Scripting Project

December 9, 2024

[14]:

```
import pandas as pd

# Loading dataset from csv containing 2020 2021 2022 data
file_path = 'Healthdata_final.csv'
data = pd.read_csv(file_path)


# Drop specified columns
columns_to_drop = [
    'RowId', 'Data_Value_Unit', 'DataValueTypeID', 'Data_Value_Type',␣
 ↪'Data_Value_Alt',
    'Low_Confidence_Limit', 'High_Confidence_Limit', 'Geolocation', 'ClassID',␣
 ↪'TopicID',
    'QuestionID', 'LocationID', 'StratificationCategory1',␣
 ↪'StratificationCategoryID1',
    'StratificationID1', 'StratificationCategoryID2', 'StratificationID2'
]
data = data.drop(columns=columns_to_drop, errors='ignore')
print(data.head(2))

# Rename columns
columns_rename = {
    'YearStart': 'Year_Start',
    'YearEnd': 'Year_End',
    'LocationAbbr': 'Location_Abbr',
    'LocationDesc': 'Location_Desc',
    'Class': 'Survey_Class',
    'Topic': 'Survey_Topic',
    'Question': 'Survey_Question',
    'Data_Value': 'Data_Value',
    'Stratification1': 'Age_Group',
    'StratificationCategory2': 'StratificationCategory2',
    'Stratification2': 'Stratification2'
}
data = data.rename(columns=columns_rename)
```

```python
# Filter 1: Only_agegroup_data
only_agegroup_data = data[data['StratificationCategory2'].isnull()]

# Filter 2: Only_start2_data
only_start2_data = data[(data['StratificationCategory2'].notnull()) & (data['Age_Group'] == 'Overall')]

# Filter 3: agegroup_and_strat2_data
agegroup_and_strat2_data = data[(data['StratificationCategory2'].notnull()) & (data['Age_Group'] != 'Overall')]

# Display the resulting dataframes
print("Main Processed DataFrame:")
print(data.head())

print("\nOnly Age Group Data:")
print(only_agegroup_data.head())

print("\nOnly Stratification Category 2 Data:")
print(only_start2_data.head())

print("\nAge Group and Stratification Category 2 Data:")
print(agegroup_and_strat2_data.head())

# # Optionally save the dataframes to CSV
# data.to_csv("processed_data.csv", index=False)
# only_agegroup_data.to_csv("only_agegroup_data.csv", index=False)
# only_start2_data.to_csv("only_start2_data.csv", index=False)
# agegroup_and_strat2_data.to_csv("agegroup_and_strat2_data.csv", index=False)
```

```
   YearStart  YearEnd LocationAbbr LocationDesc          Class  \
0       2022     2022           MD     Maryland  Mental Health
1       2022     2022           WI    Wisconsin  Mental Health

                    Topic  \
0  Frequent mental distress
1  Frequent mental distress

                                         Question  Data_Value  \
0  Percentage of older adults who are experiencin…         9.0
1  Percentage of older adults who are experiencin…         5.6

     Stratification1 StratificationCategory2      Stratification2
0  65 years or older          Race/Ethnicity  Black, non-Hispanic
1  65 years or older                  Gender                 Male
Main Processed DataFrame:
   Year_Start  Year_End Location_Abbr Location_Desc   Survey_Class  \
0        2022      2022           MD      Maryland  Mental Health
1        2022      2022           WI     Wisconsin  Mental Health
2        2022      2022           OK      Oklahoma  Mental Health
3        2022      2022           PA  Pennsylvania  Mental Health
4        2022      2022           OH          Ohio     Caregiving

                                        Survey_Topic  \
0                            Frequent mental distress
1                            Frequent mental distress
2                            Frequent mental distress
3                            Frequent mental distress
4  Expect to provide care for someone in the next…

                                     Survey_Question  Data_Value  \
0  Percentage of older adults who are experiencin…         9.0
1  Percentage of older adults who are experiencin…         5.6
2  Percentage of older adults who are experiencin…        21.5
3  Percentage of older adults who are experiencin…        10.0
4  Percentage of older adults currently not provi…        14.5

           Age_Group StratificationCategory2          Stratification2
0  65 years or older          Race/Ethnicity       Black, non-Hispanic
1  65 years or older                  Gender                      Male
2            Overall          Race/Ethnicity  Native Am/Alaskan Native
3            Overall          Race/Ethnicity       White, non-Hispanic
4        50-64 years                  Gender                      Male

Only Age Group Data:
    Year_Start  Year_End Location_Abbr Location_Desc            Survey_Class  \
6         2022      2022            NV        Nevada           Overall Health
11        2022      2022           NRE     Northeast  Screenings and Vaccines
19        2022      2022            PR   Puerto Rico  Smoking and Alcohol Use
21        2022      2022            GA       Georgia               Caregiving
51        2022      2022            UT          Utah  Screenings and Vaccines

                                    Survey_Topic  \
6   Self-rated health (good to excellent health)
11        Diabetes screening within past 3 years
19                               Current smoking
21     Duration of caregiving among older adults
51                   Colorectal cancer screening

                                      Survey_Question  Data_Value  \
6   Percentage of older adults who self-reported t…        72.9
11  Percentage of older adults without diabetes wh…        89.3
19  Percentage of older adults who have smoked at …        12.0
21  Percentage of older adults who provided care t…        73.4
51  Percentage of older adults who had either a ho…        67.8
```

