

ECON 121 FA25 Problem Set 1

Solution

Question 2

We can report summary statistics as follows:

```
summary(ssa)
```

```
##      name      sex      frequency      year
## Length:1718855 Length:1718855 Min.   :  5.0 Min.   :1940
## Class :character Class :character 1st Qu.:  7.0 1st Qu.:1976
## Mode  :character Mode  :character Median : 12.0 Median :1996
##                                     Mean  : 172.3 Mean  :1991
##                                     3rd Qu.: 31.0 3rd Qu.:2009
##                                     Max.   :99693.0 Max.   :2022
```

The dataset spans 1940 to 2022. Name frequencies range from 5 to nearly 100,000.

Question 3

We can find the all-time most common boy and girl names as follows:

```
ssa |>
  group_by(name, sex) |>
  summarise(total = sum(frequency)) |>
  group_by(sex) |>
  slice_max(total, n=1)
```

```
## `summarise()` has grouped output by 'name'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 2 x 3
## # Groups:   sex [2]
##   name    sex    total
##   <chr>  <chr>  <dbl>
## 1 Mary    F    1997025
## 2 Michael M    4264145
```

The most popular names are Michael and Mary. 20 million girls were named Mary between 1940 and 2022, and 43 million boys were named Michael.

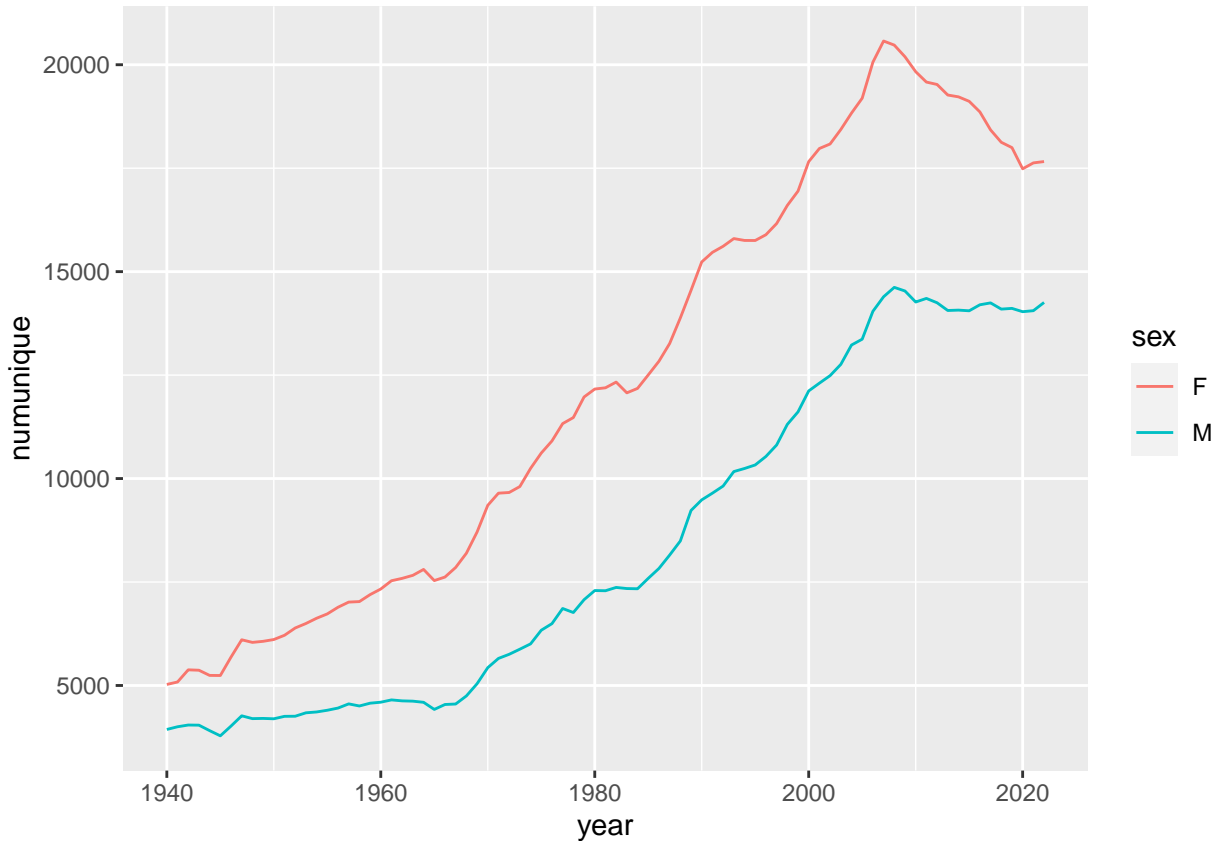
Question 4

To plot trends in the number of unique names by sex...

```
table1 <-
  ssa |>
  group_by(year, sex) |>
  summarise(numunique = n())
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
ggplot(table1, aes(x=year, y=numunique, color=sex)) +
  geom_line()
```



Both curves rise gradually from 1940 through the 1960s, then rise rapidly until 2008, and then either fall rapidly (for girls) or plateau (for boys).

