Ken Lew
Data Sci HW 7

1)

	Observations	Original Data	1 Million Data
0	Max X	1.185000e+06	1.184999e+06
1	Min X	1.182500e+06	1.182500e+06
2	Mean X	1.183805e+06	1.183804e+06
3	STD X	7.031091e+02	7.032438e+02
4	Max Y	1.867500e+06	1.867499e+06
5	Min Y	1.865000e+06	1.865000e+06
6	Mean Y	1.866243e+06	1.866243e+06
7	STD Y	6.977191e+02	6.973793e+02
8	Max Z	8.530000e+02	8.520000e+02
9	Min Z	5.680000e+02	5.680000e+02
10	Mean Z	6.230946e+02	6.231054e+02
11	STD Z	3.455555e+01	3.460462e+01
12	Max I	6.114300e+04	5.891800e+04
13	Min I	0.000000e+00	0.000000e+00
14	Mean I	1.492827e+04	1.493144e+04
15	STD I	1.379186e+04	1.379522e+04

2)

```
correlation between x and y:  0.012982643725631657
correlation between x and z:  -0.17225872545390375
correlation between y and z:  0.21963377710095003
correlation between x and intensity:  -0.04667556009151253
correlation between y and intensity:  0.07614783323928326
correlation between z and intensity:  -0.0833795629560239
```

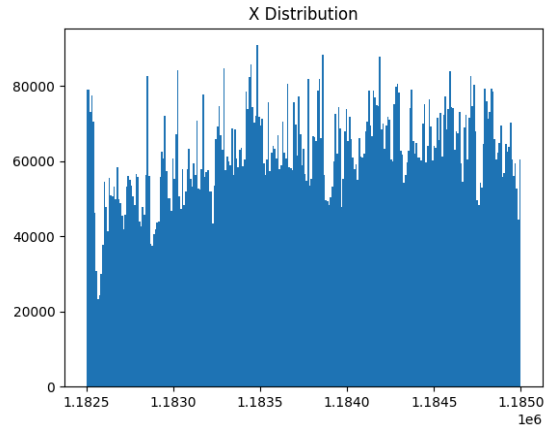
Np.corrcoef returns Pearson product-moment correlation coefficients within a range of [-1 , 1].
Based on source: [Pearson Correlation Coefficient: Calculation + Examples (questionpro.com)]

	Coefficient, r	
Strength of Association	Positive	Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.3 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to 1.0

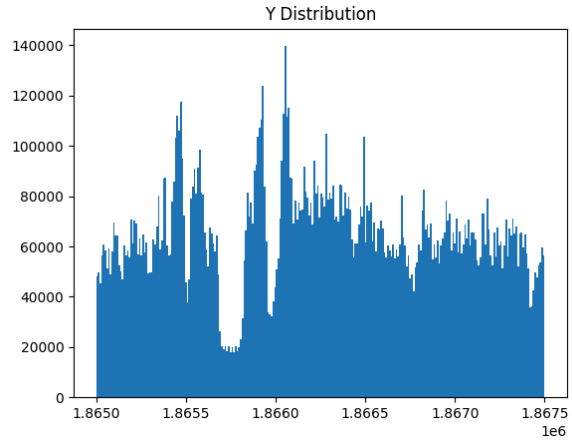
We find that the correlation between:
XY: Small Positive
XZ: Small Negative

YZ: Small Positive - highest among xyz pairings
XIntensity: Null
YIntensity: Null
ZIntensity: Null

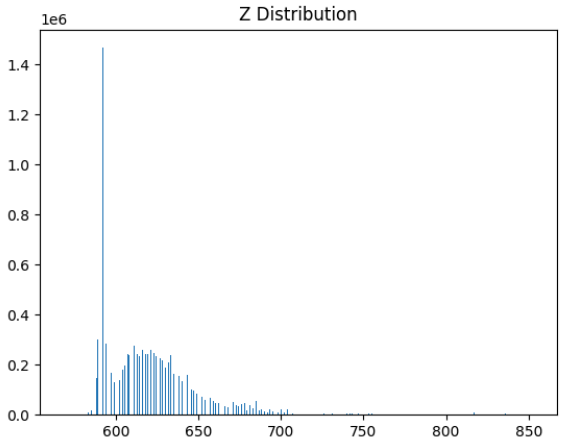
3)



