cluster

April 7, 2021

[8]: #Using the physc admit database, create clusters of three items in the dataset.

#You may select the items of most interest.

#You might consider things such as living situation with gender and age...

#Identify the clusters, and analyze results.

#Explain briefly what you see, and what it tells you about the data.

#A negative result is acceptable. (There are not clusters of this combination)

load("tedsd_2016_puf.Rdata")

bob <- tedsd_2016_puf

[13]: head(bob)

		DISYR	CASEID	STFIPS	AGE	SERVICES	SUB1	SUB2	SUB3	DETO
		<dbl></dbl>	<dbl $>$	<dbl $>$	<dbl $>$	<dbl></dbl>	<dbl $>$	<dbl $>$	<dbl $>$	<dbl></dbl>
A tibble: 6×78	1	2016	20161314323	2	6	7	2	1	1	-9
	2	2016	20161235303	2	5	7	2	1	1	-9
	3	2016	20161216159	2	4	7	4	10	1	-9
	4	2016	20161185262	2	4	7	10	4	2	-9
	5	2016	20161248228	2	5	7	2	1	1	-9
	6	2016	20161388561	2	11	7	2	4	10	6
		1								

[10]: cluster_b <-kmeans(bob,3)
cluster_b</pre>

K-means clustering with 3 clusters of sizes 491367, 475119, 491559

Cluster means:

DTSYR. CASEID AGE SERVICES SUB1 STFIPS SUB2 SUB3 1 2016 20160746037 31.54127 6.754969 5.131321 4.293351 2.828397 1.322673 2 2016 20161242490 20.65146 6.522029 5.530269 5.046611 3.331770 -2.146115 3 2016 20160248277 29.42140 6.653519 5.304134 4.423544 3.369166 1.948303 NOPRIOR PSOURCE RACE ETHNIC **EDUC** DETCRIM ARRESTS **EMPLOY** 1 -6.563567 -0.6962820 3.107327 -1.0213995 4.780067 3.512904 2.580208 2.791417 2 -5.447673 -0.1174379 3.526921 -0.5195982 4.747750 3.482378 2.128189 2.386844 3 -6.273475 -0.1626621 3.482262 -0.5452591 4.881432 3.509514 2.631684 2.750093 **METHUSE PSYPROB** PREG VET LIVARAG PRIMINC **HLTHINS** 1 1.559610 -0.4929208 -5.799295 1.4183329 2.299178 0.2068189 -4.112100

```
2 1.705331 0.1856945 -5.184851 0.8877187 1.713148 -4.1482492 -4.394783
3 1.027710 -0.4261747 -5.466699 1.2582803 2.326048 0.1486821 -3.924178
   PRIMPAY
           DETNLF
                    MARSTAT
                           DAYWAIT
                                    ROUTE1
                                           FREQ1 FRSTUSE1
1 -4.568882 -3.614197 -0.01874566 -5.601363 1.993602 1.113781 3.456496
2 -3.998678 -7.037822 -2.91003307 -1.950899 1.737375 1.533858 3.167562
3 -3.681764 -4.307851 0.67966816 -3.888542 2.021405 1.435942 3.576987
           FREQ2 FRSTUSE2
                         ROUTE3
                                  FREQ3 FRSTUSE3
1 -2.836464 -2.843553 -1.864991 -6.257136 -6.481597 -5.809808
2 -2.976901 -3.134876 -2.120942 -7.169147 -7.228311 -6.865560
3 -2.682500 -2.740674 -1.712936 -6.301150 -6.478811 -5.878407
 FREQ_ATND_SELF_HELP DSMCRIT SERVICES_D REASON
                                          SUB1_D
                                                 SUB2_D
                                                         SUB3_D
        -0.3611313 -7.657386 5.142553 2.174076 2.495312 1.340477 0.170335
1
        -2.9899983 -7.547745 5.559765 2.257329 4.870009 1.781486 -2.002633
2
3
         0.2926831 -8.249313 5.320714 2.263515 3.018008 2.066999 1.014297
  EMPLOY D
           LIVARAG_D DETNLF_D
                             FREQ1_D
                                    FREQ2_D
                                            FREQ3 D
1 1.2142533 1.034806977 -4.999831 -1.5037050 -5.397564 -7.304552
2 0.4865118 -0.007833827 -7.171189 -0.4356319 -5.779346 -7.483157
3 1.1290140 0.986577400 -5.483104 -1.2602394 -5.125960 -7.219058
 FREQ ATND SELF HELP D
                      LOS ARRESTS D
                                    ALCFLG
                                           COKEFLG
                                                    MARFLG
1
          0.8153213 20.55266 -1.4773601 0.4659002 0.2430200 0.3260272
2
          -3.0454328 22.10296 -2.0240908 0.4457262 0.1021239 0.3391719
          0.8600941 22.72626 -0.9597058 0.5159340 0.1961189 0.3728647
            METHFLG OPSYNFLG
                              PCPFLG
                                       HALLFLG
                                              MTHAMFLG
1 0.3510513 0.005899867 0.1584966 0.008378666 0.005273044 0.05927138 0.01428871
2 0.2171919 0.004321023 0.1255643 0.002931897 0.005541770 0.32600885 0.01398176
3 0.3058270 0.004400286 0.1357172 0.005928078 0.006544484 0.14172053 0.01340633
                                         SEDHPFLG
    STIMFLG
             BENZFLG
                       TRNQFLG
                                BARBFLG
1 0.003125973 0.09274127 0.0003948169 0.001927276 0.005327993 0.001076588
2 0.003171837 0.04128860 0.0005893260 0.001424906 0.005375495 0.001652218
3 0.003283431 0.07166993 0.0006265779 0.001424041 0.005602583 0.001670196
     OTCFLG
            OTHERFLG NUMSUBS
                                 IDU DIVISION
                                             REGION ALCDRUG YEAR
1 0.001416457 0.03174816 1.775365 0.095537551 3.021428 1.673450 2.082313 2016
2 0.001610123 0.05474628 1.692422 -0.004840892 7.672621 3.670729 2.060437 2016
3 0.001912283 0.02346005 1.808111 0.177752823 2.810086 1.597100 2.106718 2016
    CBSA
         GENDER
1 23519.45 1.323626
2 21032.93 1.367508
3 19567.57 1.340004
Clustering vector:
```

```
[14185] \ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2
[14437] \ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 2
[14509] \ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1
[14653] \ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1
[14689] \ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1
[14905] \ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 1\ 1\ 2\ 1
[15193] 2 2 2 2 2 1 2 1 1 2 2 1 1 1 1 1 1 2 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 2 2 1 1 1
[15229] 1 1 2 2 1 2 2 1 1 1 2 2 1 2 2 1 1 1 2 2 2 2 1 1 1 2 1 2 1 2 2 1 1 2 1 2 1 2 1 1 1 1 1 1 2 1
[15301] 1 2 2 1 2 2 2 1 2 1 2 1 2 2 1 2 1 1 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 2 2
[15337] 1 2 2 2 2 1 1 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 1 2 1 1 1
[15589] 2 1 2 2 2 2 2 2 1 2 1 1 1 1 2 1 2 2 1 1 1 1 2 2 2 1 2 2 1 2 1 1 1 1 1 2
[15661] 1 1 2 1 1 2 2 2 1 2 1 1 1 2 1 1 1 2 2 1 1 2 2 2 1 1 2 1 2 2 2 2 2 2 2
[15733] 2 1 1 2 1 2 2 1 2 1 1 1 2 1 2 2 1 1 1 2 2 1 1 1 2 2 2 1 1 1 2 1 2 2 1 2 1 2 2 1 2 1
```

 $\begin{smallmatrix} 22285 \end{smallmatrix}] \ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2$

```
[ reached getOption("max.print") -- omitted 1358046 entries ]
Within cluster sum of squares by cluster:
[1] 1.025647e+16 1.035439e+16 1.026533e+16
(between_SS / total_SS = 88.6 %)
Available components:
[1] "cluster"
 "centers"
    "withinss"
      "tot.withinss"
   "totss"
[6] "betweenss"
    "ifault"
 "size"
   "iter"
```

