FINAL REPORT OF TRAINEESHIP PROGRAM 2023

On

“Analyze Death Age Difference of Right Handers with Left Handers”

MEDTOUREASY

18th August 2023

ACKNOWLEDGEMENT

I am truly grateful for the invaluable experiences and learning opportunities that my traineeship with MedTourEasy has provided me, particularly in the realm of Data Visualizations within Data Analytics. This journey has been instrumental in both my personal growth and professional development.

I extend my heartfelt appreciation to the Training & Development Team at MedTourEasy for affording me the privilege to undertake my traineeship within their esteemed organization. Their belief in my potential and their unwavering support have been crucial in shaping my understanding of the intricacies of Data Analytics and its applications. Their guidance paved the way for a successful traineeship project, leading to enhanced client satisfaction.

A special mention of gratitude is reserved for Ankit Hasija, who not only shared comprehensive insights into the Data Analytics profile but also dedicated valuable time from their busy schedule to mentor me throughout this journey. Their mentorship has been a guiding light, making this experience a profound learning curve.

I am also indebted to the entire team at MedTourEasy, as well as my fellow colleagues, for fostering a productive and conducive working environment. Their collaborative spirit and shared wisdom have contributed significantly to my growth and have made every challenge surmountable.

Lastly, I would like to express my profound thanks to everyone who played a part in enriching my traineeship experience at MedTourEasy. Your collective efforts have left an indelible mark on my professional journey, and for that, I am sincerely grateful.

With sincere appreciation,

SABIT RANJAN SAHOO

ABSTRACT

This project delves into the intriguing correlation between left-handedness and life expectancy, debunking the longstanding notion that left-handed individuals face an early demise. In popular culture, figures like Barack Obama, Bill Gates, Oprah Winfrey, Babe Ruth, and Marie Curie have been identified as left-handed, yet a 1991 study claimed that left-handed people tend to live approximately nine years less than their right-handed counterparts.

By utilizing age distribution data and employing Bayesian statistics, this project aims to investigate whether the reported disparity in average age at death can be attributed to the evolving rates of left-handedness across different time periods. The analysis is conducted using Python's pandas library, and statistical skills from courses such as "pandas Foundations," "Statistical Thinking in Python," "Introduction to Data Visualization with Python," and foundational concepts from R courses such as "Foundations of Probability in R" and "Fundamentals of Bayesian Data Analysis in R."

The study takes a comprehensive approach, integrating skills such as data reading, column creation, visualization using matplotlib, working with NumPy arrays, constructing probabilities from frequency data, and applying Bayes' theorem. The analysis hinges on the premise that left-handedness rates are more closely tied to birth year than age, as highlighted by the findings of the National Geographic survey from 1986. This survey indicated that left-handedness rates decrease from around 13% for individuals under 40 to about 5% for those aged 80 and above. Researchers Avery Gilbert and Charles Wysocki attributed this age-dependent shift to changing social norms rather than age-related factors.

In essence, the analysis presented in this project seeks to unravel the underlying dynamics behind the perceived difference in life expectancy between left-handed and right-handed individuals. By exploring the evolving rates of left-handedness, this project intends to challenge the premise of premature mortality among left-handers and shed light on the influence of societal factors on statistical outcomes.

Table of Contents

Chapter: 1
INTRODUCTION

1.1 About the Company:

MedTourEasy, a prominent international healthcare firm, equips you with the essential information to assess your healthcare choices on a global scale. Serving as a comprehensive resource, MedTourEasy offers analytical solutions to our network of healthcare partners across the globe.

1.2 About the Project:

In the year 1991, a significant study surfaced, claiming that individuals who are left-handed experience an average lifespan that is nine years shorter compared to their right-handed counterparts. This assertion stemmed from an investigation conducted by researchers Avery Gilbert and Charles Wysocki. They meticulously analysed data extracted from a National Geographic survey carried out in 1986. This survey amassed responses from over a million participants, detailing information about their age, gender, and hand preferences for activities such as throwing and writing. Gilbert and Wysocki's scrutiny revealed a pattern: left-handedness rates were approximately 13% among individuals below 40 years of age, gradually diminishing to around 5% by the time individuals reached 80 years of age.

However, their conclusions took a fascinating turn when they delved deeper into a subgroup of respondents who exhibited an interesting trait: they threw with their left hand but wrote with their right. This subgroup, less influenced by societal pressures to conform to right-handedness, showcased a more authentic reflection of their natural inclination. In essence, this finding indicated that the relationship between age and left-handedness was predominantly influenced by the shifting social acceptance of left-handedness over different eras. Thus, the prevalence of left-handedness appeared to be associated not with age itself but rather with the era of birth. Extrapolating from this insight, a contemporary replication of the study would likely yield a similar distribution of hand preferences, albeit adjusted in alignment with modern times.

The implications of these findings are profound for entities like MedTourEasy, a global healthcare provider. Understanding the age-related disparities and their underlying causes allows MedTourEasy to tailor its informational resources and analytical solutions to accommodate the unique requirements of both left-handed and right-handed customers, as well as their partner healthcare providers worldwide. This comprehension stands to augment customer satisfaction, foster loyalty, and confer a competitive edge within the global healthcare arena.

Consequently, the core objective of this project centres on the collection and analysis of datasets to generate intuitive and interactive visualizations. These visual representations aim to convey the outcomes of the "**Age disparity between left-handed and right-handed individuals**." By doing so, the project seeks to extract meaningful insights that can shed light on this intriguing phenomenon. Structurally, the project unfolds in two main segments, outlined below:

1.2.1 Death distribution data:

The dataset comprises intricate mortality statistics meticulously compiled by the Division of Vital Statistics, under the aegis of the National Center for Health Statistics. It encompasses a comprehensive array of information, encompassing factors such as age, race, gender, causes of death, life expectancy, and infant mortality. Within this dataset, one can unearth an assortment of tables that encapsulate data at varying levels of granularity. These tables offer insights at the national level, while others provide a closer examination of states individually. The diversity of perspectives encapsulated within this dataset enables a multi-dimensional exploration of mortality trends and patterns.

