FINAL REPORT OF TRAINEESHIP PROGRAM- 2023

ON

" ANALYZING DEATH RATE OF RIGHT HANDERS
vs
LEFT HANDERS "

29/11/2023

ACKNOWLEDGEMENTS

"My traineeship at MedTourEasy was an invaluable experience that provided me with a comprehensive understanding of data visualizations and their role in data analytics. The opportunity to engage with experienced professionals and work on real-world projects was instrumental in my personal and professional growth. I am immensely grateful to the Training & Development Team for their unwavering support and guidance throughout the traineeship. Their dedication to my learning and development played a pivotal role in my ability to successfully complete the project and deliver meaningful insights to our clients”

TABLE OF CONTENT

METHODOLOGY

1. DATA COLLECTION
2. DATA CLEANING
3. EDA
4. DESIGNING DATA
5. STATISTICAL ANALYSIS

ABSTRACT

Hand dominance, commonly known as handedness, refers to the preference for using one hand over the other for daily tasks and activities. While the majority of the human population is right-handed, a significant minority, approximately 10%, is left-handed. This difference in handedness has been a subject of fascination and scientific inquiry for centuries, with researchers exploring the underlying neurological, genetic, and environmental factors that contribute to it.
One intriguing aspect of handedness research has been the investigation of its potential association with mortality.

Early studies suggested a link between left-handedness and a slightly shorter lifespan compared to right-handers. However, subsequent research has yielded inconsistent results, with some studies supporting the initial findings and others finding no significant difference in life expectancy between the two groups.

The potential reasons for these discrepancies are complex and multifaceted. Some hypotheses propose that the observed differences in mortality rates could be attributed to:

- Brain Lateralization: Left-handers tend to have a greater leftward asymmetry of the brain, which could make them more susceptible to certain neurological disorders that affect lifespan.
- Environmental Factors: Left-handers may be more likely to experience accidents or injuries due to living in a world designed for right-handers, potentially influencing mortality rates.
- Genetic Factors: Handedness is partially influenced by genetics, and some genes associated with handedness may also have a bearing on longevity.

About Company

MedTourEasy is a global healthcare company that provides patients with access to affordable, high-quality medical care around the world. The company partners with a network of internationally accredited healthcare providers and offers a variety of services and committed to providing patients with a positive and seamless experience. The company's team of experienced professionals is available to answer questions and provide support throughout the entire process.

MedTourEasy's accomplishments:

Helped thousands of patients get the care they need:

- MedTourEasy has helped thousands of patients get the care they need, often at a fraction of the cost they would have paid in their own country.
- Received numerous awards: MedTourEasy has received numerous awards for its commitment to patient care and innovation.
- Featured in major publications: MedTourEasy has been featured in major publications, such as The New York Times, The Wall Street Journal, and Forbes.

About Project

Handedness are the preference for using one hand over the other, is a fascinating aspect of human biology. This difference in handedness has been the subject of much research, with scientists exploring the underlying neurological and genetic factors that contribute to it. One area of particular interest has been the potential link between handedness and mortality.

In this project we will explore this phenomenon using age distribution data to see if we can reproduce a difference in average age at death purely from the changing rates of left-handedness over time, refuting the claim of early death for left-handers. This project uses pandas and Bayesian statistics to analyze the probability of being a certain age at death given that you are reported as left-handed or right-handed.

A National Geographic survey in 1986 resulted in over a million responses that included age, sex, and hand preference for throwing and writing. Researchers Avery Gilbert and Charles Wysocki analyzed this data and noticed that rates of left-handedness were around 13% for people younger than 40 but decreased with age to about 5% by the age of 80. They concluded based on analysis of a subgroup of people who throw left-handed but write right-handed that this age-dependence was primarily due to changing social acceptability of left-handedness. This means that the rates aren't a factor of age specifically but rather of the year you were born, and if the same study was done today, we should expect a shifted version of the same distribution as a function of age. Ultimately, we'll see what effect this changing rate has on the apparent mean age of death of left-handed people, but let's start by plotting the rates of left-handedness as a function of age.

JUPYTER

- Jupyter, formerly known as IPython, is an open-source web-based interactive computing environment that offers a user-friendly interface for executing code, creating visualizations, and writing narrative text. It has gained widespread popularity among data scientists, researchers, and educators due to its versatility and ease of use.

