

Pilot K means for general practices - version 2

Julian Flowers

7 December 2015

There are 8000 general practices in England and data for each practice is increasingly available in the public domain. There are however few tools which enable analysts and practices to compare practices with similar peers to help evaluate their performance. A previous tool developed by APHO seemingly considerably misclassified practices, and it was agreed by the National Practice Benchmarking and Indicators group that the issue of finding peer groups should be revisited.

This short note is a pilot evaluation of a simple method based on k-means analysis of an extract of the national general practice profiles. The NGPP contain about 250 publicly available indicators of the health, utilisation, demography and characteristics of general practice populations. For this analysis and to keep things simple we have chosen a few demographic variables - total population, age break down and deprivation as a proof of concept before trying the larger list of classification variables for which there is general consensus should be included in such a peer grouping exercise.

K-means analysis is a relatively simple and widely used for what is known as unsupervised machine learning. It clusters data together on the basis of similarity among all predictor variables into a pre-specified number of groups. The idea is that practices are clustered into groups which are more similar to each other on the variables of interest, and farther apart from other practice clusters.

Method

The data for this analysis was extracted from the spreadsheets downloaded from the [NGPP website](#) and included the following variables for each practice:

- Practice code
- CCG
- Total registered population (2014)
- IMD 2015 practice score (% of population with some form of deprivation)
- % population <5
- % population 5-14
- % population 65+
- % population 75+
- % population 85+
- & population under 18
- ethnic population

The analysis has been conducted in R and this report written in R Markdown using the R Studio package to allow us to embed relevant code which can then be shared and easily modified if we wish to change the variables or analysis.

It updates the previous analysis to latest data for populations and IMD scores and adds ethnicity (% population white)

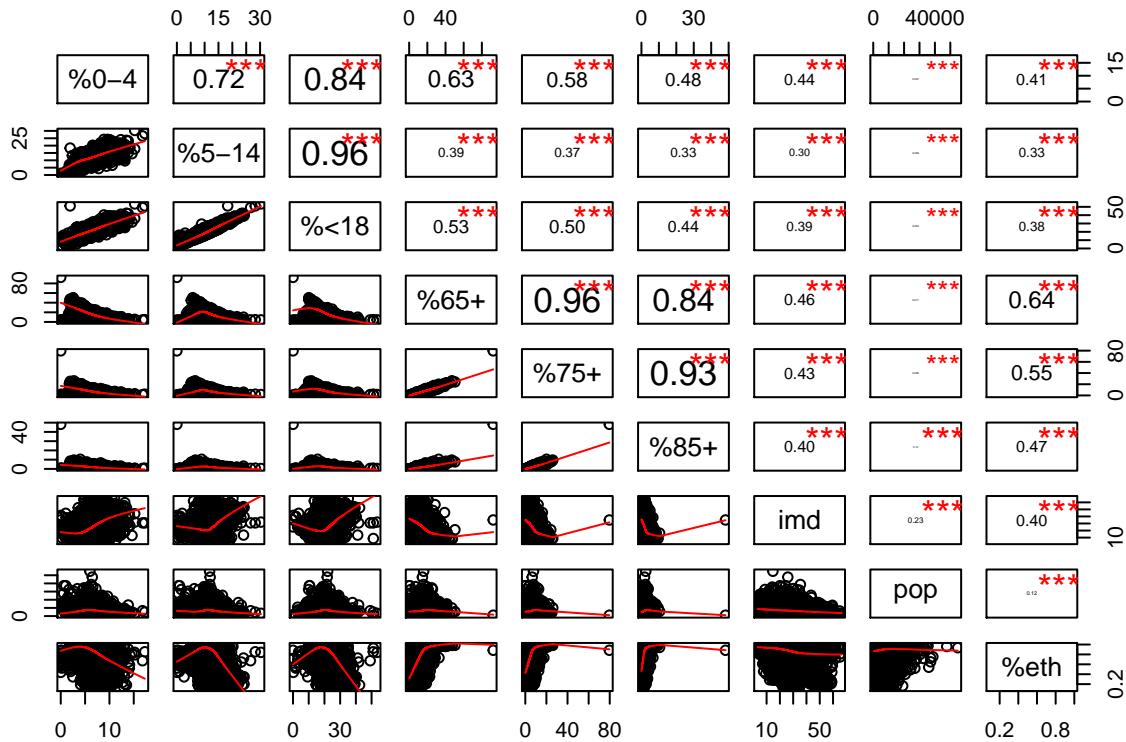
The dataset only contains 7750 practices because it excludes small practices and those with discrepancies between QOF reported practice size and the registered population.

The dataset

We can summarise the dataset and see that there are 13 missing % in work values - we won't use this variable.

%0-4	%5-14	%75+	imd	pop	%eth
Min. : 0.000	Min. : 0.00	Min. : 0.000	Min. : 3.213	Min. : 294	Min. : 0.09501
1st Qu.: 4.888	1st Qu.: 10.11	1st Qu.: 5.366	1st Qu.: 13.987	1st Qu.: 3956	1st Qu.: 0.75814
Median : 5.790	Median : 11.28	Median : 7.670	Median : 21.927	Median : 6514	Median : 0.92519
Mean : 5.985	Mean : 11.49	Mean : 7.623	Mean : 23.724	Mean : 7337	Mean : 0.83139
3rd Qu.: 6.869	3rd Qu.: 12.66	3rd Qu.: 9.671	3rd Qu.: 31.731	3rd Qu.: 9811	3rd Qu.: 0.97241
Max. : 17.268	Max. : 30.32	Max. : 79.648	Max. : 66.479	Max. : 54848	Max. : 0.99563

Exploring the relationship between the variables suggests that deprivation has a complex relationship with age structure. The under 18 variable is strongly correlated with %0-4 and %5-14 so to simplify the analysis we will exclude this. Similarly we'll use 75+ and drop 65+ and 85+.



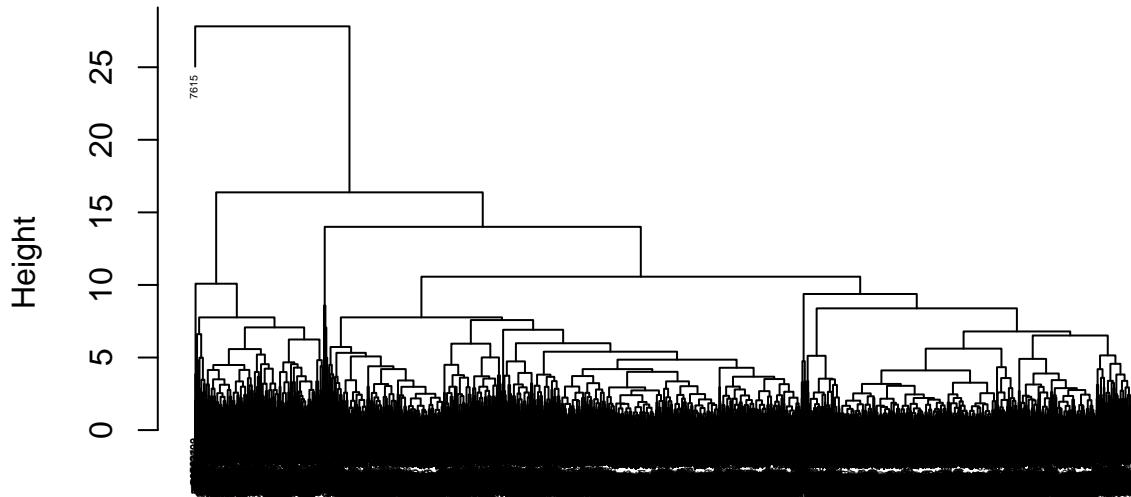
There also seems to be an outlying practice with high proportions of older people. This practice is . Looking at the characteristics of this practice shows it is small with an exclusively older population, suggesting it is probably a nursing home practice. We'll keep it in for the analysis.

```
##      practice          ccg %0-4 %5-14 %<18    %65+   %75+   %85+     imd
## 7615 Y02625 NHS Salford CCG    0     0    0 92.507 79.648 48.196 35.006
##      pop    %eth
## 7615 1081 0.884061
```

Cluster analysis

A simple hierarchical cluster analysis produces the dendrogram below. Each branch is a general practice. The dendrogram partitions the data according to similarity - the further down the tree the more similar practices are. It is hard to see all the detail (the chart shows all 7750 practices) but it picks out the outlier (single branch at the top left of the chart) and suggests there are number of practice groupings. We can use this as a basis for assigning clusters, depending on how fine grained we want them to be. Note that we have scaled the data (z-scores) because clustering is sensitive to absolute values.

