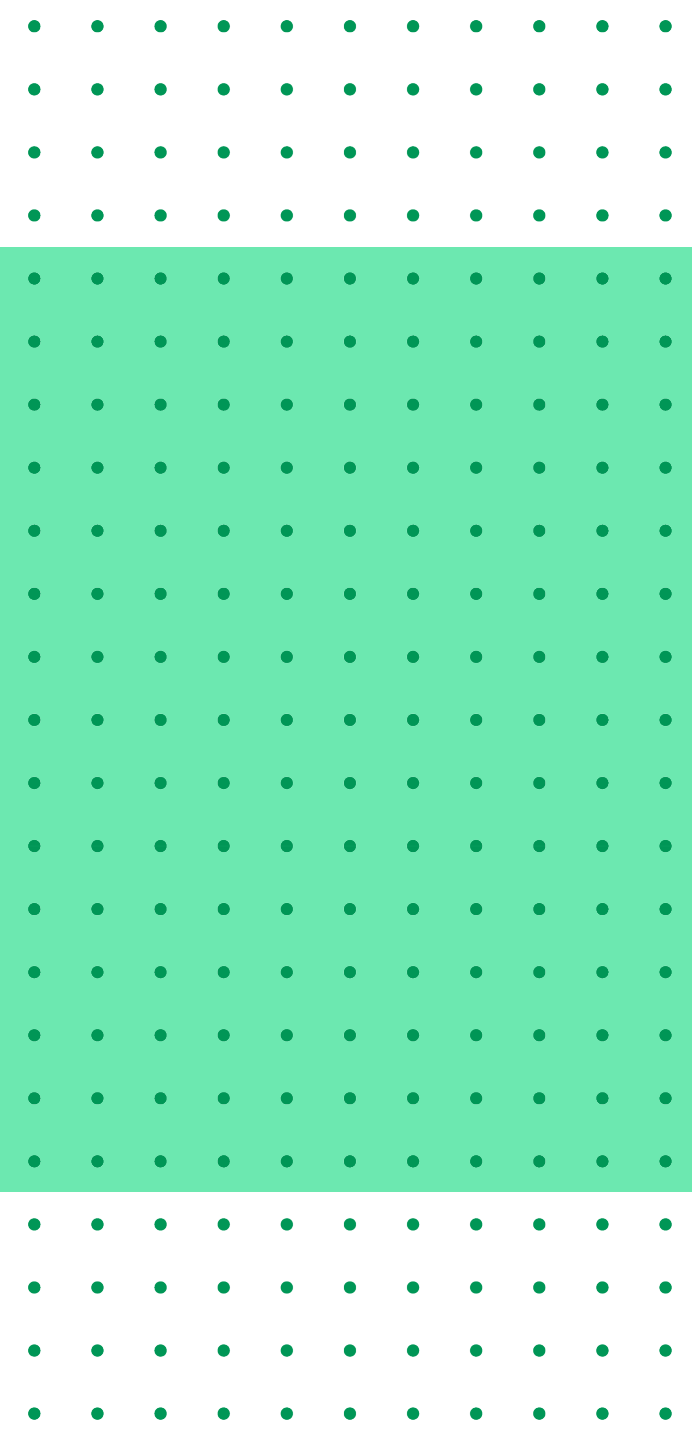


Data: where to find it, how to use it

DAVID KERNOHAN | ASSOCIATE EDITOR, WONKHE
@DKERNOHAN | DK@WONKHE.COM

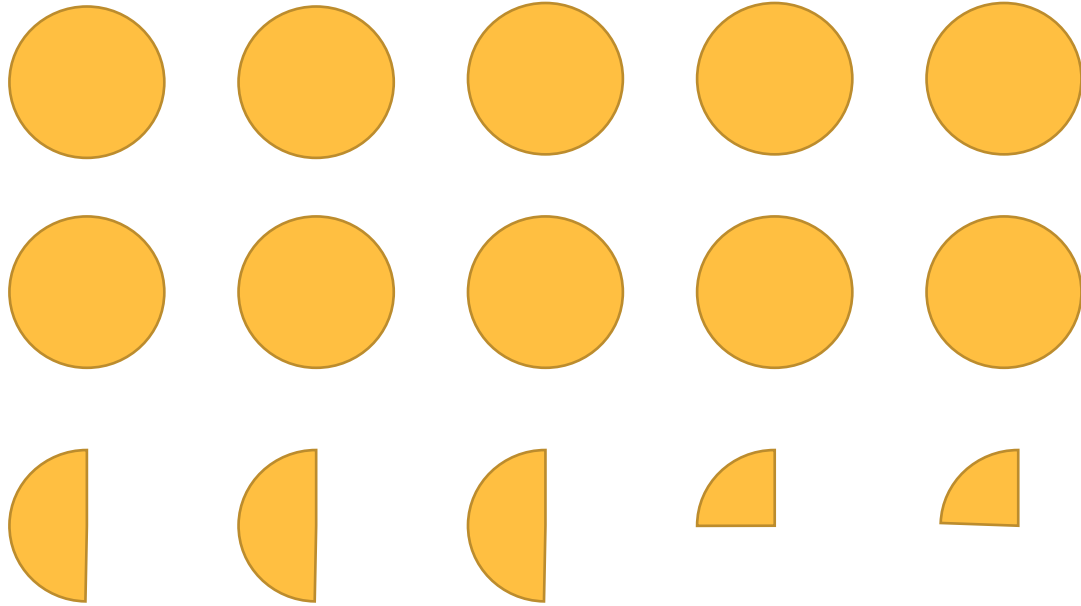
WONKHE



Data is

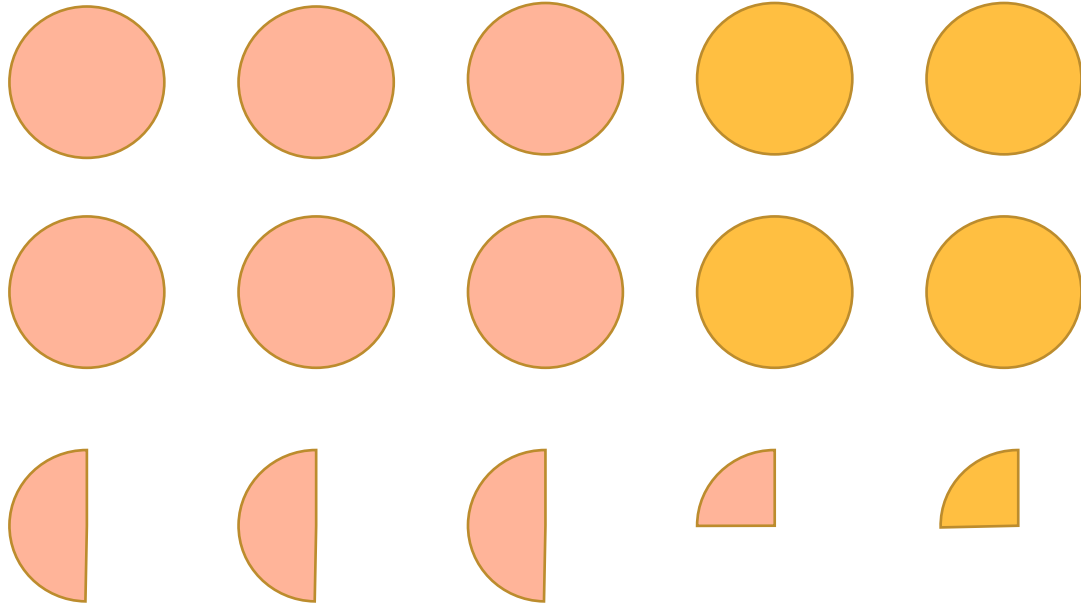
~~**Data are**~~

Counting students (I)



Students
15 (headcount)
=
12 (FTE)
10 (Full time)

Counting students (II)



