

team10

2022-12-05

```
library(ggplot2)
library(moderndiver)
library(dplyr)
```

```
##
## 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
```

```
library(readr)
library(faraway)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --

## v tibble 3.1.8      v stringr 1.4.1
## v tidyr  1.2.1      v forcats 0.5.2
## v purrr  0.3.5

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(olsrr)
```

```
##
## Attaching package: 'olsrr'
##
## The following object is masked from 'package:faraway':
##
##   hsb

## The following object is masked from 'package:datasets':
##
##   rivers
```

```
library(glmnet)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
```

```
##      expand, pack, unpack
##
## Loaded glmnet 4.1-6
```

Cleaning Data set:

```
lifeExpectancyData1 <- read_csv("Life Expectancy Data.csv")
```

```
## Rows: 2938 Columns: 22
## -- Column specification -----
## Delimiter: ","
## chr  (2): Country, Status
## dbl (20): Year, Life expectancy, Adult Mortality, infant deaths, Alcohol, pe...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
lifeExpectancyData <- lifeExpectancyData1 %>%
  rename(Life.expectancy = "Life expectancy") %>%
  rename(Adult.Mortality = "Adult Mortality") %>%
  rename(infant.deaths = "infant deaths") %>%
  rename(percentage.expenditure = "percentage expenditure") %>%
  rename(Measles = "Measles") %>%
  rename(under.five.deaths = "under-five deaths" ) %>%
  rename(Polio = "Polio") %>%
  rename(Diphtheria = "Diphtheria" ) %>%
  rename(HIV.AIDS = "HIV/AIDS")
```

```
lifeExpectancyDataUpdated <- lifeExpectancyData[which(grepl(pattern="2015", lifeExpectancyData$Year)),]
```

After looking at the dataset, only 8 of the variables had values for all the countries (some of the columns had NA entries):

```
lmod <- lm(Life.expectancy~Adult.Mortality+infant.deaths+percentage.expenditure+ Measles+under.five.deaths+Polio+Diphtheria+HIV.AIDS, data = lifeExpectancyDataUpdated)
summary(lmod)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     percentage.expenditure + Measles + under.five.deaths + Polio +
##     Diphtheria + HIV.AIDS, data = lifeExpectancyDataUpdated)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8445  -2.5420   0.3501   3.3574  10.8480
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.120e+01  1.809e+00  39.356 < 2e-16 ***
## Adult.Mortality -4.881e-02  4.517e-03 -10.806 < 2e-16 ***
## infant.deaths   6.752e-02  4.644e-02   1.454  0.14780
## percentage.expenditure 2.410e-03  1.220e-02   0.198  0.84358
## Measles         4.743e-06  8.175e-05   0.058  0.95380
## under.five.deaths -5.958e-02  3.380e-02  -1.763  0.07972 .
## Polio           4.352e-02  1.892e-02   2.300  0.02261 *
## Diphtheria       5.775e-02  2.091e-02   2.762  0.00636 **
## HIV.AIDS        -5.543e-01  3.381e-01  -1.639  0.10296
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.509 on 174 degrees of freedom
## Multiple R-squared:  0.7054, Adjusted R-squared:  0.6919
## F-statistic: 52.08 on 8 and 174 DF,  p-value: < 2.2e-16
```

```
anova(lmod)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: Life.expectancy
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Adult.Mortality	1	7291.0	7291.0	358.5390	< 2.2e-16 ***
infant.deaths	1	111.6	111.6	5.4863	0.020297 *
percentage.expenditure	1	1.2	1.2	0.0581	0.809784 .
Measles	1	63.7	63.7	3.1338	0.078436 .
under.five.deaths	1	128.2	128.2	6.3030	0.012966 *
Polio	1	612.0	612.0	30.0950	1.432e-07 ***
Diphtheria	1	210.4	210.4	10.3448	0.001548 **
HIV.AIDS	1	54.6	54.6	2.6873	0.102956
Residuals	174	3538.3	20.3		

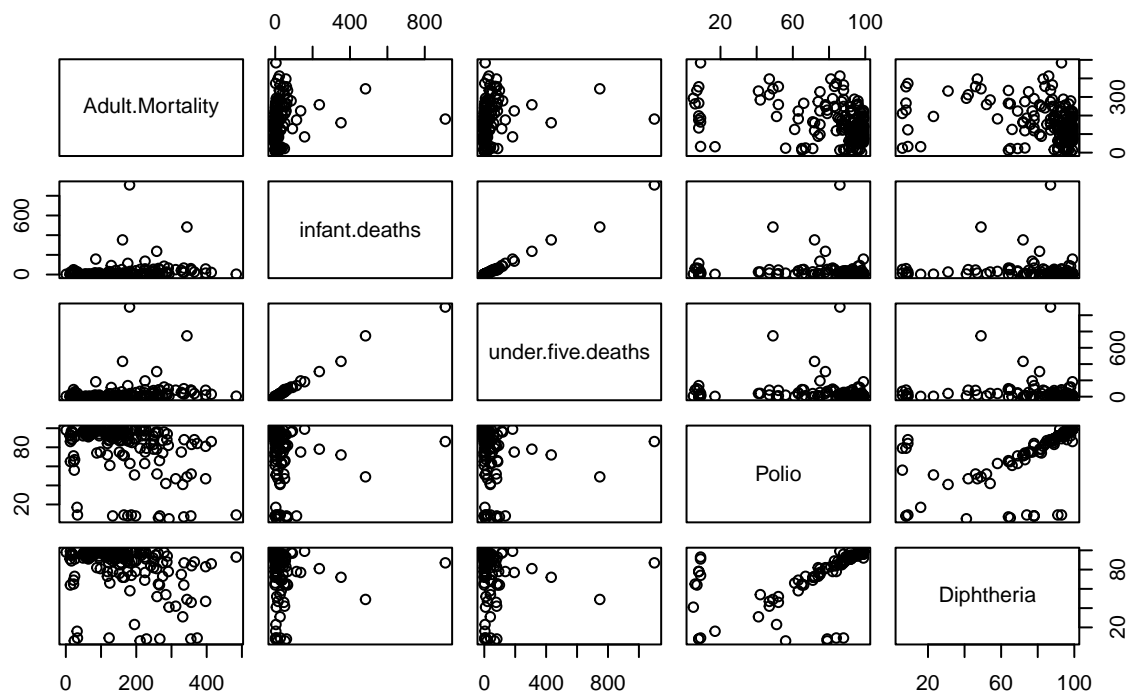
```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the ANOVA table, we see that Adult Mortality, infant deaths, under-five deaths, polio, diphtheria have significant p values

```
pairs(~ Adult.Mortality + infant.deaths + under.five.deaths + Polio + Diphtheria, data=lifeExpectancyData
```

