

Week 1 Lab

Arin Parsa

5/20/2021

Dataset 1: Dr. Arbuthnot's Baptism Records

The Arbuthnot data set refers to Dr. John Arbuthnot, an 18th century physician, writer, and mathematician. He was interested in the ratio of newborn boys to newborn girls, so he gathered the baptism records for children born in London for every year from 1629 to 1710. We can take a look at the data by typing its name into the console.

```
library(devtools)
```

```
## Warning: package 'devtools' was built under R version 4.0.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
library(shiny)
```

```
## Warning: package 'shiny' was built under R version 4.0.5
```

```
library(statsr)  
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
data(arbuthnot)
```

```
arbuthnot
```

```
## # A tibble: 82 x 3  
##   year  boys girls  
##   <int> <int> <int>  
## 1 1629  5218  4683  
## 2 1630  4858  4457  
## 3 1631  4422  4102  
## 4 1632  4994  4590  
## 5 1633  5158  4839  
## 6 1634  5035  4820  
## 7 1635  5106  4928
```

```
## 8 1636 4917 4605
## 9 1637 4703 4457
## 10 1638 5359 4952
## # ... with 72 more rows
```

#Question 1: How many variables are included in this data set?

```
dim(arbuthnot)
```

```
## [1] 82 3
```

#Exercise 1: What years are included in this dataset?

```
length(arbuthnot$year)
```

```
## [1] 82
```

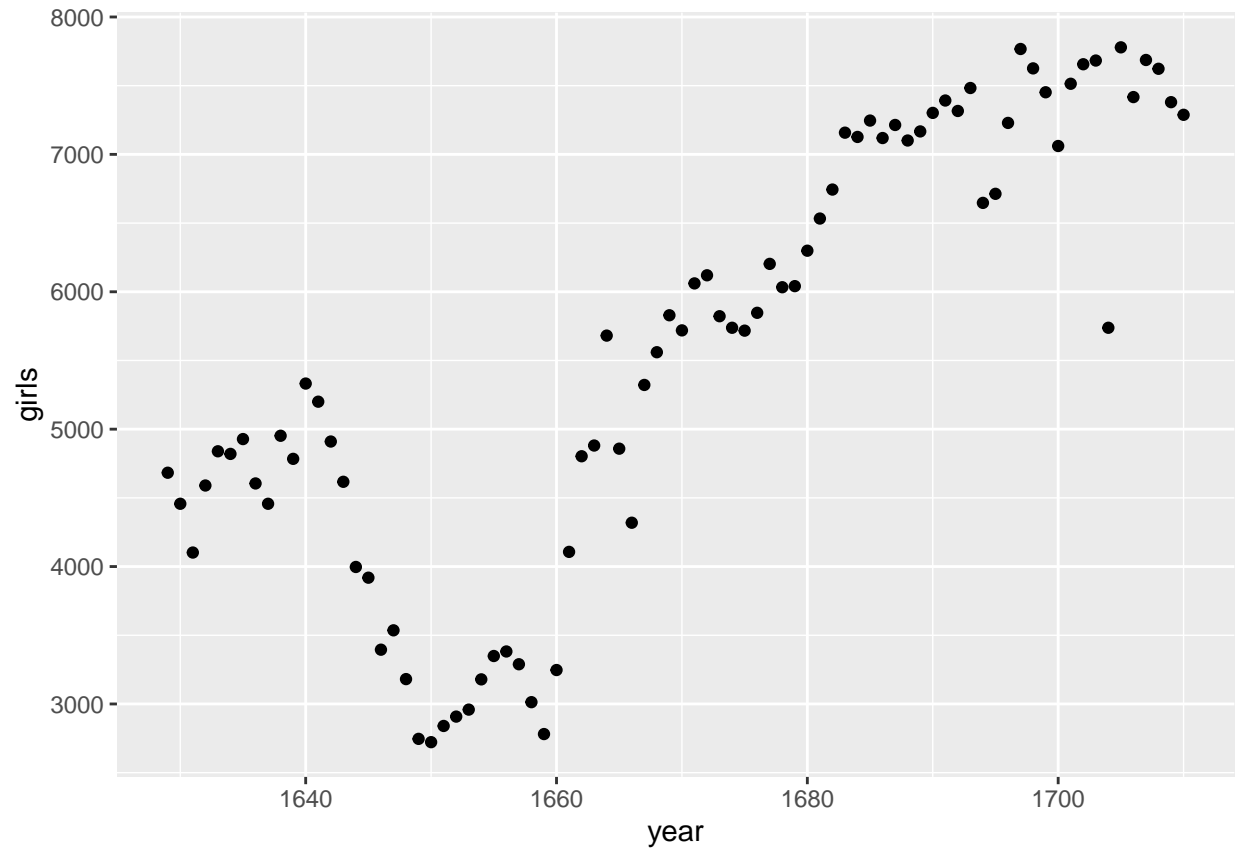
#Question 2: What command would you use to extract just the counts of girls born?

```
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
```