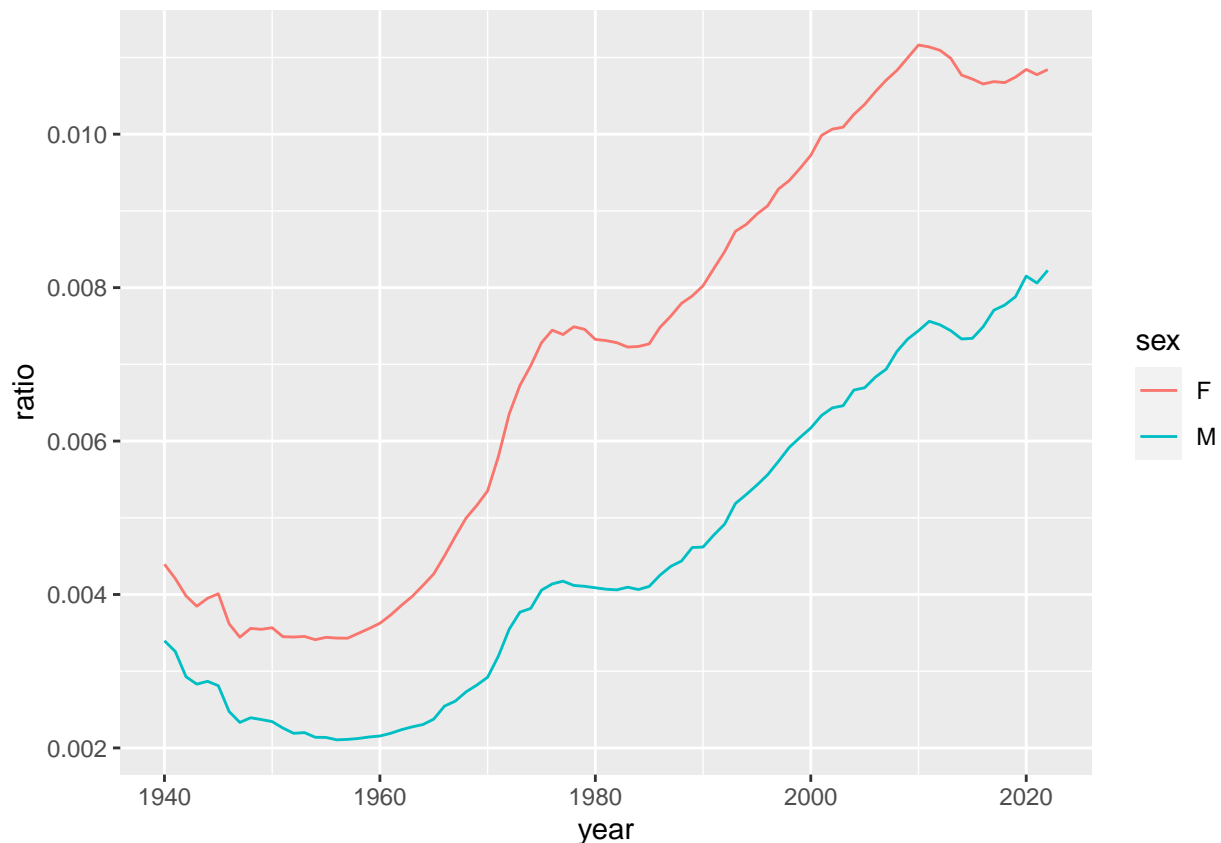
Question 5

To plot trends in the number of unique names relative to the number of babies by sex...

```
table2 <-
  ssa |>
  group_by(year, sex) |>
  summarise(numunique = n(),
            numbabies = sum(frequency)) |>
  mutate(ratio = numunique/numbabies)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
ggplot(table2, aes(x=year, y=ratio, color=sex)) +
  geom_line()
```



After a brief decline in the 1940s, the ratio rose consistently until the 2010s. The number of unique names increased faster than the number of babies. This trend is consistent with decreasing conformity over time.

Question 6

I chose Hilary and Hillary, just as in the example code. I wanted to know if naming frequencies changed after presidential elections involving the Clintons.