Scatterplot Matrix



```
lifeExpectancyDataUpdated2 <- lifeExpectancyDataUpdated[c(-1,-2,-3,-7,-8,-9, -10, -11, -14, -16, -17, -18,-19,-20,-21,-22,-23,-24,-25,-26,-27,-28,-29,-30,-31,-32,-33,-34,-35,-36,-37,-38,-39,-40,-41,-42,-43,-44,-45,-46,-47,-48,-49,-50,-51,-52,-53,-54,-55,-56,-57,-58,-59,-60,-61,-62,-63,-64,-65,-66,-67,-68,-69,-70,-71,-72,-73,-74,-75,-76,-77,-78,-79,-80,-81,-82,-83,-84,-85,-86,-87,-88,-89,-90,-91,-92,-93,-94,-95,-96,-97,-98,-99,-100,-101,-102,-103,-104,-105,-106,-107,-108,-109,-110,-111,-112,-113,-114,-115,-116,-117,-118,-119,-120,-121,-122,-123,-124,-125,-126,-127,-128,-129,-130,-131,-132,-133,-134,-135,-136,-137,-138,-139,-140,-141,-142,-143,-144,-145,-146,-147,-148,-149,-150,-151,-152,-153,-154,-155,-156,-157,-158,-159,-160,-161,-162,-163,-164,-165,-166,-167,-168,-169,-170,-171,-172,-173,-174,-175,-176,-177,-178,-179,-180,-181,-182,-183,-184,-185,-186,-187,-188,-189,-190,-191,-192,-193,-194,-195,-196,-197,-198,-199,-200,-201,-202,-203,-204,-205,-206,-207,-208,-209,-210,-211,-212,-213,-214,-215,-216,-217,-218,-219,-220,-221,-222,-223,-224,-225,-226,-227,-228,-229,-230,-231,-232,-233,-234,-235,-236,-237,-238,-239,-240,-241,-242,-243,-244,-245,-246,-247,-248,-249,-250,-251,-252,-253,-254,-255,-256,-257,-258,-259,-260,-261,-262,-263,-264,-265,-266,-267,-268,-269,-270,-271,-272,-273,-274,-275,-276,-277,-278,-279,-280,-281,-282,-283,-284,-285,-286,-287,-288,-289,-290,-291,-292,-293,-294,-295,-296,-297,-298,-299,-300,-301,-302,-303,-304,-305,-306,-307,-308,-309,-310,-311,-312,-313,-314,-315,-316,-317,-318,-319,-320,-321,-322,-323,-324,-325,-326,-327,-328,-329,-330,-331,-332,-333,-334,-335,-336,-337,-338,-339,-340,-341,-342,-343,-344,-345,-346,-347,-348,-349,-350,-351,-352,-353,-354,-355,-356,-357,-358,-359,-360,-361,-362,-363,-364,-365,-366,-367,-368,-369,-370,-371,-372,-373,-374,-375,-376,-377,-378,-379,-380,-381,-382,-383,-384,-385,-386,-387,-388,-389,-390,-391,-392,-393,-394,-395,-396,-397,-398,-399,-400,-401,-402,-403,-404,-405,-406,-407,-408,-409,-410,-411,-412,-413,-414,-415,-416,-417,-418,-419,-420,-421,-422,-423,-424,-425,-426,-427,-428,-429,-430,-431,-432,-433,-434,-435,-436,-437,-438,-439,-440,-441,-442,-443,-444,-445,-446,-447,-448,-449,-450,-451,-452,-453,-454,-455,-456,-457,-458,-459,-460,-461,-462,-463,-464,-465,-466,-467,-468,-469,-470,-471,-472,-473,-474,-475,-476,-477,-478,-479,-480,-481,-482,-483,-484,-485,-486,-487,-488,-489,-490,-491,-492,-493,-494,-495,-496,-497,-498,-499,-500,-501,-502,-503,-504,-505,-506,-507,-508,-509,-510,-511,-512,-513,-514,-515,-516,-517,-518,-519,-520,-521,-522,-523,-524,-525,-526,-527,-528,-529,-530,-531,-532,-533,-534,-535,-536,-537,-538,-539,-540,-541,-542,-543,-544,-545,-546,-547,-548,-549,-550,-551,-552,-553,-554,-555,-556,-557,-558,-559,-560,-561,-562,-563,-564,-565,-566,-567,-568,-569,-570,-571,-572,-573,-574,-575,-576,-577,-578,-579,-580,-581,-582,-583,-584,-585,-586,-587,-588,-589,-590,-591,-592,-593,-594,-595,-596,-597,-598,-599,-600,-601,-602,-603,-604,-605,-606,-607,-608,-609,-610,-611,-612,-613,-614,-615,-616,-617,-618,-619,-620,-621,-622,-623,-624,-625,-626,-627,-628,-629,-630,-631,-632,-633,-634,-635,-636,-637,-638,-639,-640,-641,-642,-643,-644,-645,-646,-647,-648,-649,-650,-651,-652,-653,-654,-655,-656,-657,-658,-659,-660,-661,-662,-663,-664,-665,-666,-667,-668,-669,-670,-671,-672,-673,-674,-675,-676,-677,-678,-679,-680,-681,-682,-683,-684,-685,-686,-687,-688,-689,-690,-691,-692,-693,-694,-695,-696,-697,-698,-699,-700,-701,-702,-703,-704,-705,-706,-707,-708,-709,-710,-711,-712,-713,-714,-715,-716,-717,-718,-719,-720,-721,-722,-723,-724,-725,-726,-727,-728,-729,-730,-731,-732,-733,-734,-735,-736,-737,-738,-739,-740,-741,-742,-743,-744,-745,-746,-747,-748,-749,-750,-751,-752,-753,-754,-755,-756,-757,-758,-759,-760,-761,-762,-763,-764,-765,-766,-767,-768,-769,-770,-771,-772,-773,-774,-775,-776,-777,-778,-779,-780,-781,-782,-783,-784,-785,-786,-787,-788,-789,-790,-791,-792,-793,-794,-795,-796,-797,-798,-799,-800,-801,-802,-803,-804,-805,-806,-807,-808,-809,-810,-811,-812,-813,-814,-815,-816,-817,-818,-819,-820,-821,-822,-823,-824,-825,-826,-827,-828,-829,-830,-831,-832,-833,-834,-835,-836,-837,-838,-839,-840,-841,-842,-843,-844,-845,-846,-847,-848,-849,-850,-851,-852,-853,-854,-855,-856,-857,-858,-859,-860,-861,-862,-863,-864,-865,-866,-867,-868,-869,-870,-871,-872,-873,-874,-875,-876,-877,-878,-879,-880,-881,-882,-883,-884,-885,-886,-887,-888,-889,-890,-891,-892,-893,-894,-895,-896,-897,-898,-899,-900,-901,-902,-903,-904,-905,-906,-907,-908,-909,-910,-911,-912,-913,-914,-915,-916,-917,-918,-919,-920,-921,-922,-923,-924,-925,-926,-927,-928,-929,-930,-931,-932,-933,-934,-935,-936,-937,-938,-939,-940,-941,-942,-943,-944,-945,-946,-947,-948,-949,-950,-951,-952,-953,-954,-955,-956,-957,-958,-959,-960,-961,-962,-963,-964,-965,-966,-967,-968,-969,-970,-971,-972,-973,-974,-975,-976,-977,-978,-979,-980,-981,-982,-983,-984,-985,-986,-987,-988,-989,-990,-991,-992,-993,-994,-995,-996,-997,-998,-999,-1000,-1001,-1002,-1003,-1004,-1005,-1006,-1007,-1008,-1009,-1010,-1011,-1012,-1013,-1014,-1015,-1016,-1017,-1018,-1019,-1020,-1021,-1022,-1023,-1024,-1025,-1026,-1027,-1028,-1029,-1030,-1031,-1032,-1033,-1034,-1035,-1036,-1037,-1038,-1039,-1040,-1041,-1042,-1043,-1044,-1045,-1046,-1047,-1048,-1049,-1050,-1051,-1052,-1053,-1054,-1055,-1056,-1057,-1058,-1059,-1060,-1061,-1062,-1063,-1064,-1065,-1066,-1067,-1068,-1069,-1070,-1071,-1072,-1073,-1074,-1075,-1076,-1077,-1078,-1079,-1080,-1081,-1082,-1083,-1084,-1085,-1086,-1087,-1088,-1089,-1090,-1091,-1092,-1093,-1094,-1095,-1096,-1097,-1098,-1099,-1100,-1101,-1102,-1103,-1104,-1105,-1106,-1107,-1108,-1109,-1110,-1111,-1112,-1113,-1114,-1115,-1116,-1117,-1118,-1119,-1120,-1121,-1122,-1123,-1124,-1125,-1126,-1127,-1128,-1129,-1130,-1131,-1132,-1133,-1134,-1135,-1136,-1137,-1138,-1139,-1140,-1141,-1142,-1143,-1144,-1145,-1146,-1147,-1148,-1149,-1150,-1151,-1152,-1153,-1154,-1155,-1156,-1157,-1158,-1159,-1160,-1161,-1162,-1163,-1164,-1165,-1166,-1167,-1168,-1169,-1170,-1171,-1172,-1173,-1174,-1175,-1176,-1177,-1178,-1179,-1180,-1181,-1182,-1183,-1184,-1185,-1186,-1187,-1188,-1189,-1190,-1191,-1192,-1193,-1194,-1195,-1196,-1197,-1198,-1199,-1200,-1201,-1202,-1203,-1204,-1205,-1206,-1207,-1208,-1209,-1210,-1211,-1212,-1213,-1214,-1215,-1216,-1217,-1218,-1219,-1220,-1221,-1222,-1223,-1224,-1225,-1226,-1227,-1228,-1229,-1230,-1231,-1232,-1233,-1234,-1235,-1236,-1237,-1238,-1239,-1240,-1241,-1242,-1243,-1244,-1245,-1246,-1247,-1248,-1249,-1250,-1251,-1252,-1253,-1254,-1255,-1256,-1257,-1258,-1259,-1260,-1261,-1262,-1263,-1264,-1265,-1266,-1267,-1268,-1269,-1270,-1271,-1272,-1273,-1274,-1275,-1276,-1277,-1278,-1279,-1280,-1281,-1282,-1283,-1284,-1285,-1286,-1287,-1288,-1289,-1290,-1291,-1292,-1293,-1294,-1295,-1296,-1297,-1298,-1299,-1300,-1301,-1302,-1303,-1304,-1305,-1306,-1307,-1308,-1309,-1310,-1311,-1312,-1313,-1314,-1315,-1316,-1317,-1318,-1319,-1320,-1321,-1322,-1323,-1324,-1325,-1326,-1327,-1328,-1329,-1330,-1331,-1332,-1333,-1334,-1335,-1336,-1337,-1338,-1339,-1340,-1341,-1342,-1343,-1344,-1345,-1346,-1347,-1348,-1349,-1350,-1351,-1352,-1353,-1354,-1355,-1356,-1357,-1358,-1359,-1360,-1361,-1362,-1363,-1364,-1365,-1366,-1367,-1368,-1369,-1370,-1371,-1372,-1373,-1374,-1375,-1376,-1377,-1378,-1379,-1380,-1381,-1382,-1383,-1384,-1385,-1386,-1387,-1388,-1389,-1390,-1391,-1392,-1393,-1394,-1395,-1396,-1397,-1398,-1399,-1400,-1401,-1402,-1403,-1404,-1405,-1406,-1407,-1408,-1409,-1410,-1411,-1412,-1413,-1414,-1415,-1416,-1417,-1418,-1419,-1420,-1421,-1422,-1423,-1424,-1425,-1426,-1427,-1428,-1429,-1430,-1431,-1432,-1433,-1434,-1435,-1436,-1437,-1438,-1439,-1440,-1441,-1442,-1443,-1444,-1445,-1446,-1447,-1448,-1449,-1450,-1451,-1452,-1453,-1454,-1455,-1456,-1457,-1458,-1459,-1460,-1461,-1462,-1463,-1464,-1465,-1466,-1467,-1468,-1469,-1470,-1471,-1472,-1473,-1474,-1475,-1476,-1477,-1478,-1479,-1480,-1481,-1482,-1483,-1484,-1485,-1486,-1487,-1488,-1489,-1490,-1491,-1492,-1493,-1494,-1495,-1496,-1497,-1498,-1499,-1500,-1501,-1502,-1503,-1504,-1505,-1506,-1507,-1508,-1509,-1510,-1511,-1512,-1513,-1514,-1515,-1516,-1517,-1518,-1519,-1520,-1521,-1522,-1523,-1524,-1525,-1526,-1527,-1528,-1529,-1530,-1531,-1532,-1533,-1534,-1535,-1536,-1537,-1538,-1539,-1540,-1541,-1542,-1543,-1544,-1545,-1546,-1547,-1548,-1549,-1550,-1551,-1552,-1553,-1554,-1555,-1556,-1557,-1558,-1559,-1560,-1561,-1562,-1563,-1564,-1565,-1566,-1567,-1568,-1569,-1570,-1571,-1572,-1573,-1574,-1575,-1576,-1577,-1578,-1579,-1580,-1581,-1582,-1583,-1584,-1585,-1586,-1587,-1588,-1589,-1590,-1591,-1592,-1593,-1594,-1595,-1596,-1597,-1598,-1599,-1600,-1601,-1602,-1603,-1604,-1605,-1606,-1607,-1608,-1609,-1610,-1611,-1612,-1613,-1614,-1615,-1616,-1617,-1618,-1619,-1620,-1621,-1622,-1623,-1624,-1625,-1626,-1627,-1628,-1629,-1630,-1631,-1632,-1633,-1634,-1635,-1636,-1637,-1638,-1639,-1640,-1641,-1642,-1643,-1644,-1645,-1646,-1647,-1648,-1649,-1650,-1651,-1652,-1653,-1654,-1655,-1656,-1657,-1658,-1659,-1660,-1661,-1662,-1663,-1664,-1665,-1666,-1667,-1668,-1669,-1670,-1671,-1672,-1673,-1674,-1675,-1676,-1677,-1678,-1679,-1680,-1681,-1682,-1683,-1684,-1685,-1686,-1687,-1688,-1689,-1690,-1691,-1692,-1693,-1694,-1695,-1696,-1697,-1698,-1699,-1700,-1701,-1702,-1703,-1704,-1705,-1706,-1707,-1708,-1709,-1710,-1711,-1712,-1713,-1714,-1715,-1716,-1717,-1718,-1719,-1720,-1721,-1722,-1723,-1724,-1725,-1726,-1727,-1728,-1729,-1730,-1731,-1732,-1733,-1734,-1735,-1736,-1737,-1738,-1739,-1740,-1741,-1742,-1743,-1744,-1745,-1746,-1747,-1748,-1749,-1750,-1751,-1752,-1753,-1754,-1755,-1756,-1757,-1758,-1759,-1760,-1761,-1762,-1763,-1764,-1765,-1766,-1767,-1768,-1769,-1770,-1771,-1772,-1773,-1774,-1775,-1776,-1777,-1778,-1779,-1780,-1781,-1782,-1783,-1784,-1785,-1786,-1787,-1788,-1789,-1790,-1791,-1792,-1793,-1794,-1795,-1796,-1797,-1798,-1799,-1800,-1801,-1802,-1803,-1804,-1805,-1806,-1807,-1808,-1809,-1810,-1811,-1812,-1813,-1814,-1815,-1816,-1817,-1818,-1819,-1820,-1821,-1822,-1823,-1824,-1825,-1826,-1827,-1828,-1829,-1830,-1831,-1832,-1833,-1834,-1835,-1836,-1837,-1838,-1839,-1840,-1841,-1842,-1843,-1844,-1845,-1846,-1847,-1848,-1849,-1850,-1851,-1852,-1853,-1854,-1855,-1856,-1857,-1858,-1859,-1860,-1861,-1862,-1863,-1864,-1865,-1866,-1867,-1868,-1869,-1870,-1871,-1872,-1873,-1874,-1875,-1876,-1877,-1878,-1879,-1880,-1881,-1882,-1883,-1884,-1885,-1886,-1887,-1888,-1889,-1890,-1891,-1892,-1893,-1894,-1895,-1896,-1897,-1898,-1899,-1900,-1901,-1902,-1903,-1904,-1905,-1906,-1907,-1908,-1909,-1910,-1911,-1912,-1913,-1914,-1915,-1916,-1917,-1918,-1919,-1920,-1921,-1922,-1923,-1924,-1925,-1926,-1927,-1928,-1929,-1930,-1931,-1932,-1933,-1934,-1935,-1936,-1937,-1938,-1939,-1940,-1941,-1942,-1943,-1944,-1945,-1946,-1947,-1948,-1949,-1950,-1951,-1952,-1953,-1954,-1955,-1956,-1957,-1958,-1959,-1960,-1961,-1962,-1963,-1964,-1965,-1966,-1967,-1968,-1969,-1970,-1971,-1972,-1973,-1974,-1975,-1976,-1977,-1978,-1979,-1980,-1981,-1982,-1983,-1984,-1985,-1986,-1987,-1988,-1989,-1990,-1991,-1992,-1993,-1994,-1995,-1996,-1997,-1998,-1999,-2000,-2001,-2002,-2003,-2004,-2005,-2006,-2007,-2008,-2009,-2010,-2011,-2012,-2013,-2014,-2015,-2016,-2017,-2018,-2019,-2020,-2021,-2022,-2023,-2024,-2025,-2026,-2027,-2028,-2029,-2030,-2031,-2032,-2033,-2034,-2035,-2036,-2037,-2038,-2039,-2040,-2041,-2042,-2043,-2044,-2045,-2046,-2047,-2048,-2049,-2050,-2051,-2052,-2053,-2054,-2055,-2056,-2057,-2058,-2059,-2060,-2061,-2062,-2063,-2064,-2065,-2066,-2067,-2068,-2069,-2070,-2071,-2072,-2073,-2074,-2075,-2076,-2077,-2078,-2079,-2080,-2081,-2082,-2083,-2084,-2085,-2086,-2087,-2088,-2089,-2090,-2091,-2092,-2093,-2094,-2095,-2096,-2097,-2098,-2099,-2100,-2101,-2102,-2103,-2104,-2105,-2106,-2107,-2108,-2109,-2110,-2111,-2112,-2113,-2114,-2115,-2116,-2117,-2118,-2119,-2120,-2121,-2122,-2123,-2124,-2125,-2126,-2127,-2128,-2129,-2130,-2131,-2132,-2133,-2134,-2135,-2136,-2137,-2138,-2139,-2140,-2141,-2142,-2143,-2144,-2145,-2146,-2147,-2148,-2149,-2150,-2151,-2152,-2153,-2154,-2155,-2156,-2157,-2158,-2159,-2160,-2161,-2162,-2163,-2164,-2165,-2166,-2167,-2168,-2169,-2170,-2171,-2172,-2173,-2174,-2175,-2176,-2177,-2178,-2179,-2180,-2181,-2182,-2183,-2184,-2185,-2186,-2187,-2188,-2189,-2190,-2191,-2192,-2193,-2194,-2195,-2196,-2197,-2198,-2199,-2200,-2201,-2202,-2203,-2204,-2205,-2206,-2207,-2208,-2209,-2210,-2211,-2212,-2213,-2214,-2215,-2216,-2217,-2218,-2219,-2220,-2221,-2222,-2223,-2224,-2225,-2226,-2227,-2228,-2229,-2230,-2231,-2232,-2233,-2234,-2235,-2236,-2237,-2238,-2239,-2240,-2241,-2242,-2243,-2244,-2245,-2246,-2247,-2248,-2249,-2250,-2251,-2252,-2253,-2254,-2255,-2256,-2257,-2258,-2259,-2260,-2261,-2262,-2263,-2264,-2265,-2266,-2267,-2268,-2269,-2270,-2271,-2272,-2273,-2274,-2275,-2276,-2277,-2278,-2279,-2280,-2281,-2282,-2283,-2284,-2285,-2286,-2287,-2288,-2289,-2290,-2291,-2292,-2293,-2294,-2295,-2296,-2297,-2298,-2299,-2300,-2301,-2302,-2303,-2304,-2305,-2306,-2307,-2308,-2309,-2310,-2311,-2312,-2313,-2314,-2315,-2316,-2317,-2318,-2319,-2320,-2321,-2322,-2323,-2324,-2325,-2326,-2327,-2328,-2329,-2330,-2331,-2332,-2333,-2334,-2335,-2336,-2337,-2338,-2339,-2340,-2341,-2342,-2343,-2344,-2345,-2346,-2347,-2348,-2349,-2350,-2351,-2352,-2353,-2354,-2355,-2356,-2357,-2358,-2359,-2360,-2361,-2362,-2363,-2364,-2365,-2366,-2367,-2368,-2369,-2370,-2371,-2372,-2373,-2374,-2375,-2376,-2377,-2378,-2379,-2380,-2381,-2382,-2383,-2384,-2385,-2386,-2387,-2388,-2389,-2390,-
```

```
##           Adult.Mortality infant.deaths under.five.deaths   Polio
## Adult.Mortality           1.0000           0.1882           0.2120 -0.3724
## infant.deaths           0.1882           1.0000           0.9942 -0.1297
## under.five.deaths       0.2120           0.9942           1.0000 -0.1475
## Polio                   -0.3724          -0.1297          -0.1475  1.0000
```