#Question 3: Which of the following best describes the number of girls baptised over the years included

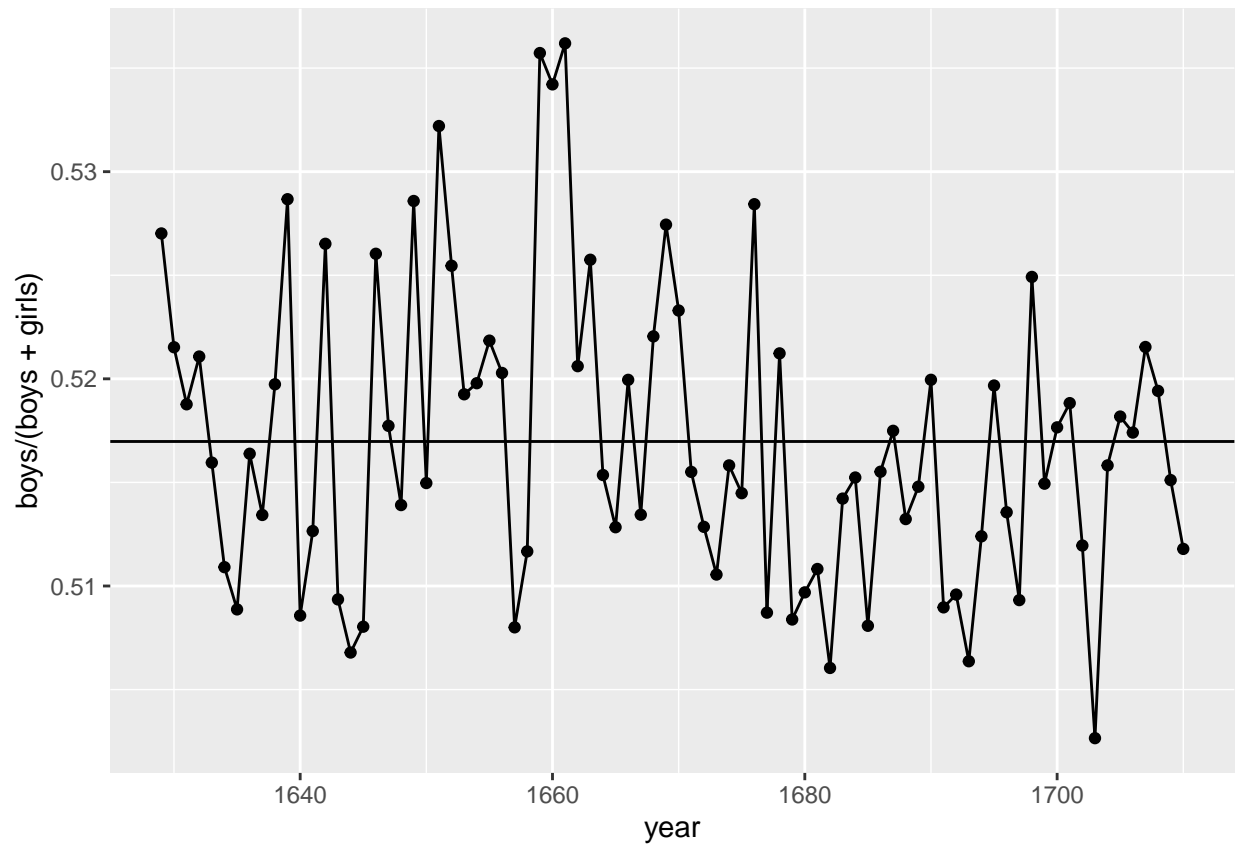
```
ggplot(arbuthnot, aes(year, girls)) + geom_point()
```



*#There is initially an increase in the number of girls baptised, which peaks around 1640.
 #After 1640 there is a decrease in the number of girls baptised, but the number begins to
 #increase again in 1660. Overall the trend is an increase in the number of girls baptised.*

#Exercise 2: Now, generate a plot of the proportion of boys born over time. What do you see?