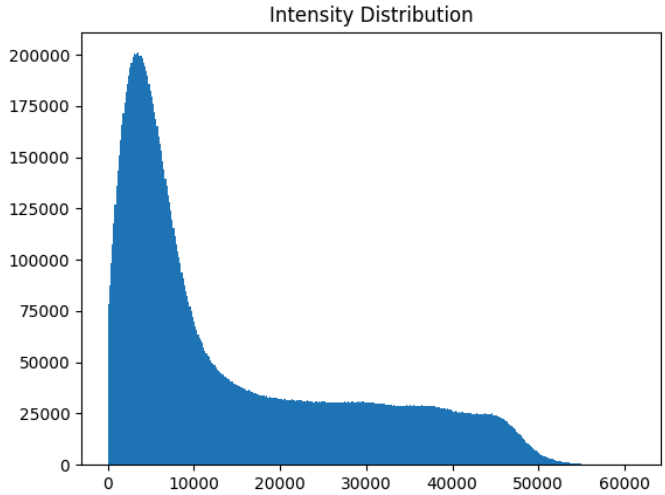
Regular spikes in occurrences at certain intervals. A slight decrease in number of data near the beginning at approximately 1.1826 e6.



A lot more volatile compared to X data. Large gaps in data followed by a spike in data occurring at 1.8657 e6 and 1.866 e6.



Very large spike in data at 592 in an otherwise bell curve-shaped distribution with a right tail.



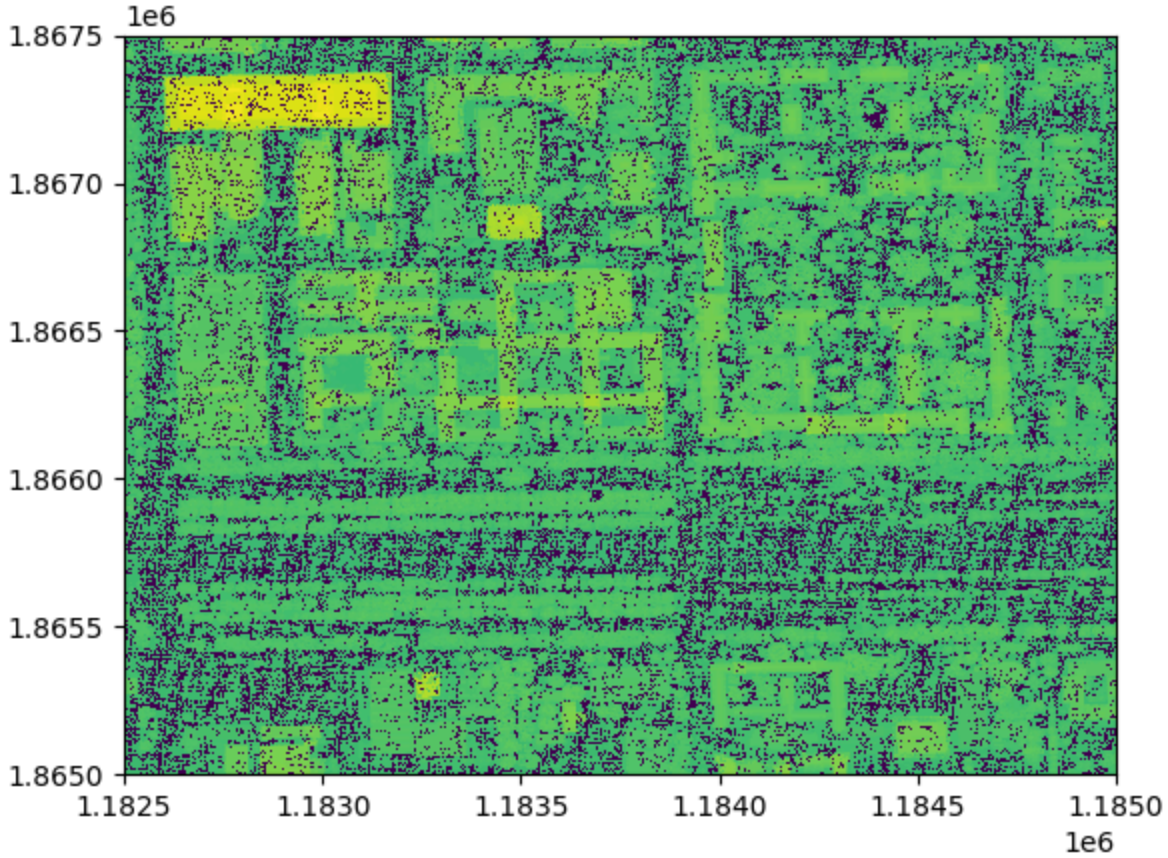
Intensity grouped around 3.2 e3, similar to z with a bell-curved distribution and a lagging right tail.