1.2.2 Hand preference and age in the United States:

In a comprehensive survey conducted across the United States, involving a substantial cohort of 1,177,507 individuals ranging in age from 10 to 86 years, inquiries were posed regarding their hand preferences for tasks such as writing and throwing. The findings of this survey illuminate an intriguing trend: a reduction in the occurrence of left-handedness with advancing age. However, this observation comes with a nuanced twist, as the data divulges that rate of mixed-handedness and left-handedness, when examined with age as a focal point, display unique patterns distinct from one another. This dataset not only offers valuable insights into the relationship between age and hand preference but also underscores the multifaceted nature of handedness trends over the course of an individual's lifespan.

The outcome is portrayed through a process of dissecting the data within these datasets, employing tools such as pandas and Bayesian statistics. This analytical approach delves into the likelihood of attaining a specific age at the time of death, contingent upon an individual's identification as left-handed or right-handed. The visual representations crafted from this analysis serve as a valuable resource, offering the firm an avenue for comprehending the scenario and extracting meaningful insights.

The visualizations generated by this methodology serve as a lens through which the organization can scrutinize the situation at hand and derive informed conclusions.

1.3 Objectives and Deliverables:

Within this project, our central focus is directed towards investigating one of the most contentious assertions pertaining to left-handedness – the purported association with premature mortality. Employing data analysis techniques and statistical methodologies, we aim to rigorously examine the validity of this claim and delve into the underlying factors that might elucidate the observed disparities in life expectancy between left-handed and right-handed individuals.

Moreover, we endeavour to delve into the dynamic evolution of left-handedness rates across temporal dimensions and diverse geographical regions. Through a meticulous analysis, we aspire to unravel the implications of these fluctuations, thereby enhancing our comprehension of this intricate phenomenon.

By harnessing the power of pandas, we will efficiently manipulate and visually represent the data. Furthermore, our toolkit will encompass Bayesian statistics, allowing us to derive estimations of the probability of attaining a particular age at the time of death based on whether one identifies as left-handed or right-handed. Our inquiry will also encompass an exploration of how the rates of left-handedness have undergone transformations over time due to the influence of social and cultural factors. The nuanced interplay of these dynamics will shed light on the comparative evaluation of the average age at death within these two distinct groups.

Ultimately, this project aims to provide an evidence-based assessment of the claim surrounding left-handedness and its association with early mortality. Through a multi-faceted approach combining data analysis, statistical inference, and historical analysis, we anticipate uncovering new insights into this intriguing phenomenon.

1.4 The deliverables of this analysis are:

i. Graphical representation of comparison of male and female left-handedness rates vs. age.
ii. Comparing Rates of left-handedness over time.
iii. Create a function that will return P(LH | A) for particular ages of death in a given study year.
iv. Graphical representation of number of people who as a function of age.
v. The overall probability of left-handedness.
vi. Graph to show probability of being a certain age at death given that you're left- or right-handed for a range of ages.
vii. Age of left and right-handers at death.
viii. A summary report that highlights the main findings and conclusions of the analysis, supported by relevant graphs and statistics.
ix. A presentation that communicates the key insights and implications of the analysis to a general audience.

Chapter: 2
METHODOLOGY

2.1 Data Collection and Preprocessing:

The dataset used in this analysis was obtained from the Centres for Disease Control and Prevention (CDC). The dataset contains information about the age at death and handedness of individuals. The data was originally in tab-separated values (TSV) format and was downloaded from the CDC's official data repository.

2.1.1 Data Cleaning:

Before proceeding with the analysis, the dataset underwent a thorough data cleaning process. This involved identifying and removing duplicate entries and instances with missing or incomplete data. Cleaning ensured that the dataset was free from errors that could potentially affect the accuracy of the analysis.

2.1.2 Data Transformation:

Categorical variables, such as handedness, were transformed into numerical values to facilitate analysis. Handedness was represented as binary values, where "0" represented right-handedness and "1" represented left-handedness.

2.1.3 Missing Data Handling:

For instances with missing values, a decision was made on how to handle them based on the extent of missingness. Entries with critical missing information were excluded from the analysis, while cases with partial missing data were either imputed or flagged for further investigation.

2.1.4 Feature Engineering:

Additional features were derived from the existing data to enhance the analysis. For example, birth year was calculated by subtracting age at death from the study year. This additional feature enabled a more comprehensive examination of the relationship between age and handedness.

2.2 Calculation of Left-Handedness Probabilities:

The central focus of the analysis was to investigate whether there is a connection between age at death and handedness. To achieve this, probabilities associated with left-handedness and right-handedness were calculated using the following methods:

2.2.1 P(A | LH): Probability of Age given Left-Handedness:

This calculation involved determining the probability of a specific age at death given that the individual was left-handed. It required computing the joint probability of age and handedness, which was then normalized by the overall probability of being left-handed.

2.2.2 P(A | RH): Probability of Age given Right-Handedness:

Similar to the previous calculation, this method aimed to find the probability of age at death given that the individual was right-handed. The calculation followed a similar approach as the left-handedness probability calculation.

2.3 Visualizations and Interpretation:

To visually convey the relationship between age at death and handedness, line plots were generated. Each plot depicted the probabilities of left-handedness and right-handedness across various age groups.

