

# STAT 100B Lab 1

Wesley Chang

Summer 2020 Session B

## Lab Exercises

### Setup code for Lab Assignment

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

*Get sourcefile from openintro.org*

### Exercise 1

*What command would you use to extract just the counts of girls baptized?*

**Answer**

```
arbuthnot$girls
```

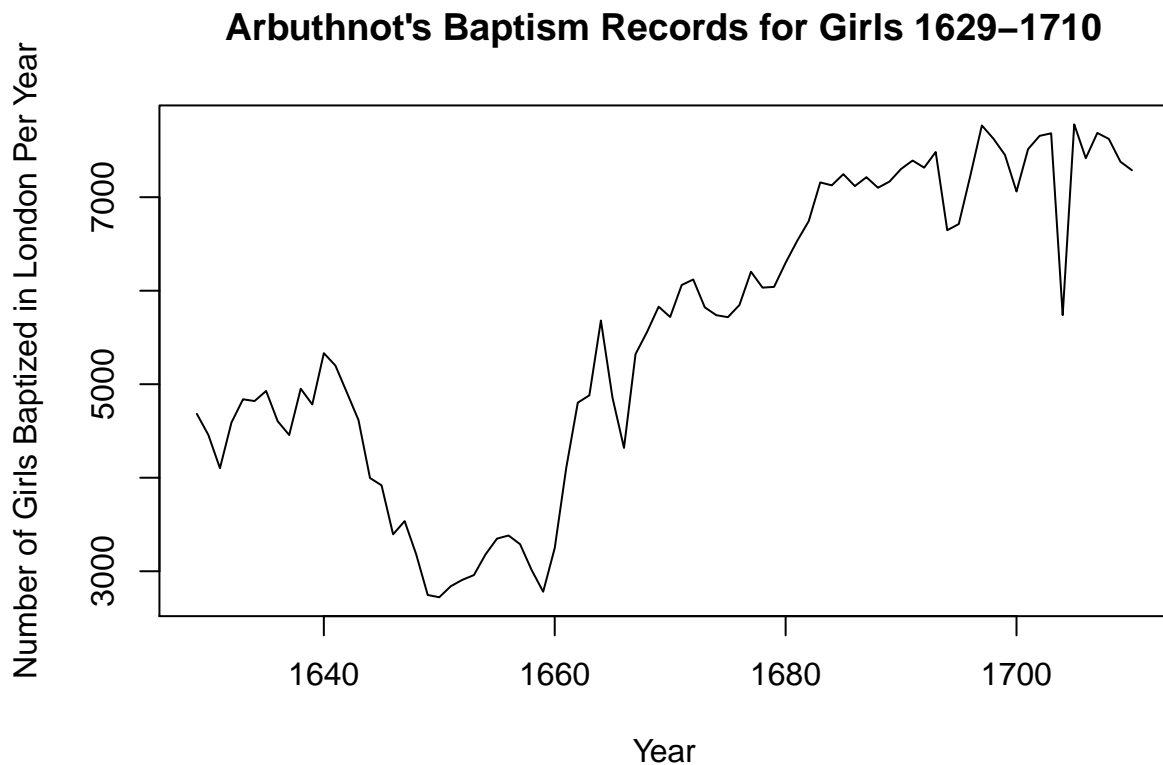
```
## [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910 4617
## [16] 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382 3289 3013
## [31] 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719 6061 6120 5822
## [46] 5738 5717 5847 6203 6033 6041 6299 6533 6744 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
```

### Exercise 2

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

**Code to display plot of number of girls baptized vs year**

```
plot(x = arbuthnot$year,
     y = arbuthnot$girls,
     type = "l",
     xlab = "Year",
     ylab = "Number of Girls Baptized in London Per Year",
     main = "Arbuthnot's Baptism Records for Girls 1629-1710"
)
```



#### Answer

There appears to be an apparent trend between the number of girls baptized over the years. As time passes, the number girls baptized per year tends to increase with each passing year.

