare *BETWEEN* data points

Relationships Between Variables

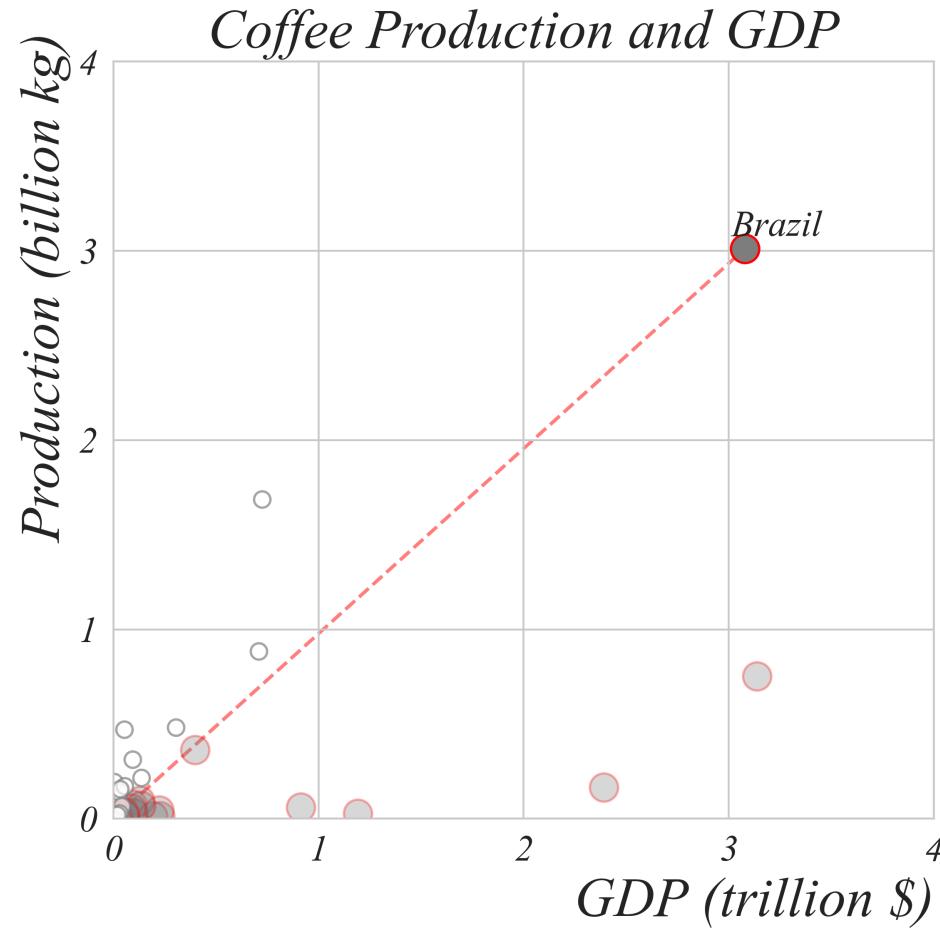
Which countries produce less coffee per dollar than Brazil?



> separating lines can help make comparisons between ratios

Relationships Between Variables

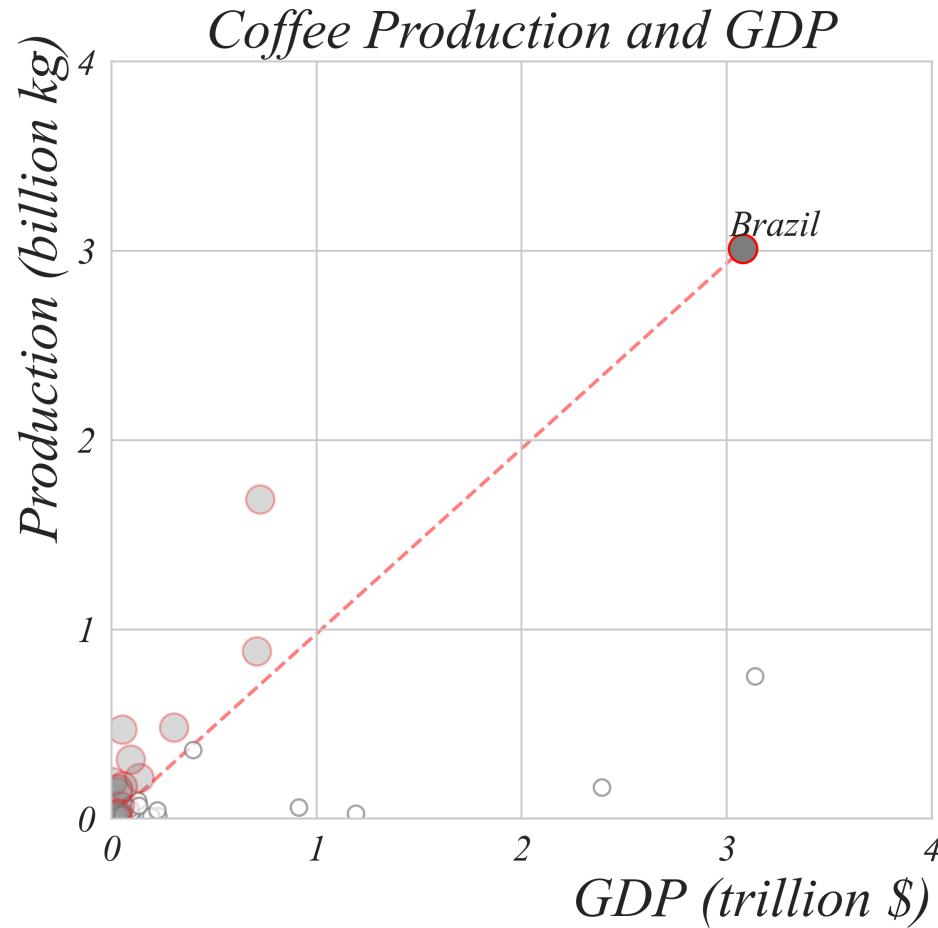
Which countries produce less coffee per dollar than Brazil?



> separating lines can help make comparisons between ratios

Relationships Between Variables

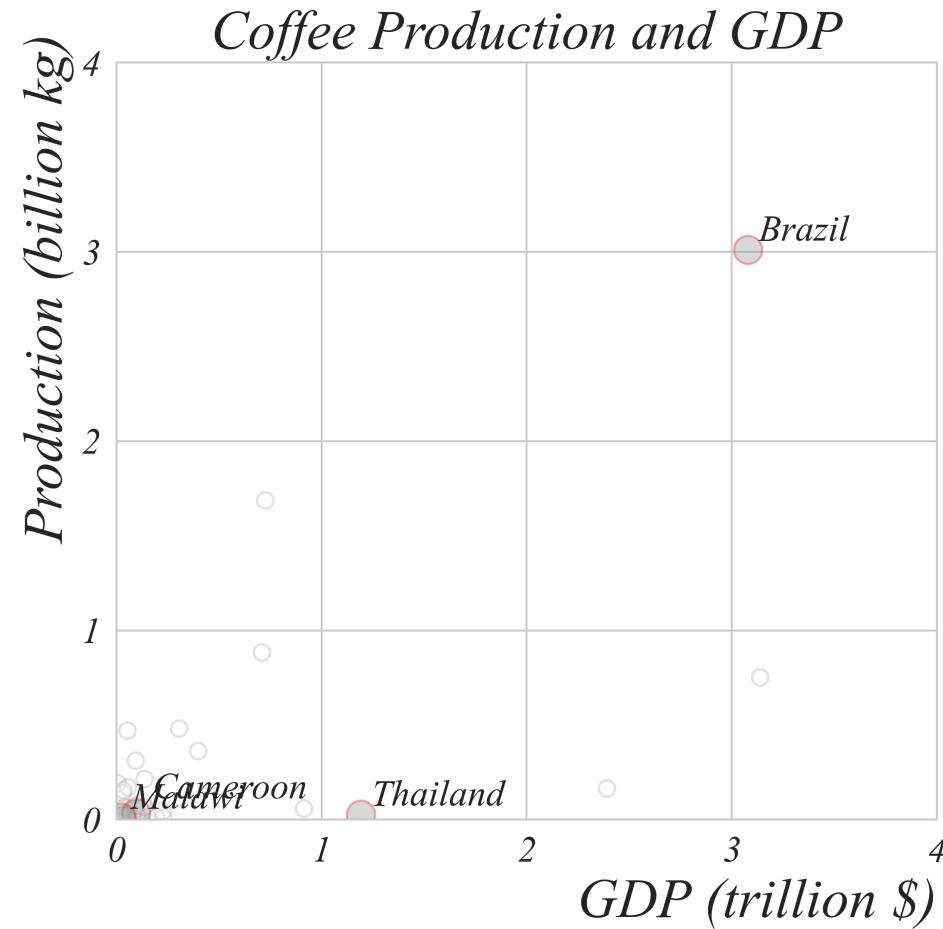
Which countries produce more coffee per dollar than Brazil?



> separating lines can help make comparisons between ratios

Relationships Between Variables

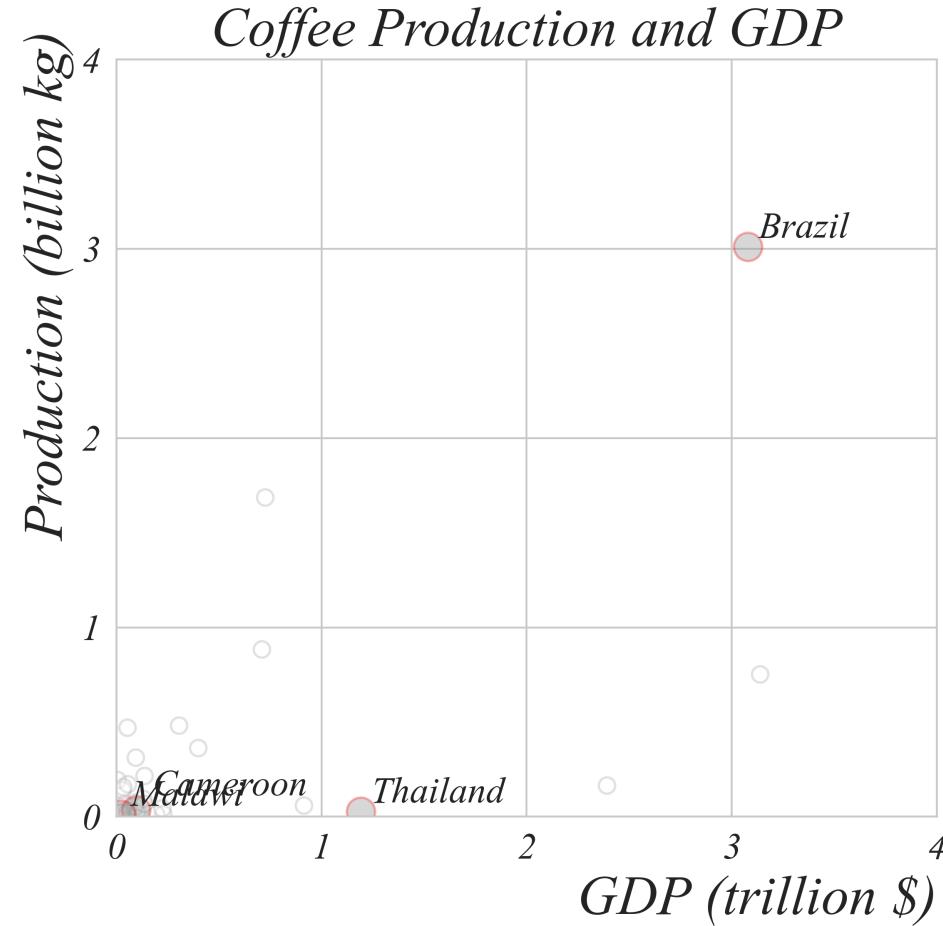
Do the GDPs of the upper or lower pair differ by a larger amount?



> we can also use the scale on the axis to measure differences

Relationships Between Variables

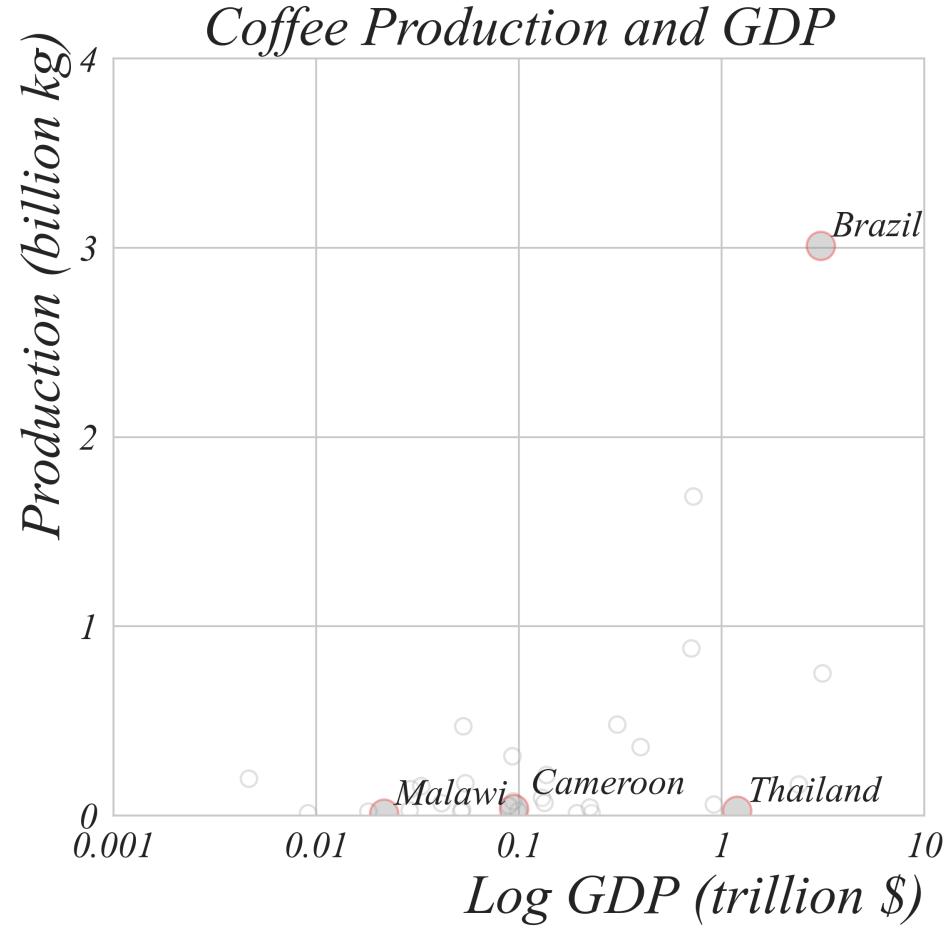
Which is larger: the ratio of GDPs of the upper or lower pair?



> but the scale matters - this question is difficult to answer

Relationships Between Variables

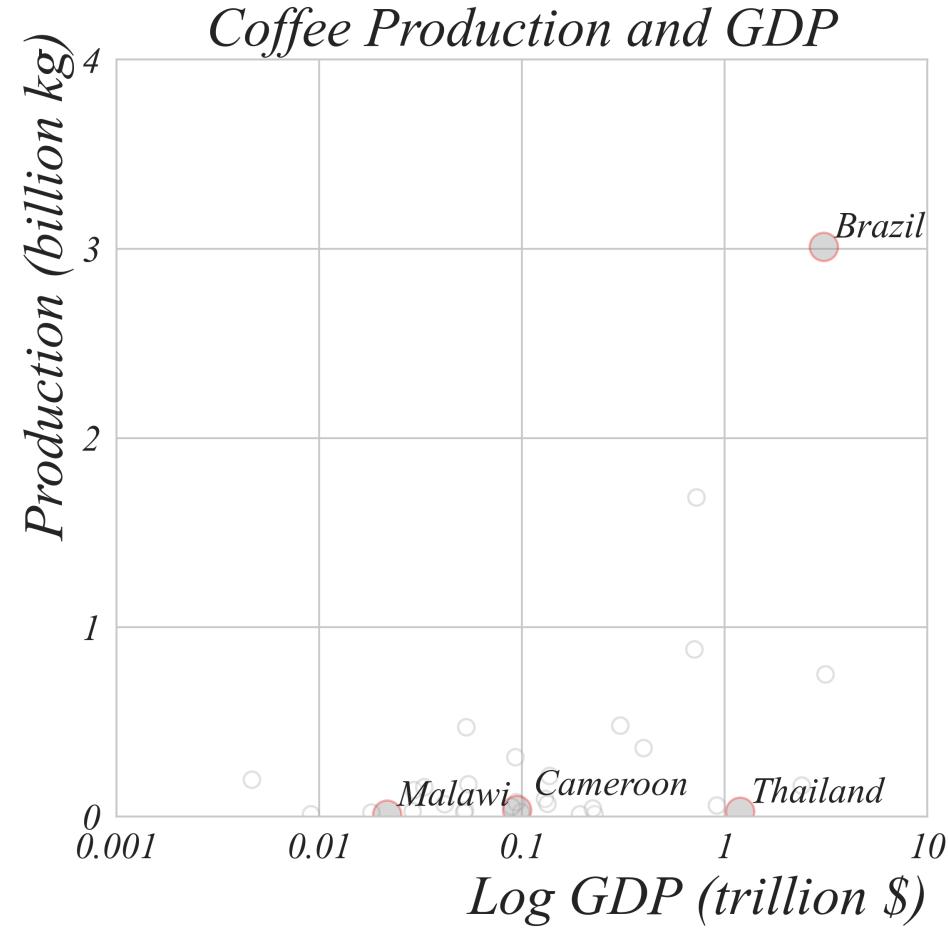
Which is larger: the ratio of GDPs of the upper or lower pair?



> a log scale makes RATIOS easier to visualize

Relationships Between Variables

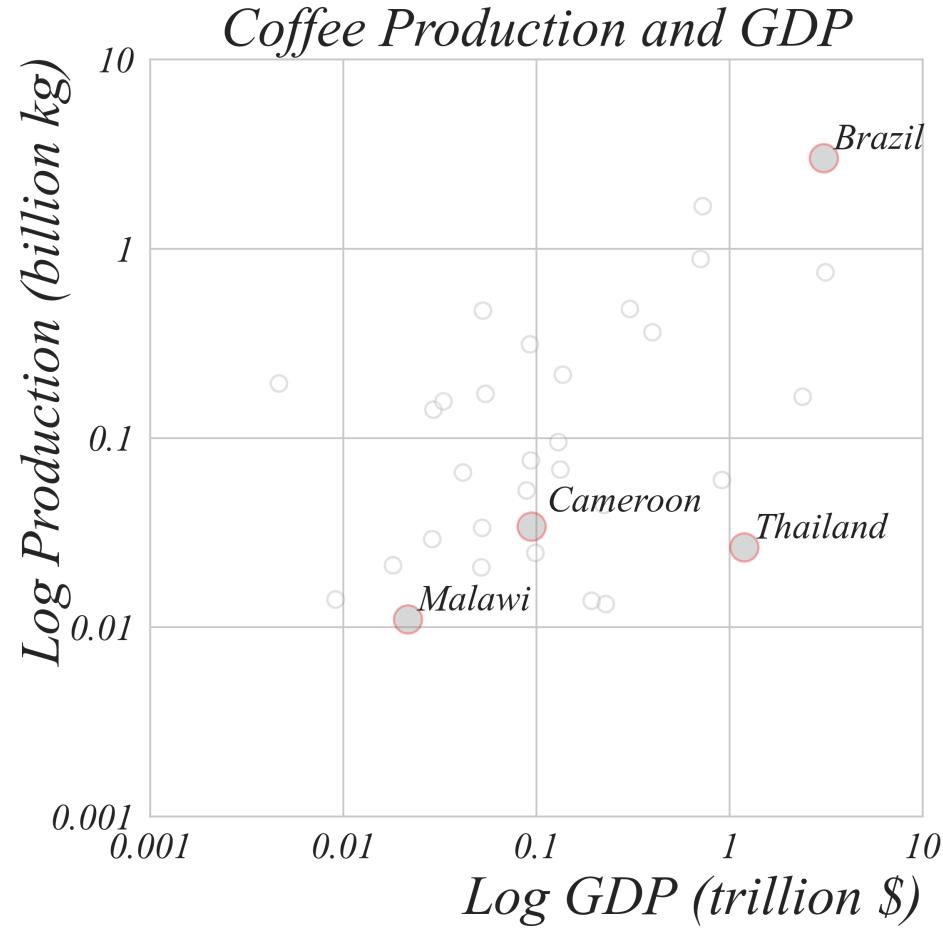
Which country produces the second highest output of coffee?



