A Study of Drug Use by Age Groups

Stephanie Chiang
DATA 602 Fall 2024
Final Project

What patterns, if any, exist in drug use across different age groups?

For example, are there age groups that exhibit higher use rates of a particular type of drug?

Are there noticeable relationships between the frequency or popularity certain drugs?

For example, does marijuana correlate with opioid use?

```
for i in range(len(new_idx)):
    if i <= 7:
        new_idx[i] = "12-19"
    elif i <= 13:
        new_idx[i] = "20-34"
    else:
        new_idx[i] = drug_data.iloc[i]["age"]

# add the new index column for aggregation
drug_data["age_range"] = new_idx

# confirm the age_range column was added correctly
print(drug_data.head().iloc[:, -3:])
```

✓ 0.0s

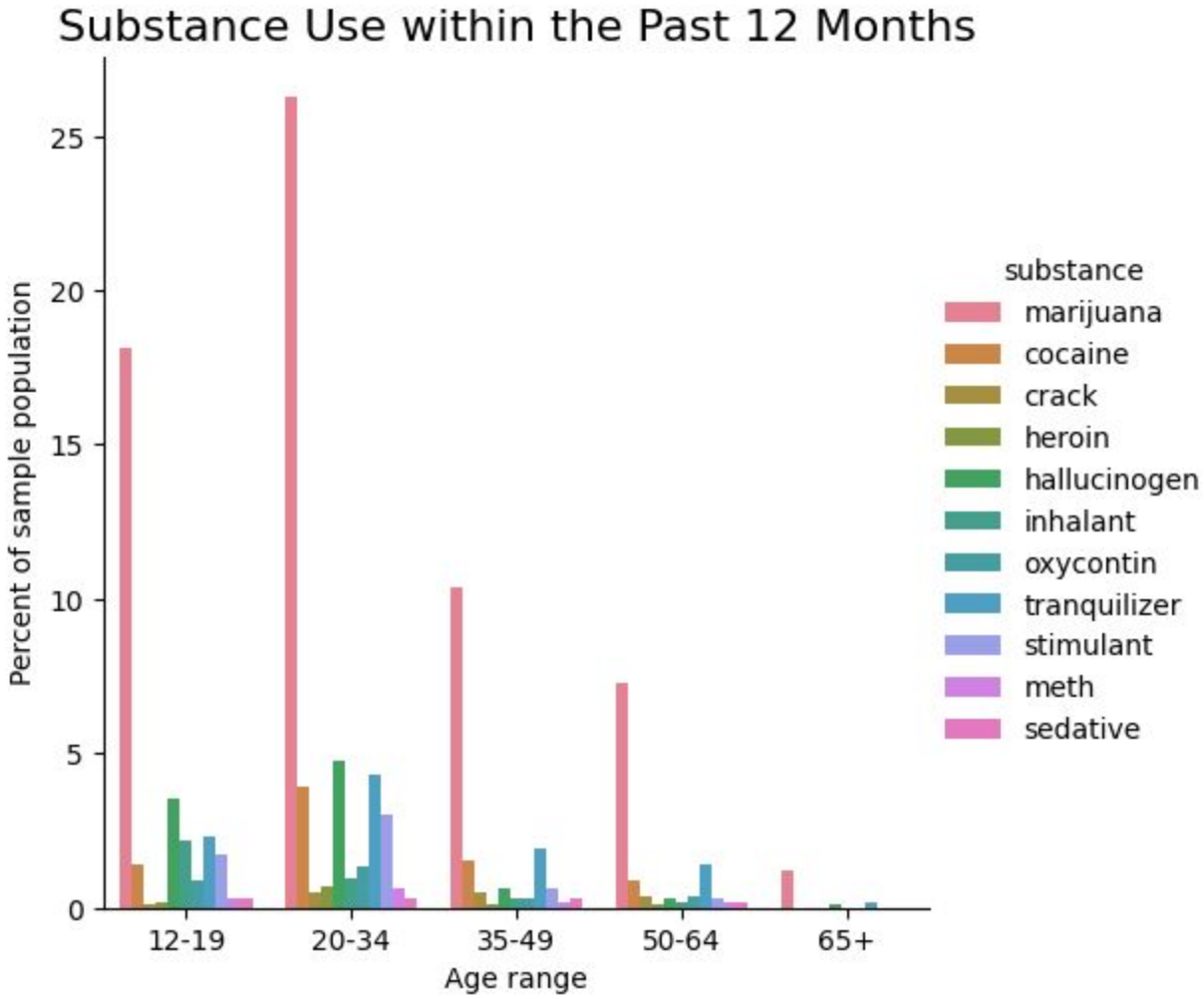
	sedative_use	sedative_frequency	age_range
0	0.2	13.0	12-19
1	0.1	19.0	12-19
2	0.2	16.5	12-19
3	0.4	30.0	12-19
4	0.2	3.0	12-19

