

多元-11-2020270026

2020270026 王姿文

5/17/2021

1. 数据

- 数据叙述：广东省各地区（21区）农村经济发展指标，其中变量为 农业产值 (x1)、林业产值 (x2)、牧业产值 (x3)、企业人数 (x4)、企业总产值 (x5)、利润总额 (x6)
- 目标：利用古典多维标度法替六个农村经济发展指标做分析评价。由于多维数据如果想做二维或三维图，需要进行降维，而常见的主成分、因子分析虽可以降维，但是其关注的保留信息不同。本次报告希望保留原来的距离信息或相似度信息，因此采用多维标度分析。

下表为数据，以及数据的结构：

```
df <- read_excel("ex10.2.xls")
a <- df$...1
df <- df[,-1]
rownames(df) <- a
colnames(df) <- c('农业产值','林业产值','牧业产值','企业人数','企业总产值','利润总额')

kbl(df) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_size = 7)
```

	农业产值	林业产值	牧业产值	企业人数	企业总产值	利润总额
广州市	97.84	1.28	38.86	141.98	2089.55	121.07
深圳市	11.20	0.66	12.59	156.52	418.16	50.12
珠海市	5.67	0.11	3.60	17.39	360.58	10.58
汕头市	29.87	0.57	17.26	52.45	673.74	24.07
佛山市	52.39	0.29	32.14	90.77	1649.81	62.74
韶关市	47.82	4.47	18.44	27.91	144.51	16.14
河源市	33.57	3.10	12.84	12.62	51.25	4.73
梅州市	57.10	2.74	28.02	44.12	226.65	19.75
惠州市	61.57	4.70	25.20	70.38	568.79	40.39
汕尾市	29.82	1.70	12.09	30.52	189.00	6.78
东莞市	20.97	0.14	20.35	134.63	1380.42	74.01
中山市	16.87	0.21	5.33	91.43	1148.14	52.10
江门市	57.33	1.79	39.21	85.64	1252.07	32.68
阳江市	47.72	3.27	21.39	19.52	191.64	11.08
湛江市	87.20	4.72	34.07	40.60	390.06	20.96
茂名市	112.00	7.85	81.36	76.47	739.34	40.85

肇庆市	76.06	16.45	46.77	52.97	569.93	19.40
清远市	57.35	6.67	28.47	17.95	75.29	6.76
潮州市	27.05	1.63	14.88	35.22	501.63	20.97
揭阳市	71.08	2.09	26.43	50.52	891.76	17.79
云浮市	44.07	4.65	38.97	22.23	188.47	8.70

```
str(df)
```

```
## tibble[,6] [21 × 6] (S3: tbl_df/tbl/data.frame)
## $ 农业产值 : num [1:21] 97.84 11.2 5.67 29.87 52.39 ...
## $ 林业产值 : num [1:21] 1.28 0.66 0.11 0.57 0.29 4.47 3.1 2.74 4.7 1.7 ...
## $ 牧业产值 : num [1:21] 38.9 12.6 3.6 17.3 32.1 ...
## $ 企业人数 : num [1:21] 142 156.5 17.4 52.5 90.8 ...
## $ 企业总产值: num [1:21] 2090 418 361 674 1650 ...
## $ 利润总额 : num [1:21] 121.1 50.1 10.6 24.1 62.7 ...
```

2. 古典多维标度分析

首先将21个省份的6个经济指标转换为距离矩阵，以便后续古典多维标度分析

```
d <- dist(df)
D <- round(as.matrix(d), 0)
kbl(D) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_size = 7)
```

	广州市	深圳市	珠海市	汕头市	佛山市	韶关市	河源市	梅州市	惠州市	汕尾市	东莞市	中山市	江门市	阳江市	湛江市	茂名市	肇庆市	清远市	潮州市	揭阳市	云浮市
广州市	0	1675	1740	1424	449	1952	2047	1869	1525	1909	715	949	845	1906	1705	1355	1526	2022	1596	1206	1909
深圳市	1675	0	156	278	1234	307	397	229	181	266	963	733	839	270	146	353	201	376	151	490	273
珠海市	1740	156	0	317	1294	221	311	148	225	174	1029	792	897	175	95	407	229	291	145	537	180
汕头市	1424	278	317	0	978	530	624	448	113	486	713	477	580	484	290	127	119	600	173	222	487
佛山市	449	1234	1294	978	0	1507	1602	1425	1082	1463	275	504	399	1461	1262	914	1082	1577	1151	761	1464
韶关市	1952	307	221	530	1507	0	96	85	427	49	1242	1007	1109	48	250	604	428	72	358	748	50
河源市	2047	397	311	624	1602	96	0	181	523	139	1337	1101	1204	142	345	700	523	38	451	842	140

