Module_1: *Neurodegeneration*

Team Members:

Neil Shroff and Melanie Lam

Project Title:

Alzheimer's Disease: How cerebral pH and beta-amyloid 40/42 concentration correlate with the diagnosis of Alzheimer's Disease

Project Goal:

This project seeks to analyze the relationship between brain pH and amyloid-beta plaque concentration to determine if either are prominent precursors to the development of Alzheimer's Disease.

Hypothesis:

If an individual is diagnosed with Alzheimer's Disease then they will have a lower brain pH and higher levels of amyloid-beta plaque concentration compared to a healthy individual.

Null Hypothesis:

If an individual is diagnosed with Alzheimer's Disease then there will be no correlation between their brain pH and levels of amyloid-beta plaque concentration compared to that of a healthy individual.

Disease Background:

- Prevalence & incidence
    - ~7.2 million Americans have Alzheimer’s Disease
    - ~61.2 million Americans over the age of 65
    - Prevalence = 7.2/61.2×100≈11.76%
    - Incidence = 910,000/(61.2 million) ×100≈1.4%
    - Alzheimer’s Association. (2025). *2025 ALZHEIMER’S DISEASE FACTS AND FIGURES*. https://www.alz.org/getmedia/ef8f48f9-ad36-48ea-87f9-b74034635c1e/alzheimers-facts-and-figures.pdf.

- Economic burden
    - Cost of unpaid caregivers: Give ~18.4 billion hours of care in 2023, equating to 346.6 billion dollars
    - Health care, long-term care, hospice services = $360 billion in 2024
    - 2024 Alzheimer's disease facts and figures. (2024). *Alzheimer's & dementia : the journal of the Alzheimer's Association, 20*(5), 3708–3821. https://doi.org/10.1002/alz.13809.
- Risk factors (genetic, lifestyle)
    - Greatest risk factor is the age of the patient, with older patients having a higher likelihood of developing Alzheimer's
    - Family history of the disease may increase risk of developing Alzheimer's
    - Genetics, both genes and deterministic genes, can increase the likelihood of developing Alzheimer's
    - Head injury increases one's risk of developing Alzheimer's
    - Alzheimer's Association. (n.d.). *What are the Causes and Risk Factors of Alzheimer's and Other Dementias?* https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors?utm_source=google-grant&utm_medium=paidsearch&utm_campaign=google_grant&gad_source=1&gad_
    s0W8bHAiksvVJEzfMx&gclid=Cj0KCQjw267GBhCSARIsAOjVJ4F4vKHAtoWfvK5_Uc0TlFvI
- Societal determinants
    - Lower education level
        - Worsening memory loss is doubled for those without a high school diploma and is halved for those with college degrees
    - Decreased access to healthcare
        - Without consistent access to healthcare, early diagnoses cannot be made so care cannot be provided leading to more severe cases of pathogenesis
    - Social isolation: The lack of human interaction results in a decrease in brain activity and stimulation, leading to a greater risk of loss of cognitive functions
    - CDC Alzheimer's Disease and Dementia. (2024). *Non-Medical Factors that Affect Alzheimer's Disease and Related Dementias Risk*. Centers for Disease Control and Prevention. https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html.
    - Kuiper, J. S., Zuidersma, M., Oude Voshaar, R. C., Zuidema, S. U., van den Heuvel, E. R., Stolk, R. P., & Smidt, N. (2015). *Social relationships and risk of dementia: A systematic review and meta-analysis of longitudinal cohort studies.* Ageing Research Reviews, 22, 39–57. https://doi.org/10.1016/j.arr.2015.04.006
    - African Americans and Hispanic patients have almost a doubled risk of developing Alzheimer's compared to their counterparts likely due to higher prevalency of high blood pressure and diabetes in these demographics
    - Alzheimer's Impact Movement. (2020). *FACTSHEET: Race, Ethnicity, and Alzheimer's.* alzimpact. Alzheimer's Association. Retrieved 2025, from

https://aaic.alz.org/downloads2020/2020_Race_and_Ethnicity_Fact_Sheet.pdf.
    - Women are at a higher risk of developing Alzheimer's due to women living longer than men on average, as well as biological and gender role differences which would increase the risk of dementia
    - Alzheimer's Society. (2025). *Why is dementia different for women?* https://www.alzheimers.org.uk/blog/why-dementia-different-women.
- Symptoms
    - Memory loss is the main symptom of Alzheimer's
    - Trouble concentrating and thinking
    - Making judgement and decisions
    - Behavioral changes
        - Depression, loss of inhibitions, mood swings, etc.
    - Certain skills such as reading, singing, dancing, drawing etc. are preserved
    - Mayo Clinic Staff. (2024, November 8). *Alzheimer's disease*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447.
- Diagnosis
    - Diagnosed through an exam provided by a physician, involving:
        - Impaired thinking
        - Behavioral changes
        - How severe is memory loss
        - How mental impairment impacts daily life
        - Family History
        - Comorbidities
    - Mayo Clinic Staff. (2024a, March 13). *Diagnosing Alzheimer's: How Alzheimer's is diagnosed*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075.
- Standard of care treatments (& reimbursement)
    - Drugs which alleviate symptoms
    - Drugs which "slow disease progression" (donanemab, lecanemab)
        - Both are covered by Medicare and Medicaid with proper enrollment
        - Only treats up to mild dementia
    - Insurance coverage not ensured b/c of newness of treatments; talk to doctor or file appeals
    - Alzheimer's Association. (n.d.-a). *Navigating Treatment Options*. https://www.alz.org/alzheimers-dementia/treatments/navigating-treatment-options.
- Disease progression & prognosis
    - Progressive brain disease = Impairments get worse over time; Stages:
        - Asymptomatic

        - Mild cognitive impairment (MCI)
        - Early stage (mild)
        - Middle stage (moderate)
        - Late stage (severe)
    - No concrete timeline of how the disease progresses across all demographics, though those with preexisting health conditions and a younger age have shown quicker progressions in Alzheimer's
    - Alzheimer's Society. (2013). *The progression, signs and stages of dementia*. https://www.alzheimers.org.uk/about-dementia/stages-and-symptoms/progression-stages-dementia
    - Alzheimer's Association. (n.d.-a). *Navigating Treatment Options*. https://www.alz.org/alzheimers-dementia/treatments/navigating-treatment-options.
- Continuum of care providers
    - Family members and other caregivers contribute to the care of Alzheimer's patients, though the number of them are declining, resulting in more responsibilities for each individual caretaker
    - Healthcare professionals and community-based workforces diagnose, treat, and care for Alzheimer's patients, though they are also experiencing a shortage in staff with the increase of people diagnosed with dementia
    - 2024 Alzheimer's disease facts and figures. (2024). *Alzheimer's & dementia : the journal of the Alzheimer's Association, 20*(5), 3708–3821. https://doi.org/10.1002/alz.13809.
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)
    - Alzheimer's damages the neurons in the brain due to an increase in proteins which clog up neurons, thus disrupting communication between brain cells and negatively impacting human function such as memory, language, and thinking
    - 2024 Alzheimer's disease facts and figures. (2024). *Alzheimer's & dementia : the journal of the Alzheimer's Association, 20*(5), 3708–3821. https://doi.org/10.1002/alz.13809.
- Clinical Trials/next-gen therapies
    - Newest treatments target amyloid-beta protein plaques in the brain to clear up neuron pathways
        - Lecanemab and donanemab
    - More than 100 clinical trials that are available are testing different mechanisms behind the pathogenesis of the disease to find more treatments
    - Landhuis, E., Now Medical Studios, & Christiansen, J. (2025, September 16). *Alzheimer's Drugs Are Finally Tackling the Disease Itself. Here's How.* Scientific American. https://www.scientificamerican.com/article/a-new-generation-of-alzheimers-treatments-explained-in-graphics/.

