Weather data is prevalent, yet messy. Most likely not the easiest project to kick off a data science career, an analysis of Hurricane Harvey seemed a challenging and rewarding project that quite literally hits close to home. Many lessons were learned in this short journey with regard to data sourcing and acquisition, cleaning and parsing, visualizing and conveying – Data has a story to tell. We all greatly enjoyed using some newly acquired skills to piece together out own small bit of this st. The realization grasped fully that there is vast power and much to be learned in this exciting field.

The Gulf of Mexico Coastal Ocean Research “is the heart of data collection for ocean and control waters of the Gulf”. This resource was used to pull rig and weather station web-accessible-folder data in .csv format for cleaning and analysis. The Harvey path was collected by BEST!? Path. The actual data pull relied on a mixture of availability of sensors, stations, and rigs in a proximity that was cohesive to showing result, and completeness of data acquisition in the actual timeframe of the hurricane passage.

All .csv data (air temperatures, water temperatures, air pressures, water salinities, and wind speeds) was obtained and sorted by platform/station. It was then merged via an outer join technique via acquisition date. Much of the data actually aligned quite nicely due the to the apparent standardization of half-hourly interval. Even though it aligned nicely, the cleaning process was not as straightforward.

As stated, the learning curves was fast and brutal. Weather data is rarely complete, valid, etc etc etc