Project\_healthcare.R

sumas

2021-11-24

setwd("C:/Users/sumas/Documents/Data science/R progamming/Project/Projects for Submission/Healthcare/Healthcare")  
library(dplyr) #### loading dplyr package ###

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Hospital\_data <- read.csv("HospitalCosts.csv")  
print(Hospital\_data)

## AGE FEMALE LOS RACE TOTCHG APRDRG  
## 1 17 1 2 1 2660 560  
## 2 17 0 2 1 1689 753  
## 3 17 1 7 1 20060 930  
## 4 17 1 1 1 736 758  
## 5 17 1 1 1 1194 754  
## 6 17 0 0 1 3305 347  
## 7 17 1 4 1 2205 754  
## 8 16 1 2 1 1167 754  
## 9 16 1 1 1 532 753  
## 10 17 1 2 1 1363 758  
## 11 17 1 2 1 1245 758  
## 12 15 0 2 1 1656 753  
## 13 15 1 2 1 1379 751  
## 14 15 1 4 1 2346 758  
## 15 15 1 7 1 4006 753  
## 16 15 1 4 1 2181 758  
## 17 14 1 1 1 628 754  
## 18 14 1 4 1 2463 758  
## 19 15 1 3 1 1956 753  
## 20 14 1 3 1 1802 758  
## 21 13 1 1 1 3188 812  
## 22 17 1 2 1 2129 566  
## 23 12 0 1 1 7421 249  
## 24 15 1 1 1 1122 422  
## 25 13 1 2 4 1173 754  
## 26 12 0 2 1 3625 812  
## 27 11 1 2 1 3908 50  
## 28 15 0 1 1 3994 139  
## 29 11 0 0 1 1033 753  
## 30 10 0 2 1 2860 141  
## 31 11 0 2 1 3814 420  
## 32 7 0 0 1 1132 139  
## 33 16 1 2 6 1163 751  
## 34 17 1 1 1 610 751  
## 35 6 0 3 1 9530 97  
## 36 15 1 1 1 1268 811  
## 37 17 1 4 1 2582 753  
## 38 16 1 2 1 1287 755  
## 39 17 1 3 1 6594 930  
## 40 13 1 0 1 909 755  
## 41 7 0 0 1 2530 347  
## 42 11 1 2 2 1534 753  
## 43 3 0 5 1 14243 720  
## 44 16 1 3 1 1699 754  
## 45 2 0 2 1 7298 53  
## 46 16 1 1 1 636 754  
## 47 15 1 1 1 626 754  
## 48 1 0 2 1 3782 53  
## 49 14 1 2 1 1444 753  
## 50 14 1 2 1 1183 754  
## 51 14 1 5 1 3045 754  
## 52 14 1 5 1 3624 754  
## 53 14 1 12 1 6810 760  
## 54 1 0 1 1 1409 249  
## 55 13 0 2 1 1211 754  
## 56 1 0 4 1 9606 53  
## 57 1 1 1 1 1411 249  
## 58 15 1 0 1 607 754  
## 59 1 0 1 1 2932 249  
## 60 1 0 3 1 5075 139  
## 61 14 1 1 1 762 753  
## 62 16 1 6 1 6329 753  
## 63 17 1 1 1 1226 753  
## 64 3 1 4 1 8223 710  
## 65 17 0 2 1 1193 776  
## 66 13 1 2 1 1076 754  
## 67 12 1 6 1 17434 115  
## 68 12 1 2 1 1647 753  
## 69 14 1 7 1 3865 754  
## 70 13 1 1 1 628 754  
## 71 15 1 1 1 806 755  
## 72 0 1 41 1 29188 602  
## 73 0 0 2 1 4717 138  
## 74 0 0 12 1 15129 137  
## 75 0 1 2 1 1085 640  
## 76 0 0 3 1 1607 640  
## 77 0 1 3 1 1499 640  
## 78 0 1 3 1 7648 53  
## 79 0 1 2 1 1527 640  
## 80 0 0 2 1 1483 640  
## 81 0 1 4 1 2844 640  
## 82 0 1 3 1 3124 640  
## 83 0 0 3 1 1760 640  
## 84 0 1 2 1 1278 640  
## 85 0 1 2 1 1620 640  
## 86 0 1 2 1 1220 640  
## 87 0 1 2 1 1134 640  
## 88 16 1 0 1 1235 754  
## 89 0 0 3 1 1656 640  
## 90 0 0 4 5 4072 639  
## 91 0 0 2 5 1393 143  
## 92 0 0 0 5 615 254  
## 93 16 1 1 1 779 755  
## 94 0 0 2 1 1385 640  
## 95 0 0 2 1 1224 640  
## 96 0 1 3 1 1779 640  
## 97 0 0 2 1 1526 640  
## 98 15 1 1 1 882 754  
## 99 0 0 1 1 2075 581  
## 100 0 0 17 1 12042 633  
## 101 0 0 2 1 1309 640  
## 102 0 0 2 1 1290 640  
## 103 0 0 2 1 1280 640  
## 104 0 0 3 1 1719 640  
## 105 0 1 2 1 1102 640  
## 106 0 1 3 1 1543 640  
## 107 0 1 2 1 1174 640  
## 108 0 1 2 1 1105 640  
## 109 0 0 2 1 1335 640  
## 110 0 0 2 1 1550 640  
## 111 0 0 4 1 2473 640  
## 112 0 0 2 1 1322 640  
## 113 0 0 4 1 2553 640  
## 114 15 0 5 1 2835 753  
## 115 0 1 2 1 1191 640  
## 116 0 0 2 1 1439 640  
## 117 0 1 2 1 