ECON 0150 | Economic Data Analysis

The economist's data analysis skillset.

Part 1.5 | Panel Data (Wide Format)

Recap Part 1.4 | Long Format Panel Data

Each row represents an observation of an entity at one point in time

Code	Year	Consumption
AUT	1990	10.47
AUT	1991	10.07
AUT	1992	9.27
AUT	1993	10.13
AUT	1994	8.21
…	…	…

> this is data on coffee consumption per capita, 34 countries, 1990 to 2019

Question

Is the world drinking more coffee in 2019 than in 1999?

Code	Year	Consumption
AUT	1990	10.47
AUT	1991	10.07
AUT	1992	9.27
AUT	1993	10.13
AUT	1994	8.21
…	…	…

> *how can we use this data to answer our question?*

> *it's challenging... lets instead use one row per* ***Entitiy***

Wide Format Panel Data

A row for each entity across many column time periods

Code	1999	2004	2009	2014	2019
AUT	8.43	7.31	6.37	7.97	7.93
BGR	2.65	2.83	3.30	3.12	3.64
HRV	4.48	5.16	5.10	5.21	5.62
CYP	3.48	3.53	4.05	4.13	5.62
CZE	3.26	3.56	3.02	5.69	4.74
…	…	…	…	…	…

> now comparing 1999 vs 2019 is just comparing two columns!

Wide Format

Each year is a column

Wide Format: Coffee Consumption (kg per capita)

Code	1999	2004	2009	2014	2019
FRA	5.5	4.7	5.3	5.4	5.5
DEU	7.1	7.6	6.5	6.4	6.3
JPN	3.0	3.3	3.3	3.5	3.6
GBR	2.3	2.5	3.1	2.7	3.4
USA	4.2	4.3	4.2	4.5	5.0

Long Format

Each observation is a row

Long Format: Coffee Consumption

Code	Year	Consumption
FRA	1999	5.5
DEU	1999	7.1
JPN	1999	3.0
GBR	1999	2.3
USA	1999	4.2
FRA	2004	4.7
DEU	2004	7.6
JPN	2004	3.3
GBR	2004	2.5
USA	2004	4.3
FRA	2009	5.3
DEU	2009	6.5

Two Formats, Same Data

Panel data can be stored in two ways

- ***Wide format:*** *Each time period is a separate column*
- ***Long format:*** *Each observation is a separate row*
- *Same information, different shapes*
- *Different shapes make different tasks easier*

Panel Data: Coffee Consumption Per Capita

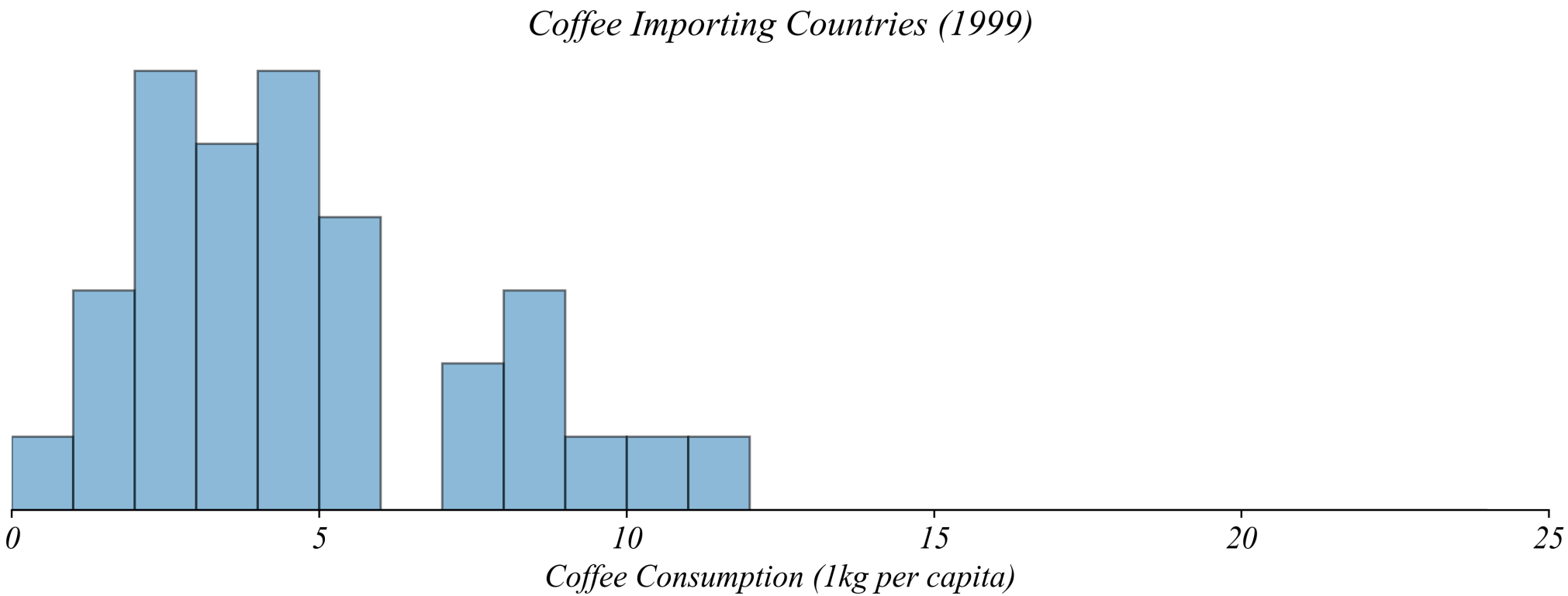
Is the world drinking more coffee?

Lets examine whether the world is drinking more coffee today than in the 1990s.

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

Panel Data: Coffee Consumption Per Capita

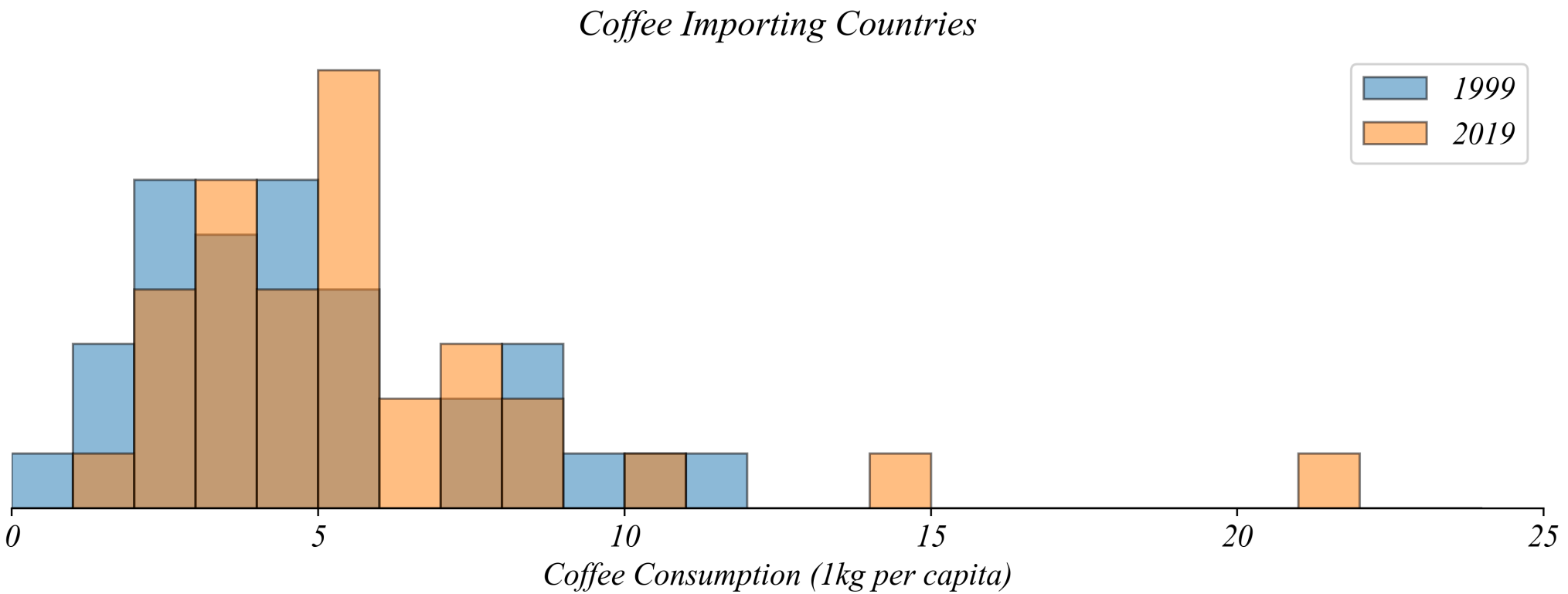
Is the world drinking more coffee?



> compared to what…?

Panel Data: Coffee Consumption Per Capita

Is the world drinking more coffee?

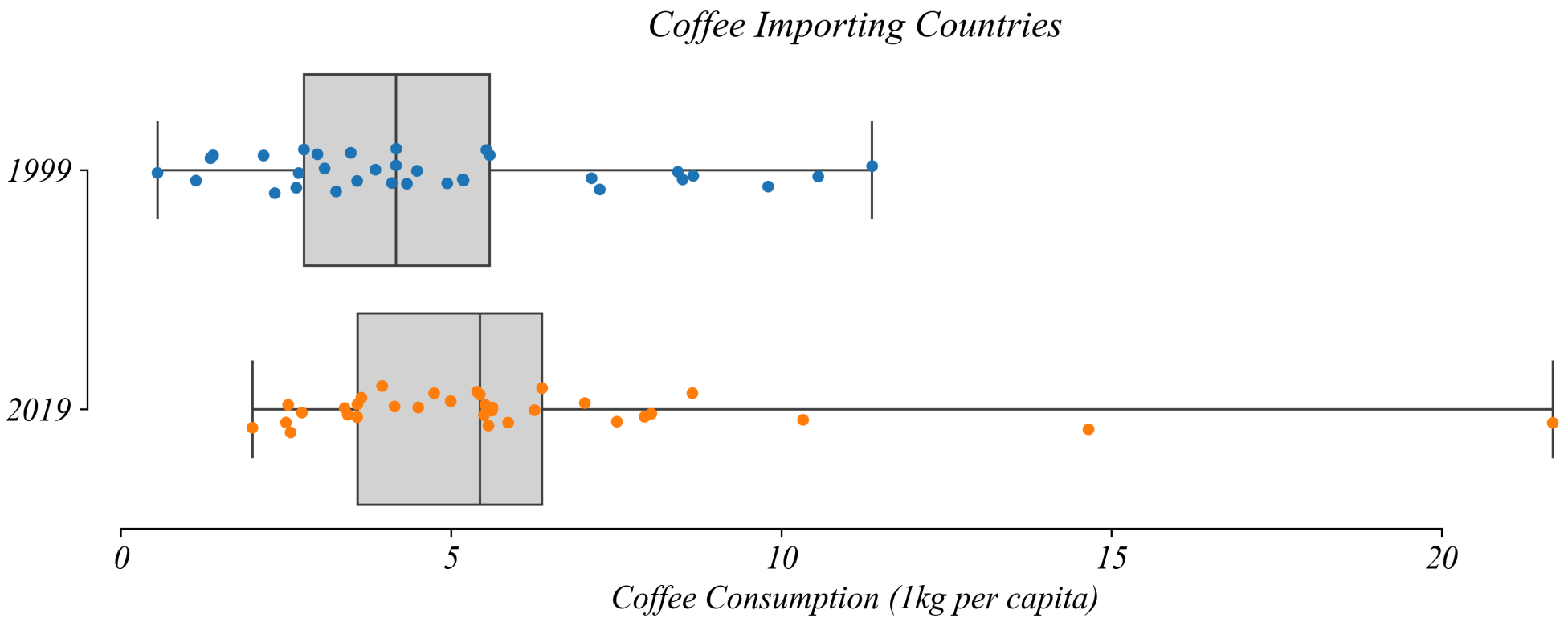


> *this is still pretty unclear: histograms aren't great for comparison*

> *lets use a multi-boxplot*

Panel Data: Multi-Boxplots

Is the world drinking more coffee?