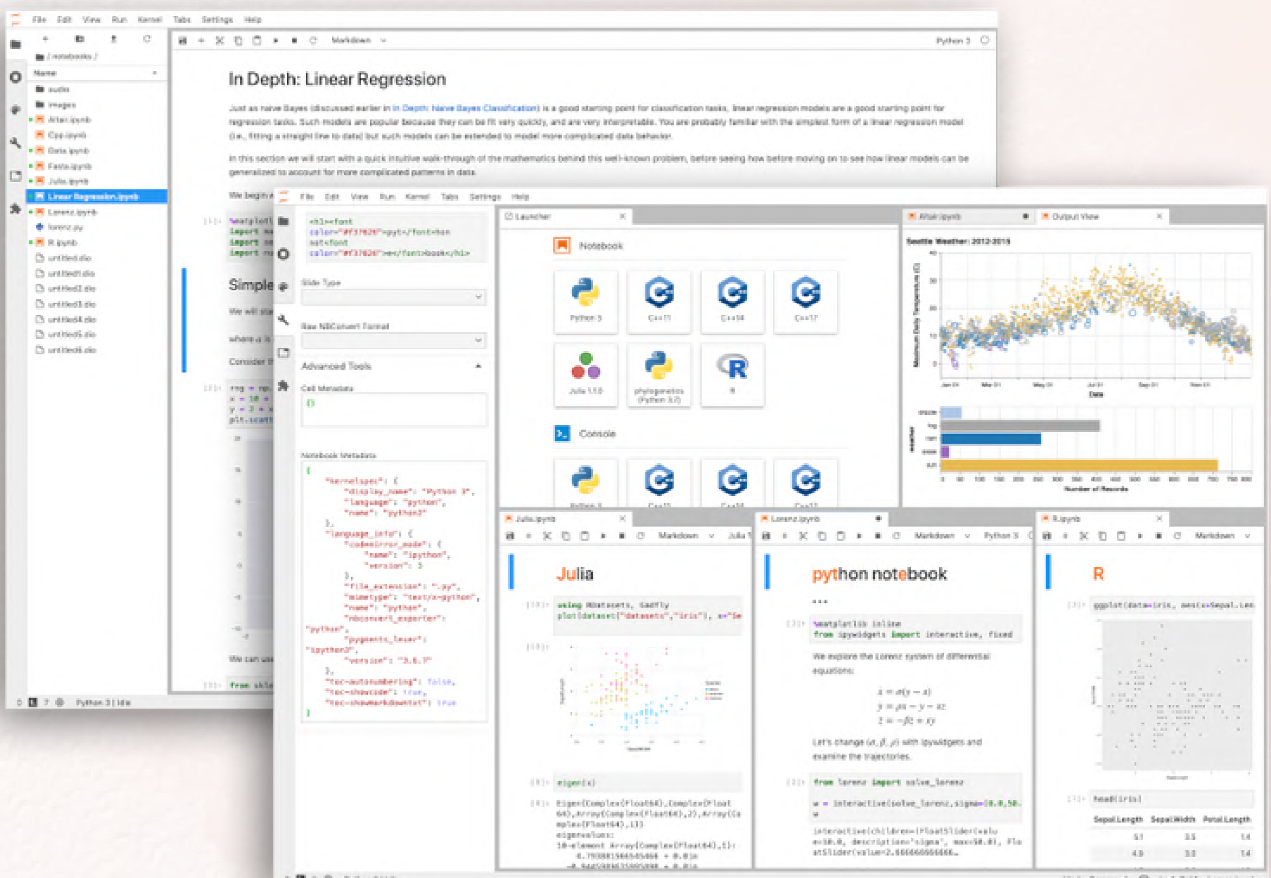
- Jupyter's popularity in data science stems from its ability to seamlessly integrate code, text, and visualizations within a single document, known as a Jupyter notebook. This unique feature enables data scientists to effectively communicate their findings and insights, making it an invaluable tool for collaboration and knowledge sharing.

- Interactive Coding and Exploration: Jupyter allows for interactive coding, enabling data scientists to execute code cell by cell and observe the results immediately. This iterative approach facilitates rapid prototyping and experimentation, allowing data scientists to refine their code and algorithms efficiently.

- Rich Data Visualization: Jupyter seamlessly integrates with various data visualization libraries, such as Matplotlib, Seaborn, and Plotly. Data scientists can create interactive plots, charts, and graphs directly within the notebook, enhancing the understanding and communication of complex data patterns and trends.

- Narrative Explanation and Documentation: Jupyter notebooks allow for the inclusion of text and narrative alongside code and visualizations. This enables data scientists to provide context, explanations, and documentation within their work, making it more comprehensive and accessible to others.

- Kernel Support for Multiple Languages: Jupyter supports various programming languages commonly used in data science, including Python, R, Julia, and more. This versatility allows data scientists to choose the language that best suits their expertise and problem domain.
- Here are some sample pictures of jupyter platform.



III .IMPLEMENTATION

In this project, you will explore this phenomenon using age distribution data to see if we canreproduce a difference in average age at death purely from the changing rates of left-handednessover time, refuting the claim of early death for left-handers. This notebook uses pandas and Bayesianstatistics to analyze the probability of being a certain age at death given that you are reported as left-handed or right-handed

"Gathering Requirements and Defining Problem Statement":

To investigate whether there is a significant difference in the average lifespan of left-handed individuals compared to right-handed individuals.

