Problem Set 1

Joe Ornstein

2022-08-09

```r
# In this R script, we'll plot the number of babies born each
# year in the United States with the name "Joseph"

## Step 1: Load Packages --------------

library(tidyverse) # a package with lots of useful functions
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(babynames) # a package with the baby names data from US Census


## Step 2: Clean Up The Data ------------------

# here's what the dataset looks like
head(babynames)
```

```
## # A tibble: 6 x 5
##    year sex   name           n   prop
##   <dbl> <chr> <chr>      <int>  <dbl>
## 1  1880 F     Mary        7065 0.0724
## 2  1880 F     Anna        2604 0.0267
## 3  1880 F     Emma        2003 0.0205
## 4  1880 F     Elizabeth   1939 0.0199
## 5  1880 F     Minnie      1746 0.0179
## 6  1880 F     Margaret    1578 0.0162
```

```r
# create an object called 'd' with just the male Andrew
d <- babynames |>
  filter(name == 'Andrew',
         sex == 'M')

head(d)
```
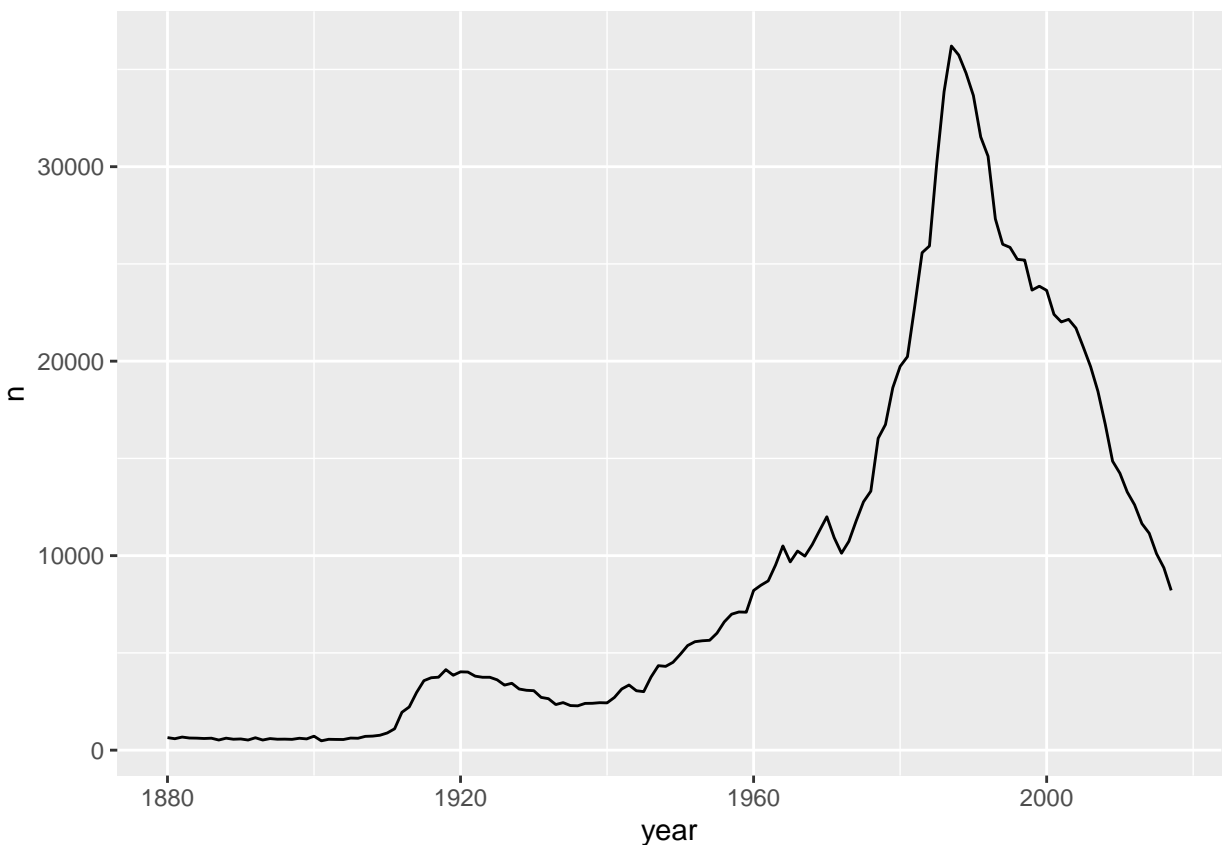
```
## # A tibble: 6 x 5
##    year sex   name        n    prop
##   <dbl> <chr> <chr>   <int>   <dbl>
## 1  1880 M     Andrew    644 0.00544
## 2  1881 M     Andrew    588 0.00543
## 3  1882 M     Andrew    673 0.00551
## 4  1883 M     Andrew    623 0.00554
## 5  1884 M     Andrew    616 0.00502
## 6  1885 M     Andrew    599 0.00517
```

```r
## Step 3: Plot The Data ------------------

# create a ggplot object called p
p <- ggplot(data = d,
            mapping = aes(x=year, y=n)) +
  # add a line geometry
  geom_line()

# return the plot
p
```



```r
# we'll learn how to make this prettier in the coming weeks :-)
```