PCA

KUSUM

2022-10-14

Sys.setenv(RGL\_USE\_NULL=TRUE)  
library(matlib)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(datasets)  
library(readr)  
library(base)  
library(dplyr)  
data("Seatbelts")   
head(Seatbelts)

## DriversKilled drivers front rear kms PetrolPrice VanKilled law  
## [1,] 107 1687 867 269 9059 0.1029718 12 0  
## [2,] 97 1508 825 265 7685 0.1023630 6 0  
## [3,] 102 1507 806 319 9963 0.1020625 12 0  
## [4,] 87 1385 814 407 10955 0.1008733 8 0  
## [5,] 119 1632 991 454 11823 0.1010197 10 0  
## [6,] 106 1511 945 427 12391 0.1005812 13 0

na.omit(Seatbelts)

## DriversKilled drivers front rear kms PetrolPrice VanKilled law  
## Jan 1969 107 1687 867 269 9059 0.10297181 12 0  
## Feb 1969 97 1508 825 265 7685 0.10236300 6 0  
## Mar 1969 102 1507 806 319 9963 0.10206249 12 0  
## Apr 1969 87 1385 814 407 10955 0.10087330 8 0  
## May 1969 119 1632 991 454 11823 0.10101967 10 0  
## Jun 1969 106 1511 945 427 12391 0.10058119 13 0  
## Jul 1969 110 1559 1004 522 13460 0.10377398 11 0  
## Aug 1969 106 1630 1091 536 14055 0.10407640 6 0  
## Sep 1969 107 1579 958 405 12106 0.10377398 10 0  
## Oct 1969 134 1653 850 437 11372 0.10302640 16 0  
## Nov 1969 147 2152 1109 434 9834 0.10273011 13 0  
## Dec 1969 180 2148 1113 437 9267 0.10199719 14 0  
## Jan 1970 125 1752 925 316 9130 0.10127456 14 0  
## Feb 1970 134 1765 903 311 8933 0.10070398 6 0  
## Mar 1970 110 1717 1006 351 11000 0.10013961 8 0  
## Apr 1970 102 1558 892 362 10733 0.09862110 11 0  
## May 1970 103 1575 990 486 12912 0.09834929 7 0  
## Jun 1970 111 1520 866 429 12926 0.09808018 13 0  
## Jul 1970 120 1805 1095 551 13990 0.09727921 13 0  
## Aug 1970 129 1800 1204 646 14926 0.09741062 11 0  
## Sep 1970 122 1719 1029 456 12900 0.09742524 11 0  
## Oct 1970 183 2008 1147 475 12034 0.09638063 14 0  
## Nov 1970 169 2242 1171 456 10643 0.09573896 16 0  
## Dec 1970 190 2478 1299 468 10742 0.09510631 14 0  
## Jan 1971 134 2030 944 356 10266 0.09673597 17 0  
## Feb 1971 108 1655 874 271 10281 0.09610922 16 0  
## Mar 1971 104 1693 840 354 11527 0.09536725 15 0  
## Apr 1971 117 1623 893 427 12281 0.09470959 13 0  
## May 1971 157 1805 1007 465 13587 0.09411762 13 0  
## Jun 1971 148 1746 973 440 13049 0.09353215 15 0  
## Jul 1971 130 1795 1097 539 16055 0.09295405 12 0  
## Aug 1971 140 1926 1194 646 15220 0.09283979 6 0  
## Sep 1971 136 1619 988 457 13824 0.09272474 9 0  
## Oct 1971 140 1992 1077 446 12729 0.09226965 13 0  
## Nov 1971 187 2233 1045 402 11467 0.09170669 14 0  
## Dec 1971 150 2192 1115 441 11351 0.09126207 15 0  
## Jan 1972 159 2080 1005 359 10803 0.09071160 14 0  
## Feb 1972 143 1768 857 334 10548 0.09027633 3 0  
## Mar 1972 114 1835 879 312 12368 0.08995192 12 0  
## Apr 1972 127 1569 887 427 13311 0.08909964 13 0  
## May 1972 159 1976 1075 434 13885 0.08867919 12 0  
## Jun 1972 156 1853 1121 486 14088 0.08815929 8 0  
## Jul 1972 138 1965 1190 569 16932 0.08890206 8 0  
## Aug 1972 120 1689 1058 523 16164 0.08818133 15 0  
## Sep 1972 117 1778 939 418 14883 0.08894029 8 0  
## Oct 1972 170 1976 1074 452 13532 0.08772661 5 0  
## Nov 1972 168 2397 1089 462 12220 0.08742885 17 0  
## Dec 1972 198 2654 1208 497 12025 0.08703543 14 0  
## Jan 1973 144 2097 903 354 11692 0.08644992 13 0  
## Feb 1973 146 1963 916 347 11081 0.08587264 5 0  
## Mar 1973 109 1677 787 276 13745 0.08539822 8 0  
## Apr 1973 131 1941 1114 472 14382 0.08382198 5 0  
## May 1973 151 2003 1014 487 14391 0.08459078 12 0  
## Jun 1973 140 1813 1022 505 15597 0.08413690 11 0  
## Jul 1973 153 2012 1114 619 16834 0.08377841 13 