> a log scale also makes it easier to see SCALING

Relationships Between Variables

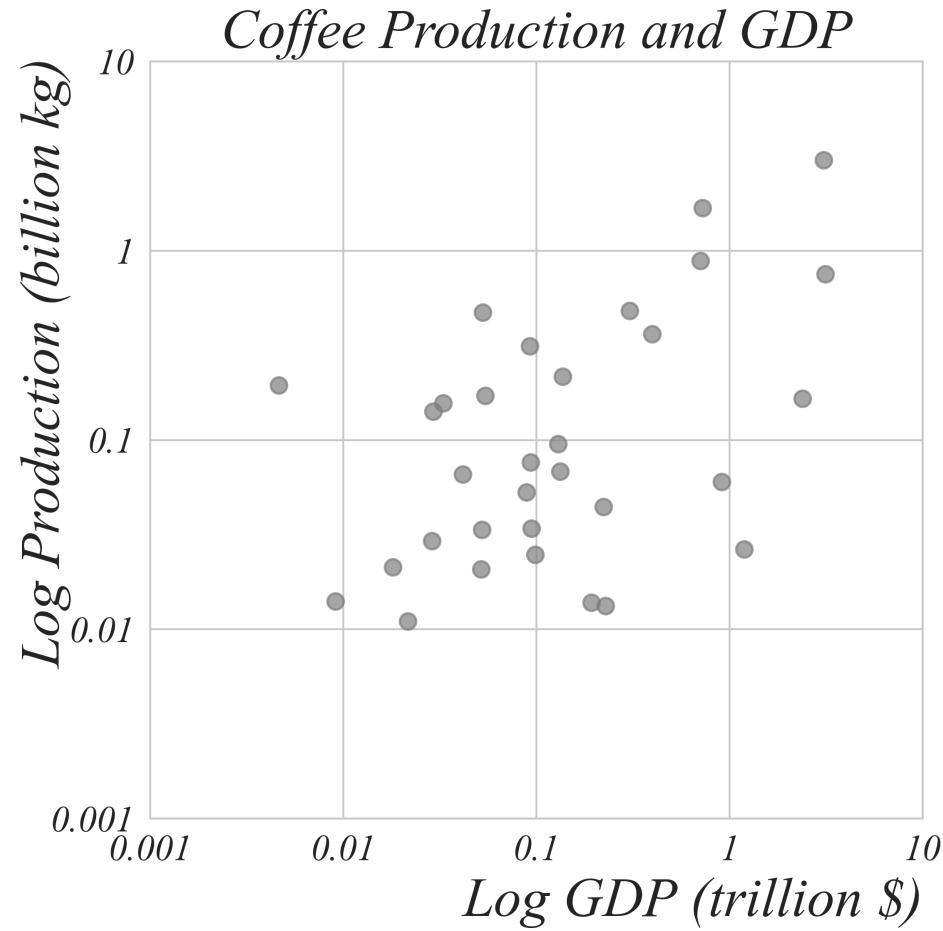
Which country produces the second highest output of coffee?



> scaling the vertical axis in logs clarifies both small and large variation

Relationships Between Variables

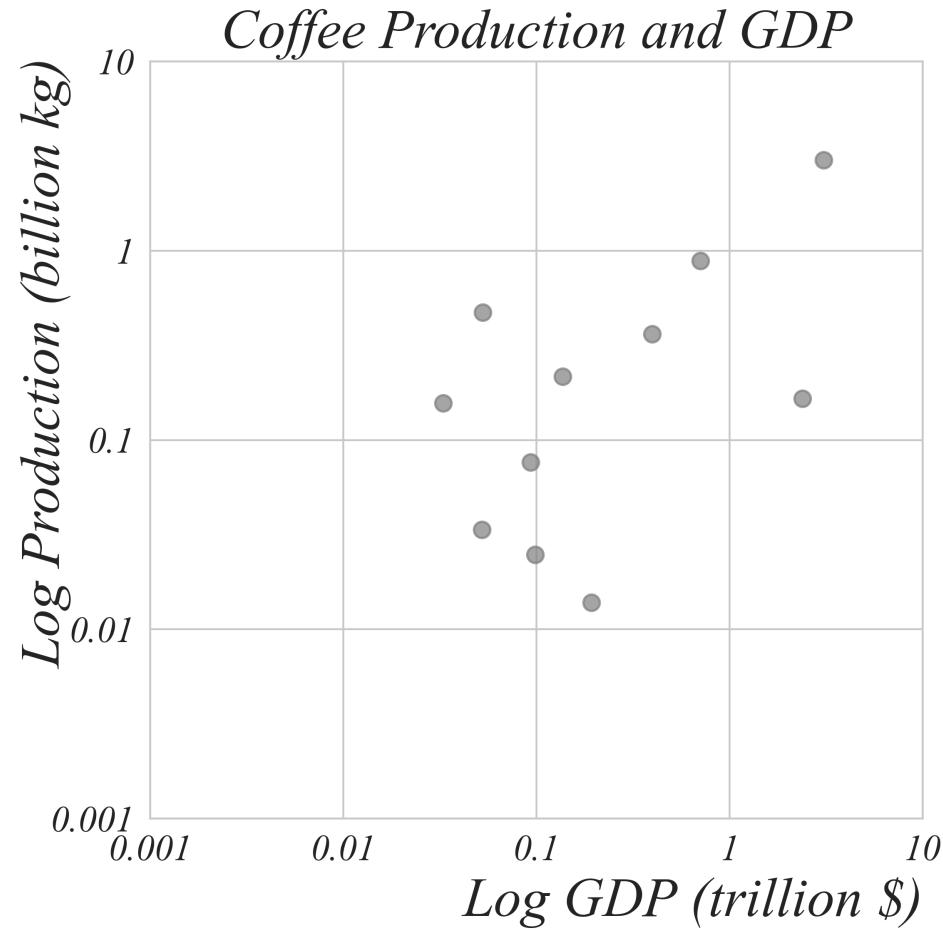
How does GDP relate to coffee production in the Americas?



> lets use a filter with this data

Relationships Between Variables

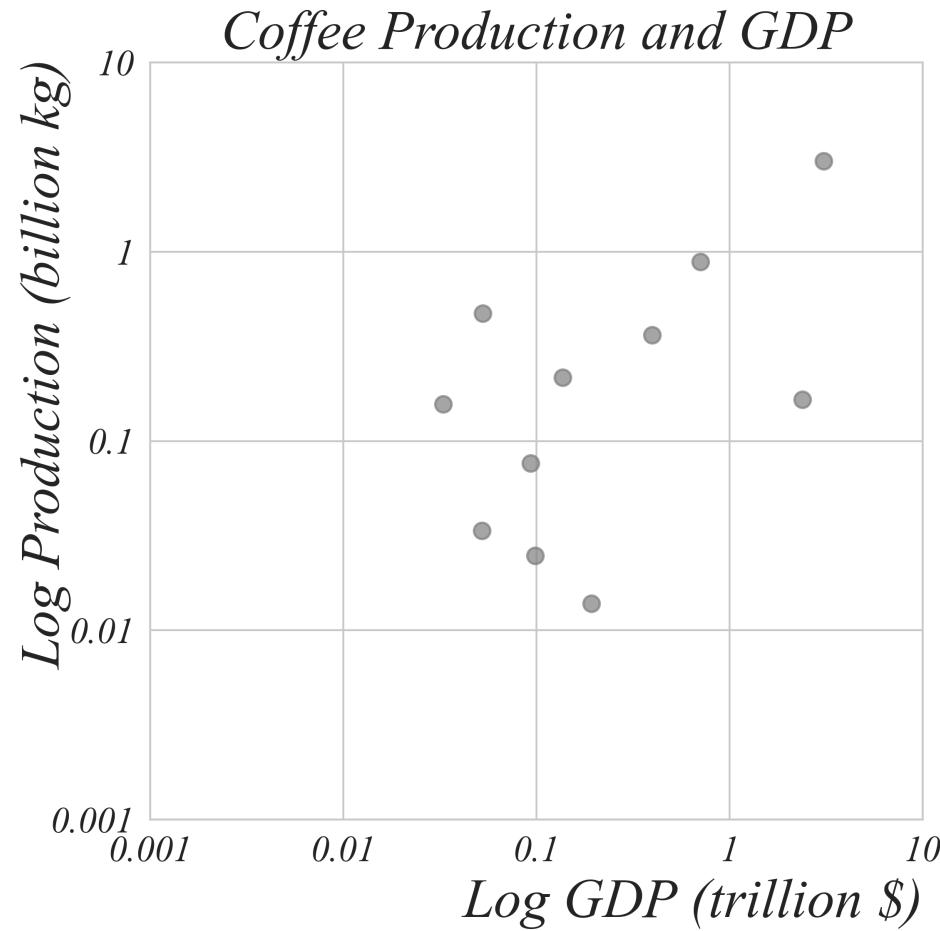
How does GDP relate to coffee production in the Americas?



> looks positive, but we'll formally test this in Part 4

Relationships Between Variables

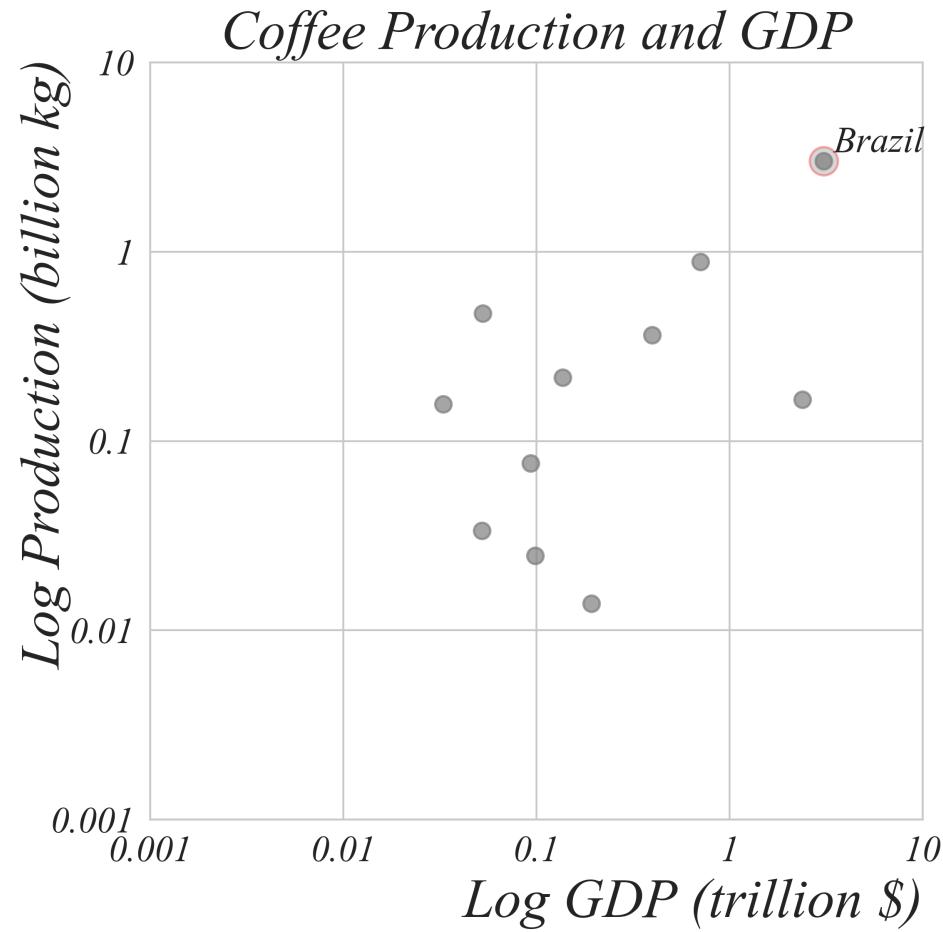
Which country in the Americas produces the most coffee?



> looks positive, but we'll formally test this in Part 4

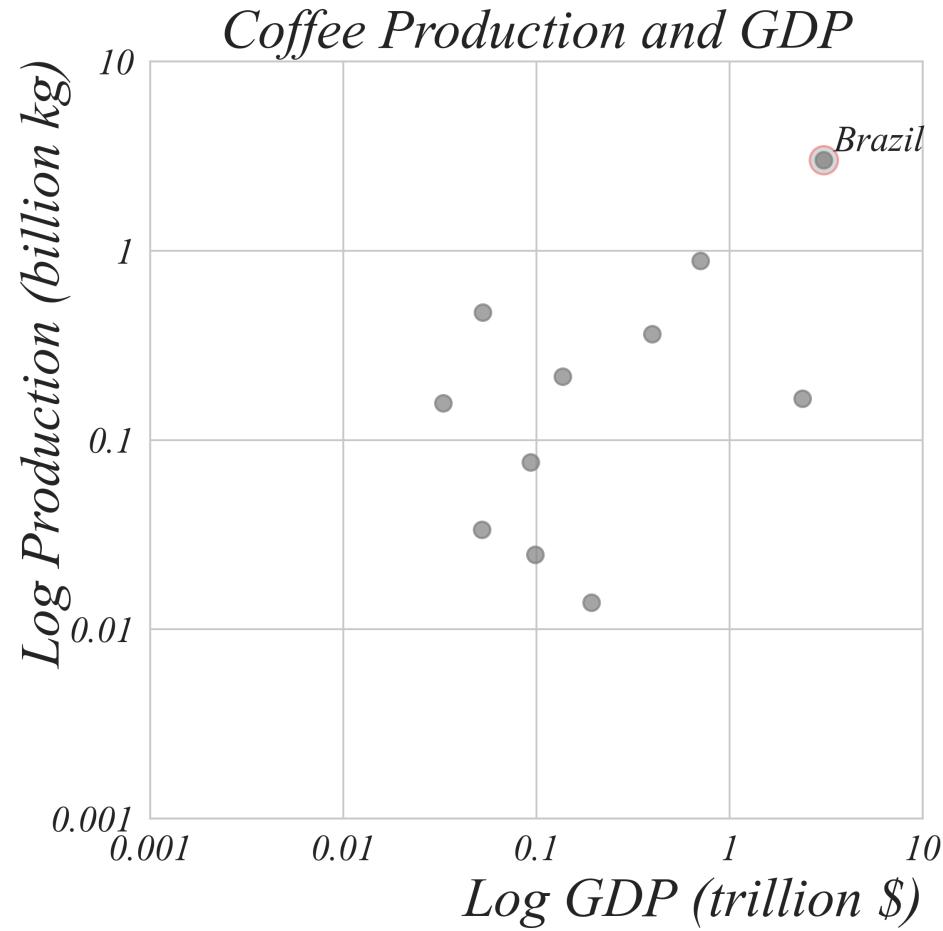
Relationships Between Variables

Which country in the Americas produces the most coffee?



Relationships Between Variables

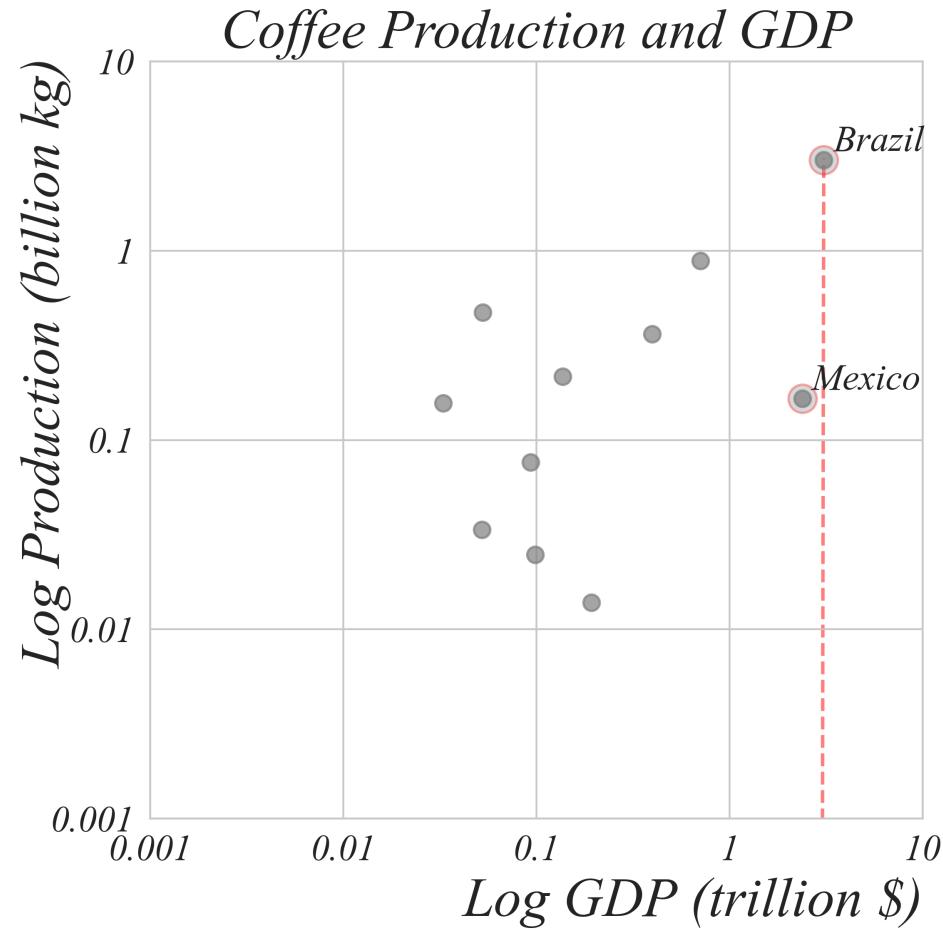
Which country's GDP is closest to Brazil's GDP?



> we can use the horizontal axis

Relationships Between Variables

Which country's GDP is closest to Brazil's GDP?



> we can use a vertical line here to find the closest on the horizontal axis

Exercise: Scatterplots

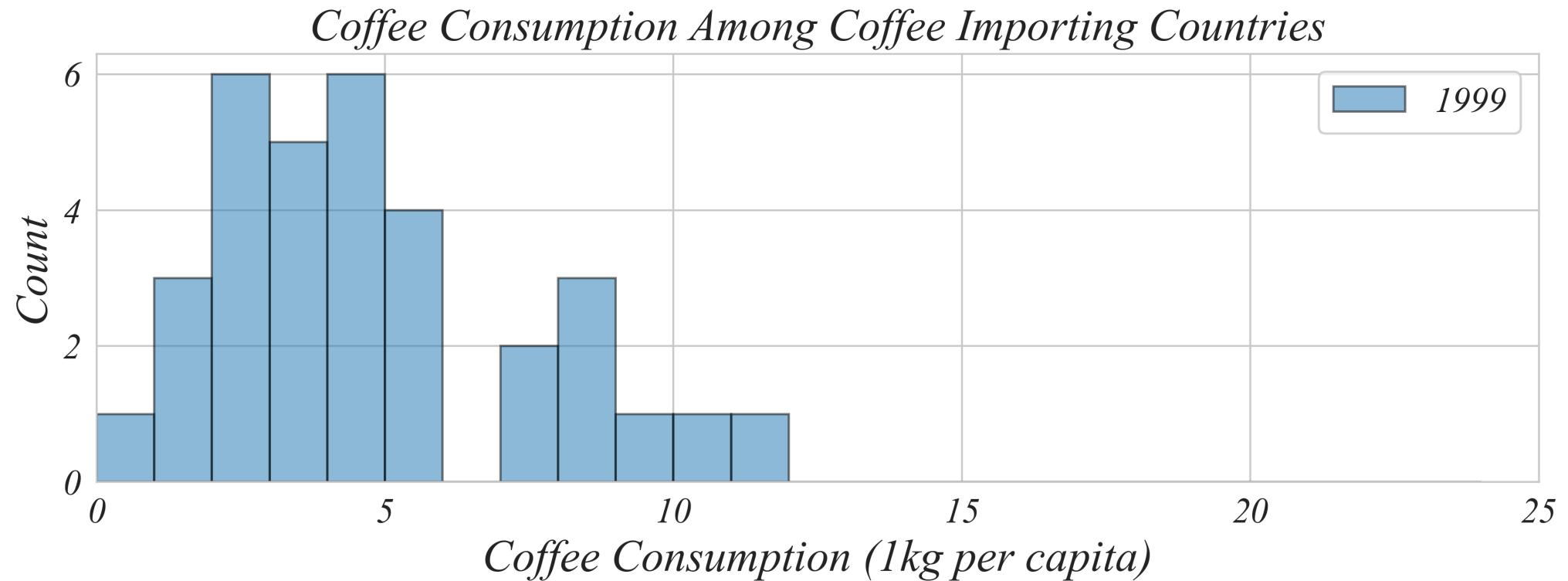
Visualizing GDP and Coffee Production Relationships

We're going to use a scatterplot to visually examine the relationship between coffee production and GDP.