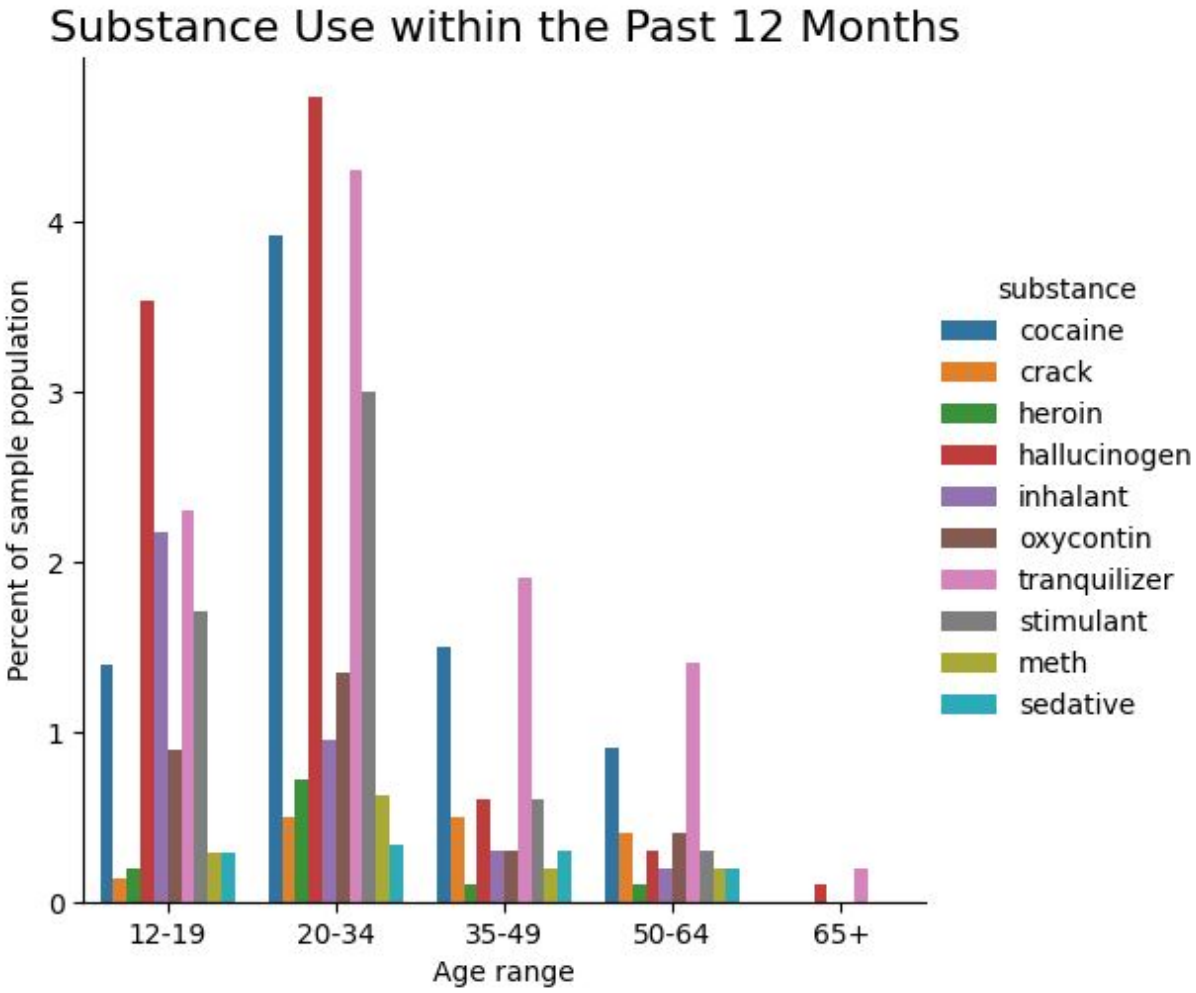
```
# column split and cleanup
temp = grp_drug["substance"].str.split("_", n=1, expand=True)
grp_drug["substance"] = temp[0]
grp_drug["value_type"] = temp[1]

grp_drug["n"] = grp_drug["n"].round(0)
grp_drug["result"] = grp_drug["result"].round(2)
grp_drug = grp_drug.iloc[:, [0, 1, 2, 4, 3]]

print(grp_drug)
```