4)

Aggregated by taking the sum of each z/intensity value and dividing by the count (Mean).

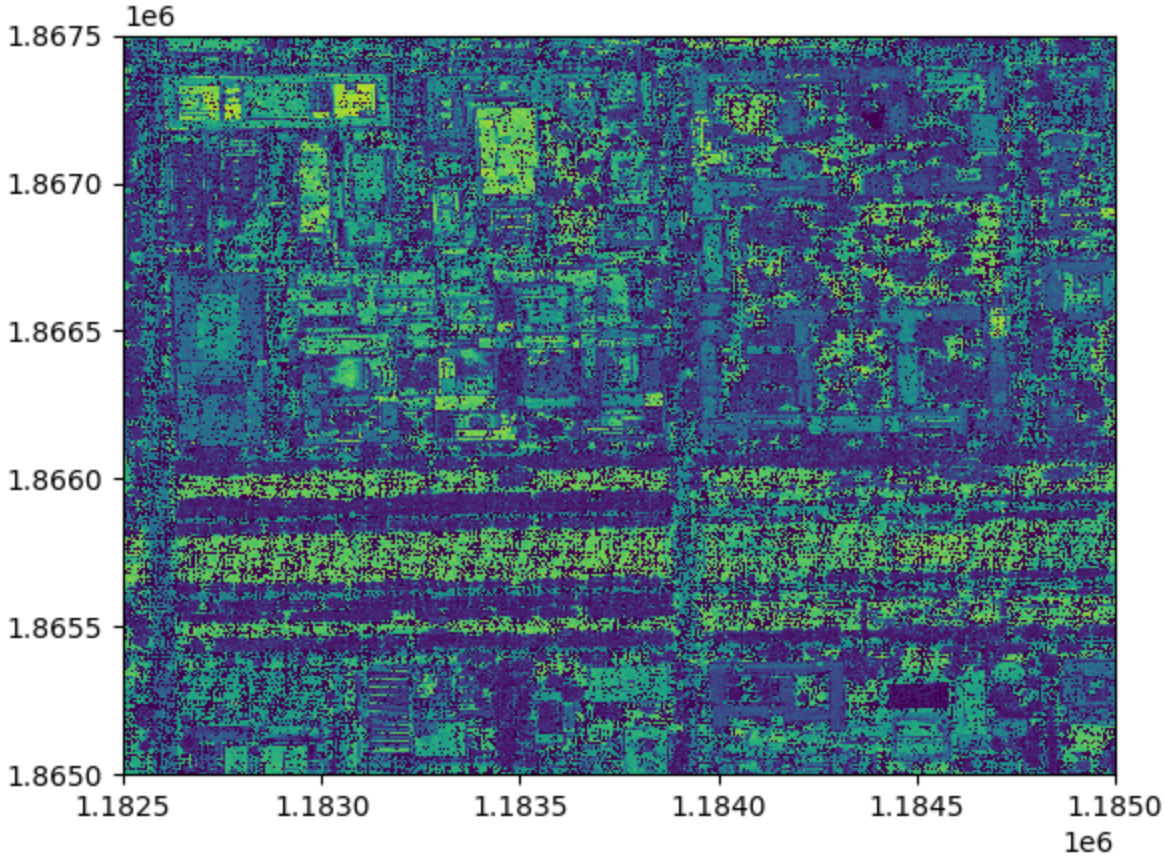
5)

Z Value Mean



Looks like a map of campus. Assuming Z values are the vertical values, the greener areas represent the elevation where buildings or trees are. The redder areas are the lower areas, where the ground may be. There seems to be a scattering of missing values across the map.

