607 -Project 2 - Oscars

Jose Mawyin

10/6/2019

Oscar Winners Demographics

How has ethnicity shaped the movie industry?

We will look at how the ethnic mix of Oscar winners has changed or nor in time by looking at a dataset that has collated Oscar Winner information from the years 1927 to 2014.

```
Oscars <- read.csv("/Users/josemawyin/Downloads/Oscars-demographics-DFE.csv")
```

The dataset has many fields but of interest are race_ethnicity year_of_award

```
str(Oscars)
```

```
## 'data.frame':    441 obs. of  27 variables:
##  $ X_unit_id                : int  670454353 670454354 670454355 670454356 670454357 670454358 67
##  $ X_golden                 : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ X_unit_state             : Factor w/ 2 levels "finalized","golden": 1 1 1 1 1 1 1 1 1 1 ...
##  $ X_trusted_judgments      : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ X_last_judgment_at       : Factor w/ 50 levels "","2/10/15 1:43",..: 36 18 20 19 5 38 22 3 4 3
##  $ birthplace               : Factor w/ 233 levels "Arlington, Va",..: 32 61 32 30 172 61 14 26 3
##  $ birthplace.confidence    : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ date_of_birth            : Factor w/ 346 levels "1-Apr-1885","1-Aug-65",..: 263 114 263 163 10
##  $ date_of_birth.confidence : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ race_ethnicity           : Factor w/ 6 levels "Asian","Black",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ race_ethnicity.confidence: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ religion                 : Factor w/ 22 levels "Agnostic","Anglican/episcopalian",..: 16 16 16
##  $ religion.confidence      : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ sexual_orientation       : Factor w/ 6 levels "Bisexual","Gay",..: 6 6 6 6 6 6 6 1 6 6 ...
##  $ sexual_orientation.confidence: num  1 0.684 1 1 1 ...
##  $ year_of_award            : int  1927 1930 1931 1932 1933 1934 1935 1936 1937 1938 ...
##  $ year_of_award.confidence : num  1 1 0.667 1 1 ...
##  $ award                    : Factor w/ 5 levels "Best Actor","Best Actress",..: 3 3 3 3 3 3 3 3
##  $ biourl                   : Factor w/ 348 levels "http://www.nndb.com/people/001/000022932/",.
##  $ birthplace_gold          : Factor w/ 9 levels "","Bascom, Fl",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ date_of_birth_gold       : Factor w/ 9 levels "","11-Dec-67",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ movie                    : Factor w/ 336 levels "12 Years a Slave",..: 322 246 22 213 35 60 12
##  $ person                   : Factor w/ 348 levels "Adrien Brody",..: 210 94 210 258 92 94 93 175
##  $ race_ethnicity_gold      : Factor w/ 3 levels "","Black","White": 1 1 1 1 1 1 1 1 1 1 ...
##  $ religion_gold            : Factor w/ 6 levels "","Born-Again Christian",..: 1 1 1 1 1 1 1 1 1 1
##  $ sexual_orientation_gold  : Factor w/ 3 levels "","Bisexual",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year_of_award_gold       : int  NA NA NA NA NA NA NA NA NA NA ...
```

```
dim(Oscars)
```

```
## [1] 441  27
```

