

Lab0__Arbuthnot

Professor Soares

9/9/2021

The Arbuthnot Data

The Arbuthnot data set refers to Dr. John Arbuthnot, an 18th century physician, writer, and mathematician. He was interested in the ratio of newborn boys to newborn girls, so he gathered the baptism records for children born in London for every year from 1629 to 1710.

First we import the Arbuthnot data:

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

We can then determine its dimensions and identify the names of the columns in the data frame `arbuthnot`:

```
dim(arbuthnot)
```

```
## [1] 82  3
```

```
names(arbuthnot)
```

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

It appears there are 82 cases and for each case, we record 3 variables: year, number of boys baptized, and number of girls baptized. These baptismal records are surrogates for the number of births.

Exploration

To extract information from the data frame like the number of boys born per year, we either have to type `arbuthnot$boys`, or we can run the `attach` command on the data frame and then simply use the column names:

```
attach(arbuthnot)
```

```
boys
```

```
## [1] 5218 4858 4422 4994 5158 5035 5106 4917 4703 5359 5366 5518 5470 5460 4793
## [16] 4107 4047 3768 3796 3363 3079 2890 3231 3220 3196 3441 3655 3668 3396 3157
## [31] 3209 3724 4748 5216 5411 6041 5114 4678 5616 6073 6506 6278 6449 6443 6073
## [46] 6113 6058 6552 6423 6568 6247 6548 6822 6909 7577 7575 7484 7575 7737 7487
## [61] 7604 7909 7662 7602 7676 6985 7263 7632 8062 8426 7911 7578 8102 8031 7765
## [76] 6113 8366 7952 8379 8239 7840 7640
```

To extract the counts for the girls, we take a similar approach:

```
girls
```

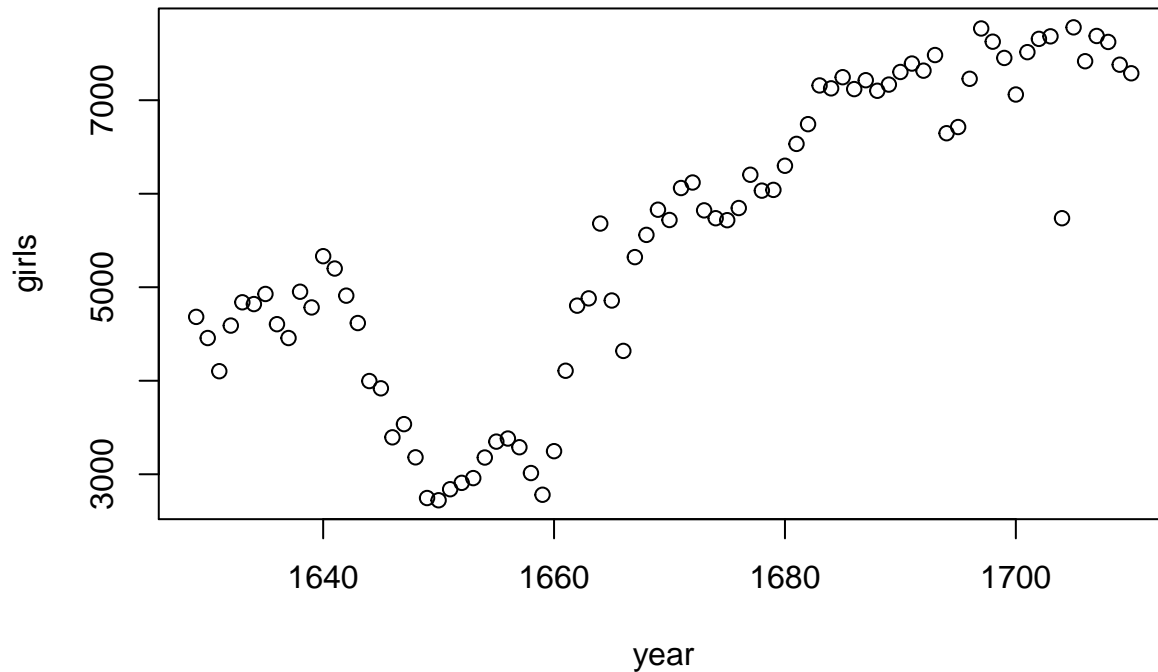
```
## [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910 4617
## [16] 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382 3289 3013
## [31] 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719 6061 6120 5822
## [46] 5738 5717 5847 6203 6033 6041 6299 6533 6744 7158 7127 7246 7119 7214 7101
## [61] 7167 7302 7392 7316 7483 6647 6713 7229 7767 7626 7452 7061 7514 7656 7683
```

```
## [76] 5738 7779 7417 7687 7623 7380 7288
```

