Secondary Data in Education Research

Access to Education Data in England

Dr Yiyang Gao

Durham University Evidence Centre for Education

Epiphany Term Jan 11 2024

Lectures

Prof Nadia Siddiqui

Prof Stephen Gorard

Workshops

Keji Fan

Yiyang Gao

Logistics

Date	Topic	Instructor	Location
2024.01.11	Access education data in England	Yiyang	D/W007
2024.01.18	OECD PISA	Keji	D/W007
2024.01.25	National Pupil Database (NPD)	Yiyang	D/W007
2024.02.01	TALIS and PRILS	Keji	D/W007
2024.02.08	Longitudinal data	Keji	D/W007

Logistics

- Please bring a laptop for data analysis.

Logistics

- Please bring a laptop for data analysis.
- Laptop loan service available at Bill Bryson Library via the self-service laptop locker on Level 2.

- Thursdays 9:00 – 10:45
- 30 minutes introduction
- 45 minutes practice
- 30 minutes Q&A

Logistics

- Please bring a laptop for data analysis.
- Laptop loan service available at Bill Bryson Library via the self-service laptop locker on Level 2.

- Thursdays 9:00 – 10:45
- 30 minutes introduction
- 45 minutes practice
- 30 minutes Q&A

Logistics

- Please bring a laptop for data analysis.
- Laptop loan service available at Bill Bryson Library via the self-service laptop locker on Level 2.

- Thursdays 9:00 – 10:45
- 30 minutes introduction
- 45 minutes practice
- 30 minutes Q&A

Logistics

- Attendance to workshops is not mandatory - but highly recommended!

Logistics

- Attendance to workshops is not mandatory - but highly recommended!
- No prerequisite knowledge of programming required

Logistics

- Attendance to workshops is not mandatory - but highly recommended!
- No prerequisite knowledge of programming required
- Data in England and EU countries

Logistics

- Attendance to workshops is not mandatory - but highly recommended!
- No prerequisite knowledge of programming required
- Data in England and EU countries
- How to find and access to the data of interest

Logistics

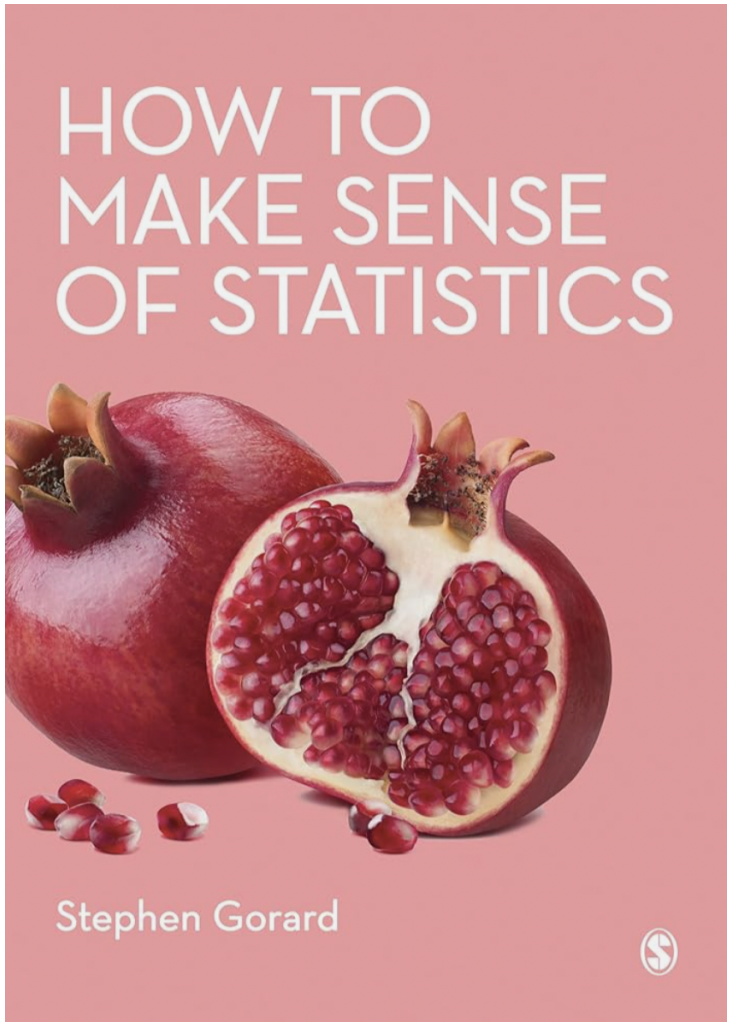
- Attendance to workshops is not mandatory - but highly recommended!
- No prerequisite knowledge of programming required
- Data in England and EU countries
- How to find and access to the data of interest
- Office hours and enquiry via emails

Logistics

- To produce a high-quality dissertation on secondary data analysis, you will need...