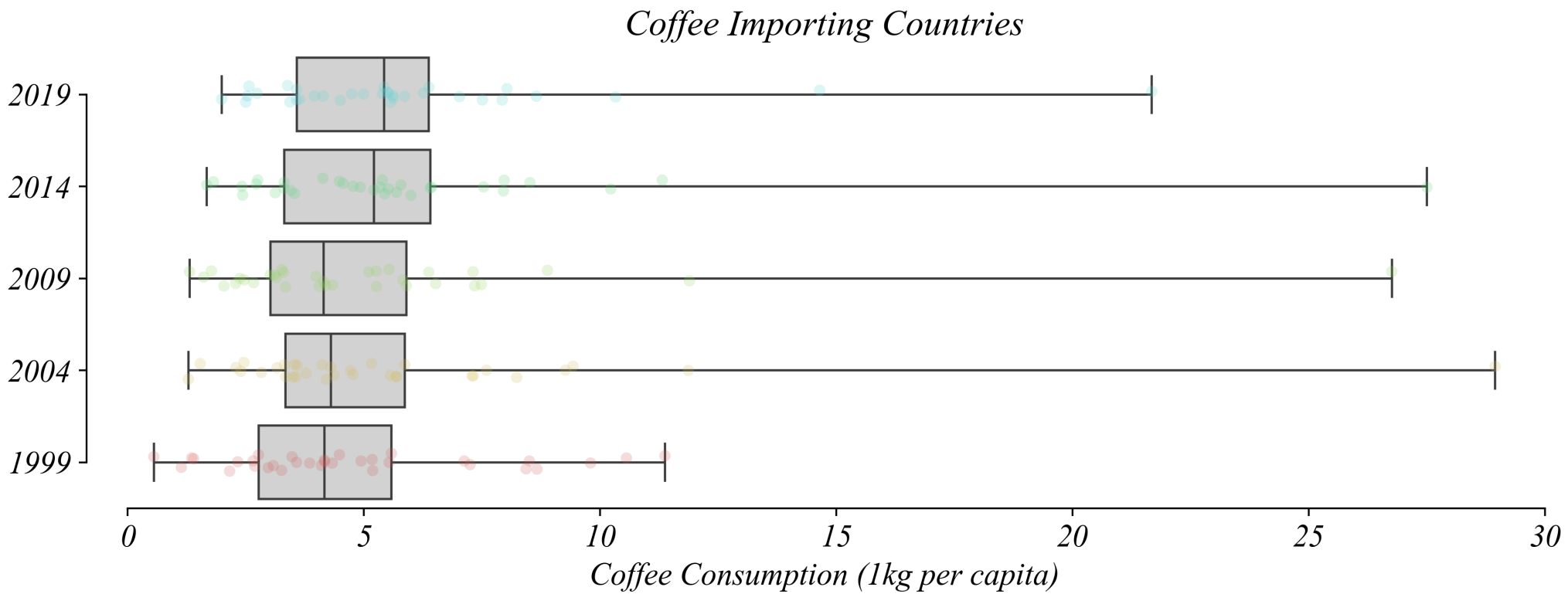


> *this is better: it looks like the distribution is shifted higher!*

> *lets examine the years in between to see how the distribution evolved*

Panel Data: Multi-Boxplots

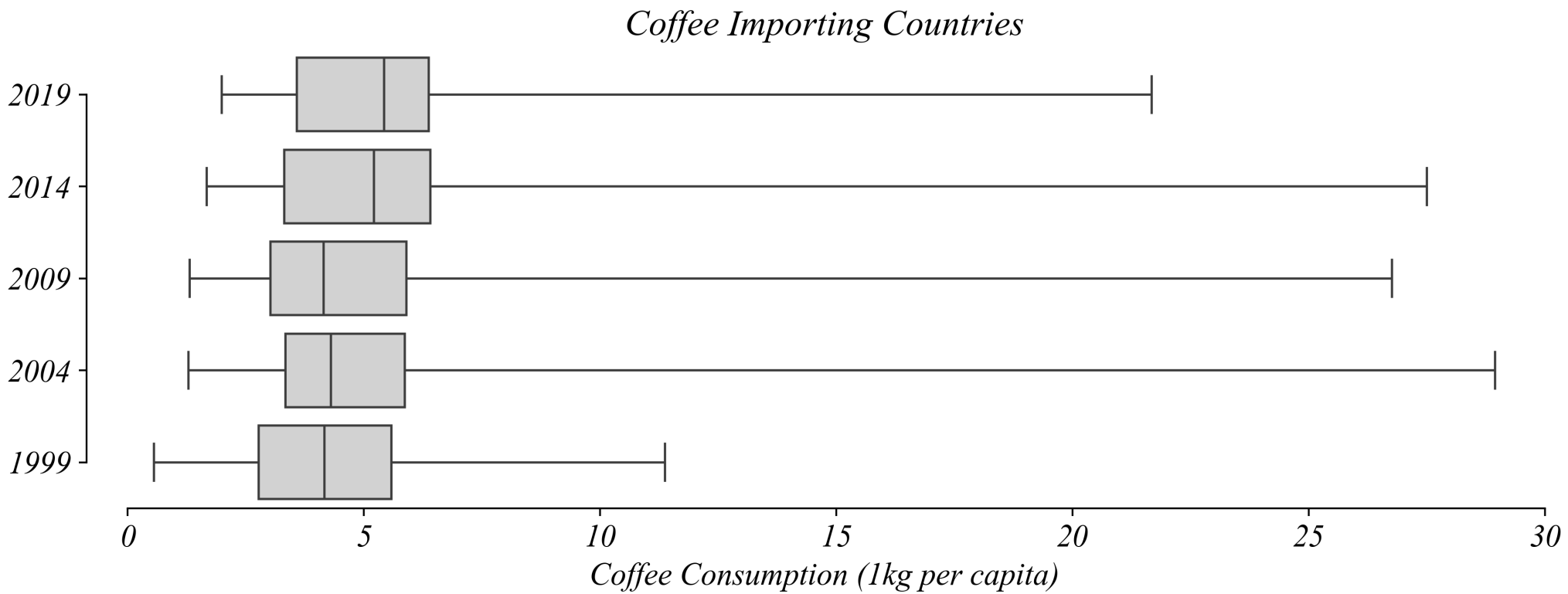
Is the world drinking more coffee?



> lets ask some smaller more focussed questions

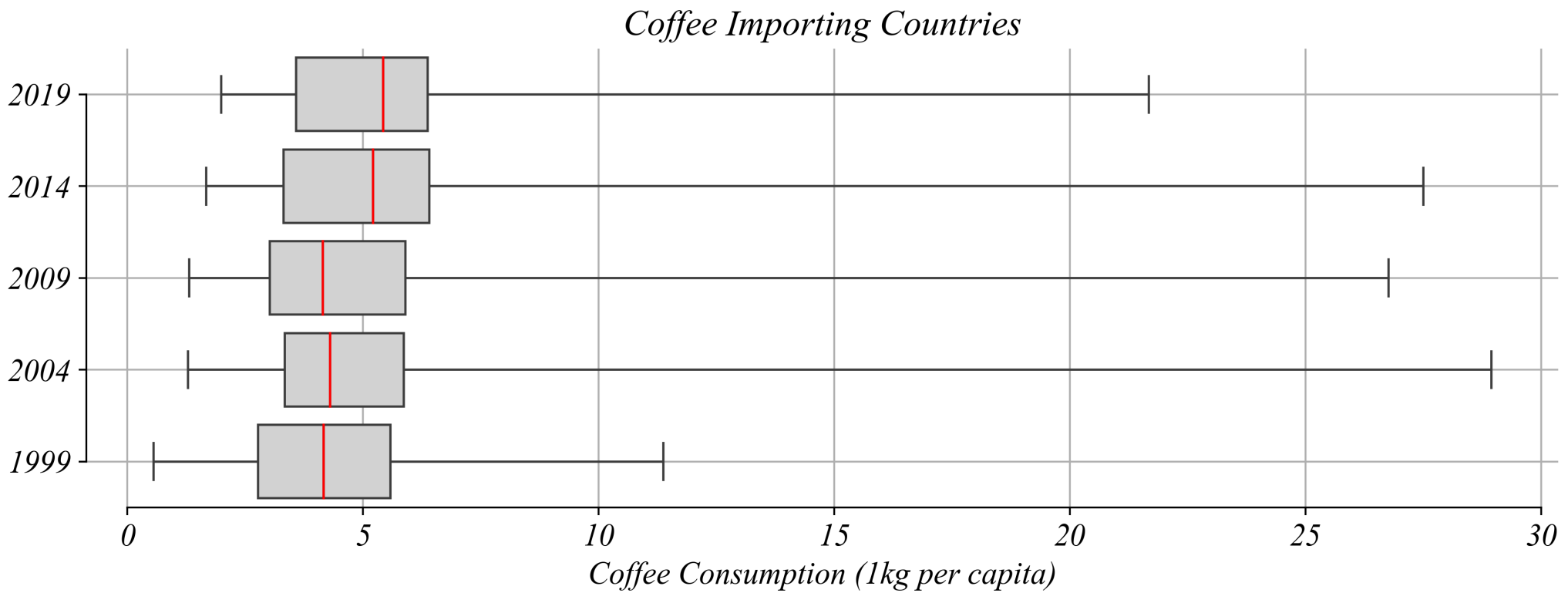
Panel Data: Multi-Boxplots

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



Panel Data: Multi-Boxplots

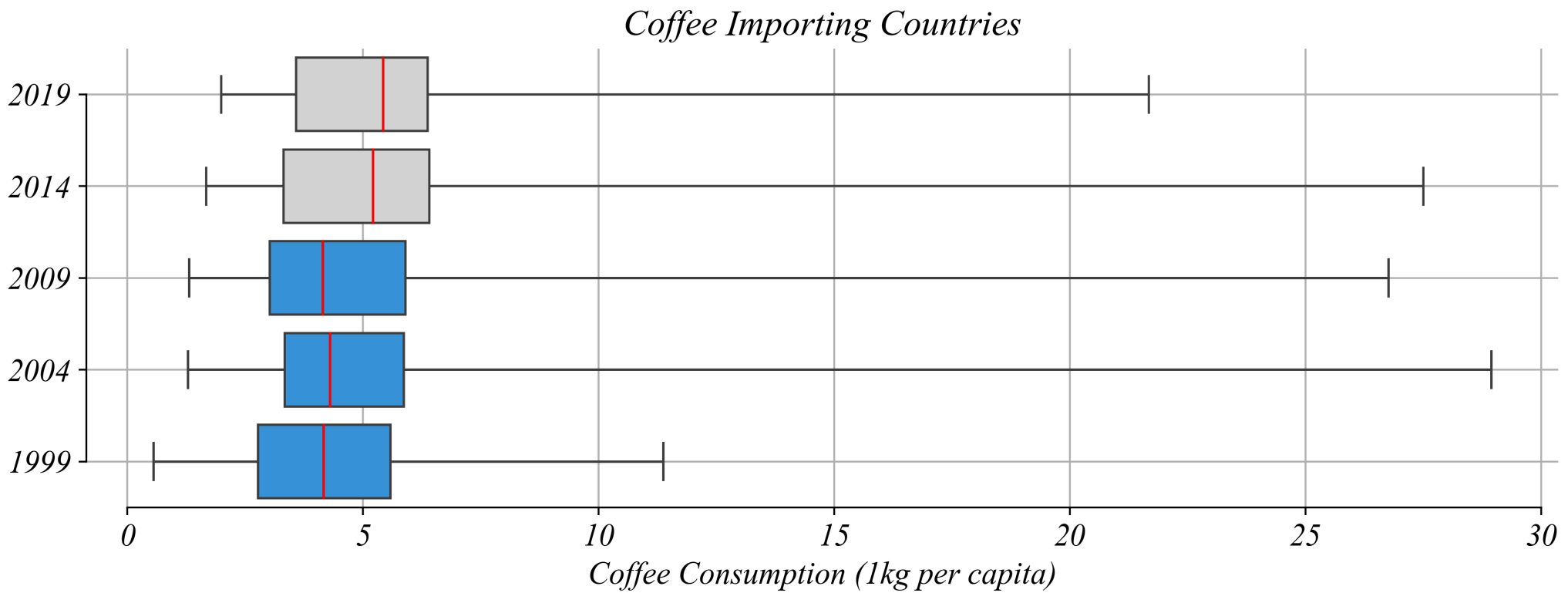
Which years show at least half consuming less than 5 kg per cap?



> *focus on the medians*

Panel Data: Multi-Boxplots

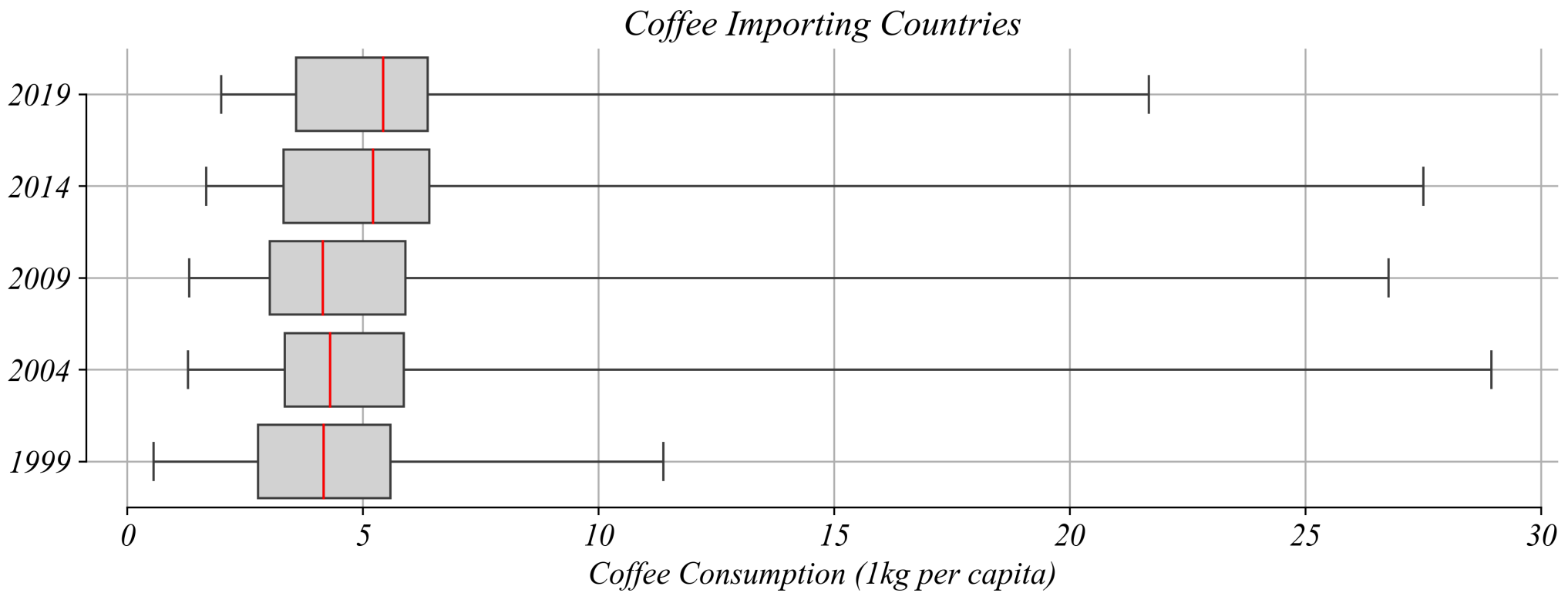
Which years show at least half consuming less than 5 kg per cap?



