DSCI Imports

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn import preprocessing
from sklearn.ensemble import RandomForestRegressor
```

Data Cleaning

```
dfall = pd.read_csv("data/MentalHealthSurvey.csv")

print(dfall.columns)
print(pd.unique(dfall.values.ravel("K")))

df = dfall.drop(['university', 'degree_level', 'degree_major', 'stress_relief_activities'], a

df =df.replace(to_replace="Male", value=1)
df =df.replace(to_replace="Female", value=-1)
df =df.replace(to_replace="1st year", value=1)
df =df.replace(to_replace="2nd year", value=2)
df =df.replace(to_replace="3rd year", value=3)
df =df.replace(to_replace="4th year", value=4)
df =df.replace(to_replace="0.0-0.0", value=0)
df =df.replace(to_replace="1.5-2.0", value=1)
df =df.replace(to_replace="2.0-2.5", value=2)
df =df.replace(to_replace="2.5-3.0", value=3)
df =df.replace(to_replace="3.0-3.5", value=4)
df =df.replace(to_replace="3.5-4.0", value=5)
df =df.replace(to_replace="Off-Campus", value=-1)
df =df.replace(to_replace="On-Campus", value=1)
df =df.replace(to_replace="Yes", value=1)
df =df.replace(to_replace="No", value=-1)
df =df.replace(to_replace="No Sports", value=0)
df =df.replace(to_replace="1-3 times", value=1)
df =df.replace(to_replace="4-6 times", value=2)
df =df.replace(to_replace="7+ times", value=3)
df =df.replace(to_replace="4-6 hrs", value=5)
```

```
df =df.replace(to_replace="2-4 hrs", value=3)
df =df.replace(to_replace="7-8 hrs", value=8)

print(pd.unique(df.values.ravel("K")))
```

Creating Regressions

```
x = df.drop(['depression', 'anxiety', 'isolation', 'future_insecurity', 'gender', 'age', 'aca

depression = df['depression']
anxiety = df['anxiety']
isolation = df['isolation']
insecurity = df['future_insecurity']
```

Train Test Split

```
dx_train, dx_test, dy_train, dy_test = train_test_split(x, depression, test_size=0.15, rando
ax_train, ax_test, ay_train, ay_test = train_test_split(x, anxiety, test_size=0.15, random_st
ix_train, ix_test, iy_train, iy_test = train_test_split(x, isolation, test_size=0.15, random_
fx_train, fx_test, fy_train, fy_test = train_test_split(x, insecurity, test_size=0.15, rando
```

Running Regressions & Accuracy Outputs

```
model_depression = LinearRegression()
rf_depression = RandomForestRegressor(n_estimators=10, random_state=0, oob_score=True)

model_depression.fit(dx_train, dy_train)
rf_depression.fit(dx_train, dy_train)

print('Depression LIN REG Variance Score: {}'.format(model_depression.score(dx_test,dy_test))
print('Depression Random Forest R^2 Score: {}'.format(rf_depression.score(dx_test,dy_test)))

model_anxiety = LinearRegression()
rf_anxiety = RandomForestRegressor(n_estimators=10, random_state=0, oob_score=True)

model_anxiety.fit(ax_train, ay_train)
rf_anxiety.fit(ax_train, ay_train)

print('Anxiety LIN REG Variance Score: {}'.format(model_anxiety.score(ax_test,ay_test)))
print('Anxiety Random Forest R^2 Score: {}'.format(rf_anxiety.score(ax_test,ay_test)))
```

```
model_isolation = LinearRegression()
rf_isolation = RandomForestRegressor(n_estimators=10, random_state=0, oob_score=True)

model_isolation.fit(ix_train, iy_train)
rf_isolation.fit(ix_train, iy_train)

print('Isolation LIN REG Variance Score: {}'.format(model_isolation.score(ix_test,iy_test)))
print('Isolation Random Forest R^2 Score: {}'.format(rf_isolation.score(ix_test,iy_test)))

model_insecurity = LinearRegression()
rf_insecurity = RandomForestRegressor(n_estimators=10, random_state=0, oob_score=True)

model_insecurity.fit(fx_train, fy_train)
rf_insecurity.fit(fx_train, fy_train)

print('Insecurity LIN REG Variance Score: {}'.format(model_insecurity.score(fx_test,fy_test)))
print('Insecurity Random Forest R^2 Score: {}'.format(rf_insecurity.score(fx_test,fy_test)))

models = [model_depression, model_anxiety, model_isolation, model_insecurity]
y = [depression, anxiety, isolation, insecurity]
y_names = ["depression", "anxiety", "isolation", "future_insecurity"]
```

```
Depression LIN REG Variance Score: 0.19516854604834477
Depression Random Forest R^2 Score: 0.31976851851851873
Anxiety LIN REG Variance Score: 0.4013545618687956
Anxiety Random Forest R^2 Score: 0.09548192771084363
Isolation LIN REG Variance Score: 0.5431250523573483
Isolation Random Forest R^2 Score: 0.5751595744680852
Insecurity LIN REG Variance Score: 0.14098686977821573
Insecurity Random Forest R^2 Score: 0.035624999999999796
```

Model Coeffecients

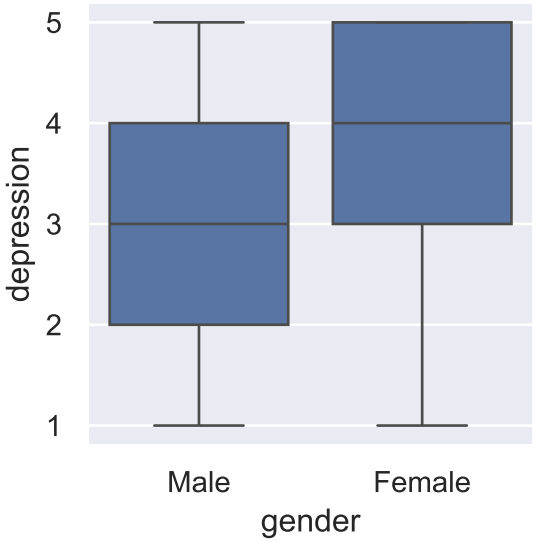
```
for i in range(len(x.columns)):
  print()
  print(x.columns[i])
  for model in models:
    print(round(model.coef_[i], 7), end="\t")
```

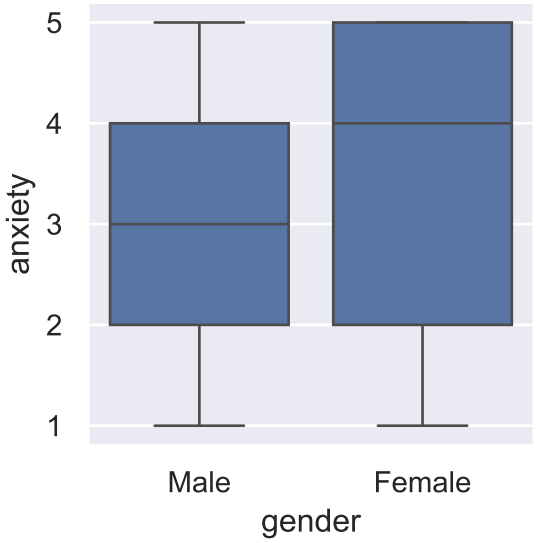
```
cgpa
```

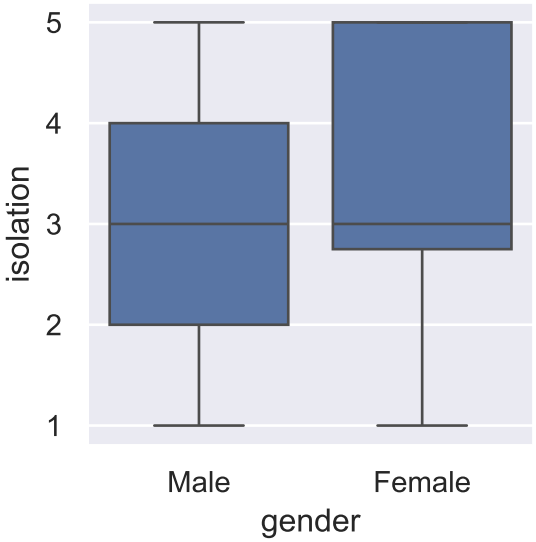
```
0.225703    0.0955451   0.0606372   0.2420488
residential_status
-0.2213241  -0.1448405  -0.1548686  -0.1799232
campus_discrimination
0.1220193   0.0849769   -0.0286518  0.3089447
sports_engagement
-0.0372295  -0.0516192  -0.0338199  0.1073513
average_sleep
0.0871455   0.0499156   0.024717    0.0985726
study_satisfaction
-0.2839526  -0.1853568  -0.2389771  -0.4686896
academic_workload
0.4296254   0.3916275   0.0446256   0.2888843
academic_pressure
0.3592741   0.2860356   0.3205623   0.2526749
financial_concerns
0.336914    0.1711097   0.1372198   0.0807601
social_relationships
-0.1204017  -0.2568048  -0.5619233  -0.1589357
```

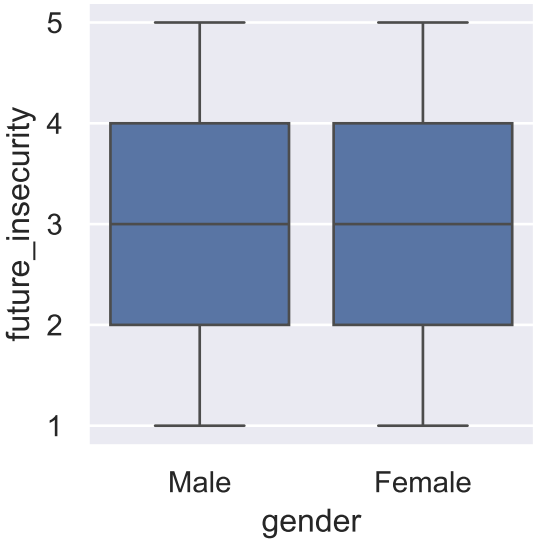
Exploratory Data Analysis Box Plots

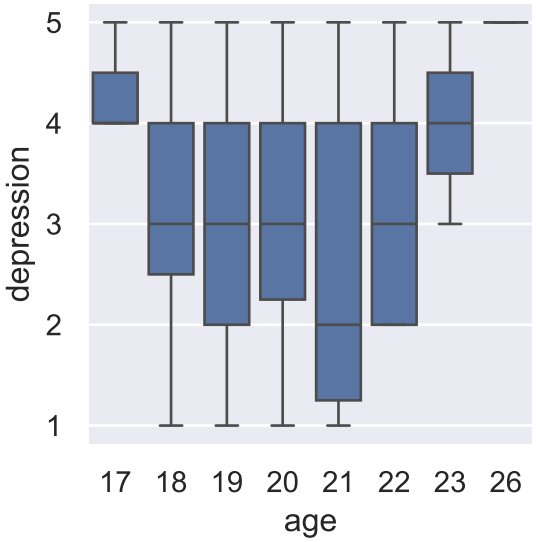
```
sns.set_theme(style="darkgrid")