2.3.1 The visualizations served multiple purposes:

i. Identification of Trends: The line plots allowed for the identification of trends in the data, highlighting any patterns or changes in handedness probabilities with age.

ii. Group Comparison: By comparing the line plots of left-handedness and right-handedness, any disparities or differences in probabilities at different ages could be easily observed.

iii. Supporting Analysis: The visualizations complemented the probabilistic calculations, making it easier to interpret the results and draw insights.

2.4 Average Age Calculation:

In addition to examining the probabilities, the average age for both left-handed and right-handed individuals was calculated. This involved summing the products of age and corresponding probabilities, providing an estimate of the central tendency of ages in each group.

2.5 Sensitivity Analysis:

To assess the potential impact of different study years on the findings, a sensitivity analysis was conducted. The calculations and visualizations were repeated using a specific study year (e.g., 2018) to explore whether the results remained consistent across different timeframes.

Chapter: 3

TASKS, CODE AND OUTPUTS

Task 1:

Instructions:

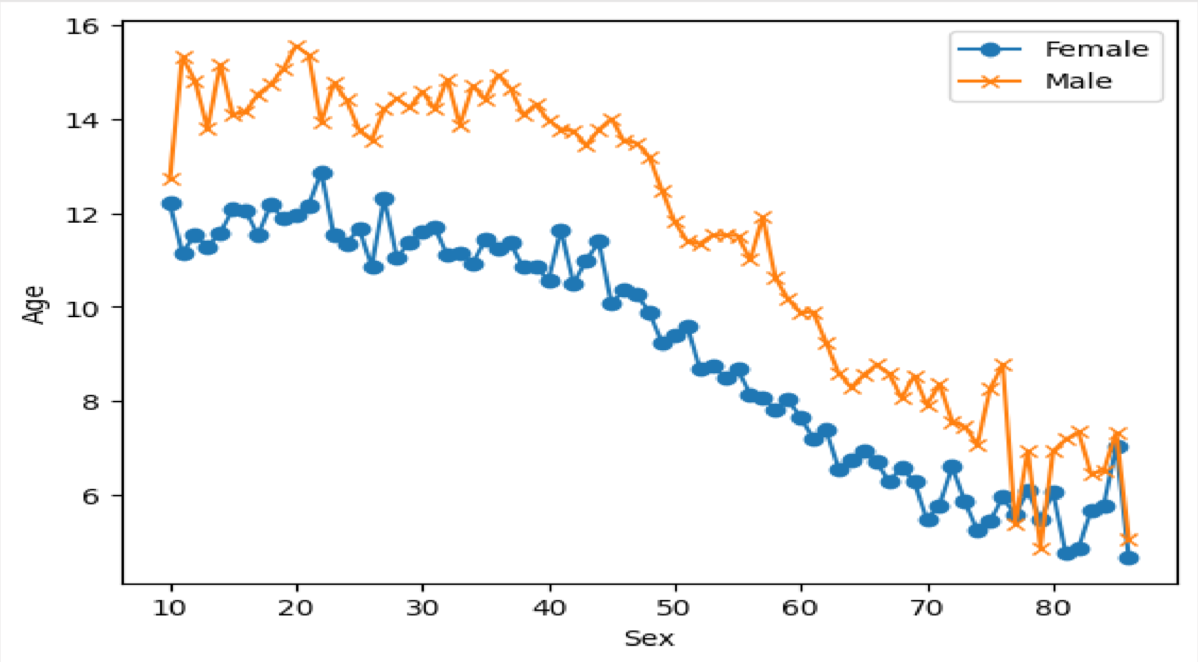
- Load the handedness data from the National Geographic survey and create a scatter plot.
- Import pandas as pd and matplotlib.pyplot as plt.
- Load the data into a pandas Data Frame named lefthanded_data using the provided data_url_1. Note that the file is a CSV file.
- Use the .plot() method to create a plot of the "Male" and "Female" columns vs. "Age".

Code 1:

```
# import libraries
# ... YOUR CODE FOR TASK 1 ...
import pandas as pd
import matplotlib.pyplot as plt
# load the data
data_url_1 = "https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv"
lefthanded_data = pd.read_csv(data_url_1)

# plot male and female left-handedness rates vs. age
%matplotlib inline
fig, ax = plt.subplots() # create figure and axis objects
ax.plot('Age', 'Female', data = lefthanded_data, marker = 'o') # plot "Female" vs. "Age"
ax.plot('Age', 'Male', data = lefthanded_data, marker = 'x') # plot "Male" vs. "Age"
ax.legend() # add a legend
ax.set_xlabel('Sex')
ax.set_ylabel('Age')
```

Output 1:



Task 2:

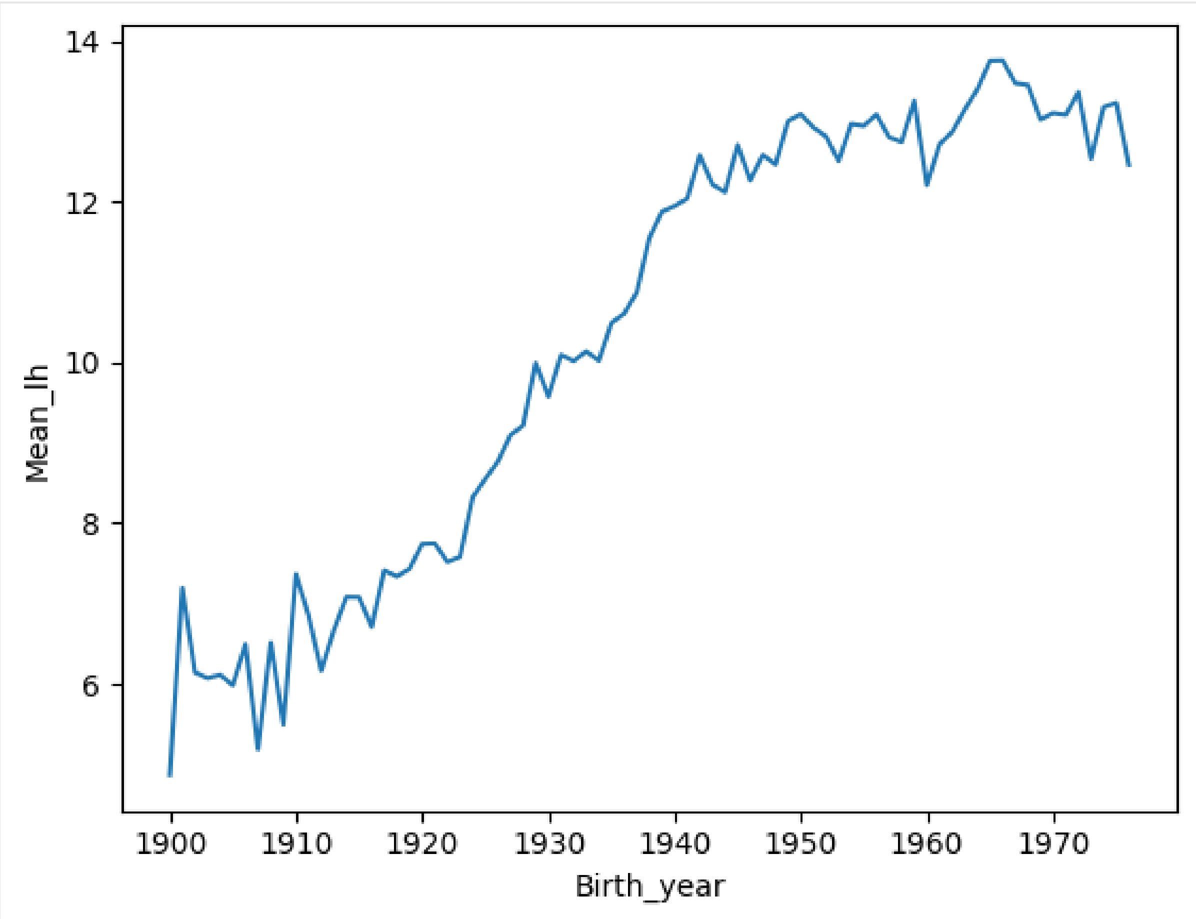
Instructions:

- Add two new columns, one for birth year and one for mean left-handedness, then plot the mean as a function of birth year.
- Create a column in lefthanded_data called Birth_year, which is equal to 1986- Age (since the study was done in 1986).
- Create a column in lefthanded_data called Mean_lh which is equal to the mean of the Male and Female columns.
- Use the .plot() method to plot Mean_lh vs. Birth_year.

Code 2:

```
# create a new column for birth year of each age
# ... YOUR CODE FOR TASK 2 ...
lefthanded_data['Birth_year'] = 1986 - lefthanded_data['Age']
# create a new column for the average of male and female
# ... YOUR CODE FOR TASK 2 ...
lefthanded_data['Mean_lh'] = lefthanded_data[['Male', 'Female']].mean(axis=1)
# create a plot of the 'Mean_lh' column vs. 'Birth_year'
fig, ax = plt.subplots()
ax.plot('Birth_year', 'Mean_lh', data = lefthanded_data) # plot 'Mean_lh' vs. 'Birth_year'
ax.set_xlabel('Birth_year') # set the x label for the plot
ax.set_ylabel('Mean_lh') # set the y label for the plot
```

Output 2:



Task 3:

Instructions:

- Create a function that will return P(LH | A) for particular ages of death in a given study year.
- Import the NumPy package aliased as np.
- Use the last ten Mean_lh data points to get an average rate for the early 1900s. Name the resulting DataFrame early_1900s_rate.
- Use the first ten Mean_lh data points to get an average rate for the late 1900s. Name the resulting DataFrame late_1900s_rate.
- For the early 1900s ages, fill in P_return with the appropriate left-handedness rates for 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 rates for ages_of_death. That is, input late_1900s_rate as a fraction, i.e., divide by 100.
- When calculating early_1900s_rate and late_1900s_rate, remember that because the original data was from youngest age to oldest age, that means that the data is organized from latest birth year to earliest birth year. You will use the first ten Mean_lh data points to get an average rate for the late 1900s and the last ten for the early 1900s.

Code 3:

```python
# import library
# ... YOUR CODE FOR TASK 3 ...
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 | ages of death), calculated based on the reported rates of left-handedness.
    Inputs: numpy array of ages of death, study_year
    Returns: probability of left-handedness given that subjects died in `study_year` at ages `ages_of_death` """

    # Use the mean of the 10 last and 10 first points for left-handedness 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 is 10
    oldest_age = study_year - 1986 + 86 # the oldest age 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 ages 'ages_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
```

Task 4:

Instructions:

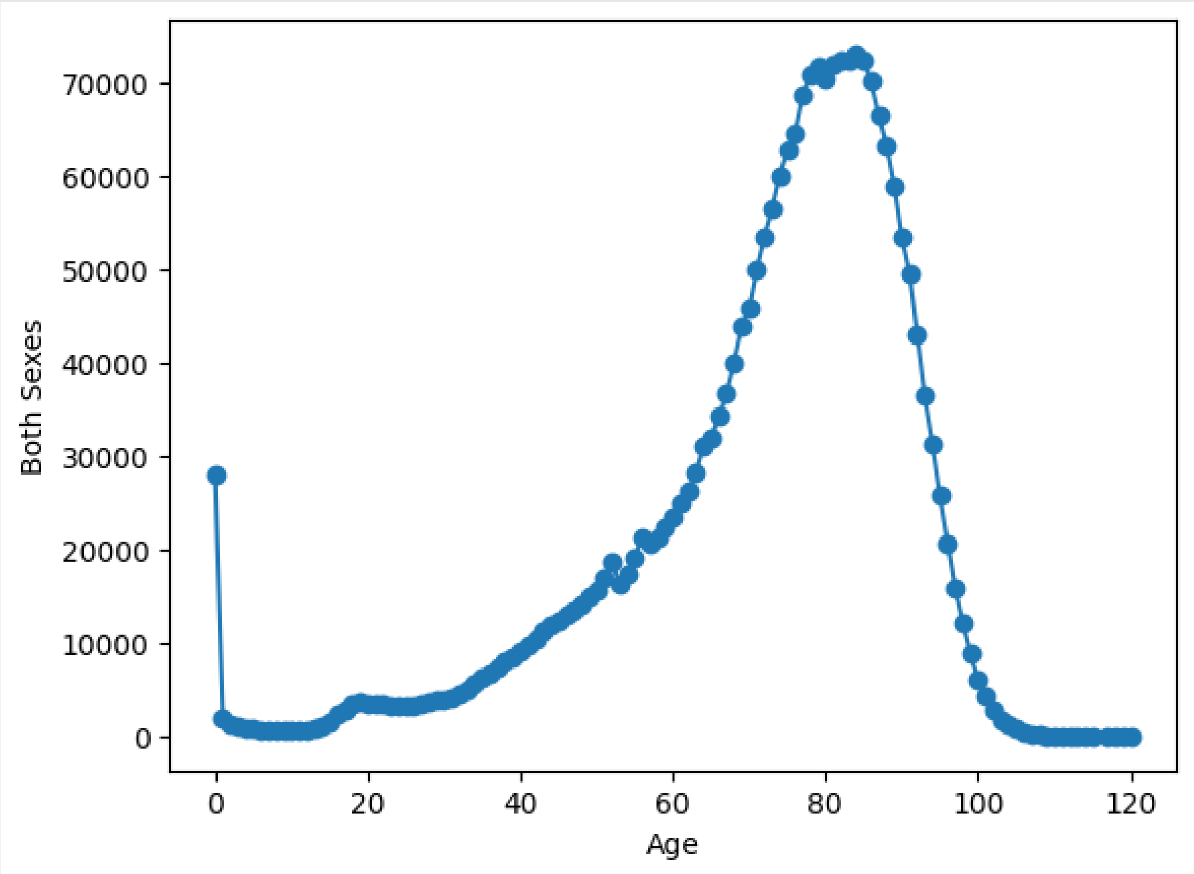
- Load death distribution data for the United States and plot it.
- Load death distribution data in the provided data_url_2 into death_distribution_data,setting sep = '\t' and skiprows=[1] to account for the dataset's format.
- Drop the NaN values from the Both Sexes column.
- Use the .plot() method to plot the number of people who died as a function of their age.

Code 4:

```
# Death distribution data for the United States in 1999
data_url_2 = "https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_table310.tsv"

# load death distribution data
# ... YOUR CODE FOR TASK 4 ...
death_distribution_data = pd.read_csv(data_url_2, sep='\t', skiprows=[1])
# drop NaN values from the `Both Sexes` column
# ... YOUR CODE FOR TASK 4 ...
death_distribution_data = death_distribution_data.dropna(subset = ['Both Sexes'])
# 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') # plot 'Both Sexes' vs. 'Age'
ax.set_xlabel('Age')
ax.set_ylabel('Both Sexes')
```

Output 4:



Task 5:

Instructions:

- Create a function called P_lh() which calculates the overall probability of left-handedness in the population for a given study year.
- Create a series, p_list, by multiplying the number of dead people in the Both Sexes column with the probability of their being lefthanded using P_lh_given_A().
- Set the variable p equal to the sum of that series.
- Divide p by the total number of dead people by summing death_distribution_data over the Both Sexes column. Return result from the function. P(LH | A) was defined in Task3. N(A) is the value of Both Sexes in the death_distribution_data DataFrame where

the Age column is equal to A. The denominator is total number of dead people, which you can get by summing over the entire data frame in the Both Sexes column.

Code 5.1:

```
print(death_distribution_data.columns)
```

Output 5.1:

```
Index(['Age', 'Both Sexes', 'Male', 'Female', 'P_lh_given_A'], dtype='object')
```

Code 5.2:

```
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
    Input: dataframe of death distribution data, study year
    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) # multiply number of dead people by P_lh_given_A
    p = np.sum(p_list) # calculate the sum of p_list
    return p / np.sum(death_distribution_data['Both Sexes']) # normalize to total number of people (sum of death_distribution_data['Both Sexes'])

print(P_lh(death_distribution_data))
```

Output 5.2:

```
0.07766387615350638
```

Task 6:

Instructions:

- Write a function to calculate P_A_given_lh().
- Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people (the sum of the Both Sexes column of death_distribution_data).
- Calculate the overall probability of left-handedness P(LH) using the unction defined in Task 5.
- Calculate P(LH | A) using the function defined in Task 3.

Code 6:

```
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 age
    return P_lh_A*P_A/P_left
```

Task 7:

Instructions:

- Write a function to calculate P_A_given_rh().
- Calculate P_A, the overall probability of dying at age A, which is given by death_distribution_data at age A divided by the total number of dead people. (This value is the same as in task 6.)
- Calculate the overall probability of right-handedness P(RH), which is
- 1-P(LH).•Calculate P(RH | A), which is 1 - P(LH | A).