Data-Set:

- Source: Gabitto, M. I., Travaglini, K. J., Rachleff, V. M., Kaplan, E. S., Long, B., Ariza, J., Ding, Y., Mahoney, J. T., Dee, N., Goldy, J., Melief, E. J., Agrawal, A., Kana, O., Zhen, X., Barlow, S. T., Brouner, K., Campos, J., Campos, J., Carr, A. J., … Lein, E. S. (2024). *Integrated Multimodal Cell Atlas of Alzheimer's disease*. Nature Neuroscience, 27(12), 2366–2383. https://doi.org/10.1038/s41593-024-01774-5.
- How was the data acquired?
  - The participants' cerebral tissue was acquired within 7.0 hours postmortem. This tissue was then able to undergo high-quality snRNA-seq, snATAC-seq, and MERFISH profiling to obtain various biomarkers such as amyloid-beta 42, amyloid-beta 40, pTau, tTau, etc.
- When was the data acquired?
  - The data was acquired postmortem by individuals who participated in the longitudinal study starting in 1994.
- Who acquired the data?
  - The researchers that acquired the data via a longitudinal study were associated with the Kaiser Permanente Washington in partnership with the University of Washington
- What data will be analyzed?
  - Brain pH, beta-amyloid 40 concentration, beta-amyloid 42 concentration, and the cognitive status of the participants will be analyzed in this computational study.

Data Analysis:

The purpose of this code is to use patient data that is contained within two different published .csv files to create a "Patient" class characterized by attributes that describe demographic data (like sex, age at death, Thal Score) and biochemical data (like amyloid-beta levels and tau levels in their brains).

In [12]:
```
# Import the necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import linregress
from scipy import stats
from scipy.stats import ttest_ind
import numpy as np
```

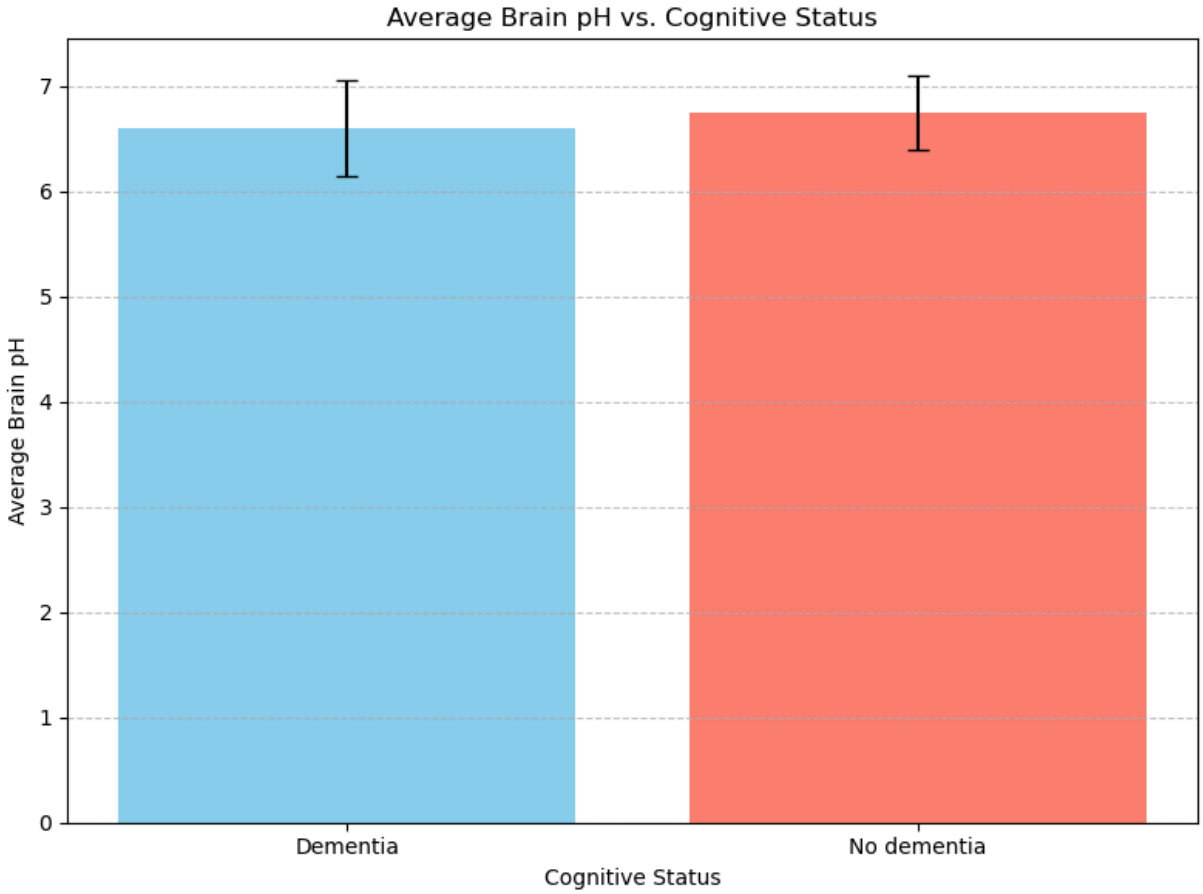
In [13]:
```
# Create a class combining the metadata and the luminex csv files
class DonorData:
    def __init__(self, luminex_path, metadata_path):
        self.luminex_df = pd.read_csv(luminex_path)
        self.metadata_df = pd.read_csv(metadata_path)
        self.merged_df = self._merge_data()

    def _merge_data(self):
        # Merge on 'Donor ID'
        merged = pd.merge(self.luminex_df, self.metadata_df, on='Donor ID', how='in
        return merged

    def get_donor_info(self, donor_id):
        # Retrieve full record for a specific donor
        record = self.merged_df[self.merged_df['Donor ID'] == donor_id]
        return record.to_dict(orient='records') if not record.empty else None

    def get_biomarker_summary(self):
        # Summary statistics for biomarker columns
        biomarker_cols = ['ABeta40 pg/ug', 'ABeta42 pg/ug', 'tTAU pg/ug', 'pTAU pg/
        return self.merged_df[biomarker_cols].describe()

    def filter_by_condition(self, condition_col, condition_value):
        # Filter donors by a specific clinical condition
        filtered = self.merged_df[self.merged_df[condition_col] == condition_value]
        return filtered

    def export_merged_data(self, output_path):
        # Save merged data to a new CSV
        self.merged_df.to_csv(output_path, index=False)
```

In [14]:
```
# Print out the dataset with selected columns, ensuring that the CSV files are down
luminex_df = pd.read_csv('UpdatedLuminex.csv')
metadata_df = pd.read_csv('UpdatedMetaData.csv')
merged_df = pd.merge(luminex_df, metadata_df, on='Donor ID', how='inner')
selected_columns = ['Donor ID', 'ABeta40 pg/ug', 'ABeta42 pg/ug', 'Brain pH', 'Cogn
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
print(merged_df[selected_columns].to_string(index=False))
```

