Microplant Paper: Begun in Spring 2020, will be completed spring 2021!

Tom Potential Abstract:

Museums contain some of the world's greatest treasures. Priceless specimens collected over many decades wait to be analyzed, studied, and have their secrets revealed. Biological collections present unique opportunities to better understand and preserve biodiversity thus any acceleration of their accessibility could have a great impact on species conservation and preservation. To this end, many efforts have aimed to increase digitization of museum specimens and increase database access. In an earlier manuscript, we reported on a community science effort at the Field Museum to not only digitize specimens, but also to aid in the study and classification of a cryptic species of liverworts via an online platform. Here we outline a novel approach of a natural history museum using an unaided interactive kiosk with touch screen technology to engage patrons in the collection of biological data. We demonstrate that interactive kiosks are a potential route to engage museum patrons and collect biologically relevant data on cryptic plants that can be used to classify and identify similar species.

Parts 1 & 2 on Museum & biology background here.

1. Ref to earlier paper, big picture stuff, specimens, etc.
2. Intro to interactive exhibits, most exist for fun, few exist for science/data generation.

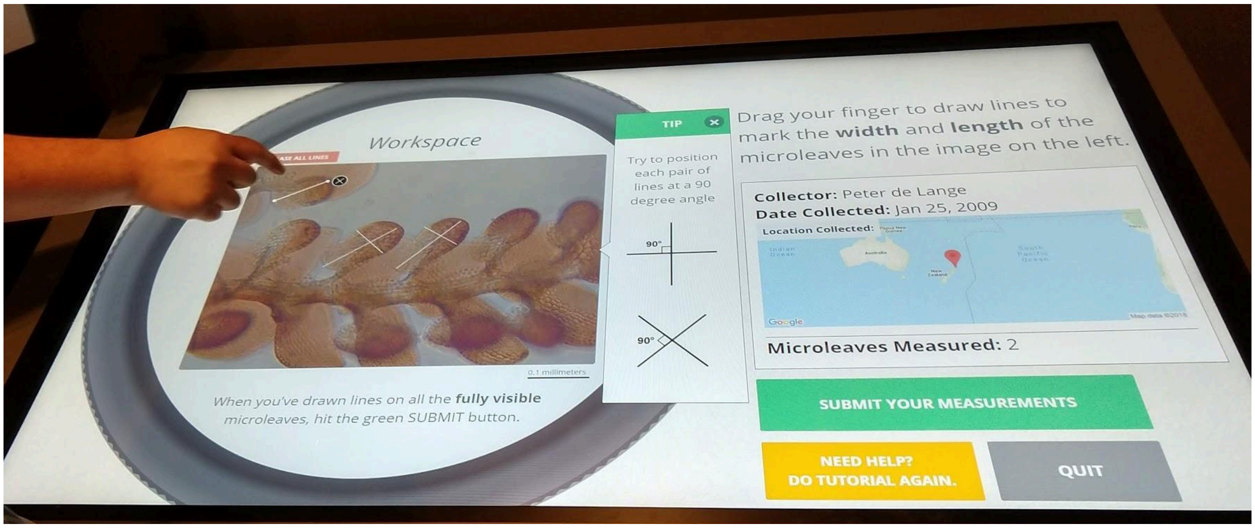
The Field Museum of Chicago houses over 40 million items in their collection, with less than 1% on display in their 19 acres of exhibit space. Microplants are among the tiniest specimens, with small lobules that are on a scale of microns.

Microplant Kiosk at the Field Museum

The Field Museum has a Science Hub, where patrons can participate in community science activities. The exhibit overall saw XXX people. In this area, for many years there was a kiosk displaying images of microplants. During the forty-eight day period of June 21-August 7, 2018, there were over a thousand individuals or groups who used the kiosk, with an average of approximately 22 uses per day.