- *Data: Beans_GDP.csv*

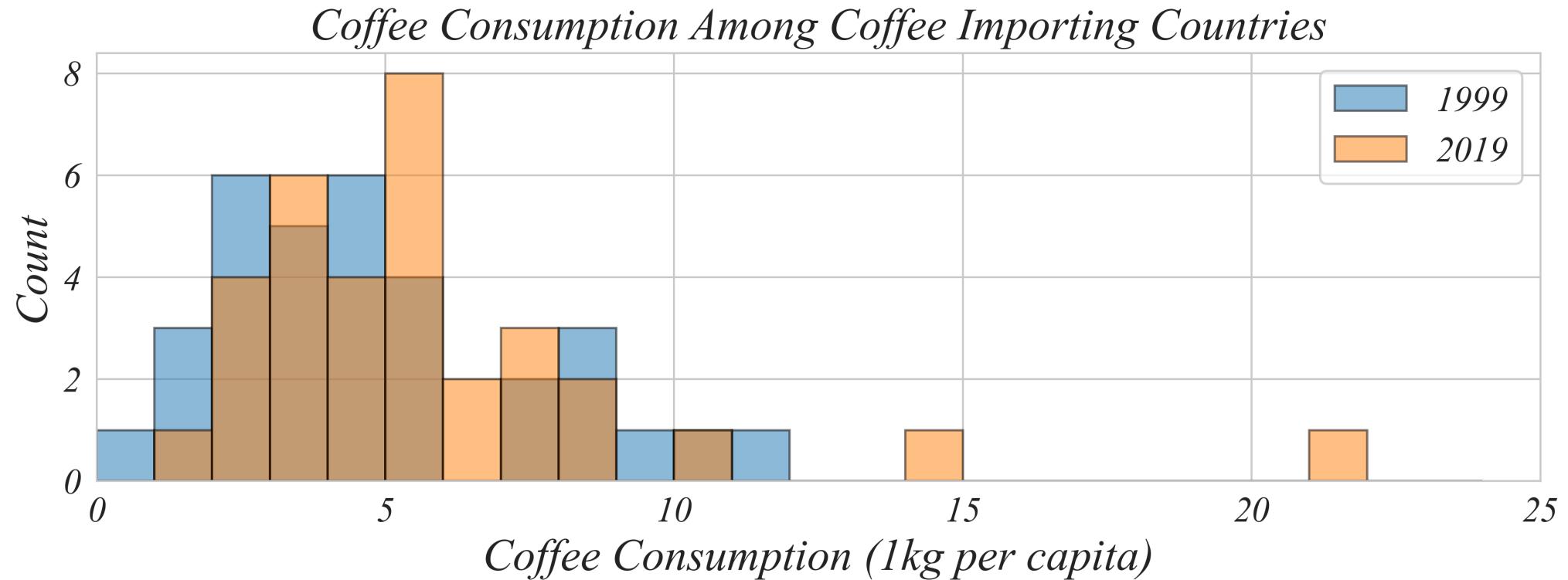
Relationships Between Variables

Does the data confirm that the world is drinking more coffee?



Relationships Between Variables

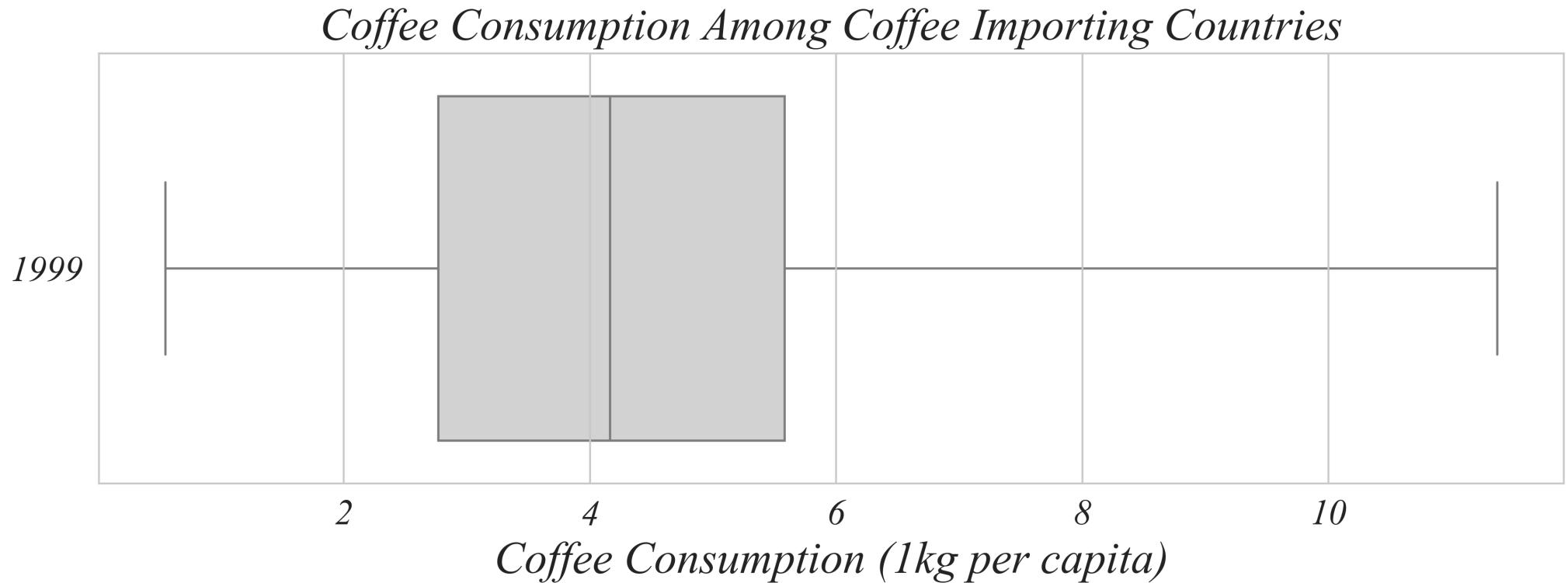
Does the data confirm that the world is drinking more coffee?



- > this is still pretty unclear
- > lets focus on one year for a sec

Relationships Between Variables

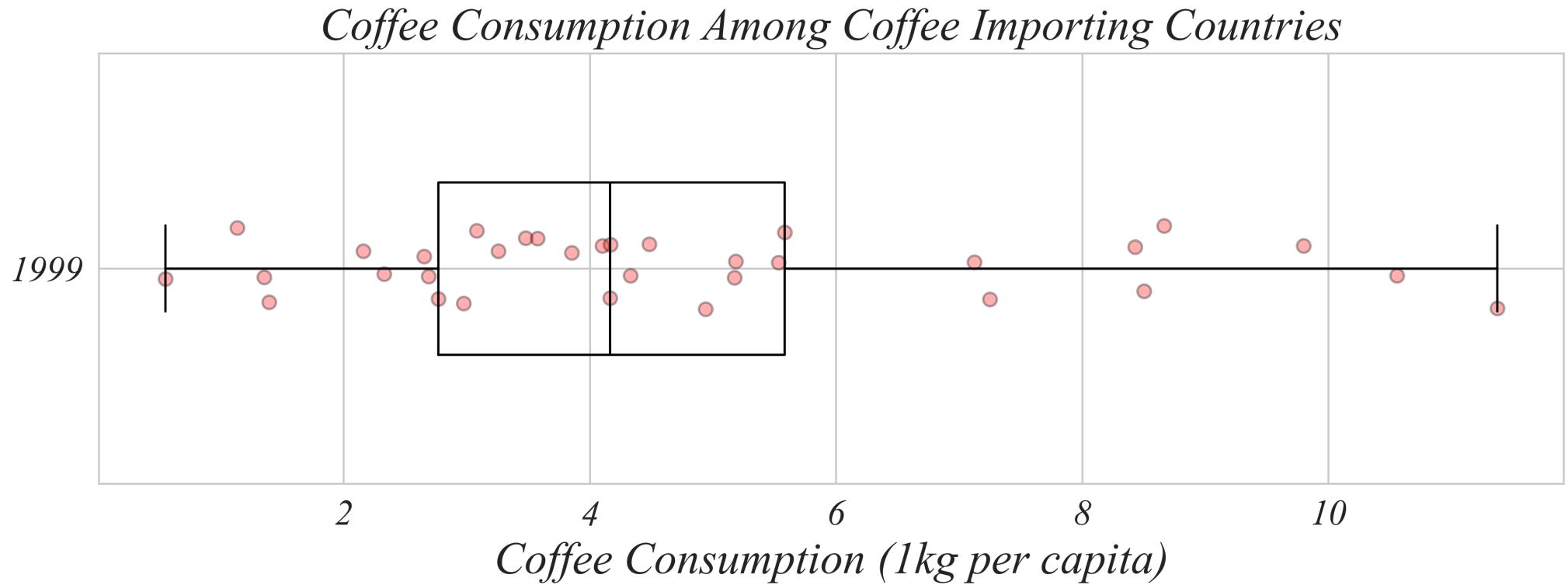
Does the data confirm that the world is drinking more coffee?



> boxplots can tell us about quartiles

Relationships Between Variables

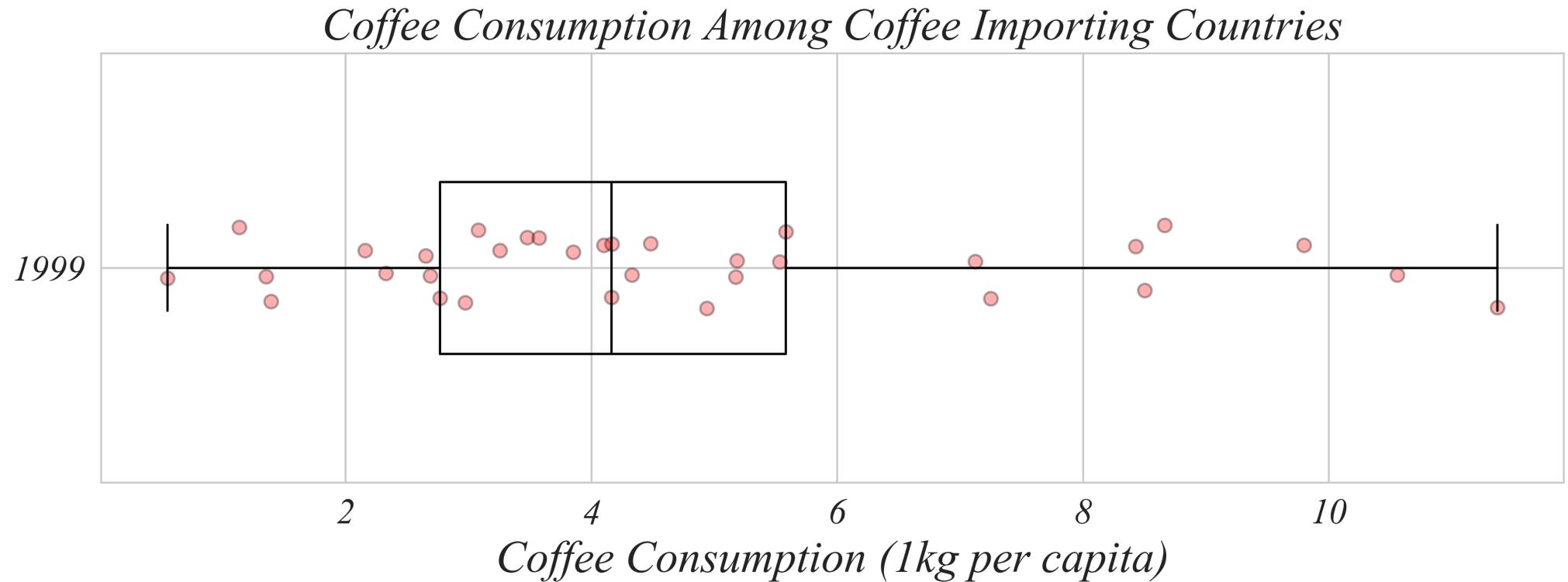
Does the data confirm that the world is drinking more coffee?



> adding a jittered scatter makes it even clearer

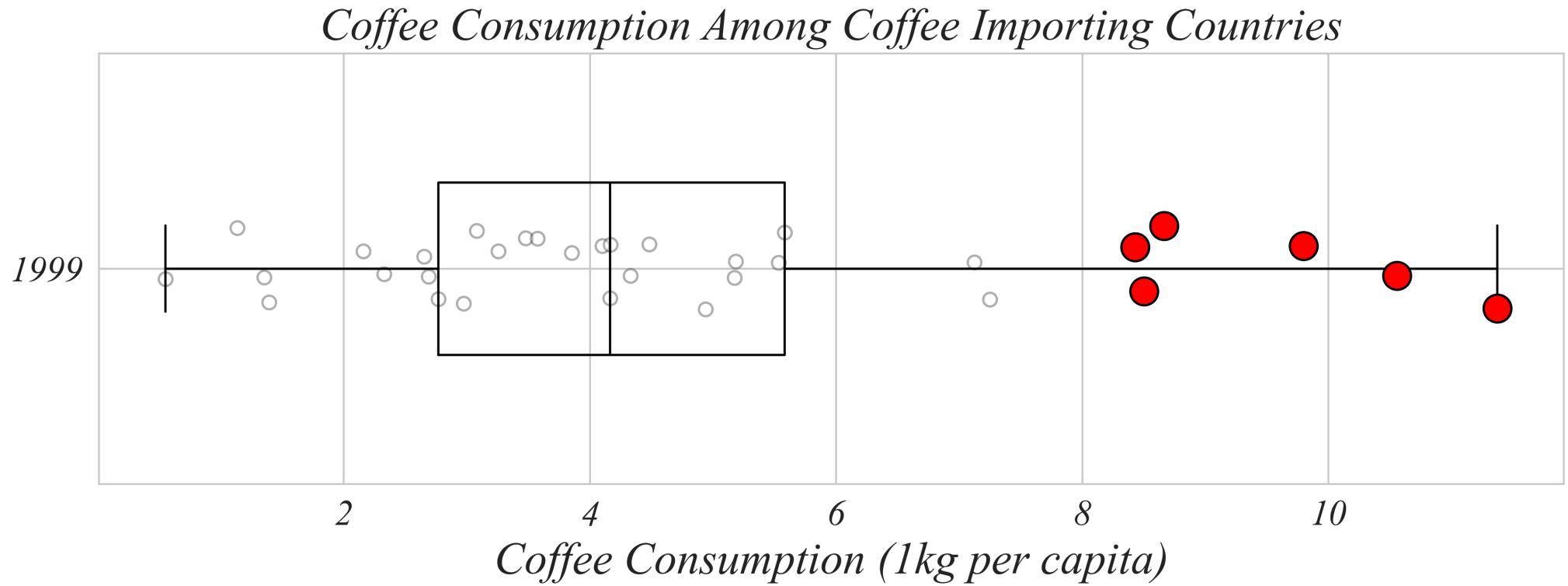
Relationships Between Variables

Which countries consume more than 8 kg per capita?



Relationships Between Variables

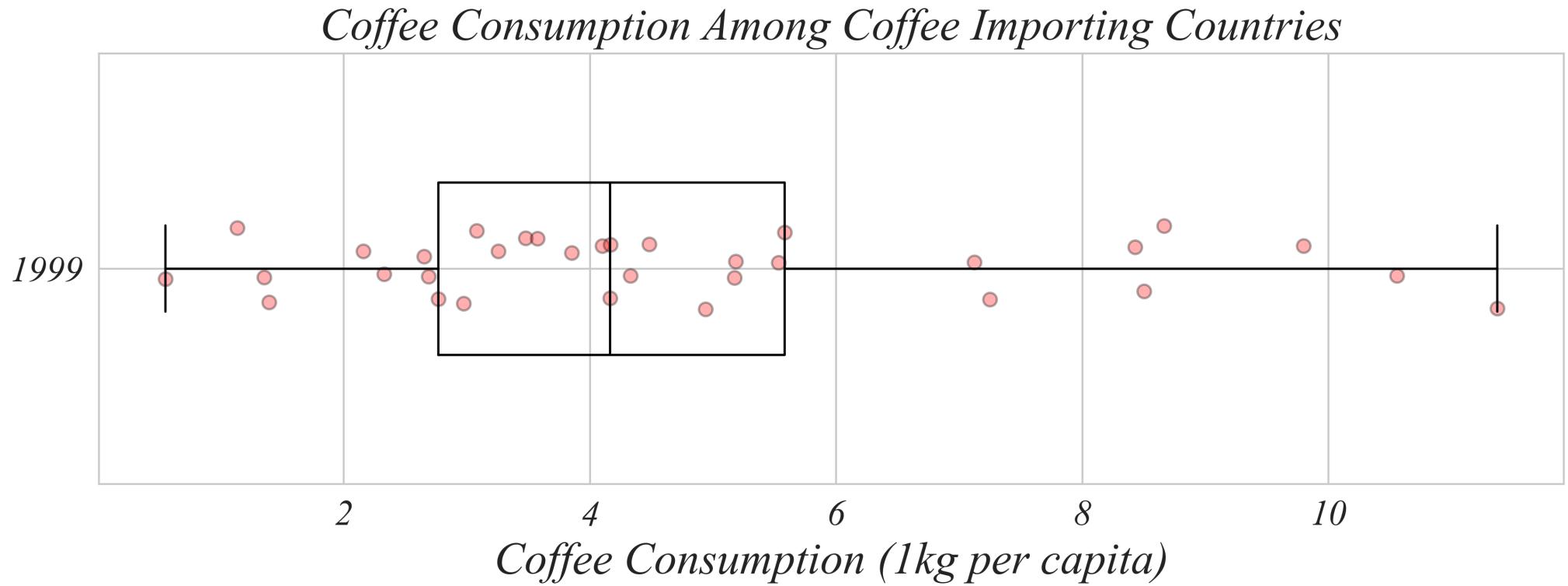
Which countries consume more than 8 kg per capita?



> we can highlight the relevant subsets of the data

Relationships Between Variables

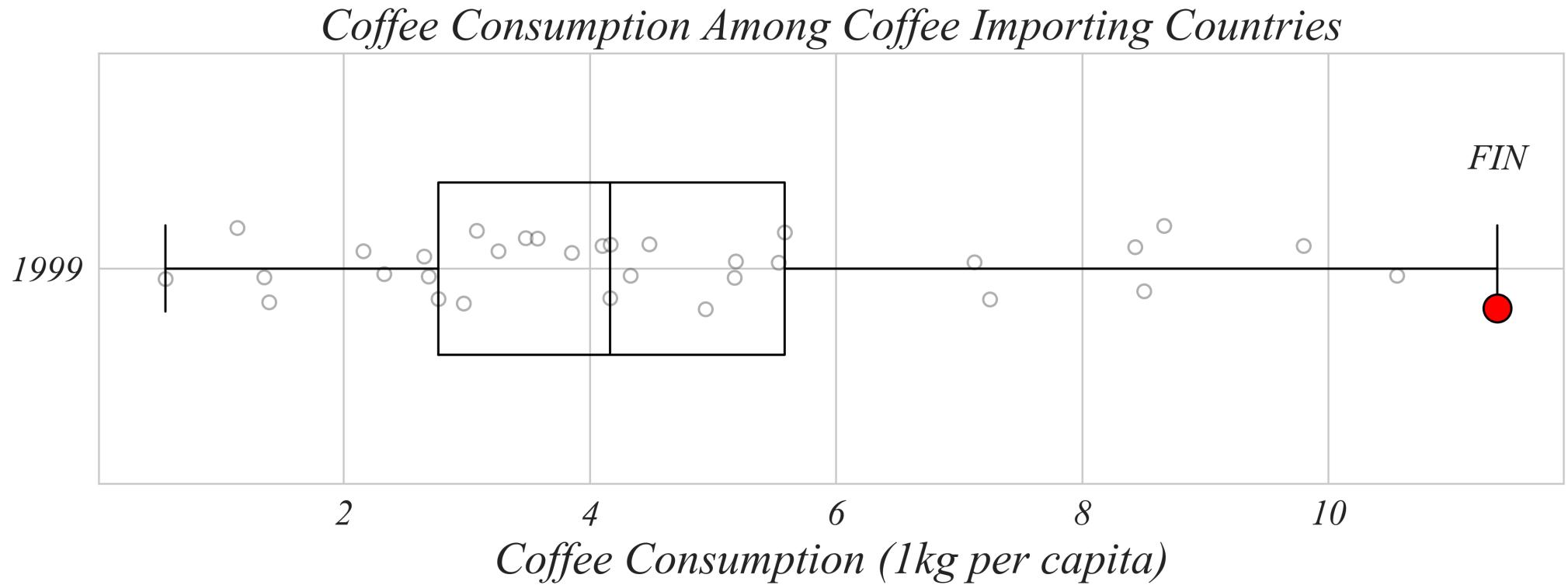
Which country consumes the most coffee per capita?