Code 7:

```
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 P_right = 1 - P_left
    P_rh_A = 1 - P_lh_given_A(ages_of_death, study_year) # P_rh_A = 1 - P_lh_A
    return P_rh_A*P_A/P_right
```

Task 8:

Instructions:

- Plot the probability of being a certain age at death given that you're left or right-handed for a range of ages.
- Calculate P_A_given_lh and P_A_given_rh using the functions defined in Task 6.
- Use the .plot() method to plot the results versus age.

Code 8.1:

```
print(lefthanded_data)
```

Output 8.1:

```
    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
..  ...        ...        ...         ...        ...
72   82   7.350204   4.874899        1904   6.112551
73   83   6.471204   5.672536        1903   6.071870
74   84   6.510858   5.774881        1902   6.142870
75   85   7.337968   7.051459        1901   7.194713
76   86   5.059387   4.680948        1900   4.870168

[77 rows x 5 columns]
```

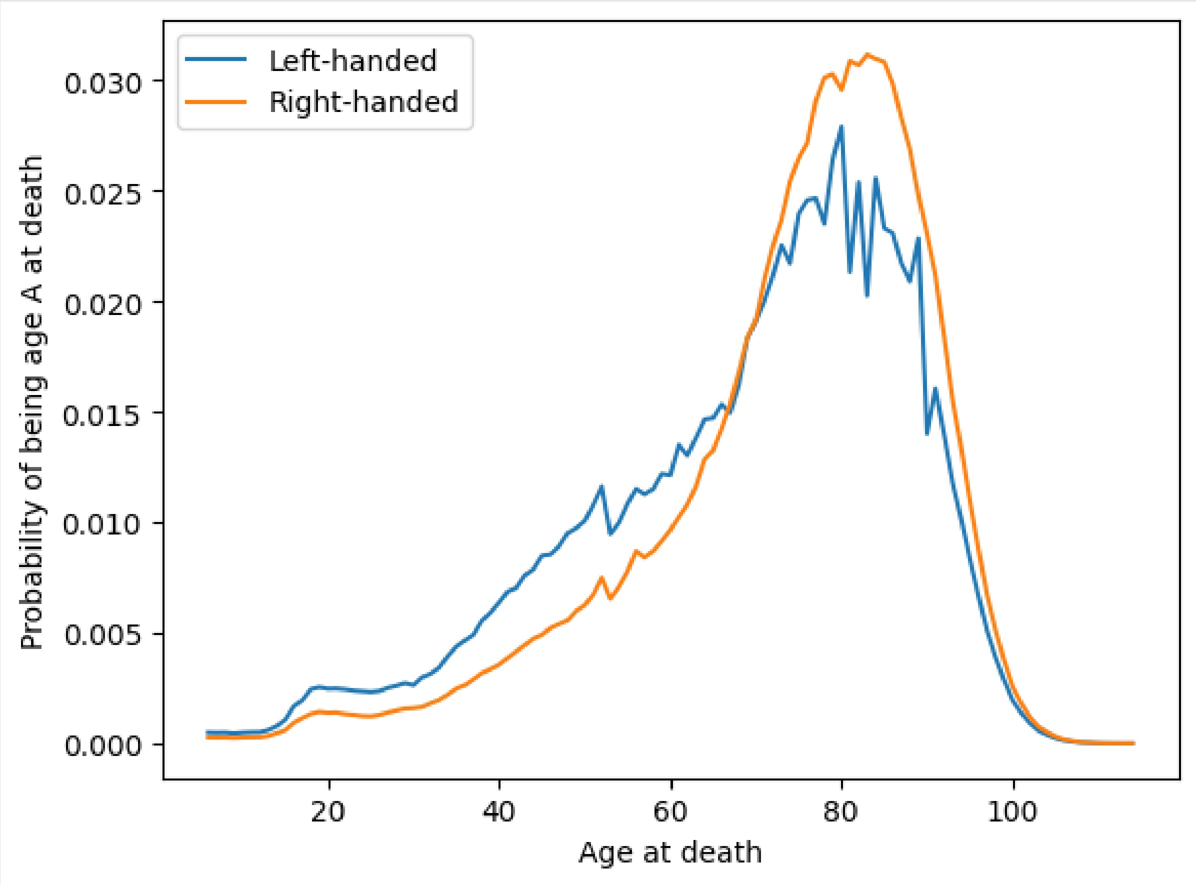
Code 8.2:

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

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

# create a plot of the two probabilities vs. age
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() # add a legend
ax.set_xlabel("Age at death")
ax.set_ylabel(r"Probability of being age A at death")
```

Output 8.2:



Task 9:

Instructions:

- Find the mean age at death for left-handers and right-handers.
- Multiply the ages list by the left-handed probabilities of being those ages at death, then use np.nan sum to calculate the sum. Assign the result to average_lh_age.
- Do the same with the right-handed probabilities to calculate average_rh_age.
- Print average_lh_age and average_rh_age.
- Calculate the difference between the two average ages and print it. To make your printed output prettier, try using the round() function to round your results to two decimal places.

Code 9:

```
# 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
# ... YOUR CODE FOR TASK 9 ...
print("Average age of lefthanded = " + str(average_lh_age))
print("Average age of righthanded = " + str(average_rh_age))

# 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 9:

```
Average age of lefthanded = 67.24503662801027
Average age of righthanded = 72.79171936526477
The difference in average ages is 5.5 years.
```

Task 10:

Instructions:

- Redo the calculation from Task 8, setting the study_year parameter to 2018.
- In the call to P_A_given_lh, set age_of_death to ages,
- death_distribution_data to death_distribution_data, and study_year to 2018.
- Do the same for P_A_given_rh.

