**Homework 1**

**Hypothesis:**

The number of pedestrians compared to people using bikes has significantly increased from the year 2023 to 2024 in Somerville, MA.

**Problem Statement:**

By analyzing trends in the number of pedestrians compared to people using bikes, insights into shifts in modes of transportation can be gained. This analysis can lead to improved safety measures and more effective planning of future infrastructures.

The mode of transportation data from the years 2023 and 2024 will be analyzed to determine the counts for pedestrians and cyclists. Comparing these counts will help identify whether there was a significant increase in pedestrians compared to cyclists in 2024 relative to 2023.

**How the hypothesis will be proved or disproved:**

The hypothesis will be proved or disproved by using data from the mode, count, and year columns from the dataset. The year column will be used to analyze data from 2023 and 2024. The mode column indicates whether the entry is for a bike or pedestrian (ped) and corresponds to the count column. By retrieving the bike entries along with their corresponding count entries, the total number of people who rode bikes can be calculated. Similarly, the pedestrian entries and their counts will be collected to determine the total number of pedestrians. The difference in pedestrian counts between 2023 and 2024 will be calculated, and the same process will be applied to the bike counts. The differences in totals for bikes and pedestrians will then be compared. By comparing the differences obtained from the totals, the hypothesis can be validated or disproved.

**Documentation:**

* The data set being used is from the data.gov website: <https://data.somervillema.gov/api/views/qu9x-4xq5/rows.csv?accessType=DOWNLOAD>
* curl ‘https://data.somervillema.gov/api/views/qu9x-4xq5/rows.csv?accessType=DOWNLOAD’ > bike-pedestrian.csv was used to copy the data from the website and add it as a csv file to my current directory
* Columns being used are 1,7, and 8
* The awk command was used to create a file called “selected-bike-ped.csv” which contains only columns of interest
* Used head to copy the column names to a new file called “filtered-years-bp.csv”
* Used an awk command to obtain only data from years 2023 and 2024 then added it the filtered-years-bp.csv file
* Created an awk command that showed the total pedestrians in 2023 which was 23, 932
* Modified the previous awk command to get the total number of cyclists in 2023 which was 12, 468
* Modified the awk statement again to get the total number of cyclists in 2024 which was 12, 095
* Modified the awk command to get the total number pedestrians in 2024 which was 23, 743

**Conclusion:**

The hypothesis was disproven. The number of both bike riders and pedestrians decreased in 2024 compared to 2023. The decrease in bike riders was more significant than the decrease in pedestrians.