

Artist Project 1

Diana

February 17, 2019

Today I learned about creating a density plot to show the distribution of birth decades for genders (male/female). These are the steps to produce a nice density plot :

Step 1: Set working directory, load libraries, and read the .csv file

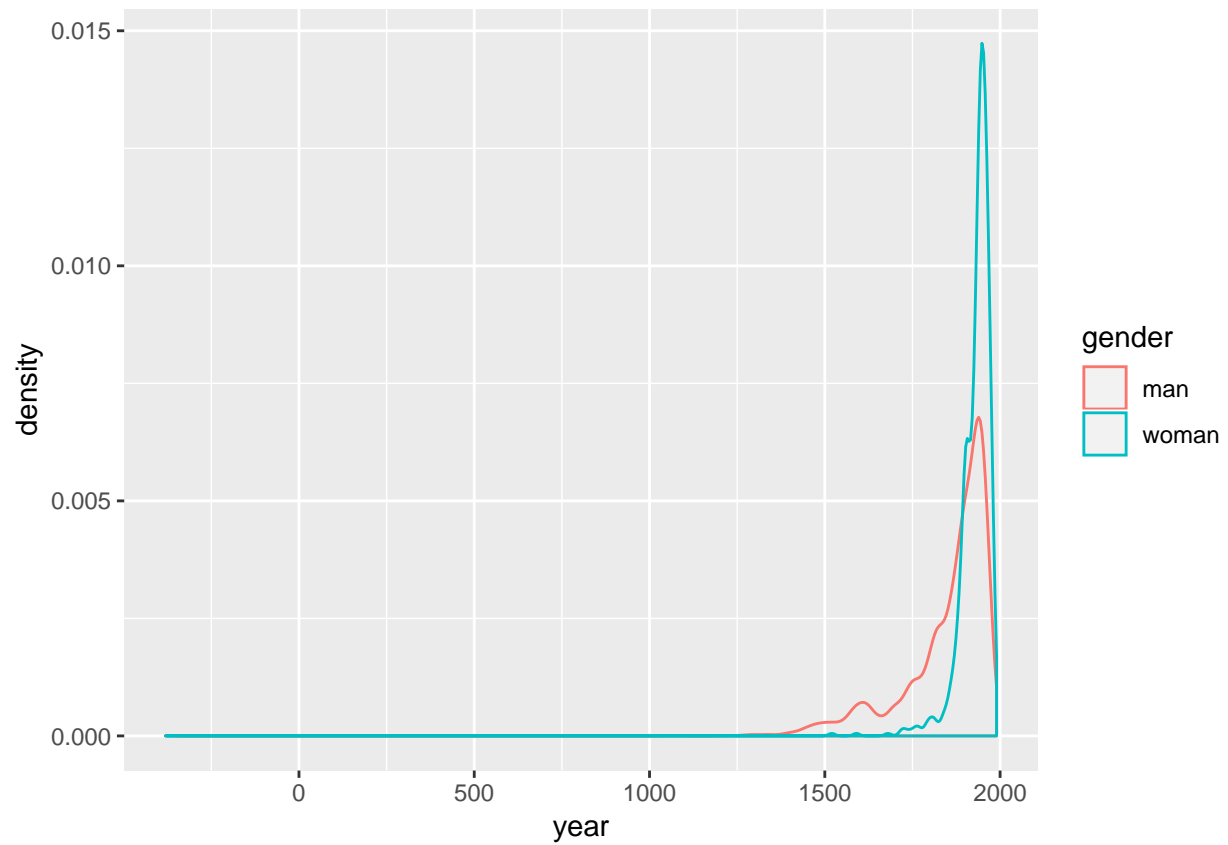
```
## [1] "C:/Users/diana/DataViz"
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Step 2: Omit NA values and create a dataframe to store the gender and year info

