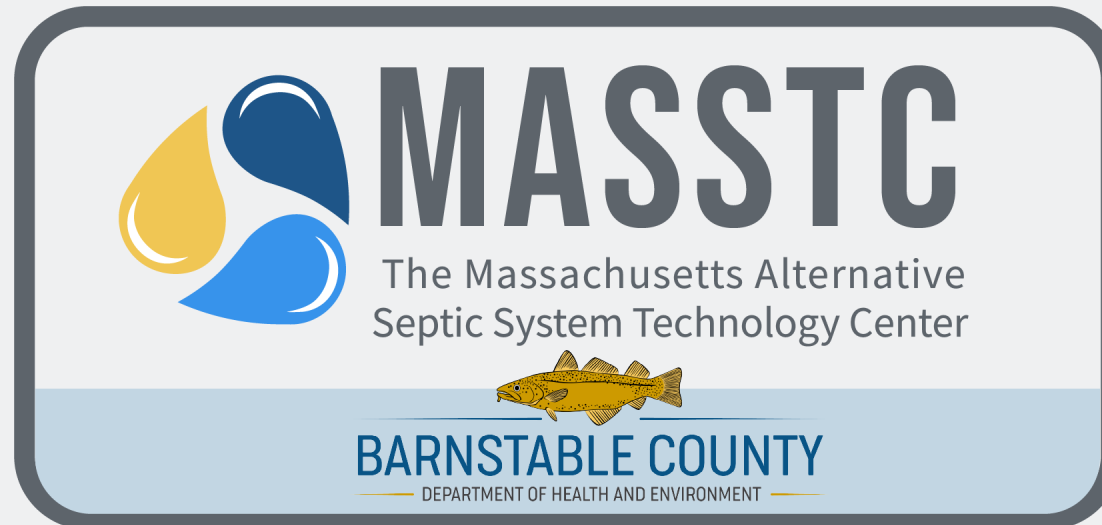


Dancing Through Data

Emily Michele Olmsted

Massachusetts Health Officers Association

Thursday, November 6, 2025

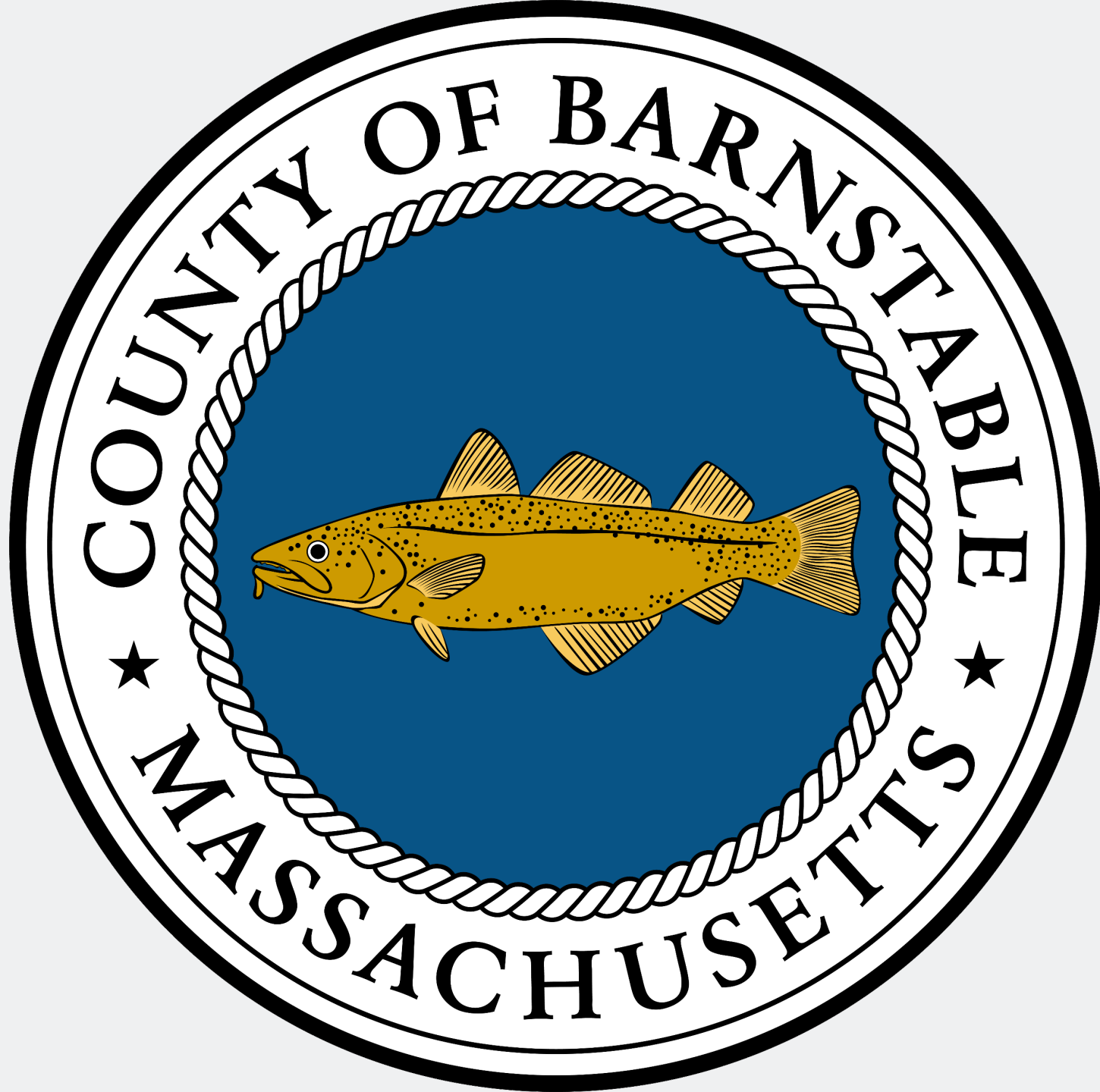


This project may have been financed in whole or in part with funds (competitive grants or otherwise) from the Environmental Protection Agency (EPA) and/or the Massachusetts Department of Environmental Protection (MASSDEP). The contents do not necessarily reflect the views and policies of EPA or MASSDEP, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.