Intensity Mean

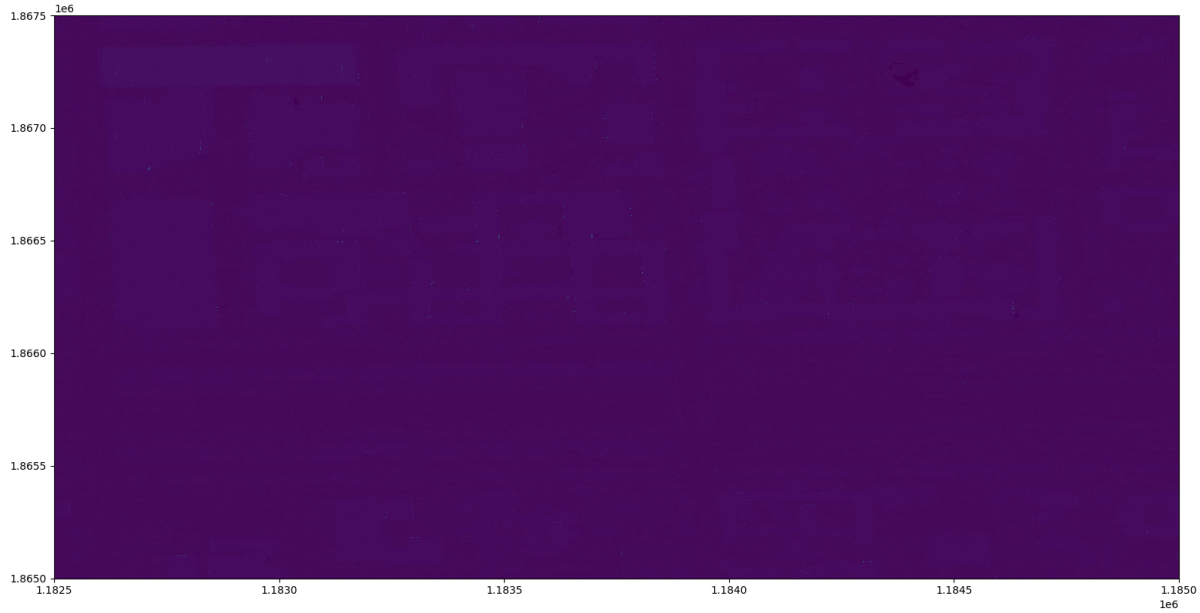


Similarly to the Z values, the intensity displays a map of the campus. This time, elevated surfaces are blueish in color where elevation closer to the ground is greener in color. Similarly, the missing values seemed to be scattered throughout.

6/7)

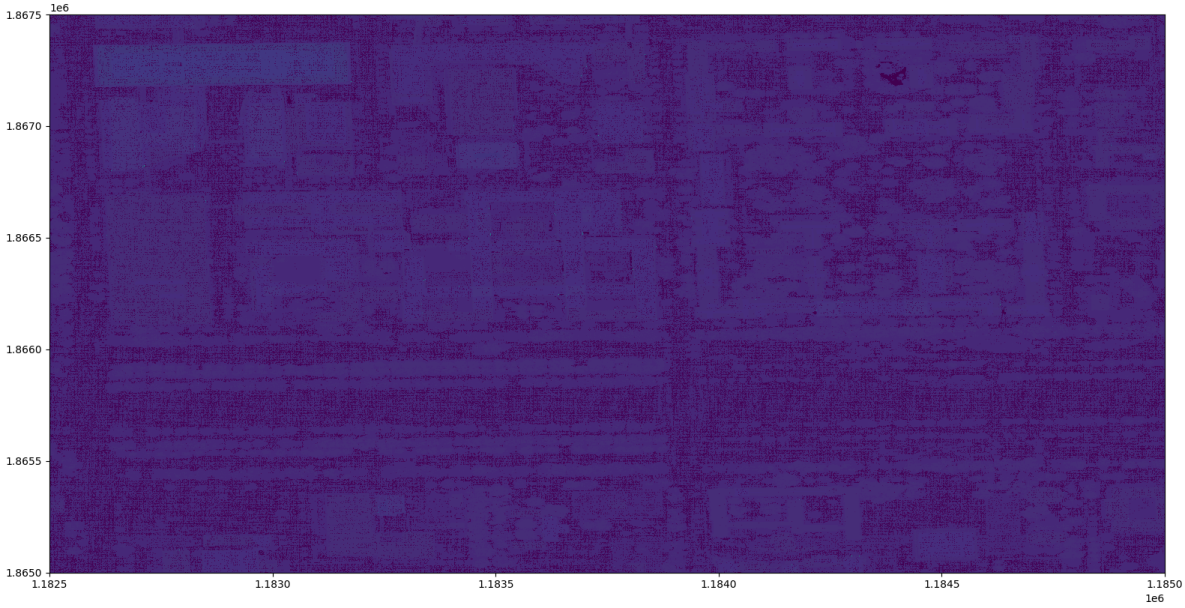
The method is to take left, right,top, and bottom pixel values around the missing pixel and average them. If a directional pixel is missing, it takes the average of the non-missing pixels.

