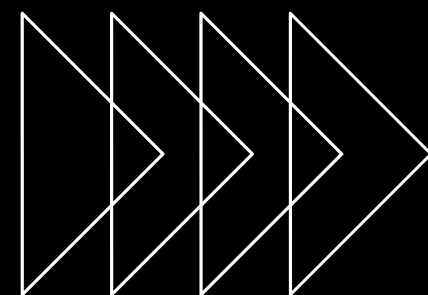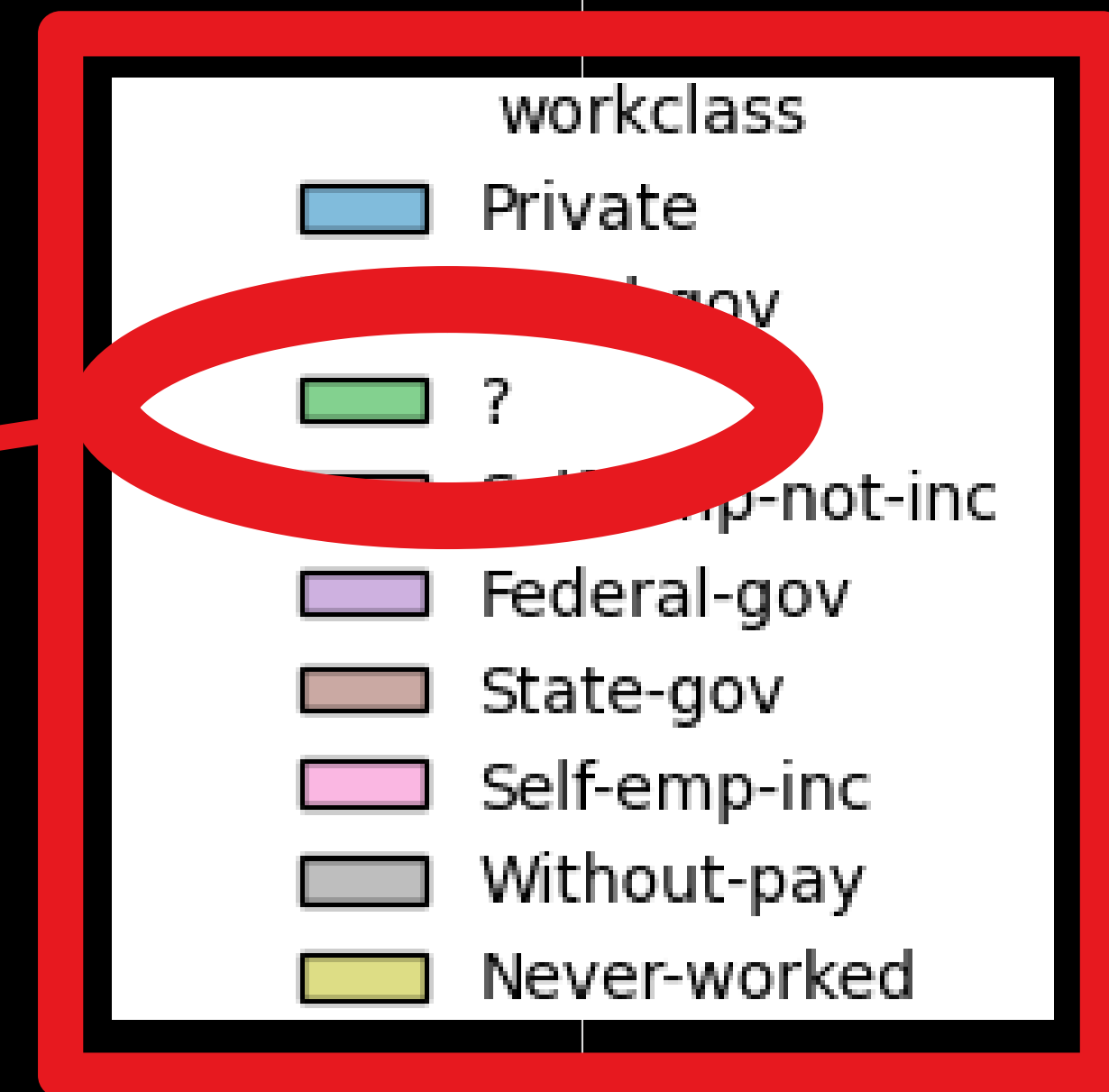