梅州市	1869	229	148	448	1425	85	181	0	344	53	1159	924	1026	45	166	520	345	154	277	665	49
惠州市	1525	181	225	113	1082	427	523	344	0	385	816	582	684	382	184	187	39	497	86	325	385
汕尾市	1909	266	174	486	1463	49	139	53	385	0	1198	962	1066	24	211	564	386	119	313	704	32
东莞市	715	963	1029	713	275	1242	1337	1159	816	1198	0	238	149	1196	999	654	819	1313	886	502	1199
中山市	949	733	792	477	504	1007	1101	924	582	962	238	0	118	961	764	427	585	1077	650	268	964
江门市	845	839	897	580	399	1109	1204	1026	684	1066	149	118	0	1063	864	518	684	1179	753	363	1066
阳江市	1906	270	175	484	1461	48	142	45	382	24	1196	961	1063	0	204	558	382	117	311	701	19
湛江市	1705	146	95	290	1262	250	345	166	184	211	999	764	864	204	0	356	181	317	128	502	207
茂名市	1355	353	407	127	914	604	700	520	187	564	654	427	518	558	356	0	180	672	265	171	560
肇庆市	1526	201	229	119	1082	428	523	345	39	386	819	585	684	382	181	180	0	497	93	323	384
清远市	2022	376	291	600	1577	72	38	154	497	119	1313	1077	1179	117	317	672	497	0	428	817	115
潮州市	1596	151	145	173	1151	358	451	277	86	313	886	650	753	311	128	265	93	428	0	393	315
揭阳市	1206	490	537	222	761	748	842	665	325	704	502	268	363	701	502	171	323	817	393	0	705
云浮市	1909	273	180	487	1464	50	140	49	385	32	1199	964	1066	19	207	560	384	115	315	705	0

接着来看多维标度分析结果，这次分析设定降维的维度为两维：

首先能看出第一对典型相关系数很大，高达0.9717098

1. points
- 为矩阵，每行是一个降维坐标，两个维度的座标点相似则代表距离相近，也就是该区间的经济发展相似

```
res <- cmdscale(d, k = 2, eig=TRUE)
p <- as.data.frame(res$points)
colnames(p) <- c('dim1','dim2')
kbl(p) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_size = 7)
```

	dim1	dim2
广州市	1442.53029	17.821289

深圳市	-227.90609	-100.399426
珠海市	-295.12444	-18.027296
汕头市	20.16907	-10.265173
佛山市	997.87159	11.283411
韶关市	-509.15290	2.366295
河源市	-603.74766	-1.207035
梅州市	-425.97126	4.148167
惠州市	-82.31511	-2.347869
汕尾市	-465.31107	-11.021836
东莞市	731.73974	-51.523424
中山市	496.50661	-36.958806
江门市	599.68498	12.996520
阳江市	-462.83165	10.696532
湛江市	-262.74283	34.200087
茂名市	89.13344	53.880161
肇庆市	-82.80124	30.081550
清远市	-578.98821	18.051838
潮州市	-152.64444	-7.404210
揭阳市	237.77499	30.540053
云浮市	-465.87381	13.089172

2. eig
是21个特征值，会发现前几个特征值极高以外，其余特征值都非常低

```
e <- as.data.frame(res$eig)
row.names(e) <- rownames(df)
colnames(e) <- c('eigenvalue')
kbl(e) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_siz
e = 7)
```

eigenvalue	
广州市	6316312.0403
深圳市	21872.8443
珠海市	16135.6782
汕头市	1987.8471
佛山市	795.5720
韶关市	125.5068
河源市	0.0000
梅州市	0.0000