x = ["gender", "age", "average_sleep", "cgpa", "residential_status", "campus_discrimination"

for i in x:
  for m in y:
    plt.figure(figsize=(3,3))
    sns.boxplot(x=dfall[i], y=m)
    plt.show()
```

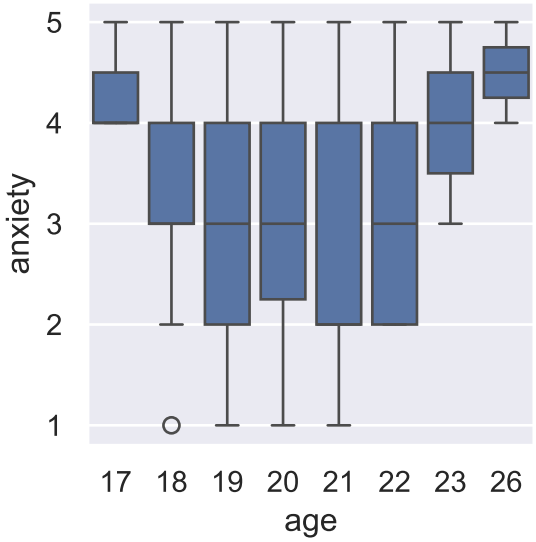

5
4
3
2
1
depression
Male
Female
gender

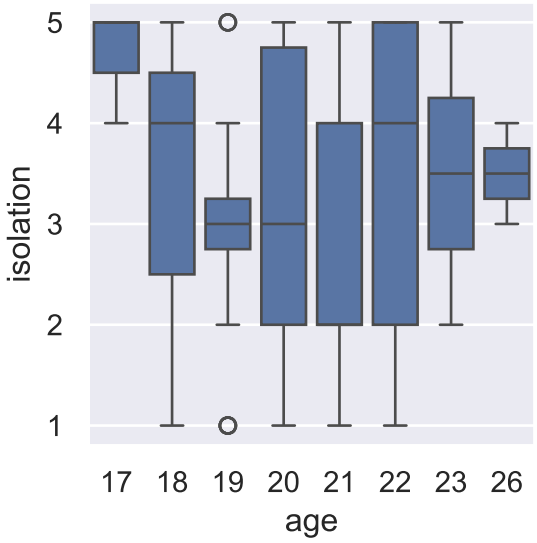

5
4
3
2
1
anxiety
Male
Female
gender

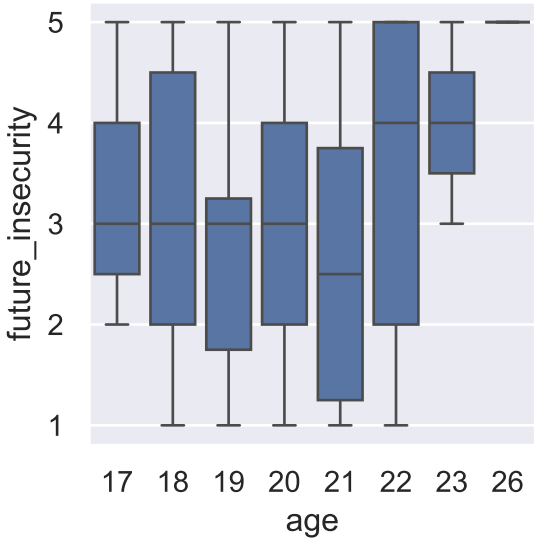

5
4
3
2
1
isolation
Male
Female
gender

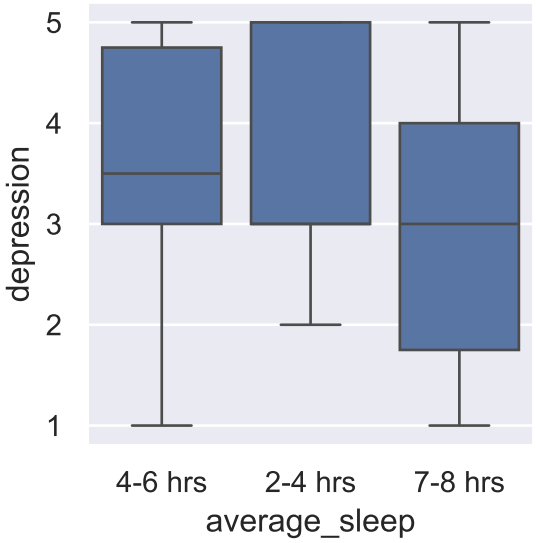

5
4
3
2
1
future_insecurity
Male
Female
gender

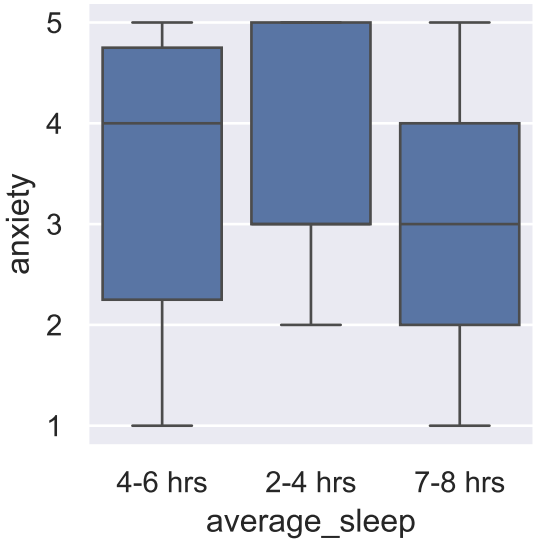

depression
5
4
3
2
1
17
18
19
20
21
22
23
26
age

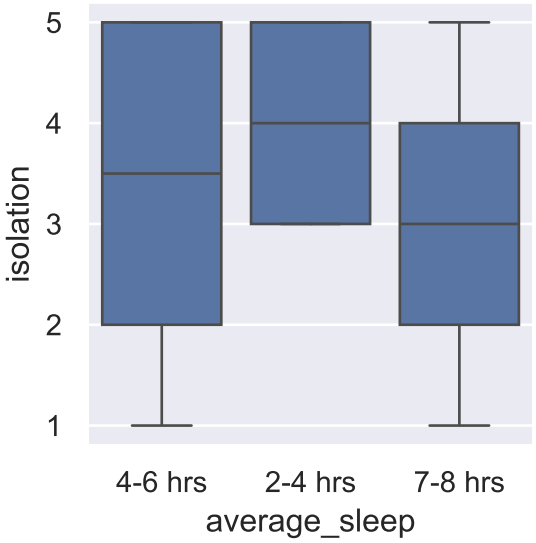

anxiety
5
