Henry Hitch

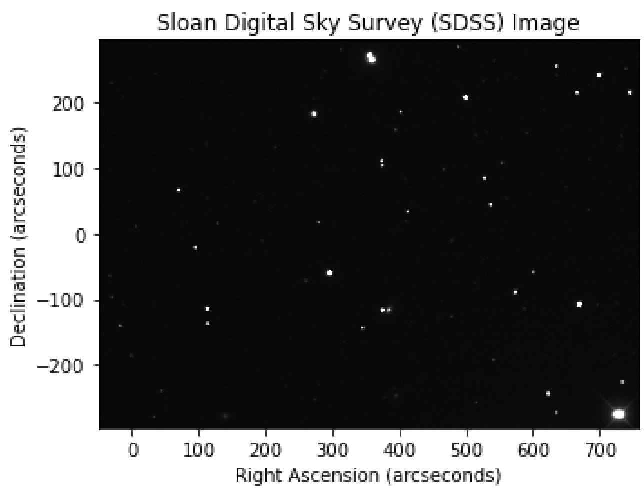
Mark Rast

ASTR 3800

Project 2

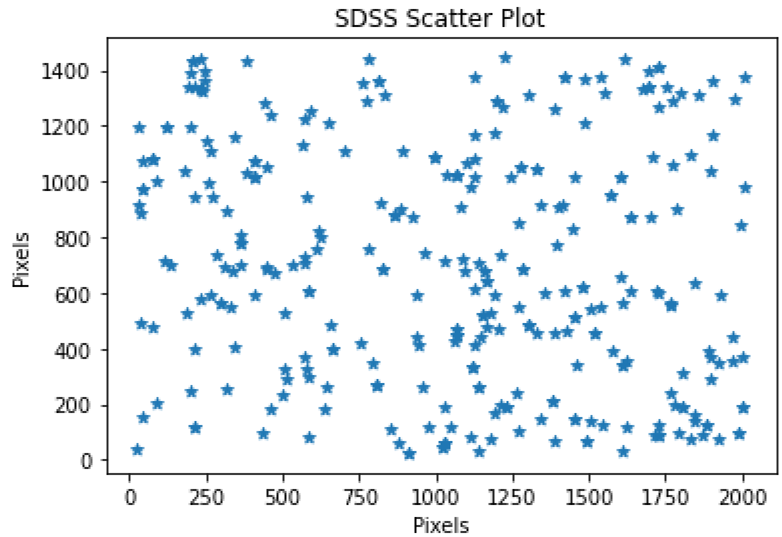
Write Up

After successfully reading in a fits file from the Sloan Digital Sky Survey and messing with the axes and color gradients, I created this grey-scale image:



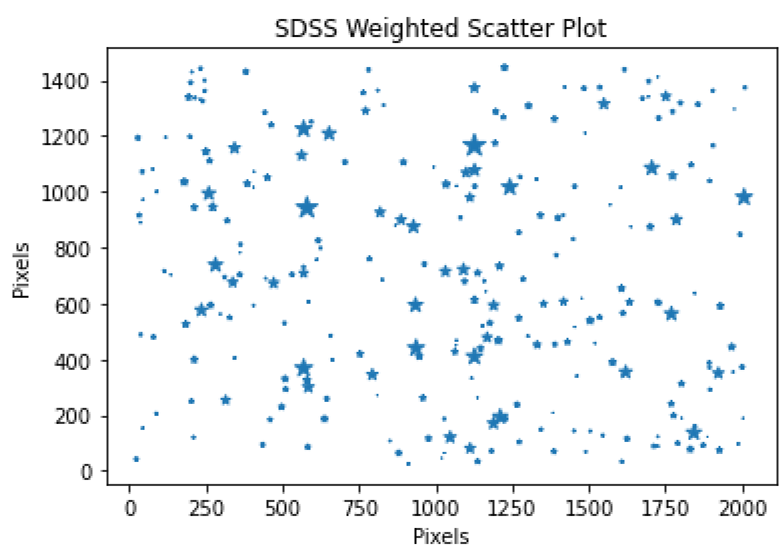
This grey-scale image looks very similar to the one in the Project 2 file, however it appears to have a little less detail, and is also upside down (due to how I plotted it with imshow).

Next, I extracted the same image data from a csv file, which allowed me to create a scatter plot of the X and Y positions:



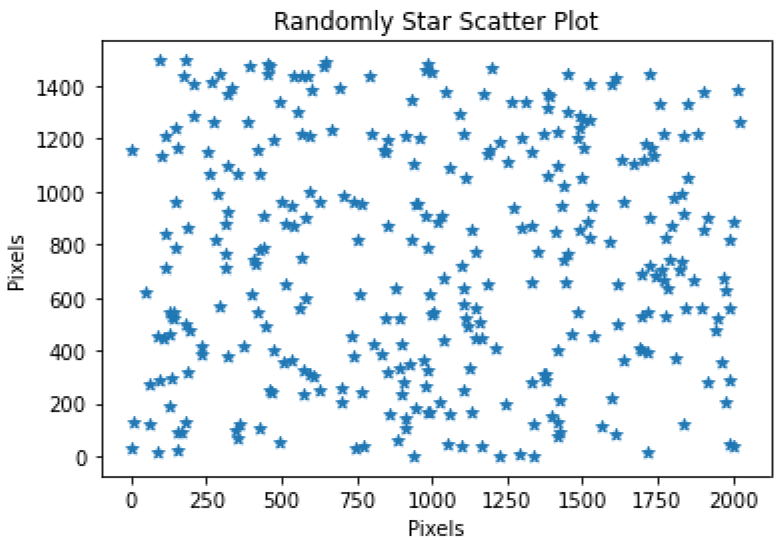
The largest difference that's apparent between my grey-scale image and this scatter plot is the number of visible stars. Many stars in the grey-scale image hardly show up and are so faint that you can't even see them. This scatter plot helps see all the stars in the image, but also makes it difficult to tell you're even looking at the same data.

Then, using light intensity data from the same csv file, I was able to create a weighted scatter plot for the star positions, which turned out way better:



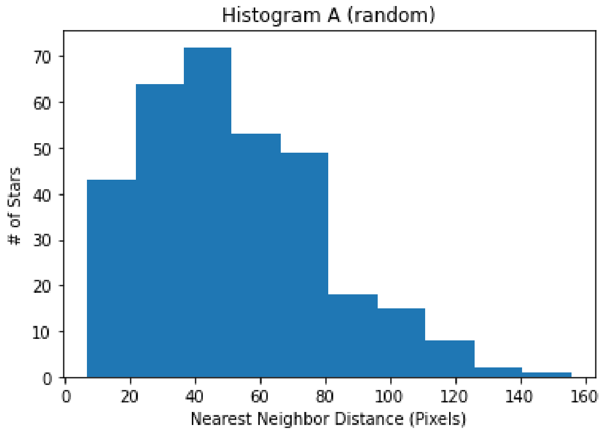
The differences between my scatter plot and my grey-scale image are mostly accounted for, but it's still kinda hard to tell they're the same image.

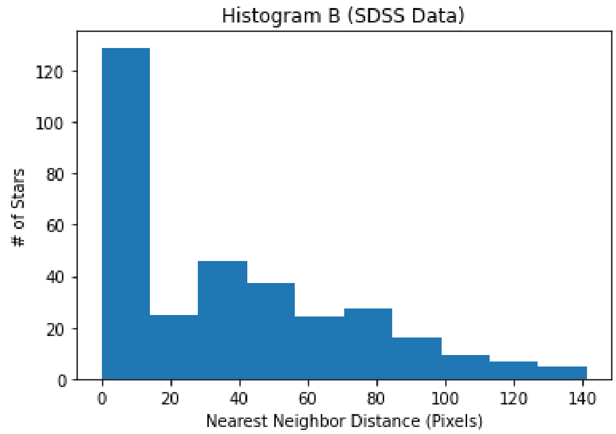
Next, I generated my own random star positions using the function numpy.random.uniform(), and scatter plotted them:



Other than the physical locations of the stars, this scatter plot is qualitatively AND quantitatively indistinguishable from my scatter plot in part A.ii.

Additionally, I wrote a function to calculate the nearest neighboring star (for each star) and plotted the distances for my randomly generated stars and the SDSS data in 2 histograms:





There is a large visual difference between these two histograms; the real star data from the image has many more stars that are closer together than my random star locations, which is evident by the huge count in bin 1 of Histogram B.

Finally, I analyzed these histograms through a skewness test and chi-squared test. I found that Histogram A has a skewness of 0.61 and Histogram B has a skewness of 0.77. In light of the visual differences between the two histograms, these values of skew make sense because Histogram B has many more values on the left side, which results in a higher value of skewness. Histogram A is still skewed quite a bit, but less than the real SDSS data.

<u>Thought Questions</u>

1. For aesthetics, you would want to use image processing that shows the true color of an image, as it's much more attractive than a scatter plot (for instance). For scientific analysis, it kind of depends on your goal, but ideally you want to use image processing that captures and shows all the relevant information. In this case, it was much more helpful to look at a scatter plot of the stars instead of the real image, because you could actually count every star in the scatter plot, which you couldn't do with the "aesthetic" image. If you used the images produced in A.i and A.iii in a scientific paper, it should be accompanied by a title, axes, dimensions, and information on how one could recreate the image.
2. The coordinate system when comparing the star fields shouldn't matter, because the distances to the stars in the image are extremely large. In theory, uniformly distributed star fields should be uniform in RA and Dec, as well as cartesian.
3. When computing a Chi-squared value, you're effectively comparing the inconsistencies between 2 data sets, and calculating the probability that the data sets are correlated based on the number of data points you have and the number of variables.
4. I'm not entirely confident of my conclusion about the non-randomness of the star field in Part B, because I know how many stars are in a galaxy, which I believe should be approximately a uniform distribution when looking at a small section of the sky.

<u>Comment</u>

This project was cool! My favorite part was plotting the weighted star locations and seeing the result.