Analysis of Staffing in Two China Government Agencies

Jenne Hui — Katie Li — Harry Xu

STAT 450

Abstract

This project aims to understand the distribution and trends of employment for two of Chinas civil government branches, Taxation Administration Services (TAS) and Environmental Protection Agencies (EPA). To better visualize the data, plots were generated to show how the composition of staff for different education and work unit levels has changed over time. A model-based clustering analysis forms homogeneous groups of provinces that have similar trends in the proportion of staff at each variable level.

1. Introduction

This report presents a graphical visualization and clustering analysis of employment levels for two of Chinas civil government branches, Taxation Administration Services (TAS) and Environmental Protection Agencies (EPA). Civil services in China are very decentralized and exist on several levels, from provincial to city. Several years of data on staffing numbers for these two government agencies have been provided.

The contents of this report addresses two main objectives. The first objective is to explore meaningful ways to visualize trends and changes within government staffing. The second objective is to determine which provinces have similar variation in staffing both between and within the two agencies.

The report is structured as follows: Section 2 is an overview of the datasets used throughout the case. Section 3 explains both the visualization and the clustering methodology. Section 4 summarizes and interprets the results of the analysis. Lastly, Section 5 provides concluding remarks about the project.

2. Data

Employment data for the two civil service branches (TAS and EPA) were separately hand-collected by research students as part of a study. These panel data include staff numbers for each of the 30 provinces. Tables 1 and 2 show an overview of the datasets. While both datasets are structured as panel data and contain staffing counts for each of the thirty provinces, they differ in the categorizations of staff and the number of years spanned. For the TAS dataset [1], the staff numbers are categorized by age, education, and work unit levels. They are also separated into two bureaus, state tax bureau (STB) and local tax bureau

[1] The client provided both the original TAS dataset and a partially-cleaned version in which he imputed data from 2005 and 2006 to account for temporary workers. In our analysis, we use the cleaned dataset.

Table 1: Summary of TAX dataset

Variable	Levels	Years Spanned
Age	8 levels (<30, 31-35, 36-40, 41-45, 46-50, 51-54, 55-59, >60)	1996, 2000-2007
Education	7 levels (Junior High, High School, Technical College, Other Post-Secondary, Bachelors, Post-Graduate)	1996, 2000-2007
Work Unit	4 levels (Admin, Directly Affiliated (DAE) Admin, Tax, Directly Affiliated (DAE) Non-Admin)	2000-2003
Bureau	2 levels (STB and LTB)	1996, 2000-2007

Table 2: Summary of EPA dataset

Variable	Levels	Years Spanned
Department	7 levels (EPA, Monitoring, Inspection, Research, Education, Information, Other)	1992-1993, 1995-2001, 2004-2011
Government Level	4 levels (Provincial, Prefectural, County, Township)	

(LTB). On the other hand, the EPA dataset has staffing counts categorized by department and the level of government.

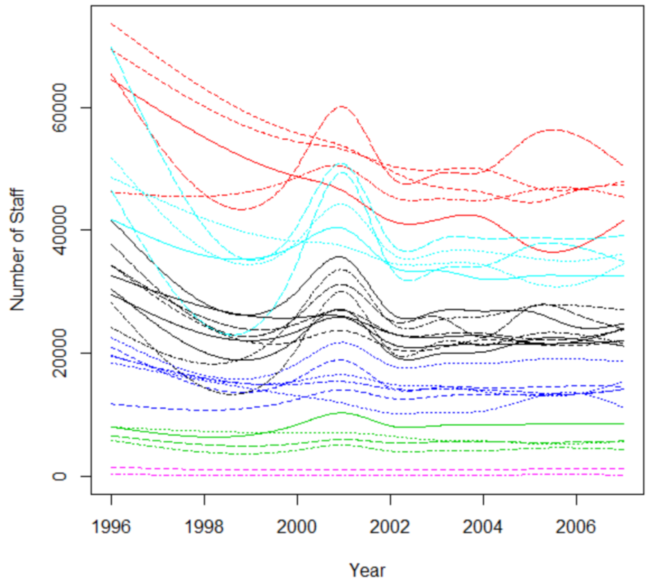
A third dataset with several economic and demographic variables at the provincial level was also provided by the client. The data were extracted from economic journals. These economic indicators were provided so that we could determine if some variables partially relate to tax and EPA staffing. Due to limitations, we were not able to conduct analysis related to this objective.

3. Methodology

Tableau, a visualization software, was extensively used to plot the data. Tableaus ability to group, filter, and select variables was utilized throughout the visualization and analysis process. The user is also able to create calculated fields such as ratios, percentage difference, and proportion of total, which is useful for manipulating data.

The main statistical analysis method of this report is functional clustering, which is motivated by a desire to find groups of provinces that behave similarly over time. First, data need to be converted into matrix format ($m \times n$) where m is the number of years in the dataset, and n is the number of provinces to cluster. Each value in the matrix contains the response variable of interest, such as the total number of staff, the proportion of staff in a given level, or the year-over-year percentage change in staffing. The latter two response variables are preferred because it allows each province to be on a similar scale. If we cluster on the total counts of staff, the algorithms will determine cluster membership by the height of the curve (as evidenced in Figure 1). Because there is nothing informative about these clusters, we convert the staffing counts into percentages or ratios so that we can reduce the amplitudes and extremes in the functions that represent the staffing changes over the years. This helps focus the clustering algorithm to compare similarities and differences across the ”shapes” of the functions as opposed to just the inherent staffing differences between each province.

Figure 1: Total number of staff for each province color-coded by cluster membership (TAS dataset)



Given the data matrix, we then convert the data into functional form using curve-smoothing techniques. Lastly we apply a model-based clustering algorithm to group provinces based on a predetermined number of clusters, k.

4. Results

4.1. Visualization Results

To look at the composition of staff based on age, education and work unit, we plotted the staff count for each level (as a percentage of total) as a stacked line graph to see how the distribution of age, education, and unit varied over time for each bureau.

Figure 2 shows the composition of staff based on education level; the staff counts are summed across all provinces. The percentage of total staff that had a Bachelors degree grew rapidly, from only 5.6% in 1996 to 46% in 2007. On the other hand, the percentage of staff with a Technical college background consistently decreased each year. The education group with the largest percentage of staff is Other Post-Secondary education - it grew from 32% in 1996 to a maximum of 53% in 2003 but then decreased yearly, reaching 40% in 2007.

Figure 3 shows the composition of staff based on work unit. There is one plot for each bureau, LTB and STB. In 2000, the Tax Office unit had the most staff, with 63%, however, this percentage decreased yearly and in 2003, the percentage of staff working in a tax office was only 34%. This decrease was most pronounced in the STB bureau group, as the percentage of STB staff in Tax offices decreased from 62% to 27% - however this decrease in STB was made up for by an increase in DAE Admin and DAE Non-Admin workers.

It was also of interest to see if these trends were the same for every province. To gain insight into provincial variation we replicated these charts for all provinces. Refer to Appendix Section 4 to see the results for Age, Education, and Work Unit. For age groups, it is immediately clear that Tibet's workers have a much higher proportion of workers under the age of 30 years, compared to other provinces. When it comes to Education group, Hainan and Yangzhou greatly vary from the other provinces. Hainan has a much higher proportion of high school workers for most years. While most provinces predominantly consist of "other post-secondary" workers, there are very few of them in Yangzhou - instead there is a much higher proportion of post-graduate workers.

Figure 2: Distribution of Staff by Education Level

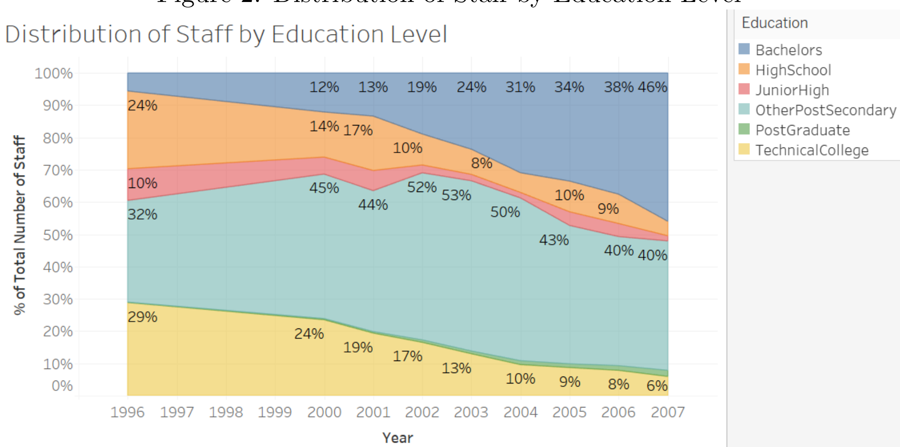
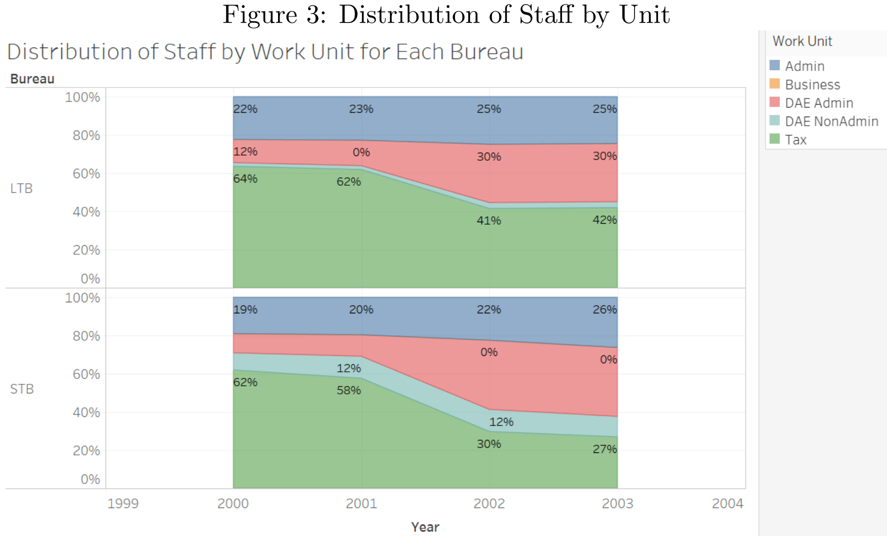


