NAME: Xander Sam E. Galapia

SECTION: CPE22S3

```
import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt

import seaborn as sns
```

```
pip install ucimlrepo
```

```
    Requirement already satisfied: ucimlrepo in /usr/local/lib/python3.10/dist-packages (0.0.6)
```

```
from ucimlrepo import fetch_ucirepo

# fetch dataset
census_income = fetch_ucirepo(id=20)

# data (as pandas dataframes)
X = census_income.data.features
y = census_income.data.targets

# metadata
print(census_income.metadata)

# variable information
print(census_income.variables)
```

```
{'uci_id': 20, 'name': 'Census Income', 'repository_url': 'https://archive.ics.uci.edu/dataset/20/census+income', 'data_url': 'https://archive.ics.uci.edu/static/public/20/data.csv', 'abstract': 'Predict whether income exceeds $50K/yr based on census da
              name     role         type      demographic  \
0              age  Feature      Integer              Age
1        workclass  Feature  Categorical           Income
2           fnlwgt  Feature      Integer             None
3        education  Feature  Categorical  Education Level
4    education-num  Feature      Integer  Education Level
5   marital-status  Feature  Categorical            Other
6       occupation  Feature  Categorical            Other
7     relationship  Feature  Categorical            Other
8             race  Feature  Categorical             Race
9              sex  Feature       Binary              Sex
10    capital-gain  Feature      Integer             None
11    capital-loss  Feature      Integer             None
12  hours-per-week  Feature      Integer             None
13  native-country  Feature  Categorical            Other
14          income   Target       Binary           Income

                                          description units missing_values
0                                                 N/A  None             no
1   Private, Self-emp-not-inc, Self-emp-inc, Feder...  None            yes
2                                                None  None             no
3    Bachelors, Some-college, 11th, HS-grad, Prof-...  None             no
4                                                None  None             no
5   Married-civ-spouse, Divorced, Never-married, S...  None             no
6   Tech-support, Craft-repair, Other-service, Sal...  None            yes
7   Wife, Own-child, Husband, Not-in-family, Other...  None             no
8   White, Asian-Pac-Islander, Amer-Indian-Eskimo,...  None             no
9                                       Female, Male.  None             no
10                                               None  None             no
11                                               None  None             no
12                                               None  None             no
13  United-States, Cambodia, England, Puerto-Rico,...  None            yes
14                                       >50K, <=50K.  None             no
```

```
combd = pd.concat([X, y], axis=1)
combd
```

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
48837	39	Private	215419	Bachelors	13	Divorced	Prof-specialty	Not-in-family	White	Female	0	0	36	United-States	<=50K.
48838	64	NaN	321403	HS-grad	9	Widowed	NaN	Other-relative	Black	Male	0	0	40	United-States	<=50K.
48839	38	Private	374983	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	50	United-States	<=50K.
48840	44	Private	83891	Bachelors	13	Divorced	Adm-clerical	Own-child	Asian-Pac-Islander	Male	5455	0	40	United-States	<=50K.
48841	35	Self-emp-inc	182148	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	60	United-States	>50K.

48842 rows × 15 columns

Next steps: View recommended plots

```
combd.rename(columns = {'marital-status': 'marital_status',
                        'native-country': 'Birthplace',
                        'hours-per-week' : 'Weekly_Hours',
                        'capital-gain' : 'capital_gain',
                        'capital-loss' : 'capital_loss'
                     }, inplace = True)
```

```
high = combd.sort_values(by=["age"], ascending=False)
high
```

	age	workclass	fnlwgt	education	education-num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	Weekly_Hours	Birthplace	income
46519	90	Private	272752	Some-college	10	Never-married	Other-service	Own-child	White	Male	0	0	10	United-States	<=50K.
39537	90	Private	250832	HS-grad	9	Married-civ-spouse	Transport-moving	Husband	White	Male	2414	0	40	United-States	<=50K.
19212	90	Private	139660	Some-college	10	Divorced	Sales	Unmarried	Black	Female	0	0	37	United-States	<=50K
24043	90	Self-emp-not-inc	82628	HS-grad	9	Never-married	Exec-managerial	Not-in-family	White	Male	2964	0	12	United-States	<=50K
5370	90	Local-gov	227796	Masters	14	Married-civ-spouse	Exec-managerial	Husband	White	Male	20051	0	60	United-States	>50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
34434	17	Private	173740	10th	6	Never-married	Sales	Not-in-family	White	Female	0	0	15	United-States	<=50K.
21943	17	Private	295991	10th	6	Never-married	Adm-clerical	Own-child	White	Female	0	0	20	United-States	<=50K
43806	17	Private	34943	10th	6	Never-married	Other-service	Own-child	White	Female	0	0	15	United-States	<=50K.
22582	17	Private	198146	11th	7	Never-married	Sales	Own-child	White	Female	0	0	16	United-States	<=50K
47879	17	Private	160029	11th	7	Never-married	Sales	Own-child	White	Female	0	0	14	United-States	<=50K.

48842 rows × 15 columns

Next steps: View recommended plots

```
t30b = combd.query('age <= 30').sort_values(by='age', ascending=True)
t30b
```

	age	workclass	fnlwgt	education	education-num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	Weekly_Hours	Birthplace	income
43398	17	Private	89259	11th	7	Never-married	Adm-clerical	Own-child	White	Female	0	0	15	United-States	<=50K.
23660	17	Private	232713	10th	6	Never-married	Craft-repair	Not-in-family	White	Male	594	0	30	United-States	<=50K
46920	17	Private	140117	11th	7	Never-married	Sales	Own-child	White	Female	0	0	14	United-States	<=50K.
44542	17	Private	58037	10th	6	Never-married	Other-service	Own-child	White	Male	0	0	40	United-States	<=50K.
15373	17	Private	95799	11th	7	Never-married	Sales	Own-child	White	Female	0	0	18	United-States	<=50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8845	30	Private	88913	Some-college	10	Separated	Other-service	Unmarried	Asian-Pac-Islander	Female	0	0	40	United-States	<=50K
24953	30	Self-emp-not-inc	196342	Some-college	10	Never-married	Sales	Own-child	White	Male	0	0	25	United-States	<=50K
41905	30	Private	183801	HS-grad	9	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	40	United-States	<=50K.
8878	30	Private	158688	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	United-States	<=50K
44219	30	Private	176064	HS-grad	9	Divorced	Craft-repair	Unmarried	White	Male	0	0	55	United-States	<=50K.