惠州市	0.0000
汕尾市	0.0000
东莞市	0.0000
中山市	0.0000
江门市	0.0000
阳江市	0.0000
湛江市	0.0000
茂名市	0.0000
肇庆市	0.0000
清远市	0.0000
潮州市	0.0000
揭阳市	0.0000
云浮市	0.0000

3. GOF

是绝对值累积贡献率和平方累积贡献率，可以看出降到两维的贡献率为0.99，因此数据信息再降维过程中几乎没有遗失

```
g <- as.data.frame(res$GOF)
row.names(g) <- c('绝对值累积贡献率', '平方累积贡献率')
colnames(g) <- c('GOF')
kbl(g) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, font_size = 7)
```

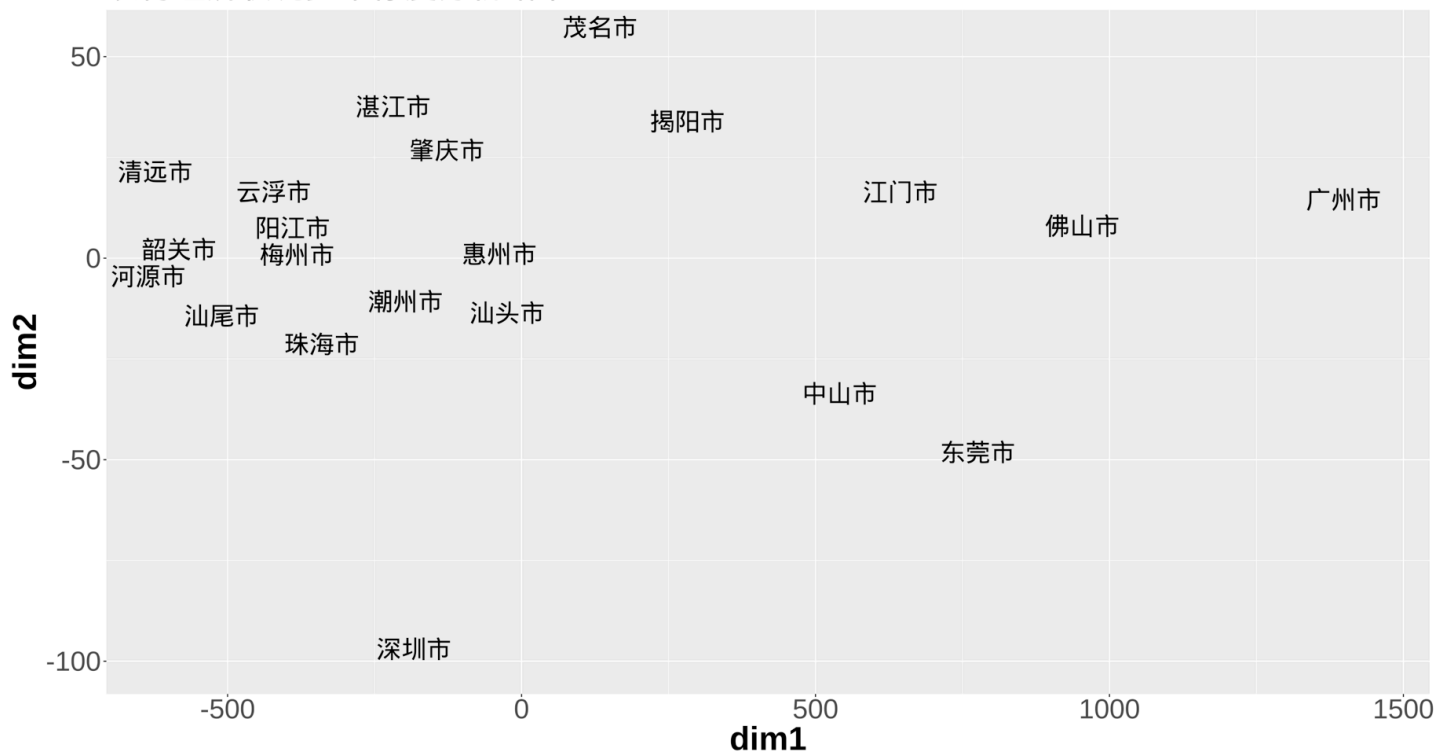
GOF	
绝对值累积贡献率	0.9970043
平方累积贡献率	0.9970043