`ggplot(arbuthnot, aes(year, boys/(boys+girls))) + geom_line() + geom_point() + geom_hline(aes(yintercept`



#Question 4: How many variables are included in this data set?

```
data(present)
dim(present)
```

```
## [1] 74 3
```

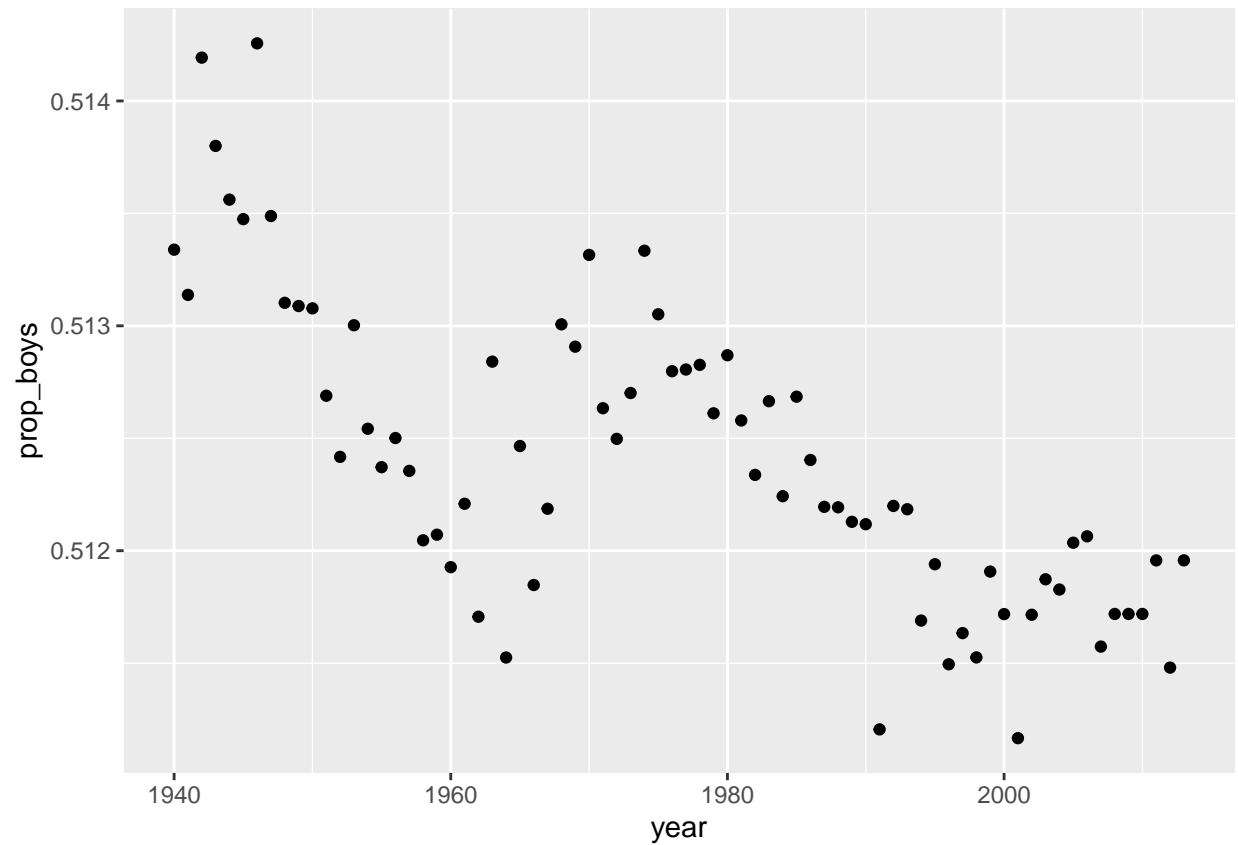
#Exercise 3: What years are included in this dataset?

```
length(present$year)
```