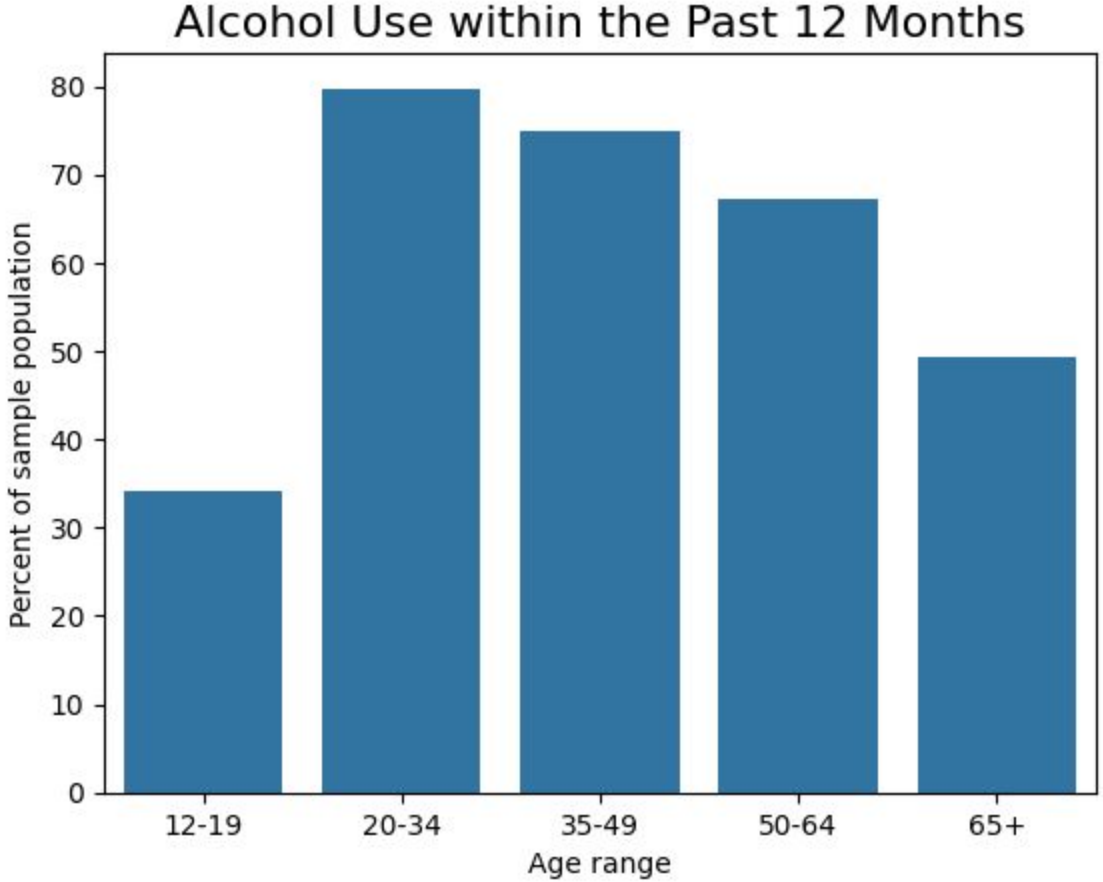
✓ 0.0s

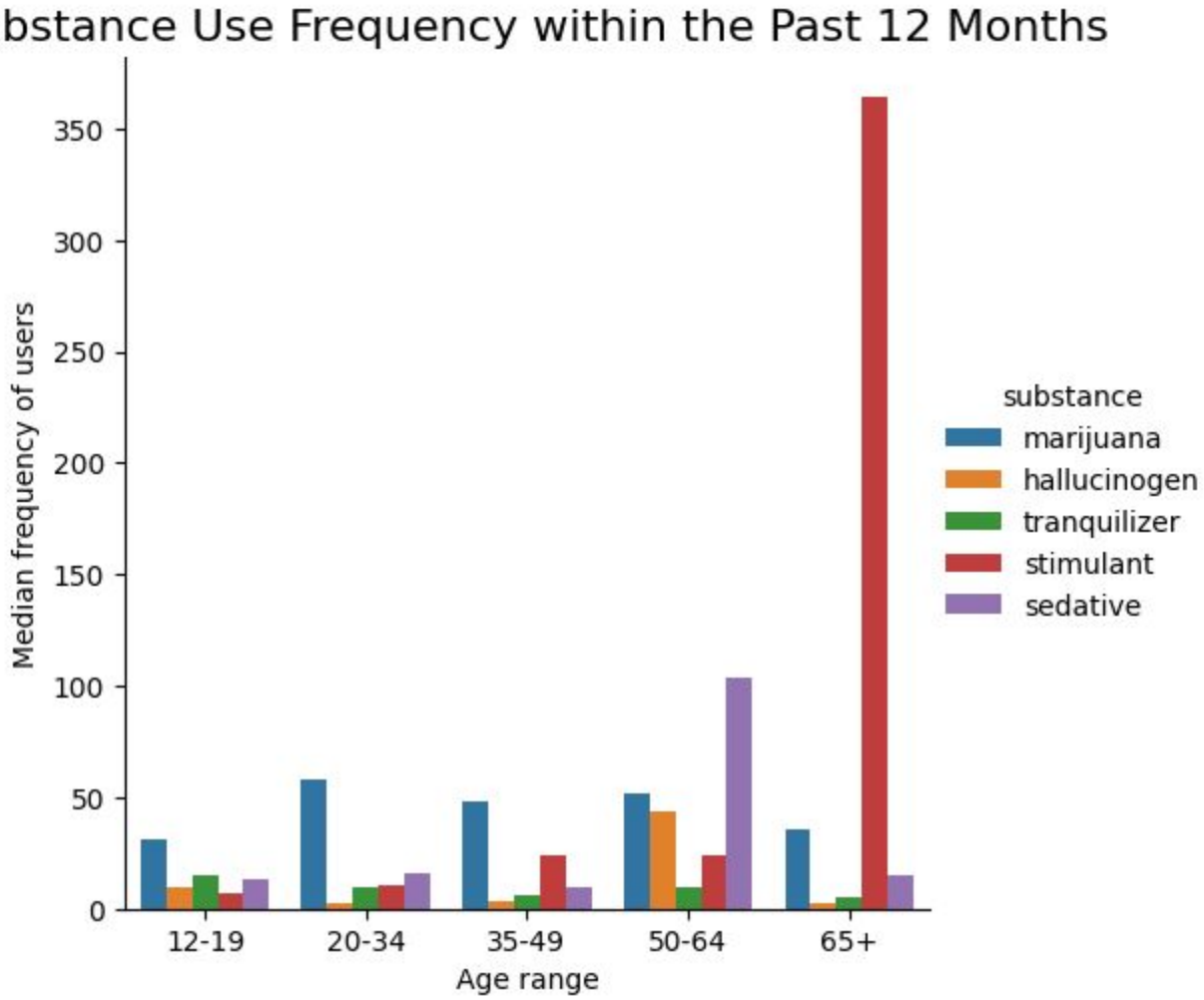
	age_range	n	substance	value_type	result
0	12-19	2761.0	alcohol	use	34.05
1	20-34	3236.0	alcohol	use	79.73
2	35-49	7391.0	alcohol	use	75.00
3	50-64	3923.0	alcohol	use	67.20
4	65+	2448.0	alcohol	use	49.30
..	...	...	...	...	...
95	12-19	2761.0	sedative	frequency	13.00
96	20-34	3236.0	sedative	frequency	16.08
97	35-49	7391.0	sedative	frequency	10.00
98	50-64	3923.0	sedative	frequency	104.00
99	65+	2448.0	sedative	frequency	15.00

[100 rows x 5 columns]


Substance Use within the Past 12 Months
Percent of sample population
0
5
10
15
20
25
12-19
20-34
35-49
50-64
65+
Age range
substance
marijuana
cocaine
crack
heroin
hallucinogen
inhalant
oxycontin
tranquilizer
stimulant
meth
sedative


Substance Use within the Past 12 Months
Percent of sample population
0
1
2
3
4
12-19
20-34
35-49
50-64
65+
Age range
substance
cocaine
crack
heroin
hallucinogen
inhalant
oxycontin
tranquilizer
stimulant
meth
sedative


Alcohol Use within the Past 12 Months
Percent of sample population
0
10
20
30
40
50
60
70
80
12-19
20-34
35-49
50-64
65+
Age range


Substance Use Frequency within the Past 12 Months
Median frequency of users
0
50
100
150
200
250
300
350
12-19
20-34
35-49
50-64
65+
Age range
substance
marijuana
hallucinogen
tranquilizer
stimulant
sedative

Age groups with the highest and lowest
median number of uses in the past year for each drug:

age_range	substance	
50-64	hallucinogen	44
20-34	marijuana	58
50-64	sedative	104
65+	stimulant	364
12-19	tranquilizer	15

