Lab 1: Intro to R

Donald Butler

2021-08-31

```
library(tidyverse)
library(openintro)
```

```
## Warning: package 'openintro' was built under R version 4.1.1
```

Exercise 1

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

```
arbuthnot$girls
```

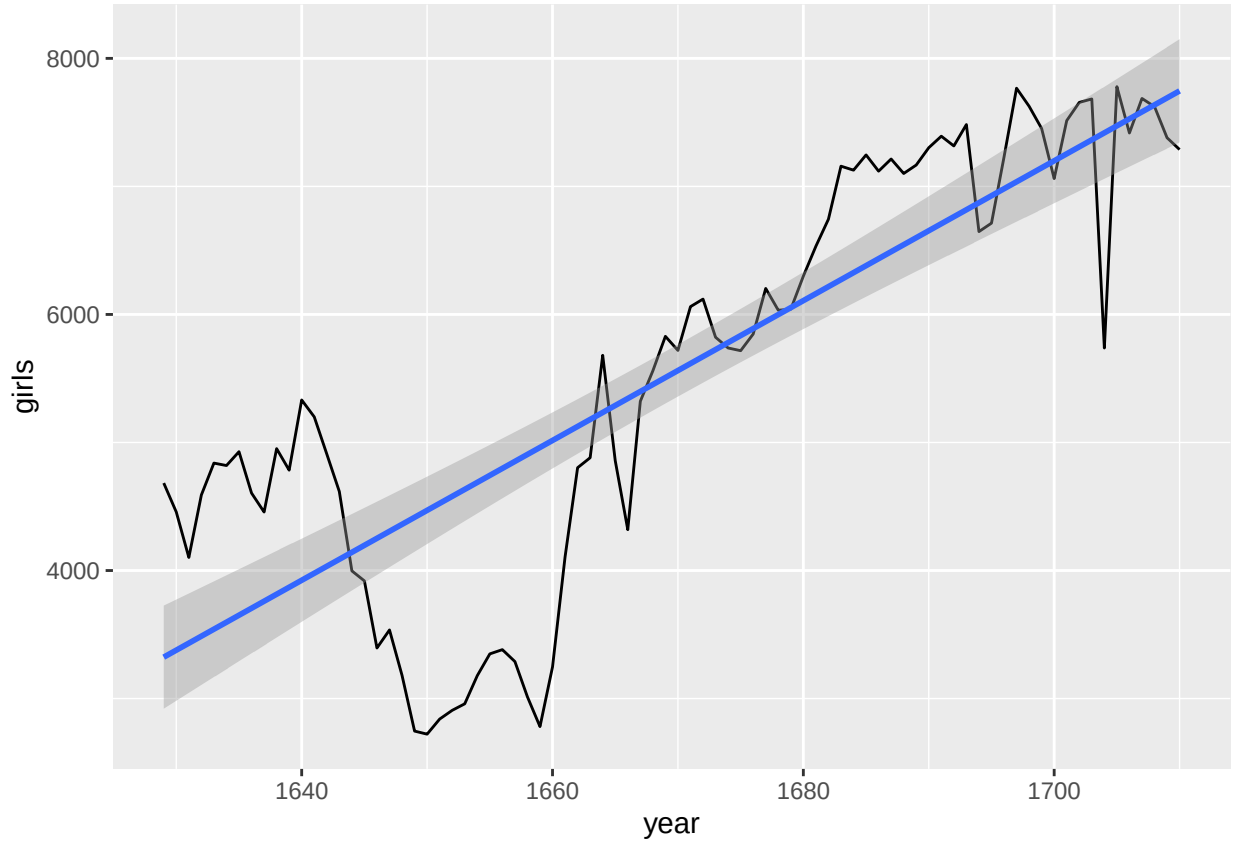
```
##  [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910 4617
## [16] 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382 3289 3013
## [31] 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719 6061 6120 5822
## [46] 5738 5717 5847 6203 6033 6041 6299 6533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

The trend in the number of girls baptized appears to be linear. The drop in baptismal rates for the period 1640-1660 is mostly likely caused by the English Civil Wars (1642-1651).