1237 640  
## 118 0 0 2 1 1265 640  
## 119 0 1 4 1 2280 640  
## 120 0 0 2 1 1096 640  
## 121 0 1 2 1 1156 640  
## 122 0 0 2 1 1199 640  
## 123 13 1 10 1 5615 754  
## 124 0 1 4 1 2518 640  
## 125 15 0 0 1 625 754  
## 126 0 1 2 1 1246 640  
## 127 0 1 3 1 1821 640  
## 128 0 0 5 1 3101 626  
## 129 12 1 2 1 1293 754  
## 130 0 1 2 1 1176 640  
## 131 0 0 3 1 1891 640  
## 132 5 1 2 1 10584 53  
## 133 13 1 3 1 2373 754  
## 134 0 0 1 1 935 640  
## 135 0 0 2 1 1395 640  
## 136 0 0 2 1 1561 640  
## 137 0 1 7 1 6912 636  
## 138 12 1 2 1 1157 754  
## 139 0 0 3 1 2197 640  
## 140 0 0 4 1 2288 640  
## 141 16 1 4 1 2348 754  
## 142 0 0 2 1 1320 640  
## 143 0 1 2 1 1139 640  
## 144 0 1 4 1 2134 639  
## 145 0 0 2 1 1407 640  
## 146 0 0 2 1 1982 640  
## 147 0 0 4 1 2539 640  
## 148 0 0 2 1 1528 640  
## 149 0 1 2 1 1513 640  
## 150 0 1 2 1 1191 640  
## 151 0 0 2 1 1280 640  
## 152 0 0 2 1 3977 139  
## 153 0 1 2 1 1269 640  
## 154 0 0 2 1 1501 640  
## 155 0 1 2 1 1396 640  
## 156 0 0 3 1 1777 640  
## 157 0 1 1 1 833 640  
## 158 0 1 1 1 715 640  
## 159 17 1 5 1 2936 751  
## 160 0 0 2 1 1375 640  
## 161 0 0 2 1 1330 640  
## 162 0 0 2 1 1628 640  
## 163 0 0 2 1 1368 640  
## 164 12 1 1 1 622 755  
## 165 17 0 2 1 14174 23  
## 166 7 0 1 1 6425 57  
## 167 3 0 1 1 8084 57  
## 168 4 1 1 1 6762 347  
## 169 0 0 39 1 26356 421  
## 170 0 0 2 1 1305 640  
## 171 0 0 2 1 1416 640  
## 172 0 0 2 1 877 640  
## 173 15 1 1 1 622 758  
## 174 0 1 4 1 2082 640  
## 175 0 1 2 1 1096 640  
## 176 0 0 3 1 2028 640  
## 177 0 0 2 1 1104 640  
## 178 0 1 2 1 1093 640  
## 179 0 1 4 1 2592 640  
## 180 13 0 1 1 1125 758  
## 181 0 0 4 1 3609 640  
## 182 0 0 3 1 2118 639  
## 183 0 0 2 1 1805 640  
## 184 0 0 1 1 2825 580  
## 185 0 1 2 1 1299 640  
## 186 0 0 2 1 1413 640  
## 187 0 1 3 1 2354 640  
## 188 16 1 0 1 622 754  
## 189 0 1 2 1 1214 640  
## 190 17 0 1 1 887 758  
## 191 0 1 3 1 1843 640  
## 192 16 1 2 1 1275 754  
## 193 0 1 4 1 2090 640  
## 194 0 1 2 1 1111 640  
## 195 15 1 5 1 3102 758  
## 196 15 1 1 1 743 758  
## 197 0 1 3 1 1505 640  
## 198 0 0 2 1 1277 640  
## 199 0 1 2 1 1096 640  
## 200 0 0 3 1 1828 640  
## 201 0 0 2 1 1581 640  
## 202 15 0 0 1 607 754  
## 203 0 1 2 1 1128 640  
## 204 0 0 2 1 1976 640  
## 205 17 0 6 1 4227 753  
## 206 17 0 5 1 3315 753  
## 207 0 0 3 1 2144 640  
## 208 0 0 2 1 1283 640  
## 209 13 1 2 1 1302 751  
## 210 0 1 3 1 1725 640  
## 211 0 1 0 1 2847 581  
## 212 0 1 3 1 1947 640  
## 213 5 0 3 1 7923 53  
## 214 0 0 3 1 1669 640  
## 215 0 0 3 1 1669 640  
## 216 14 1 1 1 707 751  
## 217 0 1 2 1 1404 640  
## 218 0 0 2 1 1272 640  
## 219 14 1 1 1 888 754  
## 220 0 1 2 1 1161 640  
## 221 16 0 1 1 2112 755  
## 222 0 0 2 1 1286 640  
## 223 0 0 2 1 1454 640  
## 224 0 0 1 1 874 640  
## 225 17 0 3 1 1753 750  
## 226 0 1 3 1 2129 626  
## 227 0 1 3 1 2149 626  
## 228 17 1 3 1 1749 753  
## 229 0 1 2 1 1317 640  
## 230 0 1 2 1 1281 640  
## 231 13 0 1 1 607 755  
## 232 0 0 3 1 2330 640  
## 233 0 0 8 1 5014 640  
## 234 0 1 3 1 2029 640  
## 235 0 1 3 1 1530 640  
## 236 0 0 3 1 2220 626  
## 237 0 0 2 1 1267 640  
## 238 0 0 3 1 2057 640  
## 239 0 0 2 1 1351 640  
## 240 0 1 18 1 10649 626  
## 241 0 1 3 1 2777 640  
## 242 0 1 3 1 1915 640  
## 243 0 0 2 1 1420 640  
## 244 0 0 2 1 1337 640  
## 245 15 0 6 1 20195 