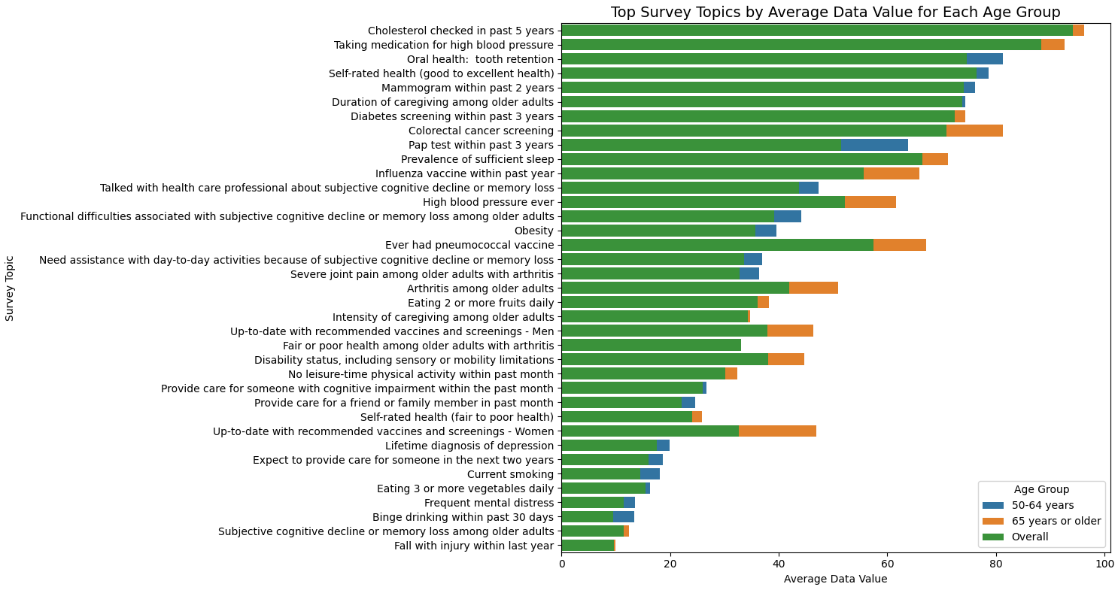
```
      Age_Group StratificationCategory2 Stratification2
6       Overall                     NaN             NaN
11  50-64 years                     NaN             NaN
19  50-64 years                     NaN             NaN
21  50-64 years                     NaN             NaN
51  50-64 years                     NaN             NaN

Only Stratification Category 2 Data:
    Year_Start  Year_End Location_Abbr Location_Desc  \
2         2022      2022            OK      Oklahoma   
3         2022      2022            PA  Pennsylvania   
7         2022      2022            GA       Georgia   
10        2022      2022           SOU         South   
12        2022      2022            PA  Pennsylvania   

                           Survey_Class  \
2                         Mental Health   
3                         Mental Health   
7                        Overall Health   
10  Nutrition/Physical Activity/Obesity   
12                       Overall Health   

                                         Survey_Topic  \
2                            Frequent mental distress   
3                            Frequent mental distress   
7        Self-rated health (good to excellent health)   
10                                            Obesity   
12  Disability status, including sensory or mobili…   

                                      Survey_Question  Data_Value Age_Group  \
2   Percentage of older adults who are experiencin…        21.5   Overall   
3   Percentage of older adults who are experiencin…        10.0   Overall   
7   Percentage of older adults who self-reported t…        70.5   Overall   
10  Percentage of older adults who are currently o…        41.3   Overall   
12  Percentage of older adults who report having a…        39.9   Overall   

   StratificationCategory2           Stratification2
2           Race/Ethnicity  Native Am/Alaskan Native
3           Race/Ethnicity       White, non-Hispanic
7           Race/Ethnicity       Black, non-Hispanic
10          Race/Ethnicity                  Hispanic
12                  Gender                    Female

Age Group and Stratification Category 2 Data:
   Year_Start  Year_End Location_Abbr Location_Desc  \
0        2022      2022            MD      Maryland   
1        2022      2022            WI     Wisconsin   
```

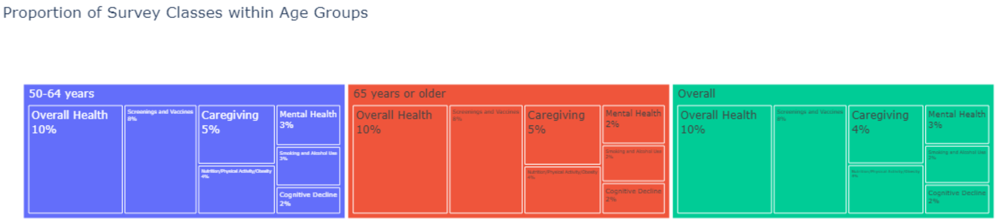
```
4        2022      2022            OH          Ohio
5        2022      2022           SOU         South
8        2022      2022            ID         Idaho

                          Survey_Class  \
0                        Mental Health
1                        Mental Health
4                           Caregiving
5  Nutrition/Physical Activity/Obesity
8                       Overall Health

                                        Survey_Topic  \
0                           Frequent mental distress
1                           Frequent mental distress
4  Expect to provide care for someone in the next…
5                                            Obesity
8                       Arthritis among older adults

                                     Survey_Question  Data_Value  \
0  Percentage of older adults who are experiencin…         9.0
1  Percentage of older adults who are experiencin…         5.6
4  Percentage of older adults currently not provi…        14.5
5  Percentage of older adults who are currently o…        32.7
8  Percentage of older adults ever told they have…        42.7

           Age_Group StratificationCategory2      Stratification2
0  65 years or older          Race/Ethnicity  Black, non-Hispanic
1  65 years or older                  Gender                 Male
4        50-64 years                  Gender                 Male
5  65 years or older          Race/Ethnicity             Hispanic
8  65 years or older          Race/Ethnicity             Hispanic
```

