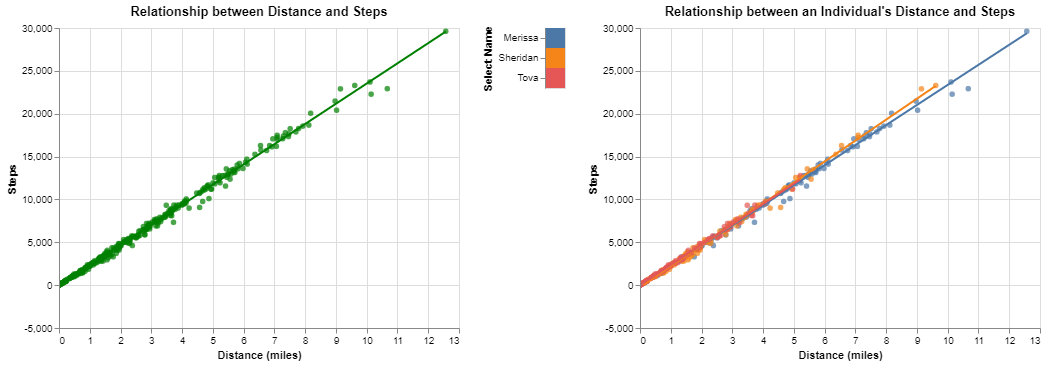
**[The Data](https://gist.githubusercontent.com/mlissad000/92d81416c1a087cd4b80d4ed2eb6a8cf/raw/1c2f1cbe1b46b850d04b1a4e414835ed91a50bd2/dataviz_project_data)**

The data we will be using for this project is daily step data combined with daily temperature data from June 1st, 2019 to September 29th, 2019. The dataset has the following variables: date, distance walked (in miles) for Merissa, Sheridan, and Tova, steps walked for Merissa, Sheridan, and Tova, and the daily temperature.

**Visualization 1 - Sheridan Kamal**

The following visualizations were made using the Altair library in Python3 and seek to explore the effect distance has on steps, both in general and on an individual basis. The fully interactive version of this visualization can be found here: <https://csc83060-project-1.skamal.repl.co/> (click on the colors in the legend to filter the visualization on the right by name). The HTML5 code for the interaction can be found here: <https://repl.it/@skamal/CSC83060-Project-1>. The Python3 code for the project can be found here: <https://repl.it/@skamal/CSC83060-Project1>.



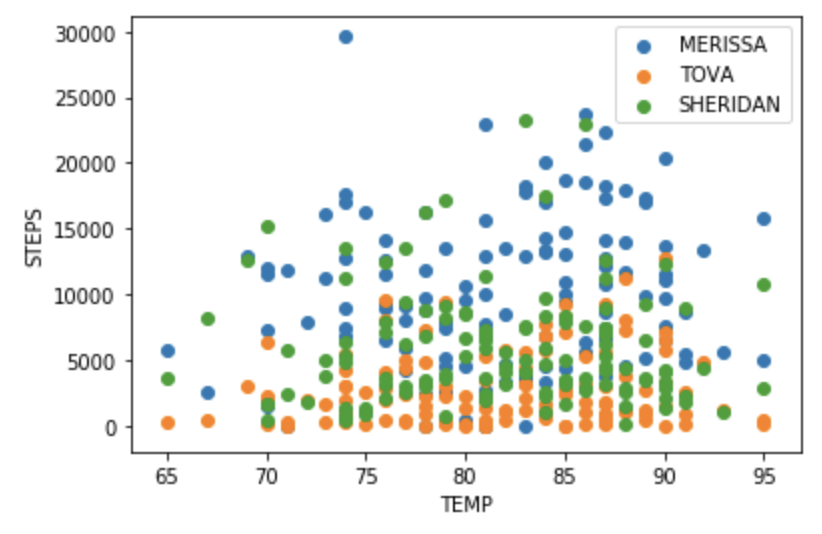
There seems to be an overall positive linear relationship between distance and the amount of steps walked as shown by the visual on the left. This assumption still holds true when we plot each individual's distance against the steps walked. By comparing each individual's trendline, we can compare the stride of the individuals. Sheridan has the highest coefficient on distance (2443.92), which means she needs to walk more steps to walk the same distance, so her stride is the smallest. Merissa has the smallest coefficient on distance (2341.15), which means she needs to walk less steps to walk the same distance, so her stride is the greatest. Tova's stride falls in the middle of the two, but is closest to Merissa's stride because the coefficient on distance is 2371.47. Each trendline (both overall and each individual’s) is significant at the 1% level having p-values < 0.0001 and having R^2 values between 0.994 and 0.996.