49  
## 246 0 1 2 1 1096 640  
## 247 0 0 3 1 2187 640  
## 248 17 1 3 1 3023 51  
## 249 0 1 3 1 1748 640  
## 250 0 1 2 1 1141 640  
## 251 0 0 3 1 1795 640  
## 252 0 1 2 1 1436 640  
## 253 0 0 2 1 1156 640  
## 254 12 1 3 1 2195 754  
## 255 0 0 2 1 1263 640  
## 256 0 1 2 1 1324 640  
## 257 16 0 1 2 8159 313  
## 258 0 0 3 1 1874 640  
## 259 0 0 2 1 1538 640  
## 260 0 0 4 1 3257 640  
## 261 0 1 2 1 1576 633  
## 262 0 0 15 1 8631 614  
## 263 0 1 2 1 1173 640  
## 264 14 0 2 1 2401 753  
## 265 0 0 3 1 2516 640  
## 266 0 0 2 1 1446 640  
## 267 0 0 3 1 1807 640  
## 268 0 0 2 1 1307 640  
## 269 0 1 2 1 1096 640  
## 270 0 0 2 1 1395 640  
## 271 0 0 3 1 1739 640  
## 272 0 1 2 1 1298 640  
## 273 0 0 1 1 921 640  
## 274 0 0 7 1 8184 634  
## 275 0 1 3 1 2051 640  
## 276 0 1 3 1 1743 640  
## 277 0 1 2 NA 1156 640  
## 278 0 0 2 1 1491 640  
## 279 0 0 2 1 1286 640  
## 280 0 1 3 1 1772 640  
## 281 0 0 2 1 1437 640  
## 282 0 1 4 1 2724 640  
## 283 0 0 2 2 1291 640  
## 284 0 1 2 1 1211 640  
## 285 0 1 2 1 1096 640  
## 286 0 0 2 1 1280 640  
## 287 13 1 1 1 735 755  
## 288 0 0 2 1 1280 640  
## 289 0 0 2 1 1383 640  
## 290 0 1 3 1 1660 640  
## 291 0 0 2 1 1458 640  
## 292 0 0 2 1 1427 640  
## 293 0 1 2 1 1217 640  
## 294 0 1 2 1 1096 640  
## 295 0 1 2 1 1099 640  
## 296 16 1 2 1 2711 812  
## 297 0 1 2 1 1099 640  
## 298 0 0 1 1 871 640  
## 299 17 0 2 1 1418 753  
## 300 17 1 4 1 2936 751  
## 301 16 0 1 1 2543 420  
## 302 17 0 1 1 4833 952  
## 303 16 1 0 1 2570 811  
## 304 16 0 2 2 10002 21  
## 305 16 0 2 1 1385 758  
## 306 16 1 3 1 6692 53  
## 307 15 1 1 1 783 758  
## 308 15 1 4 1 2255 751  
## 309 15 0 1 1 12024 92  
## 310 15 0 3 1 2089 751  
## 311 14 1 1 1 627 755  
## 312 14 1 2 1 1321 751  
## 313 14 1 1 1 763 755  
## 314 17 1 3 1 1688 751  
## 315 13 1 1 1 1760 751  
## 316 13 1 2 1 1306 754  
## 317 13 1 6 1 3674 758  
## 318 12 1 2 1 1348 753  
## 319 12 0 1 1 622 755  
## 320 12 0 1 1 767 755  
## 321 11 0 1 1 1178 422  
## 322 11 0 1 1 869 756  
## 323 10 0 7 1 17524 317  
## 324 14 1 1 1 625 756  
## 325 8 0 1 1 3588 139  
## 326 10 0 0 1 2925 344  
## 327 9 0 3 1 10562 114  
## 328 17 1 2 1 1436 758  
## 329 4 0 3 1 9230 206  
## 330 14 0 4 1 10756 53  
## 331 15 0 18 1 11685 753  
## 332 1 1 1 1 1711 723  
## 333 17 1 7 1 48388 911  
## 334 17 1 1 1 851 54  
## 335 15 1 1 1 634 753  
## 336 14 1 6 1 3656 753  
## 337 17 1 1 1 616 753  
## 338 0 1 1 1 1688 138  
## 339 15 0 9 1 16520 225  
## 340 12 1 3 1 1786 753  
## 341 14 1 1 1 972 751  
## 342 1 0 1 1 2117 58  
## 343 1 0 2 1 8050 53  
## 344 17 1 4 1 9932 53  
## 345 16 1 2 1 1422 753  
## 346 16 1 2 1 1381 753  
## 347 0 1 3 1 2877 422  
## 348 0 0 1 1 2805 138  
## 349 0 0 2 1 4412 138  
## 350 15 1 5 1 11125 740  
## 351 0 1 2 1 1818 249  
## 352 0 0 5 1 4296 640  
## 353 0 0 3 1 1887 640  
## 354 0 1 2 1 1418 640  
## 355 0 1 2 1 1136 640  
## 356 0 1 2 1 1475 640  
## 357 0 1 2 1 1099 640  
## 358 0 1 3 1 1538 640  
## 359 1 0 2 1 1651 249  
## 360 17 1 1 1 1106 753  
## 361 0 0 1 1 1097 640  
## 362 13 0 2 1 1273 754  
## 363 0 1 2 1 1362 640  
## 364 0 0 3 1 3342 640  
## 365 12 1 5 1 11877 344  
## 366 16 1 1 1 547 754  
## 367 0 0 6 1 5568 640  
## 368 0 1 7 1 4288 639  
## 369 0 1 7 1 5788 614  
## 370 16 1 4 1 2238 754  
## 371 0 0 2 1 1283 640  
## 372 0 1 0 1 2531 581  