> we can find the exact values according to quartiles

Relationships Between Variables

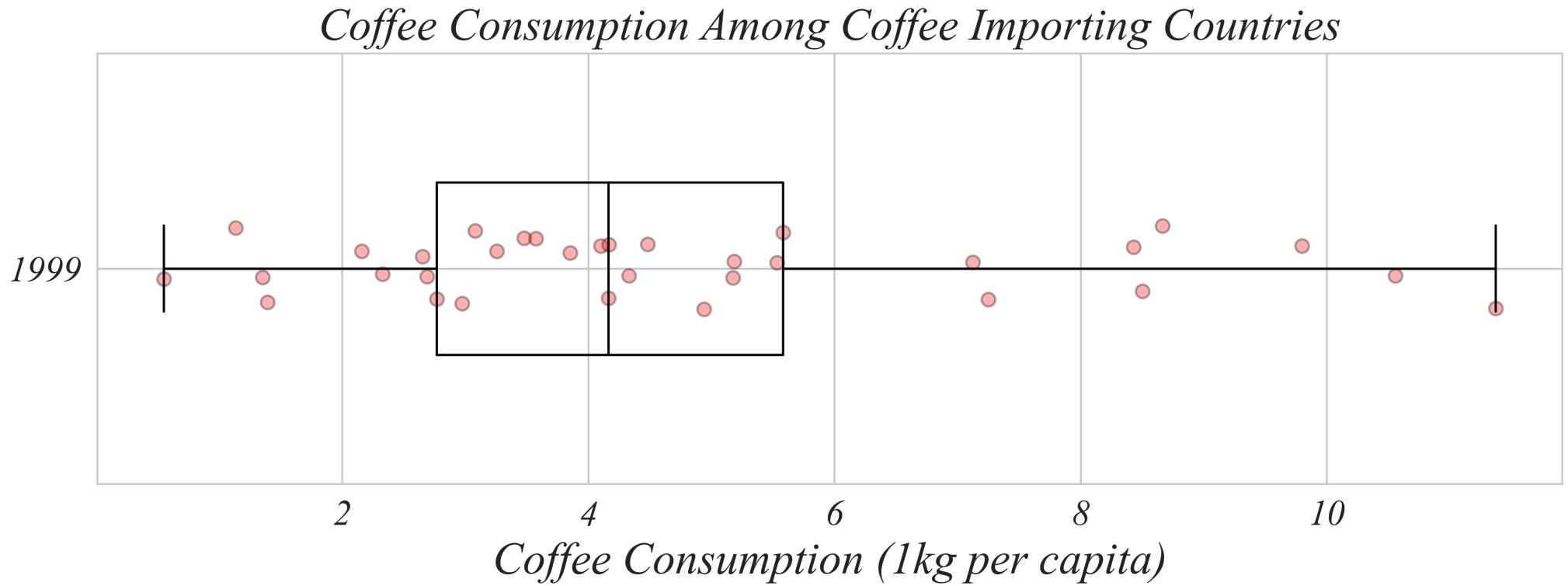
Which country consumes the most coffee per capita?



> we can find the exact values according to quartiles

Relationships Between Variables

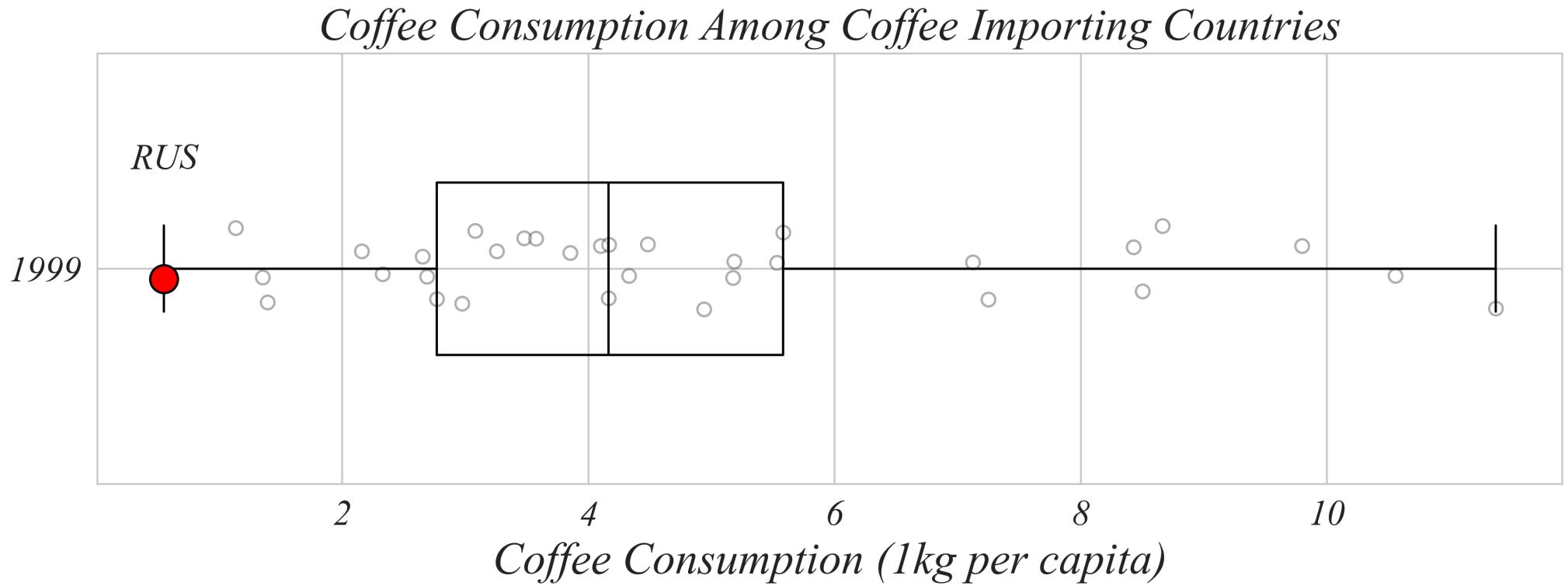
How about the least consumption per capita?



> we can find the exact values according to quartiles

Relationships Between Variables

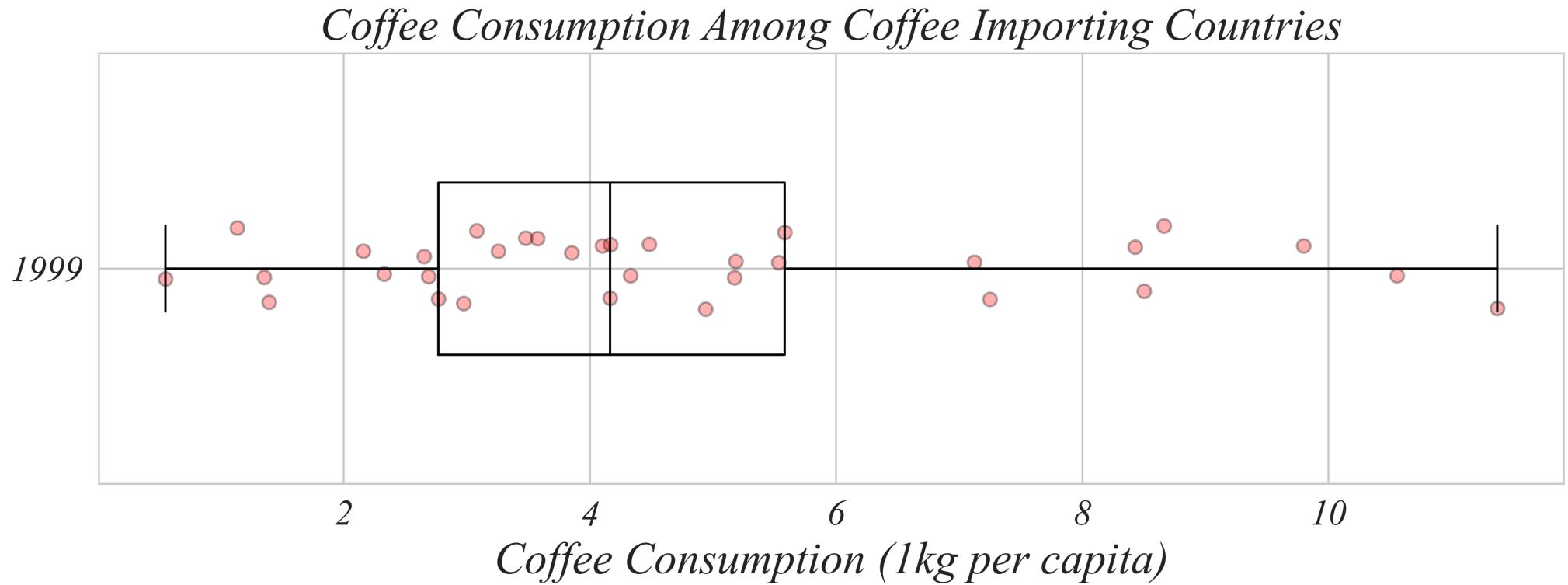
How about the least consumption per capita?



> we can find the exact values according to quartiles

Relationships Between Variables

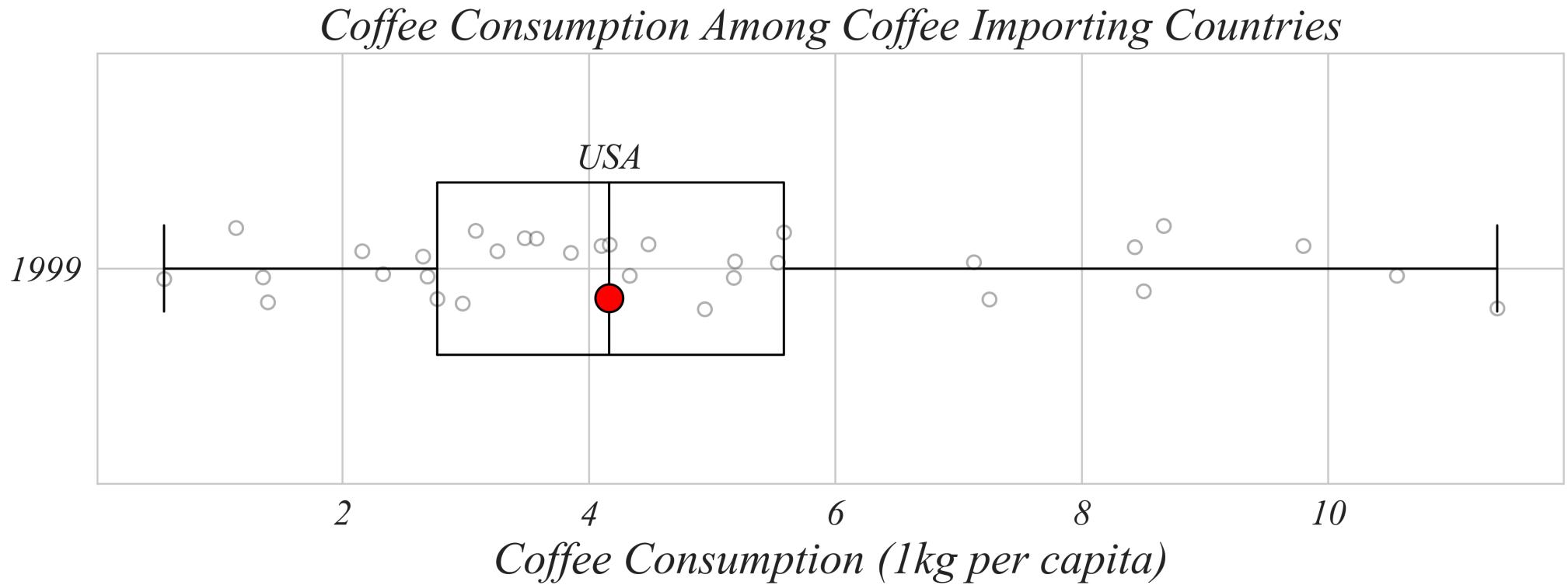
How about the median?



> we can find the exact values according to quartiles

Relationships Between Variables

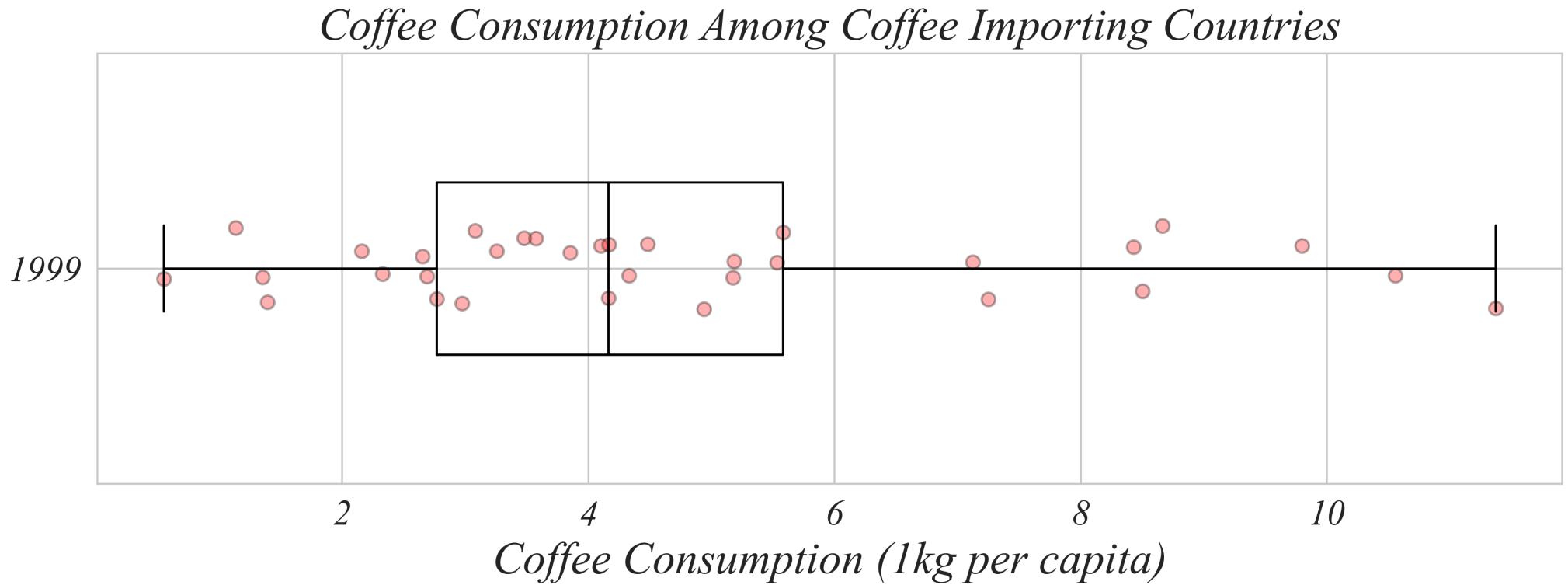
How about the median?



> we can find the exact values according to quartiles

Relationships Between Variables

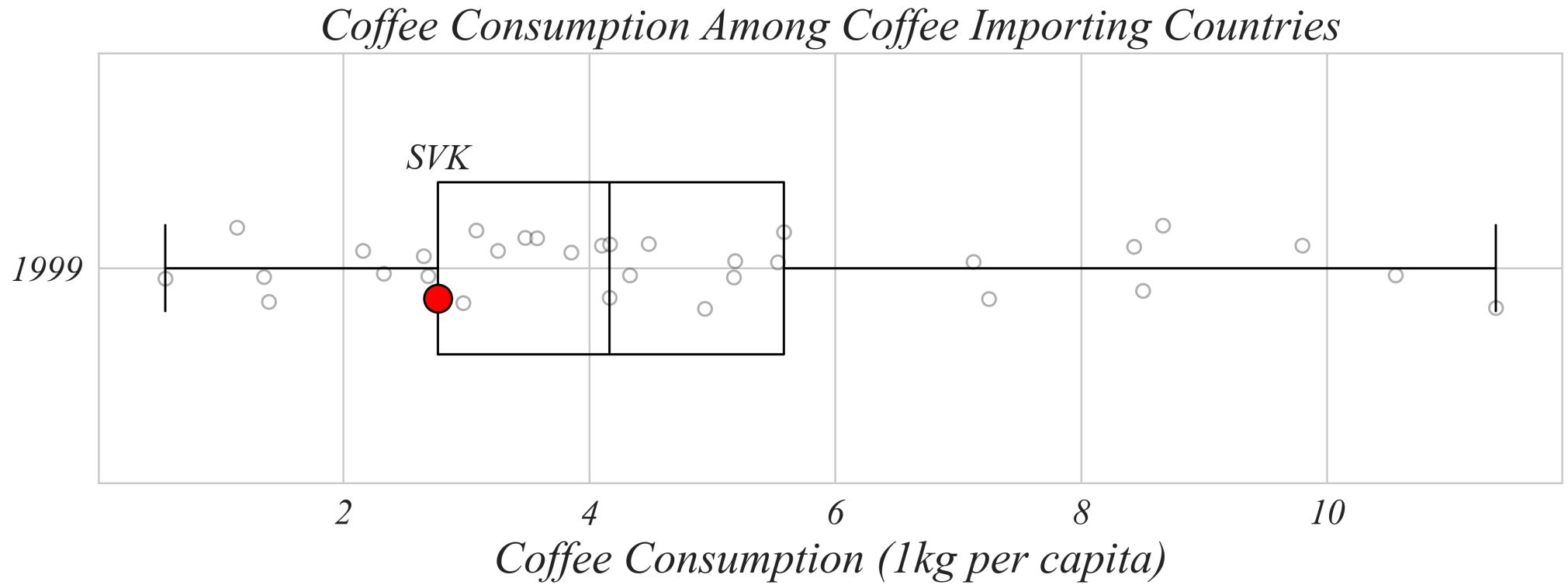
Which country consumes more than exactly 25% of countries?



> we can find the exact values according to quartiles

Relationships Between Variables

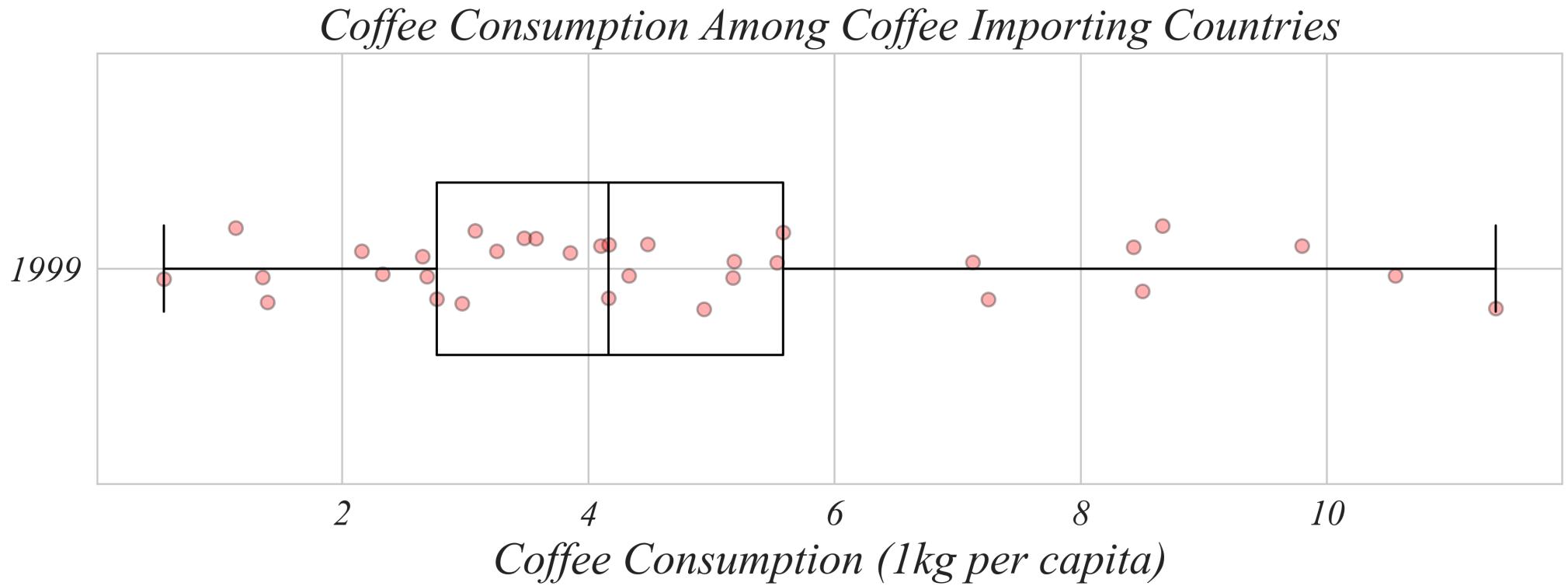
Which country consumes more than exactly 25% of countries?



