Lab 1: Intro to R

Noah Collin

2021-09-05

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

#data('arbuthnot', package='openintro')

#glimpse(arbuthnot)

Exercise 1

```
arbuthnot$girls
```

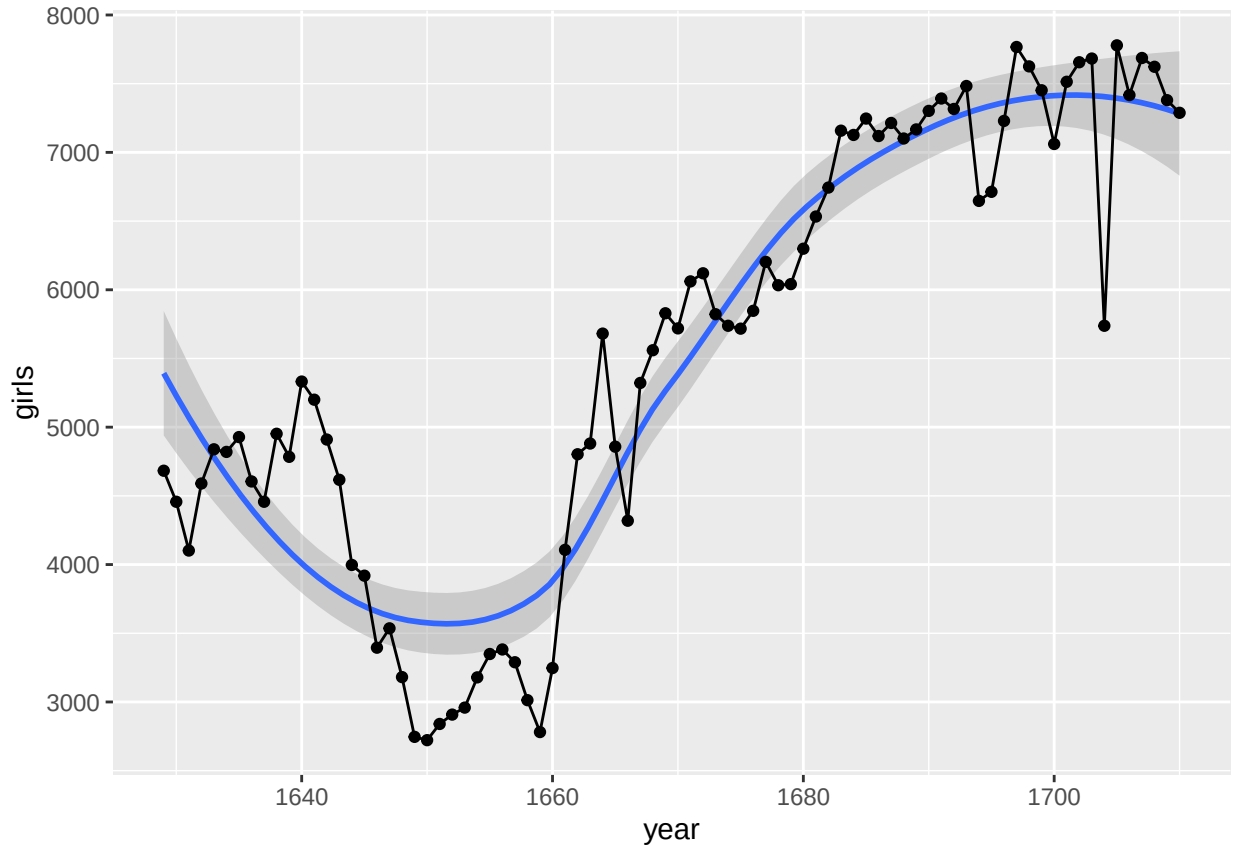
```
##  [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910 4617
## [16] 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382 3289 3013
## [31] 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719 6061 6120 5822
## [46] 5738 5717 5847 6203 6033 6041 6299 6533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

It's my opinion there's not a clear trend or even direction for the population of girls born in this time period. A broad summary would indicate that the number of girls born per year increased.

