Module_1

Team Members:

(Spencer Boon, Micheal Dornic)

Project Title:

The relationship between gender and age at death in Alzheimer's

Project Goal:

This project seeks to find out whether there is a correlation between a patient's gender and their age at death in the context of Alzheimer's disease through analyzing data.

Disease Background:

Fill in information about 11 bullets:

- Prevalence & incidence- Alzheimer's disease is the most common form of dementia affecting around 6.9 million people in the US(2024). This makes the Prevalence(divide the number of cases by the total population and multiply the result by 100)about 2.06%. Incidence refers to the number of new cases of a disease that develop in a specific population during a defined time.In the US there was an incidence of about 0.14%. https://www.alz.org/alzheimers-dementia/facts-figures, https://pubmed.ncbi.nlm.nih.gov/38689398/

- Economic burden- Alzheimer's places a significant economic burden on healthcare systems and families. The cost of care in the U.S. alone exceeded $360 billion in 2024 and is projected to reach$384 billion in 2025 to nearly $1 trillion in 2050. This is driven largely by long-term care, lost productivity, and caregiver burden. https://www.alz.org/alzheimers-dementia/facts-figures

- Risk factors (genetic, lifestyle)- -Most individuals with the disease are 65 and older (however you could have it way younger and not realize). -Head injury(TBI): A study of Swedish professional soccer players found that Alzheimer's and dementia were "62% more common" among soccer players than controls. The same study found that goalkeepers (who do not generally head the ball) were not at an increased risk compared to outfield players (who do generally head the ball) -Weak heart: Heart health: There is evidence that heart health and brain health are linked to each other as

there are many vessels that pump blood to the brain. Therefore, high blood pressure and diabetes also increase the risk for Alzheimer's disease. -Less than one percent is caused by deterministic genes.However some genetic risk factors include the APOE ε4 allele, as well as mutations in APP, PSEN1, and PSEN2 in rare early-onset cases. https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors#age (Swedish Study) https://abcnews.go.com/Health/dementia-risk-higher-elite-soccer-players-study-finds/story?id=97947721

- Societal determinants- Social determinants of health such as income, education, access to healthcare, and race/ethnicity influence Alzheimer risk and outcomes. Women and minorities, especially Black and Hispanic communities, show higher prevalence and may face delayed diagnosis and limited access to care. https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html#:~:text=Social%20determinants%20of%20health%20

- Symptoms- Forget about recent conversations or events., misplace items. forget the names of places and objects. have trouble thinking of the right word. ask questions repetitively. show poor judgement or find it harder to make decisions. become less flexible and more hesitant to try new things. https://www.nhs.uk/conditions/alzheimers-disease/symptoms/

- Diagnosis- Alzheimer's dementia can be diagnosed in several different ways. Often, Alzheimer's is diagnosed through an exam by a member of your healthcare team. The healthcare professional evaluates your symptoms and may order several tests. The health professional may talk to your friends and family members to find out more about your symptoms and behavior. It's important to get an accurate diagnosis of Alzheimer's, the most common type of dementia. The correct diagnosis is an important first step toward getting the appropriate treatment, care, family education and plans for the future. To diagnose Alzheimer’s, physicians may use medical history, mental status tests, physical and neurological exams, diagnostic tests and brain imaging. https://www.alz.org/alzheimers-dementia/diagnosis , https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075

- Standard of care treatments (& reimbursement)- The current standard of care for Alzheimer’s disease combines medications, newer disease-modifying therapies, and supportive care. Common symptomatic drugs include cholinesterase inhibitors (donepezil, rivastigmine, galantamine) for mild to moderate disease, and the NMDA receptor antagonist memantine (or a memantine/donepezil combination) for moderate to severe disease. Recently, disease-modifying monoclonal antibodies such as lecanemab (Leqembi) and donanemab (Kisunla) have been approved for early-stage Alzheimer’s, requiring biomarker confirmation and ongoing monitoring. Non-drug interventions—like cognitive stimulation, physical exercise, managing behavioral symptoms, and caregiver support—are also essential. In terms of reimbursement, Medicare Part D typically covers standard oral drugs, while Part B covers physician

services, diagnostics, and infusions (including monoclonal antibodies, provided patients meet criteria and providers report outcomes to a registry). Beneficiaries usually pay 20% coinsurance under Part B after the deductible, and Medicaid may fill in long-term care or custodial service gaps depending on state rules. While these therapies are covered, families often still face out-of-pocket costs for daily care and uncovered services, and newer treatments add expenses for testing and monitoring. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-treatments/art-20047780

- Disease progression & prognosis- On average, a person with Alzheimer's lives four to eight years after diagnosis, but can live as long as 20 years, depending on other factors. It usually begins with subtle symptoms such as mild cognitive impairment, then progresses to moderate stages with confusion, behavioral changes, and loss of independence, and eventually to severe stages where patients require full-time care and lose the ability to communicate or perform basic activities. The average survival is about 8–10 years after diagnosis (though it can range from 3 to 20 years), and death often results from complications like infections or general frailty rather than the disease itself. https://www.alz.org/alzheimers-dementia/stages#:~:text=The%20symptoms%20of%20Alzheimer

