

EP DataKind UK July DataDive 2015

GC **Main WIFI: "USWGuest", password  mple way to switch"**

Find us on:

- **Hackpad** for **reference documentation** (you're here already), at <http://bit.ly/centreport-datadive>
- **GitHub** for **data and code** at <https://github.com/DataKind-UK/Centreport> (it's a private repository, ask giacecco@ dico.im to be added to the team)
- **Slack** for **chatting** at <https://datakinduk.slack.com/messages/centreport/> (as above, ask giacecco@ dico.im to be added to the group)
- **Google Drive** for **anything else** at <https://drive.google.com/open?id=0B3i-XvikxKzAfmXIYW96OVM5eTR5dXFUQ2hZNkxmeGRIYU51VjdkSTRzeF9ZVG VULWILWmM>

Go back to the general DataDive [here](#).

EP Centreport

JM Charity Leads: Jesse Mears & Gaia Marcus ([@yhatabank](#), [@la_gaia](#))

EP Data Ambassadors:

- Emma Deraze
- GC • Gianfranco Cecconi (giacecco@ dico.im, [@giacecco](#))
- EP • Giasone Salati

About Centreport:

JM Centreport is the UK's leading charity for homeless young people. It supports over 8,400 16-25 year olds into housing and employment every year. Working directly in London, Yorkshire and the North East of England, Centreport also partners with other organisations across the UK and gives homeless young people a voice through the Centreport Parliament. Centreport conducts research and influences government policy with the overall aim of ending youth homelessness.

Currently Centreport uses data to argue for new business, evaluate their

services (in the short-term), pilot and introduce new interventions, evaluate their campaigns, and lobby government for policy change.

The Youth Homelessness Databank

Centrepont would like to help end youth homelessness by collecting, analysing, and sharing all data on youth homelessness.

" We want to be able to answer questions we haven't thought of yet to solve a social problem that should not exist in one of the richest countries in the world. Ultimately, we want to end youth homelessness in the UK and we know we can't do that without collecting and understanding the data."

This was the pitch that won Centrepont the Google Impact Challenge in 2014 and has led to the [Youth Homelessness Databank](#). Hello! This is a two year project to equip the youth homelessness sector with the digital tools to collate and understand their data.

Our first step was to try to gather together the freely available data on youth homelessness. We were **amazed** by how little data there is (how much information is lost in the system), and how the data that are available are not broken down by age. Getting an informative national picture of youth homelessness is currently impossible.

If we could understand the scale of youth homelessness Centrepont and the rest of the sector could better argue for services for young people who need support; if we could understand who is homeless and where, we could commission services better; and if we could understand what interventions are most effective we could allocate scarce resources in a more effective and responsible way.

Why do we need your help? Although we are a data project, we are a team of 3 (with only one data dude), we believe in working openly and collaboratively, and we want to explore innovative solutions to our data problems.

We can't do this alone! With your help we can kick-start the Youth Homelessness Databank's contribution to the youth homelessness sector, by making sense of the new data we have collected, answering what should be a simple and important question, presenting it in a clear and effective way, and making it available to all.

GC **Here's something we prepared earlier:** if useful, these are the two digital tools we have out there- <http://centrepont.nqminds.com/> (aimed at MPs and local

activists) and <http://yhd.meteor.com/> (a *very* minimum, minimum viable product in development)

EP The Challenge:

JM No one knows how many young people are facing homelessness. Official data are not reported by age, are unreliable and underestimate the problem. As a first step, we want to answer the following question:

How many young people in the UK turn up to their local authority because they are homeless?

EP What we're going to do to solve it:

JM We have asked every local authority in the UK (England, Northern Ireland, Scotland and Wales) for:

- the number of young people that turned up because they were homeless or at risk of homelessness, and
- how many were accepted as 'officially homeless'.

As you might expect, the data collected are patchy and of variable quality. Using this raw data, and other data sources that are available, we would like to build a model to calculate an estimate of youth homelessness in the UK (represented by the number of young people presenting to their local authorities) and present this beautifully to a lay audience.

EP The Data

- JM 1. Centrepoin data from local authorities in England (by either **District** or **Unitary Local Authority**), Wales and Northern Ireland.
2. The Children's Society data for 16-17 year olds from local authorities in England.
- ED 3. **Annual** England official homelessness data (P1E data), by **District** or **Unitary Local Authority**.
- JM 4. **Annual** Scottish official homelessness data, by local authority
5. English social housing data (CORE data), by **Region**.
 - ED a. Affordable - general
 - b. Affordable - supported housing
 - c. Social - general
 - d. Social - supported housing
6. Census data from 2011 census.
- JB 7. Please add data from <http://centrepoin.ngminds.com/explore>

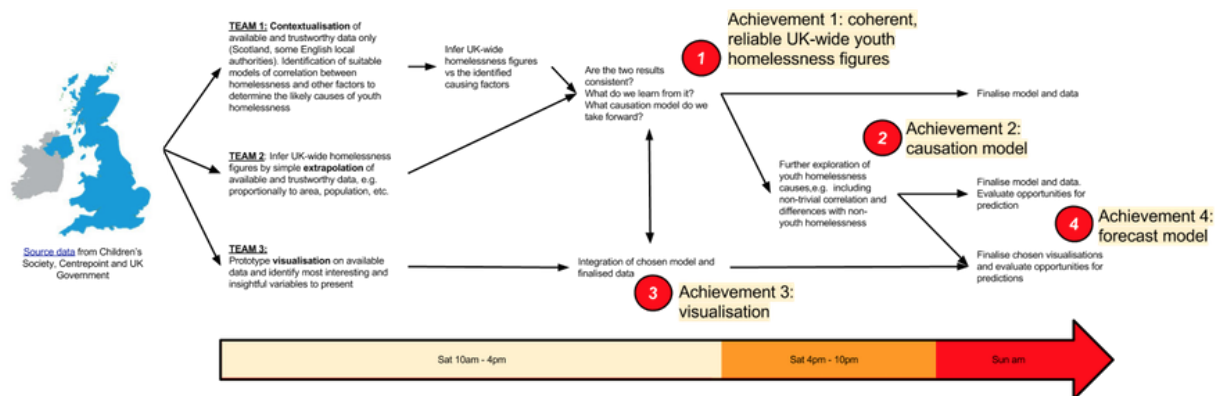
GC The (proposed) workflow

- **How (much)?** How big is the youth homelessness problem in the UK? Can

we build a reliable, coherent estimate for all country at local authority level?

- **Why?** What causes youth homelessness?
- **When?** Looking at the future, can we make forecasts of how youth homelessness will change, e.g. as a consequence to changes to UK regulation or budget?

... and prepare to communicate our findings!



(see the full-size original [here](#))

Need to know

'Youth homelessness'

- JM
- 'Youth' is defined as 16-24 year olds (or <25 year olds).
 - Youth homelessness is seen as a problem of transition between childhood and adulthood. It is different to adult homelessness in that it is most often the result of a relationship breakdown.
- GC
- Different age ranges imply different contexts, regulation etc., as described below:
- JM
- 16-17

16-17 year olds have different rights because they are still legally children.

When they approach their local authority they should generally be assessed by children's services (not housing departments) and are entitled to housing *and support* from their local authority. (There are only a small minority of cases that should go through housing rather than children's services e.g. if they are deemed to to not be a 'child in need', which is unlikely if they are homeless, or if they refuse children's services support.).

Homelessness' is not a category within the children's services assessment so these 16-17 year olds are immediately 'lost' from the homelessness statistics as they will instead be recorded along with other children in care and cannot be pulled out separately.

As you will see from the data, this isn't what always happens! Some 16-17 year olds are assessed by the housing department rather than

children's services, where they should fall into the priority need category 'vulnerable due to being 16-17 years old'.

More information on the rights of homeless 16-17 year olds can be found [here](#). We have tried to collect data on this age group from both housing departments and Children's Services within each local authority.

b. 18-24

When 18-24 year olds approach their local authority because they are homeless or at risk of homelessness, they should be assessed by the housing department. They have fewer rights than 16-17 year olds and >25 year olds (an chilling example of this is the recent [housing benefit cut for <25 year olds](#)). They only have a legal right to housing support if they are in a 'priority need' group e.g. if they have children of their own, are a care leaver aged up to 21, or are particularly vulnerable due to severe health problems or domestic violence. (Other 18-24s may still get some housing support from their local authority even if they are not priority need by they will not have a legal right to it and it won't be capture by the official figures.)