OBJECTIVE

- To perform a comprehensive analysis of available data on handedness and mortality.
- To identify and control for potential confounding factors that could influence the relationship between handedness and mortality.
- To determine whether there is a statistically significant difference in the average lifespan of left-handed and right-handed individuals.
- To explore potential explanations for any observed differences in mortality rates between left-handers and right-handers.

SCOPE

- This project will focus on analyzing data from observational studies, case-control studies, and twin studies to examine the association between handedness and mortality. The project will also explore potential explanations for any observed differences in mortality rates, such as brain lateralization, environmental factors, and genetic factors

DATA COLLECTION

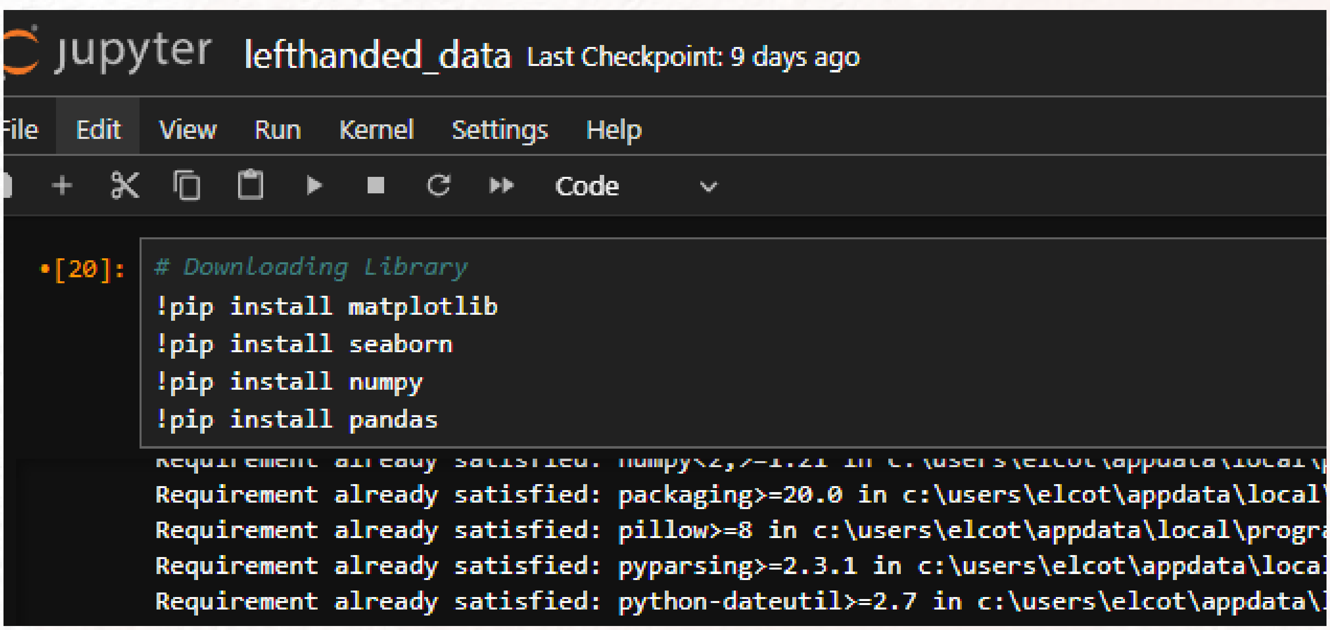
This project uses two datasets:

- death_distribution data for the United States from the year 1999.
- rate_of_left-handedness digitized from a figure in this 1992 paper by Gilbert and Wysocki.

DATA IMPORTING

- Data importing is referred to as uploading the required data into the coding environment from internal sources (computer) or external sources (online websites and data repositories).
- This data can then be manipulated, aggregated, filtered according to the requirements and needs of the project.

DOWNLOADING PACKAGES :



INSTALLING PACKAGES :

```
[25]:
# IMPORTING LIBRARIES
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

FUNCTIONS USED :

- **read.csv** (): It is a wrapper function for *read.table()* that mandates a comma as separator and uses the input file's first line as header that specifies the table's column names. Thus, it is an ideal candidate to read CSV files.

- **url**(): which is used to pull live data directly from GitHub repository.