> we can find the exact values according to quartiles

Relationships Between Variables

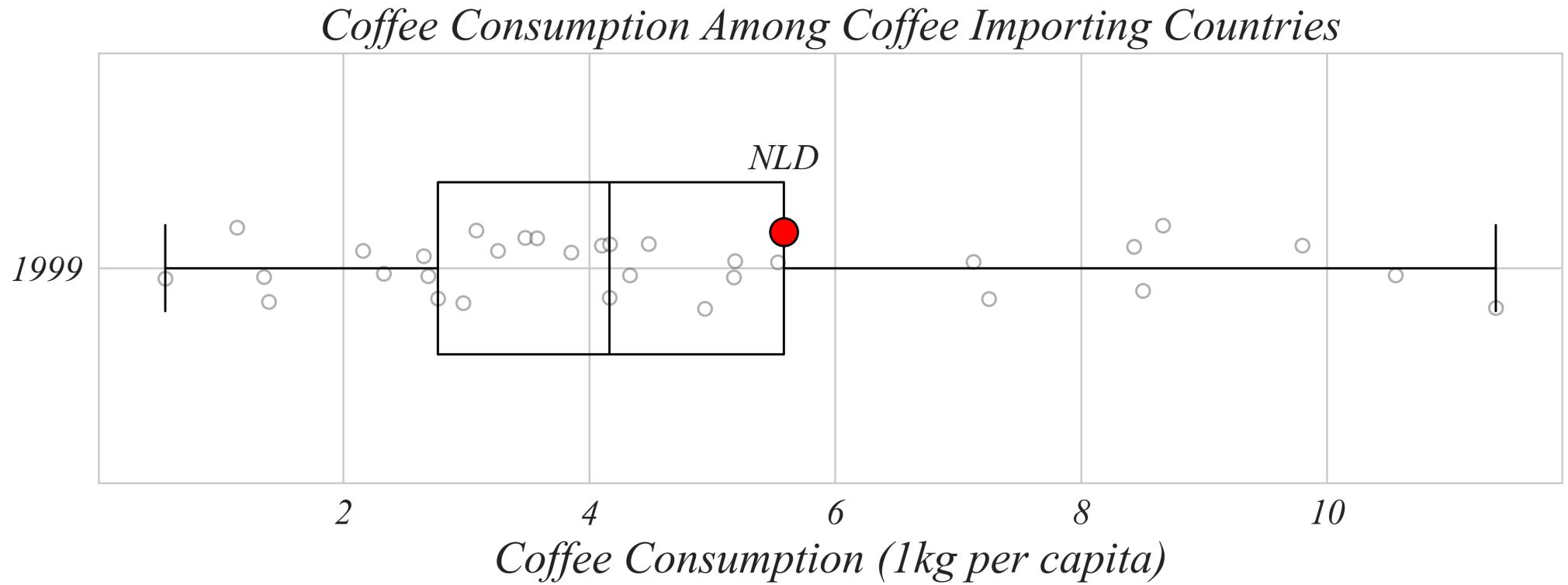
Which country consumes more than exactly 75% of countries?



> we can find the exact values according to quartiles

Relationships Between Variables

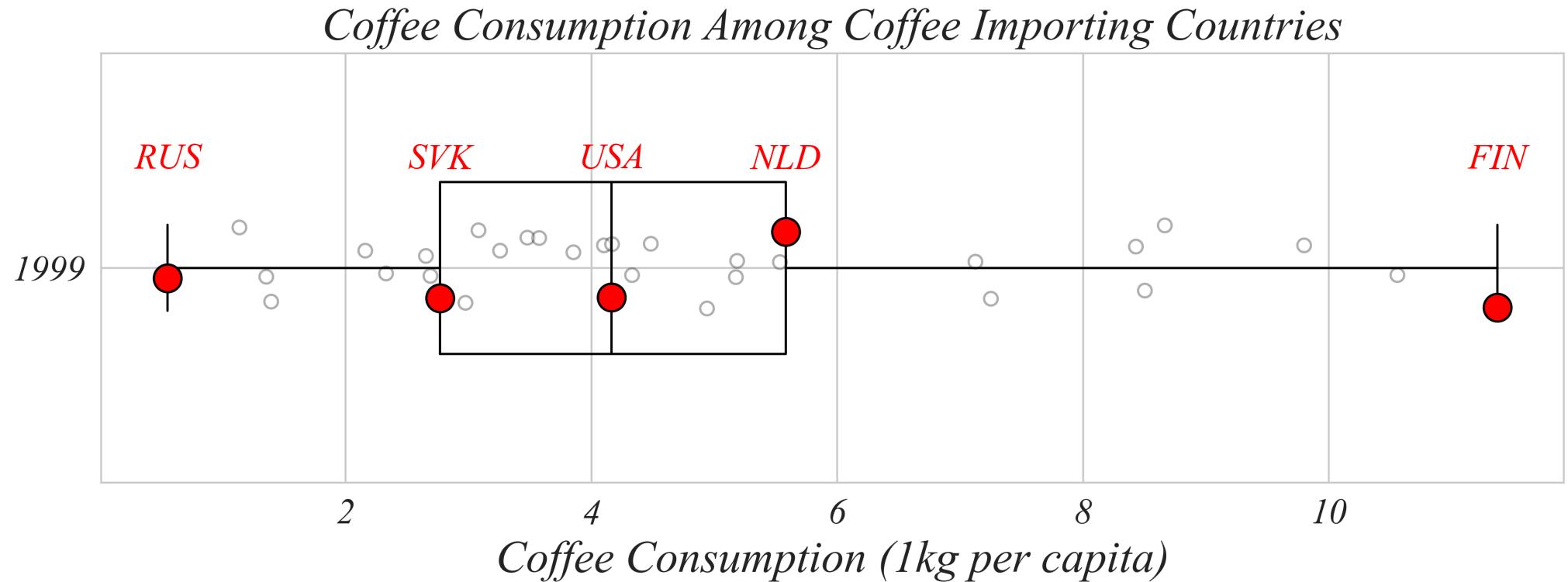
Which country consumes more than exactly 75% of countries?



> we can find the exact values according to quartiles

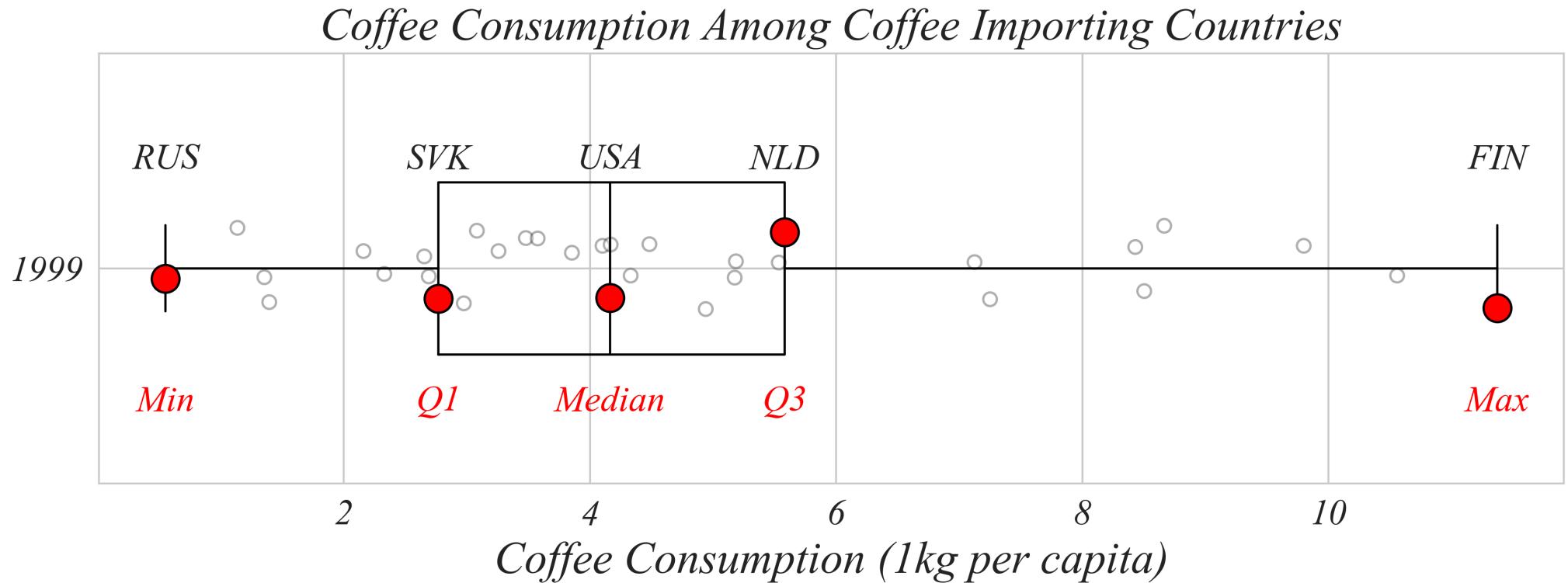
Relationships Between Variables

Does the data confirm that the world is drinking more coffee?



Relationships Between Variables

Does the data confirm that the world is drinking more coffee?



> no - boxplots can visualize the distribution of the data, but we need something else for comparison through time

Exercise: Boxplots

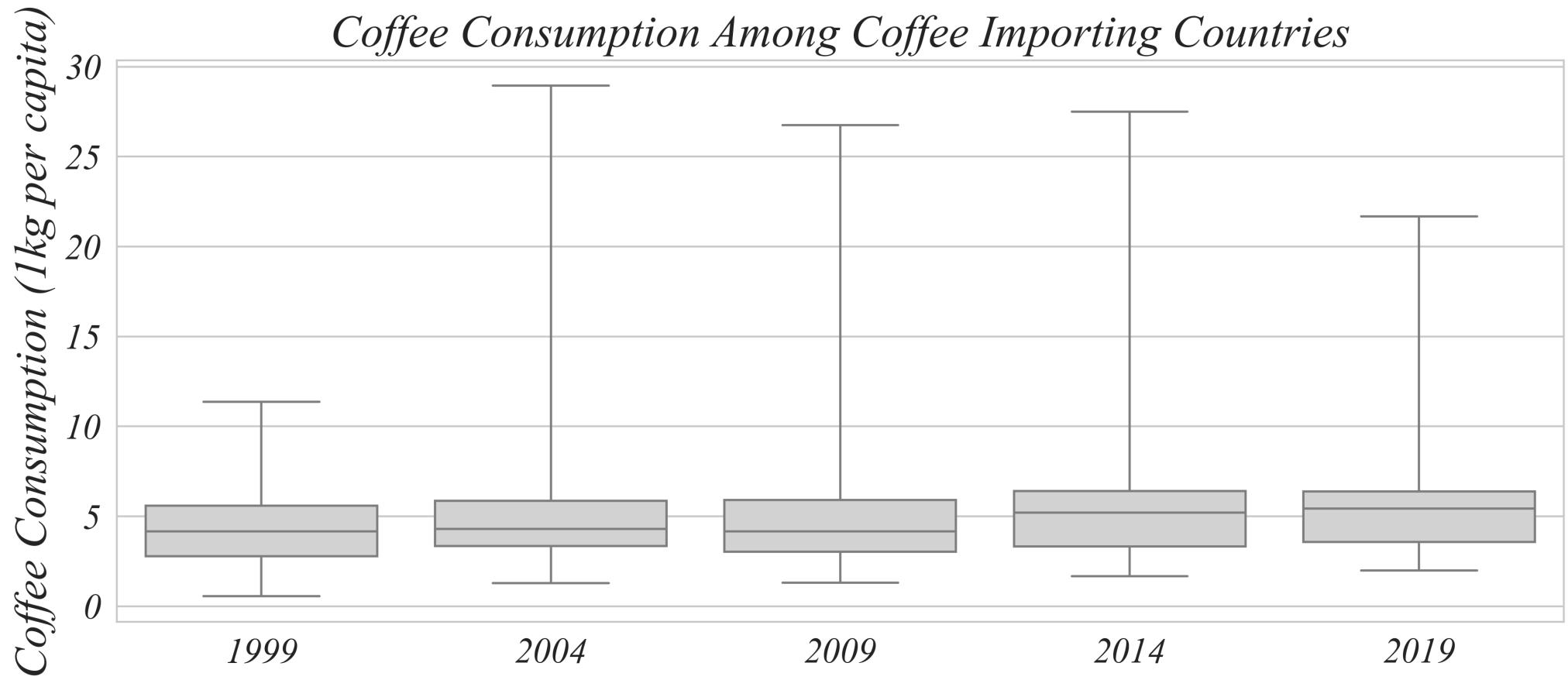
Exploring Coffee Consumption Distribution

We're going to use a boxplot to visually examine the distribution of coffee consumption per capita among coffee-importing countries.

- **Data:** *Coffee_Per_Cap.csv*

Relationships Between Variables

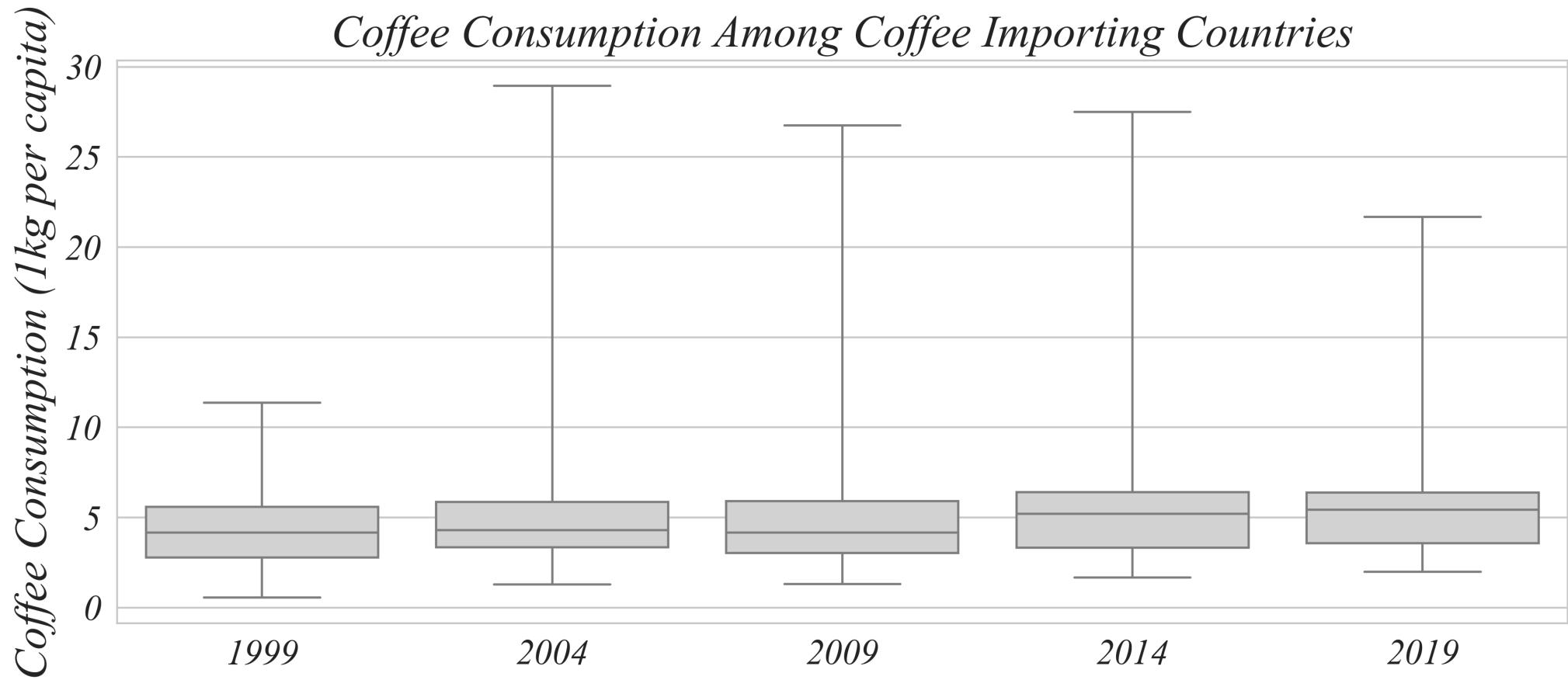
Does the data confirm that the world is drinking more coffee?



> lets maybe start with a smaller question

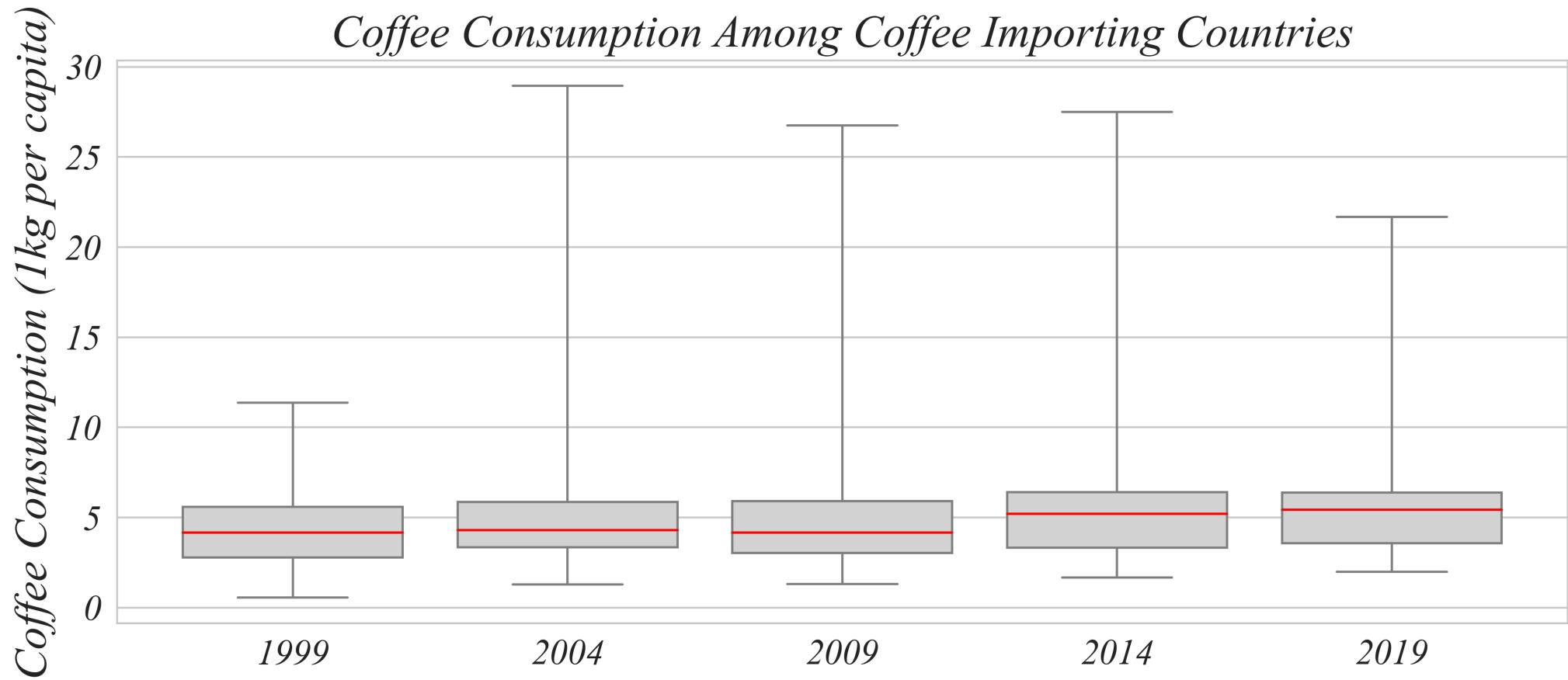
Relationships Between Variables

Which years show at least half consuming less than 5 kg per cap?