GC • **Types of homelessness relevant to the DataDive investigation**

JM ◦ A young person is '**presented as homeless**' or '**at risk of homelessness**' when he or she turns up at their local authority and asks for help because they are homeless or at risk of becoming homeless.

GC ◦ To be accepted as 'officially homeless' and given full housing support from the local authority, one needs to be assessed as '**eligible, unintentionally homeless and in priority need**'.

In more detail:

JM ◦ **eligible** - this will depend on immigration status

GC ◦ **unintentionally homeless** - prove that it is not the 'fault' of the individual that they are homeless - this is often a highly subjective measure

JM ◦ **in priority need** - a pregnant woman; have dependent children; someone vulnerable as a result of old age; mental illness or handicap or physical disability or other special reason; someone homeless or threatened with homelessness as a result of an emergency such as flood, fire or other disaster; aged 16 and 17 years old; aged under 21 years old who were in local authority care between the ages of 16 and 18; aged 21 and over who are vulnerable as a result of leaving local authority care; vulnerable as a result of leaving the armed forces; vulnerable as a result of leaving prisoner; vulnerable as

a result of fleeing domestic violence or the threat of domestic violence

They must also be able to prove they have a '**local connection**' to the area

GC Sources of youth homelessness data for the DataDive and known issues

JM • Data sources

- Centrepont research (FOI requests to local authorities in England and Wales)
- Another charity's FOI data on 16-17 year olds (The Children's Society)
- English Department for Communities and Local Government (DCLG; 'official data' from the P1E form)
- Welsh Government's online repository for statistical data for Wales (StatsWales)
- Scottish Government statistics ('official data')
- Northern Ireland Housing Executive

GC Data sources for context:

See also: <http://www.crisis.org.uk/data/files/publications/Crisis%20briefing%20-%20youth%20homelessness.pdf>

Unemployment data:

Sourcing this data by local authority was actually more annoying than I expected! The only way to get it is by doing a query on Nomis. I've done one for the 16-24 claimant count by local authority (please see attached). I've included both district and unitary local authorities which is why there are so many, and cut it by duration of claim. I couldn't get different date ranges to show in the same table but it would be easy to run it again for earlier dates if needed. If you want to run a new query of your own, you can do so here:

<https://www.nomisweb.co.uk/query/select/getdatasetbytheme.asp?opt=3&theme=&subgrp=>

You could run another query for headline unemployment but unfortunately the wizard tool does not allow you to break age down further than 16-64. I think we could probably request the data directly from ONS though.

Educational attainment:

The table entitled 'local authority tables 16-24' includes several measures of

educational attainment. I think the 18a tab is probably the best one – the proportion of 19 year olds who have a level 3 qualification.

<https://www.gov.uk/government/statistics/level-2-and-3-attainment-by-young-people-aged-19-in-2014>

Apprenticeship starts

You want the table entitled ‘apprenticeships geography, age and level starts’ available at this link:

[Apprenticeships geography age and level: starts 2005/06 to 2013/14](#)

The local authority level data is available in the ‘local education authority’ tab. There are figures for under 19s and 19-24s. Ideally it would be good to amalgamate these two.

Statutory youth homelessness:

All the statutory homeless data is available in its raw form by local authority, but each quarter is in a separate spreadsheet. The figure for young people is in the ‘section 1’ tab of each spreadsheet, in the column labelled e1b1a (normally in column AY). All the spreadsheets are available at this link:

<https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness#detailed-local-authority-level-responses>

Homelessness prevention and relief statistics:

The other potential source of homelessness data is the homelessness prevention statistics. Unfortunately these do not break down by age, but could still be useful to demonstrate relative levels of need in different areas. I would just use the ‘total prevention and relief’ stats for each local authority.

[Table 792: total reported cases of homelessness prevention and relief by outcome and local authority, 2009-10 to 2013-14](#)

Indices of multiple deprivation

The data available at local authority level is available here:

[English indices of deprivation 2010: local authority summaries](#)

CORE data

Social landlords submit data on who they let to as part of the CORE database. Unfortunately, you have to register to be able to cut this in a useful way. You can register here:

<https://core.communities.gov.uk/AnalyseCOREdataMain.aspx>

The best thing to do is probably to use the number of young people that

received a letting who were formerly homeless (which is a tag on the system). It could be good to split this by general needs and supported housing lettings if possible.

JM • Data quality

◦ Collected at different **geographical levels**

- In **England**, there are 3 types of local authority: Counties, Districts and Unitary Authorities. **Counties** are responsible for **children's services** and are known as 'upper tier authorities'; they will be coterminous **but not institutionally linked** to a number of **districts**, which are responsible for **housing services**. Think Oxfordshire County Council vs. Oxford City Council (a district). **Unitary** authorities are responsible for both housing and children's services, but the data are kept in different departments and in different databases.
 - The data on 16-17 year olds should mainly be held by counties or Unitary authorities because they are responsible for children's services, however this does not always happen. The data on 16-17 year olds from counties and the sum of districts within the county **will not** equal each other. This is because they are from different departments and different pathways/processes, and are likely to represent different people.
 - If attempting to get an aggregate number for counties, the 16-17 year olds reported by the districts within the county and those reported by the county should be summed. This is further complicated by missing data.
- **Missing data!** There is so much missing data. This is because local authorities only *have* to report the data on the P1E form, and therefore many of them only collect the data they are required to report. The data on this form is only based on people who are 'accepted' as homeless. The only field that is reported by age is the number 'accepted' as homeless. This is only reported at the local authority level by quarter, rather than annually (which is only available for the whole country). Due to the reporting practices described below (e.g. <5), the quarterly figures don't add up to the annual figures...
- Due to missing data we may not be able to distinguish successfully between age ranges (16-17 and 18-24) and we may have to either consider the total only or make an assumption around the two ranges.

- The data quality is also poor due to a **lack of resources** at the local authority level, lack of user-friendly databases that can be freely interrogated (they are often locked into expensive contracts with database providers who charge the thousands of pounds to perform a new query), and a lack of data- and database-literacy/capacity.
- **It is what it is.**
- Data reporting issues
 - <5 -- where fewer than 5 cases are reported in a field, it is the norm for it to be reported as '<5' in order to protect the identity of the individual(s). This causes a problem when aggregating the data and means that quarterly data rarely adds up to annual data, and totals aren't necessary the sum of their parts.
 - Wales' official data is rounded to the nearest 5 -- This causes a problem when aggregating the data and means that quarterly data rarely adds up to annual data, and totals aren't necessary the sum of their parts.

Acronyms

- FOI - Freedom of Information request - a data collection method to gain access to government data
- P1E - The statutory homelessness data reported by all local authorities in England to create the 'official figures'
- CORE - Continuous Recording of Lettings and Sales of Social Housing in England - Social housing data for England

EP **People, twitter and GitHub IDs:**

ED [Main GitHub](#) repo for the project.

[CORE data Dropbox](#) , explanation of fields [here](#) .

GC **Skills at the table (PLEASE EDIT!):**

KD Kevin - GIS analysis & visualisation, open data, web developement, CartoDB, D3

BS Brian - Solution design, data governance / data quality / data management in general. Excel, SQL, JavaScript. User experience

JM Matt - Analytics and big data, vis

Tharindi - Information architect, maths, data analysis, python, scraping, bit of D3, nlp

Michael - python, R, data analysis, d3 (GitHub ambassador)

- KN** Kelvin - python, Java, open data (GitHub ambassador)
- JM** Jonathan - stats, machine learning, R, python, data wrangling
 Peter - excel, pivot tables, SQL
 Pieter - Excel, My-SQL, Python scraper, data wrangling, finding a sweet spot!
- AK** Adam - physicist, data science (python, scikit-learn; can help with github!)
- JM** Sam - research, big data, data analytics, python, mongodb
- CL** Charlotte - data science, python, research, big data, data analytics, mongodb, databases
- JM** James - statistician
 Robin - telling stories from the data, dark arts! R, SQL
Judy - strategic analytics, classical analyst, SPSS!
 Verena - data analyst, SQL, dashboard reports, tableau, cartodb
 Justina - econometrician, bayesian analysis
 Esther - epidemiological data, statistician, R
 Andrea - data analysis, modelling
 Gaia - Network analysis/vis

Project Suggestions (additional ideas for the DataDive, where time permits):

- GC**
- analysing open spending data - <https://openspending.org/> or <https://www.spend>
 - network.com/ (better?)
 - playing with the MVP (meteor,mongodb, D3, carto db)

EP Parking lot (ideas for Centrepont to explore at a later date):

- JM**
- Are FOIs the best way for Centrepont (or charities in general) to collect national data? Alternatives?