15793 rows × 15 columns

Next steps: View recommended plots

```
combd.dtypes
```

```
age                int64
workclass         object
fnlwgt             int64
education         object
education-num      int64
marital_status    object
occupation        object
relationship      object
race              object
sex               object
capital_gain       int64
capital_loss       int64
Weekly_Hours       int64
Birthplace        object
income            object
dtype: object
```

```
wc = combd.workclass.unique()
```

```
wc
```

```
array(['State-gov', 'Self-emp-not-inc', 'Private', 'Federal-gov',
       'Local-gov', '?', 'Self-emp-inc', 'Without-pay', 'Never-worked',
       nan], dtype=object)
```

```
marst = combd.marital_status.unique()
marst
```

```
array(['Never-married', 'Married-civ-spouse', 'Divorced',
       'Married-spouse-absent', 'Separated', 'Married-AF-spouse',
       'Widowed'], dtype=object)
```

```
female = combd.query('sex == "Female"')
female
```

	age	workclass	fnlwgt	education	education-num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	Weekly_Hours	Birthplace	income
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
5	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
6	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<=50K
8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K
12	23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States	<=50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
48826	50	Local-gov	139347	Masters	14	Married-civ-spouse	Prof-specialty	Wife	White	Female	0	0	40	NaN	>50K.
48827	55	Private	224655	HS-grad	9	Separated	Priv-house-serv	Not-in-family	White	Female	0	0	32	United-States	<=50K.
48828	38	Private	247547	Assoc-voc	11	Never-married	Adm-clerical	Unmarried	Black	Female	0	0	40	United-States	<=50K.
48834	25	Private	350977	HS-grad	9	Never-married	Other-service	Own-child	White	Female	0	0	40	United-States	<=50K.
48837	39	Private	215419	Bachelors	13	Divorced	Prof-specialty	Not-in-family	White	Female	0	0	36	United-States	<=50K.

16192 rows × 15 columns

Next steps: View recommended plots

```
male = combd.query('sex == "Male"')
male
```

	age	workclass	fnlwgt	education	education-num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	Weekly_Hours	Birthplace	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
7	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
48836	33	Private	245211	Bachelors	13	Never-married	Prof-specialty	Own-child	White	Male	0	0	40	United-States	<=50K.
48838	64	NaN	321403	HS-grad	9	Widowed	NaN	Other-relative	Black	Male	0	0	40	United-States	<=50K.
48839	38	Private	374983	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	50	United-States	<=50K.
48840	44	Private	83891	Bachelors	13	Divorced	Adm-clerical	Own-child	Asian-Pac-Islander	Male	5455	0	40	United-States	<=50K.
48841	35	Self-emp-inc	182148	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	60	United-States	>50K.

32650 rows × 15 columns

Next steps: View recommended plots

```
rls = combd.relationship.unique()
rls
```

```
array(['Not-in-family', 'Husband', 'Wife', 'Own-child', 'Unmarried',
       'Other-relative'], dtype=object)
```

```
inc = combd.income.unique()
inc
```

```
array(['<=50K', '>50K', '<=50K.', '>50K.'], dtype=object)
```

```
inc_up = {'>50K.' : '>50K',
              '<=50K.' : '<=50K'}
combd.replace({'income':inc_up},inplace = True)
```

```
inc1 = combd.income.unique()
inc1
```

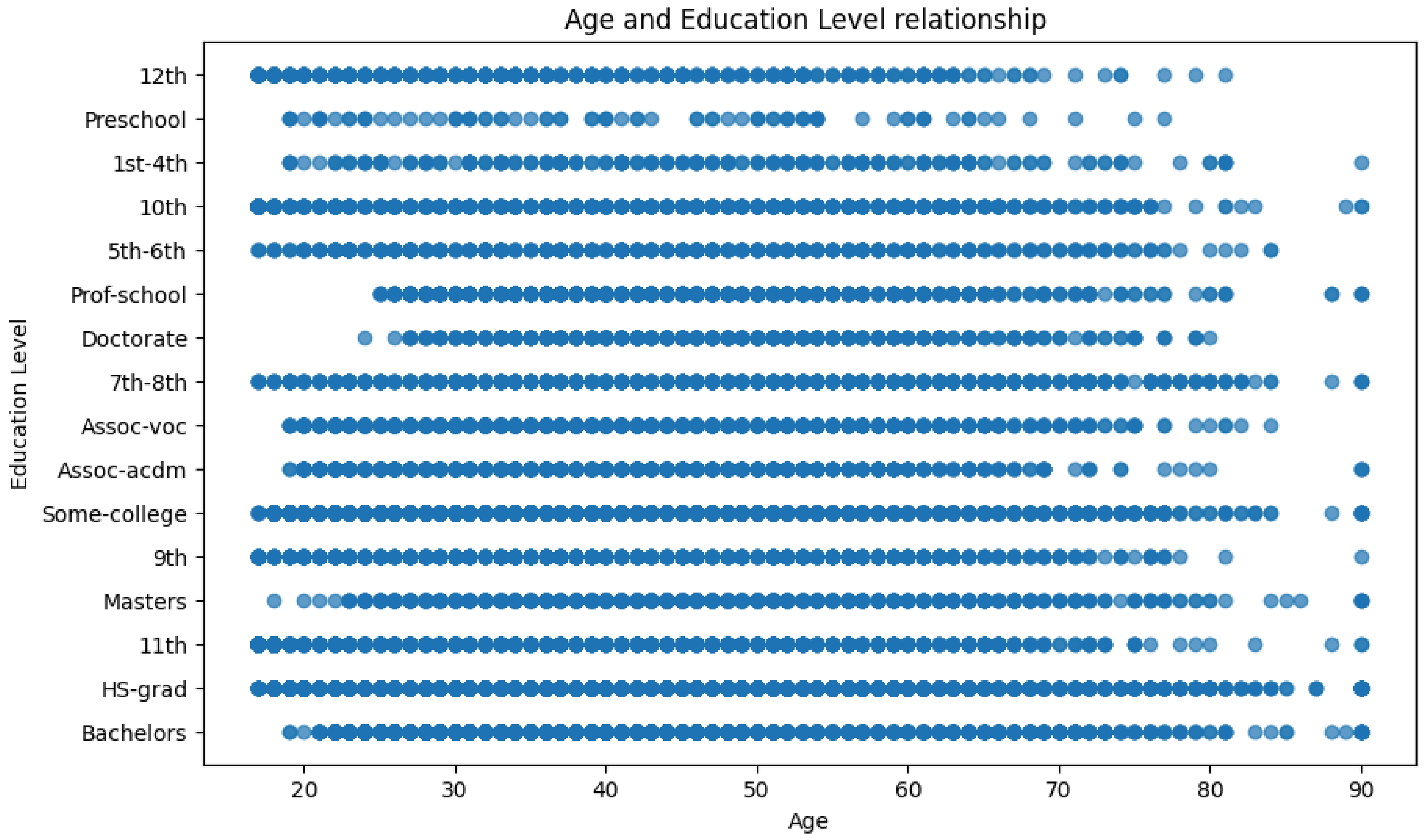
```
array(['<=50K', '>50K'], dtype=object)
```

```
t18b = combd.query('age <= 20 & education == "Bachelors"').sort_values(by='age', ascending=True)
t18b
```

