STA 210: Lab 1

Jackson Hubbard

September 3rd, 2018

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:GGally':
##
##     nasa
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Question 1

```
oscar.data = oscar_winners
glimpse(oscar_winners)
```

```
## Observations: 180
## Variables: 5
## $ award.year <int> 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 193...
## $ age        <int> 44, 38, 62, 53, 41, 34, 33, 49, 41, 37, 38, 34, 32,...
## $ name       <chr> "Emil Jannings", "Warner Baxter", "George Arliss", ...
## $ movie      <chr> "The Last Command", "In Old Arizona", "Disraeli", "...
## $ category   <chr> "Best Actor", "Best Actor", "Best Actor", "Best Act...
```

Question 2

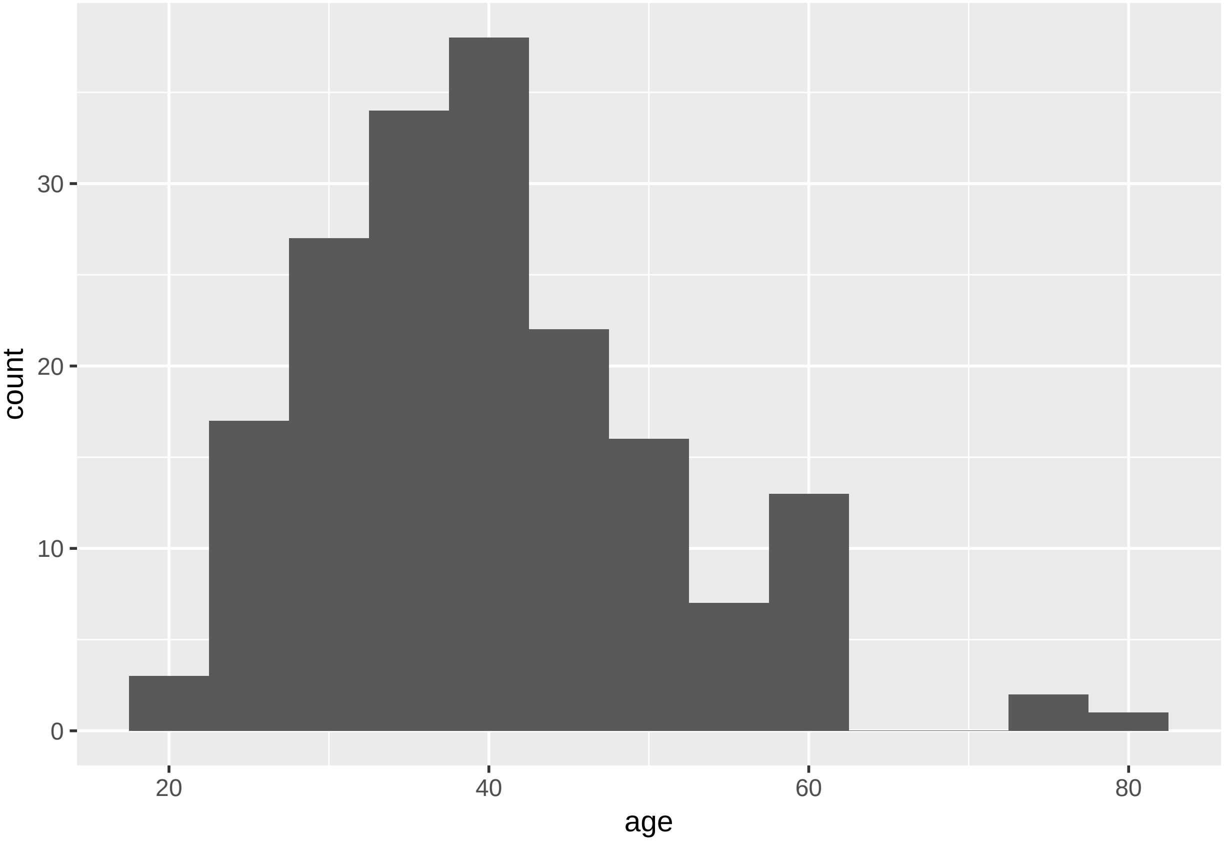
There are 180 observations in this data set