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

Panel Data: Multi-Boxplots

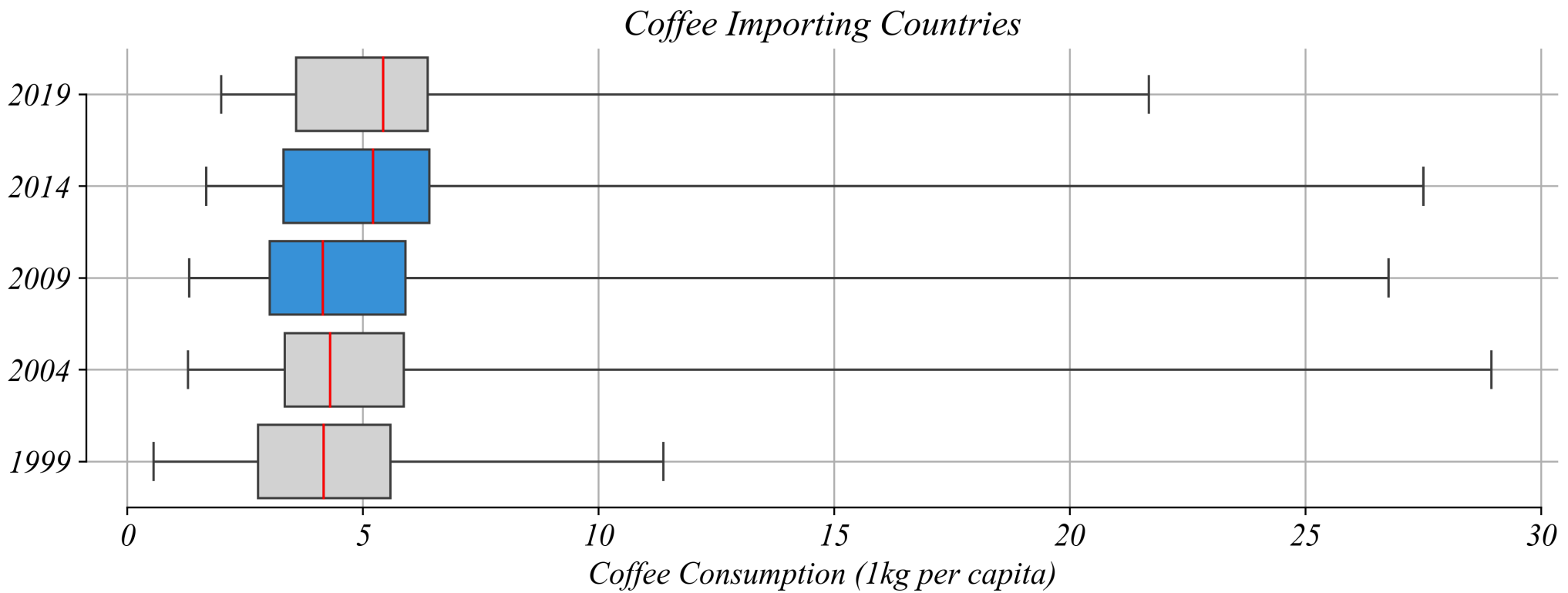
Which years saw the largest jump in the median?



> *... a little difficult to see*

Panel Data: Multi-Boxplots

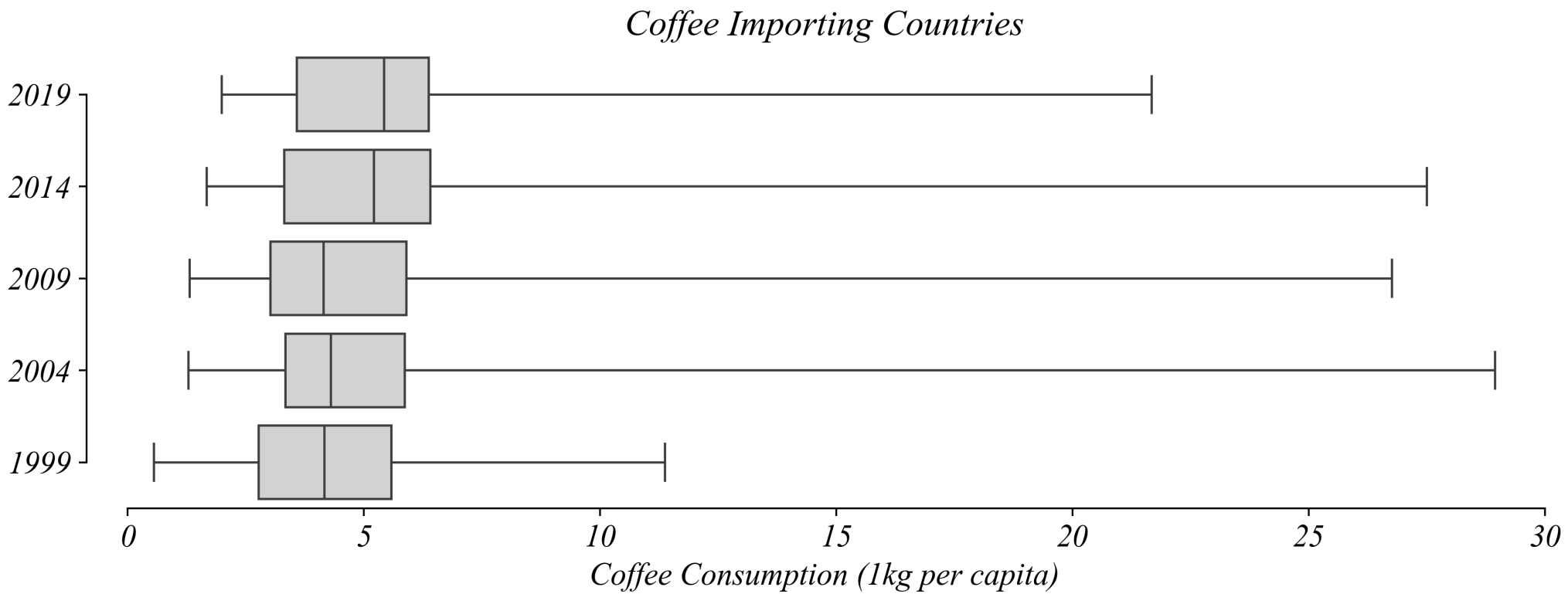
Which years saw the largest jump in the median?



> *... a little difficult to see*

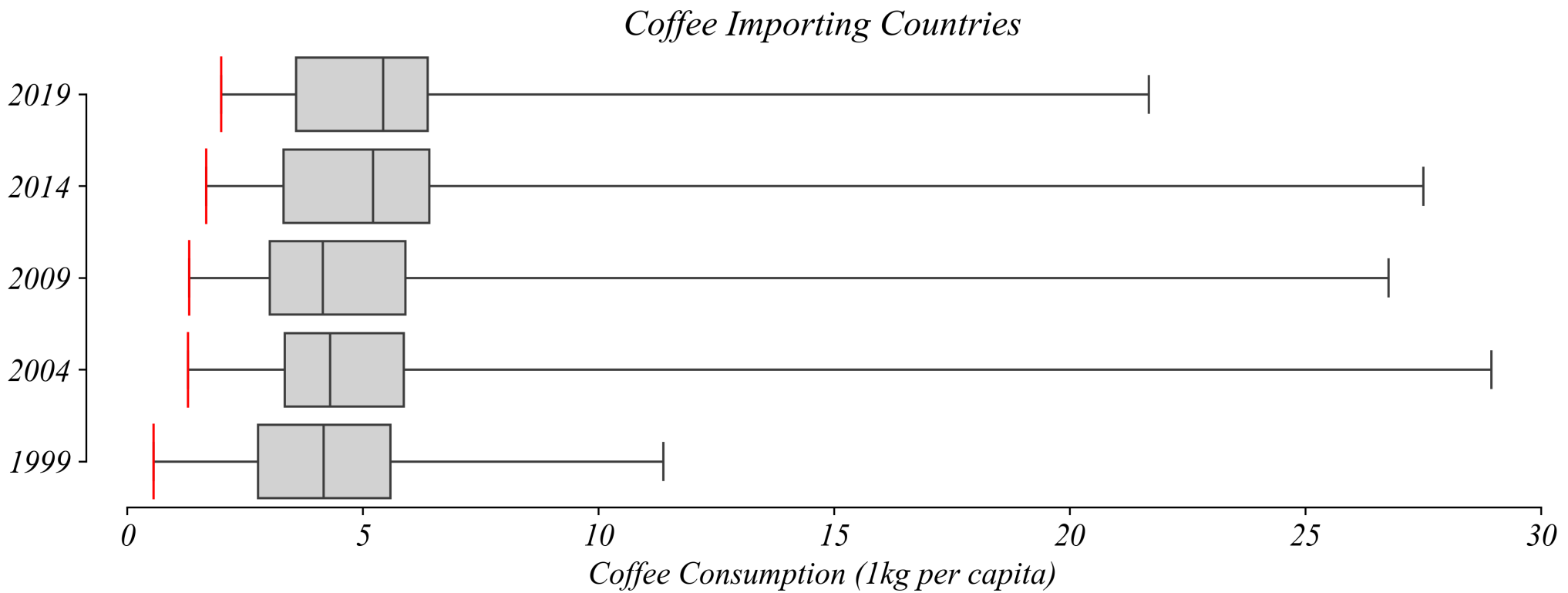
Panel Data: Multi-Boxplots

Is the country with the lowest consumption consuming more today?



Panel Data: Multi-Boxplots

Is the country with the lowest consumption consuming more today?

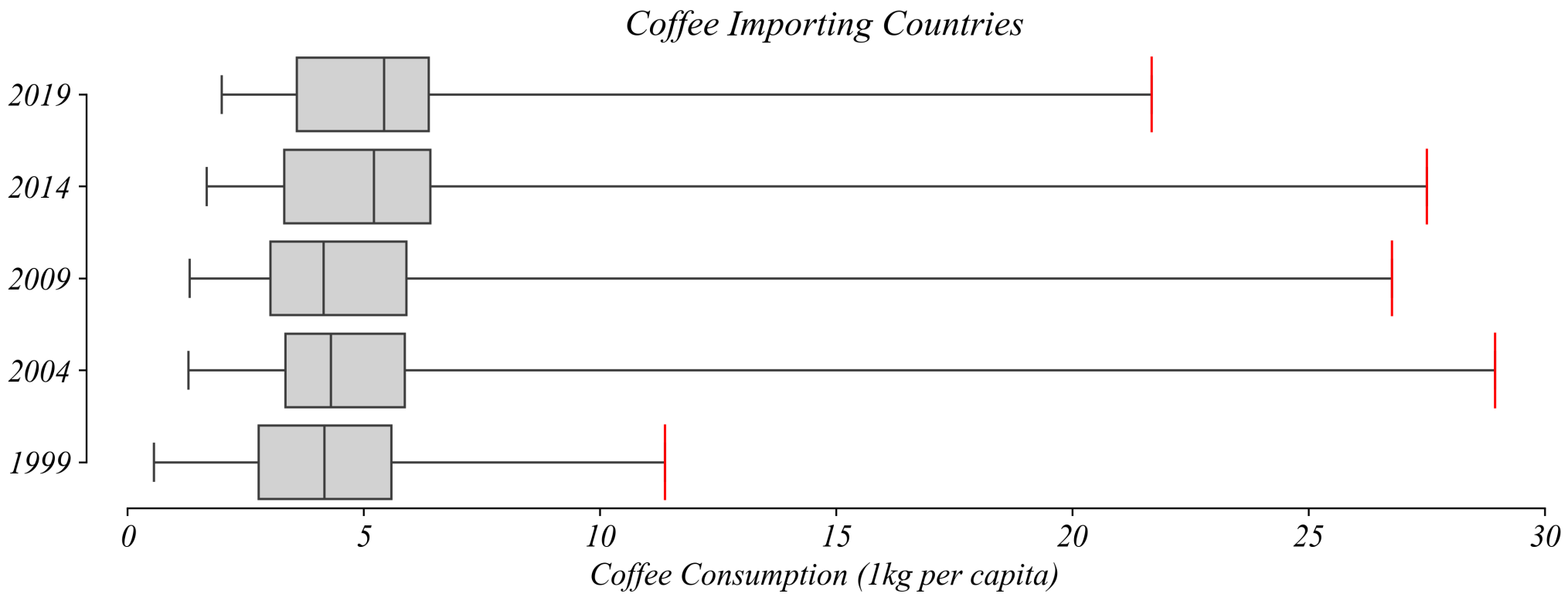


> *focus on the minimums*

> *yes!*

Panel Data: Multi-Boxplots

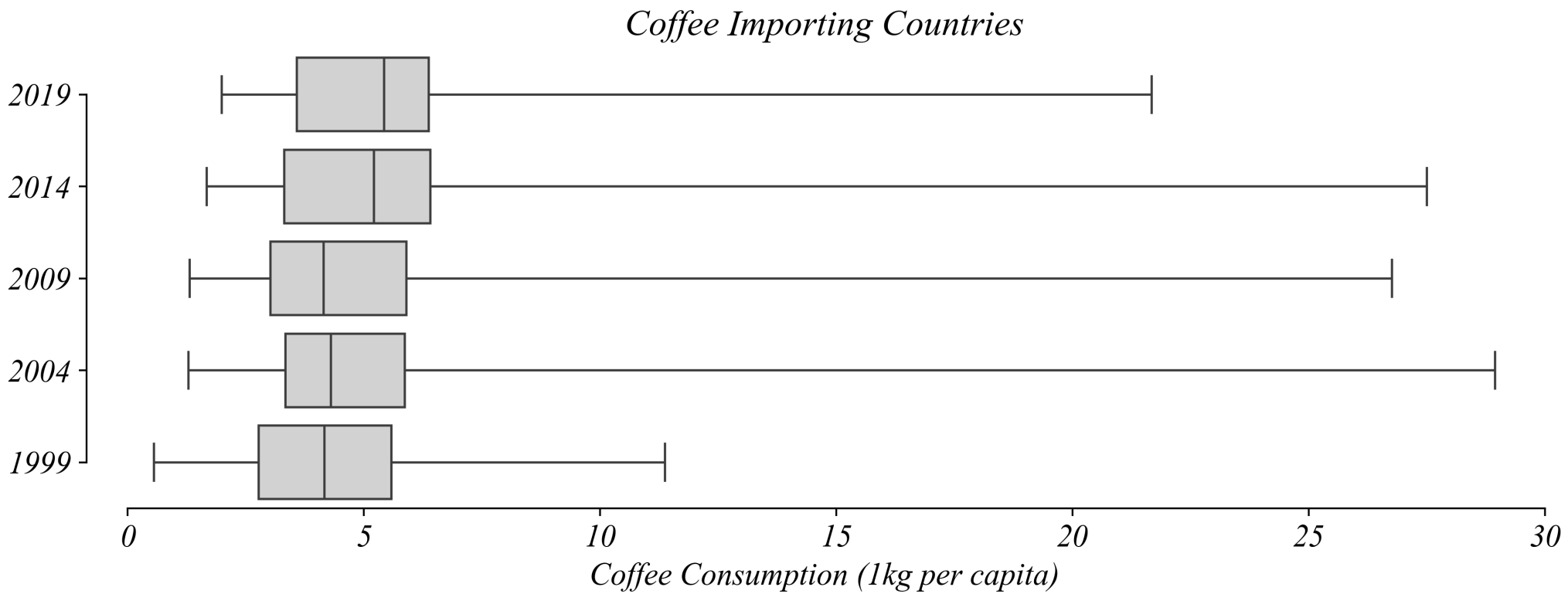
What patterns do we observe about the maximums?