Performing all possible, forward, and backward regressions

```
ourmodel <- lm(Life.expectancy~Adult.Mortality+infant.deaths+under.five.deaths+Polio+Diphtheria, data =
anova(ourmodel))
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: Life.expectancy
```

```
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## Adult.Mortality    1 7291.0   7291.0 359.0728 < 2.2e-16 ***
## infant.deaths      1  111.6    111.6   5.4945  0.020187 *
## under.five.deaths  1  188.5    188.5   9.2836  0.002666 **
## Polio              1  613.1    613.1  30.1957 1.344e-07 ***
## Diphtheria         1  212.8    212.8  10.4802  0.001441 **
## Residuals         177 3594.0     20.3
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(ourmodel)
```

```
##
```

```
## Call:
```

```
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     under.five.deaths + Polio + Diphtheria, data = lifeExpectancyDataUpdated2)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -18.3307 -2.5620   0.5472   3.1795  10.7421
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  70.656344   1.775388  39.798 < 2e-16 ***
## Adult.Mortality -0.052790   0.003815 -13.837 < 2e-16 ***
## infant.deaths   0.076029   0.038280   1.986  0.04856 *
## under.five.deaths -0.065702   0.029900  -2.197  0.02929 *
## Polio          0.044868   0.018868   2.378  0.01847 *
## Diphtheria      0.065743   0.020308   3.237  0.00144 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 4.506 on 177 degrees of freedom
```

```
## Multiple R-squared:  0.7008, Adjusted R-squared:  0.6923
```

```
## F-statistic: 82.91 on 5 and 177 DF, p-value: < 2.2e-16
```

```
k <- ols_step_all_possible(ourmodel)
```

```
k
```

```
##      Index N
```

```
## 1         1 1
```

```
Predictors
```

```
Adult.Mortality
```

