Deliverable 1:
1. Make all csv sorted/filtered by features
separate the data into smaller sets/data frames
i.e. write functions to parse csv (using data frames ) to get data we want
combine features over the years together in one spot

Categories: (Scale: 0 small, 1 medium, 2 big)

1. age (2) **-- james**
2. Gender
    a. gender- population (0) **-- james**
    b. **X** labor force (0) **-- Emily**
    c. **X** unemployment (0) **-- Emily**
3. Marital Status (0) **-- Philip**
4. Citizenship (0) **-- Philip**
5. year of entry(0) **-- Philip**
6. ability to speak english (0) **-- Philip**
7. educational attainment(1) **-- Philip**
8. Employment type (0) **-- james**
9. Employment by industry (2) **-- Emily**
10. Employment by Occupation (2) **-- james**
    **Money**
11. Poverty
    a. Poverty Status (0) **-- Philip**
    b. Poverty Status of family households (0) **-- Emily - SECOND SHEET**
12. Median Personal Earnings (0)**--Qingyang**
13. Median Household income (0)**--Qingyang**
14. Occupied Housing Units (0)**--Qingyang**
15. Total number of family households(0)**--Qingyang**
16. Housing ownership(0)**--Qingyang**
17. Selected Monthly Owner Costs As A Percentage of Household Income(1) **--Qingyang**
18. Gross Rent as a percentage of household income (1) **-- Emily - SECOND SHEET**
19. Crowded units (0) **-- Emily - SECOND SHEET**

us
2. Visualizations (generalize to work with all years):
year of entry — see repo (Team2 branch) for Yichuan's visualization (2019 only so far)
(not map) find top 5 most populated employment by industry growth over the years using a line chart(I used a stacked chart to describe the top 5 populated employment by occupation through years and I think it is a good visualization. --Qingyang)

3. Training models
predict household income
*** Determine tools for this in Deliverable 1 *** predict growth of population based on time and amount of people (First determine if  linear regression or exponential are good descriptors )

Deliverable 2:
4. get average for each feature over the years (2, 5, etc.)
5. plot of last five years showing a heat map over actual map of U.S. showing change in population
6. Correlations - check for these
women with the economy through the years ( compare change of women pop. to change in household income) i.e. run correlation function on the two columns
7. Linear Regression - on features
combine multiple columns using low rank matrix approximations and compare to year
Change in educational attainment of population over the years (potential increase in education) — see if any features correlate with this increase
Based on employment industry change see if there is a corresponding change in money categories (poverty status, median personal earnings, housing ownership, percentage of monthly costs, gross rent percentage)