```
ggplot(data = arbuthnot, aes (x = year, y = girls)) +
  geom_smooth() +
  geom_point() +
  geom_line()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



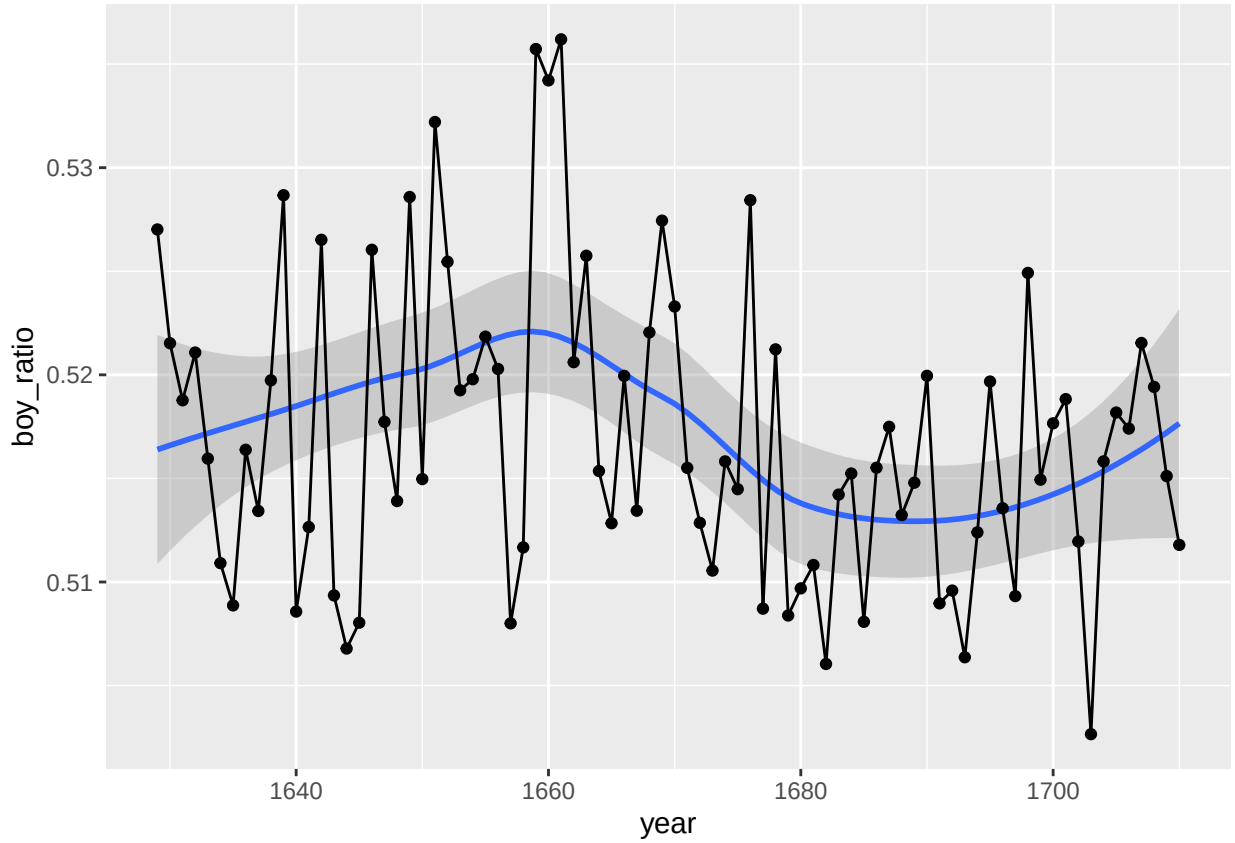
Exercise 3

Surprisingly, the ratio of boys baptized exceeds 50% for every data point observed.

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

ggplot(data = arbuthnot, aes (x = year, y = boy_ratio)) +
  geom_smooth() +
  geom_point() +
  geom_line()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Exercise 4

The "present" dataset contains data from 1940 - 2002. That dataset has 63 rows and 3 columns. The column names are "year", "boys", and "girls".

```
present %>%
  summarize(min = min(year), max = max(year))
```

```
## # A tibble: 1 x 2
##     min   max
##   <dbl> <dbl>
## 1  1940  2002
```

```
glimpse(present)
```

```
## Rows: 63
## Columns: 3
## $ year  <dbl> 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950~
## $ boys  <dbl> 1211684, 1289734, 1444365, 1508959, 1435301, 1404587, 1691220, 1~
## $ girls <dbl> 1148715, 1223693, 1364631, 1427901, 1359499, 1330869, 1597452, 1~
```

Exercise 5

The maximum births observed in the "arbuthnot" dataset is 16145, while the minimum total births observed in the "present" dataset is 2360399. The scale of the numbers is completley different.

```
arbuthnot %>%
  summarize(min = min(total), max = max(total))
```

```
## # A tibble: 1 x 2
##     min   max
##   <int> <int>
## 1  5612 16145
```

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

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

```
## # A tibble: 1 x 2
##       min     max
##     <dbl>   <dbl>
## 1 2360399 4268326
```