0  
## Aug 1973 140 1912 1132 640 17282 0.08351074 15 0  
## Sep 1973 161 2084 1111 559 15779 0.08280639 11 0  
## Oct 1973 168 2080 1008 453 13946 0.08117889 11 0  
## Nov 1973 152 2118 916 418 12701 0.08285361 10 0  
## Dec 1973 136 2150 992 419 10431 0.09419012 13 0  
## Jan 1974 113 1608 731 262 11616 0.09239984 8 0  
## Feb 1974 100 1503 665 299 10808 0.10816148 6 0  
## Mar 1974 103 1548 724 303 12421 0.10721169 8 0  
## Apr 1974 103 1382 744 401 13605 0.11404297 14 0  
## May 1974 121 1731 910 413 14455 0.11245412 12 0  
## Jun 1974 134 1798 883 426 15019 0.11131625 14 0  
## Jul 1974 133 1779 900 516 15662 0.11030125 13 0  
## Aug 1974 129 1887 1057 600 16745 0.10819718 9 0  
## Sep 1974 144 2004 1076 459 14717 0.10702744 4 0  
## Oct 1974 154 2077 919 443 13756 0.10494698 13 0  
## Nov 1974 156 2092 920 412 12531 0.11935775 6 0  
## Dec 1974 163 2051 953 400 12568 0.11762190 15 0  
## Jan 1975 122 1577 664 278 11249 0.13302742 12 0  
## Feb 1975 92 1356 607 302 11096 0.13084524 16 0  
## Mar 1975 117 1652 777 381 12637 0.12831848 7 0  
## Apr 1975 95 1382 633 279 13018 0.12354745 12 0  
## May 1975 96 1519 791 442 15005 0.11858681 10 0  
## Jun 1975 108 1421 790 409 15235 0.11633748 9 0  
## Jul 1975 108 1442 803 416 15552 0.11516148 9 0  
## Aug 1975 106 1543 884 511 16905 0.11450120 6 0  
## Sep 1975 140 1656 769 393 14776 0.11352298 7 0  
## Oct 1975 114 1561 732 345 14104 0.11193018 13 0  
## Nov 1975 158 1905 859 391 12854 0.11061053 14 0  
## Dec 1975 161 2199 994 470 12956 0.11527439 13 0  
## Jan 1976 102 1473 704 266 12177 0.11379349 14 0  
## Feb 1976 127 1655 684 312 11918 0.11234958 11 0  
## Mar 1976 125 1407 671 300 13517 0.11175347 11 0  
## Apr 1976 101 1395 643 373 14417 0.10964252 10 0  
## May 1976 97 1530 771 412 15911 0.10844090 4 0  
## Jun 1976 112 1309 644 322 15589 0.10788494 8 0  
## Jul 1976 112 1526 828 458 16543 0.10908477 9 0  
## Aug 1976 113 1327 748 427 17925 0.10757145 10 0  
## Sep 1976 108 1627 767 346 15406 0.10616402 10 0  
## Oct 1976 128 1748 825 421 14601 0.10630000 5 0  
## Nov 1976 154 1958 810 344 13107 0.10482531 13 0  
## Dec 1976 162 2274 986 370 12268 0.10345175 12 0  
## Jan 1977 112 1648 714 291 11972 0.10144992 10 0  
## Feb 1977 79 1401 567 224 12028 0.10040232 9 0  
## Mar 1977 82 1411 616 266 14033 0.09886203 7 0  
## Apr 1977 127 1403 678 338 14244 0.10249615 5 0  
## May 1977 108 1394 742 298 15287 0.10302743 10 0  
## Jun 1977 110 1520 840 386 16954 0.10217891 5 0  
## Jul 1977 123 1528 888 479 17361 0.09983664 6 0  
## Aug 1977 103 1643 852 473 17694 0.09263669 8 0  
## Sep 1977 97 1515 774 332 16222 0.09181496 6 0  
## Oct 1977 140 1685 831 391 14969 0.09072430 12 0  
## Nov 1977 165 2000 889 370 13624 0.09002121 15 0  
## Dec 1977 183 2215 1046 431 13842 0.08933071 7 0  
## Jan 1978 148 1956 889 366 12387 0.08844273 14 0  
## Feb 1978 111 1462 626 250 11608 0.08835257 4 0  
## Mar 1978 116 1563 808 355 15021 0.08675736 10 0  
## Apr 1978 115 1459 746 304 14834 0.08499524 8 0  
## May 1978 100 1446 754 379 16565 0.08456794 7 0  
## Jun 1978 106 1622 865 440 16882 0.08443190 11 0  
## Jul 1978 134 1657 980 500 18012 0.08435088 3 0  
## Aug 1978 125 1638 959 511 18855 0.08360098 5 0  
## Sep 1978 117 1643 856 384 17243 0.08341726 11 0  
## Oct 1978 122 1683 798 366 16045 0.08274514 10 0  
## Nov 1978 153 2050 942 432 14745 0.08523527 10 0  
## Dec 1978 178 2262 1010 390 13726 0.08477030 7 0  
## Jan 1979 114 1813 796 306 11196 0.08445892 10 0  
## Feb 1979 94 1445 643 232 12105 0.08535212 11 0  
## Mar 1979 128 1762 794 342 14723 0.08755921 9 0  