MW Data Acquisition Queue

We will first download as much as possible and then look to clean them later.

We will try to map all the disparate data to **LSOA geographic code**, if it isn't already.

- TH**
- Spending Data (in progress) - available through [Spend Network](#) API, but limits to 300 entries per council. Insufficient. Looking at other sources...
- KN**
- Crime Data - Huge file of crime data per incident per region (LSOA code) https://drive.google.com/file/d/0B0qBz_8LYyRwVHNpRHh3cURtWUU/view?usp=sharing
- MW**
- Social Care - Children in care https://drive.google.com/open?id=0B_1oINVHtrVqbHBTdUZaTIZrQnc

- GC • Census Data (in progress, partly accounted for in the Shelter data)
- EW • **Youth Unemployment**
- MW
 - **England and Wales** https://drive.google.com/open?id=0B_1oINVHtrVqR1U3VW1RaEF5aE0
 - KN
 - **Scotland**
 - https://drive.google.com/file/d/0B0qBz_8LYyRwYjM3V2ZnMm1hR2s
 - GC
 - **Northern Ireland** <https://drive.google.com/open?id=0B3i-XvikxKzANIZpZTRIWnR6UU0>
- MW • Approximate Social Grade
 - https://drive.google.com/open?id=0B_1oINVHtrVqdlc0SjRiVmlYX00
- Shelter (England only)

http://england.shelter.org.uk/professional_resources/housing_databank

 - Regions mapping: https://drive.google.com/open?id=0B_1oINVHtrVqdkxsMUJYNFlaNm8
 - Extract of all of the following are here: https://drive.google.com/open?id=0B_1oINVHtrVqNTRoUXRseIR6bGM
 1. Families with children accepted as homeless and in priority need
 2. Households accepted as homeless and in priority need
 3. Households found to be 'intentionally homeless'
 4. Households found to be eligible, homeless but not in priority need
 5. Households in Temporary Accommodation
 6. Households on council housing waiting lists
 7. Households with dependent children in temporary accommodation
 8. Number of children in temporary accommodation
 9. Possession claims issued by landlords
 10. Possession claims issued by Mortgage lenders
 11. Possession orders (outright plus suspended) made for landlords
 12. Possession orders (outright plus suspended) made for mortgage lenders
 13. Average (mean) selling price
 14. Average weekly council rents
 15. Average weekly housing association rents
 16. Lower quartile house price to income ratios
 17. Mean private rents (All properties, 12 month rolling average)
 18. Median full-time wages
 19. Median house price to income ratios
 20. Median private rents (All properties, 12 month rolling average)
 21. Median selling price
 22. Council homes sold through Right To Buy

23. Homes owned by council
24. New homes built by council (Completions)
25. New homes built by council (Starts)
26. New homes built by housing associations (Completions)
27. New homes built by housing associations (Starts)
28. New homes built by private developers (Completions)
29. New homes built by private developers (Starts)
30. Number of affordable homes delivered
31. Total number of new homes built (Completions)
32. Total number of new homes built (Starts)
33. Vacant dwellings (all)
34. Housing Benefit claimants - all
35. Housing Benefit claimants - council and housing association tenants
36. Housing Benefit claimants - private tenants
37. Population density (people per sq km)
38. Working age unemployment rate (%)

EW • x social housing waiting list

TH ◦ In above (Households on council housing waiting lists)

MW • **Deprivation:**

GC ◦ **England and Wales:** <https://drive.google.com/open?id=0B3i-XvikxKzANjJPczBiTnBfUE0>

MW (see also https://drive.google.com/open?id=0B_1oINVHtrVqajd4ak13ckhOMTg)

GC ◦ **Scotland:** <https://drive.google.com/open?id=0B3i-XvikxKzAbjYUtUV1piS1k>

◦ **Northern Ireland:** likely not available, see [here](#)

EW • educational attainment

TH **England, Wales and Scotland:**

<https://drive.google.com/drive/folders/0B4hYdHWH499efl8zeTN1akJRY1UyeUItNHdjUjN4RHRpRDNPcDR0eXBJSk5Qb0tldlBOcnc>

- EW • income
- non-traditional family structure
 - bedroom tax

TH ◦ In the Shelter set (housing benefit claimants)

EW • number of school exclusions

TH ◦

MW • youth crime... put more here

KN ◦ source: <https://www.gov.uk/government/collections/youth-justice->

statistics

- By region: [2011](#), [2012](#), [2013](#), [2014](#)
- Let me know if you need LSOA code for the region
- ED ▪ crime-related index of deprivation in [above 'deprivation' link](#) as well
- KN • Add more to this queue!

JM **Saturday 4pm update!**

Data Modelling Team

- Defining the format towards which all of the data sources will converge for the rest of the day.
- Solving administrative geography problems.
- Centrepont's dataset has been cleansed and turned into a usable data model
- Next step - bringing in the wider contextualising variables

Strategy Team

- Filling in the missing data - building simple models to impute the missing data so ready for when the full dataset is ready
- Identifying opportunities for studying correlation and causation (statisticians)
- Documenting the assumptions
- Will look at - deprivation data, education data, investigate (unexpected) social consequences of gentrification (changes in wealth - housing prices)

Data Gathering Team

- Pulling in all the data that need to be used to fill in all the blanks, mostly from the Census 2011
- Frustrated by API limits!

These 3 teams will converge when they are ready. At the end of this, we will have the first UK-wide estimate of young people presenting to their local authority as homeless!

Team Visualisation

- Exploring map-based visualisations of the data. Neutral feelings! The calm before the storm.

Interactive Flow Team

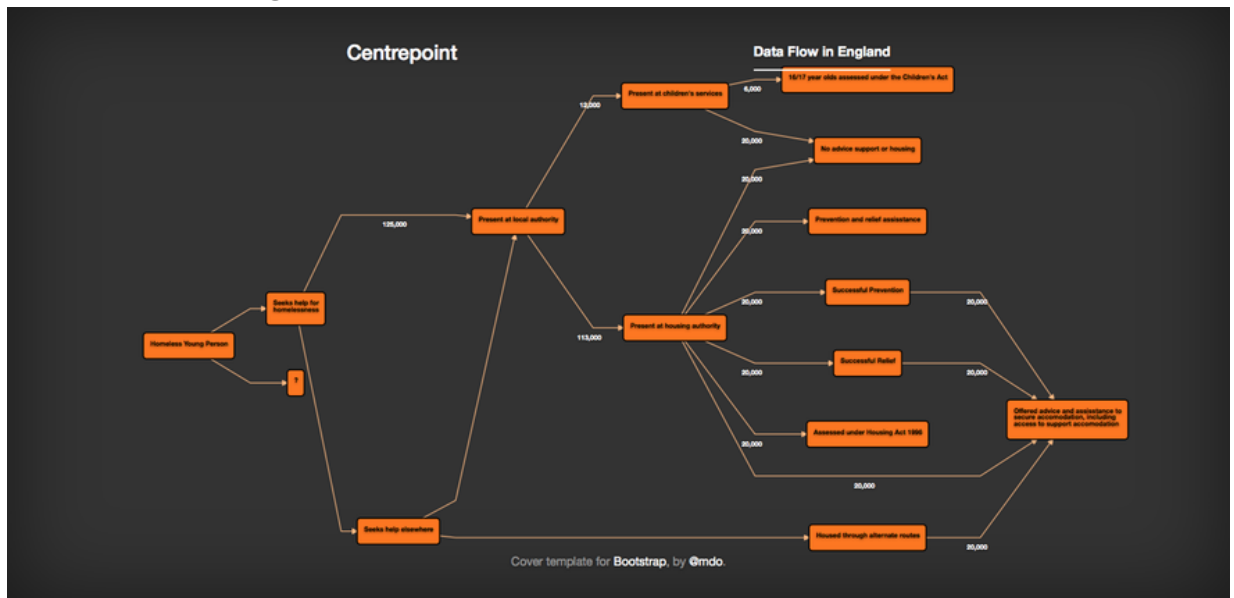
CL



- JM • Working on the interactive visualisation of the flow of data - flow diagram has been recreated in D3 - whoop! Ready for the data...

