Module 1:

Team Members:

Annie & Ana

Project Title:

APOE-ε4 and Its Role in Alzheimer's Biomarkers

Project Goal:

This project seeks to determine whether people with the APOE-ε4 gene have higher levels of Alzheimer's biomarkers, like amyloid and tau, compared to those without the gene. The goal is to understand how APOE-ε4 relates to these biomarkers using existing data.

Questions:

- Do patients with the APOE-ε4 gene show higher levels of certain biomarkers (like amyloid or tau) compared to non-carriers?

Other options:

- Which one, Aβ40 or Aβ42, lines up better with other disease markers like tau or inflammation?
- Do lifestyle factors in the dataset (exercise, smoking, education level, etc.) appear to slow or speed up decline?

Disease Background:

- Prevalence & incidence (for USA)
    - prevalence: About 7.2 million people aged 65+ have alzheimers disease in the US. (2025)
    - Incidence: It is estimated that there are 0.4% new cases around the age of 65 to 74, 3.2% around the age of 75 to 84, 7.6% for ages 85+.
- Economic burden
    - The total projected economic burden for the US in 2025 is $781 billion. This includes indirect (unpaid care, etc) and direct care (medical and long term ca$ 232 to $384 billion - Indirect care is projected to be a total of $233 billion for unpaid care and $302 billion for other indirect factors
        - The estimated economic burden from 2020 to 2022 was $600 billion. Therefore, It is estimated to have an increase of $181 billion by the end of 2025.

- Risk factors (genetic, lifestyle)
    - Genetic: Having a family history of Alzheimer's or certain genes, like APOE-ε4, can increase risk.
    - Lifestyle: Things like little exercise, unhealthy diet, smoking, heavy drinking, or not staying mentally and socially active can raise risk.
    - Health conditions: Problems such as high blood pressure, diabetes, obesity, or heart disease can also make Alzheimer's more likely.
- Societal determinants
    - Access to healthcare can affect early diagnosis and treatment.
    - Education and cognitive engagement may lower risk or delay symptoms.
    - Socioeconomic status can influence nutrition, healthcare access, and support systems.
    - Social support from family, friends, and community helps manage symptoms and improve quality of life.
- Symptoms
    - Early symptoms include memory loss, trouble remembering recent events, and difficulty planning or solving problems.
    - As the disease progresses, people may become confused, have mood changes, and struggle with daily tasks.
    - In later stages, severe memory loss, difficulty communicating, and loss of independence are common.
    - Other possible symptoms include disorientation, personality changes, and sleep disturbances.
- Diagnosis
    - Diagnosis is usually based on a combination of medical history, cognitive tests, and physical exams.
    - Doctors may use brain imaging (like MRI or PET scans) to look for changes in the brain.
    - Biomarker tests, including amyloid and tau measurements from blood or cerebrospinal fluid, can help confirm the disease.
    - Early detection is important for planning care and starting treatments that can help manage symptoms.
- Standard of care treatments (& reimbursement)
    - Common medicines include donepezil and memantine, which help with memory and thinking.
    - Newer drugs, like lecanemab, aim to slow the disease in its early stages.
    - These treatments don't cure Alzheimer's but can manage symptoms or slow progression.
    - Most medicines are covered by Medicare, Medicaid, or private insurance, though coverage can vary.
    - Support services, such as therapy or caregiver programs, may also be partly covered.
- Disease progression & prognosis
    - Alzheimer's disease progresses gradually, usually over several years.
    - Early stages involve mild memory loss and difficulty with thinking or planning.
    - Middle stages bring increased confusion, mood changes, and trouble performing daily tasks.

    - Late stages can lead to severe cognitive decline, loss of independence, and difficulty communicating.
    - The rate of progression varies by individual and can be influenced by genetics, overall health, and lifestyle.
    - There is no cure, but treatments and support can help manage symptoms and improve quality of life.
- Continuum of care providers
    - primary care doctors often are the first to notice symptoms, provive checkups, and manage overall health
    - neurologists and geriatric specialists diagnoce and treat memory loss and brain changes
    - nurses and nurse practitioners help monitor symptons, medications, and daily needs
    - social workers and care coordinators connect families with resources, support groups, and long-term care options.
    - long-term care facilities and memory care units provide 24/7 support as the disease progresses.
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)
    - anatomy (brain regions): alzheimer's starts in the hippocampus (memory) and spreads through the cortex, causing the brain to shrink.
    - organ physiology (brain functions): signals between brain cells get weaker, making it harder to remember, think, and carry out daily tasks.
    - cell physiology: nerve cells lose connections and die, while suppoer cells create inflammation.
    - molecular physiology:sticky amyloid plaques form outside cells, while twisted tau tangles build up inside, leading to damage and cell death.
- Clinical Trials/next-gen therapies
    - Lecanemab (Leqembi): newest treatment which slows early disease by reducing amyloid
    - Donanemab: next-gen drug which helps in late-stage trials
    - Various therapies are being tested which target tau protiens, inflamation, and brain cell health
    - clinical trials use blood tests and brain scans to find the disease earlier and help to monitor the disease and treatment safetly.

