Data Visualisation

CMP020L013A

Week 9: Time-Series Visualisation

Mohammad Javaheri
(Dr Mohammad Ali Javaheri Javid)

Agenda

- Time-series visualisation
- Time-series definition, data types
- Applications, examples
- Analysis, models
- Case study

Time-series

- Time series data often generated when monitoring industrial processes or tracking corporate business metrics.
- Time series analysis accounts for the fact that data points taken over time may have an internal structure
  - such as autocorrelation, trend or seasonal variation that should be accounted for
- Definition:
  - An ordered sequence of values of a variable at equally spaced time intervals.
  - What if they are is not equally spaced?
- Change over time
- Line graphs are the most common method

Time-series – Applications

1. Obtain an understanding of the underlying forces, factors and structure that produced the observed data
2. Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control.

- Anything involving measurements over time

- A random sample of 4,000 graphics from 15 of the world's newspapers published from 1974 to 1989 found that more than 75% of all graphics were time series

Time-series – Applications

- Economic Forecasting
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Yield Projections
- Process and Quality Control
- Inventory Studies
- Workload Projections
- Census Analysis

Examples of time-series data

- IoT and sensor data:
  - Monitoring and analysing sensor data from devices, machinery, or infrastructure to predict maintenance needs, optimize performance, and detect anomalies.
- Weather forecasting:
  - Utilising historical weather data to forecast future meteorological conditions, such as temperature, precipitation, and wind patterns.
- E-commerce and retail:
  - Tracking sales data over time to identify seasonal trends, forecast demand, and optimize inventory management and pricing strategies.

Examples of time-series data

- Healthcare:
  - Analysing patient vital signs, medical records, and treatment outcomes to improve healthcare delivery, disease surveillance, and patient care.
- Energy consumption:
  - Studying electricity or energy usage patterns to optimise consumption, forecast demand, and support energy efficiency initiatives.
- Manufacturing and supply chain:
  - Monitoring production processes, inventory levels, and supply chain data to enhance operational efficiency and demand forecasting.

Looking for trends

- Seasonal variation
- Significant events
- Natural disasters, changes of government, tax regulation…
- Correlations
- Do variables change in similar ways over time?
- Understanding the past and predicting the future

Different kinds of patterns

- Variability
- Rate of change
- Co-variance and correlation
- Cycles
- Exceptions

Sampling rate

- Some variable are constantly changing over time
  - Temperature, physiological response, stock market
- Rate of measurement is important
  - e.g. temperature – one measurement per day, or one measurement per second?
  - e.g. stock market – trend over a whole year, or changes over milliseconds?

Time series Analysis Models

- Classification:
  - Identifies and assigns categories to the data.
- Curve fitting:
  - Plots the data along a curve to study the relationships of variables within the data.
- Descriptive analysis:
  - Identifies patterns in time series data, like trends, cycles, or seasonal variation.
- Explanatory analysis:
  - Attempts to understand the data and the relationships within it, as well as cause and effect.

Time series Analysis Models

- Exploratory analysis:
  - Highlights the main characteristics of the time series data, usually in a visual format.
- Forecasting:
  - Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.
- Intervention analysis:
  - Studies how an event can change the data.
- Segmentation:
  - Splits the data into segments to show the underlying properties of the source information.

Variations in Data

- In time series data, variations can occur periodically throughout the data:
- Functional analysis can pick out the patterns and relationships within the data to identify notable events.
- Trend analysis means determining consistent movement in a certain direction. T
- Two types of trends:
  - deterministic, where we can find the underlying cause,
  - stochastic, which is random and unexplainable.

Variations in Data

- Seasonal variation describes events that occur at specific and regular intervals during the course of a year.
- Serial dependence occurs when data points close together in time tend to be related.
- Cross-sectional data consists of several variables recorded at the same time.
- Pooled data is a combination of both time series data and cross-sectional data.

Variations in Data

- Time series analysis and forecasting models must define the data types relevant to answering the business question.

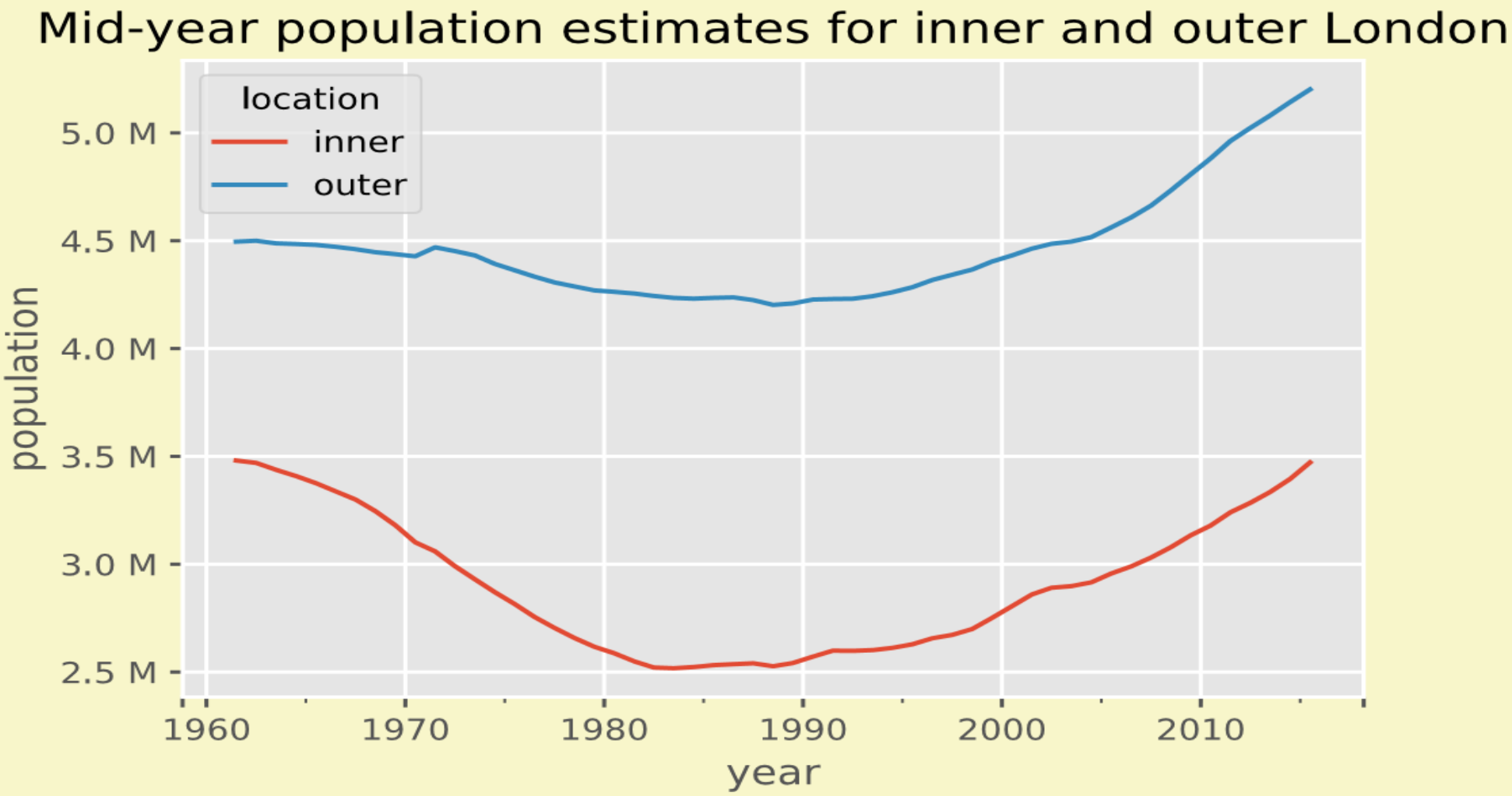
- Once analysts have chosen the relevant data they want to analyse, they choose the types of analysis and techniques that best fit.