**Visualization 2 - Tova Schwartz**

The second visualization seeks to incorporate the daily temperature in NYC into the analysis. The visualization was created using the Python Matplotlib Library. The original intention for this plot was to map all three lines on one plot. However, the constant overlapping of the lines made the visualization highly uninterpretable without interaction. This visualization is a time-series analysis: the number of steps taken daily (y-axis) is plotted over the course of the summer (x-axis). This format for a visualization lends itself greatly to an interaction between several visualizations. It provides the base work for indexing connected graphs based on time or temperature.

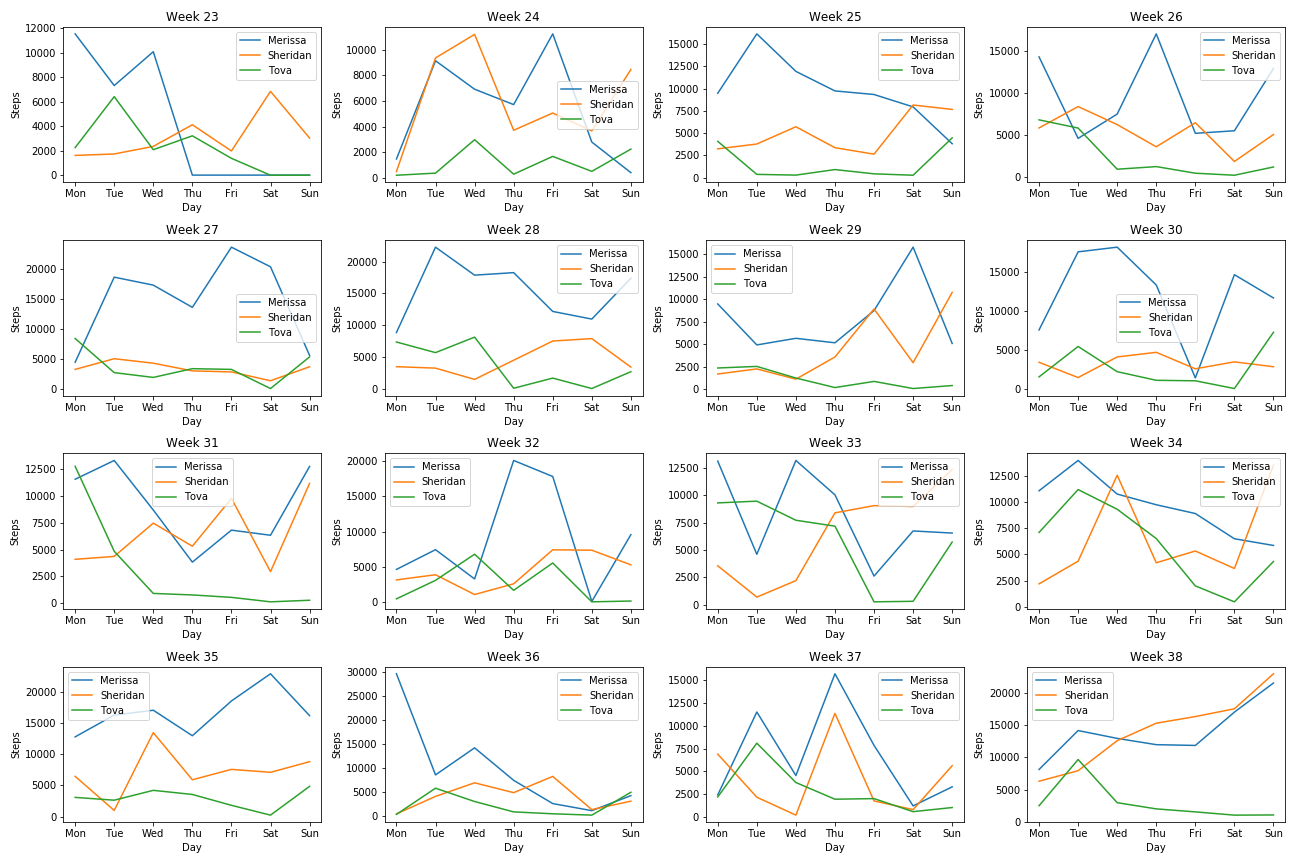


The original assumption was that days with milder weather would result in a greater number of steps taken. This is because in very high temperatures people opt to stay in door or commute by car in contrast to taking a walk when the weather is beautiful. However, a cursory analysis of this visualization provides no support for said hypothesis. Infact, after plotting all the points with temperature on the x-axis and steps on the y-axis it seems like temperature and steps have a zero or slightly positive correlation.