Relationships Between Variables

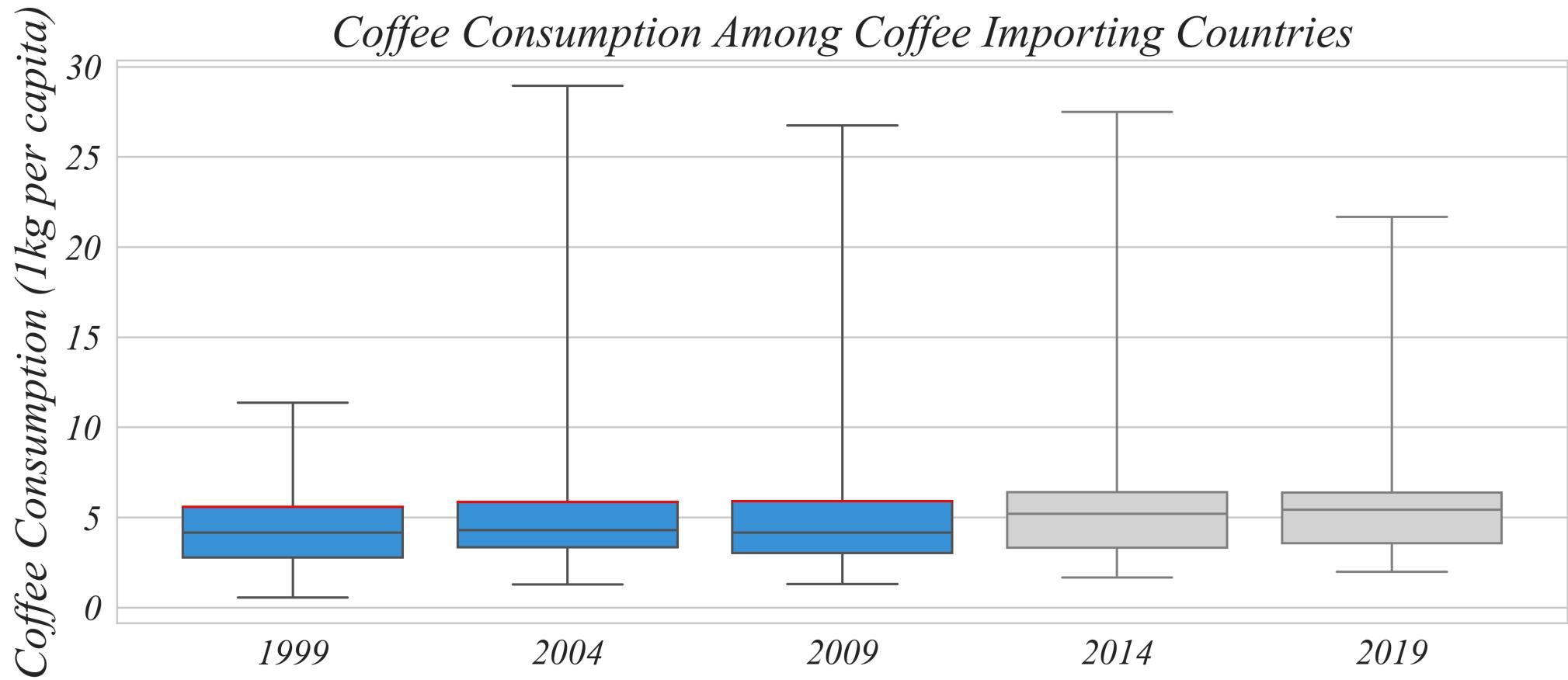
Which years show at least half consuming less than 5 kg per cap?



> focus on the medians

Relationships Between Variables

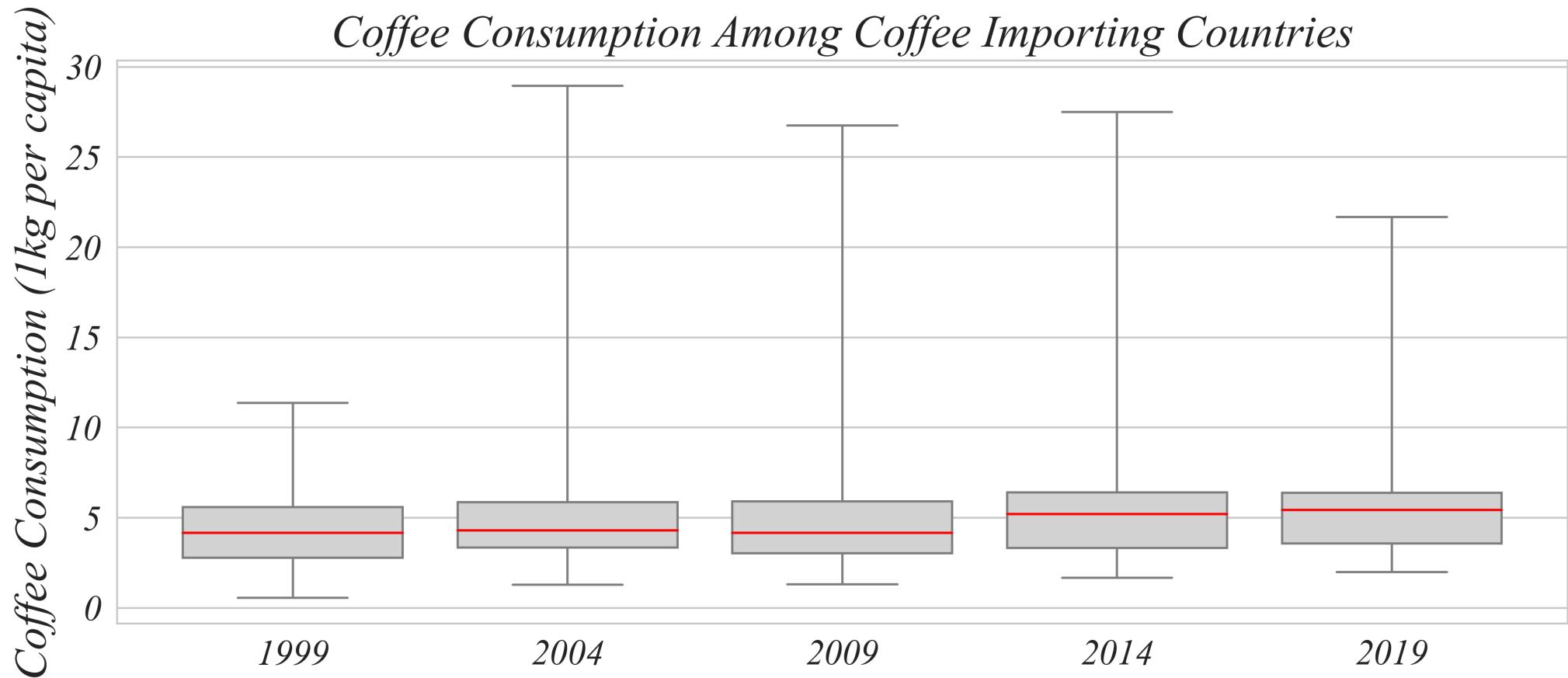
Which years show at least half consuming less than 5 kg per cap?



> ... when the median is above 5 kg per cap

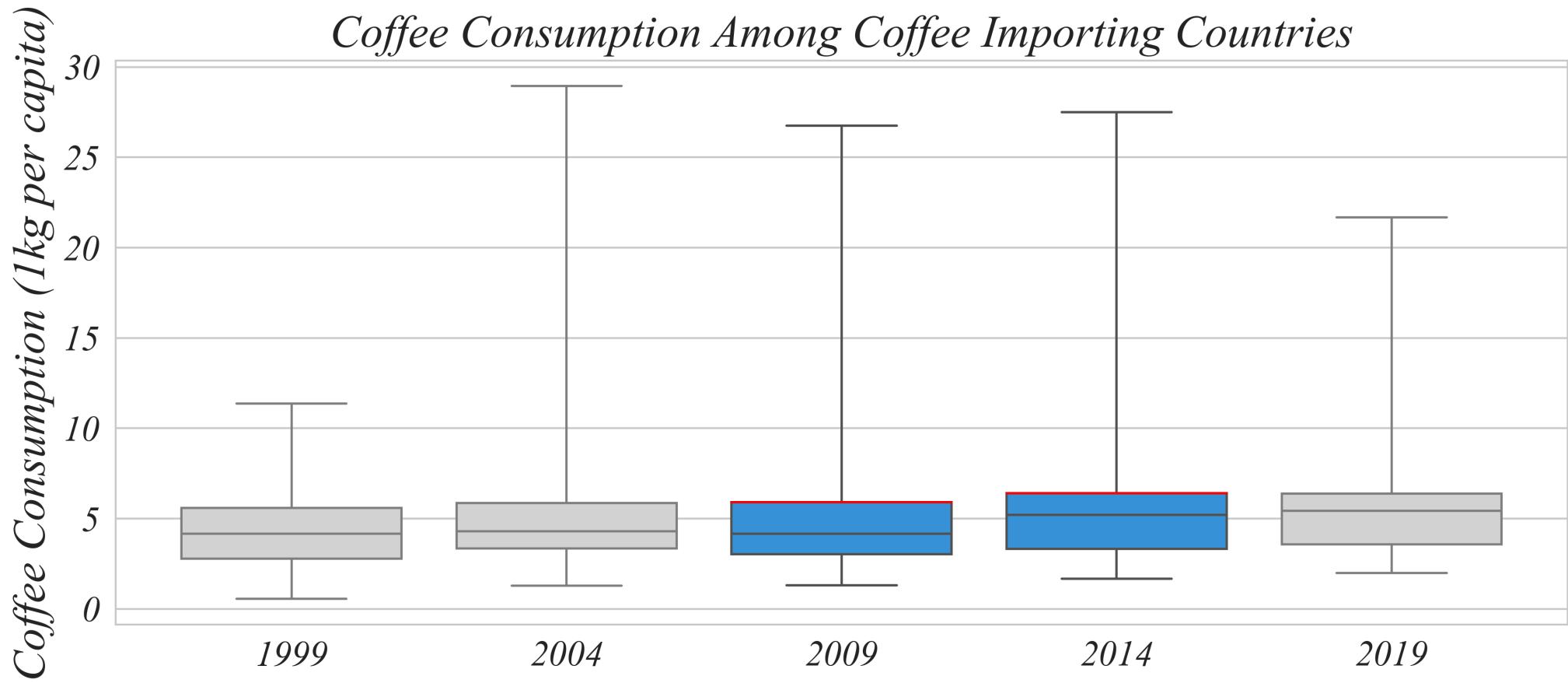
Relationships Between Variables

Between which two years was there the largest jump in the median?



Relationships Between Variables

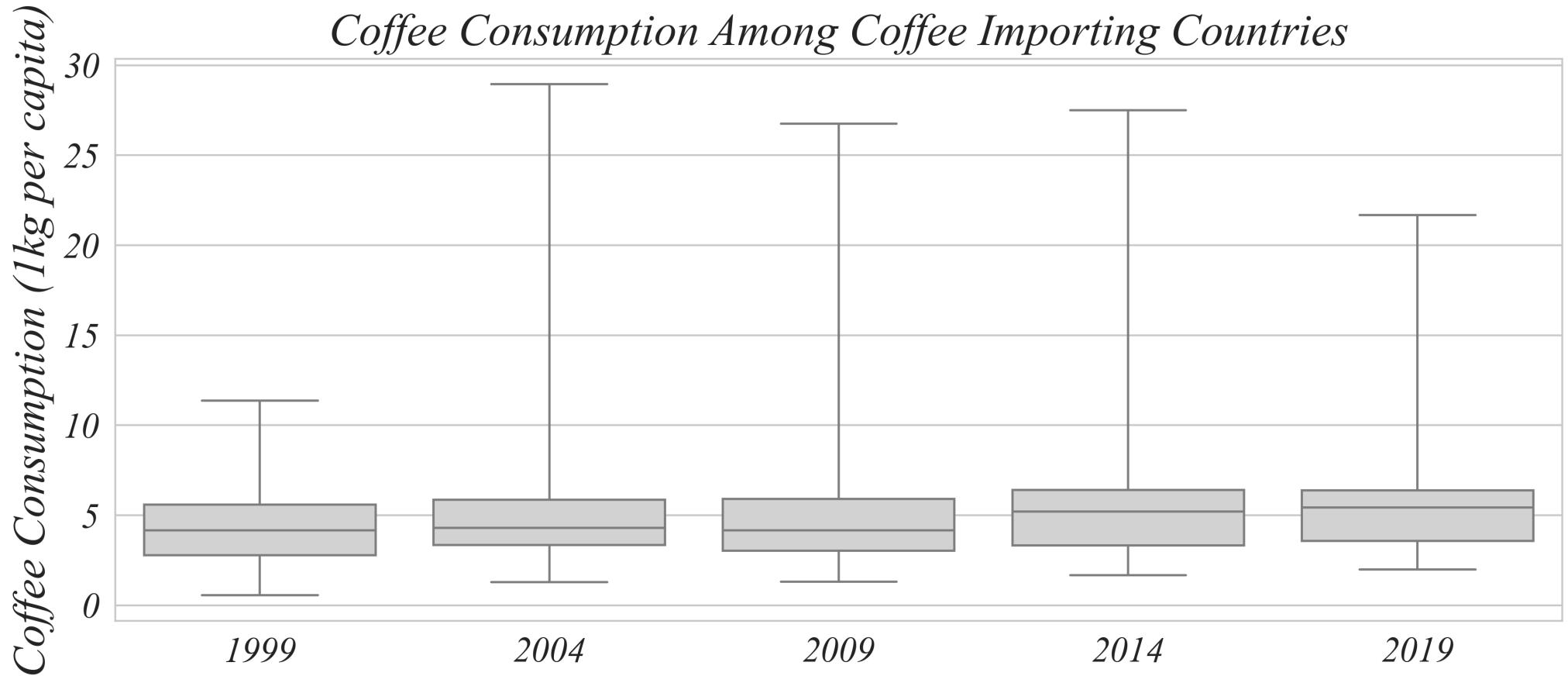
Between which two years was there the largest jump in the median?



> it's maybe a little difficult to see

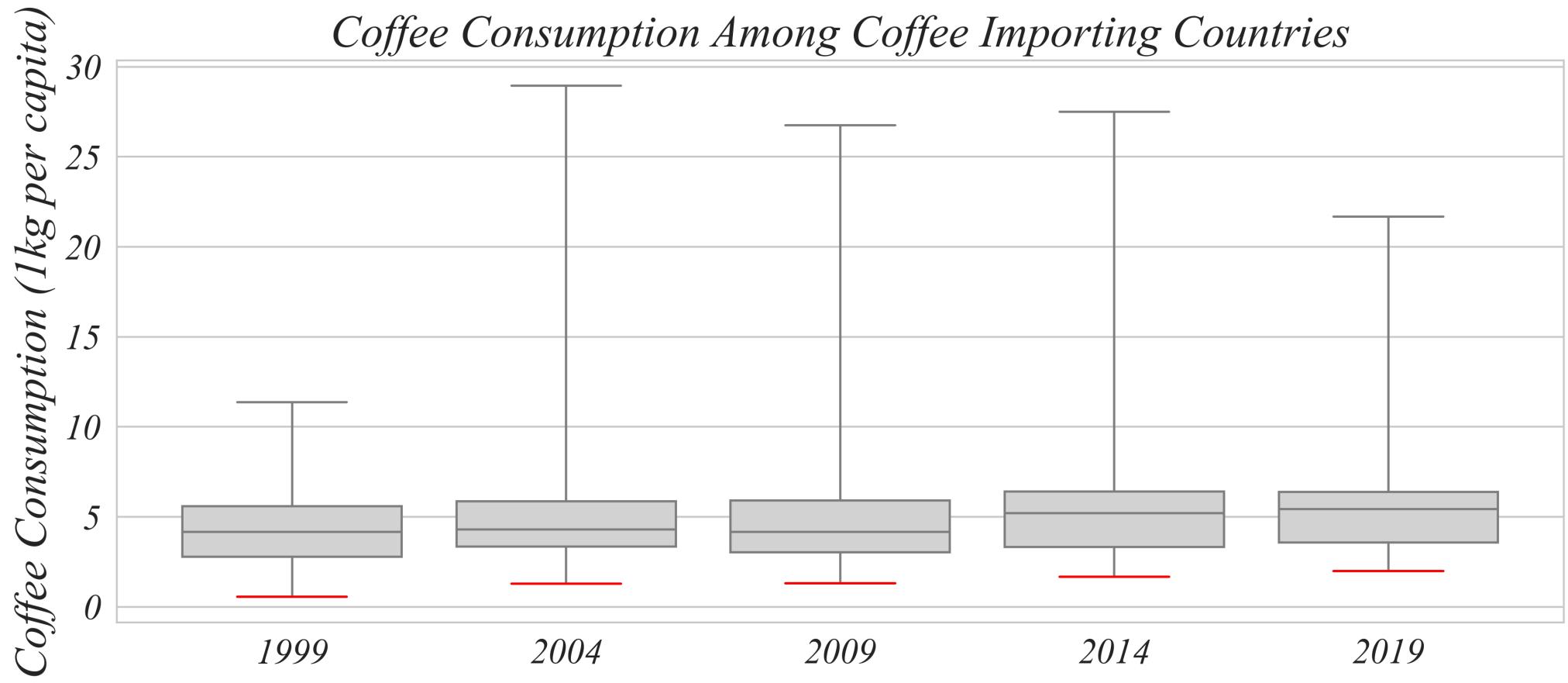
Relationships Between Variables

Is the country with the lowest consumption consuming more today?



Relationships Between Variables

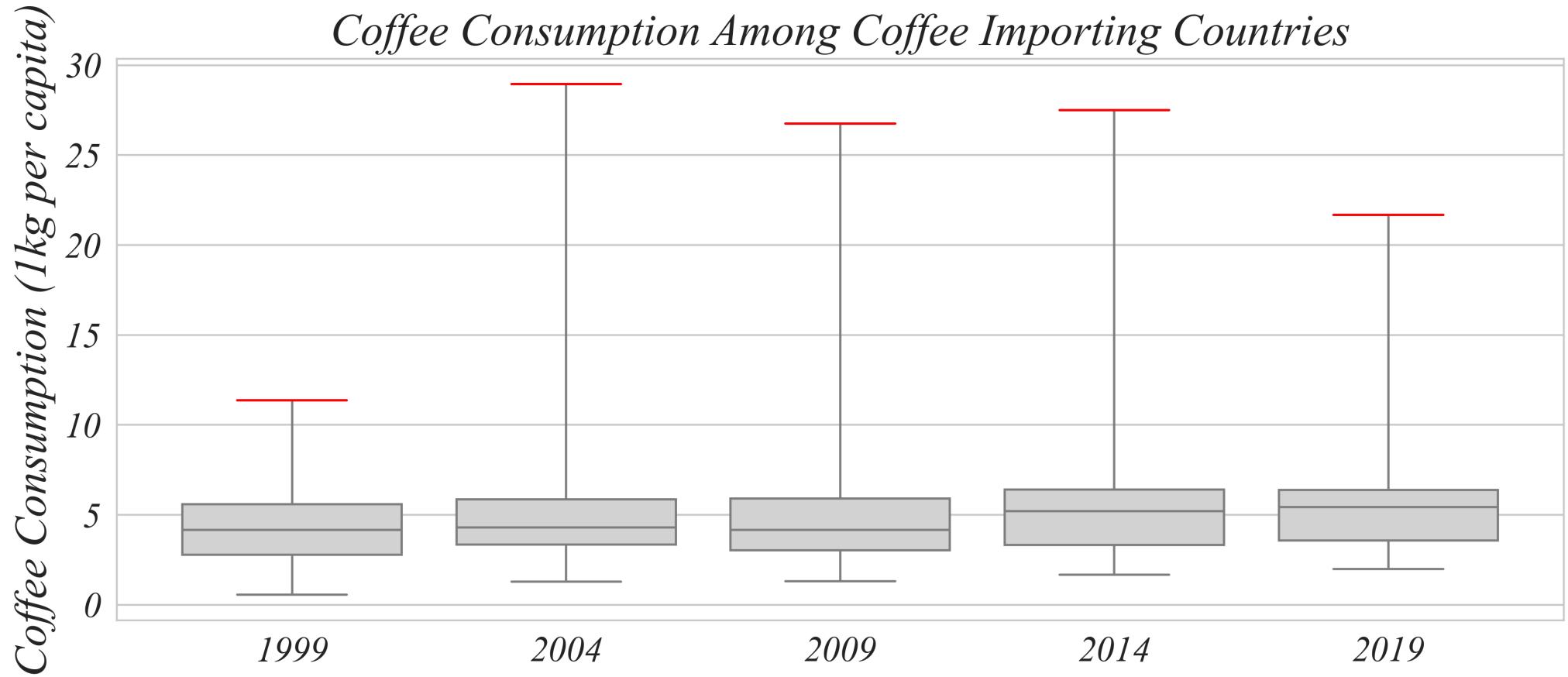
Is the country with the lowest consumption consuming more today?