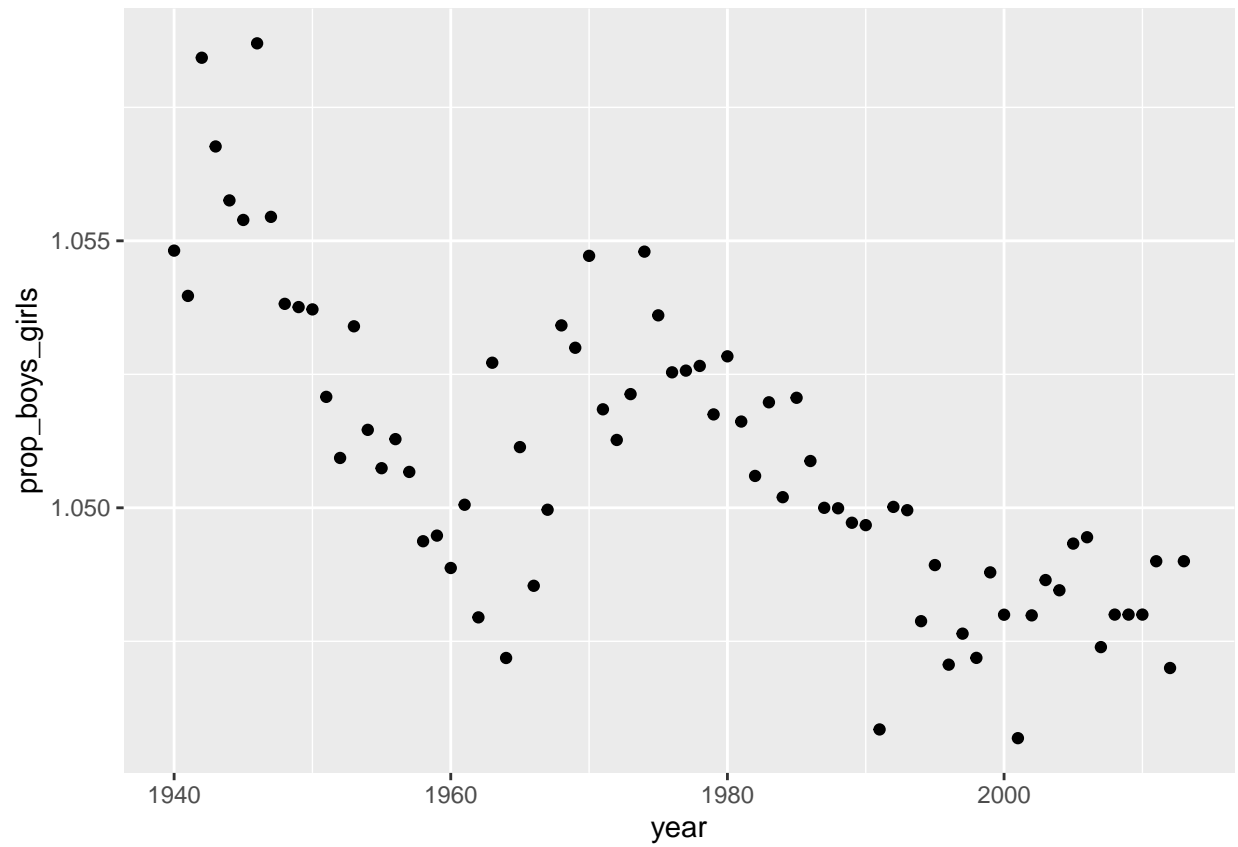
```
## [1] 74
```

#Question 5

```
present <- present %>% mutate(total = boys + girls) %>% mutate(prop_boys = boys/total)
ggplot(present, aes(year, prop_boys)) + geom_point()
```



```
#Question 6  
present <- present %>% mutate(more_boys = boys > girls) %>% mutate(prop_boys_girls = boys/girls)  
  
#Question 7  
ggplot(present, aes(year, prop_boys_girls)) + geom_point()
```



#Question 8

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