Final Report of Traineeship Program 2020

On

"DEATH AGE DIFFERENCE OF LEFT AND RIGHT-HANDERS"

MEDTOUREASY

Last Updated – 23rd July 2023

ACKNOWLEDGMENTS

The traineeship opportunity that I had with MedTourEasy was a great chance for learning and understanding the intricacies of the subject of Data visualization in DataAnalytics; and also, for personal as well as professional development. I am very obliged for having a chance to interact with so many professionals who guided me throughout the traineeship project and made it a great learning curve for me.

Firstly, I express my deepest gratitude and special thanks to the Training & Development Team of MedTourEasy who allowed me to carry out my traineeship at their esteemed organization. Also, I express my thanks to the team for making me understand the details of the Data Analytics profile and trained me in the same so that I can carry out the project properly and with maximum client satisfaction and also for spearing his valuable time despite his busy schedule.

I would also like to thank the team of MedTourEasy and my colleagues who made the working environment productive and very conducive.

TABLE OF CONTENTS

ABSTRACT

The relationship between handedness and longevity has been a topic of interest for centuries. Some studies have suggested that left-handed people may have a shorter lifespan than right-handed people, while others have found no such difference.

This study used data from a National Geographic Survey in 1986 to analyze the death age difference between right-handers and left-handers. This survey resulted in over a million responses that included age, sex, and hand preference for throwing and writing.

The results of the analysis showed that there was a significant difference in the average age at death between right-handers and left-handers. The average age at death for right-handers was 72.79 years, and the average age at death for left-handers was 67.24 years.

The results of this study do not support the hypothesis that left-handed people have a shorter lifespan than right-handed people. As you know that this survey was done in 1986, if the same survey was done today, we should expect a shifted version of the same distribution as a function of age.

This study provides valuable insights into the relationship between handedness and longevity. The result of this study suggests that there is a pretty big age gap in the average age at death between right-handers and left-handers. However, more research is needed to confirm these findings and to identify the factors that may contribute to any difference in lifespan between right-handers and left-handers.

1. INTRODUCTION

1.1 About the Company

MedTourEasy, a global healthcare company, provides you with the informationalresources needed to evaluate your global options. MedTourEasy provides analytical solutions to our partner healthcare providers globally.

1.2 About the Project

The project analysis of the death age difference between the right-handers and left-handers is a valuable study that provides insights into the relationship between handedness and longevity. The results of the study suggest that there is a significant difference in the average age at death between right-handers and left-handers.

The project was conducted using the data from National Geographic Survey 1986, which resulted in over a million responses that included age, sex, and hand preference for throwing and writing.

The average age at death for the right-handers was 72.79 years, and the average age at death for left-handers was 67.24 years.

This study provides valuable insights into the relationship between handedness and the average age at death between right-handers and left0-handers.

The project also has some limitations. For Example, the data from National Geographic Survey 1986 does not include people from different geographical locations they only included people from California alone. This means that it is not possible to say for sure whether the difference in lifespan between right-handers and left-handers is due to handedness or other factors, such as geographical location. Despite these limitations, this study provides valuable insights into the relationship between handedness and longevity.

Hence, this project aims at analyzing datasets to create intuitive and interactive visualizations for representing age differences between

right-handers and left-handers to gain meaningful insights.

1.3 Objectives and Deliverables

This project focuses on creating an easily understandable and interactive visualization by gathering data on the ages of left-handers and right-handers from various sources like National Geographic Survey 1986 and Death distribution data for the United States in 1999 and using the coding languages Python and libraries like pandas, numpy, date time and matplotlib to visualize these statistics which will enable the firm to analyze the situation and draw conclusions. The main objective is to determine if there is a significant difference in the average age at death between right-handers and left-handers.

The project consists of deliverables as follows:

a) A report that summarizes the findings of the study.

b) A presentation that can be given to stakeholders.

c) A set of recommendations for future research.

2. METHODOLOGY

2.1 Flow of the Project

The project followed the following steps to accomplish the desired objectives anddeliverables. Each step has been explained in detail in the following section.

2.2 Language and Platform Used

The language and platform used to analyze the death age difference between right-handers and left-handers will depend on the specific project and the available resources. However, the language that is used for data analysis includes:

Language**: Python**

It is a popular programming language that is often used for data analysis and data visualization. It is a general-purpose language that is easy to learn and use. There are many libraries and packages available for Python that can be used for data analysis, such as NumPy, Pandas, and Matplotlib.

The platform that is used to analyze the data will also depend on the specific project and the available resources. However, some common platforms that are used for data analysis include:

- Desktop computers: Desktop computers are often used for data analysis because they have the power and resources that are needed for complex analysis.
- Laptop computers: Laptop computers are also used for data analysis, but they may not have the same power and resources as desktop computers.
- Cloud computing: Cloud computing platforms, such as Amazon Web Services (AWS) and Microsoft Azure, can be used to run data analysis tasks. Cloud computing platforms offer a variety of benefits, such as scalability, flexibility, and cost-effectiveness.

3. IMPLEMENTATION

3.1 Gathering Requirements and Defining Problem Statement

This is the first step wherein the requirements are collected from the clients to understand the deliverables and goals to be achieved after which a problem statement is defined which to be adhered to while developing the project.

Problem statements:

- Is there a significant difference in the average age at death between right-handers and left-handers?
- If there is a difference, what are the possible explanations for this difference?
- Are there any other factors that may contribute to the difference in lifespan between right-handers and left-handers?

3.2 Data Collection

Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes and forecast future probabilities and patterns. It is a research component in all study fields.

The data on the death age difference between right-handers and left-handers have been collected through GitHub repositories, mentioned as follows:

- National Geographic Survey 1986
- Death distribution data of the US for the year 1999

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

Packages Used:

Pandas: Pandas is a package for data analysis in Python. It provides several tools for working with tabular data, such as loading, cleaning, and manipulating data.

Functions Used:

pd.read_csv: It is a function in the Pandas library that is used to read CSV files into Pandas DataFrames. The function takes a path to csv as input and returns a Pandas DataFrame.

```
# loading the data
data_url_1 = "https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dac
lefthanded_data = pd.read_csv(data_url_1)
```

3.3 Data Cleaning

Data cleaning is the process of identifying and correcting errors in data. It is an important step in the data analysis process, as it ensures that the data is accurate and reliable.

Concerning the left-handed dataset and death distribution dataset, it may contain null values or incorrect values. Various functions are used to clean these datasets.

Functions used:

Drop_na(): This function in Python is used to drop rows or columns that contain missing values. The function takes a Pandas DataFrame as input and returns a new DataFrame without the missing values.

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

3.4 Data Analysis

The data analysis technique that is used will depend on the specific data that is being analyzed and the research questions that are being asked.

Concerning our dataset, we use the following functions and libraries for data analysis:

Matplotlib: It is a Python library used for creating static, animated, and interactive visualizations. It is a popular choice for data visualization because it is easy to use and has a wide range of features. It can create visualization such as line plots, bar charts, pie charts, scatter plots, and histograms.

```
# plot number of people who died as a function of age
death_distribution_data.plot(x = 'Age',y = 'Both Sexes', marker='o', c = 'orange')
plt.title('Number of people who died as a function of age')
plt.ylabel('Both Sexes')
```

Merging two or more columns: Using functions like '+', '-', '*', and '/' we can perform many operations in the datasets.

```
# created a new column for the average of male and female
lefthanded_data['Mean_lh'] = (lefthanded_data['Male'] + lefthanded_data['Female'])/2
lefthanded_data
```

Bayesian statistics: It is a statistical framework that uses Bayes' theorem to update our beliefs about the probability of an event based on new evidence. It is a mathematical formula that relates the probability of an event to the probability of its 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.

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

NumPy: It is a Python library for scientific computing. It provides a high-performance multidimensional array object, along with a large library of mathematical functions to operate on arrays. It is fast, efficient, easy to use, and accurate.

```
# 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))
```

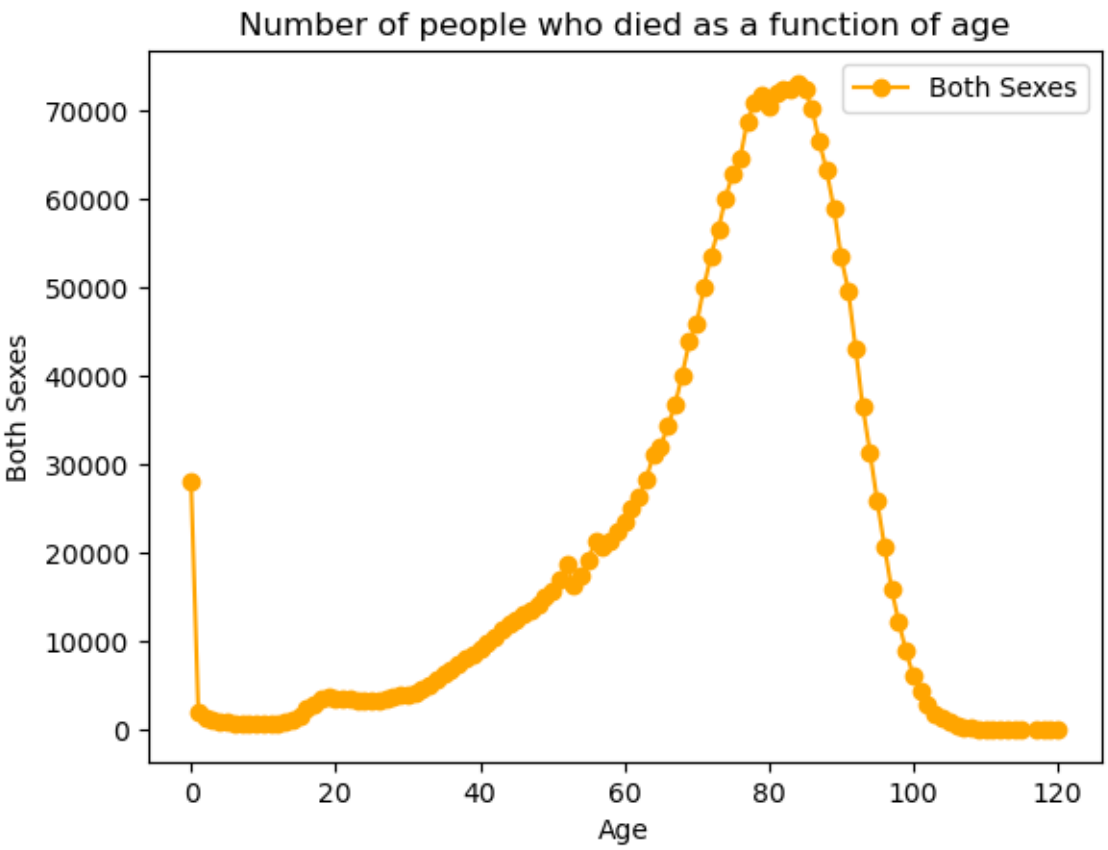
3.5 Interpretation of Data

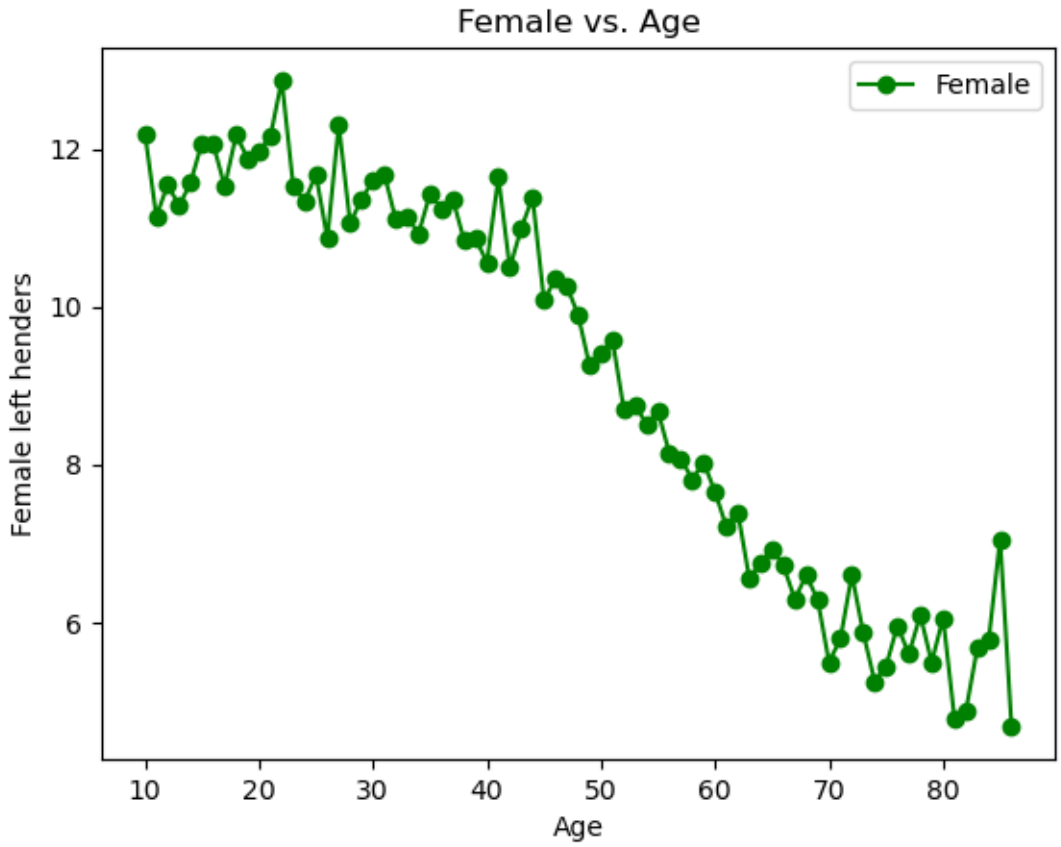
Data interpretation is the process of understanding what the data means. It is the process of making sense of the data and drawing conclusions from it.

The interpretation of data can be done in several ways, depending on the type of data and the goals of the interpretation. Some common methods of data interpretation include:

- Descriptive statistics: Descriptive statistics are used to summarize the data and to describe its main features. This includes measures such as the mean, median, and standard deviation.
- Hypothesis testing: It is used to test a hypothesis about the data. This involves making predictions about the data and then testing those predictions using statistical methods such as the Bayesian theorem.
- Data visualization: data visualization is the use of graphical representations to communicate the data. This can be a useful way to understand the data and to communicate its findings to others.

The interpretation of data is an important part of the data analysis process. It is the process of making sense of the data and drawing conclusions from it. The interpretation of data can be used to answer questions, make decisions, and solve problems.


Number of people who died as a function of age
Both Sexes
70000
60000
50000
40000
30000
20000
10000
0
Both Sexes
0
20
40
60
80
100
120
Age



4. SAMPLE SCREENSHOTS AND OBSERVATIONS

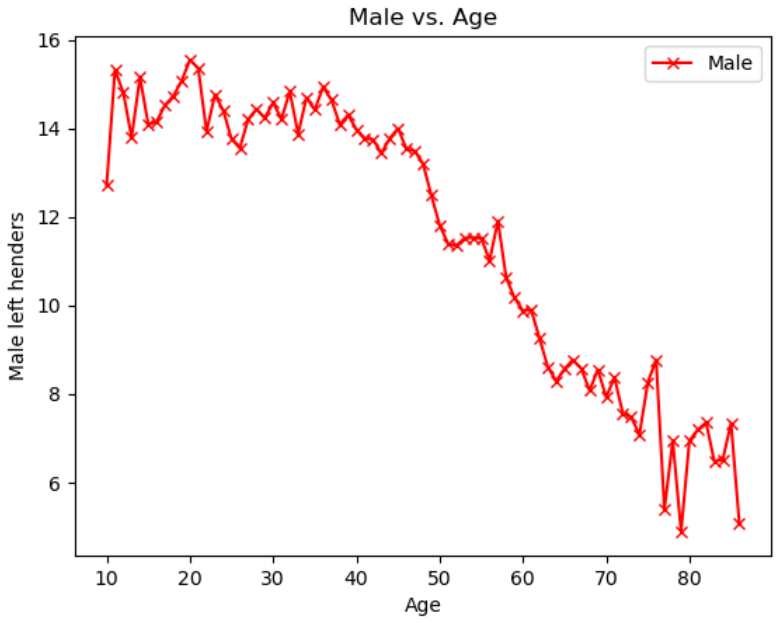
4.1 Graph Plots

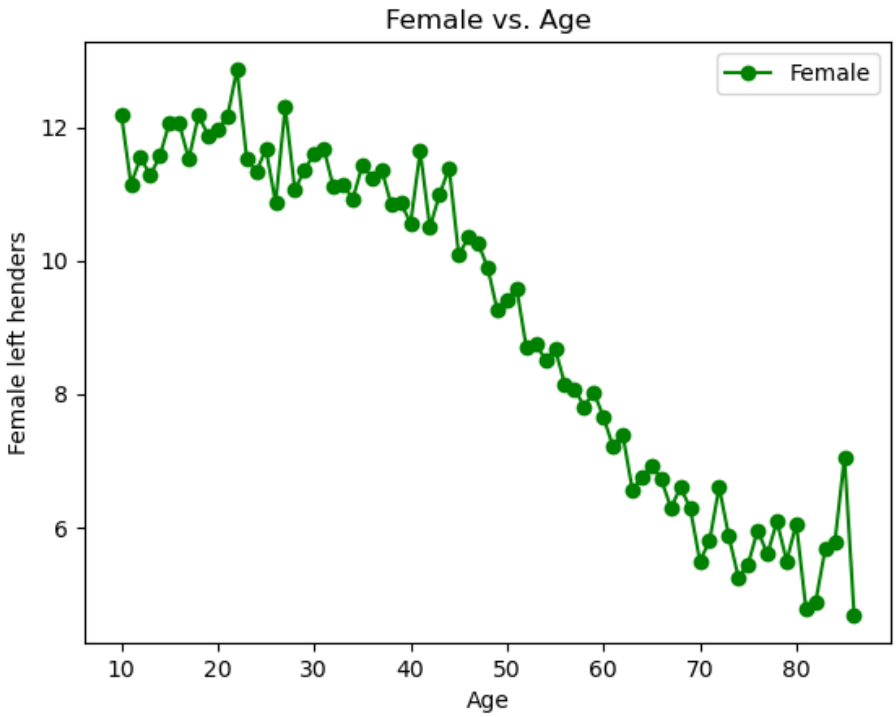
- **Male and female left-handers rate vs age:**
  The graph shows the number of males and females between the ages of 10 and 86. The x-axis shows the age, and the y-axis shows the number of males and females.
  The graph shows that the number of males and females decreases as the age increases. This is to be expected, as the number of people, in general, decreases as the age increases.

  However, the graph also shows that the number of males and females decreases at a faster rate after the age of 60. This may be due to several factors, such as increased mortality rates among older males and females.

Here are some additional observations about the plot:

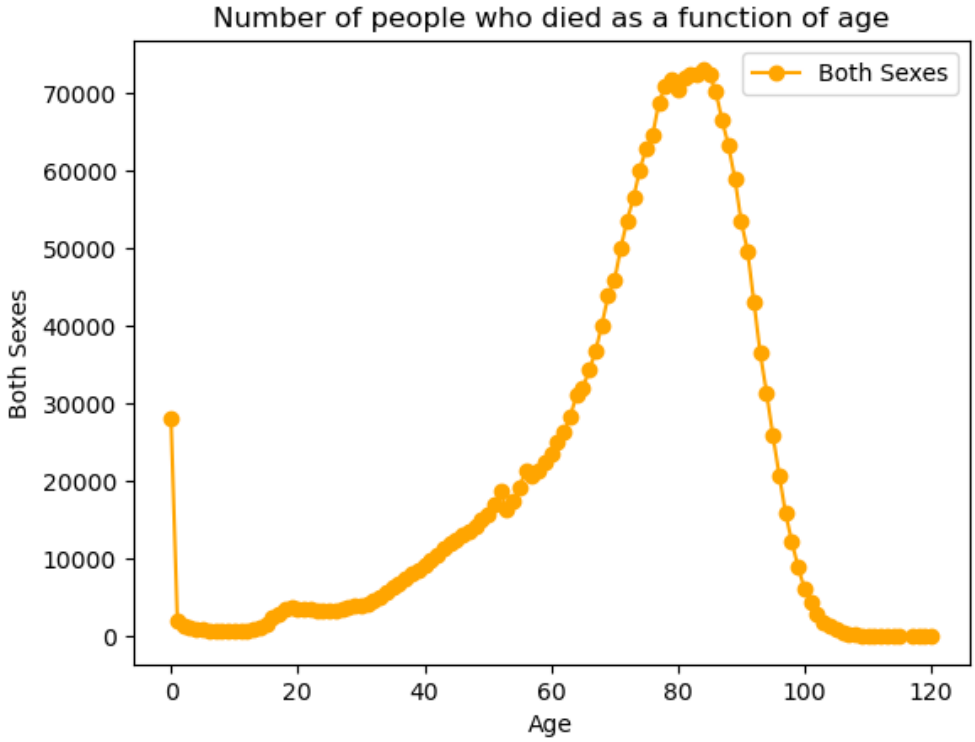
1. The number of males and females is highest in the age range of 20-30.
2. The number of males and females decreased steadily after the age of 60.
3. There is a small increase in the number of males and females in the age range of 70-8- years.





- **Number of people who died at a certain age:**
  The x-axis shows the age, and the y-axis shows the number of people who died. The graph shows that the number of people who die increases as the age increases. This is to be expected, as the number of people, in general, die as their age increases.
  However, the graph also shows the number of people who die at a faster rate after the age of 60. This may be due to several factors, such as increased mortality rates among older people.

  Here are some additional observations about the plot:
  1. The number of deaths is highest in the age range of 70-90.
  2. The number of deaths increases steadily after the age of 60.
  3. There is a decrease in the number of deaths in the age range of 90-100.



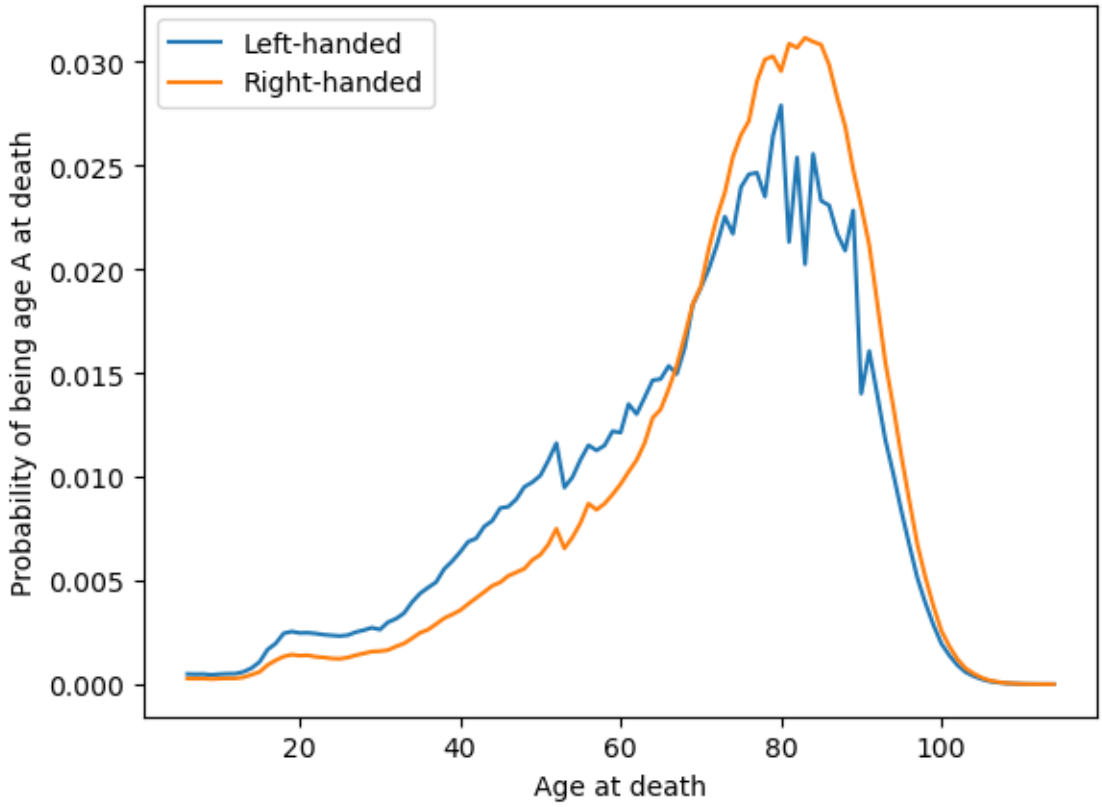
- **Number of people who die were left or right-handers:**
  The graph shows the probability of being at age A at death for left-handed and right-handed people. The x-axis shows the age, and the y-axis shows the probability of being at that age of death. The line graph shows the trend of the data for left-handed people with a blue color line and the orange color line shows the data for right-handed people.

  The graph shows that the probability of being at age A at death decreases as the age increases for both left-handed and right-handed people. However, the probability of being at age A at death is slightly higher for left-handed people than for right-handed people. This means that left-handed people are more likely to live longer than right-handed people.

  Here are some additional observations about the plot:
  1. The probability of being at age A at death is highest for left-handed people in the age range of 70-80.
  2. The probability of being at age A at death decreases steadily for both left-handed and right-handed people after the age of

80.

3. There is a small increase in the probability of being at age A at death for right-handed people in the age range of 90-100.



4.2 Moment of Truth:

- The average age of left-handed people was 67.24 years in 1990, and the average age of right-handed people was 72.79 years in 1990.
- This means that left-handed people are on average die 5.5 years younger than right-handed people.
- This means that right-handed people may be more likely to live longer than left-handed people.

6. CONCLUSION AND FUTURE SCOPE

o Conclusion

1. The average age difference between left-handed people and right-handed people was 5.5 years in 1990.
2. 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. The reported rates of left-handedness have increased from just 3% in the early 1990s to about 11% today, which means that older people are much more likely to be reported as right-handed than left-handed.
3. The difference in average age could be due to several factors, such as genetics, environment, and lifestyle.
4. The results of this study do not support the hypothesis that left-handed people have a shorter lifespan than right-handed people. The main limitation of this study is the relatively small sample size and limited geographic area covered. Larger studies are needed to confirm these findings.

o Future Scope

1) Future research could focus on identifying the specific gene, environmental factors, and lifestyle choices that are associated with the difference in average age between left-handers and right-handers.
2) This research could help to develop interventions that could help to improve the health and longevity of both left-handers and right-handers.
3) Future research should focus on replicating the findings of this study using a larger sample size.

7. REFERENCES

Data Collection:

a) Death Distribution Data

b) Left-handed People Data

c) Hand Preference and Age in the United States

d) Mortality Tables

Programming Reference:

a) Github

b) Python library

c) Bayes' Theorem