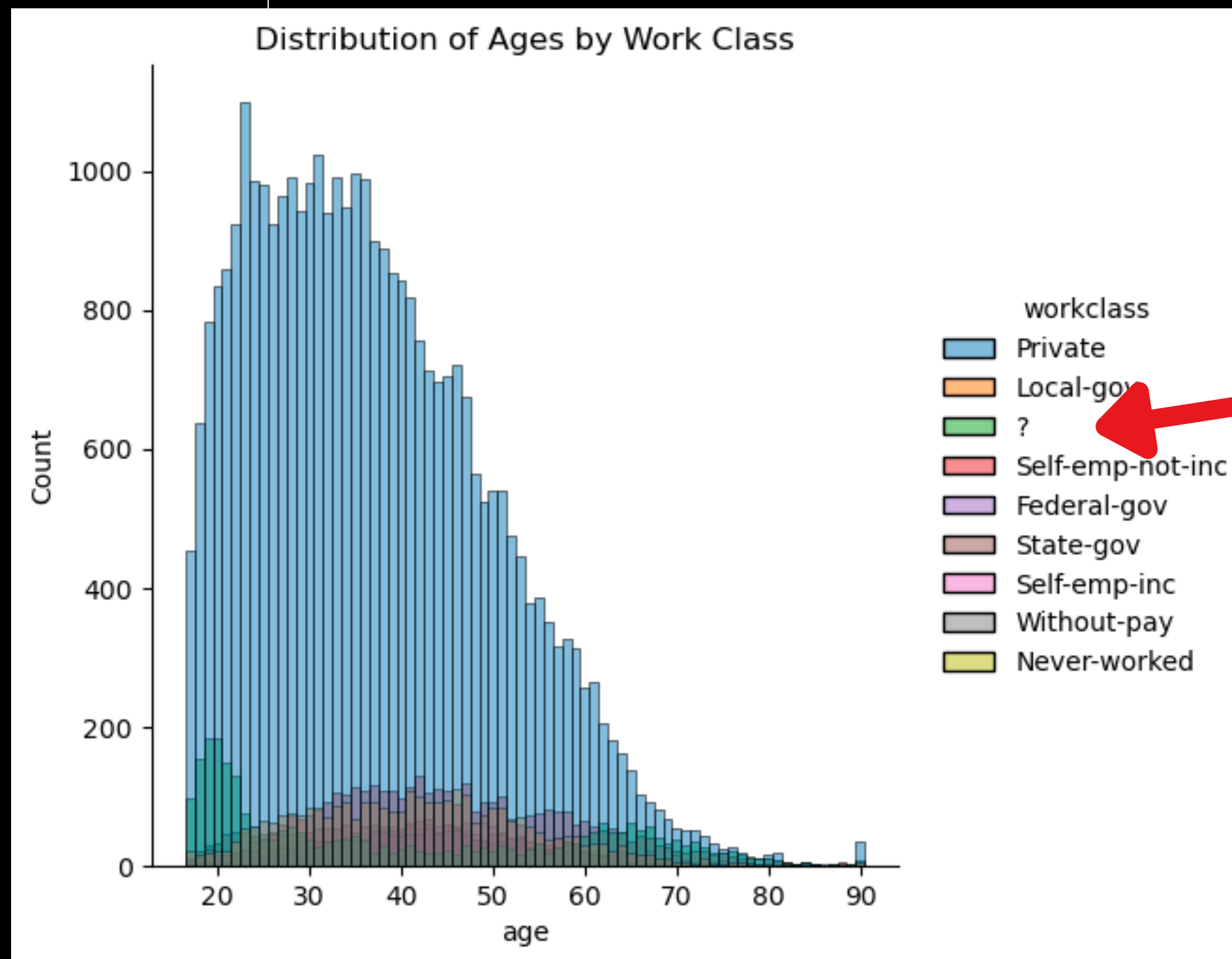# Discovering who the ? Group is