Figure 3: Distribution of Staff by Unit

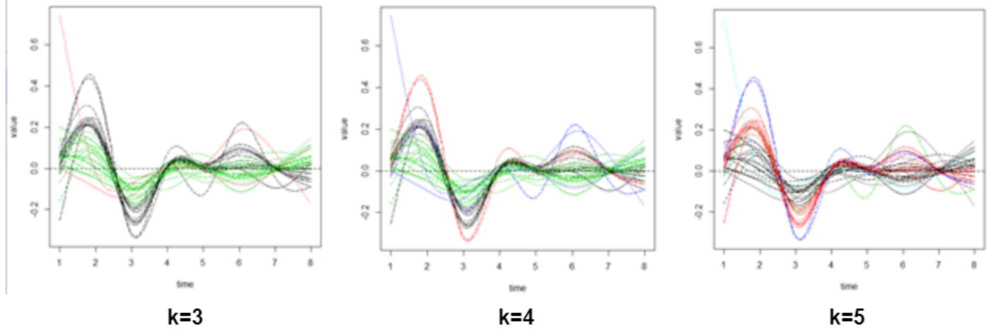


4.2. Clustering Results

As mentioned in the previous section, the objective of the clustering algorithm is to determine groups of provinces for with the trends of staffing counts behave homogeneously. For example, we can cluster the year-over-year percentage change in overall staff counts for each province. Figure 4 shows the results for $k = 3, 4, 5$ clusters. Chongqing was excluded due to missing data and Yangzhou was excluded as an outlier.

In the TAS dataset, staffing was broken down into different age groups and education levels. Due to significant amount of missing data from 1996 to 1999, we reduce our analysis

Figure 4: Year-over-year change in staff for each province color-coded by cluster membership (TAS dataset) for k=3,4,5 clusters



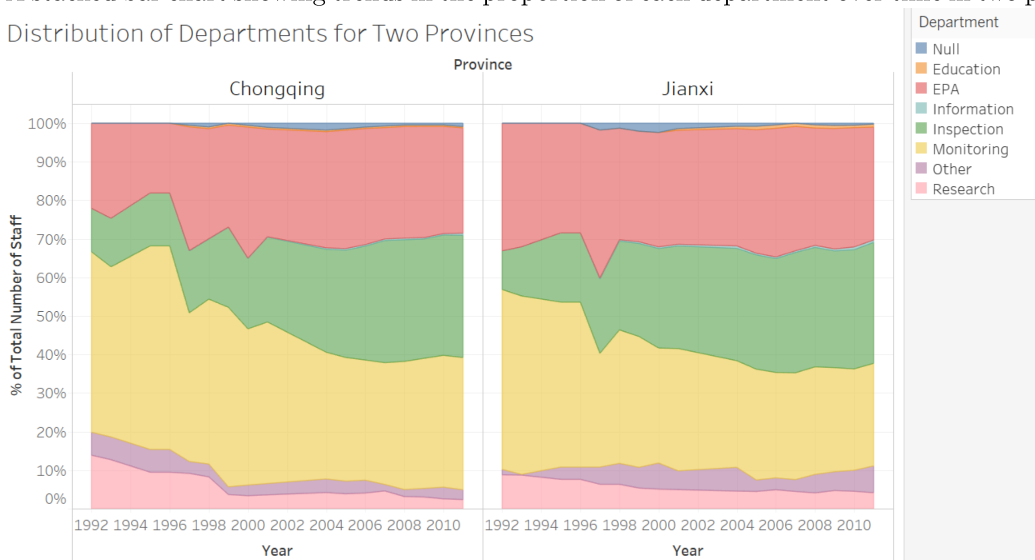
to data between 2000 to 2007. All of the cluster plots can be found in Appendix Section 2. The two main objectives were to determine how provinces are clustered for each of the different age and education levels and if there are any sets of provinces that are recurrently grouped together across most levels. A similarity matrix for each of age and education can be found in Appendix 5. Each (i,j) entry in the matrix corresponds to a pair of provinces, and the value represents the number of levels in which the provinces were clustered together. For example, since there are 8 age groups, the largest possible entry in the age similarity matrix is 8.

On a first glance, we notice that the matrix contains relatively few 0 entries and its predominant entry is 1. This implies that almost of all pairs of provinces are grouped together at least once across the different age levels. The column sums are presented in the bottom row. The largest of these values belongs to Guizhou (100), Guangxi (99), Gansu (99), Hunan (99), and Yunnan (99), meaning those provinces often occur in large, recurrent clusters. On the other hand, Yangzhou has a strikingly low value of 28. This low values implies that Yangzhou does not share many clusters with the other provinces, as its trends across the age groups are unique in comparison with other provinces. In addition, Guizhou and Yangzhou are the only provinces that were paired together in clusterings for all eight age levels. There are also a moderate number of other province pairs that are paired together in clusters for seven out of eight age levels, such as Sichuan and Guizhou, and Hubei and Shandong.

In contrast to the results from the age group analysis, the predominant values in the education similarity matrix are 0 and 1. While there are six education levels, we do not see any distinct provinces that are paired together in clusterings across the six levels. There are only two province pairs that shared 5 clusters, namely Shandong and Guangdong, and Inner Mongolia and Yunnan.

For the EPA dataset, staffing numbers were broken down by department and level of government rather than by age and education. Two provinces, Hunan and Yunnan, were not included in the analysis because data was missing for some years. After clustering the provinces for each of the seven departments and four government levels, we again performed a similarity analysis to see if some provinces appeared in the same clusters for most of the departments. We found that Chongqing and Jiangxi province were clustered together in 6 out of 7 departments, as well as Guangdong and Shandong. In addition, Tibet was grouped by itself for 4 out of 5 government levels; it is plausible that this provinces's agency has different work functions and requirements compared to other provinces, which may explain the variation. For the full cluster results, refer to Section 3 of the Appendix.

Figure 5: A stacked bar chart showing trends in the proportion of each department over time in two provinces.



For staffing at the four levels of government, it was found that Liaoning and Gansu were clustered together for all four levels. Shandong and Shanxi also appeared in the same cluster for all levels, meaning that the trends of staffing for those provinces are very similar. Provinces that were clustered together in three out of four levels are summarized in Table 3. Due to the large amount of missing data for the Township level, only a few provinces were able to be included in the cluster analysis for that level.Tibet was the only province to appear in its own cluster for two out of four levels, Provincial and Prefectural, once again showing that it exhibited unique trends in staffing.

Table 3: Groups of provinces that were clustered together in three out of four government levels for the EPA dataset

Government Levels Clustered Together	Provinces in Cluster		
Provincial Prefectural County	Anhui Gansu Guangdong Guangxi Inner Mongolia Liaoning	Chongqing Fujian Heilongjiang Jiangsu Jiangxi Zhejiang	Henan Jilin Shandong Shanxi
Provincial County Township	Hebei Shandong Shanxi		
Prefectural County Township	Xinjiang Gansu Liaoning		

5. Conclusions and Recommendations

As evidenced by the visualization and cluster analysis, a lot of provinces have similar trends in staffing and government composition. However, there is still a lot of provincial variation. In further analysis, it would be useful to explore what drives the similarity within groups and variation between groups of provinces.

In our cluster analysis, we fixed the number of clusters at 6. It is recommended to try different cluster numbers. In future analysis, one can try to methodologically determine the optimal number of clusters. Finally, a further investigation can be made for provinces that showed a unique staffing trends based on the clustering results. A similar clustering analysis can also be extended to the Work Unit factor and others.

6. Appendix

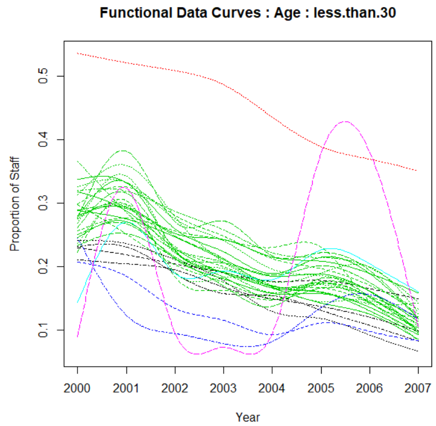
6.1. Functional Clustering in R

To do the clustering, we used the "FunFEM" package in R. Before this, we tried the "Funclustering" package but found that the results were more stable and satisfactory with "FunFEM". The reader should be advised that we found a bug in the "FunFEM" source code. If you would like to reproduce the results of this analysis and are having trouble with the "FunFEM" package, feel free to contact `katieli16@yahoo.com` for the code fix.

6.2. Cluster Results for TAX Dataset

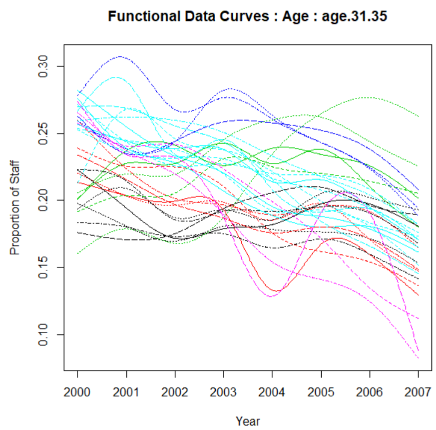
For each level of Age and Education, the results of the clustering as well as the cluster memberships are shown. The response variable that is being clustered is the proportion of staff at the given variable level.