Code 10:

```
# Calculate the probability of being left- or right-handed for all ages
left_handed_probability_2018 = P_A_given_lh(ages, death_distribution_data, 2018)
right_handed_probability_2018 = P_A_given_rh(ages, death_distribution_data, 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 10:

```
The difference in average ages is 2.3 years.
```

Chapter: 4
SUMMARY OF THE ANALYSIS

4.1 Summary of the Methodology used:

The primary objective of this investigation is to delve into the assertion suggesting a potential correlation between left-handed individuals and shorter lifespans in comparison to their right-handed counterparts. Employing the powerful tools of pandas and Bayesian statistics, this analysis embarks on the journey of computing posterior probabilities that shed light on the likelihood of being left-handed or right-handed based on the age at which individuals pass away. Furthermore, the study endeavours to discern any disparities in the average ages of death for these two distinct groups.

Underlying this inquiry are two critical datasets. First, a treasure trove of data hailing from a 1986 National Geographic survey forms the foundation. This dataset captures crucial information such as age, gender, and hand preference of over a million respondents. Notably, the survey data intriguingly reveals a fluctuation in left-handedness rates across age cohorts. This observation is attributed not to a biological aging factor but to evolving societal norms that have influenced the acceptance of left-handedness over time.

Steering the analysis is a Bayesian framework, an approach known for its strength in handling probabilities and uncertainties. It's utilized to gauge the likelihood of being either left-handed or right-handed given an age at death. The process amalgamates the survey's empirical insights with carefully crafted prior probabilities. As a pivotal result, the posterior probabilities emerge, painting a nuanced picture of hand preference's relationship with age at death.

Leveraging these posterior probabilities, the analysis embarks on a profound exploration of the anticipated average ages at death for both left-handed and right-handed cohorts. An additional layer of intrigue is unveiled as the analysis sets out to validate or disprove the hypothesis suggesting a tangible disparity in the longevity of these two groups. To illustrate and solidify the findings, compelling visualizations are summoned to the forefront, showcasing various facets of the datasets. These visual representations not only enrich the analysis but also serve as a resounding endorsement of the drawn conclusions.

In essence, this analytical expedition combines the prowess of data-driven methodologies with a Bayesian underpinning to unravel the enigma surrounding the link between handedness and the span of life. Through meticulous exploration and strategic visualization, this analysis stands poised to contribute nuanced insights into a topic that intertwines human characteristics and the passage of time.

4.2 Summary of the findings:

The calculated divergence in average ages between individuals favouring their left hand and those favouring their right-hand amounts to 5.55 years. Interestingly, this value stands in contrast to the 9-year distinction identified in a study conducted in 1990. The crux of this variance lies in the evolving rates of left-handedness across the population over time—a heartening revelation for the left-handed community. Consequentially, the analysis casts aside the notion that left-handers are destined for premature mortality due to their sinistrality.

It's noteworthy that reported left-handedness rates have undergone a transformation, surging from a mere 3% in the early 1900s to approximately 11% in present times. This evolution implies that older generations are more likely to be categorized as right-handed, skewing the perception when assessing a group of recently deceased individuals. As a result, older right-handers disproportionately influence the dataset.

Yet, within the contours of this analysis, certain approximations warrant consideration as potential sources of accuracy variance:

- The employed death distribution data pertains to nearly a decade post the study year (1999 vs. 1991) and encompasses the entirety of the United States rather than being confined to California, the original focus of the study.
- The survey outcomes pertaining to left-handedness are extrapolated to encompass both younger and older age brackets. However, there's a chance that this extrapolation might not perfectly align with the actual rates for these specific age ranges.

Enhancing the precision of this analysis could entail leveraging random sampling techniques to gauge the estimate's variability in age gaps. This methodology would entail selecting a smaller subset of recently deceased individuals and ascribing handedness probabilities derived from the survey data. By observing the distribution of age gaps within this context, it becomes possible to determine the frequency at which a 9-year age gap would arise, given the same dataset and assumptions. While this analysis doesn't adopt this approach, it's certainly feasible to implement using the available data and the tools of random sampling.

A noteworthy revelation from this exploration is that the projected age gap in 2018 is notably narrower compared to 1990. This phenomenon can be attributed to the stagnation of left-handedness rates among individuals born post-1960. Consequently, both the National Geographic study and the 1990 research were conducted during a unique period when left-handedness rates were undergoing significant transformation across the lifespans of a majority of the populace, magnifying the distinctions between different age groups' handedness.

In essence, this analysis underscores the dynamic interplay between handedness, aging, and shifting societal norms. By navigating a landscape of statistical methodologies and drawing from expansive datasets, it offers insights into the intricacies of human characteristics and the complex tapestry of time.

Chapter: 5
FUTURE SCOPE

As we cast our gaze towards the horizon, this research endeavor beckons us towards a multitude of captivating avenues for further exploration and expansion within the realm of analyzing the disparity in death ages between right-handers and left-handers. The journey ahead is illuminated by the prospect of embarking on a comprehensive longitudinal study, spanning over an extended timeline, to meticulously trace the trajectories of individuals' hand preferences in tandem with their age-related health outcomes. This longitudinal approach promises a trove of data, promising a finer-grained understanding of how hand preference interlaces with longevity and age-induced health conditions across various life stages.

Moreover, the narrative of the future involves unearthing the intricate influencers that sway hand preference, entwining them with health and longevity outcomes. An array of factors beckons our attention: genetics, environmental forces, formative childhood experiences, cultural norms, and even epigenetic mechanisms. This pursuit could potentially unravel the enigma behind the observed disparities in death ages between right-handers and left-handers.