The kiosk has a demonstration of how to properly measure microplant lobules, and then it prompts the user to measure the microplant leaves displayed on the screen. Users can erase and correct their measurements until they decide to click submit. There are multiple images that can be measured, and the images can have more than one leaf on them. The kiosk displays these images randomly.

The kiosk collected various leaf measurements which includes x,y coordinates, angles created by the measurements, the lengths of the axes (since the leaf was measured lengthwise and widthwise), and the time of the measurement. The kiosk tracked the measurements per session for each specific image. An intern stood at the kiosk at certain points during the day collecting both the ages of the people participating (10 and under, 11-17, 18+, and people who chose to not answer were labeled "skip") and the time. Ultimately, the intern's recordings helped us match the kiosk data and the age demographics into a final copy of the data used for our statistics. (This was actually a file embedded in the kiosk itself, not an intern. We should fix this writing!!!)



It is important to note that the data taken by the kiosk was in units of pixels. Each image was taken with a consistent scale, and 1.0500 pixels correspond to 1 micron. (Similarly, 0.9532 microns correspond to one pixel.) For simplicity, the lengths in this paper are given in pixels.

Processing the data

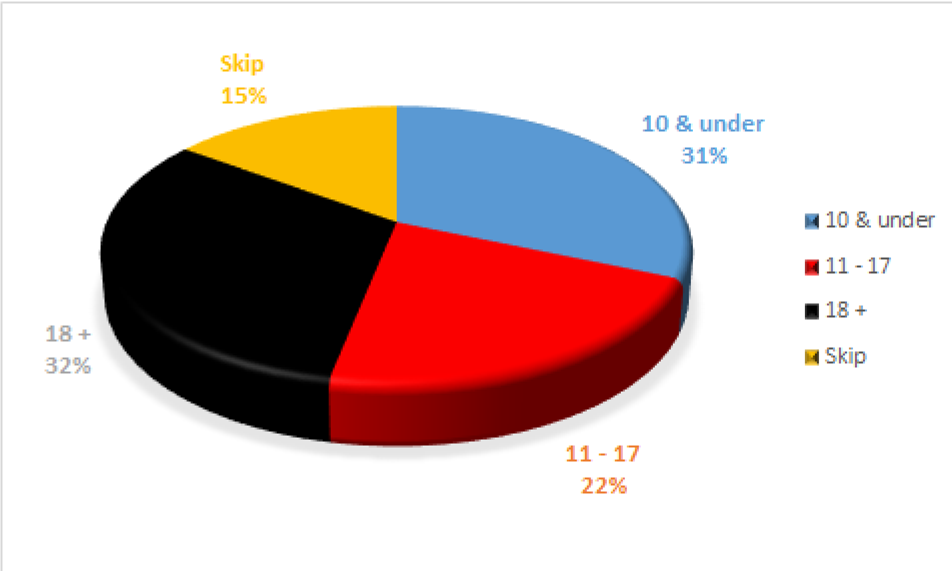
When people are at the kiosk measuring an image of the microplant leaves, the kiosk records the data and transfers the image information to a csv file with each image from a session on its own respective line. Thus, if an image was measured on ten different occasions, the kiosk would record this on ten different lines. We used the language PowerShell to do the initial data extraction and processing. All line segments that were drawn on the screen were stored in a stack of data. With the use of code, we automated the process of finding line segments that crossed. If a segment has no other segment that intersects with it, that segment is discarded from the final data. The images that had no data at all, we put it into a “bad data” file. For example, someone may have come up to the kiosk, touched the screen to submit, and walked away. After cleaning the data that way, we had a set of “valid data”. We also checked for what we considered to be good angles. A perfect angle is 90 degrees. Given the nature of the kiosk and it being very hard to get a perfect angle on a touch screen, if the angle between the two segments is less than 80° it would be added to the bad data as well. Key pieces of metadata that related to the background image and the dimensions of the screen that it was displayed on were kept. The metadata allows us to guarantee the scale. Finally, the data was split from the list of intersecting segments by putting each intersecting pair into its own row in the CSV. This final form of the data was used in Excel to create the plots and statistical analyses. For the code you may visit the link below.
https://github.com/AlexLabontu/Microplant-code

