STAT 4003 Final Project

Coles,Mai,Mehrabi,Tabler

link to comic Characters: here

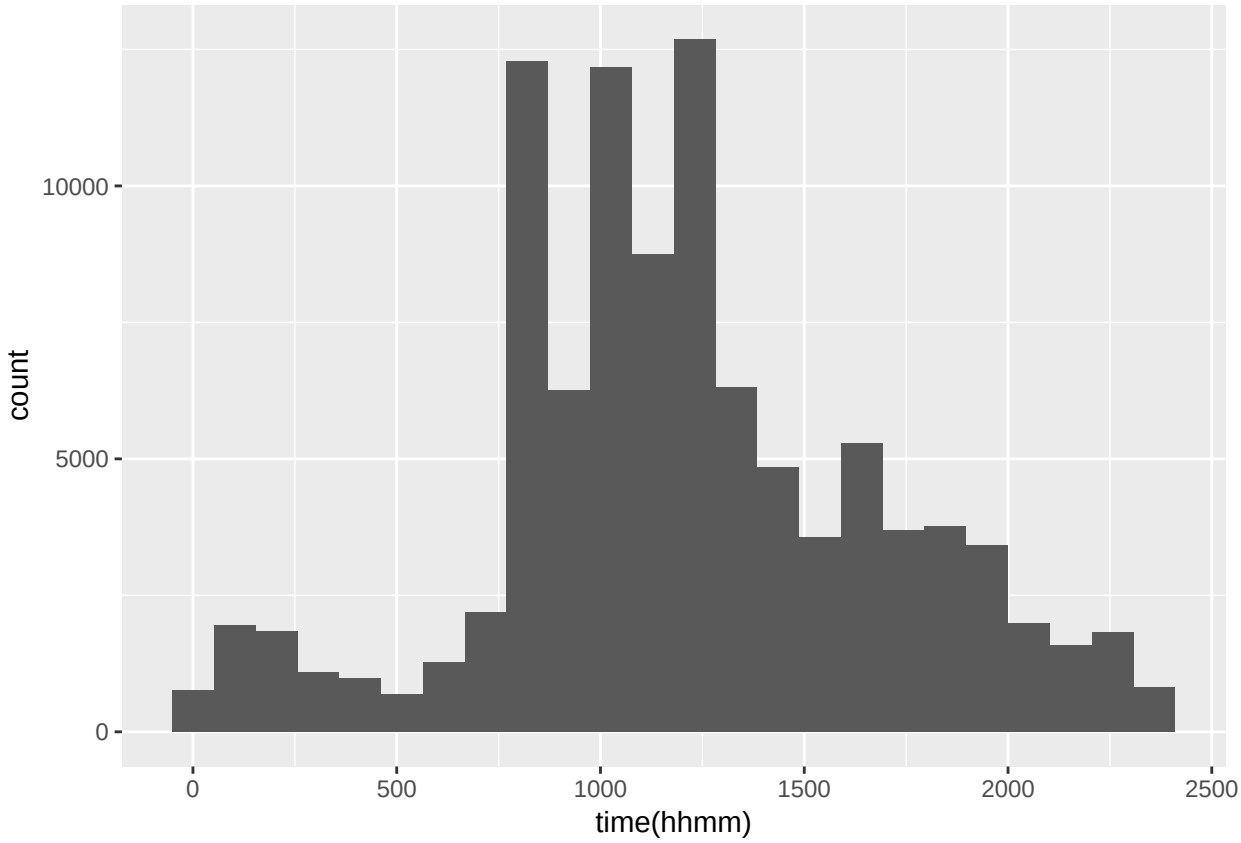
link to parking date: here

```
datcomic<-read.csv(here('Final Project','tidytuesday-comic.csv'))
#datpark<-read.csv(here('project','parking-citations.csv'))
```

There are a tone of data points in the parking data so I subsampled 100000 data points using subsample code. We can use the smaller data set to test out the code and develop statistical hypothesys.