Red students

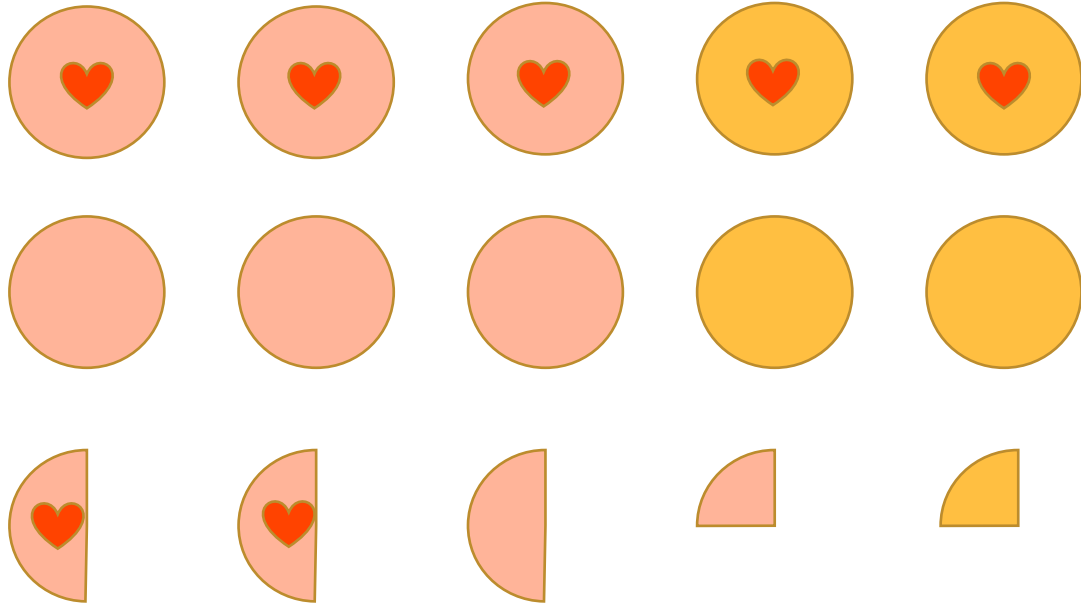
10 (headcount) = 66%

=

7.75 (FTE) = 65%

6 (Full time) = 60%

Counting students (III)



Lovely students

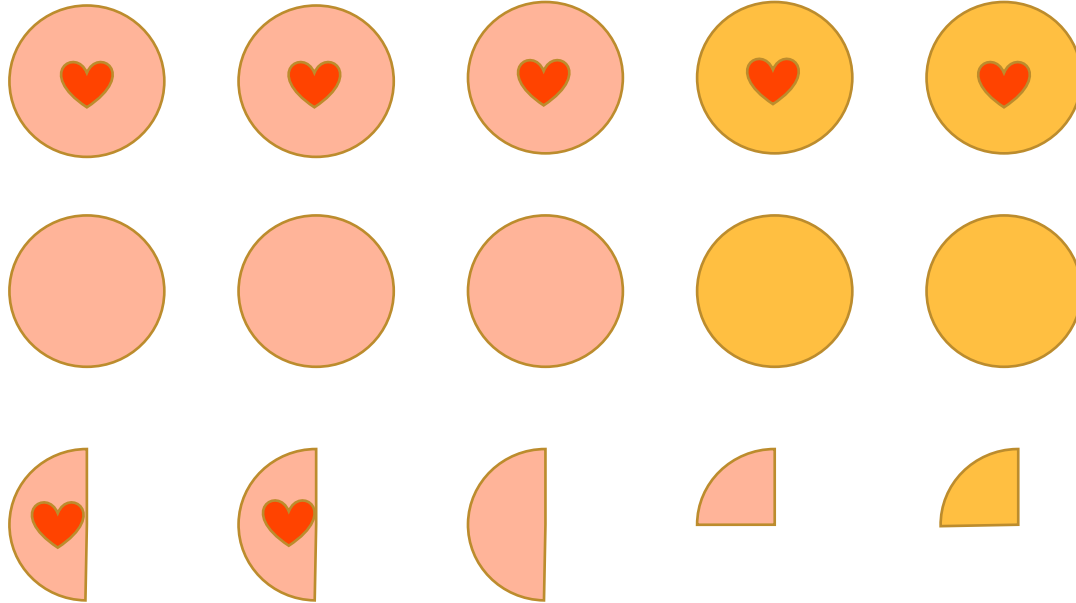
7 (headcount) = 47%

=

6 (FTE) = 50%

5 (Full time) = 50%

Counting students (IV)



Are red students more likely to be lovely?

(headcount)

50% of red students are lovely

40% of orange students are lovely

A 10 percentage point difference

(FTE)

51% of red students are lovely

47% of orange students are lovely

A 4 per centage point difference

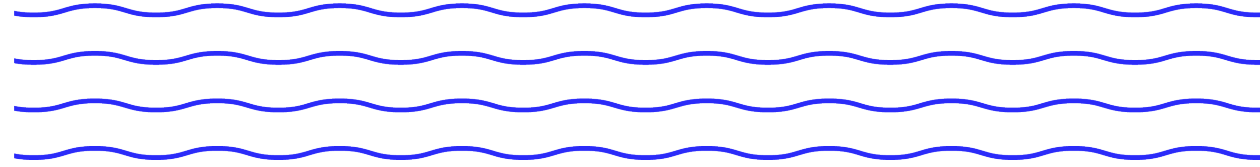
(Full time)

60% of red students are lovely

40% of orange students are lovely

A 20 percentage point difference

What is...



A student?

A course?

A subject?

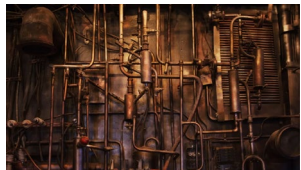
A provider?

Who counts?

Enrolled, placed, active?

UG, PG, FT, PT?

Fee-paying? Loan-taking?



Seeing study

Length – sandwich,
intercalation?

Level – UG, PG,
Foundation, HTQ?



.... or subject area

Who has decided? How?
Why?

CAH, JACS, HECOS, Cost
Centre?



And what is it like?

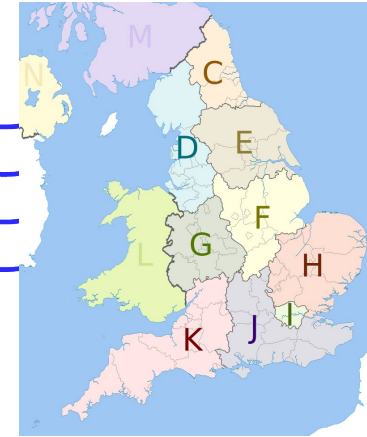
Registered?

Old HEFCE?

Degree-awarding?
Groups?



Key concepts



A year

12 months of pain

An academic year is September-August

A cohort is a group of students that all started in the same academic year.

It is NOT safe to assume all of a cohort is in the same year.

WONKHE | FEBRUARY 2022

A region

Brexit means Brexit

Official data (eg on recruitment, graduates) usually comes linked to NUTS1 (Now ITL1) codes: Scotland, Wales, NI, and 9 English regions.

Wales, Scotland, and NI are NOT regions!

Below that: UTLA and LTLA (and TTWA if you must)

UCAS Codes

AOO

Used because UCAS defines a provider differently – in practice mappable but not trivial to do.

The UKPRN

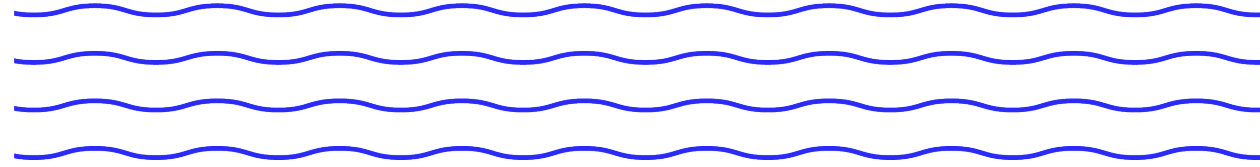
Eight figures of truth

Administered by the government (UKLRP)

Doesn't always deal well with mergers.

