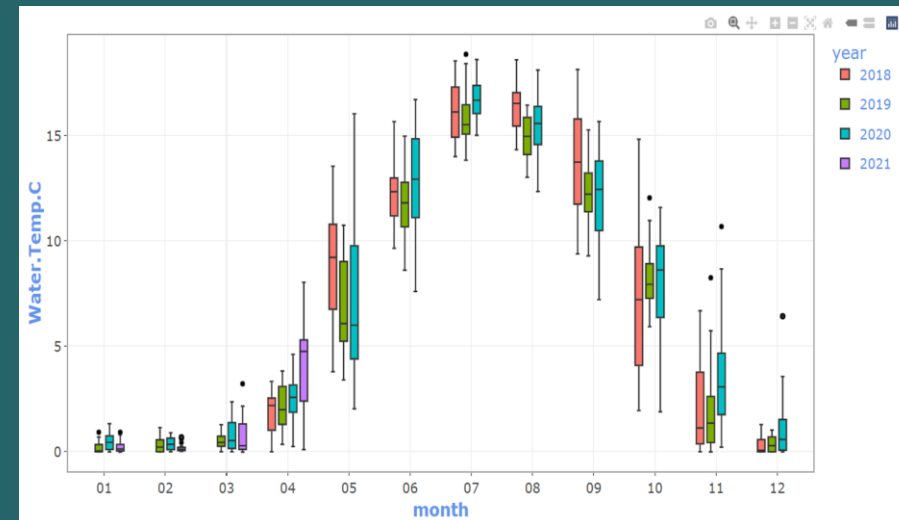
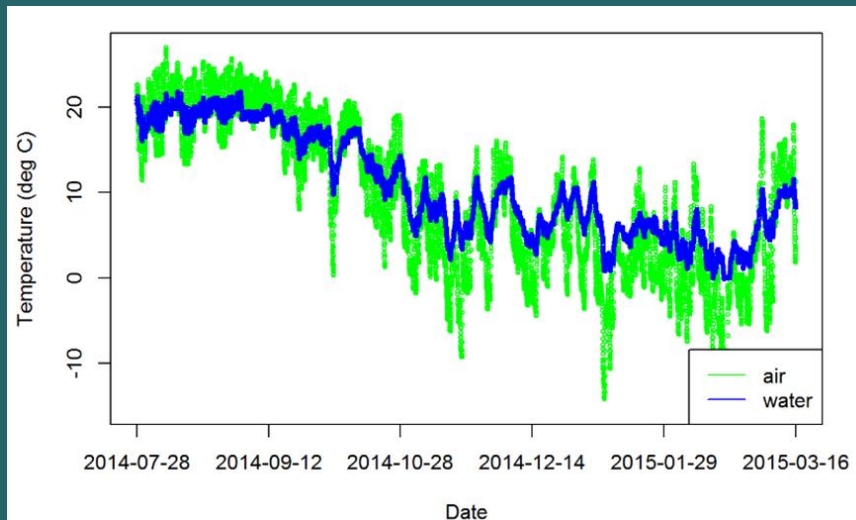


# R-based Tools to Perform Quality Control Procedures, Summarize, and Visualize Continuous Sensor Data



Jen Stamp (Tt), Erik Leppo (Tt), Michael Pennino (EPA), Britta Bierwagen (EPA), & Yadong Xu (EPA)

*The views expressed in this presentation are those of the author and they do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency or other collaborating agencies.*

*Mention of trade names or commercial products does not constitute endorsement or recommendation for use but is for descriptive purposes only.*

# Tool # 1 - ContDataQC

**Purpose:** help QC continuous sensor data

Current website (Shiny app) - <https://nalms.shinyapps.io/ContDataQC/>

The screenshot displays the EPA website header with the logo and navigation links: Environmental Topics, Laws & Regulations, Report a Violation, and About EPA. A search bar is located in the top right corner. Below the header, a navigation bar for the ContDataQC app includes links for Continuous Data QC, Summary, and Statistics; About; Data Preparation; Main Functions; Download USGS Gage Data; and Troubleshooting. A secondary navigation bar contains links for Overview (which is highlighted), Test Files, Basic Information, Tips, Advanced, and Related Apps. The main content area shows the text "Version 2.0.7.9032." followed by the heading "About ContDataQC". The text below the heading describes the tool as a free R-based application for water quality monitoring, detailing its capabilities in merging and summarizing sensor data, and downloading USGS gage data. It also mentions the tool's development by Tetra Tech in support of the EPA and its availability on GitHub.