Any mention of a name, product, service, company, or institution does not constitute an endorsement by MASSTC or Barnstable County Regional Government.

Key Takeaways

- My work is primarily focused on data for I/A septic systems
 - There are ***many*** different ways to look at data
- My team and I usually follow certain patterns when looking at I/A septic system data
- Microsoft Excel can help cut down on time spent working with data
- I have some favorite formulas I recommend in Excel, and have a tool I have developed using them (that will hopefully inspire you)





MASSTC

The Massachusetts Alternative
Septic System Technology Center



BARNSTABLE COUNTY

DEPARTMENT OF HEALTH AND ENVIRONMENT

MASSTC



The Massachusetts Alternative
Septic System Test Center



A Division of

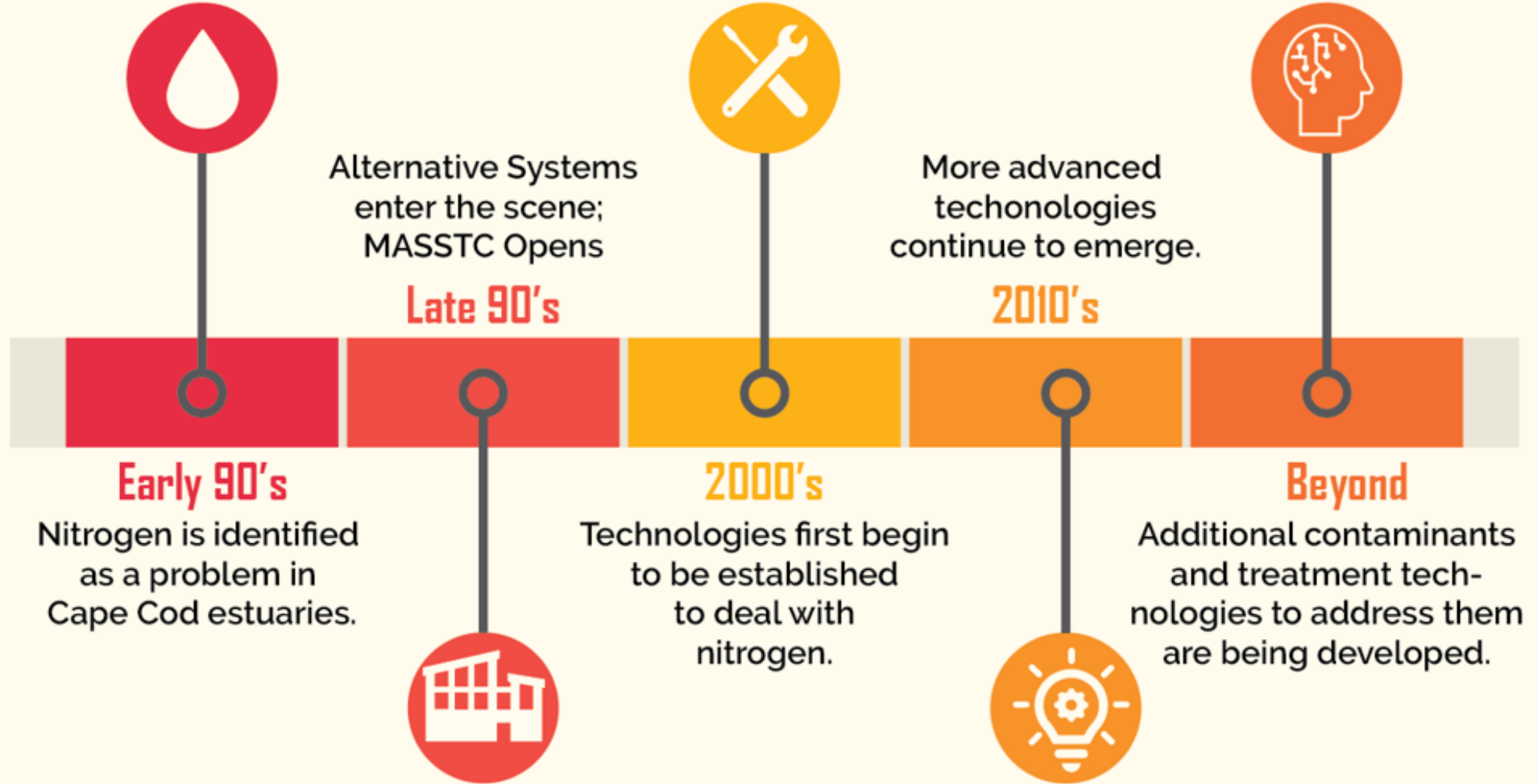
**Barnstable County Department
of Health and Environment**

