Lab 0

Calvin Wong

9/2/2018

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

Exercise 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
```

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

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

There is a big dip from 1640 to 1660. However, since 1660, the amount of girls baptized has been steadily increasing year to year.

```
arbuthnot$girls / (arbuthnot$boys + arbuthnot$girls)
```

```
##  [1] 0.4729825 0.4784756 0.4812295 0.4789232 0.4840452 0.4890918 0.4911302
##  [8] 0.4836169 0.4865721 0.4802638 0.4713300 0.4914286 0.4873477 0.4734812
## [15] 0.4906482 0.4932132 0.4919659 0.4739634 0.4822695 0.4860941 0.4714163
## [22] 0.4850321 0.4677977 0.4745431 0.4807474 0.4802115 0.4781553 0.4797163
## [29] 0.4919970 0.4883306 0.4642738 0.4657868 0.4638058 0.4793892 0.4742518
## [36] 0.4846443 0.4871641 0.4800489 0.4865606 0.4779507 0.4725578 0.4767025
## [43] 0.4844924 0.4871448 0.4894493 0.4841786 0.4855202 0.4715703 0.4912878
## [50] 0.4787715 0.4916178 0.4903090 0.4891801 0.4939574 0.4857822 0.4847640
## [57] 0.4919212 0.4844835 0.4825095 0.4867699 0.4852075 0.4800473 0.4910323
## [64] 0.4904143 0.4936341 0.4876027 0.4803234 0.4864410 0.4906817 0.4750810
## [71] 0.4850615 0.4823417 0.4811732 0.4880474 0.4973459 0.4841786 0.4818210
## [78] 0.4825948 0.4784638 0.4805825 0.4848883 0.4882101
```

Here is the proportion of newborns that are girls. Suprisingly, girls consistently make up less than half the population.

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

```
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
```

On Your Own

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

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

```
str(present)
```

```
## 'data.frame':    63 obs. of  3 variables:
##  $ year : num  1940 1941 1942 1943 1944 ...
##  $ boys : num  1211684 1289734 1444365 1508959 1435301 ...
##  $ girls: num  1148715 1223693 1364631 1427901 1359499 ...
```

```
present$year
```

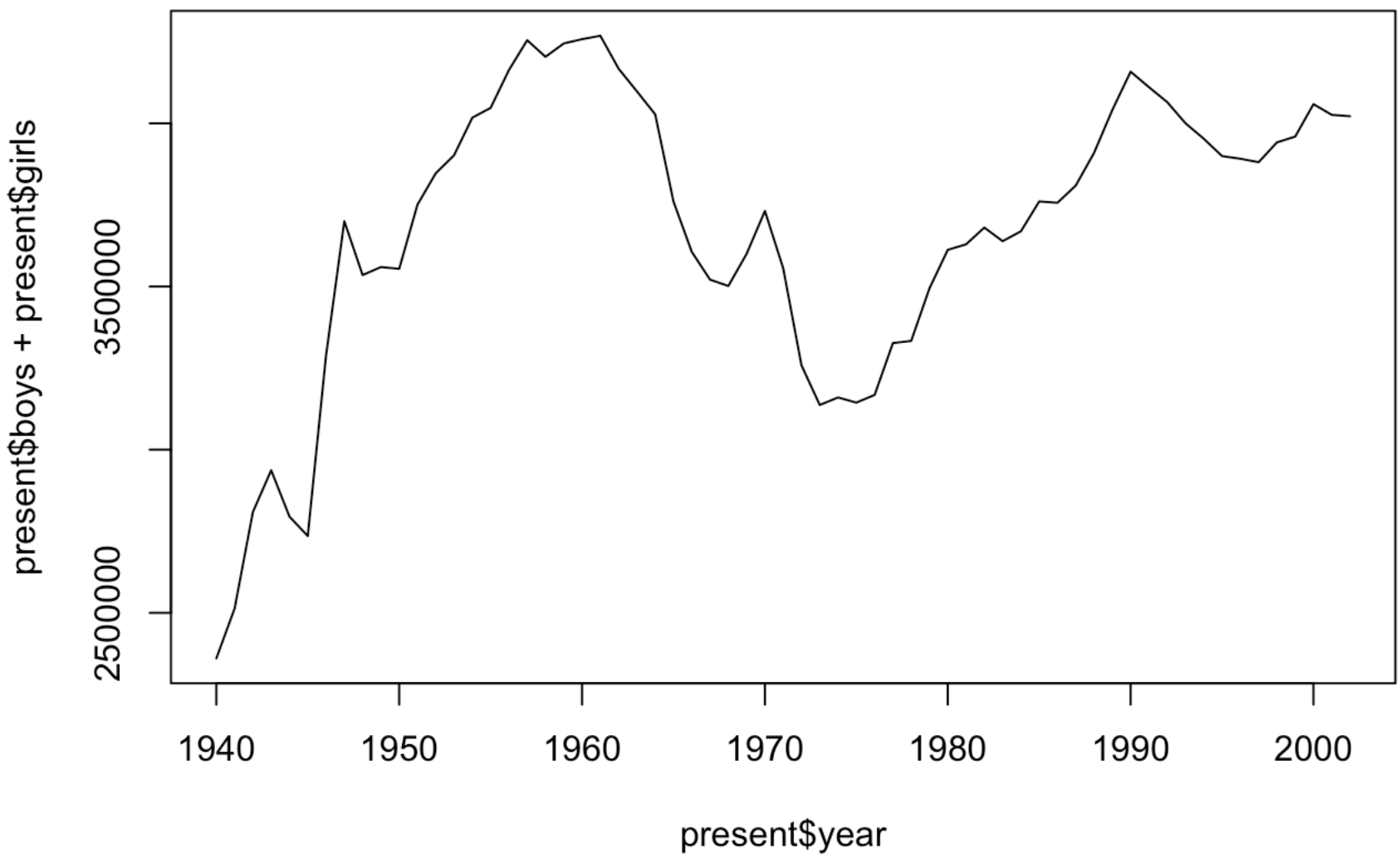
```
##  [1] 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953
## [15] 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967
## [29] 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981
## [43] 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992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