Cluster Dendrogram



```
gpd  
hclust (*, "complete")
```

K means analysis

For this part of the analysis, we'll exclude practices with no IMD scores. The k in k means to be specified and we'll arbitrarily start with 10. For context the average of each variable is:

```
require(knitr)  
round(apply(kgp[,-(1:2)], 2, function(x) mean(x, na.rm = TRUE)),2)  
  
##      %0-4    %5-14    %75+     imd      pop      %eth  
##      5.99   11.49    7.62   23.72  7336.64     0.83
```

Running the k means analysis

With 10 clusters

```
set.seed(1) ## this is needed because there is an element of random sampling  
k <- kmeans(scale(kgp[,-(1:2)]), 10)  
  
## Warning: did not converge in 10 iterations
```

Summary of results

```

require(knitr)
k$size

## [1] 1406 517 1162 668 164 1117 326 814 585 991

agg <- round.aggregate(kgp[,-(1:2)], by = list(cluster=k$cluster), mean),2) ## summary of the results
agg$cluster <-as.factor(agg$cluster)
kable(agg[,-8])

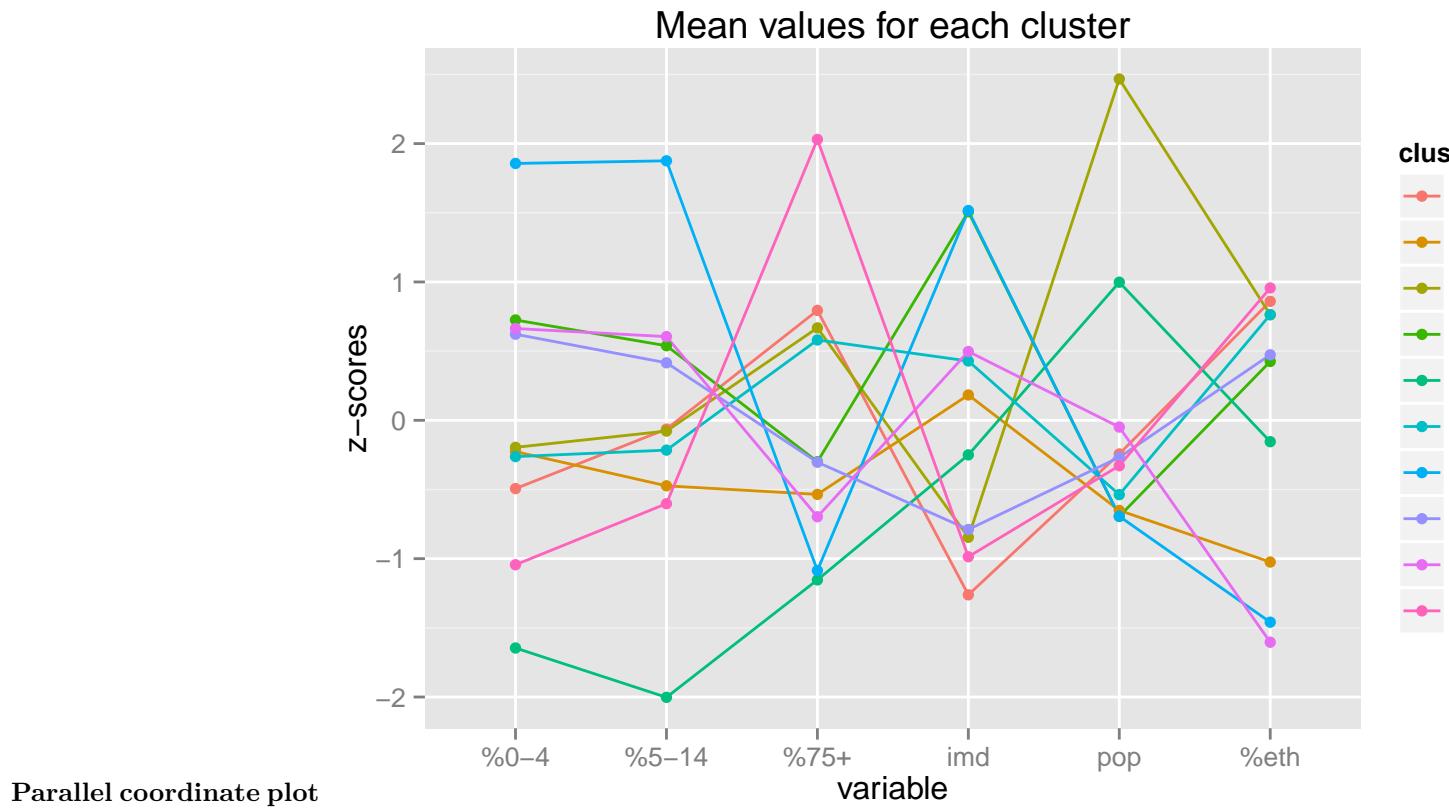
```

cluster	%0-4	%5-14	%75+	imd	pop	%eth
1	5.19	11.24	8.93	13.25	6457.22	0.94
2	5.71	9.81	4.82	28.43	5285.68	0.55
3	5.77	11.19	8.54	17.62	14237.14	0.92
4	7.56	13.34	5.55	42.35	5161.78	0.85
5	2.95	4.48	2.91	23.88	10020.24	0.73
6	5.64	10.71	8.27	31.01	5612.78	0.92
7	9.76	18.00	3.12	42.44	5164.04	0.46
8	7.36	12.91	5.54	18.21	6387.54	0.86
9	7.44	13.57	4.32	31.73	7014.81	0.43
10	4.12	9.36	12.75	16.14	6212.32	0.96

```

require(ggplot2)
require(GGally)
ggparcoord(agg, columns = c(2:7), scale = 'std',scaleSummary = mean, showPoints = TRUE, groupColumn = 1)
  ggttitle("Mean values for each cluster")+
  ylab("z-scores") ## parallel coorindates plot of results of mean values for each cluster

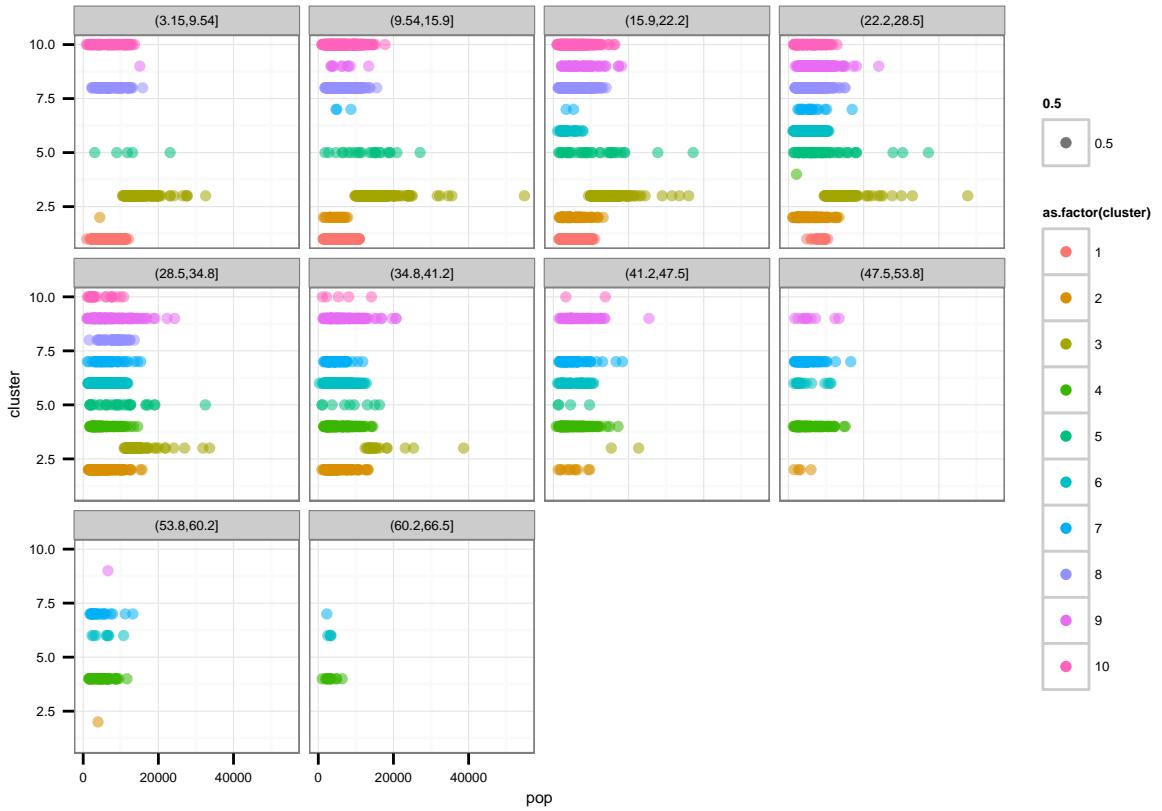
```



We can begin to see the nature of the clusters this method identifies. For example, the second smallest is cluster 6 with 283 practices. These are on average characterised by average practice size, slightly higher levels of deprivation, and a “middle aged” age distribution (the proportion of the population either young or old is much lower than average). The smallest, cluster 9 with 190 practices, clusters the larger practices with average levels of the other variables. Cluster 5 is the largest group - these are smaller practices with a “typical” age distribution and so on. Cluster 1 and 2 are larger practices - cluster 2 tends to be older and more deprived than cluster 1.

