

Lab - 1

Irene Jacob

30/08/2020

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

Exercise 1

Command to extract the count of girls baptized

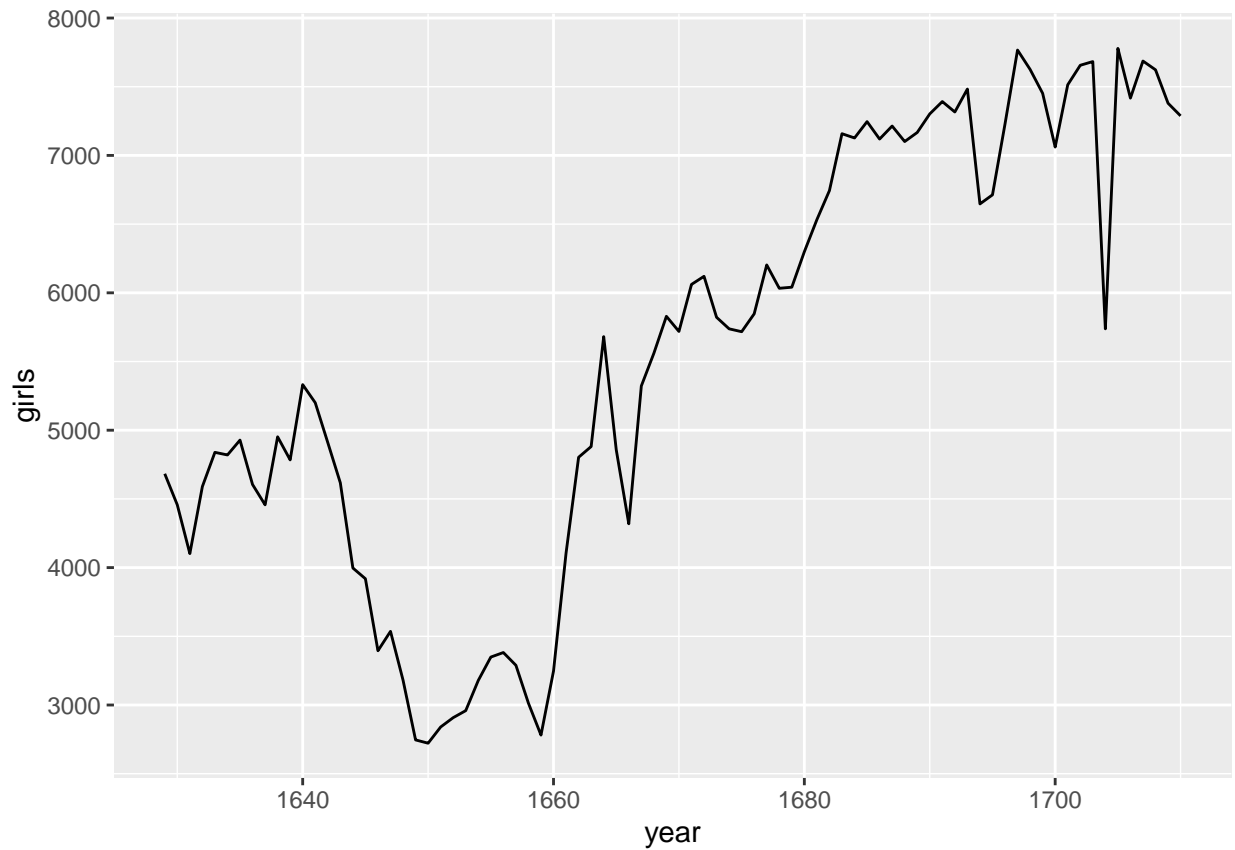
```
arbuthnot$girls
```

```
## [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

During the period from 1629 till 1640 the number of girls baptized were more than in the period from 1640 till 1660 but after 1660 the number kept increasing.