Let's first find how many Oscars have been awarded based on ethnicity:

```
summary(Oscars$race_ethnicity)
```

```
##          Asian          Black      Hispanic Middle Eastern     Multiracial
##              4             15              8              1               2
##          White
##            411
```
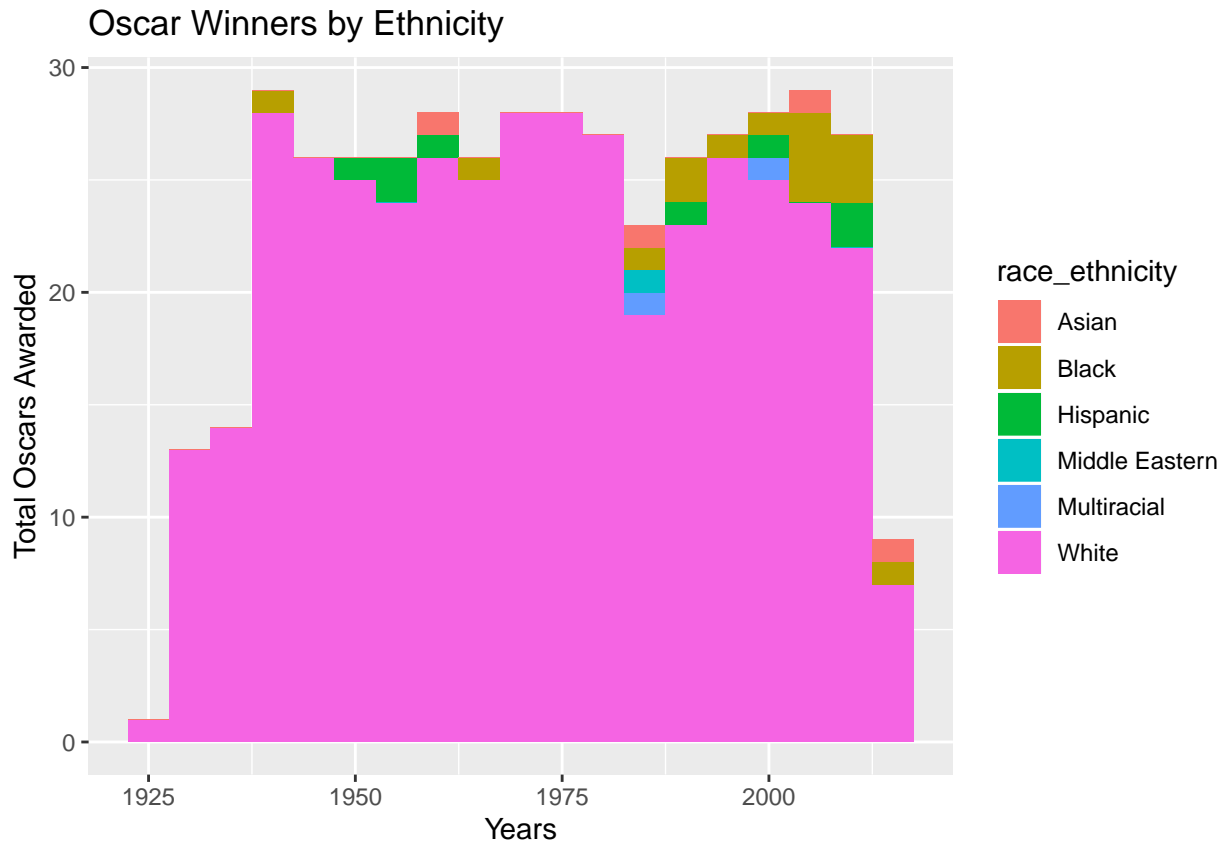
Then let's find when was the first Oscar awarded to an ethnicy other than "White":

```
count(Oscars, year_of_award, race_ethnicity) %>% filter(race_ethnicity != "White") %>%  arrange(year_of_
```

```
## # A tibble: 27 x 3
##    year_of_award race_ethnicity       n
##            <int> <fct>            <int>
##  1          1940 Black                1
##  2          1951 Hispanic             1
##  3          1953 Hispanic             1
##  4          1957 Hispanic             1
##  5          1958 Asian                1
##  6          1962 Hispanic             1
##  7          1964 Black                1
##  8          1983 Black                1
##  9          1983 Multiracial          1
## 10          1985 Asian                1
## # ... with 17 more rows
```

We can see that in the period of 1927 to 1940 no Oscars were awarded to non-White winners.

Below we can see how the ethnic inclusivity has increased since the later quarter of the 20th century until the present.

```
ggplot(Oscars, aes(year_of_award, fill = race_ethnicity)) +
  geom_histogram(binwidth = 5) + ylab("Total Oscars Awarded") + xlab("Years") + ggtitle("Oscar Winners I
```

Oscar Winners by Ethnicity

Lack of inclusion for unrepresented members of the movie community is evident in the data shown in this study. However, there is been a change in the number of Oscar's awarded to other-than-white Further study could see how does the number of winners compared to the total number of movie industry workers broke down by ethnicity. Are other-than-white represented?