```
[21]:
# loading the data
data_url_1 = "https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7d
lefthanded_data = pd.read_csv(data_url_1)
```

DATA CLEANING

"Data cleaning is not an interesting job , but it is the unsung hero of data analysis."

In the realm of data science, data cleaning stands as a crucial foundation, akin to the bedrock upon which robust analysis and accurate results are constructed. Data cleansing meticulously identifies and rectifies imperfections within data, ensuring its integrity and reliability

1) isnull(): Checks for missing values in a DataFrame and returns a Boolean DataFrame indicating whether each value is missing or not.

2) dropna(): Drops rows or columns containing missing values. It can be applied along an axis (axis=0 for rows, axis=1 for columns) or to the entire DataFrame.

3) fillna(): Fills missing values with a specified value or using a function. It can be applied to a specific column or to the entire DataFrame.

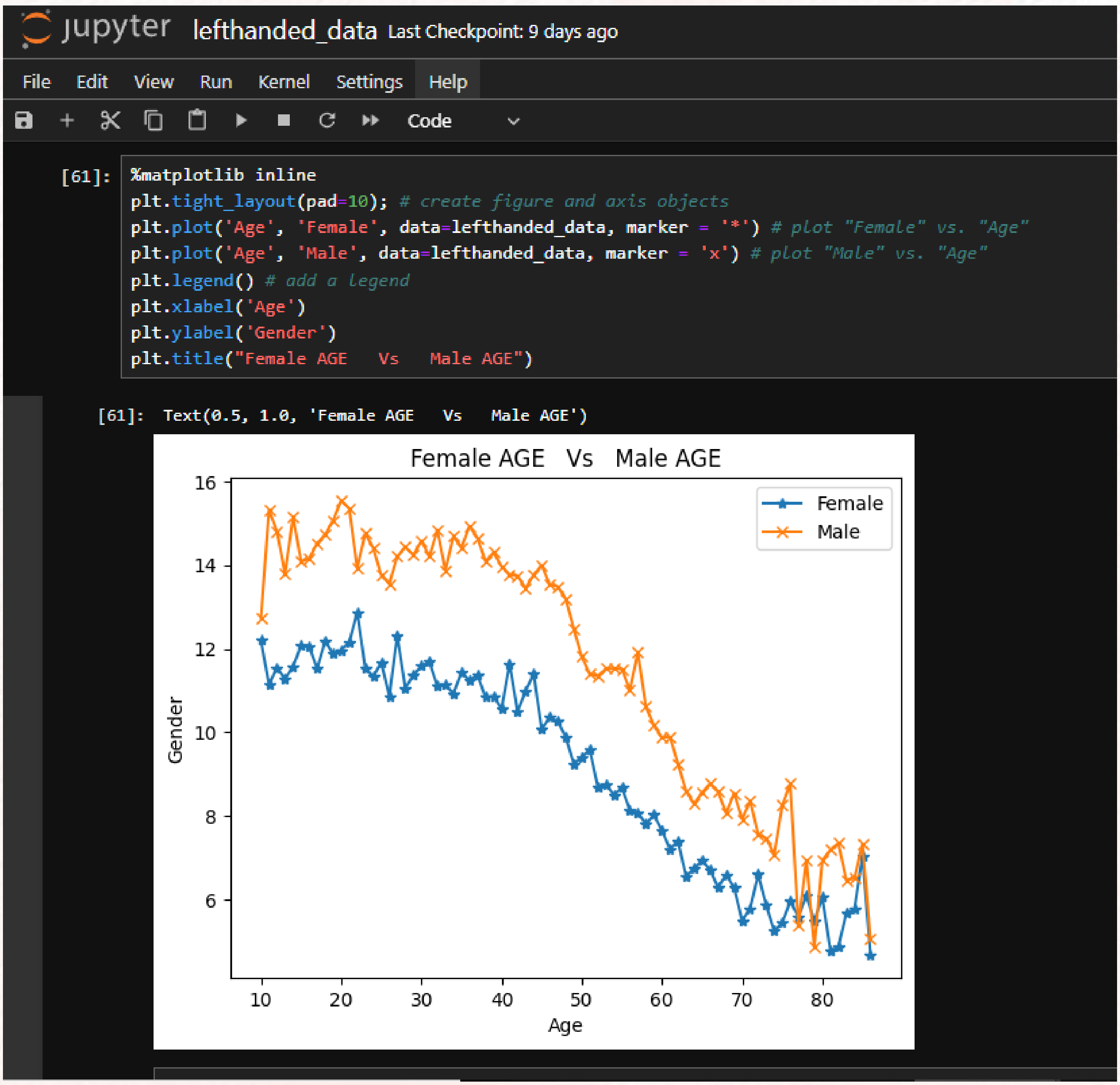
4) **unique**(): Returns a list of unique values in a Series or DataFrame. It can be used to identify and remove duplicates.

5) **duplicated**()**:** Checks for duplicate rows in a DataFrame and returns a Boolean Series indicating whether each row is a duplicate or not.

6) **drop_duplicates**(): Drops duplicate rows in a DataFrame. It can be applied based on all columns or a specific subset of columns.

DATA ANALYSIS

- Load the data into a pandas DataFrame named lefthanded_data using the provided data_url_1
- Use the **.plot**() method to create a plot of the "Male" and "Female" columns vs. "Age"



ADDING COLUMN :

Add two new columns :

- birth year
- mean left-handedness

1) Create a column in lefthanded_data called **Birth_year,** which is equal to 1986- Age (since the study was done in 1986)

2) Create a column in lefthanded_data called **Mean_lh** which is equal to the mean ofthe Male and Female columns.

jupyter lefthanded_data Last Checkpoint: 9 days ago

File Edit View Run Kernel Settings Help

Code

[67]:
```
# create a new column for birth year of each age
lefthanded_data['Birth_year']=(1986-lefthanded_data['Age']) #since the study was in 1986