An experiential movement through and about one demographic

WHAT IS THE ? GROUP

From a cursory look at age demographics broken up by work sector, there is a curious category called ?

*all data is from Adult Income dataset

The ? Group

Distribution of Ages

Box and Whisker for Age of ?

1 This group is left-skewed in relation to age distribution. Although the average age of the ? Group is 40.1 years (compared to the sample's 38.6), they also have a median age of 34 years (below that of the samples 37 years)
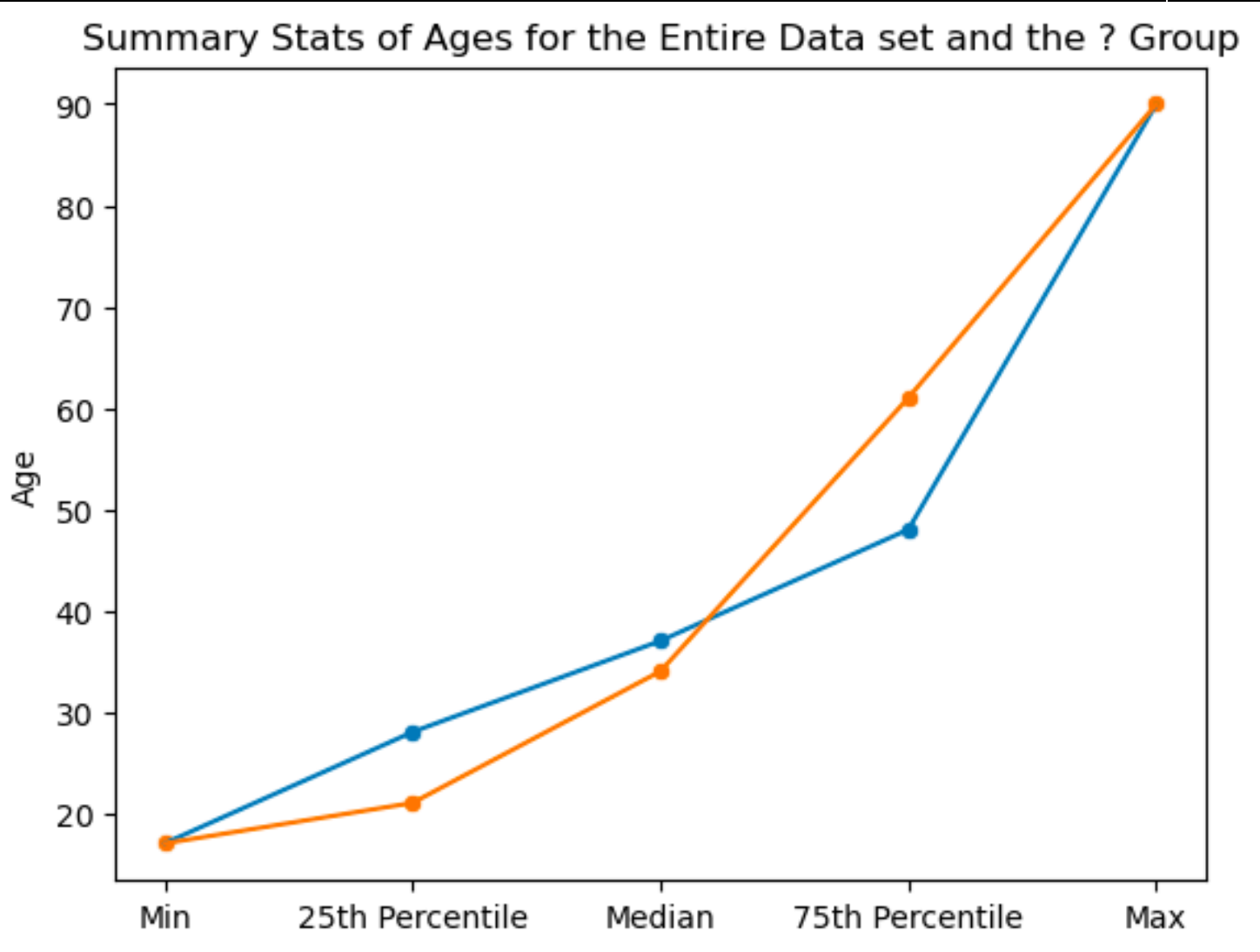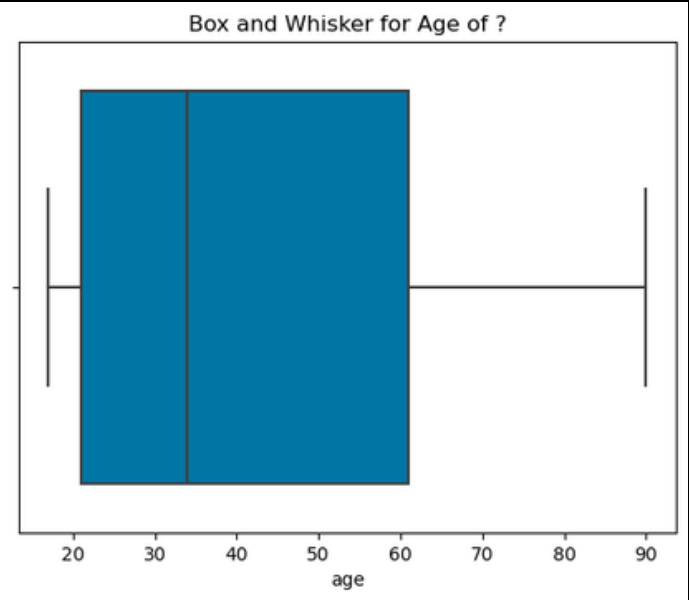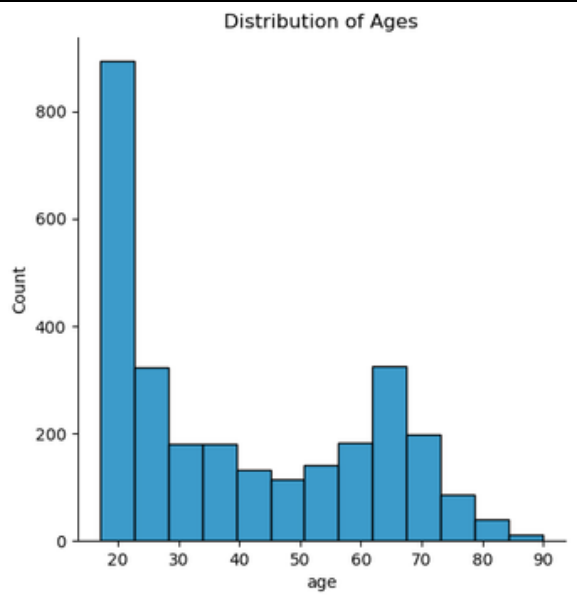
2 The ? Group is as 1/20th the size of the entire dataset. It closely mirrors the racial demographics of the whole. 82.6% of ? are White (compared to 85.5% (-2.9%)), 11.6% are Black (9.6% (+2%)), 3.4% are Asian/Pacific-Islander (3.1% (+0.3%)), 1.3% are American-Indian/Eskimo (1.0% (+ 0.3%)), and 1.1% are Other (0.8% (+ 0.3%)).

3 The gender differences are quite large. The ? Group is close to evenly split between binary genders: 54.6% male, 45.4% female. The overall dataset is 66.8% male and 33.2% female. This is the closest split for any other work class. The second closest is the local goverment (59.9%M/40.1%F) and the group with the biggest split is self employed (87.6%M/12.4%F).

4 For the entire data set, 76.1% of the people make $50,000 or less. For the ? Group, 90.53% make that $50K or less. The second closest split is the without-pay group (90.47%), while the never-worked group logically are all within that income category. The group with the highest percentage of workers making more than $50K is self-employeed, with 44.7%.

