1. (10) I wish that I had Jessie’s girl. Or maybe Jessie’s boy?

Use the `babynames` data from the `babynames` library to plot the number of male and female babies named Jessie over time.

Three time periods show up in the history of Jessie. Approximately what range of years does each time period span?

- More female than male Jessie, from 1880 to 1950.
- About the same number of male and female Jessie, from 1980 to 2015.

2. (10) Consider `babynames` but only the names from 2015. Say a name is “popular” in 2015 if it was given to more than 1000 babies of the same sex.

(a) Write R code that filters just the popular names from 2015.

**Solution:**
```
babynames %>% filter(year == 2015, n >= 1000)
```

(b) How many of these popular names are there?

**Solution:** 667

(c) How many of these popular names end in the letter ‘a’?

**Solution:** 123.
```
babynames %>% filter(year == 2015, n >= 1000, str_ends(name,'a'))
```

(d) What percentage of popular names ending in ‘a’ are girls names?

**Solution:** $\frac{120}{123} = 97.6\%$
(10) 3. Consider the `Pitching` data frame from the `Lahman` library. There is a variable `IPouts` which gives the number of outs pitched by a player that season.

Which player had the most outs pitched from 2000-2015? How many outs did they pitch?

Solution:

```r
Pitching %>% filter(yearID >= 2000 & yearID <= 2015) %>%
  group_by(playerID) %>% summarize(outs = sum(IPouts)) %>%
  arrange(desc(outs))

buehrma01 (Mark Buehrle) pitched 9850 outs in this period.
```

(10) 4. Consider the `Pitching` data frame from the `Lahman` library, and let’s talk about career statistics - the sum of a player’s statistics over all the seasons they played.

Which pitcher has the most outs pitched (`IPouts`) of all the pitchers who have more bases on balls (`BB`) than hits allowed (`H`)?

Who has the second most?

Solution:

```r
Pitching %>% group_by(playerID) %>%
  summarize(H = sum(H), BB = sum(BB), outs = sum(IPouts)) %>%
  filter(BB > H) %>%
  arrange(desc(outs))

First and second are willimi02 and marmoca01. Mitch Williams, the leader, was known as “Wild Thing”. Number two on the list is Carlos Mármol.
```
5. The data set `faithful` is built in to R, and shows the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park.

Use `ggplot` to produce a plot showing how eruption duration (`eruptions`) affects waiting time (`waiting`).

Write your code below, and then describe the main interesting feature of this graph.

```
Solution: The points appear in two clusters.
ggplot(faithful, aes(x=eruptions, y=waiting)) + geom_point()
```