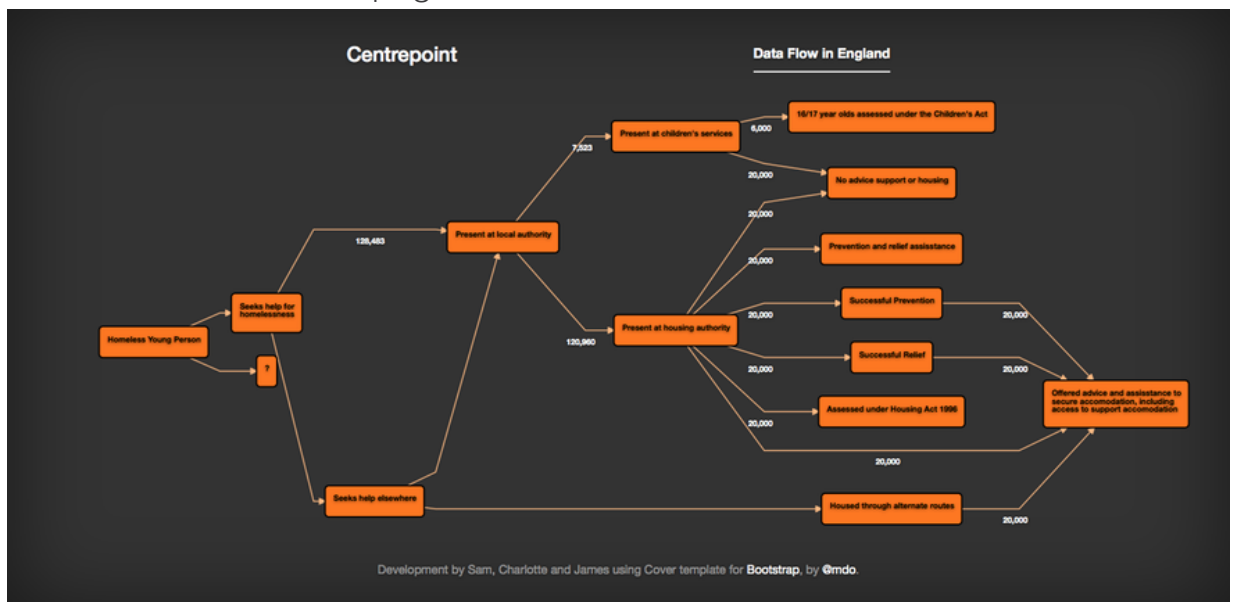
- CL • Initial flow diagram

SC



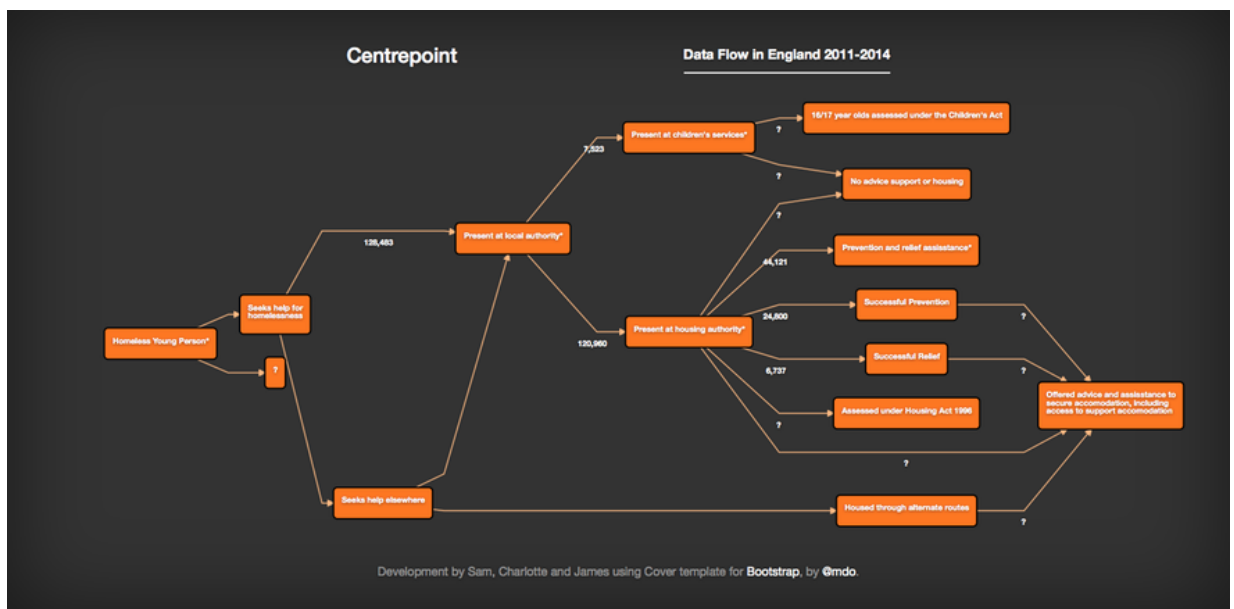
- CL • Updated flow diagram
- Figures updated on the flow diagram to match the new data presented to us.
 - Better fits the page!

SC



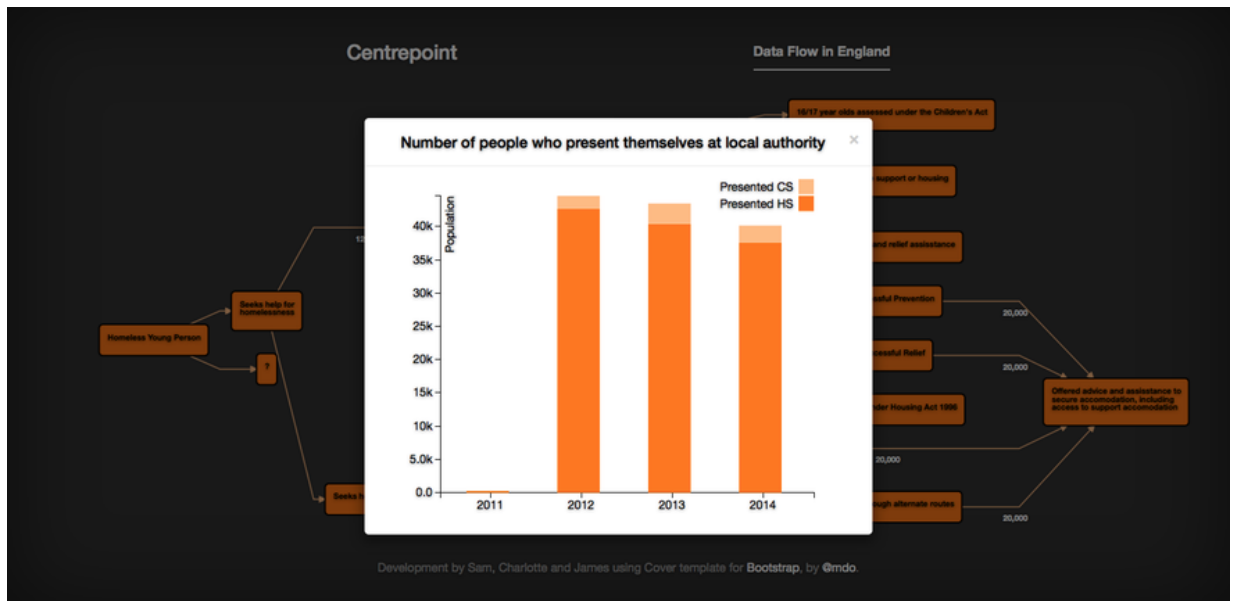
- CL • Final Flow diagram
- A number of data items are missing which are represented by ?

SC



- CL • Number of people who presented themselves at a local authority

SC



- CL • Number of people present at housing authority
 - Example:-
 - Erewash has a low population but a high homelessness figure
 - HS - 1791 and population 112081
 - You can hover over blobs on the graph to display local authority, population and homelessness figure.

The bubble chart displays the number of people who present themselves at housing authorities in 2014, categorized by year (2011, 2012, 2013, 2014). The Y-axis represents Population Density in 2014, ranging from 0 to 350,000. The X-axis represents the Year. The chart shows a general trend of decreasing population density over time, with a notable peak in 2012. The bubbles are color-coded and sized to represent different categories of homelessness, with orange being the most prominent.