Donor ID	ABeta40 pg/ug	ABeta42 pg/ug	Brain pH	Cognitive Status
H20.33.045	981.444000	142.778000	7.3	Dementia
H20.33.044	0.007088	0.245263	7.6	No dementia
H21.33.045	21.423158	53.878947	7.2	Dementia
H20.33.046	25.295789	69.988421	6.2	Dementia
H20.33.014	0.526168	16.137895	6.8	No dementia
H21.33.046	2.421053	19.195789	6.4	Dementia
H21.33.047	0.000981	0.049053	7.2	No dementia
H20.33.011	11.845263	60.511579	7.0	Dementia
H19.33.004	0.019621	0.971579	7.0	No dementia
H21.33.005	1.655789	6.554737	6.8	Dementia
H21.33.001	0.000883	0.405263	7.2	Dementia
H20.33.024	5.095789	105.104210	7.2	No dementia
H21.33.007	11.418947	287.412000	7.0	Dementia
H20.33.012	2.529474	47.709474	6.6	No dementia
H20.33.025	3.532632	95.792632	7.4	No dementia
H20.33.004	60.766316	80.266316	6.7	Dementia
H20.33.017	52.642105	209.434737	7.2	Dementia
H20.33.013	1.127368	24.781053	7.6	No dementia
H20.33.015	1.944211	27.609474	6.8	Dementia
H20.33.018	196.732000	1412.566961	6.8	Dementia
H20.33.008	3.991579	101.830526	6.4	No dementia
H20.33.005	5.136842	16.156842	6.8	No dementia
H20.33.026	31.565263	63.374737	7.2	Dementia
H20.33.041	5.522105	242.586316	6.8	Dementia
H20.33.027	1.843158	29.955789	6.7	No dementia
H21.33.008	18.994000	18.994000	6.7	Dementia
H20.33.019	1.718947	28.813684	7.0	No dementia
H20.33.001	0.215789	2.744211	6.8	No dementia
H20.33.002	0.000598	0.147158	7.3	No dementia
H20.33.028	1.127368	18.947368	6.6	Dementia
H21.33.006	12.876842	82.972632	6.8	No dementia
H20.33.029	1.633684	28.854737	6.5	Dementia
H20.33.030	17.221053	58.266316	6.7	No dementia
H20.33.031	2.004211	42.513684	6.4	Dementia
H20.33.020	145.254737	45.728421	6.8	Dementia
H20.33.032	91.748421	44.256842	6.8	No dementia
H20.33.033	20.211579	123.368421	6.8	Dementia
H20.33.034	4.794737	4.960000	6.8	No dementia
H20.33.035	0.030147	0.525263	6.6	No dementia
H20.33.036	3.594737	102.455789	6.4	No dementia
H20.33.037	53.012632	67.654737	6.4	Dementia
H20.33.043	97.800000	60.953684	7.0	No dementia
H20.33.016	2.671579	21.273684	6.8	Dementia
H21.33.012	0.215789	3.502105	6.4	Dementia
H21.33.011	0.000688	0.137347	6.8	No dementia
H20.33.038	5.176842	81.137895	6.0	Dementia
H21.33.010	66.775789	24.637895	6.7	Dementia
H20.33.039	2.062105	27.334737	6.8	No dementia
H21.33.009	189.290526	40.198947	6.4	Dementia
H21.33.017	1.412632	20.182105	6.4	Dementia
H21.33.016	0.009427	0.525263	6.4	Dementia
H21.33.015	0.661053	10.095789	6.8	No dementia
H20.33.040	1.412632	12.697895	6.4	Dementia
H21.33.002	93.676842	74.776842	6.8	Dementia
H21.33.003	0.000805	0.405263	6.4	No dementia

H21.33.014	4.864211	35.275789	6.8	No dementia
H21.33.013	43.233684	68.366316	6.8	Dementia
H21.33.021	0.001261	2.672632	6.9	Dementia
H21.33.020	1.547368	38.158947	6.2	Dementia
H21.33.004	0.001155	0.670526	6.4	No dementia
H21.33.022	0.000130	7.666316	6.8	No dementia
H21.33.019	0.001078	0.019621	6.4	No dementia
H21.33.018	0.263158	10.988421	6.4	Dementia
H21.33.023	0.000598	0.114737	6.6	No dementia
H21.33.025	21.209474	8.842105	6.4	No dementia
H21.33.044	7.130526	33.637895	6.6	Dementia
H21.33.026	76.917895	263.536842	6.6	No dementia
H21.33.027	31.711579	39.837895	6.4	Dementia
H21.33.028	0.072507	0.204386	6.4	No dementia
H21.33.029	17.827368	146.862105	6.4	Dementia
H21.33.030	1.827368	18.547368	6.7	No dementia
H21.33.031	34.429474	124.434737	6.4	Dementia
H21.33.032	1.375789	6.777895	6.4	No dementia
H21.33.033	3.967368	31.763158	6.0	No dementia
H21.33.034	12.153684	161.094737	4.5	Dementia
H21.33.035	7.491579	143.464211	6.4	No dementia
H21.33.036	5.302105	75.789474	6.8	No dementia
H21.33.037	0.036217	0.449649	7.0	No dementia
H21.33.038	0.079679	0.122632	6.2	No dementia
H21.33.039	1.450526	76.226316	6.5	Dementia
H21.33.040	0.065192	0.490526	6.7	No dementia
H21.33.041	5.010526	88.169474	6.4	No dementia
H21.33.042	20.538947	47.932632	6.4	Dementia
H21.33.043	1.593684	29.892632	6.6	Dementia

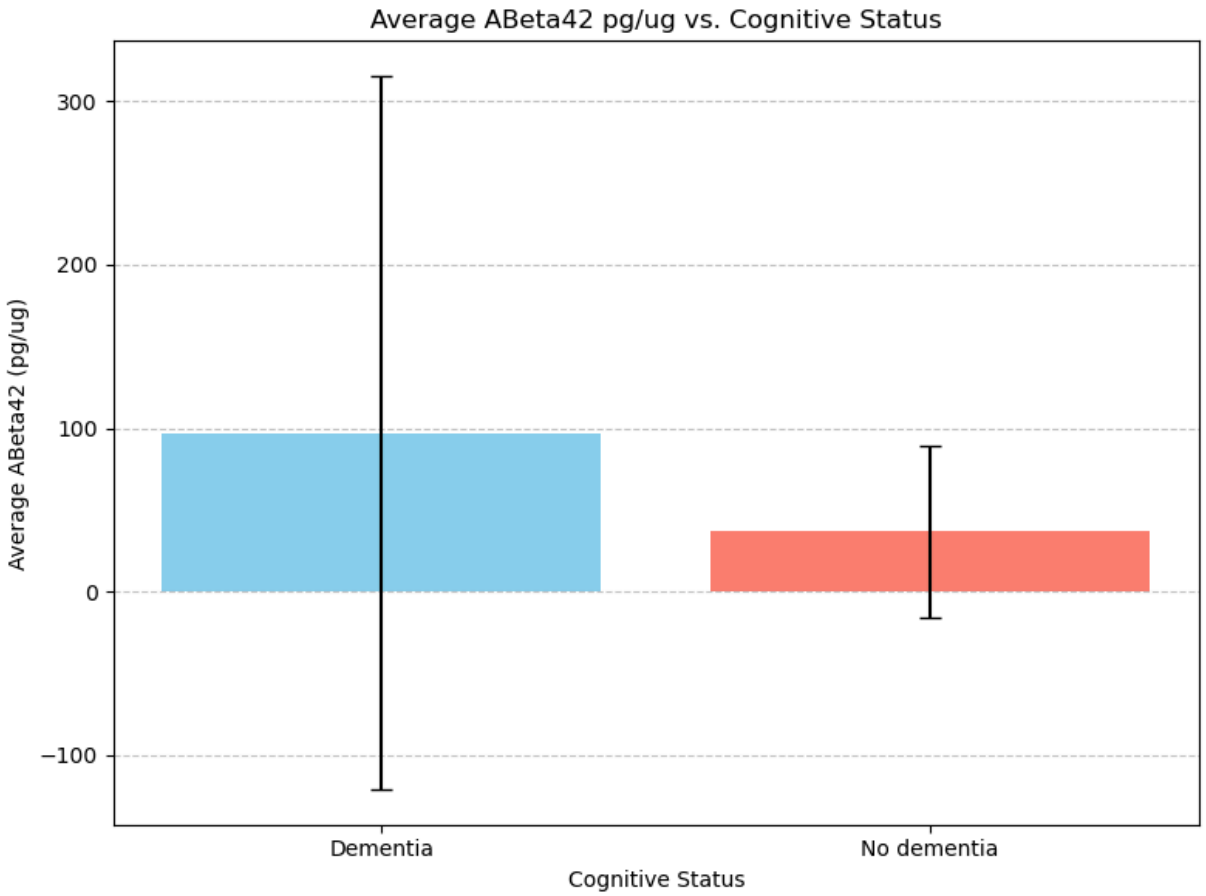
In [15]:
```
#Create a Bar graph comparing Cognitive Status to Average Brain pH
group_stats = merged_df.groupby('Cognitive Status')['Brain pH'].agg(['mean', 'std']
plt.figure(figsize=(8, 6))
plt.bar(group_stats.index, group_stats['mean'], yerr=group_stats['std'], capsize=5,
plt.ylabel('Average Brain pH')
plt.xlabel('Cognitive Status')
plt.title('Average Brain pH vs. Cognitive Status')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



In [16]:
```
#Statistically analyze the bar graph's data using a t-test
group1 = merged_df[merged_df['Cognitive Status'] == 'No dementia']['Brain pH'].drop
group2 = merged_df[merged_df['Cognitive Status'] == 'Dementia']['Brain pH'].dropna(
t_stat, p_value = ttest_ind(group1, group2, equal_var=False)
print(f"T-statistic: {t_stat:.4f}")
print(f"P-value: {p_value:.4f}")
alpha = 0.05
if p_value < alpha:
    print("There is a statistically significant difference in Brain pH between the
else:
    print("There is no statistically significant difference in Brain pH between the
```

