

Lab 1 - Intro to R and RStudio

Sanaz Saadatifar

1/29/2022

Lab report

Load Arbuthnot data

```
arbuthnot <- read_csv("https://dyurovsky.github.io/85309/data/lab1/arbuthnot.csv")
```

```
## Rows: 82 Columns: 3
```

```
## -- Column specification -----  
## Delimiter: ","  
## dbl (3): year, boys, girls
```

```
##  
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Exercise 1:

the first line of code is what is required for the exercise 1. I intentionally added the second line of code for boys because I need it to answer the exercise 5.

```
# code for Exercise 1 is already entered below as an example  
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
```

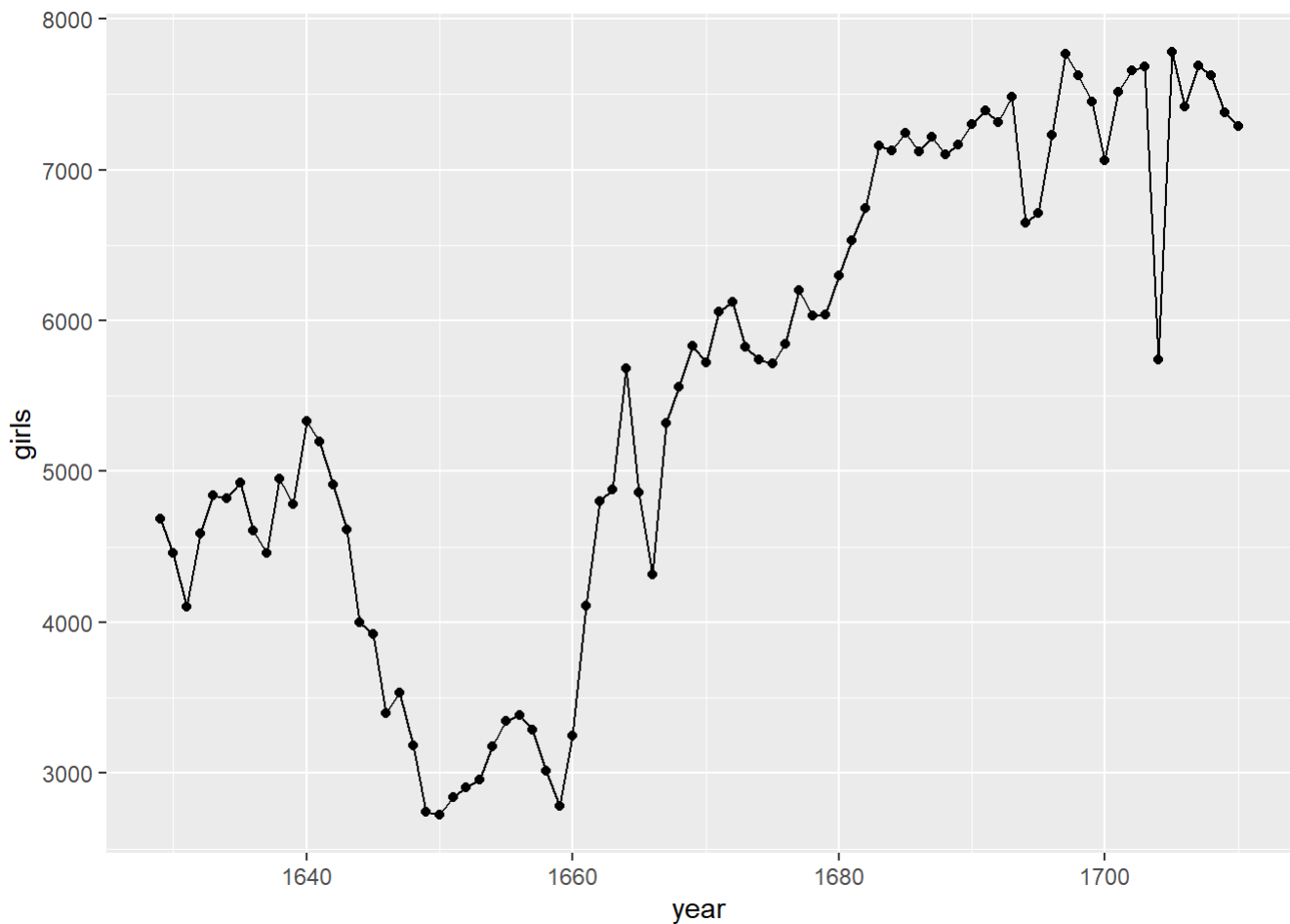
```
arbuthnot$boys
```