Plots

R has some powerful functions for making graphics. We can create a simple plot of the number of girls baptized per year with the command:

```
plot(year,girls,type = "p")
```



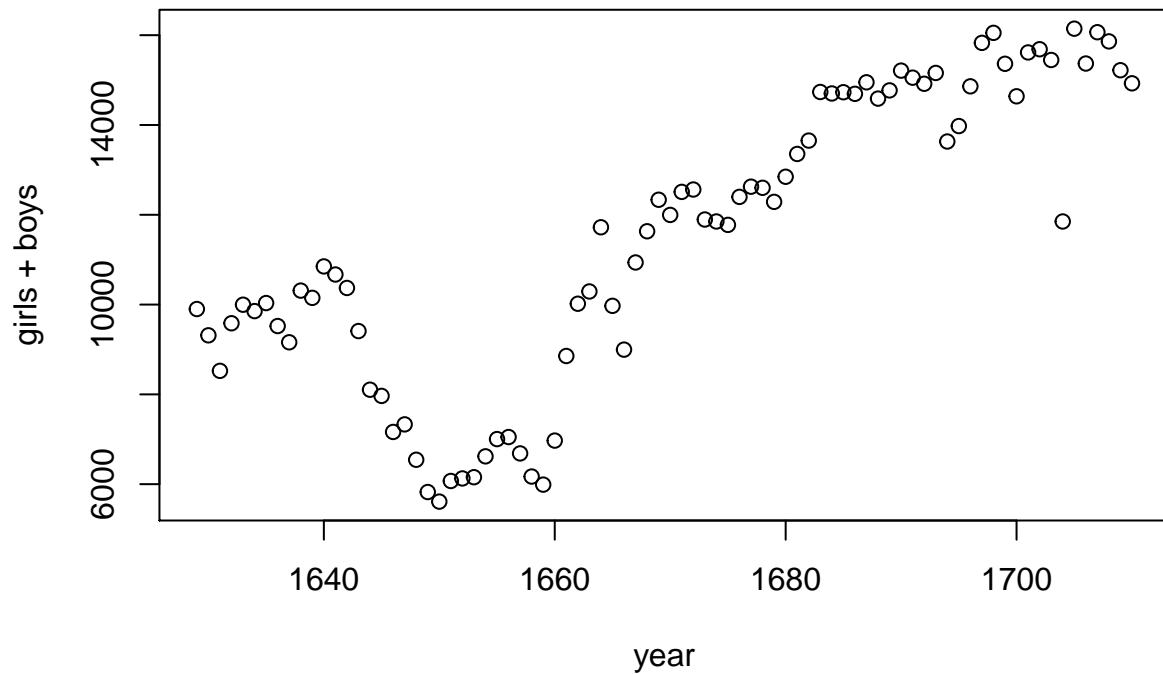
and we can use `type = "l"` to plot using lines. More details about plot can be found by typing `?plot` in the command window.

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

There is a general increase in the number of girl births, except for a dip between 1640 and 1660.

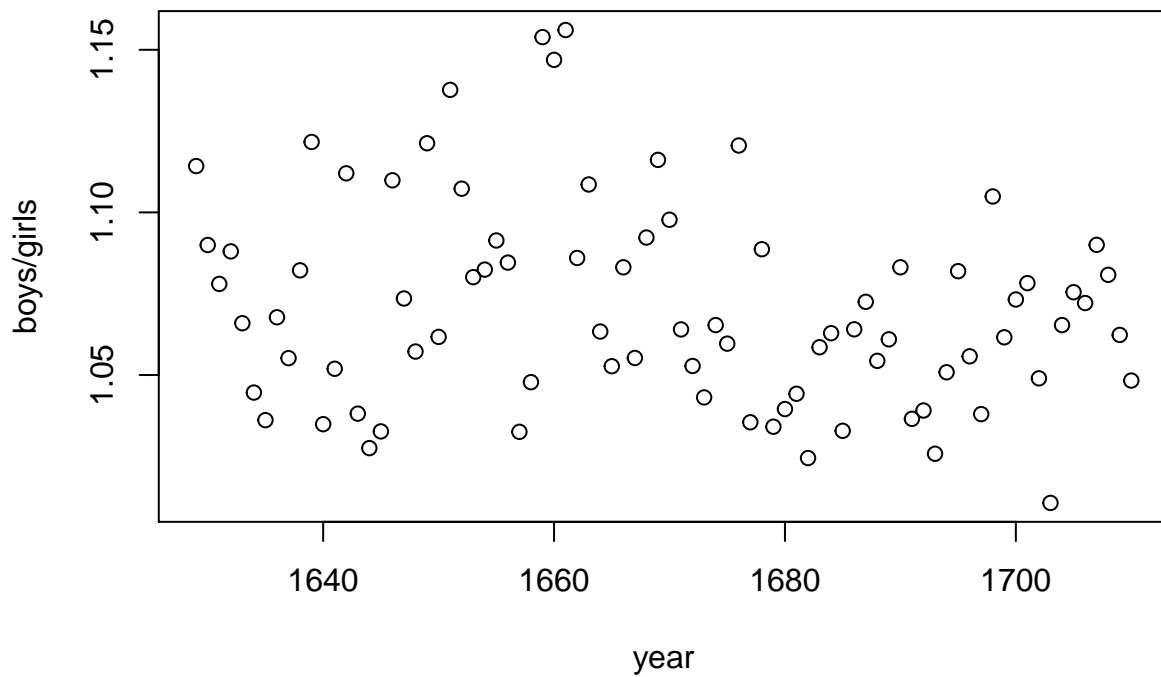
We can make a plot of the total number of baptisms per year with the command:

```
plot(year,girls+boys,type = "p")
```