Time-series - London population

- London population between 1961 and 2015

year	inner	outer
1961-06-30	3.481 (+06)	4.496 (+06)
1962-06-30	3.47 (+06)	4.5 (+06)
1963-06-30	3.438 (+06)	4.488 (+06)
1964-06-30	3.409 (+06)	4.485 (+06)
1965-06-30	3.376 (+06)	4.481 (+06)

Time-series - London population



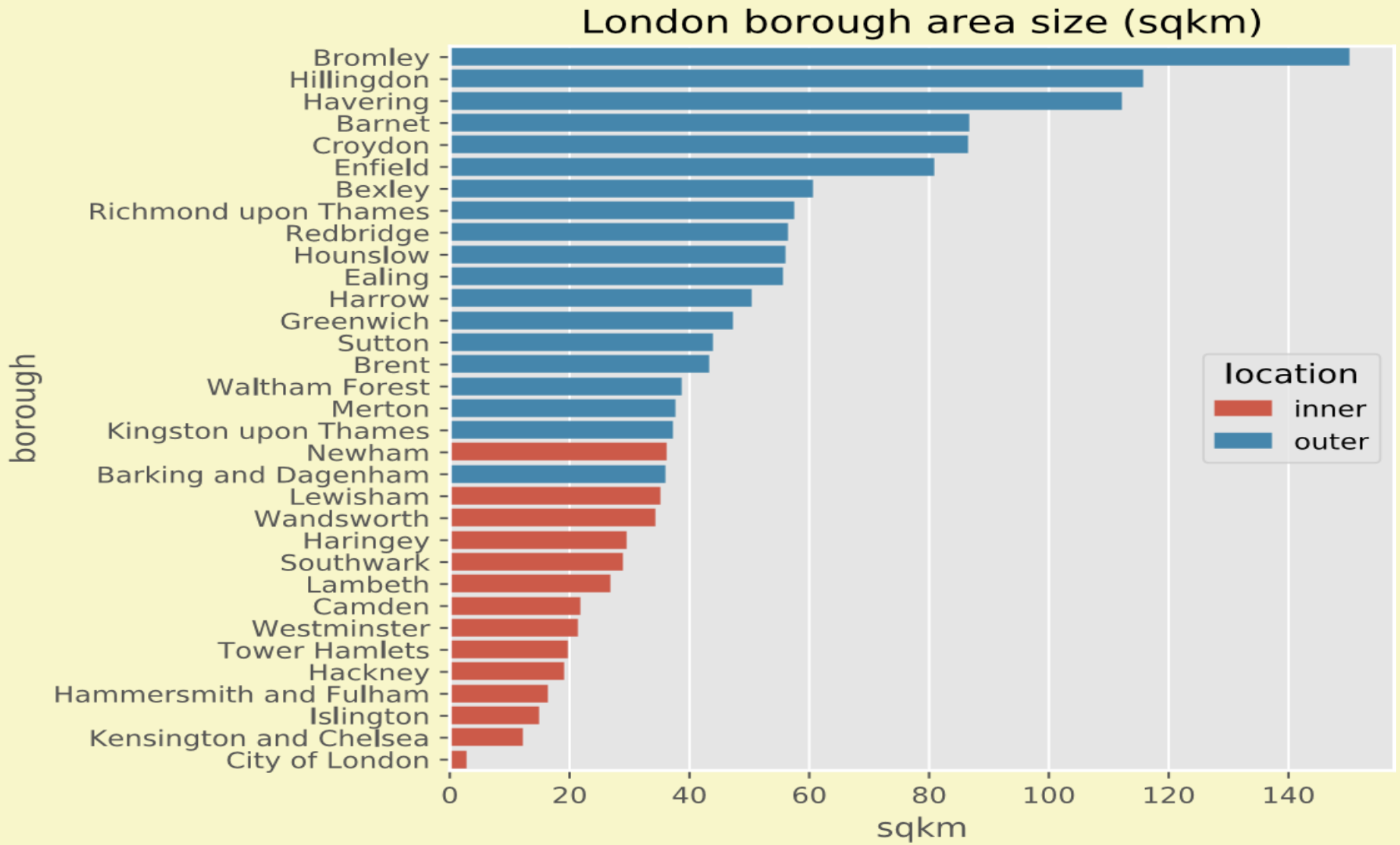
Time-series - London population between 1961 and 2015

- Variables
- What are the variables?
- Is the data tidy?