## Apr 1979 119 1461 750 329 15582 0.09038292 7 0  
## May 1979 111 1556 809 394 16863 0.09078329 8 0  
## Jun 1979 110 1431 716 355 16758 0.10874278 13 0  
## Jul 1979 114 1427 851 385 17434 0.11414223 8 0  
## Aug 1979 118 1554 931 463 18359 0.11299293 5 0  
## Sep 1979 115 1645 834 453 17189 0.11132071 8 0  
## Oct 1979 132 1653 762 373 16909 0.10912623 7 0  
## Nov 1979 153 2016 880 401 15380 0.10769846 12 0  
## Dec 1979 171 2207 1077 466 15161 0.10760157 10 0  
## Jan 1980 115 1665 748 306 14027 0.10377502 7 0  
## Feb 1980 95 1361 593 263 14478 0.10711417 4 0  
## Mar 1980 92 1506 720 323 16155 0.10737477 10 0  
## Apr 1980 100 1360 646 310 16585 0.11169537 4 0  
## May 1980 95 1453 765 424 18117 0.11063818 8 0  
## Jun 1980 114 1522 820 403 17552 0.11185521 8 0  
## Jul 1980 102 1460 807 406 18299 0.10974234 7 0  
## Aug 1980 104 1552 885 466 19361 0.10819393 10 0  
## Sep 1980 132 1548 803 381 17924 0.10625536 8 0  
## Oct 1980 136 1827 860 369 17872 0.10419303 14 0  
## Nov 1980 117 1737 825 378 16058 0.10193397 8 0  
## Dec 1980 137 1941 911 392 15746 0.10279382 9 0  
## Jan 1981 111 1474 704 284 15226 0.10476034 8 0  
## Feb 1981 106 1458 691 316 14932 0.10400254 6 0  
## Mar 1981 98 1542 688 321 16846 0.11665552 7 0  
## Apr 1981 84 1404 714 358 16854 0.11516148 6 0  
## May 1981 94 1522 814 378 18146 0.11298954 5 0  
## Jun 1981 105 1385 736 382 17559 0.11386064 4 0  
## Jul 1981 123 1641 876 433 18655 0.11911808 5 0  
## Aug 1981 109 1510 829 506 19453 0.12448999 10 0  
## Sep 1981 130 1681 818 428 17923 0.12322295 7 0  
## Oct 1981 153 1938 942 479 17915 0.12067793 10 0  
## Nov 1981 134 1868 782 370 16496 0.12104898 12 0  
## Dec 1981 99 1726 823 349 13544 0.11696857 7 0  
## Jan 1982 115 1456 595 238 13601 0.11275026 4 0  
## Feb 1982 104 1445 673 285 15667 0.10807931 5 0  
## Mar 1982 131 1456 660 324 17358 0.10883852 6 0  
## Apr 1982 108 1365 676 346 18112 0.11129177 4 0  
## May 1982 103 1487 755 410 18581 0.11130401 4 0  
## Jun 1982 115 1558 815 411 18759 0.11545436 8 0  
## Jul 1982 122 1488 867 496 20668 0.11476830 8 0  
## Aug 1982 122 1684 933 534 21040 0.11720743 3 0  
## Sep 1982 125 1594 798 396 18993 0.11907640 7 0  
## Oct 1982 137 1850 950 470 18668 0.11796586 12 0  
## Nov 1982 138 1998 825 385 16768 0.11744913 2 0  
## Dec 1982 152 2079 911 411 16551 0.11698846 7 0  
## Jan 1983 120 1494 619 281 16231 0.11261054 8 0  
## Feb 1983 95 1057 426 300 15511 0.11365702 3 1  
## Mar 1983 100 1218 475 318 18308 0.11314445 2 1  
## Apr 1983 89 1168 556 391 17793 0.11849553 6 1  
## May 1983 82 1236 559 398 19205 0.11796940 3 1  
## Jun 1983 89 1076 483 337 19162 0.11768661 7 1  
## Jul 1983 60 1174 587 477 20997 0.12005924 6 1  
## Aug 1983 84 1139 615 422 20705 0.11943775 8 1  
## Sep 1983 113 1427 618 495 18759 0.11888127 8 1  
## Oct 1983 126 1487 662 471 19240 0.11846236 4 1  
## Nov 1983 122 1483 519 368 17504 0.11801660 3 1  
## Dec 1983 118 1513 585 345 16591 0.11770662 5 1  
## Jan 1984 92 1357 483 296 16224 0.11777609 5 1  
## Feb 1984 86 1165 434 319 16670 0.11479699 3 1  
## Mar 1984 81 1282 513 349 18539 0.11573525 4 1  
## Apr 1984 84 1110 548 375 19759 0.11535626 3 1  
## May 1984 87 1297 586 441 19584 0.11481536 6 1  
## Jun 1984 90 1185 522 465 19976 0.11477748 6 1  
## Jul 1984 79 1222 601 472 21486 0.11493598 7 1  
## Aug 1984 96 1284 644 521 21626 0.11479699 5 1  
## Sep 1984 122 1444 643 429 20195 0.11409316 7 1  
## Oct 1984 120 1575 641 408 19928 0.11646552 7 1  
## Nov 1984 137 1737 711 490 18564 0.11602611 4 1  
## Dec 1984 154 1763 721 491 18149 0.11606673 7 1