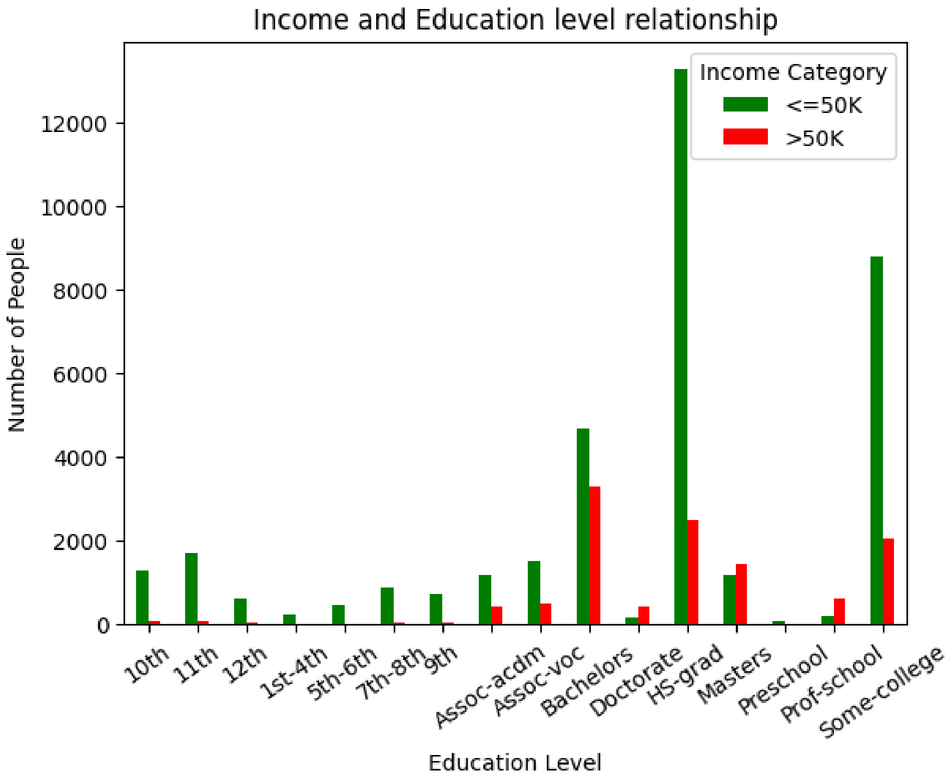
	age	workclass	fnlwgt	education	education-num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	Weekly_Hours	Birthplace	income
1570	19	?	62534	Bachelors	13	Never-married	?	Own-child	Black	Female	0	0	40	Jamaica	<=50K
3591	19	Private	100999	Bachelors	13	Never-married	Prof-specialty	Own-child	White	Female	0	0	30	United-States	<=50K
8415	20	Private	216436	Bachelors	13	Never-married	Sales	Other-relative	Black	Female	0	0	30	United-States	<=50K

Next steps: View recommended plots

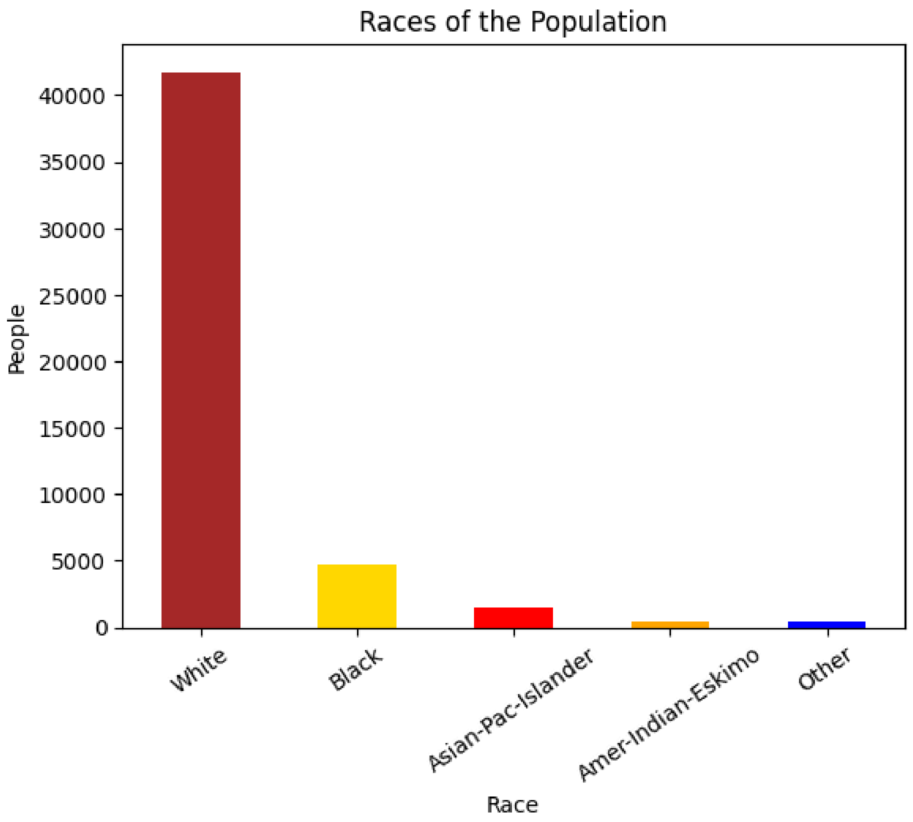
```
df = pd.DataFrame(combd)

plt.figure(figsize=(10, 6))
plt.scatter(df["age"], df["education"], alpha=0.7)

# Label axis
plt.xlabel("Age")
plt.ylabel("Education Level")

#Adding Title
plt.title("Age and Education Level relationship ")

#Displaying the plot
plt.show()
```



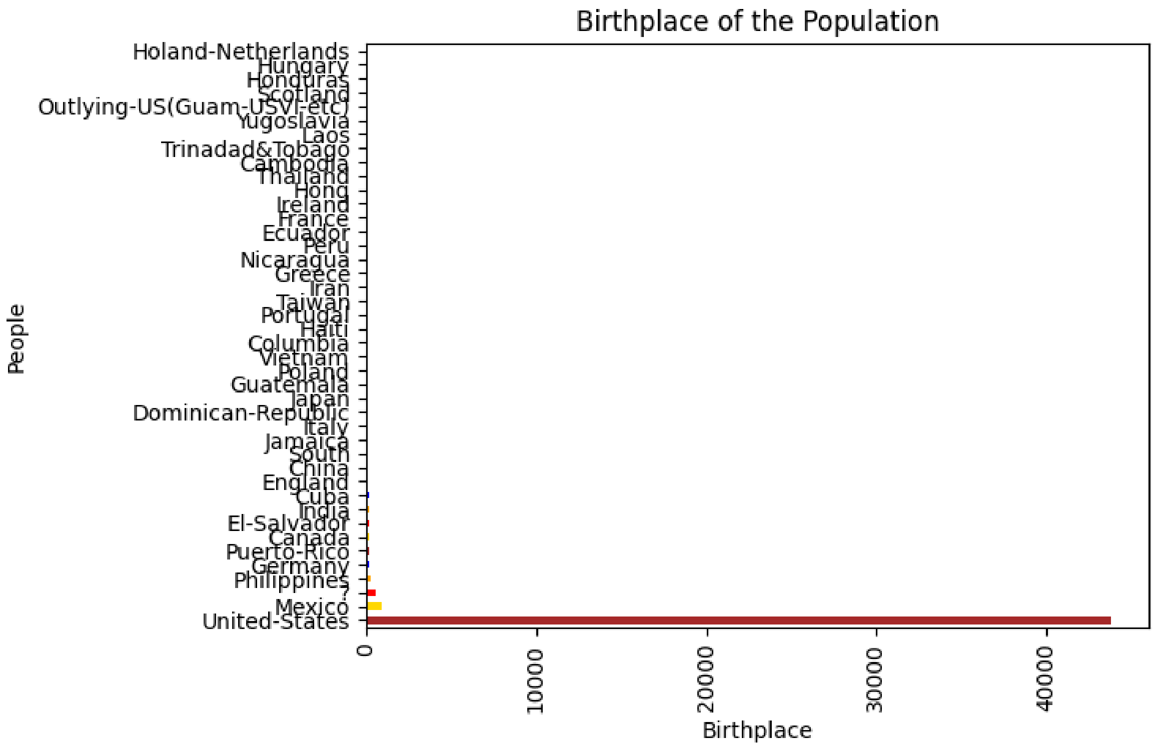
```
IncEd = df.groupby(['education', 'income']).size().unstack()


# Create the bar chart
IncEd.plot(kind='bar', stacked=False, color=['green', 'red'])
plt.xlabel("Education Level")
plt.ylabel("Number of People")
plt.title("Income and Education level relationship")
plt.legend(title="Income Category")
plt.xticks(rotation=35)
plt.show()
```



