Real Data Are Messy! Cleaning, Organizing, and Drawing Meaning from Raw Data on Maple Trees

Nicole Williams
St. Lawrence University
nwill15@stlawu.edu

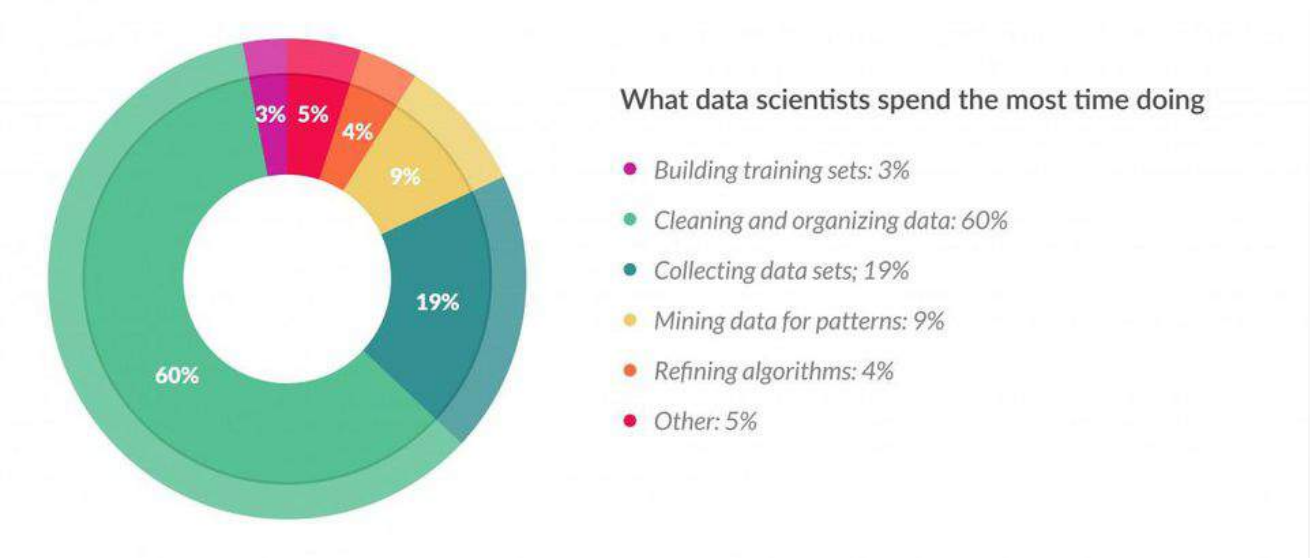
Jeffrey McLean
St. Lawrence University
jmclean@stlawu.edu

What is Real Data?

GAISE recommendation

- Integrate real data with a context and a purpose.

- How do Data Scientists work with real data?



NatureUpNorth

A community-based organization whose mission is to foster a deeper sense of appreciation for, and connection to, the North Country environment.

Nature Up North
natureupnorth.org

Fall Maple Monitoring

Please also enter data at:
http://natureupnorth.org/form/maple-monitoring

Date_____

Name(s) _____ Tree ID (if applicable)_____

Maple Species (Circle One): Sugar Maple Red Maple Silver Maple Norway Maple

Latitude, Longitude _____

Is your tree within 100 feet of buildings, concrete, or asphalt?

- o Yes
- o No

What kind of habitat is your site?

- o Home lawn
- o Home garden
- o School lawn
- o School garden
- o School paved area
- o City or Community Park (developed)
- o Natural Setting (forest, open space)
- o Other

Describe the shading at this site:

- o Open (more than 5hr per day of direct sun)
- o Partially Shaded (2-5hr per day of direct sun)
- o Shaded (less than 2hr per day of direct sun)

Tree Circumference (inches): _____

Damages/Diseases: _____

Leaves Changing Colors

- o Early: Only a few leaves have turned color (less than 10%)
- o Middle: Many leaves have turned color
- o Late: Most or all leaves have turned color (more than 90%)

Leaves Dropping

- o Early: Only a few leaves have dropped (less than 10%)
- o Middle: Many leaves have dropped
- o Late: Most or all leaves have fallen (over 90%)