ChatGPT

Data-Set:

For this project, we will use two datasets. The first, UpdatedMetaData.csv, contains participant information such as diagnosis (healthy, mild cognitive impairment, or Alzheimer’s), basic demographics (age, sex, education), and genetic details including APOE status. The second, UpdatedLuminex.csv, includes biomarker measurements from lab tests, such as Aβ40, Aβ42, tau, and several inflammation-related proteins. These biomarkers were measured using Luminex immunoassays in units of picograms per milliliter (pg/mL). Together, these datasets allow us to analyze whether APOE-ε4 carriers have higher biomarker levels compared to non-carriers, directly addressing our research question. To analyze the data, we merged the two datasets by patient ID so that each participant’s biomarker measurements were matched with their demographic and genetic information. Then, organized the

data into two groups: APOE-ε4 carriers and non-carriers. For each biomarker, we calculated group means, standard deviations, and variances. Using a Welch's two-sample t-test, we compared the biomarker levels between carriers and non-carriers. This provided t-statistics, approximate degrees of freedom, and two-tailed p-values to assess whether observed differences were statistically significant. The results were summarized by reporting descriptive statistics for each group, the test values, and whether differences reached significance at the 0.05 level. UpdatedMetaData.csv UpdatedLuminex.csv

Code Description:

The purpose of this code is to use patient data to explore how APOE-ε4 status might relate to dementia risk. It organizes the data so that carriers and non-carriers are clearly identified, and then performs statistical analyses to assess differences between these groups. The code also generates visualizations, including bar graphs and scatter plots, to illustrate these patterns. Each section of the code is labeled and structured for clarity, making it easy for someone reading the notebook to follow the steps and reproduce the results.

Data Analyis:

In [47]:
```
import csv

file_path = r'C:\Users\Annie\Downloads\comp bme\module 1\UpdatedMetaData.csv'

with open(file_path, newline='') as f:
    reader = csv.reader(f)
    headers = next(reader)  # get the header row
    print("Headers for UpdatedMetaData.csv:")
    for h in headers:
        print(h)

print("\n")  # blank line for readability

# headers for data set UpdatedLuminex.csv
import csv

file_path = r'C:\Users\Annie\Downloads\comp bme\module 1\UpdatedLuminex.csv'

with open(file_path, newline='') as f:
    reader = csv.reader(f)
    headers = next(reader)  # get the header row
    print("Headers for UpdatedLuminex.csv:")
    for h in headers:
        print(h)
```

Headers for UpdatedMetaData.csv:
Donor ID
Primary Study Name
Secondary Study Name
Age at Death
Sex
Race (choice=White)
Race (choice=Black/ African American)
Race (choice=Asian)
Race (choice=American Indian/ Alaska Native)
Race (choice=Native Hawaiian or Pacific Islander)
Race (choice=Unknown or unreported)
Race (choice=Other)
specify other race
Hispanic/Latino
Highest level of education
Years of education
APOE Genotype
Cognitive Status
Age of onset cognitive symptoms
Age of Dementia diagnosis
Known head injury
Have they had neuroimaging
Consensus Clinical Dx (choice=Alzheimers disease)
Consensus Clinical Dx (choice=Alzheimers Possible/ Probable)
Consensus Clinical Dx (choice=Ataxia)
Consensus Clinical Dx (choice=Corticobasal Degeneration)
Consensus Clinical Dx (choice=Control)
Consensus Clinical Dx (choice=Dementia with Lewy Bodies/ Lewy Body Disease)
Consensus Clinical Dx (choice=Frontotemporal lobar degeneration)
Consensus Clinical Dx (choice=Huntingtons disease)
Consensus Clinical Dx (choice=Motor Neuron disease)
Consensus Clinical Dx (choice=Multiple System Atrophy)
Consensus Clinical Dx (choice=Parkinsons disease)
Consensus Clinical Dx (choice=Parkinsons Cognitive Impairment - no dementia)
Consensus Clinical Dx (choice=Parkinsons Disease Dementia)
Consensus Clinical Dx (choice=Prion)
Consensus Clinical Dx (choice=Progressive Supranuclear Palsy)
Consensus Clinical Dx (choice=Taupathy)
Consensus Clinical Dx (choice=Vascular Dementia)
Consensus Clinical Dx (choice=Unknown)
Consensus Clinical Dx (choice=Other)
If other Consensus dx, describe
Last CASI Score
Interval from last CASI in months
Last MMSE Score
Interval from last MMSE in months
Last MOCA Score
Interval from last MOCA in months
PMI
Rapid Frozen Tissue Type
Ex Vivo Imaging
Fresh Brain Weight
Brain pH
Overall AD neuropathological Change
Thal
Braak
CERAD score
Overall CAA Score
Highest Lewy Body Disease

```
Total Microinfarcts (not observed grossly)
Total microinfarcts in screening sections
Atherosclerosis
Arteriolosclerosis
LATE
RIN
Severely Affected Donor


Headers for UpdatedLuminex.csv:
Donor ID
ABeta40 pg/ug
ABeta42 pg/ug
tTAU pg/ug
pTAU pg/ug
```