4 Kittridge Road | JBCC Building #599

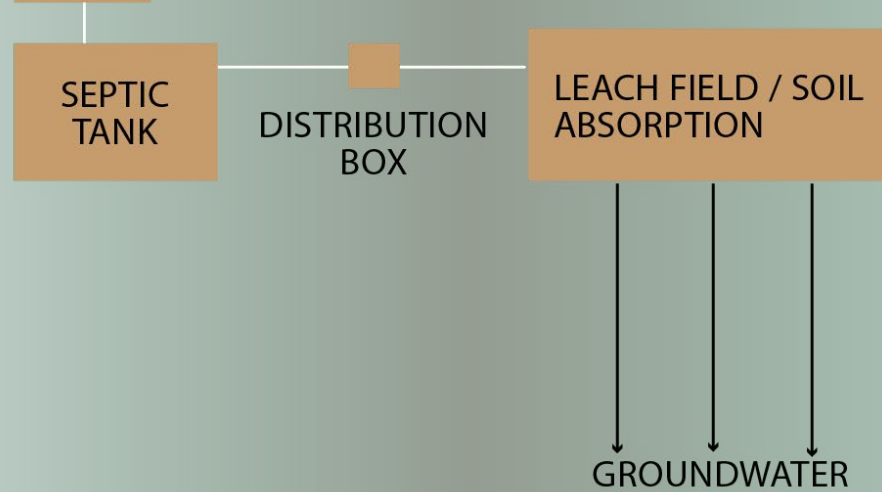
DELIVERIES
&
VISITORS



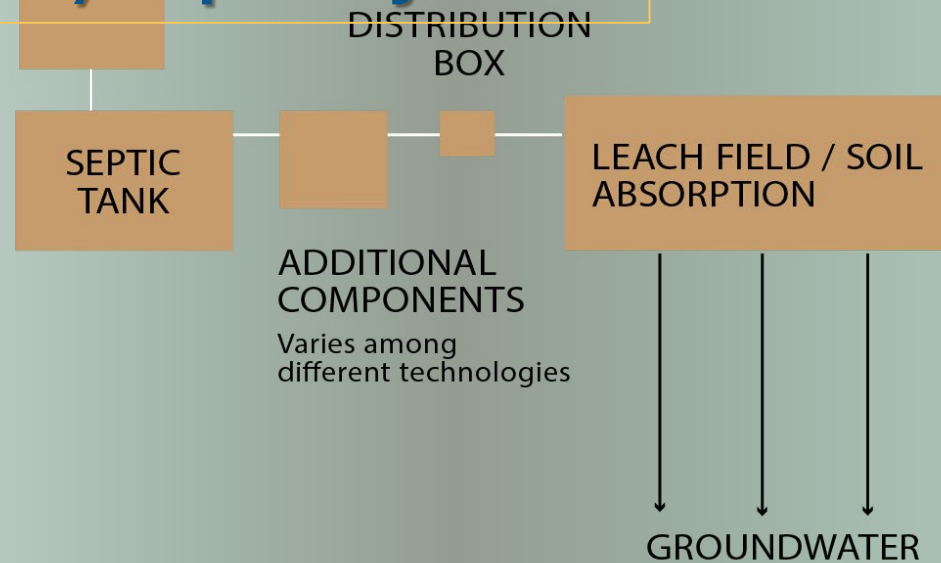
About MASSTC



Traditional Title 5 Septic System



Innovative/Alternative (I/A) Septic System



MASSTC

The Massachusetts Alternative Septic System Test Center (MASSTC) is the nation's premier third-party testing and research facility for innovative/alternative (I/A) onsite septic treatment technologies. Since 1999, MASSTC has tested dozens of technologies for industry standard certifications, provided space for technology vendors to do R&D work, and conducted original research on topics from constructed wetlands to nutrient and virus removal.

Beginning in the early 1990s, Barnstable County, Massachusetts (Cape Cod) began realizing that nitrogen from septic systems was dramatically affecting our marine environments, causing accelerated eutrophication and nuisance algae blooms.

Searching for technologies to address the issue, the Barnstable County Department of Health and Environment, in conjunction with Massachusetts Coastal Zone Management through the Buzzards Bay Project, began what is now called the Massachusetts Alternative Septic System Test Center (MASSTC). The Test Center began in 1999 testing advanced onsite septic treatment technologies in triplicate under a program of the U.S. Environmental Protection Agency (EPA) called the Environmental Technology Initiative (ETI). Working with EPA and the National Sanitation Foundation (NSF) of Ann Arbor, Michigan, MASSTC conducted a refined nutrient testing protocol in 2002 referred to as the Environmental Technology Verification (ETV). Further refinement of the nutrient standards was completed in 2007 by NSF and has resulted in the NSF Standard 245.

Following over two decades of successful operation thanks to Barnstable County staff, state and federal grant funding and industry support (private companies are issued fees for space, testing and staff oversight), MASSTC continues to serve as an economic driver of the onsite wastewater treatment industry by fostering local and out-of-state investment in innovative/alternative septic system technology. Its facility can accommodate over 20 concurrent tests, allowing companies to conduct research and development on their products or complete any number of standardized test protocols. MASSTC has since become a premier onsite wastewater treatment research facility, hosting numerous research and development efforts by private manufacturers, states and municipalities and conducting protocols for the National Sanitation Foundation International (NSFI).

About Me

I serve as the Quality Assurance Manager for MASSTC.

I help organize our processes to try to make things go more smoothly.

I also handle large amounts of data and spend a lot of my time trying to figure out how to use it, how to let others use it, how to make sure it's as accurate as possible (and what we can do better as a staff to provide the same experience to our clients for all systems).