Age 30 and below



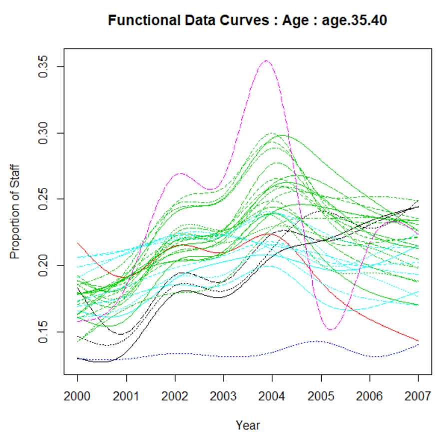

```
The best model is AjB with K = 6 ( bic = 417.7199 )
[1] "Clusters: "
[[1]]
[1] "Beijing"  "Hainan"   "Jilin"    "Liaoning"

[[2]]
[1] "Tibet"

[[3]]
 [1] "Anhui"          "Chongqing"      "Fujian"         "Gansu"          "Guangdong"      "Guangxi"
 [7] "Guizhou"        "Hebei"          "Henan"          "Hubei"          "Hunan"          "Inner.Mongolia"
[13] "Jiangsu"        "Jiangxi"        "Ningxia"        "Qinghai"        "Shaanxi"        "Shandong"
[19] "Shanxi"         "Sichuan"        "Xinjiang"       "Yunnan"         "Zhejiang"

[[4]]
[1] "Shanghai" "Tianjin"

[[5]]
[1] "Heilongjiang"

[[6]]
[1] "Yangzhou"
```

Age 31-35



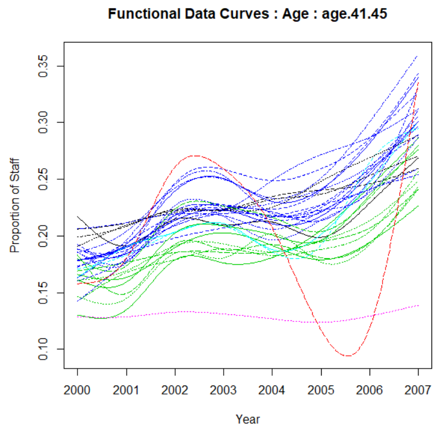

```
The best model is AkBk with K = 6 ( bic = 651.2045 )
[1] "Clusters: "
[[1]]
[1] "Guizhou"  "Hainan"   "Hebei"    "Jiangsu"  "Jilin"    "Liaoning"

[[2]]
[1] "Beijing"      "Chongqing"    "Fujian"       "Heilongjiang" "Shanxi"

[[3]]
[1] "Anhui"     "Guangdong" "Shaanxi"   "Shandong"  "Tibet"

[[4]]
[1] "Hubei"          "Inner.Mongolia" "Xinjiang"

[[5]]
 [1] "Gansu"    "Guangxi"  "Henan"    "Hunan"    "Jiangxi"  "Ningxia"  "Qinghai"  "Sichuan"  "Yunnan"   "Zhejiang"

[[6]]
[1] "Shanghai" "Tianjin"  "Yangzhou"
```

Age 35-40



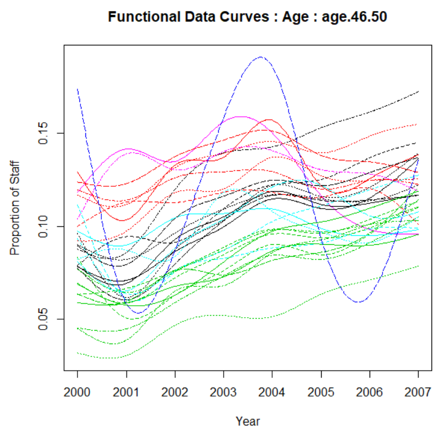

```
The best model is DB with K = 6 ( bic = 634.8192 )
[1] "Clusters: "
[[1]]
[1] "Anhui"          "Hubei"          "Inner.Mongolia"

[[2]]
[1] "Heilongjiang"

[[3]]
 [1] "Chongqing" "Fujian"    "Gansu"     "Guangxi"   "Guizhou"   "Henan"     "Hunan"     "Jiangsu"   "Jiangxi"
[10] "Ningxia"   "Qinghai"   "Shanghai"  "Shanxi"    "Sichuan"   "Tianjin"   "Xinjiang"  "Yunnan"    "Zhejiang"

[[4]]
[1] "Tibet"

[[5]]
[1] "Beijing"   "Guangdong" "Hainan"    "Hebei"     "Jilin"     "Liaoning"  "Shaanxi"   "Shandong"

[[6]]
[1] "Yangzhou"
```

Age 41-45



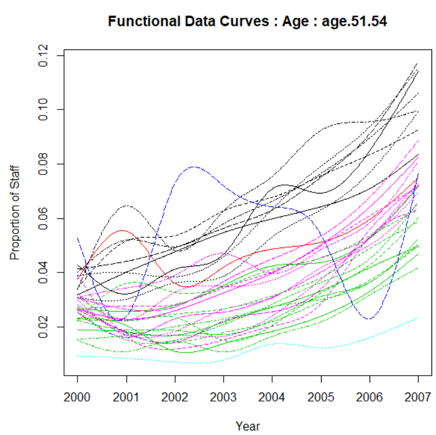

```
The best model is DB with K = 6 ( bic = 684.2996 )
[1] "Clusters: "
[[1]]
[1] "Beijing"      "Hainan"       "Heilongjiang" "Jilin"

[[2]]
[1] "Yangzhou"

[[3]]
[1] "Anhui"          "Guangdong"      "Hebei"          "Hubei"          "Hunan"          "Inner.Mongolia"
[7] "Shandong"       "Shanxi"         "Sichuan"

[[4]]
 [1] "Chongqing" "Fujian"    "Gansu"     "Guangxi"   "Guizhou"   "Jiangxi"   "Liaoning"  "Ningxia"   "Qinghai"
[10] "Shaanxi"   "Shanghai"  "Tianjin"   "Xinjiang"  "Yunnan"    "Zhejiang"

[[5]]
[1] "Henan"   "Jiangsu"

[[6]]
[1] "Tibet"
```

Age 46-50



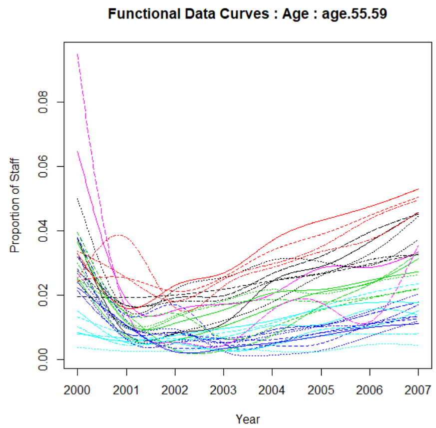

```
The best model is DB with K = 6 ( bic = 707.9864 )
[1] "Clusters: "
[[1]]
[1] "Beijing"   "Gansu"     "Guangdong" "Guizhou"   "Hainan"    "Shandong"  "Shanxi"    "Sichuan"

[[2]]
[1] "Chongqing"    "Hebei"        "Heilongjiang" "Jilin"        "Liaoning"     "Tianjin"

[[3]]
 [1] "Fujian"         "Guangxi"        "Hunan"          "Inner.Mongolia" "Jiangxi"        "Qinghai"
 [7] "Tibet"          "Xinjiang"       "Yunnan"         "Zhejiang"

[[4]]
[1] "Yangzhou"

[[5]]
[1] "Henan"   "Hubei"   "Jiangsu" "Ningxia" "Shaanxi"

[[6]]
[1] "Anhui"    "Shanghai"
```

Age 51-54



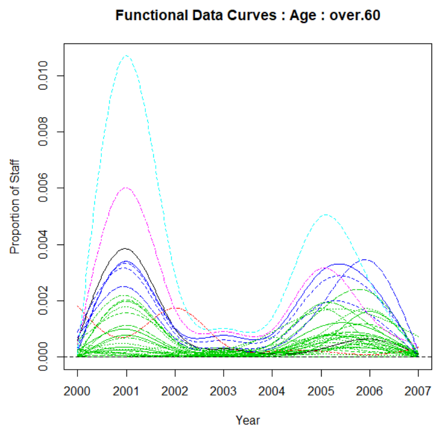

```
The best model is DB with K = 6 ( bic = 877.241 )
[1] "Clusters: "
[[1]]
[1] "Chongqing"    "Hebei"        "Heilongjiang" "Jiangsu"      "Jilin"        "Liaoning"     "Shanghai"
[8] "Tianjin"

[[2]]
[1] "Anhui"

[[3]]
 [1] "Fujian"         "Guangdong"      "Guangxi"        "Hunan"          "Inner.Mongolia" "Jiangxi"
 [7] "Qinghai"        "Shaanxi"        "Xinjiang"       "Yunnan"         "Zhejiang"

[[4]]
[1] "Yangzhou"

[[5]]
[1] "Tibet"

[[6]]
 [1] "Beijing"  "Gansu"    "Guizhou"  "Hainan"   "Henan"    "Hubei"    "Ningxia"  "Shandong" "Shanxi"   "Sichuan"
```