Question 3

Age is an int variable Movie is a character variable

Question 4

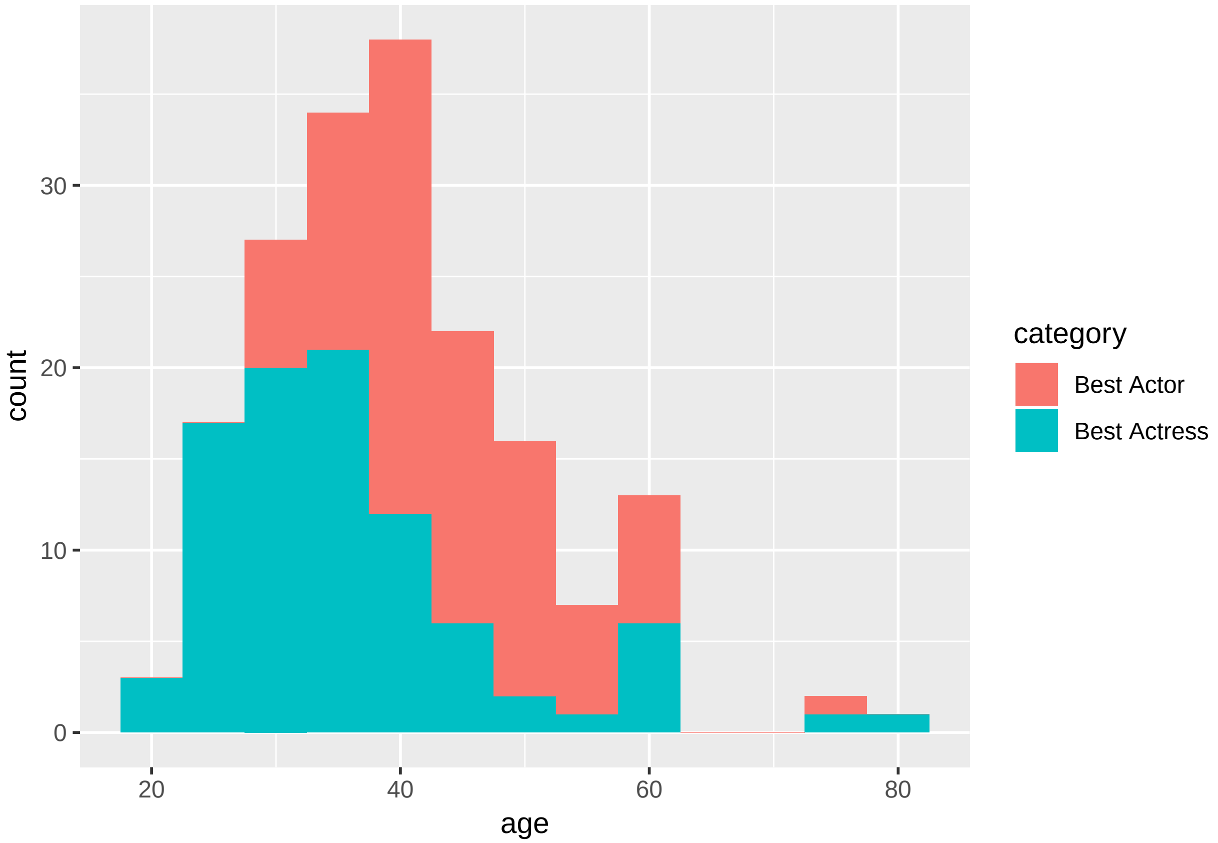
```r
ggplot(oscar_winners, aes(age)) + geom_histogram(binwidth=5)
```