```
T-statistic: 1.6462
P-value: 0.1038
There is no statistically significant difference in Brain pH between the groups (p ≥
0.05).
```

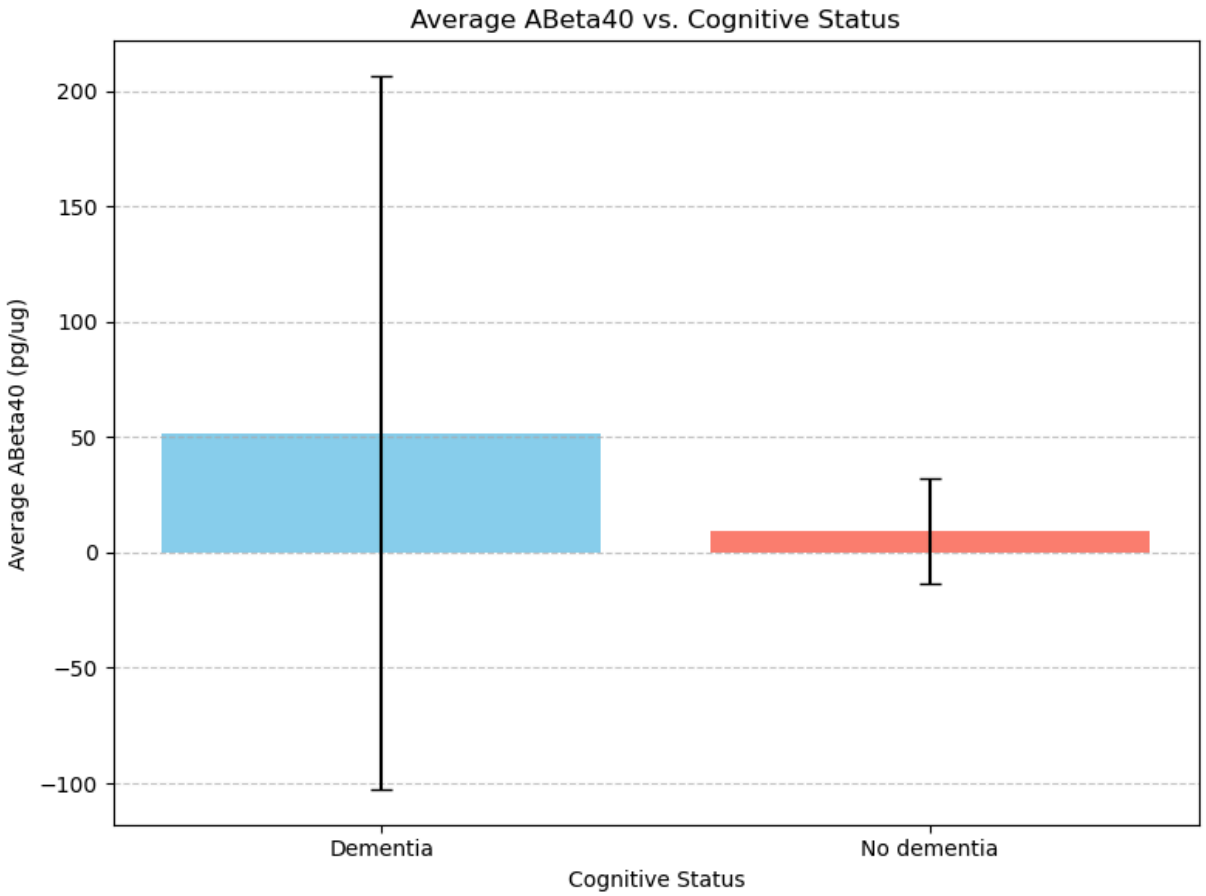
In [17]:
```
#Create a Bar graph comparing Cognitive Status to Average ABeta42 pg/ug
group_stats = merged_df.groupby('Cognitive Status')['ABeta42 pg/ug'].agg(['mean', '
plt.figure(figsize=(8, 6))
plt.bar(group_stats.index, group_stats['mean'], yerr=group_stats['std'], capsize=5,
plt.ylabel('Average ABeta42 (pg/ug)')
plt.xlabel('Cognitive Status')
plt.title('Average ABeta42 pg/ug vs. Cognitive Status')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



In [18]:
```
#Statistically analyze the bar graph's data using a t-test
group1 = merged_df[merged_df['Cognitive Status'] == 'No dementia']['ABeta42 pg/ug']
group2 = merged_df[merged_df['Cognitive Status'] == 'Dementia']['ABeta42 pg/ug'].dr
t_stat, p_value = ttest_ind(group1, group2, equal_var=False)
print(f"T-statistic: {t_stat:.4f}")
print(f"P-value: {p_value:.4f}")
alpha = 0.05
if p_value < alpha:
    print("There is a statistically significant difference in Amyloid-Beta 42 betwe
else:
    print("There is no statistically significant difference in Amyloid-Beta 42 betw
```

```
T-statistic: -1.7457
P-value: 0.0876
There is no statistically significant difference in Amyloid-Beta 42 between the grou
ps (p ≥ 0.05).
```

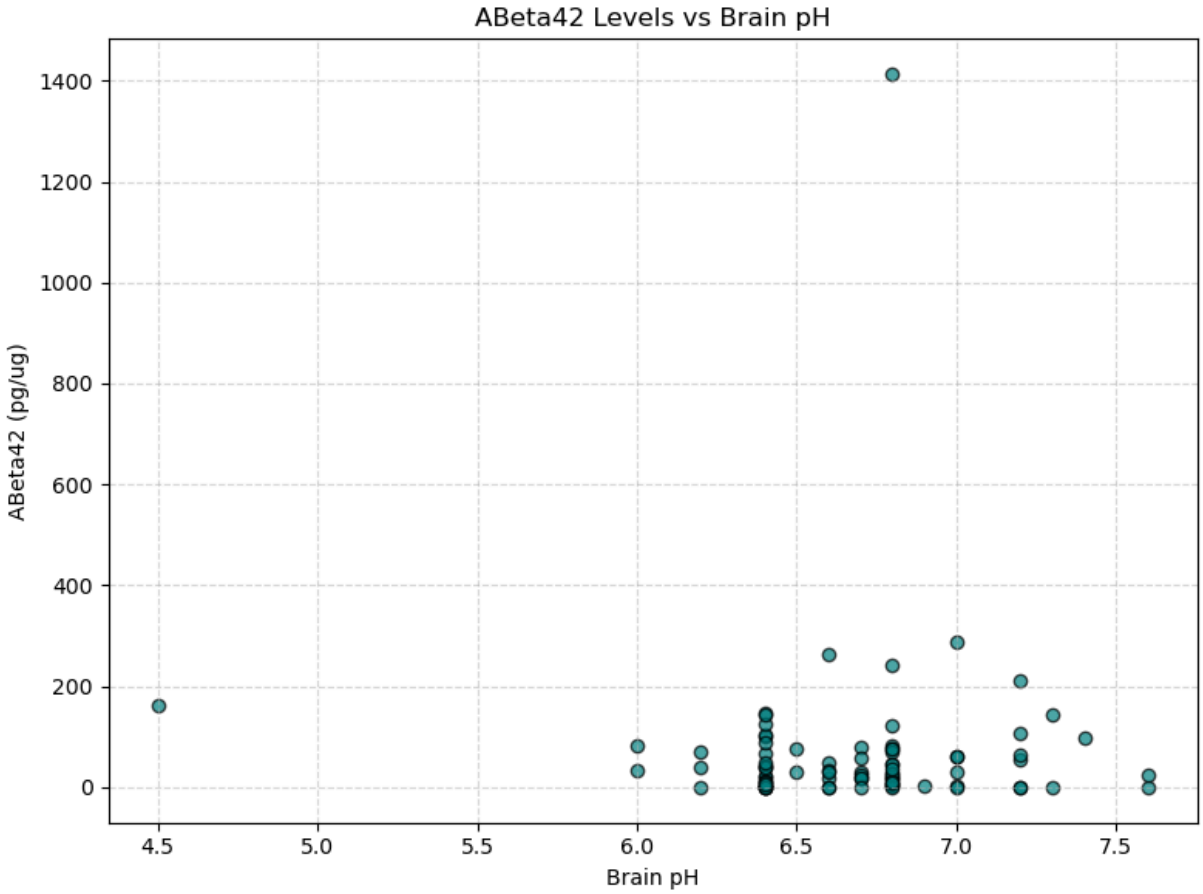
In [21]:
```
#Create a Bar graph comparing Cognitive Status to Average ABeta40 pg/ug
group_stats = merged_df.groupby('Cognitive Status')['ABeta40 pg/ug'].agg(['mean', '
plt.figure(figsize=(8, 6))
plt.bar(group_stats.index, group_stats['mean'], yerr=group_stats['std'], capsize=5,
plt.ylabel('Average ABeta40 (pg/ug)')
plt.xlabel('Cognitive Status')
plt.title('Average ABeta40 vs. Cognitive Status')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



In [20]:
```
#Statistically analyze the bar graph's data using a t-test
group1 = merged_df[merged_df['Cognitive Status'] == 'No dementia']['ABeta40 pg/ug']
group2 = merged_df[merged_df['Cognitive Status'] == 'Dementia']['ABeta40 pg/ug'].dr
t_stat, p_value = ttest_ind(group1, group2, equal_var=False)
print(f"T-statistic: {t_stat:.4f}")
print(f"P-value: {p_value:.4f}")
alpha = 0.05
if p_value < alpha:
    print("There is a statistically significant difference in Amyloid-Beta 40 betwe
else:
    print("There is no statistically significant difference in Amyloid-Beta 40 betw
```

