R Notebook

Author: Andres Felipe Alba Hernández
Department: Electrical Engineering
Date: October 1st, 2018
Course: ISYE670 Data Science for Engineers
Professor: Dr. Christine Nguyen
Northern Illinois University

```
rm(list = ls())
library(dplyr,lib.loc="/home/leasanspy/DataScience_NIU/Rpackages") #Loading the library
```

```
##
## 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
```

1) (2 points) Import the data with the headers. How is the data organized? (i.e. what are the column names? What does each row of the data represent?)

From the information below I infered that the data is organized in a way that every observation collected an index which reference the first letter of the state, the state, and the total income by year.

```
h_data <- read.csv("historical_data.csv")
names(h_data)
```

```
##  [1] "Index" "State" "Y2002" "Y2003" "Y2004" "Y2005" "Y2006" "Y2007"
##  [9] "Y2008" "Y2009" "Y2010" "Y2011" "Y2012" "Y2013" "Y2014" "Y2015"
```

```
str(h_data)
```

```
## 'data.frame':    51 obs. of  16 variables:
##  $ Index: Factor w/ 19 levels "A","C","D","F",..: 1 1 1 1 2 2 2 3 3 4 ...
##  $ State: Factor w/ 51 levels "Alabama","Alaska",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Y2002: int  1296530 1170302 1742027 1485531 1685349 1343824 1610512 1330403 1111437 1964626 ...
##  $ Y2003: int  1317711 1960378 1968140 1994927 1675807 1878473 1232844 1268673 1993741 1468852 ...
##  $ Y2004: int  1118631 1818085 1377583 1119299 1889570 1886149 1181949 1706751 1374643 1419738 ...
##  $ Y2005: int  1492583 1447852 1782199 1947979 1480280 1236697 1518933 1403759 1827949 1362787 ...
##  $ Y2006: int  1107408 1861639 1102568 1669191 1735069 1871471 1841266 1441351 1803852 1339608 ...
##  $ Y2007: int  1440134 1465841 1109382 1801213 1812546 1814218 1976976 1300836 1595981 1278550 ...
##  $ Y2008: int  1945229 1551826 1752886 1188104 1487315 1875146 1764457 1762096 1193245 1756185 ...
##  $ Y2009: int  1944173 1436541 1554330 1628980 1663809 1752387 1972730 1553585 1739748 1818438 ...
##  $ Y2010: int  1237582 1629616 1300521 1669295 1624509 1913275 1968730 1370984 1707823 1198403 ...
##  $ Y2011: int  1440756 1230866 1130709 1928238 1639670 1665877 1945524 1318669 1353449 1497051 ...
##  $ Y2012: int  1186741 1512804 1907284 1216675 1921845 1491604 1228529 1984027 1979708 1131928 ...
##  $ Y2013: int  1852841 1985302 1363279 1591896 1156536 1178355 1582249 1671279 1912654 1107448 ...
##  $ Y2014: int  1558906 1580394 1525866 1360959 1388461 1383978 1503156 1803169 1782169 1407784 ...
##  $ Y2015: int  1916661 1979143 1647724 1329341 1644607 1330736 1718072 1627508 1410183 1170389 ...
```

```
summary(h_data)
```

```
##      Index           State        Y2002             Y2003
##  M      : 8   Alabama   : 1   Min.   :1111437   Min.   :1110625
##  N      : 8   Alaska    : 1   1st Qu.:1374180   1st Qu.:1292390
##  A      : 4   Arizona   : 1   Median :1584734   Median :1485909
##  I      : 4   Arkansas  : 1   Mean   :1566034   Mean   :1509193
##  W      : 4   California: 1   3rd Qu.:1776054   3rd Qu.:1686698
##  C      : 3   Colorado  : 1   Max.   :1983285   Max.   :1994927
##  (Other):20   (Other)   :45
##      Y2004             Y2005             Y2006             Y2007
##  Min.   :1118631   Min.   :1122030   Min.   :1102568   Min.   :1109382
##  1st Qu.:1268292   1st Qu.:1267340   1st Qu.:1337236   1st Qu.:1322419
##  Median :1522230   Median :1480280   Median :1531641   Median :1563062
##  Mean   :1540554   Mean   :1522064   Mean   :1530969   Mean   :1553219
##  3rd Qu.:1808109   3rd Qu.:1778170   3rd Qu.:1732259   3rd Qu.:1780589
##  Max.   :1979395   Max.   :1990062   Max.   :1985692   Max.   :1983568
##
##      Y2008             Y2009             Y2010             Y2011
##  Min.   :1112765   Min.   :1116168   Min.   :1103794   Min.   :1116203
##  1st Qu.:1254244   1st Qu.:1553958   1st Qu.:1328439   1st Qu.:1371730
##  Median :1545621   Median :1658551   Median :1498662   Median :1575533
##  Mean   :1538398   Mean   :1658519   Mean   :1504108   Mean   :1574968
##  3rd Qu.:1779538   3rd Qu.:1857746   3rd Qu.:1639186   3rd Qu.:1807766
##  Max.   :1990431   Max.   :1993136   Max.   :1999102   Max.   :1992996
##
##      Y2012             Y2013             Y2014             Y2015
##  Min.   :1108281   Min.   :1100990   Min.   :1110394   Min.   :1110655
##  1st Qu.:1360654   1st Qu.:1285738   1st Qu.:1385703   1st Qu.:1372523
##  Median :1643855   Median :1531212   Median :1580394   Median :1627508
##  Mean   :1591135   Mean   :1530078   Mean   :1583360   Mean   :1588298
##  3rd Qu.:1866322   3rd Qu.:1725377   3rd Qu.:1791594   3rd Qu.:1848316
##  Max.   :1988270   Max.   :1994022   Max.   :1990412   Max.   :1996005
##
```