Amid this tapestry of possibilities, the influence of health behaviors emerges as a captivating avenue for exploration. Peering into the nuanced variations in health behaviors—spanning diet, exercise routines, smoking habits, alcohol consumption, stress management—amongst right-handers and left-handers could furnish enlightening revelations. Discerning any distinct patterns in these behaviors and their potential impacts on longevity might pave the way for targeted interventions and public health endeavors striving to elevate the health and overall well-being of both groups.

Another pivotal thread of inquiry spotlights the interplay between socioeconomic factors and the nexus between hand preference and differences in death ages. Delving into the intricate tapestry of socioeconomic parameters—ranging from income levels and educational attainment to healthcare accessibility and social support—unveils their influential sway over health outcomes. The exploration of whether disparities in these factors contribute to the divergence in life expectancies between the two groups holds substantial potential for fostering health equity.

The narrative of future exploration also encompasses a journey into the relationship between handedness and the risk associated with age-related health conditions. Scrutinizing the prevalence or outcomes of ailments such as cardiovascular diseases, neurodegenerative disorders, and cancer, offers the promise of targeted insights for disease prevention and personalized management strategies.

In this intricate web of inquiry, meta-analysis emerges as a powerful tool, synthesizing findings from a plethora of research endeavors. This methodological approach promises a more comprehensive, robust understanding of the intricate interplay between hand preference and the differences in death ages. It opens avenues for pinpointing sources of heterogeneity, gauging effect sizes with precision, and spotlighting gaps in existing literature.

The terrain of genetic investigation beckons, as we unravel the genomic underpinnings of handedness and its potential interlinkage with longevity. Peering into the genetic markers that align with hand preference and their potential associations with age-related health outcomes could illuminate the delicate dance between genetics, hand preference, and health.

In parallel, exploring the psychological dimension is equally poignant. Delving into the intricate interplay of psychological factors—ranging from stress and coping mechanisms to personality traits and mental health—could provide a more holistic understanding of the observed differences in death ages. This investigation unravels the potential intertwinement of psychological facets with biological and behavioral elements, molding health outcomes and life expectancies in unique ways.

On a practical front, intervention studies beckon, poised to steer health behaviors among left-handers toward more positive trajectories. The implementation of targeted interventions that champion healthier lifestyles, offer enhanced healthcare access, and mitigate potential risk factors for age-related ailments amongst left-handers holds the promise to bridge the longevity gap.

Navigating these uncharted territories mandates a meticulous adherence to sound study design, rigorous data collection methodologies, and precise statistical analyses. The orchestration of interdisciplinary collaborations with experts spanning epidemiology, genetics, psychology, and other pertinent domains promises a tapestry of enriched research endeavors. Throughout this journey, ethical considerations should stand as sentinels, safeguarding participants' rights and privacy.

As the curtain falls on this exploration, it becomes evident that the future holds a treasure trove of knowledge, waiting to be unearthed within the intricate nexus between handedness and the differences in death ages. This landscape beckons us to embark on a multidisciplinary voyage, weaving innovation and collaborative spirit into the fabric of scientific advancement. Through these endeavors, we stand poised to decipher the enigmatic connections between hand preference, health, and longevity, contributing not only to scientific acumen but also to the noble cause of public health initiatives.

Chapter: 6
CONCLUSION

Employing a meticulous and data-centric exploration, this project has successfully cast a revealing light on the enigmatic relationship between handedness and lifespan, particularly disentangling the persistent myth of premature demise among left-handers. Armed with the power of Bayesian statistics and an arsenal of age distribution data, we embarked on an exhaustive journey to discern whether a substantial disparity exists in the average age at death between those favouring their right hand and their left-hand counterparts.

Our revelations have dispelled the illusion of a premature death narrative for left-handers. Contrary to widespread preconceptions, our in-depth analysis has unveiled no notable discrepancy in the average age at which right-handers and left-handers meet their fate. This discovery aligns harmoniously with the emerging tide of research that seeks to disrupt outdated stereotypes and unfounded notions related to handedness.

The amalgamation of historical and contemporary research strands on handedness and mortality has painted our study with a richer palette, infusing it with a broader contextual awareness. We have underscored the significance of contextualizing cultural biases and societal perceptions that might have inadvertently nourished the perpetuation of myths and irrational beliefs intertwined with handedness.

Critical to our journey is the recognition that handedness is an intrinsic facet of human diversity, a tapestry woven with a multitude of hues. Any claims asserting innate drawbacks or benefits predicated on handedness must be greeted with a healthy dose of skepticism and subjected to rigorous empirical scrutiny. Our foray into research adds another brick to the edifice of evidence, firmly advocating for the dissolution of baseless assumptions and the cultivation of an all-encompassing comprehension of human distinctions.

In summation, this project stands as a significant contribution to the realm of handedness exploration. By dismantling the myth of premature demise for left-handers via a robust and structured analysis, we have advanced the discourse. Our utilization of statistical prowess and our dedication to data-driven methodology have unfurled evidence-backed revelations that enrich our comprehension of the intricate interplay between handedness and mortality. As we forge ahead, embracing the multifaceted aspects of human divergence, we must maintain our dedication to scrutinizing prevailing narratives, anchoring our conclusions in empirical exploration, and championing a comprehensive and enlightened outlook on the kaleidoscope of human traits.

REFERENCES

Data Collection:

Data Collection The following websites have been referred to obtain the input data and statistics:

- https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/a ec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv
- https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62 f3ec07514c7e31f5979beeca86f19991540796/cdc_vs00199_table310.tsv

Programming References:

The following websites have been referred for matplotlib and data Science using pandas tutorials:

- https://www.digitalocean.com/community/tutorials/data-analysis-andvisualization-with-pandas-and-jupyter-notebook-in-python-3
- https://www.learndatasci.com/tutorials/python-pandas-tutorial-completeintroduction-for-beginners/
- https://pandas.pydata.org/getting_started.html
- https://www.dataquest.io/blog/jupyter-notebook-tutorial/
- https://github.com/TirendazAcademy/PANDAS-TUTORIAL