Week 1- Lab I

Amol Gote

09/06/2019

source("http://www.openintro.org/stat/data/arbuthnot.R")  
arbuthnot

## year boys girls  
## 1 1629 5218 4683  
## 2 1630 4858 4457  
## 3 1631 4422 4102  
## 4 1632 4994 4590  
## 5 1633 5158 4839  
## 6 1634 5035 4820  
## 7 1635 5106 4928  
## 8 1636 4917 4605  
## 9 1637 4703 4457  
## 10 1638 5359 4952  
## 11 1639 5366 4784  
## 12 1640 5518 5332  
## 13 1641 5470 5200  
## 14 1642 5460 4910  
## 15 1643 4793 4617  
## 16 1644 4107 3997  
## 17 1645 4047 3919  
## 18 1646 3768 3395  
## 19 1647 3796 3536  
## 20 1648 3363 3181  
## 21 1649 3079 2746  
## 22 1650 2890 2722  
## 23 1651 3231 2840  
## 24 1652 3220 2908  
## 25 1653 3196 2959  
## 26 1654 3441 3179  
## 27 1655 3655 3349  
## 28 1656 3668 3382  
## 29 1657 3396 3289  
## 30 1658 3157 3013  
## 31 1659 3209 2781  
## 32 1660 3724 3247  
## 33 1661 4748 4107  
## 34 1662 5216 4803  
## 35 1663 5411 4881  
## 36 1664 6041 5681  
## 37 1665 5114 4858  
## 38 1666 4678 4319  
## 39 1667 5616 5322  
## 40 1668 6073 5560  
## 41 1669 6506 5829  
## 42 1670 6278 5719  
## 43 1671 6449 6061  
## 44 1672 6443 6120  
## 45 1673 6073 5822  
## 46 1674 6113 5738  
## 47 1675 6058 5717  
## 48 1676 6552 5847  
## 49 1677 6423 6203  
## 50 1678 6568 6033  
## 51 1679 6247 6041  
## 52 1680 6548 6299  
## 53 1681 6822 6533  
## 54 1682 6909 6744  
## 55 1683 7577 7158  
## 56 1684 7575 7127  
## 57 1685 7484 7246  
## 58 1686 7575 7119  
## 59 1687 7737 7214  
## 60 1688 7487 7101  
## 61 1689 7604 7167  
## 62 1690 7909 7302  
## 63 1691 7662 7392  
## 64 1692 7602 7316  
## 65 1693 7676 7483  
## 66 1694 6985 6647  
## 67 1695 7263 6713  
## 68 1696 7632 7229  
## 69 1697 8062 7767  
## 70 1698 8426 7626  
## 71 1699 7911 7452  
## 72 1700 7578 7061  
## 73 1701 8102 7514  
## 74 1702 8031 7656  
## 75 1703 7765 7683  
## 76 1704 6113 5738  
## 77 1705 8366 7779  
## 78 1706 7952 7417  
## 79 1707 8379 7687  
## 80 1708 8239 7623  
## 81 1709 7840 7380  
## 82 1710 7640 7288

**Excercise 1**

What command would you use to extract just the counts of girls baptized?

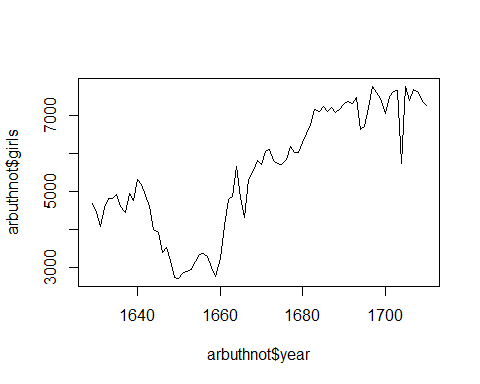
arbuthnot$girls

## [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910  
## [15] 4617 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382  
## [29] 3289 3013 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719  
## [43] 6061 6120 5822 5738 5717 5847 6203 6033 6041 6299 6533 6744 7158 7127  
## [57] 7246 7119 7214 7101 7167 7302 7392 7316 7483 6647 6713 7229 7767 7626  
## [71] 7452 7061 7514 7656 7683 5738 7779 7417 7687 7623 7380 7288

**Excercise 2**

Is there an apparent trend in the number of girls baptized over the years? How would you describe it?

plot(x = arbuthnot$year, y = arbuthnot$girls, type = "l")



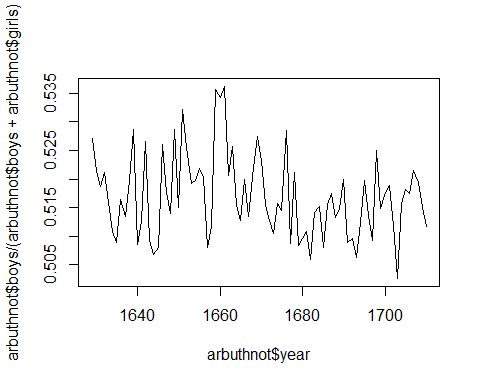
Answer:

1. Number of girls baptized and years are associated with each other.
2. There has been steady increase in the number of girls baptized as years have progressed, barring the time period of 140 to 1660 where there was steep decline, in the number of girls baptized.