Summary Stats of Ages for the Entire Data set and the ? Group

The ? Group

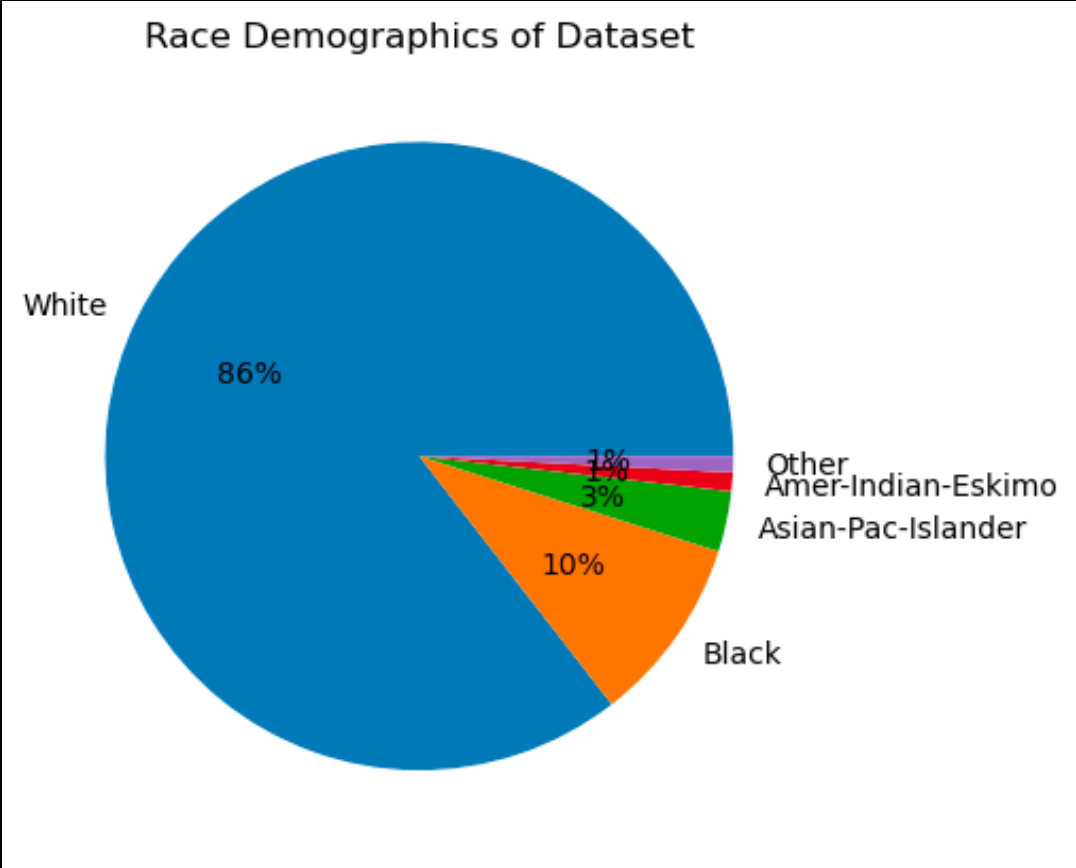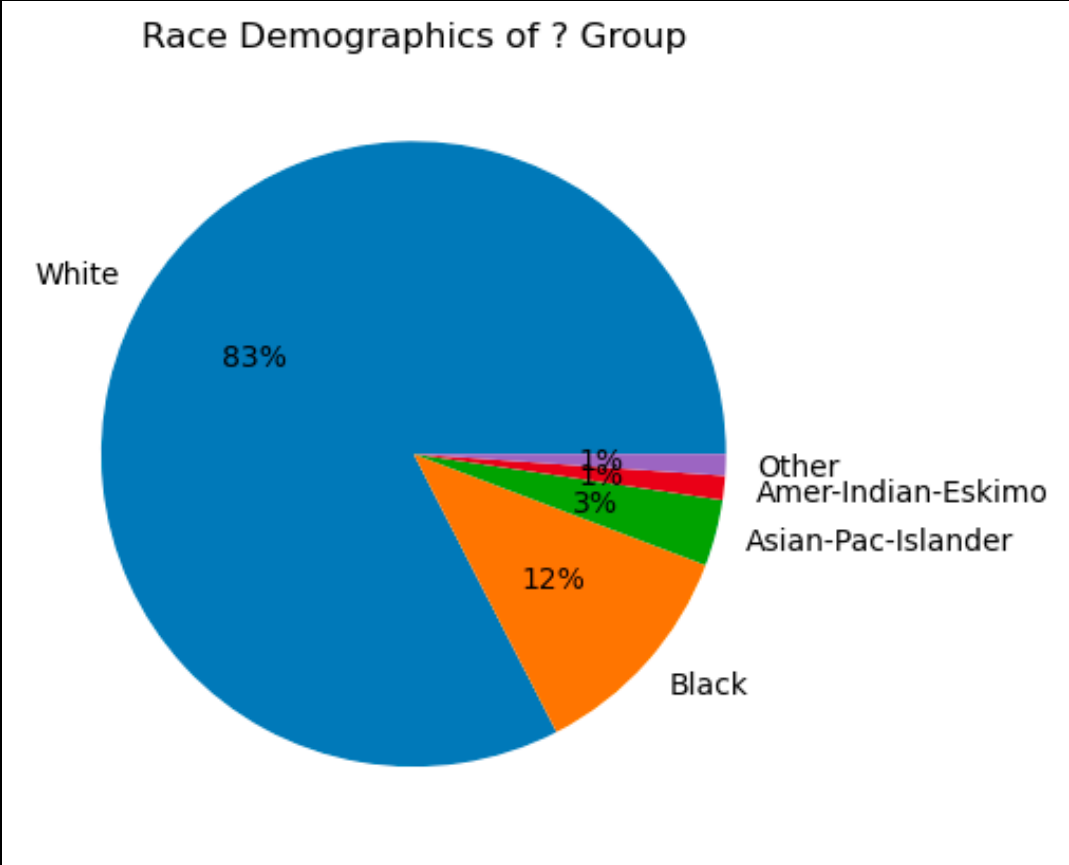|	
1	This group is left-skewed in relation to age distribution. Although the average age of the ? Group is 40.1 years (compared to the sample's 38.6), they also have a median age of 34 years (below that of the samples 37 years)
2	The ? Group is as 1/20th the size of the entire dataset. It closely mirrors the racial demographics of the whole. 82.6% of ? are White (compared to 85.5% (-2.9%)), 11.6% are Black (9.6% (+2%)), 3.4% are Asian/Pacific-Islander (3.1% (+0.3%)), 1.3% are American-Indian/Eskimo (1.0% (+ 0.3%)), and 1.1% are Other (0.8% (+ 0.3%)).
3	The gender differences are quite large. The ? Group is close to evenly split between binary genders: 54.6% male, 45.4% female. The overall dataset is 66.8% male and 33.2% female. This is the closest split for any other work class. The second closest is the local goverment (59.9%M/40.1%F) and the group with the biggest split is self employed (87.6%M/12.4%F).
4	For the entire data set, 76.1% of the people make $50,000 or less. For the ? Group, 90.53% make that $50K or less. The second closest split is the without-pay group (90.47%), while the never-worked group logically are all within that income category. The group with the highest percentage of workers making more than $50K is self-employeed, with 44.7%.