```
ggplot(arbuthnot,aes(year,girls))+geom_line()+geom_smooth(method = 'lm',formula = y ~ x)
```



Exercise 3

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

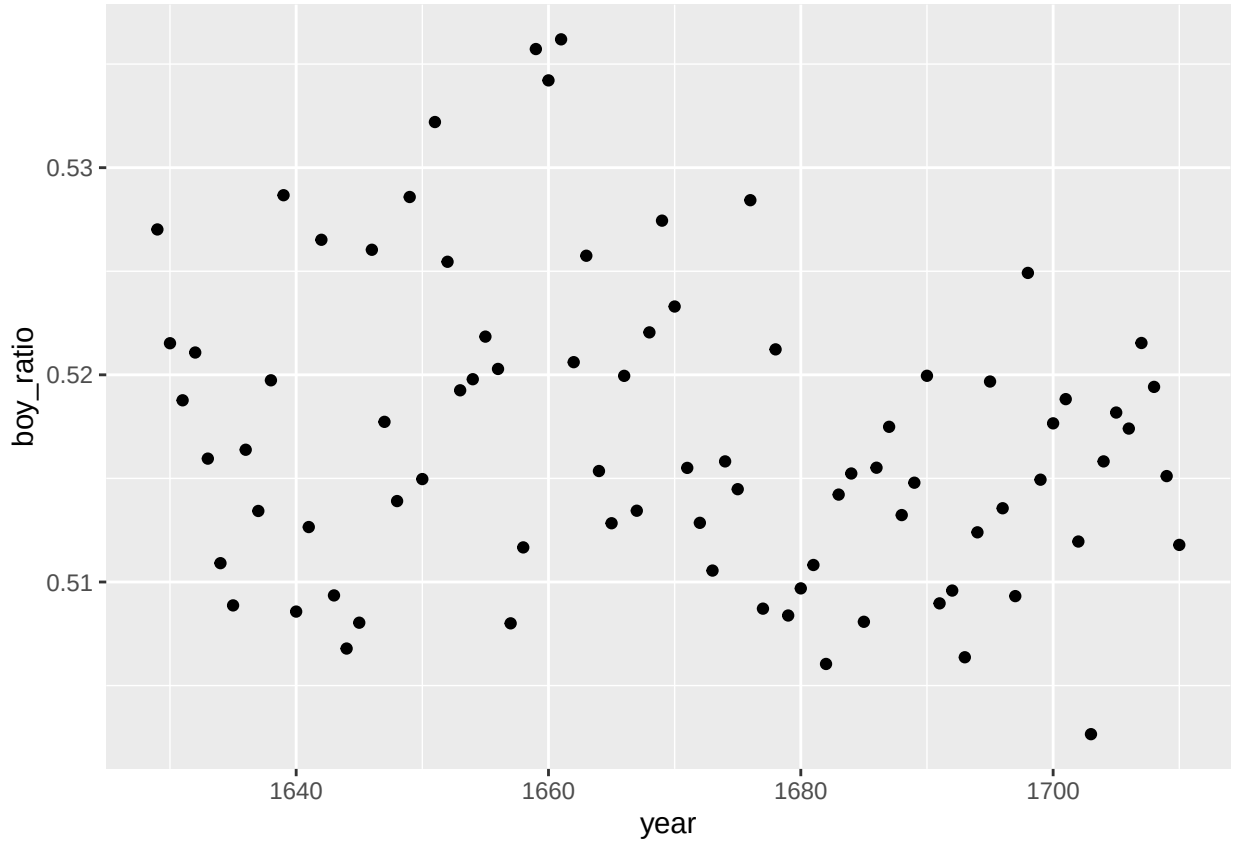
I would have expected a 50/50 split between boys and girls, but the number of boys baptized were slightly higher than girls.

```
arbuthnot <- arbuthnot %>% mutate(total = boys + girls)
arbuthnot <- arbuthnot %>% mutate(boy_to_girl_ratio = boys / girls)
arbuthnot <- arbuthnot %>% mutate(boy_ratio = boys / total)

mean(arbuthnot$boy_ratio)
```

```
## [1] 0.5169751
```

```
ggplot(arbuthnot,aes(year,boy_ratio))+geom_point()
```



Exercise 4

What years are included in this data set? What are the dimensions of the data frame? What are the variable (column) names?

The present data set contains 63 observations in years 1940 through 2002 of 3 variables (year, boys, and girls).

```
present <- present %>% mutate(total = boys + girls)
present <- present %>% mutate(boy_to_girl_ratio = boys / girls)
present <- present %>% mutate(boy_ratio = boys / total)

summary(present)
```

```
##       year           boys             girls            total
##  Min.   :1940   Min.   :1211684   Min.   :1148715   Min.   :2360399
##  1st Qu.:1956   1st Qu.:1799857   1st Qu.:1711405   1st Qu.:3511262
##  Median :1971   Median :1924868   Median :1831679   Median :3756547
##  Mean   :1971   Mean   :1885600   Mean   :1793915   Mean   :3679515
##  3rd Qu.:1986   3rd Qu.:2058524   3rd Qu.:1965538   3rd Qu.:4023830
##  Max.   :2002   Max.   :2186274   Max.   :2082052   Max.   :4268326
##  boy_to_girl_ratio   boy_ratio
##  Min.   :1.046     Min.   :0.5112
##  1st Qu.:1.050     1st Qu.:0.5121
##  Median :1.051     Median :0.5125
```

```
##  Mean   :1.051     Mean   :0.5125
##  3rd Qu.:1.053     3rd Qu.:0.5130
##  Max.   :1.059     Max.   :0.5143
```

Exercise 5

How do these counts compare to Arbuthnot's? Are they of a similar magnitude?

The present dataset contains birth rates in the United States which is significantly larger than London. THe present dataset contains birth rates over 300 times Arbuthnot's data.