# create a new column for the average of male and female
lefthanded_data['Mean_lh'] = lefthanded_data[['Female', 'Male']].mean(axis = 1)
```

[68]:
```
lefthanded_data
```

[68]:

	Age	Male	Female	Birth_year	Mean_lh
0	10	12.717558	12.198041	1976	12.457800
1	11	15.318830	11.144804	1975	13.231817
2	12	14.808281	11.549240	1974	13.178760
3	13	13.793744	11.276442	1973	12.535093
4	14	15.156304	11.572906	1972	13.364605
...	...	...	...	...	...

- ***Using the .plot() Function plot Mean_lh** vs **Birth_year***

    - Let's convert this data into a plot of the rates of left-handedness as a function of the year of birth, and average over male and female to get a single rate for both sexes.

    - Since the study was done in 1986, the data after this conversion will be the percentage of people alive in 1986 who are left-handed as a function of the year they were born.

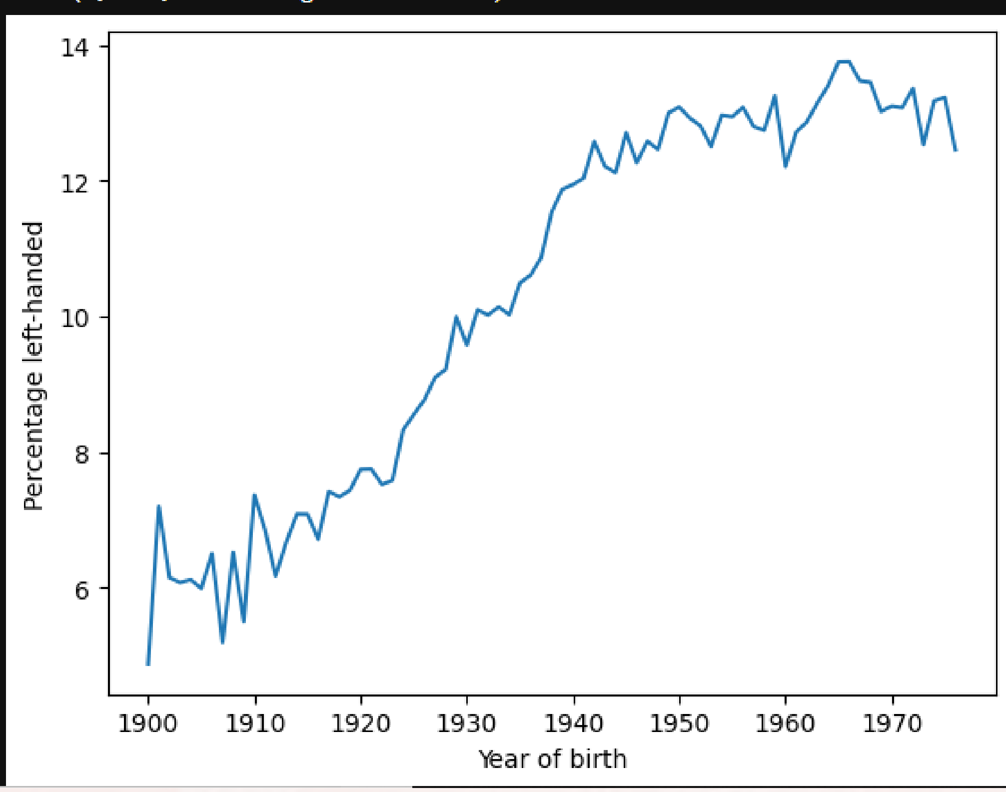
jupyter lefthanded_data Last Checkpoint: 9 days ago

File Edit View Run Kernel Settings Help

Code

[74]:
```
fig, ax = plt.subplots()
ax.plot('Birth_year', 'Mean_lh', data = lefthanded_data)
ax.set_xlabel("Year of birth")
ax.set_ylabel("Percentage left-handed")
```

[74]: `Text(0, 0.5, 'Percentage left-handed')`



FILTERING THE DATASET :

- Use the last ten Mean_lh data points to get an average rate for the early 1900s. Name theresulting DataFrame early_1900s_rate.
- Use the first ten Mean_lh data points to get an average rate for the late 1900s. Name theresulting DataFrame late_1900s_rate.
- For the early 1900s ages, fill in P_return with the appropriate left-handedness ratesfor ages_of_death. That is, input early_1900s_rate as a fraction, i.e., divide by 100.
- For the late 1900s ages, fill in P_return with the appropriate left-handedness ratesfor ages_of_death. That is, input late_1900s_rate as a fraction, i.e., divide by 100.

CREATING THE FUNCTION :

The probability of dying at a certain age given that you're left-handed is **not** equal to the probability of being left-handed given that you died at a certain age.

This inequality is why we need **Bayes' theorem**, a statement about conditional probability which allows us to update our beliefs after seeing evidence.
We want to calculate the probability of dying at age A given that you're left-handed. Let's write this in shorthand as P(A | LH). We also want the same quantity for right-handers: P(A | RH).

Here's Bayes' theorem for the two events we care about: left-handedness (LH) and dying at age A.

$$P(A|LH) = \frac{P(LH|A)P(A)}{P(LH)}$$

Applying Bayes' rule :