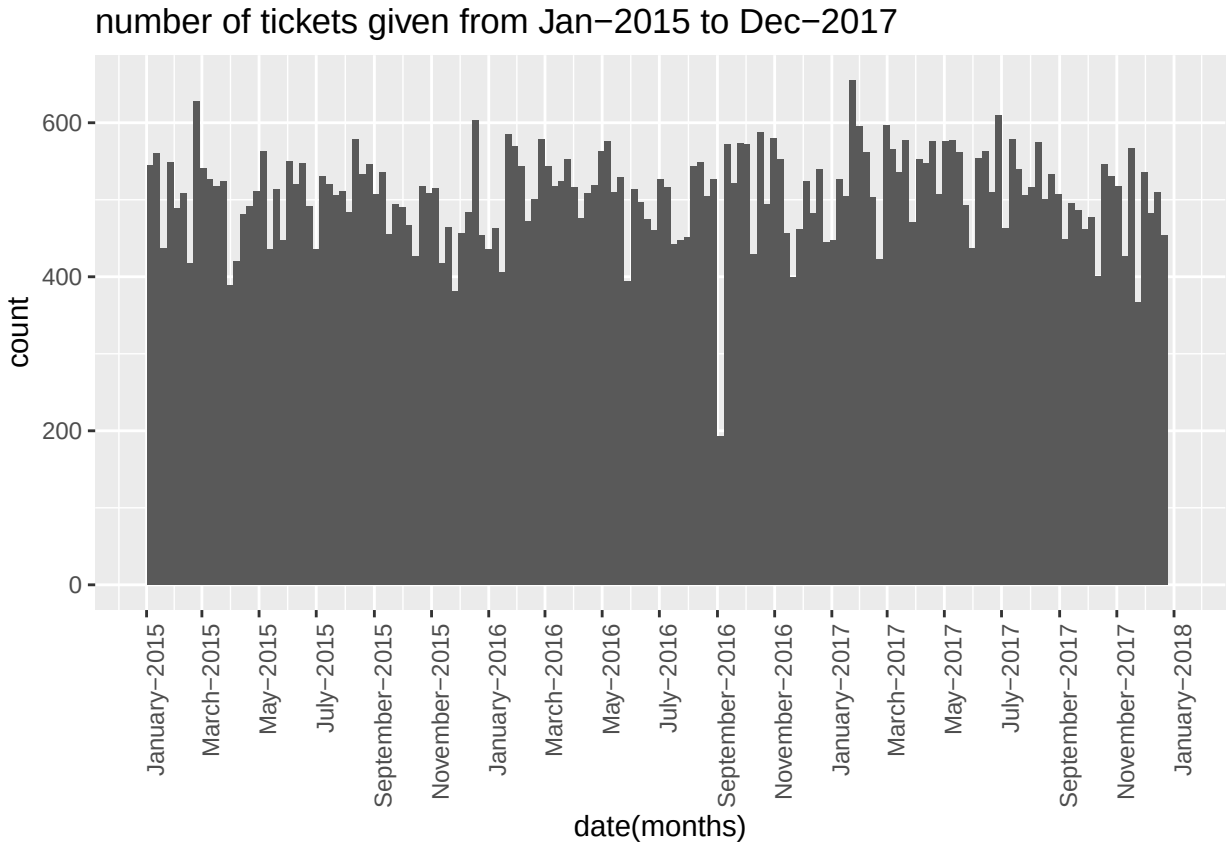
```
#n=100000
#smalldat<- datpark[sample(1:nrow(datpark), n, replace=FALSE),]
#write.csv(smalldat,file=here('Final Project','parking-sub-dat.csv'))
smalldat<-read.csv(here('Final Project','parking-sub-dat.csv'))
```

```
## Warning: Removed 25 rows containing non-finite values (stat_bin).
```

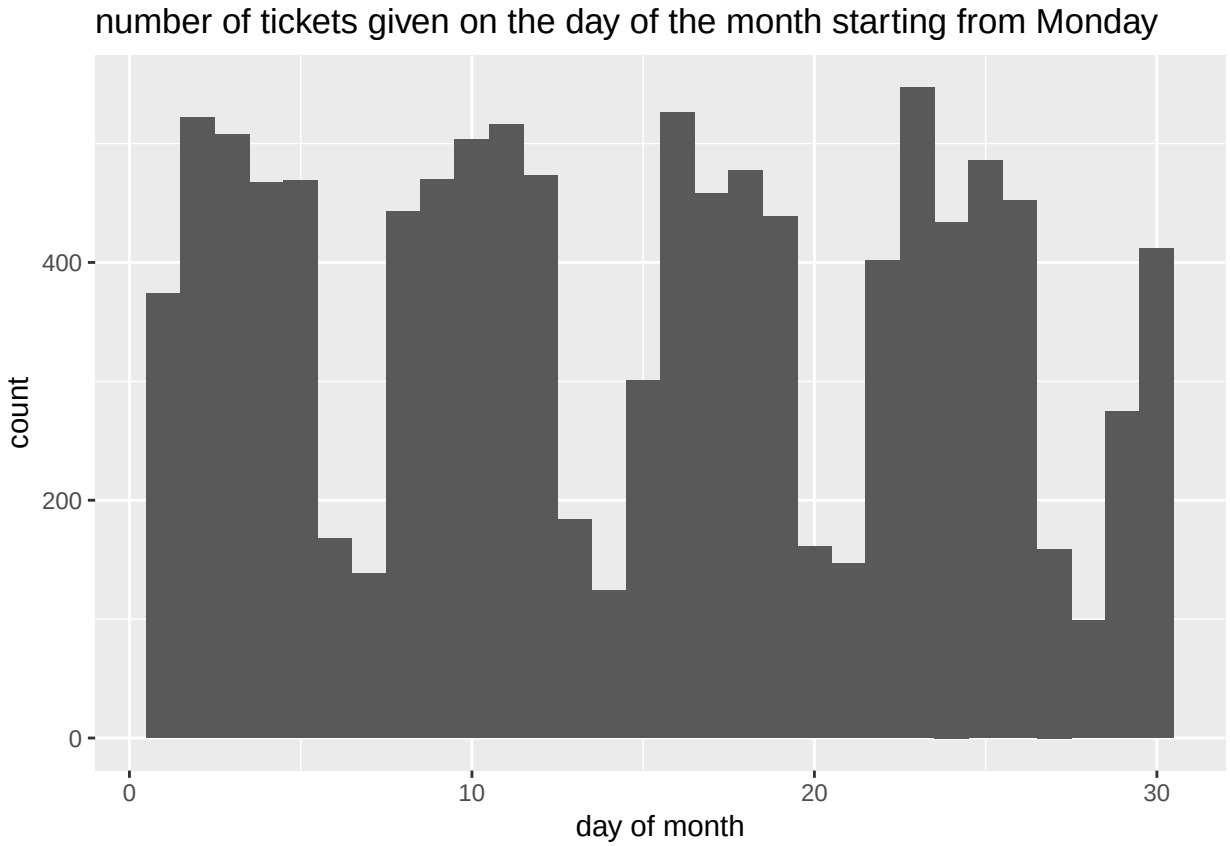



```
## Warning: Removed 21555 rows containing non-finite values (stat_bin).
```

