

Exercises

Exercises 1-4 deal with the Tips data set (`tips.csv`).

Exercise 1. Make a visualization that displays the relationship between the day of the week and party size.

```
In [ ]: # ENTER YOUR CODE HERE
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

tips_df = pd.read_csv("tips.csv")
tips_df.head()
```

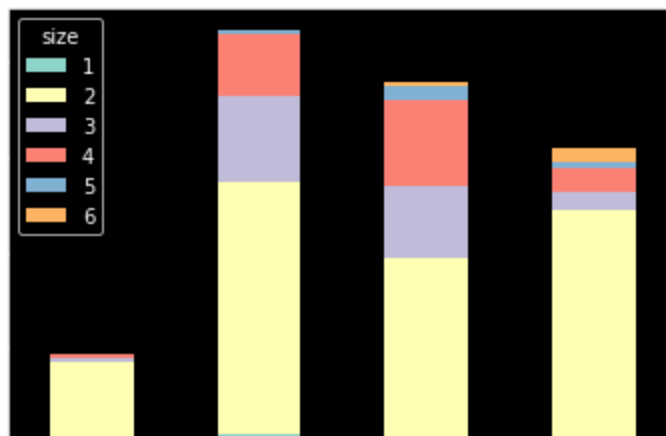
```
Out[ ]:   obs  totbill   tip sex smoker day  time  size
0    1    16.99  1.01  F   No   Sun  Night    2
1    2    10.34  1.66  M   No   Sun  Night    3
2    3    21.01  3.50  M   No   Sun  Night    3
3    4    23.68  3.31  M   No   Sun  Night    2
4    5    24.59  3.61  F   No   Sun  Night    4
```

```
In [ ]: relationship = pd.crosstab(tips_df["day"], tips_df["size"])
relationship

relationship.plot.bar(stacked = True)
```

```
Out[ ]: size  1   2   3   4   5   6
day
Fri  1  16   1   1   0   0
Sat  2  53  18  13   1   0
Sun  0  39  15  18   3   1
Thu  1  48   4   5   1   3
```

```
Out[ ]: <AxesSubplot:xlabel='day'>
```



Exercise 2. Calculate the marginal distribution of day of week in two different ways.

```
In [ ]: # ENTER YOUR CODE HERE
print(relationship.sum().sum())
joint = relationship / relationship.sum().sum()
joint

joint.sum(axis = 1)
```

244

```
Out[ ]: size      1      2      3      4      5      6
      day
      ---
Fri  0.004098  0.065574  0.004098  0.004098  0.000000  0.000000
Sat  0.008197  0.217213  0.073770  0.053279  0.004098  0.000000
Sun  0.000000  0.159836  0.061475  0.073770  0.012295  0.004098
Thu  0.004098  0.196721  0.016393  0.020492  0.004098  0.012295
```

```
Out[ ]: day
Fri    0.077869
Sat    0.356557
Sun    0.311475
Thu    0.254098
dtype: float64
```

```
In [ ]: pd.crosstab(tips_df.time, tips_df.day,
                    normalize=True, margins=True)
```

```
Out[ ]: day      Fri      Sat      Sun      Thu      All
      time
Day  0.028689  0.000000  0.000000  0.250000  0.278689
Night 0.049180  0.356557  0.311475  0.004098  0.721311
All  0.077869  0.356557  0.311475  0.254098  1.000000
```

```
In [ ]: days = tips_df.groupby("day").count()["size"]
days
days2 = days / days.sum()
days2
```

```
Out[ ]: day
Fri    19
Sat    87
Sun    76
Thu    62
Name: size, dtype: int64
```

```
Out[ ]: day
Fri    0.077869
Sat    0.356557
Sun    0.311475
Thu    0.254098
Name: size, dtype: float64
```

Exercise 3. Make a visualization that displays the conditional distribution of party size, given the

day of the week.