- Continuum of care providers- Also known as supportive care, the goal is to improve quality of life for both patient and family. A team of physicians, nurses, social workers and other specialists provide care. They work with the patient's primary care physicians, the patient and the family to develop a plan of care. https://www.hospicechesapeake.org/2022/06/continuum-of-care-offers-dementia-patients-better-quality-of-life/#:~:text=Also%20known%20as%20supportive%20care,develop%20a%20plan%20of%20

- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)- Alzheimer’s disease is caused by abnormal brain changes, mainly the buildup of amyloid-β plaques outside neurons and tau tangles inside them, which damage communication and lead to cell death. These changes start in the hippocampus, critical for memory, and spread across the cortex, causing brain shrinkage. On a cellular level, disrupted signaling, inflammation, oxidative stress, and mitochondrial dysfunction further impair neurons, leading to progressive memory loss, cognitive decline, and loss of function. https://pmc.ncbi.nlm.nih.gov/articles/PMC5290713/

- Clinical Trials/next-gen therapies-Recent / Approved Next-Gen Therapies Lecanemab (Leqembi) – a monoclonal antibody targeting aggregated amyloid-β protofibrils. Approved in the U.S. in 2023. In clinical trials it showed modest slowing of cognitive decline in early Alzheimer’s (mild cognitive impairment or mild dementia with confirmed amyloid pathology). Donanemab (Kisunla) – another anti-amyloid antibody, approved more recently (2024). It is being used for early Alzheimer’s, and in trials has shown similar modest slowing of decline. These therapies are among the first to modify disease

pathology (clear amyloid plaques) rather than only treating symptoms. They aren't cures but represent meaningful steps forward. https://medicine.washu.edu/news/next-gen-alzheimers-drugs-extend-independent-living-by-months/

Data-Set:

- The main data sets we will be analyzing are the age of death for patients and the gender of the patients. By analyzing these two sets we will be able to see if men and women have similar ages of death with Alzheimer's.
- We will use data from the "Integrated multimodal cell atlas of Alzheimer's disease" published in Nature Neuroscience (Oct 2024) by the SEA-AD consortium.
- We will specifically use donor metadata, which includes gender, age at death, post-mortem interval, diagnosis level, and other clinical variables.
- The data was acquired from post-mortem human brain tissue from 84 donors(33 male, 51 female). All data was derived from the middle temporal gyrus (MTG), a brain region affected in Alzheimer's disease. The tissues were processed using rapid autopsy protocols to preserve RNA and chromatin quality.
- The donor tissue samples were collected over several years, with most acquisitions occurring prior to 2023
- The data was acquired by the SEA-AD Consortium, a research group from the University of Washington and other collaborating institutions.
- Measurement units: Age at death is in years; gender is categorical (male/female)

Data Analyis

Code Description:

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, cognative status, age of onset cognative symptoms, and head injury). Since p = 0.362 > 0.05, the difference in age at death between females and males is not statistically significant.