Race Demographics of ? Group

White 83%
1%
1%
3%
12%
Other
Amer-Indian-Eskimo
Asian-Pac-Islander
Black

Race Demographics of Dataset

White 86%
1%
1%
3%
10%
Other
Amer-Indian-Eskimo
Asian-Pac-Islander
Black

The ? Group

1

This group is left-skewed in relation to age distribution. Although the average age of the ? Group is 40.1 years (compared to the sample's 38.6), they also have a median age of 34 years (below that of the samples 37 years)
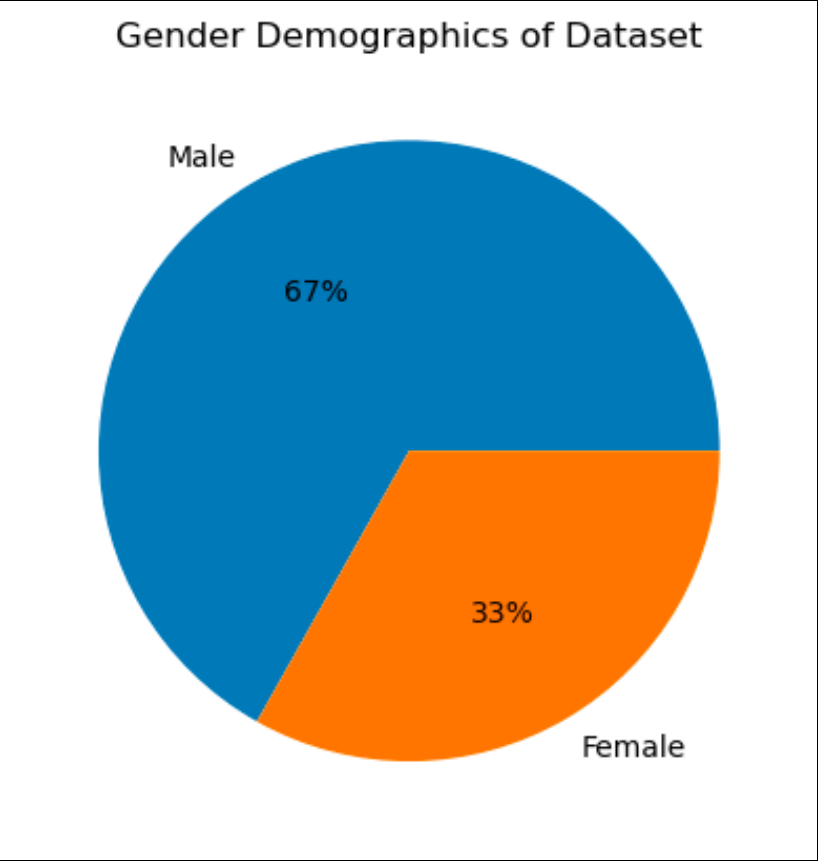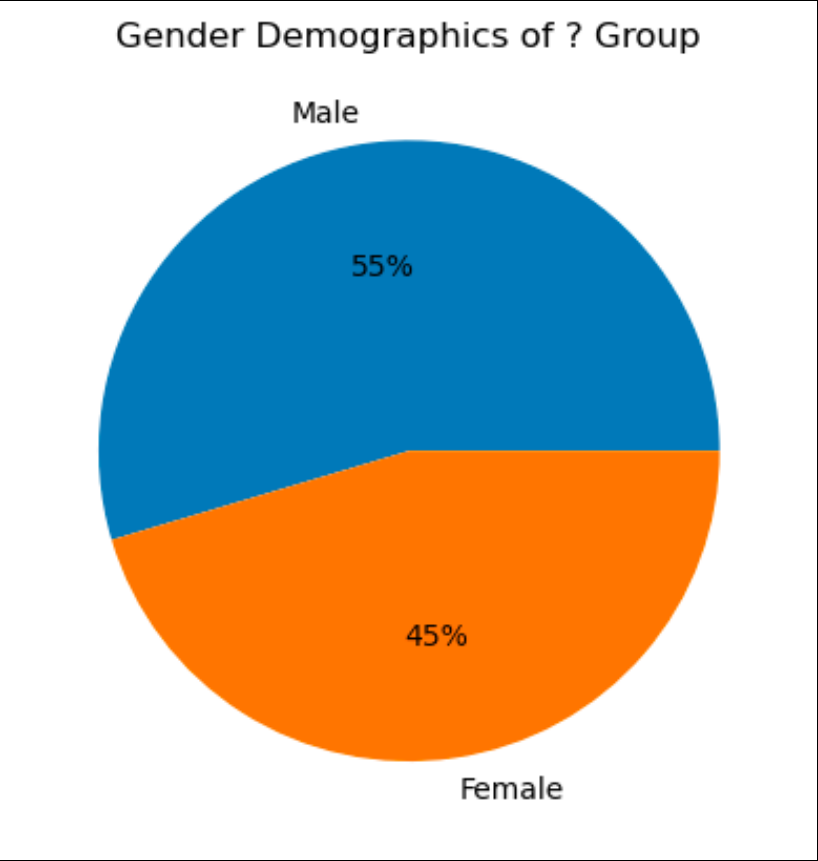
2