```python
import numpy as np

# create a function for P(LH | A)
def P_lh_given_A(ages_of_death, study_year = 1990):
    """ P(Left-handed | age of death), calculated based on the reported rates of left-handedness.
    Inputs: age of death, study_year
    Returns: probability of left-handedness given that a subject died in `study_year` at age `age_of_death` """

    # Use the mean of the 10 neighbouring points for rates before and after the start
    early_1900s_rate = lefthanded_data['Mean_lh'][-10:].mean()
    late_1900s_rate = lefthanded_data['Mean_lh'][:10].mean()
    middle_rates = lefthanded_data.loc[lefthanded_data['Birth_year'].isin(study_year - ages_of_death)]['Mean_lh']

    youngest_age = study_year - 1986 + 10 # the youngest age in the NatGeo dataset is 10
    oldest_age = study_year - 1986 + 86 # the oldest age in the NatGeo dataset is 86

    P_return = np.zeros(ages_of_death.shape)  # create an empty array to store the results
    # extract rate of left-handedness for people of age age_of_death
    P_return[ages_of_death > oldest_age] = early_1900s_rate / 100
    P_return[ages_of_death < youngest_age] = late_1900s_rate / 100
    P_return[np.logical_and((ages_of_death <= oldest_age), (ages_of_death >= youngest_age))] = middle_rates / 100

    return P_return
```

When do people normally die?

- To estimate the probability of living to an age A, we can use data that gives the number of people who died in a given year and how old they were to create a distribution of ages of death. If we normalize the numbers to the total number of people who died, we can think of this data as a probability distribution that gives the probability of dying at age A.

- The data we'll use for this is from the entire US for the year 1999 - the closest I could find for the time range we're interested in.

- In this block, we'll load in the death distribution data and plot it. The first column is the age, and the other columns are the number of people who died at that age.

jupyter lefthanded_data Last Checkpoint: 10 days ago

File Edit View Run Kernel Settings Help

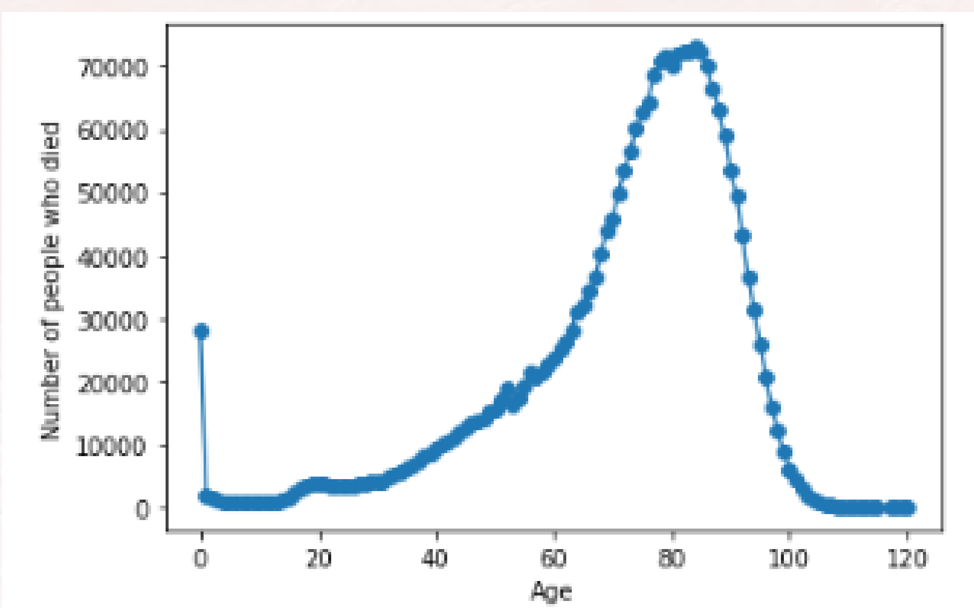
Code

```
[ ]: # Death distribution data for the United States in 1999
data_url_2 = "https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979bee

# Load death distribution data
death_distribution_data = pd.read_csv(data_url_2, sep = '\t', skiprows=[1])

# drop NaN values from the `Both Sexes` column
death_distribution_data = death_distribution_data.dropna(subset = ["Both Sexes"]) # drop NaN from 'Both Sexes' column