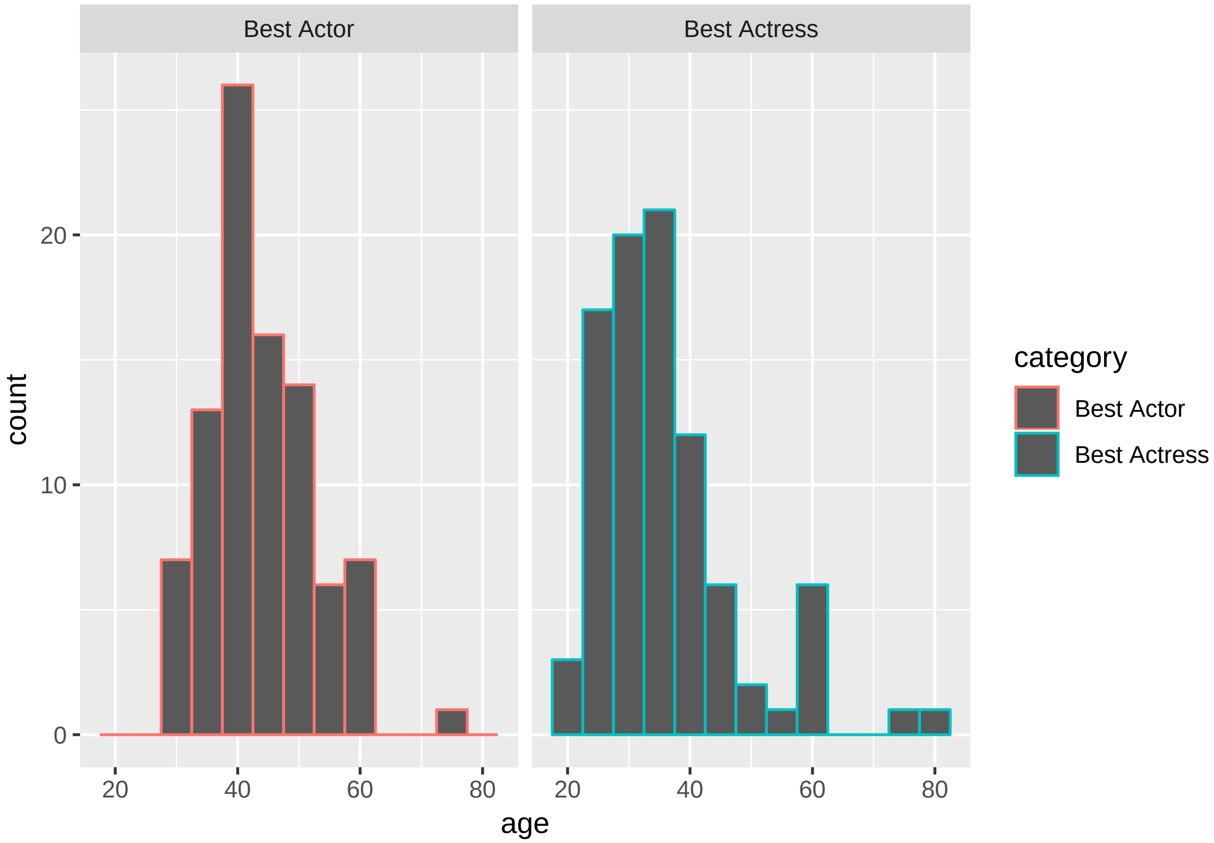
Age is approximately normal, with a skew to the right