```
T-statistic: -1.7749
P-value: 0.0830
There is no statistically significant difference in Amyloid-Beta 40 between the grou
ps (p ≥ 0.05).
```

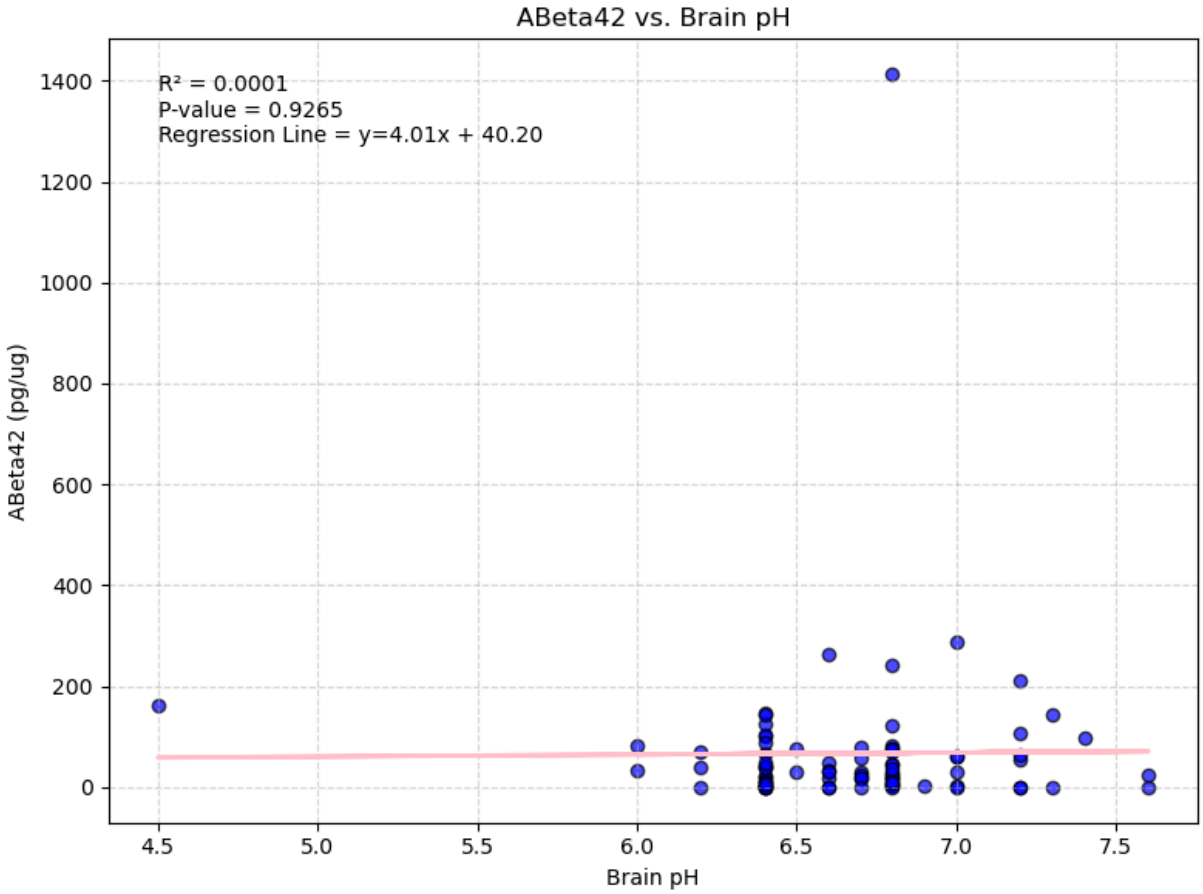
In [22]:
```
#Create a scatter plot comparing ABeta42 pg/ug to Brain pH for all participants
plot_df = merged_df[['Brain pH', 'ABeta42 pg/ug']].dropna()
plt.figure(figsize=(8, 6))
plt.scatter(plot_df['Brain pH'], plot_df['ABeta42 pg/ug'], color='teal', alpha=0.7,
plt.xlabel('Brain pH')
plt.ylabel('ABeta42 (pg/ug)')
plt.title('ABeta42 Levels vs Brain pH')
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



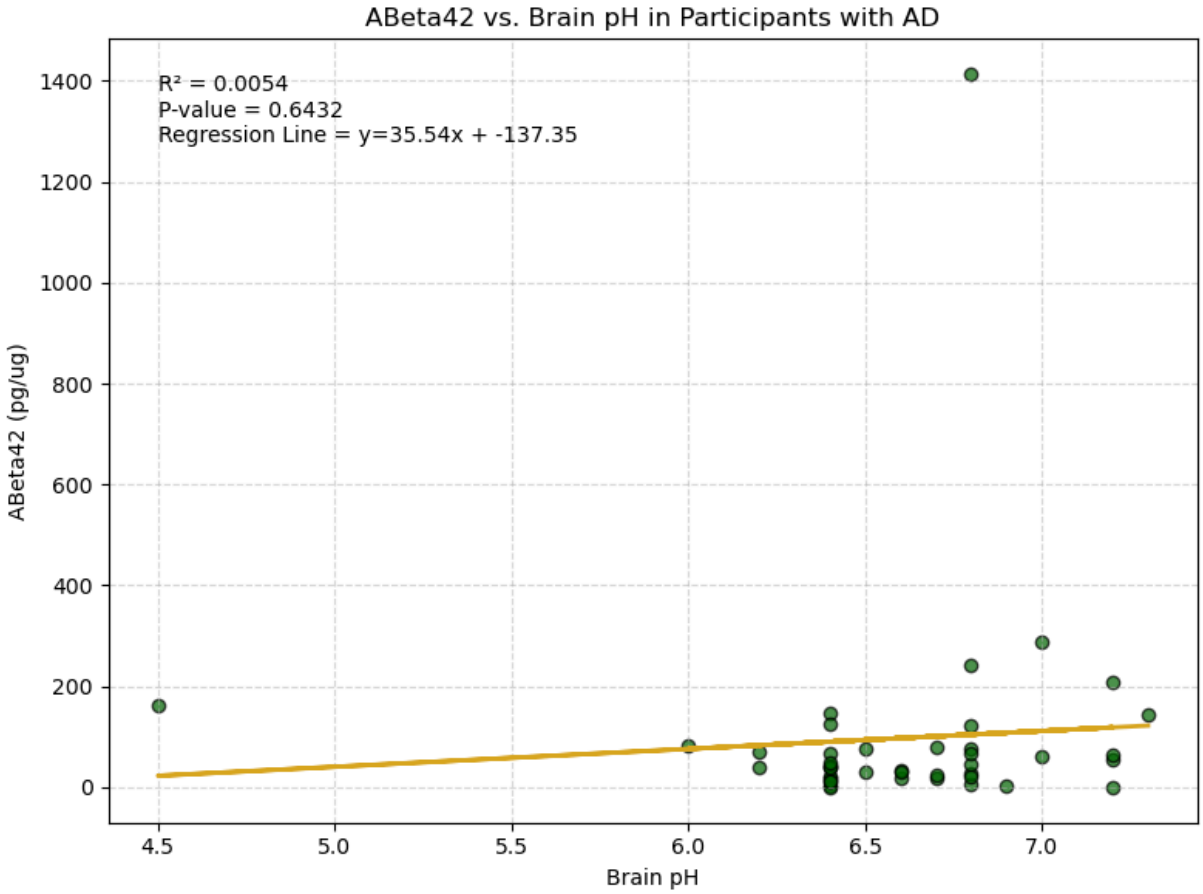
In [23]:
```
#Export the scatter plot as a CSV file
df = pd.DataFrame(plot_df)
df.to_csv('Brain_pH_vs_ABeta42.csv', index=False)
print("CSV file 'Brain_pH_vs_ABeta42.csv' has been created.")
```

```
CSV file 'Brain_pH_vs_ABeta42.csv' has been created.
```