2) Lookup the sample_n and sample_frac functions.

a) (2 points) Use the appropriate function to randomly choose 15 rows.
b) (2 points) Use the appropriate function to randomly choose 40%

```
Sn <- sample_n(h_data,15) #a
Sf <- sample_frac(h_data,0.4) #b
str(Sn) #a
```

```
## 'data.frame':    15 obs. of  16 variables:
##  $ Index: Factor w/ 19 levels "A","C","D","F",..: 1 10 11 6 10 12 19 11 10 8 ...
##  $ State: Factor w/ 51 levels "Alabama","Alaska",..: 2 25 28 12 27 36 50 32 26 17 ...
##  $ Y2002: int  1170302 1983285 1885081 1461570 1877154 1802132 1788920 1819239 1221316 1509054 ...
##  $ Y2003: int  1960378 1292558 1309769 1200280 1540099 1648498 1518578 1226057 1858368 1290700 ...
##  $ Y2004: int  1818085 1631325 1425527 1213993 1332722 1441386 1289663 1935991 1773451 1522230 ...
##  $ Y2005: int  1447852 1943311 1240465 1245931 1273327 1670280 1436888 1124400 1573967 1532094 ...
##  $ Y2006: int  1861639 1354579 1500594 1459383 1625721 1534888 1251678 1723493 1374863 1104256 ...
##  $ Y2007: int  1465841 1731643 1278272 1430465 1983568 1314824 1721874 1475985 1486197 1863278 ...
##  $ Y2008: int  1551826 1428291 1140598 1919423 1251742 1516621 1980167 1237704 1735099 1949478 ...
##  $ Y2009: int  1436541 1568049 1270585 1928416 1592690 1511460 1901394 1820856 1800620 1561528 ...
##  $ Y2010: int  1629616 1383227 1128711 1330509 1350619 1585465 1648755 1801430 1164202 1550433 ...
##  $ Y2011: int  1230866 1629132 1187207 1902816 1520064 1887714 1940943 1653384 1425363 1465812 ...
##  $ Y2012: int  1512804 1988270 1569665 1695126 1185225 1227303 1729177 1475715 1800052 1882929 ...
```

```
##  $ Y2013: int  1985302 1907777 1690920 1517184 1465705 1840898 1510119 1623388 1698105 1410249 ...
##  $ Y2014: int  1580394 1649668 1459243 1948108 1110394 1880804 1701650 1533494 1767835 1930090 ...
##  $ Y2015: int  1979143 1991232 1802211 1150882 1125903 1573117 1846238 1868612 1996005 1385528 ...
```

```
str(Sf) #b
```

```
## 'data.frame':    20 obs. of  16 variables:
##  $ Index: Factor w/ 19 levels "A","C","D","F",..: 7 12 7 2 8 13 10 1 19 10 ...
##  $ State: Factor w/ 51 levels "Alabama","Alaska",..: 16 37 13 7 17 39 25 3 48 27 ...
##  $ Y2002: int  1499269 1173918 1353210 1610512 1509054 1320191 1983285 1742027 1977749 1877154 ...
##  $ Y2003: int  1444576 1334639 1438538 1232844 1290700 1446723 1292558 1968140 1687136 1540099 ...
##  $ Y2004: int  1576367 1663622 1739154 1181949 1522230 1218591 1631325 1377583 1199490 1332722 ...
##  $ Y2005: int  1388924 1798714 1541015 1518933 1532094 1122030 1943311 1782199 1163092 1273327 ...
##  $ Y2006: int  1554813 1312574 1122387 1841266 1104256 1971479 1354579 1102568 1334864 1625721 ...
##  $ Y2007: int  1452911 1708245 1772050 1976976 1863278 1563062 1731643 1109382 1621989 1983568 ...
##  $ Y2008: int  1317983 1256746 1335481 1764457 1949478 1274168 1428291 1752886 1545621 1251742 ...
##  $ Y2009: int  1150783 1853142 1748608 1972730 1561528 1571032 1568049 1554330 1555554 1592690 ...
##  $ Y2010: int  1751389 1673831 1436809 1968730 1550433 1433835 1383227 1300521 1179331 1350619 ...
##  $ Y2011: int  1992996 1822933 1456340 1945524 1465812 1483292 1629132 1130709 1150089 1520064 ...
##  $ Y2012: int  1501879 1674707 1643855 1228529 1882929 1290329 1988270 1907284 1775787 1185225 ...
##  $ Y2013: int  1173694 1900523 1312561 1582249 1410249 1475344 1907777 1363279 1273834 1465705 ...
##  $ Y2014: int  1431705 1956742 1713718 1503156 1930090 1931500 1649668 1525866 1387428 1110394 ...
##  $ Y2015: int  1641866 1307678 1757171 1718072 1385528 1668232 1991232 1647724 1377341 1125903 ...
```

3) Output certain columns.