4
3
2
1
17
18
19
20
21
22
23
26
age

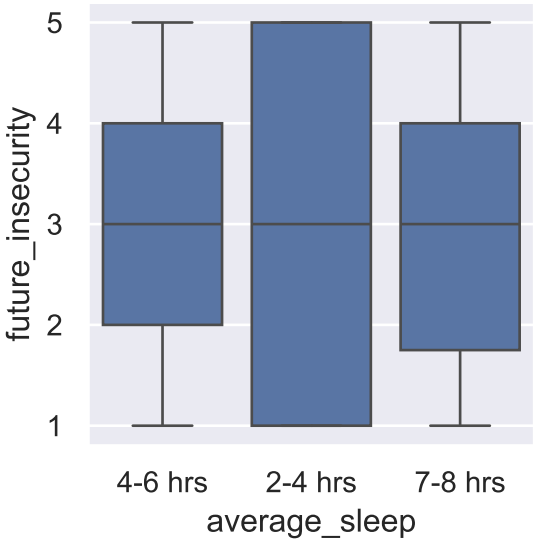

isolation
5
4
3
2
1
17
18
19
20
21
22
23
26
age

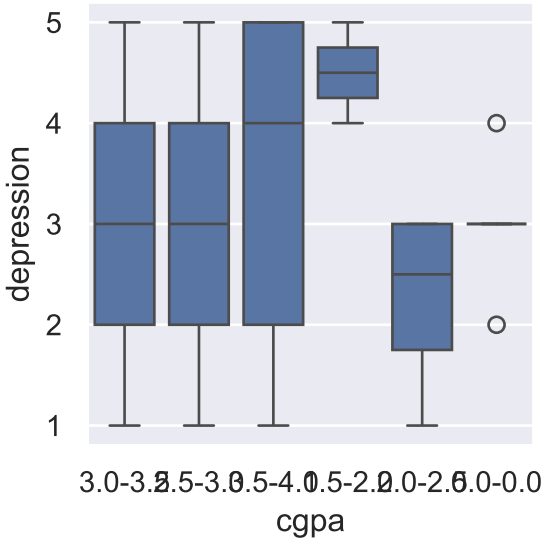

future_insecurity
5
4
3
2
1
17
18
19
20
21
22
23
26
age

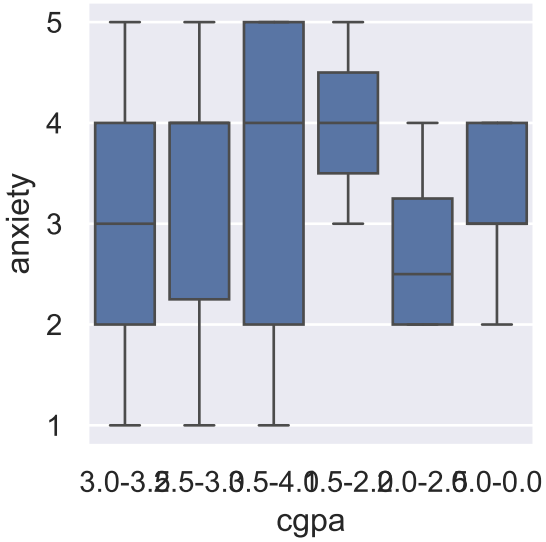

depression
5
4
3
2
1
4-6 hrs
2-4 hrs
7-8 hrs
average_sleep

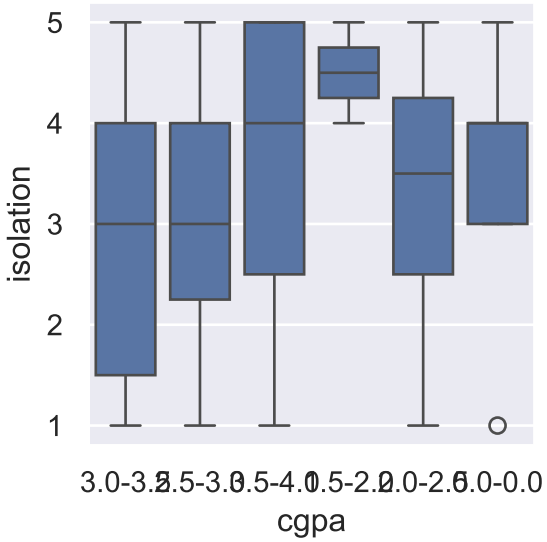

anxiety
5
4
3
2
1
4-6 hrs
2-4 hrs
7-8 hrs
average_sleep

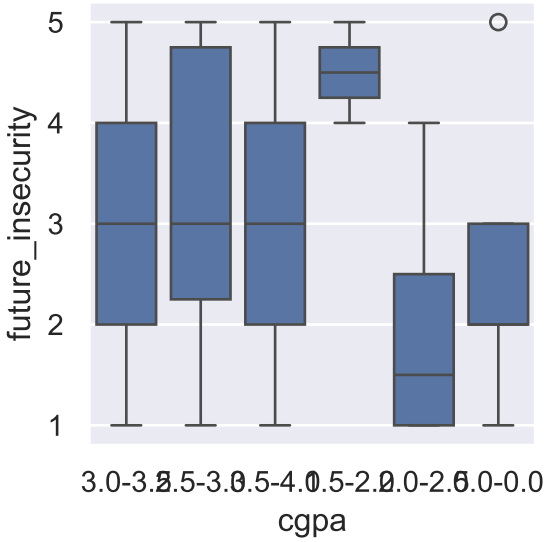

5
4
3
2
1
isolation
4-6 hrs
2-4 hrs
7-8 hrs
average_sleep

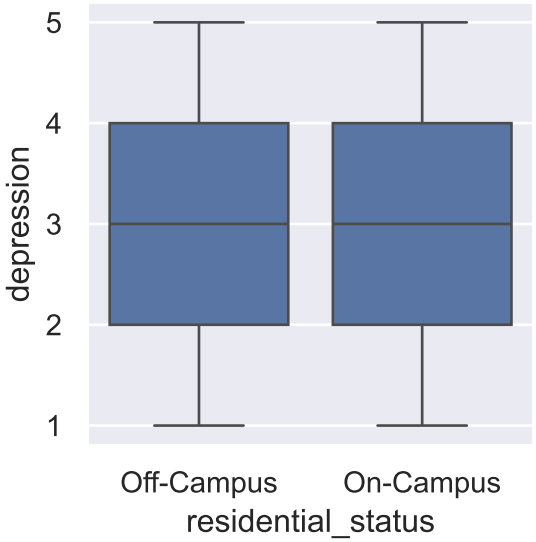

5
4
3
2
1
future_insecurity
4-6 hrs
2-4 hrs
7-8 hrs
average_sleep

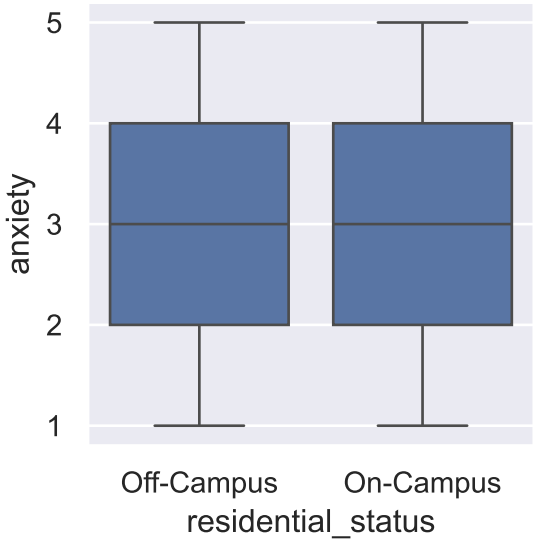

