

PresentHW

Samuel Kigamba

Sep 1, 2019

On Your Own

In the previous few pages, you recreated some of the displays and preliminary analysis of Arbuthnot's baptism data. Your assignment involves repeating these steps, but for present day birth records in the United States. Load up the present day data with the following command.

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

```
head(present, 5)
```

```
##   year   boys  girls
## 1 1940 1211684 1148715
## 2 1941 1289734 1223693
## 3 1942 1444365 1364631
## 4 1943 1508959 1427901
## 5 1944 1435301 1359499
```

```
tail(present, 5)
```

```
##   year   boys  girls
## 59 1998 2016205 1925348
## 60 1999 2026854 1932563
## 61 2000 2076969 1981845
## 62 2001 2057922 1968011
## 63 2002 2057979 1963747
```

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

```
head(arbuthnot, 5)
```

```
##   year boys girls
## 1 1629 5218 4683
## 2 1630 4858 4457
## 3 1631 4422 4102
## 4 1632 4994 4590
## 5 1633 5158 4839
```

```
tail(arbuthnot, 5)
```

```
##   year boys girls
## 78 1706 7952 7417
## 79 1707 8379 7687
## 80 1708 8239 7623
## 81 1709 7840 7380
## 82 1710 7640 7288
```

What years are included in this data set?

```
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
```

The year is 1940-2002

What are the dimensions of the data frame?

```
dim(present)
```

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

63 rows and 3 columns

What are the variable or column names?

```
names(present)
```

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

“year” “boys” “girls”

How do these counts compare to Arbuthnots? Are they on a similar scale?

both have 3 columns, “year” “boys” “girls”.

```
names(arbuthnot)
```

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

```
names(present)
```

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

Arbuthnot has 82 rows while present has 63,

```
tail(present, 1)
```

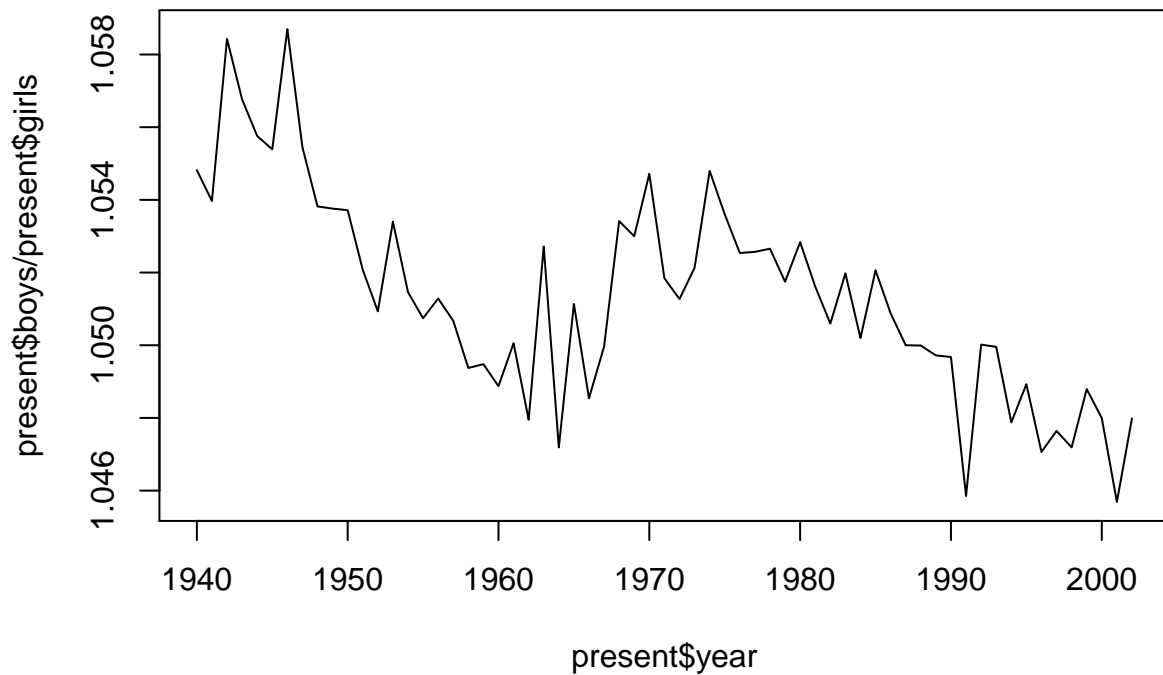
```
##   year   boys  girls
## 63 2002 2057979 1963747
```

```
tail(arbuthnot, 1)
```

```
##   year boys girls
## 82 1710 7640 7288
```