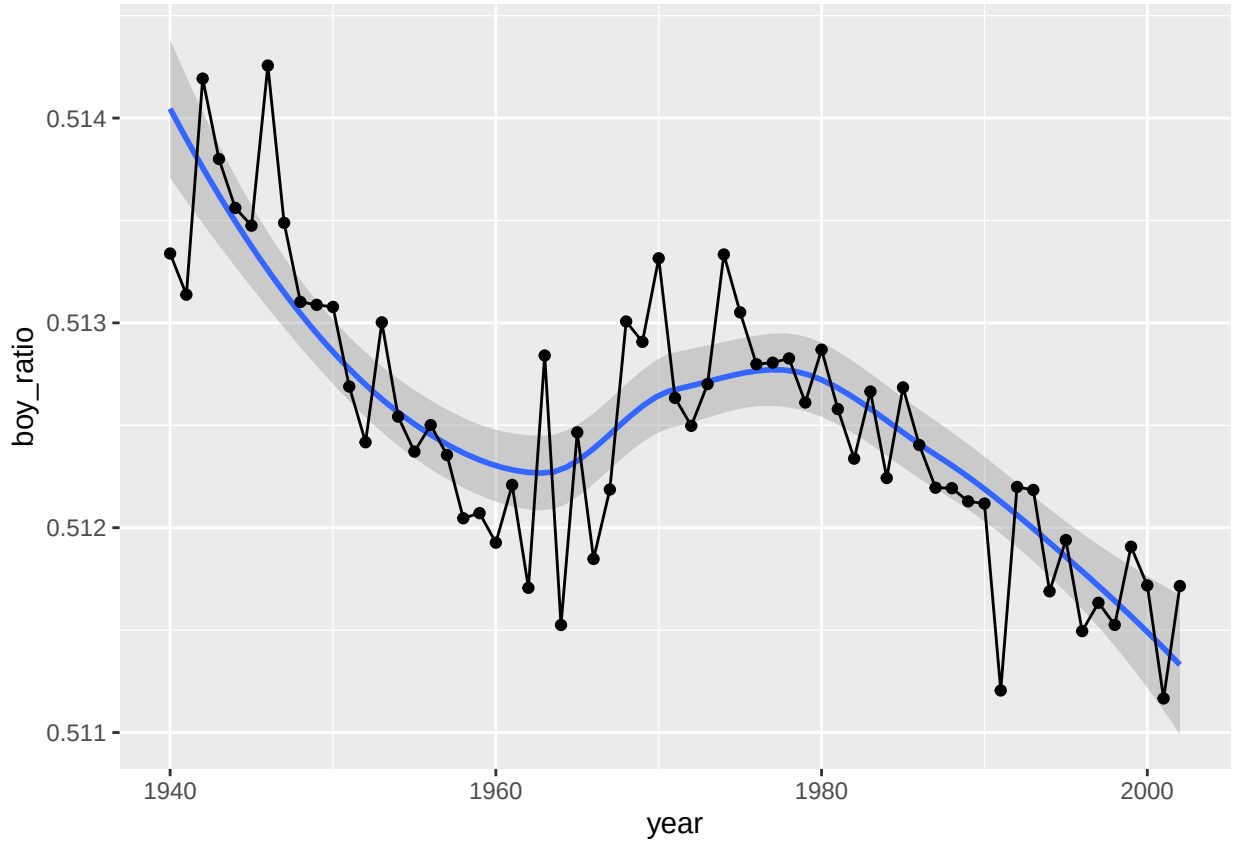
Exercise 6

There is indeed a greater ratio of boys born in the "present" dataset, just as was observed in the "arbuthnot" dataset.

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

ggplot(data = present, aes (x = year, y = boy_ratio)) +
  geom_smooth() +
  geom_point() +
  geom_line()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Exercise 7

The highest number of births, 4268326 of them, occurred in 1961.

```
totalss <-  present$total

present %>%
  arrange(desc(total))
```

```
## # A tibble: 63 x 5
##     year    boys   girls   total boy_ratio
##    <dbl>   <dbl>   <dbl>   <dbl>     <dbl>
##  1  1961 2186274 2082052 4268326     0.512
##  2  1960 2179708 2078142 4257850     0.512
##  3  1957 2179960 2074824 4254784     0.512
##  4  1959 2173638 2071158 4244796     0.512
##  5  1958 2152546 2051266 4203812     0.512
##  6  1962 2132466 2034896 4167362     0.512
##  7  1956 2133588 2029502 4163090     0.513
##  8  1990 2129495 2028717 4158212     0.512
##  9  1991 2101518 2009389 4110907     0.511
## 10  1963 2101632 1996388 4098020     0.513
## # ... with 53 more rows
```