**Hypothesis:**

The number of pedestrians compared to people using bikes has significantly increased from the 2015 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 2015 and 2024 will be analyzed to determine the counts for pedestrians and cyclists then getting the ratio of “Pedestrians-to-Cyclists. Looking at the ratio from each year will help identify whether there was a significant increase in pedestrians compared to cyclists from 2015 to 2024.

**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 2015 to 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 total number of pedestrians will then be divided by the total number of cyclists to obtain a “Pedestrian-to-Cyclists ratio.” If the ratio after each year is 2 or greater, then that means it’s a 2:1 ratio. If the ratio is less than 2, then that means it’s a 1:2 ratio. If the ratio is exactly 1, then that means that it’s a 1:1 ratio. Each ratio from all the years of the selected decade can determine a trend which will can validate or disprove the hypothesis.

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 an awk command that counted and added up the total pedestrians and cyclists each year then divided total pedestrians by total cyclists to get a ratio for “Pedestrians-to-Cyclists” for each year.
  + awk '{count[$1][$2] += $3} END {for (year in count) print year, count[year]["Ped"]/count[year]["Bike"]}' filtered-time-period.csv
* The same awk command from above was used to create the file called “pedestrian-to-cyclist-ratio.csv” to contain ratio findings from each year.

**Conclusion:**

The hypothesis was disproven. The amount of pedestrians increased slightly, reached a peak during 2021, then decreased.