When I first came into the Barnstable County Department of Health and Environment, I worked with the real-world I/A tracking database (from late 2015 until early 2020).



Welcome!

Barnstable County provides Innovative/Alternative compliance and tracking services.

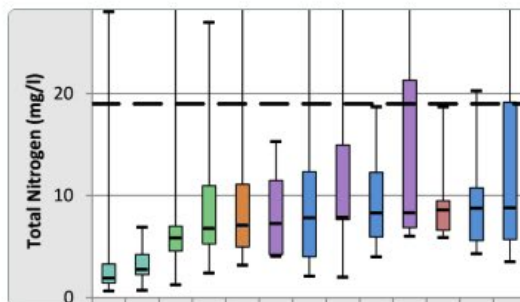
[Learn more »](#)



Barnstable County Innovative/Alternative Septic System Tracking Program

<https://septic.barnstablecountyhealth.org/>

Data and Statistics



In order to provide information on

I/A Owners Guide



I/A system owners have the responsibility of

For I/A Operators

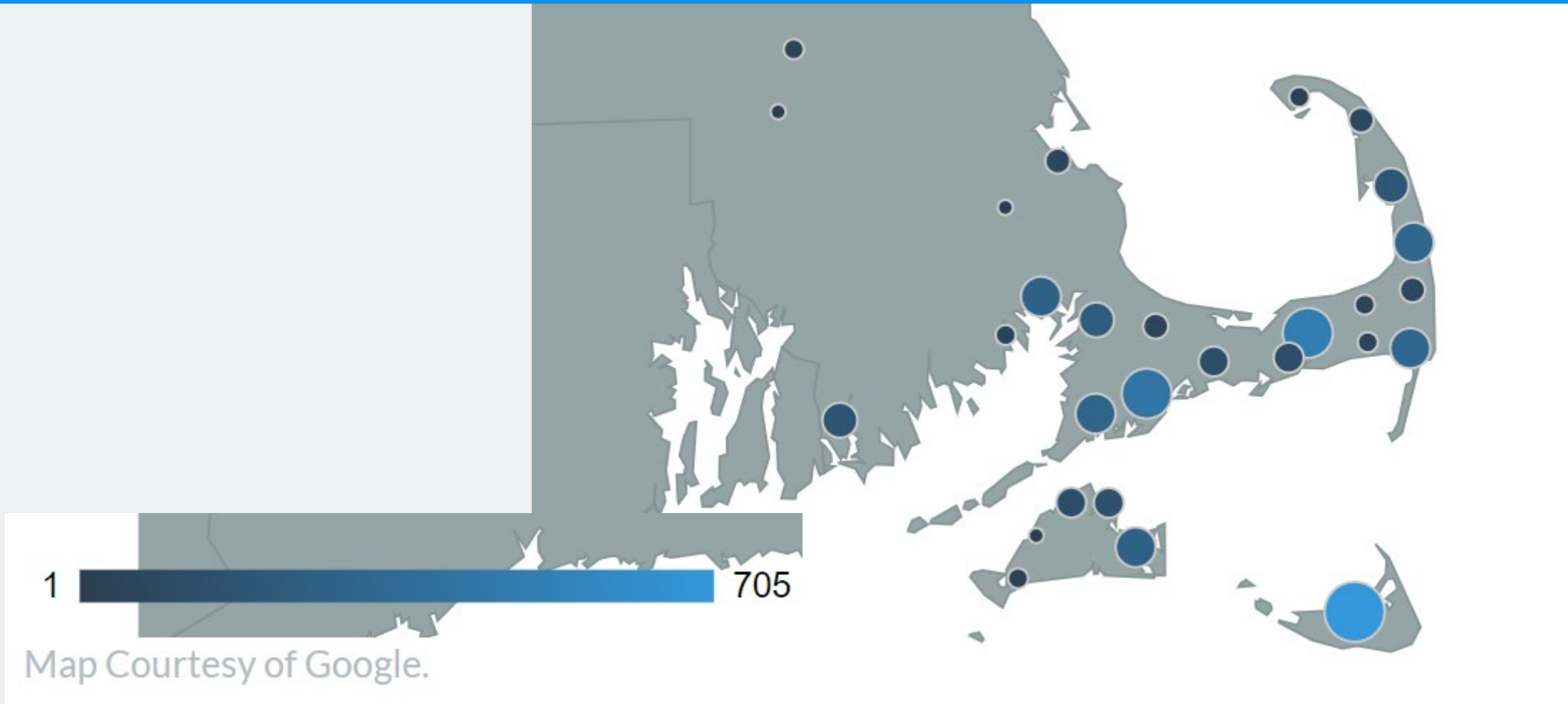


If you are an O&M Provider and need help with



Barnstable County Innovative/Alternative Septic System Tracking Program

<https://septic.barnstablecountyhealth.org/>



Barnstable County I/A Tracking Program

- Tracks real-world I/A systems on Cape Cod and many of the surrounding towns
 - All 15 towns on Cape Cod
 - Nantucket and most of Martha's Vineyard
 - 7 towns “over the bridge”
- Operators report sample results to this database and the town Health Agents access it
- Currently more than 67,000 sample reports (as of October 14, 2025)

Barnstable County I/A Tracking Program

- Operators submit samples
- We do not check that the samples were submitted correctly
- Samples can be submitted for more than once
- Data could also be missing or the permit information could have been given to us incorrectly
 - For example, a Bioclere may have been called an Advantex by the town, and so it will appear as an Advantex when we export large datasets

Barnstable County I/A Tracking Program

Full Disclaimer

Data in the I/A Tracking Database is submitted by wastewater operators, usually those who hold the Operation and Maintenance agreement, and/or their staff. Barnstable County does not check this data to ensure that samples were taken properly, nor that data have been submitted correctly. It is possible that data is missing from this database if it has not been submitted by the operator, or that the data that has been submitted is false or incorrect.