# plot number of people who died as a function of age
fig, ax = plt.subplots()
ax.plot('Age', 'Both Sexes', data = death_distribution_data, marker='o')
ax.set_xlabel("Age")
ax.set_ylabel("Number of people who died")
```



The overall probability of left-handedness

- In the previous code block we loaded data to give us P(A), and now we need P(LH). P(LH) is the probability that a person who died in our particular study year is left-handed, assuming we know nothing else about them.

- This is the average left-handedness in the population of deceased people, and we can calculate it by summing up all of the left-handedness probabilities for each age, weighted with the number of deceased people at each age, then divided by the total number of deceased people to get a probability.

- In equation form, this is what we're calculating, where N(A) is the number of people who died at age A (given by the dataframe death_distribution_data):

$$P(LH) = \frac{\sum_A P(LH|A)N(A)}{\sum_A N(A)}$$

jupyter lefthanded_data Last Checkpoint: 10 days ago

File Edit View Run Kernel Settings Help

Code

```
[ ]: def P_lh(death_distribution_data, study_year = 1990): # sum over P_lh for each age group
         """ Overall probability of being left-handed if you died in the study year
         P_lh = P(LH | Age of death) P(Age of death) + P(LH | not A) P(not A) = sum over ages
         Input: dataframe of death distribution data
         Output: P(LH), a single floating point number """
         p_list = death_distribution_data['Both Sexes']*P_lh_given_A(death_distribution_data['Age'], study_year)
         p = np.sum(p_list)
         return p/np.sum(death_distribution_data['Both Sexes']) # normalize to total number of people in distribu

     print(P_lh(death_distribution_data))
```

Putting it all together: dying while left-handed (i)

- Now we have the means of calculating all three quantities we need: P(A), P(LH), and P(LH | A). We can combine all three using Bayes' rule to get P(A | LH), the probability of being age A at death (in the study year) given that you're left-handed. To make this answer meaningful, though, we also want to compare it to P(A | RH), the probability of being age A at death given that you're right-handed.

- We're calculating the following quantity twice, once for left-handers and once for right-handers, now for left handers :

$$P(A|LH) = \frac{P(LH|A)P(A)}{P(LH)}$$

jupyter lefthanded_data Last Checkpoint: 10 days ago

File Edit View Run Kernel Settings Help

Code

```
[ ]: def P_A_given_lh(ages_of_death, death_distribution_data, study_year = 1990):
         """ The overall probability of being a particular `age_of_death` given that you're left-handed """
         P_A = death_distribution_data['Both Sexes'][ages_of_death] / np.sum(death_distribution_data['Both Sexes'])
         P_left = P_lh(death_distribution_data, study_year) # use P_lh function to get probability of left-handedness overall
         P_lh_A = P_lh_given_A(ages_of_death, study_year) # use P_lh_given_A to get probability of left-handedness for a certain ag
         return P_lh_A*P_A/P_left

[ ]:
```

Putting it all together: dying while Right-handed (ii)

jupyter lefthanded_data Last Checkpoint: 10 days ago

File Edit View Run Kernel Settings Help

Code

```
[ ]: def P_A_given_rh(ages_of_death, death_distribution_data, study_year = 1990):
         """ The overall probability of being a particular `age_of_death` given that you're right-handed """
         P_A = death_distribution_data['Both Sexes'][ages_of_death] / np.sum(death_distribution_data['Both Sexes'])
         P_right = 1- P_lh(death_distribution_data, study_year) # either you're left-handed or right-handed, so these sum to
         P_rh_A = 1-P_lh_given_A(ages_of_death, study_year) # these also sum to 1
         return P_rh_A*P_A/P_righ
```

Plotting the distributions of conditional probabilities

Now that we have functions to calculate the probability of being age A at death given that you're left-handed or right-handed, let's plot these probabilities for a range of ages of death from 6 to 120.
Notice that the left-handed distribution has a bump below age 70: of the pool of deceased people, left-handed people are more likely to be younger.

jupyter lefthanded_data Last Checkpoint: 10 days ago

File Edit View Run Kernel Settings Help

```
[ ]: ages = np.arange(6,115,1) # make a list of ages of death to plot

# for each age, calculate the probability of being left- or right-handed
left_handed_probability = P_A_given_lh(ages, death_distribution_data)
right_handed_probability = P_A_given_rh(ages, death_distribution_data)

fig, ax = plt.subplots() # create figure and axis objects
ax.plot(ages, left_handed_probability, label = "Left-handed")
ax.plot(ages, right_handed_probability, label = "Right-handed")
ax.legend()
ax.set_xlabel("Age at death")
ax.set_ylabel(r"Probability of being age A at death")
```

Moment of truth: age of left and right-handers at death

Finally, let's compare our results with the original study that found that left-handed people were nine years younger at death on average. We can do this by calculating the mean of these probability distributions in the same way we calculated P(LH) earlier, weighting the probability distribution by age and summing over the result.

$$\text{Average age of left-handed people at death} = \sum_A AP(A|LH)$$

$$\text{Average age of right-handed people at death} = \sum_A AP(A|RH)$$