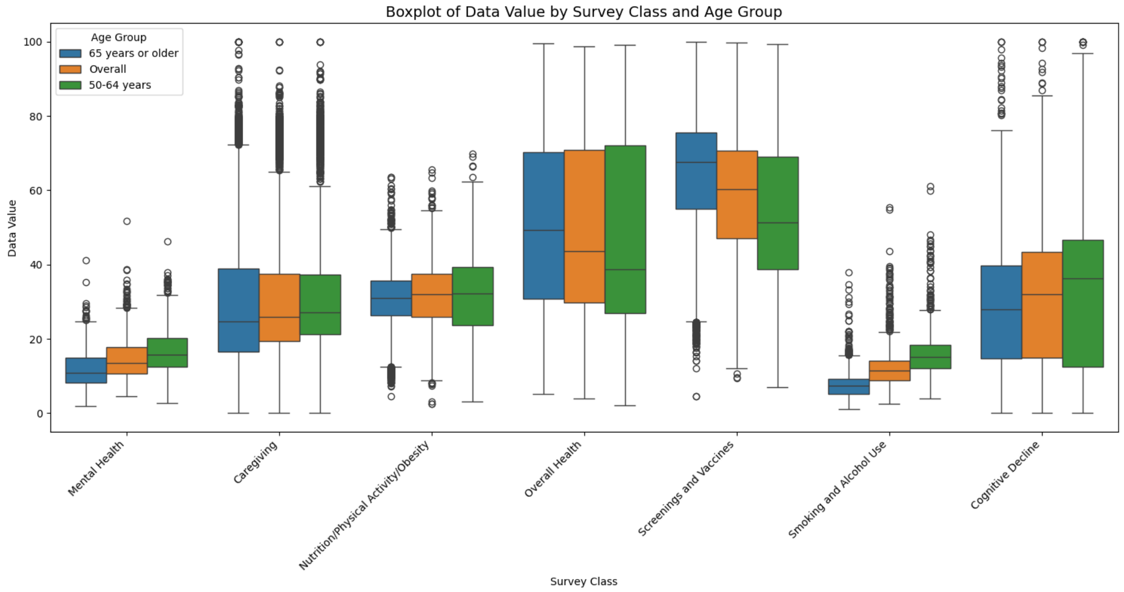
[5]:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Top Survey Topics by Average Data Value for Each Age Group
top_topics = data.groupby(['Age_Group', 'Survey_Topic'])['Data_Value'].mean().
 ↪reset_index()
top_topics = top_topics.sort_values(by=['Age_Group', 'Data_Value'],␣
 ↪ascending=[True, False])

plt.figure(figsize=(15, 8))
sns.barplot(
    data=top_topics,
    x='Data_Value',
    y='Survey_Topic',
    hue='Age_Group',
```

```python
    dodge=False
)
plt.title('Top Survey Topics by Average Data Value for Each Age Group', fontsize=14)
plt.xlabel('Average Data Value')
plt.ylabel('Survey Topic')
plt.legend(title='Age Group')
plt.tight_layout()
plt.show()
#############################################################3

# Boxplot of Data Value by Survey Class and Age Group
plt.figure(figsize=(15, 8))
sns.boxplot(
    data=data,
    x='Survey_Class',
    y='Data_Value',
    hue='Age_Group'
)
plt.title('Boxplot of Data Value by Survey Class and Age Group', fontsize=14)
plt.xlabel('Survey Class')
plt.ylabel('Data Value')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Age Group')
plt.tight_layout()
plt.show
#########################################################333
import plotly.express as px

