

Lab 1: Intro to R

Nick Climaco

2/1/2023

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

Exercise 1

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

```
arbuthnot$girls
```

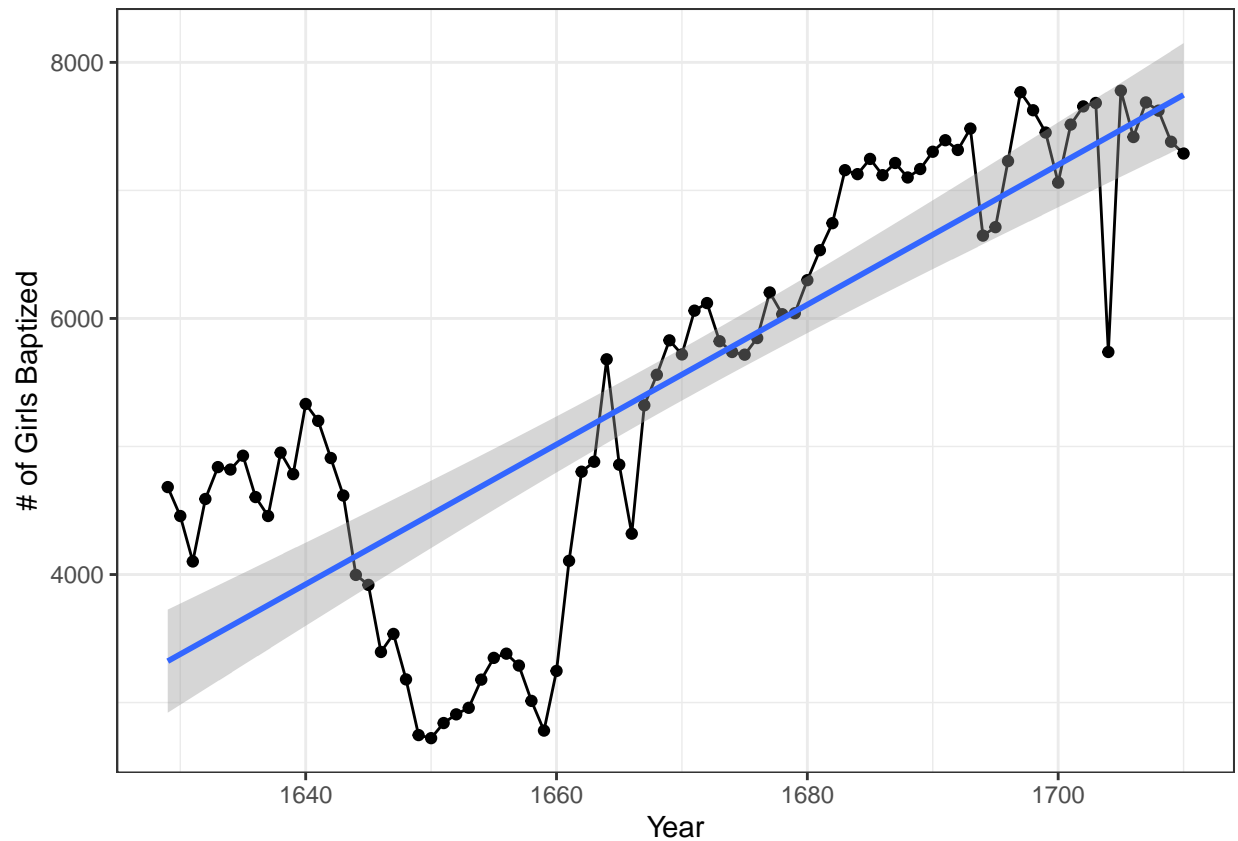
```
## [1] 4683 4457 4102 4590 4839 4820 4928 4605 4457 4952 4784 5332 5200 4910 4617
## [16] 3997 3919 3395 3536 3181 2746 2722 2840 2908 2959 3179 3349 3382 3289 3013
## [31] 2781 3247 4107 4803 4881 5681 4858 4319 5322 5560 5829 5719 6061 6120 5822
## [46] 5738 5717 5847 6203 6033 6041 6299 6533 6744 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? (To ensure that your lab report is comprehensive, be sure to include the code needed to make the plot as well as your written interpretation.)