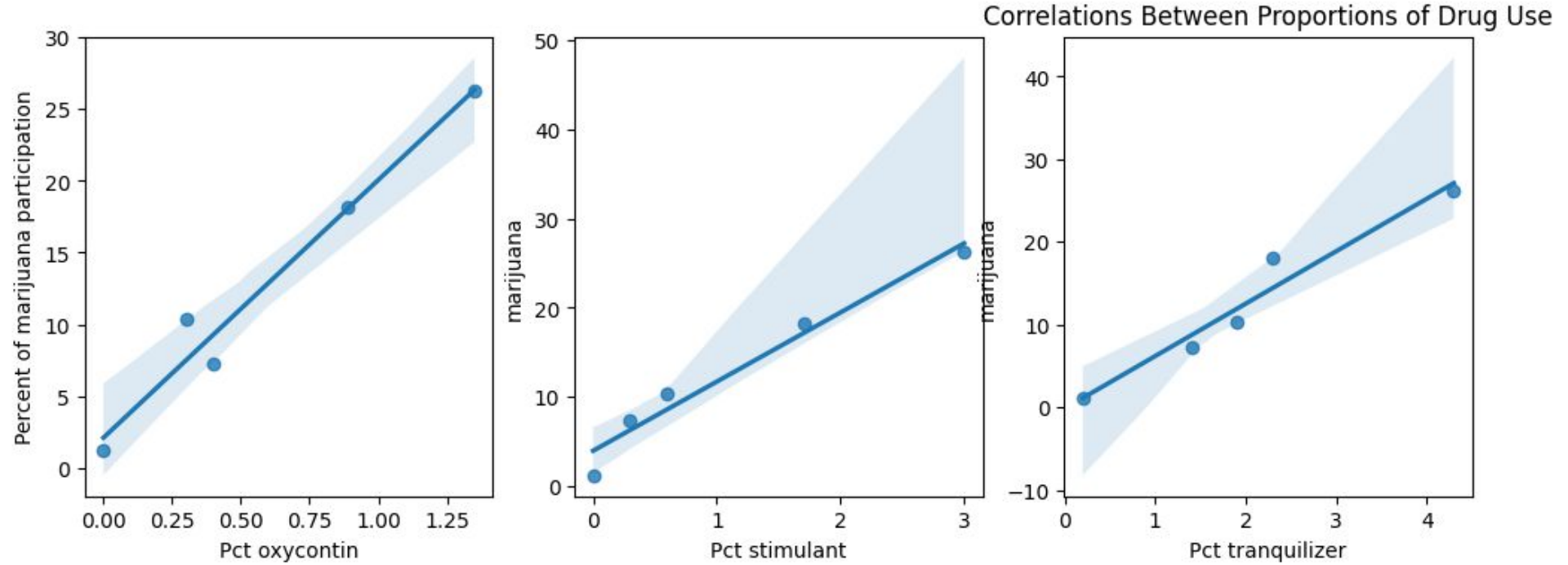
age_range	substance	
65+	hallucinogen	2
12-19	marijuana	31
35-49	sedative	10
12-19	stimulant	7
65+	tranquilizer	5

Correlation of substances with marijuana

```python
# pivot wider again for correlation
use_corr = subuse_df.pivot(index="age_range", columns="substance", values="result")
correlation_matrix = use_corr.corr()
marijuana_corr = correlation_matrix["marijuana"].sort_values(ascending=False)

print(marijuana_corr)
```

✓ 0.0s

```
substance
marijuana       1.000000
oxycontin       0.982428
stimulant       0.981854
tranquilizer    0.974796
meth            0.960199
hallucinogen    0.957875
cocaine         0.928600
heroin          0.901355
sedative        0.826518
inhalant        0.670386
crack           0.485423
Name: marijuana, dtype: float64
```


Correlations Between Proportions of Drug Use
Percent of marijuana participation
Pct oxycontin
marijuana
Pct stimulant
marijuana
Pct tranquilizer

In conclusion..

To answer the questions posed at the start of this analysis, there are some notable patterns in drug use across different age groups. Though every group had the highest percentage of participants with alcohol and marijuana (likely due to legality), the following trends were of note:

Younger Adults (20-34 years): Highest proportion of participants for every substance, except for `inhalants` (which had a higher teen percentage).

Older Adults (50-64 years): Highest frequency of use for nearly every substance, with particular "preference" for hallucinogens and sedatives.

Elderly Adults (65+ years): Lowest rates of participation and frequency, but with outliers.

The data was also able to illuminate noticeable relationships between marijuana and nearly every other substance. And in fact, only alcohol had a low correlation coefficient: 0.236558. Therefore, marijuana correlates positively with increases in the proportion of the population that use illicit substances.