df1 <- as.data.frame(Seatbelts)  
head(df1)

## DriversKilled drivers front rear kms PetrolPrice VanKilled law  
## 1 107 1687 867 269 9059 0.1029718 12 0  
## 2 97 1508 825 265 7685 0.1023630 6 0  
## 3 102 1507 806 319 9963 0.1020625 12 0  
## 4 87 1385 814 407 10955 0.1008733 8 0  
## 5 119 1632 991 454 11823 0.1010197 10 0  
## 6 106 1511 945 427 12391 0.1005812 13 0

## Question 1

Consider the inbuilt dataset Seatbelts. Please use help() to learn more about the data.

### Task 1: Generate the covariance and correlation matrix. What can you say about the variables in the data

mtrx1 <- as.matrix(df1)  
head(mtrx1)

## DriversKilled drivers front rear kms PetrolPrice VanKilled law  
## [1,] 107 1687 867 269 9059 0.1029718 12 0  
## [2,] 97 1508 825 265 7685 0.1023630 6 0  
## [3,] 102 1507 806 319 9963 0.1020625 12 0  
## [4,] 87 1385 814 407 10955 0.1008733 8 0  
## [5,] 119 1632 991 454 11823 0.1010197 10 0  
## [6,] 106 1511 945 427 12391 0.1005812 13 0

covM1=cov(mtrx1)  
covM1

## DriversKilled drivers front rear  
## DriversKilled 6.441386e+02 6.533134e+03 3.140834e+03 745.2613438  
## drivers 6.533134e+03 8.387451e+04 4.099501e+04 8271.1764834  
## front 3.140834e+03 4.099501e+04 3.065965e+04 9024.9594241  
## rear 7.452613e+02 8.271176e+03 9.024959e+03 6905.9773124  
## kms -2.394370e+04 -3.784451e+05 -1.838551e+05 81306.4232112  
## PetrolPrice -1.194695e-01 -1.613852e+00 -1.149645e+00 -0.1341974  
## VanKilled 3.757161e+01 5.112650e+02 3.008460e+02 36.7995201  
## law -2.714387e+00 -4.197941e+01 -3.206299e+01 0.7864311  
## kms PetrolPrice VanKilled law  
## DriversKilled -23943.69655 -0.1194695154 3.757161e+01 -2.714386998  
## drivers -378445.06621 -1.6138524988 5.112650e+02 -41.979412086  
## front -183855.12238 -1.1496454293 3.008460e+02 -32.062990838  
## rear 81306.42321 -0.1341973554 3.679952e+01 0.786431065  
## kms 8632133.14092 13.7333452831 -5.321710e+03 469.225676265  
## PetrolPrice 13.73335 0.0001482509 -1.277804e-02 0.001548726  
## VanKilled -5321.71019 -0.0127780401 1.322707e+01 -0.467631981  
## law 469.22568 0.0015487260 -4.676320e-01 0.105993674