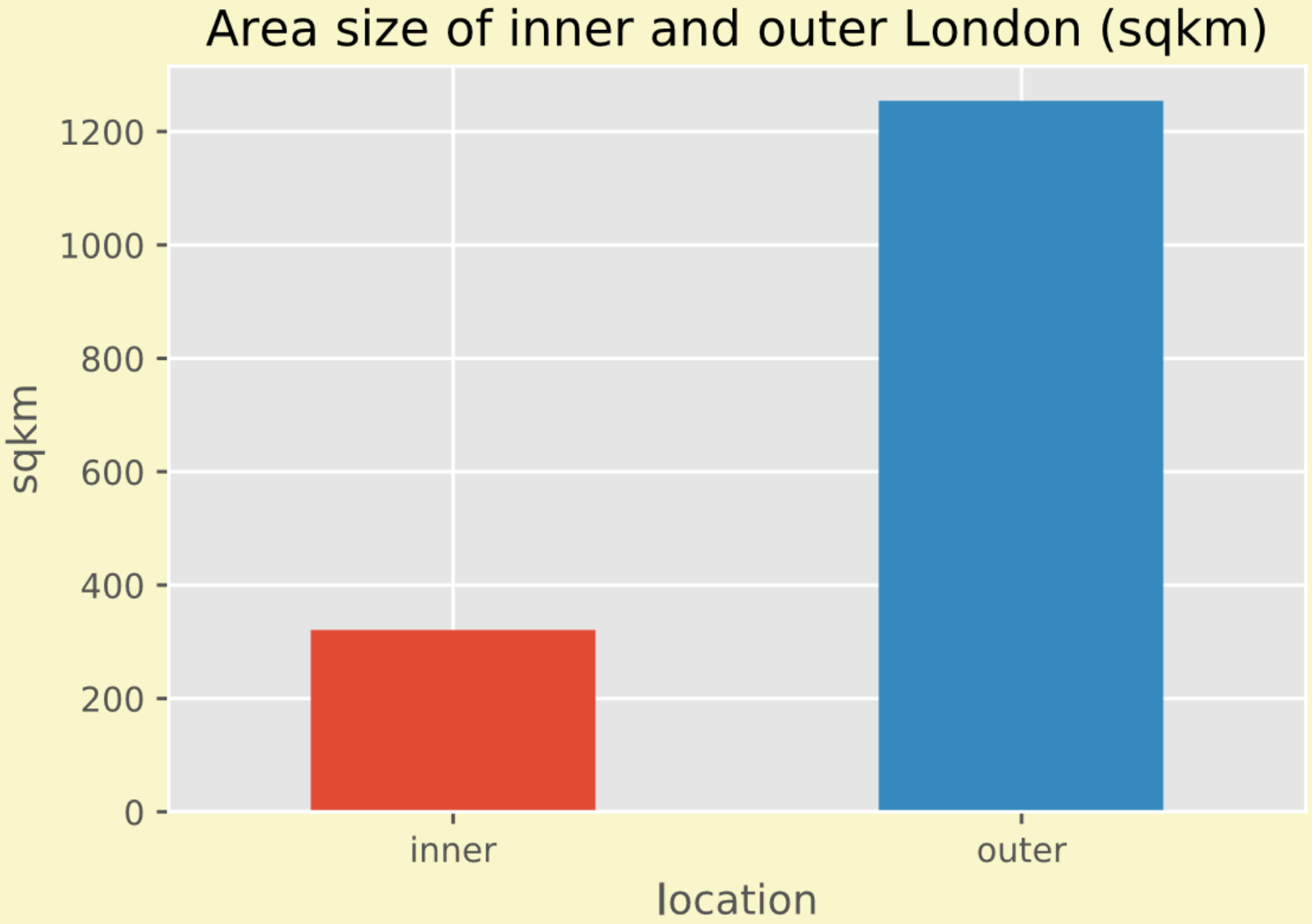
	A	B	C	T	U	V	W	X
1								
2	**Code**	**New Code**	**Area name**	**1976**	**1977**	**1978**	**1979**	**1980**
6	00AC	E09000003	Barnet	300,900	298,900	296,900	295,800	294,800
7	00AD	E09000004	Bexley	218,100	217,100	216,500	216,200	216,600
8	00AE	E09000005	Brent	263,300	260,600	259,900	256,900	255,000
9	00AF	E09000006	Bromley	300,500	299,000	298,700	298,100	298,300
10	00AG	E09000007	Camden	190,000	188,400	189,700	188,000	182,400
11	00AH	E09000008	Croydon	327,800	325,900	324,800	324,000	322,200
12	00AJ	E09000009	Ealing	294,600	291,700	291,400	287,600	283,100
13	00AK	E09000010	Enfield	263,100	261,400	260,400	259,500	260,500
14	00AL	E09000011	Greenwich	214,800	214,200	213,400	214,200	216,100
15	00AM	E09000012	Hackney	198,100	192,400	188,800	186,800	186,000
16	00AN	E09000013	Hammersmith and Fulham	172,100	166,800	163,100	157,900	154,200
17	00AP	E09000014	Haringey	224,700	222,800	221,800	217,600	214,700
18	00AQ	E09000015	Harrow	202,400	201,100	199,600	199,100	199,400
19	00AR	E09000016	Havering	243,300	243,200	243,300	243,300	243,500

Borough area size (sq km)

Borough area size (sq km)



Borough area size (sq km) -Aggregate location



Borough population over time

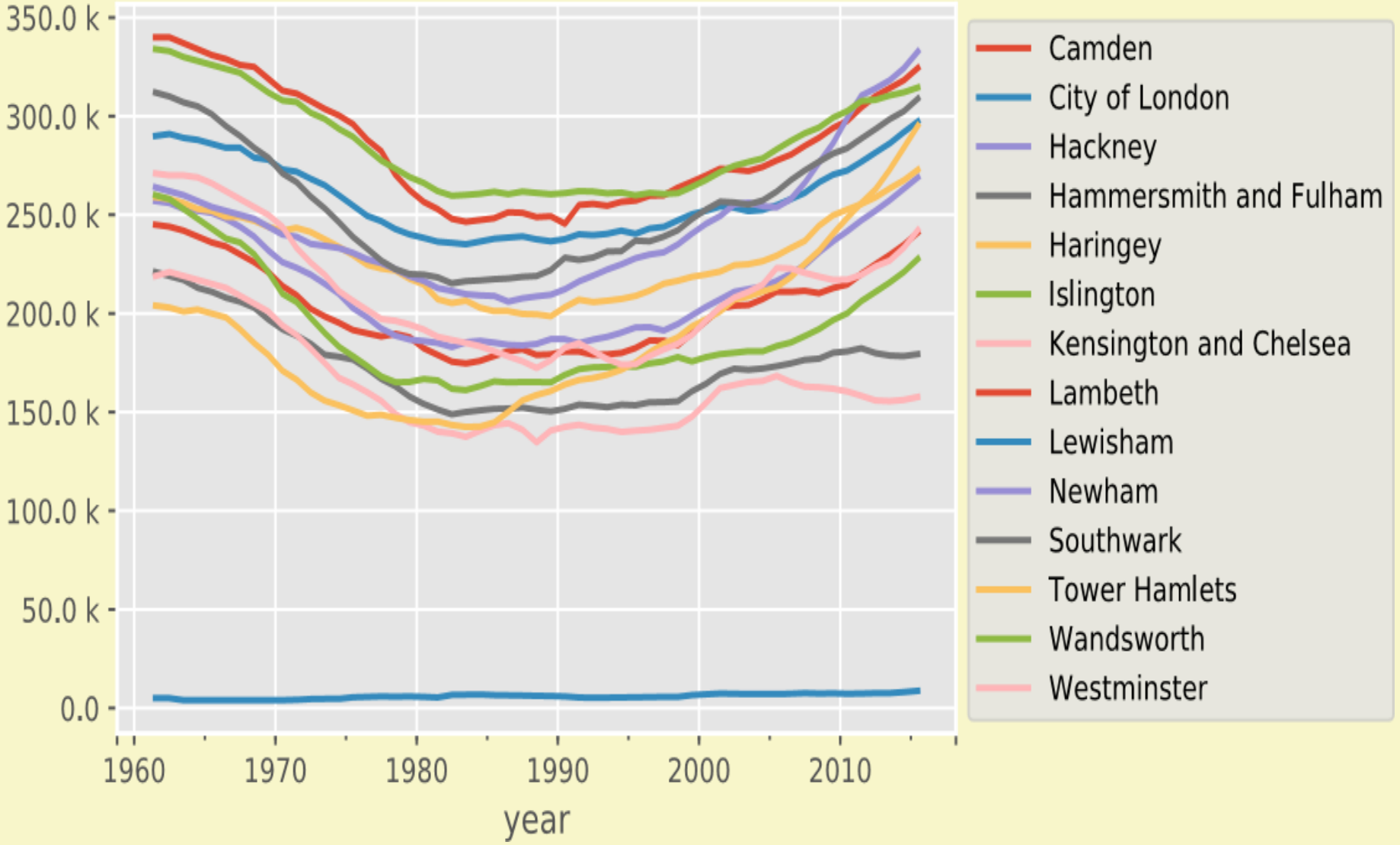
borough	location	1961	1962	1963
0City of London	inner	5000	5000	4000
1 Barking and Dagenham	outer	176000	176000	174000
2Barnet	outer	318000	318000	316000
3Bexley	outer	210000	211000	212000