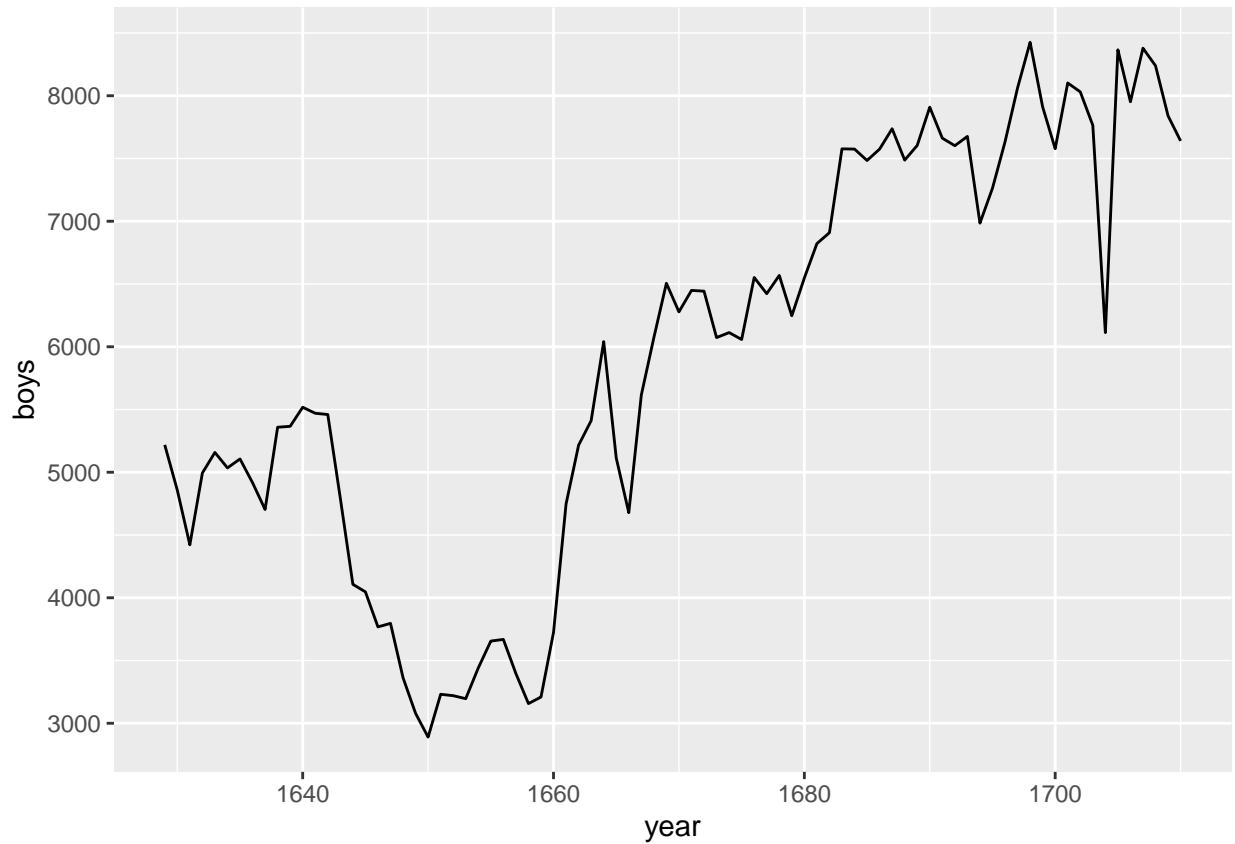
```
ggplot(data = arbutnot, aes(x = year, y = girls)) +  
  geom_line()
```



Exercise 3

A graph showing proportion of boys over time. Here we see that during the period from 1629 till 1640 the number of boys baptized were more than in the period from 1640 till 1660 but after 1660 the number kept increasing.