a) (2 points) Output only the values for the year 2005 and States
b) (2 points) Output only the values for all the years. (Hint, use the -)
c) (2 points) Output only the values for columns that begin with a Y

```
#Answer a)
select(h_data,State,Y2005)
```

```
##                   State   Y2005
## 1               Alabama 1492583
## 2                Alaska 1447852
## 3               Arizona 1782199
## 4              Arkansas 1947979
## 5            California 1480280
## 6              Colorado 1236697
## 7           Connecticut 1518933
## 8              Delaware 1403759
## 9  District of Columbia 1827949
## 10              Florida 1362787
## 11              Georgia 1779091
## 12               Hawaii 1245931
## 13                Idaho 1541015
## 14             Illinois 1261353
## 15              Indiana 1204117
## 16                 Iowa 1388924
## 17               Kansas 1532094
## 18             Kentucky 1250524
## 19            Louisiana 1751920
## 20                Maine 1912040
## 21             Maryland 1397738
## 22        Massachusetts 1777250
## 23             Michigan 1340716
```

```
## 24            Minnesota 1561839
## 25          Mississippi 1943311
## 26             Missouri 1573967
## 27              Montana 1273327
## 28             Nebraska 1240465
## 29               Nevada 1758830
## 30        New Hampshire 1990062
## 31           New Jersey 1181452
## 32           New Mexico 1124400
## 33             New York 1446810
## 34       North Carolina 1532347
## 35         North Dakota 1443172
## 36                 Ohio 1670280
## 37             Oklahoma 1798714
## 38               Oregon 1133510
## 39         Pennsylvania 1122030
## 40         Rhode Island 1961923
## 41       South Carolina 1458191
## 42         South Dakota 1417141
## 43            Tennessee 1157059
## 44                Texas 1907326
## 45                 Utah 1241662
## 46              Vermont 1579265
## 47             Virginia 1853855
## 48           Washington 1163092
## 49        West Virginia 1888948
## 50            Wisconsin 1436888
## 51              Wyoming 1881688
```

```
#Answer b
select(h_data,-Index,-State)
```

```
##      Y2002   Y2003   Y2004   Y2005   Y2006   Y2007   Y2008   Y2009   Y2010
## 1  1296530 1317711 1118631 1492583 1107408 1440134 1945229 1944173 1237582
## 2  1170302 1960378 1818085 1447852 1861639 1465841 1551826 1436541 1629616
## 3  1742027 1968140 1377583 1782199 1102568 1109382 1752886 1554330 1300521
## 4  1485531 1994927 1119299 1947979 1669191 1801213 1188104 1628980 1669295
## 5  1685349 1675807 1889570 1480280 1735069 1812546 1487315 1663809 1624509
## 6  1343824 1878473 1886149 1236697 1871471 1814218 1875146 1752387 1913275
## 7  1610512 1232844 1181949 1518933 1841266 1976976 1764457 1972730 1968730
## 8  1330403 1268673 1706751 1403759 1441351 1300836 1762096 1553585 1370984
## 9  1111437 1993741 1374643 1827949 1803852 1595981 1193245 1739748 1707823
## 10 1964626 1468852 1419738 1362787 1339608 1278550 1756185 1818438 1198403
## 11 1929009 1541565 1810773 1779091 1326846 1223770 1773090 1630325 1145473
## 12 1461570 1200280 1213993 1245931 1459383 1430465 1919423 1928416 1330509
## 13 1353210 1438538 1739154 1541015 1122387 1772050 1335481 1748608 1436809
## 14 1508356 1527440 1493029 1261353 1540274 1747614 1871645 1658551 1422021
## 15 1776918 1734104 1269927 1204117 1848073 1129546 1139551 1883976 1999102
## 16 1499269 1444576 1576367 1388924 1554813 1452911 1317983 1150783 1751389
## 17 1509054 1290700 1522230 1532094 1104256 1863278 1949478 1561528 1550433
## 18 1813878 1448846 1800760 1250524 1137913 1911227 1301848 1956681 1350895
## 19 1584734 1110625 1868456 1751920 1233709 1920301 1185085 1124853 1498662
## 20 1582720 1678622 1208496 1912040 1438549 1330014 1295877 1969163 1627262
## 21 1579713 1404700 1849798 1397738 1310270 1789128 1112765 1967225 1486246
## 22 1647582 1686259 1620601 1777250 1531641 1380529 1978904 1567651 1761048
```

```
## 23 1295635 1149931 1601027 1340716 1729449 1567494 1990431 1575185 1267626
## 24 1729921 1675204 1903907 1561839 1985692 1148621 1328133 1890633 1995304
## 25 1983285 1292558 1631325 1943311 1354579 1731643 1428291 1568049 1383227
## 26 1221316 1858368 1773451 1573967 1374863 1486197 1735099 1800620 1164202
## 27 1877154 1540099 1332722 1273327 1625721 1983568 1251742 1592690 1350619
## 28 1885081 1309769 1425527 1240465 1500594 1278272 1140598 1270585 1128711
## 29 1426117 1114500 1119707 1758830 1694526 1765826 1903270 1231480 1526066
## 30 1419776 1854370 1195119 1990062 1645430 1286967 1762936 1763211 1265642
## 31 1605532 1141514 1613550 1181452 1541327 1156804 1568034 1357418 1443718
## 32 1819239 1226057 1935991 1124400 1723493 1475985 1237704 1820856 1801430
## 33 1395149 1611371 1170675 1446810 1426941 1463171 1732098 1426216 1604531
## 34 1616742 1292223 1482792 1532347 1158716 1827420 1267737 1116168 1791535
## 35 1618807 1510193 1876940 1443172 1425030 1868788 1720352 1671468 1534571
## 36 1802132 1648498 1441386 1670280 1534888 1314824 1516621 1511460 1585465
## 37 