## 373 0 1 23 1 13112 614  
## 374 0 0 3 1 3497 640  
## 375 0 0 3 1 1832 640  
## 376 0 1 3 1 1685 640  
## 377 0 0 2 1 1135 640  
## 378 0 1 2 1 1156 640  
## 379 13 1 1 1 548 751  
## 380 0 1 2 1 1303 640  
## 381 0 1 3 1 1890 640  
## 382 11 0 2 1 1285 753  
## 383 0 0 3 1 1674 640  
## 384 0 1 2 1 1128 640  
## 385 0 0 4 1 2595 640  
## 386 0 0 3 1 1782 640  
## 387 0 1 2 1 1251 640  
## 388 0 0 3 1 1997 640  
## 389 16 1 1 1 532 755  
## 390 0 0 2 1 1758 640  
## 391 0 0 2 1 1369 640  
## 392 0 0 2 1 1096 640  
## 393 0 0 2 1 1477 640  
## 394 16 1 1 1 840 753  
## 395 0 0 2 1 1340 640  
## 396 0 0 2 1 1236 640  
## 397 0 0 5 1 4677 640  
## 398 17 0 2 1 3285 754  
## 399 9 0 1 1 10585 308  
## 400 12 0 3 1 1808 753  
## 401 0 0 1 1 1051 640  
## 402 16 1 3 1 2201 758  
## 403 0 1 4 3 3041 626  
## 404 0 0 1 1 550 640  
## 405 0 1 2 1 1156 640  
## 406 17 1 1 1 3800 753  
## 407 0 0 2 1 1039 640  
## 408 0 1 3 1 1769 633  
## 409 0 0 2 1 1794 640  
## 410 0 0 2 1 1578 640  
## 411 8 0 2 2 1153 758  
## 412 0 0 3 1 1883 640  
## 413 0 0 2 6 1535 640  
## 414 0 1 2 1 1200 640  
## 415 0 1 3 1 1946 640  
## 416 0 1 2 1 1875 640  
## 417 0 1 3 1 1908 640  
## 418 0 0 3 1 2080 640  
## 419 0 0 2 1 1408 640  
## 420 0 0 3 1 2008 640  
## 421 17 0 4 1 2977 754  
## 422 0 0 2 1 1278 640  
## 423 0 0 2 1 1408 640  
## 424 0 1 2 1 1083 640  
## 425 0 0 3 1 1964 640  
## 426 0 0 2 1 1231 640  
## 427 0 0 3 1 1845 640  
## 428 0 0 2 1 1428 640  
## 429 14 1 5 4 2735 753  
## 430 0 1 7 1 5934 640  
## 431 0 1 3 1 1888 640  
## 432 0 1 3 1 1431 640  
## 433 14 0 2 1 9129 225  
## 434 0 0 1 1 651 640  
## 435 0 1 4 1 3969 640  
## 436 0 0 2 1 1363 640  
## 437 0 1 2 1 1118 640  
## 438 0 1 2 1 1543 640  
## 439 14 0 1 1 678 754  
## 440 14 1 2 1 1463 760  
## 441 17 0 3 1 8439 204  
## 442 0 1 5 1 4304 640  
## 443 0 1 3 1 1890 640  
## 444 0 0 3 1 1933 640  
## 445 0 1 2 1 1281 640  
## 446 10 1 2 1 1160 753  
## 447 0 0 2 1 1120 640  
## 448 16 1 2 1 1164 754  
## 449 0 1 2 1 1083 640  
## 450 0 0 2 1 1465 640  
## 451 0 1 2 1 1083 640  
## 452 0 0 2 1 1768 634  
## 453 14 1 4 1 2296 561  
## 454 0 0 2 1 1284 640  
## 455 13 1 4 1 2632 758  
## 456 0 0 3 1 1952 640  
## 457 0 0 4 1 2100 640  
## 458 17 1 2 1 2217 560  
## 459 0 0 2 1 1281 640  
## 460 0 1 2 1 1250 640  
## 461 6 0 2 1 8398 115  
## 462 0 0 2 1 1263 640  
## 463 0 1 2 1 1178 640  
## 464 0 1 24 1 13040 863  
## 465 0 1 3 1 1517 640  
## 466 0 0 3 4 3126 640  
## 467 0 1 4 1 2049 640  
## 468 0 0 2 1 1245 640  
## 469 16 0 1 1 3578 723  
## 470 0 1 3 1 1603 640  
## 471 0 1 3 1 2840 640  
## 472 0 1 2 1 1411 640  
## 473 0 1 2 1 1175 640  
## 474 0 0 2 1 1689 640  
## 475 0 1 3 1 2023 640  
## 476 12 0 2 1 1310 753  
## 477 0 0 2 1 2540 640  
## 478 0 0 7 1 10431 636  
## 479 0 0 3 1 2204 633  
## 480 0 0 3 1 2218 640  
## 481 0 0 2 1 1382 640  
## 482 0 1 2 1 1115 640  
## 483 11 0 1 1 629 753  
## 484 0 0 1 1 825 640  
## 485 0 0 3 1 1701 640  
## 486 0 0 2 1 1273 640  
## 487 0 0 2 1 1270 640  
## 488 0 1 2 1 1106 640  
## 489 0 1 2 1 1065 640  
## 490 0 0 2 1 1264 640  
## 491 0 0 2 1 1282 640  
## 492 0 0 2 1 1393 640  
## 493 15 1 4 2 3074 758  
## 494 0 0 2 1 1266 640  
## 495 0 1 3 1 1886 640  
## 496 0 1 6 1 5881 636  
## 497 0 1 2 1 1171 640  
## 498 0 1 2 1 1171 640  
## 499 0 1 2 1 1086 640  
## 500 0 0 4 1 4931 640