## 4	2 1	Polio
## 5	3 1	Diphtheria
## 3	4 1	under.five.deaths
## 2	5 1	infant.deaths
## 9	6 2	Adult.Mortality Diphtheria
## 8	7 2	Adult.Mortality Polio
## 7	8 2	Adult.Mortality under.five.deaths
## 6	9 2	Adult.Mortality infant.deaths
## 15	10 2	Polio Diphtheria
## 13	11 2	under.five.deaths Polio
## 11	12 2	infant.deaths Polio
## 14	13 2	under.five.deaths Diphtheria
## 12	14 2	infant.deaths Diphtheria
## 10	15 2	infant.deaths under.five.deaths
## 21	16 3	Adult.Mortality Polio Diphtheria
## 20	17 3	Adult.Mortality under.five.deaths Diphtheria
## 18	18 3	Adult.Mortality infant.deaths Diphtheria
## 19	19 3	Adult.Mortality under.five.deaths Polio
## 17	20 3	Adult.Mortality infant.deaths Polio
## 16	21 3	Adult.Mortality infant.deaths under.five.deaths
## 25	22 3	under.five.deaths Polio Diphtheria
## 22	23 3	infant.deaths under.five.deaths Polio
## 24	24 3	infant.deaths Polio Diphtheria
## 23	25 3	infant.deaths under.five.deaths Diphtheria
## 29	26 4	Adult.Mortality under.five.deaths Polio Diphtheria
## 28	27 4	Adult.Mortality infant.deaths Polio Diphtheria
## 27	28 4	Adult.Mortality infant.deaths under.five.deaths Diphtheria
## 26	29 4	Adult.Mortality infant.deaths under.five.deaths Polio
## 30	30 4	infant.deaths under.five.deaths Polio Diphtheria
## 31	31 5	Adult.Mortality infant.deaths under.five.deaths Polio Diphtheria
##	R-Square Adj. R-Square Mallow's Cp	
## 1	0.60702715 0.60485603 53.453978	
## 4	0.26584521 0.26178911 255.272250	
## 5	0.25030006 0.24615807 264.467625	
## 3	0.07394106 0.06882471 368.788705	
## 2	0.05822064 0.05301744 378.087750	
## 9	0.67617183 0.67257374 14.553047	
## 8	0.66604570 0.66233510 20.542926	
## 7	0.61895787 0.61472407 48.396640	
## 6	0.61631578 0.61205263 49.959505	
## 15	0.31090042 0.30324376 230.620886	
## 13	0.30507170 0.29735027 234.068730	
## 11	0.29678159 0.28896805 238.972552	
## 14	0.29228605 0.28442256 241.631787	
## 12	0.28398757 0.27603188 246.540558	
## 10	0.14686529 0.13738602 327.652061	
## 21	0.68671445 0.68146386 10.316804	
## 20	0.68422466 0.67893234 11.789582	
## 18	0.68262532 0.67730619 12.735638	
## 19	0.67435180 0.66889401 17.629646	
## 17	0.67255075 0.66706278 18.695013	
## 16	0.63201008 0.62584265 42.675907	
## 25	0.34579411 0.33482977 211.980329	
## 22	0.34360456 0.33260352 213.275508	

```
## 24 0.33874374    0.32766123  216.150815
## 23 0.32818330    0.31692380  222.397600
## 29 0.69410563    0.68723160    7.944727
## 28 0.69261147    0.68570387    8.828558
## 27 0.69121458    0.68427558    9.654857
## 26 0.68305720    0.67593489   14.480168
## 30 0.37707642    0.36307813  195.476008
## 31 0.70077435    0.69232164    6.000000
```