Flow Diagram on the Right:

- Homeless Young Person
- Seeking help for homelessness
- Offered advice and assistance to secure accommodation, including access to support accommodation

- SC

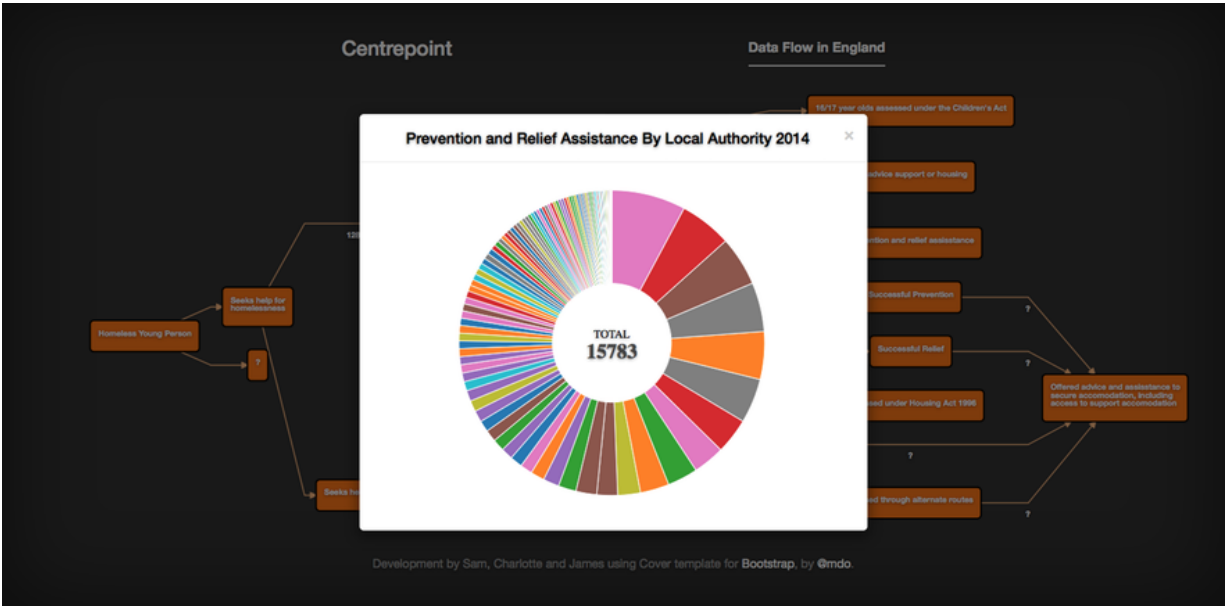


- SC

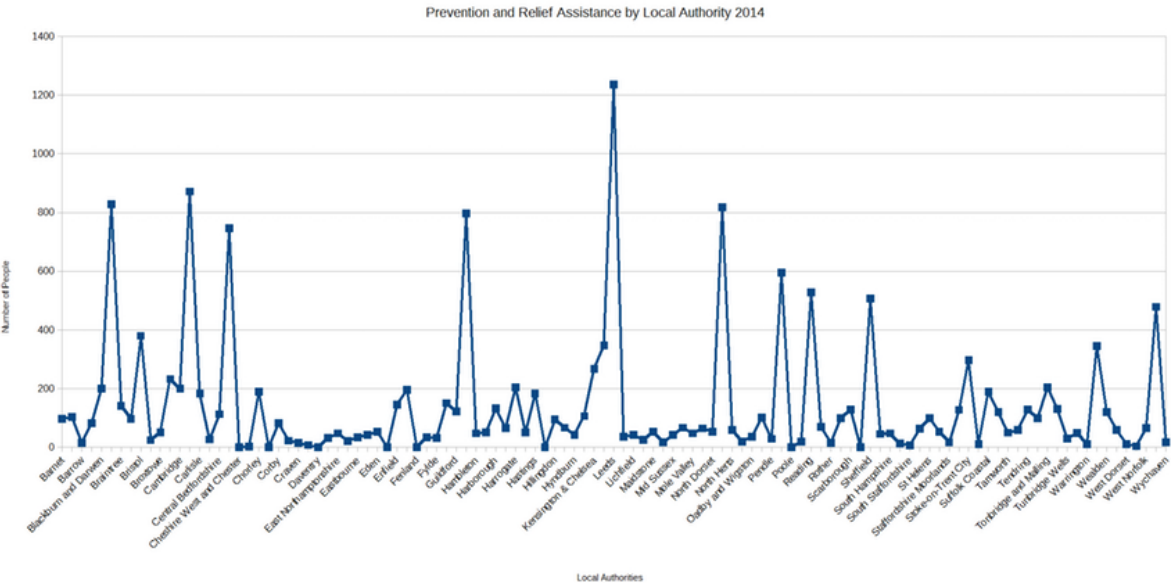


- CL
- Prevention and relief assistance by local authority 2014
 - Pie Chart

SC

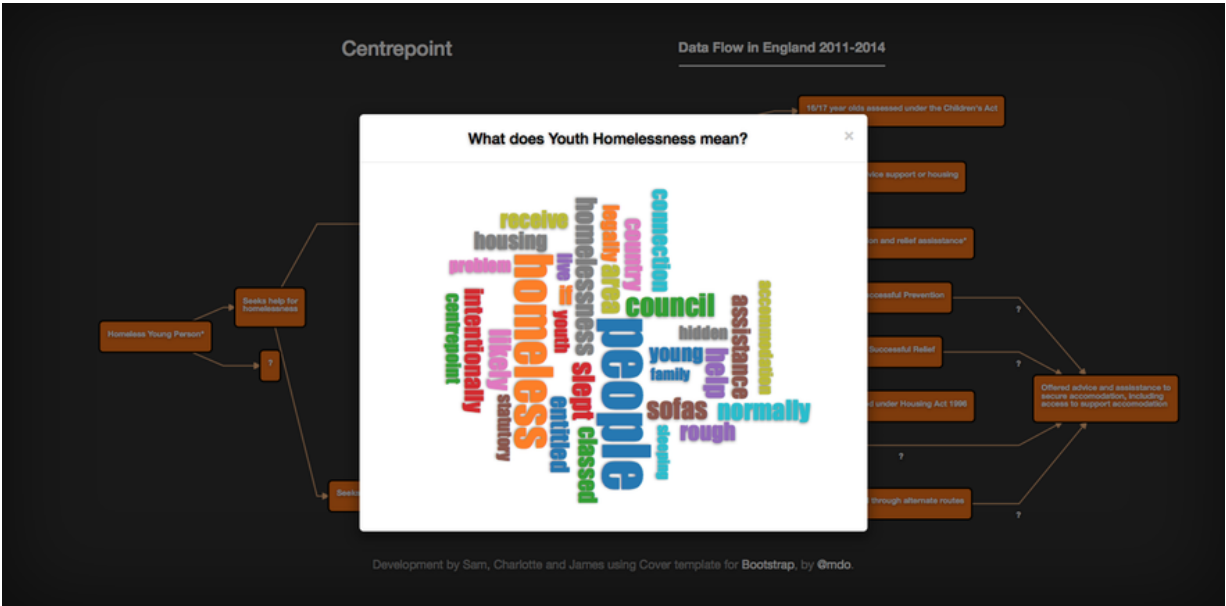


- CL
- Line Chart

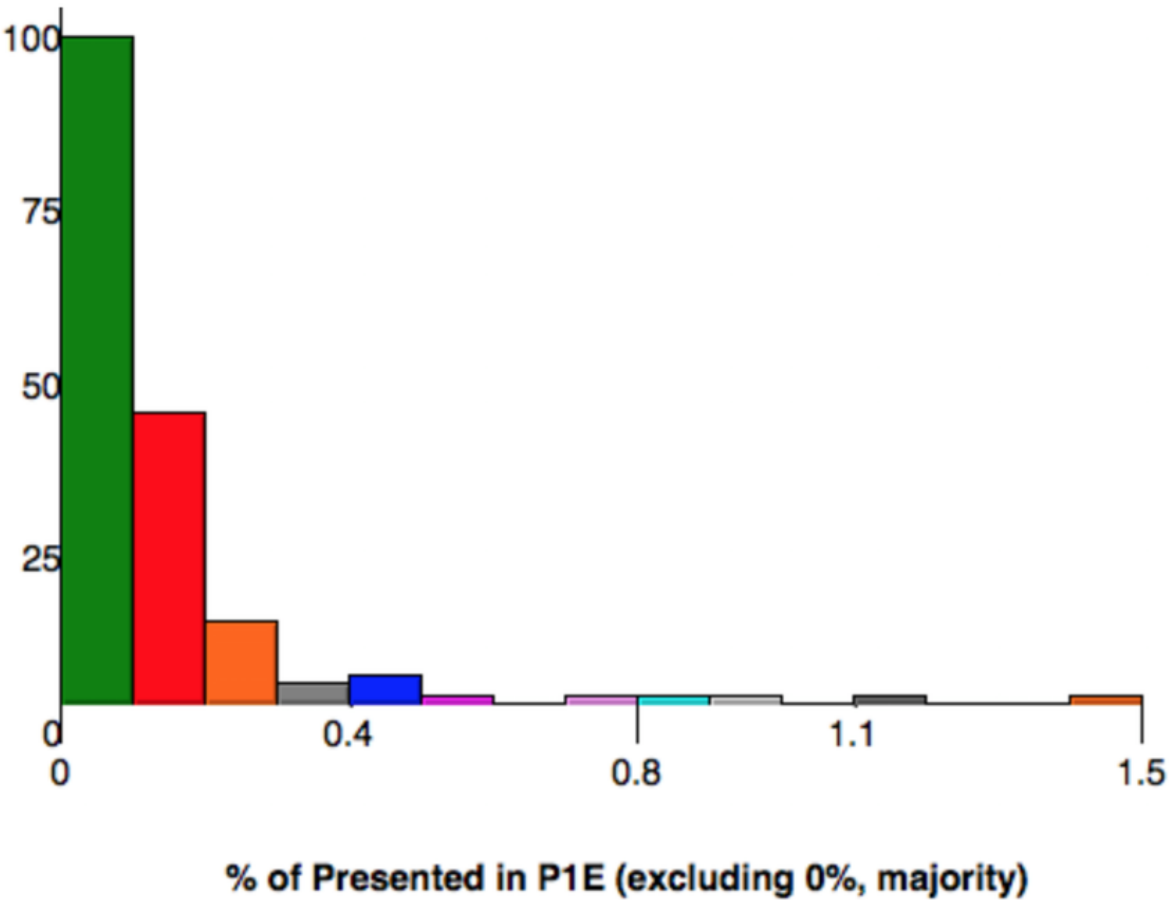


- Further:-
- Word Cloud list has been updated by Centrepont

SC



- ED