Z Gap Filled 1x Multiplier



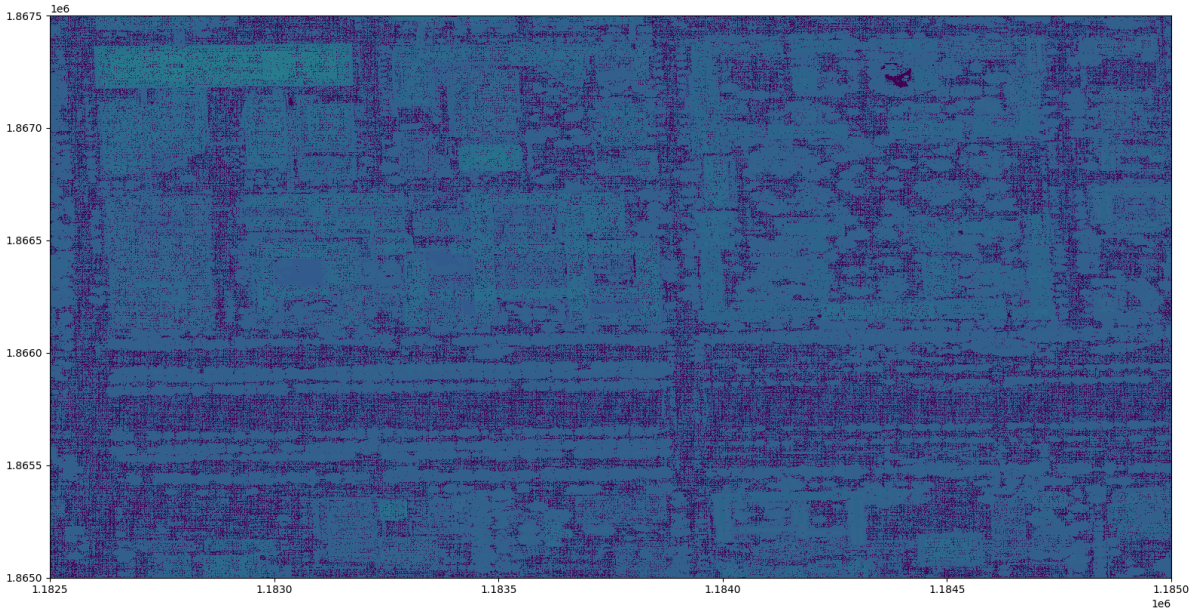
It didn't work so well initially for the Z map with a 1x average multiplier.

Z Gap Filled 0.25x Linear Multiplier



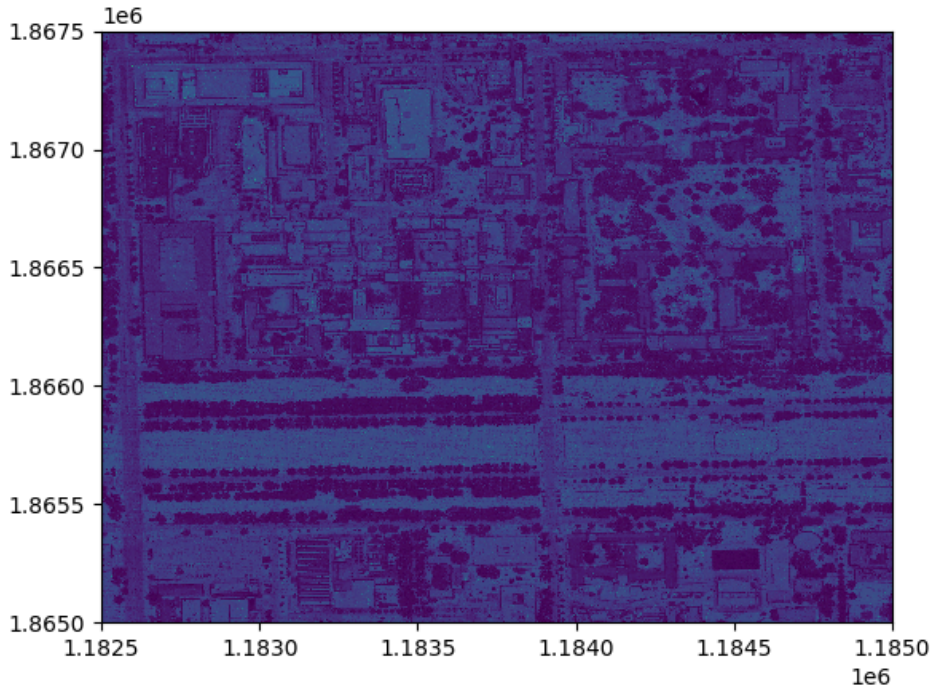
If a 0.25 linear multiplier is applied to the missing pixels, it gains a bit more resolution.

Z Gap Filled 0.1x Linear Multiplier



Similarly, if a 0.1 linear multiplier is applied to the missing pixels, it gains a bit more resolution but it isn't considerably better than the 0.25 multiplier.

Intensity Gap Filled 1x Multiplier

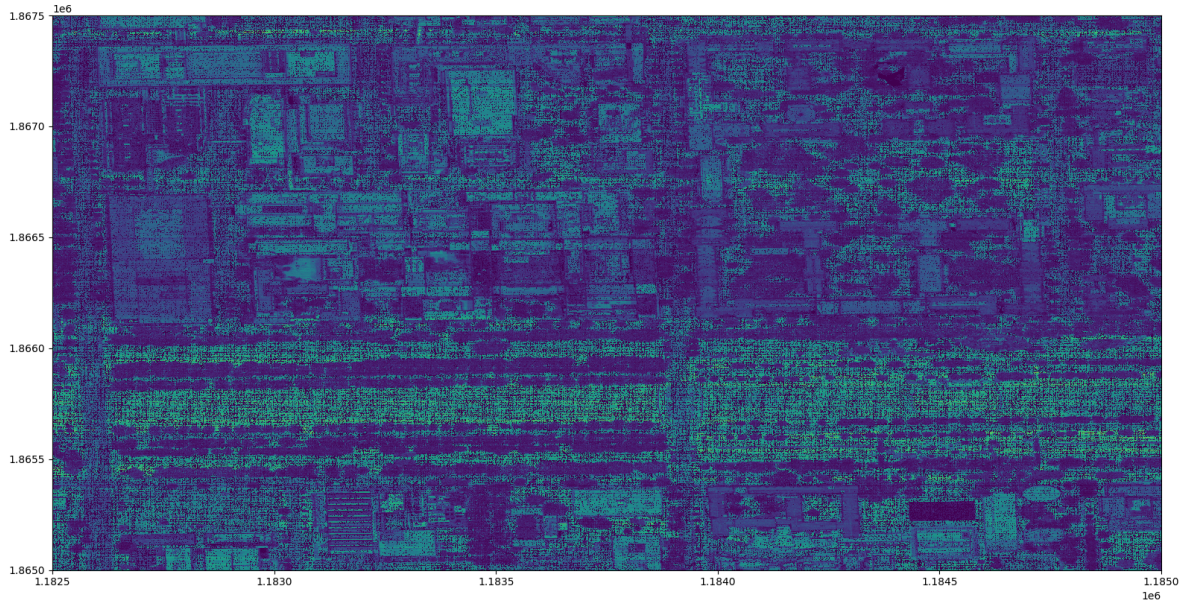


Filling the gaps for the intensity map produces the most detailed map yet. No linear multipliers were needed.

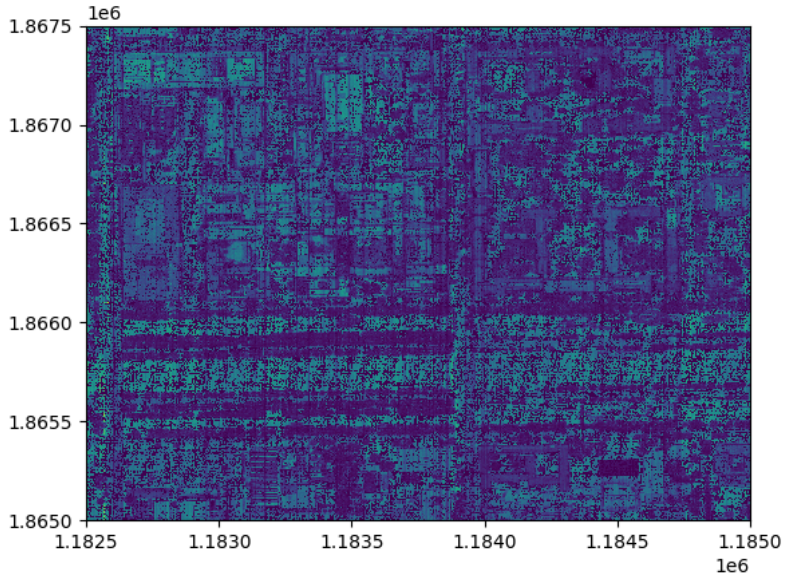
8)

Transforming the col and rows to have the same variance improved the resolution of the map. Note: Gap filling wasn't performed on the transformed values - too computationally intensive for the computer.

Variance Intensity (COL)



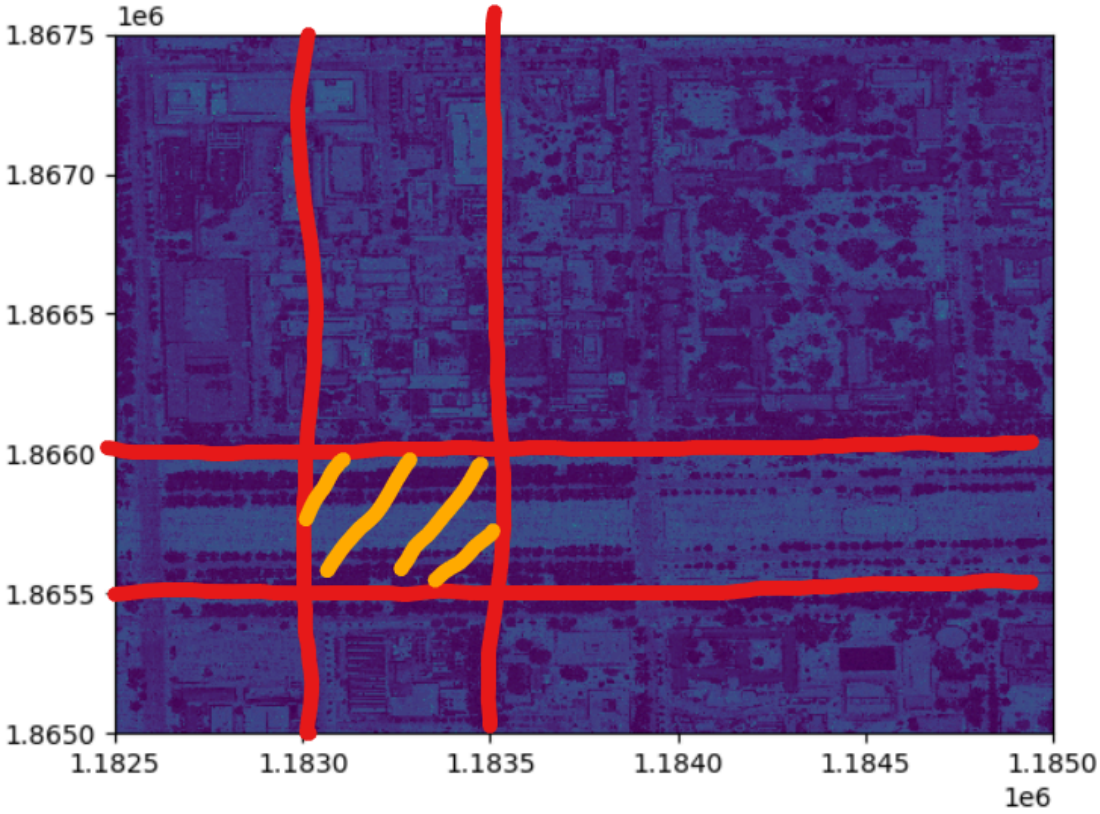
Variance Intensity (Row)