4. plot

可以此看出广东省21个区的经济发展相似程度，像是几个我们认知里发展较为好的市确实在图内比较特别，例如广州市、深圳市、东莞市...等

```
dp <- as.data.frame(res$points)
colnames(dp) <- c("dim1", "dim2")
dp[["label"]] <- rownames(df)
ggplot(data = dp, mapping = aes(x=dim1, y=dim2, label=label)) +
  geom_text_repel(size=8) +
  labs(title="农村经济状况多维标度分析结果") +
  theme(plot.title = element_text(size=30, face="bold"),
        axis.title = element_text(size=30, face="bold"),
        axis.text = element_text(size=25))
```

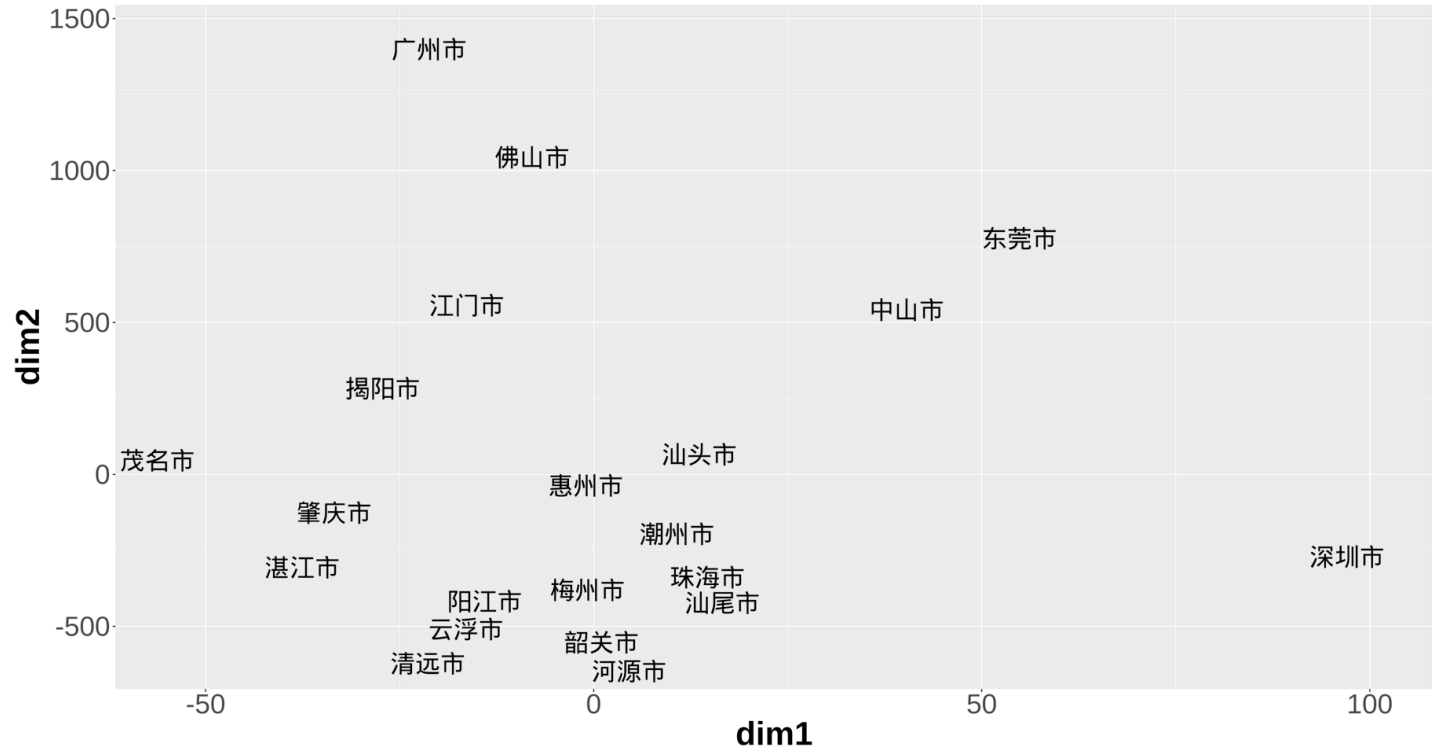
农村经济状况多维标度分析结果



接着来逆时针移动此图，使此图和地图上的地理位置类似来比较，可以看出以广州市、东莞市、深圳市三者在地上的连线来看，地图上的左下半部位置和本图相似，但地图上的右上半部则否，地图上的右上半部则大都集中在本图的正下方，由此可知经济发展相似度或多或少会受地理位置影响，可以大致判别以广州市、东莞市、深圳市三者在地上的连线来看，其在地图上的左下角和右上角经济发展相似，而本三个市则独树一帜

```
rotate.mat <- function(theta){
  rbind(c(cos(theta), sin(theta)), c(-sin(theta), cos(theta)))
}
deg <- 90
theta <- pi/180 * deg
MX <- res$points %*% t(rotate.mat(-theta))
dp <- as.data.frame(MX)
colnames(dp) <- c("dim1", "dim2")
dp[["label"]] <- rownames(df)
ggplot(data = dp, mapping = aes(x=dim1, y=dim2, label=label)) +
  geom_text_repel(size=8) +
  labs(title="农村经济状况多维标度分析结果") +
  theme(plot.title = element_text(size=30, face="bold"),
        axis.title = element_text(size=30, face="bold"),
        axis.text = element_text(size=25))
```

农村经济状况多维标度分析结果



```
knitr::include_graphics('map-guangdong.png')
```



5. 原始 v.s. 降维 比较原始经济指标矩阵和降维到2维的近似距离矩阵，数值确实非常接近
原始经济指标矩阵:

```
as.dist(d)
```

##	广州市	深圳市	珠海市	汕头市	佛山市	韶关市
## 深圳市	1675.40647					

##	珠海市	1739.77237	156.03759				
##	汕头市	1423.73788	277.85012	316.62725			
##	佛山市	448.89638	1234.31074	1293.52874	977.95965		
##	韶关市	1951.95595	306.54406	221.00608	530.17761	1507.40740	
##	河源市	2046.88712	397.36236	310.82914	624.09446	1601.74943	96.42458
##	梅州市	1868.69423	229.30634	148.27045	448.15118	1424.58903	84.87440
##	惠州市	1525.07725	181.42944	225.07869	112.63801	1081.51348	427.36412
##	汕尾市	1908.64333	265.73802	174.02403	485.57259	1463.43460	49.45480
##	东莞市	715.11324	962.88637	1028.76473	713.25530	275.22341	1242.15737
##	中山市	949.34003	732.93689	792.20264	477.14980	503.75280	1006.85104
##	江门市	844.98591	838.79193	896.57155	580.41247	399.00334	1109.42530
##	阳江市	1905.77485	270.22656	175.04337	483.75252	