Question 5

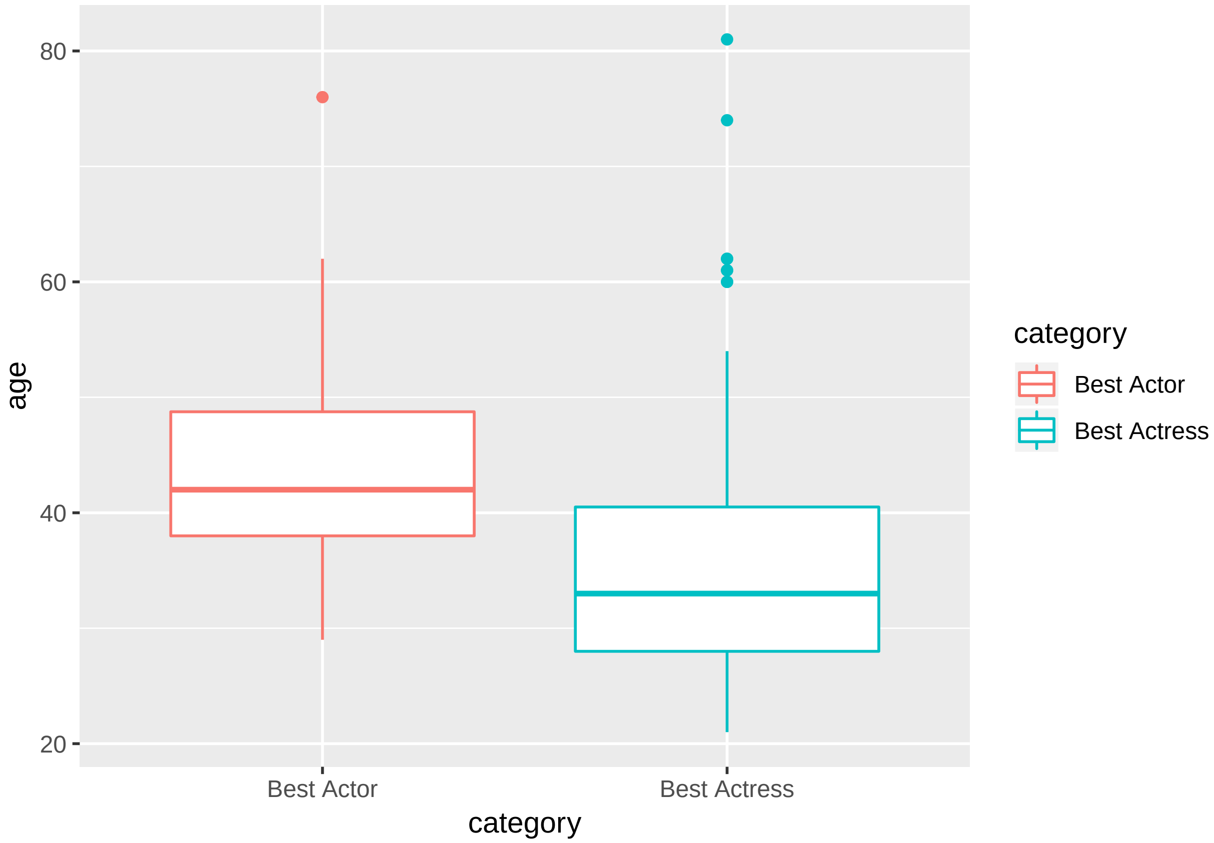
```r
ggplot(oscar_winners,aes(age,fill=category)) + geom_histogram(binwidth=5)
```