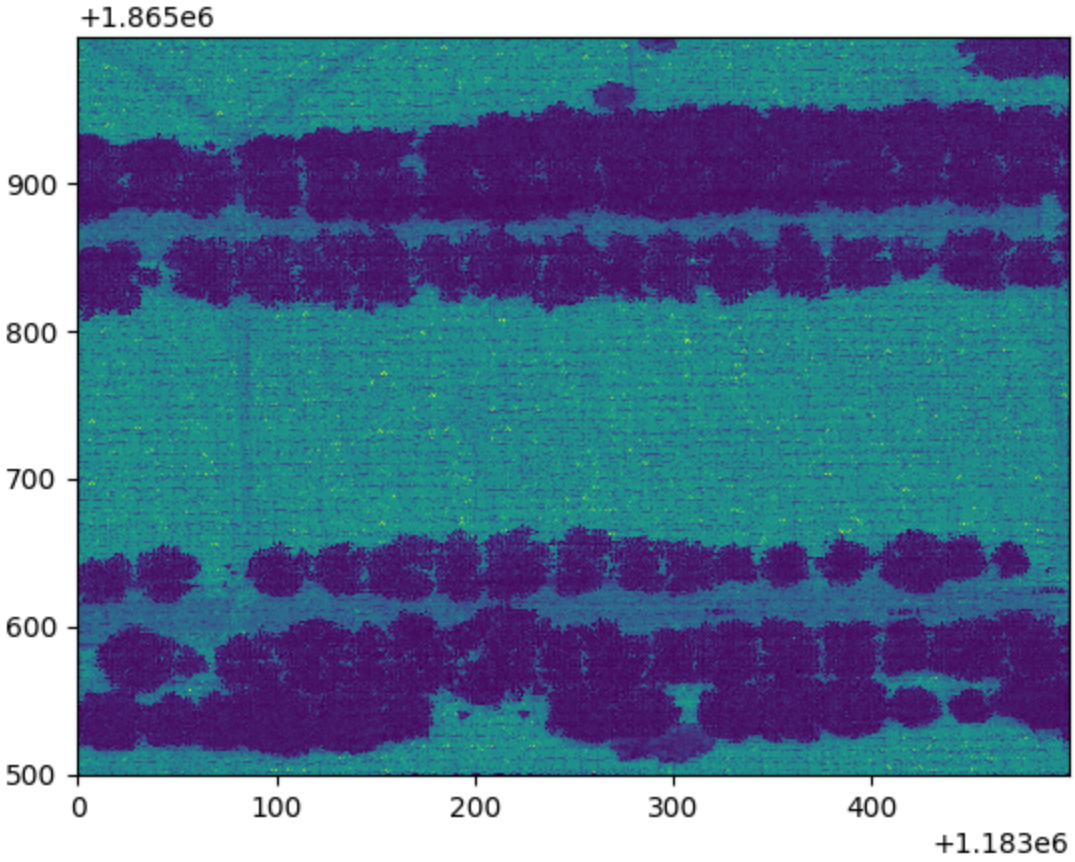


9)

SUBSECTION (x 1183000 - 1183500) (y 1865500 - 1866000)

The following subsection dictated by the bottom left coordinate and top right coordinate were chosen and zoomed into. Image 1 is the approximate area on the larger graph and Image 2 is the zoomed-in area.





10)

Name
Cloudimage - converts csv data files to images

Synopsys
Cloud image [-n input] [-c other1.2] [-a type] [-f jp] [-r dir] [-p] [-z lx ly rx rx]

Description
Cloudimage - converts csv data files with 4 columns of numerical [x,y,other1,other2] data to a x,y coordinate map, mapped to either other1 or other2 with various filtering and refinement features.

Options
-n –input
Input file name of **CSV** file containing data
-c –other1/2

Sets whether coordinates map to either other1 or other2. Where other1.2 is one of **1,2** with 1 being the default

-a –type

Sets analysis type for repeated x,y coordinate other 1.2 values. Where type is one of: **Mean, Max, or min,** with mean being the default

-f –jp

Downloads output image as either jpg or png. Where jp is one of **JPG, PNG** with PNG being the default

-r –dir

Refines image quality by averaging variance along direction dir. Where dir is one of **row,col, both** with None being the default

-p

Patches missing data points in other1.2 by averaging neighboring data points

-z –lx ly rx ry

Slices subimage governed by bottom left and top right coordinate points. Where lx is the left x coordinate, ly is the left y coordinate, rx is the right x coordinate, and ry is the right y coordinate