

Assignment 14

March 7, 2021

0.0.1 Portfolio assignment 14

Perform a bivariate analysis on at least 1 combination of 2 columns with categorical data in the dataset that you chose in portfolio assignment 4. - Do you expect there to be a correlation between the two columns? - Create a contingency table. Do you observe different ratios between categories here? - Create a bar plot for this contingency table. Do you observe different ratios between categories here? - Do a chi-squared test. What does the result say? What's the chance of there being a correlation between the two columns?

```
[1]: import pandas as pd
import seaborn as sns
```

```
[2]: vg = pd.read_csv('vgsales.csv')
vg.head()
```

```
[2]:
```

	Rank	Name	Platform	Year	Genre	Publisher	\
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	41.49	29.02	3.77	8.46	82.74
1	29.08	3.58	6.81	0.77	40.24
2	15.85	12.88	3.79	3.31	35.82
3	15.75	11.01	3.28	2.96	33.00
4	11.27	8.89	10.22	1.00	31.37

I'm going to choose Genre and Platform. There might be a correlation between which platform can make use of a certain genre. For example, loads of people prefer a mouse for shooting games, or a controller for platformers.

```
[3]: table = vg.groupby(['Genre', 'Platform']).size().unstack('Platform', fill_value=0)
table
```

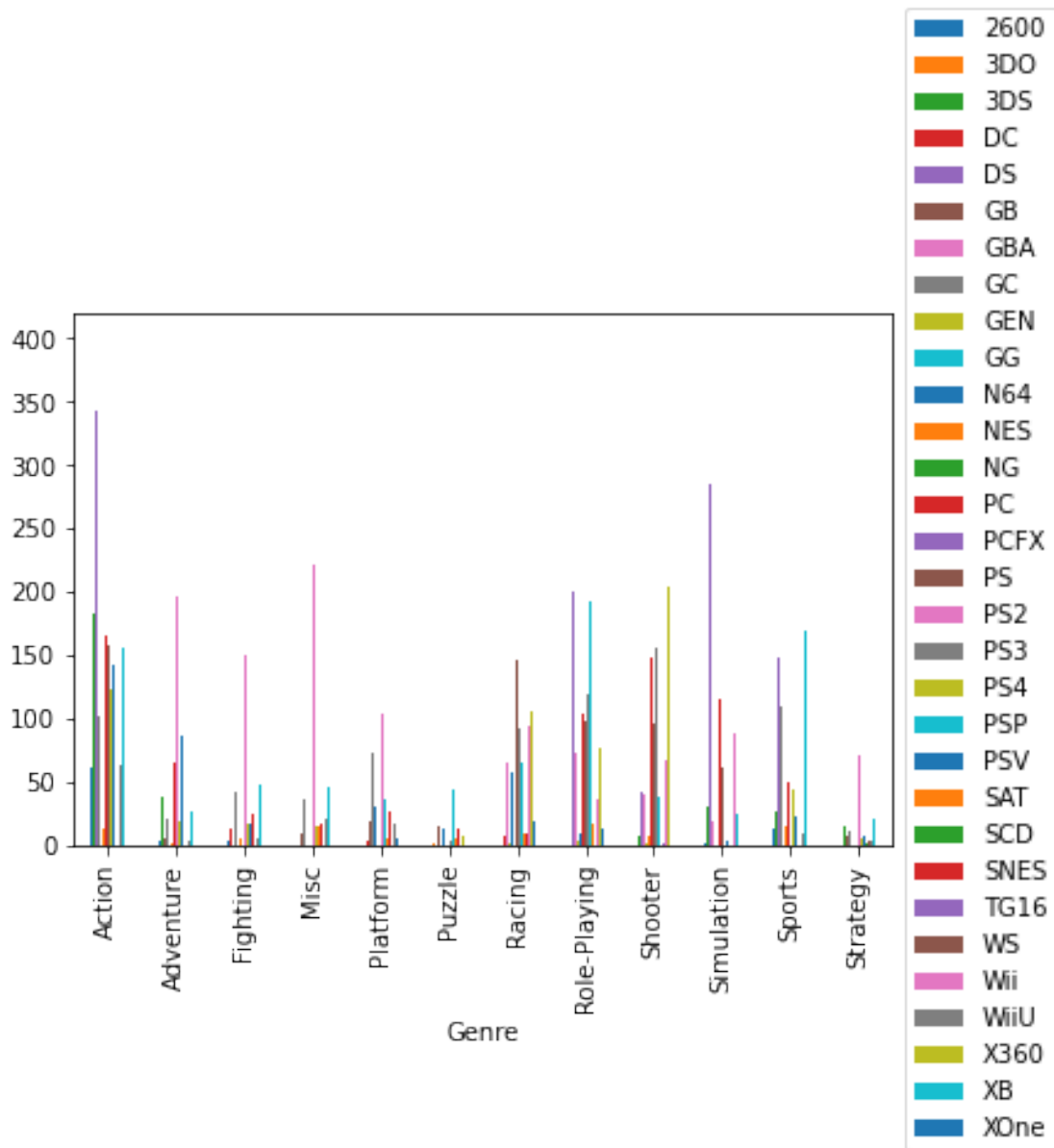
```
[3]: Platform    2600  3DO  3DS  DC  DS  GB  GBA  GC  GEN  GG  ...  SAT  SCD  \
Genre
```