In [48]:

```python
# Displays paitents, carier vs noncarier, and biomarkers

import csv
import warnings
import pandas as pd
import matplotlib.pyplot as plt

class Patient:
    all_patients = []

    def __init__(self, DonorID, ABeta40=None, ABeta42=None, tTau=None, pTau=None):
        self.DonorID = DonorID
        self.ABeta40 = ABeta40
        self.ABeta42 = ABeta42
        self.tTau = tTau
        self.pTau = pTau

        # metadata variables (will be filled later)
        self.sex = None
        self.death_age = None
        self.ed_lvl = None
        self.cog_stat = None
        self.age_symp_on = None
        self.age_diag = None
        self.head_inj = None
        self.thal_score = None

        # NEW: APOE fields
        self.apoe_genotype = None
        self.apoe_e4_status = None  # Carrier / Non-carrier

        Patient.all_patients.append(self)

    def __repr__(self):
        return (
            f"{self.DonorID} | sex: {self.sex} | APOE: {self.apoe_genotype} "
            f"({self.apoe_e4_status}) | ABeta40 {self.ABeta40} | ABeta42 {self.ABeta42} | "
            f"tTau {self.tTau} | pTau {self.pTau} | "
            f"Death Age {self.death_age} | Thal Score {self.thal_score}"
        )

    @classmethod
    def instantiate_from_csv(cls, luminex_file: str, metadata_file: str):
        """Create patient objects from the Luminex file, then enrich with Metadata."""
```

```
        with open(luminex_file, encoding="utf8") as f:
            reader = csv.DictReader(f)
            rows = list(reader)
            for row in rows:
                Patient(
                    DonorID=row['Donor ID'],
                    ABeta40=float(row['ABeta40 pg/ug']) if row['ABeta40 pg/ug'] else None,
                    ABeta42=float(row['ABeta42 pg/ug']) if row['ABeta42 pg/ug'] else None,
                    tTau=float(row['tTAU pg/ug']) if row['tTAU pg/ug'] else None,
                    pTau=float(row['pTAU pg/ug']) if row['pTAU pg/ug'] else None
                )

        # Sort by DonorID for consistent merging
        Patient.all_patients.sort(key=lambda x: x.DonorID)

        # Add metadata info
        Patient.combine_data(metadata_file)

    @classmethod
    def combine_data(cls, metadata_file: str):
        """Merge metadata into already-created patient objects."""
        with open(metadata_file, encoding="utf8") as f:
            reader = csv.DictReader(f)
            rows = list(reader)

            metadata_dict = {row["Donor ID"]: row for row in rows}

            for patient in Patient.all_patients:
                meta = metadata_dict.get(patient.DonorID)
                if meta:
                    if meta.get("Sex"):
                        patient.sex = meta["Sex"]

                    if meta.get("Age at Death"):
                        patient.death_age = int(meta["Age at Death"])

                    if meta.get("Highest level of education"):
                        patient.ed_lvl = meta["Highest level of education"]

                    if meta.get("Cognitive Status"):
                        patient.cog_stat = meta["Cognitive Status"]

                    if meta.get("Age of onset cognitive symptoms"):
                        try:
                            patient.age_symp_on = int(meta["Age of onset cognitive symptoms"
                        except:
                            pass

                    if meta.get("Age of Dementia diagnosis"):
                        try:
                            patient.age_diag = int(meta["Age of Dementia diagnosis"])
                        except:
                            pass

                    if meta.get("Known head injury"):
                        patient.head_inj = meta["Known head injury"]

                    if meta.get("Thal"):
                        try:
                            patient.thal_score = int(meta["Thal"].split()[1])  # "Thal 3" →
```

```
                        except:
                            patient.thal_score = None

                    # ✅ NEW: APOE genotype + Carrier/Non-carrier logic
                    if meta.get("APOE Genotype"):
                        patient.apoe_genotype = meta["APOE Genotype"]

                        if "4" in meta["APOE Genotype"]:
                            patient.apoe_e4_status = "Carrier"
                        else:
                            patient.apoe_e4_status = "Non-carrier"
                else:
                    warnings.warn(f"No metadata found for DonorID {patient.DonorID}")


# -------------------------
# Run the merge on your files
# -------------------------
if __name__ == "__main__":
    Patient.instantiate_from_csv(
        r"C:\Users\Annie\Downloads\comp bme\module 1\UpdatedLuminex.csv",
        r"C:\Users\Annie\Downloads\comp bme\module 1\UpdatedMetaData.csv"
    )

    # print first few patients to check
    for p in Patient.all_patients:
        print(p)
```

H19.33.004 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.019621053 | ABeta42 0.971578947 | tTau 1552.414737 | pTau 1.901052632 | Death Age 80 | Thal Score 0
H20.33.001 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.215789474 | ABeta42 2.744210526 | tTau 756.0905263 | pTau 2.737894737 | Death Age 82 | Thal Score 2
H20.33.002 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 0.000597895 | ABeta42 0.147157895 | tTau 313.5252632 | pTau 2.615789474 | Death Age 97 | Thal Score 0
H20.33.004 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 60.76631579 | ABeta42 80.26631579 | tTau 318.5284211 | pTau 7.412631579 | Death Age 86 | Thal Score 5
H20.33.005 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 5.136842105 | ABeta42 16.15684211 | tTau 107.3484211 | pTau 1.327368421 | Death Age 99 | Thal Score 3
H20.33.008 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 3.991578947 | ABeta42 101.8305263 | tTau 125.9336842 | pTau 2.569473684 | Death Age 92 | Thal Score 4
H20.33.011 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 11.84526316 | ABeta42 60.51157895 | tTau 1141.492355 | pTau 8.536842105 | Death Age 93 | Thal Score 5
H20.33.012 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 2.529473684 | ABeta42 47.70947368 | tTau 950.7410526 | pTau 4.545263158 | Death Age 91 | Thal Score 1
H20.33.013 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 1.127368421 | ABeta42 24.78105263 | tTau 272.5084211 | pTau 3.106315789 | Death Age 94 | Thal Score 3
H20.33.014 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.526168105 | ABeta42 16.13789474 | tTau 258.6242105 | pTau 3.398947368 | Death Age 82 | Thal Score 3
H20.33.015 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 1.944210526 | ABeta42 27.60947368 | tTau 393.1831579 | pTau 1.827368421 | Death Age 88 | Thal Score 3
H20.33.016 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 2.671578947 | ABeta42 21.27368421 | tTau 488.8989474 | pTau 2.282105263 | Death Age 93 | Thal Score 4
H20.33.017 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 52.64210526 | ABeta42 209.4347368 | tTau 239.3778947 | pTau 5.881052632 | Death Age 69 | Thal Score 4
H20.33.018 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 196.732 | ABeta42 1412.566961 | tTau 177.5663158 | pTau 5.110526316 | Death Age 81 | Thal Score 5
H20.33.019 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 1.718947368 | ABeta42 28.81368421 | tTau 312.7442105 | pTau 2.884210526 | Death Age 87 | Thal Score 4
H20.33.020 | sex: Male | APOE: 4/4 (Carrier) | ABeta40 145.2547368 | ABeta42 45.72842105 | tTau 21.71894737 | pTau 3.873684211 | Death Age 81 | Thal Score 5
H20.33.024 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 5.095789474 | ABeta42 105.1042105 | tTau 309.08 | pTau 5.222105263 | Death Age 90 | Thal Score 4
H20.33.025 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 3.532631579 | ABeta42 95.79263158 | tTau 384.84 | pTau 3.691578947 | Death Age 94 | Thal Score 4
H20.33.026 | sex: Female | APOE: 4/4 (Carrier) | ABeta40 31.56526316 | ABeta42 63.37473684 | tTau 191.0505263 | pTau 12.56736842 | Death Age 75 | Thal Score 4
H20.33.027 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.843157895 | ABeta42 29.95578947 | tTau 224.2431579 | pTau 3.365263158 | Death Age 99 | Thal Score 3
H20.33.028 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.127368421 | ABeta42 18.94736842 | tTau 192.0284211 | pTau 2.927368421 | Death Age 94 | Thal Score 4
H20.33.029 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.633684211 | ABeta42 28.85473684 | tTau 302.2315789 | pTau 3.191578947 | Death Age 91 | Thal Score 4
H20.33.030 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 17.22105263 | ABeta42 58.26631579 | tTau 114.6231579 | pTau 6.56 | Death Age 86 | Thal Score 4
H20.33.031 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 2.004210526 | ABeta42 42.51368421 | tTau 335.7452632 | pTau 7.827368421 | Death Age 87 | Thal Score 4
H20.33.032 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 91.74842105 | ABeta42 44.25684211 | tTau 156.6284211 | pTau 12.47052632 | Death Age 98 | Thal Score 5
H20.33.033 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 20.21157895 | ABeta42 123.3684211 | tTau 92.80210526 | pTau 3.712631579 | Death Age 68 | Thal Score 5
H20.33.034 | sex: Female | APOE: 2/2 (Non-carrier) | ABeta40 4.794736842 | ABeta42 4.96 | tTau 569.2336842 | pTau 2.593684211 | Death Age 85 | Thal Score 3
H20.33.035 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.030147368 | ABeta42 0.525263158 | tTau 533.5926316 | pTau 4.036842105 | Death Age 99 | Thal Score 0
H20.33.036 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 3.594736842 | ABeta42 102.4557895 | tTau 345.8894737 | pTau 1.28 | Death Age 100 | Thal Score 5
H20.33.037 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 53.01263158 | ABeta42 67.65473684 | tTau 283.24 | pTau 4.569473684 | Death Age 96 | Thal Score 5