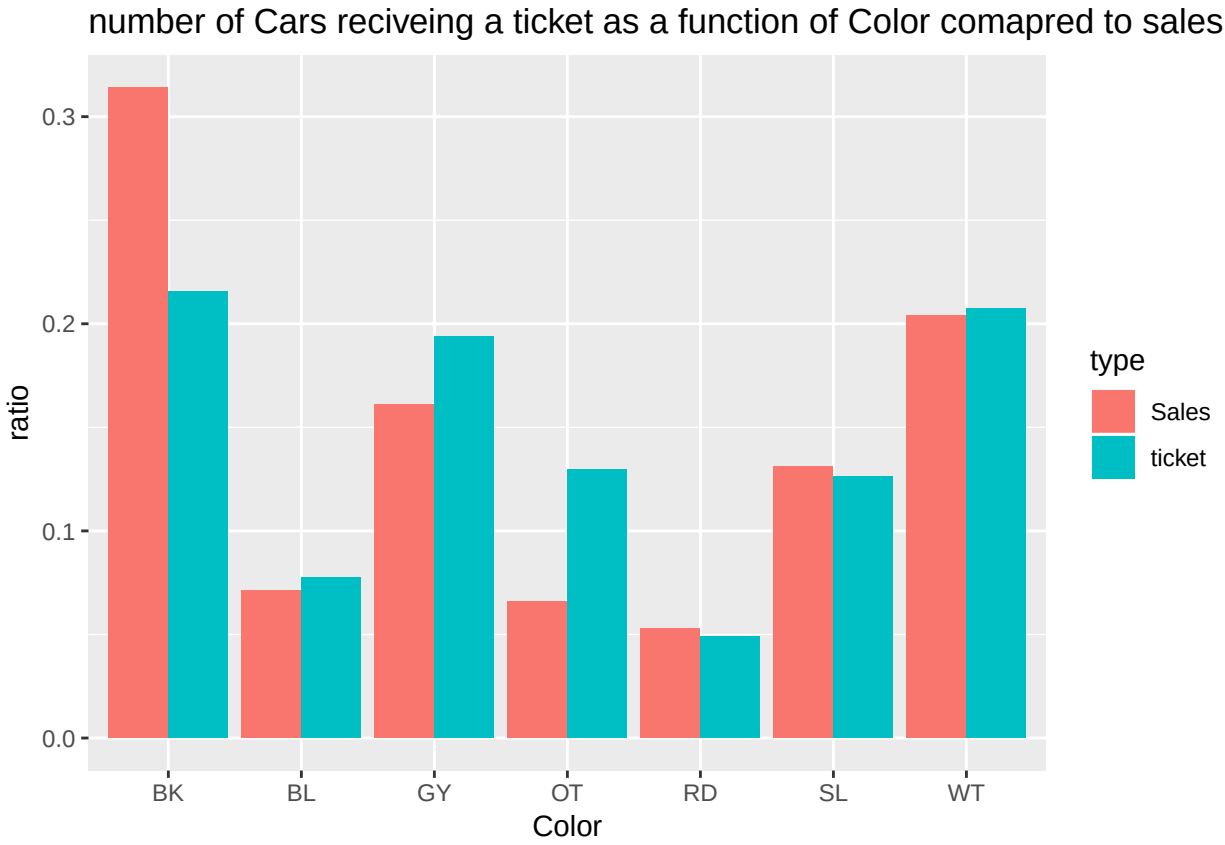
```
## Warning: Removed 2 rows containing missing values (geom_bar).
```