Action	61	0	182	3	343	6	167	101	3	0	...	3	0
Adventure	2	1	37	11	240	5	38	20	2	0	...	26	0
Fighting	2	0	14	12	36	0	23	42	5	0	...	31	0
Misc	5	0	53	0	393	8	110	36	1	0	...	15	2
Platform	9	0	28	2	92	19	142	73	7	1	...	5	1
Puzzle	11	1	20	0	238	15	41	13	0	0	...	5	0
Racing	6	0	11	6	67	2	64	63	1	0	...	8	1
Role-Playing	0	0	86	4	200	21	73	27	3	0	...	17	1
Shooter	24	0	7	3	42	1	40	48	1	0	...	22	0
Simulation	1	1	30	1	285	5	18	12	0	0	...	7	0
Sports	12	0	26	10	148	9	88	110	3	0	...	16	0
Strategy	0	0	15	0	79	7	18	11	1	0	...	18	1

Platform	SNES	TG16	WS	Wii	WiiU	X360	XB	XOne
Genre								
Action	12	0	0	238	63	324	155	68
Adventure	4	1	0	84	3	47	26	12
Fighting	25	0	0	42	5	65	48	7
Misc	17	0	0	280	21	126	46	15
Platform	26	0	0	58	16	24	49	4
Puzzle	13	0	0	55	4	7	7	0
Racing	9	0	0	94	3	105	123	19
Role-Playing	50	0	4	35	6	76	23	13
Shooter	10	1	0	66	10	203	132	33
Simulation	9	0	0	87	1	40	24	3
Sports	49	0	0	261	8	220	170	36
Strategy	15	0	2	25	3	28	21	3

[12 rows x 31 columns]

```
[4]: table.plot.bar().legend(loc='center left',bbox_to_anchor=(1.0, 0.5));
```



Reading data from this is going to be very hard, but you can definitely see some correlation.

```
[5]: from scipy.stats import chi2_contingency
     chi2_contingency(table)
```

```
[5]: (5909.978728846863,
     0.0,
     330,
     array([[2.65711532e+01, 5.99349319e-01, 1.01689601e+02, 1.03887215e+01,
            4.32130859e+02, 1.95787444e+01, 1.64221713e+02, 1.11079407e+02,
```