> *same with the maximums*

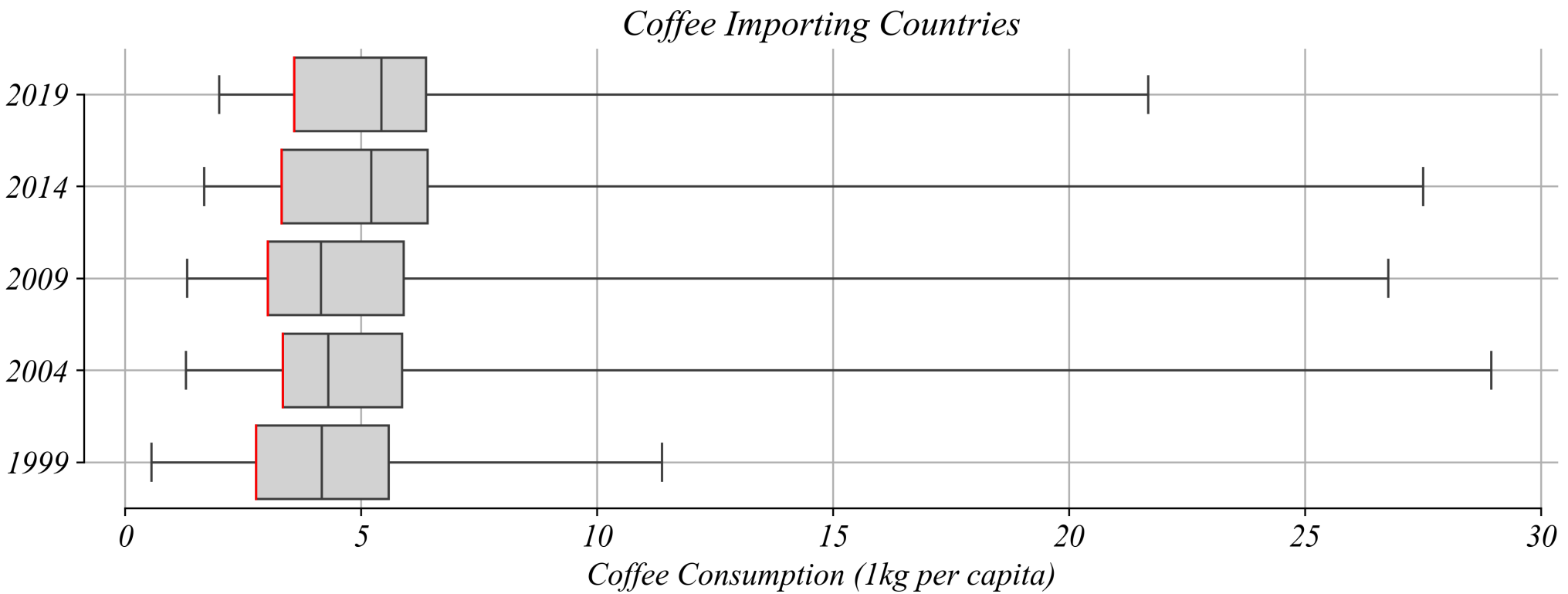
Panel Data: Multi-Boxplots

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



Panel Data: Multi-Boxplots

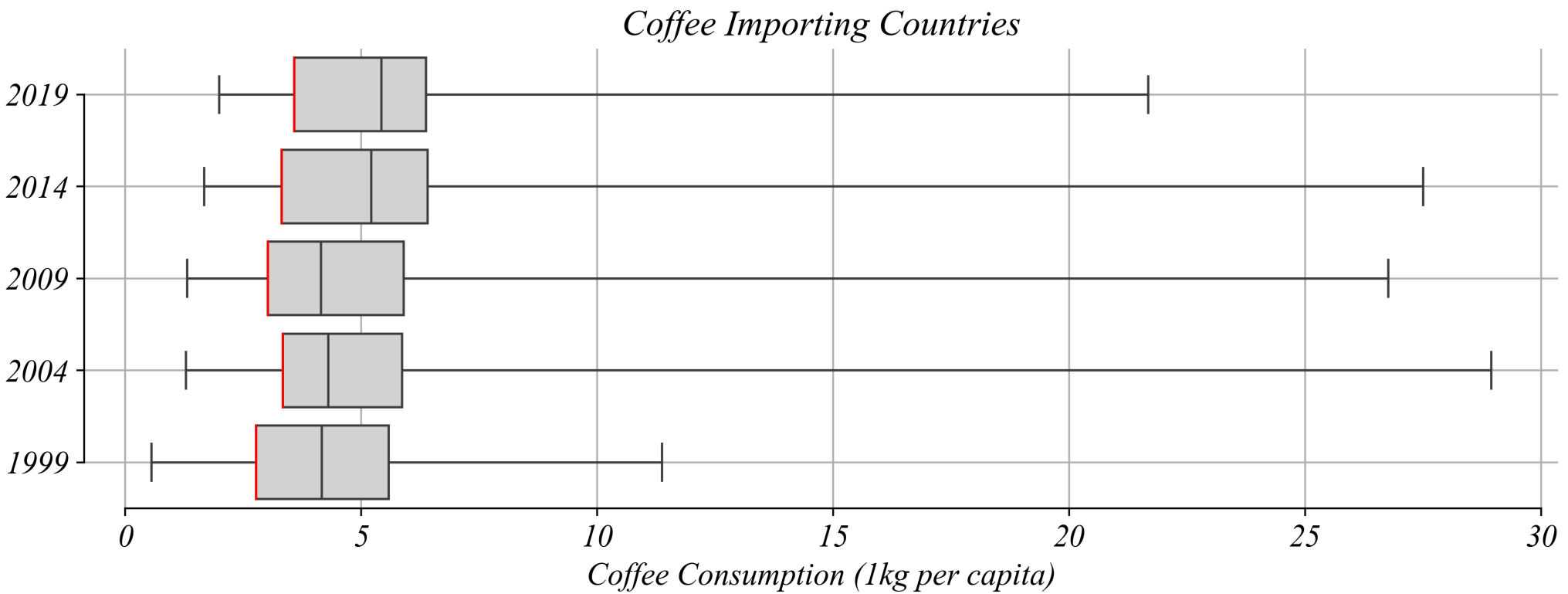
Which years did more than 25% consume less than 5 kg?



> *look at the 25%*

Panel Data: Multi-Boxplots

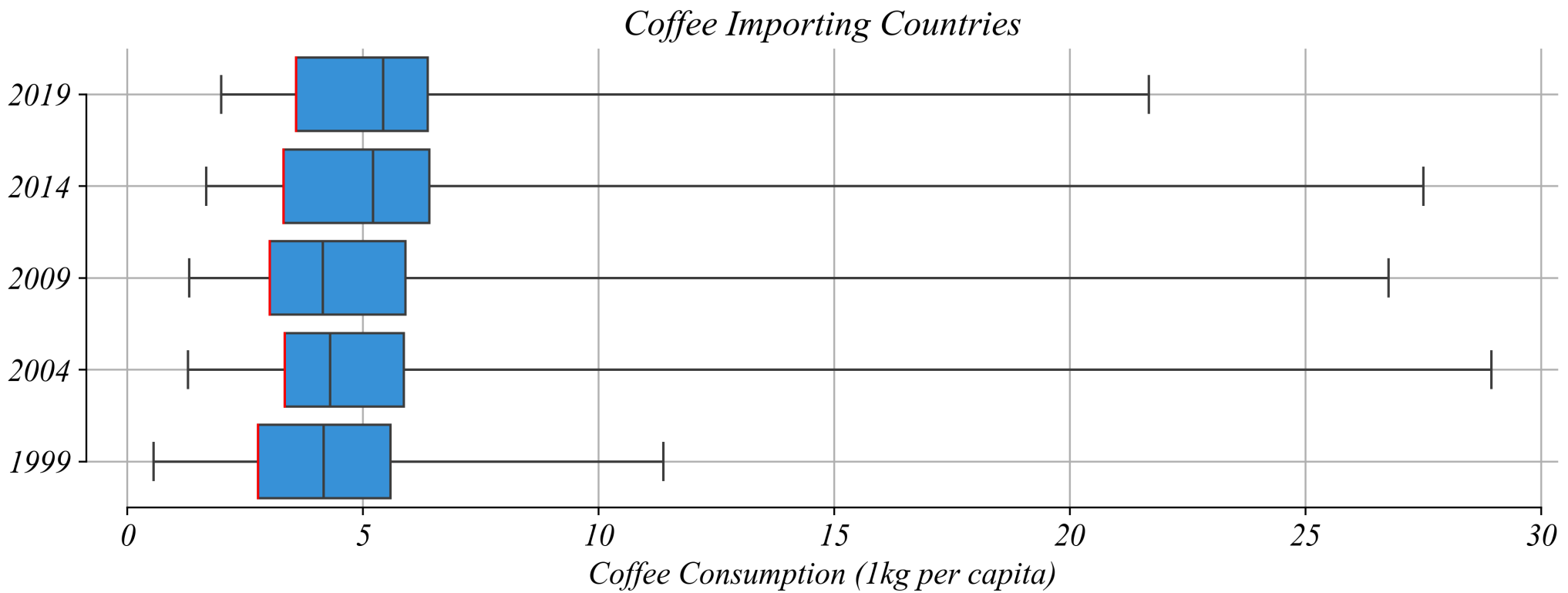
Which years did more than 25% consume less than 5 kg?



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

Panel Data: Multi-Boxplots

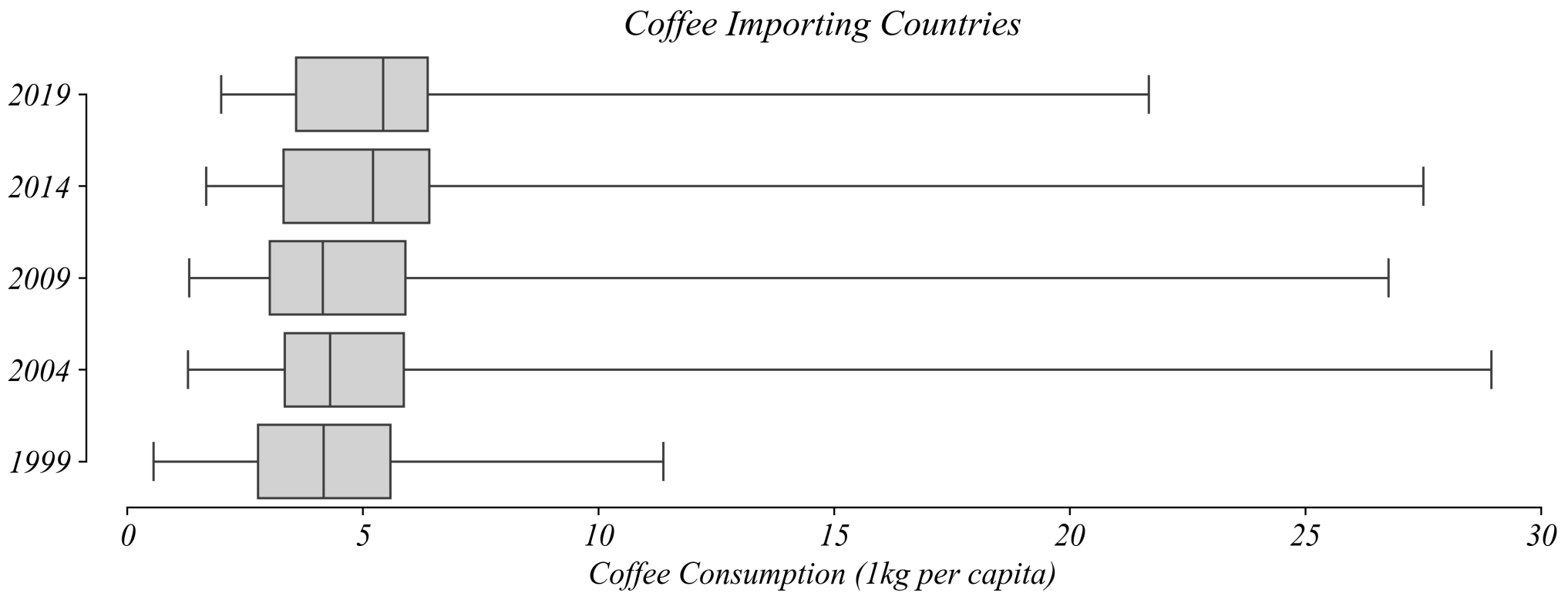
Which years did more than 25% consume less than 5 kg?