Logistics

- Statistics 101

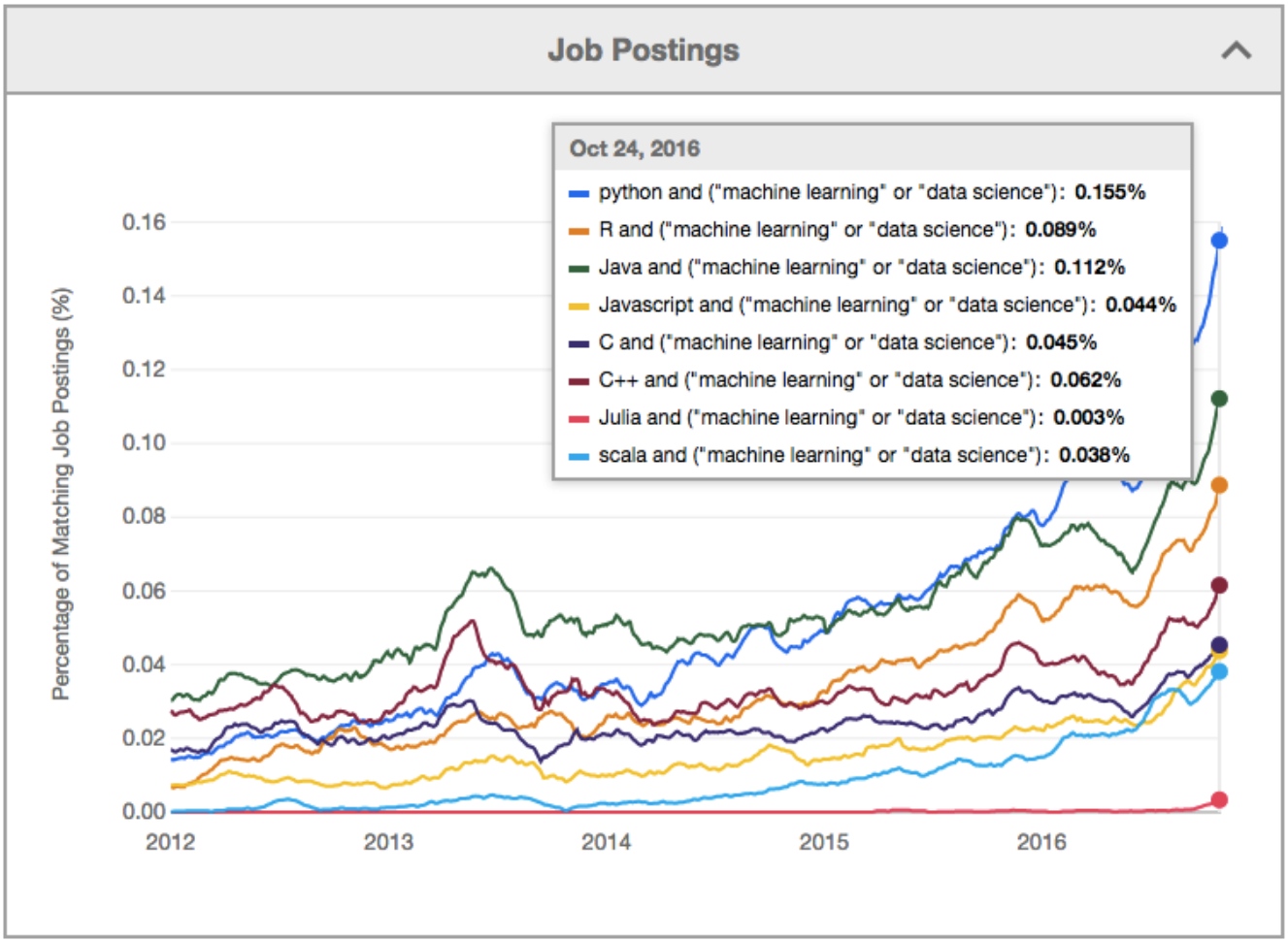


Logistics

- Data source

Logistics

- Software/programming languages



Source: CEAT

Outline

- Background
- Open access data: GOV.UK
- Summary

The National Curriculum in England

- The national curriculum is organised into blocks of years called ‘key stages’ (KS). At the end of each key stage, the teacher will formally assess your child’s performance.
- https://www.gov.uk/national-curriculum

The National Curriculum in England

- The national curriculum is organised into blocks of years called ‘key stages’ (KS). At the end of each key stage, the teacher will formally assess your child’s performance.
- https://www.gov.uk/national-curriculum

The National Curriculum in England

Table 1: Organisation of Schools in England

Stage	Year/Grade	Typical age	Type of Institution
Early Years Foundation Stage	Pre-school and nursery education	0 – 5	Pre-school settings
	Reception	4 – 5	Primary school
Key Stage 1	Year 1	5 – 6	
	Year 2	6 – 7	
Key Stage 2	Year 3	7 – 8	
	Year 4	8 – 9	Primary school; Middle school
	Year 5	9 – 10	
	Year 6	10 – 11	
Key Stage 3	Year 7	11 – 12	Secondary school; Middle school
	Year 8	12 – 13	
	Year 9	13 – 14	Secondary school
Key Stage 4	Year 10	14 – 15	
	Year 11	15 – 16	
Post compulsory	Year 12	16 – 17	Secondary school; 6th Form college; Further Education college
	Year 13	17 – 18	

Source: Semantic Scholar

The National Curriculum in England

Age	England	Scotland	Wales	Northern Ireland
3-4	Foundation Stage		Foundation Phase	
4-5				
5-6	Key Stage 1	Primary 1		Foundation
6-7		Primary 2		Key Stage 1
7-8	Key Stage 2	Primary 3	Key Stage 2	
8-9		Primary 4		Key Stage 2
9-10		Primary 5		
10-11		Primary 6		
11-12	Key Stage 3	Primary 7	Key Stage 3	Key Stage 3
12-13		Secondary 1		
13-14		Secondary 2		
14-15	Key Stage 4	Secondary 3	Key Stage 4	Key Stage 4
15-16		Secondary 4		
16-17	Key Stage 5	Secondary 5	Key Stage 5	Key Stage 5
17-18		Secondary 6		

Source: Curriculum-exams

UK Geography

Regions of England

Source: Maps of regions

Data Wrangling

Source: Data Wrangling

GOV.UK

- Explore our statistics and data

GOV.UK

- Explore our statistics and data
- https://explore-education-statistics.service.gov.uk/
- Go to Data Catalogue

GOV.UK

- Explore our statistics and data
- https://explore-education-statistics.service.gov.uk/
- Go to Data Catalogue

GOV.UK

- Download the dataset of Key Stage 4 Performance

GOV.UK

Select a theme

- ○ Children's social care
- ○ COVID-19
- ○ Destination of pupils and students
- ○ Early years
- ○ Finance and funding
- ○ Further education
- ○ Higher education
- ○ Pupils and schools
- ◉ School and college outcomes and performance
- ○ Teachers and school workforce

Select a publication

- ○ A level and other 16 to 18 results
- ○ Key stage 1 and phonics screening check attainment
- ○ Key stage 2 attainment
- ○ Key stage 2 attainment: National headlines
- ◉ Key stage 4 performance
- ○ Level 2 and 3 attainment age 16 to 25
- ○ Multi-academy trust performance measures at key stage 2
- ○ Multi-academy trust performance measures (Key stages 2, 4 and 5)
- ○ Multiplication tables check attainment

GOV.UK

- Select "Academic year 2022/23"
- Download the Local authority data (csv, 1 Mb)

GOV.UK

- Decompress the zip file
- What can you find in this folder?