Borough population over time - Cross-section

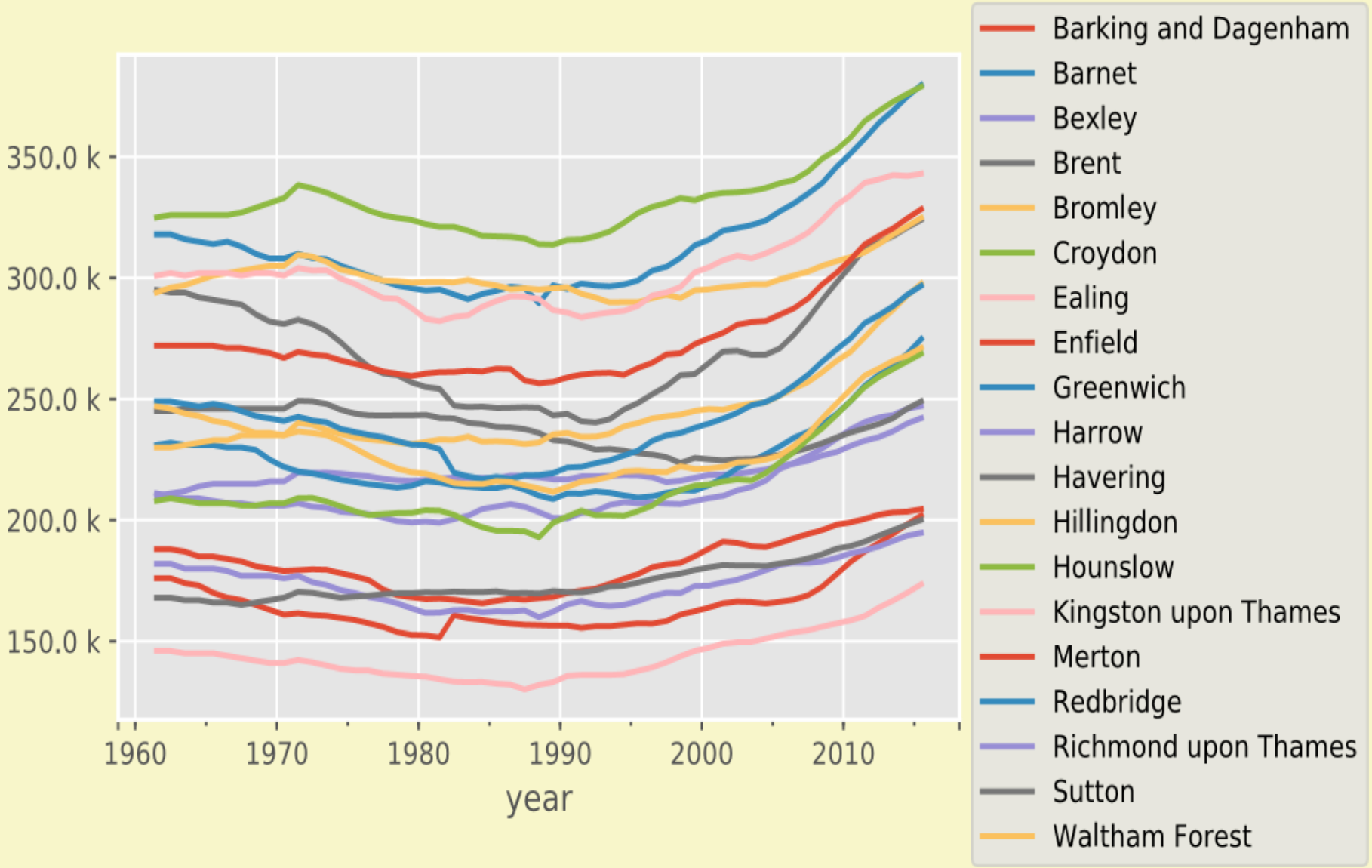
borough	population
Camden	209300
City of London	4200
Hackney	223200
Hammersmith and Fulham	188800
Haringey	243400

- Cross-section can be used across multiple levels
  - but: only one value can be sliced for each level