It is possible to create qualitative labels for each group, and enrich the clustering with additional variables - ethnicity and rurality are of particular interest.

We can add the clusters back to the original data to identify which practice is in which cluster and do some more sense checking.



We can look at the distribution of practices within clusters by CCG. For example, most practices in Barking and Dagenham are in cluster 4, and most in Bradford are in cluster 3.

```
require(knitr)
t<- (with(kgp,table(ccg, cluster)))

kable(t)
```

	1	2	3	4	5	6	7	8	9	10
NHS Airedale, Wharfdale And Craven CCG	3	0	6	1	0	0	1	0	1	5
NHS Ashford CCG	4	0	3	0	0	0	0	5	0	2
NHS Aylesbury Vale CCG	10	0	6	0	0	0	1	2	0	0
NHS Barking And Dagenham CCG	0	1	0	6	0	1	16	0	15	0
NHS Barnet CCG	14	16	4	0	1	0	2	21	8	1
NHS Barnsley CCG	1	0	7	8	0	18	0	1	0	1
NHS Basildon And Brentwood CCG	7	0	7	12	0	6	0	7	0	5
NHS Bassetlaw CCG	3	0	3	1	0	1	0	0	0	4
NHS Bath And North East Somerset CCG	12	0	3	0	1	0	0	4	0	7
NHS Bedfordshire CCG	16	0	16	0	0	1	1	17	2	2
NHS Bexley CCG	9	0	6	0	0	0	2	7	1	2
NHS Birmingham Crosscity CCG	4	5	10	27	2	20	29	2	9	5
NHS Birmingham South And Central CCG	2	3	0	6	1	8	16	4	5	0
NHS Blackburn With Darwen CCG	1	0	2	5	0	9	5	2	4	0
NHS Blackpool CCG	0	0	3	6	0	11	0	0	0	2
NHS Bolton CCG	5	0	6	11	0	10	9	5	4	0
NHS Bracknell And Ascot CCG	5	0	3	0	0	0	0	7	0	0
NHS Bradford City CCG	0	1	0	0	1	0	24	0	1	0

	1	2	3	4	5	6	7	8	9	10
NHS Bradford Districts CCG	4	0	7	8	0	5	9	4	3	1
NHS Brent CCG	0	30	0	0	0	0	3	0	33	1
NHS Brighton And Hove CCG	9	0	5	3	12	9	0	5	0	2
NHS Bristol CCG	3	0	12	7	4	13	1	9	4	1
NHS Bromley CCG	15	3	6	3	0	2	0	9	2	5
NHS Bury CCG	6	0	2	5	0	5	1	12	0	2
NHS Calderdale CCG	3	0	7	6	0	5	2	2	0	1
NHS Cambridgeshire and Peterborough CCG	37	0	22	6	6	6	4	18	1	6
NHS Camden CCG	2	13	4	1	9	1	0	4	2	0
NHS Cannock Chase CCG	10	0	4	0	0	6	0	2	0	4
NHS Canterbury And Coastal CCG	6	0	8	0	1	0	0	1	0	5
NHS Castle Point And Rochford CCG	7	0	4	0	0	4	0	1	0	11
NHS Central London (Westminster) CCG	1	13	2	0	12	0	1	0	3	0
NHS Central Manchester CCG	0	4	0	3	4	0	6	2	13	0
NHS Chiltern CCG	16	0	9	0	0	0	0	9	0	0
NHS Chorley And South Ribble CCG	16	0	3	0	0	6	0	4	0	3
NHS City And Hackney CCG	0	23	0	0	2	0	5	0	13	0
NHS Coastal West Sussex CCG	5	0	15	0	0	3	0	2	0	29
NHS Corby CCG	0	0	1	0	0	2	0	2	0	0
NHS Coventry And Rugby CCG	11	5	8	6	3	12	5	7	8	6
NHS Crawley CCG	1	1	4	0	0	0	0	7	0	0
NHS Croydon CCG	3	12	6	5	0	0	2	4	23	3
NHS Cumbria CCG	12	0	9	0	0	25	0	0	0	34
NHS Darlington CCG	3	0	5	1	0	1	0	1	0	0
NHS Dartford, Gravesham And Swanley CCG	9	0	8	0	0	3	2	11	0	1
NHS Doncaster CCG	7	0	4	8	0	22	0	1	0	1
NHS Dorset CCG	20	0	14	0	2	7	0	4	0	52
NHS Dudley CCG	10	0	4	8	0	14	2	2	0	7
NHS Durham Dales, Easington And Sedgefield CCG	2	0	9	3	0	22	1	0	0	4
NHS Ealing CCG	1	28	2	0	1	0	3	9	35	0
NHS East And North Hertfordshire CCG	19	0	22	0	1	0	0	17	0	0
NHS East Lancashire CCG	5	0	7	8	0	27	7	2	1	1
NHS East Leicestershire And Rutland CCG	15	1	11	0	0	1	0	2	0	2
NHS East Riding Of Yorkshire CCG	12	0	7	1	0	5	0	0	0	12
NHS East Staffordshire CCG	8	0	3	0	0	1	0	5	0	2
NHS East Surrey CCG	7	0	6	0	0	0	0	5	0	0
NHS Eastbourne, Hailsham And Seaford CCG	3	0	6	0	0	1	0	1	0	10
NHS Eastern Cheshire CCG	8	0	5	0	0	0	0	2	0	7
NHS Enfield CCG	5	7	4	0	0	1	8	8	16	0
NHS Erewash CCG	4	0	3	0	0	3	0	1	0	1
NHS Fareham And Gosport CCG	8	0	3	0	0	0	0	3	0	7
NHS Fylde & Wyre CCG	3	0	1	0	0	4	0	0	0	12
NHS Gloucestershire CCG	36	0	13	3	1	1	0	4	0	25
NHS Great Yarmouth And Waveney CCG	1	0	6	3	0	7	0	1	0	8
NHS Greater Huddersfield CCG	14	2	2	1	1	7	2	3	5	2
NHS Greater Preston CCG	8	4	3	5	1	5	1	1	2	3
NHS Greenwich CCG	3	13	2	0	0	2	4	7	11	0
NHS Guildford And Waverley CCG	9	0	9	0	0	0	0	1	0	2
NHS Halton CCG	0	0	4	7	0	4	0	1	0	0
NHS Hambleton, Richmondshire And Whitby CCG	4	0	2	0	0	0	0	2	0	14
NHS Hammersmith And Fulham CCG	1	14	0	0	7	0	0	5	4	0
NHS Hardwick CCG	4	0	2	0	0	8	0	0	0	2