(Formerly “HESAcodes”)

Data sources: In a nutshell



HESA

Gold-standard data on students, graduates, staff, finances, estates, and business/community interaction.

UCAS

Data on each recruitment cycle – released at set points each year.

OfS, other regulators

Data used in regulation – often published as open data. Area based measures. Complex dashboards.

SLC

Data about student fee and maintenance loans.

ONS

Survey data on students and the wider workforce. Detailed data on place

DfE, other governments

LEO data on graduate destinations. Survey data – student expenditure and income survey.

League tables

Compiled from numerous data sources (including specially collected data).

Other sources

General medical council. Various think tanks. House of commons library. Other government departments.

Common pitfalls

DEFINITIONS

What the data is

The label is not enough:

- how is it counted?
- how credible is it?
- who does it include?

IDENTIFIERS

What it refers to

The name is not enough:

- Who does it refer to?
- Who did it refer to last time?
- What else do we know about them?

ROUNDING

Does $2+2 = 5$?

- does rounding make small figures unreliable?
- How are small figures presented?
- How are these used in proportions?

TIMING

When does it cover

- Academic year?
- Financial year (which one)?
- UCAS cycle?
- Cohort?

Statistics about (most*) undergraduate entrants

(* not in Scotland)



<https://www.ucas.com/data-and-analysis/>

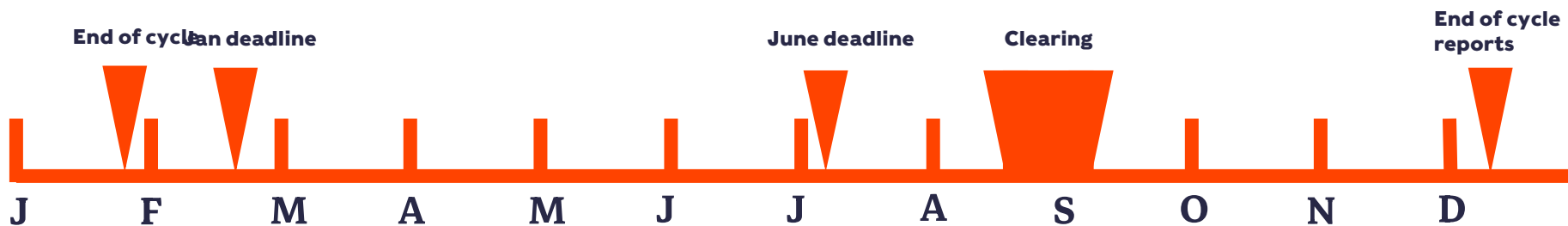
Presented as applications, applicants, placed students (main scheme, or all).

An institution is “something you apply to”

(sometimes) includes RPAs – self-notified by providers

Disproportionally represents young applicants

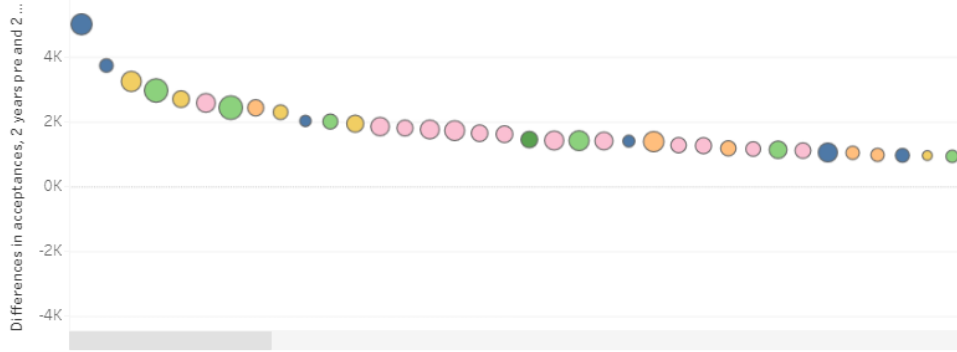
Scottish articulation issue



High level domicile: Region: Group: Highlight Name:

Difference in acceptances (2018 + 2019 vs 2020 + 2021)

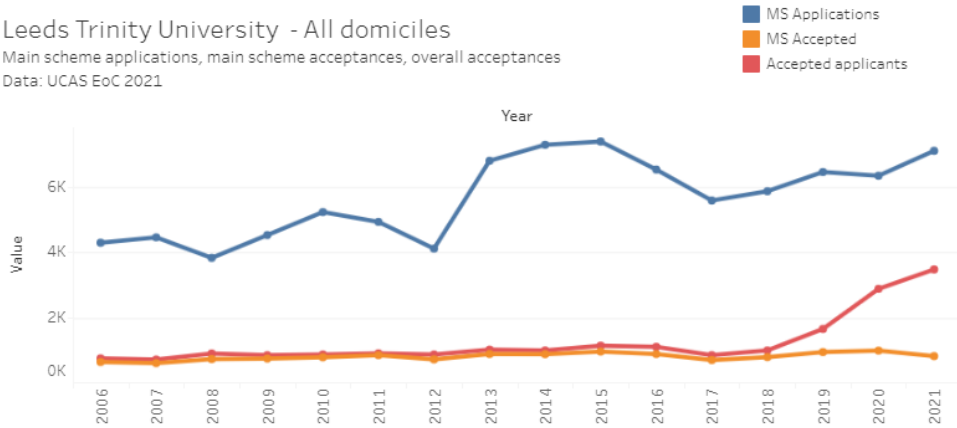
Domicile: UK, Data: UCAS EoC



Leeds Trinity University - All domiciles

Main scheme applications, main scheme acceptances, overall acceptances

Data: UCAS EoC 2021

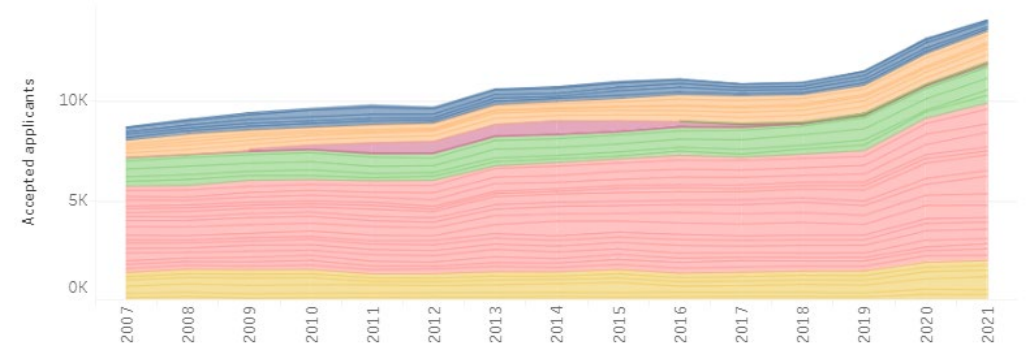


Highlight Name: JACS3 subject group: High level domicile:

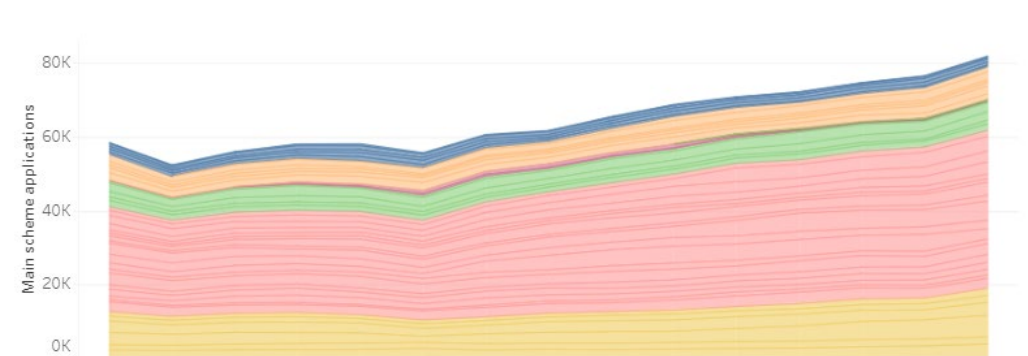
Acceptances: M Law

Domicile: All

Data: UCAS EoS



MS applications: M Law



<https://wonkhe.com/blogs/ucas-end-of-cycle-2021-provider-data/>

Statistics about fee and maintenance loans



Student Loans Company

<https://www.gov.uk/government/organisations/student-loans-company/about/statistics>

Two main series:

- Student loans (outlays, repayments, borrower activity)
- Student support (amounts by provider)

Doesn't separate out new loans from existing ones. Contains a lot of historic loans.