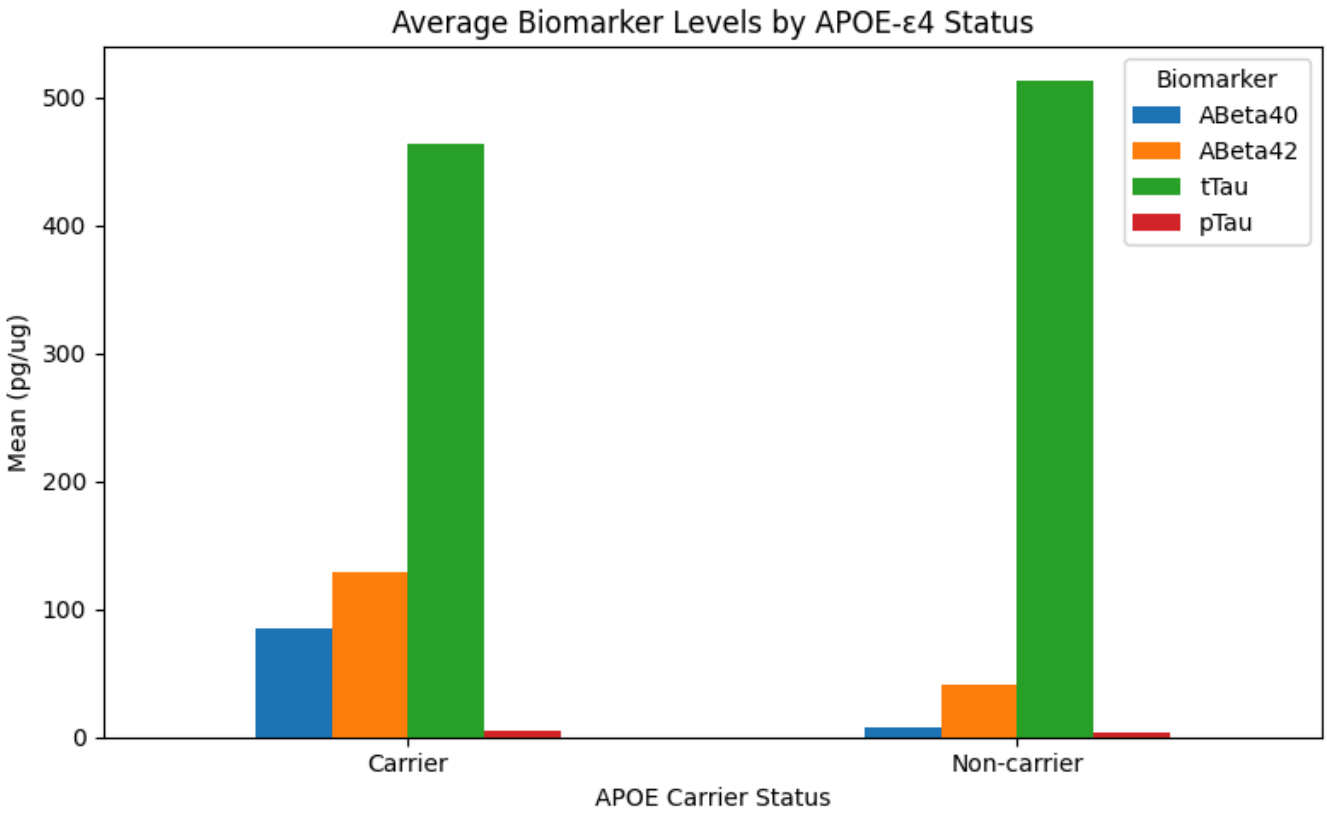
H20.33.038 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 5.176842105 | ABeta42 81.1378947
4 | tTau 121.4084211 | pTau 4.016842105 | Death Age 90 | Thal Score 4
H20.33.039 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 2.062105263 | ABeta42 27.3347368
4 | tTau 482.5421053 | pTau 3.865263158 | Death Age 96 | Thal Score 4
H20.33.040 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 1.412631579 | ABeta42 12.69789474
| tTau 401.9305263 | pTau 1.809473684 | Death Age 98 | Thal Score 4
H20.33.041 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 5.522105263 | ABeta42 242.586315
8 | tTau 196.9957895 | pTau 2.406315789 | Death Age 91 | Thal Score 4
H20.33.043 | sex: Male | APOE: 4/4 (Carrier) | ABeta40 97.8 | ABeta42 60.95368421 | tTau 709.
8136842 | pTau 5.782105263 | Death Age 85 | Thal Score 4
H20.33.044 | sex: Male | APOE: 2/3 (Non-carrier) | ABeta40 0.007088421 | ABeta42 0.245263158
| tTau 7005.543158 | pTau 5.630526316 | Death Age 81 | Thal Score 0
H20.33.045 | sex: Female | APOE: 4/4 (Carrier) | ABeta40 981.444 | ABeta42 142.778 | tTau 112
2.432229 | pTau 5.415789474 | Death Age 77 | Thal Score 5
H20.33.046 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 25.29578947 | ABeta42 69.98842105
| tTau 283.4368421 | pTau 15.91789474 | Death Age 94 | Thal Score 5
H21.33.001 | sex: Male | APOE: 2/3 (Non-carrier) | ABeta40 0.000882947 | ABeta42 0.405263158
| tTau 452.1894737 | pTau 3.038947368 | Death Age 80 | Thal Score 2
H21.33.002 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 93.67684211 | ABeta42 74.77684211 |
tTau 200.3842105 | pTau 7.317894737 | Death Age 70 | Thal Score 5
H21.33.003 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.000804526 | ABeta42 0.405263158
| tTau 393.8768421 | pTau 3.092631579 | Death Age 78 | Thal Score 0
H21.33.004 | sex: Male | APOE: 2/3 (Non-carrier) | ABeta40 0.001155368 | ABeta42 0.670526316
| tTau 324.3410526 | pTau 3.475789474 | Death Age 93 | Thal Score 0
H21.33.005 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 1.655789474 | ABeta42 6.554736842
| tTau 549.82 | pTau 3.131578947 | Death Age 95 | Thal Score 3
H21.33.006 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 12.87684211 | ABeta42 82.97263158 | tT
au 160.5831579 | pTau 3.169473684 | Death Age 97 | Thal Score 4
H21.33.007 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 11.41894737 | ABeta42 287.412 |
tTau 1179.673684 | pTau 6.410526316 | Death Age 86 | Thal Score 4
H21.33.008 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 18.994 | ABeta42 18.994 | tTau 1
26.1673684 | pTau 6.175789474 | Death Age 91 | Thal Score 4
H21.33.009 | sex: Female | APOE: 4/4 (Carrier) | ABeta40 189.2905263 | ABeta42 40.19894737 |
tTau 130.4147368 | pTau 4.948421053 | Death Age 65 | Thal Score 5
H21.33.010 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 66.77578947 | ABeta42 24.6378947
4 | tTau 290.8684211 | pTau 1.922105263 | Death Age 93 | Thal Score 5
H21.33.011 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.000688421 | ABeta42 0.13734736
8 | tTau 276.5368421 | pTau 3.052631579 | Death Age 83 | Thal Score 0
H21.33.012 | sex: Female | APOE: 2/4 (Carrier) | ABeta40 0.215789474 | ABeta42 3.502105263 |
tTau 238.6705263 | pTau 3.669473684 | Death Age 93 | Thal Score 3
H21.33.013 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 43.23368421 | ABeta42 68.36631579 |
tTau 599.8652632 | pTau 1.630526316 | Death Age 94 | Thal Score 4
H21.33.014 | sex: Male | APOE: 2/3 (Non-carrier) | ABeta40 4.864210526 | ABeta42 35.27578947
| tTau 197.3589474 | pTau 1.598947368 | Death Age 92 | Thal Score 4
H21.33.015 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.661052632 | ABeta42 10.09578947
| tTau 322.6021053 | pTau 5.006315789 | Death Age 98 | Thal Score 2
H21.33.016 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.009426737 | ABeta42 0.52526315
8 | tTau 303.0031579 | pTau 4.090526316 | Death Age 94 | Thal Score 1
H21.33.017 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.412631579 | ABeta42 20.1821052
6 | tTau 164.9431579 | pTau 2.075789474 | Death Age 92 | Thal Score 5
H21.33.018 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.263157895 | ABeta42 10.9884210
5 | tTau 170.9052632 | pTau 1.995789474 | Death Age 89 | Thal Score 3
H21.33.019 | sex: Male | APOE: 2/3 (Non-carrier) | ABeta40 0.001077758 | ABeta42 0.019621053
| tTau 122.2210526 | pTau 2.208421053 | Death Age 75 | Thal Score 1