	1	2	3	4	5	6	7	8	9	10
NHS Haringey CCG	1	22	4	0	0	0	2	4	11	1
NHS Harrogate And Rural District CCG	11	0	3	0	0	0	0	0	0	4
NHS Harrow CCG	2	13	2	0	0	0	0	0	17	0
NHS Hartlepool And Stockton-On-Tees CCG	2	0	7	5	0	13	1	4	0	3
NHS Hastings And Rother CCG	1	0	1	4	0	16	0	1	0	9
NHS Havering CCG	12	2	5	4	0	5	0	9	1	10
NHS Herefordshire CCG	4	0	4	0	0	1	0	2	0	13
NHS Herts Valleys CCG	21	3	24	0	0	2	1	17	1	0
NHS Heywood, Middleton And Rochdale CCG	2	0	1	12	0	12	7	3	0	0
NHS High Weald Lewes Havens CCG	14	0	2	0	0	2	0	0	0	3
NHS Hillingdon CCG	11	2	3	0	1	0	2	12	17	0
NHS Horsham And Mid Sussex CCG	10	0	8	0	0	0	0	3	0	2
NHS Hounslow CCG	2	19	0	0	0	0	2	12	18	1
NHS Hull CCG	1	2	2	21	1	23	1	2	0	4
NHS Ipswich And East Suffolk CCG	10	0	15	1	0	1	0	0	0	14
NHS Isle Of Wight CCG	0	0	4	0	0	3	0	0	0	10
NHS Islington CCG	0	26	1	0	7	0	0	1	1	0
NHS Kernow CCG	7	0	15	1	0	20	0	0	0	26
NHS Kingston CCG	6	1	5	0	1	0	0	12	0	1
NHS Knowsley CCG	0	0	1	13	0	13	0	1	0	1
NHS Lambeth CCG	0	28	1	0	5	0	0	2	11	0
NHS Lancashire North CCG	1	0	5	0	1	4	0	0	0	2
NHS Leeds North CCG	7	2	3	1	1	0	0	3	7	3
NHS Leeds South And East CCG	5	0	5	15	0	5	7	1	0	4
NHS Leeds West CCG	7	1	7	4	5	8	0	3	0	3
NHS Leicester City CCG	1	9	1	11	3	4	7	3	22	1
NHS Lewisham CCG	0	19	1	0	0	1	2	1	17	0
NHS Lincolnshire East CCG	2	0	6	0	0	3	0	1	0	18
NHS Lincolnshire West CCG	11	0	0	0	0	14	0	2	0	6
NHS Liverpool CCG	6	2	3	38	3	34	0	2	0	5
NHS Luton CCG	2	2	1	0	0	0	9	5	12	0
NHS Mansfield And Ashfield CCG	2	0	5	2	0	16	0	2	0	1
NHS Medway CCG	11	0	3	9	0	11	1	17	0	4
NHS Merton CCG	4	4	3	0	0	0	0	6	8	0
NHS Mid Essex CCG	19	0	14	0	0	1	0	8	0	6
NHS Milton Keynes CCG	2	0	7	2	0	0	2	13	1	0
NHS Nene CCG	18	0	27	7	1	7	1	7	0	1
NHS Newark & Sherwood CCG	1	0	6	0	0	3	0	1	0	3
NHS Newbury And District CCG	5	0	4	0	0	0	0	2	0	0
NHS Newcastle And Gateshead CCG	12	0	7	5	4	27	2	3	2	3
NHS Newham CCG	0	15	0	0	0	0	5	0	40	0
NHS North & West Reading CCG	2	0	5	0	0	0	0	3	0	0
NHS North Derbyshire CCG	14	0	6	0	0	3	0	1	0	12
NHS North Durham CCG	6	0	6	0	1	13	0	2	0	3
NHS North East Essex CCG	6	0	6	1	1	7	0	10	0	11
NHS North East Hampshire And Farnham CCG	8	0	8	0	0	1	0	7	0	0
NHS North East Lincolnshire CCG	1	0	4	8	0	13	0	0	0	3
NHS North Hampshire CCG	5	0	9	0	0	0	0	5	0	1
NHS North Kirklees CCG	3	2	2	3	0	8	6	3	2	0
NHS North Lincolnshire CCG	4	0	7	3	0	4	0	1	0	1
NHS North Manchester CCG	0	2	0	21	2	2	6	0	3	0
NHS North Norfolk CCG	3	0	2	0	0	0	0	0	0	14

	1	2	3	4	5	6	7	8	9	10
NHS North Somerset CCG	4	0	3	1	0	1	0	5	0	11
NHS North Staffordshire CCG	11	0	4	2	1	3	0	2	0	10
NHS North Tyneside CCG	7	0	5	1	0	6	0	5	0	5
NHS North West Surrey CCG	15	0	13	0	0	0	1	9	1	3
NHS Northern, Eastern And Western Devon CCG	23	0	21	7	1	22	0	6	0	41
NHS Northumberland CCG	10	0	6	0	0	6	0	0	0	22
NHS Norwich CCG	4	0	5	3	2	5	0	2	0	2
NHS Nottingham City CCG	1	10	0	19	4	12	7	2	4	0
NHS Nottingham North And East CCG	6	0	5	0	0	3	0	2	0	5
NHS Nottingham West CCG	5	0	3	0	0	0	0	2	0	2
NHS Oldham CCG	3	1	5	14	0	3	12	5	2	0
NHS Oxfordshire CCG	30	2	21	1	7	1	0	15	0	2
NHS Portsmouth CCG	3	1	4	4	1	8	0	2	0	1
NHS Redbridge CCG	3	4	2	0	0	0	3	10	24	0
NHS Redditch And Bromsgrove CCG	7	0	7	1	0	1	0	3	0	3
NHS Richmond CCG	7	0	5	0	0	0	0	15	0	2
NHS Rotherham CCG	2	0	8	6	0	17	2	0	0	1
NHS Rushcliffe CCG	5	0	3	0	0	0	0	2	0	2
NHS Salford CCG	1	1	3	15	2	14	2	7	0	2
NHS Sandwell And West Birmingham CCG	1	13	5	17	1	8	20	6	28	2
NHS Scarborough And Ryedale CCG	2	0	2	1	0	5	0	0	0	5
NHS Sheffield CCG	18	3	9	18	2	14	6	4	3	10
NHS Shropshire CCG	12	0	4	0	0	0	0	3	0	25
NHS Slough CCG	0	1	0	0	0	0	2	0	13	0
NHS Solihull CCG	11	0	5	5	0	5	0	3	0	3
NHS Somerset CCG	20	0	12	0	0	3	0	9	0	31
NHS South Cheshire CCG	6	0	7	0	0	0	0	1	0	4
NHS South Devon And Torbay CCG	6	0	5	2	0	7	0	1	0	13
NHS South East Staffs And Seisdon Peninsular CCG	7	0	6	1	0	2	0	8	0	7
NHS South Eastern Hampshire CCG	9	0	5	2	0	2	0	1	0	6
NHS South Gloucestershire CCG	7	0	12	0	0	1	0	3	0	2
NHS South Kent Coast CCG	2	0	1	2	0	10	0	2	0	13
NHS South Lincolnshire CCG	4	0	7	0	0	0	0	0	0	4
NHS South Manchester CCG	0	2	2	12	2	3	0	1	2	0
NHS South Norfolk CCG	4	0	8	0	0	0	0	2	0	11
NHS South Reading CCG	1	4	1	0	1	0	0	12	1	0
NHS South Sefton CCG	4	0	0	9	0	9	0	0	0	9
NHS South Tees CCG	1	0	4	14	0	16	3	2	0	2
NHS South Tyneside CCG	1	0	3	1	0	18	0	0	0	4
NHS South Warwickshire CCG	16	1	8	0	0	0	0	3	0	8
NHS South West Lincolnshire CCG	11	0	3	0	0	0	0	2	0	3
NHS South Worcestershire CCG	10	0	9	0	1	0	0	2	0	10
NHS Southampton CCG	3	0	5	4	5	9	0	6	1	0
NHS Southend CCG	5	0	4	5	0	9	0	2	0	9
NHS Southern Derbyshire CCG	16	0	17	2	0	8	5	4	1	3
NHS Southport And Formby CCG	4	0	1	0	0	1	0	2	0	11
NHS Southwark CCG	0	21	0	0	4	0	0	4	14	1
NHS St Helens CCG	4	0	2	3	0	22	0	1	0	3
NHS Stafford And Surrounds CCG	4	0	7	0	0	0	0	0	0	3
NHS Stockport CCG	16	0	7	2	0	14	0	5	0	3
NHS Stoke On Trent CCG	5	0	4	20	1	15	1	2	0	4
NHS Sunderland CCG	3	1	3	4	0	29	0	5	0	6