The next task was to match the age demographic file with the leaf measurement data. This was done using Excel. We looked at how long it took each person at the kiosk in the demographic file. We only considered the kiosk data to match with the demographic data if they were within 3 minutes of one another on the same date. Anything that was above our time frame made it harder to determine which image the person worked on. We also focused on the different age groups: 10 and under, 11-17, 18+, or “skip”. A person listed in the interns’ demographics could have measured more than one image. In that case, each image was counted as a separate line of data that was done by that one person. At the end, this affected our percentages as to which age group is the best at making measurements.

Total numbers by different demographic groups.

Demographics	**Number**
10 & under	**319**
11-17	**220**
18+	**324**
Skip	**151**
Total	**1014**

Percentage of images measured by each age group:



Types of data

All data: Data collected any time a person pushed the submit button on the kiosk, regardless of quality or even existence of measurements.

Valid data: A set of pairs of line segments that intersected for an image. If a person put five line segments into an image with two pairs intersecting, that would lead to two lines of valid data.

Good data: A set of pairs of line segments that intersected for an image and which formed an angle of at least 80 degrees.

IQR cut: A set of good data for a particular image where the IQR (interquartile range) is calculated for that image specifically and a cut is made based on it.

Expert bounds 1: A subset of good data from for a particular image that an expert has also measured multiple times (typically 6) which has both small and large axis lengths within the expert's minimum and maximum measurements for those respective axes.

Expert bounds 2: A subset of good data for a particular image that an expert has also measured multiple times (typically 6) which has both small and large axis lengths within 10 pixels of the expert's average measurements for those respective axes. (Note that in this particular file, this was equivalent to 3 standard deviations of the expert's measurements.)

Separating the Good Data from the Bad Data

As the data was processed, we had many ways of increasing the accuracy of the leaf measurements. We automated the restriction from the full data set to a good data set, where leaves have a major axis and minor axis measurement with an intersection that's above 80 degrees. After some experimentation, we found that the IQR (Interquartile Range) cut was able to remove most of the inaccurate data. We performed this cut for the set of all data that corresponded to an individual image, using the IQR that the data gives for that particular image. To calculate the IQR cut, we first split the dataset into quartiles. IQR stands for the difference between Q3 and Q1 where Q1 is the cutoff of the first 25% of the data and Q3 is the beginning of the last 25% of the data. We keep all data that is in the range from Q1-1.5IQR to Q3+1.5IQR. In this manner, we are able to remove the majority of outliers. Note that this process does not involve human judgement, and so it can be automated. It is also simpler to implement than a cluster analysis. We have graphed the data overlaid on the images from the kiosk in order to illustrate the effectiveness of this approach.

Regarding age group, no trends were discovered in the accuracy or consistency of the measurements. Our assumed thought was that children 10 and under would have more inconsistent measurements than the other age groups simply due to their young age, but we found that on average, children's measurements varied very little compared to other demographics. We can see from the percentages of "good data" below that in most cases the lowest percentages went to the group that decided to skip the demographic entry. One can presumably attribute this to questionable effort during the measuring process.

Due to the lack of trends found within age groups, we decided to see if there were any trends from image to image. Comparing the third (8735460) and fourth (8735428) images first, as they are drastically different, is a good place to start.

