**Census Data Extraction, Preprocessing and Exploration**

Question 1: Geography and variable selection (20 pts)
Public transportation in San Francisco massively spreads across different neighborhoods. Given the convenience of buses, I want to figure out if people are taking advantage of the resources. I will use census data about people's means of transportation to work and focus on "car"(including car, truck, or van) and "bus"(including all public transportation, excluding taxicab). I will also look at how the data vary across sex.

Question 2: Map of Variables (10 pts)

	By car	By bus
All		
Male		

Female
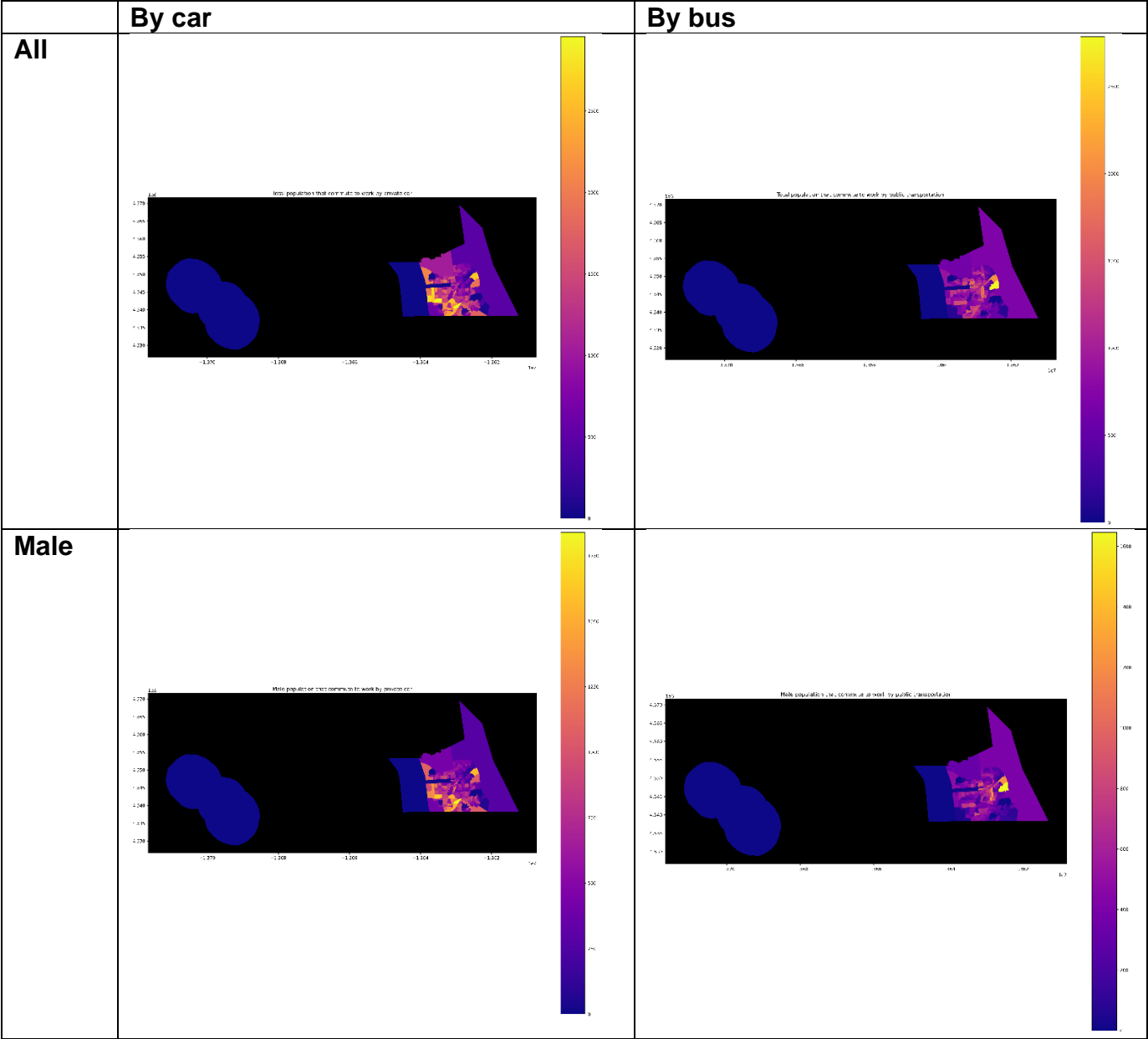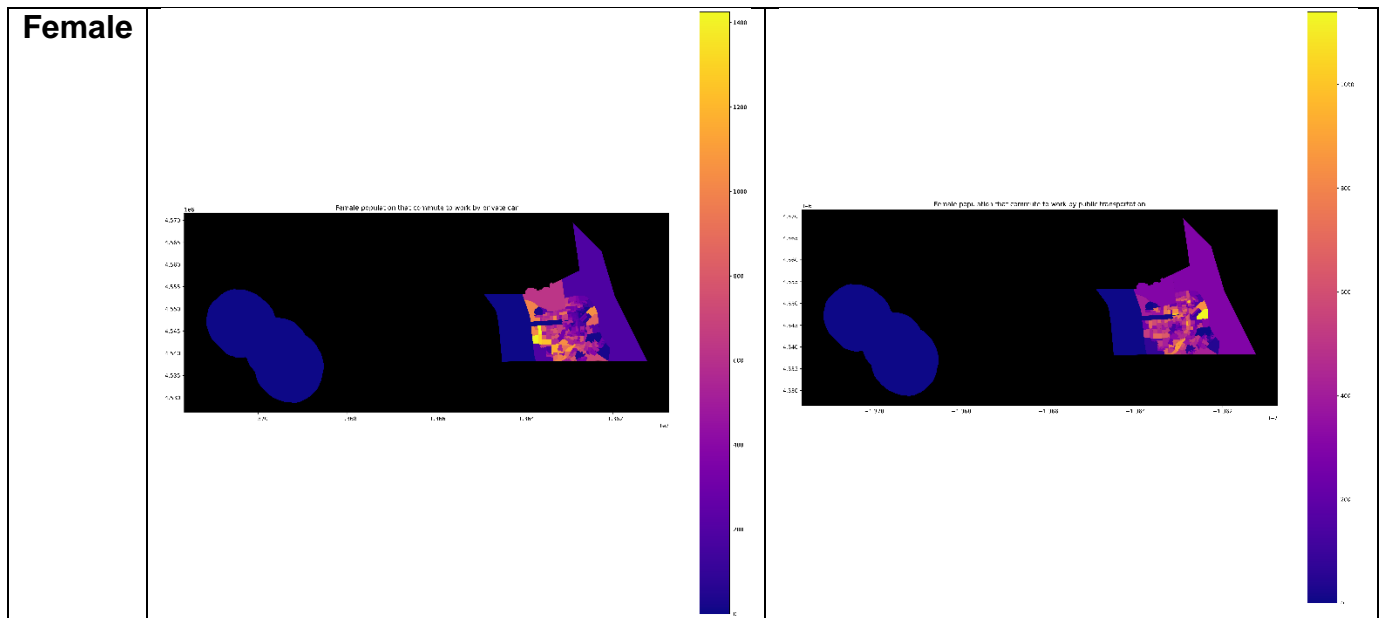
Question 3: Data exploration (20 pts)