```
## [1] 5218 4858 4422 4994 5158 5035 5106 4917 4703 5359 5366 5518 5470 5460 4793
## [16] 4107 4047 3768 3796 3363 3079 2890 3231 3220 3196 3441 3655 3668 3396 3157
## [31] 3209 3724 4748 5216 5411 6041 5114 4678 5616 6073 6506 6278 6449 6443 6073
## [46] 6113 6058 6552 6423 6568 6247 6548 6822 6909 7577 7575 7484 7575 7737 7487
## [61] 7604 7909 7662 7602 7676 6985 7263 7632 8062 8426 7911 7578 8102 8031 7765
## [76] 6113 8366 7952 8379 8239 7840 7640
```

Exercise 2:

[Enter your response for Exercise 2 here.] In general more girls are baptized each year, except the dip between the year 1640 to year 1660.

```
# enter your code for Exercise 2 here
ggplot(arbutnot, aes(x = year, y = girls)) +
  geom_point() +
  geom_line()
```

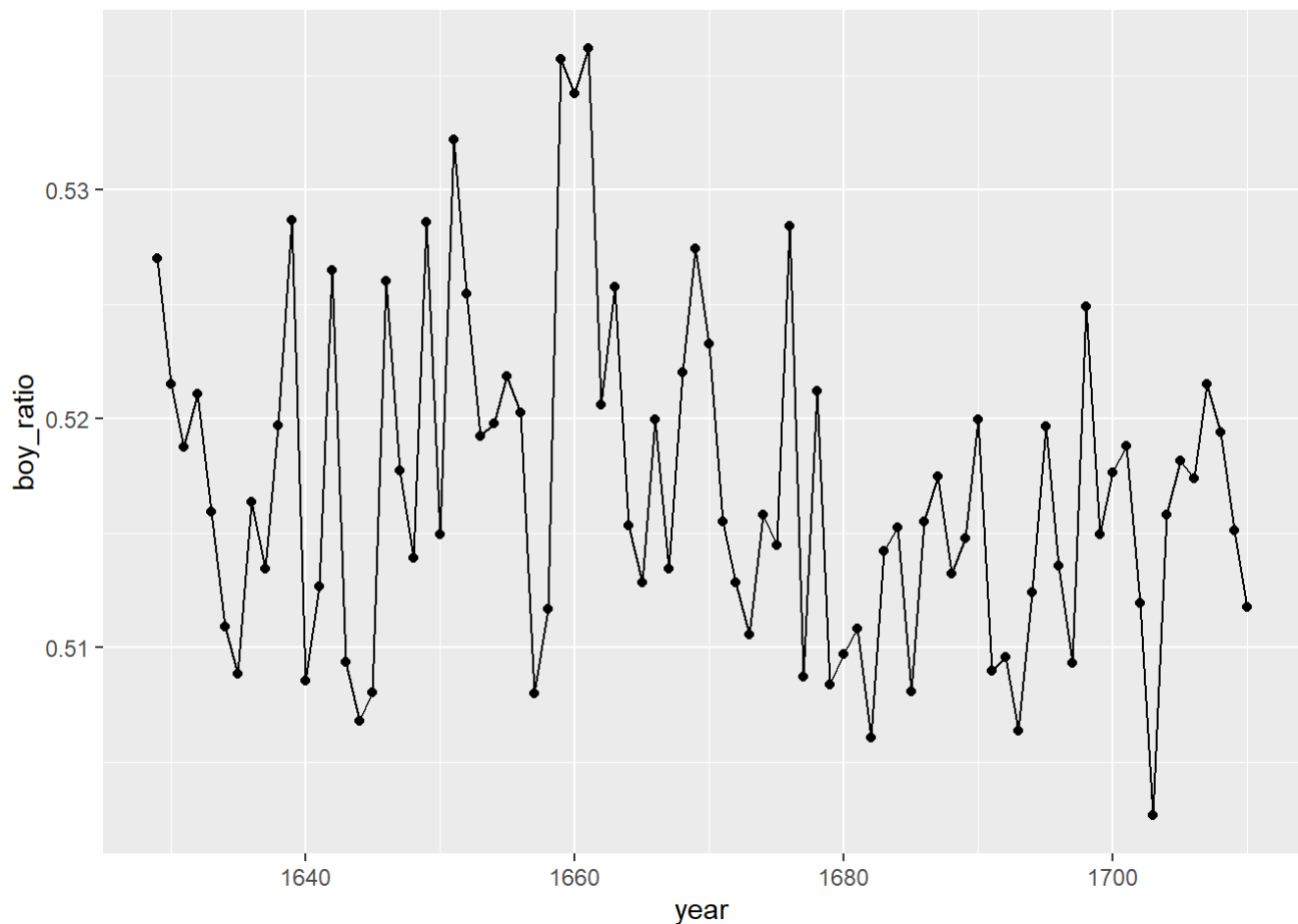


Exercise 3:

The ratio of the boys goes up and down a little bit from year to year, but it does not change significantly.