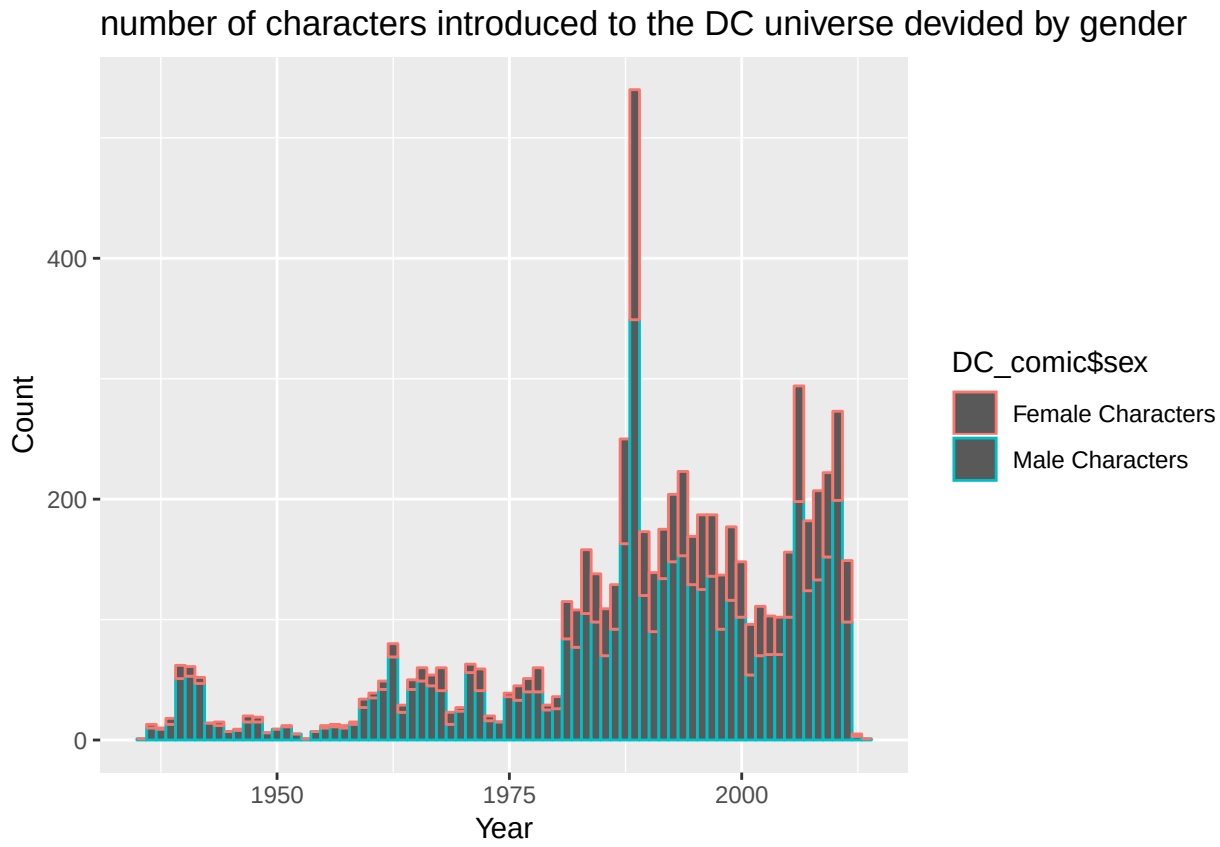
months begining with moday



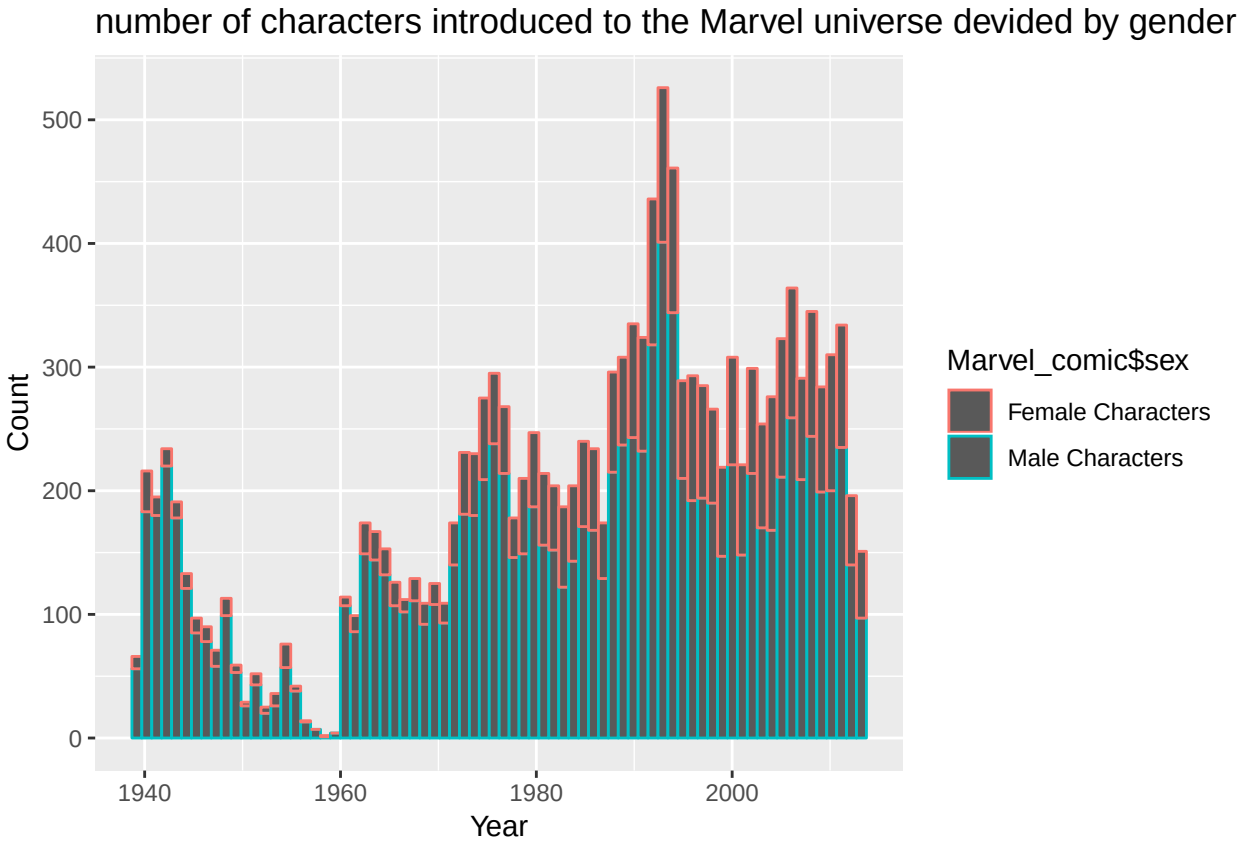

```
##
##  Chi-squared test for given probabilities
##
## data:  TicketCarCol$number
## X-squared = 64231, df = 6, p-value < 2.2e-16
```