forward

```
f <- ols_step_forward_p(ourmodel, penter= 0.3, details = TRUE)
```

```
## Forward Selection Method
```

```
## -----
```

```
##
```

```
## Candidate Terms:
```

```
##
```

```
## 1. Adult.Mortality
```

```
## 2. infant.deaths
```

```
## 3. under.five.deaths
```

```
## 4. Polio
```

```
## 5. Diphtheria
```

```
##
```

```
## We are selecting variables based on p value...
```

```
##
```

```
##
```

```
## Forward Selection: Step 1
```

```
##
```

```
## - Adult.Mortality
```

```
##
```

```
##                               Model Summary
```

```
## -----
```

```
## R                0.779          RMSE                5.107
```

```
## R-Squared        0.607          Coef. Var            7.130
```

```
## Adj. R-Squared   0.605          MSE                26.077
```

```
## Pred R-Squared   0.596          MAE                3.581
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
##                               ANOVA
```

```
## -----
```

```
##                Sum of
##                Squares      DF      Mean Square      F      Sig.
```

```
## -----
```

```
## Regression      7291.014        1      7291.014    279.592    0.0000
```

```
## Residual        4720.004       181        26.077
```

```
## Total          12011.017       182
```

```
## -----
```

```
##
```

```
##                               Parameter Estimates
```

```
## -----
```

```
##                model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
```

```

## -----
##      (Intercept)      81.534      0.703      115.974      0.000      80.147      82.921
## Adult.Mortality     -0.065      0.004      -0.779      -16.721      0.000      -0.073      -0.057
## -----
##
##
##
## Forward Selection: Step 2
##
## - Diphtheria
##
##                               Model Summary
## -----
## R                          0.822      RMSE                          4.648
## R-Squared                   0.676      Coef. Var                     6.491
## Adj. R-Squared              0.673      MSE                          21.608
## Pred R-Squared              0.656      MAE                          3.473
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      8121.512      2      4060.756      187.925      0.0000
## Residual        3889.506      180      21.608
## Total           12011.017      182
## -----
##
##                               Parameter Estimates
## -----
##                               model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upper
## -----
##      (Intercept)      71.774      1.699      42.236      0.000      68.421      75.128
## Adult.Mortality     -0.057      0.004      -0.689      -15.386      0.000      -0.065      -0.050
##      Diphtheria      0.102      0.016      0.278      6.200      0.000      0.069      0.134
## -----
##
##
##
## Forward Selection: Step 3
##
## - Polio
##
##                               Model Summary
## -----
## R                          0.829      RMSE                          4.585
## R-Squared                   0.687      Coef. Var                     6.402
## Adj. R-Squared              0.681      MSE                          21.022
## Pred R-Squared              0.664      MAE                          3.471
## -----

```

```
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression    8248.139        3      2749.380    130.788    0.0000
## Residual      3762.878       179        21.022
## Total        12011.017       182
```

```
## -----
```

```
##
```

```
## Parameter Estimates
```

```
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
##      (Intercept)  70.175        1.798          -0.665      39.020    0.000    66.626    73.723
## Adult.Mortality  -0.055        0.004          -0.665     -14.654    0.000    -0.063    -0.048
##      Diphtheria   0.071        0.021          0.193       3.436    0.001     0.030     0.111
##      Polio        0.047        0.019          0.141       2.454    0.015     0.009     0.085
```

```
## -----
```

```
##
```

```
##
```

```
##
```

```
## Forward Selection: Step 4
```

```
##
```

```
## - under.five.deaths
```

```
##
```

```
## Model Summary
```

```
## -----
## R              0.833      RMSE              4.543
## R-Squared      0.694      Coef. Var        6.344
## Adj. R-Squared 0.687      MSE              20.641
## Pred R-Squared 0.662      MAE              3.412
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression    8336.915        4      2084.229    100.975    0.0000
## Residual      3674.103       178        20.641
## Total        12011.017       182
```

```
## -----
```

```
##
```

```
## Parameter Estimates
```

```
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
```



```

## -----
##      (Intercept)      70.437      1.787      39.426      0.000      66.912      73.963
##      Adult.Mortality -0.054      0.004      -0.649      -14.237      0.000      -0.062      -0.047
##      Diphtheria      0.069      0.020      0.189      3.390      0.001      0.029      0.109
##      Polio      0.046      0.019      0.136      2.398      0.018      0.008      0.083
##      under.five.deaths -0.007      0.003      -0.088      -2.074      0.040      -0.013      0.000
## -----
##
##
##
## Forward Selection: Step 5
##
## - infant.deaths
##
##
##      Model Summary
## -----
##      R      0.837      RMSE      4.506
##      R-Squared      0.701      Coef. Var      6.292
##      Adj. R-Squared      0.692      MSE      20.305
##      Pred R-Squared      0.374      MAE      3.384
## -----
##      RMSE: Root Mean Square Error
##      MSE: Mean Square Error
##      MAE: Mean Absolute Error
##
##
##      ANOVA
## -----
##      Sum of
##      Squares      DF      Mean Square      F      Sig.
## -----
##      Regression      8417.013      5      1683.403      82.905      0.0000
##      Residual      3594.005      177      20.305
##      Total      12011.017      182
## -----
##
##
##      Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upper
## -----
##      (Intercept)      70.656      1.775      39.798      0.000      67.153      74.160
##      Adult.Mortality      -0.053      0.004      -0.634      -13.837      0.000      -0.060      -0.045
##      Diphtheria      0.066      0.020      0.179      3.237      0.001      0.026      0.106
##      Polio      0.045      0.019      0.134      2.378      0.018      0.008      0.082
##      under.five.deaths      -0.066      0.030      -0.872      -2.197      0.029      -0.125      -0.007
##      infant.deaths      0.076      0.038      0.783      1.986      0.049      0.000      0.152
## -----
##
##
##
## Variables Entered:
##
## + Adult.Mortality
## + Diphtheria