```
# enter your code for Exercise 3 here
arbuthnot <- arbuthnot %>%
  mutate(total = boys + girls) %>%
  mutate(boy_ratio = boys / total)

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



On your own:

Load present day data

```
present <- read_csv("https://dyurovsky.github.io/85309/data/lab1/present.csv")
```

```
## Rows: 76 Columns: 3
```

```
## -- Column specification -----
## Delimiter: ","
## dbl (3): year, boys, girls
```

```
##  
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

4:

Year 1940 to year 2015 (76 years) are included in this data, as listed below. Dimensions of the tibble are Rows: 76 Columns: 3. variable (column) names are year, boys, girls.

```
# enter your code for Exercise 4 here  
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 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014  
## [76] 2015
```

5:

The Arbuthnot and present in terms of the years, in which data is collected for both , are completely different. However, to compare the Counts of girls and boys in the present data set with Arbuthnot, it is possible to conclude that the number of boys and girls are growing (like a thousand time). Because during the Arbuthnot's time, the counts had about 4 digits, but in present data, the counts have about 7 digits.

```
# enter your code for Exercise 5 here  
present$girls
```

```
## [1] 1148715 1223693 1364631 1427901 1359499 1330869 1597452 1800064 1721216  
## [10] 1733177 1730594 1827830 1875724 1900322 1958294 1973576 2029502 2074824  
## [19] 2051266 2071158 2078142 2082052 2034896 1996388 1967328 1833304 1760412  
## [28] 1717571 1705238 1753634 1816008 1733060 1588484 1528639 1537844 1531063  
## [37] 1543352 1620716 1623885 1703131 1759642 1768966 1794861 1773380 1789651  
## [46] 1832578 1831679 1858241 1907086 1971468 2028717 2009389 1982917 1951379  
## [55] 1930178 1903234 1901014 1895298 1925348 1932563 1981845 1968011 1963747  
## [64] 1996415 2007391 2019367 2081318 2108162 2074069 2016926 1952825 1929522  
## [73] 1931041 1919073 1947375 1942625
```