```
dist <- artist %>% na.omit() %>% select(gender, year)
```

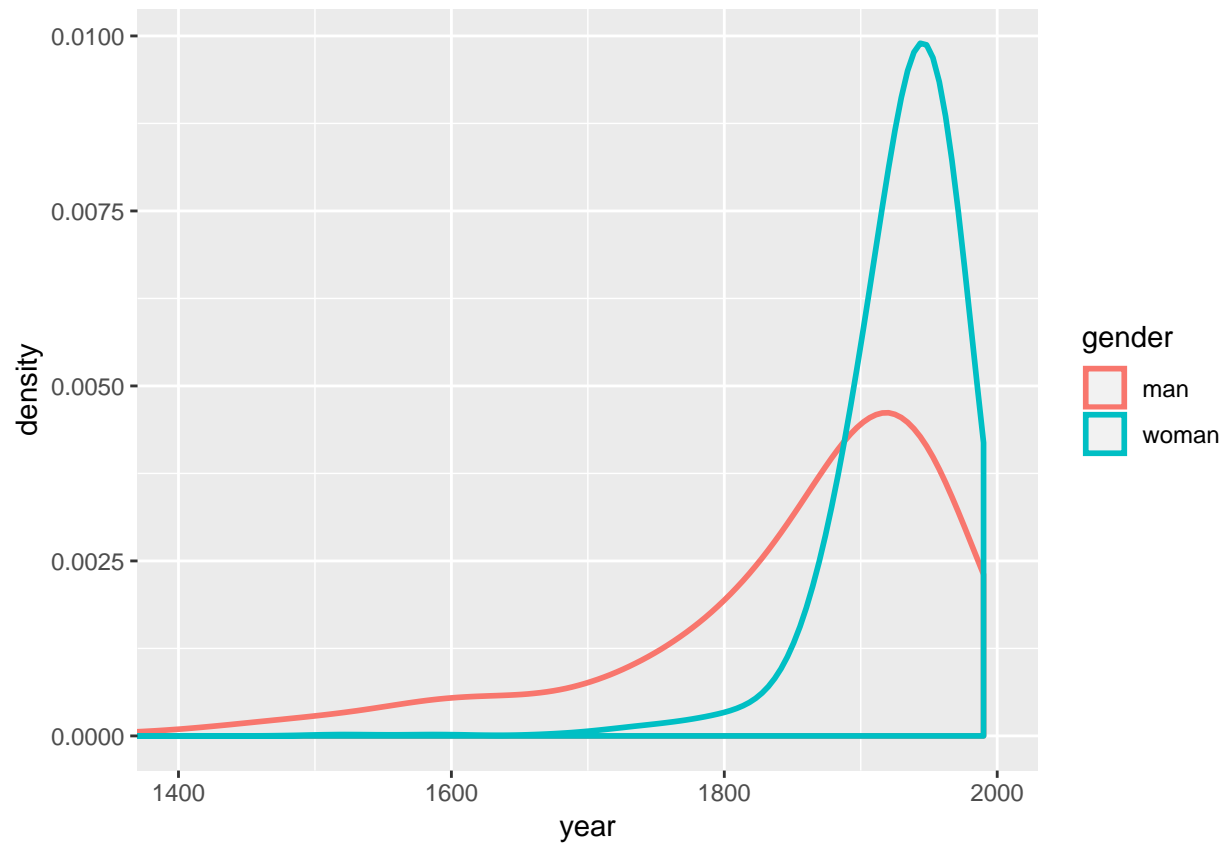
Step 3: Create a simple density plot to show the distribution of birth decades for artist's gender

```
ggplot(data=dist, aes(x=year, color=gender)) + geom_density()
```



Step 4 : Create a smoother density plot

```
ggplot(data=dist, aes(x=year, color=gender)) + geom_density(size=1, adjust=3) +  
  coord_cartesian(xlim=c(1400, 2000))
```



This is the density plot for each gender category

FEMALE ARTISTS

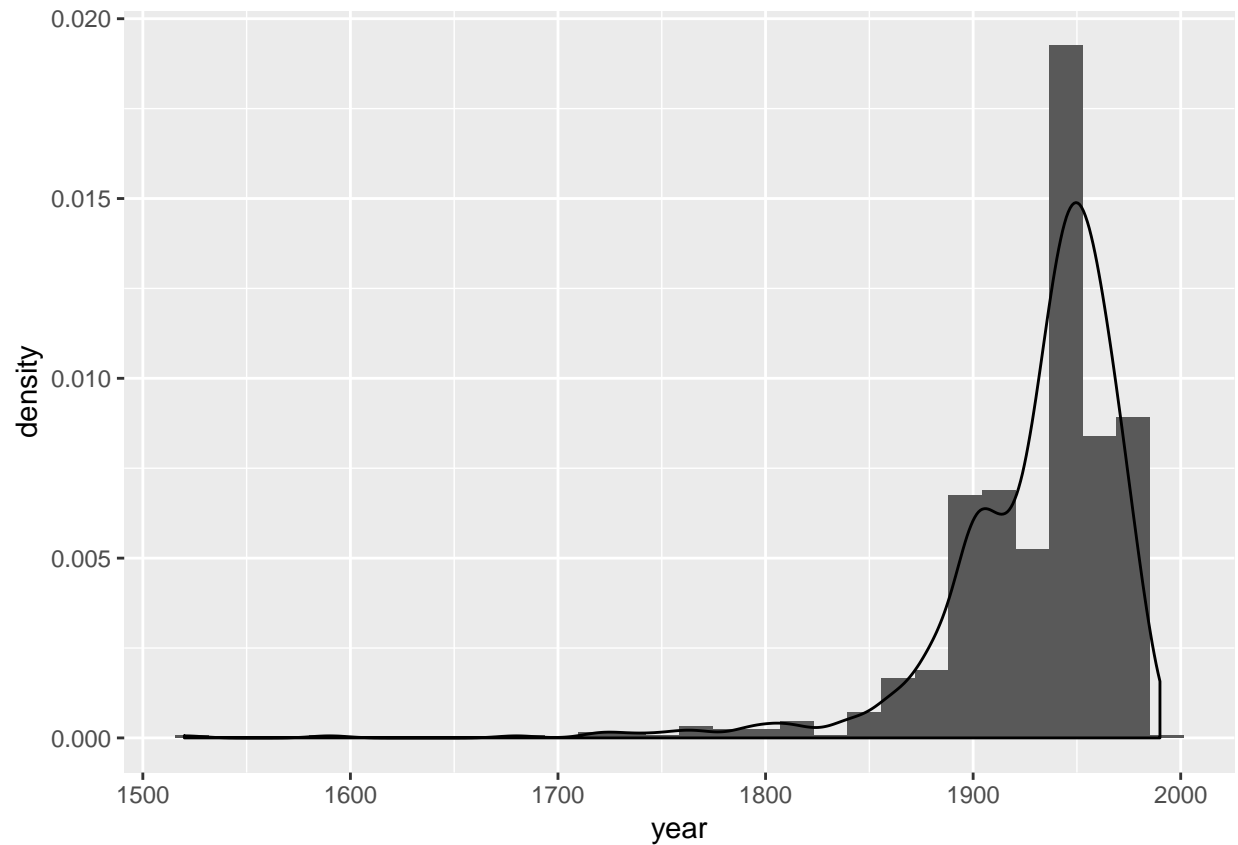
```
female.year <- dist %>% filter(gender == "woman")