Fruit

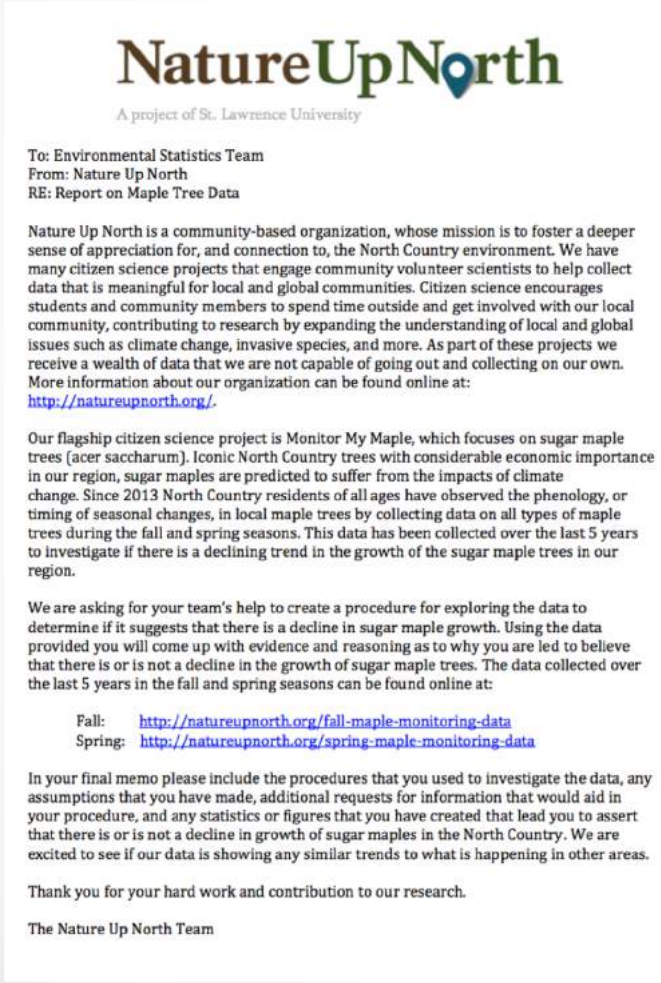
- o None: Tree is not fruiting this year
- o Early: Only a few ripe fruits are visible (less than 10%)
- o Middle: Many fruits are ripe
- o Late: Most or all fruits or seeds have been dispersed from tree (over 90%)