> *all of them*

Panel Data: Multi-Boxplots

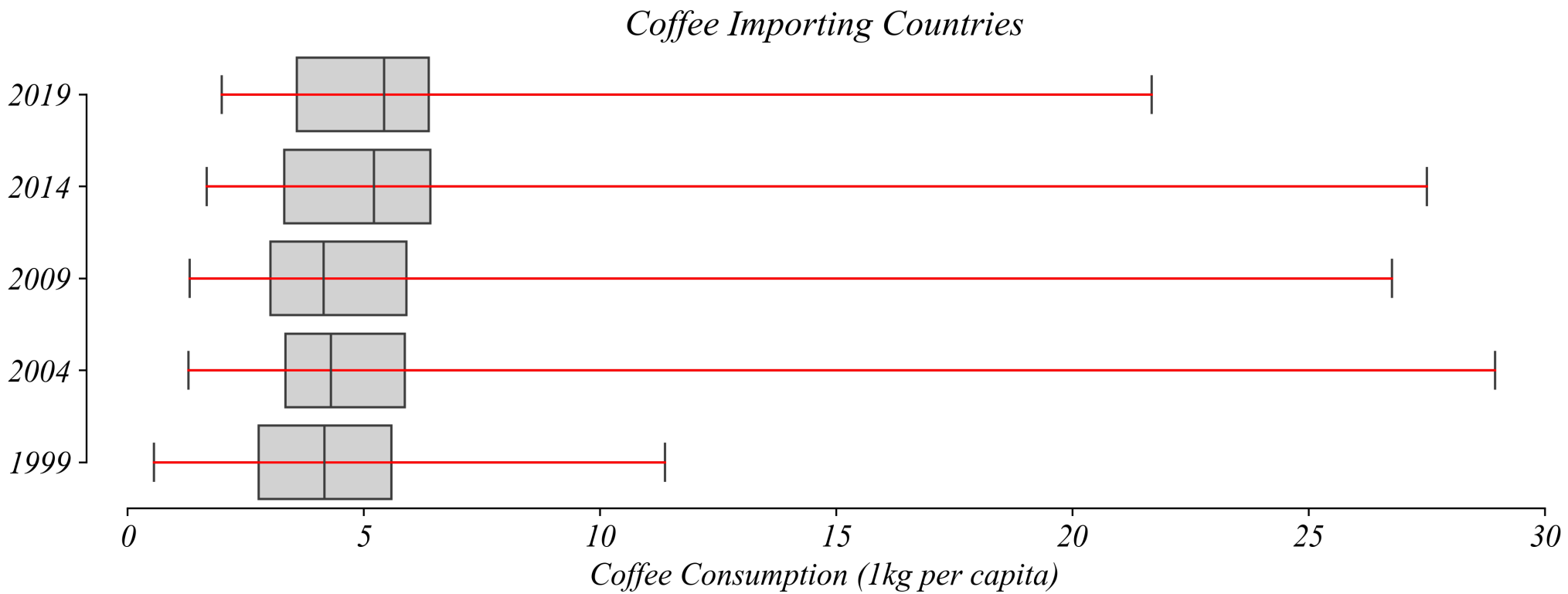
Which year saw the greatest difference between any two countries?



> look at the range

Panel Data: Multi-Boxplots

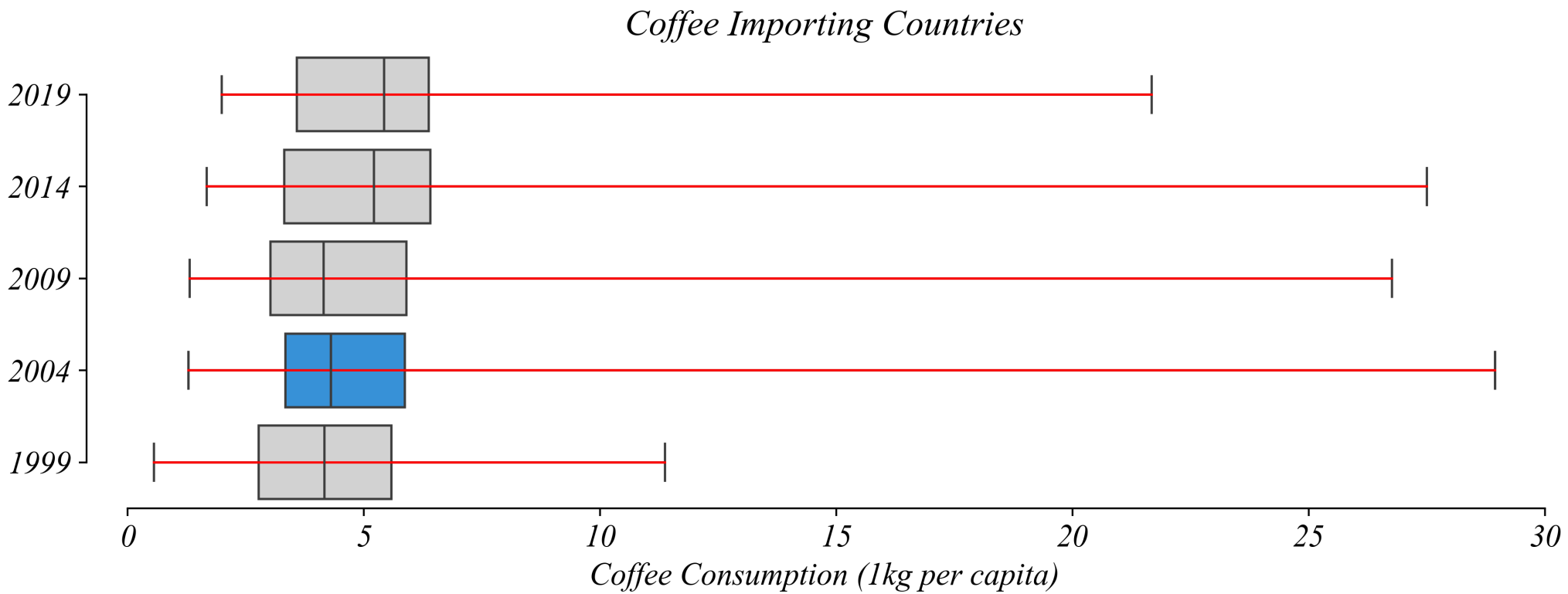
Which year saw the greatest difference between any two countries?



> look at the range

Panel Data: Multi-Boxplots

Which year saw the greatest difference between any two countries?



> look at the range and select the largest

Exercise 1.5 | Multi-Boxplots

Is the world drinking more coffee?

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`

Code	1999	2009	2019
AUT	8.43	6.37	7.93
BGR	2.65	3.30	3.64
HRV	4.48	5.10	5.62
CYP	3.48	4.05	5.62
CZE	3.26	3.02	4.74
…	…	…	…

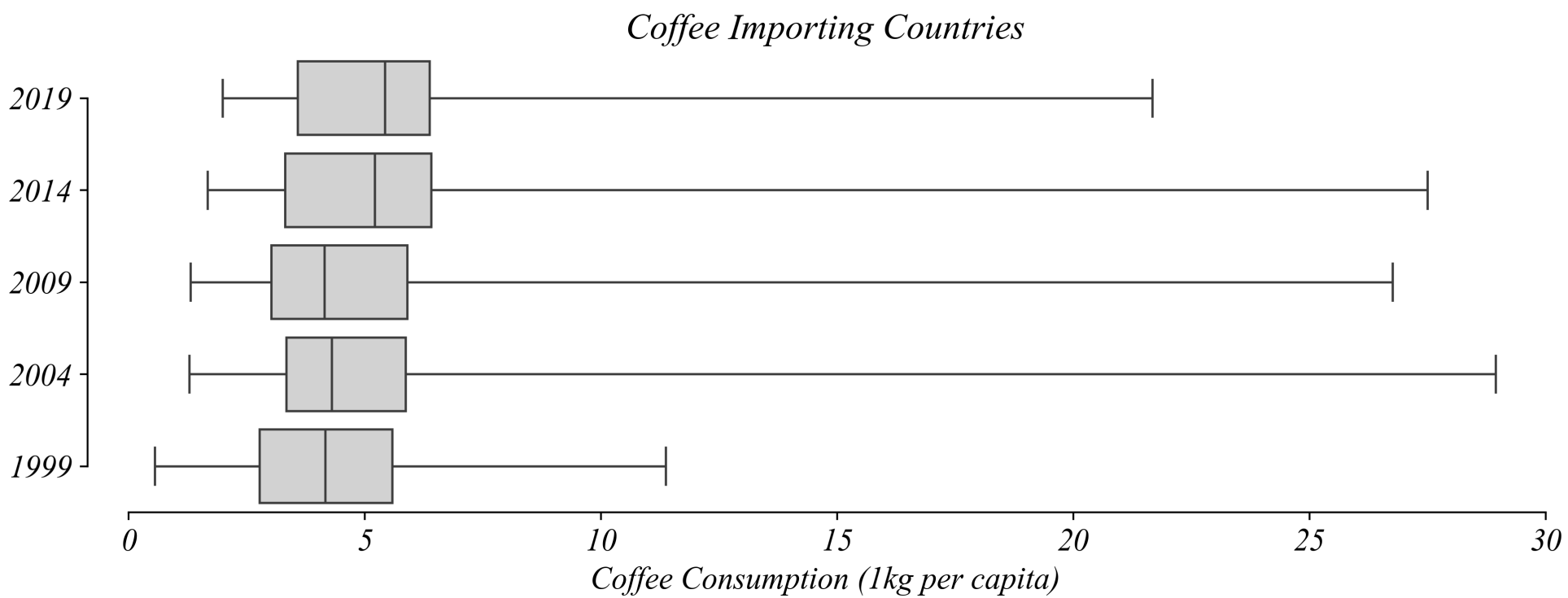
> *this is **Wide-Format Panel Data**: each year is in a separate column*

Exercise 1.5 | Multi-Boxplots

Is the world drinking more coffee?

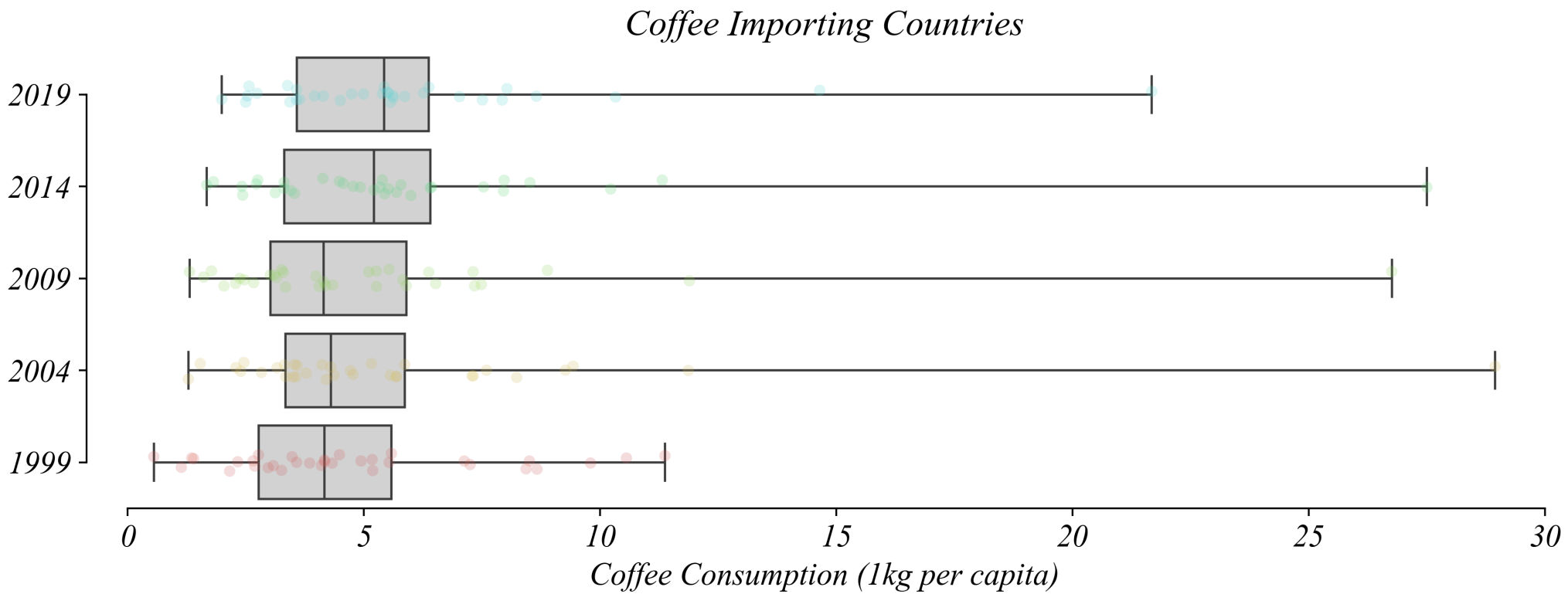
With wide-format panel data seaborn looks a little different.

```
# Wide Format Multi-Boxplot
percap_selected = percap[['1999','2009','2019']]
sns.boxplot(percap_selected, orient='h', whis=(0, 100))
```