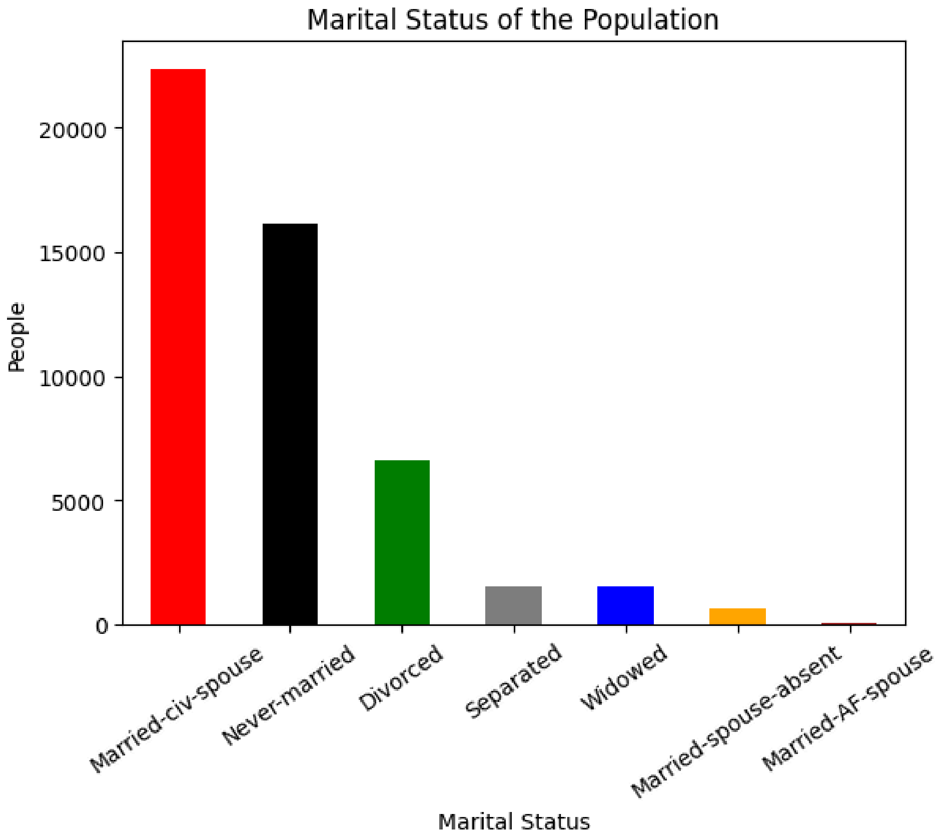
```
df.race.value_counts().plot(
    kind='bar', title='Races of the Population', rot=0, color = ['Brown', 'Gold', 'Red', 'Orange', 'Blue']
)


plt.xlabel('Race')
plt.ylabel('People')
plt.xticks(rotation=35)


plt.show()
```



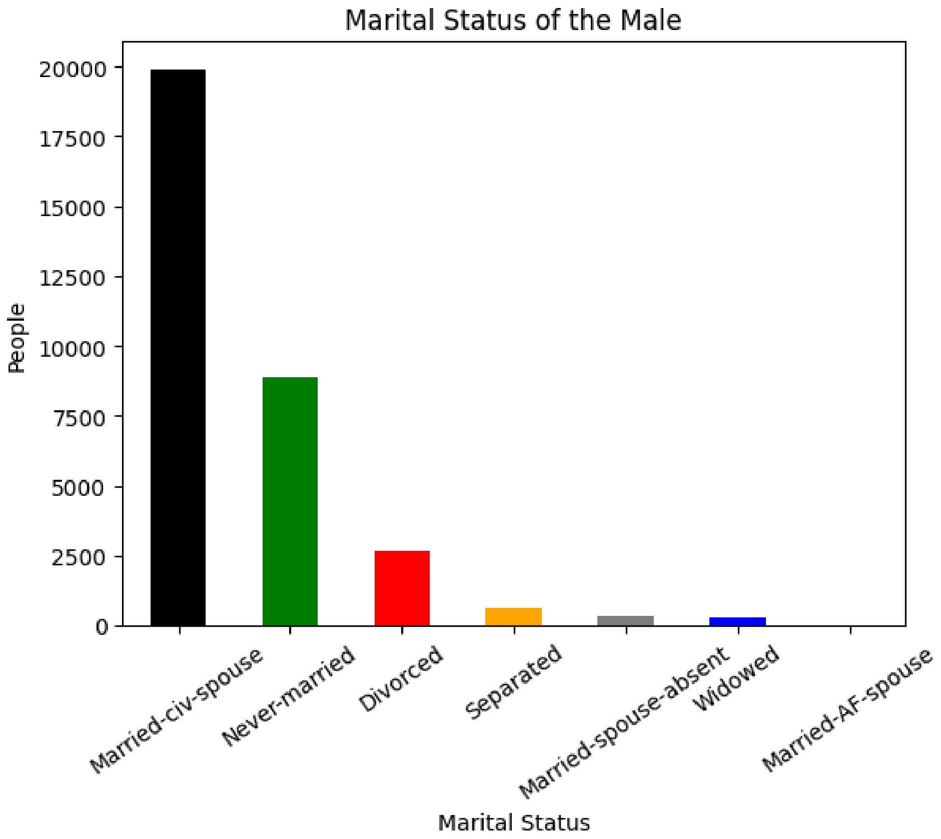
```
df.Birthplace.value_counts().plot(
    kind='barh', title='Birthplace of the Population', rot=0, color = ['Brown', 'Gold', 'Red', 'Orange', 'Blue']
)


plt.xlabel('Birthplace')
plt.ylabel('People')
plt.xticks(rotation=90)


plt.show()
```