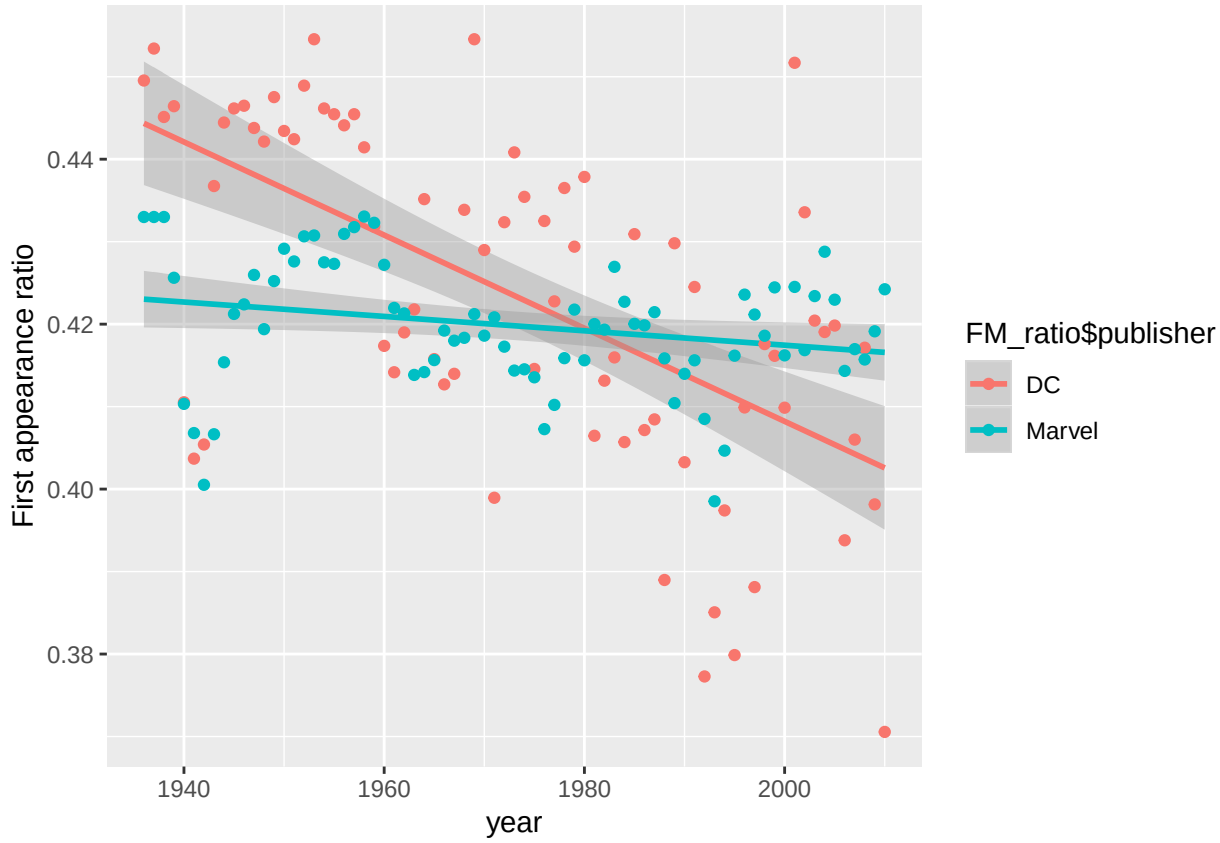



```
## Warning: Removed 1047 rows containing non-finite values (stat_bin).
```