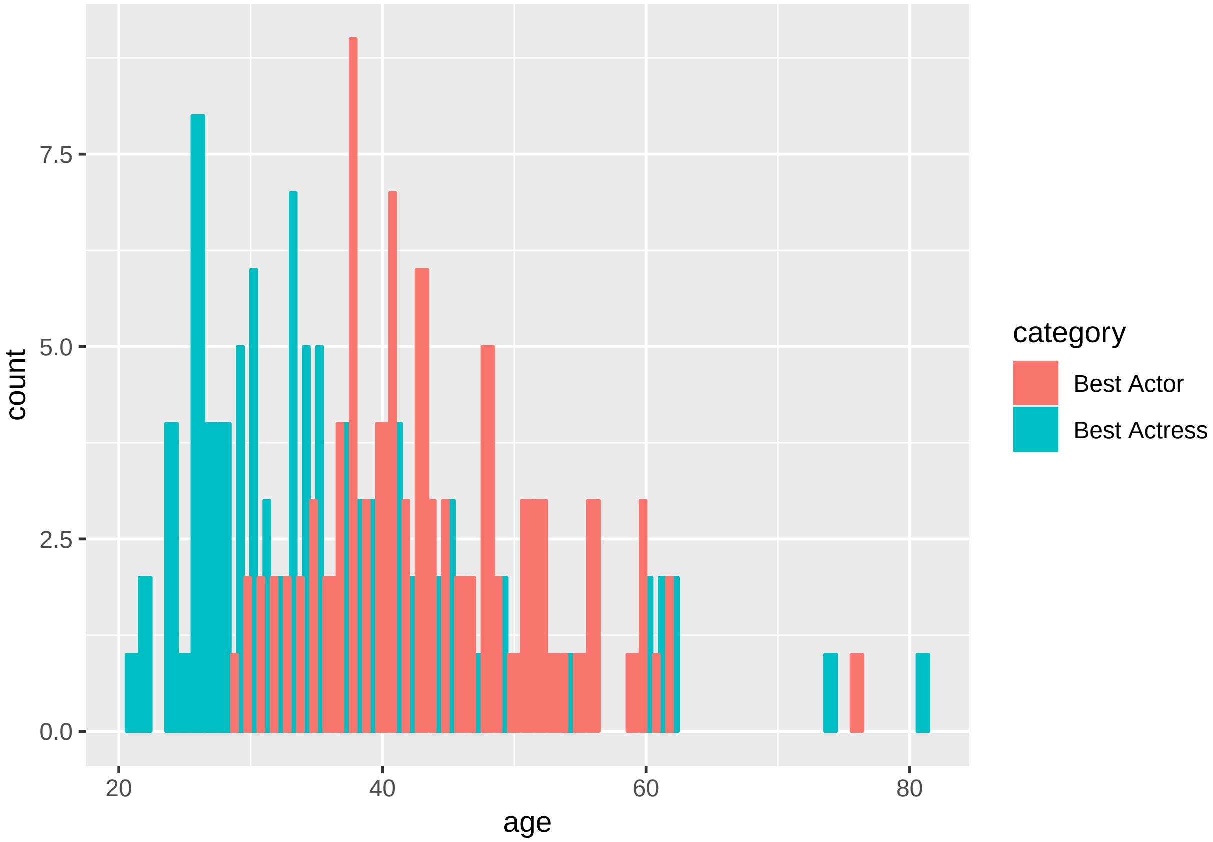
```
ggplot(oscar_winners,aes(age, colour= category)) + geom_histogram(binwidth=5) +
  facet_wrap(~category)
```



```r
ggplot(oscar_winners, aes(x= category, y= age, colour= category)) +
 geom_boxplot()
```



```
ggplot(oscar_winners,aes(age, colour= category)) +
 geom_bar(aes(fill=category), position = 'dodge')
```



Question 6

ACTORS n= 90 mean= 43.822 median= 42 standard error= 8.782

ACTRESS's n=90 mean= 36.0 median= 33 s= 11.58 90th percentile = 49.5 IQR = 12.5

```
oscar_winners %>%
  filter(category=="Best Actor") %>%
  summarise(n=n(), mean=mean(age), median = median(age), s=sd(age),
            nintieth= quantile(age, .9), IQR= IQR(age))
```

```
## # A tibble: 1 x 6
##       n  mean median     s nintieth   IQR
##   <int> <dbl>  <dbl> <dbl>    <dbl> <dbl>
## 1    90  43.8     42  8.88       56  10.8
```

```
oscar_winners %>%
  filter(category=="Best Actress") %>%
  summarise(n=n(), mean=mean(age), median = median(age), s=sd(age),
            nintieth= quantile(age, .9), IQR= IQR(age))
```

```
## # A tibble: 1 x 6
##       n  mean median     s nintieth   IQR
##   <int> <dbl>  <dbl> <dbl>    <dbl> <dbl>
## 1    90  36.0     33  11.6     49.5  12.5
```

Question 7

The distributions of age for Best Actress and Best Actor are different. The winner of Best Actress tends to be younger, since the mean age is 36.0 (compared to 43.8), the median is 33 (compared to 42), and the 90th percentile is 49.5 (compared to 56). This can be seen in the histogram divided by the category as it is clear that the actress datapoints are centered around a mean smaller than the actor's mean. This is also seen in the summary table.