

# Lab 1.1: Arbuthnot

## Understanding the Context of the Data

1. What question did John Arbuthnot set out to answer in collecting this data?
2. What do *you* think the probability is that a newborn child is recorded as a girl? What form of evidence or reasoning did you use to come to that determination?
3. Sketch below a simplified version of the christening records (see the lab slides) as a data frame. What is the unit of observation?

When Arbuthnot published his study of the christening records, he presented his data like this (we show here just the first 10 rows).

```
# 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
# i 72 more rows
```

4. What does every row in this data frame published by John Arbuthnot correspond to? For each variable, state its type and the sort of values you would expect it to take (and also values it *should not* take).

5. What operations were necessary to get from the original christenings data frame to Arbuthnot's data frame?