Sometimes you get in year data on withdrawals – but you need to bear in mind what this indication means.



Academic Year 2019 2020

Highlight Name

Average fee and maintenance loans - 2020

Paid to FT UG England-domiciled students

Data: SLC

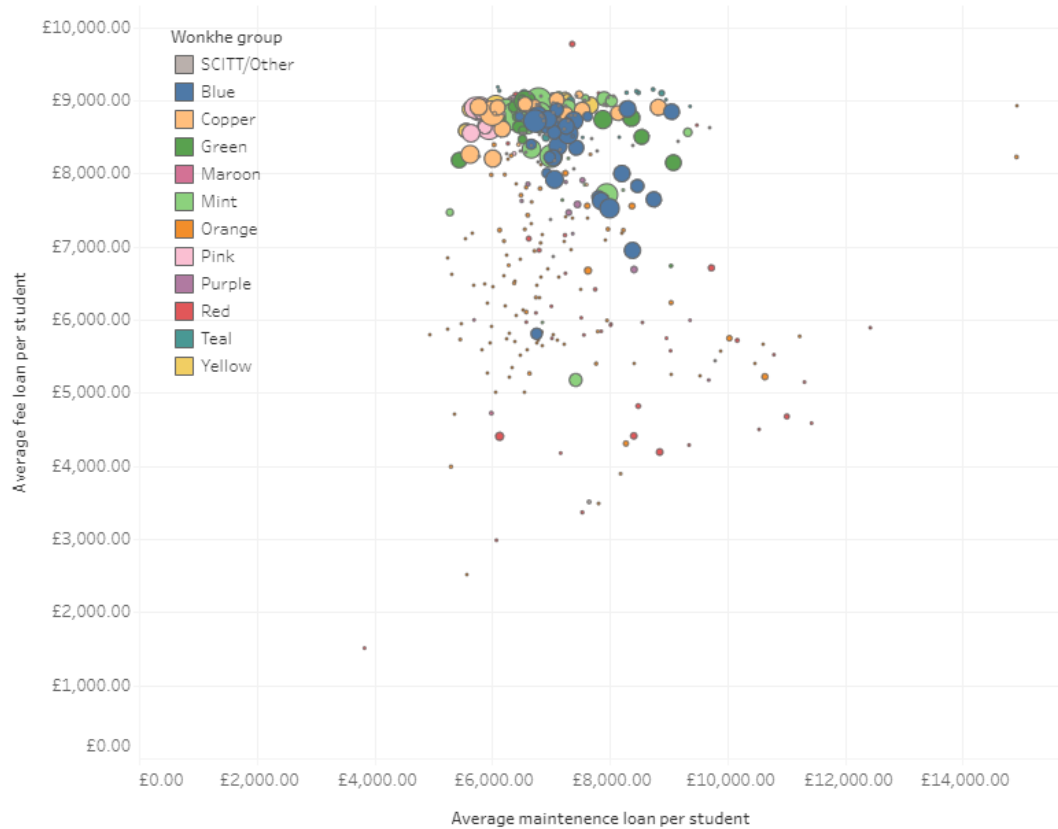


tableau navigation icons

Academic Year 2019 2020

Highlight Name

Number of students in receipt of fee loans and not maintenance loans - 2020

FT UG England domiciled students

Data: SLC

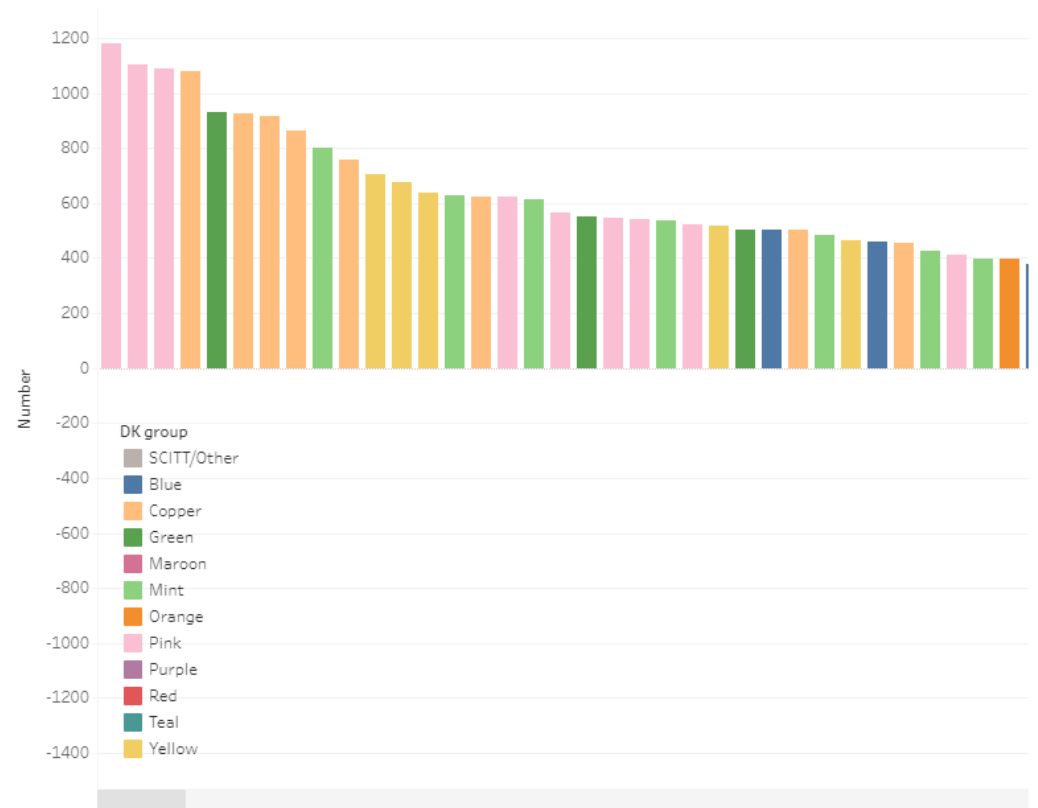


tableau navigation icons

<https://wonkhe.com/blogs/diving-into-slc-provider-level-data/>

Statistics used in regulation

National student survey

Dashboards for:

- **Access and participation**
- (soon) **Outcomes (B3)**
- (soon) **TEF**

<https://www.officeforstudents.org.uk/data-and-analysis/>

(other regulators are available, but release less data as a rule)

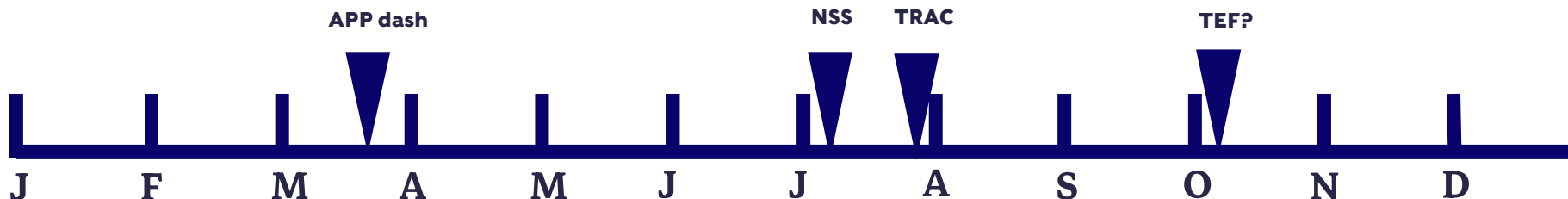
HESES... and the **funding allocations** this drives

TRAC ... (top level only)