Age 55-59



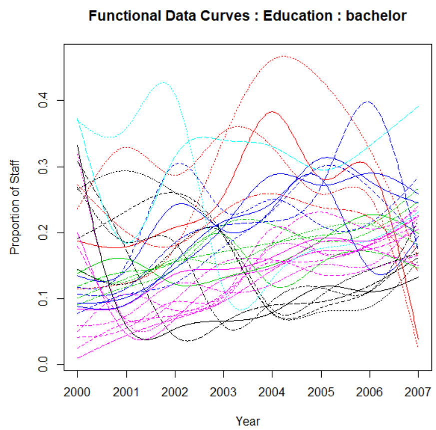

```
The best model is AkBk with K = 6 ( bic = 1066.094 )
[1] "Clusters: "
[[1]]
[1] "Chongqing"    "Hainan"       "Hebei"        "Heilongjiang" "Jilin"        "Zhejiang"

[[2]]
[1] "Jiangsu"  "Liaoning" "Shanghai" "Tianjin"

[[3]]
[1] "Fujian"    "Guangdong" "Guizhou"   "Jiangxi"   "Shanxi"    "Sichuan"

[[4]]
[1] "Gansu"    "Guangxi"  "Henan"    "Hubei"    "Hunan"    "Shaanxi"  "Shandong" "Yunnan"

[[5]]
[1] "Beijing"        "Inner.Mongolia" "Ningxia"        "Qinghai"        "Tibet"          "Xinjiang"

[[6]]
[1] "Anhui"    "Yangzhou"
```

Age 60 and above



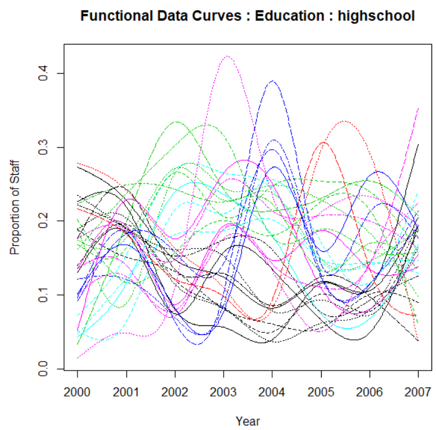

```
[1] "Clusters: "
[[1]]
[1] "Jiangsu"

[[2]]
[1] "Hainan"

[[3]]
 [1] "Anhui"          "Chongqing"      "Fujian"         "Gansu"          "Guangdong"      "Guangxi"
 [7] "Guizhou"        "Hebei"          "Heilongjiang"   "Henan"          "Hubei"          "Hunan"
[13] "Inner.Mongolia" "Jilin"          "Liaoning"       "Shaanxi"        "Shandong"       "Shanxi"
[19] "Sichuan"        "Tibet"          "Xinjiang"       "Yangzhou"       "Yunnan"         "Zhejiang"

[[4]]
[1] "Beijing" "Jiangxi" "Ningxia" "Qinghai"

[[5]]
[1] "Tianjin"

[[6]]
[1] "Shanghai"
```

Bachelors Education



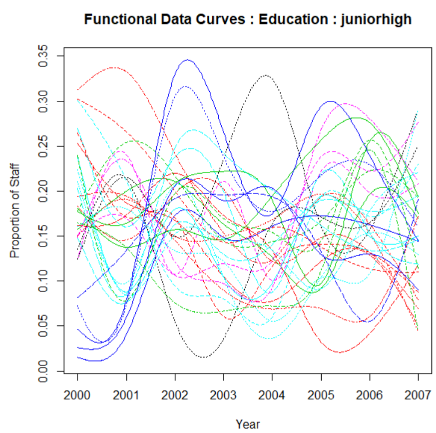

```
The best model is AkBk with K = 6 ( bic = 651.2045 )
[1] "Clusters: "
[[1]]
[1] "Guizhou"  "Hainan"   "Hebei"    "Jiangsu"  "Jilin"    "Liaoning"

[[2]]
[1] "Beijing"      "Chongqing"    "Fujian"       "Heilongjiang" "Shanxi"

[[3]]
[1] "Anhui"     "Guangdong" "Shaanxi"   "Shandong"  "Tibet"

[[4]]
[1] "Hubei"          "Inner.Mongolia" "Xinjiang"

[[5]]
 [1] "Gansu"    "Guangxi"  "Henan"    "Hunan"    "Jiangxi"  "Ningxia"  "Qinghai"  "Sichuan"  "Yunnan"   "Zhejiang"

[[6]]
[1] "Shanghai" "Tianjin"  "Yangzhou"
```

High School Education



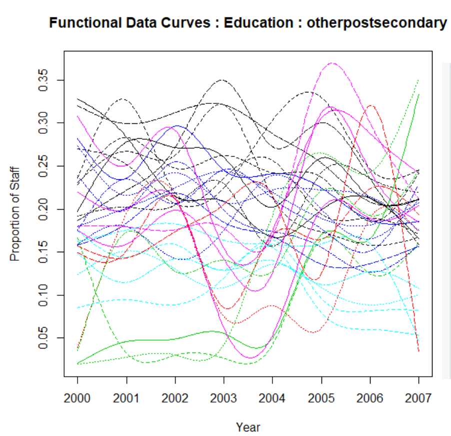

```
The best model is AkBk with K = 6 ( bic = 301.5188 )
[1] "Clusters: "
[[1]]
[1] "Guangdong" "Henan"     "Hubei"     "Liaoning" "Qinghai"   "Shandong"  "Shanxi"    "Sichuan"   "Zhejiang"

[[2]]
[1] "Chongqing" "Yangzhou"

[[3]]
[1] "Beijing"  "Gansu"    "Guizhou"  "Jilin"    "Shanghai" "Tianjin"  "Xinjiang"

[[4]]
[1] "Hunan"          "Inner.Mongolia" "Ningxia"        "Yunnan"

[[5]]
[1] "Guangxi"      "Heilongjiang" "Jiangxi"      "Shaanxi"

[[6]]
[1] "Anhui"   "Fujian"  "Hainan"  "Hebei"   "Jiangsu" "Tibet"
```

Junior High Education



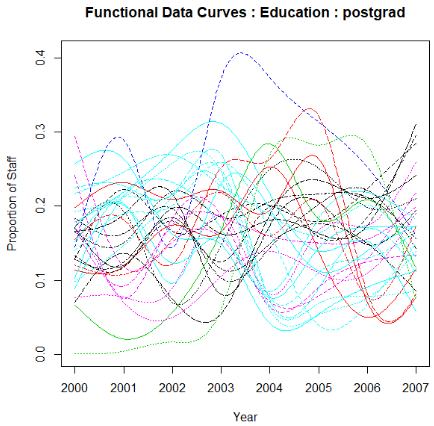

```
The best model is DBk with K = 6 ( bic = 341.0245 )
[1] "Clusters: "
[[1]]
[1] "Chongqing" "Jilin"

[[2]]
[1] "Fujian"  "Gansu"   "Guizhou" "Hainan"  "Hebei"   "Hunan"   "Jiangxi" "Shaanxi"

[[3]]
[1] "Anhui"        "Heilongjiang" "Henan"        "Qinghai"      "Xinjiang"     "Zhejiang"

[[4]]
[1] "Guangdong" "Shandong"  "Sichuan"   "Yangzhou"  "Yunnan"

[[5]]
[1] "Guangxi"        "Hubei"          "Inner.Mongolia" "Jiangsu"        "Liaoning"       "Ningxia"
[7] "Shanxi"         "Tibet"

[[6]]
[1] "Beijing"  "Shanghai" "Tianjin"
```

Other Post-Secondary Education



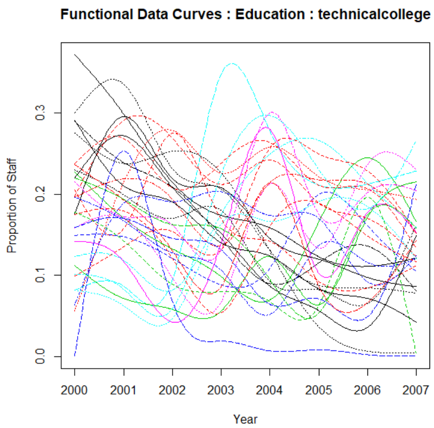

```
The best model is AkjBk with K = 6 ( bic = 336.5954 )
[1] "Clusters: "
[[1]]
[1] "Guangxi"        "Heilongjiang" "Hunan"          "Inner.Mongolia" "Jiangxi"        "Shaanxi"
[7] "Shanxi"         "Yunnan"       "Zhejiang"

[[2]]
[1] "Hebei"    "Hubei"    "Yangzhou"

[[3]]
[1] "Guangdong" "Henan"     "Liaoning"  "Shandong"

[[4]]
[1] "Anhui"    "Fujian"   "Gansu"    "Guizhou"  "Jilin"    "Tibet"    "Xinjiang"

[[5]]
[1] "Beijing"   "Chongqing" "Hainan"    "Shanghai"  "Tianjin"

[[6]]
[1] "Jiangsu" "Ningxia" "Qinghai" "Sichuan"
```

Post-Graduate Education




```
The best model is DBk with K = 6 ( bic = 320.2369 )
[1] "Clusters: "
[[1]]
[1] "Fujian"    "Gansu"     "Hubei"     "Jiangxi"   "Liaoning" "Shanxi"    "Tianjin"   "Xinjiang" "Yangzhou"

[[2]]
[1] "Anhui"          "Hebei"          "Heilongjiang" "Shanghai"

[[3]]
[1] "Qinghai" "Tibet"

[[4]]
[1] "Henan"

[[5]]
 [1] "Chongqing"      "Guangdong"      "Hunan"          "Inner.Mongolia" "Jiangsu"        "Jilin"
 [7] "Ningxia"        "Shaanxi"        "Sichuan"        "Yunnan"         "Zhejiang"

[[6]]
[1] "Beijing"   "Guangxi"   "Guizhou"   "Hainan"    "Shandong"
```

