Introduction to R and RStudio

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

Exercise 1 What command would you use to extract just the counts of girls baptized? Try

```
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
```

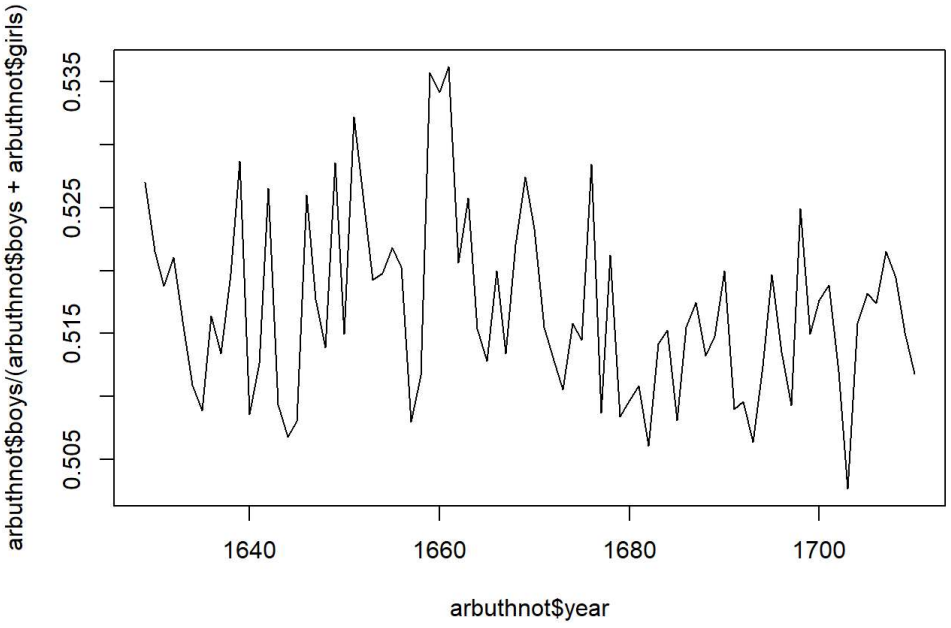
Exercise 2 Is there an apparent trend in the number of girls baptized over the years? How would you describe it?

since 1660, the number of girls have been increased gradually.

Exercise 3 Now, make a plot of the proportion of boys over time. What do you see?

The ratio is greater than 0.5 all along years , that means there were more new born boys than girls.

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



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")
```

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? ##years included in the 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
```

dimension of the data frame

```
dim(present)
```

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

column name

```
names(present)
```

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

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

the present data set have much larger new born compare to Arbuthnot's.

```
summary(present)
```

```
##       year           boys             girls
##  Min.   :1940   Min.   :1211684   Min.   :1148715
##  1st Qu.:1956   1st Qu.:1799857   1st Qu.:1711405
##  Median :1971   Median :1924868   Median :1831679
##  Mean   :1971   Mean   :1885600   Mean   :1793915
##  3rd Qu.:1986   3rd Qu.:2058524   3rd Qu.:1965538
##  Max.   :2002   Max.   :2186274   Max.   :2082052
```

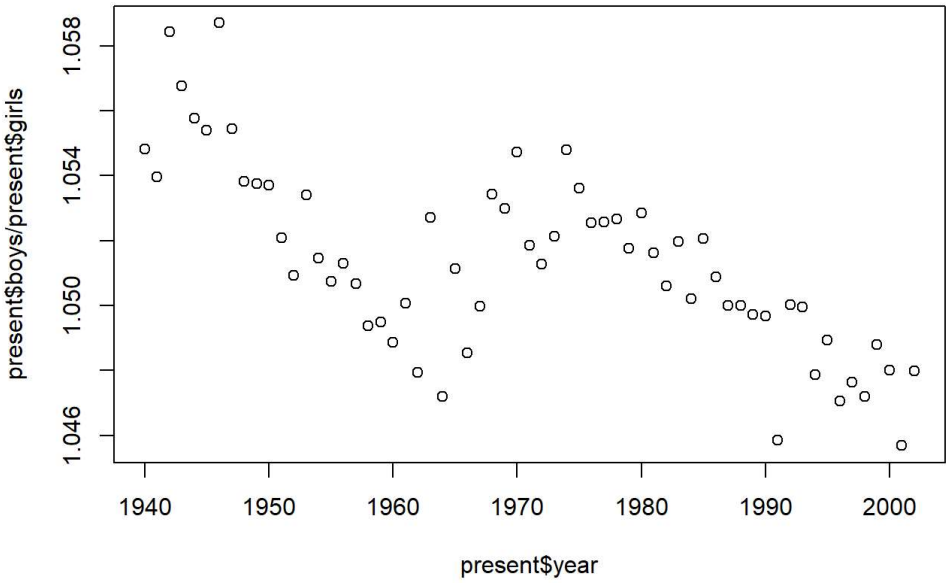
```
summary(arbuthnot)
```

```
##       year           boys          girls
##  Min.   :1629   Min.   :2890   Min.   :2722
##  1st Qu.:1649   1st Qu.:4759   1st Qu.:4457
##  Median :1670   Median :6073   Median :5718
##  Mean   :1670   Mean   :5907   Mean   :5535
##  3rd Qu.:1690   3rd Qu.:7576   3rd Qu.:7150
##  Max.   :1710   Max.   :8426   Max.   :7779
```

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 displays the boy-to-girl ration of present data set.It shows that the ratio were always greater than 1, and the trend was downhill.

```
plot(x=present$year, y=present$boys/present$girls)
```



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

```
present$total=present$boys + present$girls
which.max(present$total)
```

```
## [1] 22
```

```
present[22, ]
```

```
##    year    boys   girls   total
## 22 1961 2186274 2082052 4268326
```