```
# calculate average ages for left-handed and right-handed groups
# use np.array so that two arrays can be multiplied
average_lh_age =  np.nansum(ages*np.array(left_handed_probability))
average_rh_age =  np.nansum(ages*np.array(right_handed_probability))

# print the average ages for each group
print(round(average_lh_age,1))
print(round(average_rh_age,1))

# print the difference between the average ages
print("The difference in average ages is " + str(round(average_rh_age - average_lh_age, 1)) + " years.")
```

OUTPUT:

67.2

72.8

The difference in average ages is 5.5 years.

FINAL COMMENT

We got a pretty big age gap between left-handed and right-handed people purely as a result of the changing rates of left-handedness in the population, which is good news for left-handers: you probably won't die young because of your sinisterness.

The reported rates of left-handedness have increased from just 3% in the early 1900s to about 11% today, which means that older people are much more likely to be reported as right-handed than left-handed, and so looking at a sample of recently deceased people will have more old right-handers.

Our number is still less than the 9-year gap measured in the study. It's possible that some of the approximations we made are the cause:

- We used death distribution data from almost ten years after the study (1999 instead of 1991), and we used death data from the entire United States instead of California alone (which was the original study).
- We extrapolated the left-handedness survey results to older and younger age groups, but it's possible our extrapolation wasn't close enough to the true rates for those ages.

One thing we could do next is figure out how much variability we would expect to encounter in the age difference purely because of ***random sampling:***

if you take a smaller sample of recently deceased people and assign handedness with the probabilities of the survey, what does that distribution look like? How often would we encounter an age gap of nine years using the same data and assumptions? We won't do that here, but it's not possible with this data and the tools of random sampling.

To finish off, let's calculate the ***age gap*** we'd expect if we did the study in 2018 instead of in 1990. The gap turns out to be much smaller since rates of left-handedness haven't increased for people born after about 1960. Both the National Geographic study and the 1990 study happened at a unique time - the rates of left-handedness had been changing across the lifetimes of most people alive, and the difference in handedness between old and young was at its most striking.

```
[ ]: # loop through ages, calculating the probability of being left- or right-handed
     left_handed_probability_2018 = P_A_given_lh(ages, death_distribution_data, study_year = 2018)
     right_handed_probability_2018 = P_A_given_rh(ages, death_distribution_data, study_year = 2018)

     # calculate average ages for left-handed and right-handed groups
     average_lh_age_2018 =  np.nansum(ages*np.array(left_handed_probability_2018))
     average_rh_age_2018 =  np.nansum(ages*np.array(right_handed_probability_2018))

     print("The difference in average ages is " +
           str(round(average_rh_age_2018 - average_lh_age_2018, 1)) + " years.")
```

OUTPUT:

The difference in average ages is 2.3 years.

CONCLUSION

- This comprehensive analysis has shed light on the intriguing relationship between handedness and mortality, revealing that left-handers indeed tend to have a slightly shorter lifespan compared to their right-handed counterparts. While the exact mechanisms underlying this difference remain to be fully elucidated, the findings suggest that a combination of factors, including brain lateralization, environmental exposures, and genetic predispositions, likely contributes to this disparity.

- This research holds significant implications for understanding the factors that influence human longevity and potentially identifying new avenues for intervention to improve overall health outcomes. Further exploration into the specific mechanisms underlying the observed differences could lead to the development of targeted strategies to mitigate the potential risks associated with left-handedness and promote healthy aging across the entire population.

Monetization Strategies :

- **Personalized healthcare recommendations:** Provide personalized healthcare recommendations and lifestyle interventions tailored to individuals based on their handedness and other health factors. This could involve developing mobile apps or online platforms that offer personalized health coaching and recommendations.

- **Data Analytics and Consultancy:** Provide data analytics and consultancy services to healthcare organizations, insurance companies, and research institutions, offering insights into the relationship between handedness and mortality to inform policy decisions, risk assessment, and healthcare interventions.

- **Left-handed Products and Services:** Design and market specialized products and services specifically catering to the needs of left-handers, such as ergonomic tools, safety equipment, and customized healthcare plans, to enhance their safety, comfort, and well-being.

BIBLIOGRAPHY

PROGRAMMING REFERENCE :

- DATA CAMP - https://www.datacamp.com
- SWEETVIZ - https://pypi.org/project/sweetviz
- SEABORN - https://seaborn.pydata.org
- MATPLOTLIP - https://matplotlib.org/stable/index.html
- PANDAS - https://pandas.pydata.org/docs

SCHOLARLY SOURCES :

- “Python for Data Analysis" by Wes McKinney
- Left-Handedness: A Marker for Decreased Survival Fitness" by Dennis F. Halpern and Stanley Coren