Technical College Education

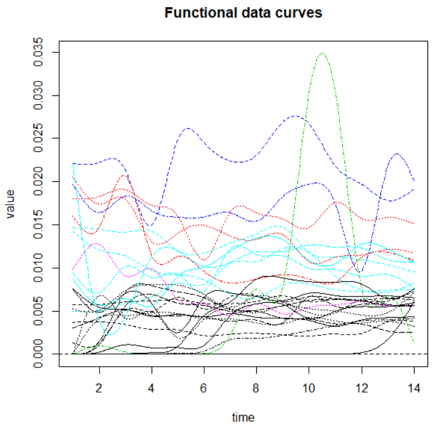



```
The best model is AkjBk with K = 6 ( bic = 311.2238 )
[1] "Clusters: "
[[1]]
[1] "Guangdong" "Hebei"     "Hubei"     "Qinghai"   "Shandong"  "Sichuan"   "Tibet"     "Yunnan"

[[2]]
[1] "Beijing"   "Chongqing" "Guangxi"   "Hainan"    "Henan"     "Ningxia"   "Tianjin"   "Zhejiang"

[[3]]
[1] "Anhui"          "Heilongjiang" "Shaanxi"       "Shanxi"

[[4]]
[1] "Hunan"           "Inner.Mongolia" "Jiangxi"        "Liaoning"       "Yangzhou"

[[5]]
[1] "Fujian"    "Gansu"     "Jilin"     "Xinjiang"

[[6]]
[1] "Guizhou"   "Jiangsu"   "Shanghai"
```

6.3. *Cluster Results for EPA Dataset*

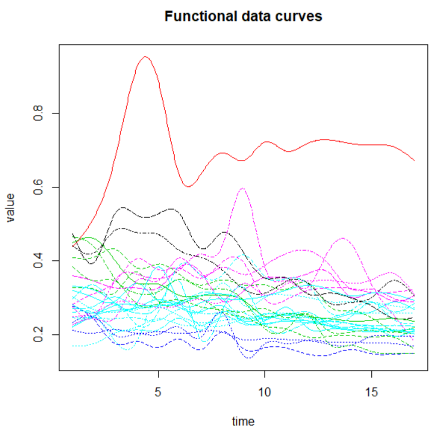
For each level of Department and Gov't Level, the results of the clustering as well as the cluster memberships are shown. The response variable that is being clustered is the proportion of staff at the given variable level.

Education Department



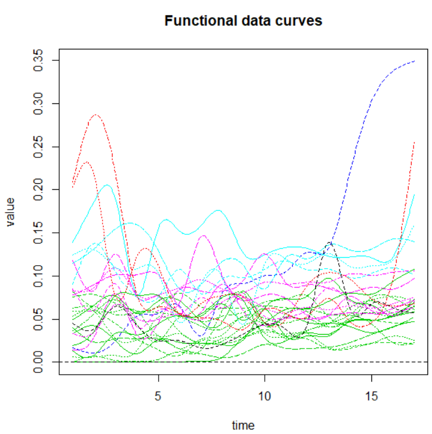

```
The best model is AjB with K = 6 ( bic = 1659.73 )
[[1]]
 [1] "Anhui"      "Chongqing" "Gansu"     "Guandong"  "Guizhou"   "Hebei"     "Henan"     "Jiangsu"   "Jianxi"
[10] "Shaanxi"    "Shandong"  "Shanxi"    "Tibet"     "Xinjiang"  "Zhejiang"

[[2]]
[1] "Fujian"    "Liaoning" "Shanghai"

[[3]]
[1] "Hainan"

[[4]]
[1] "Beijing" "Ningxia"

[[5]]
[1] "Guanxi"          "Heilongjiang"    "Inner.Mongolia" "Jilin"          "Qinghai"        "Tianjin"

[[6]]
[1] "Hubei"
```

EPA Department



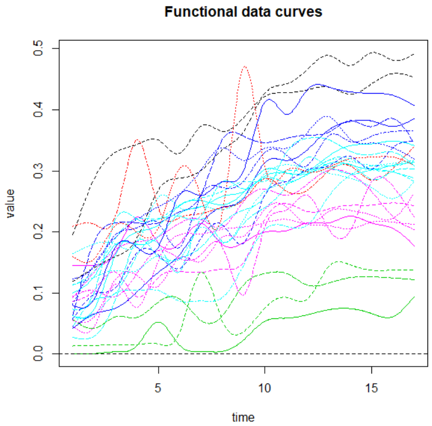

```
The best model is DB with K = 6 ( bic = 458.75 )
[[1]]
[1] "Hainan" "Hebei"

[[2]]
[1] "Tibet"

[[3]]
[1] "Beijing"      "Heilongjiang" "Henan"        "Shandong"     "Shanxi"

[[4]]
[1] "Jilin"     "Liaoning" "Shanghai"

[[5]]
 [1] "Anhui"      "Chongqing" "Fujian"     "Guandong"  "Hubei"      "Jiangsu"   "Jianxi"     "Qinghai"   "Shaanxi"
[10] "Tianjin"    "Xinjiang"  "Zhejiang"

[[6]]
[1] "Gansu"          "Guanxi"         "Guizhou"        "Inner.Mongolia" "Ningxia"
```

Information Department



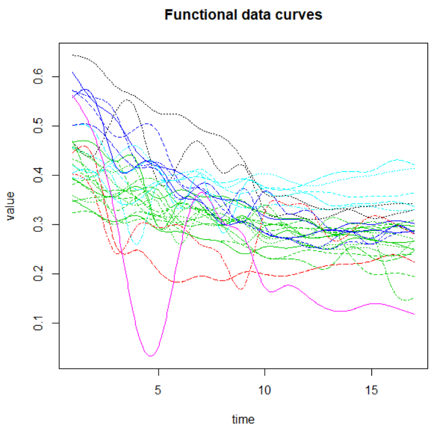

```
The best model is DB with K = 6 ( bic = 1681.372 )
[[1]]
[1] "Guizhou"

[[2]]
 [1] "Beijing"      "Chongqing"    "Gansu"        "Guandong"     "Guanxi"       "Hebei"
 [7] "Heilongjiang" "Henan"        "Hubei"        "Inner.Mongolia" "Jiangsu"    "Jianxi"
[13] "Jilin"        "Ningxia"      "Shaanxi"      "Shandong"     "Tibet"        "Xinjiang"
[19] "Zhejiang"

[[3]]
[1] "Qinghai"

[[4]]
[1] "Anhui"     "Fujian"    "Liaoning" "Shanghai" "Shanxi"

[[5]]
[1] "Hainan"

[[6]]
[1] "Tianjin"
```

Inspection Department



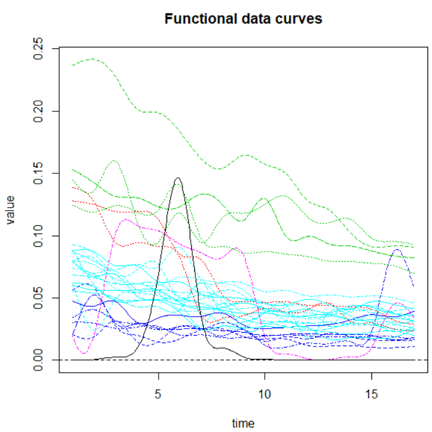

```
The best model is AkB with K = 6 ( bic = 591.4327 )
[[1]]
[1] "Henan" "Jilin"

[[2]]
[1] "Fujian"   "Liaoning"

[[3]]
[1] "Beijing" "Tianjin" "Tibet"

[[4]]
[1] "Anhui"    "Gansu"    "Hubei"    "Shaanxi"  "Shanxi"   "Xinjiang"

[[5]]
[1] "Chongqing"      "Guizhou"        "Hebei"          "Heilongjiang"   "Inner.Mongolia" "Jiangsu"
[7] "Jianxi"         "Qinghai"        "Shandong"

[[6]]
[1] "Guandong" "Guanxi"   "Hainan"   "Ningxia"  "Shanghai" "Zhejiang"
```

Monitoring Department



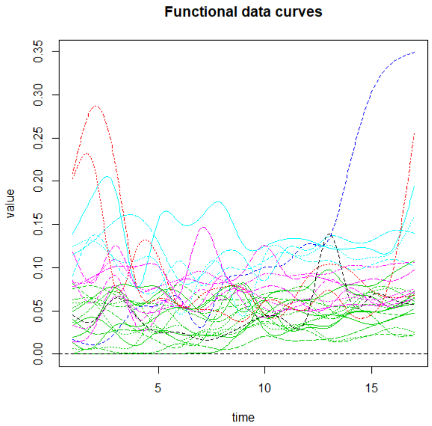

```
The best model is DB with K = 6 ( bic = 601.8183 )
[[1]]
[1] "Chongqing" "Zhejiang"

[[2]]
[1] "Hebei"    "Ningxia"

[[3]]
 [1] "Beijing"      "Gansu"        "Guandong"     "Guizhou"      "Hainan"       "Heilongjiang" "Henan"
 [8] "Hubei"        "Jianxi"       "Jilin"        "Liaoning"     "Shandong"     "Shanxi"

[[4]]
[1] "Anhui"          "Inner.Mongolia" "Qinghai"        "Shaanxi"        "Xinjiang"

[[5]]
[1] "Fujian"   "Guanxi"   "Jiangsu"  "Shanghai" "Tianjin"

[[6]]
[1] "Tibet"
```