and a plot of the proportion of boys to girls:

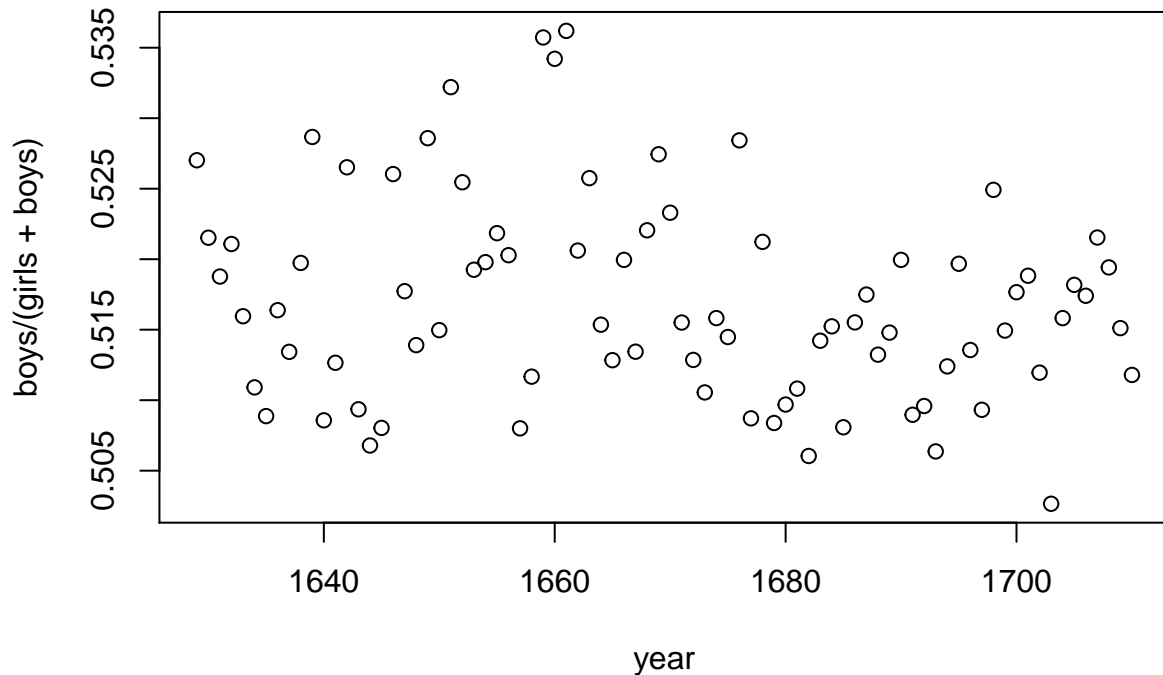
```
plot(year,boys/girls,type = "p")
```



Question: What conclusion can be drawn from this plot?

We can make a plot of the proportion of boy baptisms per year with the command:

```
plot(year, boys/(girls+boys), type = "p")
```



Question: What can be inferred from this plot?

Other Operations

Finally, in addition to simple mathematical operators like subtraction and division, you can ask R to make comparisons like greater than, $>$, less than, $<$, and equality, $==$. For example, we can ask if boys outnumber girls in each year with the expression:

```
(boys > girls)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [31] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [46] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [61] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

which shows us that boy baptisms always exceeds baptisms for girls.

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

The data are stored in a data frame called **present**.

- What years are included in this data set? What are the dimensions of the data frame and what are the variable or column names?
- How do these counts compare to Arbuthnot's? Are they on a similar scale?
- 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.
- 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> 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.