GOV.UK

- Decompress the zip file
- What can you find in this folder?

Preparation

- Can you please answer the questions?
  - Q1: Where is the codebook?
  - Q2: What is covered in this dataset?
  - Q3: What do the following variables mean?

Preparation

Variable name	Variable description
avg_att8	
avg_p8score	
t_pupils	
t_schools	

Preparation

Variable name	Variable description
avg_att8	Average Attainment 8 score of all pupils
avg_p8score	Average Progress 8 score of all pupils
t_pupils	Total number of pupils at the end of key stage 4
t_schools	Total number of schools

Preparation

- Now open the dataset in Excel

Analysis

- Q4: In **2022/23** academic year, how many state-funded schools are there in each region? And how many students in each region?

Analysis

- Q4: In the 2022/23 academic year, how many state-funded schools are there in each region? And how many students in each region?

- Hint: related variables: **time_period**, **geographic_level**
- Hint: Use the **filter** function in Excel

Analysis

time_period	region_name	t_schools	t_pupils
202223	North East	193	27,722
202223	North West	568	83,315
202223	Yorkshire and The Humber	383	61,024
202223	East Midlands	351	52,761
202223	West Midlands	470	68,236
202223	East of England	451	68,373
202223	London	612	89,791
202223	South East	627	96,417
202223	South West	392	56,009
202223	Inner London	226	30,224
202223	Outer London	386	59,567

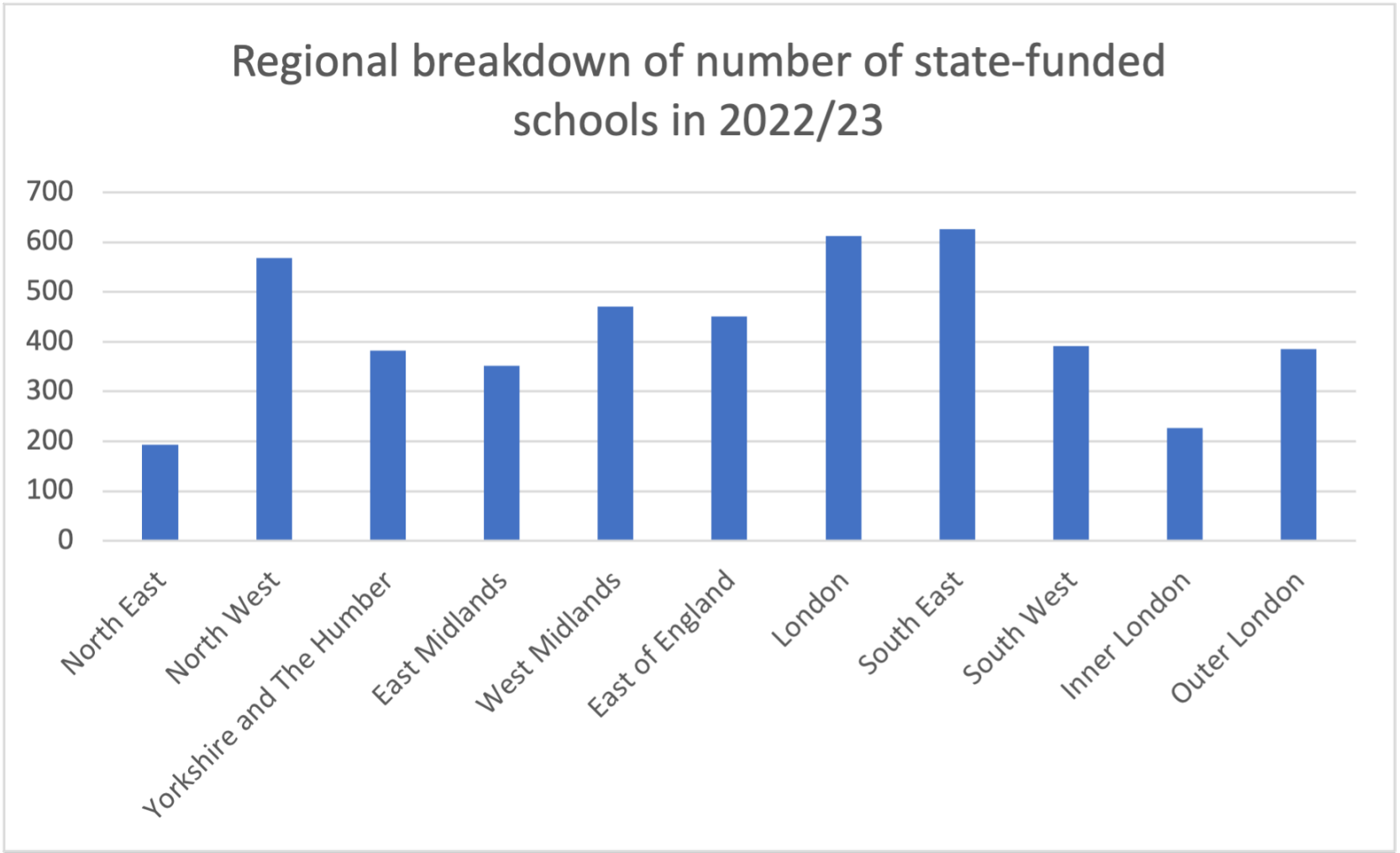
Analysis

- Q5: Can you please produce two charts to show the regional breakdown of number of schools and students separately?

Analysis

- Select the regions and the number of schools
- Go to **Insert**
- What type of chart do you need?