corM1 = cor(mtrx1)  
corM1

## DriversKilled drivers front rear kms  
## DriversKilled 1.0000000 0.8888264 0.7067596 0.35335102 -0.3211016  
## drivers 0.8888264 1.0000000 0.8084114 0.34366850 -0.4447631  
## front 0.7067596 0.8084114 1.0000000 0.62022476 -0.3573823  
## rear 0.3533510 0.3436685 0.6202248 1.00000000 0.3330069  
## kms -0.3211016 -0.4447631 -0.3573823 0.33300689 1.0000000  
## PetrolPrice -0.3866061 -0.4576675 -0.5392394 -0.13262721 0.3839004  
## VanKilled 0.4070412 0.4853995 0.4724207 0.12175808 -0.4980356  
## law -0.3285051 -0.4452269 -0.5624455 0.02906753 0.4905494  
## PetrolPrice VanKilled law  
## DriversKilled -0.3866061 0.4070412 -0.32850510  
## drivers -0.4576675 0.4853995 -0.44522689  
## front -0.5392394 0.4724207 -0.56244554  
## rear -0.1326272 0.1217581 0.02906753  
## kms 0.3839004 -0.4980356 0.49054938  
## PetrolPrice 1.0000000 -0.2885584 0.39069323  
## VanKilled -0.2885584 1.0000000 -0.39494121  
## law 0.3906932 -0.3949412 1.00000000

The matrix is asymmetric, the covariance suggests how each value is differs from the other. The covariance matrix is semi-definite. Through covariance we can determine how each variable is related (for eg., the rear and kms have a negative covariance as the values in kms has negative value) For correlation: this matrix is symmetric and the matrix is positive definite. Some variables in the above matrix are inversely proportional to each other.

### Task 2: Check if the covariance matrix is orthognal

covM1T = t(covM1)  
covM1T

## DriversKilled drivers front rear  
## DriversKilled 6.441386e+02 6.533134e+03 3.140834e+03 745.2613438  
## drivers 6.533134e+03 8.387451e+04 4.099501e+04 8271.1764834  
## front 3.140834e+03 4.099501e+04 3.065965e+04 9024.9594241  
## rear 7.452613e+02 8.271176e+03 9.024959e+03 6905.9773124  
## kms -2.394370e+04 -3.784451e+05 -1.838551e+05 81306.4232112  
## PetrolPrice -1.194695e-01 -1.613852e+00 -1.149645e+00 -0.1341974  
## VanKilled 3.757161e+01 5.112650e+02 3.008460e+02 36.7995201  
## law -2.714387e+00 -4.197941e+01 -3.206299e+01 0.7864311  
## kms PetrolPrice VanKilled law  
## DriversKilled -23943.69655 -0.1194695154 3.757161e+01 -2.714386998  
## drivers -378445.06621 -1.6138524988 5.112650e+02 -41.979412086  
## front -183855.12238 -1.1496454293 3.008460e+02 -32.062990838  
## rear 81306.42321 -0.1341973554 3.679952e+01 0.786431065  
## kms 8632133.14092 13.7333452831 -5.321710e+03 469.225676265  
## PetrolPrice 13.73335 0.0001482509 -1.277804e-02 0.001548726  
## VanKilled -5321.71019 -0.0127780401 1.322707e+01 -0.467631981  
## law 469.22568 0.0015487260 -4.676320e-01 0.105993674

omat=covM1T %\*% covM1  
omat

## DriversKilled drivers front rear  
## DriversKilled 6.268190e+08 9.748488e+09 4.775055e+09 -1858765527  
## drivers 9.748488e+09 1.520475e+11 7.436971e+10 -29644287322  
## front 4.775055e+09 7.436971e+10 3.651472e+10 -14268145385  
## rear -1.858766e+09 -2.964429e+10 -1.426815e+10 6808846000  
## kms -2.096901e+11 -3.305554e+12 -1.607556e+12 697601859953  
## PetrolPrice -3.431589e+05 -5.381705e+06 -2.627944e+06 1091869  
## VanKilled 1.317586e+08 2.069747e+09 1.009061e+09 -425462710  
## law -1.161114e+07 -1.824230e+08 -8.897510e+07 37517866  
## kms PetrolPrice VanKilled law  
## DriversKilled -2.096901e+11 -3.431589e+05 1.317586e+08 -1.161114e+07  
## drivers -3.305554e+12 -5.381705e+06 2.069747e+09 -1.824230e+08  
## front -1.607556e+12 -2.627944e+06 1.009061e+09 -8.897510e+07  
## rear 6.976019e+11 1.091869e+06 -4.254627e+08 3.751787e+07  
## kms 7.469796e+13 1.193622e+08 -4.618449e+10 4.072332e+09  
## PetrolPrice 1.193622e+08 1.925634e+02 -7.426545e+04 6.548873e+03  
## VanKilled -4.618449e+10 -7.426545e+04 2.867544e+07 -2.528271e+06  
## law 4.072332e+09 6.548873e+03 -2.528271e+06 2.229713e+05

is\_orthogonal\_matrix(omat)

## [1] FALSE

This suggests that the matrix is not orthogonal, since the resulting matrix of the product of covariance matrix and its transpose is not an identity matrix

### Task 3: Compute the eigenvalues and eigenvectors for covariance and correlation matrix. What did you observe from your analysis?