1460.87348	48.24298
##	湛江市	1705.49528	146.05217	95.45854	290.19126	1261.93017	249.54802
##	茂名市	1354.93312	352.97617	406.54453	126.76369	914.15119	604.05365
##	肇庆市	1525.86748	200.81053	228.59691	118.55432	1081.88884	428.21071
##	清远市	2021.74924	375.57914	291.09734	600.45842	1577.22177	71.93647
##	潮州市	1596.40033	150.95715	144.59466	173.04068	1150.69095	357.85974
##	揭阳市	1206.06958	490.26258	536.75448	222.17234	760.70068	748.00202
##	云浮市	1908.91981	272.57876	179.98617	487.16123	1463.99028	49.55475
##		河源市	梅州市	惠州市	汕尾市	东莞市	中山市
##	深圳市						
##	珠海市						
##	汕头市						
##	佛山市						
##	韶关市						
##	河源市						
##	梅州市	181.01722					
##	惠州市	522.87172	343.81267				
##	汕尾市	138.98295	52.62840	384.89977			
##	东莞市	1336.63868	1159.17746	815.90091	1197.91042		
##	中山市	1100.89339	924.42862	581.92759	962.25209	237.78709	
##	江门市	1203.88680	1026.40316	683.65678	1065.51356	149.20584	118.31935
##	阳江市	141.67088	45.14700	381.96377	23.57593	1196.30047	960.71065
##	湛江市	345.20722	166.32262	184.23995	210.97987	998.53117	764.21941
##	茂名市	699.79864	519.83463	186.63241	563.68075	653.86074	427.03214
##	肇庆市	523.45691	344.70076	39.47500	386.43154	818.86726	585.11515
##	清远市	37.85474	154.20496	497.44800	118.90700	1312.60549	1077.34766
##	潮州市	451.29311	277.07687	86.20548	313.01186	886.02289	649.84665
##	揭阳市	842.41127	665.29280	324.52052	704.48635	501.57308	268.27034
##	云浮市	140.47374	48.50071	385.30747	31.73312	1199.39344	964.12132
##		江门市	阳江市	湛江市	茂名市	肇庆市	清远市
##	深圳市						
##	珠海市						
##	汕头市						
##	佛山市						
##	韶关市						
##	河源市						
##	梅州市						
##	惠州市						
##	汕尾市						
##	东莞市						
##	中山市						
##	江门市						
##	阳江市	1062.90277					


```
## 湛江市 863.80230 204.04419
## 茂名市 517.53748 558.43902 355.72457
## 肇庆市 683.50680 381.98490 181.47080 179.65025
## 清远市 1179.06917 117.10198 317.36572 671.81433 496.82542
## 潮州市 753.22530 311.30439 128.34577 265.09923 92.84935 428.24543
## 揭阳市 362.80973 701.24641 502.13194 170.76871 322.84341 817.32452
## 云浮市 1065.84469 18.63708 207.38974 560.22763 384.44064 114.55341
##     潮州市     揭阳市
## 深圳市
## 珠海市
## 汕头市
## 佛山市
## 韶关市
## 河源市
## 梅州市
## 惠州市
## 汕尾市
## 东莞市
## 中山市
## 江门市
## 阳江市
## 湛江市
## 茂名市
## 肇庆市
## 清远市
## 潮州市
## 揭阳市 393.08762
## 云浮市 315.06762 704.55172
```