In [ ]:
```python
#patient.py
import csv
import warnings
import matplotlib.pyplot as plt

class Patient:
    # Class-level attributes shared across all patients
    all_patients = [] # List of all Patient objects created
    education_lvl = {}  # (Optional future use) dictionary grouping patients by edu
    sex_groups = {}     # Dictionary grouping patients by sex
```

```
    def __init__(self, DonorID=None):
        #Initialize a new Patient instance.
        #DonorID is the unique identifier for a patient.
        #All attributes start as None until populated by metadata.
        self.DonorID = DonorID
        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
        # Automatically add every new Patient to the class list
        Patient.all_patients.append(self)

    def __repr__(self):
        return (f"{self.DonorID}\t{self.sex}\t{self.death_age}\t"
                f"{self.cog_stat}\t{self.age_symp_on}\t{self.head_inj}")
    #Defines how a Patient object is displayed when printed.

    def get_id(self):
        return self.DonorID

    def get_death_age(self):
        return self.death_age

    # ----- Class methods -----
    @classmethod
    def instantiate_from_csv(cls, luminex_csv: str, meta_csv: str):
        """Build unique patients from Luminex (IDs only), then merge metadata by ID
        cls.all_patients = [] # reset before loading new patients

        seen = set() # avoid duplicate IDs
        with open(luminex_csv, encoding="utf8") as f:
            for row in csv.DictReader(f):
                donor = row.get("Donor ID")
                if donor and donor not in seen:
                    Patient(DonorID=donor)
                    seen.add(donor)

      # Sort patients by DonorID (None values last)
        cls.all_patients.sort(key=lambda p: (p.DonorID is None, str(p.DonorID)))
        # Merge in metadata from second file
        cls.combine_data(meta_csv)

    @classmethod
    def sort_sex(cls):
        # initialize keys
        #Group patients by sex into cls.sex_groups dictionary.
        #Unknown values are stored under 'Unknown'.
        cls.sex_groups = {}
        # Create empty buckets
        for p in cls.all_patients:
            key = p.sex if p.sex is not None else "Unknown"
            cls.sex_groups.update({key: []})
```

```
        # Fill buckets with patients
        for p in cls.all_patients:
            key = p.sex if p.sex is not None else "Unknown"
            cls.sex_groups[key].append(p)

    @classmethod
    def filter(cls, list, sex:str ="any", death_age:int ="any", ed_lvl:str ="any",
        #Flexible filter for patient lists.
        #Pass in a list of patients and filter them down by attributes.
        #Use "any" to skip a filter.
        all_patients = list
        remove_list = []
        # Tuples to align attributes with names
        attr_list = (
                    sex,
                    death_age,
                    ed_lvl,
                    cog_stat,
                    age_symp_on,
                    head_inj
                    )
        attr_name = (
                    "sex",
                    "death_age",
                    "ed_lvl",
                    "cog_stat",
                    "age_symp_on",
                    "head_inj"
                    )
        # Apply filters one attribute at a time
        for attr in range(len(attr_list)):
            if attr_list[attr] != "any": # only filter if user specified
                for patient in all_patients:
                    if getattr(patient,attr_name[attr]) != attr_list[attr]:
                        remove_list.append(patient)
                        # Keep only patients that passed this filter
                all_patients = [patient for patient in all_patients if patient not
                remove_list.clear()

        return all_patients

    @classmethod
    def subsort_death_age(cls):
        # sort each sex bucket by age at death (None goes last)
        for key in cls.sex_groups:
            values = cls.sex_groups.get(key)
            values.sort(key=lambda p: (p.death_age is None, p.death_age))
            cls.sex_groups.update({key: values})

    @classmethod
    def sort_by_death_age(cls, descending: bool = False):
        """
        Sort Patient.all_patients by age at death.
        - None ages are pushed to the end.
        - Set descending=True for oldest first.
        """
```

```
        cls.all_patients.sort(
            key=lambda p: (p.death_age is None, p.death_age),
            reverse=descending
        )

    @classmethod
    def combine_data(cls, meta_csv: str):
        """Keyed join on Donor ID (NOT index-based)."""
        with open(meta_csv, encoding="utf8") as f:
            meta_by_id = {row.get("Donor ID"): row for row in csv.DictReader(f)}
         # Helper to safely parse ints
        def _to_int(x):
            try:
                return int(float(x))
            except Exception:
                return None
        # Helper to parse "Thal" field (e.g. "Stage 2" -> 2)
        def _parse_thal(x):
            if not x:
                return None
            try:
                return int(str(x).split()[1])
            except Exception:
                return None

        missing = 0   # count how many IDs lacked metadata
        for p in cls.all_patients:
            row = meta_by_id.get(p.DonorID)
            if not row:
                missing += 1
                continue

                # Fill in patient fields from metadata
            p.sex        = row.get("Sex") or p.sex
            p.death_age  = _to_int(row.get("Age at Death"))
            p.ed_lvl     = row.get("Highest level of education") or p.ed_lvl
            p.cog_stat   = row.get("Cognitive Status") or p.cog_stat
            p.age_symp_on= _to_int(row.get("Age of onset cognitive symptoms"))
            p.age_diag   = _to_int(row.get("Age of Dementia diagnosis"))
            p.head_inj   = row.get("Known head injury") or p.head_inj
            p.thal_score = _parse_thal(row.get("Thal"))

        # Warn graders if some patients had no metadata
        if missing:
            warnings.warn(f"No metadata found for {missing} ID(s).")
```