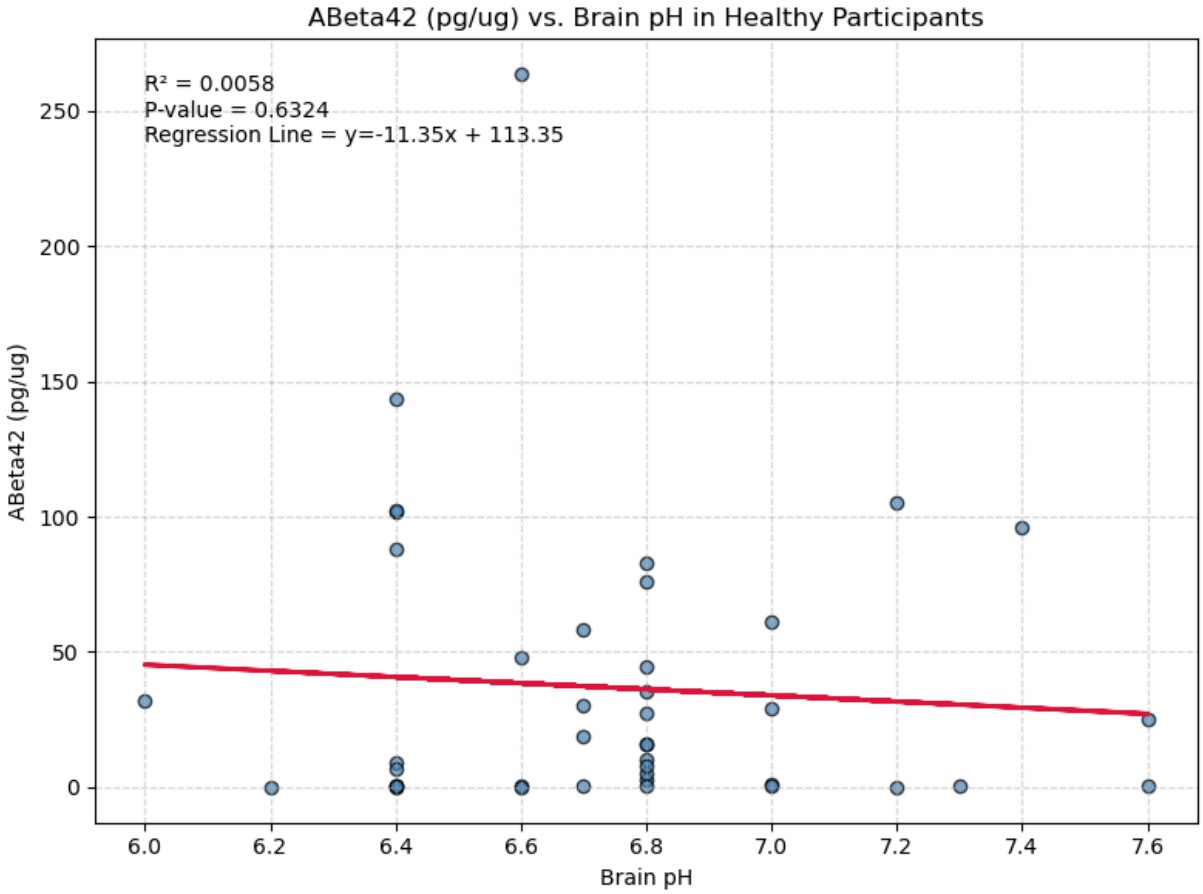
In [24]:
```
#Create a scatter plot with the linear regression line
plot_df = merged_df[['Brain pH', 'ABeta42 pg/ug']].dropna()
y = plot_df['ABeta42 pg/ug']
x = plot_df['Brain pH']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='blue', alpha=0.7, edgecolors='k', label='Data points')
plt.plot(x, regression_line, color='pink', linewidth=2, label=f'Fit line: y={slope:
plt.ylabel('ABeta42 (pg/ug)')
plt.xlabel('Brain pH')
plt.title('ABeta42 vs. Brain pH')
plt.grid(True, linestyle='--', alpha=0.5)
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.tight_layout()
plt.show()
```