	1	2	3	4	5	6	7	8	9	10
NHS Surrey Downs CCG	15	0	8	0	0	0	0	4	0	6
NHS Surrey Heath CCG	4	0	3	0	0	0	0	2	0	0
NHS Sutton CCG	10	0	6	1	0	1	1	8	0	0
NHS Swale CCG	5	0	1	4	0	5	0	5	0	0
NHS Swindon CCG	5	0	6	0	0	4	0	9	1	1
NHS Tameside And Glossop CCG	3	0	3	8	0	20	1	6	0	0
NHS Telford And Wrekin CCG	1	0	6	5	0	4	0	6	0	0
NHS Thanet CCG	1	0	3	5	0	4	0	1	0	6
NHS Thurrock CCG	7	0	2	5	0	3	4	8	0	4
NHS Tower Hamlets CCG	0	14	0	0	2	0	2	0	18	0
NHS Trafford CCG	16	0	6	2	0	0	0	6	5	0
NHS Vale Of York CCG	12	0	9	0	1	0	0	2	0	6
NHS Vale Royal CCG	5	0	1	1	0	2	0	3	0	0
NHS Wakefield CCG	5	1	15	3	0	15	0	1	0	0
NHS Walsall CCG	2	2	3	17	0	13	9	3	5	7
NHS Waltham Forest CCG	1	10	2	0	0	2	0	1	28	1
NHS Wandsworth CCG	0	15	9	0	6	1	0	7	3	0
NHS Warrington CCG	6	0	6	2	0	9	0	3	0	0
NHS Warwickshire North CCG	4	0	5	1	1	11	0	2	0	3
NHS West Cheshire CCG	14	0	4	2	1	5	0	1	0	9
NHS West Essex CCG	15	0	6	0	0	2	0	12	0	3
NHS West Hampshire CCG	13	0	19	0	0	0	0	5	0	14
NHS West Kent CCG	32	0	16	1	0	1	0	9	0	2
NHS West Lancashire CCG	7	0	1	4	0	1	0	0	0	9
NHS West Leicestershire CCG	26	1	12	0	1	2	0	6	0	2
NHS West London (K&C & QPP) CCG	3	28	1	0	7	1	0	4	5	2
NHS West Norfolk CCG	3	0	3	0	0	3	0	1	0	11
NHS West Suffolk CCG	4	0	9	0	0	0	0	4	0	8
NHS Wigan Borough CCG	9	0	4	10	0	30	1	7	0	2
NHS Wiltshire CCG	22	0	17	0	0	0	0	5	0	12
NHS Windsor, Ascot And Maidenhead CCG	6	0	7	0	0	0	0	6	0	0
NHS Wirral CCG	7	0	2	12	0	22	0	0	0	12
NHS Wokingham CCG	2	0	8	0	0	0	0	2	0	1
NHS Wolverhampton CCG	1	6	3	12	0	7	5	1	6	6
NHS Wyre Forest CCG	3	0	4	0	0	2	0	0	0	3

With 15 clusters

```
kgp<-kgp[, -c(9:10)]
set.seed(2) ## this is needed because there is an element of random sampling
k <- kmeans(scale(kgp[,-(1:2)]), 15)
```

Summary of results

```
require(knitr)
k$size

## [1] 481 167 905 948 945 251 758 604 657 110 489 754 159 235 287
```

```

agg <- round(aggregate(kgp[,-(1:2)], by = list(cluster=k$cluster), mean),2) ## summary of the results
agg$cluster <-as.factor(agg$cluster)
kable(agg[,-8])

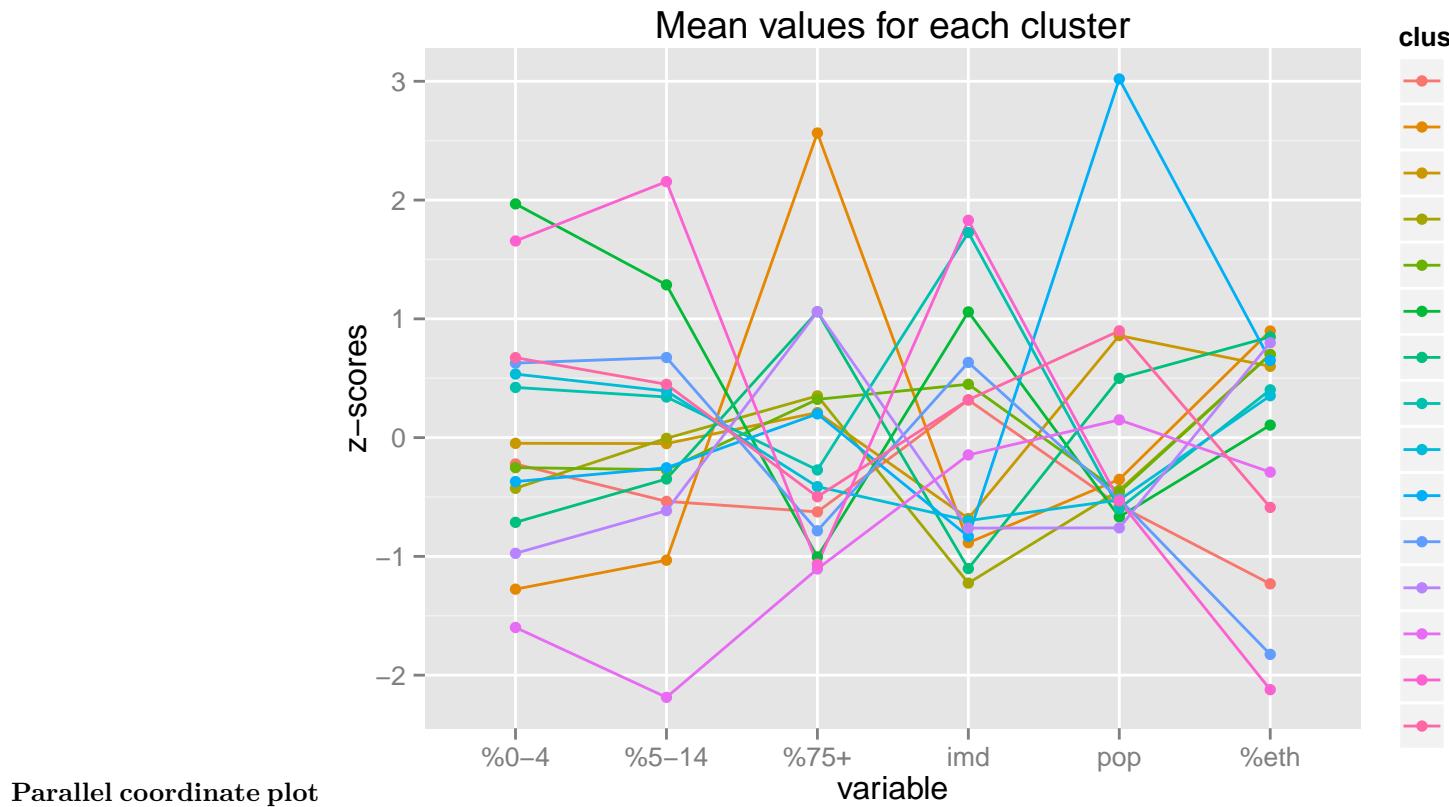
```