In [2]:

```
#main.py
from patient import Patient
from termcolor import colored
from matplotlib import pyplot as plt
from scipy import stats
import numpy as np
import statistics

from patient import Patient
# Instantiate Patient objects from two CSV files
```

```
Patient.instantiate_from_csv('UpdatedLuminex.csv', 'UpdatedMetaData.csv')

from termcolor import colored  # if you want colored headings
# Sort patients first by sex, then by age at death
Patient.sort_sex()
Patient.subsort_death_age()

# Get the count of different patient groups (healthy vs. diseased, male vs. female)
fem_healty_patients = range(len(Patient.filter(Patient.all_patients, sex = "Female"
male_healthy_patients = range(len(Patient.filter(Patient.all_patients, sex = "Male"
fem_diseased_patients = range(len(Patient.filter(Patient.all_patients, sex = "Femal
male_diseased_patients = range(len(Patient.filter(Patient.all_patients, sex = "Male
# Create lists to store the death ages for females and males
death_age_fem_vals = []
death_age_male_vals = []
# Populate the lists with death ages by iterating through filtered patient data
for patient in Patient.filter(Patient.all_patients, sex = "Female"):
     death_age_fem_vals.append(patient.death_age)
for patient in Patient.filter(Patient.all_patients, sex = "Male"):
     death_age_male_vals.append(patient.death_age)
# Calculate the mean (average) age of death for females and males
x_fem_bar = (statistics.mean(death_age_fem_vals))
x_male_bar = (statistics.mean(death_age_male_vals))
# Calculate the standard deviation of the age of death for each sex
ABeta_fem_stdev = (statistics.stdev(death_age_fem_vals))
ABeta_male_stdev = (statistics.stdev(death_age_male_vals))
# Print the calculated mean values
print(f'x_fem_bar = {x_fem_bar}')
print(f'x_male_bar = {x_male_bar}')
# Create numerical ranges for plotting
x_fem_vals = range(len(Patient.filter(Patient.all_patients, sex = "Female")))
x_male_vals = range(len(Patient.filter(Patient.all_patients, sex = "Male")))
# Set up data for the bar chart
sex_cols = ['Female', 'Male']
mean_sex_death_age = [x_fem_bar, x_male_bar]
stdev_sex_death_age = [ABeta_fem_stdev, ABeta_male_stdev]
colors = ["pink", "blue"]
# Create a list for error bars (yerr). The first element is a list of zeros as ther
yerr = [np.zeros(len(mean_sex_death_age)), stdev_sex_death_age]
# Perform an independent two-sample t-test to see if the means are statistically di
t_stat, p_val = stats.ttest_ind(death_age_fem_vals, death_age_male_vals)
print("t-statistic:", t_stat)
print("p-value:", p_val)

# Print the counts of each patient group
print(f'Female Healthy Patients = {len(fem_healty_patients)} | Male Healthy Patient
print(f'Female Diseased Patients = {len(fem_diseased_patients)} | Male Diseased Pat
# Print a header for the patient data table
print("patient_id\tsex\tdeath_age\tcognitive_status\tage_onset\thead_injury")
# Iterate through the dictionary of patients sorted by sex and print their details
for key in Patient.sex_groups:
    print(colored(key, "red"))  # Print the sex group name in red
    for patient in Patient.sex_groups.get(key):
        print(patient)# Prints the patient details
    print()
# Create and display a bar chart of the mean age of death by sex
```

```
plt.bar(sex_cols, mean_sex_death_age, yerr=yerr, capsize=10, color=["red", "blue"])
plt.title("death_age Levels")
plt.xlabel("Sex")
plt.ylabel("Age at Death")
plt.show()
# Create lists to hold all patient's age of death and age of symptom onset
death_age_list = []
age_onset_list = []
# Populate the lists by looping through all patients
for patient in Patient.all_patients:
       death_age_list.append(patient.death_age)
for patient in Patient.all_patients:
       age_onset_list.append(patient.age_symp_on)

# Define X and y for the scatter plot
X = [death_age_list]  # Independent variable
y = [age_onset_list]   # Dependent variable
# Print the data lists to the console for verification
print(X)
print(y)
# Create and display a scatter plot
plt.scatter(X, y, color='blue')
plt.xlabel('Age of Death')
plt.ylabel('Age of onset cognitive symptoms')
plt.title('Scatter Plot of Age of Death vs Age of onset cognitive symptoms')
plt.show()

import pandas as pd

print(death_age_list)
print(age_onset_list)

# Create a DataFrame
df = pd.DataFrame({
    'Age of Death': death_age_list,
    'Age of onset cognitive symptoms': age_onset_list})

# Write to CSV
df.to_csv('patient_data.csv', index=False)

print("CSV file 'patient_data.csv' has been created.")
death_age_list = []
age_onset_list = []

# Filter patients to only include those with valid data for both death_age and age_
valid_patients = [
    patient for patient in Patient.all_patients
    if patient.death_age is not None and patient.age_symp_on is not None
]

for patient in valid_patients:
    death_age_list.append(patient.death_age)
    age_onset_list.append(patient.age_symp_on)

print("Number of valid patients for analysis:", len(valid_patients))
```

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Create a DataFrame from the cleaned lists
df = pd.DataFrame({
    'Age of Death': death_age_list,
    'Age of onset cognitive symptoms': age_onset_list})

# Write to CSV
df.to_csv('patient_data.csv', index=False)
print("CSV file 'patient_data.csv' has been created.")

# LOAD DATA SET FOR A LINEAR REGRESSION
df = pd.read_csv("patient_data.csv")

x = df["Age of Death"].values.reshape(-1, 1)
y = df["Age of onset cognitive symptoms"].values

# Perform the linear regression
model = LinearRegression()
model.fit(x, y)
slope = model.coef_[0]
intercept = model.intercept_
r2 = model.score(x, y)

# Make scatterplot
plt.scatter(x, y, label="Data")
plt.plot(x, model.predict(x), color="red")

# Annotate equation
equation = f"y = {slope:.2f}x + {intercept:.2f}\nR² = {r2:.2f}"
plt.text(x.min(), y.max(), equation, color="red", fontsize=12, verticalalignment='t