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

I created matching plot graphs. Based on the x-axis, can quickly see that they are not on a similar scale.

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

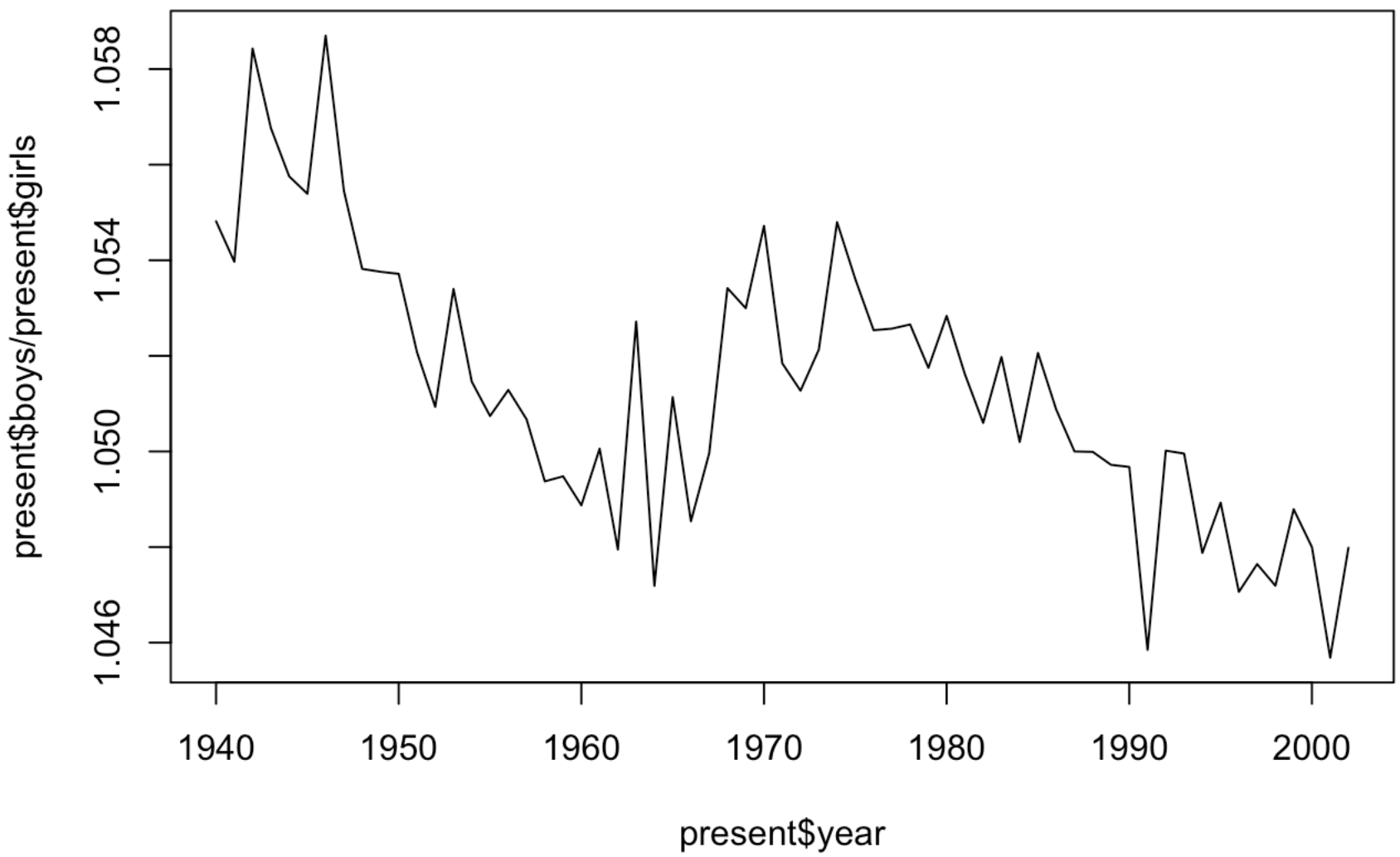

```
plot(present$year, present$boys + present$girls, type = "l")
```