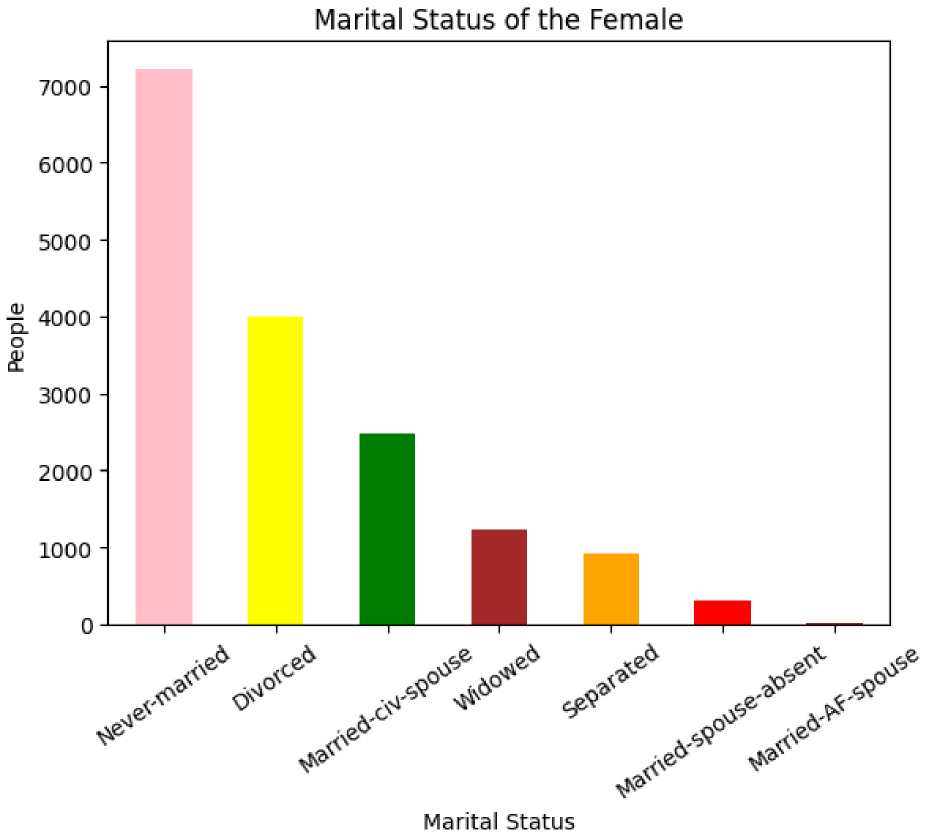
```
df.marital_status.value_counts().plot(
    kind='bar', title='Marital Status of the Population', rot=0, color = ['Red', 'Black', 'Green', 'Gray', 'Blue', 'Orange', 'Brown']
)

plt.xlabel('Marital Status')
plt.ylabel('People')
plt.xticks(rotation=35)

plt.show()
```



```
male.marital_status.value_counts().plot(
    kind='bar', title='Marital Status of the Male', rot=0, color = ['Black', 'Green', 'Red', 'Orange', 'Gray', 'Blue', 'Brown']
)

plt.xlabel('Marital Status')
plt.ylabel('People')
plt.xticks(rotation=35)

plt.show()
```

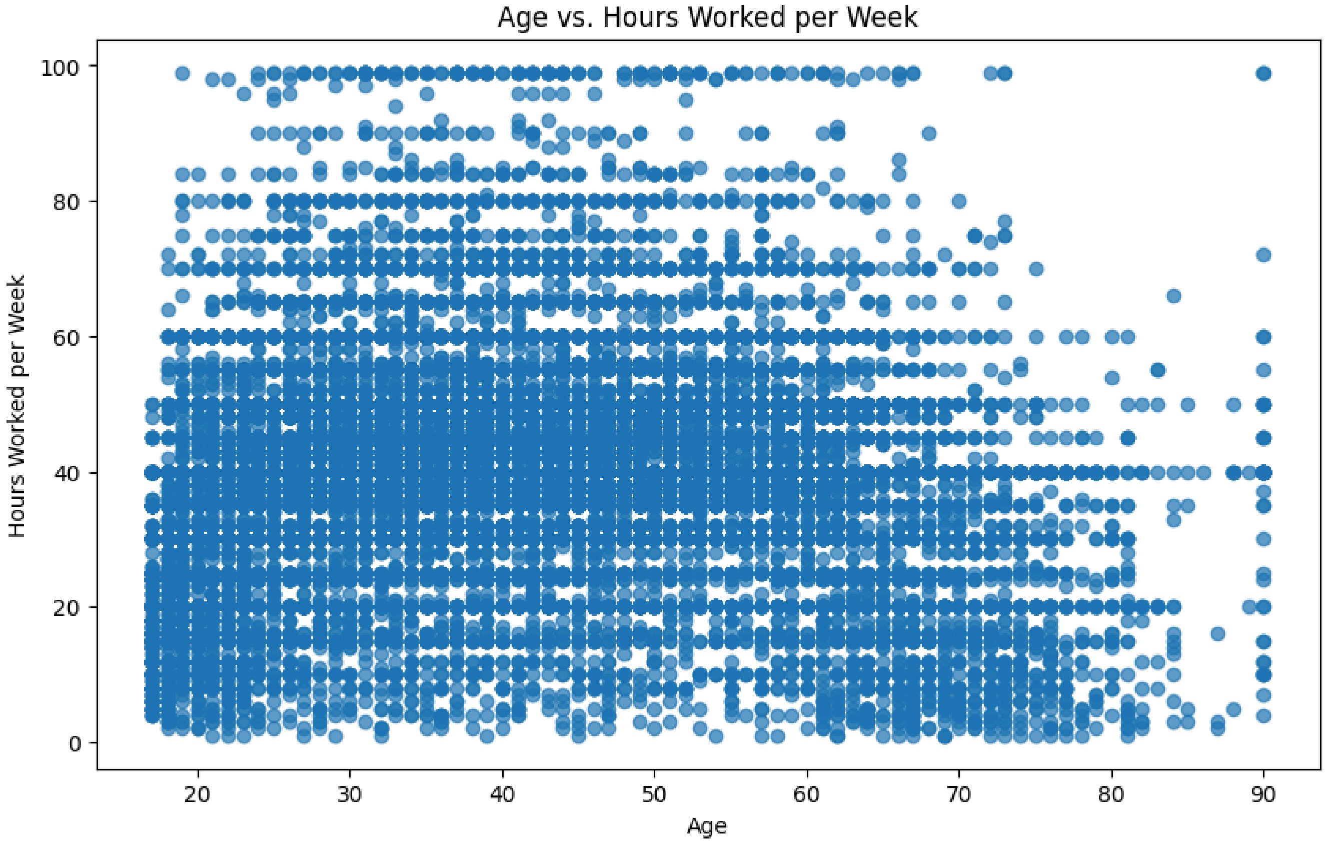


```
female.marital_status.value_counts().plot(
    kind='bar', title='Marital Status of the Female', rot=0, color = ['Pink', 'Yellow', 'Green', 'Brown', 'Orange', 'Red', 'Brown']
)

plt.xlabel('Marital Status')
plt.ylabel('People')
plt.xticks(rotation=35)

plt.show()
```