Panel Data: Multi-Boxplots

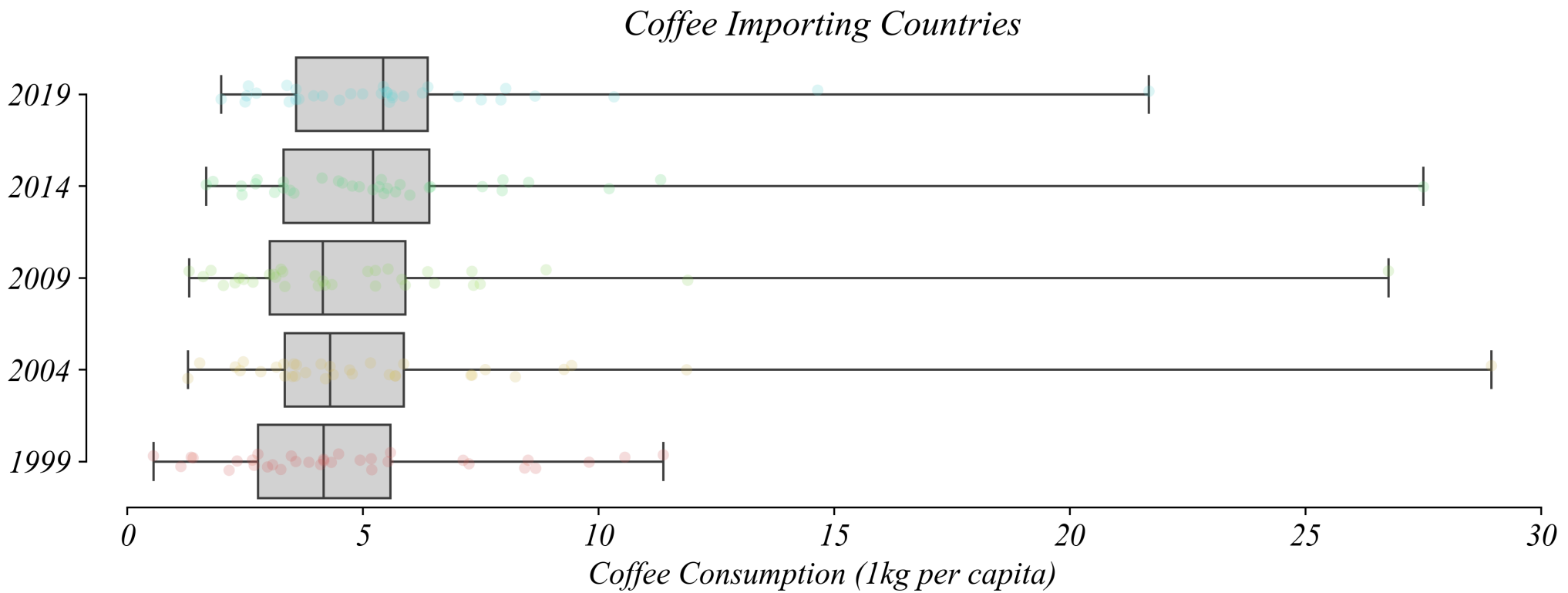
In which year did most countries increase their coffee consumption?



> not visible in the figure!

Panel Data: Relationships Between Years

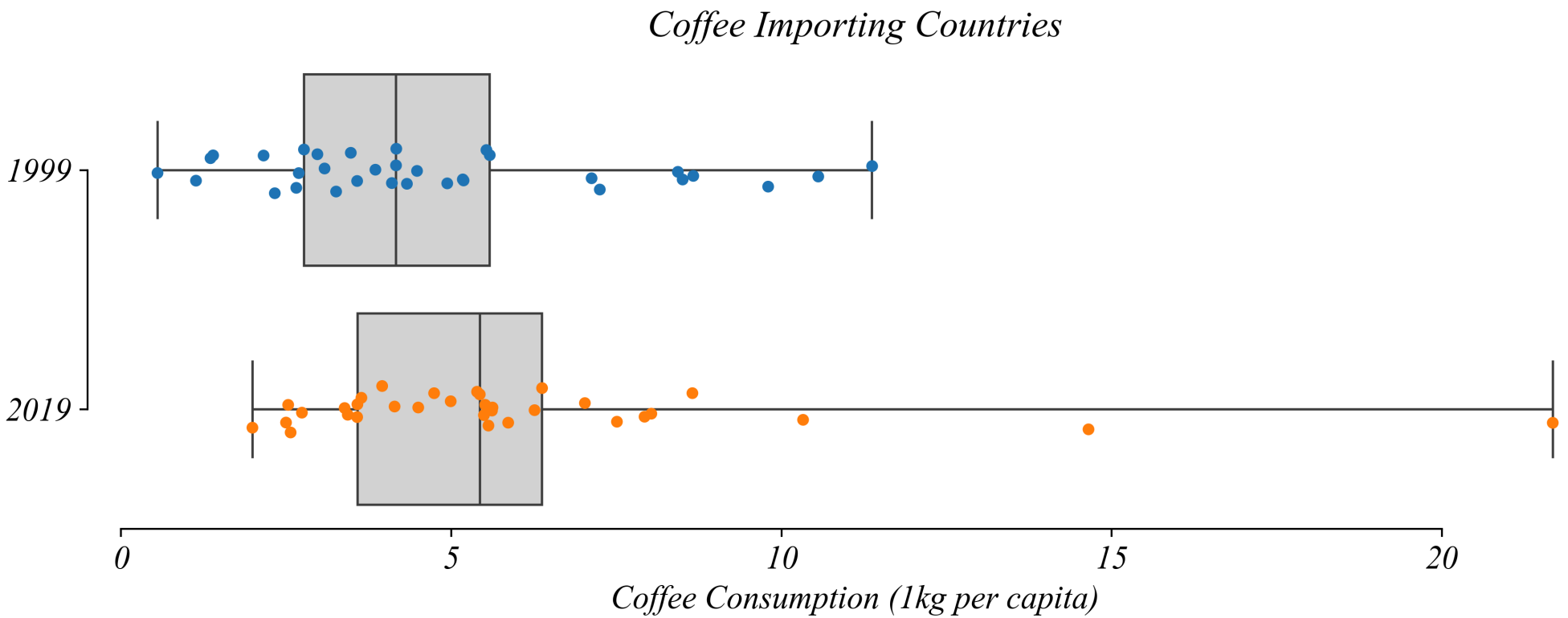
How many countries increased their coffee consumption between 1999 and 2019?



> *also not visible with this figure!*

Panel Data: Relationships Between Years

How many countries increased their coffee consumption between 1999 and 2019?



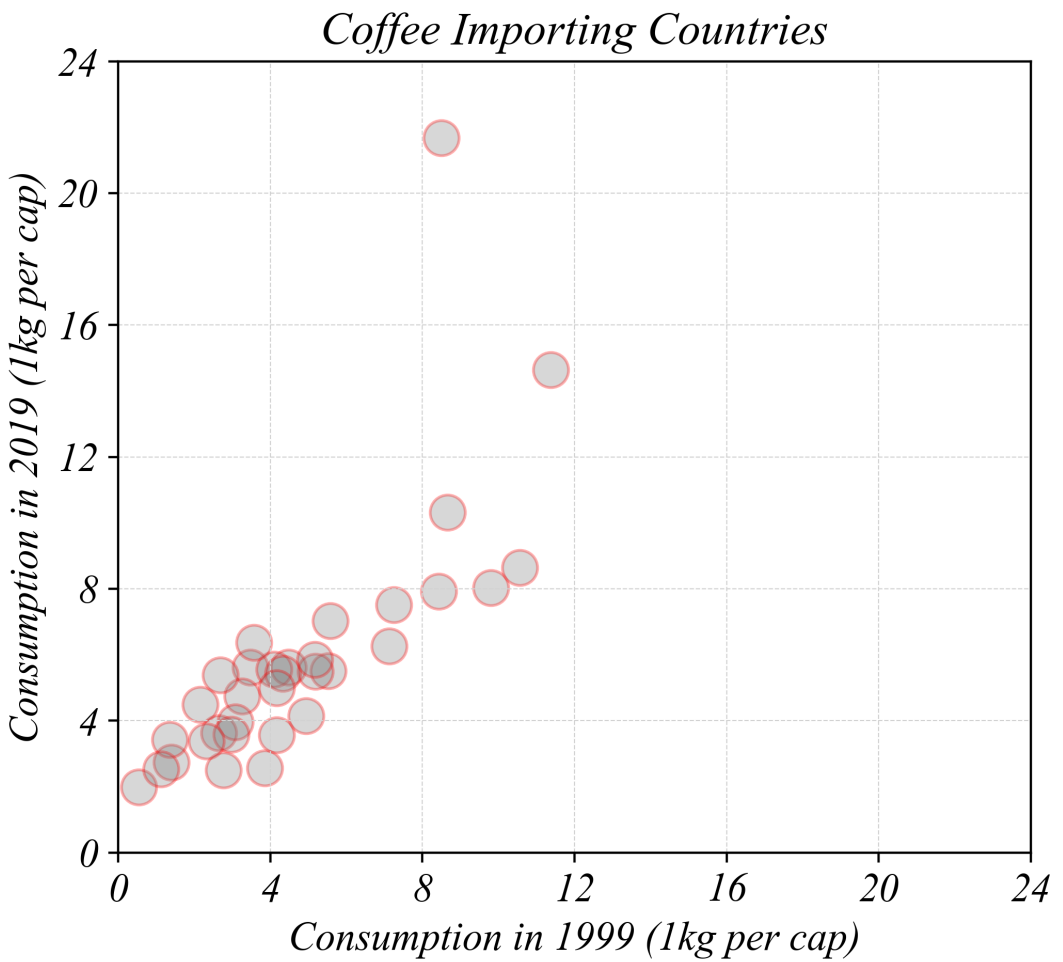
> better, but this figure still doesn't let us keep track of countries between years…

Panel Data: Relationships Between Years

How many countries increased their coffee consumption between 1999 and 2019?

Panel Data: Relationships Between Years

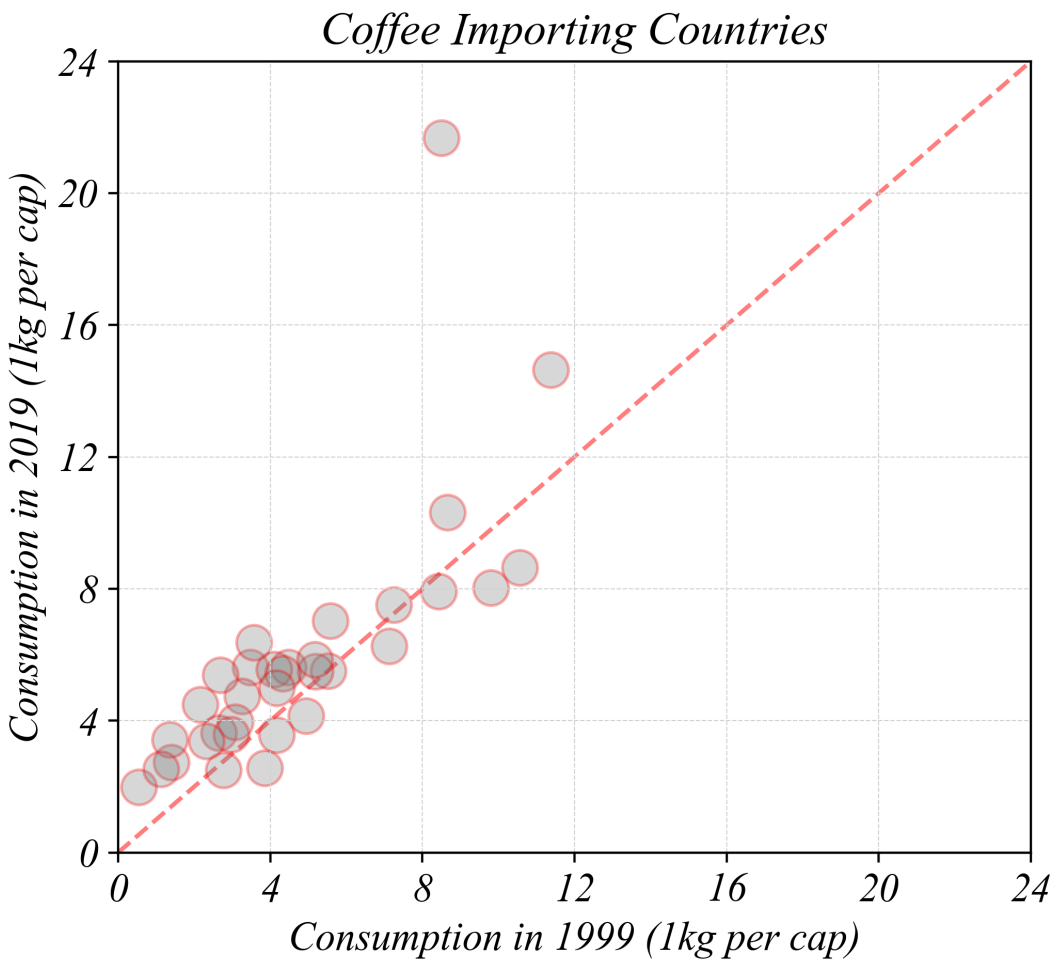
How many countries increased their coffee consumption between 1999 and 2019?