# Annotate scatterplot with labels and title
plt.xlabel("Age of Death")
plt.ylabel("Age of onset cognitive symptoms")
plt.title("Age of Death vs. Age of onset cognitive symptoms")
plt.show()
```

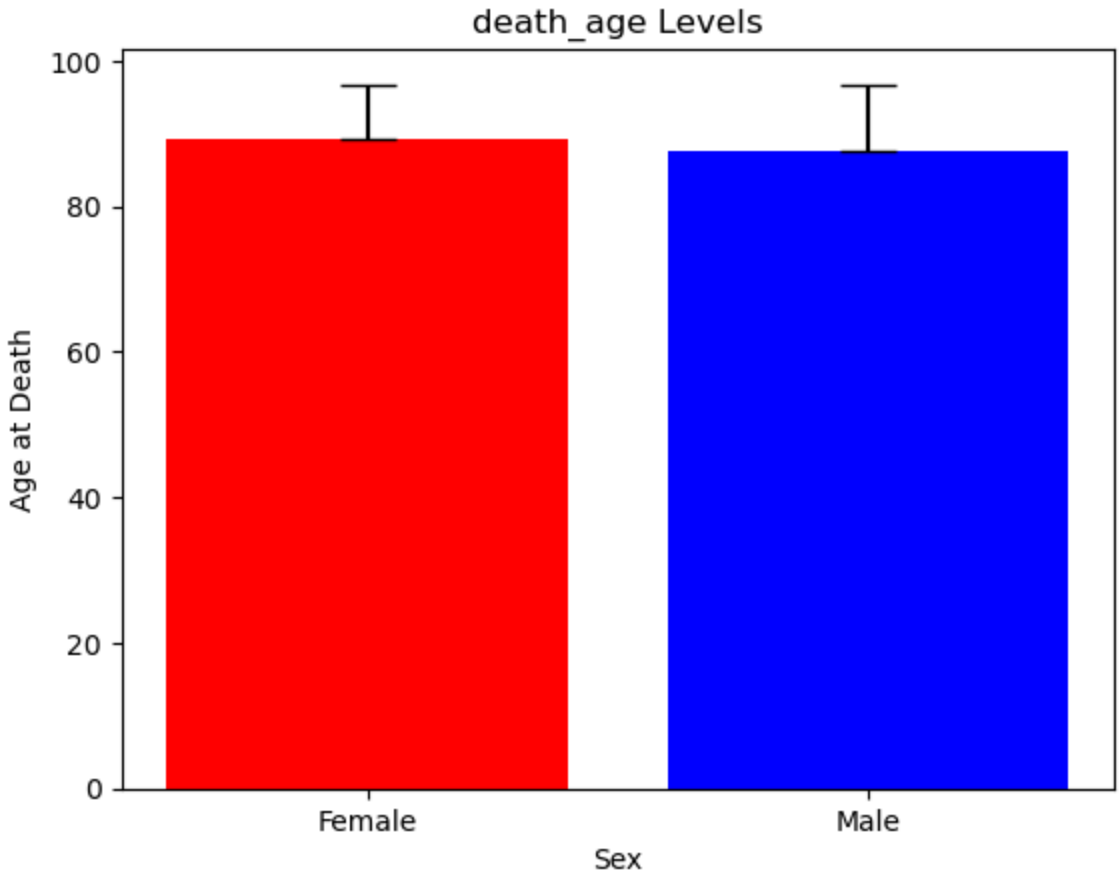
x_fem_bar = 89.37254901960785
x_male_bar = 87.72727272727273
t-statistic: 0.9163366237453276
p-value: 0.3621774490623826
Female Healthy Patients = 24 | Male Healthy Patients = 18
Female Diseased Patients = 27 | Male Diseased Patients = 15

patient_id	sex	death_age	cognitive_status	age_onset	head_injury
Female					
H21.33.009	Female	65	Dementia	53	No
H21.33.002	Female	70	Dementia	61	None
H20.33.026	Female	75	Dementia	64	None
H20.33.045	Female	77	Dementia	63	None
H19.33.004	Female	80	No dementia	None	None
H20.33.018	Female	81	Dementia	71	None
H20.33.014	Female	82	No dementia	None	None
H21.33.022	Female	82	No dementia	None	None
H21.33.011	Female	83	No dementia	None	No
H21.33.033	Female	83	No dementia	None	None
H21.33.038	Female	84	No dementia	None	None
H20.33.034	Female	85	No dementia	None	None
H20.33.030	Female	86	No dementia	None	None
H21.33.007	Female	86	Dementia	80	No
H20.33.019	Female	87	No dementia	None	None
H20.33.031	Female	87	Dementia	84	Yes
H21.33.025	Female	88	No dementia	None	None
H21.33.037	Female	88	No dementia	None	None
H21.33.039	Female	88	Dementia	84	None