```
summary(arbuthnot)
```

```
##       year           boys          girls          total       boy_to_girl_ratio
##  Min.   :1629   Min.   :2890   Min.   :2722   Min.   : 5612   Min.   :1.011
##  1st Qu.:1649   1st Qu.:4759   1st Qu.:4457   1st Qu.: 9199   1st Qu.:1.048
##  Median :1670   Median :6073   Median :5718   Median :11813   Median :1.065
##  Mean   :1670   Mean   :5907   Mean   :5535   Mean   :11442   Mean   :1.071
##  3rd Qu.:1690   3rd Qu.:7576   3rd Qu.:7150   3rd Qu.:14723   3rd Qu.:1.088
##  Max.   :1710   Max.   :8426   Max.   :7779   Max.   :16145   Max.   :1.156
##    boy_ratio
##  Min.   :0.5027
##  1st Qu.:0.5118
##  Median :0.5157
##  Mean   :0.5170
##  3rd Qu.:0.5210
##  Max.   :0.5362
```

```
mean(present$total)/mean(arbuthnot$total)
```

```
## [1] 321.5869
```

Exercise 6

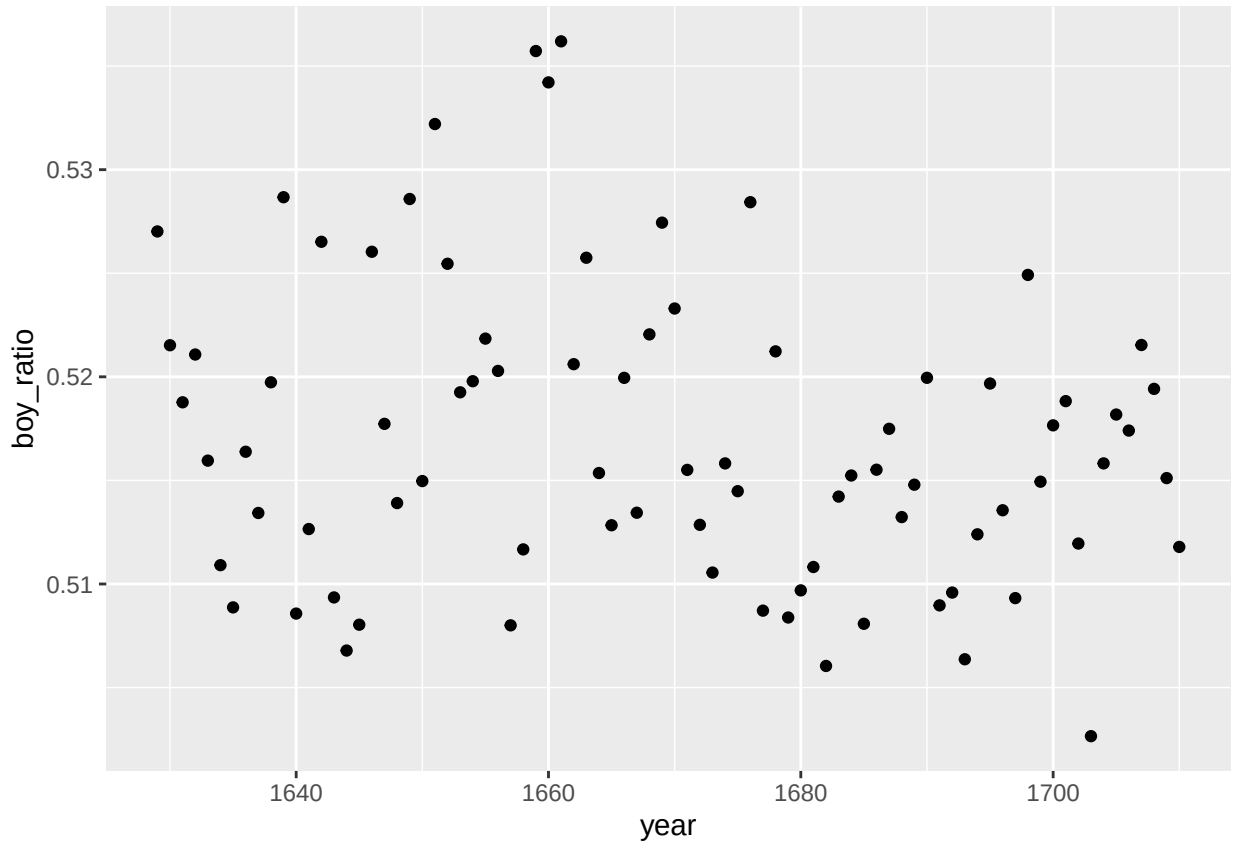
Make a plot that displays the proportion of boys born over time. What do you see? Does Arbuthnot's observation about boys being born in greater proportion than girls hold up in the U.S.?

THe present data set shows the same increased birth rates of boys over girls that was observed in the Arbuthnot data set.

```
mean(present$boy_ratio)
```

```
## [1] 0.512516
```

```
ggplot(arbuthnot,aes(year,boy_ratio))+geom_point()
```



Exercise 7

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

The most total births in the present data set was 1961.

```
filter(present,total == max(total))
```

```
## # A tibble: 1 x 6
##    year    boys   girls   total boy_to_girl_ratio boy_ratio
##   <dbl>   <dbl>   <dbl>   <dbl>             <dbl>     <dbl>
## 1  1961 2186274 2082052 4268326              1.05     0.512
```