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

Panel Data: Relationships Between Years

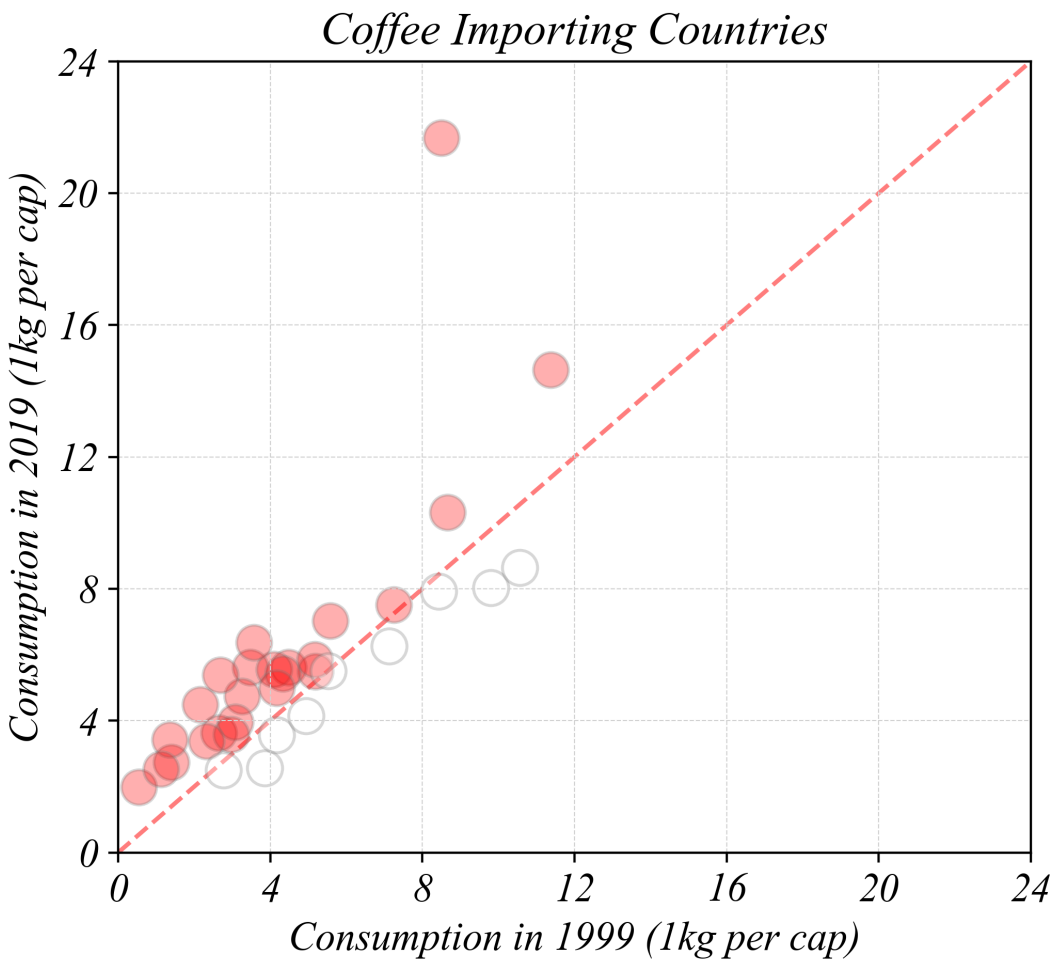
How many countries increased their coffee consumption between 1999 and 2019?



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

Panel Data: Relationships Between Years

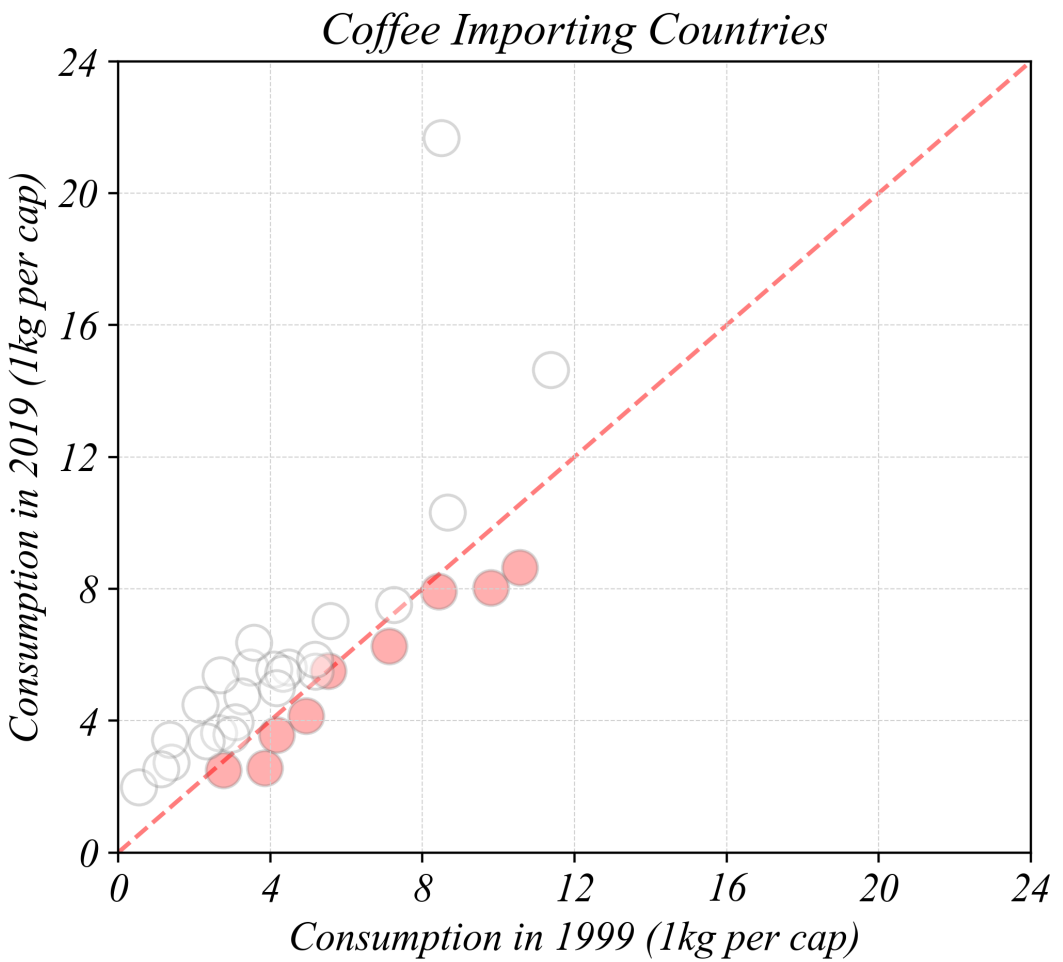
How many countries increased their coffee consumption between 1999 and 2019?



> points above the line increased

Panel Data: Relationships Between Years

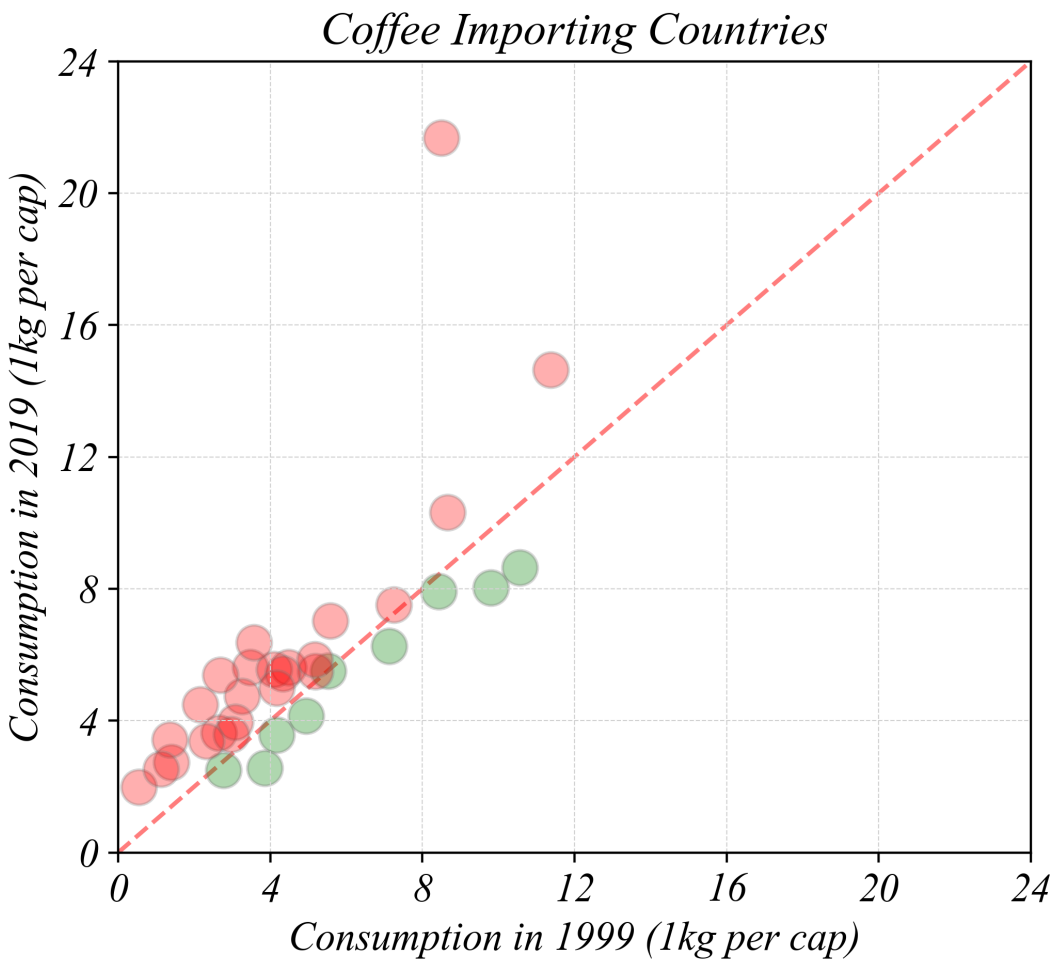
How many countries decreased their coffee consumption between 1999 and 2019?



> *points below the line decreased*

Panel Data: Relationships Between Years

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



> *we can use colors to visualize both increases and decreases*

Exercise 1.5 | Scatterplots

Is the world drinking more coffee?

We're going to use a scatterplot to visually examine how countries' coffee consumption changed between 1999 and 2019.

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

Exercise 1.5 | Scatterplots

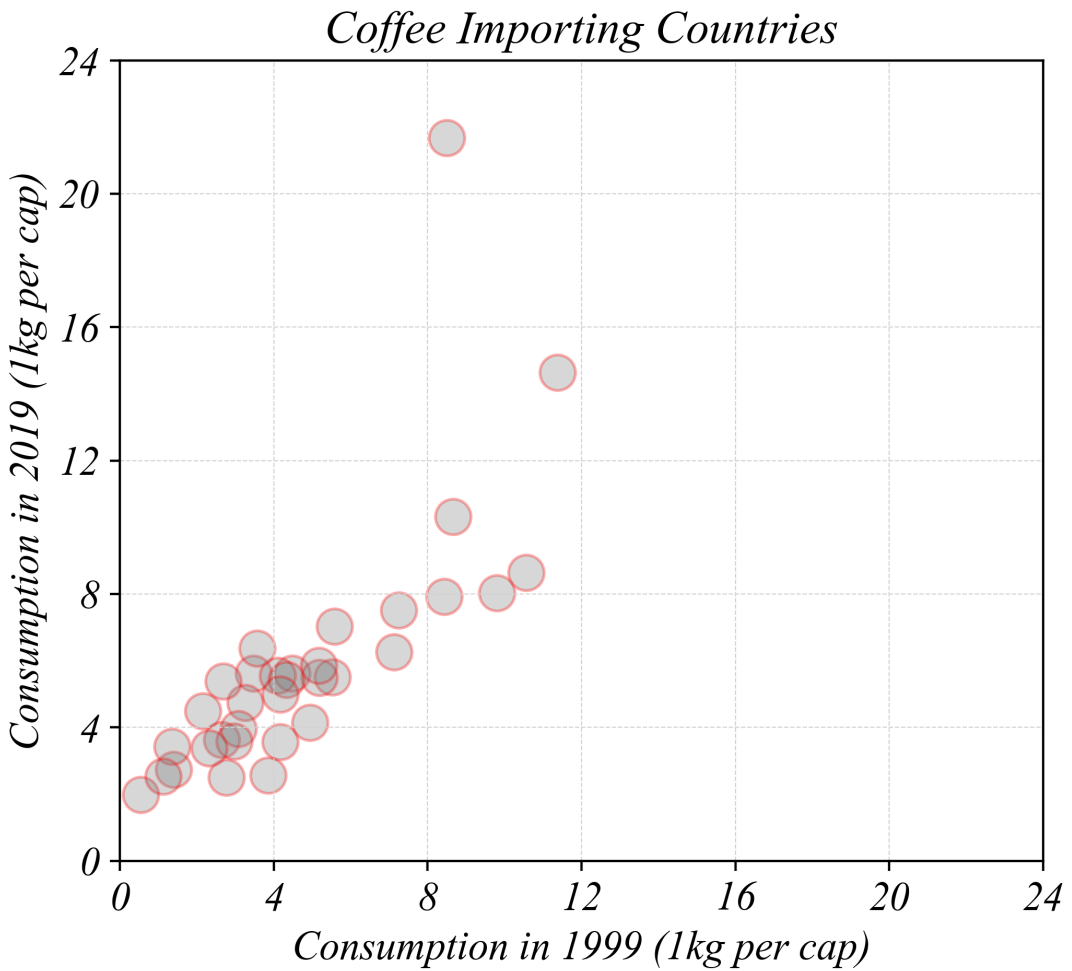
Is the world drinking more coffee?

```
# Wide Format Scatterplot
sns.scatterplot(percap, x='1999', y='2019')
```

Exercise 1.5 | Scatterplots

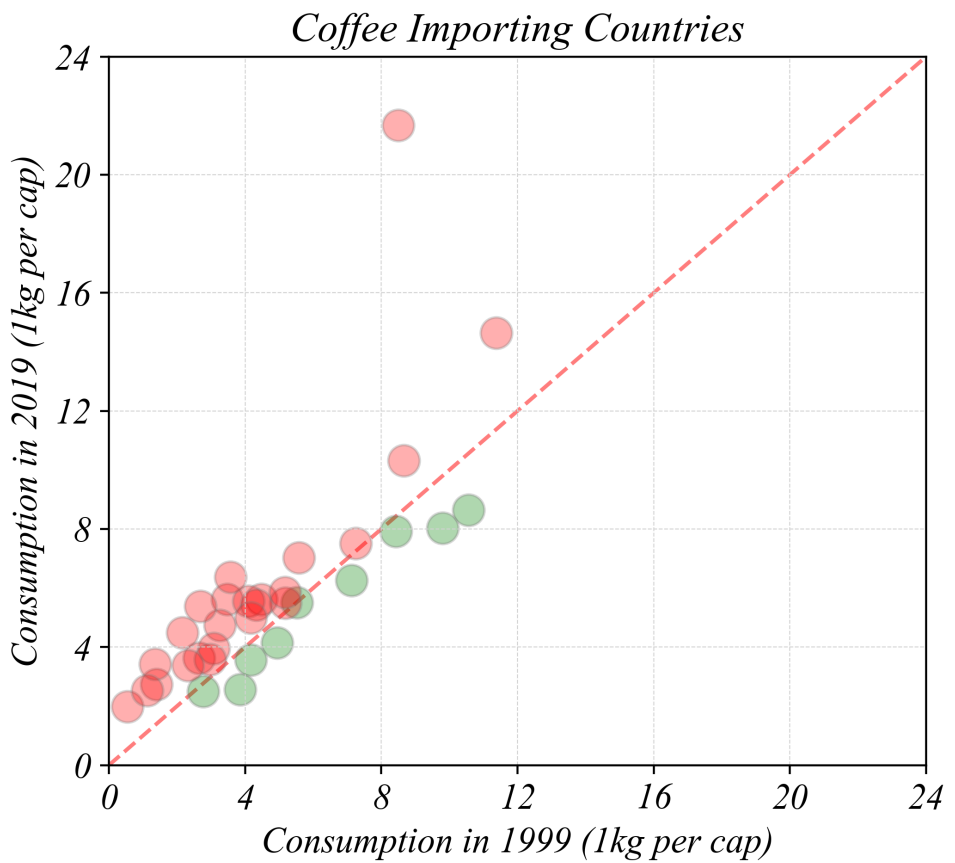
Is the world drinking more coffee?

```
# Wide Format Scatterplot
sns.scatterplot(percap, x='1999', y='2019')
```



Counting Changes

How many countries increased vs decreased?