depression
5
4
3
2
1
3.0-3.2
2.5-3.0
3.5-4.0
1.5-2.0
2.0-2.5
0.0-0.0
cgpa

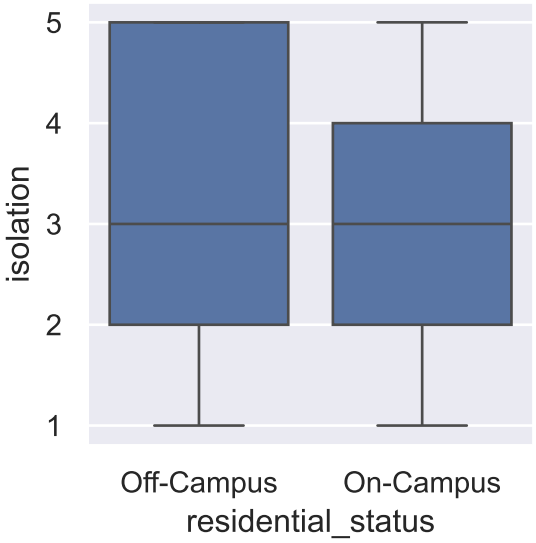

anxiety
5
4
3
2
1
3.0-3.2
2.5-3.0
3.5-4.0
1.5-2.0
2.0-2.5
0.0-0.0
cgpa

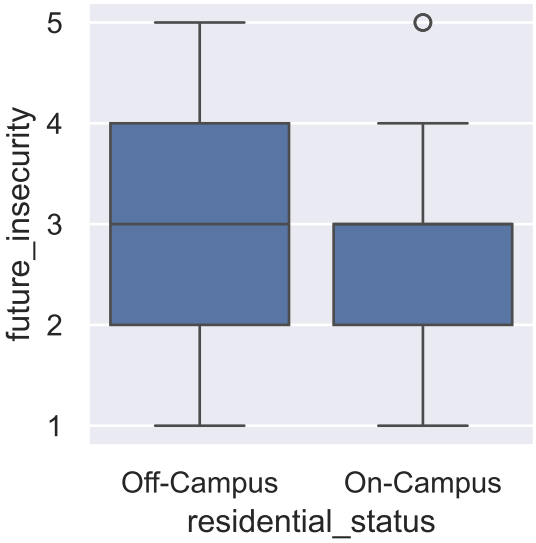

isolation
5
4
3
2
1
3.0-3.2
2.5-3.0
3.5-4.0
1.5-2.0
2.0-2.5
0.0-0.0
cgpa

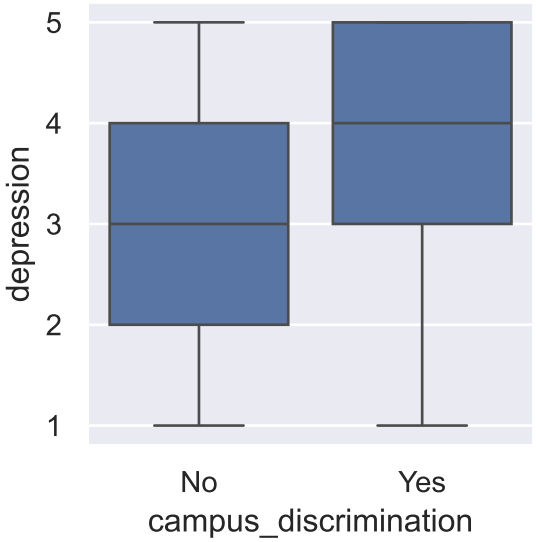

future_insecurity
5
4
3
2
1
3.0-3.2
2.5-3.0
3.5-4.0
1.5-2.0
2.0-2.5
0.0-0.0
cgpa

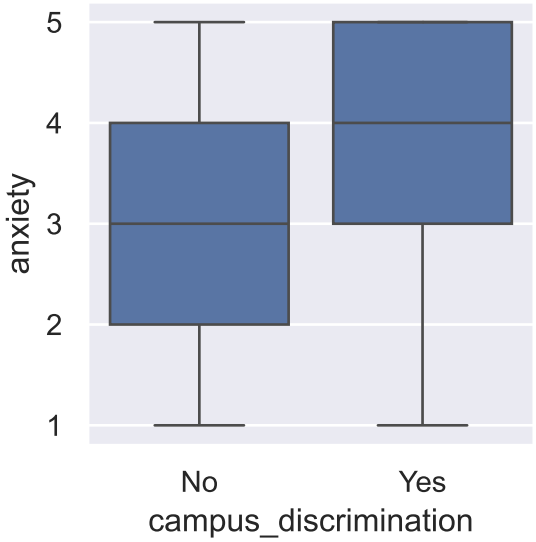

5
4
3
2
1
depression
Off-Campus
On-Campus
residential_status

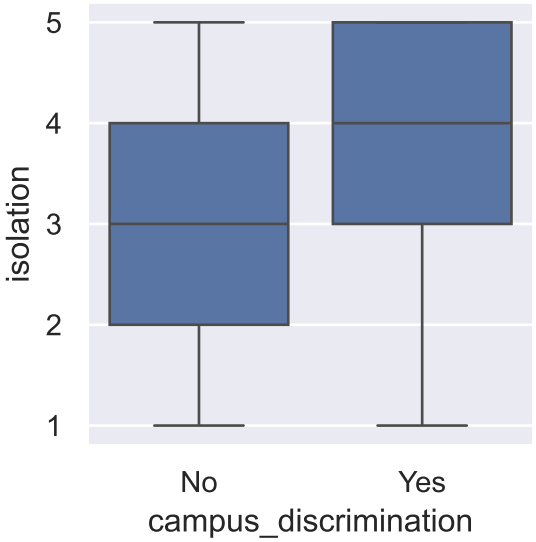

5
4
3
2
1
anxiety
Off-Campus
On-Campus
residential_status

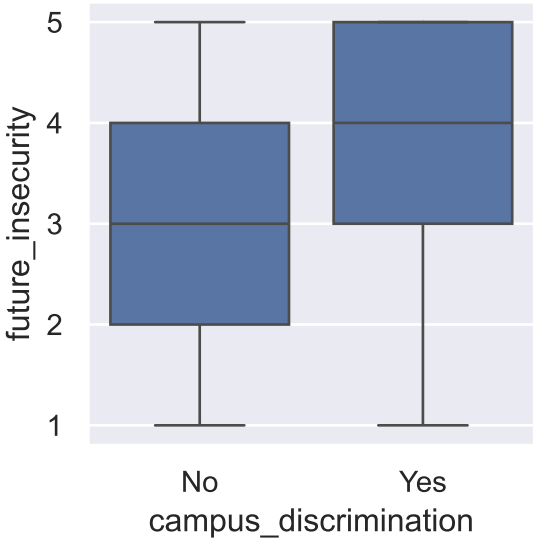

5
4
3
2
1