- Looking at the P1E (the forbidden data-set!), here's what the oversight looks like (warning, ugly graph ahead - nicer one coming soon) = the proportion of what we have now as Presented numbers accounted for in the official stat. In a nutshell: not much!!
- [excluding the 0% values i.e. when P1E say 'none' and we say otherwise: this is 50% of the districts!]



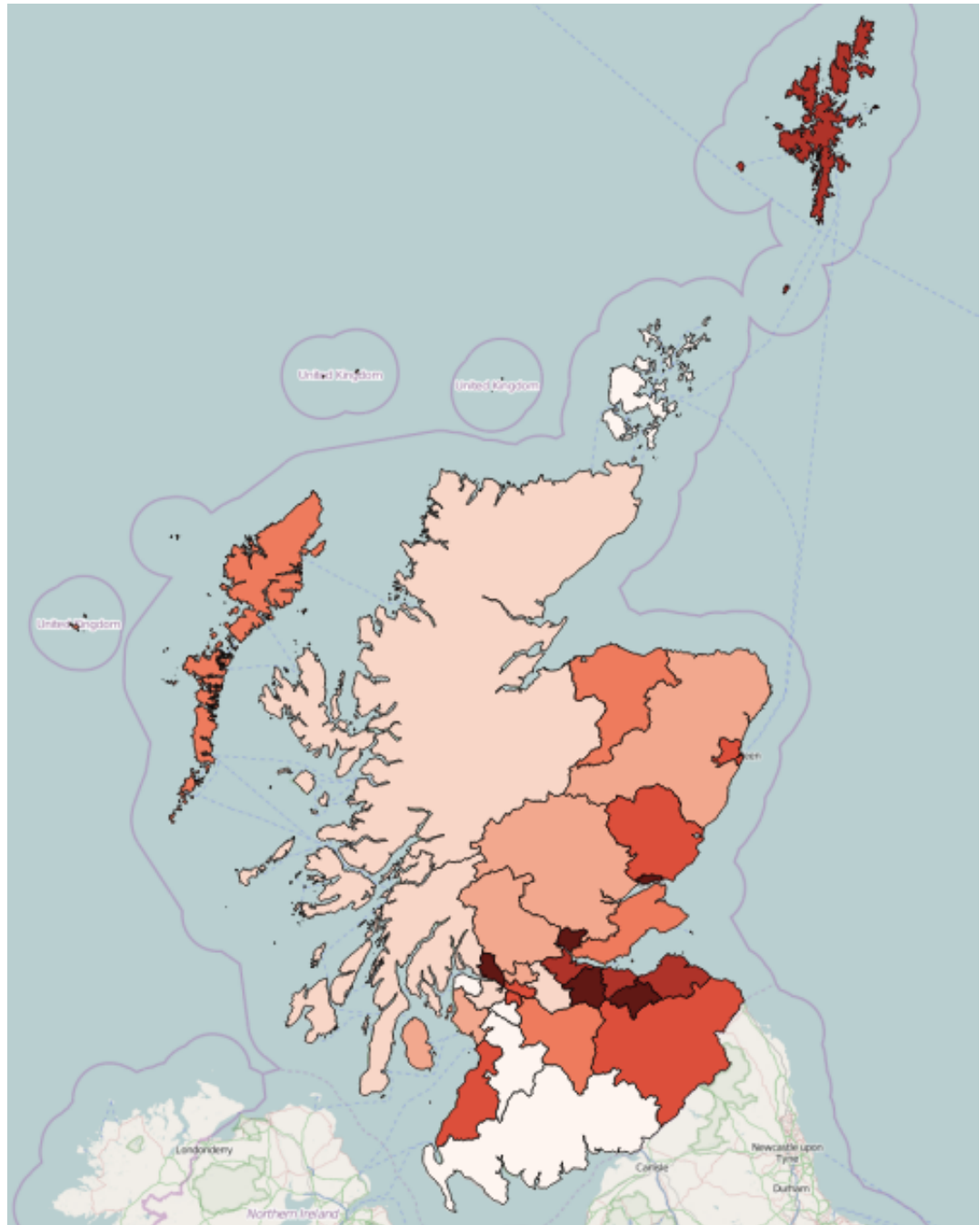
- CL
- Further data has been located and figures have been updated where appropriate on the flow diagram and this can be found above

- Code will be uploaded onto Github!