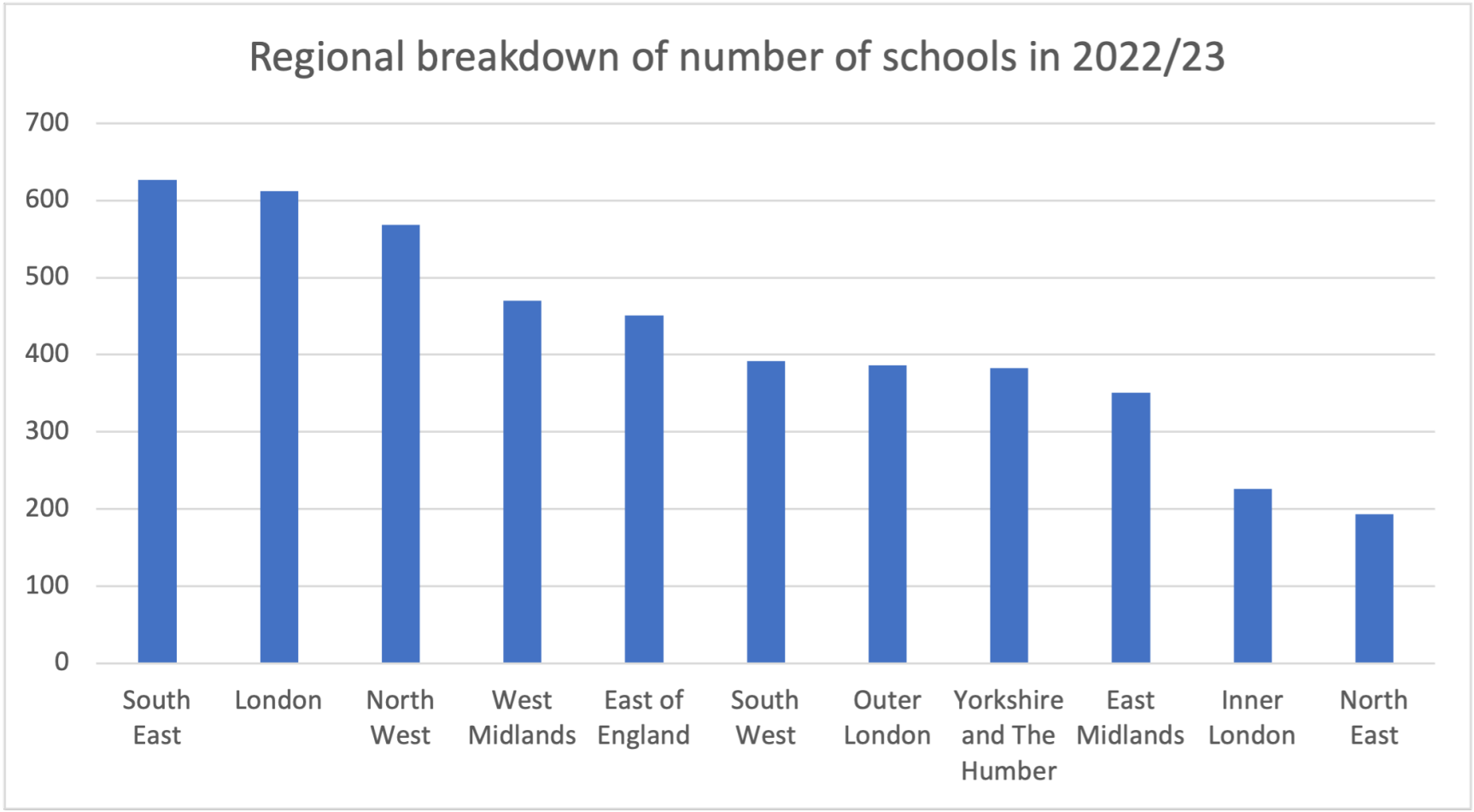
Analysis


Regional breakdown of number of state-funded schools in 2022/23
700
600
500
400
300
200
100
0
North East
North West
Yorkshire and The Humber
East Midlands
West Midlands
East of England
London
South East
South West
Inner London
Outer London

Analysis

- Which regions has the most schools?