Plot of inner boroughs



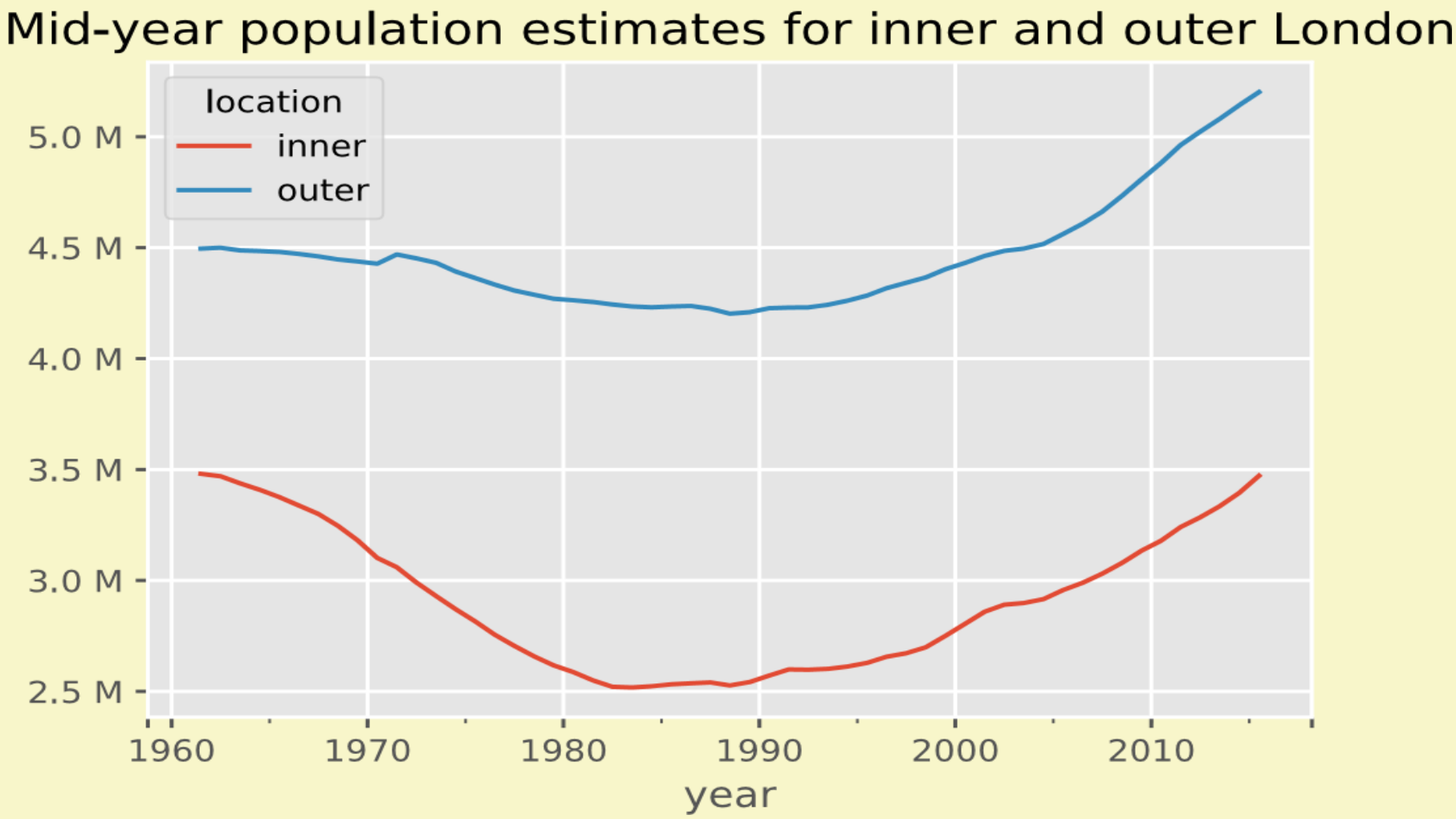
Plot of outer boroughs



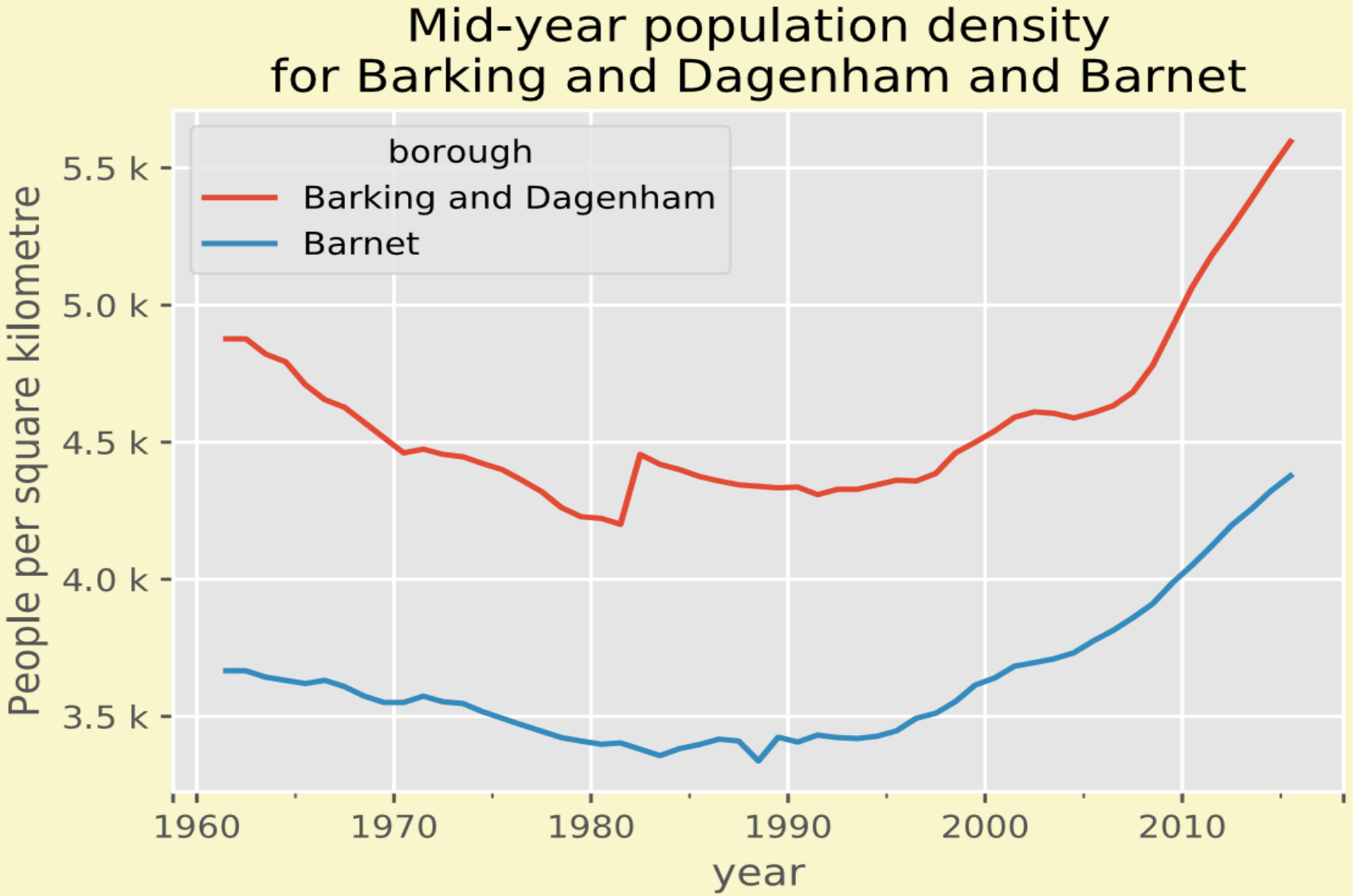
Aggregate inner/outer boroughs

	population
(Period('1961-06-30', 'D'), 'inner')	3.481 (+06)
(Period('1961-06-30', 'D'), 'outer')	4.496 (+06)
(Period('1962-06-30', 'D'), 'inner')	3.47 (+06)
(Period('1962-06-30', 'D'), 'outer')	4.5 (+06)
(Period('1963-06-30', 'D'), 'inner')	3.438 (+06)