Comments: _____

Enter More Observations on Back →

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	•A_•A_"Fall Maple Monitoring"																					
2	Submission Details											Upload a photo of your maple, taken within the last 24 hours										
3	Serial	SID	Time	Draft	IP Address	UID	Username	Observer Nam	Tree ID N	Date	Maple Species	Name	Filesize (I	Latitude	Longitude	Is your s	What kind of	Describe the	Tree Circumf	Dise	Oth	Lea
4	553	###	09/16/2015	0	163.153.124.	3878	tyra skidds	Tyra Skidds		9/9/15	Silver Maple			44415188	75292200	Yes	Home lawn	Partially Shac	57	my tree is healthy and has		Early: Only a
5	554	###	09/17/2015	1	163.153.124.	3192	Nancy Putma	brandon stamper		9/10/15	Sugar Maple			44372599	75242788	No	Natural Setti	Partially Shaded (2-5hr	per day of direct sun)			Early: Only a
6	555	###	09/18/2015	0	163.153.160.	3868	Melanie Peal	melanie and chandra		9/18/15	Red Maple			44.166678	-75.048459	Yes	School lawn	Partially Shac	23	broken/missing limbs		Middle: Mar
7	556	###	09/18/2015	0	163.153.160.	3867	Emma Leigh	Emma and Bryant		10/13/15	Sugar Maple			44.167031	-75.04949	No	School lawn	Shaded (less	18	Leaves seem to have mani		Late: Most c
8	557	###	09/18/2015	0	163.153.160.	3874	braelyn tebo	mallory peabody		9/18/15	Sugar Maple			44.166344	75.048097	No	Natural Setti	Open (more	36	leaves have lots of holes ir		Early: Only a
9	558	###	09/21/2015	0	163.153.124.	3882	matthew kin	matthew kinch		9/18/15	Sugar Maple			44372670	75242560	No	Natural Setti	Partially Shac	7	split trunk otherwise healt		Early: Only a
10	559	###	09/22/2015	0	70.42.29.3	3797	Joshua Jame	Joshua J. Loug	68	9/22/15	Red Maple			44.56082	-74.950115	Yes	School lawn	Open (more	14			Early: Only a
11	560	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 17 mjc	17	9/22/15	Sugar Maple			44.611465	75.169554	No	Natural Setti	Partially Shac	35			Early: Only a
12	561	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 31 LRH	31	9/22/15	Sugar Maple			44.36412	75.10104	No	Natural Setti	Partially Shac	45			Early: Only a
13	562	###	09/23/2015	0	163.153.19.1	2147	Creurer	CANTON 25 Gi	25	9/22/15	Sugar Maple			44.61455	75.169544	No	Natural Setti	Partially Shac	9			Early: Only a
14	563	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 28 LH I	28	9/22/15	Sugar Maple			44.611443	-75.16954	No	Natural Setti	Partially Shac	25			Early: Only a
15	564	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 30 SPG	30	9/22/15	Sugar Maple			44.610219	75.170224	No	Natural Setti	Partially Shac	22	No damage	Young tree, h	Early: Only a
16	565	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 33 HT	33	9/22/15	Sugar Maple			44.611735	75.169559	No	Natural Setti	Shaded (less	24	moss		Early: Only a
17	566	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 27 ahjl	27	9/22/15	Sugar Maple			44.610527	75.169689	No	Natural Setti	Shaded (less	31			Early: Only a
18	567	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 26 LFM	26	9/22/15	Sugar Maple			44.611458	-75.169593	No	Natural Setti	Shaded (less	8		Was rubbed	Early: Only a
19	568	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 29 KM	29	9/22/15	Sugar Maple			44.611203	75.169027	No	Natural Setti	Shaded (less	12	moss on trunk, leaves are		Early: Only a
20	569	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 34 CY	34	9/22/15	Sugar Maple			44.611451	-75.169546	No	Natural Setti	Partially Shac	48	broken missing limbs		Early: Only a
21	570	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 40 XJ	40	9/22/15	Sugar Maple			44.610247	-75.169787	No	Natural Setti	Partially Shac	26			Early: Only a
22	571	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 44 ewa	44	9/22/15	Sugar Maple			44.611459	75.16959	No	Natural Setti	Partially Shac	21		branches	Early: Only a
23	572	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 35 AB	35	9/22/15	Sugar Maple			44.611436	75.16954	No	Natural Setti	Partially Shac	37	none		Early: Only a
24	573	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 37 CCH	37	9/22/15	Sugar Maple			44.611433	75.169584	No	Natural Setti	Partially Shac	17	Broken branches and no n		Early: Only a
25	574	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 38 EFH	38	9/22/15	Sugar Maple			44.611188	-75.169001	No	Natural Setti	Partially Shac	14	No		Early: Only a
26	575	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 41 CO.	41	9/22/15	Sugar Maple			44.608917	75.171201	No	Natural Setti	Partially Shac	27	None		Early: Only a
27	576	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 36 pc a	36	9/22/15	Sugar Maple			44.611451	75.169559	No	Natural Setti	Partially Shac	26			Early: Only a
28	577	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 42 HPL	42	9/22/15	Sugar Maple			44.611236	-75.169032	No	Natural Setti	Partially Shac	38	moss on the tree and a fe		Early: Only a
29	578	###	09/23/2015	0	163.153.19.1	2147	Creurer	kyle j. hughes	39	9/22/15	Sugar Maple			44.611434	-75.169531	No	Natural Setti	Shaded (less	39	some dead branches at to		Early: Only a
30	579	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 43 ETNR		9/22/15	Sugar Maple			44.611236	-75.169032	No	Natural Setti	Shaded (less	4	some insect damege		Early: Only a
31	580	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 39 KH	39	9/22/15	Sugar Maple			44.611434	75.169532	No	Natural Setti	Shaded (less	39	broken branches at top an		Early: Only a
32	581	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 46 NC	46	9/22/15	Norway Maple			44.605492	-75.167624	Yes	School lawn	Open (more	119	Fungus on le	cool!	Early: Only a
33	582	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 51 KL S	51	9/22/15	Sugar Maple			44.605511	75.167999	Yes	School lawn	Open (more	71	Fungus, broken branches,		Early: Only a
34	583	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 49 HI E	49	9/22/15	Norway Maple			44.606267	-75.167331	Yes	School lawn	Open (more	70	black fungi on leaves		Early: Only a
35	584	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 48 PH	48	9/22/15	Sugar Maple			44.605362	75.167645	Yes	School lawn	Partially Shac	104	Spots on leaves		Early: Only a
36	585	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 53 TM	53	9/22/15	Norway Maple			44.605998	-75.167856	Yes	School lawn	Open (more	68	none		Early: Only a
37	586	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 45 oa	45	9/22/15	Silver Maple			44.605323	75.167859	Yes	School garde	Open (more	80	black splotches		Early: Only a
38	587	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 55 HS	55	9/22/15	Norway Maple			44.605696	75.167442	Yes	School lawn	Open (more	104			Early: Only a
39	588	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 55 HS	55	9/22/15	Norway Maple			44.605696	75.167442	Yes	School lawn	Open (more	104			Early: Only a
40	589	###	09/23/2015	0	163.153.19.1	2147	Creurer	Canton 47 GD	47	9/22/15	Silver Maple			44606298	-75.167596	Yes	School lawn	Open (more	4			