Policy witch-hunt of the week (deep analysis of HESA data, usually)

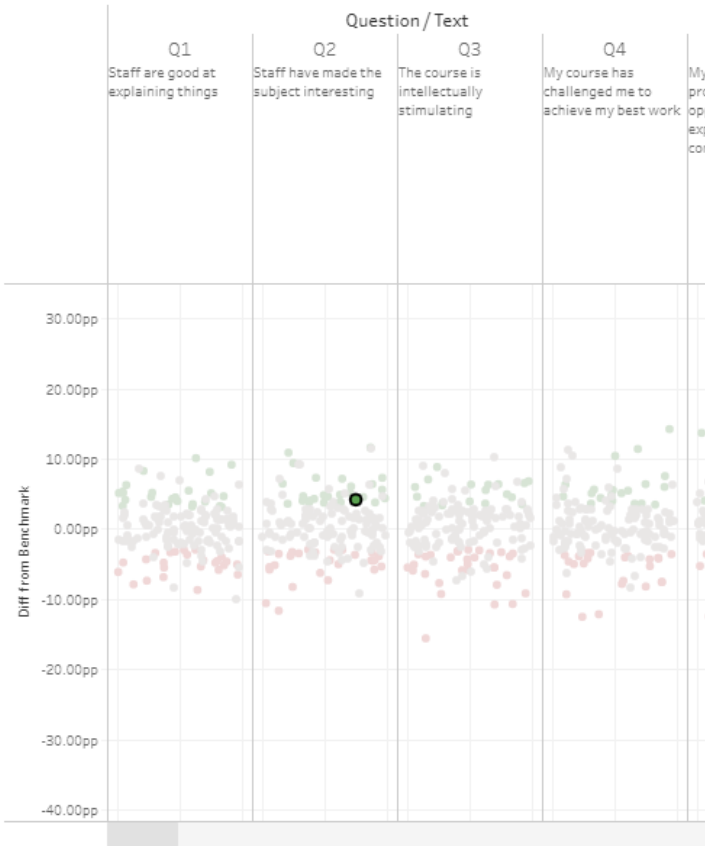
POLAR, TUNDRA, ABCs, Proceed

Why is it collected? Why is it published?



Difference from benchmark

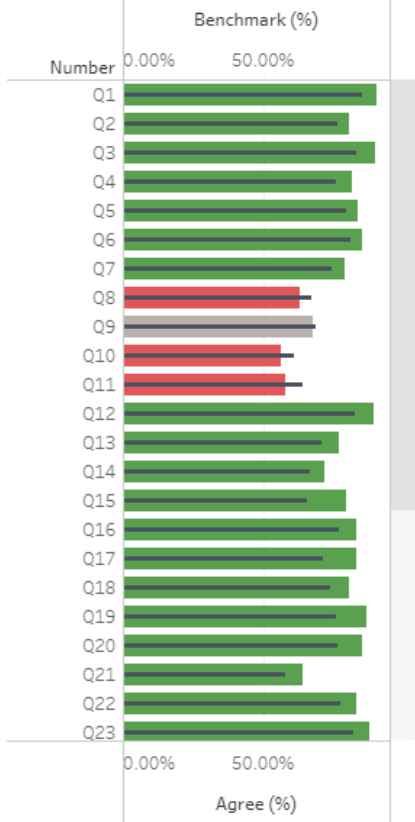
All modes at teaching provider, NSS 2021



Question: (All) | Survey year: 2020 (), 2021 (●) | Significance: None (), - (), + (●) | Highlight Provider: | Region: (All) | Group: (Multiple values)

The University of Bath

Click on a mark in the main graph to see question data by provider



<https://wonkhe.com/blogs/what-can-we-see-in-the-2021-national-student-survey/>

Subject name

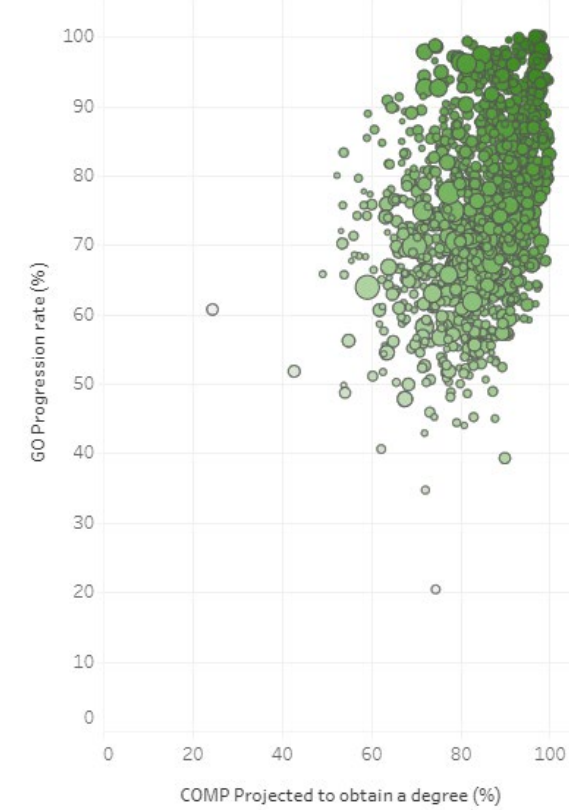
(All)

Highlight Name

Highlight Name

Proceed score

Showing progression rate and completion rate
Data: 0/5



Region: (All) | Group: (All)

<https://wonkhe.com/blogs/measuring-positive-outcomes-needs-context/>

Statistics used in regulation



Department
for Education

LEO – longitudinal educational outcomes. Based on mapping administrative data from DWP/HMRC to HESA student records.

<https://explore-education-statistics.service.gov.uk/data-catalogue> (other depts are available, but release less data as a rule)

Participation measures (HESA recut) – **HEIP** more like LEO approach

Exam results (much later, but in depth)