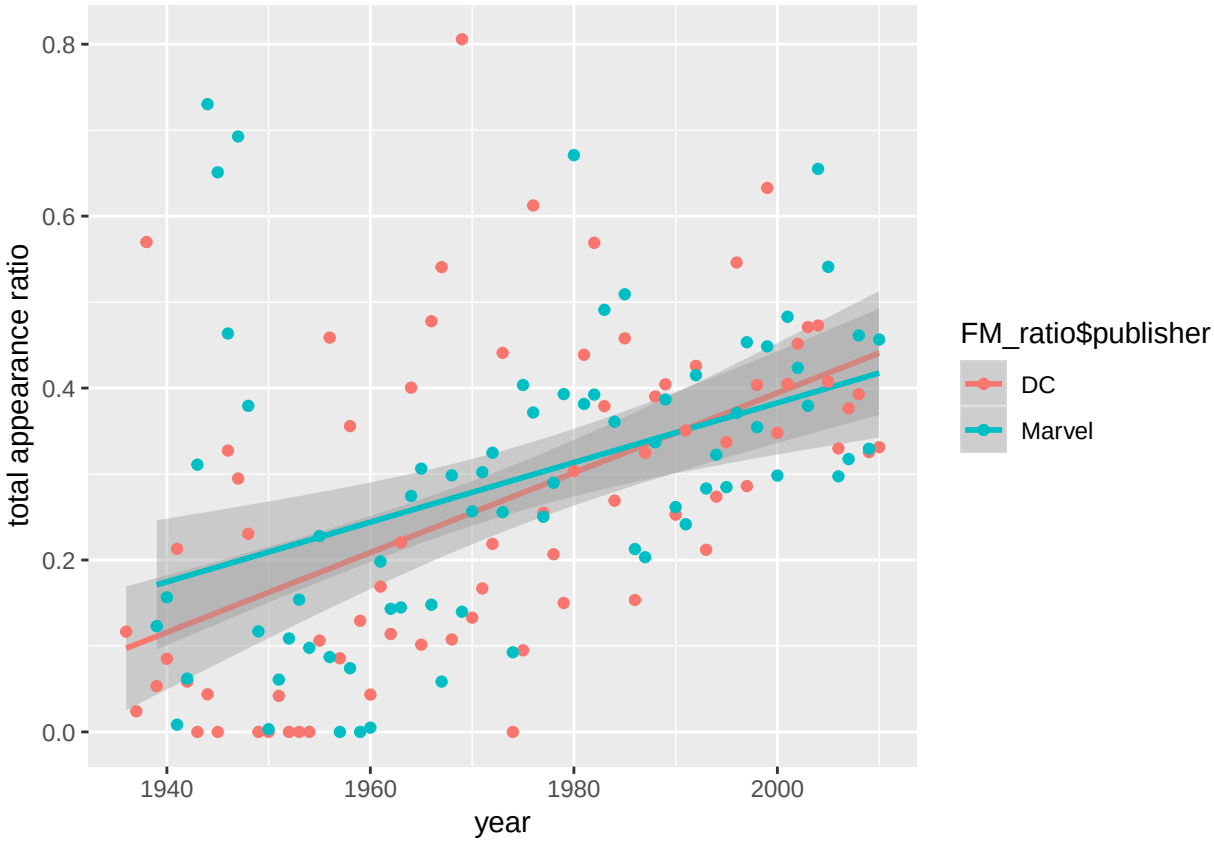



```
## Warning: Removed 1726 rows containing non-finite values (stat_bin).
```


number of characters introduced to the Marvel universe devided by gender
Count
500
400
300
200
100
0
1940
1960
1980
2000
Year
Marvel_comic$sex
Female Characters
Male Characters




```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
## Warning: Removed 3 rows containing missing values (geom_point).
```



Marvel develops female charachters less. This can be supported by the fact that Marvel and DC both appea

```
##
## Call:
## lm(formula = DC_FM_ratio_df$app_ratio ~ DC_FM_ratio_df$year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27363 -0.10928 -0.04427  0.06614  0.55536
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -8.8869702  1.6650636  -5.337 1.02e-06 ***
## DC_FM_ratio_df$year   0.0046406  0.0008439   5.499 5.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1582 on 73 degrees of freedom
## Multiple R-squared:  0.2929, Adjusted R-squared:  0.2832
## F-statistic: 30.24 on 1 and 73 DF,  p-value: 5.346e-07
```

```
##
## Call:
## lm(formula = Marvel_FM_ratio_df$app_ratio ~ Marvel_FM_ratio_df$year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.24049 -0.10817 -0.01964  0.06583  0.54188
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -6.5566421  1.8086050  -3.625 0.000544 ***
## Marvel_FM_ratio_df$year  0.0034697  0.0009159   3.788 0.000318 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1615 on 70 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.1701, Adjusted R-squared:  0.1583
## F-statistic: 14.35 on 1 and 70 DF,  p-value: 0.0003178
```

```
##
## Call:
## lm(formula = Marvel_FM_ratio_df$ratio ~ Marvel_FM_ratio_df$year)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0219862 -0.0035343  0.0005367  0.0059456  0.0119365
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              5.918e-01  7.920e-02   7.472 1.37e-10 ***
## Marvel_FM_ratio_df$year -8.719e-05  4.014e-05  -2.172   0.0331 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007526 on 73 degrees of freedom
## Multiple R-squared:  0.0607, Adjusted R-squared:  0.04783
## F-statistic: 4.717 on 1 and 73 DF,  p-value: 0.03311
```

```
##
## Call:
## lm(formula = DC_FM_ratio_df$ratio ~ DC_FM_ratio_df$year)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.037834 -0.010856  0.004561  0.011131  0.044041
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.538e+00  1.730e-01   8.887 3.04e-13 ***
## DC_FM_ratio_df$year -5.646e-04  8.769e-05  -6.439 1.12e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01644 on 73 degrees of freedom
## Multiple R-squared:  0.3623, Adjusted R-squared:  0.3535
## F-statistic: 41.47 on 1 and 73 DF,  p-value: 1.12e-08
```

```
## Warning in chisq.test(datcomic$eye, datcomic$align): Chi-squared
## approximation may be incorrect
```

```
##
```

```
##  Pearson's Chi-squared test
##
## data:  datcomic$eye and datcomic$align
## X-squared = 423.74, df = 50, p-value < 2.2e-16
```

```
##
##                      Bad Characters Good Characters Reformed Criminals
##   Amber Eyes                      4               4                   0
##   Auburn Hair                     2               2                   0
##   Black Eyeballs                  0               2                   0
##   Black Eyes                    381             315                   0
##   Blue Eyes                     954            1402                   2
##   Brown Eyes                    871            1218                   0
##   Compound Eyes                   1               0                   0
##   Gold Eyes                       7              10                   0
##   Green Eyes                    365             324                   0
##   Grey Eyes                      55              42                   0
##   Hazel Eyes                     38              36                   0
##   Magenta Eyes                    1               1                   0
##   Multiple Eyes                   2               2                   0
##   No Eyes                         4               0                   0
##   One Eye                         9               4                   0
##   Orange Eyes                    21               8                   0
##   Photocellular Eyes             15              28                   0
##   Pink Eyes                       8              13                   0
##   Purple Eyes                    20              13                   0
##   Red Eyes                      436             124                   0
##   Silver Eyes                     5               4                   0
##   Variable Eyes                  25               6                   0
##   Violet Eyes                     6              13                   0
##   White Eyes                    238             150                   0
##   Yellow Eyeballs                 3               2                   0
##   Yellow Eyes                   179              79                   0
```

```
## Warning in chisq.test(datcomic$hair, datcomic$align): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  datcomic$hair and datcomic$align
## X-squared = 534.43, df = 50, p-value < 2.2e-16
```

```
##
##                        Bad Characters Good Characters Reformed Criminals
##   Auburn Hair                      21              34                   0
##   Bald                           1137             329                   0
##   Black Hair                     2205            1863                   0
##   Blond Hair                      719             980                   2
##   Blue Hair                        44              24                   0
##   Bronze Hair                       0               1                   0
##   Brown Hair                     1257            1315                   0
##   Dyed Hair                         0               0                   0
##   Gold Hair                         3               7                   0
##   Green Hair                       68              54                   0
##   Grey Hair                       262             227                   0
```

```
##   Light Brown Hair                  0               5                  0
##   Magenta Hair                      2               2                  0
##   No Hair                           0               0                  0
##   Orange Hair                      13              25                  0
##   Orange-brown Hair                 3               0                  0
##   Pink Hair                        15              14                  0
##   Platinum Blond Hair               1               0                  0
##   Purple Hair                      28              33                  0
##   Red Hair                        345             433                  1
##   Reddish Blond Hair                3               3                  0
##   Reddish Brown Hair                1               2                  0
##   Silver Hair                       5               7                  0
##   Strawberry Blond Hair            17              40                  0
##   Variable Hair                    15               6                  0
##   Violet Hair                       1               2                  0
##   White Hair                      450             404                  0
##   Yellow Hair                       9               4                  0
```

```
## Warning in chisq.test(Chi_hair$hair, Chi_hair$align): Chi-squared
## approximation may be incorrect
```