```
present$boys
```

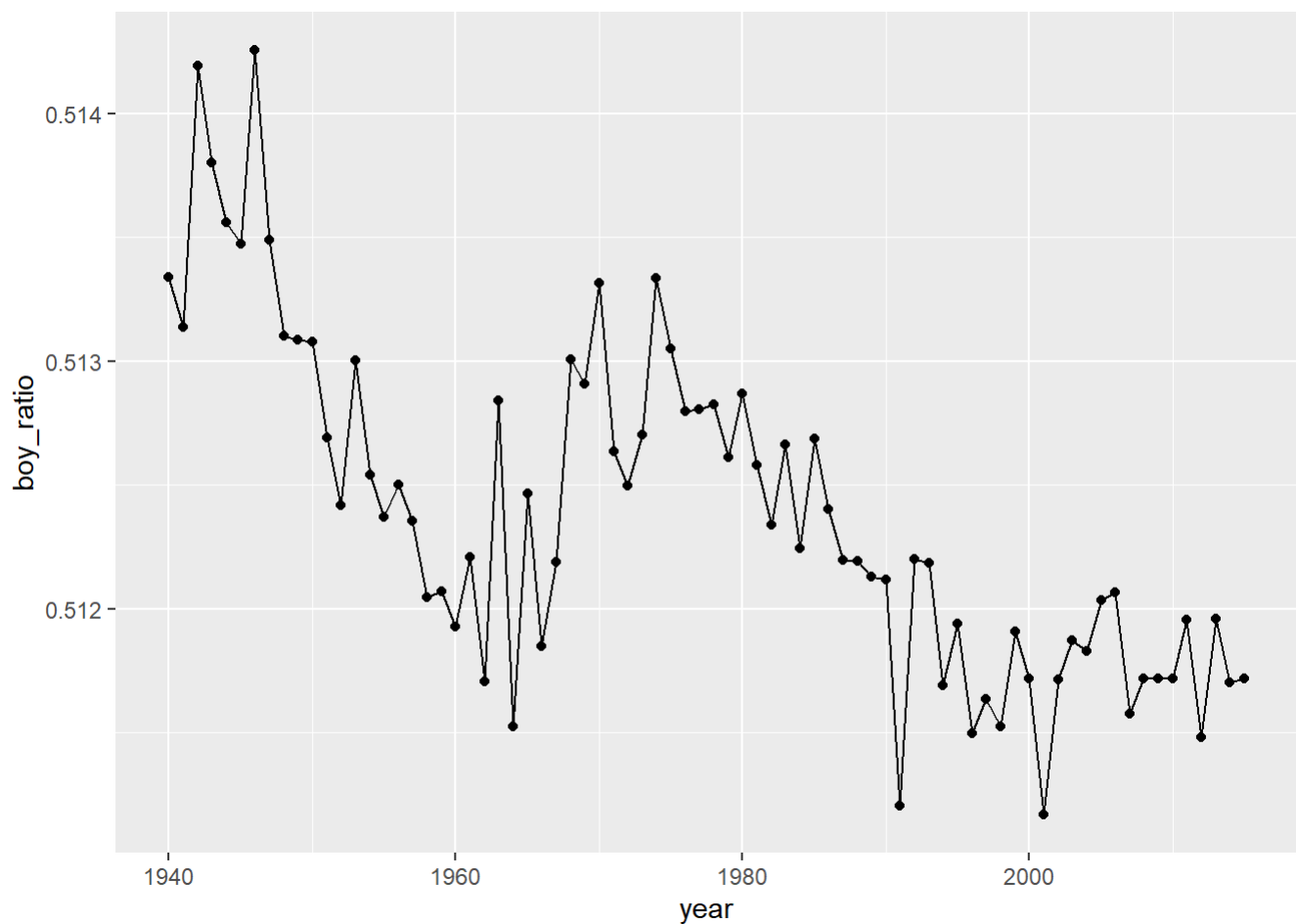
```
## [1] 1211684 1289734 1444365 1508959 1435301 1404587 1691220 1899876 1813852
## [10] 1826352 1823555 1923020 1971262 2001798 2059068 2073719 2133588 2179960
## [19] 2152546 2173638 2179708 2186274 2132466 2101632 2060162 1927054 1845862
## [28] 1803388 1796326 1846572 1915378 1822910 1669927 1608326 1622114 1613135
## [37] 1624436 1705916 1709394 1791267 1852616 1860272 1885676 1865553 1879490
## [46] 1927983 1924868 1951153 2002424 2069490 2129495 2101518 2082097 2048861
## [55] 2022589 1996355 1990480 1985596 2016205 2026854 2076969 2057922 2057979
## [64] 2093535 2104661 2118982 2184237 2208071 2173625 2113739 2046561 2024068
## [73] 2021800 2013108 2040701 2035872
```

6:

- Although the boys-ratio is decreasing in the US based on data from 1940 to 2015, still the number of boys being born each year is greater than number of girls being born in the same year.
- Arbuthnot's observation about the boy ratio was partially similar to the present's data (in terms of boy-ratio > 0.5), but that didn't have a descending order. However, overall we cannot directly relate these two data sets to each other and infer that the reason for having more boys being baptized in 1940 to 2015 is the higher number of boys being born compared to girls. And the main reason for this is not having population-born data from 1940 to 2015.

```
# enter your code for Exercise 6 here
present <- present %>%
  mutate(total = boys + girls) %>%
  mutate(boy_ratio = boys / total)

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



7:

2007 is the year with the most total number of births in the U.S. (4316233).

```
# enter your code for Exercise 7 here
present %>%
  arrange(desc(total))
```

```
## # A tibble: 76 x 5
##   year    boys  girls  total boy_ratio
##   <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1  2007 2208071 2108162 4316233    0.512
## 2  1961 2186274 2082052 4268326    0.512
## 3  2006 2184237 2081318 4265555    0.512
## 4  1960 2179708 2078142 4257850    0.512
## 5  1957 2179960 2074824 4254784    0.512
## 6  2008 2173625 2074069 4247694    0.512
## 7  1959 2173638 2071158 4244796    0.512
## 8  1958 2152546 2051266 4203812    0.512
## 9  1962 2132466 2034896 4167362    0.512
## 10 1956 2133588 2029502 4163090    0.513
## # ... with 66 more rows
```