1173918 1334639 1663622 1798714 1312574 1708245 1256746 1853142 1673831
## 38 1794912 1726665 1805445 1133510 1502242 1419251 1482786 1862351 1103794
## 39 1320191 1446723 1218591 1122030 1971479 1563062 1274168 1571032 1433835
## 40 1501744 1942942 1266657 1961923 1835983 1234040 1151409 1993136 1983569
## 41 1631522 1803455 1425193 1458191 1538731 1825195 1250499 1864685 1345102
## 42 1159037 1150689 1660148 1417141 1418586 1279134 1171870 1852424 1554782
## 43 1811867 1485909 1974179 1157059 1786132 1399191 1826406 1326460 1231739
## 44 1520591 1310777 1957713 1907326 1873544 1655483 1785986 1827503 1447457
## 45 1771096 1195861 1979395 1241662 1437456 1859416 1939284 1915865 1619186
## 46 1146902 1832249 1492704 1579265 1332048 1563537 1123567 1618583 1326369
## 47 1134317 1163996 1891068 1853855 1708715 1197698 1803330 1590043 1516758
## 48 1977749 1687136 1199490 1163092 1334864 1621989 1545621 1555554 1179331
## 49 1677347 1380662 1176100 1888948 1922085 1740826 1238174 1539322 1539603
## 50 1788920 1518578 1289663 1436888 1251678 1721874 1980167 1901394 1648755
## 51 1775190 1498098 1198212 1881688 1750527 1523124 1587602 1504455 1282142
##      Y2011   Y2012   Y2013   Y2014   Y2015
## 1  1440756 1186741 1852841 1558906 1916661
## 2  1230866 1512804 1985302 1580394 1979143
## 3  1130709 1907284 1363279 1525866 1647724
## 4  1928238 1216675 1591896 1360959 1329341
## 5  1639670 1921845 1156536 1388461 1644607
## 6  1665877 1491604 1178355 1383978 1330736
## 7  1945524 1228529 1582249 1503156 1718072
## 8  1318669 1984027 1671279 1803169 1627508
## 9  1353449 1979708 1912654 1782169 1410183
## 10 1497051 1131928 1107448 1407784 1170389
## 11 1851245 1850111 1887157 1259353 1725470
## 12 1902816 1695126 1517184 1948108 1150882
## 13 1456340 1643855 1312561 1713718 1757171
## 14 1751422 1696729 1915435 1645465 1583516
## 15 1559924 1905760 1129794 1988394 1467614
## 16 1992996 1501879 1173694 1431705 1641866
## 17 1465812 1882929 1410249 1930090 1385528
## 18 1512894 1916616 1878271 1722762 1913350
## 19 1210385 1234234 1287663 1908602 1403857
## 20 1706080 1437088 1318546 1116792 1529233
## 21 1872327 1175819 1314343 1979529 1569566
## 22 1658538 1482203 1731917 1669749 1963337
## 23 1274673 1709853 1815596 1965196 1646634
## 24 1575533 1910216 1972021 1515366 1864553
```

```
## 25 1629132 1988270 1907777 1649668 1991232
## 26 1425363 1800052 1698105 1767835 1996005
## 27 1520064 1185225 1465705 1110394 1125903
## 28 1187207 1569665 1690920 1459243 1802211
## 29 1143343 1980195 1283813 1225348 1903804
## 30 1704297 1131298 1197576 1242623 1963313
## 31 1390010 1202326 1100990 1850165 1183568
## 32 1653384 1475715 1623388 1533494 1868612
## 33 1683687 1500089 1718837 1619033 1367705
## 34 1553750 1472258 1104893 1596452 1229085
## 35 1271132 1430978 1529024 1563898 1604118
## 36 1887714 1227303 1840898 1880804 1573117
## 37 1822933 1674707 1900523 1956742 1307678
## 38 1935687 1905378 1522129 1509171 1893515
## 39 1483292 1290329 1475344 1931500 1668232
## 40 1781016 1909119 1531212 1990412 1611730
## 41 1116203 1532332 1591735 1188417 1110655
## 42 1647245 1811156 1147488 1302834 1136443
## 43 1469785 1849041 1560887 1349173 1162164
## 44 1978374 1882532 1698698 1646508 1705322
## 45 1288285 1108281 1123353 1801019 1729273
## 46 1792600 1714960 1146278 1282790 1565924
## 47 1171686 1262342 1647032 1706707 1850394
## 48 1150089 1775787 1273834 1387428 1377341
## 49 1872519 1462137 1683127 1204344 1198791
## 50 1940943 1729177 1510119 1701650 1846238
## 51 1881814 1673668 1994022 1204029 1853858
```

```
#Answer c
select(h_data,starts_with("Y"))
```

```
##      Y2002   Y2003   Y2004   Y2005   Y2006   Y2007   Y2008   Y2009   Y2010
## 1  1296530 1317711 1118631 1492583 1107408 1440134 1945229 1944173 1237582
## 2  1170302 1960378 1818085 1447852 1861639 1465841 1551826 1436541 1629616
## 3  1742027 1968140 1377583 1782199 1102568 1109382 1752886 1554330 1300521
## 4  1485531 1994927 1119299 1947979 1669191 1801213 1188104 1628980 1669295
## 5  1685349 1675807 1889570 1480280 1735069 1812546 1487315 1663809 1624509
## 6  1343824 1878473 1886149 1236697 1871471 1814218 1875146 1752387 1913275
## 7  1610512 1232844 1181949 1518933 1841266 1976976 1764457 1972730 1968730
## 8  1330403 1268673 1706751 1403759 1441351 1300836 1762096 1553585 1370984
## 9  1111437 1993741 1374643 1827949 1803852 1595981 1193245 1739748 1707823
## 10 1964626 1468852 1419738 1362787 1339608 1278550 1756185 1818438 1198403
## 11 1929009 1541565 1810773 1779091 1326846 1223770 1773090 1630325 1145473
## 12 1461570 1200280 1213993 1245931 1459383 1430465 1919423 1928416 1330509
## 13 1353210 1438538 1739154 1541015 1122387 1772050 1335481 1748608 1436809
## 14 1508356 1527440 1493029 1261353 1540274 1747614 1871645 1658551 1422021
## 15 1776918 1734104 1269927 1204117 1848073 1129546 1139551 1883976 1999102
## 16 1499269 1444576 1576367 1388924 1554813 1452911 1317983 1150783 1751389
## 17 1509054 1290700 1522230 1532094 1104256 1863278 1949478 1561528 1550433
## 18 1813878 1448846 1800760 1250524 1137913 1911227 1301848 1956681 1350895
## 19 1584734 1110625 1868456 1751920 1233709 1920301 1185085 1124853 1498662
## 20 1582720 1678622 1208496 1912040 1438549 1330014 1295877 1969163 1627262
## 21 1579713 1404700 1849798 1397738 1310270 1789128 1112765 1967225 1486246
## 22 1647582 1686259 1620601 1777250 1531641 1380529 1978904 1567651 1761048
## 23 1295635 1149931 1601027 1340716 1729449 1567494 1990431 1575185 1267626
```