The ? Group is as 1/20th the size of the entire dataset. It closely mirrors the racial demographics of the whole. 82.6% of ? are White (compared to 85.5% (-2.9%)), 11.6% are Black (9.6% (+2%)), 3.4% are Asian/Pacific-Islander (3.1% (+0.3%)), 1.3% are American-Indian/Eskimo (1.0% (+ 0.3%)), and 1.1% are Other (0.8% (+ 0.3%)).

3

The gender differences are quite large. The ? Group is close to evenly split between binary genders: 54.6% male, 45.4% female. The overall dataset is 66.8% male and 33.2% female. This is the closest split for any other work class. The second closest is the local goverment (59.9%M/40.1%F) and the group with the biggest split is self employed (87.6%M/12.4%F).

4

For the entire data set, 76.1% of the people make $50,000 or less. For the ? Group, 90.53% make that $50K or less. The second closest split is the without-pay group (90.47%), while the never-worked group logically are all within that income category. The group with the highest percentage of workers making more than $50K is self-employeed, with 44.7%.

Gender Demographics of ? Group

Male

55%

45%

Female

Gender Demographics of Dataset

Male

67%

33%

Female

The ? Group

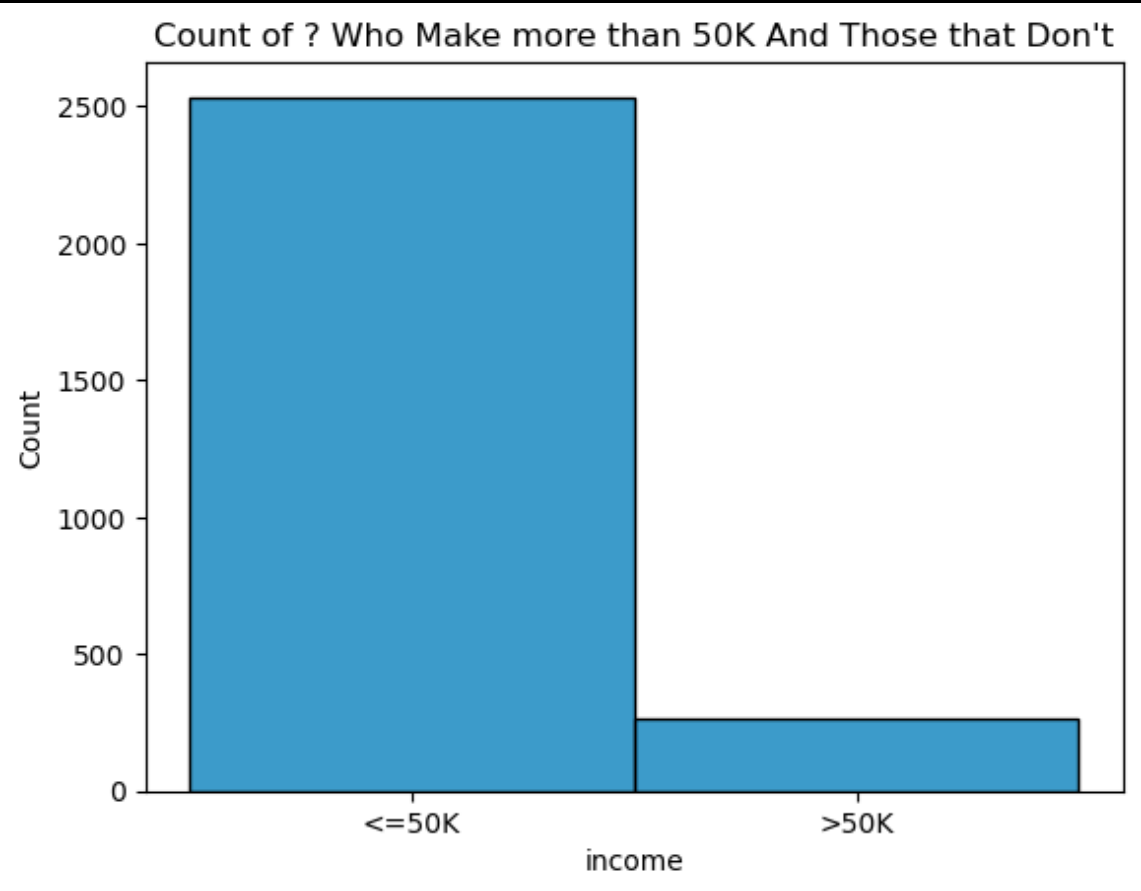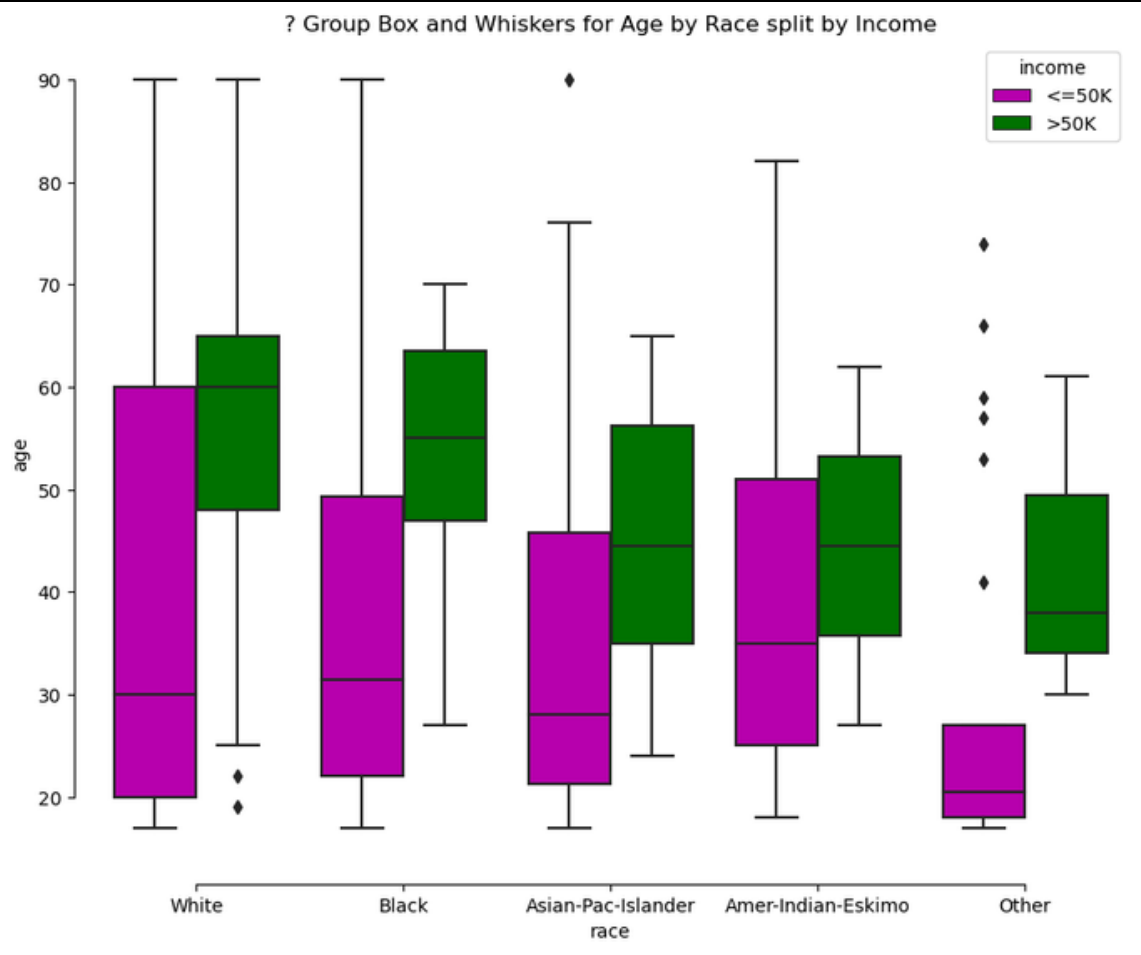|	
1	This group is left-skewed in relation to age distribution. Although the average age of the ? Group is 40.1 years (compared to the sample's 38.6), they also have a median age of 34 years (below that of the samples 37 years)
2	The ? Group is as 1/20th the size of the entire dataset. It closely mirrors the racial demographics of the whole. 82.6% of ? are White (compared to 85.5% (-2.9%)), 11.6% are Black (9.6% (+2%)), 3.4% are Asian/Pacific-Islander (3.1% (+0.3%)), 1.3% are American-Indian/Eskimo (1.0% (+ 0.3%)), and 1.1% are Other (0.8% (+ 0.3%)).
3	The gender differences are quite large. The ? Group is close to evenly split between binary genders: 54.6% male, 45.4% female. The overall dataset is 66.8% male and 33.2% female. This is the closest split for any other work class. The second closest is the local goverment (59.9%M/40.1%F) and the group with the biggest split is self employed (87.6%M/12.4%F).
4	For the entire data set, 76.1% of the people make $50,000 or less. For the ? Group, 90.53% make that $50K or less. The second closest split is the without-pay group (90.47%), while the never-worked group logically are all within that income category. The group with the highest percentage of workers making more than $50K is self-employeed, with 44.7%.