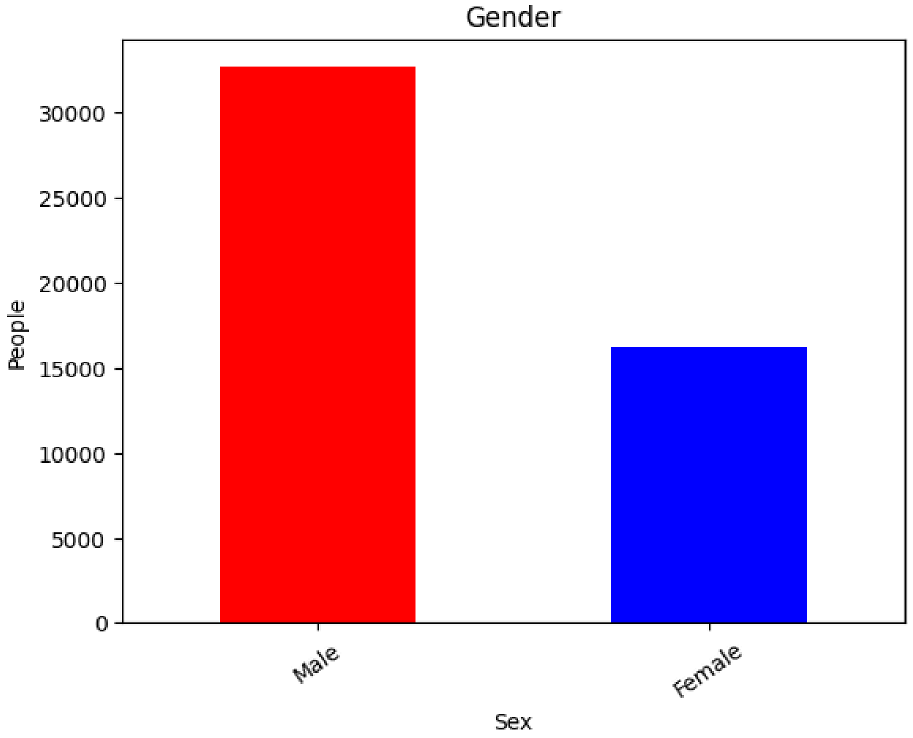


```
plt.figure(figsize=(10, 6))
plt.scatter(df["age"], df["Weekly_Hours"], alpha=0.7)

# Label axis and add title
plt.xlabel("Age")
plt.ylabel("Hours Worked per Week")
plt.title("Age vs. Hours Worked per Week")
```

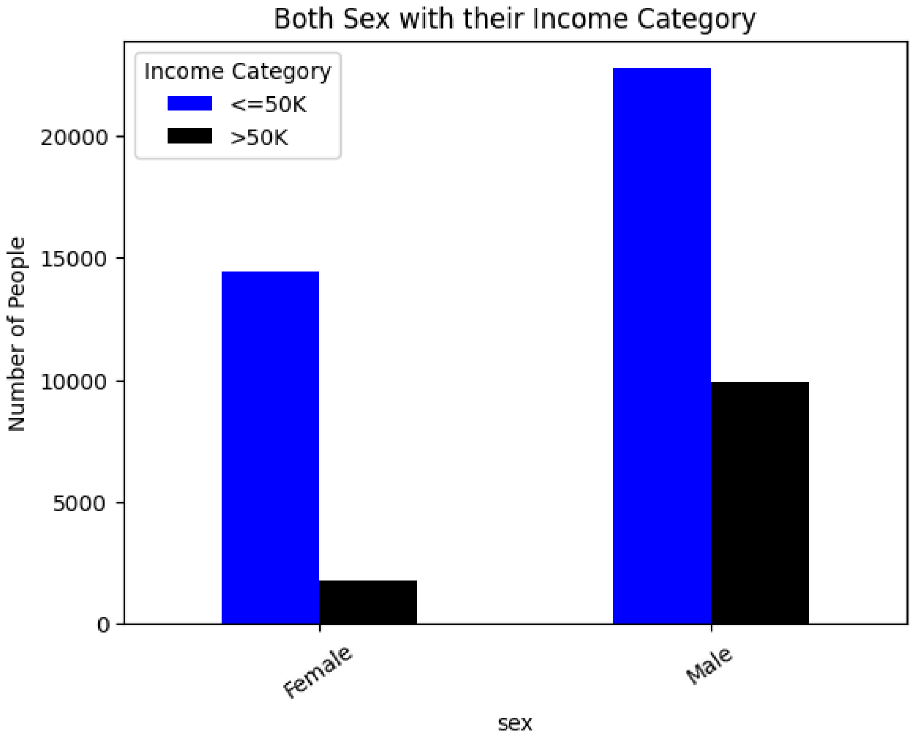
```
Text(0.5, 1.0, 'Age vs. Hours Worked per Week')
```



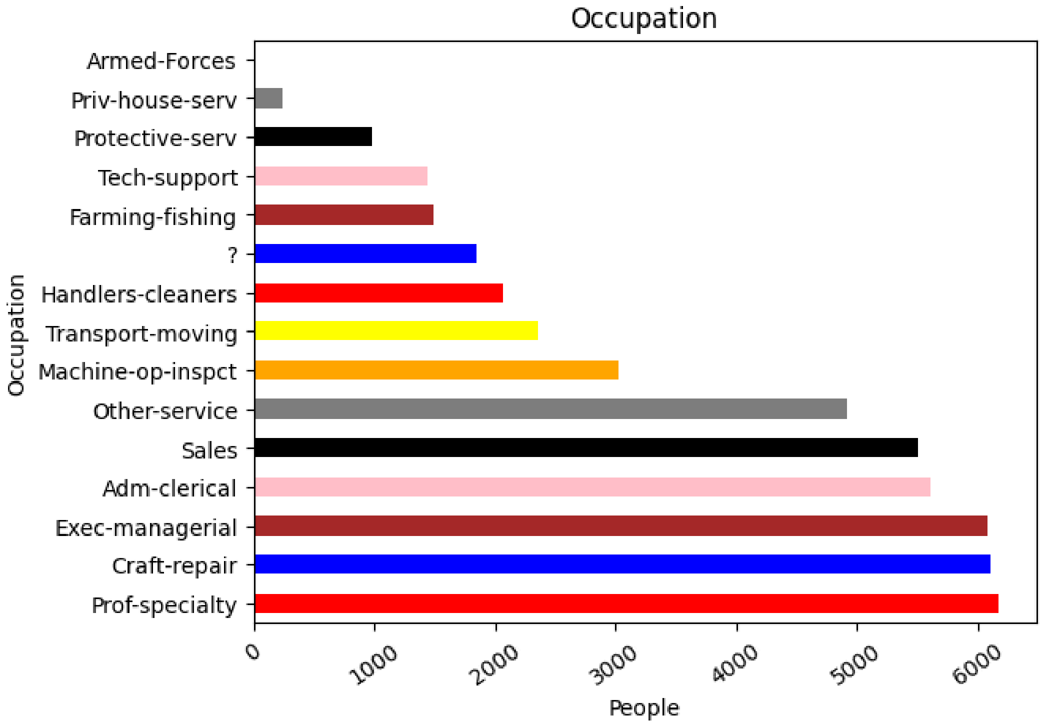
```
df.sex.value_counts().plot(
    kind='bar', title='Gender', rot=0, color = ['Red', 'Blue']
)

plt.xlabel('Sex')
plt.ylabel('People')
plt.xticks(rotation=35)


plt.show()
```