H21.33.020 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 1.547368421 | ABeta42 38.15894737
| tTau 202.5905263 | pTau 3.328421053 | Death Age 82 | Thal Score 4
H21.33.021 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.001261053 | ABeta42 2.672631579
| tTau 58.70105263 | pTau 1.324210526 | Death Age 99 | Thal Score 4
H21.33.022 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.000130411 | ABeta42 7.66631578
9 | tTau 270.3010526 | pTau 3.095789474 | Death Age 82 | Thal Score 2

```
H21.33.023 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.000597684 | ABeta42 0.114736842
| tTau 188.3642105 | pTau 1.683157895 | Death Age 102 | Thal Score 0
H21.33.025 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 21.20947368 | ABeta42 8.84210526
3 | tTau 738.4673684 | pTau 2.74 | Death Age 88 | Thal Score 3
H21.33.026 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 76.91789474 | ABeta42 263.5368421 |
tTau 386.6842105 | pTau 6.217894737 | Death Age 90 | Thal Score 4
H21.33.027 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 31.71157895 | ABeta42 39.83789474 | tT
au 222.8189474 | pTau 2.709473684 | Death Age 92 | Thal Score 5
H21.33.028 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.072506632 | ABeta42 0.204385895
| tTau 391.1515789 | pTau 3.065263158 | Death Age 72 | Thal Score 1
H21.33.029 | sex: Male | APOE: 2/4 (Carrier) | ABeta40 17.82736842 | ABeta42 146.8621053 | tT
au 953.1326316 | pTau 2.04 | Death Age 89 | Thal Score 5
H21.33.030 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 1.827368421 | ABeta42 18.54736842 | tT
au 948.1368421 | pTau 5.212631579 | Death Age 89 | Thal Score 3
H21.33.031 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 34.42947368 | ABeta42 124.4347368 | tT
au 860.3778947 | pTau 4.793684211 | Death Age 84 | Thal Score 5
H21.33.032 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.375789474 | ABeta42 6.77789473
7 | tTau 359.58 | pTau 3.14 | Death Age 98 | Thal Score 2
H21.33.033 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 3.967368421 | ABeta42 31.76315789 |
tTau 667.3905263 | pTau 4.462105263 | Death Age 83 | Thal Score 5
H21.33.034 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 12.15368421 | ABeta42 161.0947368 |
tTau 393.3484211 | pTau 6.388421053 | Death Age 90 | Thal Score 5
H21.33.035 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 7.491578947 | ABeta42 143.464210
5 | tTau 903.6189474 | pTau 5.306315789 | Death Age 97 | Thal Score 5
H21.33.036 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 5.302105263 | ABeta42 75.7894736
8 | tTau 238.4989474 | pTau 5.577894737 | Death Age 93 | Thal Score 4
H21.33.037 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 0.036216842 | ABeta42 0.44964912
6 | tTau 558.1957895 | pTau 2.334736842 | Death Age 88 | Thal Score 2
H21.33.038 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 0.079678842 | ABeta42 0.12263157
9 | tTau 131.0326316 | pTau 2.977894737 | Death Age 84 | Thal Score 1
H21.33.039 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.450526316 | ABeta42 76.2263157
9 | tTau 704.8010526 | pTau 4.146315789 | Death Age 88 | Thal Score 4
H21.33.040 | sex: Male | APOE: 3/4 (Carrier) | ABeta40 0.065191789 | ABeta42 0.490526316 | tT
au 894.1368421 | pTau 3.850526316 | Death Age 83 | Thal Score 4
H21.33.041 | sex: Female | APOE: 2/3 (Non-carrier) | ABeta40 5.010526316 | ABeta42 88.1694736
8 | tTau 740.5831579 | pTau 3.327368421 | Death Age 98 | Thal Score 0
H21.33.042 | sex: Female | APOE: 4/4 (Carrier) | ABeta40 20.53894737 | ABeta42 47.93263158 |
tTau 531.6515789 | pTau 2.507368421 | Death Age 91 | Thal Score 5
H21.33.043 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 1.593684211 | ABeta42 29.8926315
8 | tTau 611.0 | pTau 4.404210526 | Death Age 95 | Thal Score 4
H21.33.044 | sex: Female | APOE: 3/3 (Non-carrier) | ABeta40 7.130526316 | ABeta42 33.6378947
4 | tTau 417.8947368 | pTau 3.662105263 | Death Age 88 | Thal Score 3
H21.33.045 | sex: Female | APOE: 3/4 (Carrier) | ABeta40 21.42315789 | ABeta42 53.87894737 |
tTau 147.5652632 | pTau 11.48947368 | Death Age 94 | Thal Score 4
H21.33.046 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 2.421052632 | ABeta42 19.19578947
| tTau 1124.777383 | pTau 3.129473684 | Death Age 97 | Thal Score 4
H21.33.047 | sex: Male | APOE: 3/3 (Non-carrier) | ABeta40 0.000981053 | ABeta42 0.049052632
| tTau 212.9031579 | pTau 3.575789474 | Death Age 90 | Thal Score 2
```