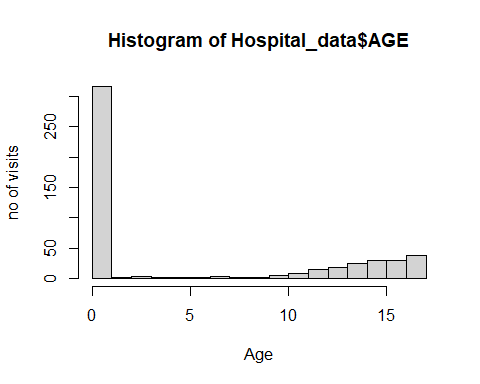
View(Hospital\_data)  
  
### the age category of people who frequently visit the hospital ####  
#### method 1 #####  
table(Hospital\_data$AGE)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17   
## 307 10 1 3 2 2 2 3 2 2 4 8 15 18 25 29 29 38

#### method 2 #####  
Hospital\_data%>% group\_by(AGE) %>% summarise(n())

## # A tibble: 18 x 2  
## AGE `n()`  
## <int> <int>  
## 1 0 307  
## 2 1 10  
## 3 2 1  
## 4 3 3  
## 5 4 2  
## 6 5 2  
## 7 6 2  
## 8 7 3  
## 9 8 2  
## 10 9 2  
## 11 10 4  
## 12 11 8  
## 13 12 15  
## 14 13 18  
## 15 14 25  
## 16 15 29  
## 17 16 29  
## 18 17 38

#### method 3 #####  
hist(Hospital\_data$AGE,xlab = "Age",ylab = "no of visits",breaks = 18)



#### method 4 #####  
summary(as.factor(Hospital\_data$AGE))

## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17   
## 307 10 1 3 2 2 2 3 2 2 4 8 15 18 25 29 29 38

### the age category of people who have maximum expenditure ####  
#### method 1 #####  
aggregate(Hospital\_data$TOTCHG, by=list(Hospital\_data$AGE), FUN=sum)

## Group.1 x  
## 1 0 678118  
## 2 1 37744  
## 3 2 7298  
## 4 3 30550  
## 5 4 15992  
## 6 5 18507  
## 7 6 17928  
## 8 7 10087  
## 9 8 4741  
## 10 9 21147  
## 11 10 24469  
## 12 11 14250  
## 13 12 54912  
## 14 13 31135  
## 15 14 64643  
## 16 15 111747  
## 17 16 69149  
## 18 17 174777

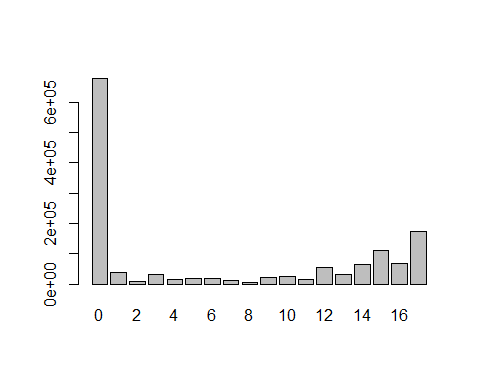
#### method 2 #####  
Hospital\_data%>% group\_by(AGE)%>% summarise(sum(TOTCHG))

## # A tibble: 18 x 2  
## AGE `sum(TOTCHG)`  
## <int> <int>  
## 1 0 678118  
## 2 1 37744  
## 3 2 7298  
## 4 3 30550  
## 5 4 15992  
## 6 5 18507  
## 7 6 17928  
## 8 7 10087  
## 9 8 4741  
## 10 9 21147  
## 11 10 24469  
## 12 11 14250  
## 13 12 54912  
## 14 13 31135  
## 15 14 64643  
## 16 15 111747  
## 17 16 69149  
## 18 17 174777

#### method 3######  
sort(tapply(Hospital\_data$TOTCHG,Hospital\_data$AGE,FUN=sum),decreasing = TRUE)

## 0 17 15 16 14 12 1 13 3 10 9   
## 678118 174777 111747 69149 64643 54912 37744 31135 30550 24469 21147   
## 5 6 4 11 7 2 8   
## 18507 17928 15992 14250 10087 7298 4741

#### method 4 ######  
barplot(tapply(Hospital\_data$TOTCHG,Hospital\_data$AGE,FUN=sum))



### diagnosis related group that has maximum expenditure.####  
#### method 1 #####  
Expenditure\_diagnosis <- aggregate(Hospital\_data$TOTCHG, by=list(Hospital\_data$APRDRG), FUN=sum)  
Expenditure\_diagnosis

## Group.1 x  
## 1 21 10002  
## 2 23 14174  
## 3 49 20195  
## 4 50 3908  
## 5 51 3023  
## 6 53 82271  
## 7 54 851  
## 8 57 14509  
## 9 58 2117  
## 10 92 12024  
## 11 97 9530  
## 12 114 10562  
## 13 115 25832  
## 14 137 15129  
## 15 138 13622  
## 16 139 17766  
## 17 141 2860  
## 18 143 1393  
## 19 204 8439  
## 20 206 9230  
## 21 225 25649  
## 22 249 16642  
## 23 254 615  
## 24 308 10585  
## 25 313 8159  
## 26 317 17524  
## 27 344 14802  
## 28 347 12597  
## 29 420 6357  
## 30 421 26356  
## 31 422 5177  
## 32 560 4877  
## 33 561 2296  
## 34 566 2129  
## 35 580 2825  
## 36 581 7453  
## 37 602 29188  
## 38 614 27531  
## 39 626 23289  
## 40 633 17591  
## 41 634 9952  
## 42 636 23224  
## 43 639 12612  
## 44 640 437978  
## 45 710 8223  
## 46 720 14243  
## 47 723 5289  
## 48 740 11125  
## 49 750 1753  
## 50 751 21666  
## 51 753 79542  
## 52 754 59150  
## 53 755 11168  
## 54 756 1494  
## 55 758 34953  
## 56 760 8273  
## 57 776 1193  
## 58 811 3838  
## 59 812 9524  
## 60 863 13040  
## 61 911 48388  
## 62 930 26654  
## 63 952 4833