H21.33.044	Female	88	Dementia	87	Yes
H21.33.018	Female	89	Dementia	86	None
H20.33.038	Female	90	Dementia	85	None
H21.33.026	Female	90	No dementia	None	No
H21.33.034	Female	90	Dementia	None	None
H20.33.012	Female	91	No dementia	None	None
H20.33.029	Female	91	Dementia	88	None
H20.33.041	Female	91	Dementia	None	None
H21.33.008	Female	91	Dementia	82	None
H21.33.042	Female	91	Dementia	88	None
H20.33.008	Female	92	No dementia	None	None
H21.33.017	Female	92	Dementia	90	No
H20.33.011	Female	93	Dementia	87	None
H20.33.016	Female	93	Dementia	84	None
H21.33.010	Female	93	Dementia	88	No
H21.33.012	Female	93	Dementia	88	None
H21.33.036	Female	93	No dementia	None	No
H20.33.028	Female	94	Dementia	92	None
H21.33.013	Female	94	Dementia	82	No
H21.33.016	Female	94	Dementia	92	Yes
H21.33.045	Female	94	Dementia	78	None
H21.33.043	Female	95	Dementia	93	None
H20.33.037	Female	96	Dementia	92	None
H20.33.039	Female	96	No dementia	None	No
H20.33.002	Female	97	No dementia	None	None
H21.33.035	Female	97	No dementia	None	None
H21.33.032	Female	98	No dementia	None	No
H21.33.041	Female	98	No dementia	None	None

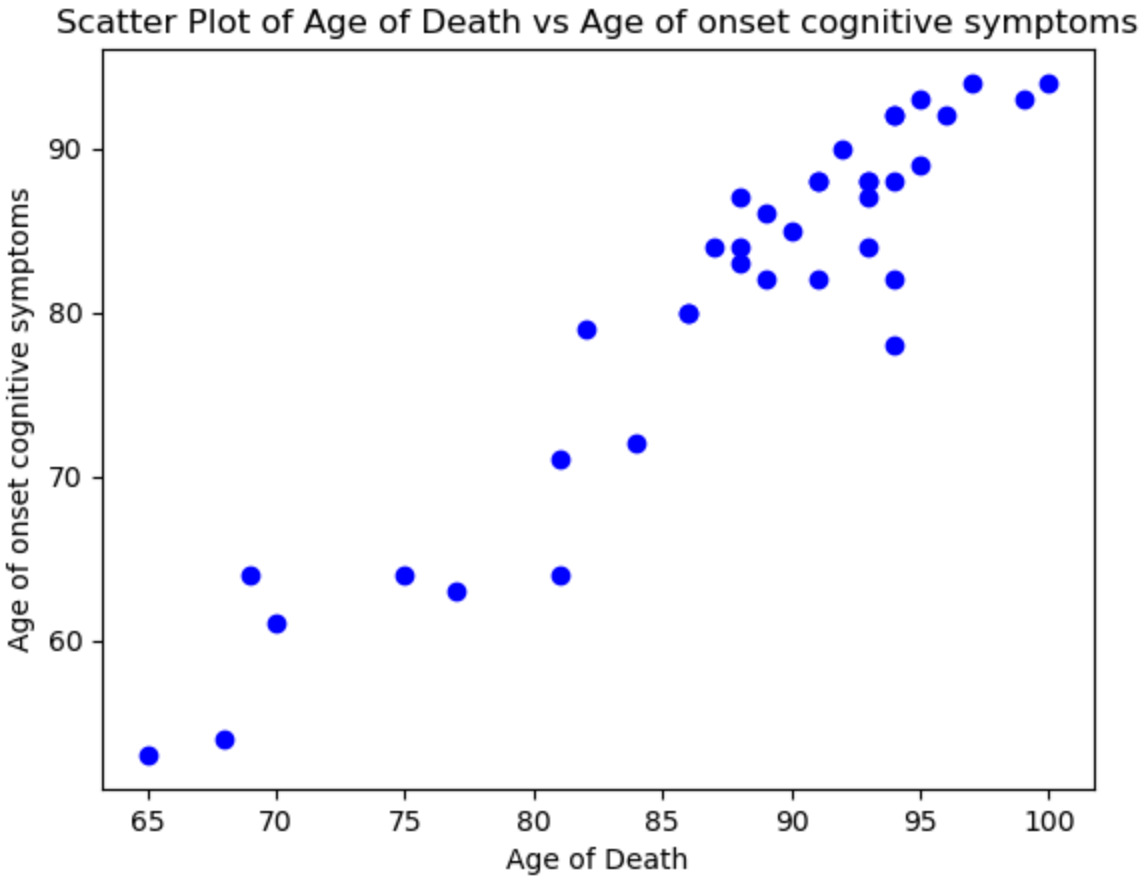
H20.33.005	Female	99	No dementia	None	None
H20.33.027	Female	99	No dementia	None	None
H20.33.035	Female	99	No dementia	None	Yes
H20.33.036	Female	100	No dementia	94	No