cluster	%0-4	%5-14	%75+	imd	pop	%eth
1	5.70	9.82	4.81	28.62	5174.72	0.54
2	3.64	8.24	16.83	16.13	6332.38	0.97
3	6.04	11.37	7.96	18.25	12694.49	0.91
4	5.30	11.51	8.49	12.60	5849.25	0.93
5	5.64	10.67	8.38	29.99	5763.80	0.93
6	9.98	15.63	3.38	36.32	4670.81	0.81
7	4.74	10.42	11.16	13.87	10803.67	0.96
8	6.96	12.62	6.14	43.26	5036.63	0.87
9	7.18	12.78	5.61	18.07	5419.71	0.86
10	5.41	10.72	7.92	16.68	24049.67	0.92
11	7.36	13.68	4.21	31.90	5462.69	0.42
12	4.23	9.57	11.16	17.41	4185.83	0.95
13	3.01	4.56	3.00	23.81	8964.45	0.73
14	9.37	18.40	3.15	44.33	5386.02	0.36
15	7.45	12.96	5.29	28.63	12904.88	0.67

```

require(ggplot2)
require(GGally)
ggparcoord(agg, columns = c(2:7), scale = 'std',scaleSummary = mean, showPoints = TRUE, groupColumn = 1,
           ggtitle("Mean values for each cluster")+
           ylab("z-scores") ## parallel coorindates plot of results of mean values for each cluster

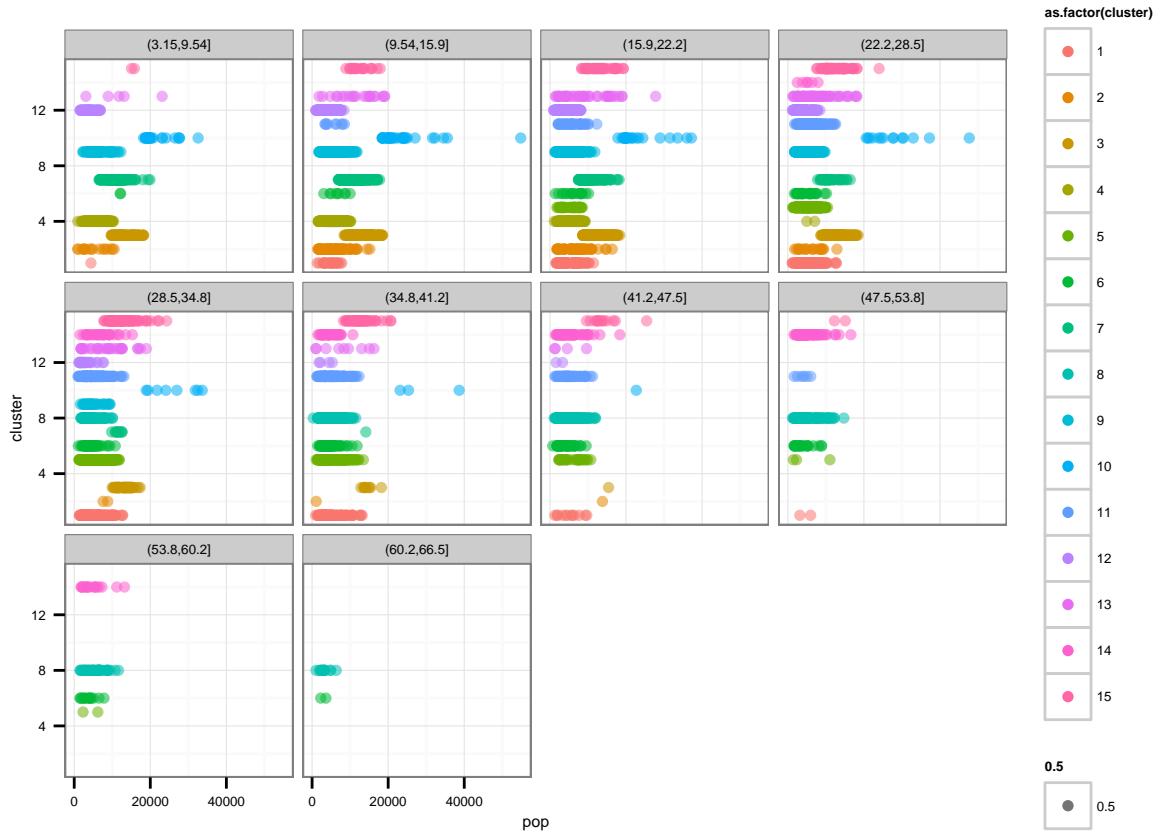
```



We can begin to see the nature of the clusters this method identifies. For example, the second smallest is cluster 6 with 283 practices. These are on average characterised by average practice size, slightly higher levels of deprivation, and a “middle aged” age distribution (the proportion of the population either young or old is much lower than average). The smallest, cluster 9 with 190 practices, clusters the larger practices with average levels of the other variables. Cluster 5 is the largest group - these are smaller practices with a “typical” age distribution and so on. Cluster 1 and 2 are larger practices - cluster 2 tends to be older and more deprived than cluster 1.

It is possible to create qualitative labels for each group, and enrich the clustering with additional variables - ethnicity and rurality are of particular interest.

We can add the clusters back to the original data to identify which practice is in which cluster and do some more sense checking.