Expenditure\_diagnosis %>% arrange(Expenditure\_diagnosis$x,desc(x))

## Group.1 x  
## 1 254 615  
## 2 54 851  
## 3 776 1193  
## 4 143 1393  
## 5 756 1494  
## 6 750 1753  
## 7 58 2117  
## 8 566 2129  
## 9 561 2296  
## 10 580 2825  
## 11 141 2860  
## 12 51 3023  
## 13 811 3838  
## 14 50 3908  
## 15 952 4833  
## 16 560 4877  
## 17 422 5177  
## 18 723 5289  
## 19 420 6357  
## 20 581 7453  
## 21 313 8159  
## 22 710 8223  
## 23 760 8273  
## 24 204 8439  
## 25 206 9230  
## 26 812 9524  
## 27 97 9530  
## 28 634 9952  
## 29 21 10002  
## 30 114 10562  
## 31 308 10585  
## 32 740 11125  
## 33 755 11168  
## 34 92 12024  
## 35 347 12597  
## 36 639 12612  
## 37 863 13040  
## 38 138 13622  
## 39 23 14174  
## 40 720 14243  
## 41 57 14509  
## 42 344 14802  
## 43 137 15129  
## 44 249 16642  
## 45 317 17524  
## 46 633 17591  
## 47 139 17766  
## 48 49 20195  
## 49 751 21666  
## 50 636 23224  
## 51 626 23289  
## 52 225 25649  
## 53 115 25832  
## 54 421 26356  
## 55 930 26654  
## 56 614 27531  
## 57 602 29188  
## 58 758 34953  
## 59 911 48388  
## 60 754 59150  
## 61 753 79542  
## 62 53 82271  
## 63 640 437978

##### method 2 #####  
sort(tapply(Hospital\_data$TOTCHG,Hospital\_data$APRDRG,FUN=sum),decreasing = TRUE)

## 640 53 753 754 911 758 602 614 930 421 115   
## 437978 82271 79542 59150 48388 34953 29188 27531 26654 26356 25832   
## 225 626 636 751 49 139 633 317 249 137 344   
## 25649 23289 23224 21666 20195 17766 17591 17524 16642 15129 14802   
## 57 720 23 138 863 639 347 92 755 740 308   
## 14509 14243 14174 13622 13040 12612 12597 12024 11168 11125 10585   
## 114 21 634 97 812 206 204 760 710 313 581   
## 10562 10002 9952 9530 9524 9230 8439 8273 8223 8159 7453   
## 420 723 422 560 952 50 811 51 141 580 561   
## 6357 5289 5177 4877 4833 3908 3838 3023 2860 2825 2296   
## 566 58 750 756 143 776 54 254   
## 2129 2117 1753 1494 1393 1193 851 615

#### diagnosis related group that has maximum hospitalization #####  
#### method 1 #####  
LOS\_diagnosis <- aggregate(Hospital\_data$LOS, by=list(Hospital\_data$APRDRG), FUN=sum)  
LOS\_diagnosis

## Group.1 x  
## 1 21 2  
## 2 23 2  
## 3 49 6  
## 4 50 2  
## 5 51 3  
## 6 53 29  
## 7 54 1  
## 8 57 2  
## 9 58 1  
## 10 92 1  
## 11 97 3  
## 12 114 3  
## 13 115 8  
## 14 137 12  
## 15 138 6  
## 16 139 7  
## 17 141 2  
## 18 143 2  
## 19 204 3  
## 20 206 3  
## 21 225 11  
## 22 249 8  
## 23 254 0  
## 24 308 1  
## 25 313 1  
## 26 317 7  
## 27 344 5  
## 28 347 1  
## 29 420 3  
## 30 421 39  
## 31 422 5  
## 32 560 4  
## 33 561 4  
## 34 566 2  
## 35 580 1  
## 36 581 1  
## 37 602 41  
## 38 614 45  
## 39 626 36  
## 40 633 25  
## 41 634 9  
## 42 636 20  
## 43 639 18  
## 44 640 652  
## 45 710 4  
## 46 720 5  
## 47 723 2  
## 48 740 5  
## 49 750 3  
## 50 751 32  
## 51 753 109  
## 52 754 85  
## 53 755 13  
## 54 756 2  
## 55 758 53  
## 56 760 14  
## 57 776 2  
## 58 811 1  
## 59 812 5  
## 60 863 24  
## 61 911 7  
## 62 930 10  
## 63 952 1

LOS\_diagnosis %>% arrange(LOS\_diagnosis$x,desc(x))

## Group.1 x  
## 1 254 0  
## 2 54 1  
## 3 58 1  
## 4 92 1  
## 5 308 1  
## 6 313 1  
## 7 347 1  
## 8 580 1  
## 9 581 1  
## 10 811 1  
## 11 952 1  
## 12 21 2  
## 13 23 2  
## 14 50 2  
## 15 57 2  
## 16 141 2  
## 17 143 2  
## 18 566 2  
## 19 723 2  
## 20 756 2  
## 21 776 2  
## 22 51 3  
## 23 97 3  
## 24 114 3  
## 25 204 3  
## 26 206 3  
## 27 420 3  
## 28 750 3  
## 29 560 4  
## 30 561 4  
## 31 710 4  
## 32 344 5  
## 33 422 5  
## 34 720 5  
## 35 740 5  
## 36 812 5  
## 37 49 6  
## 38 138 6  
## 39 139 7  
## 40 317 7  
## 41 911 7  
## 42 115 8  
## 43 249 8  
## 44 634 9  
## 45 930 10  
## 46 225 11  
## 47 137 12  
## 48 755 13  
## 49 760 14  
## 50 639 18  
## 51 636 20  
## 52 863 24  
## 53 633 25  
## 54 53 29  
## 55 751 32  
## 56 626 36  
## 57 421 39  
## 58 602 41  
## 59 614 45  
## 60 758 53  
## 61 754 85  
## 62 753 109  
## 63 640 652