# Treemap Chart
fig = px.treemap(
    survey_class_proportions,
    path=['Age_Group', 'Survey_Class'],
    values='Proportion',
    color='Age_Group',
    title='Proportion of Survey Classes within Age Groups'
)
fig.update_traces(textinfo="label+percent entry")
fig.show()
```


Top Survey Topics by Average Data Value for Each Age Group
Survey Topic
Cholesterol checked in past 5 years
Taking medication for high blood pressure
Oral health: tooth retention
Self-rated health (good to excellent health)
Mammogram within past 2 years
Duration of caregiving among older adults
Diabetes screening within past 3 years
Colorectal cancer screening
Pap test within past 3 years
Prevalence of sufficient sleep
Influenza vaccine within past year
Talked with health care professional about subjective cognitive decline or memory loss
High blood pressure ever
Functional difficulties associated with subjective cognitive decline or memory loss among older adults
Obesity
Ever had pneumococcal vaccine
Need assistance with day-to-day activities because of subjective cognitive decline or memory loss
Severe joint pain among older adults with arthritis
Arthritis among older adults
Eating 2 or more fruits daily
Intensity of caregiving among older adults
Up-to-date with recommended vaccines and screenings - Men
Fair or poor health among older adults with arthritis
Disability status, including sensory or mobility limitations
No leisure-time physical activity within past month
Provide care for someone with cognitive impairment within the past month
Provide care for a friend or family member in past month
Self-rated health (fair to poor health)
Up-to-date with recommended vaccines and screenings - Women
Lifetime diagnosis of depression
Expect to provide care for someone in the next two years
Current smoking
Eating 3 or more vegetables daily
Frequent mental distress
Binge drinking within past 30 days
Subjective cognitive decline or memory loss among older adults
Fall with injury within last year
0
20
40
60
80
100
Average Data Value
Age Group
50-64 years
65 years or older
Overall

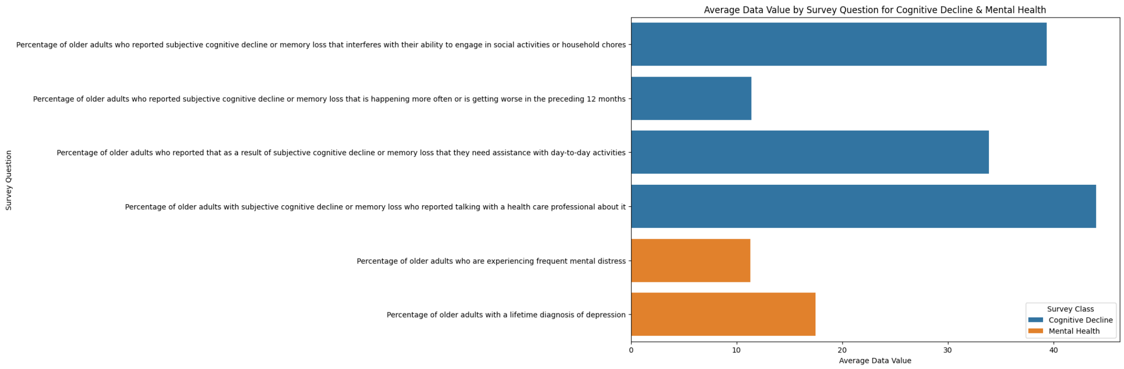

Proportion of Survey Classes within Age Groups
50-64 years
Overall Health
10%
Screenings and Vaccines
8%
Caregiving
5%
Nutrition/Physical Activity/Obesity
4%
Mental Health
3%
Smoking and Alcohol Use
3%
Cognitive Decline
2%
65 years or older
Overall Health
10%
Screenings and Vaccines
8%
Caregiving
5%
Nutrition/Physical Activity/Obesity
4%
Mental Health
2%
Smoking and Alcohol Use
2%
Cognitive Decline
2%
Overall
Overall Health
10%
Screenings and Vaccines
8%
Caregiving
4%
Nutrition/Physical Activity/Obesity
4%
Mental Health
3%
Smoking and Alcohol Use
2%
Cognitive Decline
2%



[13]:

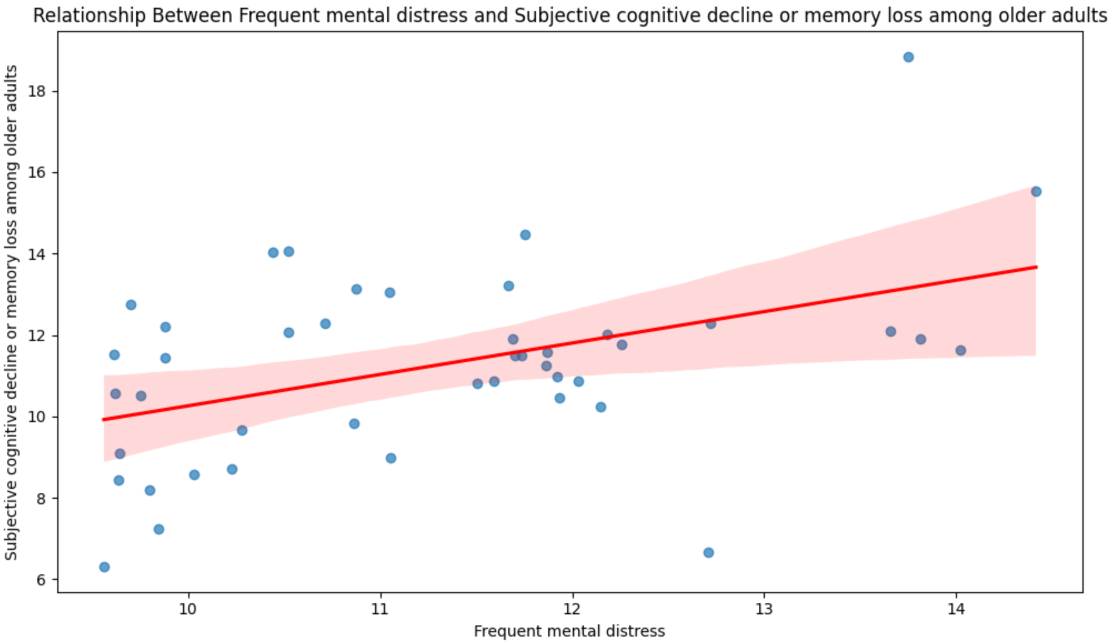
```
# Filter data for Cognitive Decline and Mental Health
filtered_data = data[data['Survey_Class'].isin(['Cognitive Decline', 'Mental␣
  ↪Health'])]
avg_data_values = filtered_data.groupby(['Survey_Class',␣
  ↪'Survey_Question'])['Data_Value'].mean().reset_index()

plt.figure(figsize=(12, 8))
sns.barplot(
    data=avg_data_values,
    x='Data_Value',
    y='Survey_Question',
    hue='Survey_Class'
)
plt.title('Average Data Value by Survey Question for Cognitive Decline & Mental␣
  ↪Health')
plt.xlabel('Average Data Value')
plt.ylabel('Survey Question')
plt.legend(title='Survey Class')
plt.tight_layout()
plt.show()
```

```
/tmp/ipykernel_310/2668823878.py:19: UserWarning:

Tight layout not applied. The left and right margins cannot be made large enough
to accommodate all axes decorations.
```

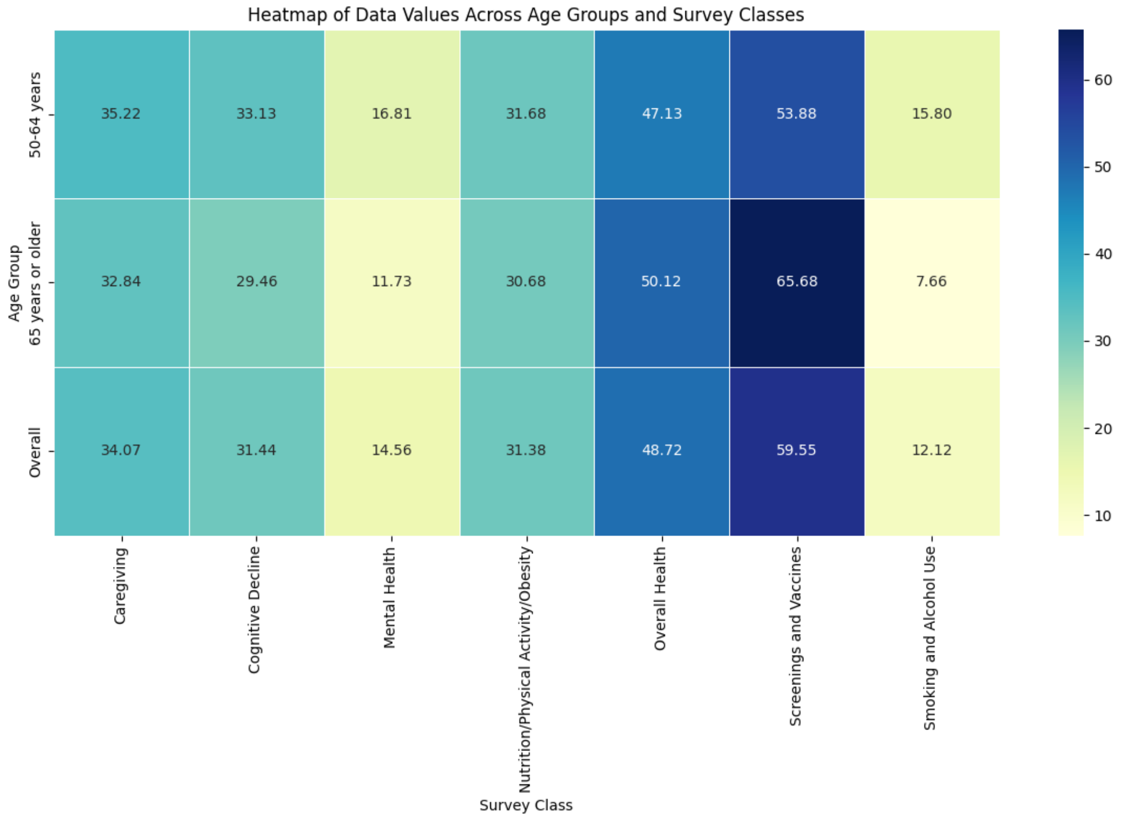


[26]:

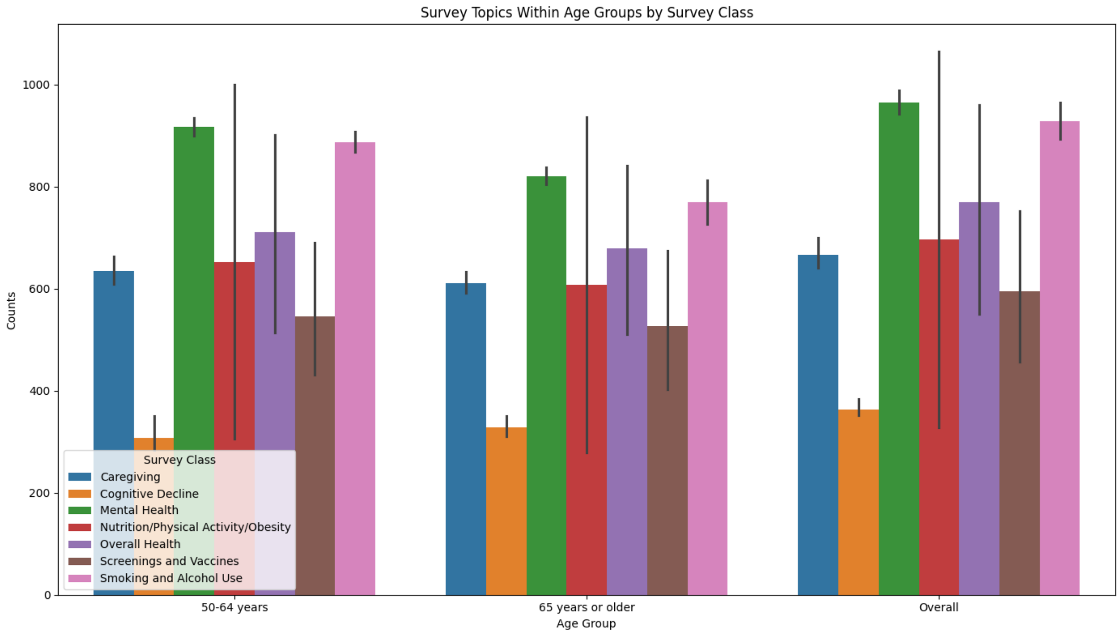
```
# Filter data for Cognitive Decline and Mental Health
related_data = data[data['Survey_Class'].isin(['Cognitive Decline', 'Mental␣
 ↪Health'])]

# Update topic names to match dataset accurately
mental_health_topic = 'Frequent mental distress'
cognitive_decline_topic = 'Subjective cognitive decline or memory loss among␣
 ↪older adults'

# Filter for specific related topics
related_topics_data = related_data[related_data['Survey_Topic'].
 ↪isin([mental_health_topic, cognitive_decline_topic])]

# Pivot data for scatter plot
scatter_data = related_topics_data.pivot_table(
    index='Location_Desc',  # Adjust to the actual column name in your dataset
    columns='Survey_Topic',
    values='Data_Value',
    aggfunc='mean'
).dropna()

# Scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(
    data=scatter_data,
    x=mental_health_topic,
    y=cognitive_decline_topic,
    scatter_kws={'alpha': 0.7},
    line_kws={'color': 'red'}
)
plt.title(f'Relationship Between {mental_health_topic} and␣
 ↪{cognitive_decline_topic}')
plt.xlabel(mental_health_topic)
plt.ylabel(cognitive_decline_topic)
plt.tight_layout()
plt.show()
```



[33]:
```
import seaborn as sns
import matplotlib.pyplot as plt

heatmap_data = data.pivot_table(
    index='Age_Group',
    columns='Survey_Class',
    values='Data_Value',
    aggfunc='mean'
).fillna(0)

plt.figure(figsize=(12, 8))
sns.heatmap(
    heatmap_data,
    annot=True,
    fmt='.2f',
    cmap='YlGnBu',
    linewidths=0.5
)
plt.title('Heatmap of Data Values Across Age Groups and Survey Classes')
plt.xlabel('Survey Class')
plt.ylabel('Age Group')
plt.tight_layout()
plt.show()
```

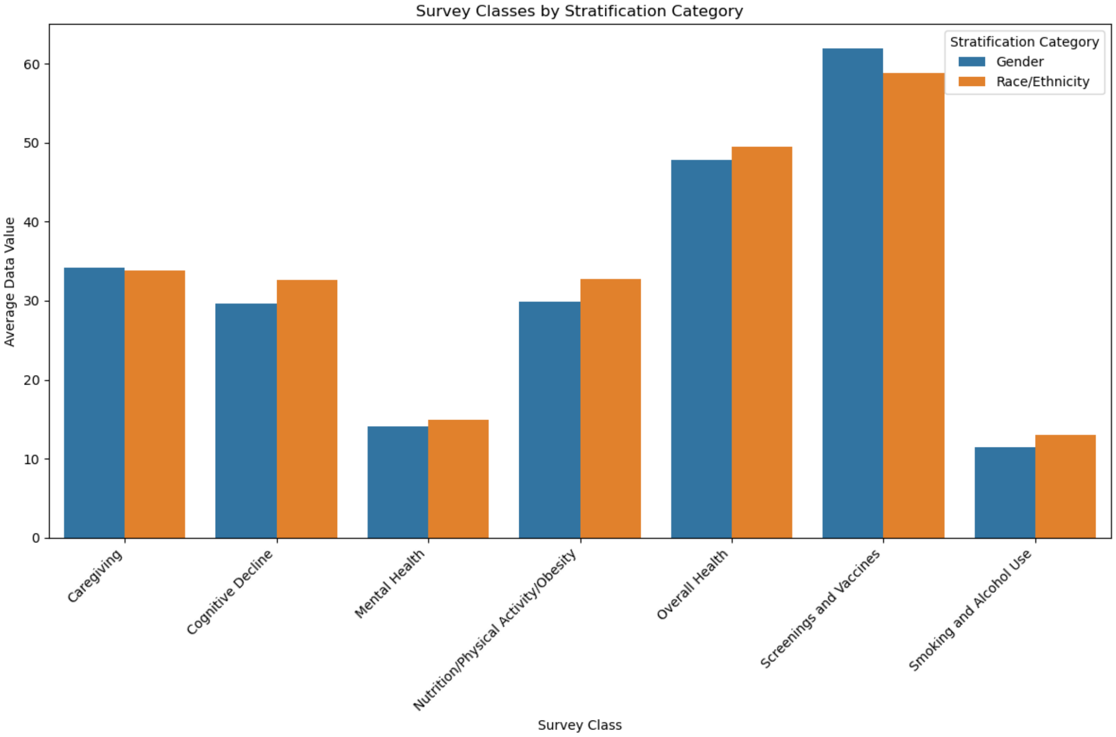


[34]:
```
topic_counts = data.groupby(['Age_Group', 'Survey_Class', 'Survey_Topic']).
 ↪size().reset_index(name='Counts')

plt.figure(figsize=(14, 8))
sns.barplot(
    data=topic_counts,
    x='Age_Group',
    y='Counts',
    hue='Survey_Class'
)
plt.title('Survey Topics Within Age Groups by Survey Class')
plt.xlabel('Age Group')
plt.ylabel('Counts')
plt.legend(title='Survey Class')
plt.tight_layout()
plt.show()
```



[11]:

```
import seaborn as sns
import matplotlib.pyplot as plt

strat_class_data = agegroup_and_strat2_data.groupby(['Survey_Class',␣
 ↪'StratificationCategory2'])['Data_Value'].mean().reset_index()