```
arbuthnot %>% ggplot(aes(year, girls)) +
  geom_point() +
  geom_line() +
  geom_smooth(method = lm) +
  labs(x = "Year",
       y = "# of Girls Baptized") +
  theme_bw()
```

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



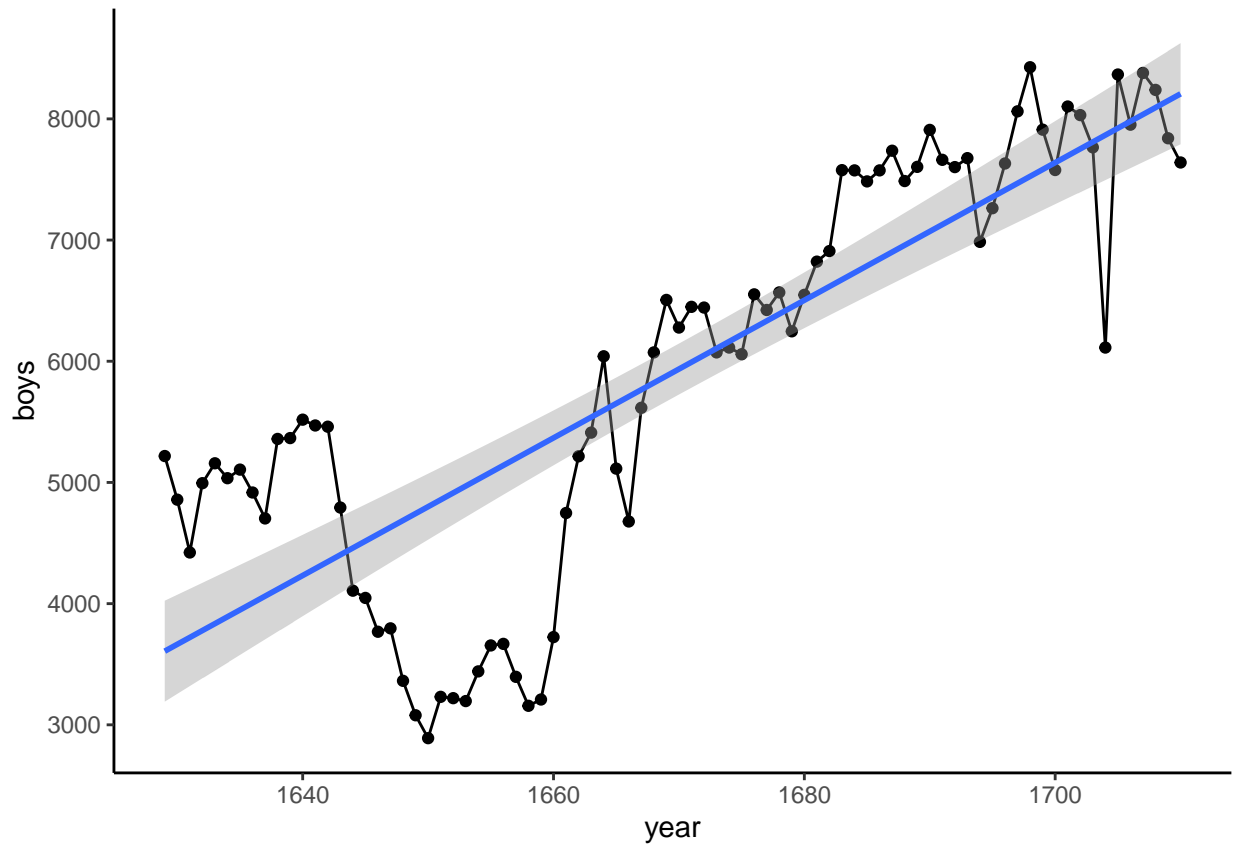
Based on the graph, we observe that there is a generally upward trends in the number of girls being baptized. Though, we see a sudden drop in the 1640s; it could be due the London revolution and various civil wars that occurred during the time period.

Exercise 3

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