? Group Box and Whiskers for Age by Race split by Income

Count of ? Who Make more than 50K And Those that Don't

Where is the ? Group from

United-States 2530
Mexico 48
? 46
Canada 19
South 14
Germany 13
Philippines 12
Taiwan 10
China 9
Puerto-Rico 9
El-Salvador 8
England 8
Poland 6
Dominican-Republic 6
Haiti 6
Cuba 5
Portugal 5
Italy 5
India 4
Jamaica 3
Iran 3
Columbia 3
Vietnam 3
Japan 3
Guatemala

France 2
Cambodia 2
Hong 2
Laos 2
Ecuador 2
Honduras 1
Nicaragua 1
Trinadad&Tobago 1
Hungary 1
Ireland 1
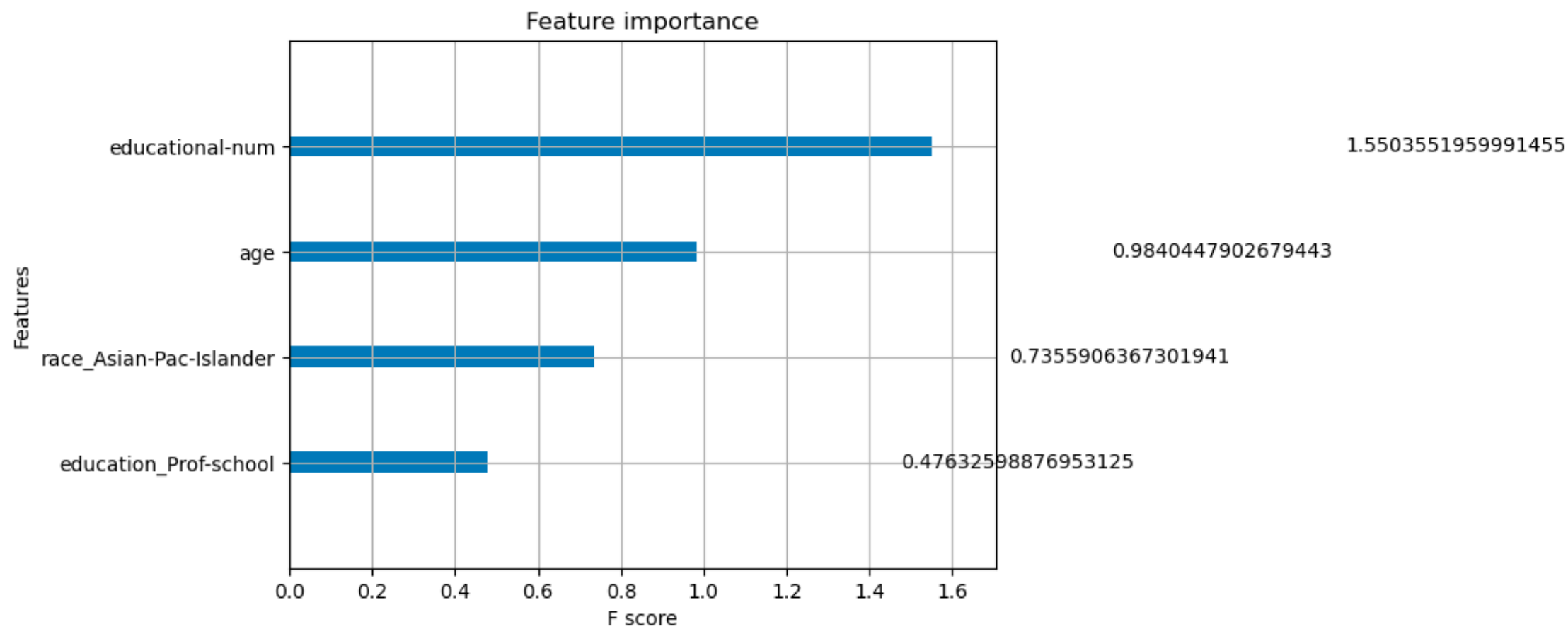Outlying-US(Guam-USVI-etc) 1
Peru 1
Thailand 1
Scotland 1

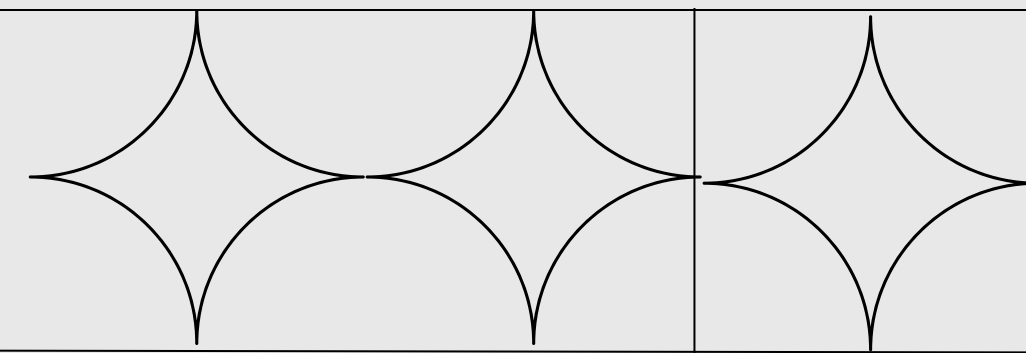2530 are from USA

45 of them are from ? (a coincidence?)

5

After running a machine learning classifier algorithm (XGBoost), it appears that the most importnt features are the number of educational years, their age, whether they identify their race as Asian/Pacific Island, and their school as professional

What's important for identifying someone from the ? Group

Feature importance

educational-num   1.5503551959991455

age   0.9840447902679443

race_Asian-Pac-Islander   0.7355906367301941

education_Prof-school   0.47632598876953125

F score

Who are ? in relation to the data set

They are not making the most money, that's for sure.

Feature importance