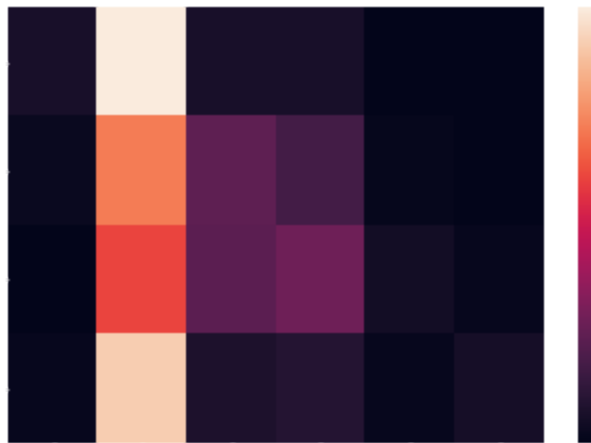
```
In [ ]: # ENTER YOUR CODE HERE
print(relationship)
pclass_counts = relationship.sum(axis=1)
relative = relationship.divide(pclass_counts, axis=0)

print(relative)

sns.heatmap(relative)
```

```
size  1   2   3   4   5   6
day
Fri   1  16   1   1   0   0
Sat   2  53  18  13   1   0
Sun   0  39  15  18   3   1
Thu   1  48   4   5   1   3
size          1          2          3          4          5          6
day
Fri   0.052632  0.842105  0.052632  0.052632  0.000000  0.000000
Sat   0.022989  0.609195  0.206897  0.149425  0.011494  0.000000
Sun   0.000000  0.513158  0.197368  0.236842  0.039474  0.013158
Thu   0.016129  0.774194  0.064516  0.080645  0.016129  0.048387
```

Out[]: <AxesSubplot:xlabel='size', ylabel='day'>



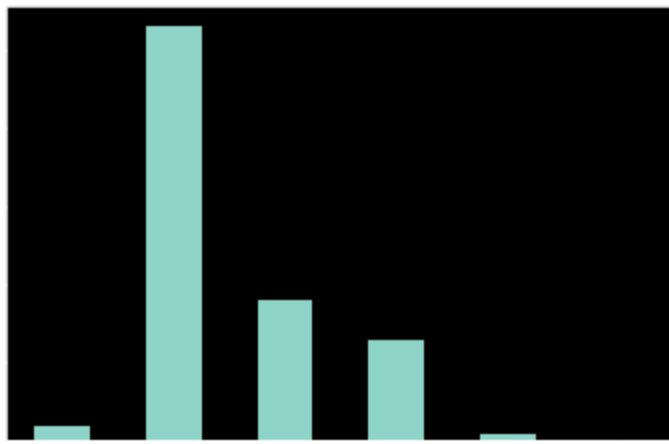
Exercise 4. What proportion of Saturday parties had 2 people? Is this the same as the proportion of 2-person parties that dined on Saturday?

```
In [ ]: # ENTER YOUR CODE HERE
saturday_distribution = relationship.loc["Sat"]
saturday_distribution

saturday_distribution.plot.bar()
```

```
Out[ ]: size
1      2
2     53
3     18
4     13
5      1
6      0
Name: Sat, dtype: int64
```

Out[]: <AxesSubplot:xlabel='size'>



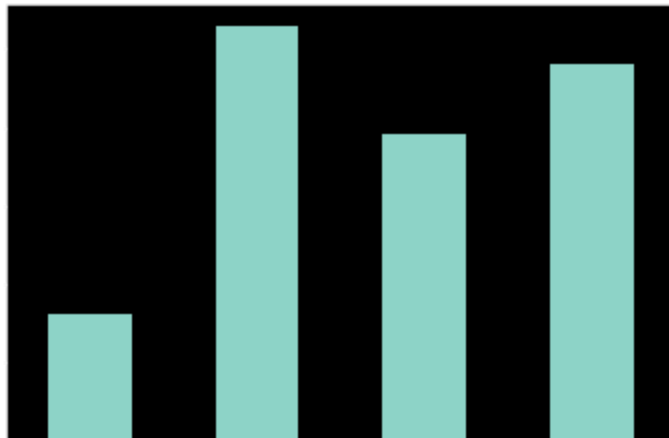
```
In [ ]: saturday_distribution[2] / saturday_distribution.sum()
```

```
Out[ ]: 0.6091954022988506
```

```
In [ ]: two_distribution = relationship[2]
two_distribution["Sat"] / two_distribution.sum()
two_distribution.plot.bar()
```

```
Out[ ]: 0.33974358974358976
```

```
Out[ ]: <AxesSubplot:xlabel='day'>
```



The proportion of Saturday parties with 2 people (60%) is NOT the same as the proportion of 2 person parties on Saturday (33%)