

2.3 Data Cubes and Pivot Tables

May 8, 2019

1 2.3 Data Cubes and Pivot Tables

```
In [1]: %matplotlib inline
import numpy as np
import pandas as pd
titanic_df = pd.read_csv("https://raw.githubusercontent.com/dlsun/data-science-book/master/titanic.csv")

titanic_df["adult"] = (titanic_df["age"] >= 18)
```

In Section 2.2, we learned to split a pandas DataFrame and apply the same analysis to each of the resulting, smaller DataFrames. For example, the following code calculates the proportion of Titanic passengers of each sex, age group, and class who survived:

```
In [2]: survivors_table = (titanic_df.
                           groupby(["sex", "adult", "pclass"]).
                           survived.
                           mean())
survivors_table.to_frame()
```

```
Out[2]:
```

			survived
sex	adult	pclass	
female	False	1	0.947368
		2	0.952381
		3	0.536364
	True	1	0.968000
		2	0.870588
		3	0.443396
male	False	1	0.400000
		2	0.464286
		3	0.147059
	True	1	0.326389
		2	0.083916
		3	0.155709

Here's another way to think about these results: there are three dimensions, `sex`, `adult`, and `pclass`, and we calculate a metric, the proportion of survivors, for each of the $2 \times 2 \times 3 = 12$ possible combinations of the dimension values.