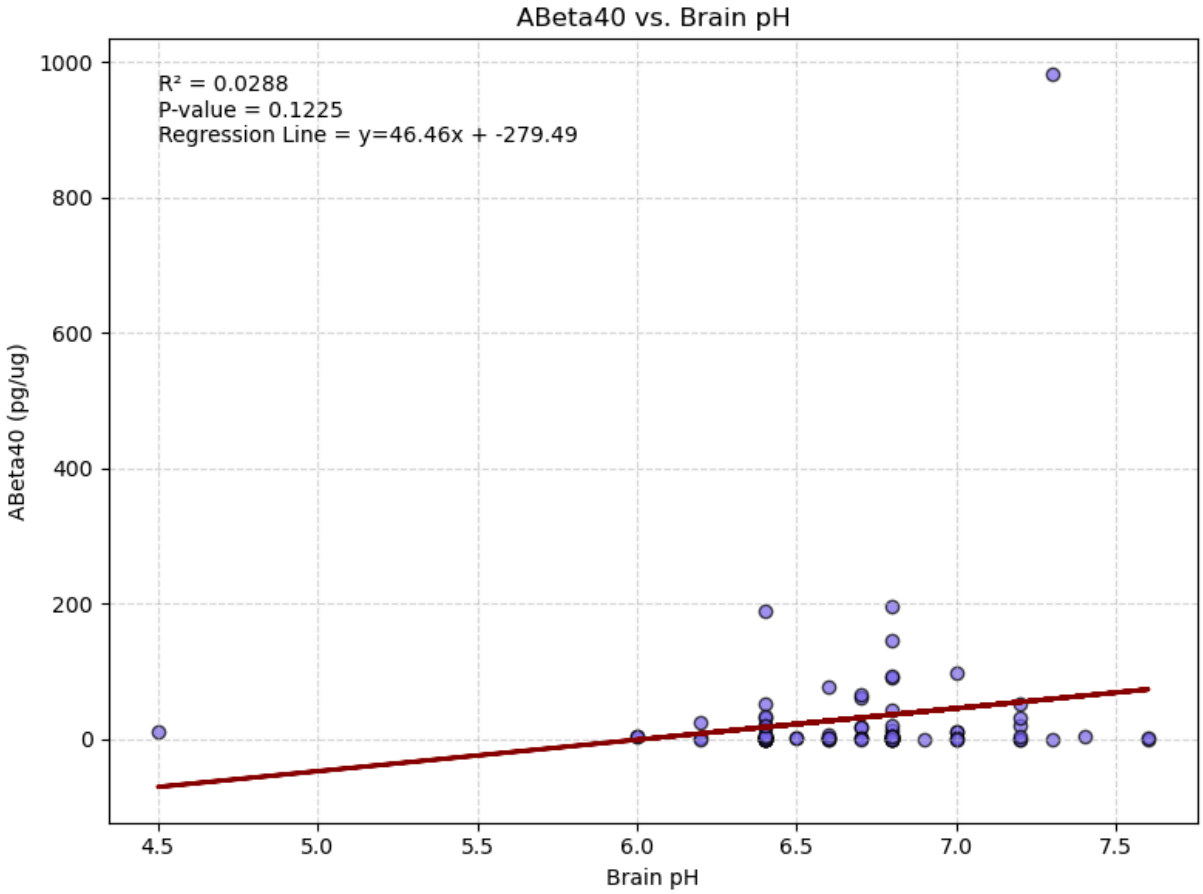
In [25]:
```
#Create a linear regression model comparing ABeta42 pg/ug to Brain pH in participan
dementia_df = merged_df[merged_df['Cognitive Status'] == 'Dementia']
plot_df = dementia_df[['Brain pH', 'ABeta42 pg/ug']].dropna()
y = plot_df['ABeta42 pg/ug']
x = plot_df['Brain pH']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='darkgreen', alpha=0.7, edgecolors='k', label='Dementia Pat
plt.plot(x, regression_line, color='goldenrod', linewidth=2, label=f'Fit line: y={s
plt.ylabel('ABeta42 (pg/ug)')
plt.xlabel('Brain pH')
plt.title('ABeta42 vs. Brain pH in Participants with AD')
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



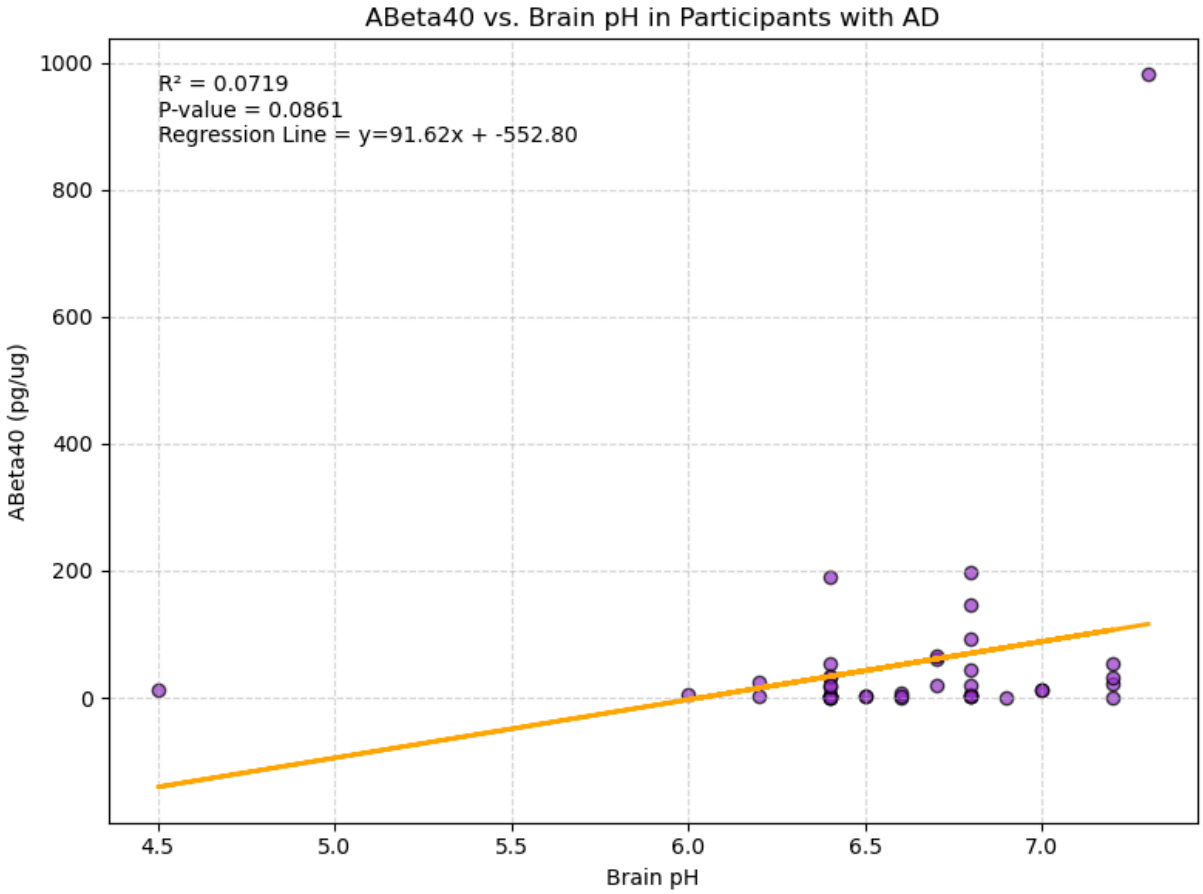
In [26]:
```
#Create a linear regression model comparing ABeta42 pg/ug to Brain pH in healthy pa
non_dementia_df = merged_df[merged_df['Cognitive Status'] != 'Dementia']
plot_df = non_dementia_df[['Brain pH', 'ABeta42 pg/ug']].dropna()
y = plot_df['ABeta42 pg/ug']
x = plot_df['Brain pH']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='steelblue', alpha=0.7, edgecolors='k', label='Non-Dementia
plt.plot(x, regression_line, color='crimson', linewidth=2, label=f'Fit line: y={slo
plt.ylabel('ABeta42 (pg/ug)')
plt.xlabel('Brain pH')
plt.title('ABeta42 (pg/ug) vs. Brain pH in Healthy Participants')
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