ecovM1=eigen(covM1)  
ecovM1

## eigen() decomposition  
## $values  
## [1] 8.653669e+06 8.887021e+04 1.017188e+04 1.382469e+03 1.290075e+02  
## [6] 8.436161e+00 4.236133e-02 9.402641e-05  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] -2.803988e-03 -6.918498e-02 4.581489e-02 -3.156923e-02 9.960454e-01  
## [2,] -4.420087e-02 -8.506155e-01 5.110515e-01 -7.787800e-02 -8.518609e-02  
## [3,] -2.149609e-02 -4.846563e-01 -7.335282e-01 4.757592e-01 1.509634e-02  
## [4,] 9.326331e-03 -1.859958e-01 -4.455659e-01 -8.753904e-01 -2.015858e-02  
## [5,] 9.987437e-01 -4.653636e-02 1.111680e-02 1.487509e-02 -4.599015e-04  
## [6,] 1.595997e-06 1.489955e-05 2.217581e-05 -6.919467e-05 -1.935179e-05  
## [7,] -6.175254e-04 -3.854373e-03 -3.271329e-03 -6.751971e-03 1.011423e-03  
## [8,] 5.445045e-05 3.314404e-04 6.693543e-04 -4.054510e-03 1.212178e-03  
## [,6] [,7] [,8]  
## [1,] 0.0013265015 -1.355558e-03 1.183774e-05  
## [2,] 0.0020711901 -2.896868e-04 -6.551244e-06  
## [3,] 0.0011078950 2.552946e-03 6.707090e-05  
## [4,] 0.0080292325 -3.240857e-03 -6.038384e-05  
## [5,] -0.0005745090 1.986889e-05 -1.120366e-07  
## [6,] -0.0001126865 -3.999820e-03 9.999920e-01  
## [7,] -0.9999198174 9.370268e-03 -7.551531e-05  
## [8,] 0.0093952450 9.999386e-01 4.000389e-03

Eigen vectors gives us vectors in principal directions.These vectors do not change the directions when transformations are performed like scaling.

ecorM1=eigen(corM1)  
ecorM1

## eigen() decomposition  
## $values  
## [1] 4.03362903 1.57815295 0.73720569 0.64061615 0.56293459 0.30569022 0.09557333  
## [8] 0.04619805  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] -0.4080022 -0.17871091 -0.25491038 0.4861386 -0.16113152 -0.3040074  
## [2,] -0.4501636 -0.10462795 -0.18642612 0.3392325 -0.15088056 -0.0661405  
## [3,] -0.4529427 -0.22338322 0.12423577 -0.1650995 -0.10215441 0.3127628  
## [4,] -0.1833141 -0.68358056 0.04508572 -0.3320193 0.11033792 0.3288694  
## [5,] 0.2878262 -0.55389084 0.12112979 -0.1335604 -0.08051612 -0.6998498  
## [6,] 0.3171550 -0.09538963 -0.72129718 -0.1780651 -0.53759412 0.2062737  
## [7,] -0.3240129 0.18802266 -0.51075334 -0.4886636 0.50574589 -0.3203784  
## [8,] 0.3220408 -0.29780228 -0.29493683 0.4719727 0.61432282 0.2494962  
## [,7] [,8]  
## [1,] 0.60921245 0.10374436  
## [2,] -0.69238591 -0.36060314  
## [3,] -0.18355907 0.74853223  
## [4,] 0.23231744 -0.46102396  
## [5,] -0.21500732 0.18801435  
## [6,] -0.02775146 0.07752596  
## [7,] -0.01777660 0.03556438  
## [8,] -0.12034254 0.20912771

Eigenvectors gives us vectors in principal directions.These vectors do not change the directions when transformations are performed like scaling. The matrix is positive definite.

### Task 4: Find the squareroot of the covariance matrix using spectral decomposition method

eigval <- eigen(covM1)$values  
eigvec <- eigen(covM1)$vectors  
  
matinv <- eigvec %\*% diag(1/eigval) %\*% t(eigvec)  
matinv

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 7.736360e-03 -6.441945e-04 2.968999e-05 -4.014583e-05 -4.544770e-06  
## [2,] -6.441945e-04 9.740092e-05 -9.084317e-05 7.035932e-05 1.954578e-07  
## [3,] 2.968999e-05 -9.084317e-05 4.228776e-04 -5.077998e-04 5.556944e-06  
## [4,] -4.014583e-05 7.035932e-05 -5.077998e-04 8.717233e-04 -1.173056e-05  
## [5,] -4.544770e-06 1.954578e-07 5.556944e-06 -1.173056e-05 3.620554e-07  
## [6,] 1.260248e-01 -6.964661e-02 7.130730e-01 -6.418898e-01 -1.193403e-03  
## [7,] -4.586306e-04 -3.047254e-04 3.775767e-04 -1.615796e-03 7.250237e-05  
## [8,] -3.148336e-02 -7.115011e-03 6.311573e-02 -7.905813e-02 4.635510e-04  
## [,6] [,7] [,8]  
## [1,] 1.260248e-01 -4.586306e-04 -0.031483356  
## [2,] -6.964661e-02 -3.047254e-04 -0.007115011  
## [3,] 7.130730e-01 3.775767e-04 0.063115730  
## [4,] -6.418898e-01 -1.615796e-03 -0.079058134  
## [5,] -1.193403e-03 7.250237e-05 0.000463551  
## [6,] 1.063514e+04 -8.039937e-01 42.450618186  
## [7,] -8.039937e-01 1.206517e-01 0.216858624  
## [8,] 4.245062e+01 2.168586e-01 23.773743846