Research Department



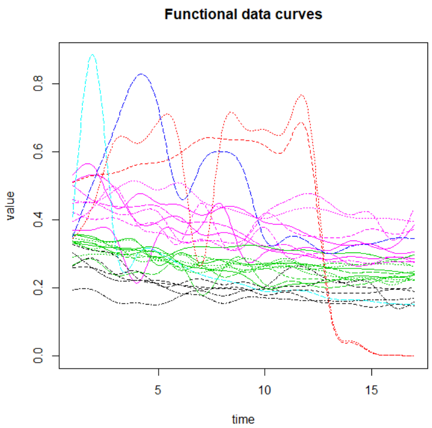

```
The best model is AB with K = 6 ( bic = 1039.063 )
[[1]]
[1] "Tibet"

[[2]]
[1] "Chongqing" "Guizhou"

[[3]]
[1] "Beijing"  "Liaoning" "Shanghai" "Tianjin"

[[4]]
[1] "Guandong" "Hainan"   "Hebei"    "Henan"    "Shandong" "Shanxi"

[[5]]
 [1] "Anhui"          "Fujian"         "Gansu"          "Guanxi"         "Heilongjiang"   "Hubei"
 [7] "Inner.Mongolia" "Jiangsu"        "Jianxi"         "Jilin"          "Qinghai"        "Shaanxi"
[13] "Xinjiang"       "Zhejiang"

[[6]]
[1] "Ningxia"
```

Other Department



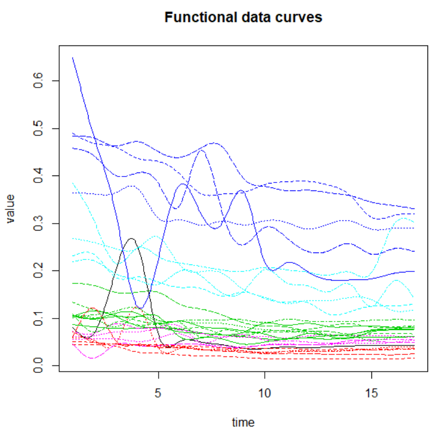

```
The best model is AB with K = 6 ( bic = 815.4727 )
[[1]]
[1] "Xinjiang"

[[2]]
[1] "Ningxia"   "Shanghai"

[[3]]
 [1] "Anhui"          "Chongqing"      "Fujian"         "Gansu"          "Guanxi"         "Guizhou"
 [7] "Hainan"         "Heilongjiang"   "Henan"          "Inner.Mongolia" "Jianxi"         "Shaanxi"
[13] "Tibet"          "Zhejiang"

[[4]]
[1] "Beijing"

[[5]]
[1] "Guandong" "Liaoning" "Shandong" "Tianjin"

[[6]]
[1] "Hebei"    "Hubei"    "Jiangsu" "Jilin"    "Qinghai" "Shanxi"
```

Prefectural Level



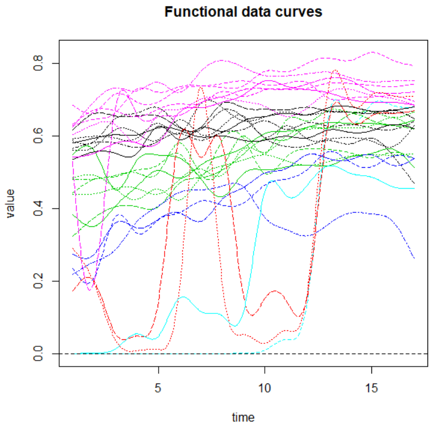

```
The best model is AB with K = 6 ( bic = 329.4629 )
[[1]]
[1] "Hainan"    "Henan"    "Jilin"    "Shandong" "Shanxi"

[[2]]
[1] "Beijing"   "Shanghai"

[[3]]
 [1] "Chongqing"    "Fujian"       "Guizhou"      "Heilongjiang" "Hubei"        "Jiangsu"      "Jianxi"
 [8] "Qinghai"      "Shaanxi"      "Zhejiang"

[[4]]
[1] "Tibet"

[[5]]
[1] "Hebei"

[[6]]
[1] "Anhui"          "Gansu"          "Guandong"       "Guanxi"         "Inner.Mongolia" "Liaoning"
[7] "Ningxia"        "Xinjiang"
```

Provincial Level



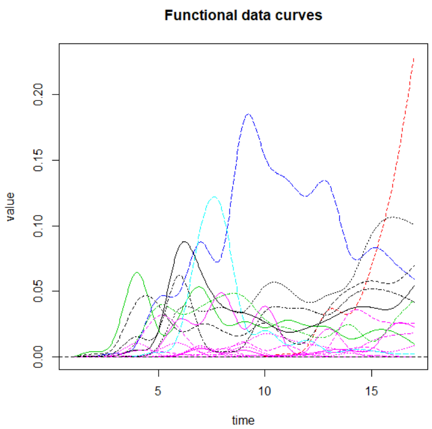

```
The best model is AkjBk with K = 6 ( bic = 951.9942 )
[[1]]
[1] "Guandong"

[[2]]
[1] "Hebei"    "Henan"    "Hubei"    "Shandong" "Shanxi"

[[3]]
[1] "Anhui"        "Fujian"       "Gansu"        "Guanxi"       "Heilongjiang" "Jianxi"       "Jilin"
[8] "Shaanxi"      "Zhejiang"

[[4]]
[1] "Beijing"  "Qinghai"  "Shanghai" "Tianjin"  "Tibet"

[[5]]
[1] "Guizhou"  "Hainan"   "Ningxia"  "Xinjiang"

[[6]]
[1] "Chongqing"      "Inner.Mongolia" "Jiangsu"        "Liaoning"
```

County Level




```
The best model is AjBk with K = 6 ( bic = 395.9709 )
[[1]]
[1] "Chongqing"    "Fujian"       "Hainan"       "Heilongjiang" "Jiangsu"      "Jianxi"       "Zhejiang"

[[2]]
[1] "Shanghai" "Tianjin"

[[3]]
[1] "Anhui"          "Gansu"          "Guandong"       "Guanxi"         "Guizhou"        "Inner.Mongolia"
[7] "Liaoning"

[[4]]
[1] "Ningxia"  "Qinghai"  "Xinjiang"

[[5]]
[1] "Beijing" "Tibet"

[[6]]
[1] "Hebei"    "Henan"    "Hubei"    "Jilin"    "Shaanxi"  "Shandong" "Shanxi"
```

Township Level




```
The best model is DB with K = 6 ( bic = 865.4535 )
[[1]]
[1] "Hebei"    "Jiangsu"  "Shandong" "Shanxi"

[[2]]
[1] "Hainan"

[[3]]
[1] "Anhui" "Henan"

[[4]]
[1] "Guandong"

[[5]]
[1] "Inner.Mongolia"

[[6]]
 [1] "Beijing"   "Chongqing" "Gansu"     "Guizhou"   "Jilin"     "Liaoning"  "Ningxia"   "Qinghai"   "Shaanxi"
[10] "Shanghai"  "Tianjin"   "Tibet"     "Xinjiang"
```

Figure 6: Percentage of staff at each age group for all provinces

Figure 7: Percentage of staff at each education group for all provinces

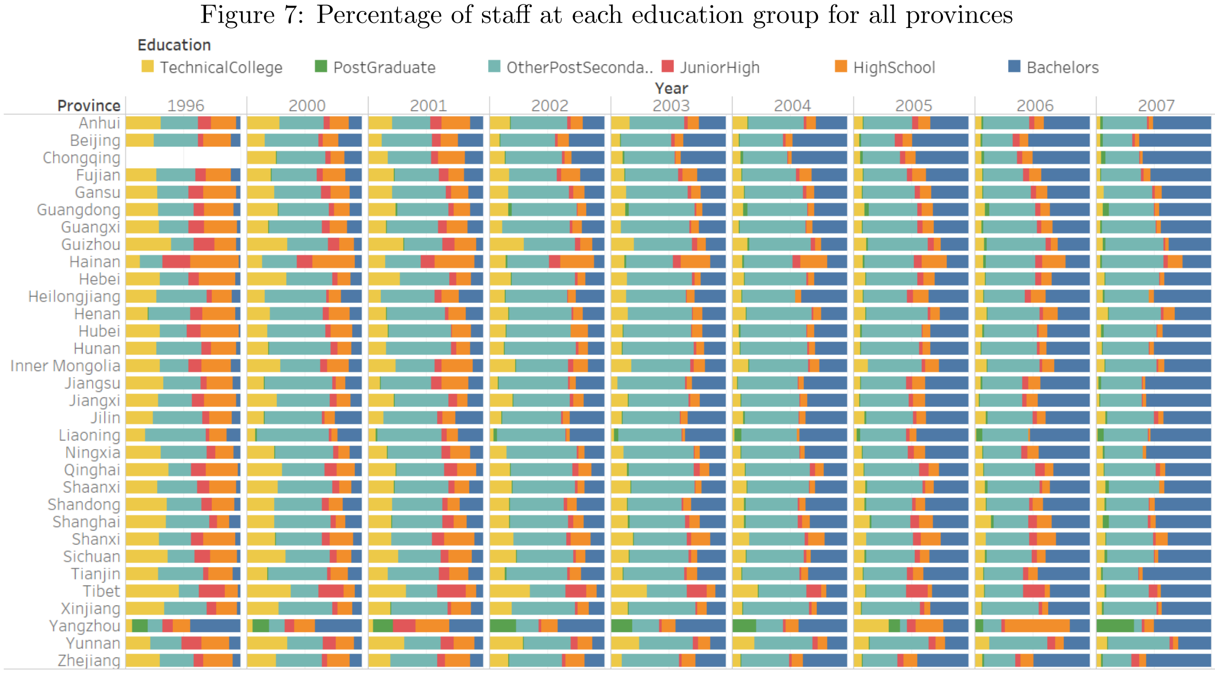
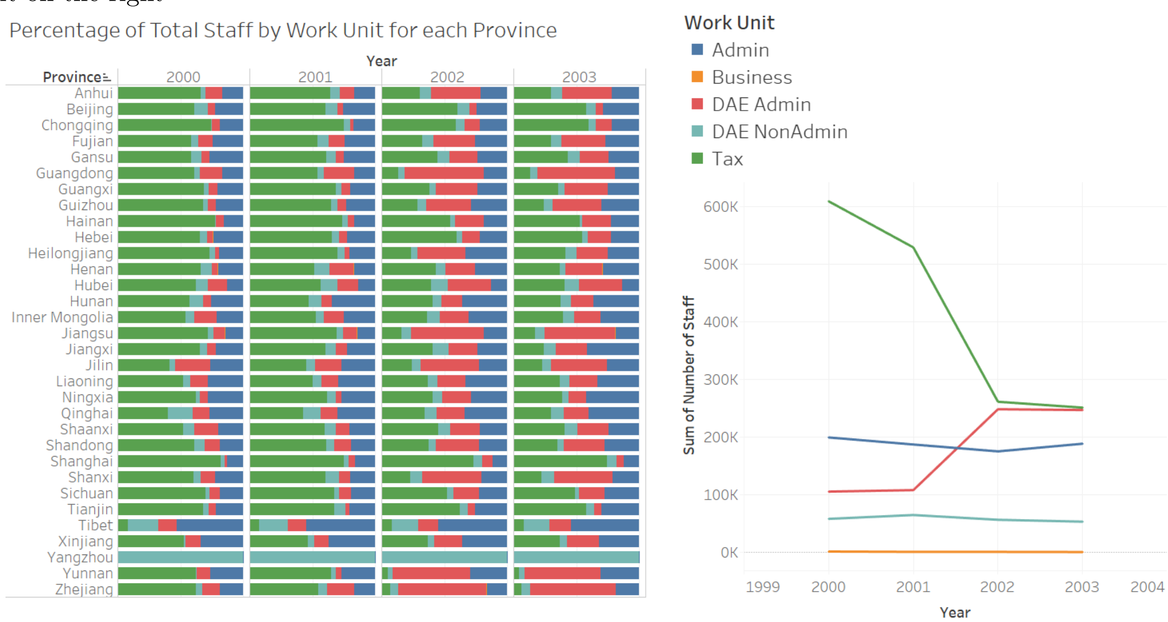