isolation
Off-Campus
On-Campus
residential_status

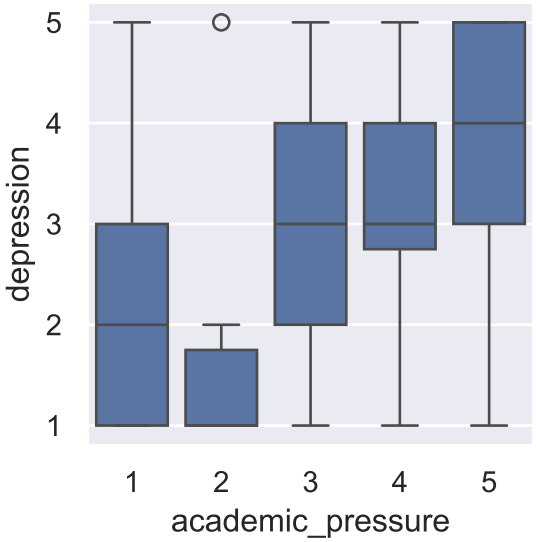

5
4
3
2
1
future_insecurity
Off-Campus
On-Campus
residential_status

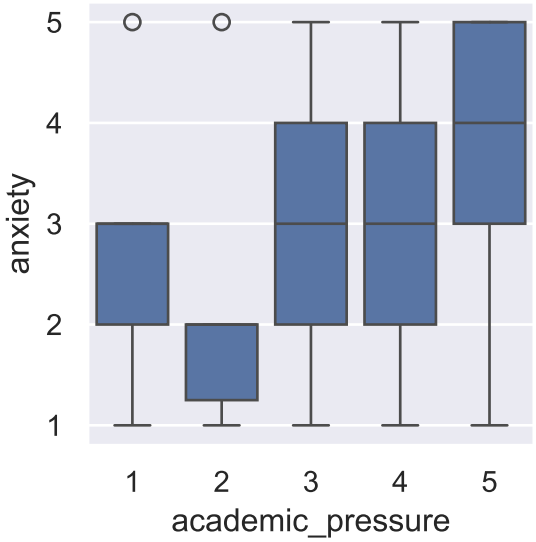

5
4
3
2
1
depression
No
Yes
campus_discrimination

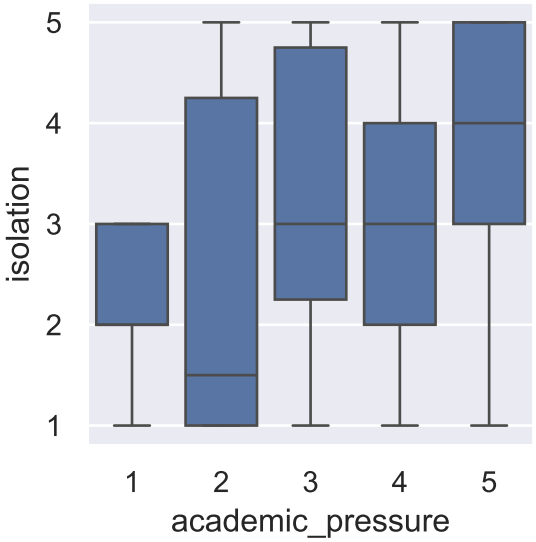

5
4
3
2
1
anxiety
No
Yes
campus_discrimination

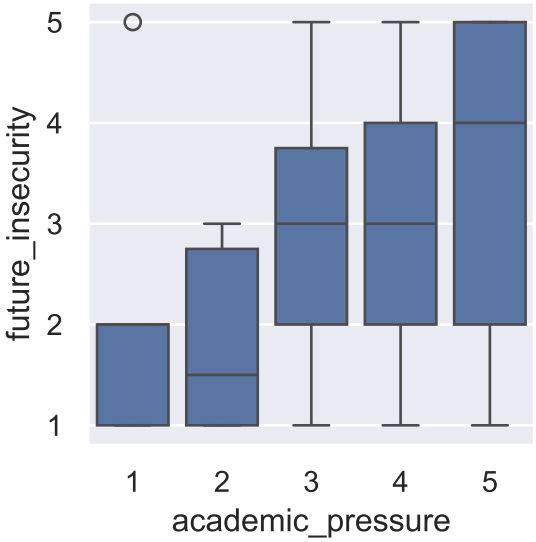

5
4
3
2
1
isolation
No
Yes
campus_discrimination

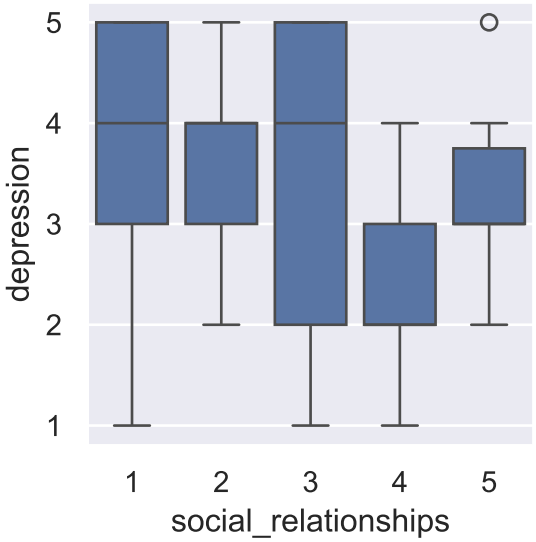

5
4
3
2
1
future_insecurity
No
Yes
campus_discrimination

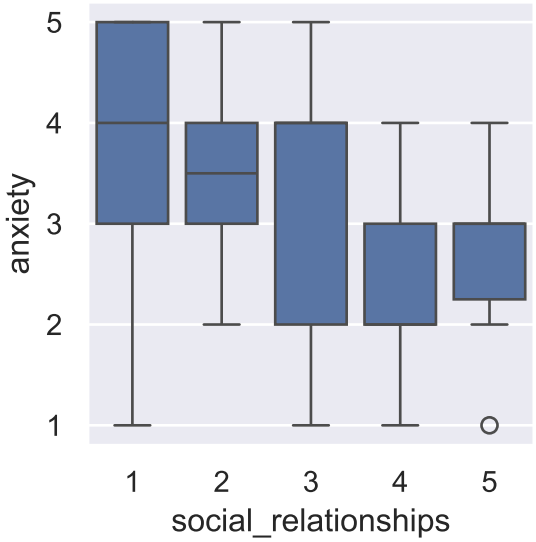

depression
5
4
3
2
1
1
2
3
4
5
academic_pressure

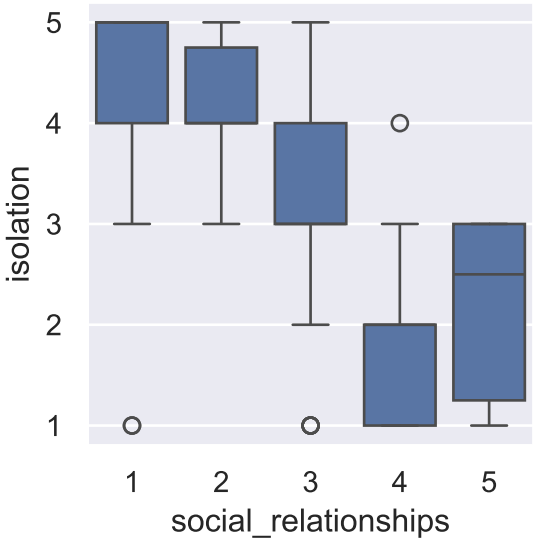

anxiety
5
4