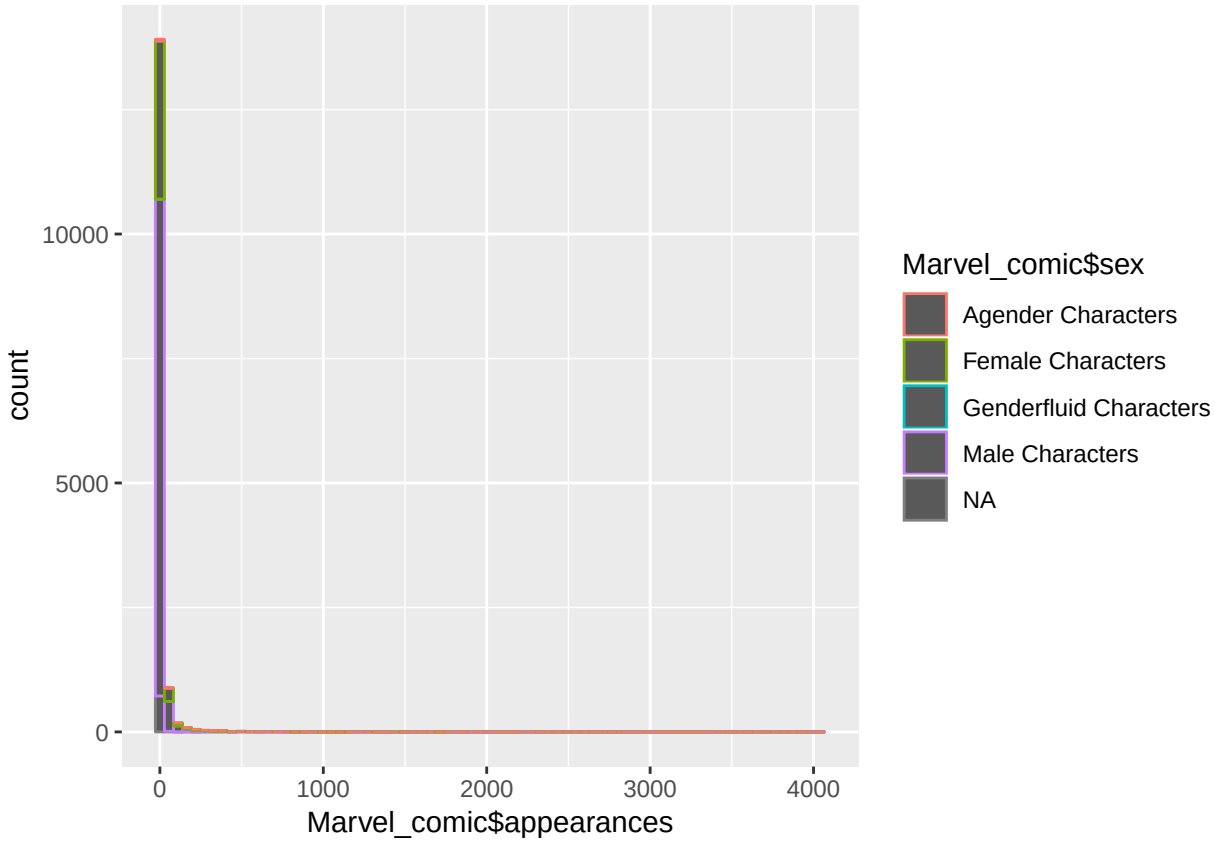
```
##
##  Pearson's Chi-squared test
##
## data:  Chi_hair$hair and Chi_hair$align
## X-squared = 487.21, df = 12, p-value < 2.2e-16
```

```
## Warning in chisq.test(Chi_eyes$eye, Chi_eyes$align): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  Chi_eyes$eye and Chi_eyes$align
## X-squared = 80.444, df = 8, p-value = 3.98e-14
```