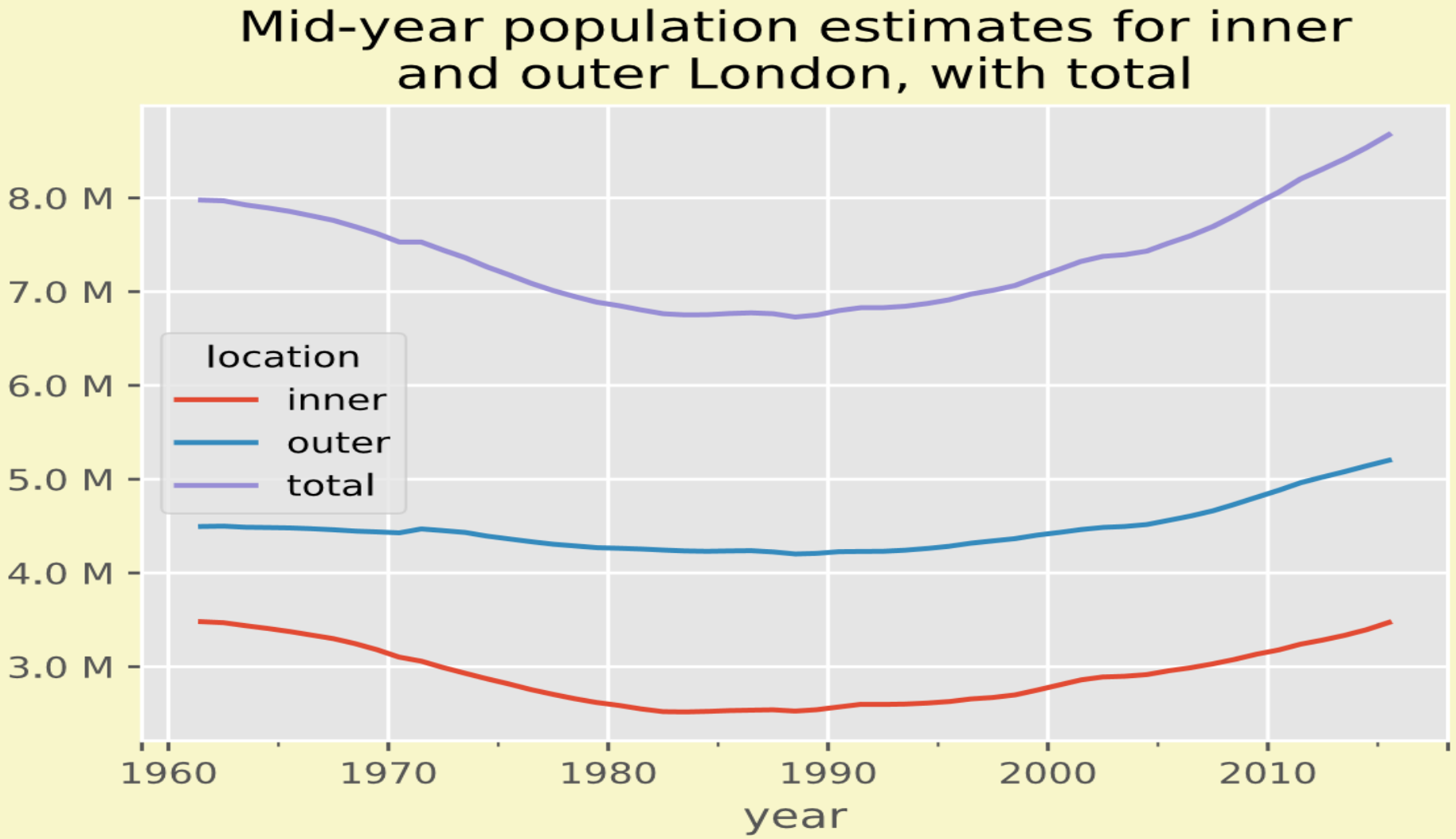
Aggregate inner/outer boroughs



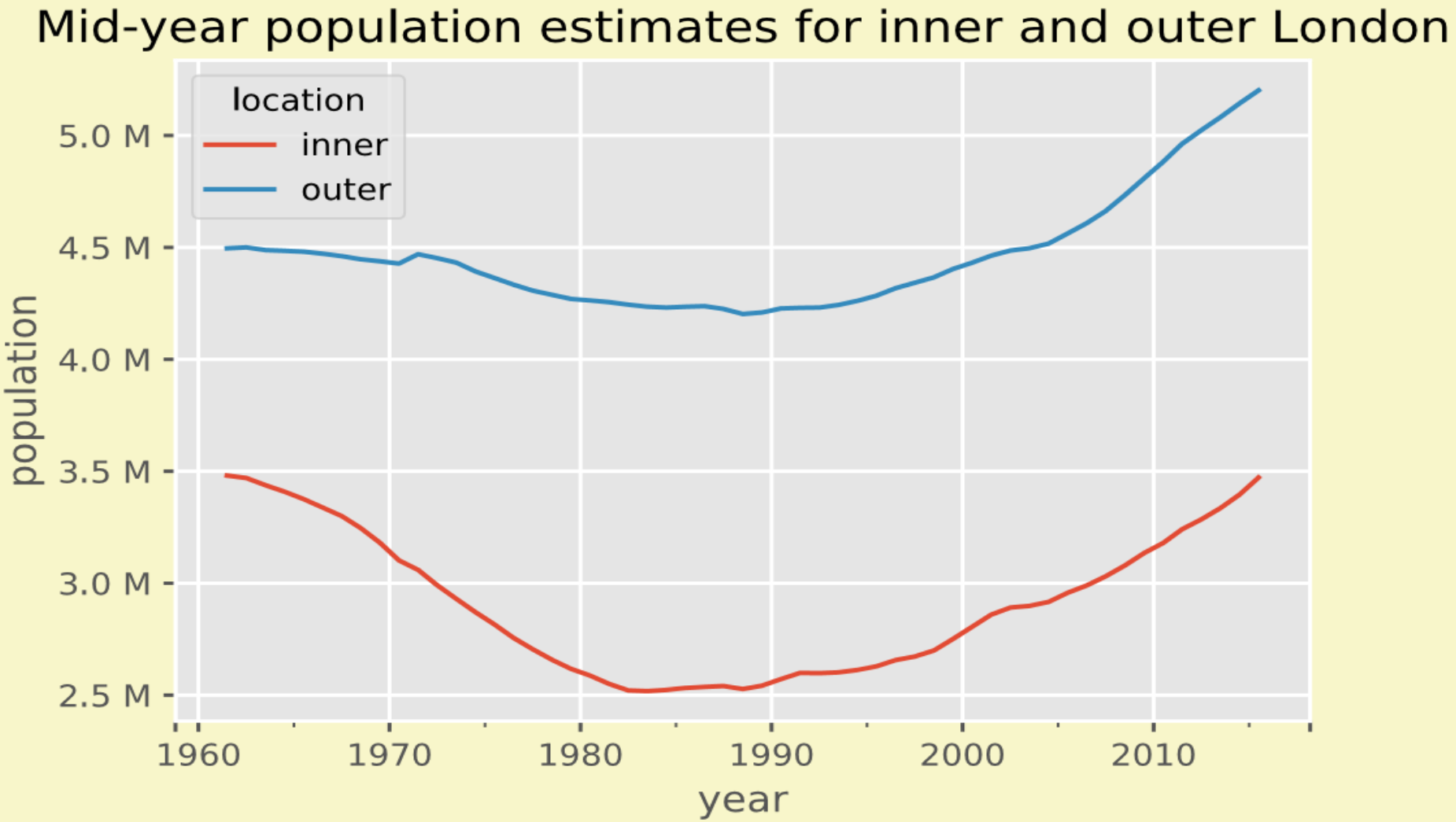
Comparing population density across boroughs