3
2
1
1
2
3
4
5
academic_pressure

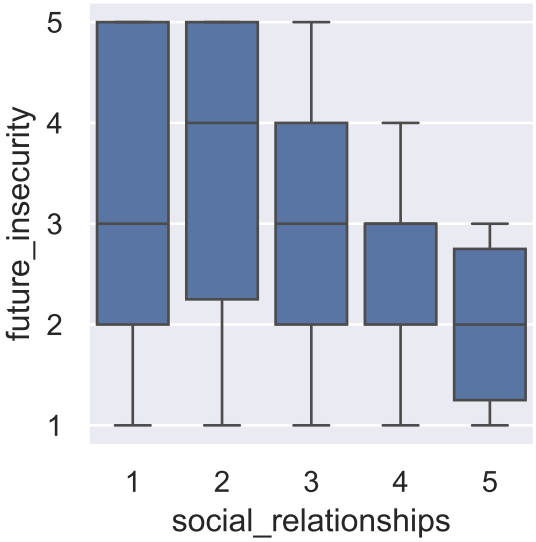

isolation
5
4
3
2
1
1
2
3
4
5
academic_pressure

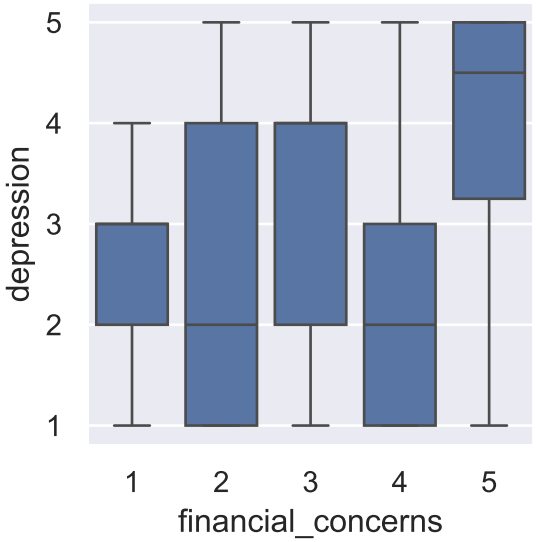

future_insecurity
5
4
3
2
1
1
2
3
4
5
academic_pressure

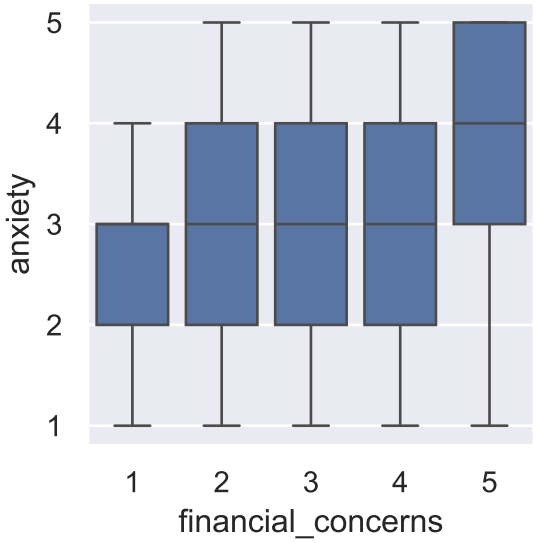

depression
5
4
3
2
1
1
2
3
4
5
social_relationships

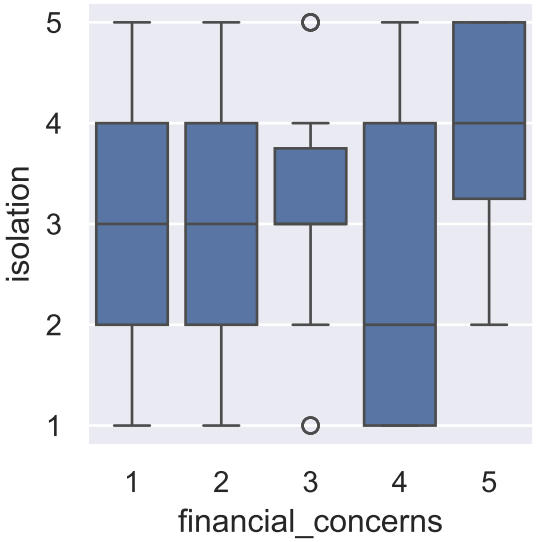

anxiety
5
4
3
2
1
1
2
3
4
5
social_relationships

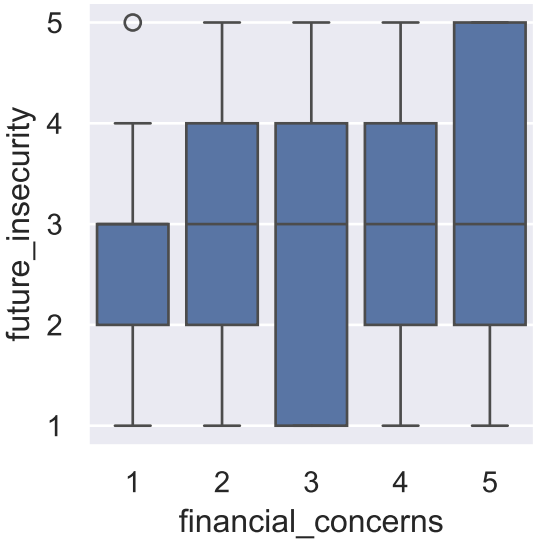

isolation
5
4
3
2
1
1
2
3
4
5
social_relationships

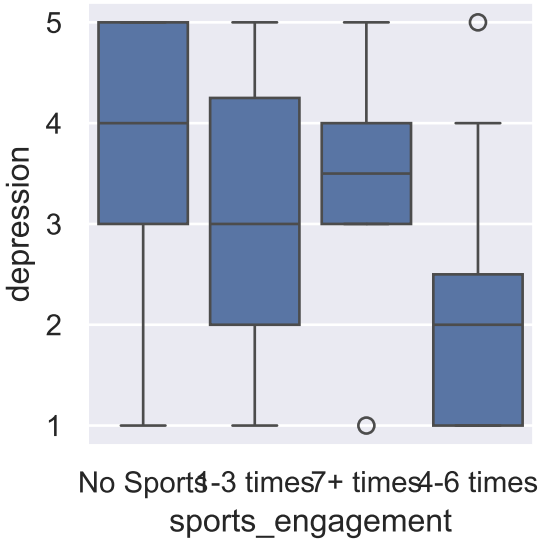

future_insecurity
5
4
3
2
1
1
2
3
4
5
social_relationships

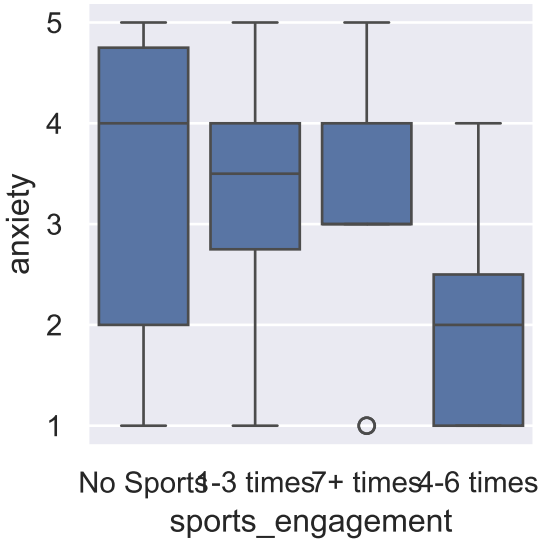

depression
5
4
3
2
1
1
2
3
4
5