Question 7

Question 3 showed that the all-time most common names were Michael and Mary. Let's subset to those names and Hilary/Hillary, and then sum up all-time totals.

```
ssa |>
  filter((name=="Michael"&sex=="M") | (name=="Mary"&sex=="F") |
         (name=="Hilary"&sex=="F") | (name=="Hillary"&sex=="F")) |>
  group_by(name) |>
  summarise(total = sum(frequency))
```

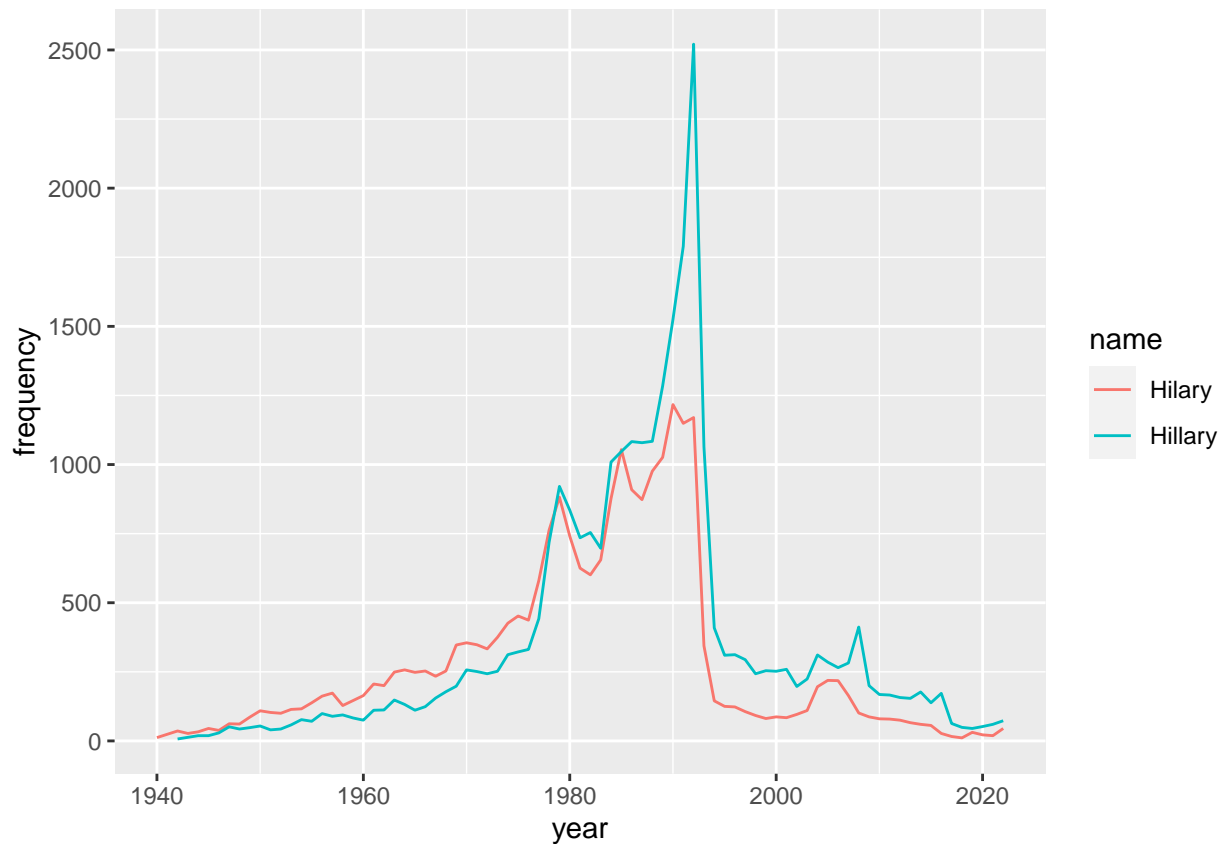
```
## # A tibble: 4 x 2
##   name      total
##   <chr>    <dbl>
## 1 Hilary    23913
## 2 Hillary   28821
## 3 Mary     1997025
## 4 Michael  4264145
```

Hilary/Hillary has been given to 53k girls since 1940, compared to 20 million Marys and 43 million Michaels.

Question 8

To draw the graph...

```
table3 <-  
  ssa |>  
  filter((name=="Hilary"&sex=="F")|(name=="Hillary"&sex=="F"))  
  
ggplot(table3, aes(x=year, y=frequency, group=name, color=name)) +  
  geom_line()
```



Hilary and Hillary steadily increased in popularity from 1940 to the early 1990s but then plummeted after the 1992 Clinton-Bush election. The names stabilized in popularity for the next decade and a half, then increased somewhat before the 2008 Clinton-Obama primary and fell again. Hillary then decreased further after the 2016 Clinton-Trump election.