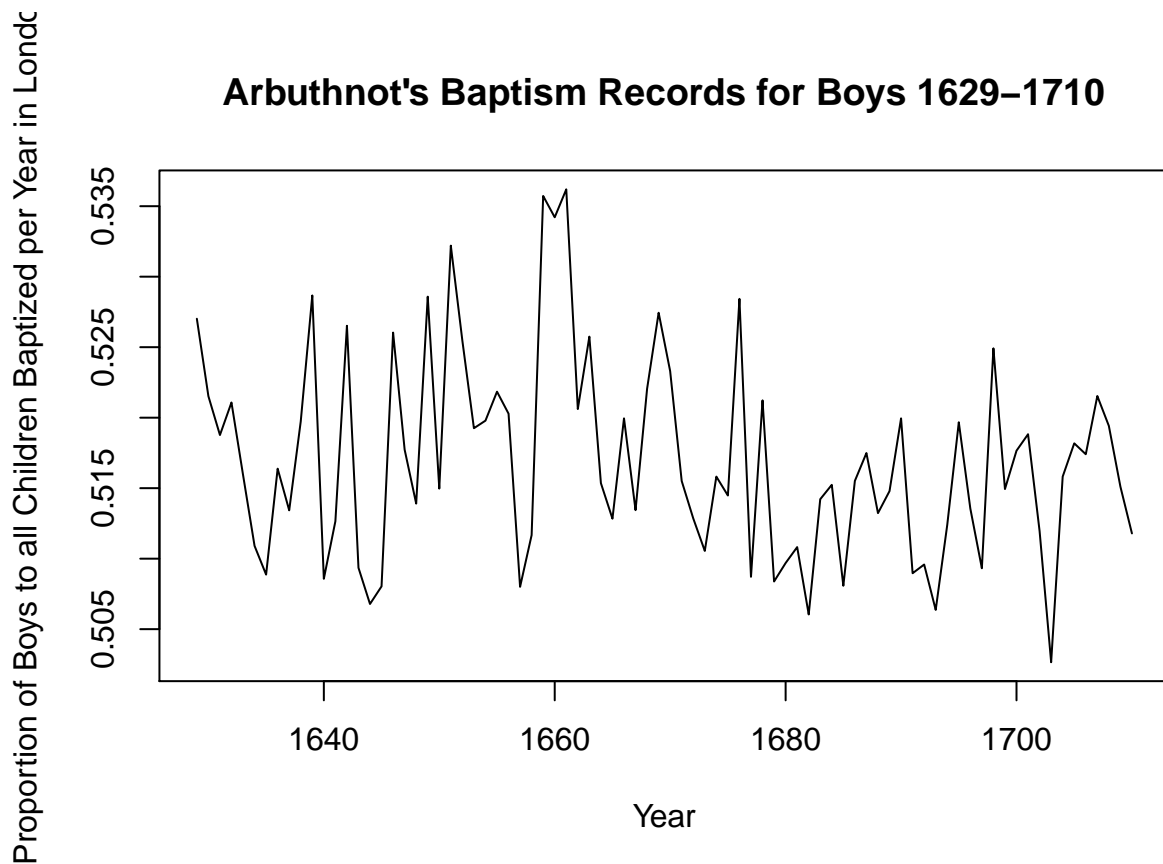
#### Exercise 3

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

```
# obtain ratio of boys to all children for each year and store in a data frame "proportionboys"
arbuthnot$proportionboys <- arbuthnot$boys / (arbuthnot$boys + arbuthnot$girls)
```

```
# generate plot of proportion of boys over time
```

```
plot(
  x = arbuthnot$year,
  y = arbuthnot$proportionboys,
  xlab = "Year",
  ylab = "Proportion of Boys to all Children Baptized per Year in London",
  main = "Arbuthnot's Baptism Records for Boys 1629-1710",
  type = "l"
)
```



## On Your Own

Set up Source for present.R

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

### Question 1

What are the dimensions of the data frame?

```
# use dim() command to view dimensions
dim(present)
```

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

Answer

The dimensions are 63 rows and 3 columns

### Question 2

What are the variable or column names?

```
# Command to get names of the variables
names(present)
```

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

Answer

The variable names are “year”, “boys”, and “girls”

### Question 3

*What years are included in this data set?*

```
# attach variables to column names
attach(present)
year
```

```
## [1] 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954
## [16] 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969
## [31] 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984
## [46] 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999
## [61] 2000 2001 2002
```

### Answer

The data set includes all years from 1940 to 2002, for a total of 63 years

### Question 4

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

```
## years

# get range of years for Arbuthnot's data and the present data
arbuthrange <- range(arbuthnot$year)
presentrange <- range(year)

# difference in years for both data sets
arbuthrange[2] - arbuthrange[1]
```

```
## [1] 81
presentrange[2] - presentrange[1]
```

```
## [1] 62
## boys
range(arbuthnot$boys)
```

```
## [1] 2890 8426
range(boys)
```

```
## [1] 1211684 2186274
## girls
range(arbuthnot$girls)
```

```
## [1] 2722 7779
range(girls)
```

```
## [1] 1148715 2082052
```