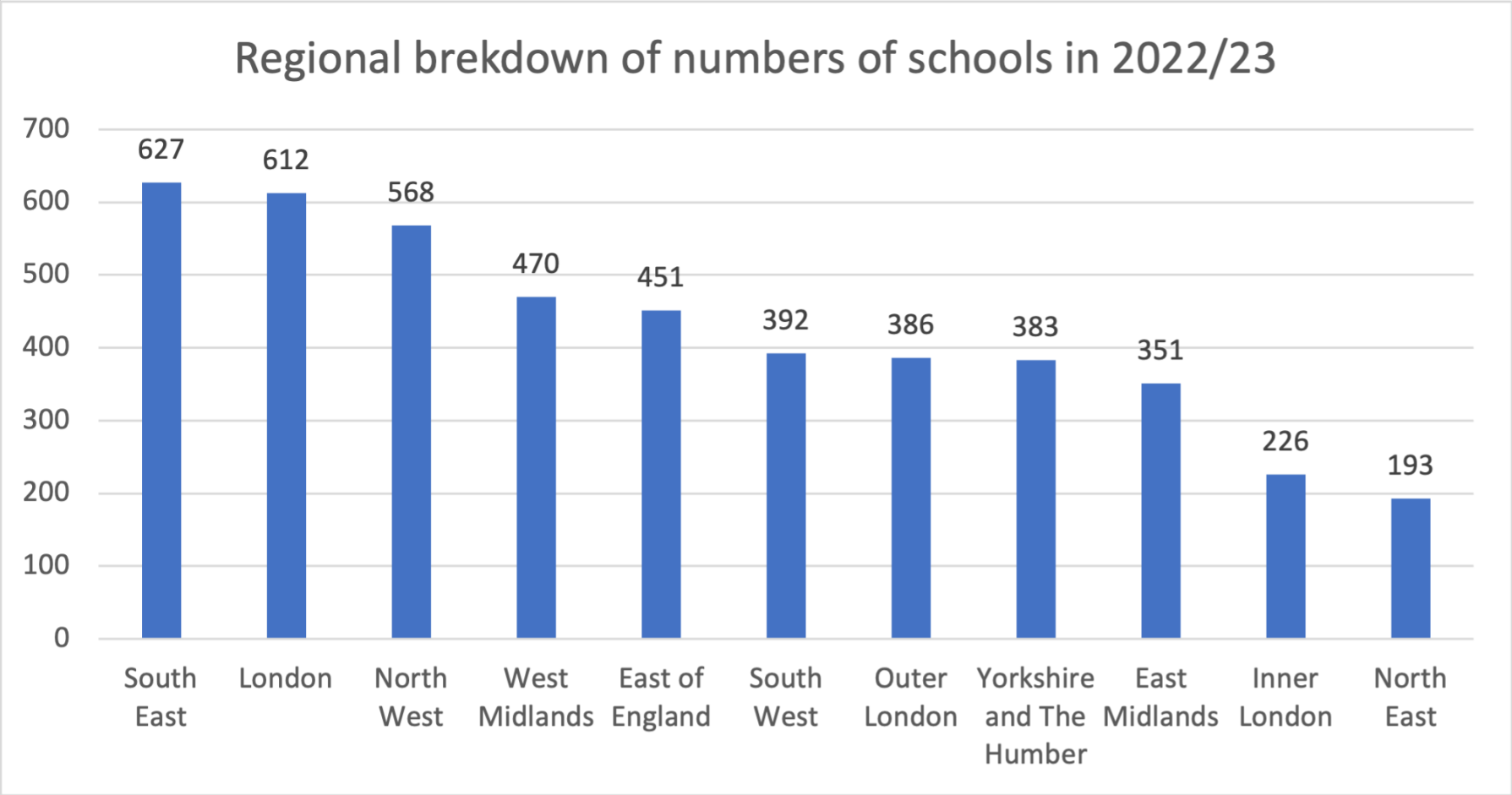
Analysis


Regional breakdown of number of schools in 2022/23
700
600
500
400
300
200
100
0
South East
London
North West
West Midlands
East of England
South West
Outer London
Yorkshire and The Humber
East Midlands
Inner London
North East

Analysis

- Add data labels

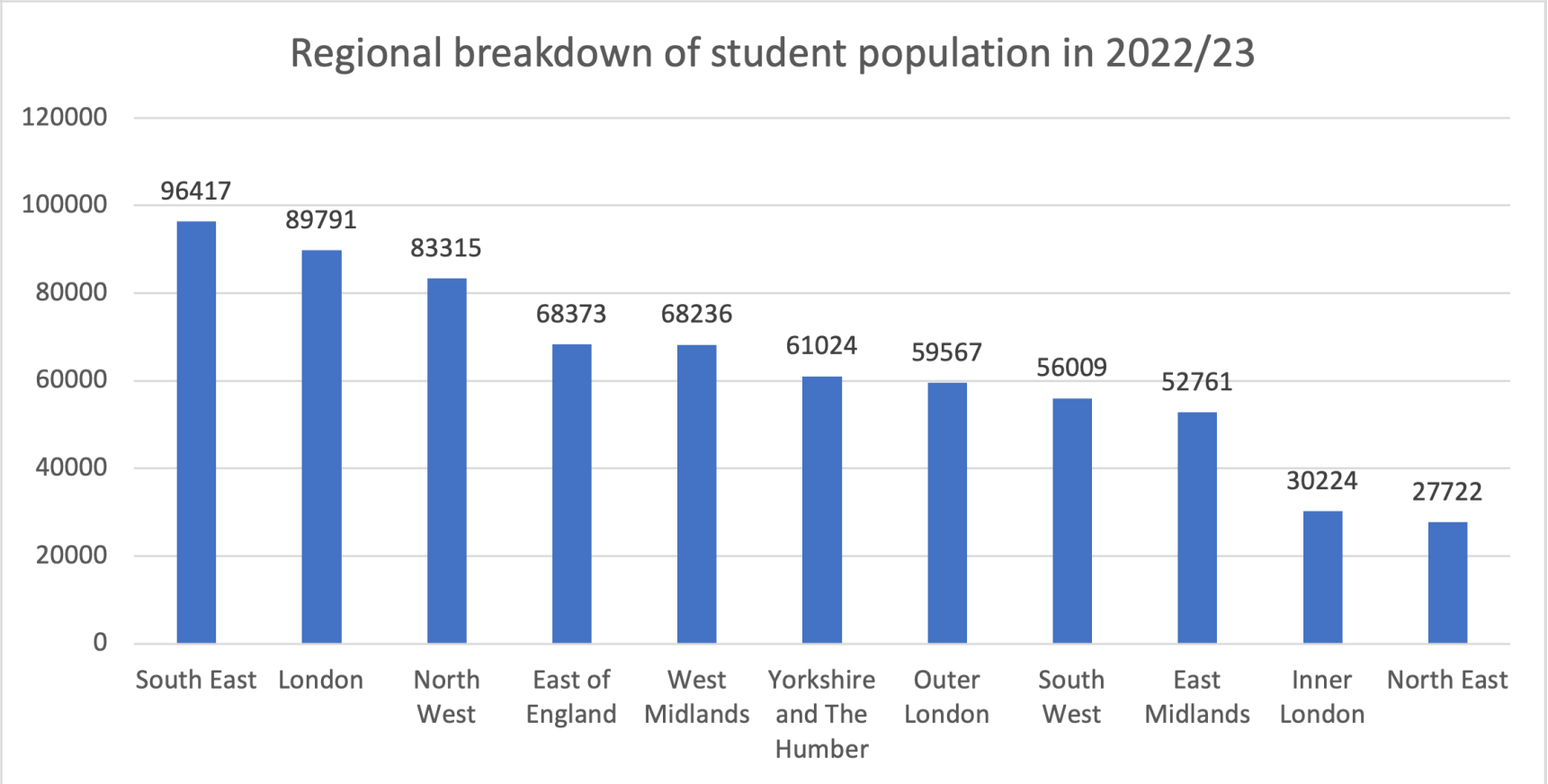
Analysis


Regional brekdown of numbers of schools in 2022/23
700
600
500
400
300
200
100
0
627
612
568
470
451
392
386
383
351
226
193
South East
London
North West
West Midlands
East of England
South West
Outer London
Yorkshire and The Humber
East Midlands
Inner London
North East

Analysis

- Do the same for the number of students...

