**PSPT Sunspots**

CU Boulder CSCI 7000 Semester Project

Jamie Mothersbaugh

**Project Overview**

Parts of the project pitch will be included in this section.

**Data Acquisition & Organization**

The PSPT data is available at <http://lasp.colorado.edu/pspt_access/>. Instructions for downloading the data are on the ‘Downloads’ tab at the top of the page. The LaTiS method is the most efficient for downloading the data.

The entire data set is ~2 TB. It is important to download the data in reasonably-sized increments- for example, downloading data in all filters in 1 month or downloading all data in 1 filter for 1 year. I obtained the data from an internal server at LASP, so I was able to bypass the LaTiS method without downloading the entire data set. However, any public user will need to obtain all the data before proceeding through the steps below.

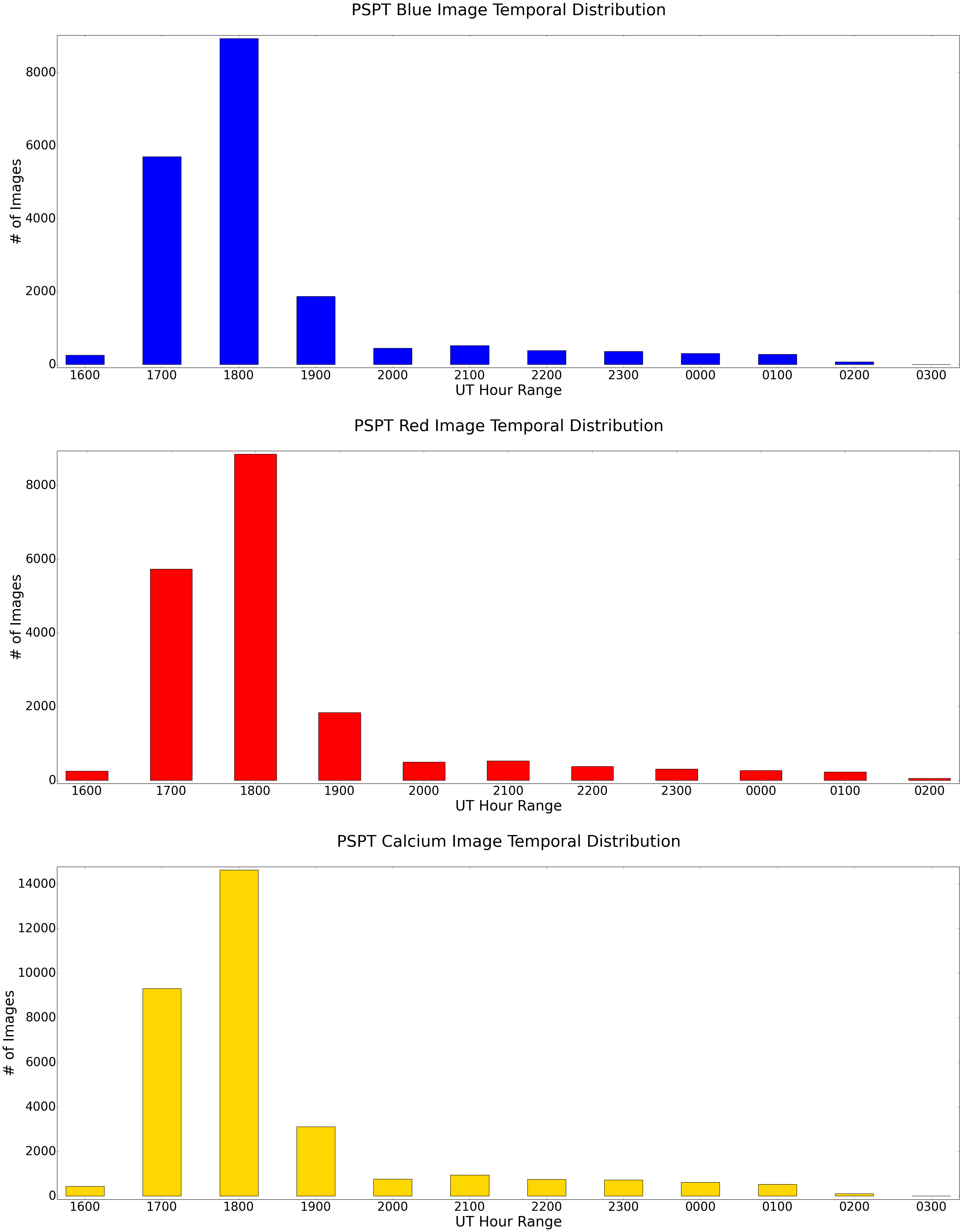
Once I obtained all the images, I looked at the release notes and the information on the PSPT website. These indicate the best images were taken between 16:00 and 19:00 UT. The next step was to trim down the data to include only the highest-quality images.

$ ls year/month/day/filter/\*.HourHour\*jpg | wc –l

This command lists the number of filenames that start with the chosen .HourHour timestamp value.

I determined the number of images in each hourly bin from 1600-0300 hours UT time. For each filter, I made a bar graph showing the image distribution as a function of hourly bin. This graph is made by running pspt\_data.py:

$ python pspt\_data.py

The script produces the plot below.

For all 3 filters, 1800-1900 hours contains the most images. However, looking at the data, there are often multiple images in this hour range on any given day, taken minutes apart. Because the goal of the project is to track the sunspot frequency over long (decades) timescales, I decided to trim the data down to contain only 1 image per filter per day on the days for which there is data. This can be done using the commands below.

$ ls \*jpg > filter\_jpg\_all.txt

This command writes all filenames into a text file. Not all images can be listed at once because there are too many files. List by subgroup instead- e.g., 199\*, 200\*, 201\*. To list the images into a file that already exists:

$ ls \*jpg >> filter\_jpg\_all.txt

$ sort -u -k1,1.8 jpg.txt > filter\_jpg\_unique\_days.txt

This command compares first 8 characters (YearMonthDay substring0 of every line and list lines with unique YearMonthDay strings into a separate text file. Only the first unique occurrence is listed into the file.

$ find . -name "\*" | grep -vFf filter\_jpg\_unique\_days.txt | xargs rm -rf

This command deletes all files whose names are NOT in the file of the filenames with unique YearMonthDay strings.

The commands above are repeated for the FITS images for the same filter, then both the JPG and FITS sorting is done on the remaining filters.

After organizing by unique year, month and day, the following data remains:

3277 unique days of blue images (1 image per day)

3267 unique days of red images (1 image per day)

3362 unique days of calcium images (1 image per day)

The final data set is now comprised of images in the 18:00-19:00 UT range, consistent with the highest-quality images.