

- Exercise 1
- Exercise 2
- Exercise 3
- Exercise 4
- Exercise 5
- Exercise 6
- Exercise 7

Lab 1: Intro to R

Yina Qiao

2023-02-02

Code ▾

Hide

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

Hide

```
arbuthnot
```

Hide

```
glimpse(arbuthnot)
```

Hide

```
arbuthnot$boys
```

Hide

```
arbuthnot$girls
```

Hide

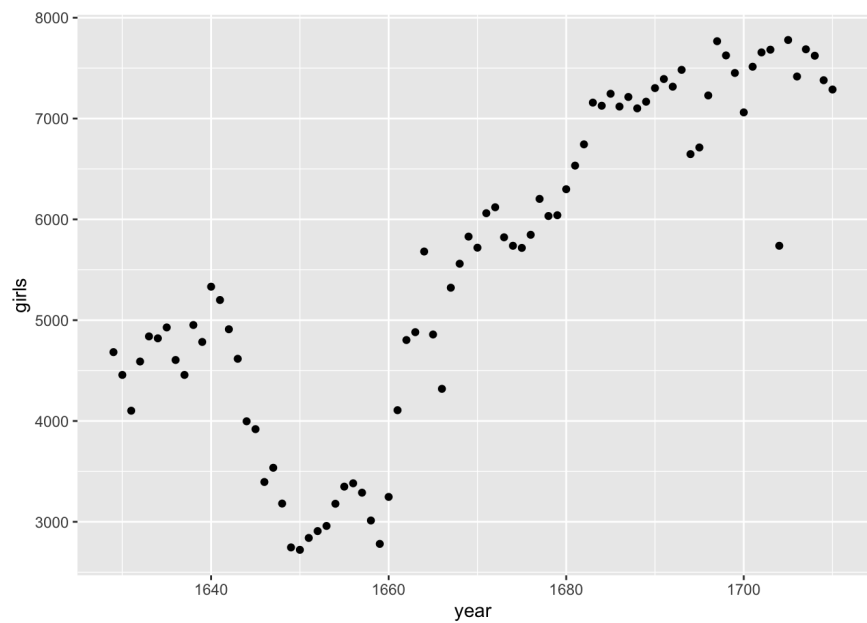
```
ggplot(data = arbuthnot, aes(x = year, y = girls)) +
  geom_point()
```

Exercise 1

Exercise 2

Insert any text here.

- Exercise 1
- Exercise 2
- Exercise 3
- Exercise 4
- Exercise 5
- Exercise 6
- Exercise 7

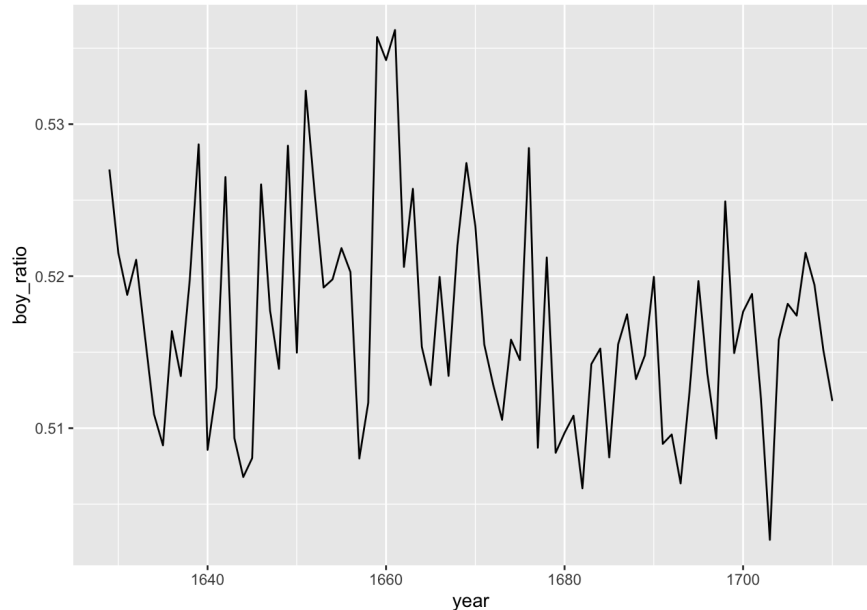


the plot above shown there is an increase in the number of girls being baptized from 1629 to 1710

Exercise 3

Hide

```
arbuthnot<-arbuthnot %>%
  mutate(total=boys+girls)
arbuthnot<-arbuthnot %>%
  mutate(boys_to_girls_ratio=boys/total)
arbuthnot<-arbuthnot %>%
  mutate(boy_ratio=boys/total)
ggplot(data=arbuthnot,aes(x=year,y=boy_ratio))+geom_line()
```



it looks like propotions of boys baptized are dropping gradually overtime.