Analysis


Regional breakdown of student population in 2022/23
120000
100000
80000
60000
40000
20000
0
96417
89791
83315
68373
68236
61024
59567
56009
52761
30224
27722
South East
London
North West
East of England
West Midlands
Yorkshire and The Humber
Outer London
South West
East Midlands
Inner London
North East

Analysis

- Now let's relax the limitation on time span...
- What if we have data of multiple years?

Analysis

- Now let's look at the regional disparities in academic attainment, and how it evolves over time.

Analysis

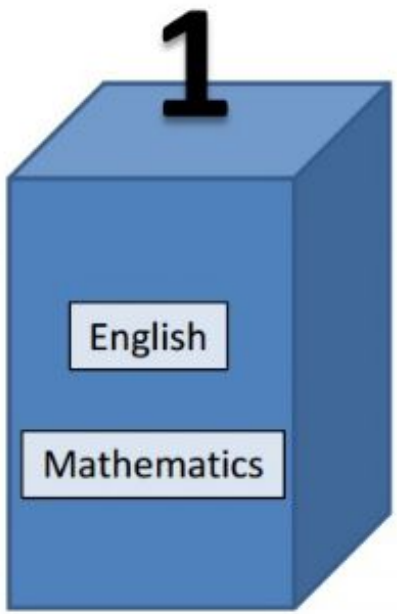
- Now let's look at the regional disparities in academic attainment, and how it evolves over time.
- Proxy of academic attainment: Attainment 8 and Progress 8

Attainment 8

- Attainment 8 is a way of measuring how well pupils do in key stage 4, which they usually finish when they are 16 years old. The 8 subjects are:
  - English and maths
  - 3 subjects from qualifications that count towards the English Baccalaureate (EBacc), like sciences, language and history
  - 3 more GCSE qualifications (including EBacc subjects) or technical awards from a list approved by the Department for Education.
  - Each grade a pupil gets is assigned a point score from 9 (the highest) to 1 (the lowest). Each pupil's Attainment 8 score is calculated by adding up the points for their 8 subjects, with English and maths counted twice.

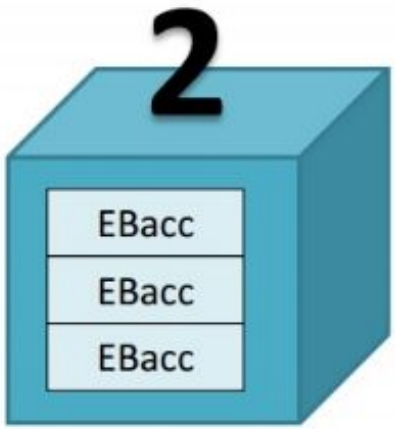
Source: GOV.UK

Attainment 8



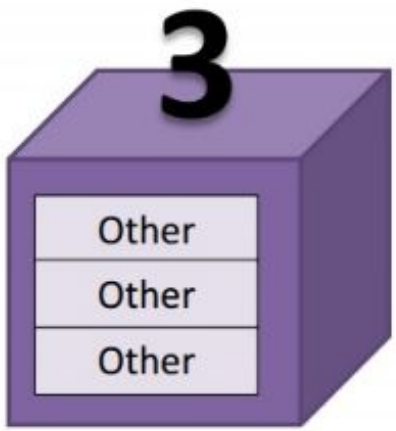
Bucket 1

- One slot for English and one for maths; double-weighted



Bucket 2

- Three EBacc qualifications
- (Sciences, computer sciences, geography, history or languages)



Bucket 3

- Three “other” slots
- Any remaining Ebacc qualifications
- Other approved academic, arts or vocational qualifications

Source: High Lane School

Progress 8

- Progress 8 tells you about the progress that pupils in a school make from the end of primary school to the end of year 11.
- It is a type of **value-added** measure, which means that pupils' results are compared to other pupils nationally with similar starting points.

Progress 8

- First, pupils' results in 8 qualifications (same as those used in Attainment 8) are converted into a point score.
- The school's average point score for these qualifications is then calculated. This is compared to the national average for pupils with similar prior attainment at key stage 2.
- A school's Progress 8 score is the average of its pupils' scores. It shows if, on average, pupils in the school are making above or below expected progress compared to pupils with similar starting points nationally.

Analysis

- Q6: Can you produce a table to show the descriptive statistics for the academic attainment (use the average Attainment 8 scores) for each region over time?

