

What we have done this week:

The datasets we're using: 2015-2019 Asians in MA

1. Preprocessed the ACS dataset from IPUMS.org, and the following is the report:

```
Total num of rows in the dataset: 22349
```

```
Number of people in laborforce: 15955
```

```
Number of people who reported their income: 18787
```

```
Number of people who reported their wage/salary: 18544
```

```
Classify by City:
```

```
Unknown      15721
```

```
Boston       3611
```

```
Lowell        1080
```

```
Cambridge     1076
```

```
Worcester     663
```

```
Springfield   198
```

```
Name: CITY, dtype: int64
```

```
Classify by Sex:
```

```
Female      11847
```

```
Male        10502
```

```
Name: SEX, dtype: int64
```

```
Classify by Asian Subgroups:
```

```
Chinese      8811
```

```
Indian       5210
```

```
Other Asians 2963
```

```
Vietnamese   2330
```

```
Korean       1384
```

```
Filipino     747
```

```
Japanese     592
```

```
Thai         186
```

```
Indonesian   96
```

```
Malaysian    30
```

```
Name: RACED, dtype: int64
```

```
Classify by Employment Status:
```

```
Employed     11603
```

```
Not in labor force 6394
```

```
Unemployed    547
```

```
Unknown       3805
```

```
Name: EMPSTAT, dtype: int64
```

```
Classify by Occupation:
```

```
10           102
```

```
20           62
```

```
30           100
```

```
100          8
```

```
110          132
```

```
...
```

```
9720         1
```

```
9810         1
```

```
9820         2
```

```
9830         4
```

```
9920        8156
```

```
Name: OCC2010, Length: 357, dtype: int64
```

```
Classify by Industry:
```

```
0           8029
```

```
170         12
```

```
180          6
```

```
280          2
```

```
370          1
```

```
...
```

```
9770         1
```

```
9780         2
```

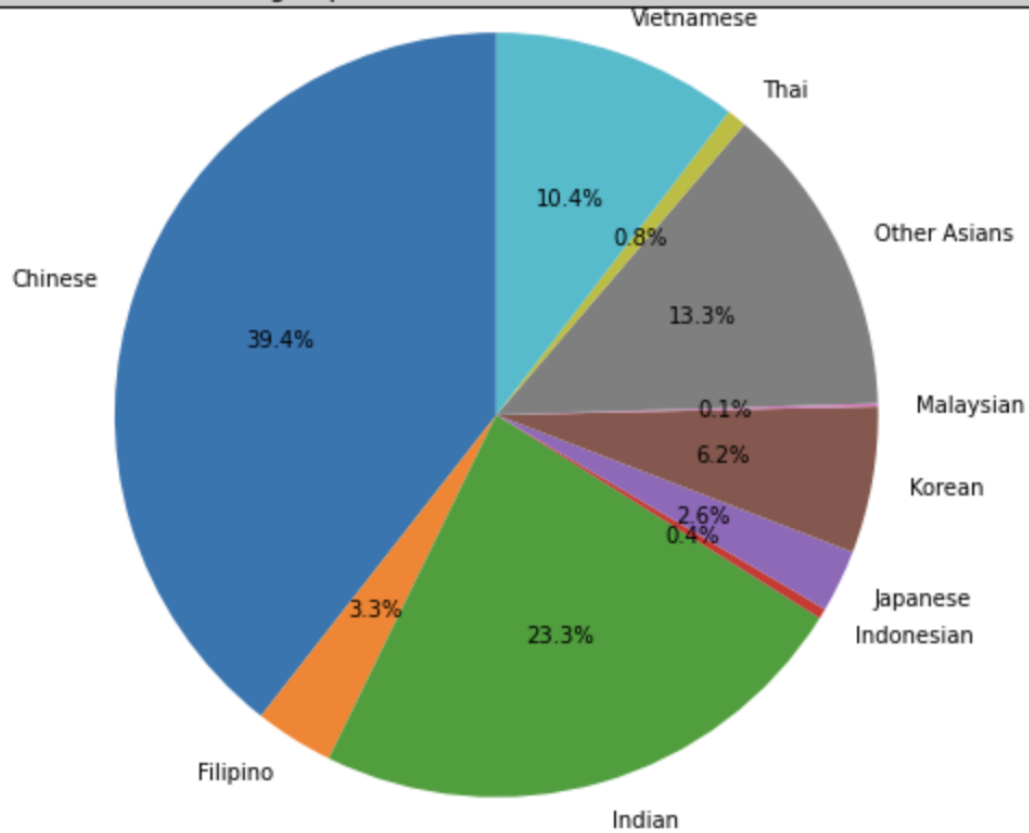
```
9790         2
```

```
9870         1
```

```
9920        127
```