5.39414387e+00, 1.99783106e-01, 6.37308109e+01, 1.95787444e+01,
 2.39739728e+00, 1.91791782e+02, 1.99783106e-01, 2.38940595e+02,
 4.31731293e+02, 2.65511748e+02, 6.71271237e+01, 2.42336908e+02,
 8.25104229e+01, 3.45624774e+01, 1.19869864e+00, 4.77481624e+01,
 3.99566213e-01, 1.19869864e+00, 2.64712616e+02, 2.85689842e+01,
 2.52725630e+02, 1.64621280e+02, 4.25538017e+01],
 [1.03047355e+01, 2.32437643e-01, 3.94369201e+01, 4.02891915e+00,
 1.67587541e+02, 7.59296301e+00, 6.36879142e+01, 4.30784432e+01,
 2.09193879e+00, 7.74792144e-02, 2.47158694e+01, 7.59296301e+00,
 9.29750572e-01, 7.43800458e+01, 7.74792144e-02, 9.26651404e+01,
 1.67432582e+02, 1.02969876e+02, 2.60330160e+01, 9.39822870e+01,
 3.19989155e+01, 1.34039041e+01, 4.64875286e-01, 1.85175322e+01,
 1.54958429e-01, 4.64875286e-01, 1.02659959e+02, 1.10795277e+01,
 9.80112062e+01, 6.38428726e+01, 1.65030727e+01],
 [6.79503555e+00, 1.53271478e-01, 2.60050609e+01, 2.65670563e+00,
 1.10508736e+02, 5.00686830e+00, 4.19963851e+01, 2.84063140e+01,
 1.37944331e+00, 5.10904928e-02, 1.62978672e+01, 5.00686830e+00,
 6.13085914e-01, 4.90468731e+01, 5.10904928e-02, 6.11042294e+01,
 1.10406555e+02, 6.78992650e+01, 1.71664056e+01, 6.19727678e+01,
 2.11003735e+01, 8.83865526e+00, 3.06542957e-01, 1.22106278e+01,
 1.02180986e-01, 3.06542957e-01, 6.76949030e+01, 7.30594047e+00,
 6.46294734e+01, 4.20985661e+01, 1.08822750e+01],
 [1.39346307e+01, 3.14314978e-01, 5.33287746e+01, 5.44812628e+00,
 2.26621099e+02, 1.02676226e+01, 8.61223039e+01, 5.82530425e+01,
 2.82883480e+00, 1.04771659e-01, 3.34221593e+01, 1.02676226e+01,
 1.25725991e+00, 1.00580793e+02, 1.04771659e-01, 1.25306904e+02,
 2.26411556e+02, 1.39241535e+02, 3.52032775e+01, 1.27088023e+02,
 4.32706953e+01, 1.81254970e+01, 6.28629955e-01, 2.50404266e+01,
 2.09543318e-01, 6.28629955e-01, 1.38822448e+02, 1.49823473e+01,
 1.32536149e+02, 8.63318472e+01, 2.23163634e+01],
 [7.09953006e+00, 1.60139776e-01, 2.71703820e+01, 2.77575612e+00,
 1.15460778e+02, 5.23123268e+00, 4.38782986e+01, 2.96792385e+01,
 1.44125798e+00, 5.33799253e-02, 1.70281962e+01, 5.23123268e+00,
 6.40559104e-01, 5.12447283e+01, 5.33799253e-02, 6.38423906e+01,
 1.15354019e+02, 7.09419207e+01, 1.79356549e+01, 6.47498494e+01,
 2.20459091e+01, 9.23472708e+00, 3.20279552e-01, 1.27578021e+01,
 1.06759851e-01, 3.20279552e-01, 7.07284010e+01, 7.63332932e+00,
 6.75256055e+01, 4.39850584e+01, 1.13699241e+01],
 [4.66357392e+00, 1.05193397e-01, 1.78478130e+01, 1.82335221e+00,
 7.58444391e+01, 3.43631763e+00, 2.88229907e+01, 1.94958429e+01,
 9.46740571e-01, 3.50644656e-02, 1.11855645e+01, 3.43631763e+00,
 4.20773587e-01, 3.36618870e+01, 3.50644656e-02, 4.19371009e+01,
 7.57743102e+01, 4.66006748e+01, 1.17816604e+01, 4.25331968e+01,
 1.44816243e+01, 6.06615255e+00, 2.10386794e-01, 8.38040728e+00,
 7.01289312e-02, 2.10386794e-01, 4.64604169e+01, 5.01421858e+00,
 4.43565490e+01, 2.88931197e+01, 7.46873117e+00],
 [1.00082540e+01, 2.25750090e-01, 3.83022653e+01, 3.91300157e+00,