```
## 24 1729921 1675204 1903907 1561839 1985692 1148621 1328133 1890633 1995304
## 25 1983285 1292558 1631325 1943311 1354579 1731643 1428291 1568049 1383227
## 26 1221316 1858368 1773451 1573967 1374863 1486197 1735099 1800620 1164202
## 27 1877154 1540099 1332722 1273327 1625721 1983568 1251742 1592690 1350619
## 28 1885081 1309769 1425527 1240465 1500594 1278272 1140598 1270585 1128711
## 29 1426117 1114500 1119707 1758830 1694526 1765826 1903270 1231480 1526066
## 30 1419776 1854370 1195119 1990062 1645430 1286967 1762936 1763211 1265642
## 31 1605532 1141514 1613550 1181452 1541327 1156804 1568034 1357418 1443718
## 32 1819239 1226057 1935991 1124400 1723493 1475985 1237704 1820856 1801430
## 33 1395149 1611371 1170675 1446810 1426941 1463171 1732098 1426216 1604531
## 34 1616742 1292223 1482792 1532347 1158716 1827420 1267737 1116168 1791535
## 35 1618807 1510193 1876940 1443172 1425030 1868788 1720352 1671468 1534571
## 36 1802132 1648498 1441386 1670280 1534888 1314824 1516621 1511460 1585465
## 37 1173918 1334639 1663622 1798714 1312574 1708245 1256746 1853142 1673831
## 38 1794912 1726665 1805445 1133510 1502242 1419251 1482786 1862351 1103794
## 39 1320191 1446723 1218591 1122030 1971479 1563062 1274168 1571032 1433835
## 40 1501744 1942942 1266657 1961923 1835983 1234040 1151409 1993136 1983569
## 41 1631522 1803455 1425193 1458191 1538731 1825195 1250499 1864685 1345102
## 42 1159037 1150689 1660148 1417141 1418586 1279134 1171870 1852424 1554782
## 43 1811867 1485909 1974179 1157059 1786132 1399191 1826406 1326460 1231739
## 44 1520591 1310777 1957713 1907326 1873544 1655483 1785986 1827503 1447457
## 45 1771096 1195861 1979395 1241662 1437456 1859416 1939284 1915865 1619186
## 46 1146902 1832249 1492704 1579265 1332048 1563537 1123567 1618583 1326369
## 47 1134317 1163996 1891068 1853855 1708715 1197698 1803330 1590043 1516758
## 48 1977749 1687136 1199490 1163092 1334864 1621989 1545621 1555554 1179331
## 49 1677347 1380662 1176100 1888948 1922085 1740826 1238174 1539322 1539603
## 50 1788920 1518578 1289663 1436888 1251678 1721874 1980167 1901394 1648755
## 51 1775190 1498098 1198212 1881688 1750527 1523124 1587602 1504455 1282142
##      Y2011   Y2012   Y2013   Y2014   Y2015
## 1  1440756 1186741 1852841 1558906 1916661
## 2  1230866 1512804 1985302 1580394 1979143
## 3  1130709 1907284 1363279 1525866 1647724
## 4  1928238 1216675 1591896 1360959 1329341
## 5  1639670 1921845 1156536 1388461 1644607
## 6  1665877 1491604 1178355 1383978 1330736
## 7  1945524 1228529 1582249 1503156 1718072
## 8  1318669 1984027 1671279 1803169 1627508
## 9  1353449 1979708 1912654 1782169 1410183
## 10 1497051 1131928 1107448 1407784 1170389
## 11 1851245 1850111 1887157 1259353 1725470
## 12 1902816 1695126 1517184 1948108 1150882
## 13 1456340 1643855 1312561 1713718 1757171
## 14 1751422 1696729 1915435 1645465 1583516
## 15 1559924 1905760 1129794 1988394 1467614
## 16 1992996 1501879 1173694 1431705 1641866
## 17 1465812 1882929 1410249 1930090 1385528
## 18 1512894 1916616 1878271 1722762 1913350
## 19 1210385 1234234 1287663 1908602 1403857
## 20 1706080 1437088 1318546 1116792 1529233
## 21 1872327 1175819 1314343 1979529 1569566
## 22 1658538 1482203 1731917 1669749 1963337
## 23 1274673 1709853 1815596 1965196 1646634
## 24 1575533 1910216 1972021 1515366 1864553
## 25 1629132 1988270 1907777 1649668 1991232
```

```
## 26 1425363 1800052 1698105 1767835 1996005
## 27 1520064 1185225 1465705 1110394 1125903
## 28 1187207 1569665 1690920 1459243 1802211
## 29 1143343 1980195 1283813 1225348 1903804
## 30 1704297 1131298 1197576 1242623 1963313
## 31 1390010 1202326 1100990 1850165 1183568
## 32 1653384 1475715 1623388 1533494 1868612
## 33 1683687 1500089 1718837 1619033 1367705
## 34 1553750 1472258 1104893 1596452 1229085
## 35 1271132 1430978 1529024 1563898 1604118
## 36 1887714 1227303 1840898 1880804 1573117
## 37 1822933 1674707 1900523 1956742 1307678
## 38 1935687 1905378 1522129 1509171 1893515
## 39 1483292 1290329 1475344 1931500 1668232
## 40 1781016 1909119 1531212 1990412 1611730
## 41 1116203 1532332 1591735 1188417 1110655
## 42 1647245 1811156 1147488 1302834 1136443
## 43 1469785 1849041 1560887 1349173 1162164
## 44 1978374 1882532 1698698 1646508 1705322
## 45 1288285 1108281 1123353 1801019 1729273
## 46 1792600 1714960 1146278 1282790 1565924
## 47 1171686 1262342 1647032 1706707 1850394
## 48 1150089 1775787 1273834 1387428 1377341
## 49 1872519 1462137 1683127 1204344 1198791
## 50 1940943 1729177 1510119 1701650 1846238
## 51 1881814 1673668 1994022 1204029 1853858
```