[11]: head(cluster_b)

```
$cluster 1. 2 2. 2 3. 2 4. 2 5. 2 6. 2 7. 2 8. 2 9. 2 10. 2 11. 2 12. 2 13. 2 14. 2 15. 2 16. 2 17. 2 18. 2
      19.\ 2\ 20.\ 2\ 21.\ 2\ 22.\ 2\ 23.\ 2\ 24.\ 2\ 25.\ 2\ 26.\ 2\ 27.\ 2\ 28.\ 2\ 29.\ 2\ 30.\ 2\ 31.\ 2\ 32.\ 2\ 33.\ 2\ 34.\ 2\ 35.\ 2
      36.\ 2\ 37.\ 2\ 38.\ 2\ 39.\ 2\ 40.\ 2\ 41.\ 2\ 42.\ 2\ 43.\ 2\ 44.\ 2\ 45.\ 2\ 46.\ 2\ 47.\ 2\ 48.\ 2\ 49.\ 2\ 50.\ 2\ 51.\ 2\ 52.\ 2
      53.\ 2\ 54.\ 2\ 55.\ 2\ 56.\ 2\ 57.\ 2\ 58.\ 2\ 59.\ 2\ 60.\ 2\ 61.\ 2\ 62.\ 2\ 63.\ 2\ 64.\ 2\ 65.\ 2\ 66.\ 2\ 67.\ 2\ 68.\ 2\ 69.\ 2
      87. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 2100. 2101. 2102. 2
      103.\ 2\ 104.\ 2\ 105.\ 2\ 106.\ 2\ 107.\ 2\ 108.\ 2\ 109.\ 2\ 110.\ 2\ 111.\ 2\ 112.\ 2\ 113.\ 2\ 114.\ 2\ 115.\ 2\ 116.\ 2
      117.\ 2\ 118.\ 2\ 119.\ 2\ 120.\ 2\ 121.\ 2\ 122.\ 2\ 123.\ 2\ 124.\ 2\ 125.\ 2\ 126.\ 2\ 127.\ 2\ 128.\ 2\ 129.\ 2\ 130.\ 2
      131.\ 2\ 132.\ 2\ 133.\ 2\ 134.\ 2\ 135.\ 2\ 136.\ 2\ 137.\ 2\ 138.\ 2\ 139.\ 2\ 140.\ 2\ 141.\ 2\ 142.\ 2\ 143.\ 2\ 144.\ 2
      145.\ 2\ 146.\ 2\ 147.\ 2\ 148.\ 2\ 149.\ 2\ 150.\ 2\ 151.\ 2\ 152.\ 2\ 153.\ 2\ 154.\ 2\ 155.\ 2\ 156.\ 2\ 157.\ 2\ 158.\ 2
      159.\ 2\ 160.\ 2\ 161.\ 2\ 162.\ 2\ 163.\ 2\ 164.\ 2\ 165.\ 2\ 166.\ 2\ 167.\ 2\ 168.\ 2\ 169.\ 2\ 170.\ 2\ 171.\ 2\ 172.\ 2
      173. 2 174. 2 175. 2 176. 2 177. 2 178. 2 179. 2 180. 2 181. 2 182. 2 183. 2 184. 2 185. 2 186. 2
      187.\ 2\ 188.\ 2\ 189.\ 2\ 190.\ 2\ 191.\ 2\ 192.\ 2\ 193.\ 2\ 194.\ 2\ 195.\ 2\ 196.\ 2\ 197.\ 2\ 198.\ 2\ 199.\ 2\ 200.\ 2
      201. 202. 2 203. 2 204. 2 205. 2 206. 2 207. 2 208. 2 209. 2 210. 2 211. 2 212. 2 213. 2 214. 2
      215.\ 2\ 216.\ 2\ 217.\ 2\ 218.\ 2\ 219.\ 2\ 220.\ 2\ 221.\ 2\ 222.\ 2\ 223.\ 2\ 224.\ 2\ 225.\ 2\ 226.\ 2\ 227.\ 2\ 228.\ 2
      229. 2 230. 2 231. 2 232. 2 233. 2 234. 2 235. 2 236. 2 237. 2 238. 2 239. 2 240. 2 241. 2 242. 2
      243.\ 2\ 244.\ 2\ 245.\ 2\ 246.\ 2\ 247.\ 2\ 248.\ 2\ 249.\ 2\ 250.\ 2\ 251.\ 2\ 252.\ 2\ 253.\ 2\ 254.\ 2\ 255.\ 2\ 256.\ 2
      257. 2 258. 2 259. 2 260. 2 261. 2 262. 2 263. 2 264. 2 265. 2 266. 2 267. 2 268. 2 269. 2 270. 2
      271. 2 272. 2 273. 2 274. 2 275. 2 276. 2 277. 2 278. 2 279. 2 280. 2 281. 2 282. 2 283. 2 284. 2
      285. 2 286. 2 287. 2 288. 2 289. 2 290. 2 291. 2 292. 2 293. 2 294. 2 295. 2 296. 2 297. 2 298. 2
      299. 2 300. 2 301. 2 302. 2 303. 2 304. 2 305. 2 306. 2 307. 2 308. 2 309. 2 310. 2 311. 2 312. 2
      313.\ 2\ 314.\ 2\ 315.\ 2\ 316.\ 2\ 317.\ 2\ 318.\ 2\ 319.\ 2\ 320.\ 2\ 321.\ 2\ 322.\ 2\ 323.\ 2\ 324.\ 2\ 325.\ 2\ 326.\ 2
      327. 2 328. 2 329. 2 330. 2 331. 2 332. 2 333. 2 334. 2 335. 2 336. 2 337. 2 338. 2 339. 2 340. 2
      341.\ 2\ 342.\ 2\ 343.\ 2\ 344.\ 2\ 345.\ 2\ 346.\ 2\ 347.\ 2\ 348.\ 2\ 349.\ 2\ 350.\ 2\ 351.\ 2\ 352.\ 2\ 353.\ 2\ 354.\ 2
      355. 2 356. 2 357. 2 358. 2 359. 2 360. 2 361. 2 362. 2 363. 2 364. 2 365. 2 366. 2 367. 2 368. 2
      369.\ 2\ 370.\ 2\ 371.\ 2\ 372.\ 2\ 373.\ 2\ 374.\ 2\ 375.\ 2\ 376.\ 2\ 377.\ 2\ 378.\ 2\ 379.\ 2\ 380.\ 2\ 381.\ 2\ 382.\ 2
      383. 2 384. 2 385. 2 386. 2 387. 2 388. 2 389. 2 390. 2 391. 2 392. 2 393. 2 394. 2 395. 2 396. 2
      397. 2 398. 2 399. 2 400. 2 401. 2
```