It is also possible that other information in the database is incorrect. Barnstable County staff do their best to get as accurate information as possible when setting up a permit, but it is possible that some of this information is incorrect or missing. This also ties into data because if, for example, we are informed that a certain technology and model are onsite (ex. The Rainbow System, model Green), and we are requested to provide data on this particular system, any data submitted for this component will appear even if we were provided incorrect information about the technology and/or model. Similar limitations apply to any dates, information about flow, etc.

This dataset is imperfect due to the nature of submissions, and so anyone using the data should be aware of the limitations and the Terms of Service:

The data you are about to review derive from a self-reporting system for alternative onsite septic systems located in 14 of the 15 towns in Barnstable County Massachusetts maintained by the Barnstable County Department of Health and Environment. The mention of any product on this website does not constitute an endorsement by Barnstable County. The validity of the data is qualified only by the licenses and integrity of Licensed Wastewater Treatment Operator who collected the sample and the Massachusetts Certified Laboratory that performed the analyses. This dataset includes samples taken at seasonally-occupied as well as full-time-occupied residences. Data during start up of the systems are also included. The reader/user is cautioned to understand the limits of these data in regard to assessing the performance of any technology and agrees that any re-representation or copying of these data will include the above qualification.

Microsoft Office

- The number of Teams users jumped from 20 million in 2019 to 115 million in 2020.
- Microsoft Office 365 controls nearly half of the office productivity software market.
- In August 2022, Windows occupied 28.79% of the Operating System market share worldwide.
- Windows runs on 1.6 billion active devices in 2022.

Microsoft: The Enduring Tech Superstar

Daniel Shvartsman

Updated August 1, 2025

<https://www.investing.com/academy/statistics/microsoft-facts/>

**With a large dataset, you may follow certain patterns
for accepting and rejecting data**

Exclusions Matter!

example

Total Nitrogen =

Nitrate-nitrogen + Nitrite-nitrogen + Total Kjeldahl Nitrogen

- Without nitrate, or without TKN, you probably do not have an accurate Total Nitrogen value.
- We would choose to exclude a result that has either nitrate or TKN not submitted.

Average vs. Median

Median is the middle number in a set of numbers.

Median is the more skew-resistant.

#	Data Set	Average	Median
1	{1, 3, 5, 9, 12}	6	5
2	{10, 16, 18, 25, 67}	27.2	18
3	{1, 24, 27, 34, 35}	24.2	27
4	{2, 6, 7, 45, 48}	21.6	7

Average vs. Median

The following is a real table of Total Nitrogen data, presented as samples taken over approximately 4 years

Sample Date	Effluent Total Nitrogen (mg/L)
2021-07-13	9.05
2021-10-13	18.40
2022-01-10	10.26
2022-04-28	14.69
2022-07-22	5.18
2023-01-04	24.20
2023-08-28	17.00
2023-11-27	19.00
2024-04-24	43.00
2024-06-20	43.21
2024-09-25	18.24
2024-12-11	8.60
2025-02-25	28.64

Average: 19.96 mg/L

Median: 18.24 mg/L

Average vs. Median

Here's another real table of Total Nitrogen data, presented as samples taken over approximately 4 years

Sample Date	Effluent Total Nitrogen (mg/L)
4/12/2018	55.6
4/12/2018	7
6/6/2018	77.7
7/17/2018	6.08
8/9/2018	8.42
9/6/2018	10.75
10/8/2018	5.7
12/7/2018	24.6
2/8/2019	30.3
3/19/2019	60.2
4/11/2019	63.4
5/15/2019	18.3
6/5/2019	5.56
7/22/2019	9.33
9/3/2019	10.4
11/18/2019	6.7
2/10/2020	11.8
8/10/2020	87
8/26/2020	8.15
10/14/2020	9.69
10/20/2020	9.03
3/16/2021	80.5
5/14/2021	12.7
5/19/2021	12
8/10/2021	7.03
11/10/2021	6.22

Average: 24.78 mg/L

Median: 10.58 mg/L



Sample Date	Effluent Total Nitrogen (mg/L)
4/12/2018	55.6
4/12/2018	7
6/6/2018	77.7
7/17/2018	6.08
8/9/2018	8.42
9/6/2018	10.75
10/8/2018	5.7
12/7/2018	24.6
2/8/2019	30.3
3/19/2019	60.2
4/11/2019	63.4
5/15/2019	18.3
6/5/2019	5.56
7/22/2019	9.33
9/3/2019	10.4
11/18/2019	6.7
2/10/2020	11.8
8/10/2020	87
8/26/2020	8.15
10/14/2020	9.69
10/20/2020	9.03
3/16/2021	80.5
5/14/2021	12.7
5/19/2021	12
8/10/2021	7.03
11/10/2021	6.22

. Median

d as samples taken over approximately 4 years

Average: 24.78 mg/L

Median: 10.58 mg/L



Excel 101

All cells have a reference point.

As an example, cell C14 is the cell that appears in Column C, in Row 14.

Excel 101

You need to use the **=** sign to tell Excel that you want it to do something

Paste

Cut

Copy

Format Painter

Clipboard

Aptos Narrow

11

A⁺

A⁻

Font

Alignment

General

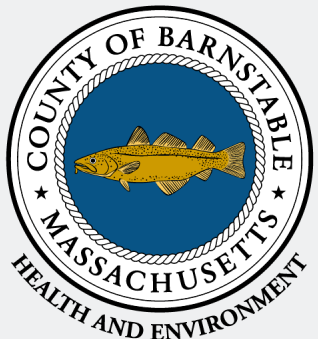
\$

%

Number

A1 : fx

	A	B	C	D	E	F	G
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							



Excel 101

=sum()

Add a group of numbers or cells together

File Home Insert Page Layout Formulas Data Review View Automate Developer New Tab Help Acrobat

Cut Copy Format Painter Clipboard

Aptos Narrow 11 B I U Font

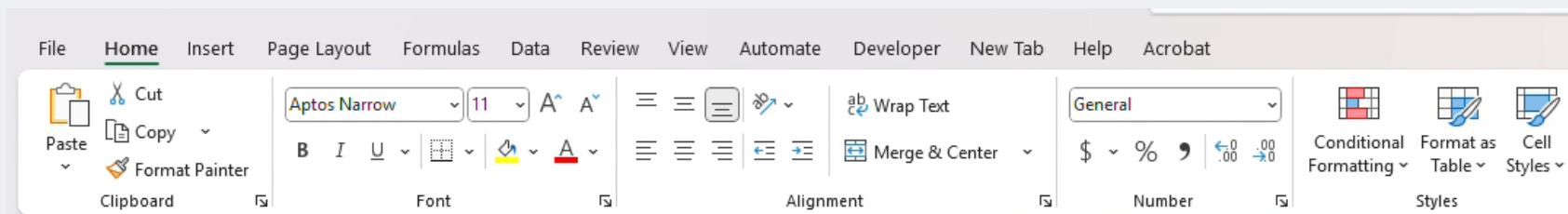
Wrap Text Merge & Center Alignment

General Number Styles

Conditional Formatting Format as Table Cell Styles

	A	B	C	D	E	F
1	List of numbers					
2	10					
3	15					
4	19					
5	25					
6	100					
7						
8						
9						
10						
11						
12						
13						
14						





	A	B	C	D	E	F
1	List of numbers		169			
2	10					
3	15					
4	19		59			
5	25					
6	100					
7						
8						
9						
10						
11						
12						
13						
14						



Excel 101

+ - * /

You can also use these signs to add, subtract, multiply, and divide

You can incorporate real numbers with cell references

FileHomeInsertPage LayoutFormulasDataReviewViewAutomateDeveloperNew TabHelpAcrobat

Paste

CutCopyFormat Painter

Clipboard

Aptos Narrow11A[^]A_v

B*I*U

Font

Wrap Text

Merge & Center

Alignment

General

\$%&#

Number

Conditional Formatting

Style

	A	B	C	D	E
1	List of numbers				
2	10				
3	15				
4	19				
5	25				
6	100				
7					
8					
9					
10					
11					
12					
13					
14					

<>

Sheet1

+

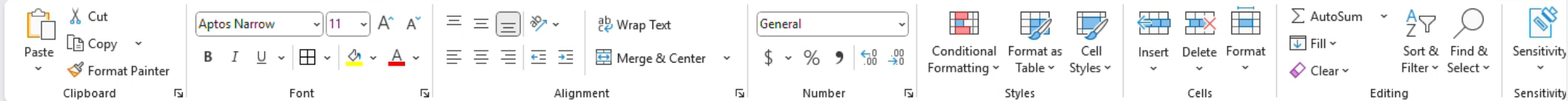
Ready

Accessibility: Good to go

Excel 101

=IF(logical_test,[value_if_true],
[value_if_false])

If (this statement is true, do this, if it's not true do this)



C5

	A	B	C	D	E	F	G	H
1	List of numbers							
2	10							
3	15							
4	19							
5	25							
6	100							
7	87							
8	45							
9	71							
10	33							
11	18							
12								
13								
14								

Excel 101

=IF(logical_test,[value_if_true],[value_if_false])

If (this statement is true, do this, if it's not true do this)

You can **nest** IF formulas to do
multiple checks

B17		
	A	B
1	Number	Formula Answer
2	1	=IF(A2>5,"Greater than 5",IF(A2=4,"It's just 4",IF(A2<10,"It's still less than 10")))
3	2	
4	3	
5	4	
6	5	
7	6	
8	7	
9	8	
10	9	
11	10	
12		



Excel 101

- =INDEX(array, row_num, [column_num])
- =MATCH(lookup_value, lookup_array, [match_type])

Use together for a more powerful result

=Return a value from this column, that matches this cell, which it should look for as appearing in this column

	C	D	E	F	G	H	I	J	K
1	Fruit/Vegetable	Quantity							
2	Apple	7	How many are there of each?						
3	Carrot	3							
4	Banana	5		+					
5	Broccoli	9							
6	Mango	3							
7	Spinach	6							
8	Blueberry	2							
9	Zucchini	5							
10	Pineapple	8							
11	Cauliflower	2							
12	Strawberry	4							
13	Kale	6							
14	Orange	7							
15	Cucumber	1							
16	Raspberry	4							
17	Lettuce	9							
18	Watermelon	3							
19	Eggplant	5							
20	Kiwi	2							
21	Bell Pepper	6							
22	Peach	8							
23	Asparagus	1							
24	Pear	7							
25	Tomato	4							
26	Grapes	2							
27	Celery	3							
28	Pomegranate	6							
29	Radish	5							