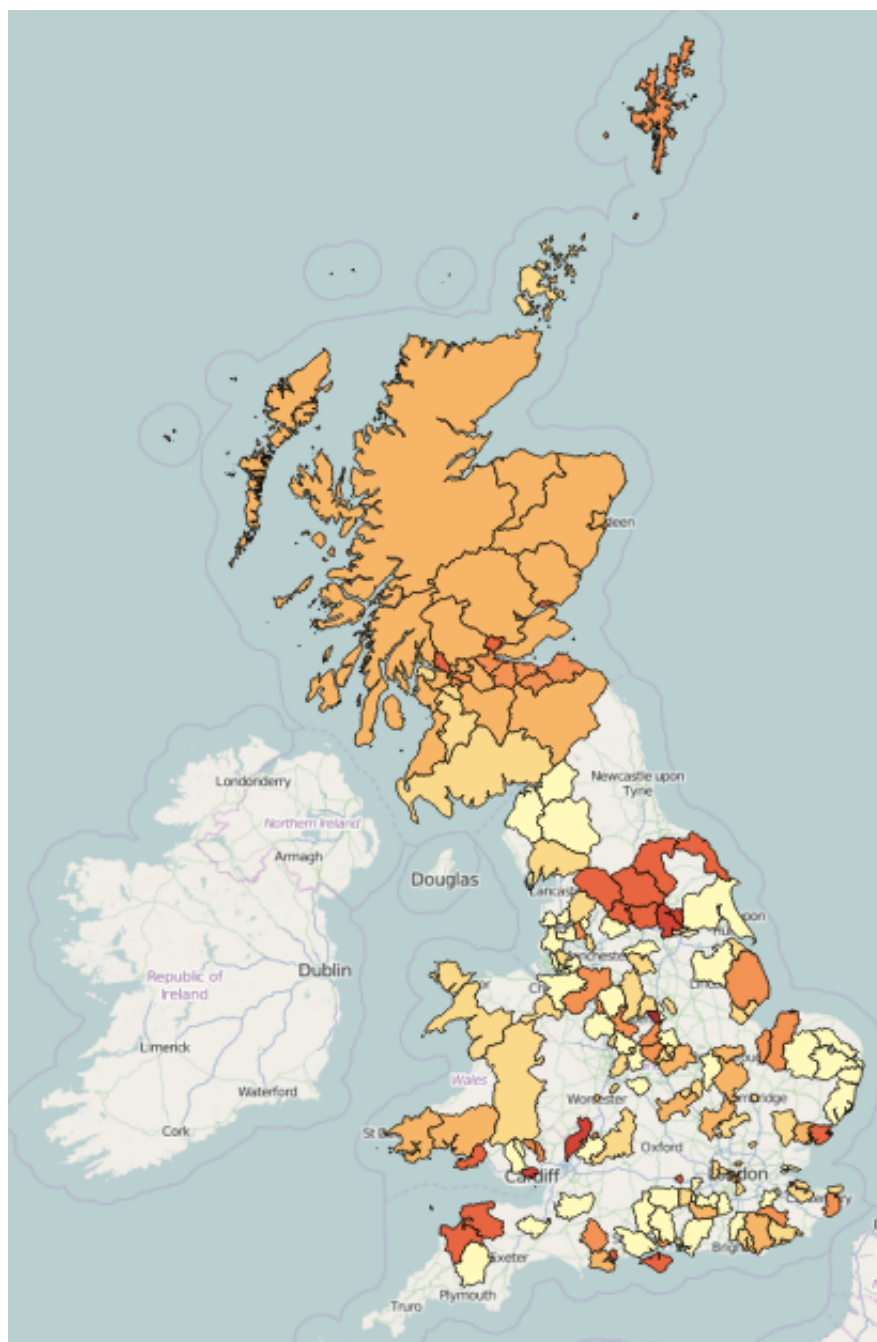
KD **Choropleth Maps**

VE *Normalised Density Ratio. We have adjusted population and presented people by size of the local authority area, created a ratio between the two factors and indexed it.*

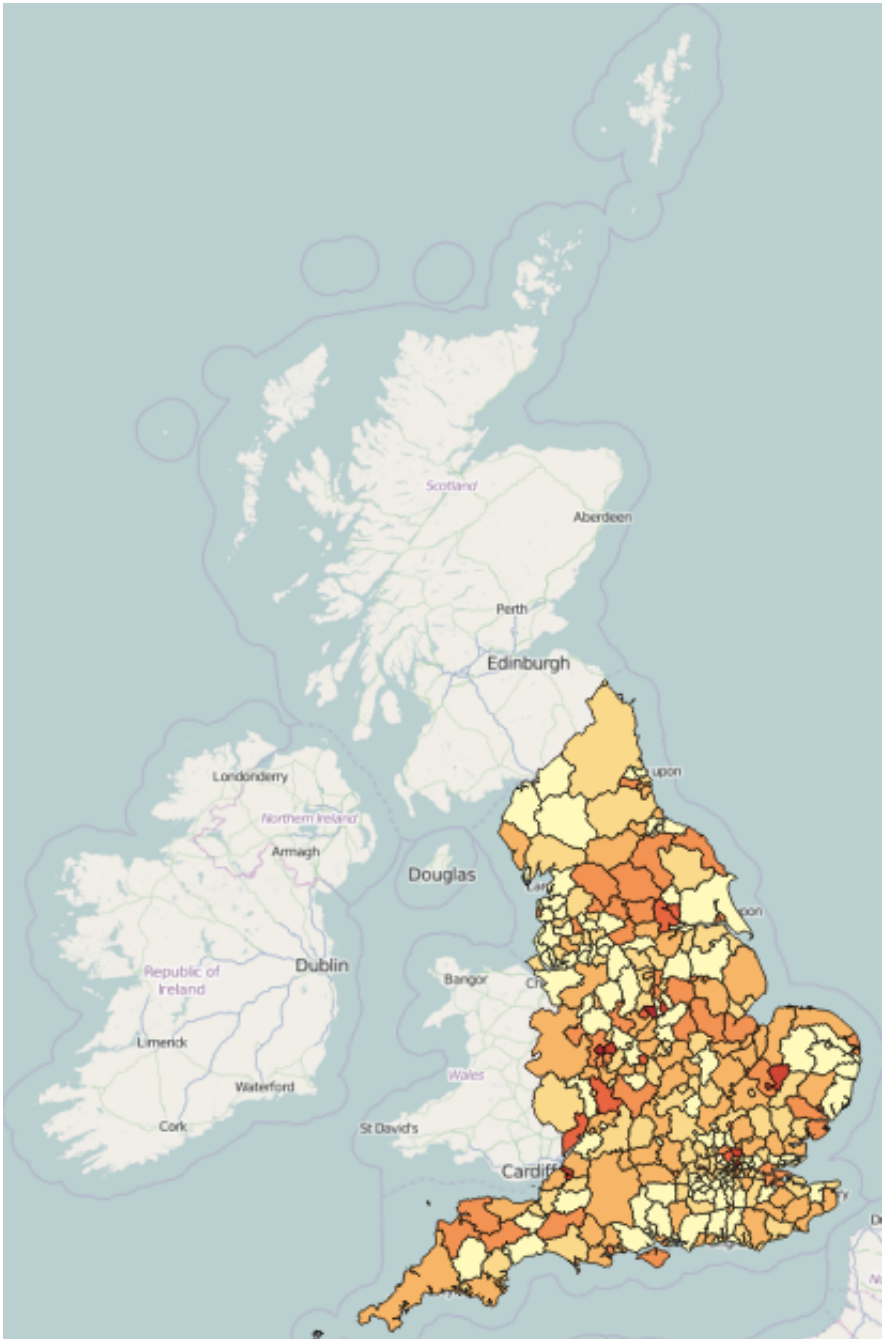
KD



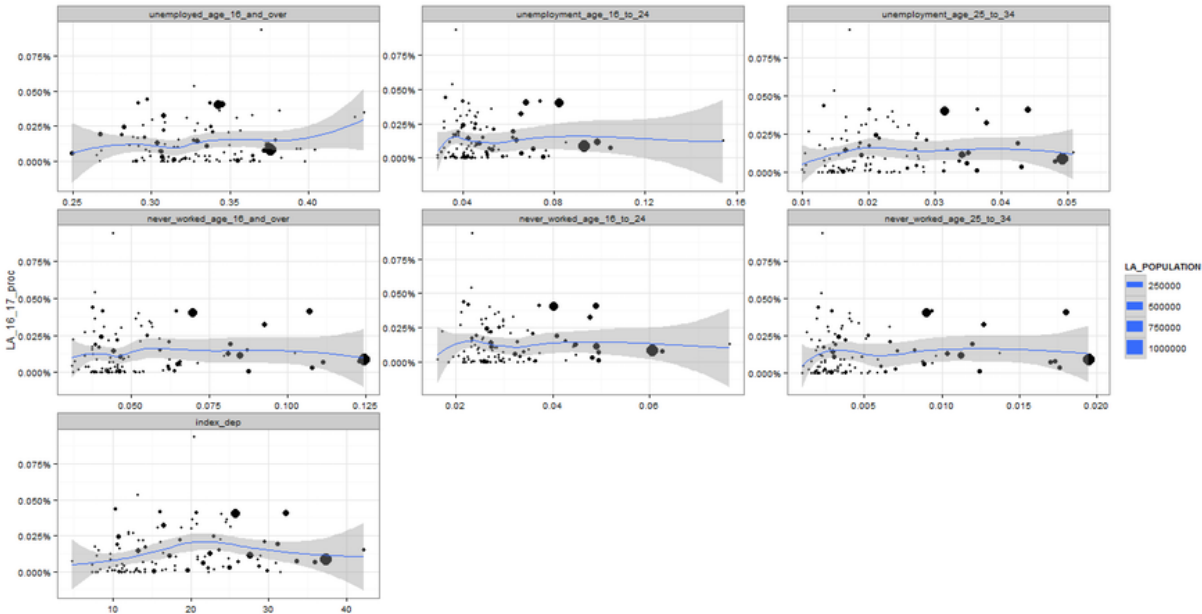
KD *presented cases (indexed) & missing data*