```

```
## + Polio
## + under.five.deaths
## + infant.deaths
##
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                0.837          RMSE                4.506
## R-Squared        0.701          Coef. Var            6.292
## Adj. R-Squared   0.692          MSE                20.305
## Pred R-Squared   0.374          MAE                3.384
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                Sum of          DF      Mean Square      F      Sig.
##                Squares
## -----
## Regression      8417.013          5      1683.403      82.905      0.0000
## Residual        3594.005         177          20.305
## Total          12011.017         182
## -----
##
##                               Parameter Estimates
## -----
##                model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
##      (Intercept)    70.656        1.775          39.798      0.000      67.153      74.160
##      Adult.Mortality -0.053        0.004          -13.837      0.000      -0.060      -0.045
##      Diphtheria      0.066        0.020           3.237      0.001       0.026       0.106
##      Polio           0.045        0.019           2.378      0.018       0.008       0.082
##      under.five.deaths -0.066        0.030          -2.197      0.029      -0.125      -0.007
##      infant.deaths   0.076        0.038           1.986      0.049       0.000       0.152
## -----
```

backward

```
b <- ols_step_backward_p(ourmodel, penter = 0.3, details = TRUE)
```

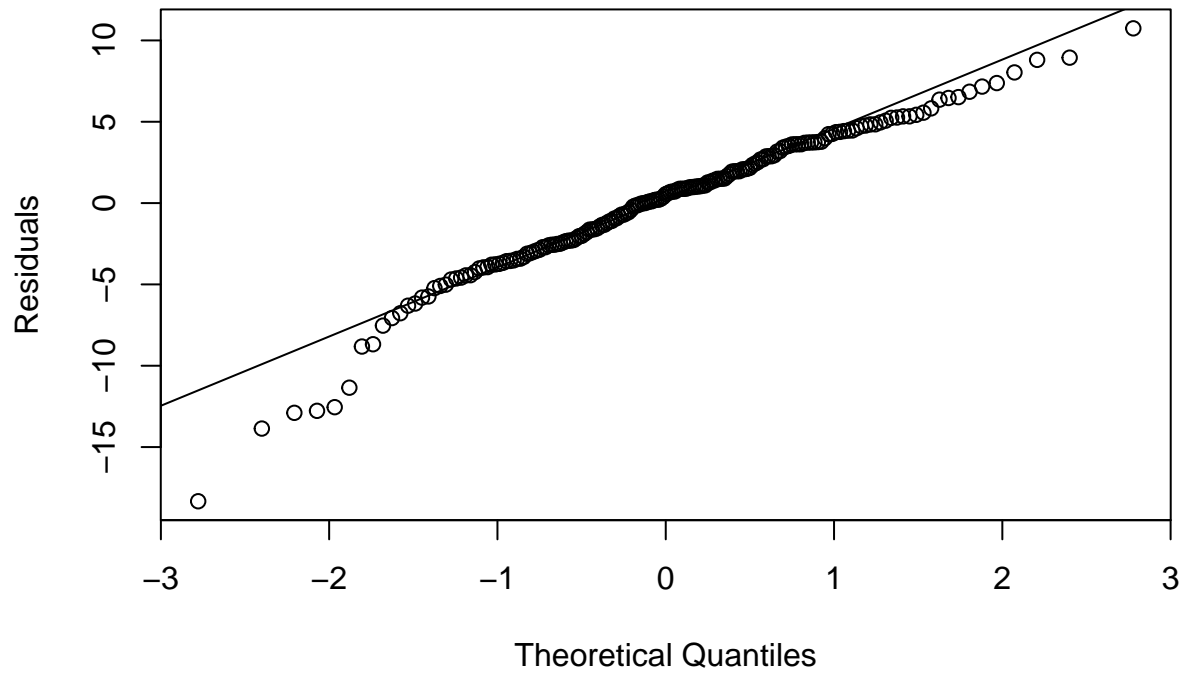
```
## Backward Elimination Method
## -----
##
## Candidate Terms:
##
## 1 . Adult.Mortality
## 2 . infant.deaths
## 3 . under.five.deaths
## 4 . Polio
## 5 . Diphtheria
```

```
##
## We are eliminating variables based on p value...
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
##
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                               0.837          RMSE                4.506
## R-Squared                       0.701          Coef. Var          6.292
## Adj. R-Squared                   0.692          MSE                20.305
## Pred R-Squared                   0.374          MAE                3.384
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      8417.013           5      1683.403      82.905      0.0000
## Residual        3594.005          177           20.305
## Total          12011.017          182
## -----
##
##                               Parameter Estimates
## -----
##                               model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upper
## -----
##      (Intercept)      70.656           1.775           39.798      0.000      67.153      74.160
##      Adult.Mortality  -0.053           0.004           -0.634     -13.837      0.000      -0.060      -0.045
##      infant.deaths     0.076           0.038           0.783       1.986      0.049       0.000       0.152
##      under.five.deaths -0.066           0.030           -0.872     -2.197      0.029      -0.125      -0.007
##      Polio             0.045           0.019           0.134       2.378      0.018       0.008       0.082
##      Diphtheria        0.066           0.020           0.179       3.237      0.001       0.026       0.106
## -----
```

Check model assumptions

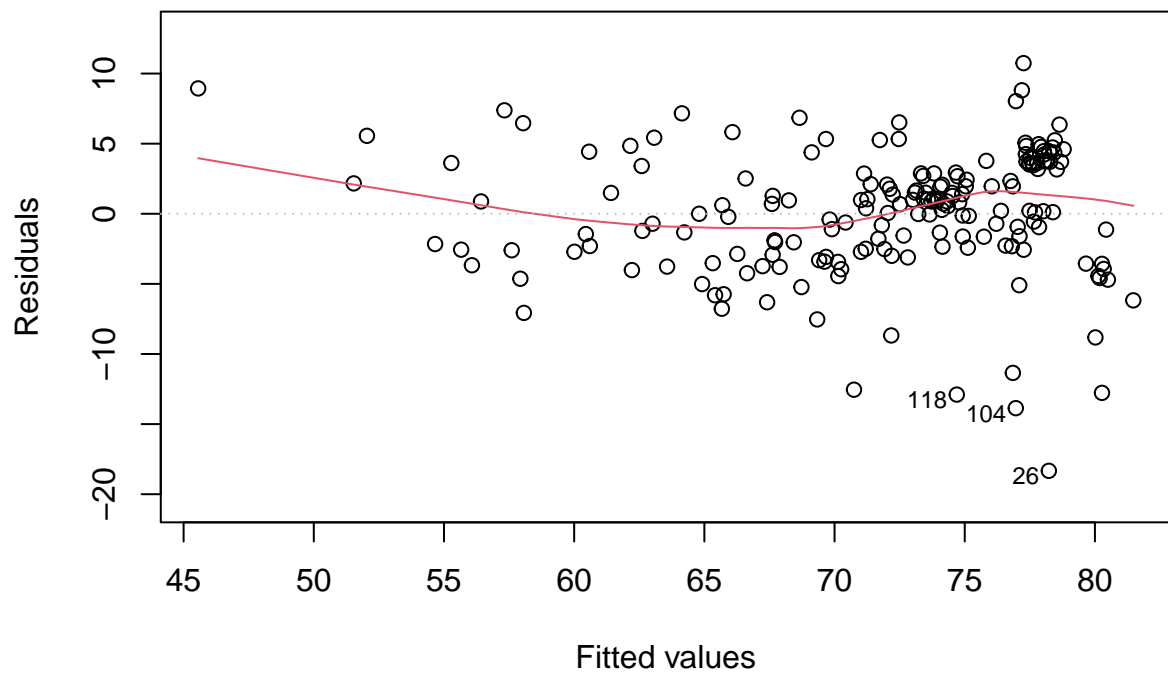
```
qqnorm(residuals(ourmodel),ylab="Residuals",main="Q-Q plot")
qqline(residuals(ourmodel))
```