Education exports

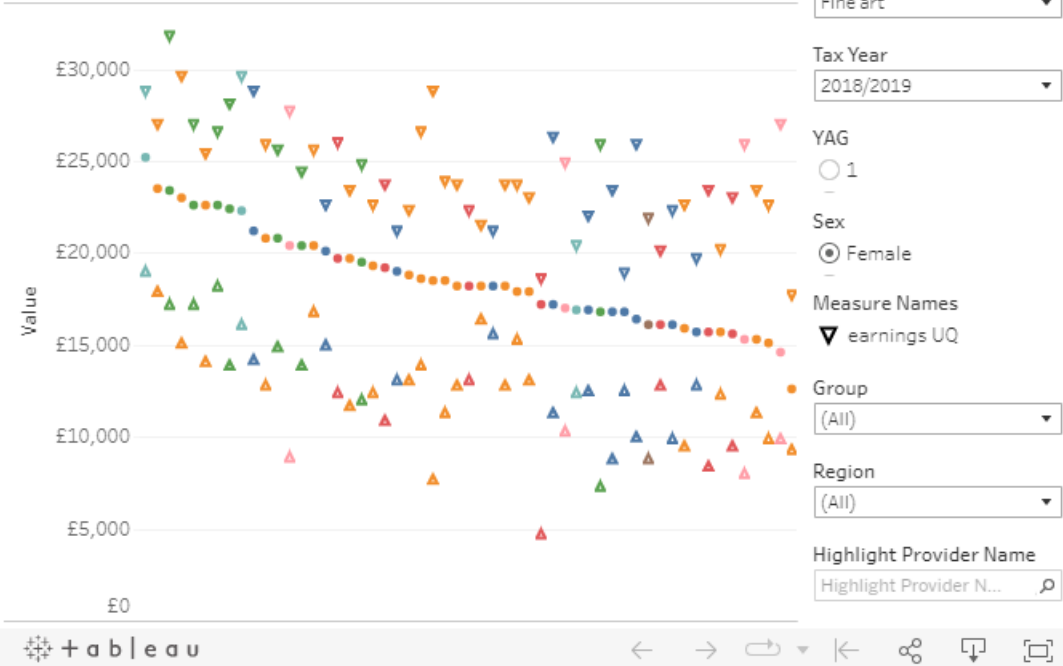
GLMS, SEIS, other bits



Creative arts and design: Fine art

Female students

Tax year: 2018/2019, 5 years(s) after graduation

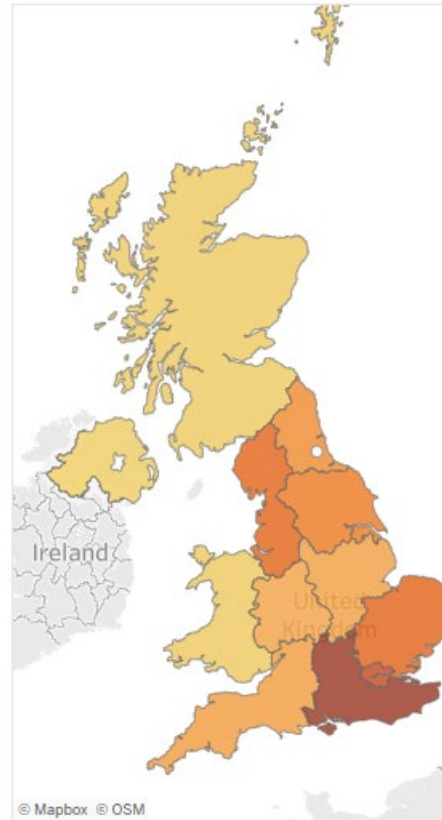


Provider Name

University of Durham

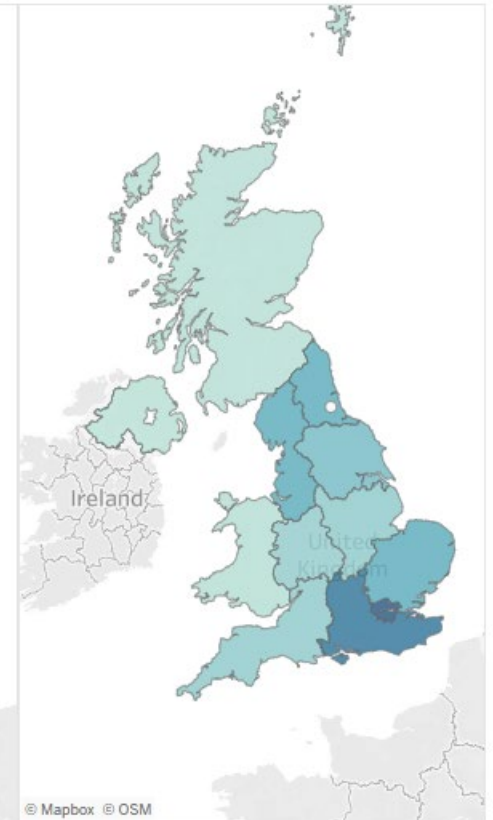
Home region

University of Durham Female + male graduates, 1 year(s) after graduation



Current region

University of Durham Female + male graduates, 1 year(s) after graduation



<https://wonkhe.com/blogs/life-with-leo-summer-2021-release/>

The motherlode!

- HESA Student
- HESA Staff
- HESA Finance
- HESA Estates
- HESA Graduate Outcomes
- HESA BCI
- UKPIs on WP and non-continuation
- Unistats
- Ad hoc reports

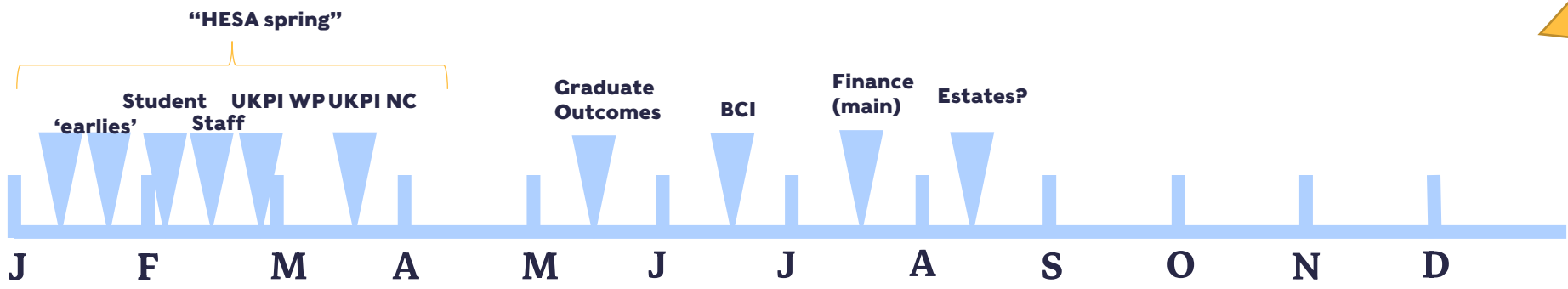
HESA



<https://www.hesa.ac.uk/data-and-analysis>

For the last time in 2022!

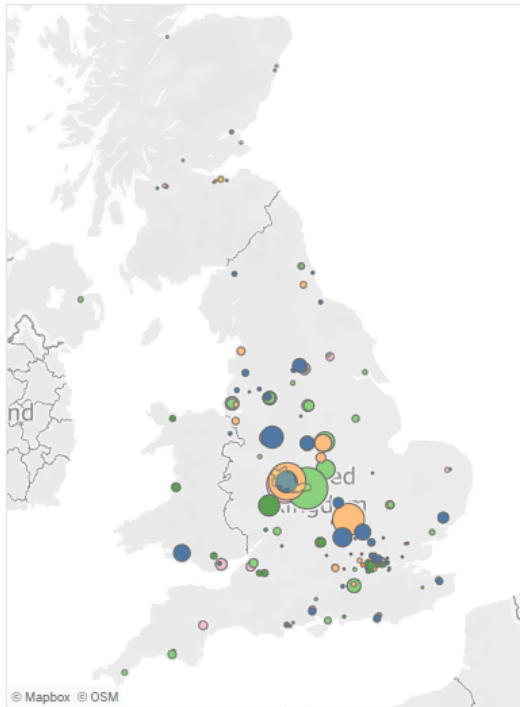
Only source of COURSE data in the UK!



County of domicile: West Midlands
 Academic year: 2020/21

Where do West Midlands students study?

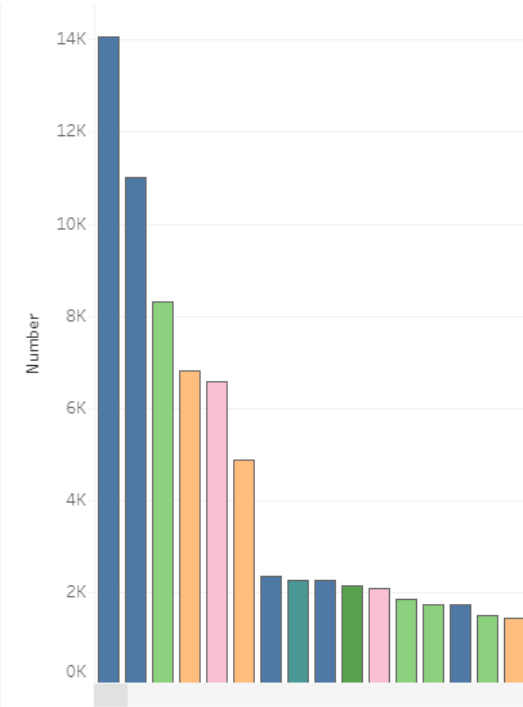
Level: All | Mode: All
 Data: HESA Student 2020/21



