data-analysis

September 4, 2017

In [1]:
```
%matplotlib inline

import pandas as pd
import matplotlib.pyplot as plt
from pandas.tools.plotting import table
from __future__ import division
```

In [2]:
```
cols = ['frame_number', 'gender', 'emotion', 'year', 'movie_name']
df = pd.read_csv('/home/mayank/Desktop/gender-bias-detection/data.csv', sep
```

In [3]:
```
df.head()
```

Out[3]:

	frame_number	gender	emotion	year	movie_name
0	28	woman	happy	2014	dedh_ishqiya
1	62	woman	happy	2014	dedh_ishqiya
2	60	man	angry	2014	dedh_ishqiya
3	60	man	sad	2014	dedh_ishqiya
4	60	man	angry	2014	dedh_ishqiya

In [4]:
```
df.gender.value_counts()
```

Out[4]:
```
man      22042
woman    10155
Name: gender, dtype: int64
```

In [5]:
```
df.emotion.value_counts()
```

Out[5]:
```
neutral     8140
sad         7809
happy       6672
angry       5396
fear        2612
surprise    1267
disgust      301
Name: emotion, dtype: int64
```

In [6]:
```
df.year.value_counts()
```

```
Out[6]: 2014    5498
        2016    4506
        2015    4441
        2017    4188
        2013    3397
        2011    2977
        2010    2307
        2012    2292
        2009    1343
        2008    1248
        Name: year, dtype: int64
```

```
In [7]: pd.crosstab(df.gender, df.emotion).apply(lambda r: r/r.sum() * 100, axis=1)
```

```
Out[7]: emotion      angry   disgust      fear      happy    neutral        sad  \
        gender
        man      19.272298  0.852917  7.394973  18.968333  26.721713  23.559568
        woman    11.304776  1.112752  9.670113  24.529788  22.156573  25.760709

        emotion  surprise
        gender
        man      3.230197
        woman    5.465288
```

```
In [17]: # Emotions vs gender
         df1 = pd.crosstab(df.gender, df.emotion).apply(lambda r: r/r.sum() * 100,
         df1
```

```
Out[17]: emotion      angry   disgust      fear      happy    neutral        sad  \
         gender
         man      19.272298  0.852917  7.394973  18.968333  26.721713  23.559568
         woman    11.304776  1.112752  9.670113  24.529788  22.156573  25.760709

         emotion  surprise
         gender
         man      3.230197
         woman    5.465288
```

```
In [16]: # Year vs Gender
         df2 = pd.crosstab(df.gender, df.year).apply(lambda r: r/r.sum() * 100, axi
         df2
```

```
Out[16]: year        2008      2009      2010      2011      2012       2013  \
         gender
         man      3.865348  4.382542  7.258869  9.069050  7.408584  10.557118
         woman    3.899557  3.712457  6.962088  9.630724  6.489414  10.536681

         year         2014       2015       2016       2017
         gender
```

```
man      17.575538  13.424372  13.442519  13.016060
woman    15.992122  14.593796  15.194485  12.988676
```

In [22]:
```
# Emotions across years & gender
df3 = pd.crosstab([df.year, df.emotion], df.gender, ).apply(lambda r:r/r.s
df3
```

Out[22]:

gender		man	woman
year	emotion		
2008	angry	80.952381	19.047619
	disgust	57.142857	42.857143
	fear	67.532468	32.467532
	happy	63.405797	36.594203
	neutral	71.000000	29.000000
	sad	65.714286	34.285714
	surprise	51.020408	48.979592
2009	angry	82.758621	17.241379
	disgust	62.500000	37.500000
	fear	70.238095	29.761905
	happy	66.938776	33.061224
	neutral	74.870466	25.129534
	sad	67.741935	32.258065
	surprise	55.319149	44.680851
2010	angry	77.806122	22.193878
	disgust	84.000000	16.000000
	fear	66.826923	33.173077
	happy	63.485477	36.514523
	neutral	72.897196	27.102804
	sad	66.610455	33.389545
	surprise	61.111111	38.888889
2011	angry	79.376499	20.623501
	disgust	54.545455	45.454545
	fear	63.008130	36.991870
	happy	61.155698	38.844302
	neutral	69.946092	30.053908
	sad	66.995074	33.004926
	surprise	49.038462	50.961538
2012	angry	83.500000	16.500000
	disgust	50.000000	50.000000
...		...	...
2013	sad	66.836735	33.163265
	surprise	57.553957	42.446043
2014	angry	78.173719	21.826281
	disgust	70.175439	29.824561
	fear	63.440860	36.559140
	happy	68.375451	31.624549
	neutral	73.352643	26.647357
	sad	67.902351	32.097649

```
          surprise  61.165049  38.834951
2015      angry     75.216853  24.783147
          disgust   73.529412  26.470588
          fear      60.695187  39.304813
          happy     59.887006  40.112994
          neutral   72.217353  27.782647
          sad       64.950495  35.049505
          surprise  47.368421  52.631579
2016      angry     78.457447  21.542553
          disgust   48.888889  51.111111
          fear      57.800512  42.199488
          happy     57.037037  42.962963
          neutral   70.807453  29.192547
          sad       64.237856  35.762144
          surprise  52.406417  47.593583
2017      angry     76.954178  23.045822
          disgust   59.375000  40.625000
          fear      60.457516  39.542484
          happy     58.169014  41.830986
          neutral   73.932384  26.067616
          sad       66.787659  33.212341
          surprise  66.279070  33.720930

[70 rows x 2 columns]
```

In [ ]: