Patching Statistics

Veer Bathwal

Table of Contents

Patch Number Distribution

Quartiles

54	97	155
First Quartile	Median	Third Quartile

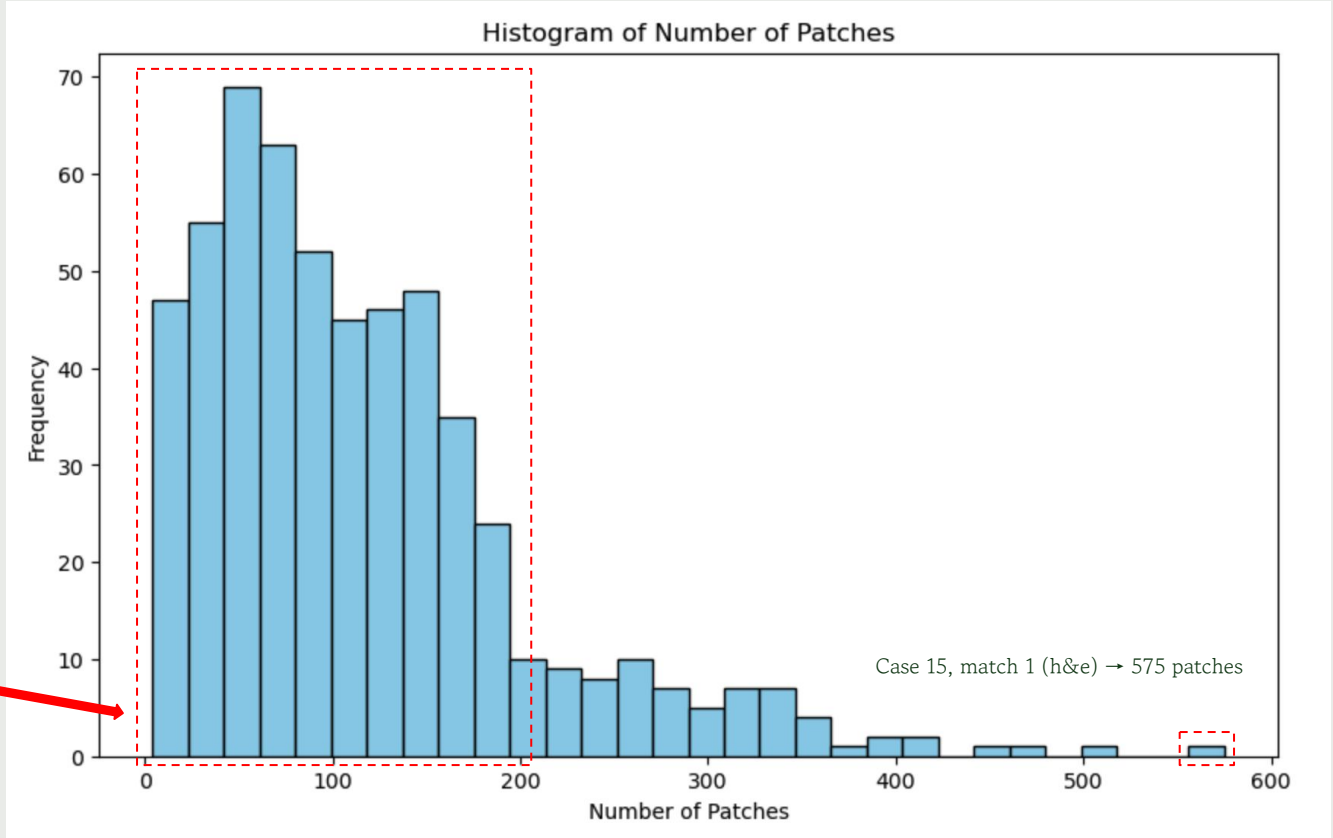
Other Statistics

4	65,372	575
Minimum Case 63, match 6 (melan)	Total Patches	Maximum Case 15, match 1 (h&e)

Histogram

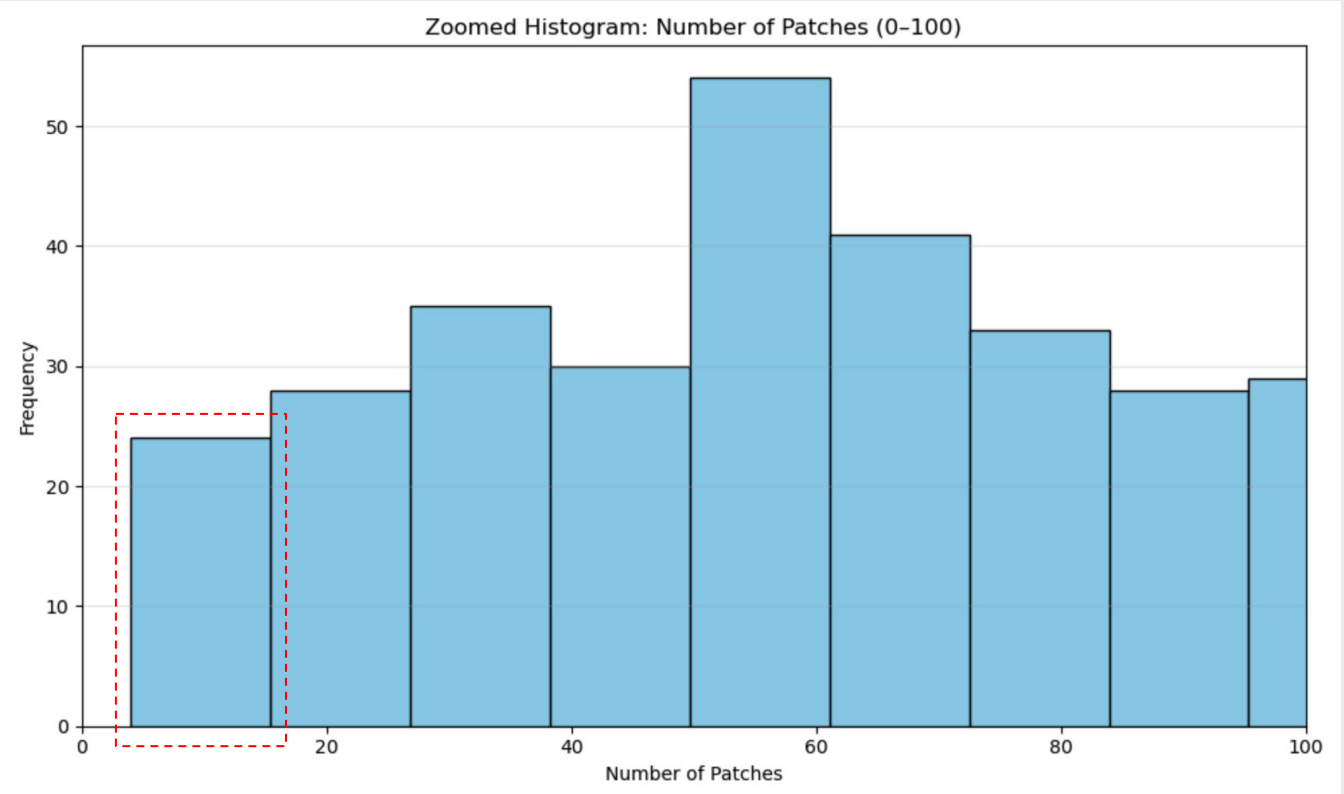
Most slices have fewer than 200 patches. We see a few outliers on the extremes, with a minimum of 4 and maximum of 575

Case 63, match 6 (melan) → 4 patches



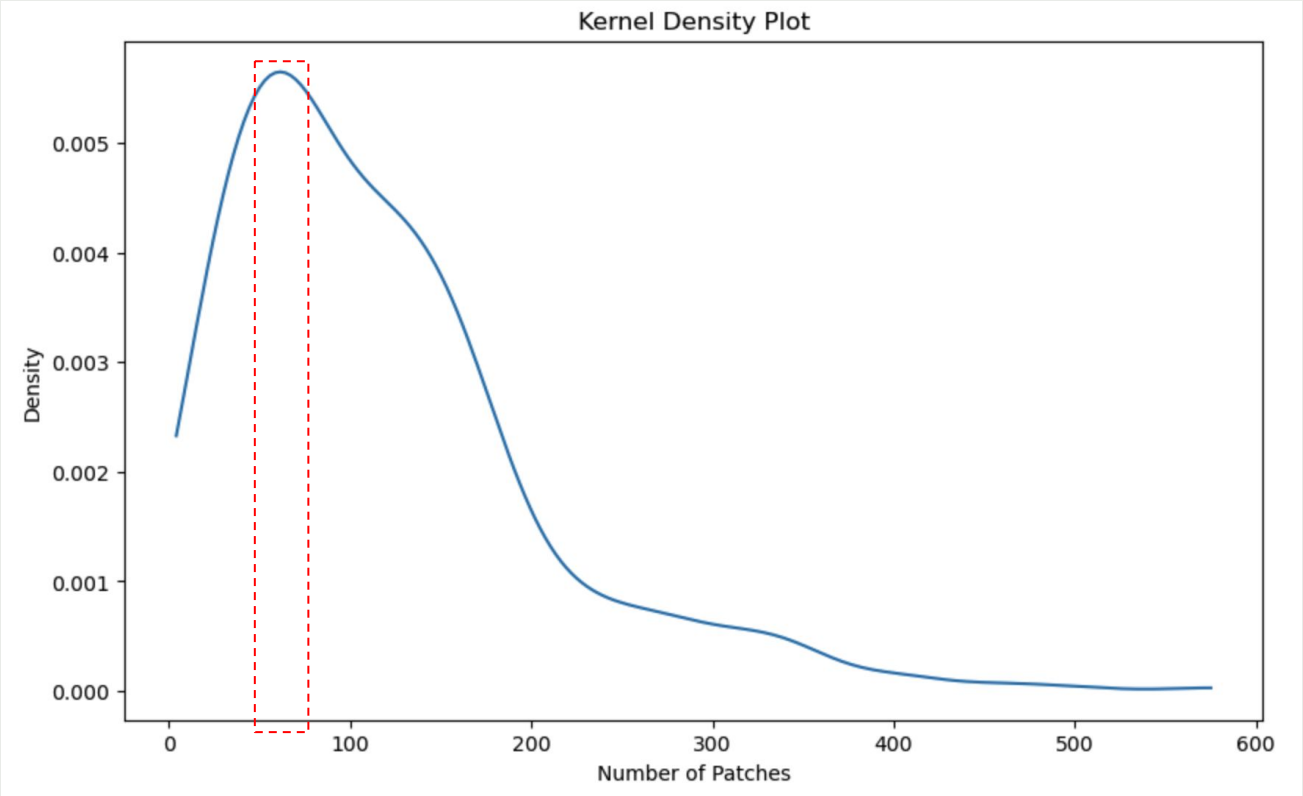
Zooming In

Looking at the lower end of the distribution, there are ~25 slices in the first bin, with fewer than ~15 patches per slice



Density Plot

We observe the highest density somewhere near 60-70 patches per slice



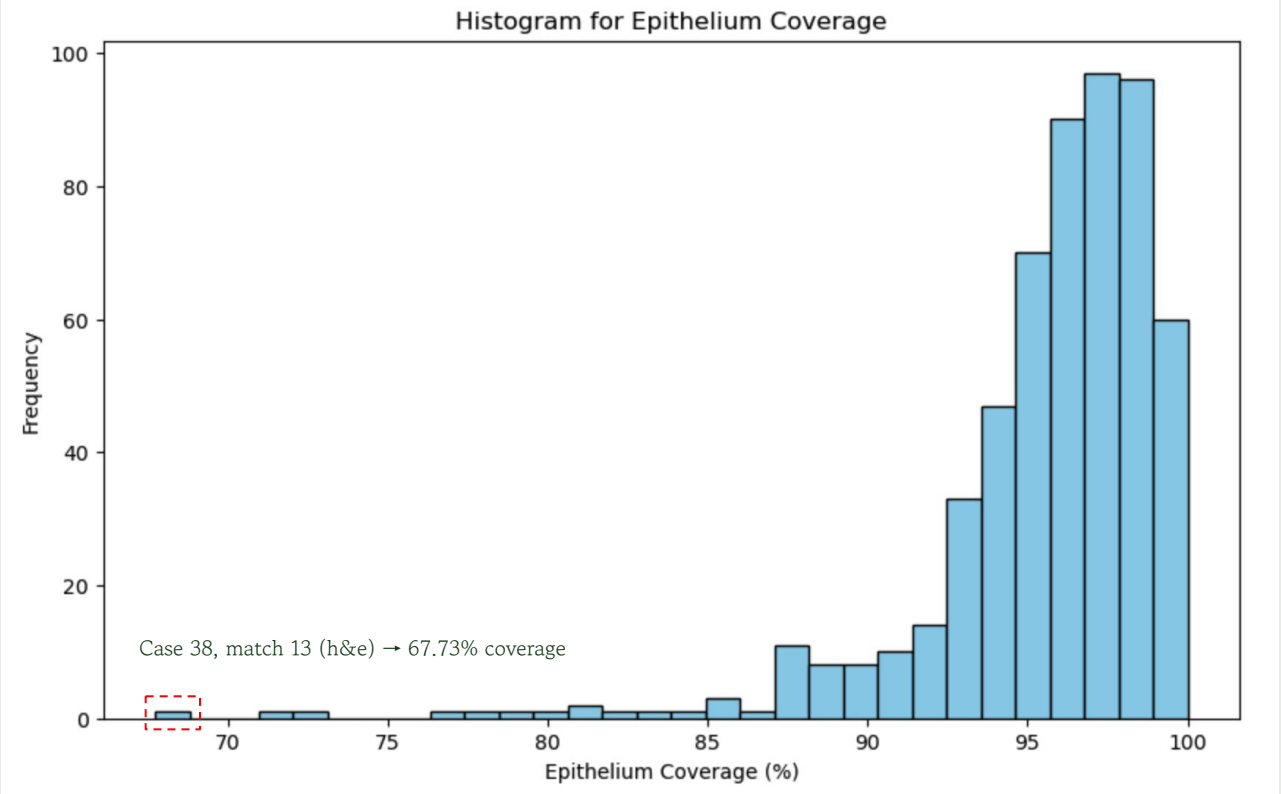
Epithelium Coverage

Quartiles

94.48	96.49	97.98
First Quartile	Median	Third Quartile

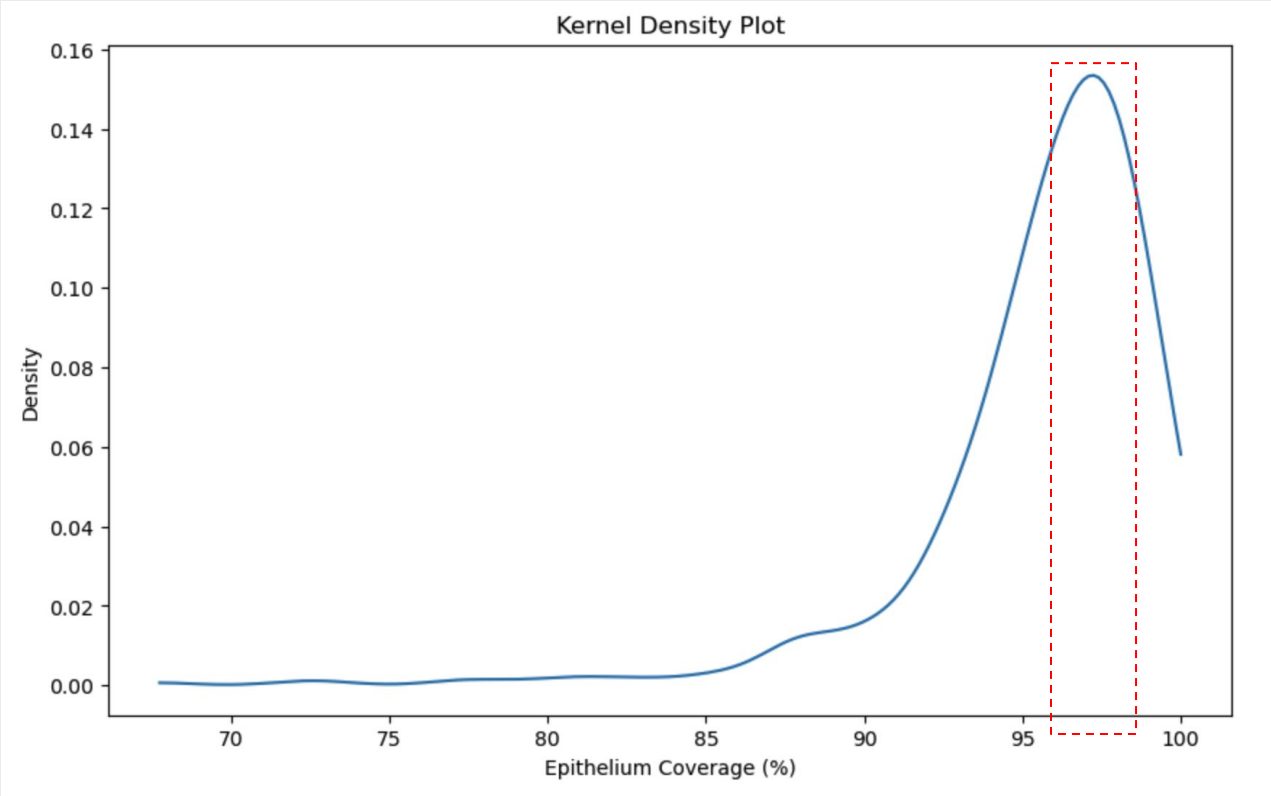
Histogram

Left-skewed distribution with the majority of observations greater than 95%. With the first and third quartiles being very close to the median, we have covered most of the epithelium across all slices, barring a few outliers



Density Plot

Density peaks within the values defining the interquartile range



Patch Length Distribution

Quartiles

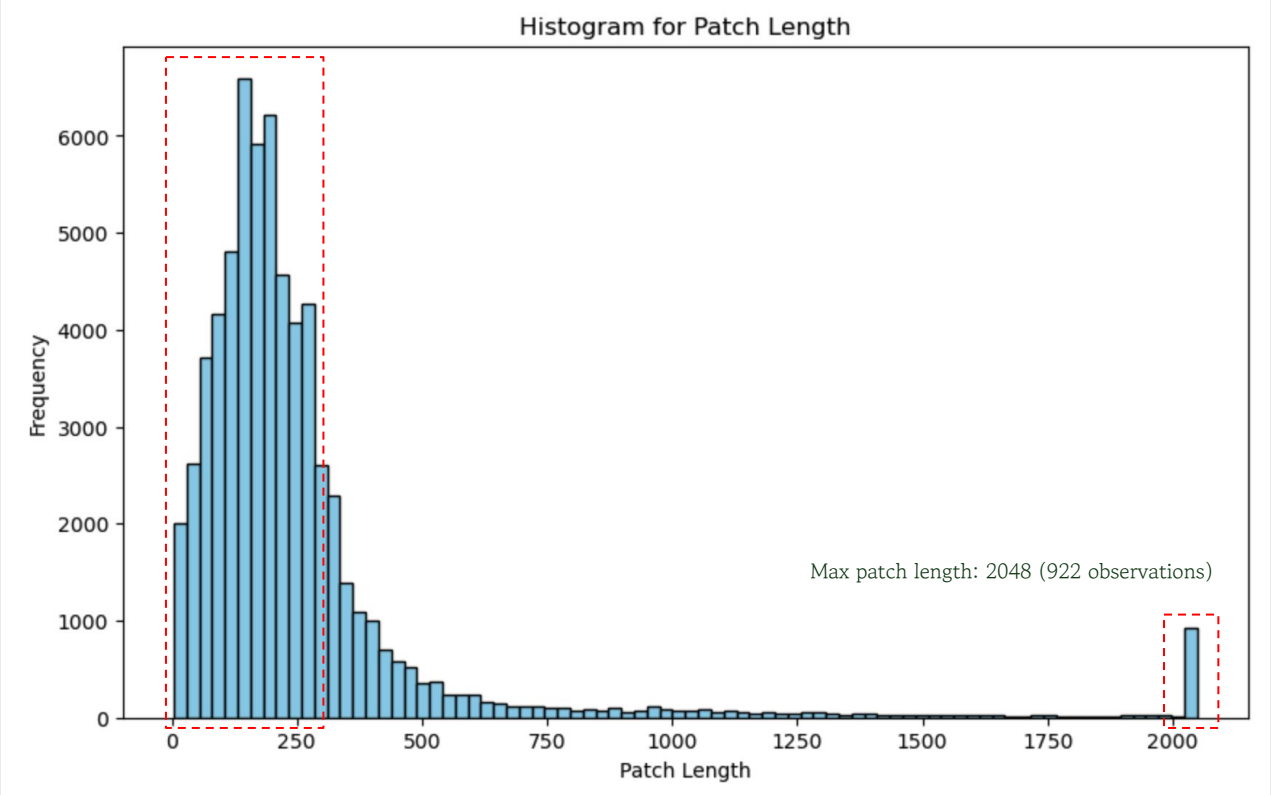
124	192	280
First Quartile	Median	Third Quartile

Other Statistics

Minimum	Mode	Maximum
4	172	2048

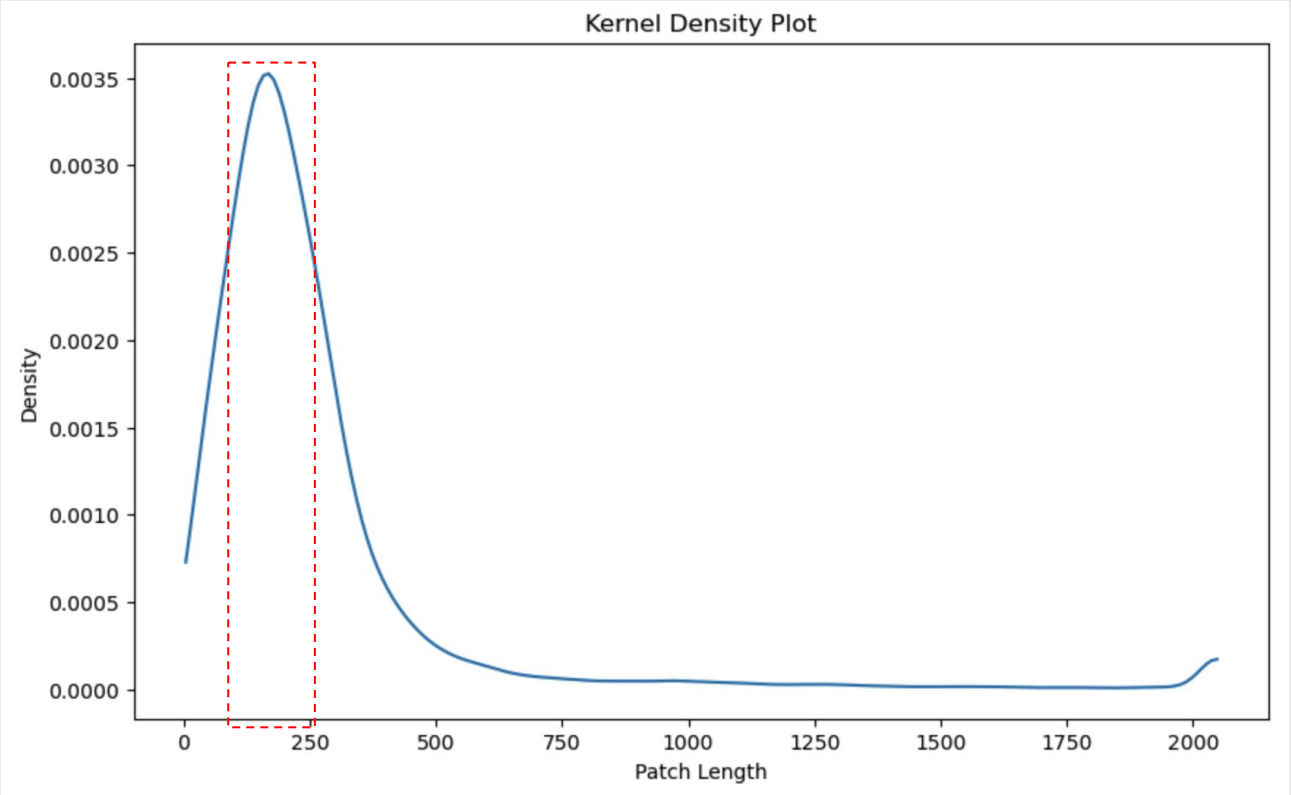
Histogram

Right-skewed distribution, except for a peak at max length of 2048 (as described in patching algorithm). Majority of the observations fall below the 250 mark (close to the third quartile value of 280)



Density Plot

Similar to epithelium coverage, we see observations tightly clustered the median rather than a loose distribution with no apparent structure



Takeaways

- Patches, in general, are well-formed and cover most of the epithelium
- There are a sufficient number of patches for pretty much all slices tested
- Characteristics of the patches themselves, in terms of length, seem to be consistent across all cases in the dataset
- However, there are a few outliers as shown through the density and histogram plots. The most significant ones being the slices that either have an extremely low number of patches (<10) or unsatisfactory epithelium coverage (<80%)
- Consider creating ad hoc thresholds on patch number and coverage ratio to filter slices by, in order to ensure our models are fed good quality images
- Consider if increasing max length would generate better results, since a lot of our current patches are bound by that upper limit
    - Although, important to consider trade-off between getting better patches for those 922 observations and by spending additional time on patching versus focusing on model creation
    - Since our current algorithm offers high epithelium coverage, it might not be worth-it to continue tweaking the process

Thank You