### Answer

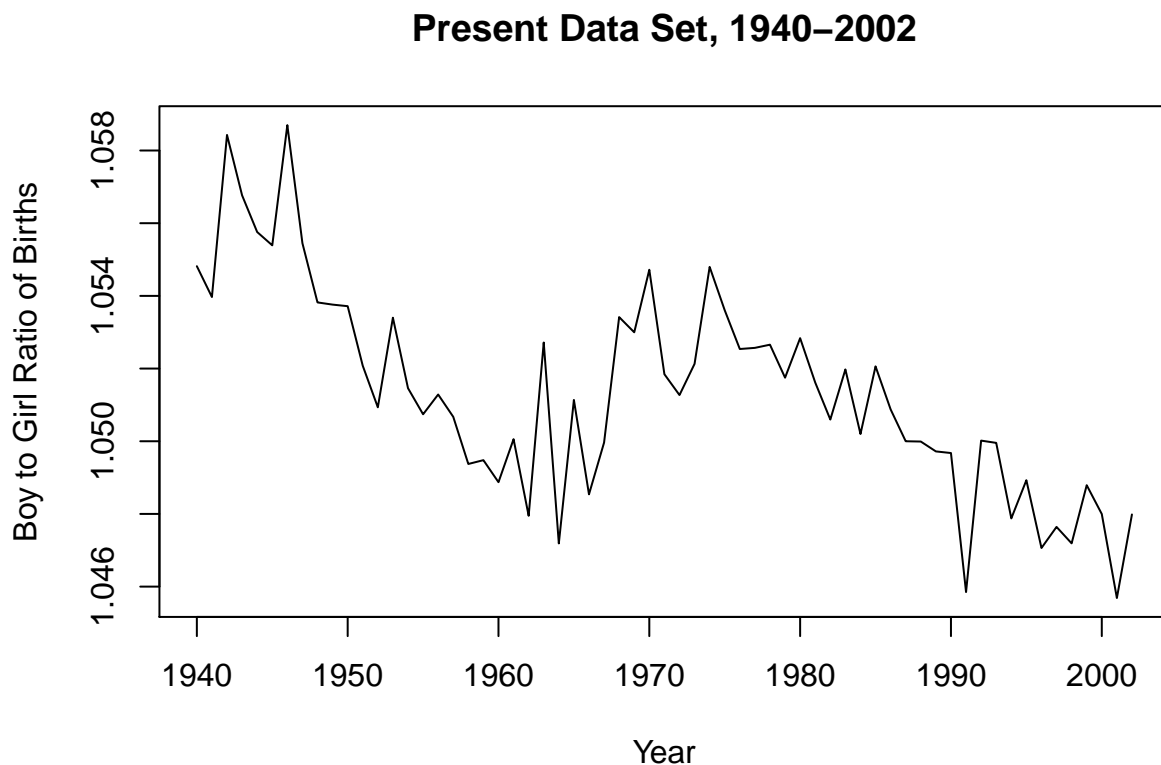
The data for years between both data sets is similar, with the total number of years being 82 for Arbuthnot's data and 63 for the present data. However, for amount of boys and girls baptized per year, the present data is significantly larger than Arbuthnot's data, being measured in millions rather than thousands.

### Question 5

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

```
# create variable for boy-to-girl ratio
boygirlratio <- boys / girls

# create plot of boy-to-girl ratio by year
plot(
  x = year,
  y = boygirlratio,
  main = "Present Data Set, 1940-2002",
  xlab = "Year",
  ylab = "Boy to Girl Ratio of Births",
  type = "l"
)
```



### Answer

Arbutnot's observation about boys being born in greater proportion than girls in the US does hold up in the earlier years of the data set, as shown by the boy to girl ratios. However, from 1940 to 2002, there is a clear trend where ratio decreases.

### Question 6

In what year did we see the most total number of births in the US?

```
# create a variable for the total number of births  
presenttotal <- boys + girls  
# find maximum value of boys + girls  
max(presenttotal)
```

```
## [1] 4268326
```

```
# find corresponding year  
year[presenttotal == 4268326]
```

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

**Answer**

1961 was the year that we saw the most total number of births in the US