Real data are messy



NatureUpNorth

A project of St. Lawrence University

To: Environmental Statistics Team
From: Nature Up North
RE: Report on Maple Tree Data

Nature Up North is a community-based organization, whose mission is to foster a deeper sense of appreciation for, and connection to, the North Country environment. We have many citizen science projects that engage community volunteer scientists to help collect data that is meaningful for local and global communities. Citizen science encourages students and community members to spend time outside and get involved with our local community, contributing to research by expanding the understanding of local and global issues such as climate change, invasive species, and more. As part of these projects we receive a wealth of data that we are not capable of going out and collecting on our own. More information about our organization can be found online at: http://natureupnorth.org/.

Our flagship citizen science project is Monitor My Maple, which focuses on sugar maple trees (acer saccharum). Iconic North Country trees with considerable economic importance in our region, sugar maples are predicted to suffer from the impacts of climate change. Since 2013 North Country residents of all ages have observed the phenology, or timing of seasonal changes, in local maple trees by collecting data on all types of maple trees during the fall and spring seasons. This data has been collected over the last 5 years to investigate if there is a declining trend in the growth of the sugar maple trees in our region.

We are asking for your team's help to create a procedure for exploring the data to determine if it suggests that there is a decline in sugar maple growth. Using the data provided you will come up with evidence and reasoning as to why you are led to believe that there is or is not a decline in the growth of sugar maple trees. The data collected over the last 5 years in the fall and spring seasons can be found online at:

Fall: http://natureupnorth.org/fall-maple-monitoring-data
Spring: http://natureupnorth.org/spring-maple-monitoring-data

In your final memo please include the procedures that you used to investigate the data, any assumptions that you have made, additional requests for information that would aid in your procedure, and any statistics or figures that you have created that lead you to assert that there is or is not a decline in growth of sugar maples in the North Country. We are excited to see if our data is showing any similar trends to what is happening in other areas.

Thank you for your hard work and contribution to our research.

The Nature Up North Team

How do you make sense of large messy data sets?

Thank You!

Comments/Questions?

Email: nwill15@stlawu.edu

References

GAISE (2016). Guidelines for assessment and instruction in statistics education. College report. Alexandria, VA: American Statistical Association. Retrieved from: www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf]

Nature up north. Retrieved from: https://www.natureupnorth.org/

Press, G. (2016, March 23). Cleaning big data: Most time-consuming, least enjoyable data science task, survey says. Retrieved from: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#1e0089a56f63