```
arbuthnot %>% ggplot(aes(year, boys)) +
  geom_line() +
  geom_point() +
  geom_smooth(method = lm) +
  theme_classic()
```

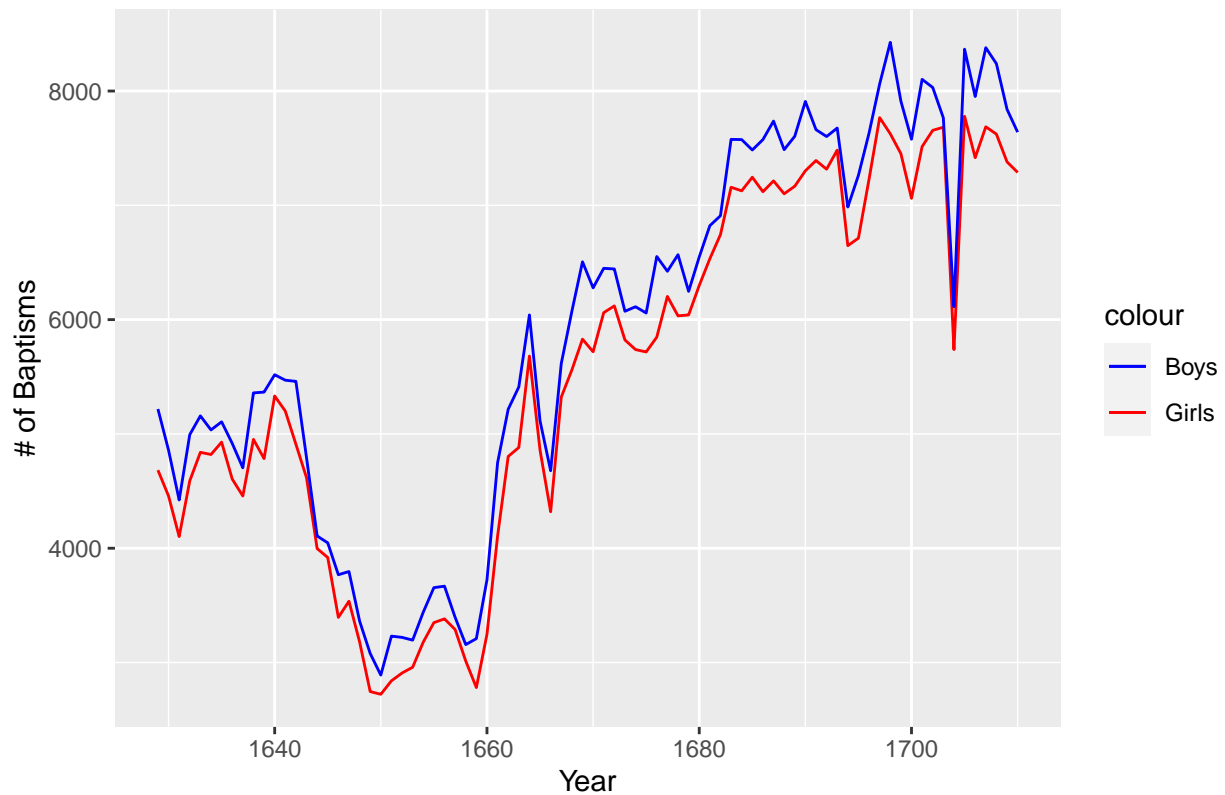
```
## 'geom_smooth()' using formula = 'y ~ x'
```



Similar to what we observed in the girls plot, we see a mainly positive trend of boys being baptized over time and notice the same sudden drop during the 1640s.

```
arbutthnot %>% ggplot(aes(x = year)) +
  geom_line(aes(y = girls, color = "Girls")) +
  geom_line(aes(y = boys, color = "Boys")) +
  scale_color_manual(values = c("Girls" = "Red", "Boys" = "Blue")) +
  labs(x = "Year",
       y = "# of Baptisms") +
  ggtitle("The Number of Boys and Girls Baptized in the 17th Century")
```

The Number of Boys and Girls Baptized in the 17th Century



```
data('present', package = 'openintro')
# present is the data set for recorded births in the US
```

```
head(present)
```

```
## # A tibble: 6 x 3
##   year    boys    girls
##   <dbl> <dbl> <dbl>
## 1  1940 1211684 1148715
## 2  1941 1289734 1223693
## 3  1942 1444365 1364631
## 4  1943 1508959 1427901
## 5  1944 1435301 1359499
## 6  1945 1404587 1330869
```

Exercise 4

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

```
range(present$year)
```

```
## [1] 1940 2002
```

```
dim(present)
```

```
## [1] 63  3
```

```
colnames(present)
```

```
## [1] "year" "boys" "girls"
```

Exercise 5

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

```
count(arbuthnot)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1     82
```

```
count(present)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1     63
```