# ContDataQC

- Developed for the Regional Monitoring Networks (RMNs) (but has applicability beyond)
- Most RMN partners did not have
  - Data management systems that could accommodate continuous data
  - Formal systems in place for QC'ing continuous sensor data



# ContDataQC

- Provides a file naming and organizational scheme
- Puts data into a standardized format that make analysis and data sharing easier
- Calculates summary statistics (which can then be uploaded to WQX)

# ContDataQC is flexible & customizable

## Parameters

- Temperature
- Water level
- Discharge
- Conductivity
- Dissolved oxygen
- pH
- Turbidity
- Chlorophyll-a
- Salinity

## Can be used on any sensor

- Formatting functions currently available for Onset HOBO sensors and PME miniDOTs

## Customization is possible

- Users can add new parameters and customize many of the requirements by editing a plain text configuration file

# ContDataQC – QC reports

QC checks	Description
Flag tests	<ul style="list-style-type: none"><li>• Unrealistic values</li><li>• Spikes</li><li>• Rate of change (RoC)</li><li>• Flat line</li></ul>
Missing observations	‘Count’ tables <ul style="list-style-type: none"><li>• Number of measurements per day</li></ul>
Time series plots	Visually check plots for errors <ul style="list-style-type: none"><li>• Individual parameters and combinations</li></ul>
	Discrete (accuracy check) measurements taken during site visits are overlaid onto the time series plot

# ContDataQC - Flag Tests

<b>Test</b>	<b>Description</b>
Unrealistic values (‘Gross range’)	Values are above or below the upper and lower thresholds
Spikes	Adjacent measurements change by more than ‘x’ amount
Rate of change (RoC)	Change exceeds a given threshold (e.g., $\geq 3$ standard deviations within 25 hrs)
Flat line	A certain number of consecutive measurements are within a certain amount of each other

# ContDataQC – thresholds

Example - water temperature defaults (based on eastern cool water streams)

Test	Flag	Default threshold
Gross I	Fail - Hi	> 30 deg C
	Fail - Lo	< 2 deg C
Spike	Suspect	$\geq 1$ deg C (+/-)
	Fail	$\geq 3$ standard deviations within 25 hours
Rate of change	Suspect	NA
	Fail	> 30 consecutive measurements within 0.01 units of one another
Flat line	Suspect	> 20 consecutive measurements within 0.01 units of one another

Flag test thresholds should be evaluated and customized if needed



# ContDataQC – aggregate QC'd data files

Choose operation to perform

QC raw data

Aggregate QC'ed data

Summary statistics

- Merge QC'd data files from different time periods
- Generate time series plots and basic summary statistics

Daily summary statistics can be uploaded to WQX

# Tool # 2 - ContDataSumViz

**Purpose:** summarize and visualize QC'd continuous sensor data

A banner image showing a natural landscape with a stream and trees. Overlaid on the image is the text "Data Visualization & Reporting" in large, blue, outlined letters, and "For Regional Monitoring Networks" in smaller, blue, outlined letters below it.

## Data Visualization & Reporting For Regional Monitoring Networks

Upload Data

Data Exploration

Create Report

Currently in user testing, then revisions and post on EPA server

# ContDataSumViz - Exploration

Upload Data

**Data Exploration**

Create Report

All parameters

Temperature

Hydrology



Summary  
statistics  
tables

Time series  
plots  
(traditional)

Time series  
plots  
(annual  
overlays)