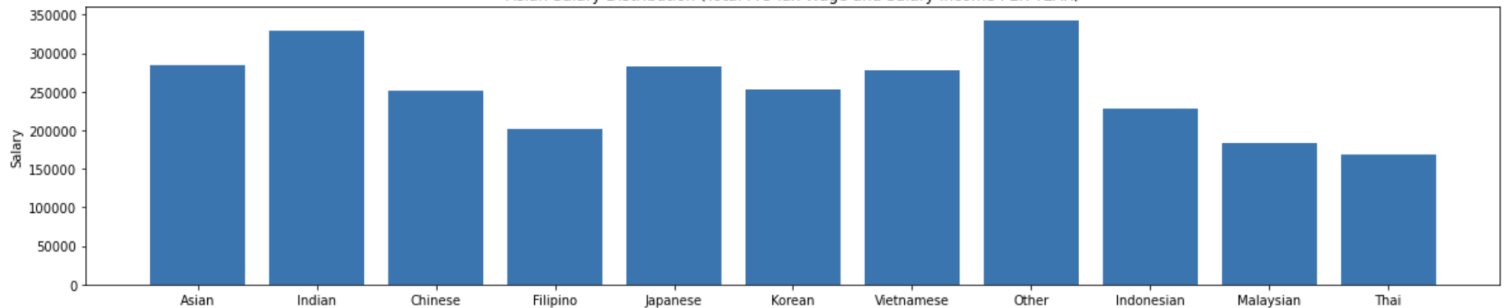
```
Name: IND, Length: 249, dtype: int64
```

Distribution of Asian subgroups between 2015-2019 in MA(A Total of 22349 Asians)



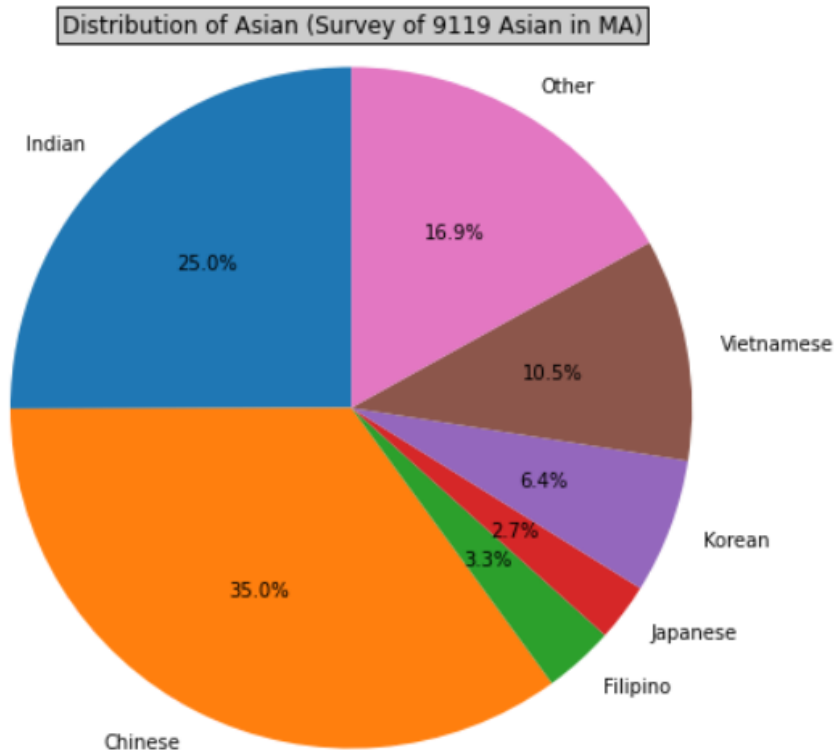
There are totally 16343 labor force in Asians

Asian Salary Distribution (Total Pre-Tax Wage and Salary Income PER YEAR)