		DISYR	CASEID	STFIPS	AGE	SERVICES	SUB1
\$centers A matrix: 3×78 of type dbl	1	2016	20160746037	31.54127	6.754969	5.131321	4.293351
	2	2016	20161242490	20.65146	6.522029	5.530269	5.046611
	3	2016	20160248277	29.42140	6.653519	5.304134	4.423544

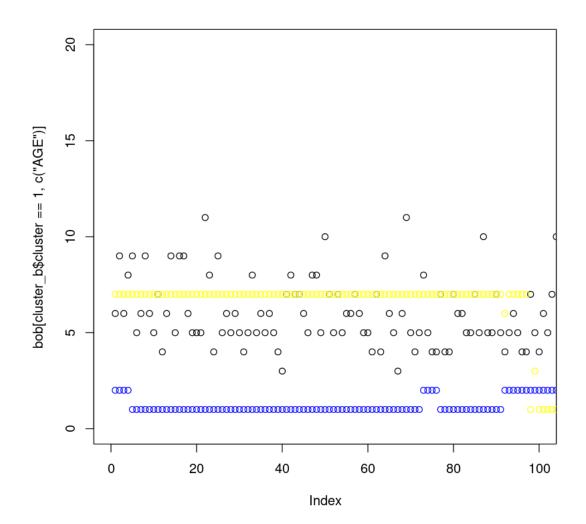
\$totss 269718615227234080

\$withinss 1. 10256467058316574 2. 10354394708591106 3. 10265331875132114

\$tot.withinss 30876193642039792

\$betweenss 238842421585194304

```
[67]: plot(bob[cluster_b$cluster==1,c('AGE')],xlim=c(0,100),ylim=c(0,20))
    points(bob[cluster_b$cluster==2,c('GENDER')],col="blue")
    points(bob[cluster_b$cluster==3,c('SERVICES')],col="yellow")
```



[]: #This data set is anazying the age of the workers in correlation with their ⇒gender and what services they provide.

#According to the chart, most of the indivduals are on the younger age range. □ ⇒Assuming that 1 = male. A majority of the

#individuals are male within this clustered data. As for services, rendered. a□ ⇒majority of them are linked to a single

#service while there are a few outliers.