Box  
plots

CDFs

Raster  
graphs

# ContDataSumViz – time series plots

Summary tables

Time series plots

Time series - Annual overlays

Box plots

CDFs

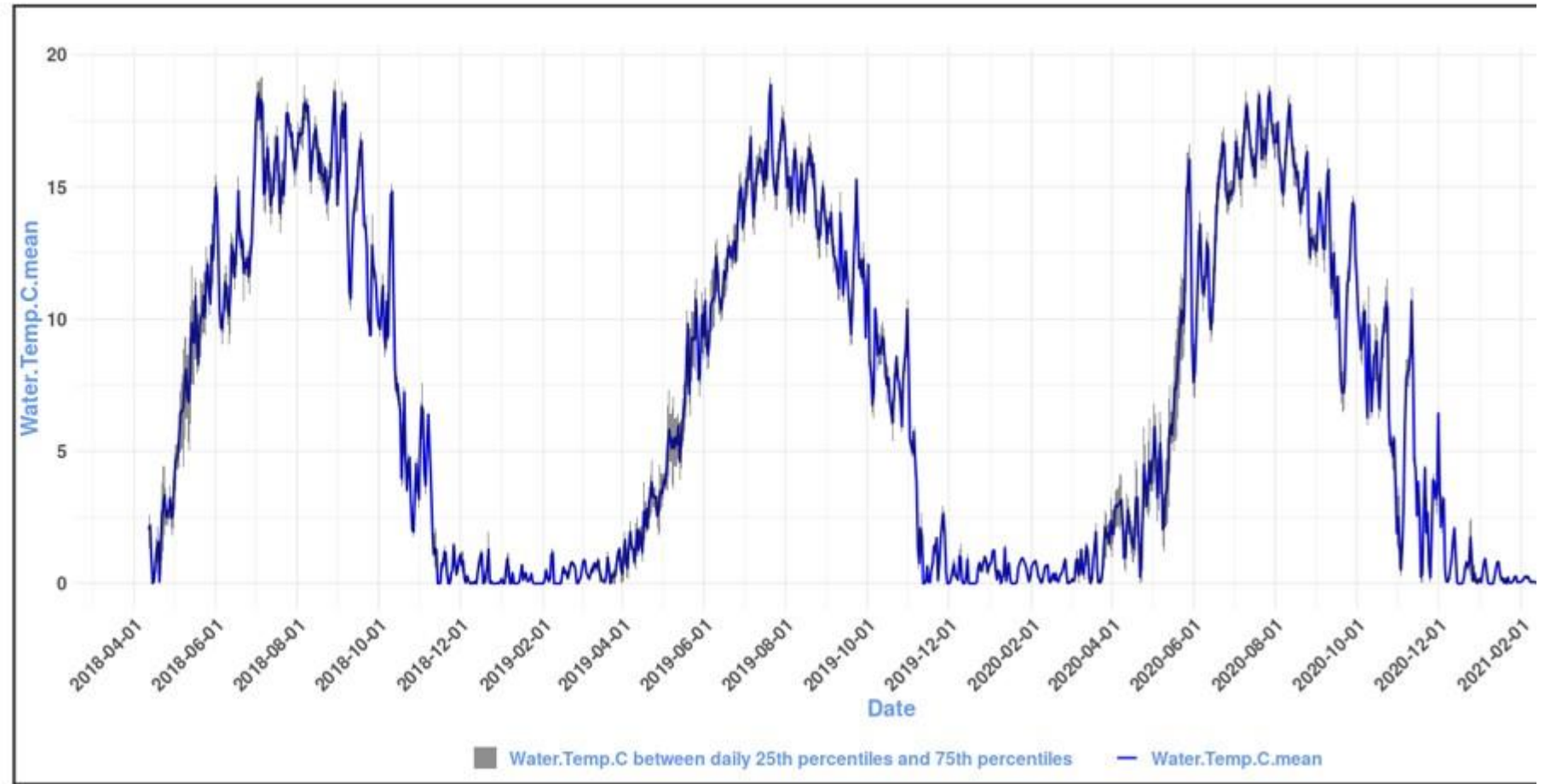
Raster graphs

Select variable name  
Water.Temp.C

Select daily statistics metrics  
mean

Add shading with  
 25th & 75th percentiles  
 minimum & maximum  
 newData

Plot title



# ContDataSumViz – time series annual overlay plots

Summary tables Time series plots **Time series - Annual overlays** Box plots CDFs Raster graphs