In [49]:

```python
# Bar Graph

if __name__ == "__main__":
    Patient.instantiate_from_csv(
        r"C:\Users\Annie\Downloads\comp bme\module 1\UpdatedLuminex.csv",
        r"C:\Users\Annie\Downloads\comp bme\module 1\UpdatedMetaData.csv"
    )

  # Build DataFrame from Patient objects
    data = []
    for p in Patient.all_patients:
```

```
        data.append({
            "PatientID": p.DonorID,
            "Sex": p.sex,
            "APOE Genotype": p.apoe_genotype,
            "APOE Carrier Status": p.apoe_e4_status,  # ✅ new col
            "ABeta40": p.ABeta40,
            "ABeta42": p.ABeta42,
            "tTau": p.tTau,
            "pTau": p.pTau,
            "Death Age": p.death_age,
            "Thal Score": p.thal_score
        })
    df = pd.DataFrame(data)


# Bar plots for group averages
    biomarkers = ["ABeta40", "ABeta42", "tTau", "pTau"]
    group_means = df.groupby("APOE Carrier Status")[biomarkers].mean()

    group_means.plot(kind="bar", figsize=(8,5))
    plt.title("Average Biomarker Levels by APOE-ε4 Status")
    plt.ylabel("Mean (pg/ug)")
    plt.xticks(rotation=0)
    plt.legend(title="Biomarker")
    plt.tight_layout()
    plt.show()
```