We can look at the distribution of practices within clusters by CCG. For example, most practices in Barking and Dagenham are in cluster 4, and most in Bradford are in cluster 3.

```
require(knitr)
t<- (with(kgp,table(ccg, cluster)))

kable(t)
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NHS Airedale, Wharfdale And Craven CCG	0	2	0	3	0	0	6	0	0	0	0	1	0	1
NHS Ashford CCG	0	0	2	4	0	1	1	0	4	1	0	1	0	0
NHS Aylesbury Vale CCG	0	0	2	7	0	1	4	0	0	2	0	0	0	0
NHS Barking And Dagenham CCG	1	0	0	0	1	14	0	2	0	0	13	0	0	6
NHS Barnet CCG	14	0	4	12	0	1	0	0	18	0	5	4	1	1
NHS Barnsley CCG	0	0	7	1	15	2	0	7	2	0	0	2	0	0
NHS Basildon And Brentwood CCG	0	1	5	3	5	3	5	11	6	0	0	5	0	0
NHS Bassetlaw CCG	0	0	1	1	0	1	1	0	0	2	0	6	0	0
NHS Bath And North East Somerset CCG	0	0	0	12	0	0	8	0	3	0	0	3	1	0
NHS Bedfordshire CCG	0	0	15	11	1	2	6	0	12	0	2	3	0	0
NHS Bexley CCG	0	0	6	4	0	2	6	0	7	0	0	1	0	0
NHS Birmingham Crosscity CCG	4	0	4	4	16	5	5	28	2	3	8	3	2	26
NHS Birmingham South And Central CCG	3	0	2	0	5	1	0	6	2	0	4	1	1	17
NHS Blackburn With Darwen CCG	0	0	3	1	7	3	0	1	2	0	4	1	0	5
NHS Blackpool CCG	0	2	2	0	9	1	1	6	0	0	0	1	0	0
NHS Bolton CCG	0	0	4	3	9	2	2	9	5	1	5	1	0	7
NHS Bracknell And Ascot CCG	0	0	2	4	0	0	1	0	7	1	0	0	0	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NHS Bradford City CCG	1	0	0	0	0	0	0	0	0	0	1	0	1	24
NHS Bradford Districts CCG	0	0	6	3	5	2	2	6	2	0	1	0	0	8
NHS Brent CCG	32	0	0	0	0	0	0	0	0	0	28	1	0	3
NHS Brighton And Hove CCG	0	0	8	6	6	0	3	4	4	0	0	2	12	0
NHS Bristol CCG	0	0	13	1	11	3	1	4	4	0	2	2	4	1
NHS Bromley CCG	3	0	6	10	2	3	8	0	6	0	2	3	0	0
NHS Bury CCG	0	0	4	4	5	4	1	1	12	0	0	2	0	0
NHS Calderdale CCG	0	0	8	0	4	2	1	3	2	1	0	2	0	2
NHS Cambridgeshire and Peterborough CCG	0	0	17	30	6	8	9	2	12	2	0	7	6	0
NHS Camden CCG	10	0	2	2	1	0	0	0	4	0	4	0	10	0
NHS Cannock Chase CCG	0	0	4	7	4	0	0	0	3	0	0	8	0	0
NHS Canterbury And Coastal CCG	0	0	5	4	0	0	4	0	1	1	0	5	1	0
NHS Castle Point And Rochford CCG	0	0	1	5	4	0	6	0	1	1	0	9	0	0
NHS Central London (Westminster) CCG	13	0	1	0	0	0	0	0	0	0	1	1	13	1
NHS Central Manchester CCG	4	0	1	0	0	1	0	2	1	0	12	0	4	6
NHS Chiltern CCG	0	0	7	12	0	0	5	0	6	1	0	0	0	0
NHS Chorley And South Ribble CCG	0	0	3	11	5	1	3	1	4	0	0	4	0	0
NHS City And Hackney CCG	21	0	0	0	0	2	0	0	0	0	11	0	2	3
NHS Coastal West Sussex CCG	0	14	10	2	5	0	17	0	2	0	0	4	0	0
NHS Corby CCG	0	0	0	0	2	0	0	0	2	1	0	0	0	0
NHS Coventry And Rugby CCG	5	1	8	8	12	5	1	5	4	1	7	5	3	1
NHS Crawley CCG	1	0	5	1	0	1	0	0	2	0	0	0	0	0
NHS Croydon CCG	8	0	4	3	1	3	1	2	4	0	19	4	0	2
NHS Cumbria CCG	0	5	6	4	21	0	8	0	0	1	0	35	0	0
NHS Darlington CCG	0	0	5	2	1	0	2	1	0	0	0	0	0	0
NHS Dartford, Gravesham And Swanley CCG	0	0	4	7	3	2	4	0	9	0	0	1	0	0
NHS Doncaster CCG	0	0	4	7	22	1	1	6	2	0	0	0	0	0
NHS Dorset CCG	0	19	11	9	8	0	19	0	3	2	0	26	2	0
NHS Dudley CCG	0	1	2	5	14	4	6	6	1	2	0	5	0	1
NHS Durham Dales, Easington And Sedgefield CCG	0	0	8	0	20	1	2	4	0	0	0	6	0	0
NHS Ealing CCG	26	0	3	1	0	0	0	0	6	0	37	0	1	2
NHS East And North Hertfordshire CCG	0	0	19	14	1	1	7	0	12	3	0	0	1	0
NHS East Lancashire CCG	0	0	5	4	19	2	4	13	2	0	0	1	0	6
NHS East Leicestershire And Rutland CCG	1	0	5	7	1	0	12	0	1	2	0	3	0	0
NHS East Riding Of Yorkshire CCG	0	5	2	8	4	0	11	1	0	1	0	5	0	0
NHS East Staffordshire CCG	0	0	4	5	1	1	2	0	2	0	0	2	0	0
NHS East Surrey CCG	0	0	9	4	0	0	1	0	3	0	0	1	0	0
NHS Eastbourne, Hailsham And Seaford CCG	0	5	2	1	1	0	7	0	1	0	0	4	0	0
NHS Eastern Cheshire CCG	0	1	4	3	0	0	11	0	1	0	0	2	0	0
NHS Enfield CCG	6	0	5	4	1	2	0	0	7	0	13	0	0	6
NHS Erewash CCG	0	0	3	1	3	0	2	0	1	0	0	2	0	0
NHS Fareham And Gosport CCG	0	1	6	4	1	1	6	0	1	0	0	1	0	0
NHS Fylde & Wyre CCG	0	6	1	0	4	0	7	0	0	0	0	2	0	0
NHS Gloucestershire CCG	0	1	12	25	1	2	16	1	3	1	0	19	1	0
NHS Great Yarmouth And Waveney CCG	0	3	2	0	6	0	7	4	1	1	0	2	0	0
NHS Greater Huddersfield CCG	2	1	3	12	6	0	1	0	2	0	4	2	1	2
NHS Greater Preston CCG	4	0	2	3	6	3	5	2	1	0	3	2	1	0
NHS Greenwich CCG	13	0	1	2	2	2	0	0	3	0	10	1	0	1
NHS Guildford And Waverley CCG	0	0	7	6	0	0	5	0	1	1	0	1	0	0
NHS Halton CCG	0	0	4	0	3	0	0	8	1	0	0	0	0	0
NHS Hambleton, Richmondshire And Whitby CCG	0	0	0	1	0	1	5	0	1	1	0	13	0	0
NHS Hammersmith And Fulham CCG	13	0	0	1	0	0	0	0	3	0	4	0	7	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NHS Hardwick CCG	0	0	3	2	8	0	1	0	0	0	0	2	0	0
NHS Haringey CCG	20	0	2	0	0	0	0	0	5	0	8	2	0	2
NHS Harrogate And Rural District CCG	0	0	0	8	0	0	8	0	0	1	0	1	0	0
NHS Harrow CCG	13	0	0	0	0	0	1	0	0	1	11	2	0	0
NHS Hartlepool And Stockton-On-Tees CCG	0	1	4	2	11	4	1	5	3	2	0	2	0	0
NHS Hastings And Rother CCG	0	3	1	1	12	0	2	4	1	0	0	8	0	0
NHS Havering CCG	2	0	6	9	3	3	3	4	7	0	1	10	0	0
NHS Herefordshire CCG	0	1	2	1	1	0	7	0	2	0	0	10	0	0
NHS Herts Valleys CCG	3	0	24	15	1	2	3	0	12	3	1	1	0	0
NHS Heywood, Middleton And Rochdale CCG	0	0	1	0	11	3	0	10	4	0	0	0	0	6
NHS High Weald Lewes Havens CCG	0	0	1	8	2	0	8	0	0	0	0	2	0	0
NHS Hillingdon CCG	1	0	3	10	0	0	1	0	11	0	16	0	1	2
NHS Horsham And Mid Sussex CCG	0	0	4	8	0	0	5	0	3	2	0	1	0	0
NHS Hounslow CCG	17	0	2	1	0	2	0	0	11	0	18	2	0	1
NHS Hull CCG	0	1	1	0	19	5	0	23	1	0	0	5	1	0
NHS Ipswich And East Suffolk CCG	0	3	10	7	2	0	10	1	0	0	0	6	0	0
NHS Isle Of Wight CCG	0	3	3	0	2	0	5	0	0	0	0	4	0	0
NHS Islington CCG	25	0	0	0	0	0	0	0	0	0	0	1	0	7
NHS Kernow CCG	0	3	12	1	14	1	14	1	0	1	0	22	0	0
NHS Kingston CCG	1	0	4	7	0	0	0	0	11	1	0	1	0	0
NHS Knowsley CCG	0	0	0	0	8	0	0	18	1	1	0	1	0	0
NHS Lambeth CCG	27	0	0	0	0	0	0	0	1	0	8	0	5	0
NHS Lancashire North CCG	0	0	3	1	4	0	2	0	0	1	0	1	1	0
NHS Leeds North CCG	1	0	2	5	0	1	2	0	2	1	3	4	1	1
NHS Leeds South And East CCG	0	0	3	3	3	3	2	13	1	0	0	5	0	5
NHS Leeds West CCG	1	0	6	5	5	0	2	4	2	2	0	5	4	0
NHS Leicester City CCG	9	0	1	0	4	1	0	8	3	0	19	2	3	6
NHS Lewisham CCG	18	0	1	0	1	1	0	0	1	0	8	0	0	1
NHS Lincolnshire East CCG	0	3	4	1	3	0	5	0	1	1	0	12	0	0
NHS Lincolnshire West CCG	0	0	2	9	12	0	4	0	2	0	0	4	0	0
NHS Liverpool CCG	1	0	3	4	20	3	1	52	2	1	0	4	2	0
NHS Luton CCG	2	0	1	2	0	2	0	0	5	0	8	0	0	6