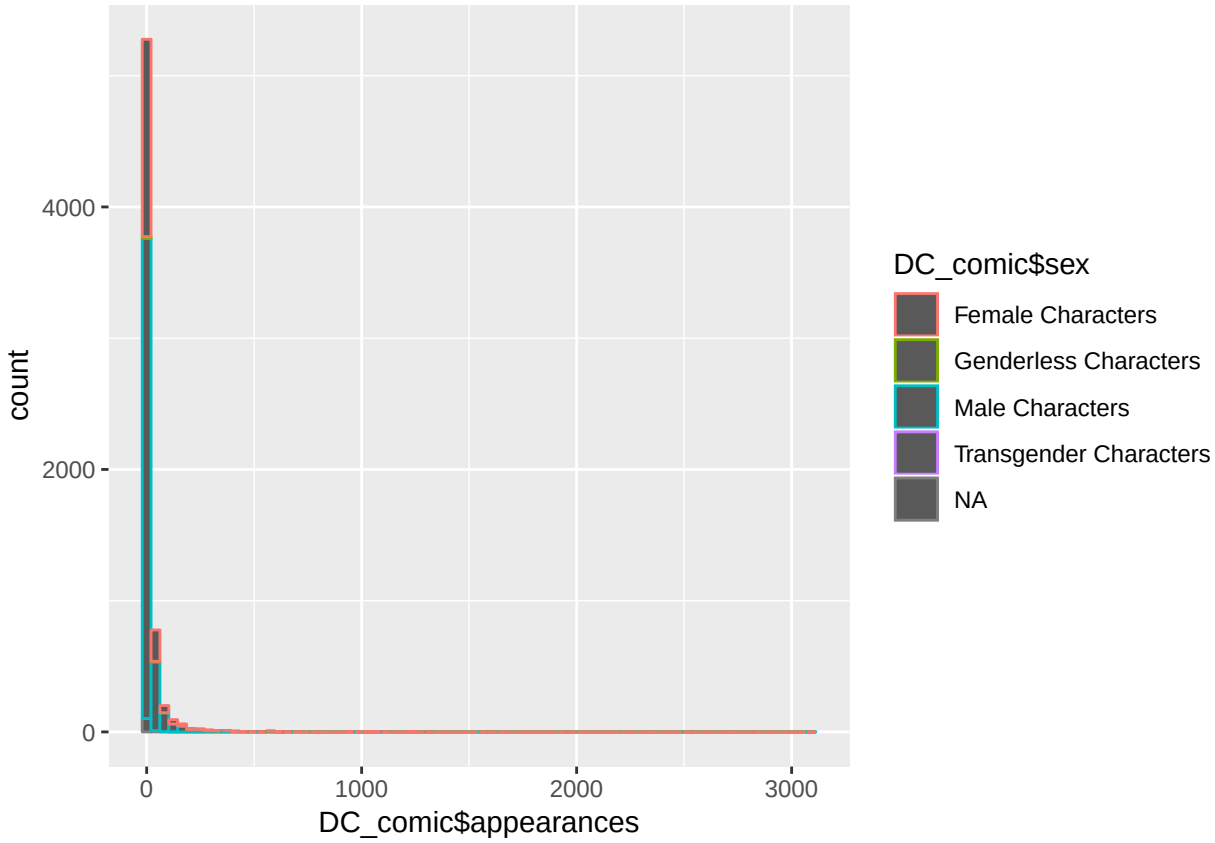
```
##Junk we can put in the appendix
##
ggplot(data=Marvel_comic ,aes(x=Marvel_comic$appearances,color=Marvel_comic$sex))+
geom_histogram(bins=74)
```

```
## Warning: Removed 1096 rows containing non-finite values (stat_bin).
```




```
ggplot(data=DC_comic ,aes(x=DC_comic$appearances,color=DC_comic$sex))+
geom_histogram(bins=76)
```

```
## Warning: Removed 355 rows containing non-finite values (stat_bin).
```


4000
2000
0
count
0
1000
2000
3000
DC_comic$appearances
DC_comic$sex
Female Characters
Genderless Characters
Male Characters
Transgender Characters
NA