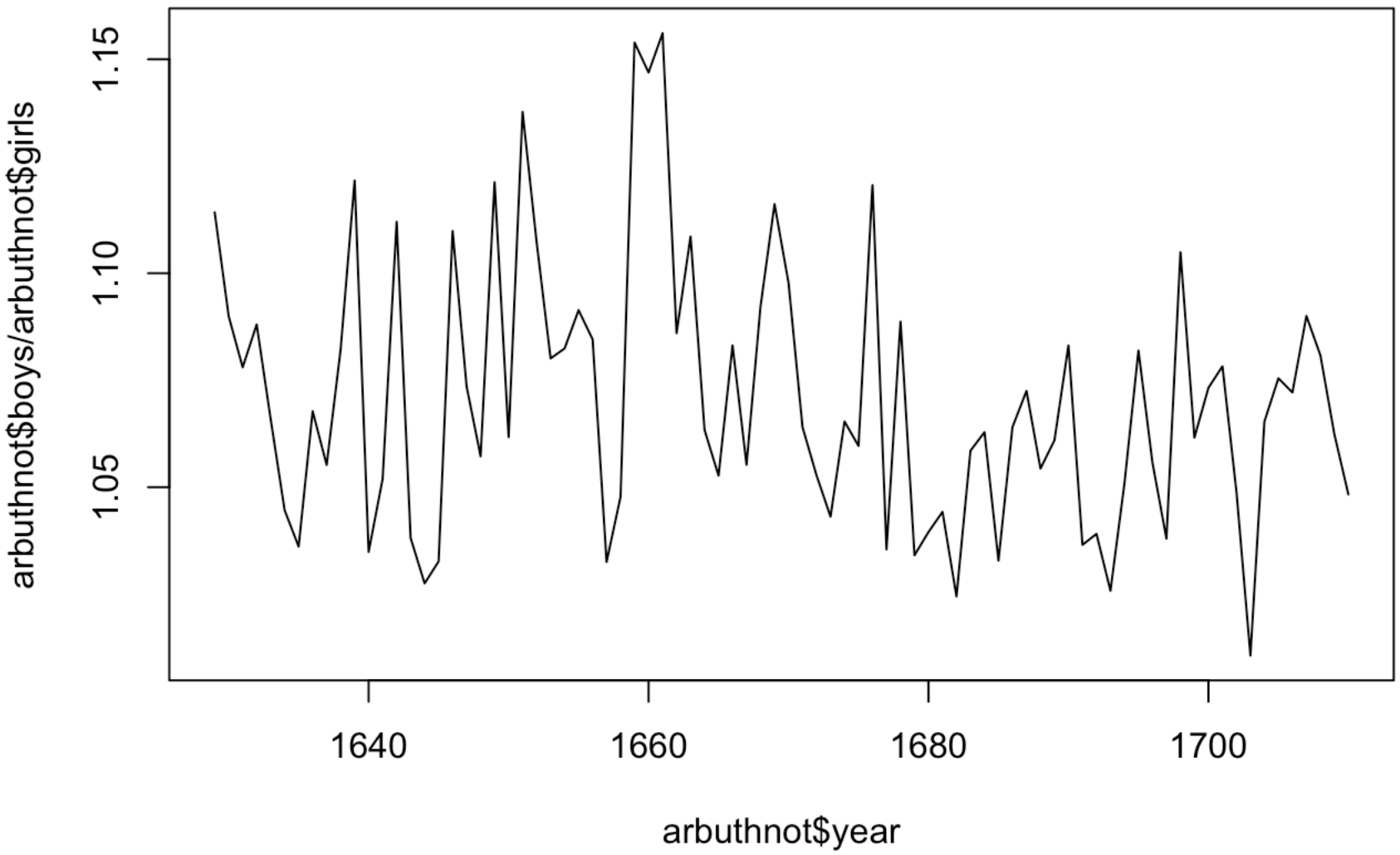
Make a plot that displays the boy-to-girl ratio for every year in the data set. What do you see? Does Arbuthnot's observation about boys being born in greater proportion than girls hold up in the U.S.? Include the plot in your response.

I have included the plots. It does seem that Arbuthnot's observation does hold up in the US. The ratio of boys born are consistently higher. However, that proportion is decreasing over time. It would be interesting to see where we are now.

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




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

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

```
## [1] 1961
```