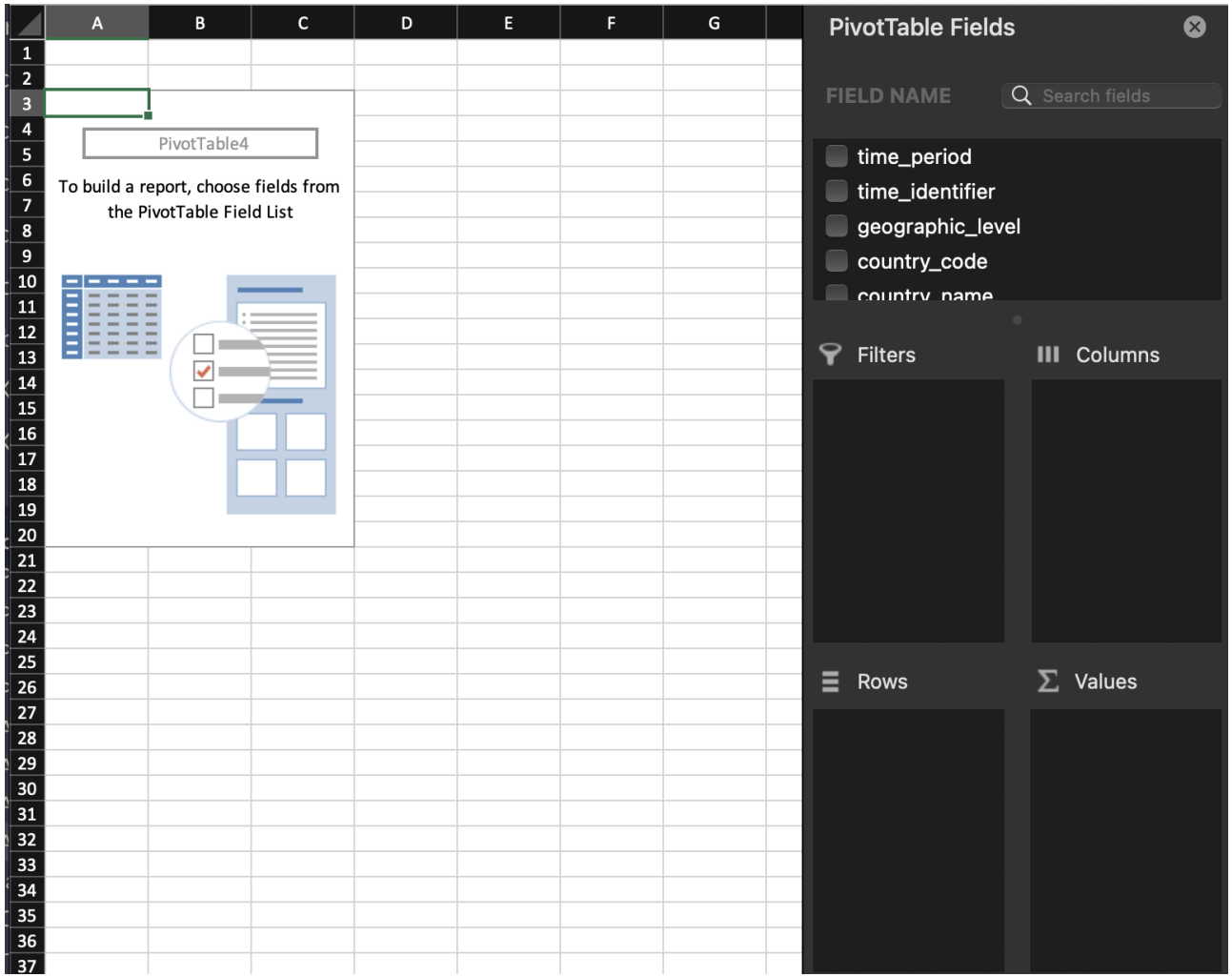
Analysis

- We can do this with Excel Formulas:
  - Minimum: =MIN()
  - 25th Percentile (Quan1): =QUARTILE.INC(, 1)
  - Median: =MEDIAN()
  - Mean: =AVERAGE()
  - Maximum: =MAX()
  - Mean: =AVERAGE()
  - 75th Percentile (Quan3): =QUARTILE.INC(,3)

Analysis

- Or, Pivot Table can quickly calculate summary statistics

Analysis


PivotTable4
To build a report, choose fields from the PivotTable Field List
PivotTable Fields
FIELD NAME
Search fields
time_period
time_identifier
geographic_level
country_code
Filters
Columns
Rows
Values

Analysis

- Step 1: Copy **time_period**, **region_name**, **avg_att8** columns into a new spreadsheet
- Step 2: Select the whole table
- Step 3: Go to **Insert** and choose "pivot_table"
- Step 4: Drag **region_name** into **Rows**
- Step 5: Drag **avg_att8** into **Values**
- Step 6: Click the small *i* in the **Values** panel and choose **Min**.
- Step 7: Drag **avg_att8** into **Values** again and choose **Average**
- Step 8: Drag **avg_att8** into **Values** again and choose **Max**

Analysis

Row Labels	Min. of avg_att8	Average of avg_att8	Max. of avg_att8	Var of avg_att8	StdDev of avg_att8
East Midlands	45.1	47.46	49.6	3.91	1.98
East of England	46.5	48.78	51	3.93	1.98
Inner London	48.4	51.14	53.4	4.27	2.07
London	49.7	52.04	54.1	3.47	1.86
North East	44.4	46.72	49.2	4.62	2.15
North West	44.5	47.16	49.6	4.79	2.19
Outer London	50.4	52.52	54.5	3.06	1.75
South East	47.4	49.8	52.1	4.23	2.06
South West	46.2	48.72	51.4	5.01	2.24
West Midlands	44.8	47.3	49.5	4.24	2.06
Yorkshire and The Humber	44.6	46.86	49.1	3.58	1.89
Grand Total	**44.4**	**48.95454545**	**54.5**	**7.54**	**2.75**

Sheet5 Sheet8 2223_la_data_provisional Sheet4 Sheet9 Sheet1 Sheet7 Sheet6

PivotTable Fields

FIELD NAME Search fields

- time_period
- region_name
- avg_att8

Filters

Columns

Values

Rows

region_name

Values

Min. of avg_att8
Average of avg_...
Max. of avg_att8
Var of avg_att8
StdDev of avg_...

Drag fields between areas

Table: Descriptive statistics for Attainment 8 by region.

Regions	Min	Mean	Max	Var	Std.Dev
East Midlands	45.1	47.46	49.6	3.91	1.98
East of England	46.5	48.78	51	3.93	1.98
Inner London	48.4	51.14	53.4	4.27	2.07
London	49.7	52.04	54.1	3.47	1.86
North East	44.4	46.72	49.2	4.62	2.15
North West	44.5	47.16	49.6	4.79	2.19
Outer London	50.4	52.52	54.5	3.06	1.75
South East	47.4	49.8	52.1	4.23	2.06
South West	46.2	48.72	51.4	5.01	2.24
West Midlands	44.8	47.3	49.5	4.24	2.06
Yorkshire and The Humber	44.6	46.86	49.1	3.58	1.89

Analysis

- Q7: How do you interpret the table? What does the variance and sd tell you?