#female.year

# histogram and density for female only
curve1 <- ggplot(data = female.year, aes(x = year)) + geom_histogram(aes(y = ..density..)) +
  geom_density()

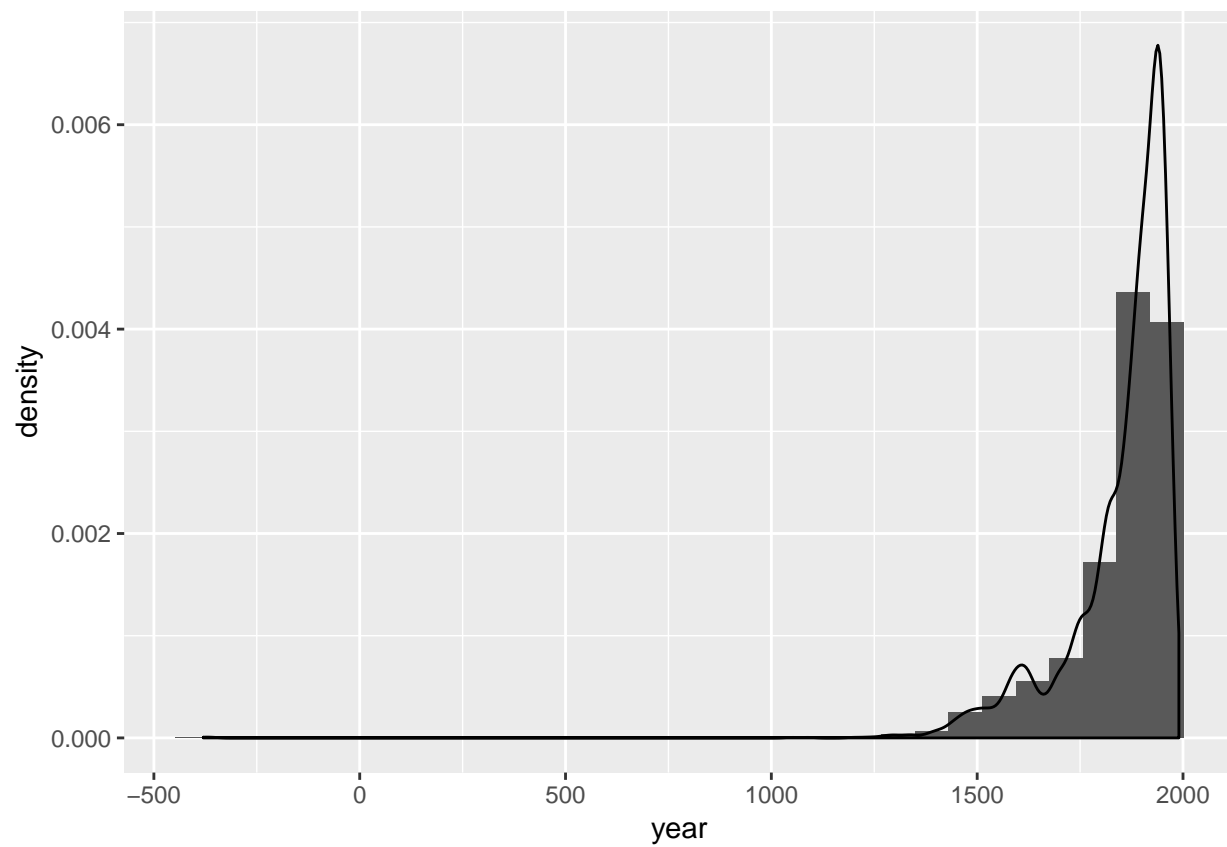
curve1

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



MALE ARTISTS

```
male.year <- dist %>% filter(gender == "man")  
  
#male.year  
  
#histogram and density for male only  
curve2 <- ggplot(data = male.year, aes(x = year)) + geom_histogram(aes(y = ..density..)) +  
  geom_density()  
  
curve2  
  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



That's all for today! See you next time!