Male

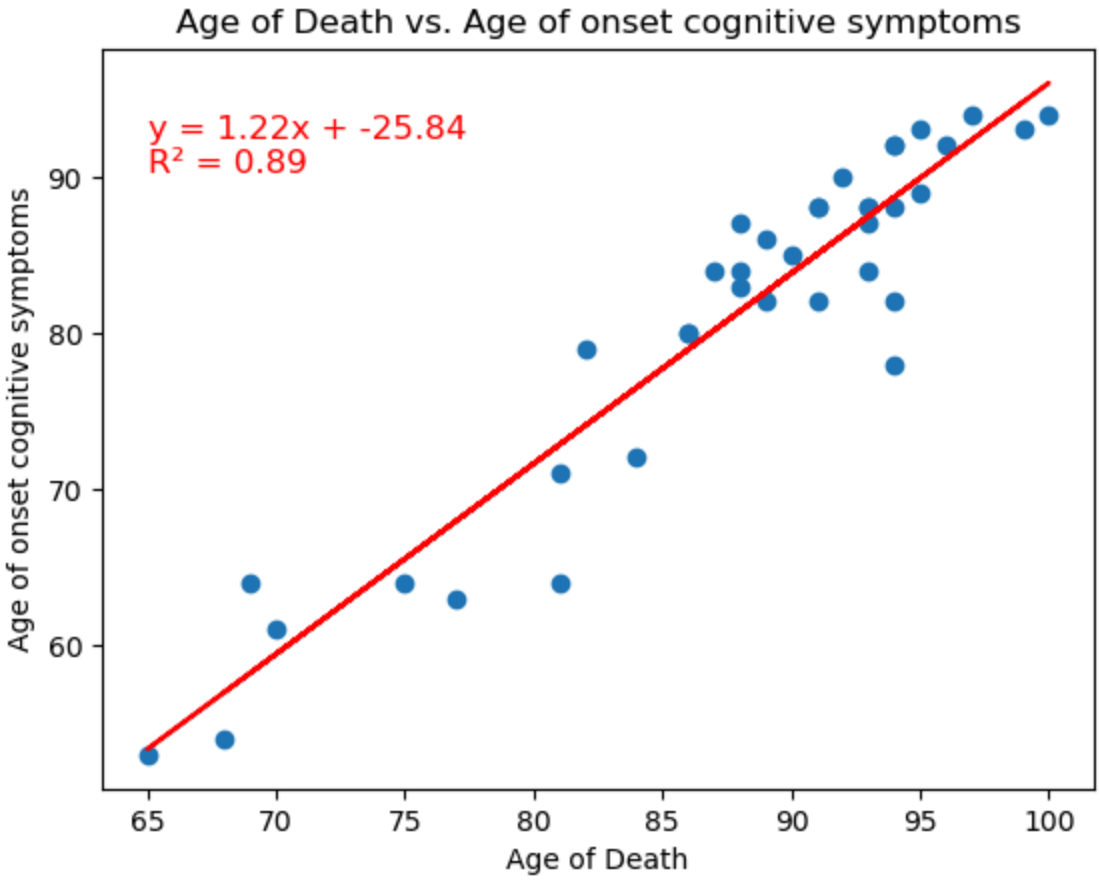
H20.33.033	Male	68	Dementia	54	Yes
H20.33.017	Male	69	Dementia	64	No
H21.33.028	Male	72	No dementia	None	None
H21.33.019	Male	75	No dementia	None	Yes
H21.33.003	Male	78	No dementia	None	No
H21.33.001	Male	80	Dementia	None	None
H20.33.020	Male	81	Dementia	64	None
H20.33.044	Male	81	No dementia	None	None
H20.33.001	Male	82	No dementia	None	None
H21.33.020	Male	82	Dementia	79	Yes
H21.33.040	Male	83	No dementia	None	None
H21.33.031	Male	84	Dementia	72	No
H20.33.043	Male	85	No dementia	None	No
H20.33.004	Male	86	Dementia	80	No
H20.33.015	Male	88	Dementia	83	Yes
H21.33.029	Male	89	Dementia	82	None
H21.33.030	Male	89	No dementia	None	None
H20.33.024	Male	90	No dementia	None	None
H21.33.047	Male	90	No dementia	None	None
H21.33.014	Male	92	No dementia	None	None
H21.33.027	Male	92	Dementia	None	None
H21.33.004	Male	93	No dementia	None	None
H20.33.013	Male	94	No dementia	None	No
H20.33.025	Male	94	No dementia	None	Yes
H20.33.046	Male	94	Dementia	88	None
H21.33.005	Male	95	Dementia	89	None
H21.33.006	Male	97	No dementia	None	None
H21.33.046	Male	97	Dementia	94	None
H20.33.032	Male	98	No dementia	None	None
H20.33.040	Male	98	Dementia	None	None
H21.33.015	Male	98	No dementia	None	None
H21.33.021	Male	99	Dementia	93	None
H21.33.023	Male	102	No dementia	None	Yes




