Illapani_Lab0

Srini Illapani

September 3, 2015

Lab0

In the Intro to R lesson, I recreated some of the displays and preliminary analysis of Arbuthnot's baptism data. Lab 0 assignment involves repeating these steps, but for present day birth records in the United States. Loading up the present day data with the following command.

```
source("more/present.R")
```

The Data: US Birth Records

We can take a look at the data by typing its name into the console.

```
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
```

What you should see are four columns of numbers, each row representing a different year: the first entry in each row is simply the row number (an index we can use to access the data from individual years if we want), the second is the year, and the third and fourth are the numbers of boys and girls born that year, respectively.

Note that the row numbers in the first column are not part of Arbuthnot's data. R adds them as part of its printout to help you make visual comparisons. You can think of them as the index that you see on the left side of a spreadsheet. In fact, the comparison to a spreadsheet will generally be helpful. R has stored the data in a kind of spreadsheet or table called a data frame. The data are stored in a data frame called `present`.

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?

The data includes birth numbers from year 1940 to 2002. One can see the dimensions of this data frame by typing:

```
dim(present)
```

```
## [1] 63  3
```

This command should output `[1] 63 3`, indicating that there are 62 rows and 3 columns, just as it says next to the object in your workspace. You can see the names of these columns (or variables) by typing:

```
names(present)
```

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

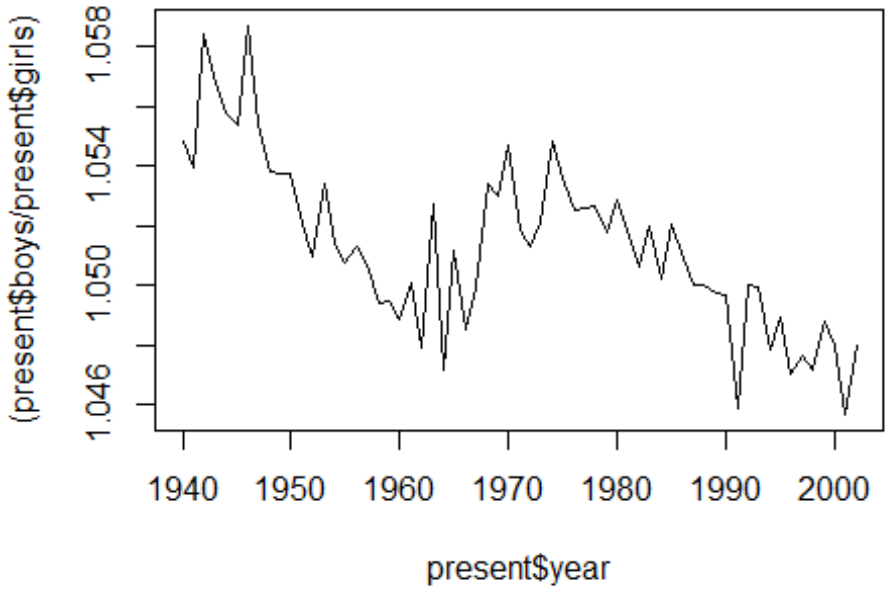
2. How do these counts compare to Arbuthnot's? Are they on a similar scale?

The counts in the 'present' data frame differ hugely compared to the ones in 'Arbuthnot's data frame. They are on a different scale, thousands vs. millions scale.

3. 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.

Here is how one can create a plot diagram using the data frame. To calculate the boy-to-girl ratio for every year in the data set, use the following plot function:

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



Compared to Arbuthnot's data, the trend is similar with the US present data. The boys to girls birth ratio is always in greater proportion for boys each year compared to girls.

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

The year is 1961. I had to use the sqldf package to use the select statement as shown below.

```
library(sqldf)

## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite
## Loading required package: DBI

DF <- data.frame(present)
top_year <- sqldf("select year, boys+girls from DF
             order by boys+girls DESC LIMIT 1",
             row.names = TRUE)

## Loading required package: tcltk

top_year

##   year boys+girls
## 1 1961    4268326
```

These data come from a report by the Centers for Disease Control http://www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_20.pdf. Check it out if you would like to read more about an analysis of sex ratios at birth in the United States.