matsqrt <- eigvec %\*% diag(sqrt(eigval)) %\*% t(eigvec)   
matsqrt

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 12.9673017814 19.39747684 6.396194106 2.499876e+00 -7.249656e+00  
## [2,] 19.3974768389 248.09333272 86.493099261 2.554054e+01 -1.175319e+02  
## [3,] 6.3961941061 86.49309926 134.068473999 4.375776e+01 -5.699160e+01  
## [4,] 2.4998759564 25.54053812 43.757756198 5.908899e+01 2.899760e+01  
## [5,] -7.2496560247 -117.53192317 -56.991598705 2.899760e+01 2.934991e+03  
## [6,] -0.0003549118 -0.00262414 -0.005123359 4.771715e-04 4.469229e-03  
## [7,] 0.0849863074 0.90162263 0.715471265 5.399883e-01 -1.766567e+00  
## [8,] 0.0140378683 -0.04606136 -0.171807992 8.427931e-02 1.538683e-01  
## [,6] [,7] [,8]  
## [1,] -0.0003549118 0.0849863074 0.0140378683  
## [2,] -0.0026241396 0.9016226255 -0.0460613640  
## [3,] -0.0051233593 0.7154712655 -0.1718079916  
## [4,] 0.0004771715 0.5399883286 0.0842793094  
## [5,] 0.0044692286 -1.7665667665 0.1538683304  
## [6,] 0.0097002013 0.0003086393 -0.0007740823  
## [7,] 0.0003086393 2.9123958939 -0.0250266972  
## [8,] -0.0007740823 -0.0250266972 0.2067645444

### Task 5: Based on the eigen decomposition in task 3, determine how many principal components you would select to reduce feature dimensions yet capture atleast 85% of the variability in the data? Perform the analysis using the correlation matrix.

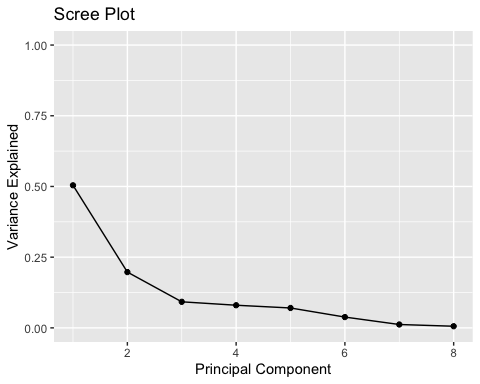
PVEcor <- eigen(corM1)$values / sum(eigen(corM1)$values)  
PVEcor

## [1] 0.504203628 0.197269118 0.092150711 0.080077019 0.070366823 0.038211278  
## [7] 0.011946666 0.005774756

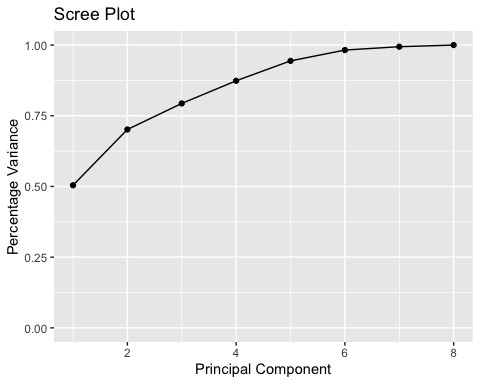
cumsumcor <- cumsum(PVEcor)  
cumsumcor

## [1] 0.5042036 0.7014727 0.7936235 0.8737005 0.9440673 0.9822786 0.9942252  
## [8] 1.0000000

library(ggplot2)  
pca<-prcomp(corM1, scale = TRUE)  
variance = pca $sdev^2 / sum(pca $sdev^2)  
qplot(c(1:8), PVEcor) +  
 geom\_line() +  
 xlab("Principal Component") +  
 ylab("Variance Explained") +  
 ggtitle("Scree Plot") +  
 ylim(0,1)



qplot(c(1:8), cumsumcor) +  
 geom\_line() +  
 xlab("Principal Component") +  
 ylab("Percentage Variance") +  
 ggtitle("Scree Plot") +  
 ylim(0,1)