降维到2维的近似距离矩阵:

```
round(dist(res$points)) #after DR
```

##	广州市	深圳市	珠海市	汕头市	佛山市	韶关市	河源市	梅州市	惠州市	汕尾市
## 深圳市	1675									
## 珠海市	1738	106								
## 汕头市	1423	264	315							
## 佛山市	445	1231	1293	978						
## 韶关市	1952	299	215	529	1507					
## 河源市	2046	389	309	624	1602	95				
## 梅州市	1869	224	133	446	1424	83	178			
## 惠州市	1525	176	213	103	1080	427	521	344		
## 汕尾市	1908	254	170	485	1463	46	139	42	383	
## 东莞市	714	961	1027	713	273	1242	1336	1159	816	1198
## 中山市	948	727	792	477	504	1006	1101	923	580	962
## 江门市	843	835	895	580	398	1109	1204	1026	682	1065
## 阳江市	1905	260	170	483	1461	47	141	37	381	22
## 湛江市	1705	139	61	286	1261	248	343	166	184	208
## 茂名市	1354	353	391	94	910	600	695	517	180	558
## 肇庆市	1525	195	218	111	1081	427	522	344	32	385
## 清远市	2022	371	286	600	1577	72	31	154	497	117
## 潮州市	1595	120	143	173	1151	357	451	274	71	313
## 揭阳市	1205	484	535	221	760	747	842	664	322	704
## 云浮市	1908	264	174	487	1464	45	139	41	384	24
##	东莞市	中山市	江门市	阳江市	湛江市	茂名市	肇庆市	清远市	潮州市	揭阳市
## 深圳市										
## 珠海市										
## 汕头市										
## 佛山市										
## 韶关市										
## 河源市										
## 梅州市										
## 惠州市										
## 汕尾市										
## 东莞市										
## 中山市	236									
## 江门市	147	115								
## 阳江市	1196	961	1063							
## 湛江市	998	763	863	201						
## 茂名市	651	417	512	554	352					
## 肇庆市	819	583	683	381	180	174				
## 清远市	1313	1077	1179	116	317	669	496			
## 潮州市	885	650	753	311	118	249	79	427		
## 揭阳市	501	267	362	701	501	150	321	817	392	
## 云浮市	1199	964	1066	4	204	557	383	113	314	704

3. 結論

整体看来廣東省經濟發展指標的古典多維標度分析，可判斷经济发展相似度或多或少会受地理位置影响，可以大致判别以广州市、东莞市、深圳市三者在地图上的连线来看，其在地图上的左下角和右上角经济发展相似，而本三个市则独树一格(且發展較好)