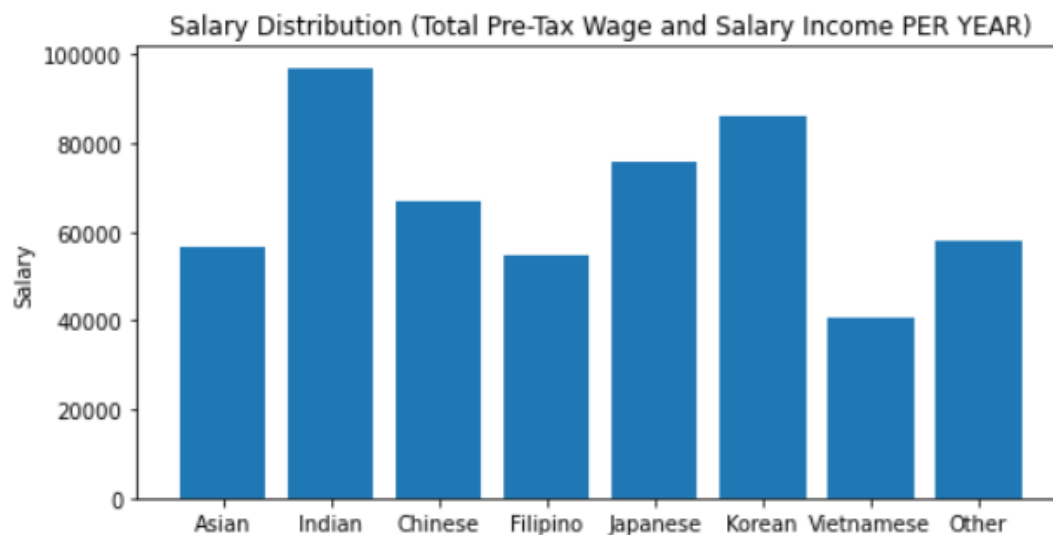
2. Preprocessed the CPS dataset from IPUMS.org, and the following is the report:

The dataset has much fewer samples after deleting non-labor force.  
This is the distribution of Asian subgroups.

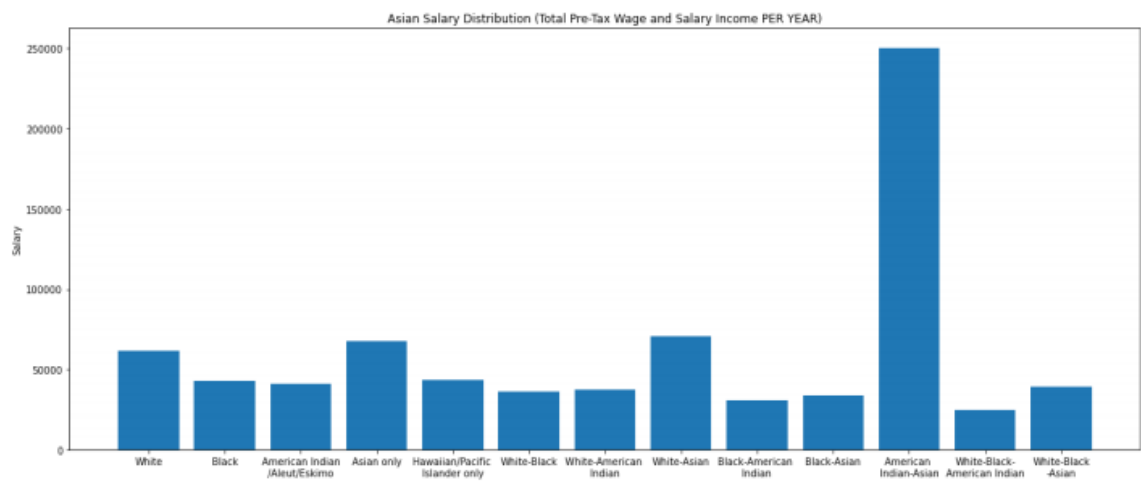


This is the salary distribution of the Asian subgroup

There are totally 5676 labor force in those Asian



This is the distribution of salary of whole races in MA



What we plan to do next:

1. Plotting line chart for each subgroup to show the change of income/wage in each year
2. Aggregate the different 357 occupations into more general categories.
3. Aggregate the different 249 industries into more general categories.
4. Treating Asians as a whole, showing the 5 most popular industries and the salary they are working in.
5. Trying to find news/research papers/reports that could back up our findings/conclusions.