financial_concerns

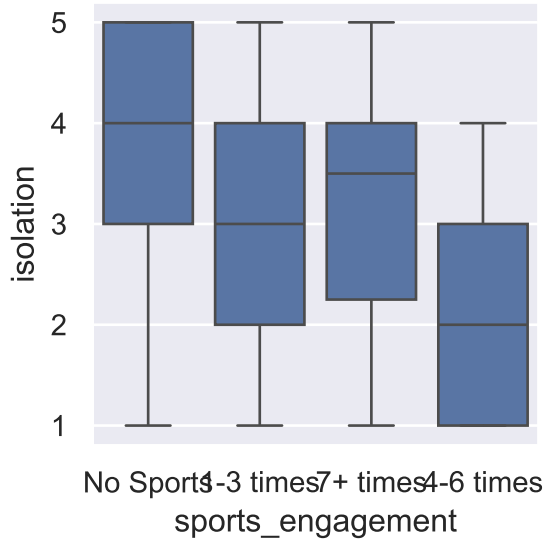

anxiety
5
4
3
2
1
1
2
3
4
5
financial_concerns

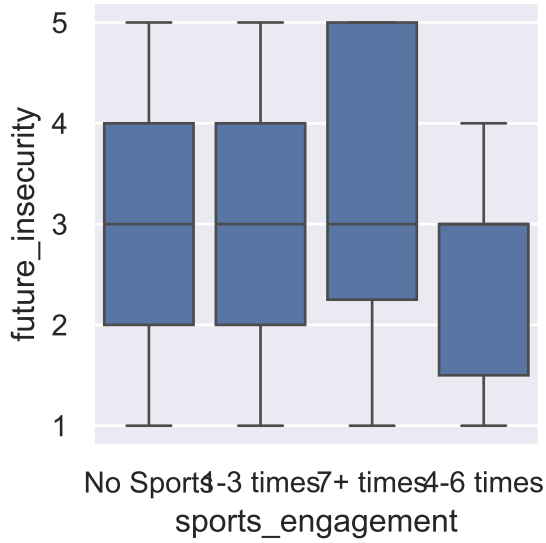

isolation
5
4
3
2
1
1
2
3
4
5
financial_concerns


future_insecurity
5
4
3
2
1
1
2
3
4
5
financial_concerns


depression
5
4
3
2
1
No Sports
1-3 times
7+ times
4-6 times
sports_engagement


anxiety
5
4
3
2
1
No Sports
1-3 times
7+ times
4-6 times
sports_engagement

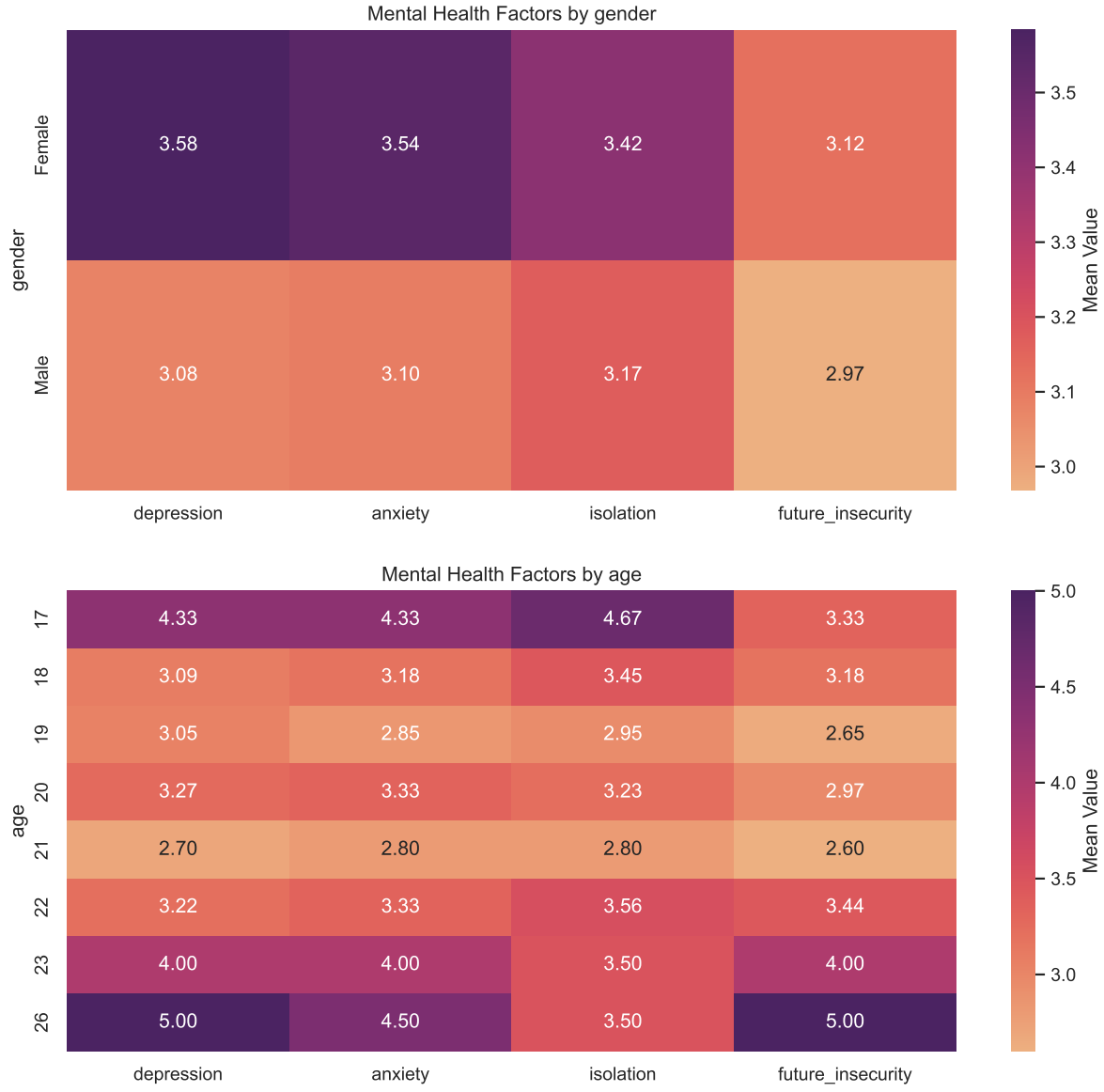


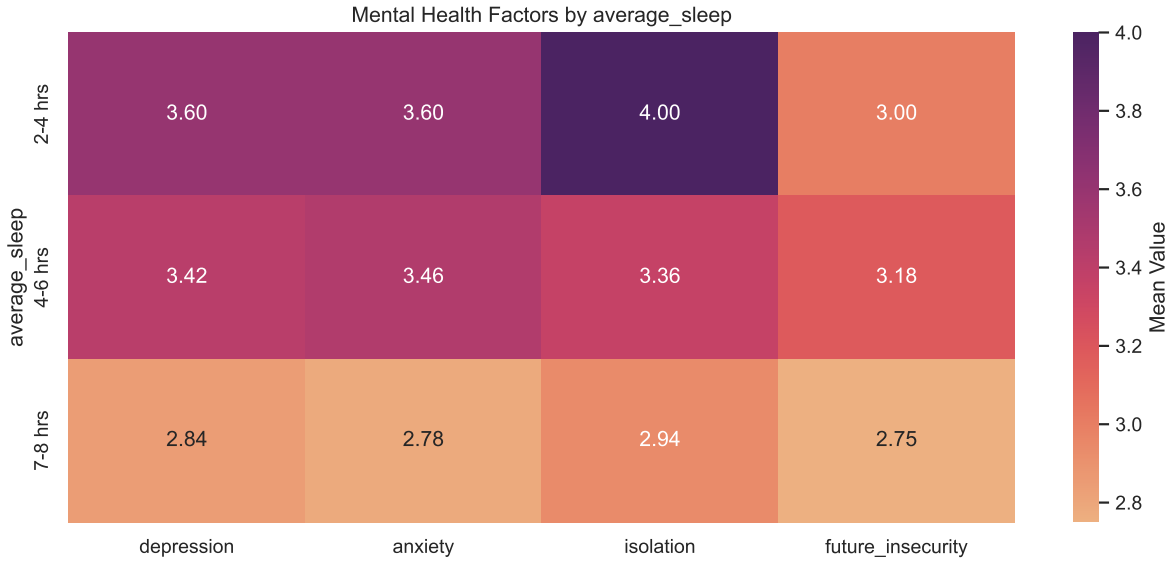



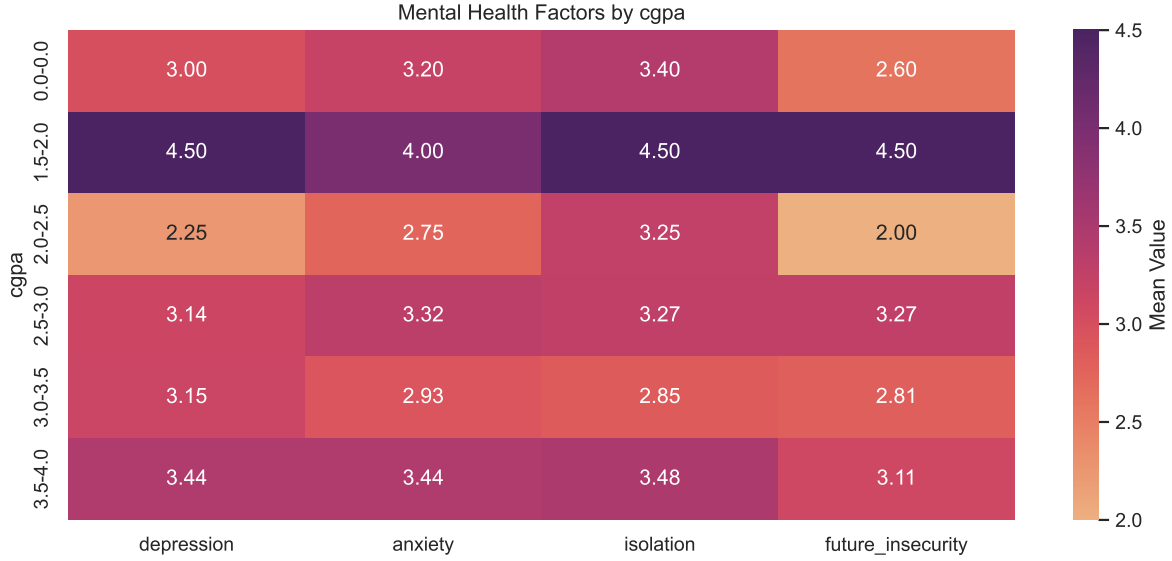

Exploratory Data Analysis Heatmaps

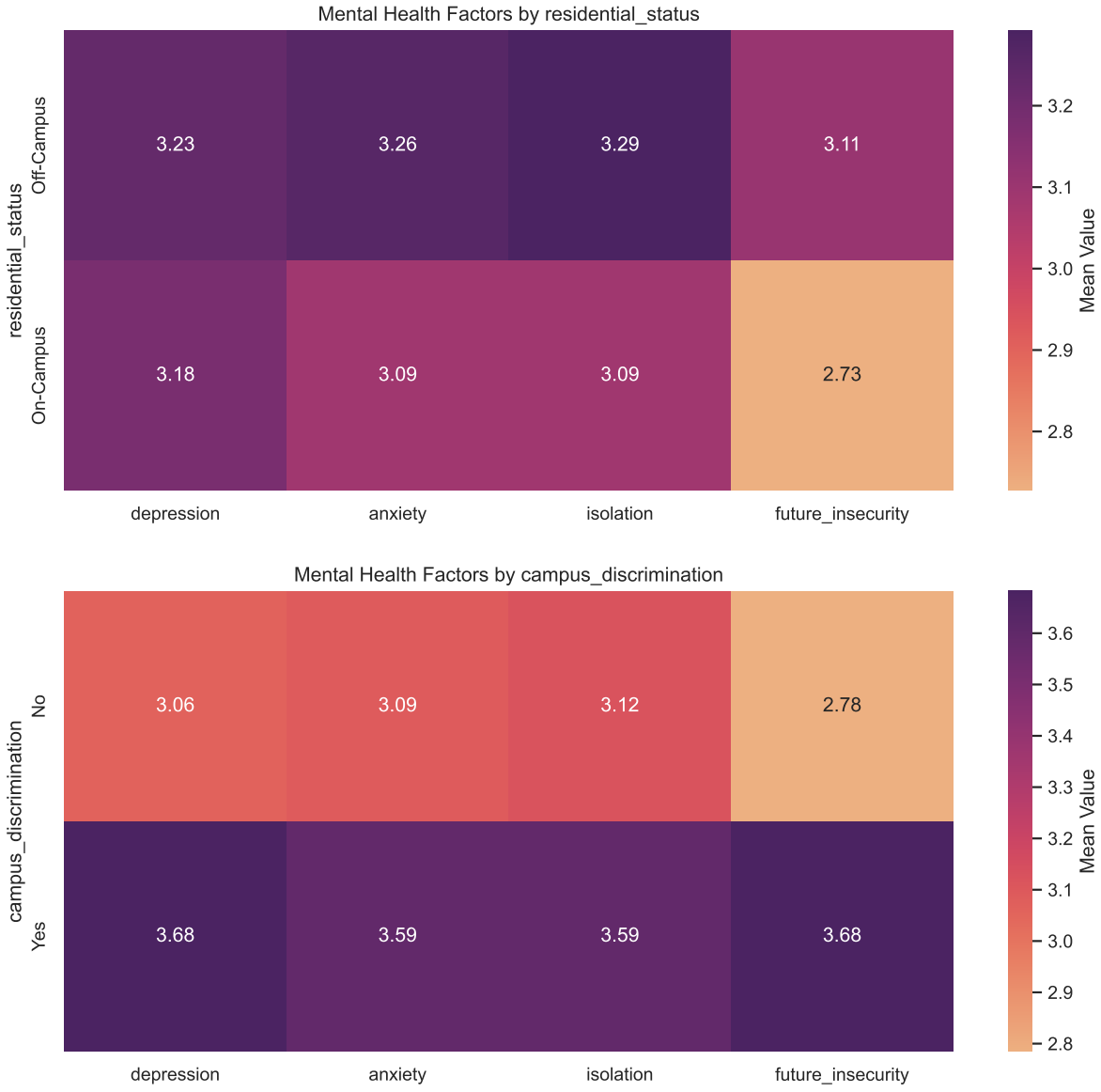
```
for m in x:
  pivot = (dfall.groupby(m)[y_names].mean())