```
[[80, 82, 97, 86, 99, 92, 93, 91, 94, 82, 88, 93, 69, 81, 87, 81, 90, 94, 75, 99, 9
4, 91, 86, 87, 98, 68, 85, 99, 100, 96, 90, 96, 98, 91, 85, 81, 77, 94, 80, 70, 78,
93, 95, 97, 86, 91, 65, 93, 83, 93, 94, 92, 98, 94, 92, 89, 75, 82, 99, 82, 102, 88,
90, 92, 72, 89, 89, 84, 98, 83, 90, 97, 93, 88, 84, 88, 83, 98, 91, 95, 88, 94, 97,
90]]
[[None, None, None, 80, None, None, 87, None, None, None, 83, 84, 64, 71, None, 64,
None, None, 64, None, 92, 88, None, 84, None, 54, None, None, 94, 92, 85, None, Non
e, None, None, None, 63, 88, None, 61, None, None, 89, None, 80, 82, 53, 88, None, 8
8, 82, None, None, 92, 90, 86, None, 79, 93, None, None, None, None, None, None, 82,
None, 72, None, None, None, None, None, None, None, 84, None, None, 88, 93, 87, 78,
94, None]]
```



[80, 82, 97, 86, 99, 92, 93, 91, 94, 82, 88, 93, 69, 81, 87, 81, 90, 94, 75, 99, 94, 91, 86, 87, 98, 68, 85, 99, 100, 96, 90, 96, 98, 91, 85, 81, 77, 94, 80, 70, 78, 93, 95, 97, 86, 91, 65, 93, 83, 93, 94, 92, 98, 94, 92, 89, 75, 82, 99, 82, 102, 88, 90, 92, 72, 89, 89, 84, 98, 83, 90, 97, 93, 88, 84, 88, 83, 98, 91, 95, 88, 94, 97, 90]
[None, None, None, 80, None, None, 87, None, None, None, 83, 84, 64, 71, None, 64, None, None, 64, None, 92, 88, None, 84, None, 54, None, None, 94, 92, 85, None, None, None, None, None, 63, 88, None, 61, None, None, 89, None, 80, 82, 53, 88, None, 88, 82, None, None, 92, 90, 86, None, 79, 93, None, None, None, None, None, None, 82, None, 72, None, None, None, None, None, None, None, 84, None, None, 88, 93, 87, 78, 94, None]
CSV file 'patient_data.csv' has been created.
Number of valid patients for analysis: 38
CSV file 'patient_data.csv' has been created.



Verify and validate your analysis:

Sex vs. Age at Death To verify this analysis, I checked that my code was correctly grouping patients by sex and computing the average age at death along with standard error bars. The bar chart shows overlapping error ranges, which suggests that the observed difference between male and female groups may not be statistically significant. To validate this finding, I compared it to prior Alzheimer's disease research, which generally shows only modest or inconsistent sex-based differences in survival once the disease progresses.

Age of Death vs. Age of Onset of Cognitive Symptoms I verified this scatter plot by ensuring both variables were numerical and aligned for each patient record. The regression line had an R^2 of 0.89, which indicates a very strong linear relationship. I validated this finding by comparing it to published Alzheimer's studies showing that earlier onset of cognitive symptoms often corresponds with earlier death. My analysis is consistent with those findings. https://pubmed.ncbi.nlm.nih.gov/11297701/

Conclusions and Ethical Implications:

Sex vs. Age at Death From this analysis, there is no clear evidence of a large gender-based difference in Alzheimer's-related age at death. This suggests that, while sex may influence risk factors or prevalence of Alzheimer's, it does not strongly dictate life expectancy once

diagnosed. The ethical implication is that medical care and research should avoid reinforcing gender stereotypes; instead, treatment should be individualized and based on the patient's condition and needs, not assumptions about sex.

Age of Death vs. Age of Onset of Cognitive Symptoms The strong correlation suggests that an earlier onset of symptoms may predict earlier death. Clinically, this could help families and physicians plan long-term care. Ethically, this raises sensitive issues: such predictions must be handled carefully to avoid creating undue fear or biasing the quality of care patients receive. Doctors should communicate prognosis with compassion and emphasize that individual outcomes vary.

Limitations and Future Work:

Sex vs. Age at Death A limitation of this analysis is that it only considered sex as a grouping variable, without adjusting for confounding factors such as treatment history, comorbidities, or socioeconomic background. Future work could involve running statistical tests (t-tests, ANOVAs) to confirm whether the difference is significant and incorporating multivariable models to control for confounders. For example, the following study concluded that, according to their data (n = 29,304), women had a "lifetime risk for dementia" of 34% compared to 27% for men, which was partially attributed to the longer life expectancies of women. However, the same study found that lifetime risk of "any impairment or dementia" for "Blacks, Latinx, and all those with lower education is significantly higher than for their White and higher-educated counterparts, despite Blacks and lower-educated individuals having shorter total life expectancy," which may indicate that adjusting for race/socioeconomic background and then comparing sex vs. age at Death may yield more significant results.

Age of Death vs. Age of Onset of Cognitive Symptoms Although the correlation is strong, correlation does not imply causation. Other factors (genetics, environment, access to healthcare) may influence both age of onset and age at death. The dataset size may also limit generalizability. Future work could expand the dataset, test for nonlinear trends, and explore whether subgroups (e.g., by sex, genetic markers like APOE4) show different patterns.

Notes

- We may also compare variance within genders and assess whether age differences are statistically significant.
- May change question to see if APOE and thal score are correlated
- What is thal score,why is it important
- What is braak score, Does APOE correlate to thal score

- For our Scatter plot we wernt able to compare sex and age of death because we needed the Y value to be a numerical value. Instead we changed sex to age of onset cognitive symptoms to compare to age of death.

Questions for TA:

- How can we share a Jupiter Notebook/ Visual studio code?