##### method 2 #####  
sort(tapply(Hospital\_data$LOS,Hospital\_data$APRDRG,FUN=sum),decreasing = TRUE)

## 640 753 754 758 614 602 421 626 751 53 633 863 636 639 760 755 137 225 930 634   
## 652 109 85 53 45 41 39 36 32 29 25 24 20 18 14 13 12 11 10 9   
## 115 249 139 317 911 49 138 344 422 720 740 812 560 561 710 51 97 114 204 206   
## 8 8 7 7 7 6 6 5 5 5 5 5 4 4 4 3 3 3 3 3   
## 420 750 21 23 50 57 141 143 566 723 756 776 54 58 92 308 313 347 580 581   
## 3 3 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1   
## 811 952 254   
## 1 1 0

###### method 3 ####  
table(Hospital\_data$APRDRG)

##   
## 21 23 49 50 51 53 54 57 58 92 97 114 115 137 138 139 141 143 204 206   
## 1 1 1 1 1 10 1 2 1 1 1 1 2 1 4 5 1 1 1 1   
## 225 249 254 308 313 317 344 347 420 421 422 560 561 566 580 581 602 614 626 633   
## 2 6 1 1 1 1 2 3 2 1 3 2 1 1 1 3 1 3 6 4   
## 634 636 639 640 710 720 723 740 750 751 753 754 755 756 758 760 776 811 812 863   
## 2 3 4 267 1 1 2 1 1 14 36 37 13 2 20 2 1 2 3 1   
## 911 930 952   
## 1 2 1

####Race and expenditure###  
  
summary(as.factor(Hospital\_data$RACE)) ### indicates presence of a missing value

## 1 2 3 4 5 6 NA's   
## 484 6 1 3 3 2 1

Hospital\_data[is.na(Hospital\_data$RACE),"RACE"] <- median(Hospital\_data$RACE,na.rm = TRUE) ### adding the missing value to the highest cost group ###  
table(Hospital\_data$RACE)

##   
## 1 2 3 4 5 6   
## 485 6 1 3 3 2

### Ho= Race is not influencing hospital costs #### NULL hypothesis  
### Ha= Race is influencing hospital costs #### Alternative hypothesis  
model <- aov(TOTCHG~RACE,Hospital\_data)  
summary(model)### reject null hypothesis ####

## Df Sum Sq Mean Sq F value Pr(>F)  
## RACE 1 2.454e+06 2453726 0.162 0.687  
## Residuals 498 7.542e+09 15145146

######Gender,age and hospital cost#####  
  
model1 <- lm(TOTCHG~AGE+FEMALE,Hospital\_data)  
model1

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE, data = Hospital\_data)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE   
## 2718.63 86.28 -748.19

summary(model1) #### depends on age and gender####

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE, data = Hospital\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3406 -1443 -869 -152 44951   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2718.63 261.14 10.411 < 2e-16 \*\*\*  
## AGE 86.28 25.48 3.387 0.000763 \*\*\*  
## FEMALE -748.19 353.83 -2.115 0.034967 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3845 on 497 degrees of freedom  
## Multiple R-squared: 0.0261, Adjusted R-squared: 0.02218   
## F-statistic: 6.66 on 2 and 497 DF, p-value: 0.001399

##### to find if the length of stay can be predicted from age, gender, and race####  
  
model2 <- lm(LOS~AGE+FEMALE+RACE,Hospital\_data)  
model2

##   
## Call:  
## lm(formula = LOS ~ AGE + FEMALE + RACE, data = Hospital\_data)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE RACE   
## 2.94209 -0.03925 0.36421 -0.09365

summary(model2)

##   
## Call:  
## lm(formula = LOS ~ AGE + FEMALE + RACE, data = Hospital\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.213 -1.213 -0.848 0.152 37.787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.94209 0.39280 7.490 3.18e-13 \*\*\*  
## AGE -0.03925 0.02227 -1.762 0.0787 .   
## FEMALE 0.36421 0.30954 1.177 0.2399   
## RACE -0.09365 0.29286 -0.320 0.7493   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.361 on 496 degrees of freedom  
## Multiple R-squared: 0.007757, Adjusted R-squared: 0.001756   
## F-statistic: 1.293 on 3 and 496 DF, p-value: 0.2763

##### to find the variable that mainly affects the hospital costs ####  
  
model3 <- lm(TOTCHG~AGE+FEMALE+RACE+LOS+APRDRG,data = Hospital\_data)  
model3

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE + RACE + LOS + APRDRG, data = Hospital\_data)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE RACE LOS APRDRG   
## 5218.766 134.683 -390.493 -212.444 743.143 -7.791

summary(model3)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE + RACE + LOS + APRDRG, data = Hospital\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6377 -700 -173 122 43378   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5218.7658 507.1007 10.291 < 2e-16 \*\*\*  
## AGE 134.6828 17.4358 7.725 6.3e-14 \*\*\*  
## FEMALE -390.4928 247.1489 -1.580 0.115   
## RACE -212.4440 227.6998 -0.933 0.351   
## LOS 743.1433 34.8826 21.304 < 2e-16 \*\*\*  
## APRDRG -7.7910 0.6809 -11.442 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2611 on 494 degrees of freedom  
## Multiple R-squared: 0.5537, Adjusted R-squared: 0.5492   
## F-statistic: 122.6 on 5 and 494 DF, p-value: < 2.2e-16