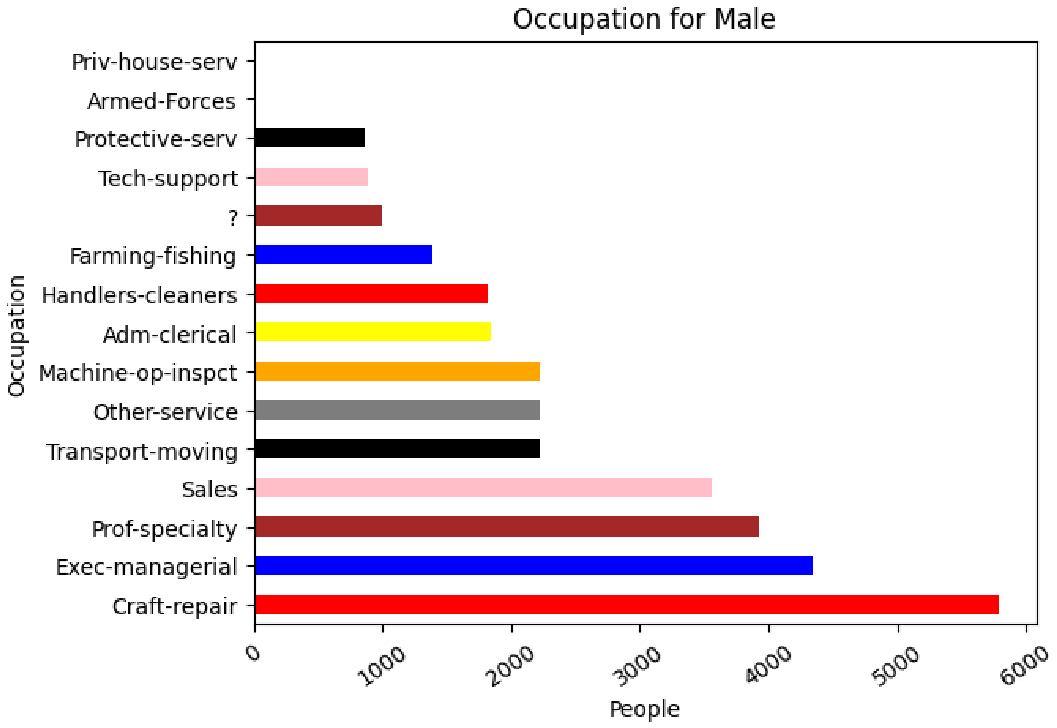
```
GIc = df.groupby(['sex', 'income']).size().unstack()


# Create the bar chart
GIc.plot(kind='bar', stacked=False, color=['Blue', 'Black'])
plt.xlabel("sex")
plt.ylabel("Number of People")
plt.title("Both Sex with their Income Category")
plt.legend(title="Income Category")
plt.xticks(rotation=35)
plt.show()
```



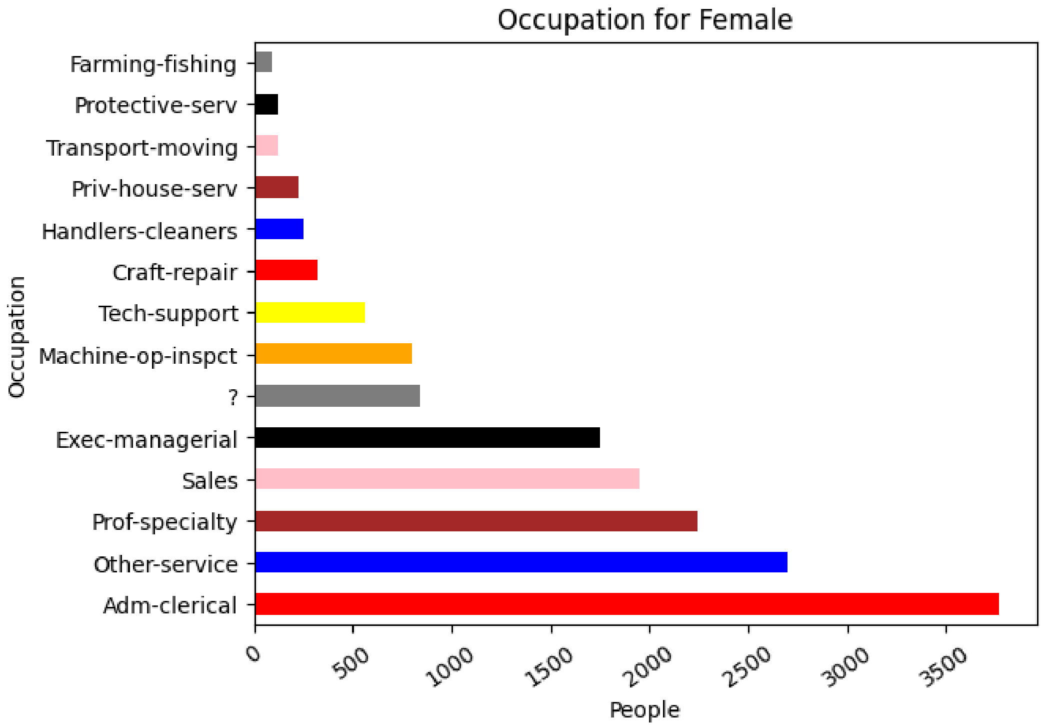
```
df.occupation.value_counts().plot(
    kind='barh', title='Occupation', rot=0, color = ['Red', 'Blue', 'Brown', 'Pink', 'Black','Gray','Orange','Yellow']
)

plt.xlabel('People')
plt.ylabel('Occupation')
plt.xticks(rotation=35)


plt.show()
```



```
male.occupation.value_counts().plot(
    kind='barh', title='Occupation for Male', rot=0, color = ['Red', 'Blue', 'Brown', 'Pink', 'Black','Gray','Orange','Yellow']
)

plt.xlabel('People')
plt.ylabel('Occupation')
plt.xticks(rotation=35)


plt.show()
```

