# Problem statement and hypothesis

Using observations and measurements of solar activity (sun spots, solar flares, coronal mass ejections) can we predict geomagnetic activity on Earth?

# Description of your data set and how it was obtained

The data was obtained from the National Geophysical Data Center which records and aggregates information from several observatory stations on Earth. Satellite observations of solar activity are from NASA’s SOHO LASCO CME CATALOG. Both sources of information are in .csv files.

I explored many catalogs of information, for example SpaceWeatherLive.com has a fantastic searchable database but it was not suitable for downloading a series of events.

# Description of any pre-processing steps you took

All data sets are in a fixed file format with footnote indicators embedded in the data. It was necessary to separate out those indicators into separate fields for interpretation.

# What you learned from exploring the data, including visualizations

It was clear from the simple visualization that the relationship between events on the sun and those on the Earth as not a simple relationship.

# How you chose which features to use in your analysis

Still a work in progress. I originally thought to focus on the CME features but am now pivoting to using counts of CME’s and solar flares.

# Details of your modeling process, including how you selected your models and validated them

Still a work in progress. Thinking that I’m going to use a clustering technique to classify the events that cause massive geomagnetic disturbances.

# Your challenges and successes

Biggest recent challenge was finding the right catalog of information on the solar flares

# Possible extensions or business applications of your project

I’ve read a couple of papers describing the use of neural networks for the prediction of sunspots but it seems like the use of these predictive techniques in this field is just emerging.

# Conclusions and key learnings