> *we can see visually that most points are above the 45° line*

> *but how do we count exactly how many?*

Counting Changes

How many countries increased vs decreased?

1. ***Create a change column:*** *Subtract old from new*

Code	1999	2019	change
AUT	8.43	7.93	-0.50
BGR	2.65	3.64	0.99
HRV	4.48	5.62	1.14
…	…	…	…

2. ***Filter:*** *Select rows where change > 0 (or < 0)*

3. ***Count:*** *Use* `len()` *to count the filtered rows*

> positive change = increased; negative change = decreased

Exercise 1.5 | Counting Changes

How many countries increased their coffee consumption?

```
# Create a change column
percap['change'] = percap['2019'] - percap['1999']
percap.head()
```

```
# Count countries that increased (change > 0)
increased = percap[percap['change'] > 0]
len(increased)
```

```
# Count countries that decreased (change < 0)
decreased = percap[percap['change'] < 0]
len(decreased)
```

Question

How has each country's consumption changed over time?

Instead of measuring changes between two years (1999 and 2019), how might we visualize countrys' full trends over all 30 years?

> lets use a line plot with Year on x-axis and Consumption on y-axis, colored by Country

Question

How has each country's consumption changed over time?

The problem is that wide format has no `Year` column!

```
# This won't work!
sns.lineplot(percap, x='Year', y='Consumption', hue='Code')
```

> we need to construct a year column

Reshaping: Wide → Long

Turn columns into rows using `melt()`.

Melting: Wide → Long

Wide Format

Code	1999	2019
FRA	5.5	5.5
DEU	7.1	6.3
JPN	3.0	3.6

melt()

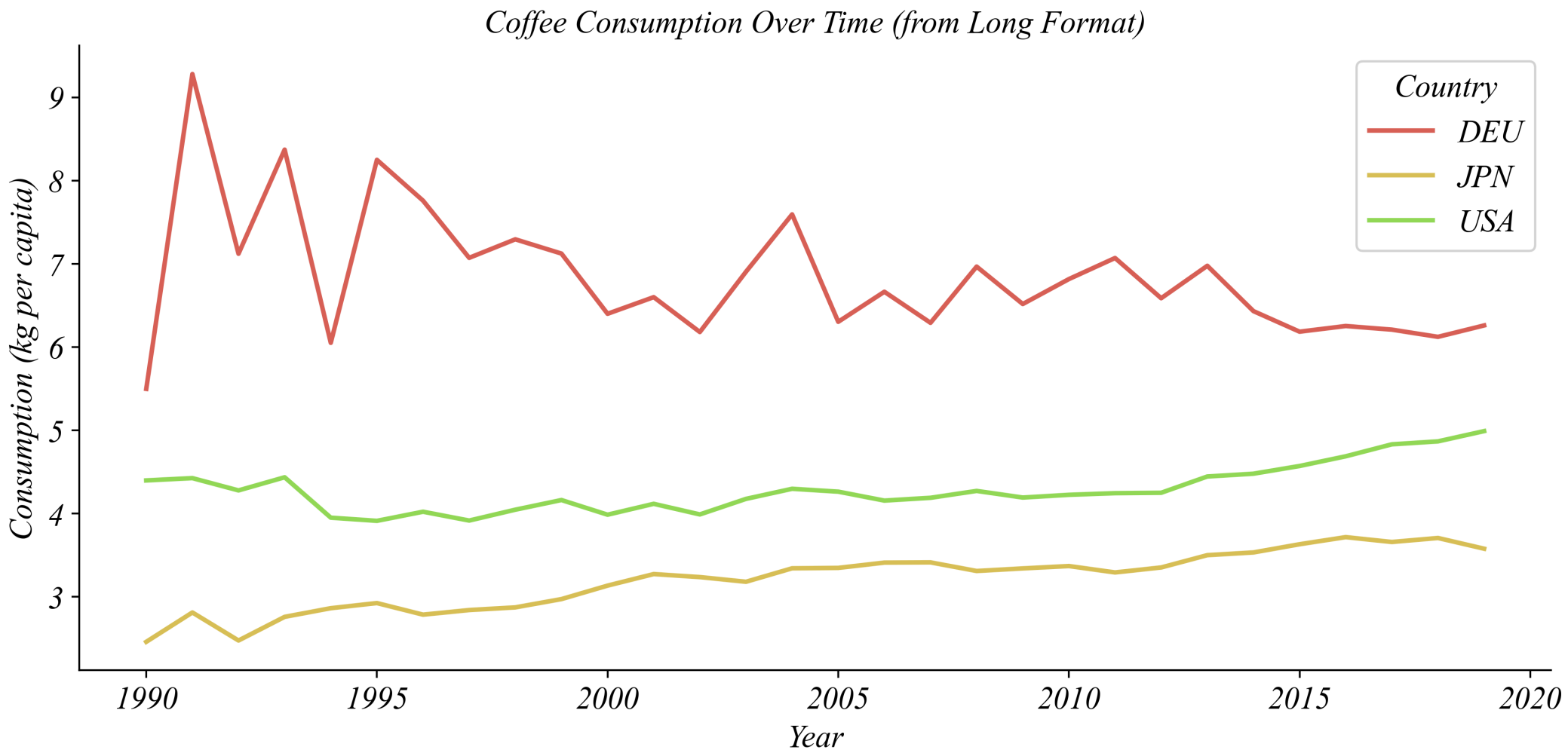
Long Format

Code	Year	Consumption
FRA	1999	5.5
DEU	1999	7.1
JPN	1999	3.0
FRA	2019	5.5
DEU	2019	6.3
JPN	2019	3.6

> *each year column becomes rows in a new `Year` column*

Question

How has each country's consumption changed over time?



> *this is possible now using the long format panel data*

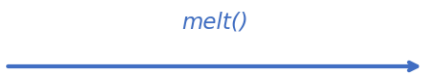
Exercise 1.5 | Reshaping

How has each country's consumption changed over time?

Use `melt()` to reshape wide → long.

Wide Format

Code	1999	2019
FRA	5.5	5.5
DEU	7.1	6.3
JPN	3.0	3.6



Long Format

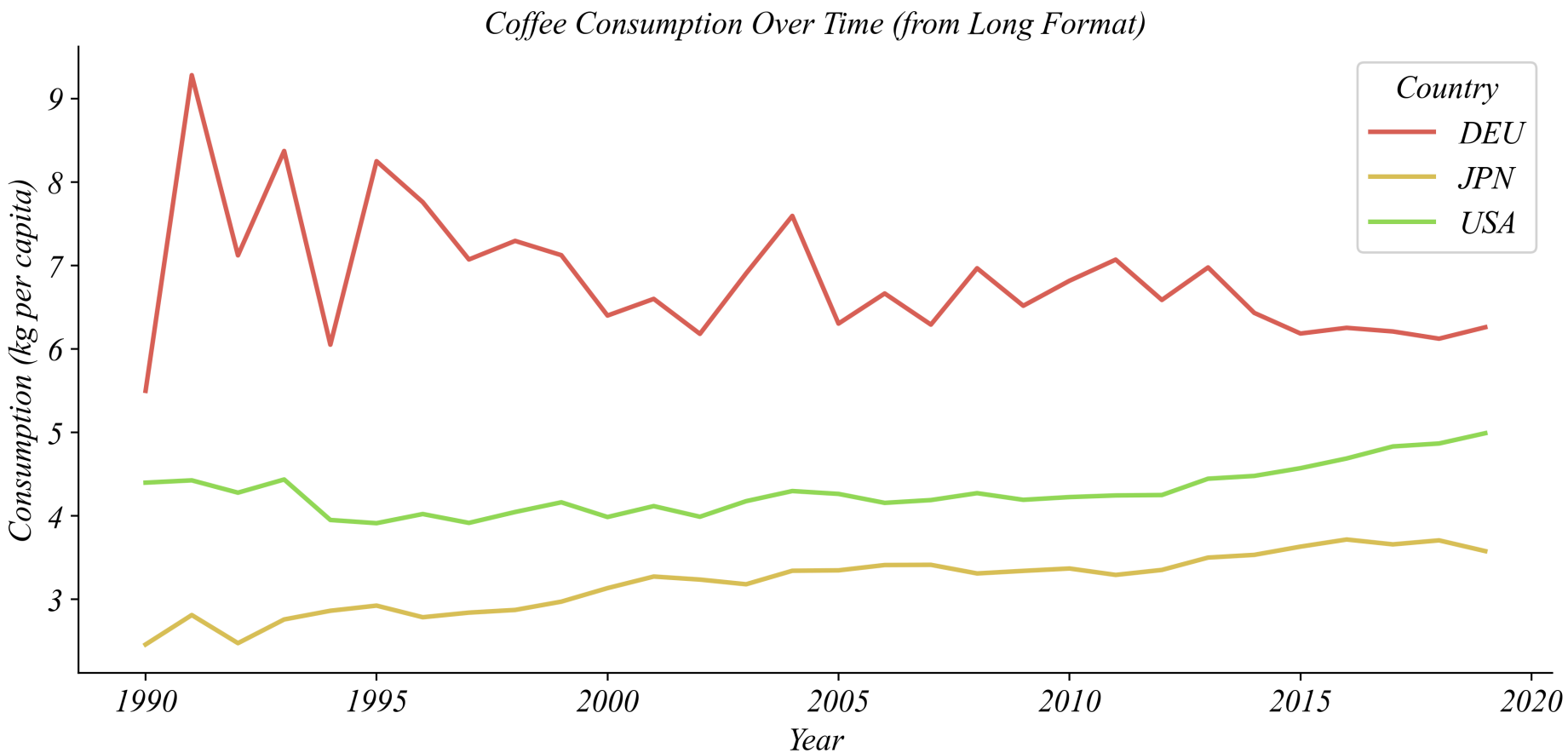
Code	Year	Consumption
FRA	1999	5.5
DEU	1999	7.1
JPN	1999	3.0
FRA	2019	5.5
DEU	2019	6.3
JPN	2019	3.6

```
# Wide to Long
percap_long = percap.melt(
    id_vars='Code',           # The identifier column
    var_name='Year',          # The new 'melted' column
    value_name='Consumption' # The new 'melted' values
)
```

Exercise 1.5 | Reshaping

How has each country's consumption changed over time?

```
# Now we can plot trends for each country
sns.lineplot(percap_long, x='Year', y='Consumption', hue='Code')
```



> *long format makes line plots natural — use `hue='Code'`*

Reshaping: Long → Wide

Turn rows into columns using `pivot()`.

Sometimes data comes in long format but we need wide format *(eg. scatterplots)*.

Long Format

Code	Year	Consumption
FRA	1999	5.5
DEU	1999	7.1
JPN	1999	3.0
FRA	2019	5.5
DEU	2019	6.3
JPN	2019	3.6

Wide Format

Code	1999	2019
FRA	5.5	5.5
DEU	7.1	6.3
JPN	3.0	3.6

Think of `pivot()` as the opposite of `melt()`.

Reshaping: The Code

Turn rows into columns using `pivot()`.

Use `pivot()` to reshape long → wide.

Long Format

Code	Year	Consumption
FRA	1999	5.5
DEU	1999	7.1
JPN	1999	3.0
FRA	2019	5.5
DEU	2019	6.3
JPN	2019	3.6

Wide Format

Code	1999	2019
FRA	5.5	5.5
DEU	7.1	6.3
JPN	3.0	3.6

```
# Long to Wide
percap_wide = percap_long.pivot(
    index='Code',          # The identifier column
    columns='Year',        # The column used to create the new column headers
    values='Consumption'   # The column used to create the new column values
)
```

> you'll use this in your homework :)

Reshaping

Choose based on your starting point and goal

The data format should match your visualization goal.

Direction	Function	When to use
Wide → Long	`melt()`	Need line plots, faceting
Long → Wide	`pivot()`	Need boxplots, scatterplots

- *`melt()` turns columns into rows (wide → long)*
- *`pivot()` turns rows into columns (long → wide)*

Part 1.5 | Panel Data (Wide Format)

Summary

- ***Wide format:*** *Each time period is a column*
- ***Multi-boxplots*** *compare distributions across time*
- ***Scatterplots with 45° lines*** *track individual changes*
- ***Filtering*** *with* `df[df['col'] > 0]` *counts subsets*
- ***melt()*** *converts wide → long when you need line plots*
- ***pivot()*** *converts long → wide when you need boxplots or scatterplots*

Building Blocks

What this unit adds to your toolkit

Block	**Part 1.5**
Variables	Numerical
Structures	Panel (wide format)
Operations	Filter, Reshape (melt, pivot)
Visualizations	Multi-boxplot, Scatterplot with 45° line