Q-Q plot

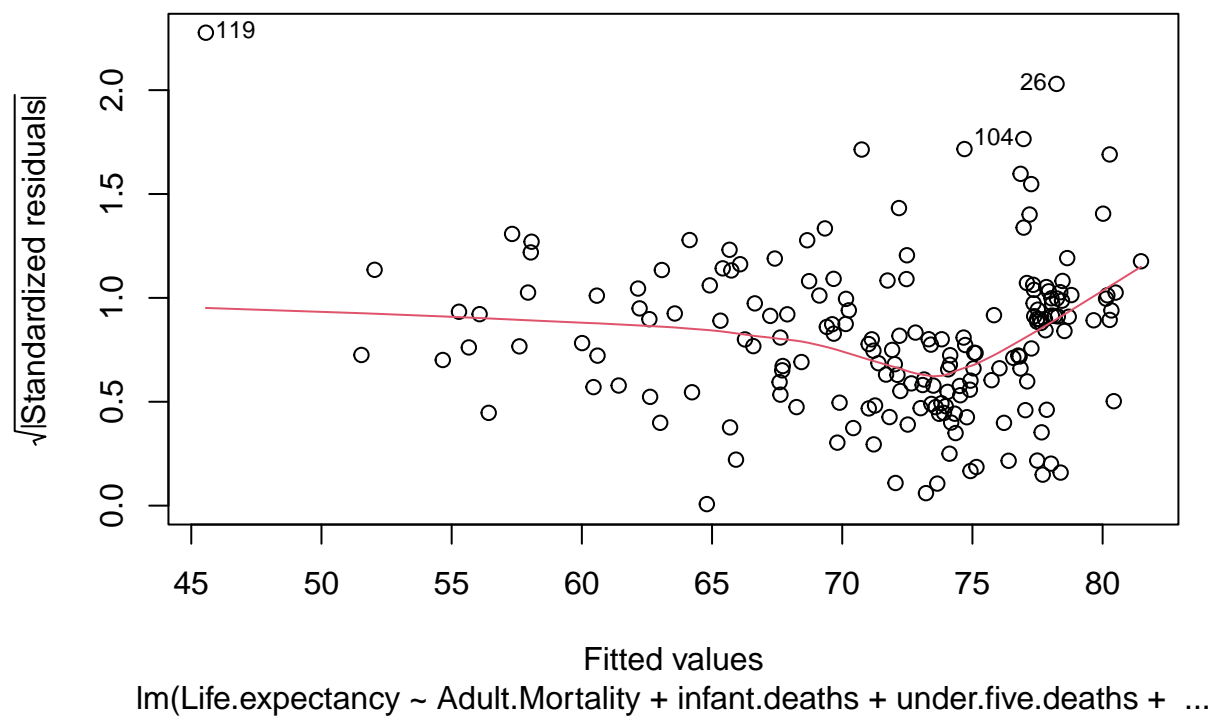
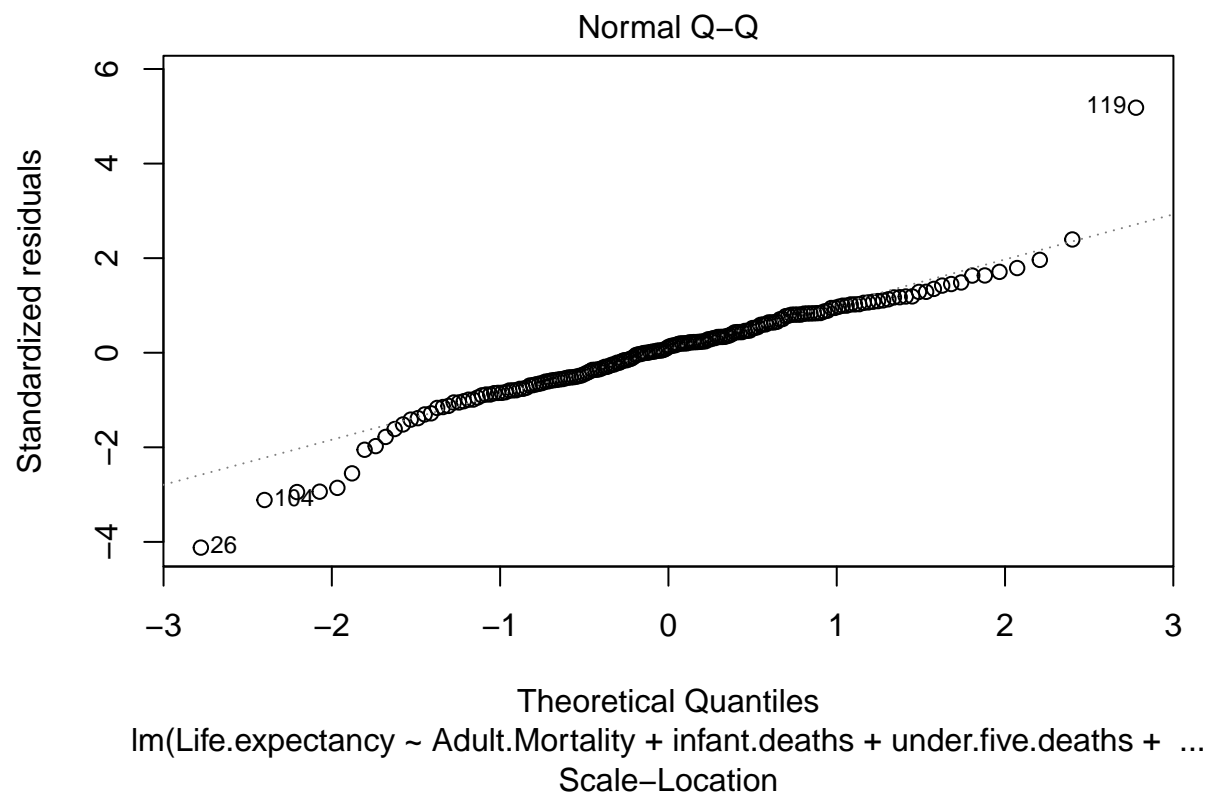