 We can choose principal components of columns 1 to 4 to capture at least 85% of the variability in data. First 4 components explain 85% of the variance

### Task 6: Compute the principal component vectors based on your selection in task 5. Comment on your interpretation of the PCs

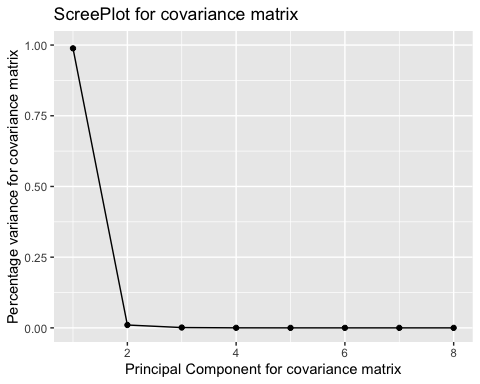
evc = eigen(corM1)$vectors[,1:4]  
colnames(evc) = c("PC1","PC2","PC3","PC4")  
row.names(evc) = colnames(Seatbelts)  
evc

## PC1 PC2 PC3 PC4  
## DriversKilled -0.4080022 -0.17871091 -0.25491038 0.4861386  
## drivers -0.4501636 -0.10462795 -0.18642612 0.3392325  
## front -0.4529427 -0.22338322 0.12423577 -0.1650995  
## rear -0.1833141 -0.68358056 0.04508572 -0.3320193  
## kms 0.2878262 -0.55389084 0.12112979 -0.1335604  
## PetrolPrice 0.3171550 -0.09538963 -0.72129718 -0.1780651  
## VanKilled -0.3240129 0.18802266 -0.51075334 -0.4886636  
## law 0.3220408 -0.29780228 -0.29493683 0.4719727

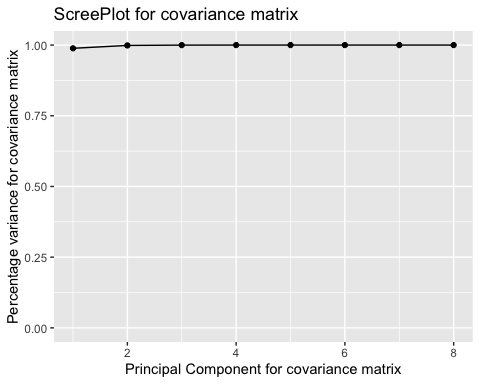
The first PC measures the variance distance driven and petrol price. This component is associated with high ratings. They’re all positively related to PC1 as they are positive values The second PC measures the number of van drivers killed. Hence, this could mean that the number of van drivers killed is more than others The third PC talks about the severity of drivers killed in either front or rear seat The fourth PC tells us that when law was nt enforced the number of deaths was more

### Task 7: Perform task 5 and 6 using covariance matrix. Compare the results with the ones obtained from correlation matrix. Do the interpretation of the components differ?

library(ggplot2)  
PVECov <- eigen(covM1)$values / sum(eigen(covM1)$values)  
cumsumcov<-cumsum(PVECov)  
  
library(ggplot2)  
qplot(c(1:8), PVECov) +  
 geom\_line() +  
 xlab("Principal Component for covariance matrix")+  
 ylab("Percentage variance for covariance matrix")+  
 ggtitle("ScreePlot for covariance matrix")+  
 ylim(0, 1)



qplot(c(1:8), cumsumcov)+  
geom\_line() +  
 xlab("Principal Component for covariance matrix")+  
 ylab("Percentage variance for covariance matrix")+  
 ggtitle("ScreePlot for covariance matrix")+  
 ylim(0, 1)



evc = eigen(covM1)$vectors[,1:4]  
colnames(evc) = c("PC1","PC2","PC3","PC4")  
row.names(evc) = colnames(Seatbelts)  
evc

## PC1 PC2 PC3 PC4  
## DriversKilled -2.803988e-03 -6.918498e-02 4.581489e-02 -3.156923e-02  
## drivers -4.420087e-02 -8.506155e-01 5.110515e-01 -7.787800e-02  
## front -2.149609e-02 -4.846563e-01 -7.335282e-01 4.757592e-01  
## rear 9.326331e-03 -1.859958e-01 -4.455659e-01 -8.753904e-01  
## kms 9.987437e-01 -4.653636e-02 1.111680e-02 1.487509e-02  
## PetrolPrice 1.595997e-06 1.489955e-05 2.217581e-05 -6.919467e-05  
## VanKilled -6.175254e-04 -3.854373e-03 -3.271329e-03 -6.751971e-03  
## law 5.445045e-05 3.314404e-04 6.693543e-04 -4.054510e-03

Yes, there’s a difference in the interpretation of covariance most of the principal components tend to be capture almost 100% of the variance. PC1 is not tending towards maximum variance