NHS Mansfield And Ashfield CCG	0	0	6	2	14	1	0	3	1	0	0	1	0	0
NHS Medway CCG	0	0	4	8	10	6	1	4	15	1	0	6	0	0
NHS Merton CCG	4	0	4	3	0	0	0	0	6	0	5	0	0	0
NHS Mid Essex CCG	0	0	11	15	1	0	7	0	7	0	0	7	0	0
NHS Milton Keynes CCG	0	0	6	2	0	3	0	1	5	1	1	0	0	0
NHS Nene CCG	0	0	17	19	6	3	2	4	5	4	0	2	1	0
NHS Newark & Sherwood CCG	0	0	5	1	3	0	2	0	1	0	0	2	0	0
NHS Newbury And District CCG	0	0	5	5	0	0	0	0	1	0	0	0	0	0
NHS Newcastle And Gateshead CCG	0	0	5	5	22	3	3	7	2	1	0	10	3	1
NHS Newham CCG	15	0	0	0	0	0	0	0	0	0	0	37	0	6
NHS North & West Reading CCG	0	0	6	1	0	0	0	0	2	0	0	0	0	0
NHS North Derbyshire CCG	0	2	5	6	3	0	11	0	1	1	0	7	0	0
NHS North Durham CCG	0	0	5	3	11	0	4	0	3	3	0	2	0	0
NHS North East Essex CCG	0	5	3	4	4	4	7	2	6	0	0	6	1	0
NHS North East Hampshire And Farnham CCG	0	0	9	5	1	0	3	0	6	0	0	0	0	0
NHS North East Lincolnshire CCG	0	0	3	1	12	1	2	8	0	0	0	2	0	0
NHS North Hampshire CCG	0	0	5	4	0	1	2	0	4	3	0	1	0	0
NHS North Kirklees CCG	2	0	3	1	8	1	0	1	3	0	2	0	0	4
NHS North Lincolnshire CCG	0	0	3	1	4	2	5	1	1	0	0	3	0	0
NHS North Manchester CCG	1	0	0	0	1	4	0	21	0	0	3	0	2	4

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NHS North Norfolk CCG	0	7	1	1	0	0	7	0	0	0	0	3	0	0
NHS North Somerset CCG	0	1	1	2	1	2	11	0	3	0	0	4	0	0
NHS North Staffordshire CCG	0	0	1	5	3	0	8	2	2	0	0	11	1	0
NHS North Tyneside CCG	0	1	3	6	6	1	3	0	5	0	0	4	0	0
NHS North West Surrey CCG	0	0	15	11	0	1	4	0	5	0	2	3	0	0
NHS Northern, Eastern And Western Devon CCG	0	8	13	13	19	0	19	9	8	2	0	29	1	0
NHS Northumberland CCG	0	1	5	6	5	0	10	1	0	1	0	15	0	0
NHS Norwich CCG	0	0	5	2	6	0	2	2	1	0	0	2	2	0
NHS Nottingham City CCG	9	0	0	1	11	8	0	17	2	1	5	0	3	1
NHS Nottingham North And East CCG	0	0	5	5	2	0	3	0	2	0	0	4	0	0
NHS Nottingham West CCG	0	0	1	3	0	0	4	0	2	0	0	2	0	0
NHS Oldham CCG	1	0	5	1	3	6	2	8	5	0	4	0	0	10
NHS Oxfordshire CCG	1	0	16	18	1	0	16	0	15	1	0	2	8	0
NHS Portsmouth CCG	1	0	2	1	8	0	1	2	1	1	0	2	1	0
NHS Redbridge CCG	4	0	2	2	0	0	0	0	9	0	24	0	0	3
NHS Redditch And Bromsgrove CCG	0	1	8	6	1	1	1	0	3	0	0	1	0	0
NHS Richmond CCG	0	0	6	7	0	0	0	0	14	0	0	2	0	0
NHS Rotherham CCG	0	0	7	2	15	2	1	6	1	1	0	0	0	1
NHS Rushcliffe CCG	0	0	2	4	0	0	3	0	2	1	0	0	0	0
NHS Salford CCG	1	1	2	1	13	3	1	13	7	0	0	1	2	1
NHS Sandwell And West Birmingham CCG	12	1	3	0	6	7	0	14	2	0	25	1	1	19
NHS Scarborough And Ryedale CCG	0	3	1	2	5	0	1	1	0	1	0	1	0	0
NHS Sheffield CCG	3	0	6	10	14	0	11	18	2	1	3	12	1	5
NHS Shropshire CCG	0	2	2	4	0	0	12	0	3	0	0	21	0	0
NHS Slough CCG	1	0	0	0	0	0	0	0	0	0	0	8	0	2
NHS Solihull CCG	0	0	5	9	3	1	6	4	2	0	0	2	0	0
NHS Somerset CCG	0	7	9	12	2	0	18	0	9	1	0	17	0	0
NHS South Cheshire CCG	0	0	4	5	0	0	3	0	0	1	0	3	0	0
NHS South Devon And Torbay CCG	0	6	0	3	6	0	13	2	1	0	0	3	0	0
NHS South East Staffs And Seisdon Peninsular CCG	0	1	5	6	1	0	3	0	9	1	0	5	0	0
NHS South Eastern Hampshire CCG	0	3	3	7	1	0	5	3	1	1	0	1	0	0
NHS South Gloucestershire CCG	0	0	10	5	1	0	6	0	1	1	0	0	0	0
NHS South Kent Coast CCG	0	3	1	1	9	1	4	1	3	0	0	7	0	0
NHS South Lincolnshire CCG	0	0	2	3	0	0	4	0	0	2	0	4	0	0
NHS South Manchester CCG	1	0	3	0	1	1	0	12	1	0	1	0	2	0
NHS South Norfolk CCG	0	0	7	2	0	1	8	0	1	1	0	5	0	0
NHS South Reading CCG	4	0	2	1	0	1	0	0	9	0	1	0	1	0
NHS South Sefton CCG	0	1	0	2	6	1	2	10	0	0	0	9	0	0
NHS South Tees CCG	0	0	4	0	11	4	1	15	2	0	0	5	0	0
NHS South Tyneside CCG	0	0	2	1	17	0	1	2	0	0	0	4	0	0
NHS South Warwickshire CCG	1	0	3	11	0	0	11	0	3	0	0	7	0	0
NHS South West Lincolnshire CCG	0	0	1	6	0	0	4	0	2	1	0	5	0	0
NHS South Worcestershire CCG	0	1	6	4	0	0	12	0	1	1	0	5	1	0
NHS Southampton CCG	0	0	6	3	7	1	0	2	6	0	1	0	4	0
NHS Southend CCG	0	2	2	1	7	2	2	4	2	1	0	11	0	0
NHS Southern Derbyshire CCG	0	0	11	6	6	1	9	1	4	1	0	6	0	4
NHS Southport And Formby CCG	0	3	1	2	1	0	4	0	2	0	0	6	0	0
NHS Southwark CCG	21	0	0	0	0	1	0	0	2	0	0	12	1	4
NHS St Helens CCG	0	0	2	2	17	0	1	6	2	0	0	5	0	0
NHS Stafford And Surrounds CCG	0	0	2	1	0	0	9	0	0	1	0	1	0	0
NHS Stockport CCG	0	0	6	13	14	1	3	2	2	0	0	5	0	0
NHS Stoke On Trent CCG	0	1	2	3	13	5	3	17	2	0	0	5	1	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NHS Sunderland CCG	1	1	2	2	23	0	2	6	5	0	0	8	1	0
NHS Surrey Downs CCG	0	0	5	15	0	0	7	0	2	2	0	2	0	0
NHS Surrey Heath CCG	0	0	1	3	0	0	2	0	2	1	0	0	0	0
NHS Sutton CCG	0	0	6	10	1	1	0	1	7	0	0	0	0	0
NHS Swale CCG	0	0	1	1	4	2	2	4	4	0	0	2	0	0
NHS Swindon CCG	0	0	5	3	3	1	3	0	7	0	0	2	0	0
NHS Tameside And Glossop CCG	0	0	3	3	17	4	0	8	5	0	0	0	0	0
NHS Telford And Wrekin CCG	0	0	6	1	3	5	1	2	3	0	0	0	0	0
NHS Thanet CCG	0	1	3	0	2	0	2	6	1	0	0	5	0	0
NHS Thurrock CCG	0	0	2	6	2	7	1	2	7	0	0	4	0	0
NHS Tower Hamlets CCG	13	0	0	0	0	0	0	0	0	0	11	0	2	5
NHS Trafford CCG	0	0	6	14	0	0	1	2	5	0	4	2	0	0
NHS Vale Of York CCG	0	2	4	8	0	0	6	0	1	4	0	4	1	0
NHS Vale Royal CCG	0	0	0	5	2	0	0	1	3	1	0	0	0	0
NHS Wakefield CCG	1	0	15	4	12	1	0	2	0	1	0	3	0	0
NHS Walsall CCG	1	0	2	3	11	7	3	15	3	0	5	5	0	6
NHS Waltham Forest CCG	11	0	0	0	3	0	0	0	1	0	22	1	0	0
NHS Wandsworth CCG	12	0	7	1	1	1	0	0	5	2	2	0	6	0
NHS Warrington CCG	0	0	7	5	6	0	2	3	3	0	0	0	0	0
NHS Warwickshire North CCG	0	0	5	2	10	0	3	1	2	0	0	4	0	0
NHS West Cheshire CCG	0	1	5	10	6	1	3	1	1	0	0	7	1	0
NHS West Essex CCG	0	0	6	11	1	0	4	0	10	1	0	3	0	0
NHS West Hampshire CCG	0	5	11	10	0	0	14	0	4	3	0	4	0	0
NHS West Kent CCG	0	2	14	27	0	1	8	0	7	1	0	1	0	0
NHS West Lancashire CCG	0	0	1	6	1	0	1	4	0	0	0	9	0	0
NHS West Leicestershire CCG	1	0	12	16	1	0	7	0	6	0	0	6	1	0
NHS West London (K&C & QPP) CCG	27	1	3	2	1	0	0	0	4	0	5	1	7	0
NHS West Norfolk CCG	0	5	1	3	3	0	0	0	1	1	0	6	0	0
NHS West Suffolk CCG	0	0	5	0	0	0	9	0	3	1	0	7	0	0
NHS Wigan Borough CCG	0	0	2	9	27	4	2	7	7	0	0	5	0	0
NHS Wiltshire CCG	0	1	11	17	0	0	12	0	4	2	0	9	0	0
NHS Windsor, Ascot And Maidenhead CCG	0	0	6	5	0	0	2	0	6	0	0	0	0	0
NHS Wirral CCG	0	1	1	6	18	2	6	13	0	0	0	8	0	0
NHS Wokingham CCG	0	0	7	1	0	0	1	0	1	2	0	1	0	0
NHS Wolverhampton CCG	6	0	1	1	7	5	3	11	1	0	4	5	0	1
NHS Wyre Forest CCG	0	0	3	1	2	0	4	0	0	0	0	2	0	0