plt.figure(figsize=(12, 8))
sns.barplot(
    data=strat_class_data,
    x='Survey_Class',
    y='Data_Value',
    hue='StratificationCategory2'
)
plt.title('Survey Classes by Stratification Category')
plt.xlabel('Survey Class')
plt.ylabel('Average Data Value')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Stratification Category')
plt.tight_layout()
plt.show()
```

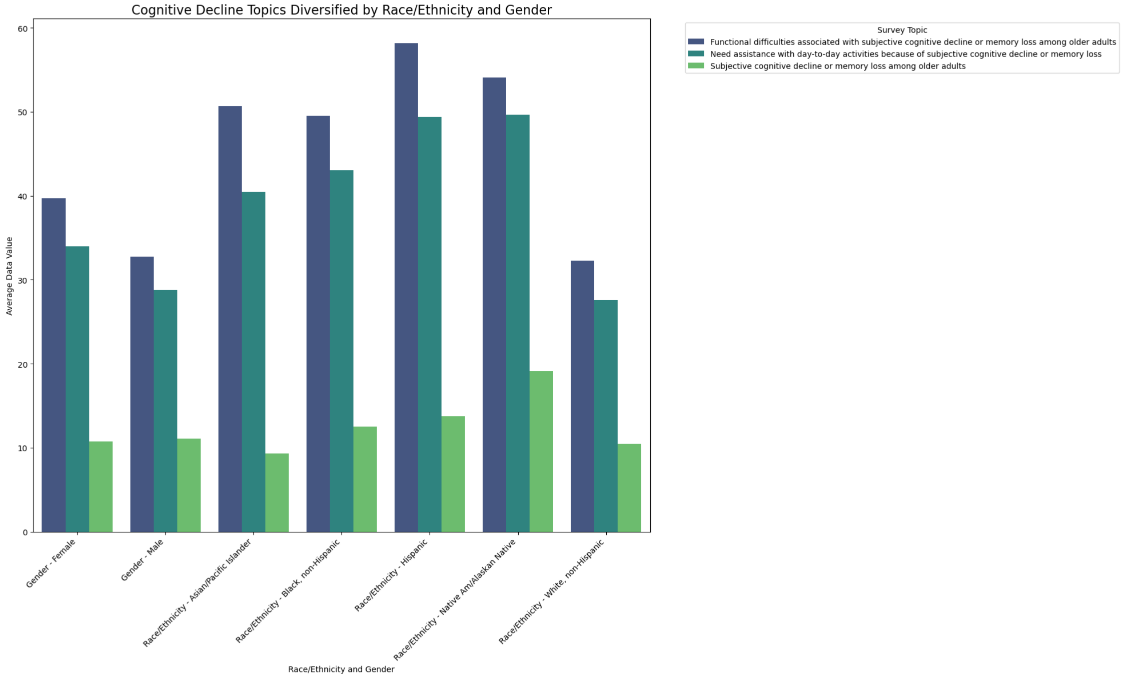


[10]:

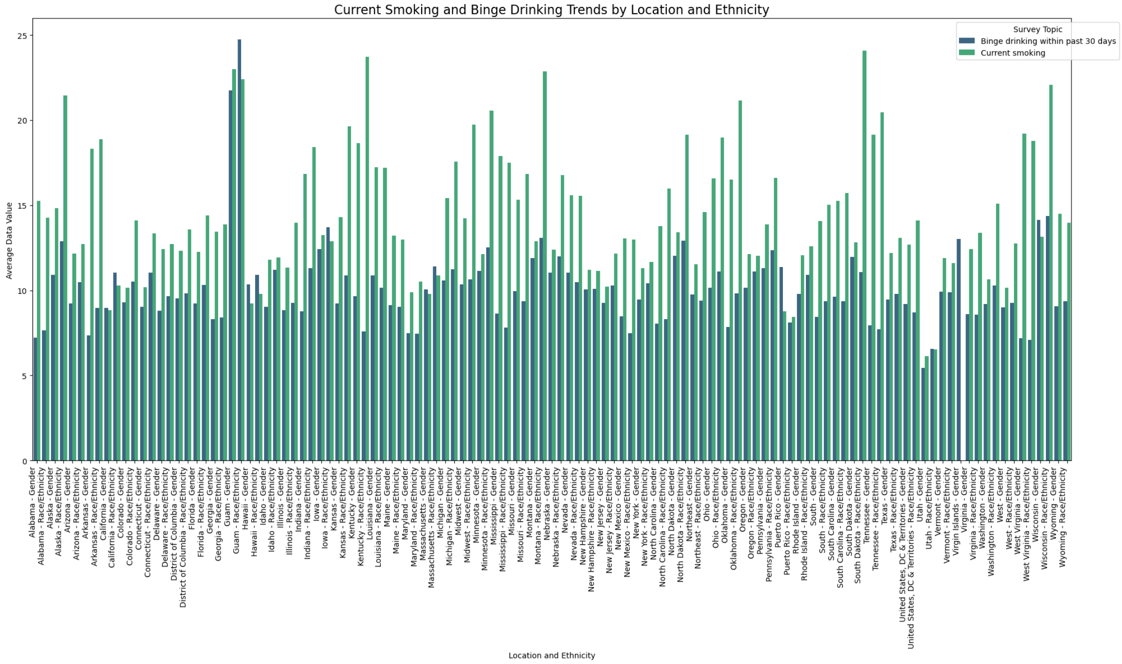
```
# Filter data for Cognitive Decline and the three topics
topics_of_interest = [
    'Functional difficulties associated with subjective cognitive decline or␣
 ↪memory loss among older adults',
    'Need assistance with day-to-day activities because of subjective cognitive␣
 ↪decline or memory loss',
    'Subjective cognitive decline or memory loss among older adults'
]
filtered_data = agegroup_and_strat2_data[
    (agegroup_and_strat2_data['Survey_Class'] == 'Cognitive Decline') &
    (agegroup_and_strat2_data['Survey_Topic'].isin(topics_of_interest))
]

# Group data by Race/Ethnicity and Topic
grouped_data = filtered_data.groupby(['StratificationCategory2',␣
 ↪'Survey_Topic'])['Data_Value'].mean().reset_index()
```

[12]:

```
import matplotlib.pyplot as plt
# Filter data for Cognitive Decline and the three topics
topics_of_interest = [
    'Functional difficulties associated with subjective cognitive decline or␣
 ↪memory loss among older adults',
    'Need assistance with day-to-day activities because of subjective cognitive␣
 ↪decline or memory loss',
    'Subjective cognitive decline or memory loss among older adults'
]
filtered_data = agegroup_and_strat2_data[
    (agegroup_and_strat2_data['Survey_Class'] == 'Cognitive Decline') &
    (agegroup_and_strat2_data['Survey_Topic'].isin(topics_of_interest))
]

# Group data by Race/Ethnicity, Gender (or Stratification2), and Topic
grouped_data = filtered_data.groupby(['StratificationCategory2',␣
 ↪'Stratification2', 'Survey_Topic'])['Data_Value'].mean().reset_index()

# Combine Race/Ethnicity and Gender into a single column for visualization
grouped_data['Race_Gender'] = grouped_data['StratificationCategory2'] + ' - ' +␣
 ↪grouped_data['Stratification2']