4) Output only certain observations.

a) (2 points) Output only the observations with an index values of A and C and N.
b) (2 points) Output only the observations related to Illinois and California.

```
#Answer a
filter(h_data, Index=="A"|Index=="C"|Index=="N")
```

```
##    Index          State   Y2002   Y2003   Y2004   Y2005   Y2006   Y2007
## 1      A        Alabama 1296530 1317711 1118631 1492583 1107408 1440134
## 2      A         Alaska 1170302 1960378 1818085 1447852 1861639 1465841
## 3      A        Arizona 1742027 1968140 1377583 1782199 1102568 1109382
## 4      A       Arkansas 1485531 1994927 1119299 1947979 1669191 1801213
## 5      C     California 1685349 1675807 1889570 1480280 1735069 1812546
## 6      C       Colorado 1343824 1878473 1886149 1236697 1871471 1814218
## 7      C    Connecticut 1610512 1232844 1181949 1518933 1841266 1976976
## 8      N       Nebraska 1885081 1309769 1425527 1240465 1500594 1278272
## 9      N         Nevada 1426117 1114500 1119707 1758830 1694526 1765826
## 10     N  New Hampshire 1419776 1854370 1195119 1990062 1645430 1286967
## 11     N     New Jersey 1605532 1141514 1613550 1181452 1541327 1156804
## 12     N     New Mexico 1819239 1226057 1935991 1124400 1723493 1475985
## 13     N       New York 1395149 1611371 1170675 1446810 1426941 1463171
## 14     N North Carolina 1616742 1292223 1482792 1532347 1158716 1827420
## 15     N   North Dakota 1618807 1510193 1876940 1443172 1425030 1868788
##      Y2008   Y2009   Y2010   Y2011   Y2012   Y2013   Y2014   Y2015
## 1  1945229 1944173 1237582 1440756 1186741 1852841 1558906 1916661
## 2  1551826 1436541 1629616 1230866 1512804 1985302 1580394 1979143
## 3  1752886 1554330 1300521 1130709 1907284 1363279 1525866 1647724
## 4  1188104 1628980 1669295 1928238 1216675 1591896 1360959 1329341
```

```
## 5  1487315 1663809 1624509 1639670 1921845 1156536 1388461 1644607
## 6  1875146 1752387 1913275 1665877 1491604 1178355 1383978 1330736
## 7  1764457 1972730 1968730 1945524 1228529 1582249 1503156 1718072
## 8  1140598 1270585 1128711 1187207 1569665 1690920 1459243 1802211
## 9  1903270 1231480 1526066 1143343 1980195 1283813 1225348 1903804
## 10 1762936 1763211 1265642 1704297 1131298 1197576 1242623 1963313
## 11 1568034 1357418 1443718 1390010 1202326 1100990 1850165 1183568
## 12 1237704 1820856 1801430 1653384 1475715 1623388 1533494 1868612
## 13 1732098 1426216 1604531 1683687 1500089 1718837 1619033 1367705
## 14 1267737 1116168 1791535 1553750 1472258 1104893 1596452 1229085
## 15 1720352 1671468 1534571 1271132 1430978 1529024 1563898 1604118
```

```
#Answer b
filter(h_data, State=="Illinois"|State=="California")
```

```
##   Index      State   Y2002   Y2003   Y2004   Y2005   Y2006   Y2007   Y2008
## 1     C California 1685349 1675807 1889570 1480280 1735069 1812546 1487315
## 2     I   Illinois 1508356 1527440 1493029 1261353 1540274 1747614 1871645
##     Y2009   Y2010   Y2011   Y2012   Y2013   Y2014   Y2015
## 1 1663809 1624509 1639670 1921845 1156536 1388461 1644607
## 2 1658551 1422021 1751422 1696729 1915435 1645465 1583516
```