a) How many census tracts your data have?
    a. 197
b) What is the Total Population in the data?
    a. Total population in the data: 495315.0
c) Calculate percentages based on your variables of interest
    a. Percentages of people going to work by private car: 0.411433128413232
    b. Percentages of people going to work by public transportation: 0.3402158222545249
    c. Percentages of male going to work by private car: 0.22857373590543392
    d. Percentages of male going to work by public transportation: 0.17095383745697182
    e. Percentages of female going to work by private car: 0.18285939250779806
    f. Percentages of female going to work by public transportation: 0.16926198479755308
d) What is the total area in square kilometers?

Question 4: Variables Creation (20 pts)

a) If you have population groups. Calculate and plot the probability density function of the ratio of each group to total population.

For Assignment 2, part 2 you will export the this data as a shape file and select 6 census tracts variables to do K-means Clustering.

Question 5: Summary Statistics (10)

Use the function describe() to summarize the statistics of the variable of interests

	total_pop	total_pop_car	total_pop_bus	male_car	male_bus	female_car	female_bus
count	197.00000	197.000000	197.000000	197.000000	197.000000	197.000000	197.000000
mean	2514.28934	1034.461929	855.401015	574.700508	429.827411	459.761421	425.573604
std	1143.86371	626.091092	419.801820	352.332558	248.146868	293.780372	205.009236
min	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1905.00000	602.000000	583.000000	335.000000	261.000000	238.000000	298.000000
50%	2408.00000	967.000000	822.000000	523.000000	396.000000	437.000000	411.000000
75%	3047.00000	1337.000000	1085.000000	738.000000	539.000000	621.000000	534.000000
max	9398.00000	2957.000000	2786.000000	1840.000000	1646.000000	1425.000000	1140.000000

Question 6: Summary Statistics (10 pts)
- Most of the pairs with the same means of transportations(car or bus) are largely correlated

Question 7: Data Interpretation (10 pts)
- More people commute to work by car than by bus.
- On average, more male drive to work than female.
- The differences of taking buses to work between male and female is small.