In [50]:
```
# Anova Tests

import math

def one_way_anova_groups(df, biomarkers, group_col="APOE Carrier Status", alpha=0.05):
    for biomarker in biomarkers:
        # collect the two groups
```

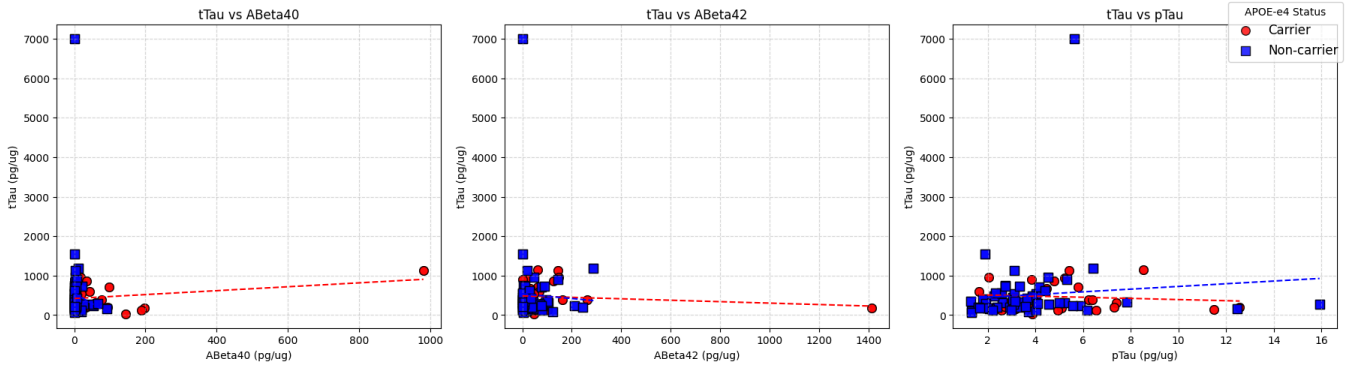
```
        groups = [list(group[biomarker].dropna()) for name, group in df.groupby(group_col)]

        if len(groups) != 2 or any(len(g) < 2 for g in groups):
            print(f"{biomarker}: Not enough data for ANOVA")
            continue

        group1, group2 = groups

        # sample sizes
        n1, n2 = len(group1), len(group2)
        N = n1 + n2

        # means
        mean1 = sum(group1)/n1
        mean2 = sum(group2)/n2
        overall_mean = (sum(group1)+sum(group2))/N

        # between-group sum of squares
        ss_between = n1*(mean1-overall_mean)**2 + n2*(mean2-overall_mean)**2

        # within-group sum of squares
        ss_within = sum((x-mean1)**2 for x in group1) + sum((x-mean2)**2 for x in group2)

        # degrees of freedom
        df_between = 1
        df_within = N - 2

        # mean squares
        ms_between = ss_between / df_between
        ms_within = ss_within / df_within

        # F-statistic
        F = ms_between / ms_within

        # Convert F to t (2 groups only)
        t = math.sqrt(F)

        # Approximate p-value using normal CDF
        def normal_cdf(x):
            return (1 + math.erf(x / math.sqrt(2))) / 2
        p = 2*(1 - normal_cdf(abs(t)))  # two-tailed

        # Print results
        print(f"\nOne-way ANOVA for {biomarker} by {group_col}")
        print(f"F({df_between}, {df_within}) = {F:.3f}, p ≈ {p:.3f}")
        if p < alpha:
            print(f"→ Statistically significant at α = {alpha}")
        else:
            print(f"→ Not statistically significant at α = {alpha}")


# Example usage:
biomarkers = ["ABeta40", "ABeta42", "tTau", "pTau"]
one_way_anova_groups(df, biomarkers)
```

```
One-way ANOVA for ABeta40 by APOE Carrier Status
F(1, 166) = 18.332, p ≈ 0.000
→ Statistically significant at α = 0.05

One-way ANOVA for ABeta42 by APOE Carrier Status
F(1, 166) = 11.173, p ≈ 0.001
→ Statistically significant at α = 0.05

One-way ANOVA for tTau by APOE Carrier Status
F(1, 166) = 0.146, p ≈ 0.702
→ Not statistically significant at α = 0.05

One-way ANOVA for pTau by APOE Carrier Status
F(1, 166) = 12.295, p ≈ 0.000
→ Statistically significant at α = 0.05
```