```
mean(arbuthnot$boys)
```

```
## [1] 5907.098
```

```
mean(present$boys)
```

```
## [1] 1885600
```

```
mean(arbuthnot$girls)
```

```
## [1] 5534.646
```

```
mean(present$girls)
```

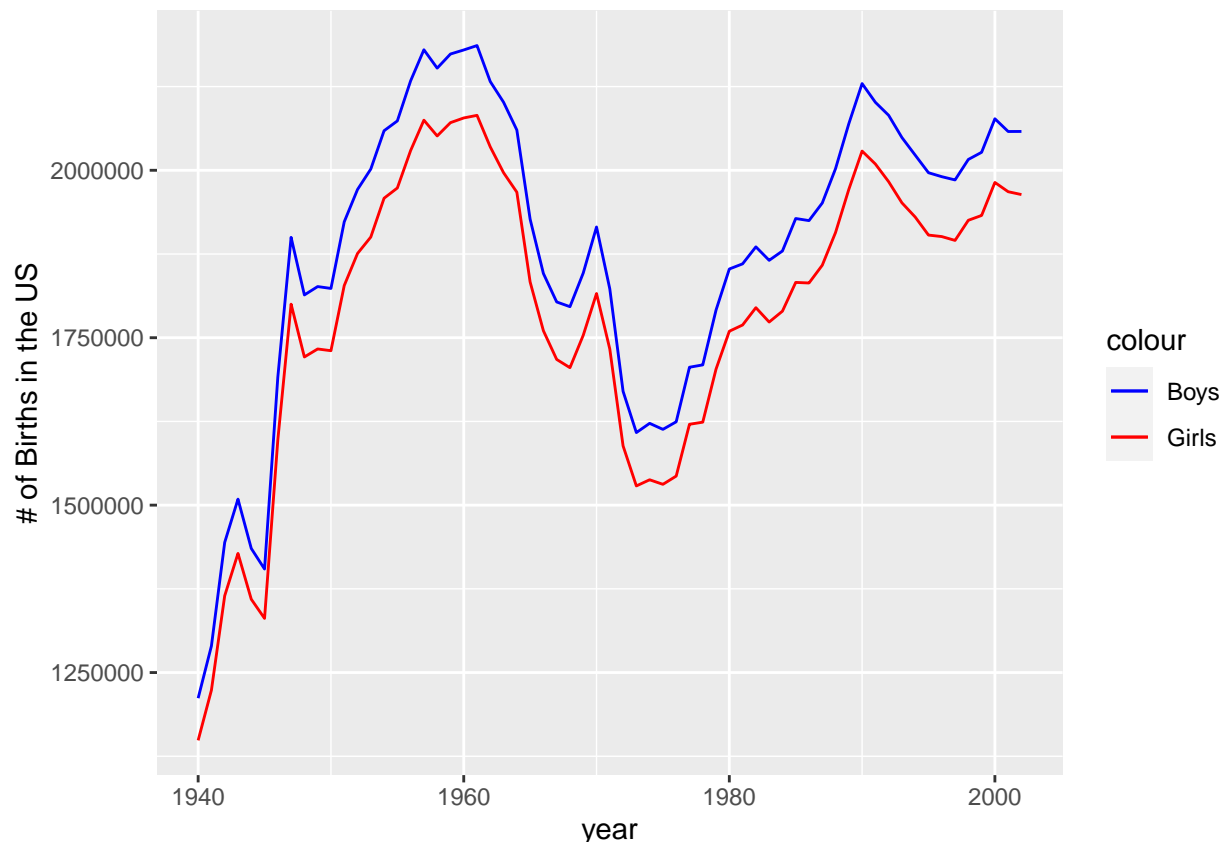
```
## [1] 1793915
```

In terms of number of observations, the arbuthnot data has 19 more entries than the present data set. As for the number of recorded births in either boys or girls, the magnitude of the present data set is 300 times larger than the magnitude of arbuthnot's data.

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.? Include the plot in your response. Hint: You should be able to reuse your code from Exercise 3 above, just replace the dataframe name.

```
present %>% ggplot(aes(x= year)) +  
  geom_line(aes(y = boys, color = "Boys")) +  
  geom_line(aes(y = girls, color = "Girls")) +  
  scale_color_manual(values = c("Boys" = "Blue", "Girls" = "Red")) +  
  labs(y = "# of Births in the US")
```



From the graph, we observe that boys were born in a greater proportion than girls similar to what we saw in Arbuthnot's data set. In addition, we notice a sudden drop in birthrate due to the introduction of birth control in the US and the women's right movement where it provides women to follow their careers and enter the workforce.

Exercise 7

In what year did we see the most total number of births in the U.S.? Hint: First calculate the totals and save it as a new variable. Then, sort your dataset in descending order based on the total column. You can do this interactively in the data viewer by clicking on the arrows next to the variable names. To include the sorted result in your report you will need to use two new functions: `arrange` (for sorting). We can arrange the data in a descending order with another function: `desc` (for descending order). The sample code is provided below.

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

```
## # A tibble: 6 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
```

In 1961, we observe from the data that it had the most total births recorded at 4,268,326.