> focus on the minimums

Relationships Between Variables

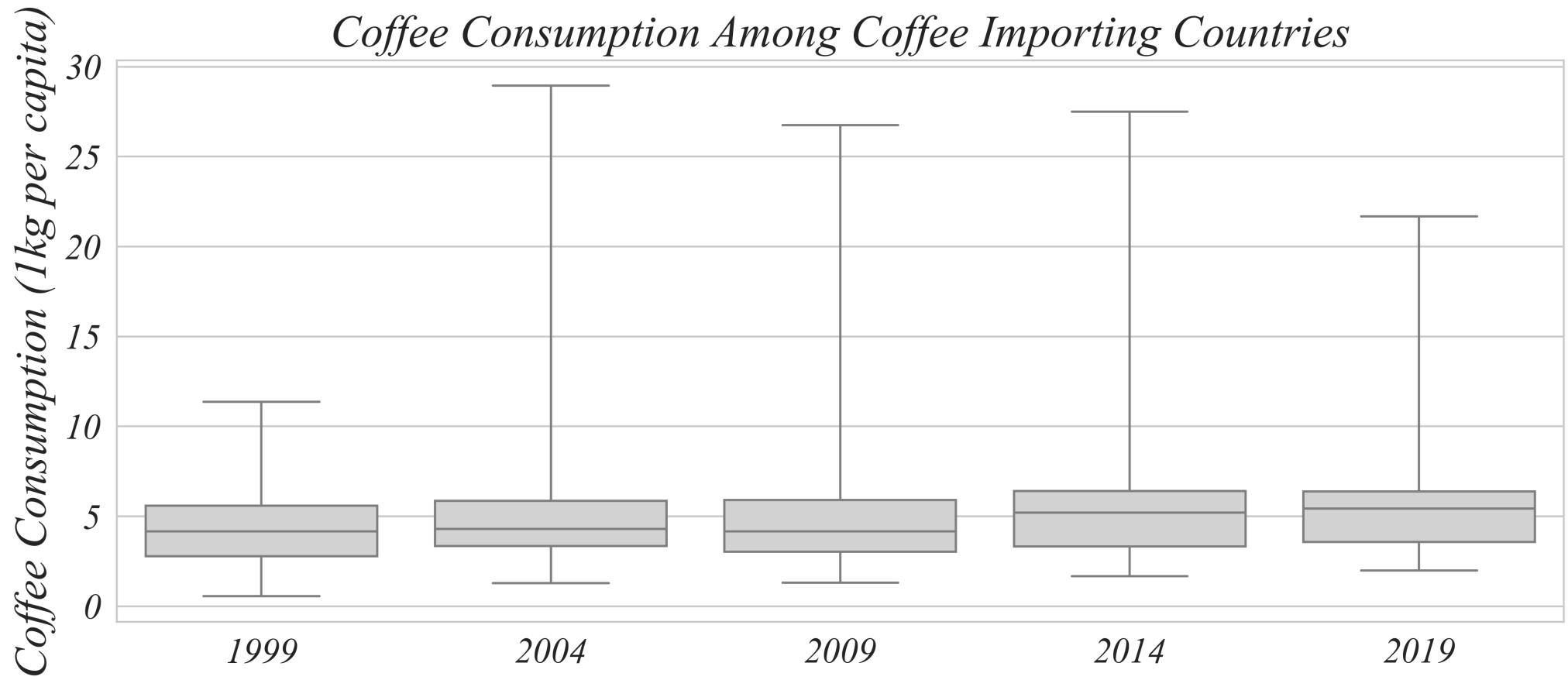
What patterns do we observe about the maximums?



> same with the maximums

Relationships Between Variables

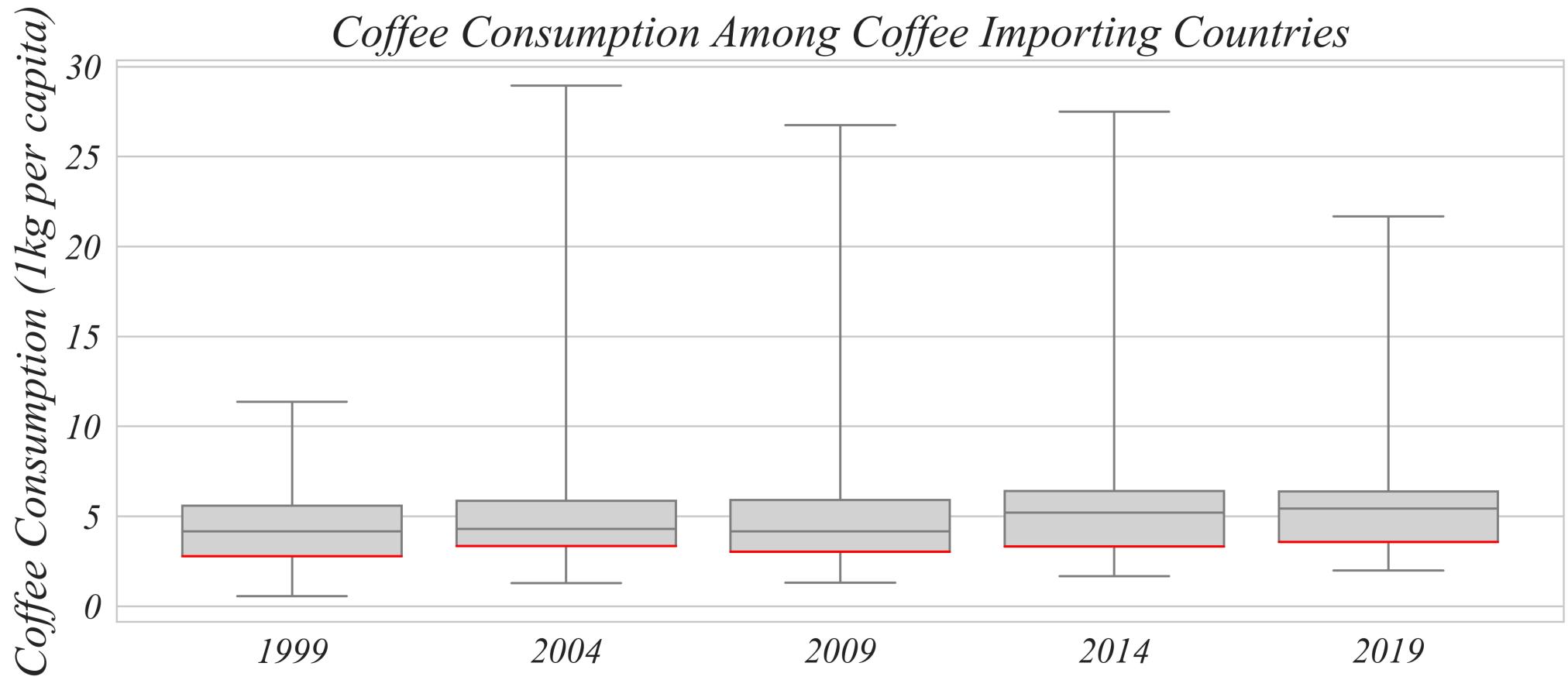
Which years did more than 25% consume less than 5 kg?



> look at the 25%

Relationships Between Variables

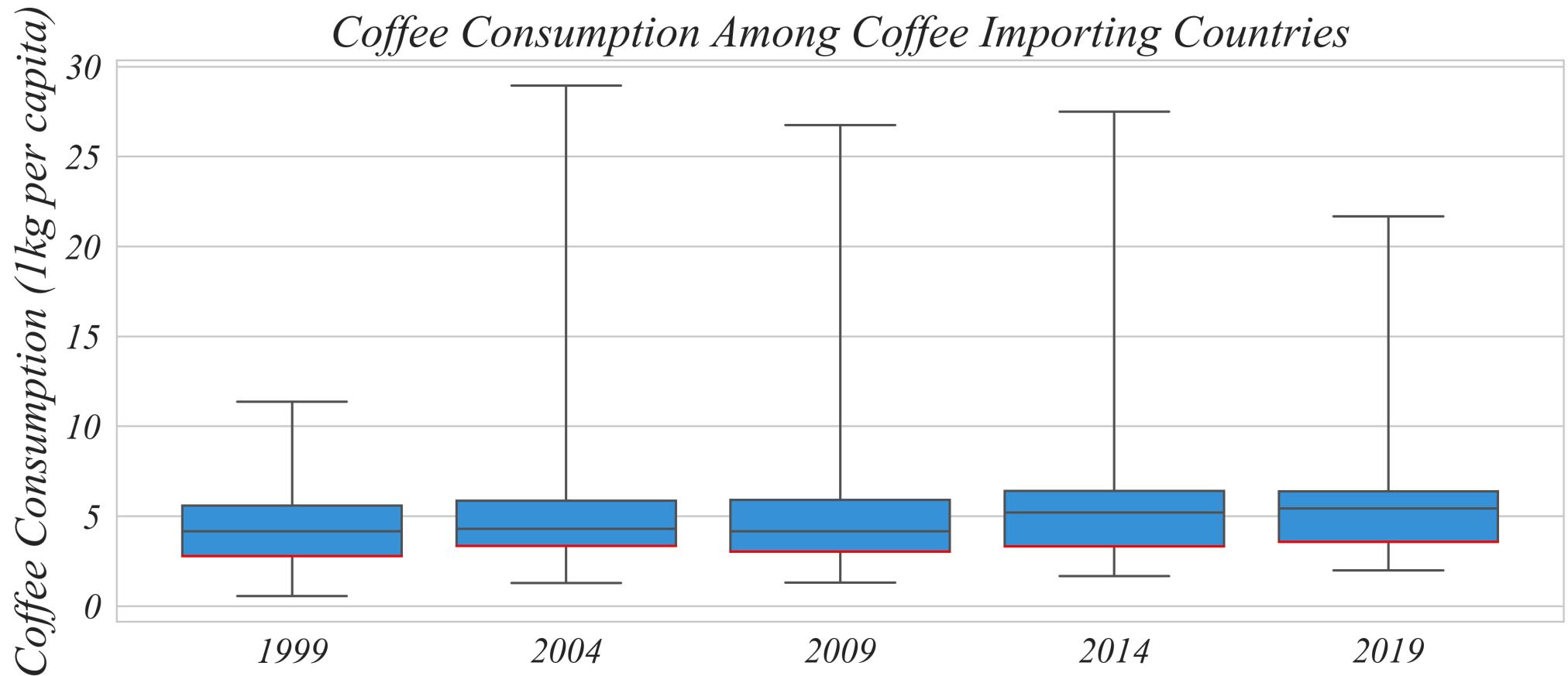
Which years did more than 25% consume less than 5 kg?



> look at the 25% and compare it to 5 kg per cap

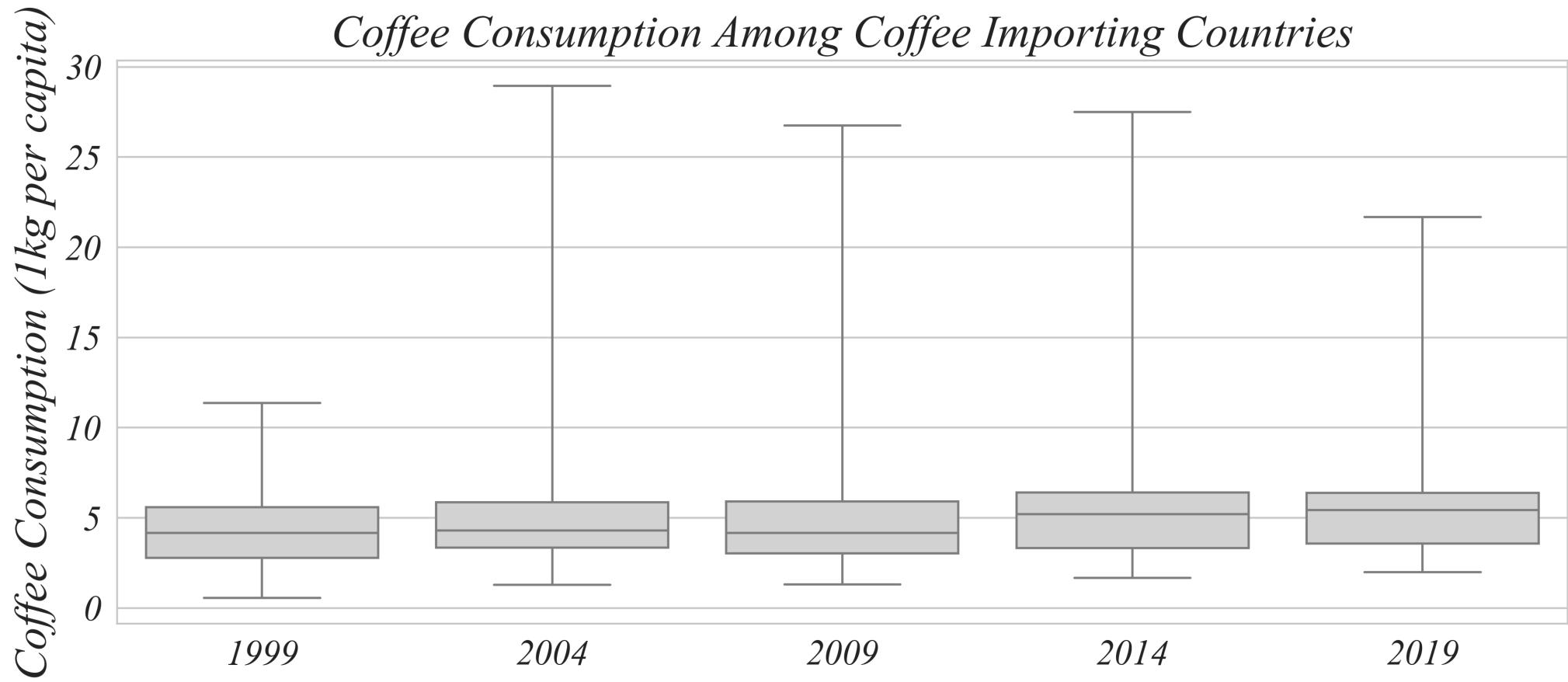
Relationships Between Variables

Which years did more than 25% consume less than 5 kg?



Relationships Between Variables

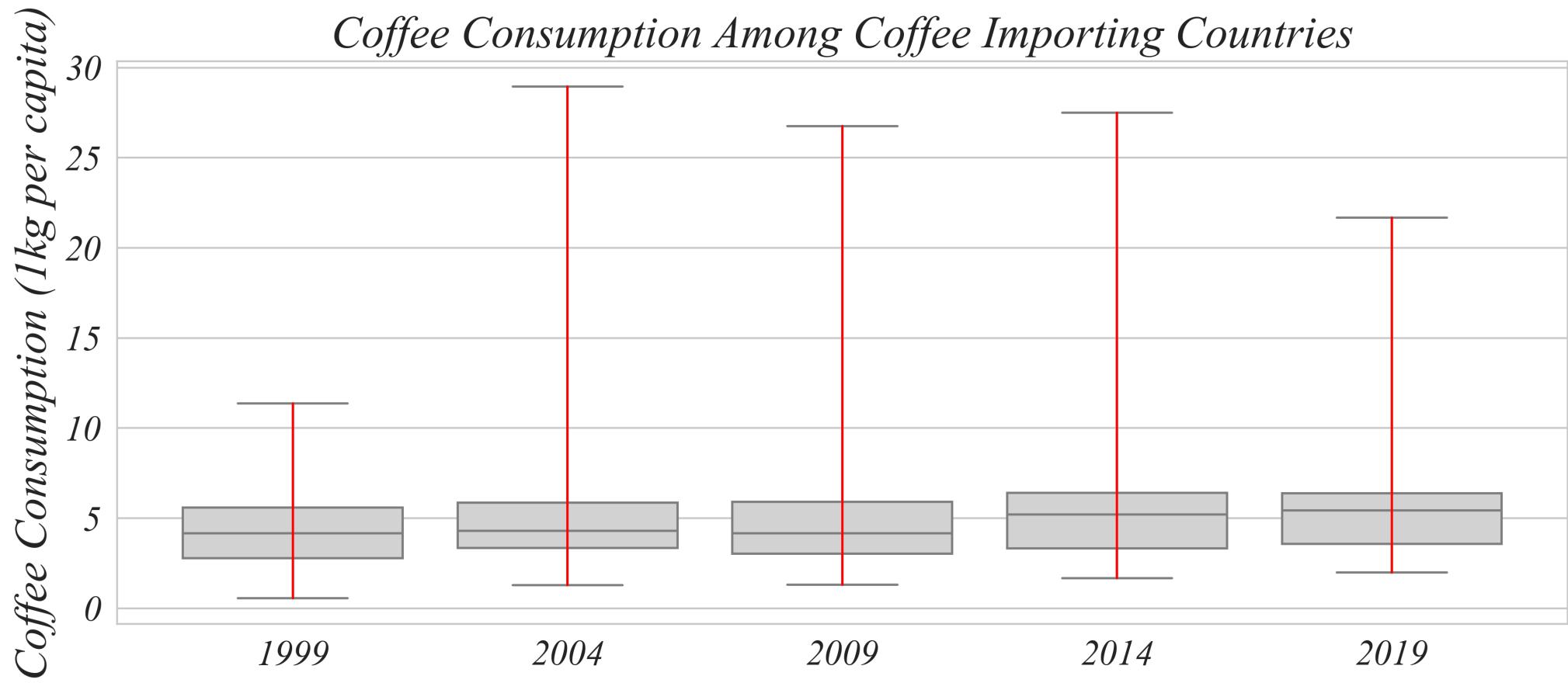
Which year saw the greatest difference between any two countries?



> look at the range

Relationships Between Variables

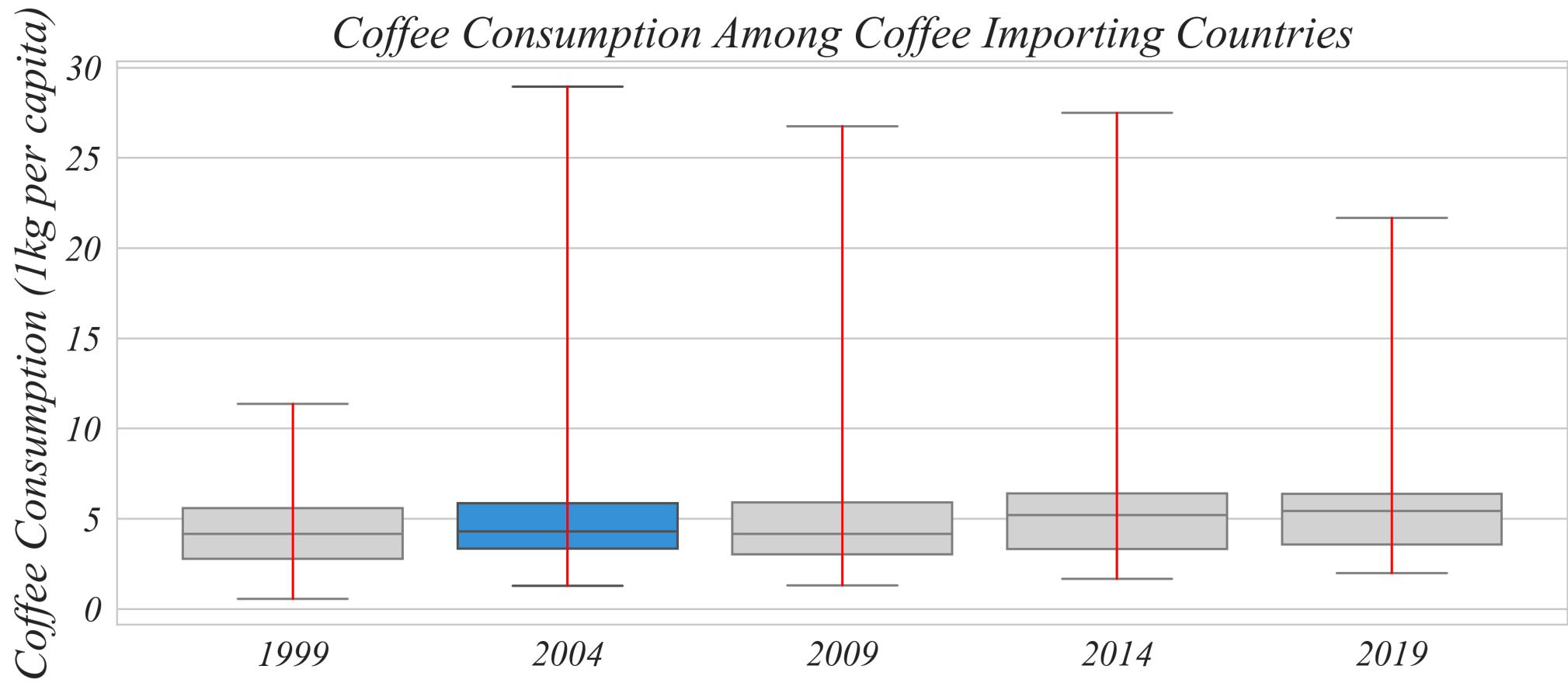
Which year saw the greatest difference between any two countries?