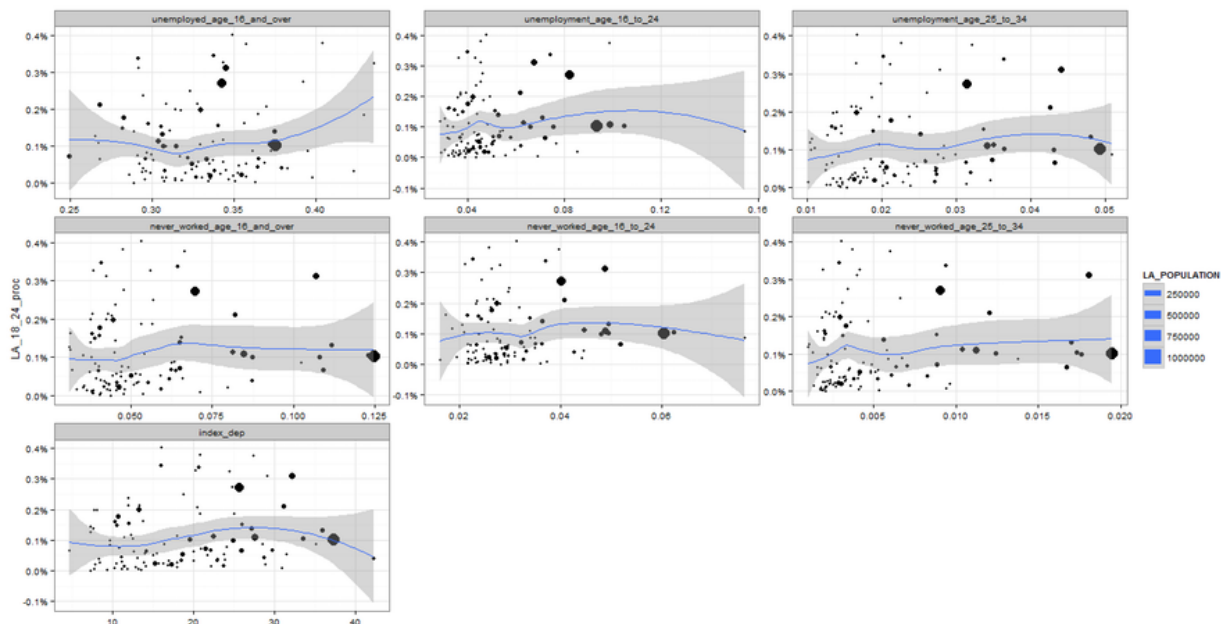
filling the gaps by populating data (presented cases)



JF



JF



AK Assumption made by the data preparation team along the way: $<5 \rightarrow 2.5$, $<10 \rightarrow 5$ (Jaz correct me if I'm wrong).

JF Youth homelessness prediction:

JF

Model type	Total estimate	Notes	Mean absolute error	Median absolute error
Mean base line (EN + WA)	75686	Using avg "presented" per district from known districts	156	107
Mice - pmm (EN + WA)	75445.74			
Mice - norm, (EN + WA)	94712.62			

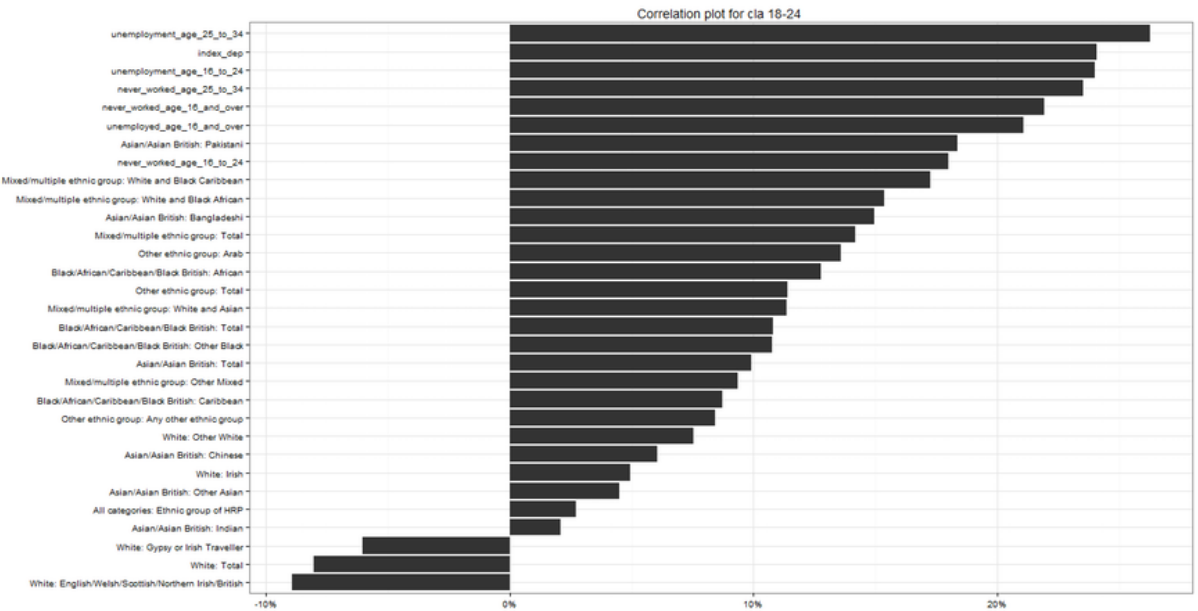
AK todo: (Adam) estimate error bars of the first model

MW

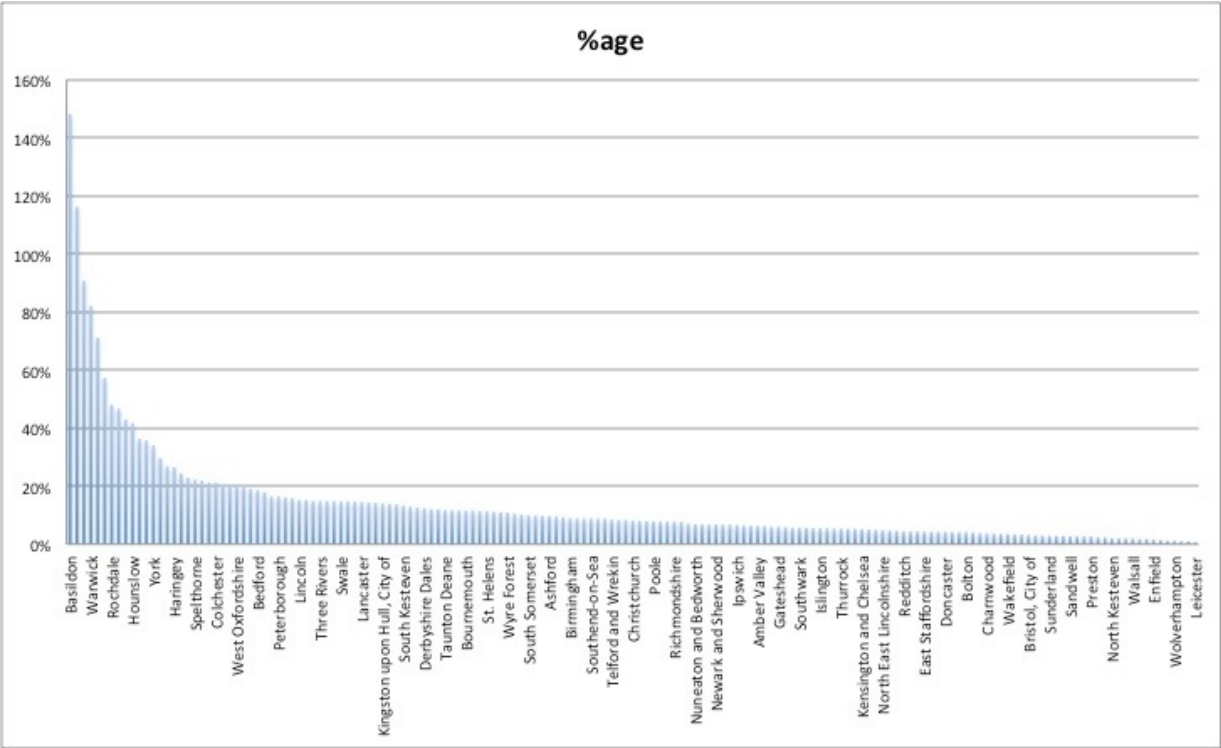


Vacant dwellings has the second highest correlation. How perverse! Also, "Households found to be 'intentionally houseless'".

JF



PN



Reported % presented cases accepted, reported by Local Authority
Almost 50% of values are shown in in the P1E as zero