Exercise 4

Hide

```
data('present', package='openintro')
glimpse(present)
```

Exercise 1

Exercise 2

Exercise 3

Exercise 4

Exercise 5

Exercise 6

Exercise 7

```
## 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

Hide

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

```
## # A tibble: 6 × 4
##   year   boys   girls total
##   <dbl> <dbl> <dbl> <dbl>
## 1  1940 1211684 1148715 2360399
## 2  1941 1289734 1223693 2513427
## 3  1942 1444365 1364631 2808996
## 4  1943 1508959 1427901 2936860
## 5  1944 1435301 1359499 2794800
## 6  1945 1404587 1330869 2735456
```

Hide

```
head(arbuthnot)
```

```
## # A tibble: 6 × 6
##   year   boys   girls total boys_to_girls_ratio boy_ratio
##   <int> <int> <int> <int>          <dbl>      <dbl>
## 1  1629   5218   4683   9901            0.527      0.527
## 2  1630   4858   4457   9315            0.522      0.522
## 3  1631   4422   4102   8524            0.519      0.519
## 4  1632   4994   4590   9584            0.521      0.521
## 5  1633   5158   4839   9997            0.516      0.516
## 6  1634   5035   4820   9855            0.511      0.511
```

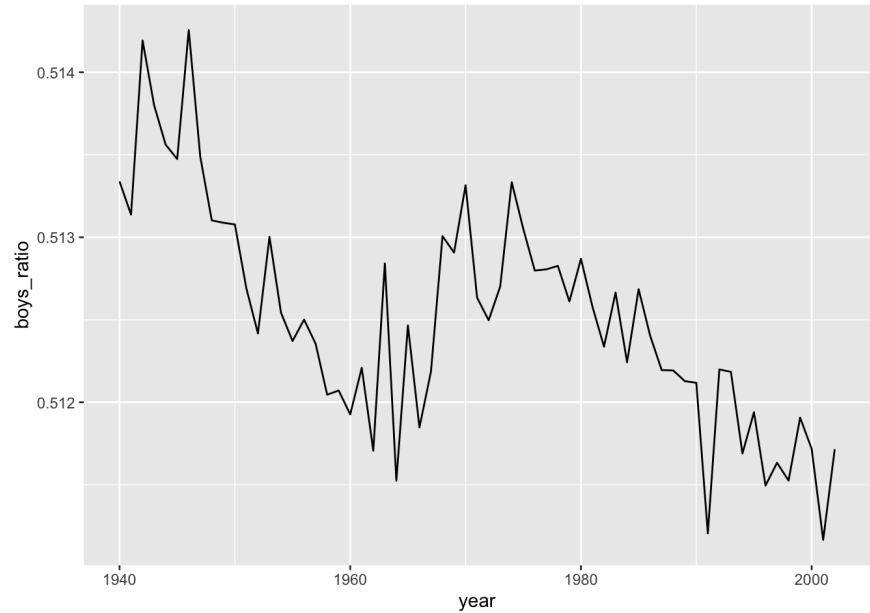
The values in present dataset are much bigger than arbuthnot by 3 orders of magnitude

Exercise 6

Hide

```
present<-present %>%
  mutate(total=boys+girls)
present<-present %>%
  mutate(boys_to_girls_ratio=boys/girls)
present<-present %>%
  mutate(boys_ratio=boys/total)
ggplot(data = present, aes(x=year,y=boys_ratio))+geom_line()
```

- Exercise 1
- Exercise 2
- Exercise 3
- Exercise 4
- Exercise 5
- Exercise 6
- Exercise 7



the boy_ratio is decreasing over time for the present data, different from Arbuthnot's observation.

Exercise 7

Hide

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

## # A tibble: 63 × 6
##   year    boys    girls  total boys_to_girls_ratio boys_ratio
##   <dbl> <dbl> <dbl> <dbl>          <dbl>      <dbl>
## 1 1961 2186274 2082052 4268326          1.05      0.512
## 2 1960 2179708 2078142 4257850          1.05      0.512
## 3 1957 2179960 2074824 4254784          1.05      0.512
## 4 1959 2173638 2071158 4244796          1.05      0.512
## 5 1958 2152546 2051266 4203812          1.05      0.512
## 6 1962 2132466 2034896 4167362          1.05      0.512
## 7 1956 2133588 2029502 4163090          1.05      0.513
## 8 1990 2129495 2028717 4158212          1.05      0.512
## 9 1991 2101518 2009389 4110907          1.05      0.511
## 10 1963 2101632 1996388 4098020          1.05      0.513
## # ... with 53 more rows
```