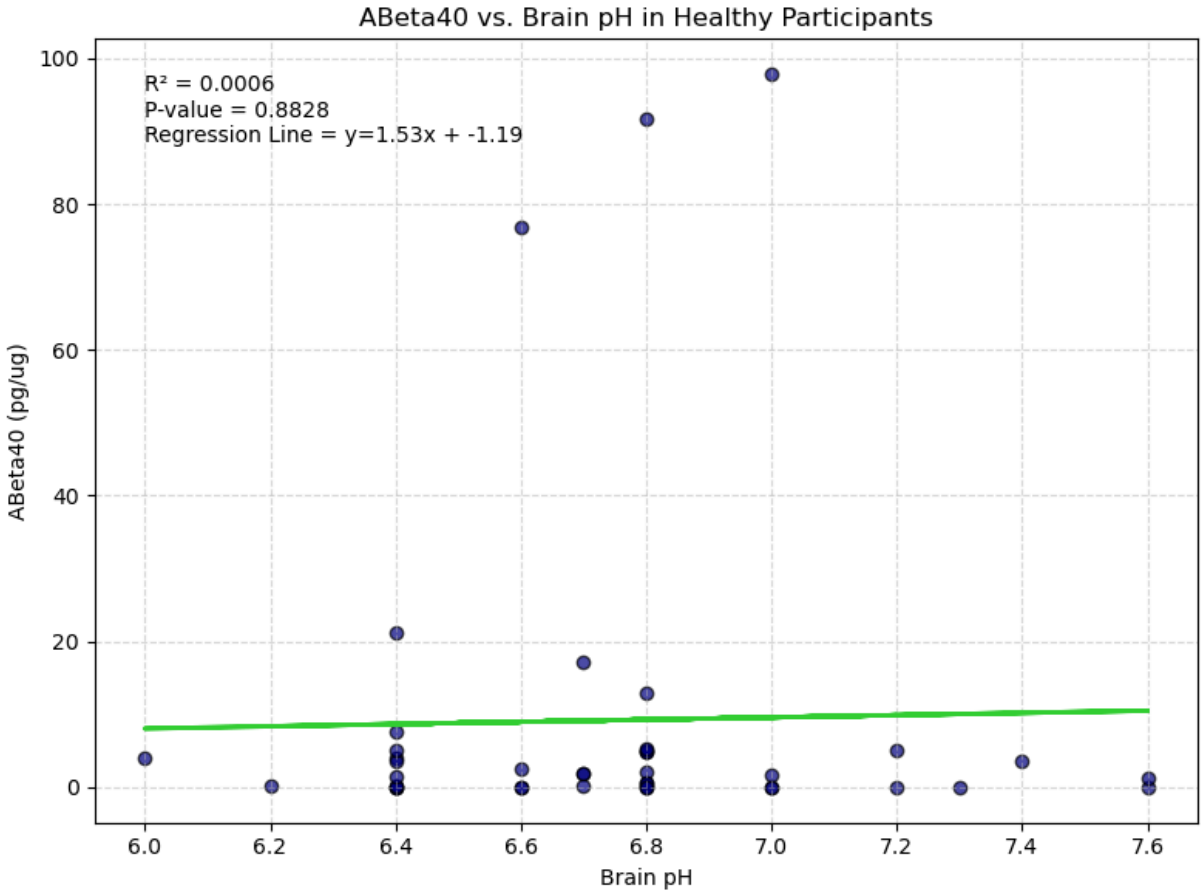
In [27]:
```
#Create a linear regression model comparing ABeta40 pg/ug to Brain pH in all patien
plot_df = merged_df[['Brain pH', 'ABeta40 pg/ug']].dropna()
y = plot_df['ABeta40 pg/ug']
x = plot_df['Brain pH']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='mediumslateblue', alpha=0.7, edgecolors='k', label='Data p
plt.plot(x, regression_line, color='darkred', linewidth=2, label=f'Fit line: y={slo
plt.ylabel('ABeta40 (pg/ug)')
plt.xlabel('Brain pH')
plt.title('ABeta40 vs. Brain pH')
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



In [28]:
```
#Create a linear regression model comparing ABeta40 pg/ug to Brain pH in patients w

plot_df = dementia_df[['Brain pH', 'ABeta40 pg/ug']].dropna()
y = plot_df['ABeta40 pg/ug']
x = plot_df['Brain pH']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='darkorchid', alpha=0.7, edgecolors='k', label='Dementia Pa
plt.plot(x, regression_line, color='orange', linewidth=2, label=f'Fit line: y={slop
plt.ylabel('ABeta40 (pg/ug)')
plt.xlabel('Brain pH')
plt.title('ABeta40 vs. Brain pH in Participants with AD')
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



In [29]:
```
#Create a linear regression model comparing ABeta40 pg/ug to Brain pH in healthy pa
plot_df = non_dementia_df[['Brain pH', 'ABeta40 pg/ug']].dropna()
x = plot_df['Brain pH']
y = plot_df['ABeta40 pg/ug']
slope, intercept, r_value, p_value, std_err = linregress(x, y)
regression_line = slope * x + intercept
r_squared = r_value**2
plt.figure(figsize=(8, 6))
plt.scatter(x, y, color='navy', alpha=0.7, edgecolors='k', label='Non-Dementia Pati
plt.plot(x, regression_line, color='limegreen', linewidth=2, label=f'Fit line: y={s
plt.xlabel('Brain pH')
plt.ylabel('ABeta40 (pg/ug)')
plt.title('ABeta40 vs. Brain pH in Healthy Participants')
plt.text(x.min(),y.max(), f'R² = {r_value**2:.4f}\nP-value = {p_value:.4f} \nRegres
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



Verify and validate your analysis:

To determine the relationship between the brain's pH and the beta-amyloid plaque concentration in participants with and without Alzheimer's disease, a computational study was conducted. Initially, a comparison of the parameters and cognitive status was performed. Bar graphs to compare brain pH, beta-amyloid 40 concentration, and beta-amyloid 42 concentration between participants with Alzheimer's and participants without Alzheimer's were made with error bars, using the MetaData provided. To determine if there was a statistical significance between the two groups in the bar graph, a T-Test was done with a p-value less than 0.05 indicating a correlation. The p-value for brain pH vs. cognitive status was 0.1038, the p-value for beta-amyloid 40 concentration vs. cognitive status was 0.0876, and the p-value for beta-amyloid 42 concentration vs. cognitive status was 0.0830. Then to observe if there was a correlation between brain pH and beta-amyloid concentration, six scatterplots were created with regression lines comparing each amyloid-beta concentration in general, with participants with Alzheimer's and with healthy participants. The data was then statistically analyzed using the python library statistics to determine a correlation. The p-value for amyloid-beta 42 vs. brain pH was 0.9265, the p-value for amyloid-beta 42 vs. brain pH in Alzheimer's participants was 0.6432, and the p-value for amyloid-beta 42 vs. brain pH in healthy participants was 0.6324. The p-value for amyloid-beta 40 vs. brain pH was 0.1225, the p-value for amyloid-beta 40 vs. brain pH in Alzheimer's participants was 0.0861, the p-value for amyloid-beta 42 vs. brain pH in healthy

participants was 0.8828. Since every p-value was greater than 0.05, the null hypothesis must be accepted and the hypothesis must be rejected for this data set. However, current literature indicates that there is a correlation with lower pH values and an increased rate of Alzheimer's development and that there is an inverse relationship with brain pH and the concentration of amyloid-beta plaques. This is linked to the increased activity of CD-68+ microglia in the brain as well as a disrupted lysosomal homeostasis (Wang et al., 2025). Thus, this data set could have had confounding factors or that the correlation between brain pH and amyloid-beta concentration was simply not seen in this group of participants.

Wang, X., Shao, X., Yu, L., Sun, J., Yin, X.-S., Chen, Z., Xu, Y., Wang, N., Zhang, D., Qiu, W., Liu, F., & Ma, C. (2025a, February 10). Changes in the ph value of the human brain in alzheimer's disease pathology correlated with CD68-positive microglia: A community-based autopsy study in Beijing, China - molecular brain. BioMed Central. https://molecularbrain.biomedcentral.com/articles/10.1186/s13041-025-01180-3

Conclusions and Ethical Implications:

Based on the computational study from the provided data set, there is no statistically significant correlation between brain pH and the incidence of Alzheimer's disease and there is no statistically significant correlation between brain pH and an increased concentration of amyloid beta plaques in participants. However, current literature has shown that there is a correlation with brain pH, Alzheimer's pathogenesis, and amyloid-beta concentration. Ethically, it would be best for physicians to be wary of a decrease in brain pH as this could potentially be a precursor to Alzheimer's; however, as to not incite panic or discrimination in patient populations, exercising caution and empathy, as well as refering to the patients symptoms and family history before communicating a potential Alzheimer's diagnosis would be best practice. Additionally, the participant's personal identifying information in this study must remain anonymous and a more robust study including a wider array of participants should be conducted before definitively relaying results to the public, as with a smaller dataset (like this one) it is harder to confidently apply these results to the entire American population.

Limitations and Future Work:

This study is limited, due to the sample size being 84 participants. Having a larger sample size with a greater representation of individuals would help to make the results of the study more credible and applicable to the general American public. Additionally, having longitudinal data from each participant could better illustrate the neuropathogenesis of Alzheimer's disease and this data would enable a more well-rounded computational analysis, as it would allow one to see the progression of the development of the disease and through this relate certain biomarkers to the pathogenesis of Alzheimer's. In the future, a longitudinal study of participants and their biomarkers as well as a larger sample size with greater

diversity reflecting the American public could be used to increase the credibility of the results. Moreover, certain neuropathological assessments, such as Thal score, could be used to create a more holistic review of the data.

Notes/Questions:

- Notes on the dataset:
  - Males: 33
  - Females: 51
  - Average age of death: 88
  - Participants were selected via varying AD pathologies
  - Single nucleus and spatial genomics were used to analyze the genes associated with the disease
  - Full cell type diversity was analyzed
  - Demographics of participants were noted
  - Cells were cross-referenced via BICCN and validated by independent datasets
- Questions:
  - N/A