  plt.figure(figsize=(12, 5))

  sns.heatmap(pivot, annot=True, cmap="flare", fmt=".2f", cbar_kws={'label': 'Mean Value'})
  plt.title('Mental Health Factors by ' + m)
```

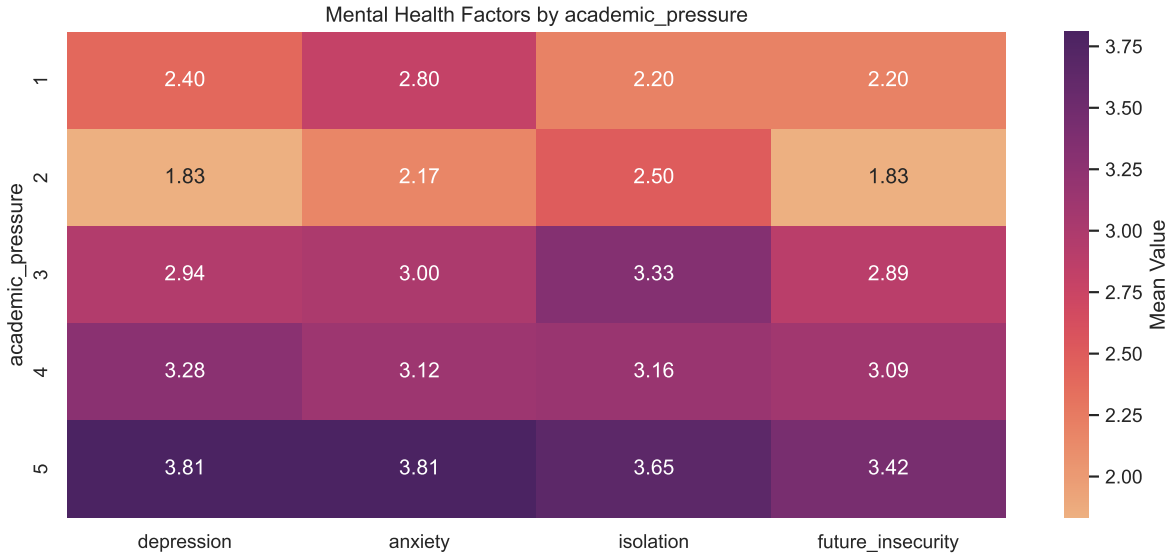
```
plt.ylabel(m)
plt.show()
```

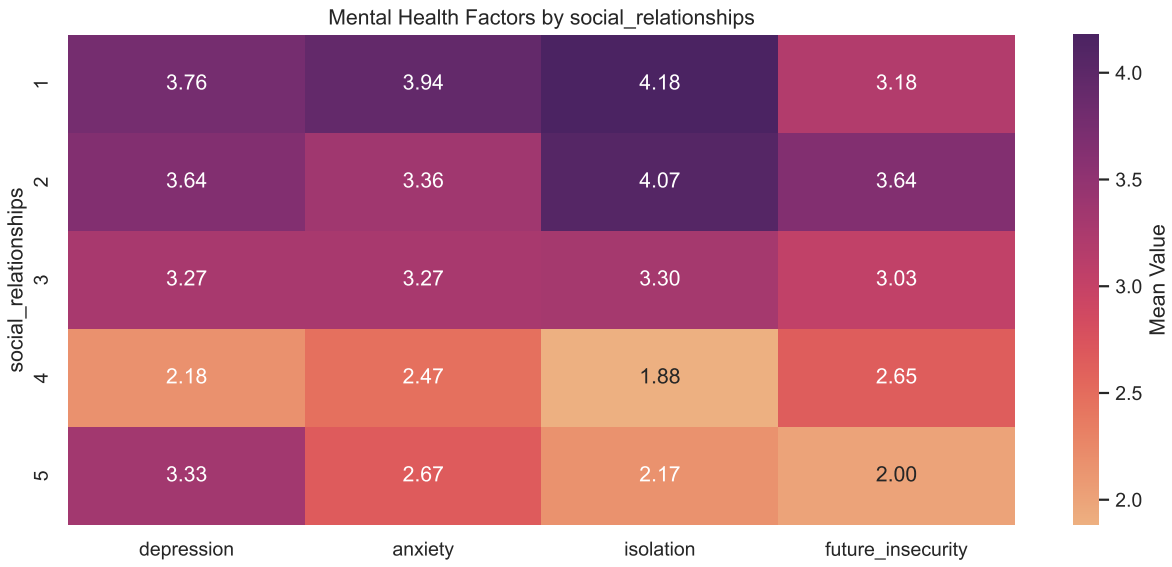


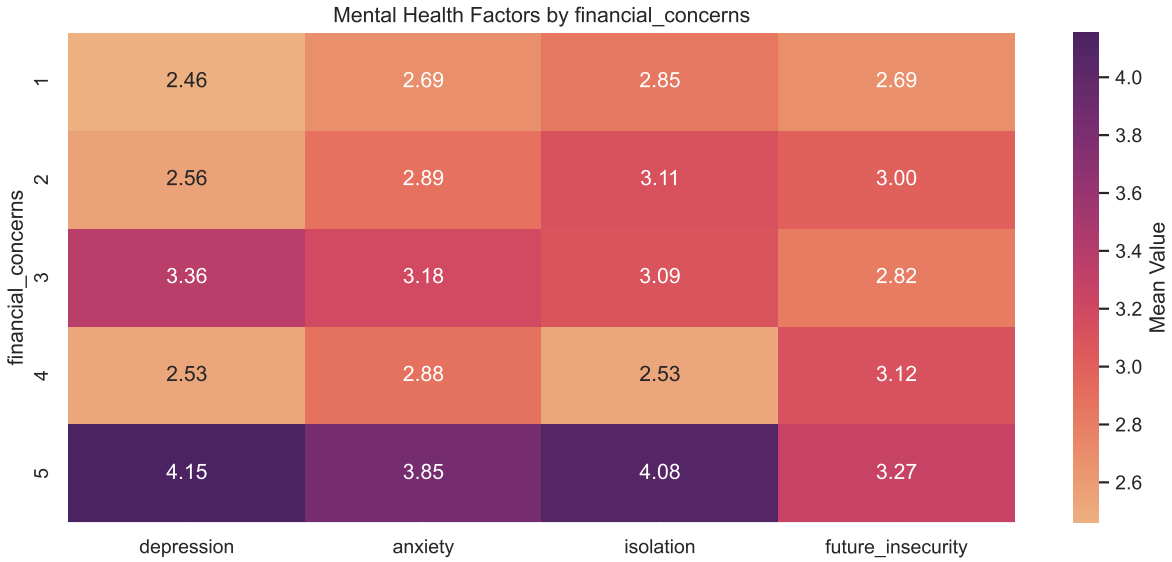

Mental Health Factors by average_sleep
average_sleep
2-4 hrs
4-6 hrs
7-8 hrs
3.60
3.60
4.00
3.00
3.42
3.46
3.36
3.18
2.84
2.78
2.94
2.75
depression
anxiety
isolation
future_insecurity
Mean Value
4.0
3.8
3.6
3.4
3.2
3.0
2.8

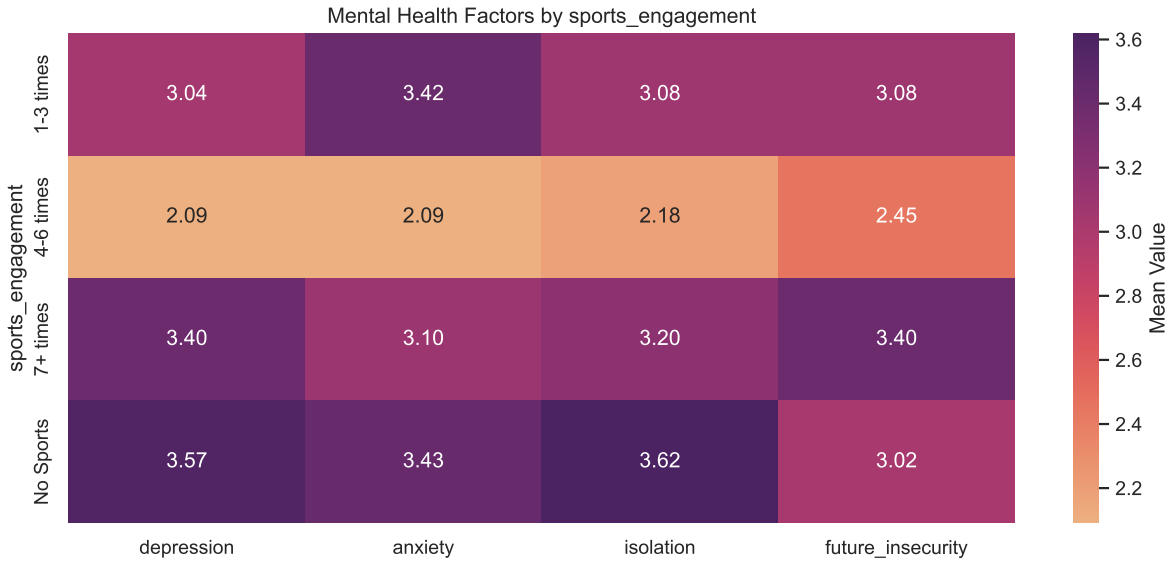

Mental Health Factors by cgpa
cgpa
0.0-0.0
1.5-2.0
2.0-2.5
2.5-3.0
3.0-3.5
3.5-4.0
3.00
3.20
3.40
2.60
4.50
4.00
4.50
4.50
2.25
2.75
3.25
2.00
3.14
3.32
3.27
3.27
3.15
2.93
2.85
2.81
3.44
3.44
3.48
3.11
depression
anxiety
isolation
future_insecurity
Mean Value
4.5
4.0
3.5
3.0
2.5
2.0


Mental Health Factors by residential_status
residential_status
Off-Campus
On-Campus
3.23
3.26
3.29
3.11
3.18
3.09
3.09
2.73
depression
anxiety
isolation
future_insecurity
Mean Value
3.2
3.1
3.0
2.9
2.8
Mental Health Factors by campus_discrimination
campus_discrimination
No
Yes
3.06
3.09
3.12
2.78
3.68
3.59
3.59
3.68
depression
anxiety
isolation
future_insecurity
Mean Value
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8


Mental Health Factors by academic_pressure
academic_pressure
1
2
3
4
5
2.40
2.80
2.20
2.20
1.83
2.17
2.50
1.83
2.94
3.00
3.33
2.89
3.28
3.12
3.16
3.09
3.81
3.81
3.65
3.42
depression
anxiety
isolation
future_insecurity
Mean Value
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00


Mental Health Factors by social_relationships
social_relationships
1
2
3
4
5
3.76
3.94
4.18
3.18
3.64
3.36
4.07
3.64
3.27
3.27
3.30
3.03
2.18
2.47
1.88
2.65
3.33
2.67
2.17
2.00
depression
anxiety
isolation
future_insecurity
Mean Value
4.0
3.5
3.0
2.5
2.0


Mental Health Factors by financial_concerns
financial_concerns
1
2
3
4
5
2.46
2.69
2.85
2.69
2.56
2.89
3.11
3.00
3.36
3.18
3.09
2.82
2.53
2.88
2.53
3.12
4.15
3.85
4.08
3.27
depression
anxiety
isolation
future_insecurity
Mean Value
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6


Mental Health Factors by sports_engagement
sports_engagement
1-3 times
4-6 times
7+ times
No Sports
3.04
3.42
3.08
3.08
2.09
2.09
2.18
2.45
3.40
3.10
3.20
3.40
3.57
3.43
3.62
3.02
depression
anxiety
isolation
future_insecurity
Mean Value
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2