Figure 8: Percentage of staff at each work unit for all provinces on left; overall staffing counts at each work unit on the right

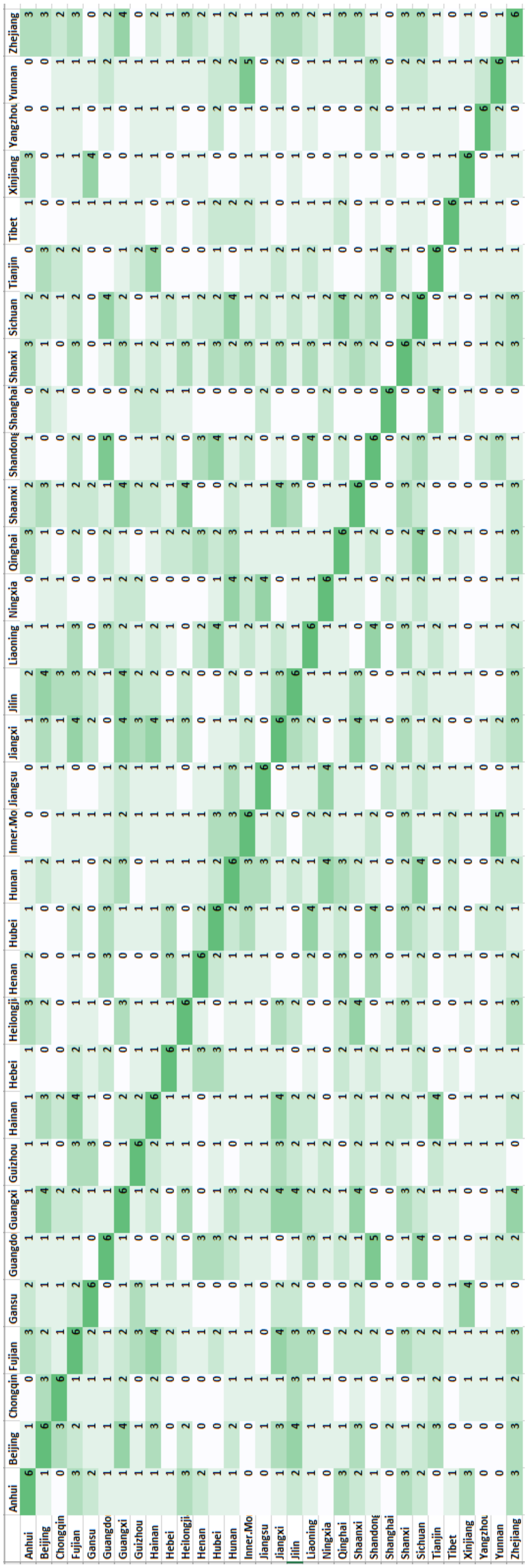


6.5) Similarity Matrices

Similarity matrix for Age Groups - TAX

	Anhui	Beijing	Chongqing	Fujian	Gansu	Guangdon	Guangx	Guizhou	Hainar	Hebei	Heilongjian	Henan	Hubei	Hunan
Anhui	8	1	1	1	3	5	3	3	1	3	1	1	6	5
Beijing	1	8	1	0	3	3	2	2	5	1	2	0	2	1
Chongqing	1	1	8	3	3	1	3	4	3	5	5	2	1	2
Fujian	1	0	3	8	3	2	4	5	1	2	1	1	1	3
Gansu	3	3	3	3	8	4	7	6	2	2	1	3	5	5
Guangdong	5	3	1	2	4	8	3	5	3	4	1	1	6	5
Guangxi	3	2	3	4	7	3	8	5	1	2	1	3	4	6
Guizhou	3	2	4	5	6	5	5	8	3	3	2	2	4	4
Hainan	1	5	3	1	2	3	1	3	8	3	4	0	2	1
Hebei	3	1	5	2	2	4	2	3	3	8	4	1	3	3
Heilongjiang	1	2	5	1	1	1	1	2	4	4	8	1	1	1
Henan	1	0	2	1	3	1	3	2	0	1	1	8	2	3
Hubei	6	2	1	1	5	6	4	4	2	3	1	2	8	6
Hunan	5	1	2	3	5	5	6	4	1	3	1	3	6	8
Inner.Mongolia	6	2	1	2	3	5	4	3	1	3	1	1	6	6
Jiangsu	1	0	3	4	2	1	2	3	1	3	1	3	1	2
Jiangxi	3	2	2	5	4	4	5	5	1	1	0	1	3	5
Jilin	1	2	5	1	1	2	1	2	4	6	5	2	1	1
Liaoning	1	2	6	2	2	2	2	3	3	5	4	1	1	1
Ningxia	1	4	2	2	4	2	3	4	2	0	0	3	2	2
Qinghai	2	4	2	4	5	2	6	4	1	1	0	1	2	4
Shaanxi	4	3	2	2	6	6	5	5	3	3	1	2	6	5
Shandong	5	3	1	1	5	7	4	4	3	4	1	2	7	6
Shanghai	1	2	3	2	3	0	3	2	0	1	0	1	0	1
Shanxi	4	2	3	4	5	6	4	7	3	4	2	2	5	5
Sichuan	5	2	2	3	5	7	4	6	2	3	1	2	6	6
Tianjin	0	2	5	2	3	0	3	2	1	2	2	1	0	1
Tibet	1	1	1	2	1	1	2	1	0	1	1	1	1	2
Xinjiang	1	1	3	4	3	1	4	3	0	1	1	4	1	3
Yangzhou	2	0	1	0	1	1	1	1	0	1	1	1	1	1
Yunnan	3	2	3	4	7	3	8	5	1	2	1	3	4	6
Zhejiang	3	1	4	3	4	3	5	4	2	2	2	3	3	5
	86	66	75	67	99	89	99	100	53	74	47	55	86	99

	Inner.Mongoli	Jiangsu	Jiangxi	Jilin	Liaoning	Ningxia	Qingha	Shaanx	Shandong	Shangha	Shanxi	Sichuar	Tianjir	Tibet	Xinjiang	Yangzhou	Yunnar	Zhejiang
Anhui	6	1	3	1	1	1	2	4	5	1	4	5	0	1	1	2	3	3
Beijing	2	0	2	2	2	4	4	3	3	2	2	2	2	1	1	0	2	1
Chongqing	1	3	2	5	6	2	2	2	1	3	3	2	5	1	3	1	3	4
Fujian	2	4	5	1	2	2	4	2	1	2	4	3	2	2	4	0	4	3
Gansu	3	2	4	1	2	4	5	6	5	3	5	5	3	1	3	1	7	4
Guangdong	5	1	4	2	2	2	2	6	7	0	6	7	0	1	1	1	3	3
Guangxi	4	2	5	1	2	3	6	5	4	3	4	4	3	2	4	1	8	5
Guizhou	3	3	5	2	3	4	4	5	4	2	7	6	2	1	3	1	5	4
Hainan	1	1	1	4	3	2	1	3	3	0	3	2	1	0	0	0	1	2
Hebei	3	3	1	6	5	0	1	3	4	1	4	3	2	1	1	1	2	2
Heilongjiang	1	1	0	5	4	0	0	1	1	0	2	1	2	1	1	1	1	2
Henan	1	3	1	2	1	3	1	2	2	1	2	2	1	1	4	1	3	3
Hubei	6	1	3	1	1	2	2	6	7	0	5	6	0	1	1	1	4	3
Hunan	6	2	5	1	1	2	4	5	6	1	5	6	1	2	3	1	6	5
Inner.Mongolia	8	1	4	1	1	2	4	4	5	0	4	5	0	3	3	1	4	4
Jiangsu	1	8	2	2	3	1	2	1	1	3	3	2	3	0	1	0	2	1
Jiangxi	4	2	8	0	1	4	6	4	3	3	4	5	2	1	3	0	5	5
Jilin	1	2	0	8	5	1	0	2	2	1	2	1	2	1	2	1	1	3
Liaoning	1	3	1	5	8	1	1	3	2	3	2	1	5	1	2	1	2	2
Ningxia	2	1	4	1	1	8	5	3	2	3	3	3	2	1	5	0	3	4
Qinghai	4	2	6	0	1	5	8	3	2	4	3	3	3	2	4	0	6	4
Shaanxi	4	1	4	2	3	3	3	8	7	1	4	5	1	1	2	1	5	4
Shandong	5	1	3	2	2	2	2	7	8	0	5	6	0	1	1	1	4	3
Shanghai	0	3	3	1	3	3	4	1	0	8	1	1	5	0	2	0	3	2
Shanxi	4	3	4	2	2	3	3	4	5	1	8	7	1	1	2	1	4	3
Sichuan	5	2	5	1	1	3	3	5	6	1	7	8	1	1	2	1	4	4
Tianjin	0	3	2	2	5	2	3	1	0	5	1	1	8	0	2	0	3	2
Tibet	3	0	1	1	1	1	2	1	1	0	1	1	0	8	4	1	2	2
Xinjiang	3	1	3	2	2	5	4	2	1	2	2	2	2	4	8	1	4	5
Yangzhou	1	0	0	1	1	0	0	1	1	0	1	1	0	1	1	8	1	1
Yunnan	4	2	5	1	2	3	6	5	4	3	4	4	3	2	4	1	8	5
Zhejiang	4	1	5	3	2	4	4	4	3	2	3	4	2	2	5	1	5	8
	84	53	85	59	66	69	81	95	91	48	96	97	52	40	71	27	99	91