Excel 101

=UNIQUE(array, [by_col], [exactly_once])

Provides a smaller list of everything in a certain part of your sheet, even if it appears more than once

[illegible]

My Favorite Formulas

- =SUM(number 1, [number 2],...)
- =IF(logical_test,[value_if_true],[value_if_false])
- =COUNTIF(range, criteria)
- =INDEX(array, row_num, [column_num])
- =MATCH(lookup_value, lookup_array, [match_type])
- =UNIQUE(array, [by_col], [exactly_once])

Also cool useful (as you will soon see)

- =AND(logical 1,[logical 2],...) especially when used with =IF(to check that multiple things are true)
- =& (puts two cells together or a cell and wording etc.)
- =MAX(number1,{number2},...)
- =AVERAGEIF(range,criteria,[average_range])

My I/A Data Spreadsheet

I have created an Excel spreadsheet that, through the use of multiple formulas, takes data exported from the Barnstable County I/A Database and automatically excludes certain data for what we consider to be greater indication of performance, while simultaneously provides a clean, organized list of properties, results, and a calculated average and median of total nitrogen.

My I/A Data Spreadsheet

Why is this helpful?

This reduces the amount of time that human eyes need to spend poring over massive lists of results, and also reduces the chance of human error or bias.

What Exclusions and Calculations Are Done?

- This tool was built primarily to look at effluent Total Nitrogen results of I/A septic systems.
 - Data should have the same date range and characteristics
 - Summed Total Nitrogen as nitrate, nitrite, and TKN.
 - Changed any reported BRLs samples (below readable levels) to 0 mg/L.
- If nitrate or TKN was not submitted, submitted as DNS (did not sample), or submitted as a negative value, that sample event was excluded.
 - If both TKN and nitrate were BRL, the sample event was excluded.
- The formulas compare the summed TN compared to the reported TN and takes whichever is larger.
 - If TN is greater than 1000 mg/L, it is excluded.
 - Duplicates were removed
 - This is based on sample event date being the same and the value being the same.
- Additional checks were performed using math. (This was done by comparing the unique value of adding nitrate + nitrite + TKN compared to the product of nitrate X nitrite X TKN against other samples and division of nitrate, nitrite, and TKN.

What Exclusions and Calculations Are Done?

- This tool was built primarily to look at effluent Total Nitrogen results of I/A septic systems.

What Exclusions and Calculations Are Done?

- Data should have the same date range (ex. the same 4 years) and characteristics (ex. single family residential systems)

What Exclusions and Calculations Are Done?

- Total Nitrogen is the sum of nitrate, nitrite, and TKN.
 - $TN = NO_3 + NO_2 + TKN$

What Exclusions and Calculations Are Done?

- Any reported BRLs sample (below readable levels) was changed to 0 mg/L – needed numerical stand-in.

What Exclusions and Calculations Are Done?

- If nitrate or TKN was not submitted, if either was submitted as DNS (did not sample), or if either was submitted as a negative value, that sample event was excluded.
- If both TKN and nitrate were both BRL, the sample event was excluded.

What Exclusions and Calculations Are Done?

- The formulas compare the summed TN compared to the TN reported by the operator, and takes whichever is larger to consider it as the “final” TN for that sample event.
- If TN is ever greater than 1000 mg/L, it is excluded.

What Exclusions and Calculations Are Done?

- Duplicates were removed
 - This is based on sample event date being the same, but the operator submitted it again (either as the same result or updated it with a different submission date and time)
 - Additional checks were performed using basic math (ex. by comparing the unique value of the product of nitrate X nitrite X TKN against other sample)

Export this from septic database:

Home > Permit Samples > Edit Permit Sample

Cancel Excel

Date Filters

Start Date: 2018-01-01

End Date: 2024-03-15

Permit Filters

Municipality: All Municipalities

Contractor: All Contractors

DEP Approval: All Approvals

Property Type: Single Family Residential

Project: All Projects

Watershed: All Watersheds

Component Filters

Technology: Bioclere

Parameters

Effluent: ☒ Influent: ☐

Select All Unselect All

☒ Alkalinity ☒ CBOD ☐ Conductance ☐ DO ☐ E. coli ☐ Fec Coliform ☐ Magnesium ☒ Nitrate ☒ Nitrite ☐ Odor ☐ Oil & Grease ☐ Orthophosphates ☐ Settleable Solids ☒ Temperature ☒ Total Dissolved Phosphorus ☐ Total Iron ☐ Total Solids ☐ Total P ☒ TSS ☐ Turbidity ☐ VOC ☐ Water Use

Parameter Tracking Filters

Tracking: All Tracking

Season Start: January

Season End: December

Events Per Season: 4

Sample Months

January February March April May June July August September October November December Clear

Select all and copy the data from the file you just exported. (Ctrl + A is select all, Ctrl + C is copy)

Paste into Original Data tab, use the headers the database gives you (for now)

Check Headings Check tab - proceed if it says ok (green)

*If it doesn't match, make sure all parameters have "Parameter name and Note" in them

*If you have to change this, do in the original exported file, then recopy and repaste

Everything should fill itself in!

Go to FINAL - copy all and paste VALUES into a new spreadsheet.

Save that spreadsheet as a .csv with initials (save as MS-DOS)

Save this sheet as the same name but as Excel (for record).