**Excercise 3**

Now, make a plot of the proportion of boys over time. What do you see?

plot(x = arbuthnot$year, y = arbuthnot$boys/(arbuthnot$boys + arbuthnot$girls), type = "l")



**Activities 1-4 from the On Your Own section of Lab 1 (Introduction to R & RStudio lab)**

source("http://www.openintro.org/stat/data/present.R")  
present

## year boys girls  
## 1 1940 1211684 1148715  
## 2 1941 1289734 1223693  
## 3 1942 1444365 1364631  
## 4 1943 1508959 1427901  
## 5 1944 1435301 1359499  
## 6 1945 1404587 1330869  
## 7 1946 1691220 1597452  
## 8 1947 1899876 1800064  
## 9 1948 1813852 1721216  
## 10 1949 1826352 1733177  
## 11 1950 1823555 1730594  
## 12 1951 1923020 1827830  
## 13 1952 1971262 1875724  
## 14 1953 2001798 1900322  
## 15 1954 2059068 1958294  
## 16 1955 2073719 1973576  
## 17 1956 2133588 2029502  
## 18 1957 2179960 2074824  
## 19 1958 2152546 2051266  
## 20 1959 2173638 2071158  
## 21 1960 2179708 2078142  
## 22 1961 2186274 2082052  
## 23 1962 2132466 2034896  
## 24 1963 2101632 1996388  
## 25 1964 2060162 1967328  
## 26 1965 1927054 1833304  
## 27 1966 1845862 1760412  
## 28 1967 1803388 1717571  
## 29 1968 1796326 1705238  
## 30 1969 1846572 1753634  
## 31 1970 1915378 1816008  
## 32 1971 1822910 1733060  
## 33 1972 1669927 1588484  
## 34 1973 1608326 1528639  
## 35 1974 1622114 1537844  
## 36 1975 1613135 1531063  
## 37 1976 1624436 1543352  
## 38 1977 1705916 1620716  
## 39 1978 1709394 1623885  
## 40 1979 1791267 1703131  
## 41 1980 1852616 1759642  
## 42 1981 1860272 1768966  
## 43 1982 1885676 1794861  
## 44 1983 1865553 1773380  
## 45 1984 1879490 1789651  
## 46 1985 1927983 1832578  
## 47 1986 1924868 1831679  
## 48 1987 1951153 1858241  
## 49 1988 2002424 1907086  
## 50 1989 2069490 1971468  
## 51 1990 2129495 2028717  
## 52 1991 2101518 2009389  
## 53 1992 2082097 1982917  
## 54 1993 2048861 1951379  
## 55 1994 2022589 1930178  
## 56 1995 1996355 1903234  
## 57 1996 1990480 1901014  
## 58 1997 1985596 1895298  
## 59 1998 2016205 1925348  
## 60 1999 2026854 1932563  
## 61 2000 2076969 1981845  
## 62 2001 2057922 1968011  
## 63 2002 2057979 1963747

1. What years are included in this data set? What are the dimensions of the data frame and what are the variable or column names?

Years included in the dataset

present$year

## [1] 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953  
## [15] 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967  
## [29] 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981  
## [43] 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995  
## [57] 1996 1997 1998 1999 2000 2001 2002

Dimensions on the data frame

dim(present)

## [1] 63 3

Variable or Column Names

names(present)

## [1] "year" "boys" "girls"

1. How do these counts compare to Arbuthnot’s? Are they on a similar scale?

Total number of boys and girls in Present datset = **231809422**

Total number of boys and girls in Arbuthnot datset = **938223**

Arbuthnot’s datset’s average boys to girls ratio = **1.070748**

Present datset’s average boys to girls ratio = **1.051353**

Boys to girl ratio seem to be constant, there is common theme across both data sets which is number of boys is greater than the number of girls. From below calculation it clearly indicates the average ratio of boys to girls is almost constant across 2 dataset (Arbuthnot and Present), difference of 0.02 in average.

