For my homework, I am visualizing a week's worth of data gathered in San Diego, CA from the National Oceanic and Atmospheric Administration's website (see link below). The data gathered contains the date and time of measurement (in 6 minute intervals), water levels (in meters), and error in prediction of water levels (in meters). The time interval for the data used begins on 01/29/2017 and ends on 02/04/2017. To begin, I exported the data from the NOAA website in the format of XML. I loaded each unique value type into its own array, knowing that accessing a specific index in any of these arrays will represent the appropriate data for a given measurement.

For my first visualization (plotA), I wanted to display the crests and peaks in the water levels over the week span, while still reflecting how accurate the predictions were versus the actual measured levels. Therefore, I plotted the points so that the y-axis represented the water level, the x-axis represented time, and the color of each point represented how accurate the water level prediction was by scaling between green and red. Green reflected data measurements that were close to the predicted levels and red reflected data measurements that were far from the predicted levels. I also included user interaction to enhance this visualization through displaying details such as the water level, date, and time of the measurement that coincided with the X position of the mouse. I choose this approach because with this additional user interaction, one is able to clearly see at what time each point occurs, as well as the actual water level at each measurement, which could not be determined through looking at the plot alone.

For my second visualization (plotC), I wanted to somehow display the total prediction error accumulated at a given time during a 24-hour day. To do this, I added animations to better convey the progression of error accumulation across time. I created ellipses that correlate to each day from the week-long data measurement. The ellipses grow depending on the prediction error from each successive time. Therefore as the time of day progresses, the size of each ellipse reflects the total prediction errors from each of the measurements, and one is able to see which days had greater prediction errors. The colors of each ellipse are scaled so that the reddest bubble has the largest total error, and the greenest bubble has the least total error at any given time. I choose this approach because this helps identify which days are experiencing the greatest prediction errors in comparison to the other days in the visualization.

I found that in my first visualization (plotA), the utility of it significantly increased when I decided to display the water levels at each point scrolled over. Without this information, there was a lot more ambiguity about what each of the points represented, and made the data/visualization a bit harder to interpret. Aesthetically, I felt like my first visualization could have been improved if I had differentiated each point on the plot better so that there was less overlap/ambiguity of points at crests and peaks.

I found that in my second visualization (plotC), the time animation is somewhat effective in showing the accumulation of total error over the course of a day. However, the time animation could have been better controlled (like possibly through some input that controls speed of time, allows pausing, etc.). This way one could better analyze the difference in time error accumulation at a specific time during the day (through the ability to pause).

https://tidesandcurrents.noaa.gov/waterlevels.html?id=9410170&units=metric&bdate=20170129&edate=20170204&timezone=GMT&datum=MLLW&interval=6&action=data