Highlight Name: Highlight Name
 Mode of study: All
 Level of study: All

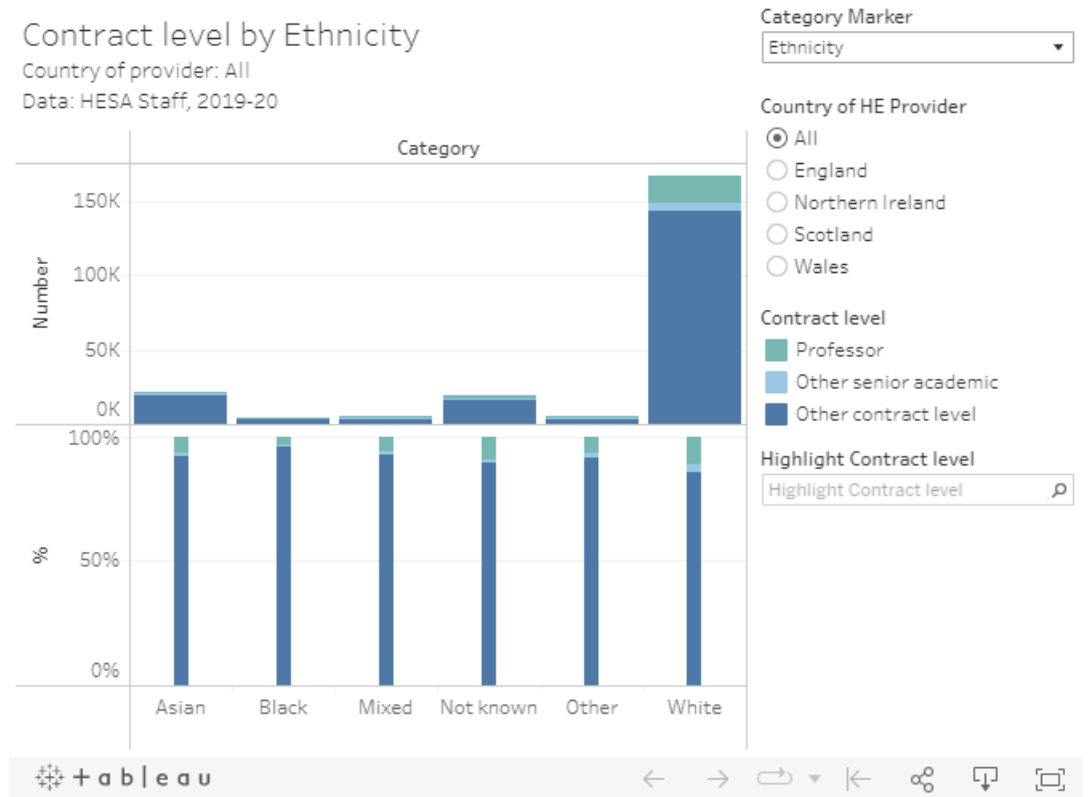
Where do West Midlands students study?

Level: All | Mode: All
 Data: HESA Student 2020/21



Contract level by Ethnicity

Country of provider: All
 Data: HESA Staff, 2019-20



<https://wonkhe.com/blogs/hesa-spring-where-students-come-from-and-what-they-study/>

<https://wonkhe.com/blogs/hesa-spring-2021-overall-staff-numbers/>

Elsewhere...



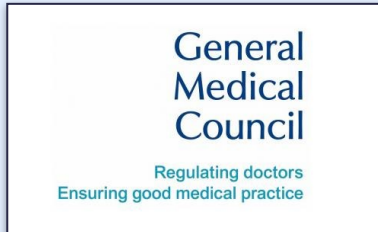
Often includes student information in releases about labour market, housing, wellbeing...
Source of truth for SIC, SOC, and geodata

<https://www.ons.gov.uk>
<https://geoportal.statistics.gov.uk/>



Results of A levels and Highers

<https://www.jcq.org.uk/examination-results/>



Information on student attainment and progression from medical schools

<https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/gmc-data-explorer>



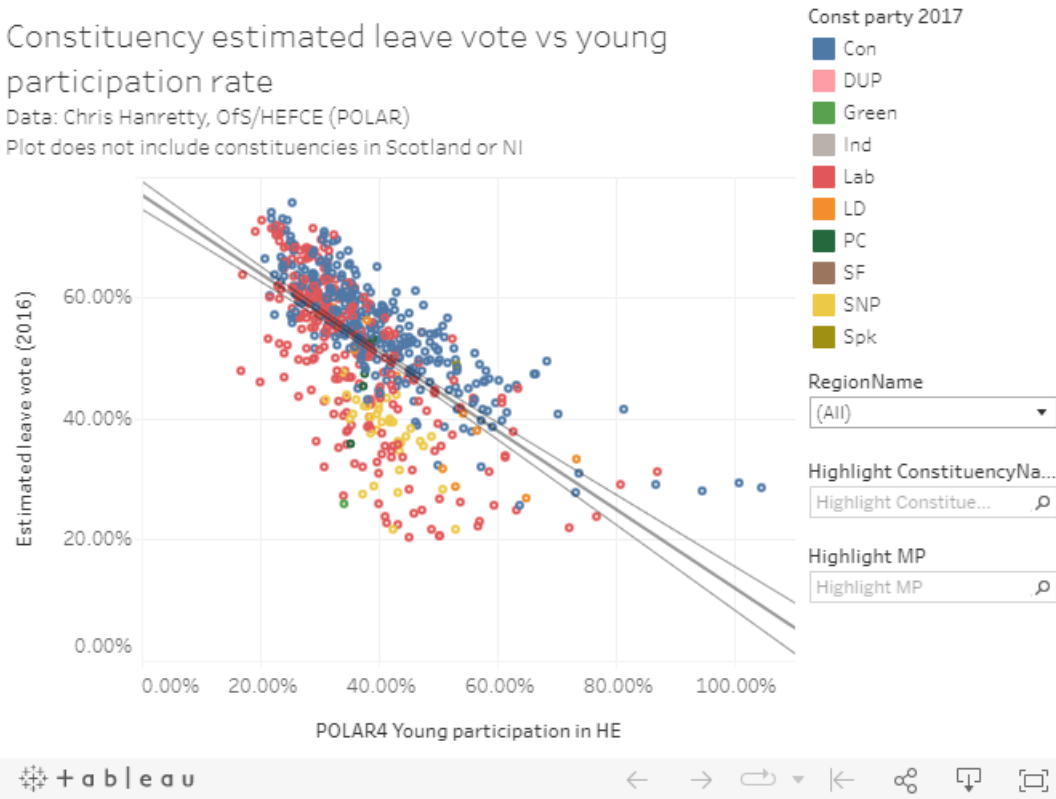
Area based data
Election results

<https://commonslibrary.parliament.uk/data-tools-and-resources/>

Constituency estimated leave vote vs young participation rate

Data: Chris Hanretty, OfS/HEFCE (POLAR)