> look at the range

Relationships Between Variables

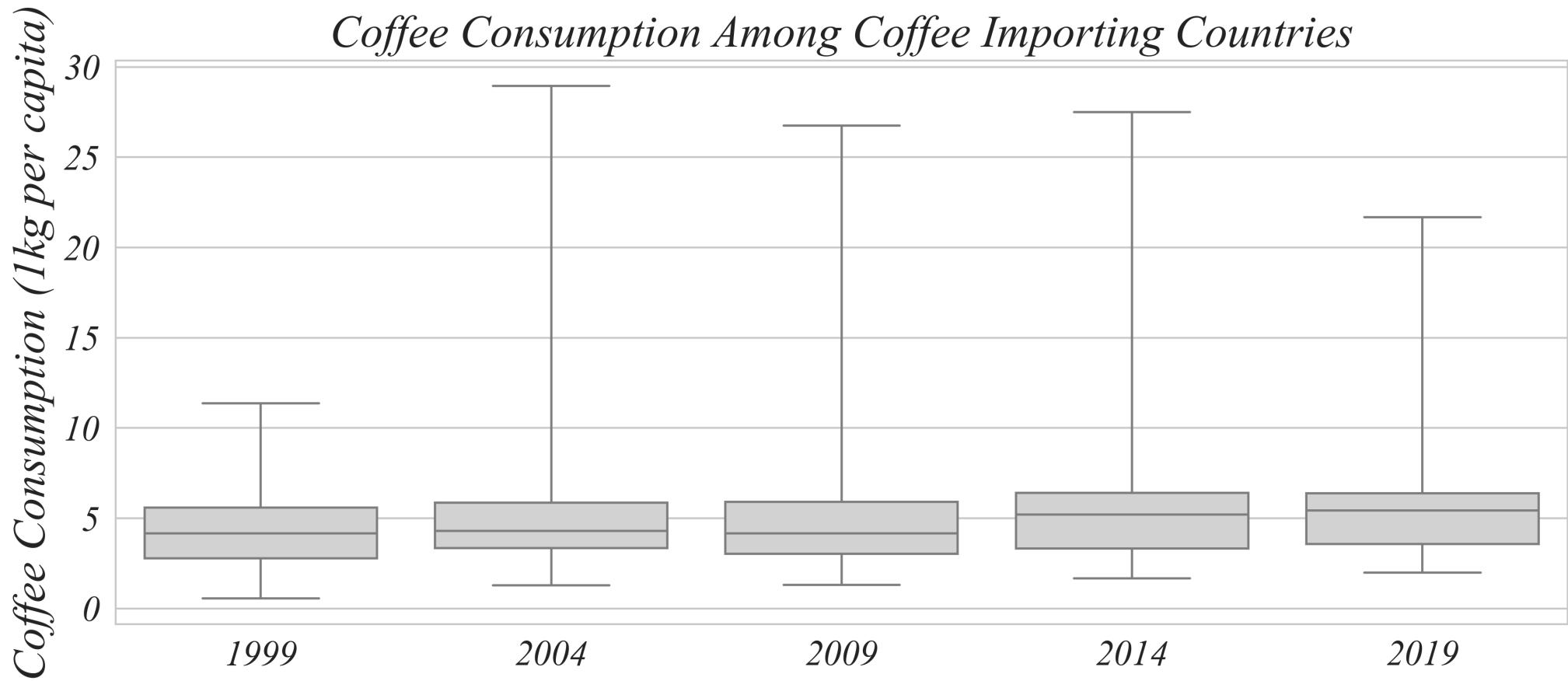
Which year saw the greatest difference between any two countries?



> look at the range and select the largest

Relationships Between Variables

In which year did all countries increase their coffee consumption?



> not visible in the figure!

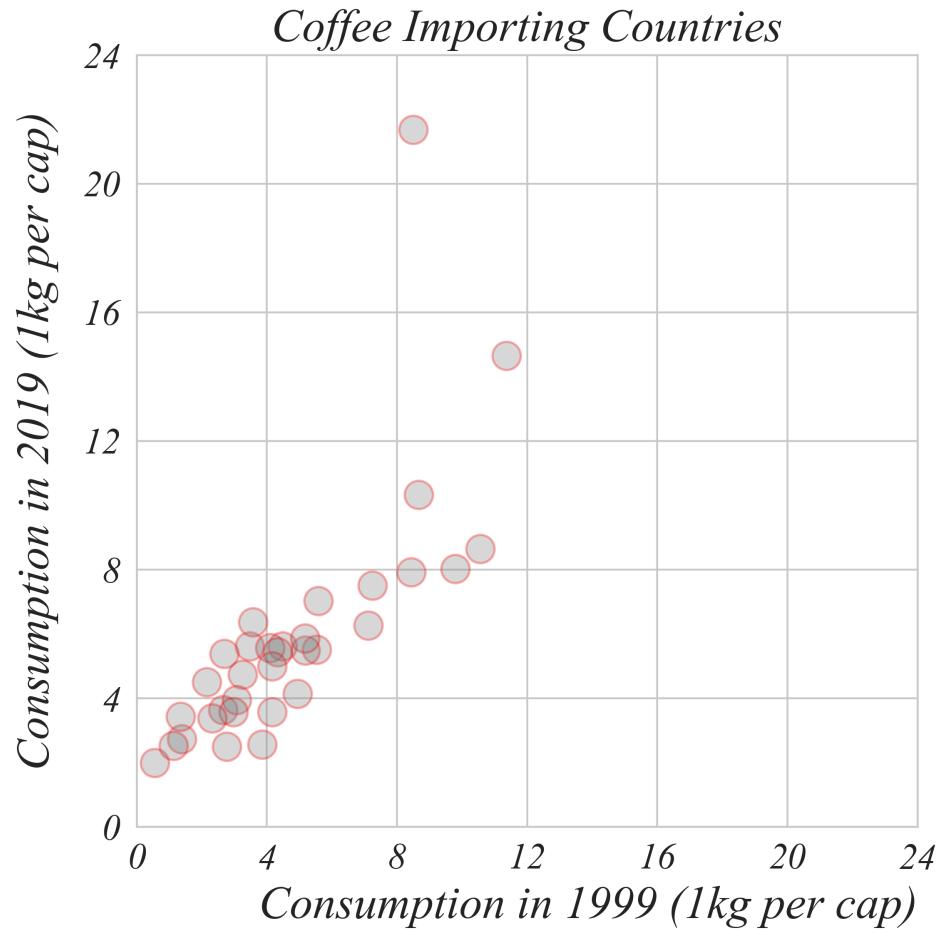
Exercise: Boxplots

We're going to use a set of boxplots to visually compare across years the distributions of coffee consumption per capital among coffee importing countries.

- Data: [Coffee_Per_Cap.csv](#)

Relationships Between Variables

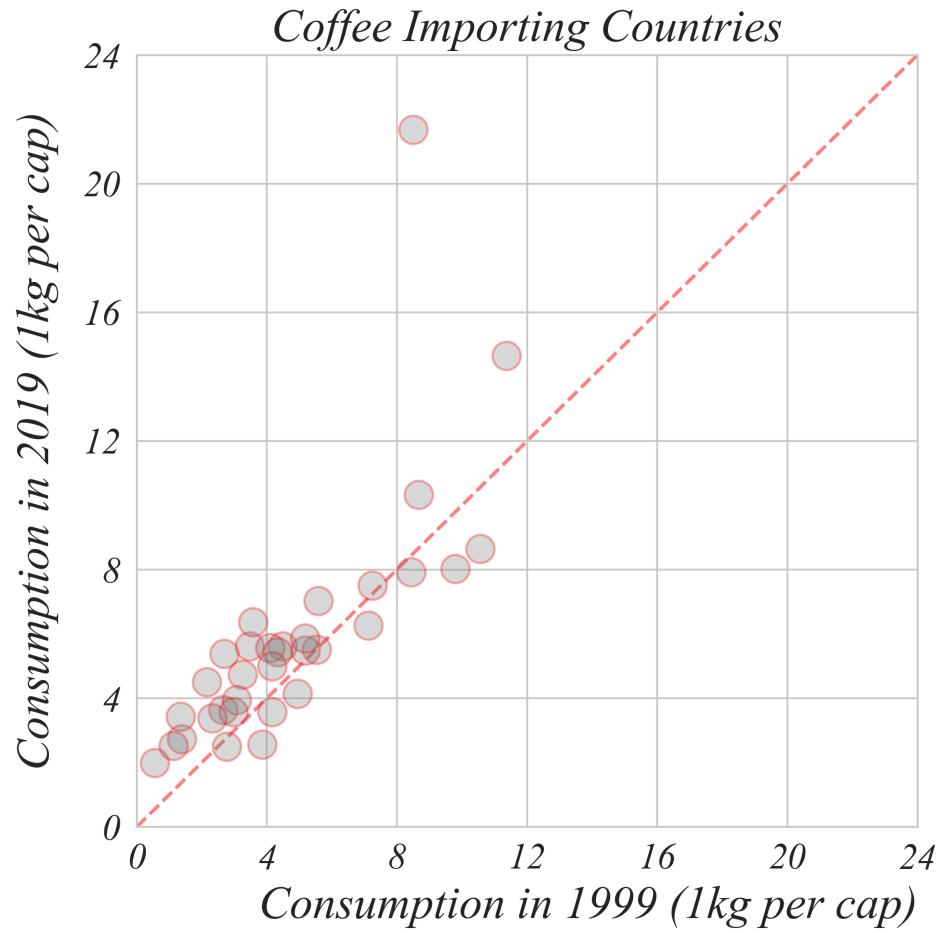
Did all countries increase their coffee consumption?



> a scatter plot can visualize changes between two points in time

Relationships Between Variables

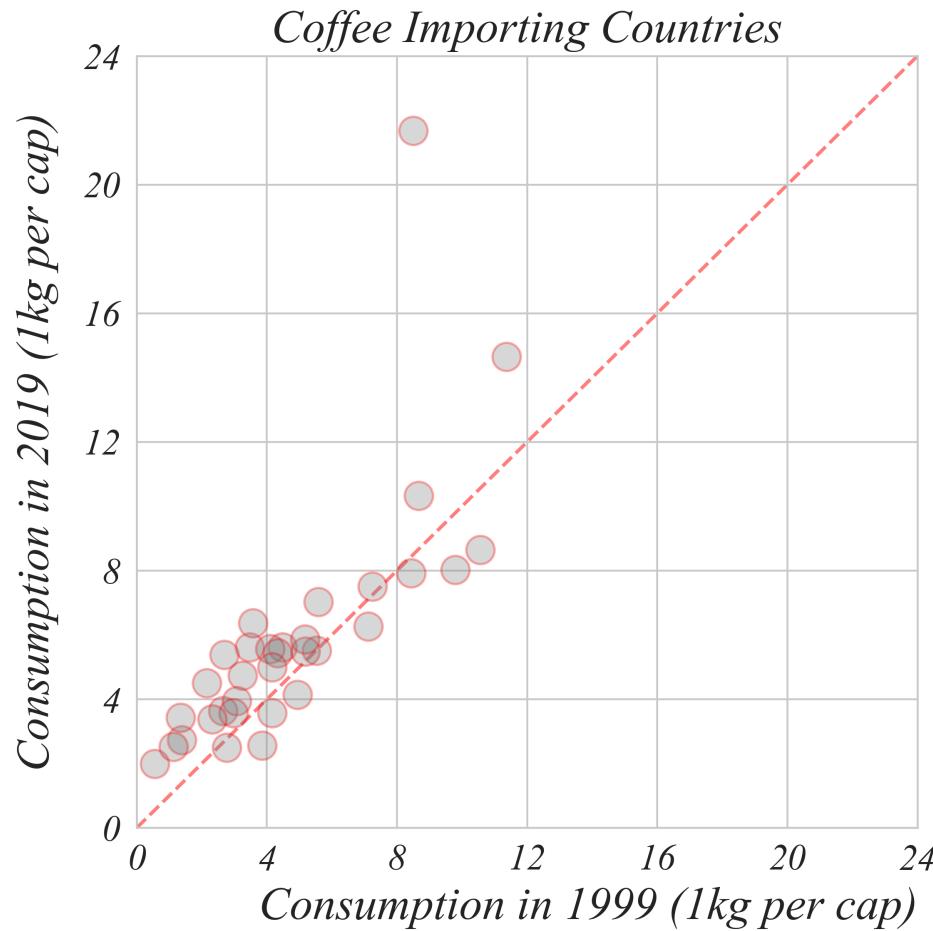
Did all countries increase their coffee consumption?



> a 45 degree line shows all the possible points with no change

Relationships Between Variables

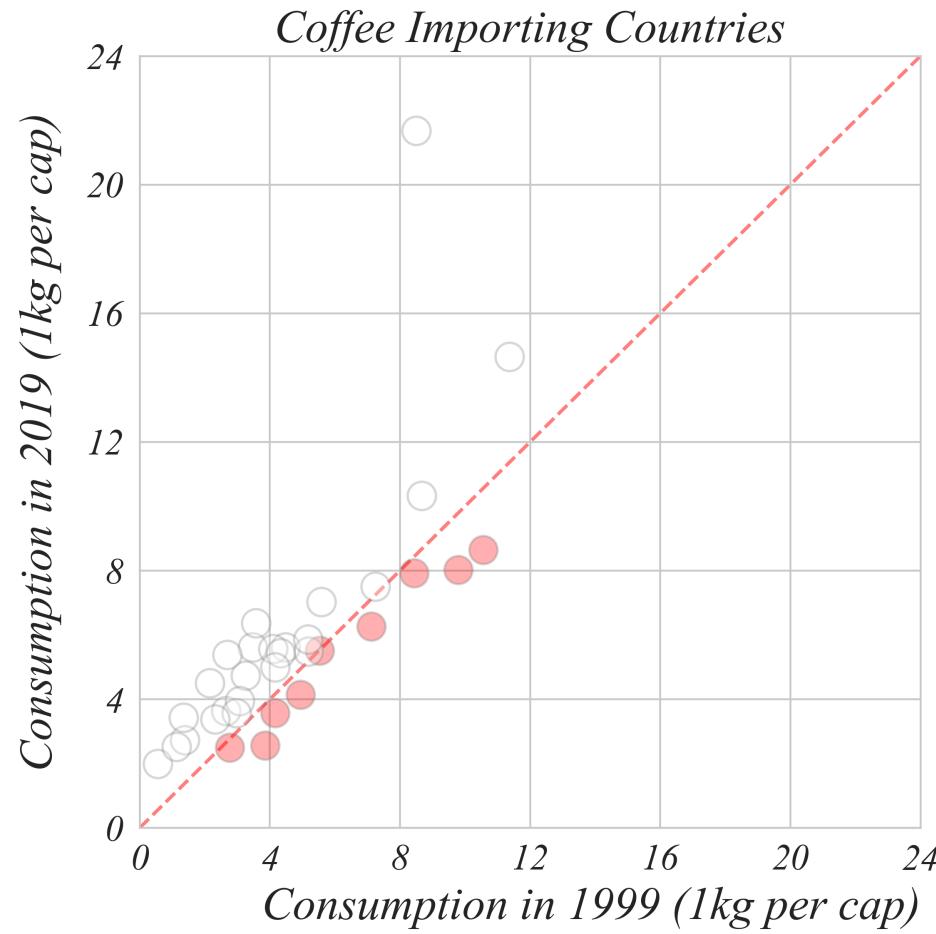
Which countries increased their coffee consumption?



> a 45 degree line shows all the possible points with no change

Relationships Between Variables

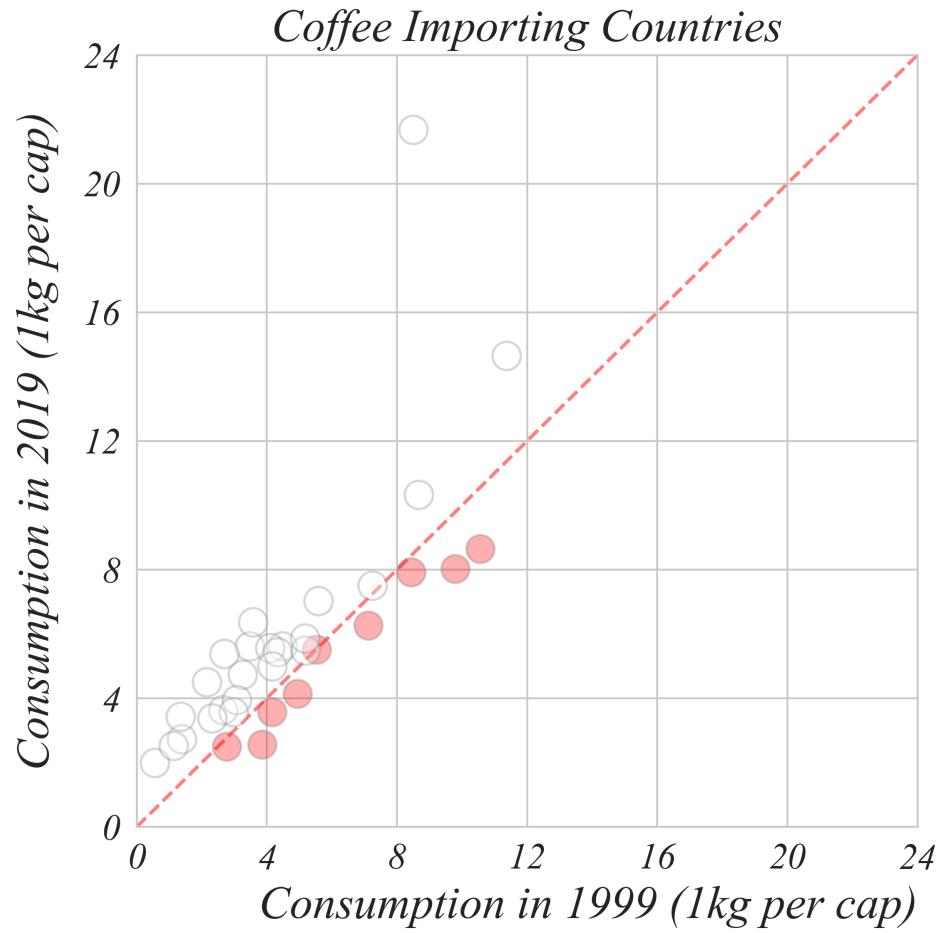
Which countries decreased their coffee consumption?



> points below the line show decreases

Relationships Between Variables

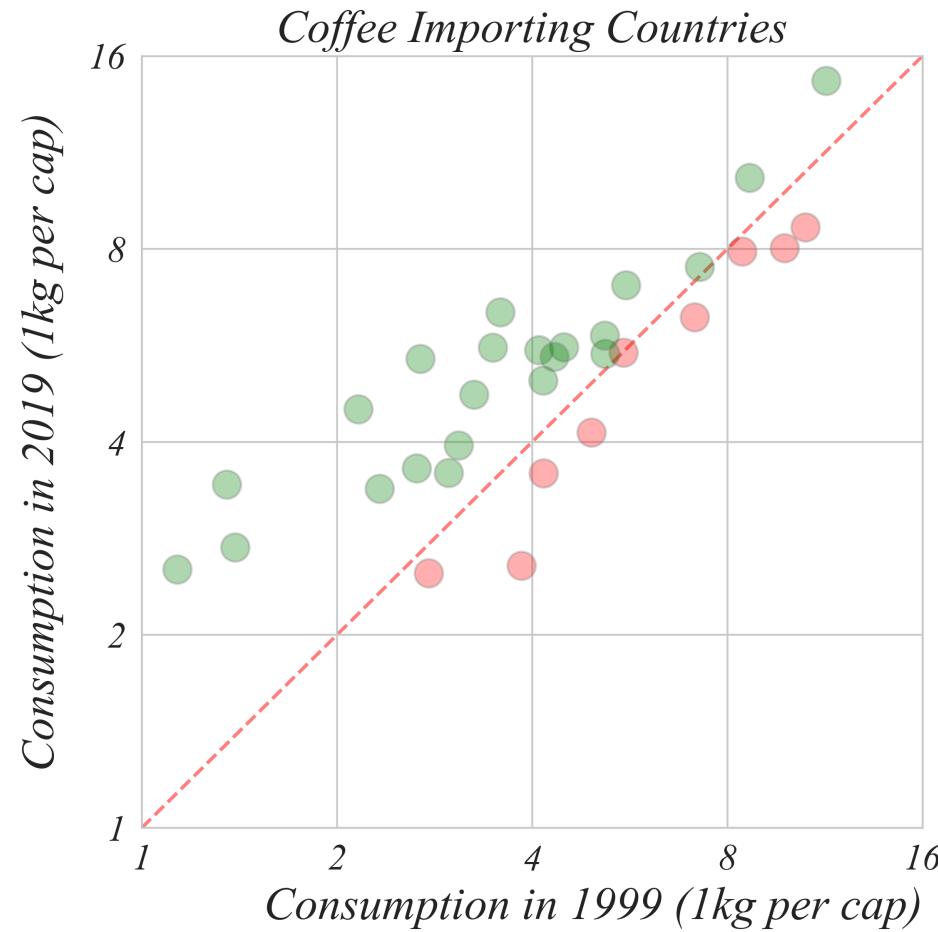
Which countries decreased their coffee consumption?



> this might also benefit from a log scale

Relationships Between Variables

Does the data confirm that the world is drinking more coffee?



> colors can visualize both increases and decreases