Analysis

- Variance is the average of the squared differences from the Mean. The formula for the sample variance s^2 is:

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

where x_i is each value, $\overline{x}$ is the mean of the data set, and n is the number of data points.

Analysis

- Standard Deviation is the square root of the variance. It is a measure of the amount of variation or dispersion of a set of values.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$

where s^2 is the variance x_i is each value, $\overline{x}$ is the mean, , and n is the number of data points.

Analysis

- The variance and standard deviation reflect the consistency of the Attainment 8 scores over time. A high variance or standard deviation in regions such as the North West and South West suggests that there has been a wider range of average scores from year to year, indicating fluctuation in academic performance.
- Conversely, regions with lower variances and standard deviations, such as Outer London, suggest that the average academic performance has remained more consistent and stable over the years.

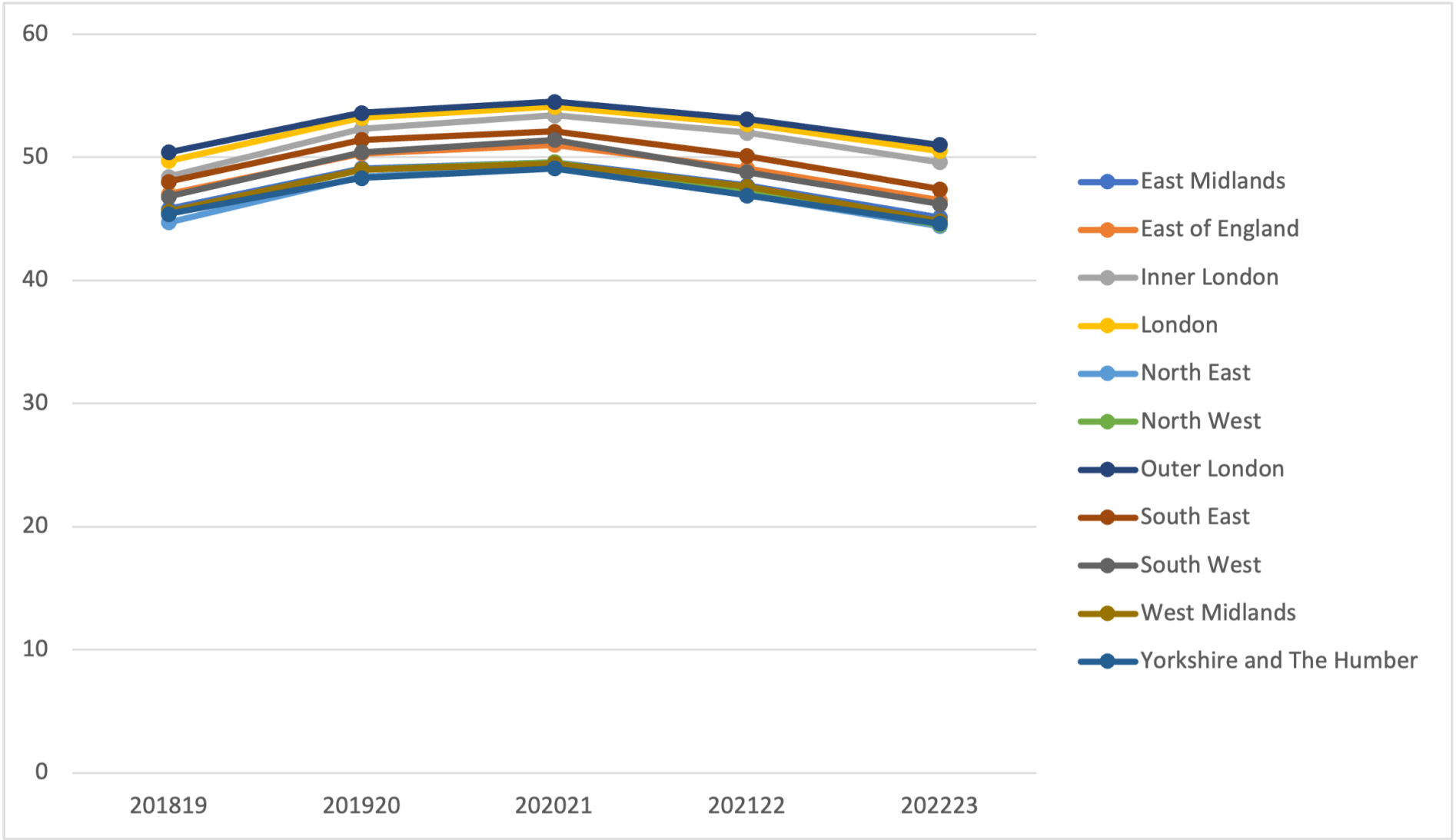
Analysis

- Q8: Can you visulise the the trend of regional differences in attainment 8 scores?

Analysis

- Step 1: Copy **time_period**, **region_name**, **avg_att8** columns into a new spreadsheet
- Step 2: Select the whole table
- Step 3: Go to **Insert** and choose "pivot_table"
- Step 4: Drag **time_period** into **Rows**
- Step 5: Drag **region_name** into **Columns**
- Step 6: Drag **avg_att8** into **Values** and choose "Average"
- Step 7: Select the whole table and insert a line chart

Analysis


60
50
40
30
20
10
0
201819
201920
202021
202122
202223
East Midlands
East of England
Inner London
London
North East
North West
Outer London
South East
South West
West Midlands
Yorkshire and The Humber

Take-home message

- National Curriculum in England
  - go
- Data Wrangling process
  - go
- KS4 data on GOV.UK
  - go
  - Long-Wide data transform
  - Descriptive statistics
  - Simple visualisation

The third session...

- Please download R and RStudio.

Icon for R

Icon for RStudio

For Windows Users

- Please visit CRAN and download R-4.3.2.exe for Windows.
- Please visit Posit and download RStudio IDE for Windows 10/11.

For Mac Users

- Please visit CRAN and download R-4.3.2.pkg for Mac
- Please visit Posit and download RStudio IDE for MacOS11+.

Thank you