# Plot clustered bar chart
plt.figure(figsize=(20, 12))
sns.barplot(
    data=grouped_data,
    x='Race_Gender',
    y='Data_Value',
    hue='Survey_Topic',
    palette='viridis'
)
plt.title('Cognitive Decline Topics Diversified by Race/Ethnicity and Gender',␣
 ↪fontsize=16)
plt.xlabel('Race/Ethnicity and Gender')
plt.ylabel('Average Data Value')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Survey Topic', bbox_to_anchor=(1.05, 1))
plt.tight_layout()
plt.show()
```



[4]:
```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Filter data for Smoking and Alcohol Use and the two topics
topics_of_interest = ['Current smoking', 'Binge drinking within past 30 days']
filtered_data = agegroup_and_strat2_data[
    (agegroup_and_strat2_data['Survey_Class'] == 'Smoking and Alcohol Use') &
    (agegroup_and_strat2_data['Survey_Topic'].isin(topics_of_interest))
]

# Group data by Location, Ethnicity, and Survey Topic
grouped_data = filtered_data.groupby(['Location_Desc',␣
 ↪'StratificationCategory2', 'Survey_Topic'])['Data_Value'].mean().
 ↪reset_index()

# Combine Location and Ethnicity for visualization
grouped_data['Location_Ethnicity'] = grouped_data['Location_Desc'] + ' - ' +␣
 ↪grouped_data['StratificationCategory2']

# Plot clustered bar chart
plt.figure(figsize=(20, 12))
sns.barplot(
    data=grouped_data,
    x='Location_Ethnicity',
    y='Data_Value',
    hue='Survey_Topic',
    palette='viridis'
)
plt.title('Current Smoking and Binge Drinking Trends by Location and␣
  ↪Ethnicity', fontsize=16)
plt.xlabel('Location and Ethnicity')
plt.ylabel('Average Data Value')
plt.xticks(rotation=90, ha='right')
plt.legend(title='Survey Topic', bbox_to_anchor=(1.05, 1))
plt.tight_layout()
plt.show()
```


Current Smoking and Binge Drinking Trends by Location and Ethnicity
Survey Topic
Binge drinking within past 30 days
Current smoking
Average Data Value
Alabama - Gender
Alabama - Race/Ethnicity
Alaska - Gender
Alaska - Race/Ethnicity
Arizona - Gender
Arizona - Race/Ethnicity
Arkansas - Gender
Arkansas - Race/Ethnicity
California - Gender
California - Race/Ethnicity
Colorado - Gender
Colorado - Race/Ethnicity
Connecticut - Gender
Connecticut - Race/Ethnicity
Delaware - Gender
Delaware - Race/Ethnicity
District of Columbia - Gender
District of Columbia - Race/Ethnicity
Florida - Gender
Florida - Race/Ethnicity
Georgia - Gender
Georgia - Race/Ethnicity
Guam - Gender
Guam - Race/Ethnicity
Hawaii - Gender
Hawaii - Race/Ethnicity
Idaho - Gender
Idaho - Race/Ethnicity
Illinois - Gender
Illinois - Race/Ethnicity
Indiana - Gender
Indiana - Race/Ethnicity
Iowa - Gender
Iowa - Race/Ethnicity
Kansas - Gender
Kansas - Race/Ethnicity
Kentucky - Gender
Kentucky - Race/Ethnicity
Louisiana - Gender
Louisiana - Race/Ethnicity
Maine - Gender
Maine - Race/Ethnicity
Maryland - Gender
Maryland - Race/Ethnicity
Massachusetts - Gender
Massachusetts - Race/Ethnicity
Michigan - Gender
Michigan - Race/Ethnicity
Midwest - Gender
Midwest - Race/Ethnicity
Minnesota - Gender
Minnesota - Race/Ethnicity
Mississippi - Gender
Mississippi - Race/Ethnicity
Missouri - Gender
Missouri - Race/Ethnicity
Montana - Gender
Montana - Race/Ethnicity
Nebraska - Gender
Nebraska - Race/Ethnicity
Nevada - Gender
Nevada - Race/Ethnicity
New Hampshire - Gender
New Hampshire - Race/Ethnicity
New Jersey - Gender
New Jersey - Race/Ethnicity
New Mexico - Gender
New Mexico - Race/Ethnicity
New York - Gender
New York - Race/Ethnicity
North Carolina - Gender
North Carolina - Race/Ethnicity
North Dakota - Gender
North Dakota - Race/Ethnicity
Northeast - Gender
Northeast - Race/Ethnicity
Ohio - Gender
Ohio - Race/Ethnicity
Oklahoma - Gender
Oklahoma - Race/Ethnicity
Oregon - Gender
Oregon - Race/Ethnicity
Pennsylvania - Gender
Pennsylvania - Race/Ethnicity
Puerto Rico - Gender
Puerto Rico - Race/Ethnicity
Rhode Island - Gender
Rhode Island - Race/Ethnicity
South - Gender
South - Race/Ethnicity
South Carolina - Gender
South Carolina - Race/Ethnicity
South Dakota - Gender
South Dakota - Race/Ethnicity
Tennessee - Gender
Tennessee - Race/Ethnicity
Texas - Gender
Texas - Race/Ethnicity
United States, DC & Territories - Gender
United States, DC & Territories - Race/Ethnicity
Utah - Gender
Utah - Race/Ethnicity
Vermont - Gender
Vermont - Race/Ethnicity
Virgin Islands - Gender
Virginia - Gender
Virginia - Race/Ethnicity
Washington - Gender
Washington - Race/Ethnicity
West - Gender
West - Race/Ethnicity
West Virginia - Gender
West Virginia - Race/Ethnicity
Wisconsin - Gender
Wisconsin - Race/Ethnicity
Wyoming - Gender
Wyoming - Race/Ethnicity
Location and Ethnicity