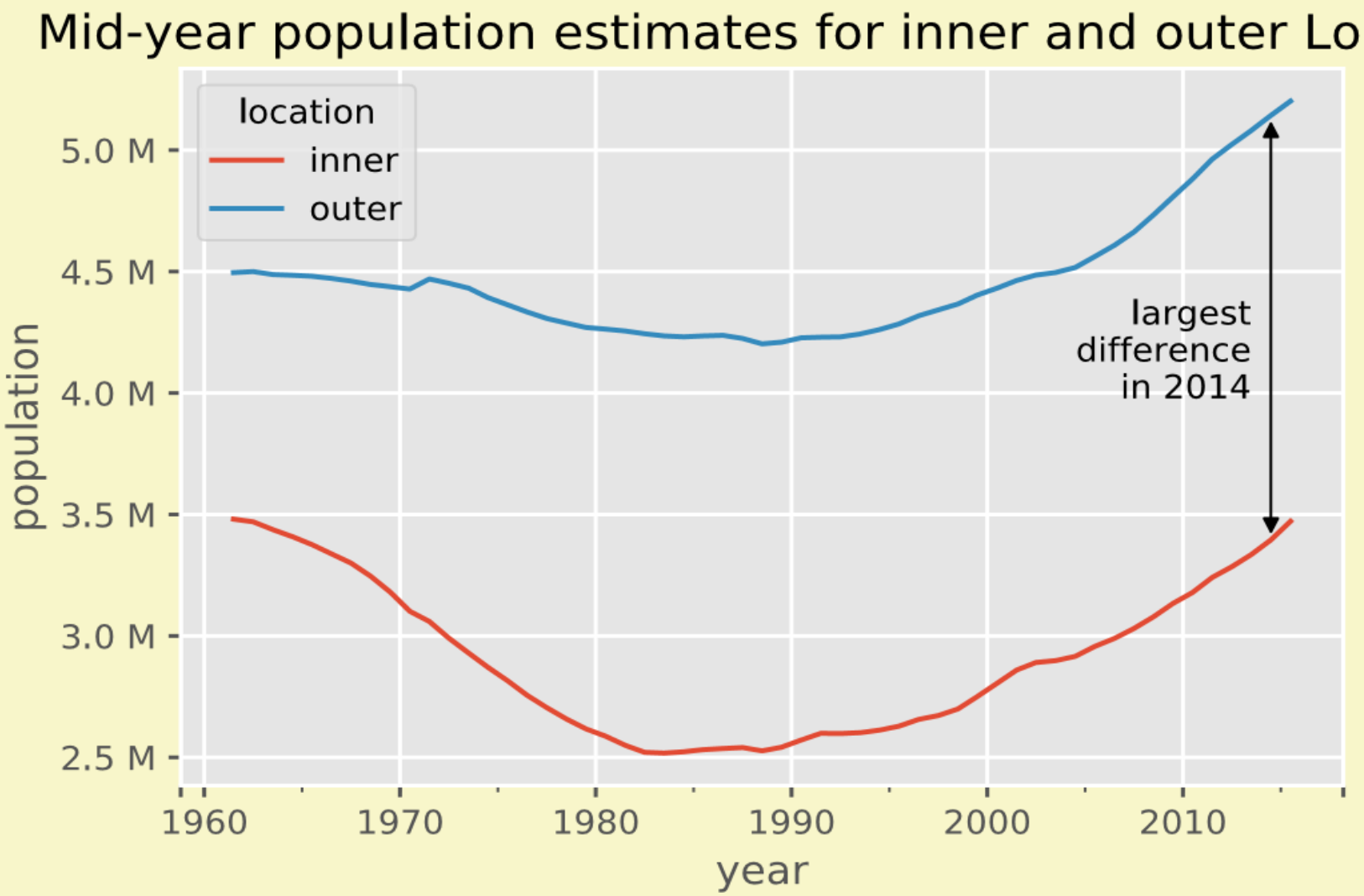
Summing variables



Difference between variables

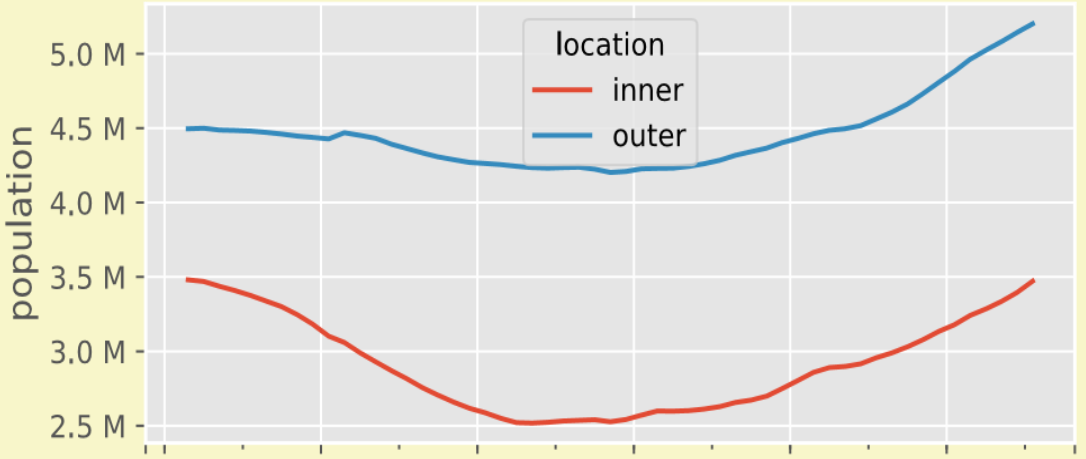


Difference between variables – biggest difference

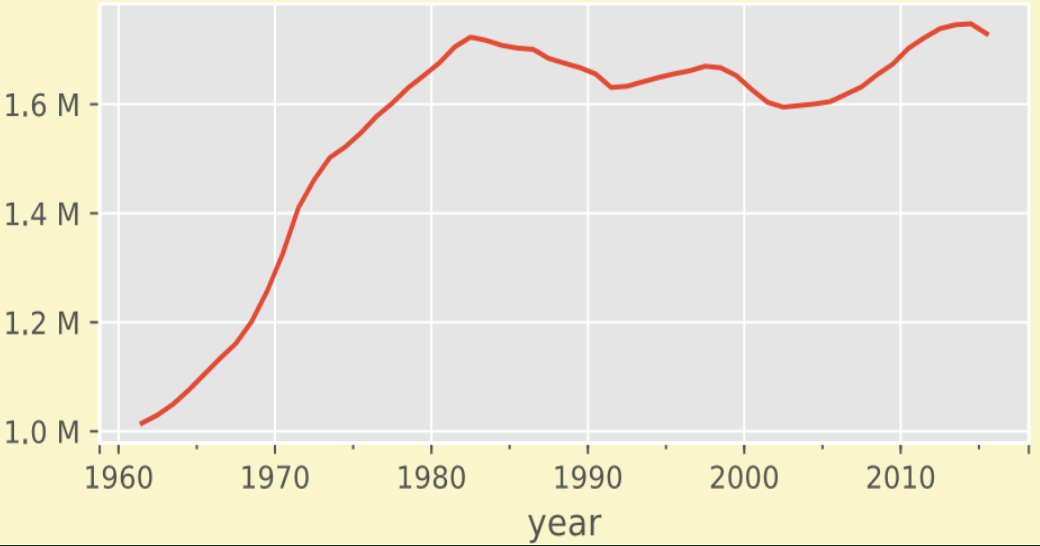


Difference between variables