Make a plot that displays the boy-to-girl ratio for every year in the data set.

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

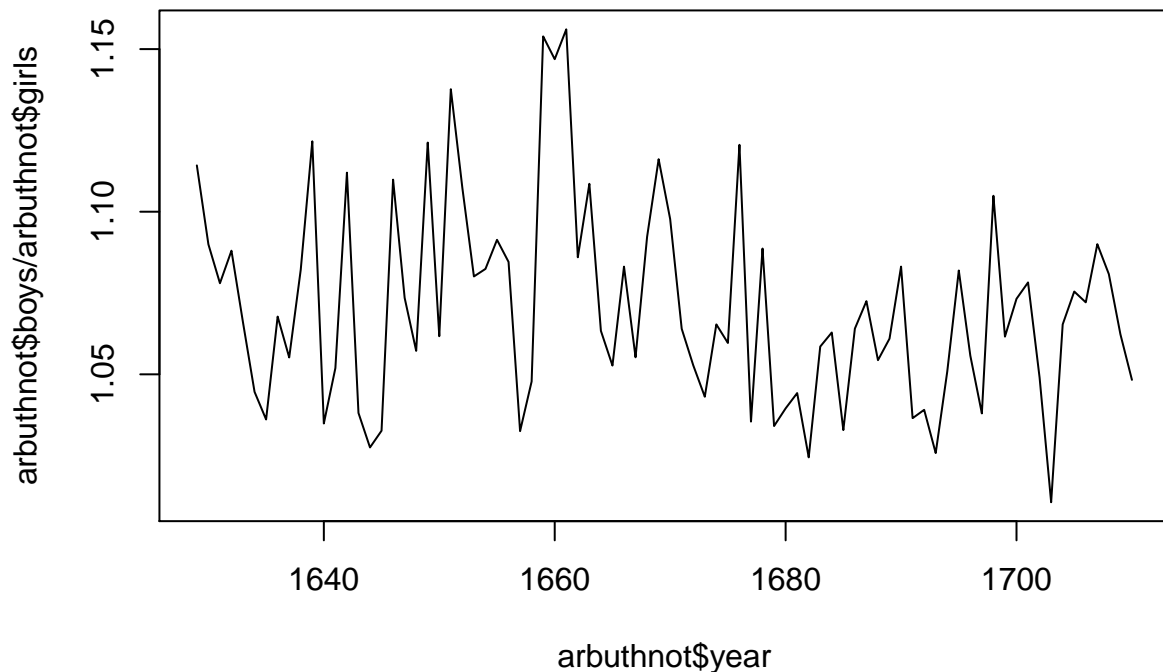


What do you see? The ratio of boys to girls is higher although reducing for all the years.

Does Arbuthnot's observation about boys being born in greater proportion than girls hold up in the U.S.? yes it does.

Include the plot in your response.

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



In what year did we see the most total number of births in the U.S.? Year 1961 total population of 4,268,326.

```
present[present$boys + present$girls == max(present$boys + present$girls),]
```

```
##   year   boys  girls
## 22 1961 2186274 2082052
```

You can refer to the help files or the R reference card <http://cran.r-project.org/doc/contrib/Short-refcard.pdf> to find helpful commands.

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.

That was a short introduction to R and RStudio, but we will provide you with more functions and a more complete sense of the language as the course progresses. Feel free to browse around the websites for R and RStudio if you're interested in learning more, or find more labs for practice at <http://openintro.org>.