The Images and Percentage of Good Data Measured

Image ID	**0-10**	**11-17**	**18+**	**Skip**	**Total (All Museum goers)**
8735447	41%	37%	22%	44%	34%
8735474	68%	37%	62%	12%	41%
8735460	69%	70%	76%	48%	68%
8735428	11%	4%	13%	7%	8%
8735482	76%	65%	83%	62%	71%

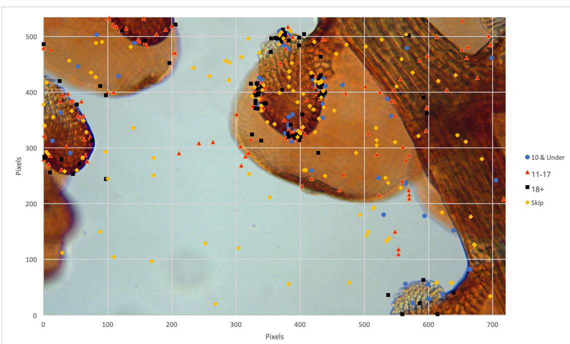
The Images and Percentage of Data within the IQR cuts

Image ID	**0-10**	**11-17**	**18+**	**Skip**	**Total (All Museum goers)**
8735447	Number that passed IQR/total number of measurement				

	s				
8735474					
8735460					
8735428					
8735482					

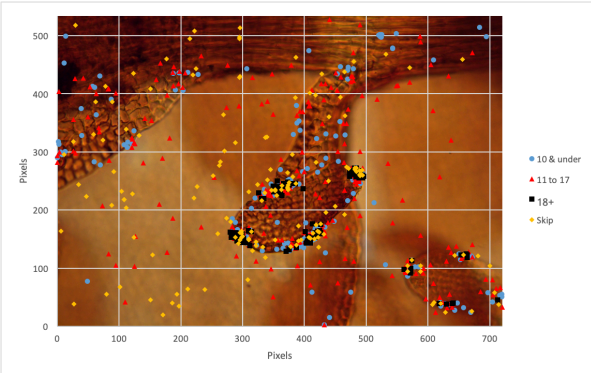
Below is our third (8735460) image. The background and the microplant have distinct colors. Within the image, the leaves to be measured are very clear, and after the data cleaning, the points are very consistent. According to our chart, 68% of the data given for this image qualified as good data. (Reminder: good data qualifications required the intersection of the points to have at least an 80 degree angle.)

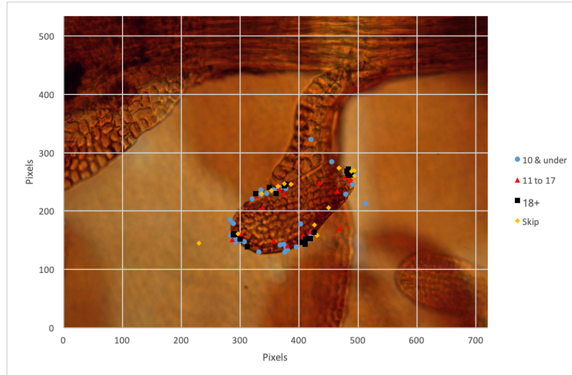
All data for image 8735460 (left) and data after angle and IQR cuts (right)



Below is our fourth (8735428) image. After a quick glance, it is fairly easy to point out the hectic nature of the points plotted on the left (bad) image. As our chart mentions, only 8% of those points were considered good data, and this can likely be attributed to the picture itself. The contrast in the color palette is low. Within the picture, the leaves are not as definitive, you can see background shades that would likely throw off some measurements, and ultimately, there are more questionable measurements.

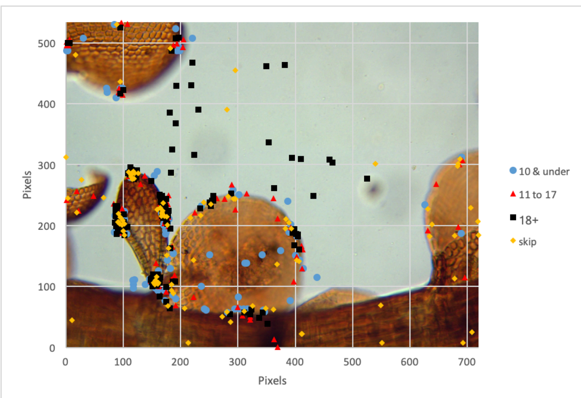
All data for image 8735428 (left) and Data after angle and IQR cuts (right)

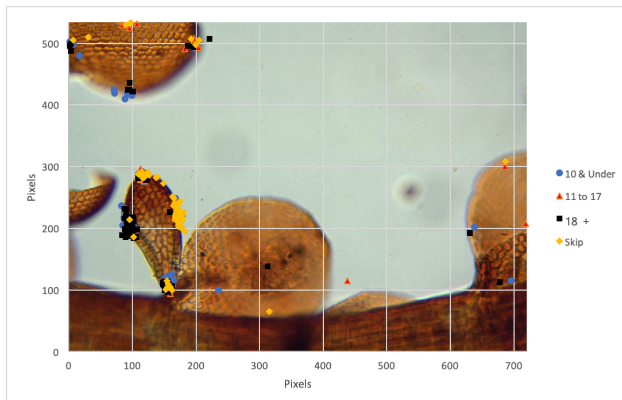




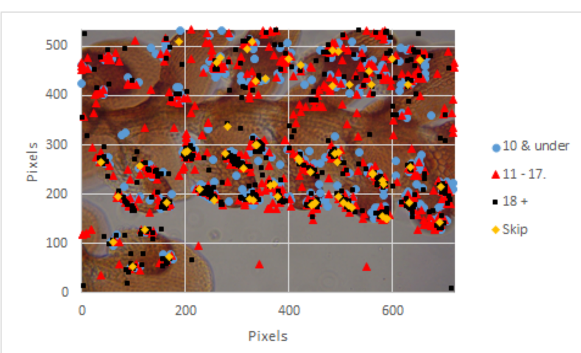
The second image ([8735474]) is quite similar to the third ([8735460]), with a high contrast in background; however, the shading within the larger leaf led many people to believe that they were valid lobules.

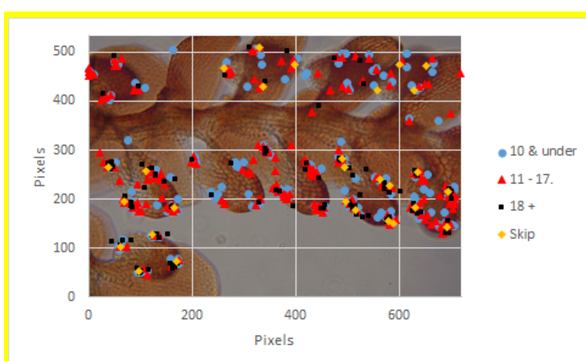
[All data for image 8735474] (left) and [Data after angle and IQR cuts] (right)



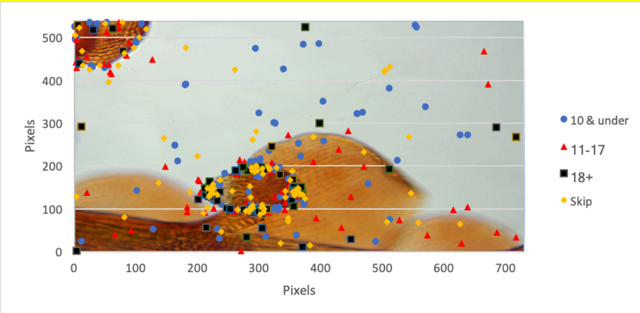


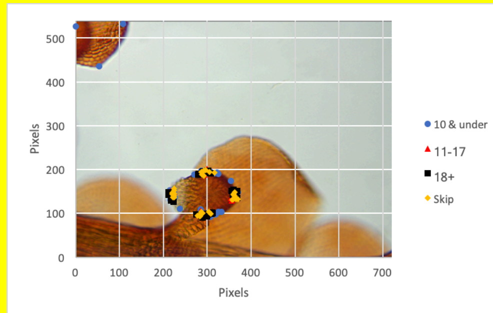
The first image (8735447) has a large number of valid lobules as well as some that are close to the edge of the screen. The contrast is medium. Visually, the measurements cover the lobules as well as parts of the stem. After cleaning, many valid measurements remained, although not as many as in the third image ([8735460]).
All data for 8735447 (left) and data after IQR cuts (right)





Joannie, Odaliz, & Ken Image (8735482)





The final microplant image (8735482) has a small part on top that resembles a lobule and on the bottom some of the shapes have more revealing scales than others. But the contrast with the background is quite large, and the overall measurements were good. For this image, we had a set of expert measurements to compare the public's measurements to. The data where we cut by both angles and IQR left almost entirely the data on the correct lobule, with the exception of one set on the portion of the leaf on the top left, as you can see from the visualization.

Comparing the general public's measurements to an expert's measurements (ID 8735482/ 25352422)

Comparing the expert data with the data that we had, we can say that (17/119) = 14.3% of all good measurements had both small and large axis lengths within the expert's min and max measurements for the first bound. As well for (47/119) = 39.5% of all good measurements had small and large axes within 10 pixels of the expert's average measurements . The good data contains data that has 80 degrees or higher.

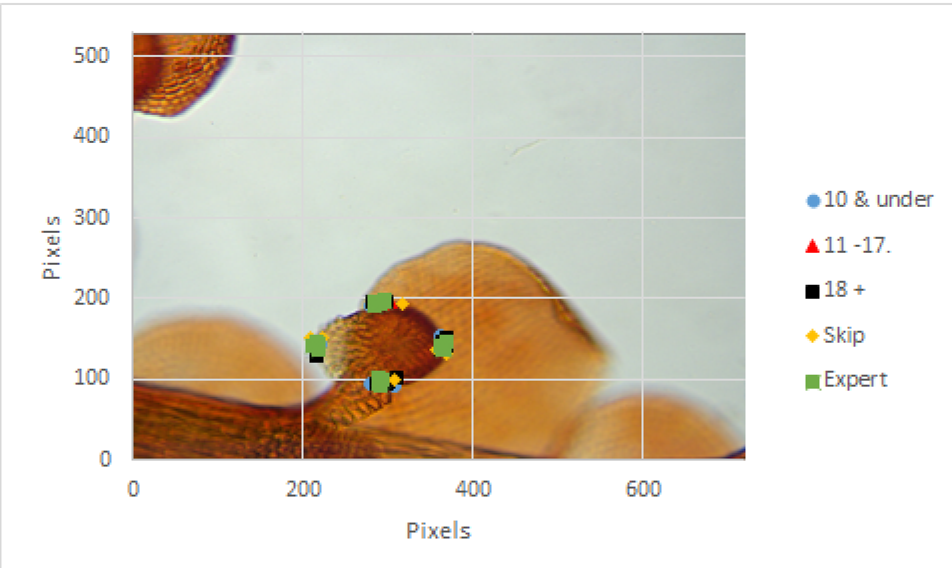
	10 and Under	11-17	18+	Skip	**Total**
Good Data (angle 80 degrees or more)	37	25	36	21	**119**
Good Data within the min and max axis lengths recorded by the Expert	2	2	11	2	**17**
Good Data with axis lengths within 3 standard deviations of the Expert's average	11	7	25	4	**47**

	10 and Under	11-17	18+	Skip	**Total**

Good Data within the min and max axis lengths recorded by the Expert	5.4%	8.0%	30.6%	9.5%	14.3%
Good Data with axis lengths within 3 standard deviations of the Expert's average	29.7%	28.0%	69.4%	19.0%	39.5%

Plotting in a graph both the expert data and the data we had, we can see that the comparison is very close. You can see below that expert was missing data but it is very close to the data that we had.

Graph one (The subset of public data for ID 8735482 that was within max/min lengths).