As far as scale is concerned the Present data set is in millions, where as the Arbuthnot’s data is in thousands that too less than 10,000. Total count of number of girls and boys in Present dataset is \*247\*\* time that of the total number bos and girls in Arbuthnot’s dataset.

arbuthnot boys to girls ratio

arbuthnot$boys / arbuthnot$girls

## [1] 1.114243 1.089971 1.078011 1.088017 1.065923 1.044606 1.036120  
## [8] 1.067752 1.055194 1.082189 1.121656 1.034884 1.051923 1.112016  
## [15] 1.038120 1.027521 1.032661 1.109867 1.073529 1.057215 1.121267  
## [22] 1.061719 1.137676 1.107290 1.080095 1.082416 1.091371 1.084565  
## [29] 1.032533 1.047793 1.153901 1.146905 1.156075 1.085988 1.108584  
## [36] 1.063369 1.052697 1.083121 1.055242 1.092266 1.116143 1.097744  
## [43] 1.064016 1.052778 1.043112 1.065354 1.059647 1.120575 1.035467  
## [50] 1.088679 1.034100 1.039530 1.044237 1.024466 1.058536 1.062860  
## [57] 1.032846 1.064054 1.072498 1.054359 1.060974 1.083128 1.036526  
## [64] 1.039092 1.025792 1.050850 1.081931 1.055748 1.037981 1.104904  
## [71] 1.061594 1.073219 1.078254 1.048981 1.010673 1.065354 1.075460  
## [78] 1.072132 1.090022 1.080808 1.062331 1.048299

Average of arbuthnot boys to girls ratio

mean(arbuthnot$boys / arbuthnot$girls)

## [1] 1.070748

present boys to girl ratio

present$boys / present$girls

## [1] 1.054817 1.053969 1.058429 1.056767 1.055757 1.055391 1.058698  
## [8] 1.055449 1.053820 1.053760 1.053716 1.052078 1.050934 1.053399  
## [15] 1.051460 1.050742 1.051286 1.050672 1.049374 1.049480 1.048873  
## [22] 1.050057 1.047948 1.052717 1.047188 1.051137 1.048540 1.049964  
## [29] 1.053417 1.052997 1.054719 1.051845 1.051271 1.052129 1.054797  
## [36] 1.053605 1.052538 1.052569 1.052657 1.051749 1.052837 1.051615  
## [43] 1.050597 1.051976 1.050199 1.052061 1.050876 1.050000 1.049991  
## [50] 1.049720 1.049676 1.045849 1.050017 1.049955 1.047877 1.048928  
## [57] 1.047062 1.047643 1.047190 1.048791 1.047998 1.045686 1.047986

Average of present boys to girl ratio

mean(present$boys / present$girls)

## [1] 1.051353

Total number of girls and boys for arbuthnot

sum(arbuthnot$boys) + sum(arbuthnot$girls)

## [1] 938223

Total number of girls and boys for present

sum(present$boys) + sum(present$girls)

## [1] 231809422

Scale ratio in total boys and girls count between arbuthnot and present

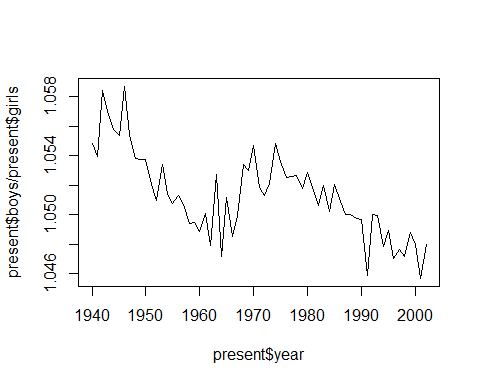
(sum(present$boys) + sum(present$girls))/(sum(arbuthnot$boys) + sum(arbuthnot$girls))

## [1] 247.0728

1. Make a plot that displays the boy-to-girl ratio for every year in the data set. What do you see? Does Arbuthnot’s observation about boys being born in greater proportion than girls hold up in the U.S.? Include the plot in your response.

Plot that displays the boy-to-girl ratio for every year in the data set

plot(x=present$year, y=present$boys/present$girls, type="l")



Boys to girl ratio almost remains constant.

present$boys > present$girls

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE

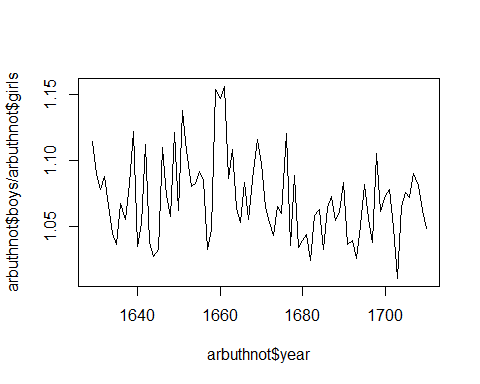
arbuthnot$boys > arbuthnot$girls

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [71] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

As you can see in both data sets number of boys is greater than number of girls. So Arbuthnot’s observation about boys being born in greater proportion than girls hold up in the U.S.

Plot that displays the boy-to-girl ratio for every year in the Arbuthnot data set

plot(x=arbuthnot$year, y=arbuthnot$boys/arbuthnot$girls, type="l")



In what year did we see the most total number of births in the U.S.?

present[which.max(present$boys + present$girls),]$year

## [1] 1961