####As it is apparent from the coefficient values, Age, Length of stay (LOS) and patient refined diagnosis related groups(APRDRG) have three stars (\*\*\*) next to it. So they are the ones with statistical significance  
### Gender and RACE are not significant. build a model after removing RACE and gender  
Hospital\_data$AGE <- as.factor(Hospital\_data$AGE)  
model4 <- lm(TOTCHG~AGE+LOS+APRDRG,data = Hospital\_data)  
model4

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS + APRDRG, data = Hospital\_data)  
##   
## Coefficients:  
## (Intercept) AGE1 AGE2 AGE3 AGE4 AGE5   
## 4281.602 -388.279 1911.682 6898.024 4166.584 3498.703   
## AGE6 AGE7 AGE8 AGE9 AGE10 AGE11   
## 3578.402 95.926 107.716 6287.811 2517.050 544.030   
## AGE12 AGE13 AGE14 AGE15 AGE16 AGE17   
## 2167.197 1005.690 922.418 1816.961 1498.882 2962.613   
## LOS APRDRG   
## 736.957 -6.966

summary(model4)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS + APRDRG, data = Hospital\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6754 -520 -100 139 42331   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4281.6024 521.1649 8.215 1.98e-15 \*\*\*  
## AGE1 -388.2789 878.0295 -0.442 0.658532   
## AGE2 1911.6818 2577.4445 0.742 0.458634   
## AGE3 6898.0243 1473.5343 4.681 3.71e-06 \*\*\*  
## AGE4 4166.5841 1818.4492 2.291 0.022379 \*   
## AGE5 3498.7033 1852.2846 1.889 0.059513 .   
## AGE6 3578.4016 1842.7379 1.942 0.052735 .   
## AGE7 95.9256 1513.6229 0.063 0.949494   
## AGE8 107.7156 1803.7380 0.060 0.952405   
## AGE9 6287.8107 1826.8043 3.442 0.000628 \*\*\*  
## AGE10 2517.0505 1288.3643 1.954 0.051320 .   
## AGE11 544.0300 909.9486 0.598 0.550209   
## AGE12 2167.1973 670.9393 3.230 0.001322 \*\*   
## AGE13 1005.6898 624.0602 1.612 0.107722   
## AGE14 922.4179 530.3730 1.739 0.082643 .   
## AGE15 1816.9612 493.1299 3.685 0.000255 \*\*\*  
## AGE16 1498.8825 496.1523 3.021 0.002654 \*\*   
## AGE17 2962.6135 436.9362 6.780 3.53e-11 \*\*\*  
## LOS 736.9570 34.0859 21.621 < 2e-16 \*\*\*  
## APRDRG -6.9660 0.7829 -8.898 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2534 on 480 degrees of freedom  
## Multiple R-squared: 0.5914, Adjusted R-squared: 0.5753   
## F-statistic: 36.57 on 19 and 480 DF, p-value: < 2.2e-16

### APRDRG has -ve t-value, dropping it ####  
  
model5 <- lm(TOTCHG~AGE+LOS,data = Hospital\_data)  
model5

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS, data = Hospital\_data)  
##   
## Coefficients:  
## (Intercept) AGE1 AGE2 AGE3 AGE4 AGE5   
## -75.66 2508.48 5883.02 7774.60 6581.02 7465.86   
## AGE6 AGE7 AGE8 AGE9 AGE10 AGE11   
## 7176.36 3189.55 1328.18 9158.52 4143.28 832.10   
## AGE12 AGE13 AGE14 AGE15 AGE16 AGE17   
## 1947.69 66.30 395.61 1564.54 1097.97 2713.69   
## LOS   
## 745.32

summary(model5)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS, data = Hospital\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4772 -635 -215 60 40533   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -75.66 192.35 -0.393 0.694237   
## AGE1 2508.48 879.22 2.853 0.004516 \*\*   
## AGE2 5883.02 2737.04 2.149 0.032100 \*   
## AGE3 7774.60 1585.22 4.904 1.28e-06 \*\*\*  
## AGE4 6581.02 1938.71 3.395 0.000744 \*\*\*  
## AGE5 7465.86 1938.43 3.851 0.000133 \*\*\*  
## AGE6 7176.36 1938.43 3.702 0.000239 \*\*\*  
## AGE7 3189.55 1588.36 2.008 0.045194 \*   
## AGE8 1328.18 1939.17 0.685 0.493723   
## AGE9 9158.52 1938.71 4.724 3.04e-06 \*\*\*  
## AGE10 4143.28 1375.08 3.013 0.002722 \*\*   
## AGE11 832.10 980.49 0.849 0.396498   
## AGE12 1947.69 722.92 2.694 0.007303 \*\*   
## AGE13 66.30 663.17 0.100 0.920402   
## AGE14 395.61 568.28 0.696 0.486669   
## AGE15 1564.54 530.82 2.947 0.003360 \*\*   
## AGE16 1097.97 532.74 2.061 0.039843 \*   
## AGE17 2713.69 470.14 5.772 1.41e-08 \*\*\*  
## LOS 745.32 36.74 20.288 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2732 on 481 degrees of freedom  
## Multiple R-squared: 0.524, Adjusted R-squared: 0.5062   
## F-statistic: 29.42 on 18 and 481 DF, p-value: < 2.2e-16

#### Removing Race and gender doesn't change the R2 value. It doesn't impact cost  
#### Removing APRDRG in model5 increases the standard error. Hence model model4 seems to be better.