1.62765815e+02, 7.37450295e+00, 6.18555248e+01, 4.18390167e+01,
 2.03175081e+00, 7.52500301e-02, 2.40047596e+01, 7.37450295e+00,
 9.03000361e-01, 7.22400289e+01, 7.52500301e-02, 8.99990360e+01,
 1.62615315e+02, 1.00007290e+02, 2.52840101e+01, 9.12782865e+01,
 3.10782624e+01, 1.30182552e+01, 4.51500181e-01, 1.79847572e+01,
 1.50500060e-01, 4.51500181e-01, 9.97062899e+01, 1.07607543e+01,
 9.51912881e+01, 6.20060248e+01, 1.60282564e+01],
 [1.19233643e+01, 2.68948066e-01, 4.56315219e+01, 4.66176648e+00,
 1.93911556e+02, 8.78563682e+00, 7.36917701e+01, 4.98450416e+01,
 2.42053259e+00, 8.96493553e-02, 2.85981444e+01, 8.78563682e+00,
 1.07579226e+00, 8.60633811e+01, 8.96493553e-02, 1.07220629e+02,
 1.93732257e+02, 1.19143993e+02, 3.01221834e+01, 1.08744668e+02,
 3.70251838e+01, 1.55093385e+01, 5.37896132e-01, 2.14261959e+01,
 1.79298711e-01, 5.37896132e-01, 1.18785396e+02, 1.28198578e+01,
 1.13406435e+02, 7.38710688e+01, 1.90953127e+01],
 [1.04970478e+01, 2.36775515e-01, 4.01729124e+01, 4.10410893e+00,
 1.70715146e+02, 7.73466683e+00, 6.48764911e+01, 4.38823955e+01,
 2.13097964e+00, 7.89251717e-02, 2.51771298e+01, 7.73466683e+00,
 9.47102060e-01, 7.57681648e+01, 7.89251717e-02, 9.43945054e+01,
 1.70557296e+02, 1.04891553e+02, 2.65188577e+01, 9.57362333e+01,
 3.25960959e+01, 1.36540547e+01, 4.73551030e-01, 1.88631160e+01,
 1.57850343e-01, 4.73551030e-01, 1.04575853e+02, 1.12862996e+01,
 9.98403422e+01, 6.50343415e+01, 1.68110616e+01],
 [6.94728281e+00, 1.56705627e-01, 2.65877214e+01, 2.71623087e+00,
 1.12984757e+02, 5.11905049e+00, 4.29373418e+01, 2.90427762e+01,
 1.41035064e+00, 5.22352091e-02, 1.66630317e+01, 5.11905049e+00,
 6.26822509e-01, 5.01458007e+01, 5.22352091e-02, 6.24733100e+01,
 1.12880287e+02, 6.94205928e+01, 1.75510302e+01, 6.33613086e+01,
 2.15731413e+01, 9.03669117e+00, 3.13411254e-01, 1.24842150e+01,
 1.04470418e-01, 3.13411254e-01, 6.92116520e+01, 7.46963490e+00,
 6.60775395e+01, 4.30418123e+01, 1.11260995e+01],
 [1.87985299e+01, 4.24026991e-01, 7.19432462e+01, 7.34980118e+00,
 3.05723461e+02, 1.38515484e+01, 1.16183396e+02, 7.85863357e+01,
 3.81624292e+00, 1.41342330e-01, 4.50882034e+01, 1.38515484e+01,
 1.69610796e+00, 1.35688637e+02, 1.41342330e-01, 1.69045427e+02,
 3.05440776e+02, 1.87843957e+02, 4.74910230e+01, 1.71448247e+02,
 5.83743825e+01, 2.44522232e+01, 8.48053982e-01, 3.37808170e+01,
 2.82684661e-01, 8.48053982e-01, 1.87278588e+02, 2.02119532e+01,
 1.78798048e+02, 1.16466080e+02, 3.01059164e+01],
 [5.45686227e+00, 1.23087119e-01, 2.08837812e+01, 2.13351006e+00,
 8.87458127e+01, 4.02084589e+00, 3.37258706e+01, 2.28121460e+01,
 1.10778407e+00, 4.10290396e-02, 1.30882636e+01, 4.02084589e+00,
 4.92348476e-01, 3.93878781e+01, 4.10290396e-02, 4.90707314e+01,
 8.86637547e+01, 5.45275937e+01, 1.37857573e+01, 4.97682251e+01,
 1.69449934e+01, 7.09802386e+00, 2.46174238e-01, 9.80594047e+00,
 8.20580793e-02, 2.46174238e-01, 5.43634775e+01, 5.86715267e+00,
 5.19017351e+01, 3.38079287e+01, 8.73918544e+00]]))

So it seems the score of my chi-squared test is 5900, that's an awfully big number. I'm assuming the test is very confident there is a correlation between these two categorical sets.