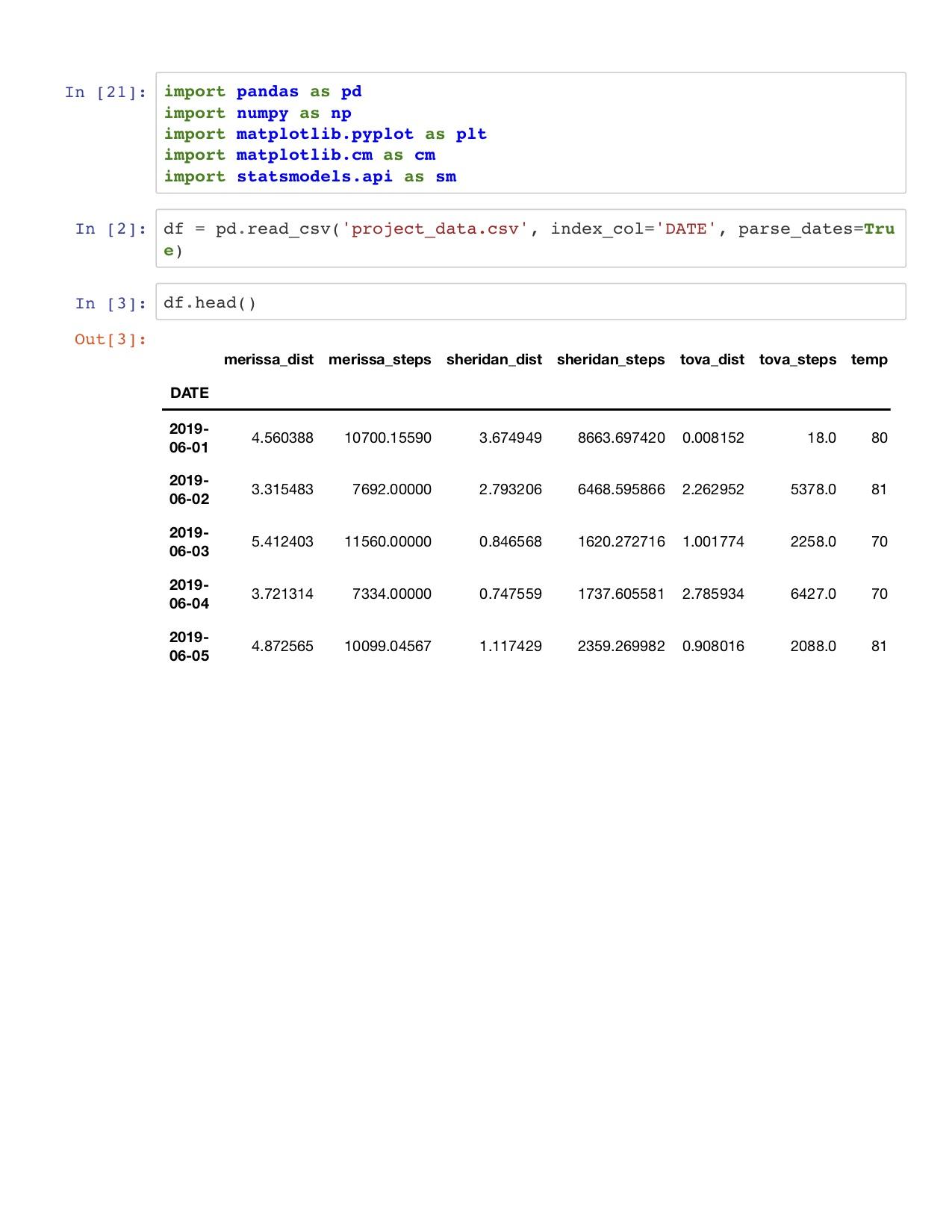


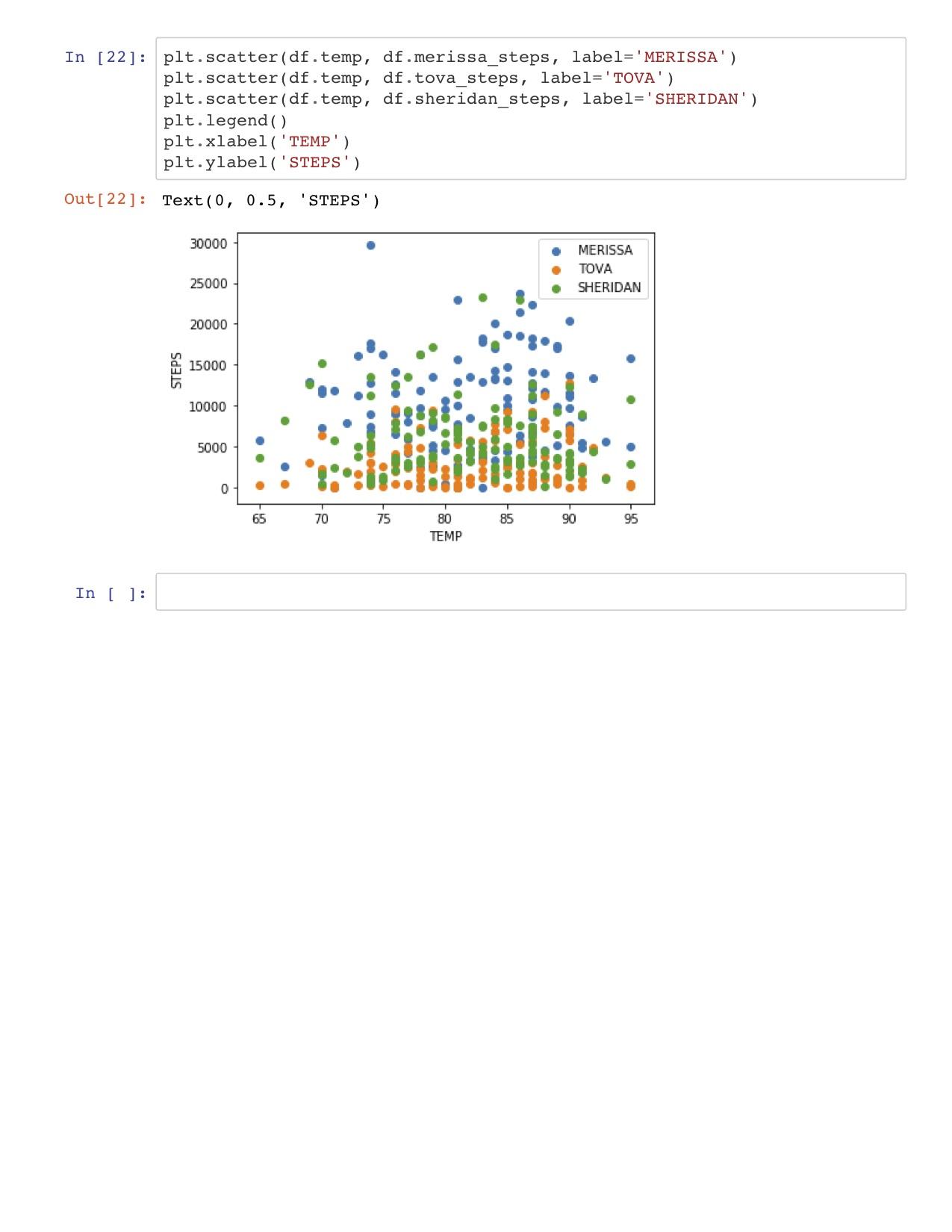
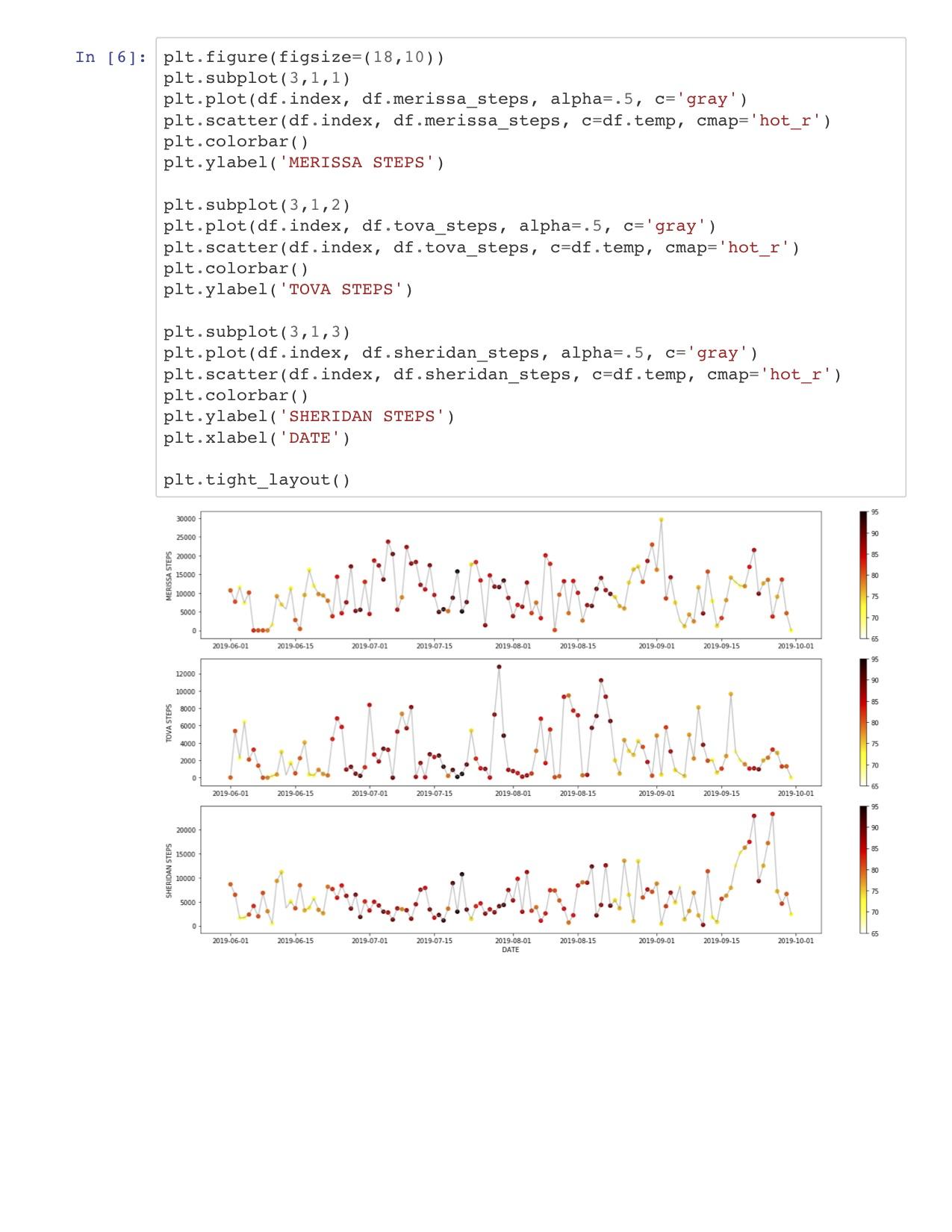
Code for the two visualizations shown above can be found at the end of this paper.

**Visualization 3 - Merissa Lissade**

The visualization in this section was created using Python Matplotlib Library via Jupyter Notebook. We wondered if there were any cyclic patterns that could be seen on a weekly basis. In this visualization, weeks 23 - 26 include the days in June, weeks 27 - 31 include the days in July, weeks 31-35 include the days in August, and weeks 35-38 include the days in September. We figured given some of us may have certain daily patterns, that maybe this would reflect in our step data as well. However, it can be seen from the visualization that there does not appear to be a particular day of the week or any type of pattern in which any individual walked the most.

**Code for Visualization 2 - Tova Schwartz**





**Code for Visualization 3 - Merissa Lissade**