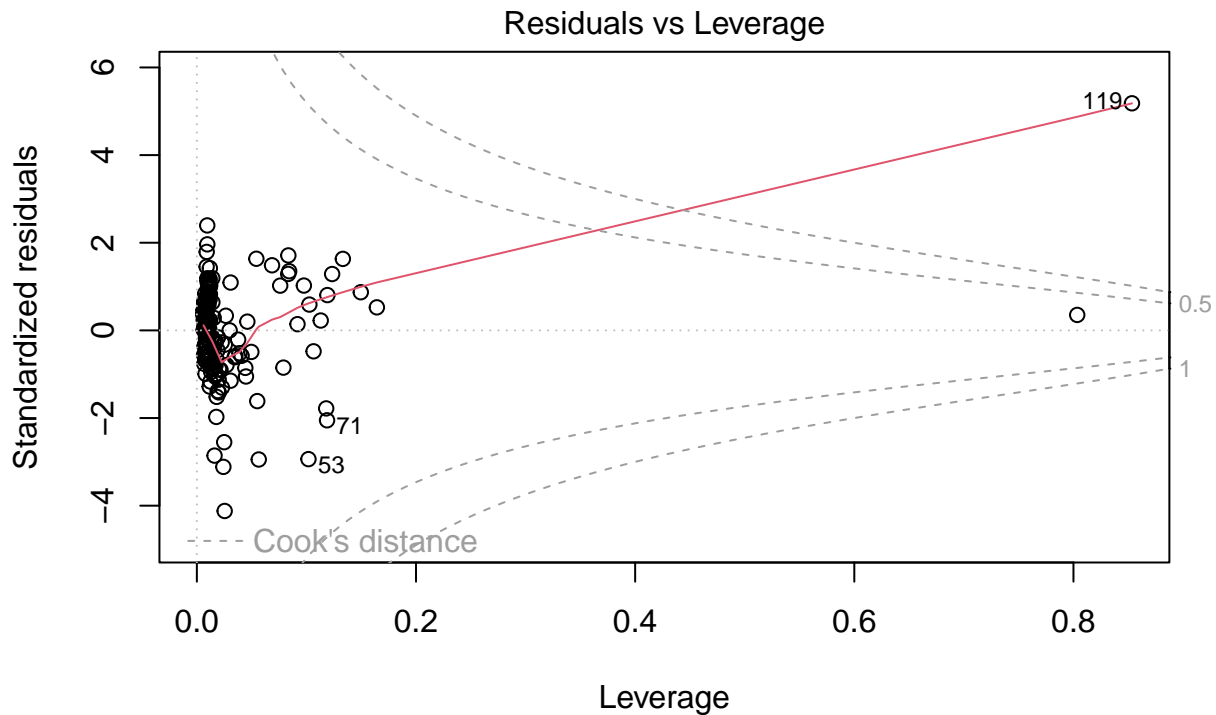
```
plot(ourmodel)
```

Residuals vs Fitted



lm(Life.expectancy ~ Adult.Mortality + infant.deaths + under.five.deaths + ...)





lm(Life.expectancy ~ Adult.Mortality + infant.deaths + under.five.deaths + ...

Check for multicollinearity

```
vif(ourmodel)
```

```
##      Adult.Mortality      infant.deaths under.five.deaths      Polio
##           1.241665           91.979835           93.084879           1.875786
##      Diphtheria
##           1.815719
```

Removed infant.deaths to remove multicollinearity? And to obtain final model, also this had higher r squared adj value and minimized Cp (error)

```
finalmod <- lm(Life.expectancy~Adult.Mortality+under.five.deaths+Polio+Diphtheria, data = lifeExpectancyDataUpdated)
summary(finalmod)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + under.five.deaths +
##      Polio + Diphtheria, data = lifeExpectancyDataUpdated)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.1710  -2.6531   0.5757   3.2032  10.6572
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   70.437108   1.786550   39.426 < 2e-16 ***
## Adult.Mortality -0.054026   0.003795  -14.237 < 2e-16 ***
## under.five.deaths -0.006654   0.003209   -2.074  0.03953 *
## Polio          0.045606   0.019020    2.398  0.01753 *
## Diphtheria     0.069168   0.020401    3.390  0.00086 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.543 on 178 degrees of freedom
## Multiple R-squared:  0.6941, Adjusted R-squared:  0.6872
## F-statistic: 101 on 4 and 178 DF,  p-value: < 2.2e-16
```

```
anova(finalmod)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: Life.expectancy
```

```
##          Df Sum Sq Mean Sq  F value    Pr(>F)
## Adult.Mortality    1 7291.0   7291.0 353.2292 < 2.2e-16 ***
## under.five.deaths    1  143.3    143.3   6.9425 0.0091594 **
## Polio                1  665.3    665.3  32.2337 5.467e-08 ***
## Diphtheria           1  237.3    237.3  11.4948 0.0008597 ***
## Residuals          178 3674.1     20.6
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
LASSO method
```

```
x<-(apply(lifeExpectancyDataUpdated2[,-1],2,as.numeric))
```

```
y<-(apply(lifeExpectancyDataUpdated2[,1],2,as.numeric))
```

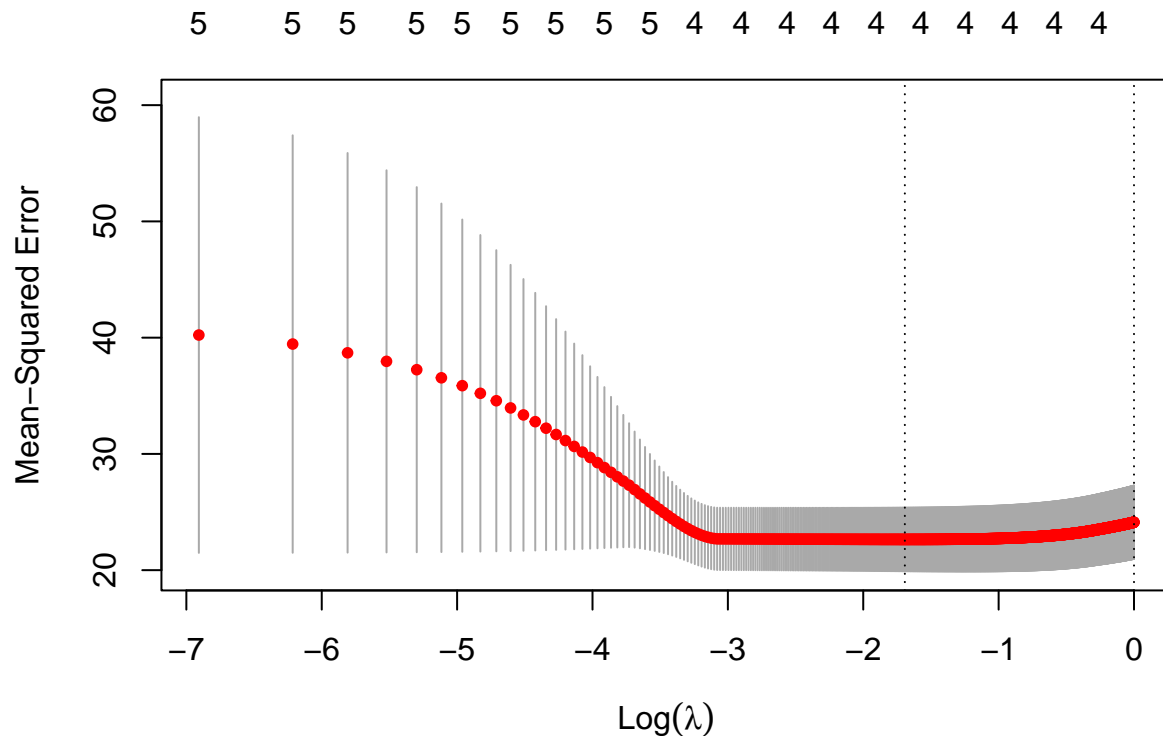
```
set.seed(43215)
```

```
lasso_cv = cv.glmnet(x, y,family=c("gaussian"),standardize=TRUE,nfolds=10,lambda=seq(0,1,0.001,alpha =
```

```
## Warning: In seq.default(0, 1, 0.001, alpha = 1) :
```

```
## extra argument 'alpha' will be disregarded
```

```
plot(lasso_cv)
```



```

(best_lambda = lasso_cv$lambda.min)

## [1] 0.184

best_lasso = glmnet(x,y, standardize=TRUE,alpha = 1, lambda = best_lambda)
coef(best_lasso)

## 6 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept)    70.804353219
## Adult.Mortality -0.052988830
## infant.deaths      .
## under.five.deaths -0.005345686
## Polio             0.042803041
## Diphtheria        0.065223126

best_lasso$dev.ratio

## [1] 0.6929939

```