In [51]:

```python
# Scatter Plot and R^2

import matplotlib.pyplot as plt
import numpy as np

# Use APOE Carrier Status directly
df["APOE_group"] = df["APOE Carrier Status"]

# Define colors and markers
group_styles = {
    "Carrier": {"color": "red", "marker": "o"},
    "Non-carrier": {"color": "blue", "marker": "s"}
}

# Only the meaningful biomarkers to plot on X-axis
x_biomarkers = ["ABeta40", "ABeta42", "pTau"]  # remove tTau vs tTau

# Y-axis biomarker
y_biomarker = "tTau"

# Create 1x3 subplot
fig, axes = plt.subplots(1, 3, figsize=(18,5))
axes = axes.flatten()

# Store R² values
r2_results = {}

for i, x_biomarker in enumerate(x_biomarkers):
    ax = axes[i]
    r2_results[x_biomarker] = {}

    for status in ["Carrier", "Non-carrier"]:
        subset = df[df["APOE_group"] == status]
        if subset.empty or subset[x_biomarker].isnull().all() or subset[y_biomarker].isnull(
            continue

        x = subset[x_biomarker].values
        y = subset[y_biomarker].values

        # Scatter plot
        ax.scatter(
            x, y,
            color=group_styles[status]["color"],
            marker=group_styles[status]["marker"],
```

```
            edgecolor='black',
            s=70,
            alpha=0.8,
            label=status
        )

        # Linear regression
        coeffs = np.polyfit(x, y, 1)  # slope and intercept
        x_sorted = np.sort(x)
        y_fit = np.polyval(coeffs, x_sorted)
        ax.plot(x_sorted, y_fit,
                color=group_styles[status]["color"], linestyle='--')

        # Calculate R²
        ss_res = np.sum((y - np.polyval(coeffs, x)) ** 2)
        ss_tot = np.sum((y - np.mean(y)) ** 2)
        r2 = 1 - ss_res/ss_tot
        r2_results[x_biomarker][status] = r2

    ax.set_xlabel(f"{x_biomarker} (pg/ug)")
    ax.set_ylabel(f"{y_biomarker} (pg/ug)")
    ax.set_title(f"{y_biomarker} vs {x_biomarker}")
    ax.grid(True, linestyle='--', alpha=0.5)

# One legend for all subplots
handles, labels = ax.get_legend_handles_labels()
fig.legend(handles, labels, title="APOE-e4 Status", loc='upper right', fontsize=12)

plt.tight_layout()
plt.show()

# Print R² values
for biomarker, groups in r2_results.items():
    for status, r2 in groups.items():
        print(f"R² for {status} in {biomarker} vs {y_biomarker}: {r2:.3f}")
```




```
R² for Carrier in ABeta40 vs tTau: 0.076
R² for Non-carrier in ABeta40 vs tTau: 0.010
R² for Carrier in ABeta42 vs tTau: 0.020
R² for Non-carrier in ABeta42 vs tTau: 0.001
R² for Carrier in pTau vs tTau: 0.013
R² for Non-carrier in pTau vs tTau: 0.008
```

Verify and validate your analysis:

To verify and validate our analysis, we first made sure our code properly organized the data and clearly separated APOE-ε4 carriers from non-carriers. We then compared the dementia incidence we observed

with findings from previous studies. According to the article, "ApoE4, the strongest genetic risk factor for Alzheimer's disease (AD), has been shown to be associated with both beta-amyloid (Aβ) and tau pathology, with the strongest evidence for effects on Aβ" (Benson et al.), which matches our observation that carriers had a higher incidence of dementia. The paper also notes that "ApoE4 impairs Aβ clearance and accelerates Aβ aggregation leading to enhanced amyloid pathology and neuritic dystrophy" (Liu et al.), which helps explain the biological mechanism behind the increased risk. Overall, comparing our results with these established findings gives us confidence that our analysis is both accurate and consistent with existing research.

Benson, Gloria S., et al. "Don't Forget about Tau: The Effects of ApoE4 Genotype on Alzheimer's Disease Cerebrospinal Fluid Biomarkers in Subjects with Mild Cognitive Impairment—Data from the Dementia Competence Network." Journal of Neural Transmission, 21 Jan. 2022, https://doi.org/10.1007/s00702-022-02461-0.

Liu, Chia-Chen, et al. "ApoE4 Accelerates Early Seeding of Amyloid Pathology." Neuron, vol. 96, no. 5, Dec. 2017, pp. 1024-1032.e3, https://doi.org/10.1016/j.neuron.2017.11.013.

Conclusions and Ethical Implications:

Based on our analysis, we determined there is a strong connection between APOE-ε4 status and dementia risk. By comparing carriers and non-carriers using bar graphs, scatter plots, and statistical tests, we observed that individuals with at least one APOE-ε4 allele showed a higher rate of dementia in our dataset. These results align with published research, such as the study from PubMed Central, which confirms APOE-ε4 as the strongest genetic risk factor for Alzheimer's disease, particularly affecting beta-amyloid accumulation. The visualizations clearly show the differences between the groups, and the statistical tests support that these differences are unlikely to be due to chance. Ethically, these findings underscore the importance of careful handling of genetic information. While knowing a patient's APOE-ε4 status can help with early monitoring, personalized interventions, and research into therapies, it also raises potential concerns regarding privacy, anxiety, and possible discrimination. Communicating genetic risk requires informed consent, confidentiality, and clear guidance so individuals can make decisions about testing and care without undue stress or misunderstanding.

Limitations and Future Work:

Although our analysis demonstrates a clear association between APOE-ε4 and dementia risk, there are limitations to consider. Our dataset is limited in size and demographic diversity, which may restrict how well these findings generalize to the broader population. Additionally, while the statistical tests show significant differences between carriers and non-carriers, our data are observational and cannot establish causation. Other factors, such as additional genetic variants, environmental exposures, and lifestyle factors, may also influence dementia risk but are not captured in this dataset. Future work could address these limitations by including larger and more diverse participant cohorts, incorporating longitudinal data to track cognitive changes over time, and integrating additional biomarkers such as tau levels or neuroimaging data. These enhancements would allow for a deeper understanding of the

mechanisms behind APOE-ε4–associated dementia risk and strengthen the robustness of conclusions drawn from the analysis.

Notes from your team:

- We worked on filling out the disease background bullet points to better understand Alzheimer's disease.
- We reviewed and organized the datasets we will be using for the project.
- We brainstormed several potential questions for our project and recorded them above.
- After examining the datasets, we decided on the specific focus for our project.
- We've been working on the code for the graphs, ttest data, and more to analyze our question.
- We have, however, experianced a ton of issues and complications during the coding process.

Questions for our TA:

- Based on the questions we considered, do you think the one we chose is the most suitable for our project?
- Currently, we were only able to get the code to work on one of our computers. However, while finishing up the code for the notebook today(9/19), the code will no longer run even though it hasn't changed. The code above in the data analysis section should work, as it did prior to today, and I have visual examples of the data the code provided us (including the graph and ttest). with this in mind, we were wondering if you would be able to tell if the code works for you and if you can see our results so far.