5) (2 points) Output data that shows only the index, state, and 2010 income values greater than $1,500,000.

a) (Extra Credit: 2 points) The same code can be re-written concisely using the pipe operator %>%. What is that command?

```
select(filter(h_data, Y2010>=1500000),Index,State,Y2010) #normal way
```

```
##    Index                State   Y2010
## 1      A               Alaska 1629616
## 2      A             Arkansas 1669295
## 3      C           California 1624509
## 4      C             Colorado 1913275
## 5      C          Connecticut 1968730
## 6      D District of Columbia 1707823
## 7      I              Indiana 1999102
## 8      I                 Iowa 1751389
## 9      K               Kansas 1550433
## 10     M                Maine 1627262
## 11     M        Massachusetts 1761048
## 12     M            Minnesota 1995304
## 13     N               Nevada 1526066
## 14     N           New Mexico 1801430
## 15     N             New York 1604531
## 16     N       North Carolina 1791535
## 17     N         North Dakota 1534571
## 18     O                 Ohio 1585465
## 19     O             Oklahoma 1673831
## 20     R         Rhode Island 1983569
## 21     S         South Dakota 1554782
## 22     U                 Utah 1619186
## 23     V             Virginia 1516758
## 24     W        West Virginia 1539603
## 25     W            Wisconsin 1648755
```

```r
select(h_data,Index,State,Y2010) %>% filter(Y2010>=1500000) #with the pipe opearator
```

```
##    Index                State   Y2010
## 1      A               Alaska 1629616
## 2      A             Arkansas 1669295
## 3      C           California 1624509
## 4      C             Colorado 1913275
## 5      C          Connecticut 1968730
## 6      D District of Columbia 1707823
## 7      I              Indiana 1999102
## 8      I                 Iowa 1751389
## 9      K               Kansas 1550433
## 10     M                Maine 1627262
## 11     M        Massachusetts 1761048
## 12     M            Minnesota 1995304
## 13     N               Nevada 1526066
## 14     N           New Mexico 1801430
## 15     N             New York 1604531
## 16     N       North Carolina 1791535
## 17     N         North Dakota 1534571
## 18     O                 Ohio 1585465
## 19     O             Oklahoma 1673831
## 20     R         Rhode Island 1983569
## 21     S         South Dakota 1554782
## 22     U                 Utah 1619186
## 23     V             Virginia 1516758
## 24     W        West Virginia 1539603
## 25     W            Wisconsin 1648755
```

6) (3 points) Output only the state and 2006 values, and arrange the values in descending order of 2006 values. What are the 3 highest income generating states?

The states of Minnesorta, Pennsylvania, and West Virginia have the highest income:

```r
select(h_data,State,Y2006) %>% arrange(desc(Y2006))
```

```
##                   State   Y2006
## 1             Minnesota 1985692
## 2          Pennsylvania 1971479
## 3         West Virginia 1922085
## 4                 Texas 1873544
## 5              Colorado 1871471
## 6                Alaska 1861639
## 7               Indiana 1848073
## 8           Connecticut 1841266
## 9          Rhode Island 1835983
## 10 District of Columbia 1803852
## 11            Tennessee 1786132
## 12              Wyoming 1750527
## 13           California 1735069
## 14             Michigan 1729449
## 15           New Mexico 1723493
## 16             Virginia 1708715
## 17               Nevada 1694526
## 18             Arkansas 1669191
```

```
## 19        New Hampshire 1645430
## 20              Montana 1625721
## 21                 Iowa 1554813
## 22           New Jersey 1541327
## 23             Illinois 1540274
## 24       South Carolina 1538731
## 25                 Ohio 1534888
## 26        Massachusetts 1531641
## 27               Oregon 1502242
## 28             Nebraska 1500594
## 29               Hawaii 1459383
## 30             Delaware 1441351
## 31                Maine 1438549
## 32                 Utah 1437456
## 33             New York 1426941
## 34         North Dakota 1425030
## 35         South Dakota 1418586
## 36             Missouri 1374863
## 37          Mississippi 1354579
## 38              Florida 1339608
## 39           Washington 1334864
## 40              Vermont 1332048
## 41              Georgia 1326846
## 42             Oklahoma 1312574
## 43             Maryland 1310270
## 44            Wisconsin 1251678
## 45            Louisiana 1233709
## 46       North Carolina 1158716
## 47              Kentucky 1137913
## 48                 Idaho 1122387
## 49               Alabama 1107408
## 50                Kansas 1104256
## 51               Arizona 1102568
```