Similarity matrix for education levels - TAX

	Anhui	Beijing	Chongqin	Fujian	Gansu	Guandon	Guanxi	Guizhou	Hainan	Hebei	Heilongjia	Henan	Hubei	Inner.Mo	Jiangsu	Jianxi	Jilin	Liaoning	Ningxia	Qingha	Shaanxi	Shandon	Shanghai	Shanxi	Tianjin	Tibet	Xinjiang	Zhejiang
Anhui	7	0	3	3	3	3	2	2	1	3	2	4	2	3	3	3	1	1	0	4	6	3	2	3	1	2	4	5
Beijing	0	7	1	0	2	2	1	1	1	1	3	2	2	1	1	2	2	2	2	0	1	2	1	2	2	1	1	1
Chongqin	3	1	7	4	4	3	3	4	2	2	3	4	3	2	4	6	1	1	1	1	4	4	2	2	2	3	2	4
Fujian	3	0	4	7	2	1	1	3	2	0	2	2	2	1	3	4	0	3	0	1	2	2	3	2	1	1	0	2
Gansu	3	2	4	2	7	4	4	5	3	3	5	5	4	4	4	5	3	2	2	1	4	5	0	4	0	3	4	4
Guandon	3	2	3	1	4	7	3	2	1	3	3	5	4	3	4	4	3	2	3	2	4	6	3	3	2	2	3	5
Guanxi	2	1	3	1	4	3	7	2	1	2	3	3	3	5	2	2	3	0	3	1	3	2	2	1	2	2	3	4
Guizhou	2	1	4	3	5	2	2	7	3	1	3	3	2	2	2	5	1	2	1	0	2	3	0	3	0	2	2	2
Hainan	1	1	2	2	3	1	1	3	7	1	3	2	2	1	1	3	1	2	0	0	1	2	0	2	0	1	0	1
Hebei	3	1	2	0	3	3	2	1	1	7	2	3	2	2	3	2	2	0	2	2	4	3	0	3	0	2	4	3
Heilongjia	2	3	3	2	5	3	3	3	3	2	7	4	4	4	3	4	3	2	1	2	3	4	0	4	0	2	2	3
Henan	4	2	4	2	5	5	3	3	2	3	4	7	4	3	4	5	4	1	1	2	5	5	1	3	1	3	3	5
Hubei	2	2	3	2	4	4	3	2	2	2	4	4	7	2	5	4	5	2	1	3	3	5	1	3	2	1	2	3
Inner.Mo	3	1	2	1	4	3	5	2	1	2	4	3	2	7	2	2	2	0	3	3	4	2	1	1	0	2	4	5
Jiangsu	3	1	4	3	4	4	2	2	1	3	3	4	5	2	7	4	3	1	1	3	4	5	1	3	1	2	3	4
Jianxi	3	2	6	4	5	4	2	5	3	2	4	5	4	2	4	7	2	2	1	1	4	5	1	3	1	3	2	4
Jilin	1	2	1	0	3	3	3	1	1	2	3	4	5	2	3	2	7	2	1	2	2	3	0	2	1	1	2	2
Liaoning	1	2	1	3	2	2	0	2	2	0	2	1	2	0	1	2	2	7	1	0	0	3	4	3	2	0	0	0
Ningxia	0	2	1	0	2	3	3	1	0	2	1	1	1	3	1	1	1	1	7	0	1	2	2	0	1	1	2	2
Qinghai	4	0	1	1	1	2	1	0	0	2	2	2	3	3	3	1	2	0	0	7	4	2	1	1	1	0	3	3
Shaanxi	6	1	4	2	4	4	3	2	1	4	3	5	3	4	4	4	2	0	1	4	7	4	1	2	1	3	5	6
Shandon	3	2	4	2	5	6	2	3	2	3	4	5	5	2	5	5	3	3	2	2	4	7	2	4	2	2	3	4
Shanghai	2	1	2	3	0	3	2	0	0	0	0	1	1	1	1	1	0	4	2	1	1	2	7	1	4	0	0	2
Shanxi	3	2	2	2	4	3	1	3	2	3	4	3	3	1	3	3	2	3	0	1	2	4	1	7	0	1	2	2
Tianjin	1	2	2	1	0	2	2	0	0	0	0	1	2	0	1	1	1	2	1	1	1	2	4	0	7	0	0	1
Tibet	2	1	3	1	3	2	2	2	1	2	2	3	1	2	2	3	1	0	1	0	3	2	0	1	0	7	2	3
Xinjiang	4	1	2	0	4	3	3	2	0	4	2	3	2	4	3	2	2	0	2	3	5	3	0	2	0	2	7	4
Zhejiang	5	1	4	2	4	5	4	2	1	3	3	5	3	5	4	4	2	0	2	3	6	4	2	2	1	3	4	7

Similarity matrix for EPA departments

	Anhui	Beijing	Chongqin	Fujian	Gansu	Guandon	Guanxi	Guizhou	Hainan	Hebei	Heilongjia	Henan	Hubei	Inner.Mo	Jiangsu	Jianxi	Ningxia	Qinghai	Shaanxi	Xinjiang
Anhui	4	0	1	1	3	3	3	1	0	1	1	2	1	3	1	1	1	1	1	2
Beijing	0	4	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1
Chongqin	1	1	4	3	2	1	1	2	1	1	3	1	2	1	3	3	1	2	3	1
Fujian	1	0	3	3	1	1	1	1	1	1	3	1	2	1	3	3	0	1	2	0
Gansu	3	1	2	1	4	3	3	2	0	1	1	1	1	3	1	1	2	2	2	3
Guandon	3	0	1	1	3	4	3	1	0	1	1	1	1	3	1	1	1	1	1	2
Guanxi	3	0	1	1	3	3	3	1	0	1	1	1	1	3	1	1	1	1	1	2
Guizhou	1	1	2	1	2	1	1	4	1	0	1	0	1	1	1	1	2	3	2	3
Hainan	0	0	1	1	0	0	0	1	4	0	1	1	0	0	1	1	1	0	0	1
Hebei	1	0	1	1	1	1	1	0	0	4	1	2	2	1	2	1	0	0	2	0
Heilongjia	1	0	3	3	1	1	1	1	1	1	3	1	2	1	3	3	0	1	2	0
Henan	2	0	1	1	1	1	1	0	1	2	1	4	2	1	1	1	0	0	2	0
Hubei	1	0	2	2	1	1	1	1	0	2	2	2	3	1	2	2	0	1	3	0
Inner.Mo	3	0	1	1	3	3	3	1	0	1	1	1	1	4	1	1	1	1	1	2
Jiangsu	1	0	3	3	1	1	1	1	1	2	3	1	2	1	4	3	0	1	2	0
Jianxi	1	0	3	3	1	1	1	1	1	1	3	1	2	1	3	3	0	1	2	0
Jilin	1	1	2	1	2	1	1	1	1	2	1	3	2	1	1	1	1	1	3	1
Liaoning	3	1	2	1	4	3	3	2	0	1	1	1	1	3	1	1	2	2	2	3
Ningxia	1	1	1	0	2	1	1	2	1	0	0	0	0	1	0	0	4	1	1	3
Qinghai	1	1	2	1	2	1	1	3	0	0	1	0	1	1	1	1	1	4	2	2
Shaanxi	1	1	3	2	2	1	1	2	0	2	2	2	3	1	2	2	1	2	4	1
Shandon	1	0	1	1	1	1	1	0	1	3	1	3	2	1	2	1	0	0	2	0
Shanghai	0	2	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	2	1	1
Shanxi	1	0	1	1	1	1	1	0	1	3	1	3	2	1	2	1	0	0	2	0
Tianjin	0	2	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1
Tibet	0	2	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1
Xinjiang	2	1	1	0	3	2	2	3	1	0	0	0	0	2	0	0	3	2	1	4
Zhejiang	1	0	3	3	1	1	1	1	1	1	3	1	2	1	3	3	0	1	2	0

Similarity matrix for EPA government level