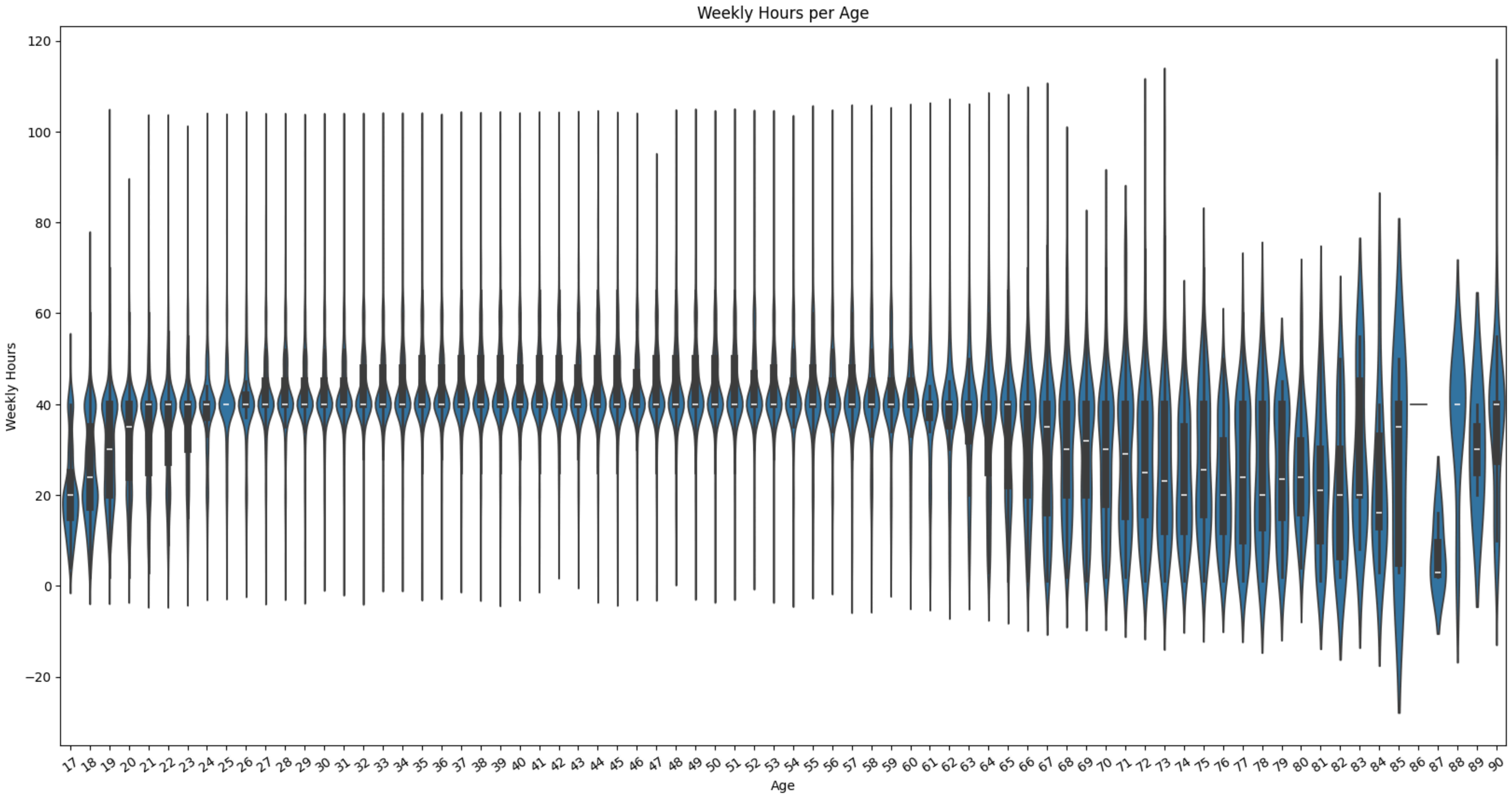


```
female.occupation.value_counts().plot(
    kind='barh', title='Occupation for Female', rot=0, color = ['Red', 'Blue', 'Brown', 'Pink', 'Black','Gray','Orange','Yellow']
)

plt.xlabel('People')
plt.ylabel('Occupation')
plt.xticks(rotation=35)


plt.show()
```



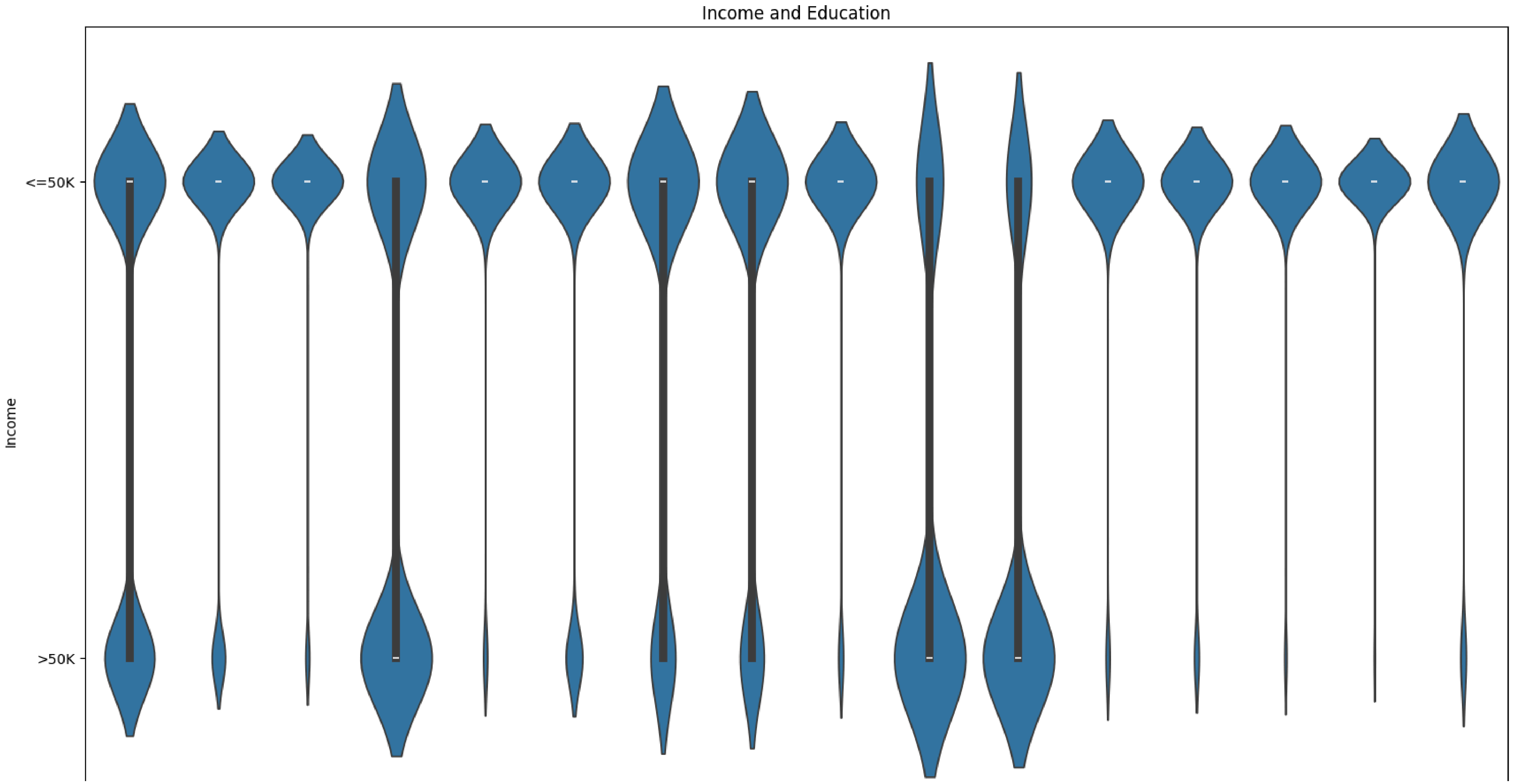
```
fig, axes = plt.subplots(figsize=(20, 10))
sns.violinplot(
    x="age",
    y="Weekly_Hours",
    data=df,        # Pass the entire DataFrame
    scale='width'   # Set same width for violins
)

plt.title('Weekly Hours per Age')

plt.xlabel('Age')
plt.ylabel('Weekly Hours')
plt.xticks(rotation=35)

plt.show()
```

```
<ipython-input-35-3bc38fd75055>:2: FutureWarning:

The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `density_norm='width'` for the same effect.
  sns.violinplot(
```



```
fig, axes = plt.subplots(figsize=(18, 10))
sns.violinplot(
    x="education",
    y="income",
    data=df,        # Pass the entire DataFrame
    scale='width'   # Set same width for violins
)

plt.title('Income and Education')

plt.xlabel('Education Level')
plt.ylabel('Income')
plt.xticks(rotation=35)

plt.show()
```

```
<ipython-input-36-8b86efdcc8ed>:2: FutureWarning:

The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `density_norm='width'` for the same effect.
  sns.violinplot(
```



```
df.workclass.value counts().plot(
```