Graph 2 (The subset of public data for ID 8735482 that was within 3 stdev of the expert)

When looking at the total good data, we see that close to forty percent of the population's axis length measurements lie within three standard deviations of the expert's average axis length measurements (both minor and major). Moreover, less than 15% of the population recorded axis length measurements both within the minimum and maximum axis length measurements recorded by the expert. However, nearly 70% of the 18+ population recorded both axis length measurements within three standard deviations of the expert's average, while less than a third of every other age group were able to do the same. Thus, we can conclude that many kiosk visitors will struggle to record measurements close to that of the expert's, with the 18+ group being relatively more successful.

Comparing what the public predicts the leaf measurements will be to what the expert predicts they will be. - look above.

Looking at just “good data” based on angle, the public predicts the average minor axis length to be 117.22 pixels with standard deviation 69.04 and the major axis length to be 159.23 with standard deviation 84.44.
Looking at just “good data” based on angle and removing points that are more than 1.5 IQR away from the median, the public predicts the average minor axis length to be 96.20 pixels with standard deviation 8.56 and the major axis length to be 135.89 with standard deviation 15.13.
The expert predicts the average minor axis length to be 93.68 pixels with standard deviation 3.00 and the major axis length to be 142.65 with standard deviation 3.26.
This tells us that the public tends to overestimate the lengths of both minor and major axes relative to the expert’s measurements

Vicky A second expert measurement: (the one ending in 026, ID 25352420)

In order to show that our process of cleaning and cutting the data leads to measurements which are close to expert measurements, we applied the process to a sixth leaf image. This image (25352420) had not been used to determine the data cleaning procedure, so it is a useful way to check that our procedure was not biased by the images used to create it. For this leaf image (25352420), the expert predicted the smaller axis length to be 96.49 pixels with standard deviation of 3.99, and the larger axis to be 193.79 pixels with a standard deviation of 2.92. When the general public measured it, they found (after cutting for angle and outliers using IQR) that the smaller axis length to be 97.34 pixels with standard deviation of 8.34, and the larger axis to be 187.83 pixels with standard deviation of 6.17.

	Major axis length	Stdev	Minor axis length	Stdev
Expert	193.79	2.92	96.49	3.99
Public without outliers	187.83	6.17	97.34	8.34

These are statistically the same. This is evidence that removing faulty data using the IQR bounds leads to a dataset which can produce a good measurement.

Future Directions

From this study, it seems like the public is capable of producing a reasonably good set of measurements. Scientists who want to distinguish between species whose size differences are significant (such as 20% different) would be able to use work from the public; however, if the difference is very subtle (such as a 1% difference in length), it wouldn’t be possible. This is partly due to the limitations of touch screen technology and partly due to the measurement skills of the public. But the limiting factor for this particular experiment appears more technological in nature.

For cases where multiple types of specimens may exist in an individual image, there could be multiple types of measurements per image. This would make the IQR analysis ineffective at finding outliers, and so more subtle methods would need to be employed. But if one found clusters of different sizes, it may be possible to have some sort of machine learning algorithm to help with the data processing.

Currently, a follow-up study to determine which properties of microplant images are most easily established by the public is being done.

It's good to have a museum exhibit that can help to generate new scientific results. This allowed the public to be active, rather than passive, participants in the creation of new scientific works during their museum visit.

Bibliography

von Konrat, M., Campbell, T., Carter, B., Greif, M., Bryson, M., Larraín, J., Trouille, L., et al. 2018. Using citizen science to bridge taxonomic discovery with education and outreach. *Applications in Plant Sciences* 6(2): e1023.

GitHub account for this project, which includes some sample Excel files, Alex work with powershell, documentation etc.
https://github.com/AlexLabontu/Microplant-code

Although the use of museums as a means of generating scientific information from the public are limited, there are other scientific groups which are performing analyses using community scientists online. For example, community scientists were able to classify particle decay in LHC collisions with efficiency similar to a computer algorithm. (ATLAS note)

'That looks weird' - evaluating citizen scientists' ability to detect unusual features in ATLAS images of LHC collisions, Barr+, 2016

Ideas: Technical paper, bio/museum one?
Matt come in next week to give some opinions on paper flow
RIO: https://riojournal.com/