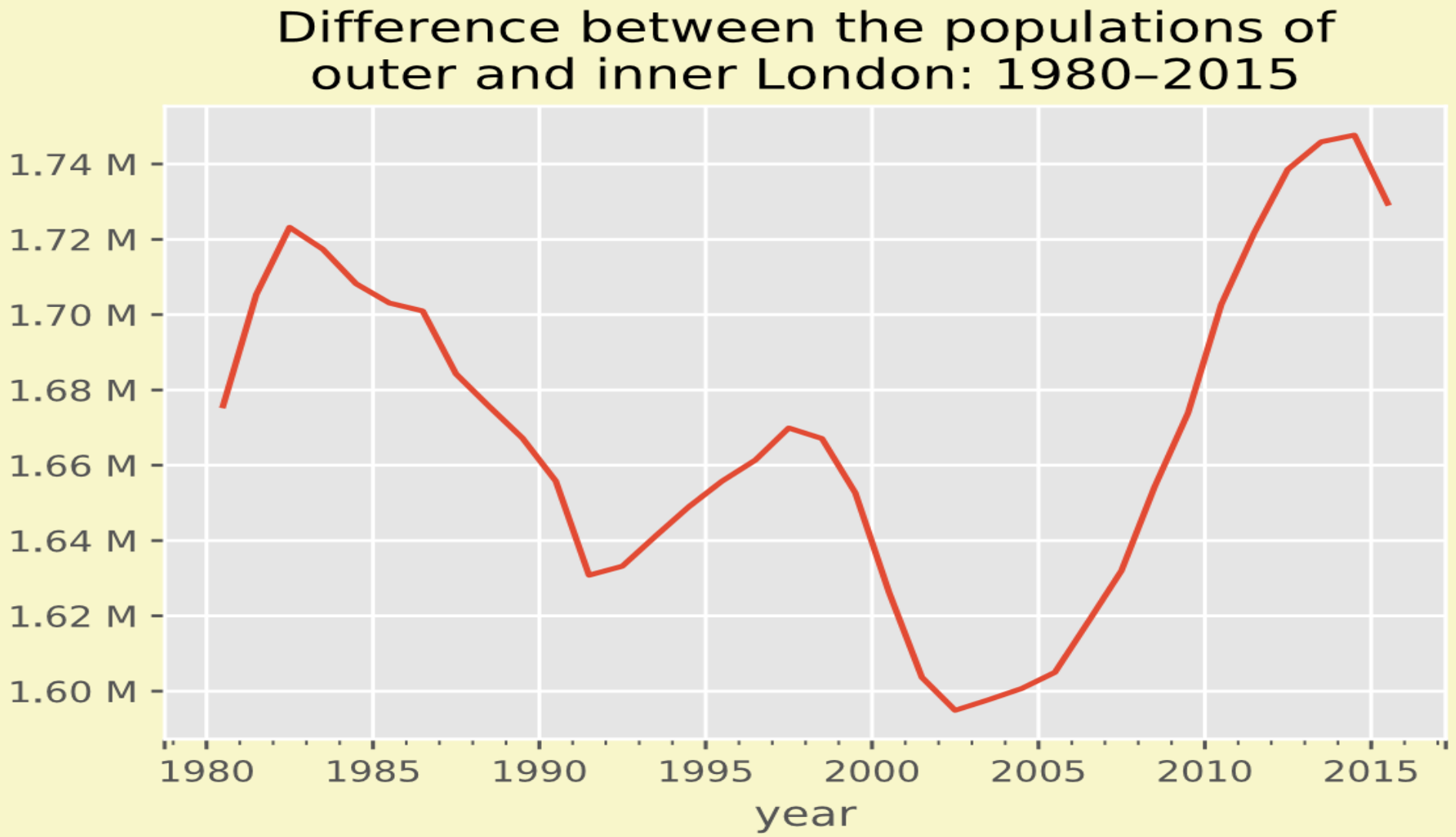
Mid-year population estimates for inner and outer London



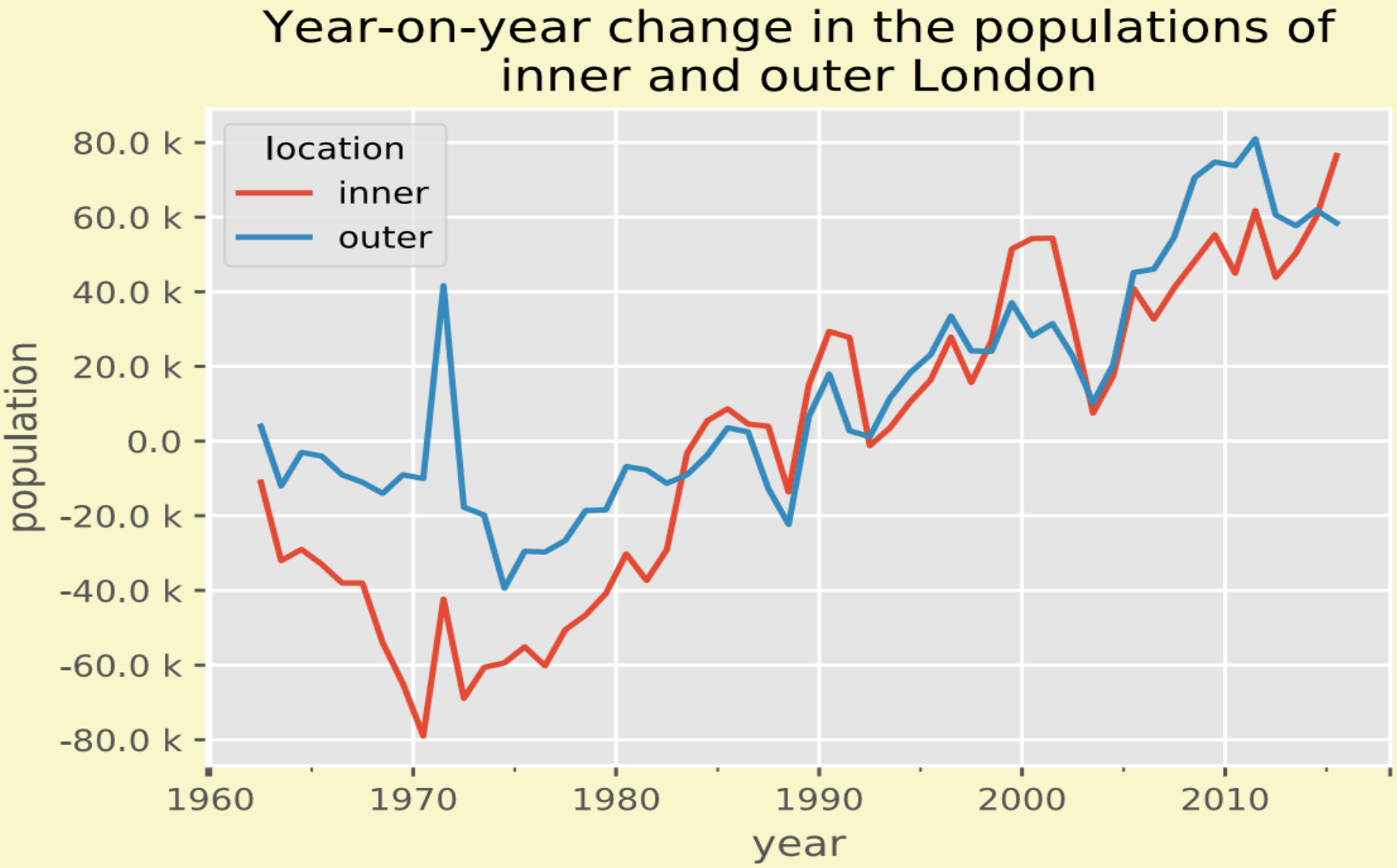
Difference between the populations of outer and inner London



Difference between variables



Difference between variables



Year-on-year change