7) (3 points) Using the mutate function, create a new column "ratio" which divides the income of 2015 by 2014. Use the appropriate combination of commands so only the state, 2014, 2015, and the new column is saved or output.

```
mutate(select(h_data,Y2014,Y2015),ratio=Y2015/Y2014)
```

```
##      Y2014   Y2015     ratio
## 1  1558906 1916661 1.2294911
## 2  1580394 1979143 1.2523099
## 3  1525866 1647724 1.0798615
## 4  1360959 1329341 0.9767679
## 5  1388461 1644607 1.1844820
## 6  1383978 1330736 0.9615297
## 7  1503156 1718072 1.1429765
## 8  1803169 1627508 0.9025821
## 9  1782169 1410183 0.7912734
## 10 1407784 1170389 0.8313697
## 11 1259353 1725470 1.3701242
## 12 1948108 1150882 0.5907691
## 13 1713718 1757171 1.0253560
## 14 1645465 1583516 0.9623517
```

```
## 15 1988394 1467614 0.7380901
## 16 1431705 1641866 1.1467907
## 17 1930090 1385528 0.7178567
## 18 1722762 1913350 1.1106293
## 19 1908602 1403857 0.7355420
## 20 1116792 1529233 1.3693087
## 21 1979529 1569566 0.7928987
## 22 1669749 1963337 1.1758276
## 23 1965196 1646634 0.8378981
## 24 1515366 1864553 1.2304308
## 25 1649668 1991232 1.2070501
## 26 1767835 1996005 1.1290675
## 27 1110394 1125903 1.0139671
## 28 1459243 1802211 1.2350315
## 29 1225348 1903804 1.5536843
## 30 1242623 1963313 1.5799748
## 31 1850165 1183568 0.6397094
## 32 1533494 1868612 1.2185323
## 33 1619033 1367705 0.8447666
## 34 1596452 1229085 0.7698853
## 35 1563898 1604118 1.0257178
## 36 1880804 1573117 0.8364067
## 37 1956742 1307678 0.6682935
## 38 1509171 1893515 1.2546723
## 39 1931500 1668232 0.8636976
## 40 1990412 1611730 0.8097469
## 41 1188417 1110655 0.9345667
## 42 1302834 1136443 0.8722853
## 43 1349173 1162164 0.8613899
## 44 1646508 1705322 1.0357204
## 45 1801019 1729273 0.9601637
## 46 1282790 1565924 1.2207173
## 47 1706707 1850394 1.0841896
## 48 1387428 1377341 0.9927297
## 49 1204344 1198791 0.9953892
## 50 1701650 1846238 1.0849693
## 51 1204029 1853858 1.5397121
```

8) (3 points) Using group_by and summarise, create a table grouped by Index and then summarizes each group by taking the mean of Y2007

```
h_data %>% group_by(Index) %>% summarise(mean_2007=mean(Y2007))
```

```
## # A tibble: 19 x 2
##    Index mean_2007
##    <fct>     <dbl>
##  1 A      1454142.
##  2 C      1867913.
##  3 D      1448408.
##  4 F      1278550
##  5 G      1223770
##  6 H      1430465
##  7 I      1525530.
##  8 K      1887252.
##  9 L      1920301
```

```
## 10 M      1552149.
## 11 N      1515404.
## 12 O      1480773.
## 13 P      1563062
## 14 R      1234040
## 15 S      1552164.
## 16 T      1527337
## 17 U      1859416
## 18 V      1380618.
## 19 W      1651953.
```

9) (3 points) Group the observations by index. Output data that shows the average ratio of 2015 divided by 2014 for each group.

```
h_data %>% group_by(Index) %>% mutate(ratio=Y2015/Y2014) #This is the answer.
```

```
## # A tibble: 51 x 17
## # Groups:   Index [19]
##    Index State  Y2002  Y2003  Y2004  Y2005  Y2006  Y2007  Y2008  Y2009
##    <fct> <fct>  <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>
##  1 A     Alab~ 1.30e6 1.32e6 1.12e6 1.49e6 1.11e6 1.44e6 1.95e6 1.94e6
##  2 A     Alas~ 1.17e6 1.96e6 1.82e6 1.45e6 1.86e6 1.47e6 1.55e6 1.44e6
##  3 A     Ariz~ 1.74e6 1.97e6 1.38e6 1.78e6 1.10e6 1.11e6 1.75e6 1.55e6
##  4 A     Arka~ 1.49e6 1.99e6 1.12e6 1.95e6 1.67e6 1.80e6 1.19e6 1.63e6
##  5 C     Cali~ 1.69e6 1.68e6 1.89e6 1.48e6 1.74e6 1.81e6 1.49e6 1.66e6
##  6 C     Colo~ 1.34e6 1.88e6 1.89e6 1.24e6 1.87e6 1.81e6 1.88e6 1.75e6
##  7 C     Conn~ 1.61e6 1.23e6 1.18e6 1.52e6 1.84e6 1.98e6 1.76e6 1.97e6
##  8 D     Dela~ 1.33e6 1.27e6 1.71e6 1.40e6 1.44e6 1.30e6 1.76e6 1.55e6
##  9 D     Dist~ 1.11e6 1.99e6 1.37e6 1.83e6 1.80e6 1.60e6 1.19e6 1.74e6
## 10 F     Flor~ 1.96e6 1.47e6 1.42e6 1.36e6 1.34e6 1.28e6 1.76e6 1.82e6
## # ... with 41 more rows, and 7 more variables: Y2010 <int>, Y2011 <int>,
## #   Y2012 <int>, Y2013 <int>, Y2014 <int>, Y2015 <int>, ratio <dbl>
```

```
#This is to show more clearly the ratio.
h_data %>% group_by(Index) %>% mutate(ratio=Y2015/Y2014) %>% select(ratio)
```

```
## Adding missing grouping variables: `Index`
```

```
## # A tibble: 51 x 2
## # Groups:   Index [19]
##    Index ratio
##    <fct> <dbl>
##  1 A     1.23
##  2 A     1.25
##  3 A     1.08
##  4 A     0.977
##  5 C     1.18
##  6 C     0.962
##  7 C     1.14
##  8 D     0.903
##  9 D     0.791
## 10 F     0.831
## # ... with 41 more rows
```