Go back to "EMO_Blank file with more parameters" and clear out Original Data.

*does not take into account DEMO sites

*final may contain "#NA"

Qualifications of Data used by George

As discussed in person at MASSTC between EMO and GH 2024-04-02

Start date 2018-01-01. EMO is going to repull data and will go until March 15, 2024.

Duplicates – someone who took a duplicate on the same date at the same location and it is reported more than once. Ex. June 15th sample is inputted in twice or it could be reported a day later or a month later. LEAVE ONE IN AND TAKE ANY DUPLICATES OUT. A duplicate is the same place on the same sample day but it's reported twice or more.

A suspect data point is one with the same data but for a different date, different collect date but the same data. TAKE BOTH OUT FOR THIS CASE.

We saw this in Advantex that even though submitted by different operators they were the same

If there was a suspected typing error (ex. 203 put in, instead of 20.3) take out the more suspect one and keep the more likely one and in this case take the more recently submitted number. DO NOT JUST AUTOMATICALLY TAKE THE MORE NEWLY SUBMITTED ONE use some judgement.

TKN or nitrate – sometimes found then by getting by TKN for nitrate and then you see a bunch of TKN that were blank and the TKN note column did not have a BRL in it. (BRL did not get discounted). Any that said DNS did not sample were also omitted.

Only nitrate, not to 800 or TSS.

Only things that were blank or sum of some of the parameters were omitted.

For BRL – what number used for stand in 0 was used. GH and EMO discussed. GH said statistic it won't make a difference between 0 and half the RL. We don't know the RL of the lab used.

*added 2024-04-09 – if both are BRLs in nitrate and TKN then exclude the data

Maybe the way to do this is to add the average if formula in there and see to remove these if average is low. GH said don't do this

2024-04-09: For those that the TN is exactly the same then do it manually. They were reported on the same day at the same time that might indicate an issue. *Add a formula to see if each parameter is the same.

GH final: sample date, permit ID, TN, operator

2024-04-11 – discussion of negative numbers. If someone enters a number above 1 (there are no TNs that are negative fractions, if they entered a -4 TKN then it's obviously above RL and the -4 will subtract from TN inappropriately)

This tab explains all of the specifics and contains instructions so you can export data the same way each time.

Let's

Try

the Spreadsheet

Data - Based on EM's built spreadsheet			
Company	Technology	Percent of systems meeting 19 mg/L, as Average TN	Percent of systems meeting 19 mg/L, as Median TN
SeptiTech, Inc. 69 Holland Street Lewiston, ME 04240	SeptiTech/Bio-Microbics of Maine, Inc.	83	82
Orenco Systems, Inc. 814 Airway Avenue Sutherlin, OR 97479	Advantex Treatment System	59	70
F.R. Mahony & Associates, Inc. 273 Weymouth Street Rockland, MA 02370	Amphidrome Process	69	88
Fuji Clean USA, LLC 41-2 Greenwood Road Brunswick, Maine 04011	Fuji Clean	20	60
Lombardo Associates, Inc. 188 Church Street Newton, MA 02458	Nitrex Filter	100	100
KleanTu LLC 300 Old Pond Road, Ste# 206 Bridgeville, PA 15017	NitROE Waste-Water Treatment System	85	91
Aquapoint.3 LLC 39 Tarkiln Place New Bedford, MA 02745	Bioclere	54	62

- For this example, I used only the technologies named in the MassDEP BANRT list
 - There are definitely more systems!
- I only used data with a sample submission falling between October 1, 2022 through October 31, 2025
- I included only single family residential systems, and did *not* exclude based on approval level (General Use, Provisional Use, etc.)

Data - Based on EM's built spreadsheet			
Company	Technology	Percent of systems meeting 19 mg/L, as Average TN	Percent of systems meeting 19 mg/L, as Median TN
SeptiTech, Inc. 69 Holland Street Lewiston, ME 04240	SeptiTech/Bio-Microbics of Maine, Inc.	83	82
Orenco Systems, Inc. 814 Airway Avenue Sutherlin, OR 97479	Advantex Treatment System	59	70
F.R. Mahony & Associates, Inc. 273 Weymouth Street Rockland, MA 02370	Amphidrome Process	69	88
Fuji Clean USA, LLC 41-2 Greenwood Road Brunswick, Maine 04011	Fuji Clean	20	60
Lombardo Associates, Inc. 188 Church Street Newton, MA 02458	Nitrex Filter	100	100
KleanTu LLC 300 Old Pond Road, Ste# 206 Bridgeville, PA 15017	NitROE Waste-Water Treatment System	85	91
Aquapoint.3 LLC 39 Tarkiln Place New Bedford, MA 02745	Bioclere	54	62

“What about *me*?”

- Start small with Excel
- Use this presentation for Excel beginning.
- The tool I created may be most useful to those who export data from Barnstable County I/A Tracking Database because of the way it is currently written, **but you can always build your own tool if you feel inspired.**

Other Tips

- Test your tool!
 - Try and use a smaller subset
 - Try to “break” it
 - Use a different formula or manually check your math

Where Might We Differ?

- We could use different datasets
- You might want to exclude different data (or include more than I did)
- You might find different reason to use average instead of median, or even do completely different calculations

Looking Forward to Future Endeavors

- Incorporation of Statistics?
- Use of AI?
- Use of Power Bi?

Thank you!

Emily Michele Olmsted

emilymichele.olmsted@capecod.gov

(one L in Michele, no A in Olmsted)

508-375-6901