```
ggplot(data = arbutnot, aes(x = year, y = boys)) +  
  geom_line()
```



Exercise 4

The years included in this data set is from 1940 till 2002. There are 63 dimensions and 3 variables in this data set. The variables are year, boys and girls.

```
glimpse(present)
```

```
## Rows: 63
## Columns: 3
## $ year  <dbl> 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1...
## $ boys  <dbl> 1211684, 1289734, 1444365, 1508959, 1435301, 1404587, 1691220...
## $ girls <dbl> 1148715, 1223693, 1364631, 1427901, 1359499, 1330869, 1597452...
```

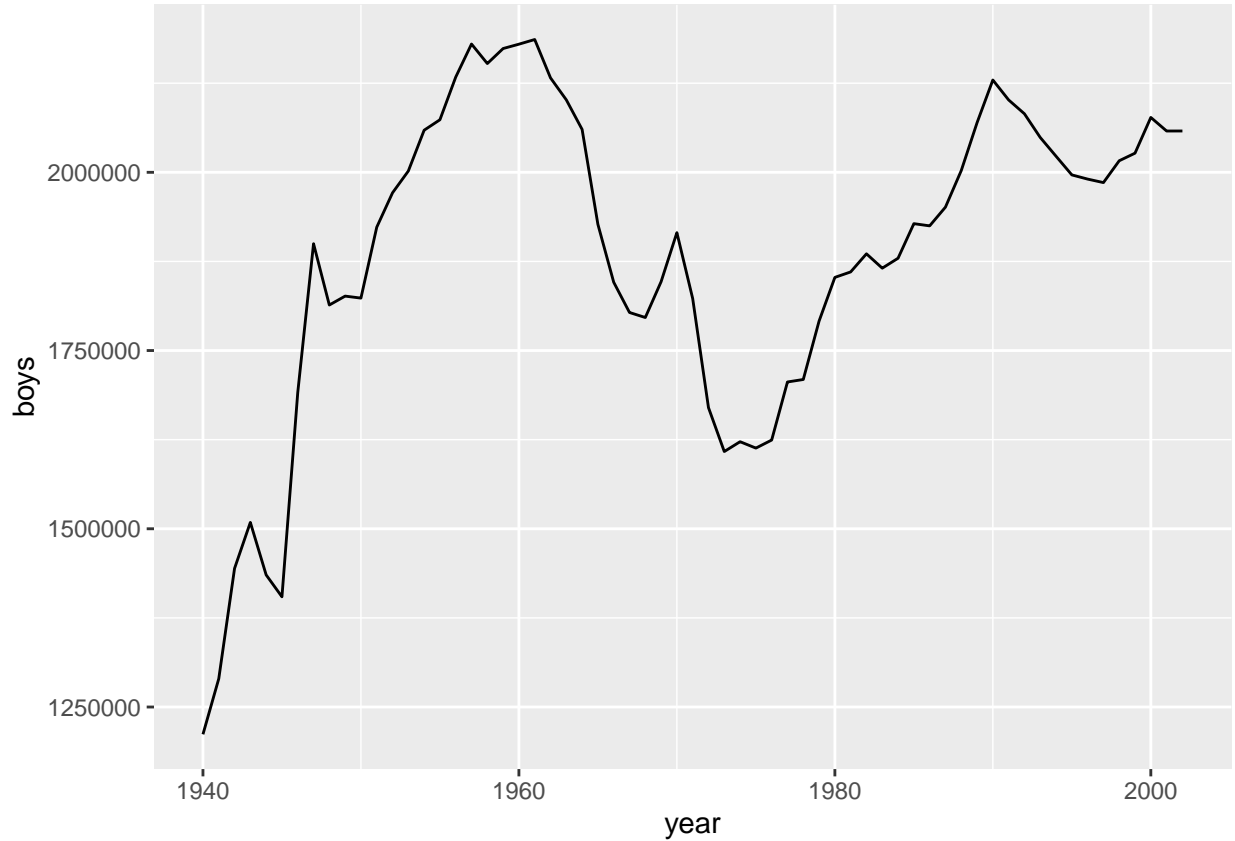
Exercise 5

These counts was very high compared to the counts in arbuthnot.

Exercise 6

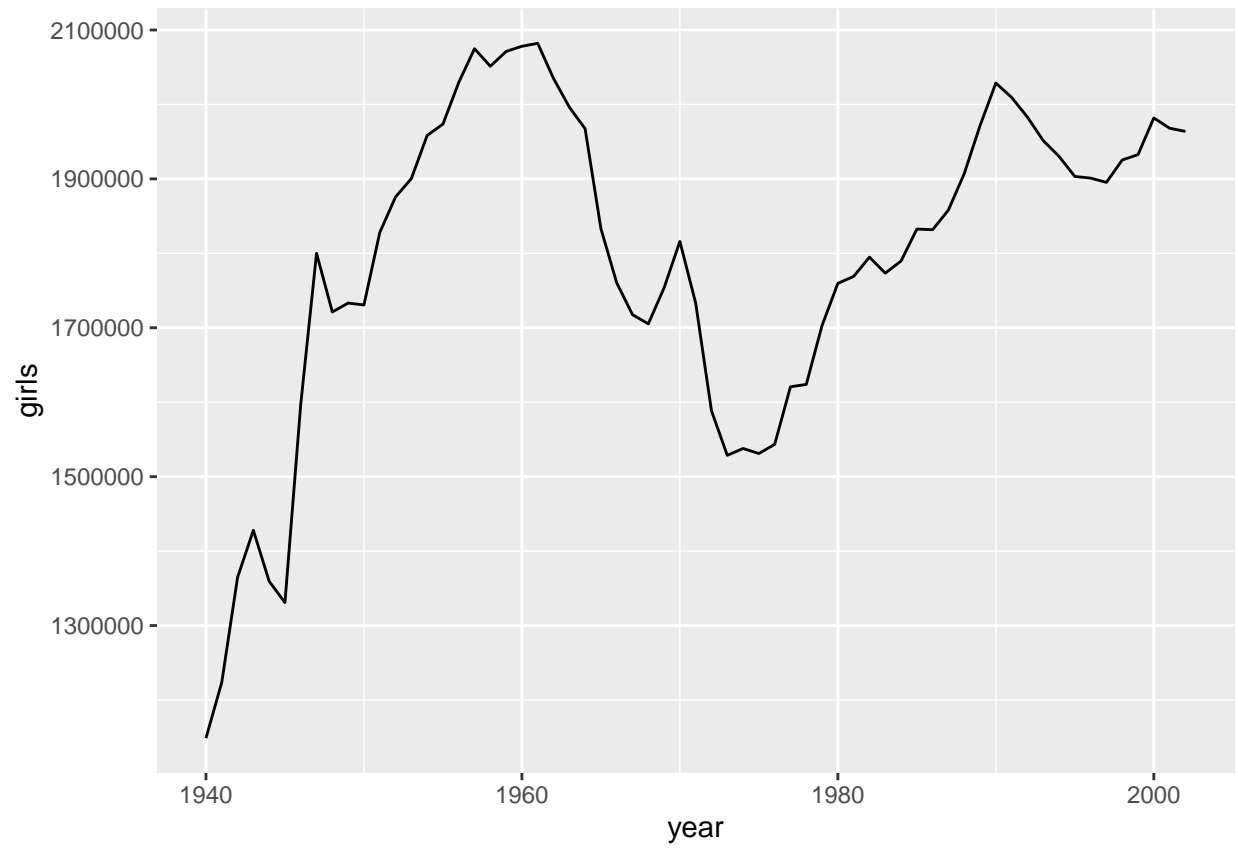
Arbuthnot's conclusion that boys are born in higher proportions when compared to girls is true in the US.

```
ggplot(data = present, aes(x = year, y = boys)) +  
  geom_line()
```



For reference check the plot for girls as well below:

```
ggplot(data = present, aes(x = year, y = girls)) +  
  geom_line()
```



Exercise 7

In the year 1961 has the maximum number of births in the US.

```
present <- present %>%  
  mutate(total = boys + girls)  
present %>%  
  arrange(desc(total))
```

```
## # A tibble: 63 x 4  
##   year    boys  girls  total  
##   <dbl>  <dbl>  <dbl>  <dbl>  
## 1  1961 2186274 2082052 4268326  
## 2  1960 2179708 2078142 4257850  
## 3  1957 2179960 2074824 4254784  
## 4  1959 2173638 2071158 4244796  
## 5  1958 2152546 2051266 4203812  
## 6  1962 2132466 2034896 4167362  
## 7  1956 2133588 2029502 4163090  
## 8  1990 2129495 2028717 4158212  
## 9  1991 2101518 2009389 4110907  
## 10 1963 2101632 1996388 4098020  
## # ... with 53 more rows
```

```
present %>%  
  summarize(max = max(total))
```

```
## # A tibble: 1 x 1  
##       max  
##   <dbl>  
## 1 4268326
```