Plot does not include constituencies in Scotland or NI

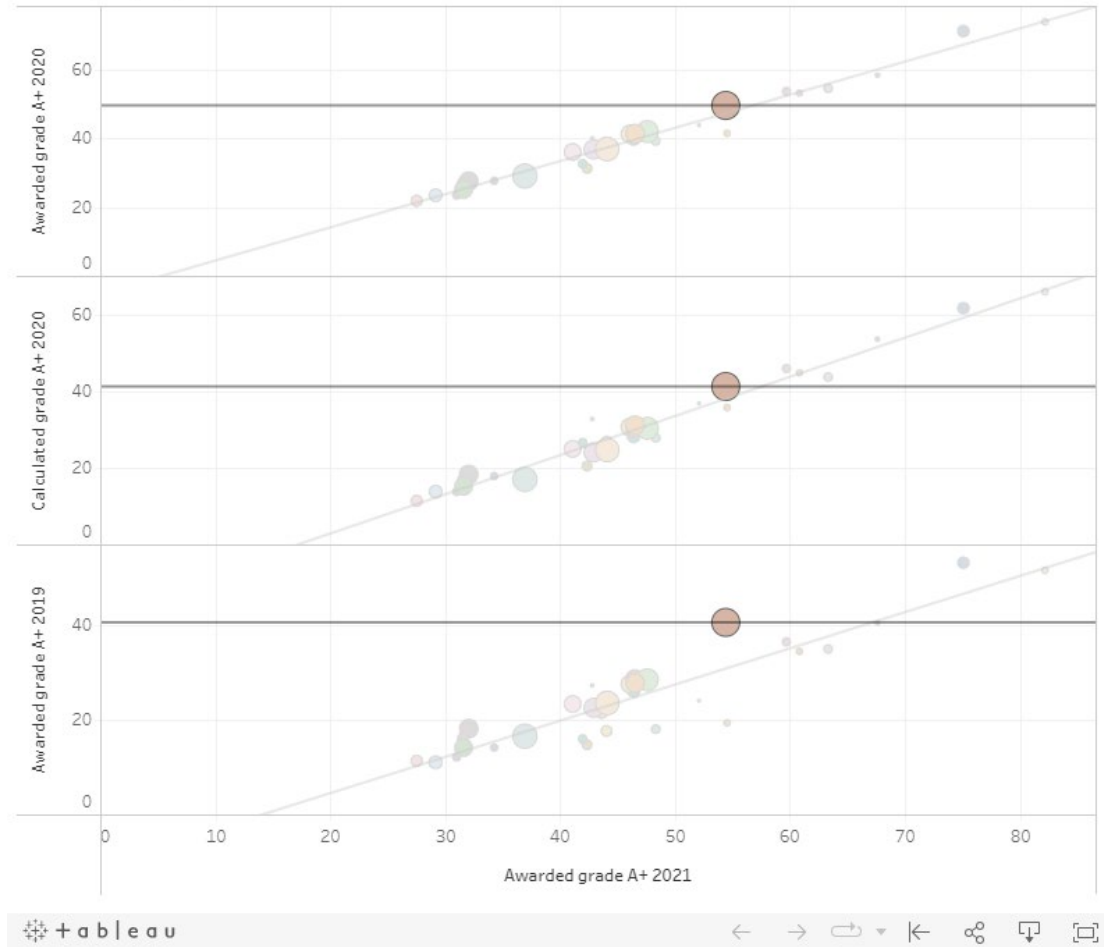


Highlight JQC Subject group

Mathematics

Grades A and above

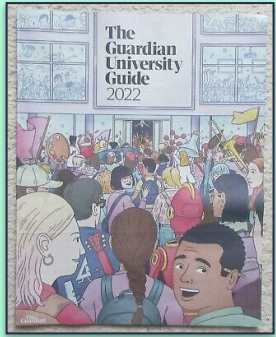
Data: Engand, JQC, 2021



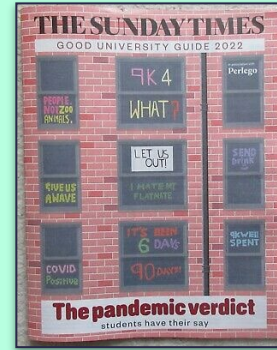
<https://wonkhe.com/blogs/could-higher-education-participation-hold-the-key-to-the-election/>

<https://wonkhe.com/blogs/a-level-results-and-university-entry-2021/>

League tables...



The Guardian



The Times/Sunday Times
(Good University Guide)



Complete University Guide



uMultirank

Decent international attempt

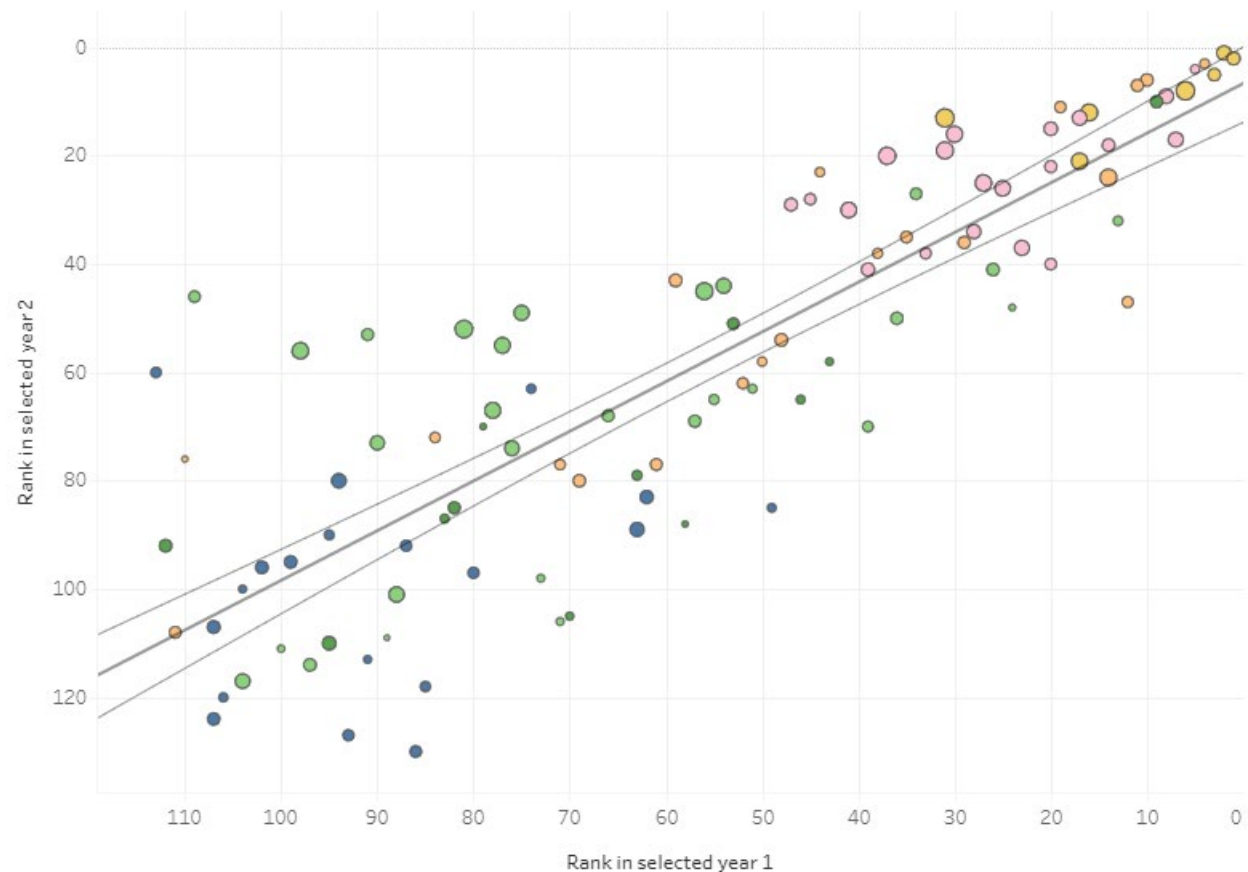
... are they as influential as we think?

Year 1 (x axis) Year 2 (y axis)
◌ ◀ ▶ ◌ ◀ ▶

Complete University Guide Rank

2008 (x) and 2022 (y)

Data: CUG



Highlight Institution Wonkhe group Region Mission group

<https://wonkhe.com/blogs/podcast-levelling-up-diversity-in-person-teaching/>

Working with data

KEY POINTS

Download .csv or .xlsx where you can – dashboards are not the best for independent analysis

Be clear about the question you are asking (including definitions, years etc)

Use column filters (Excel, Google Sheets) to drill down into complex sets. Tableau public is free if you want to try something more visual.

Join data using identifier codes (your old-fashioned VLOOKUP) where you can. It feels like a slog but it is often worth adding codes if there are none.

Document as you go! Use column titles sensibly. Don't be afraid to break down complex operations into single steps.

Check your work. Ideally have someone else check your work.

The screenshot shows an Excel spreadsheet with a formula bar at the top containing the formula: `=VLOOKUP(A2,A4:B8,2,FALSE`. Below the formula bar, the spreadsheet grid shows a table with two columns, A and B. The table data is as follows:

	A	B
1		
2	Russell Group	
3		
4	Million Plus	A
5	University Alliance	Q
6	Independent HE	Z
7	Russell Group	N
8	GuildHE	Y
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		

The formula bar also shows a tooltip for the VLOOKUP function: `VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])`. The table array in the formula is highlighted as `A4:B8` in the spreadsheet.

Thank you

& Questions?

WONKHE