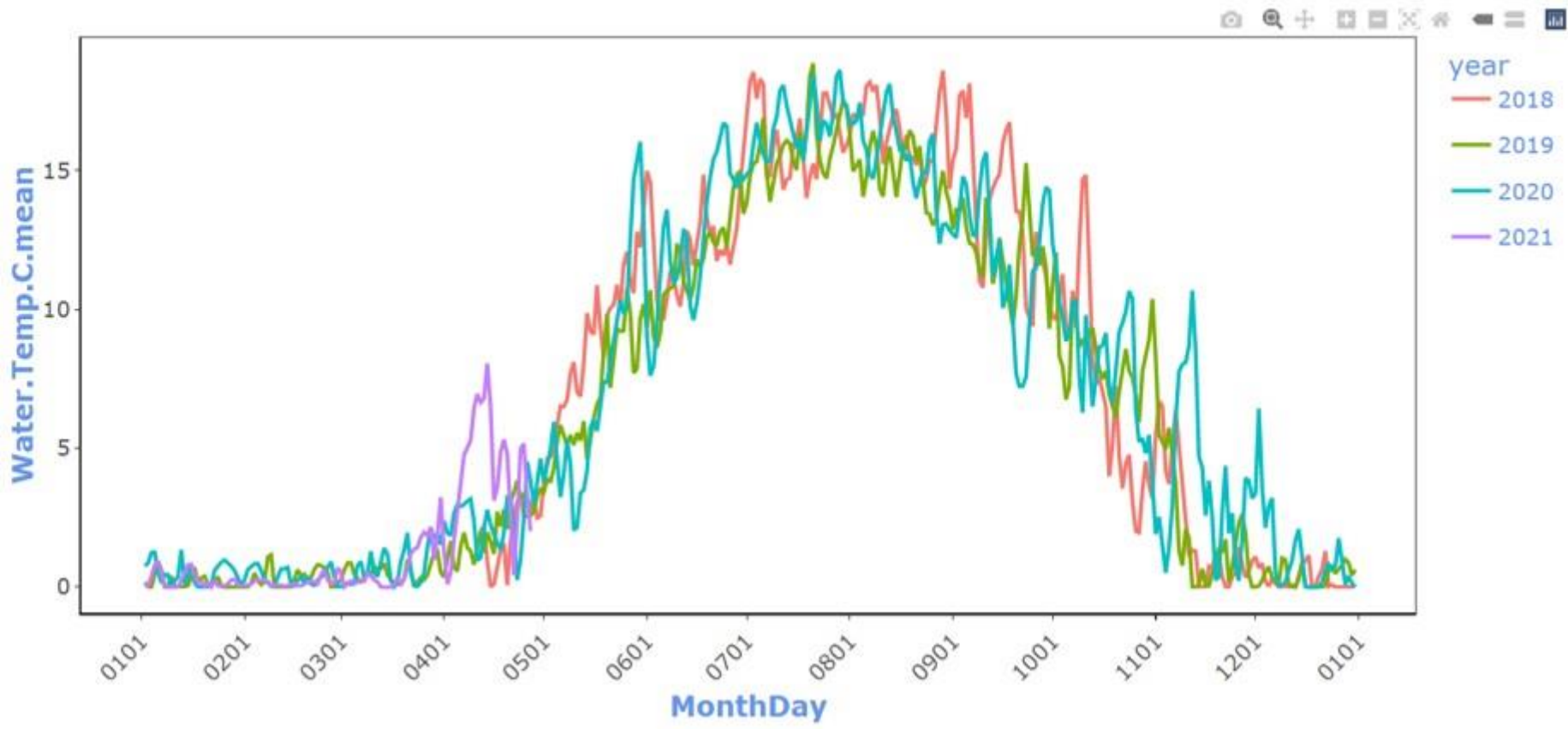
Select variable name  
Water.Temp.C

Select daily statistics metrics  
mean

Plot title

Add shading with  
 none  
 overall minimum and maximum(all years)  
 newData

**Display**



# ContDataSumViz – box plots

Summary tables

Time series plots

Time series - Annual overlays

Box plots

CDFs

Raster graphs

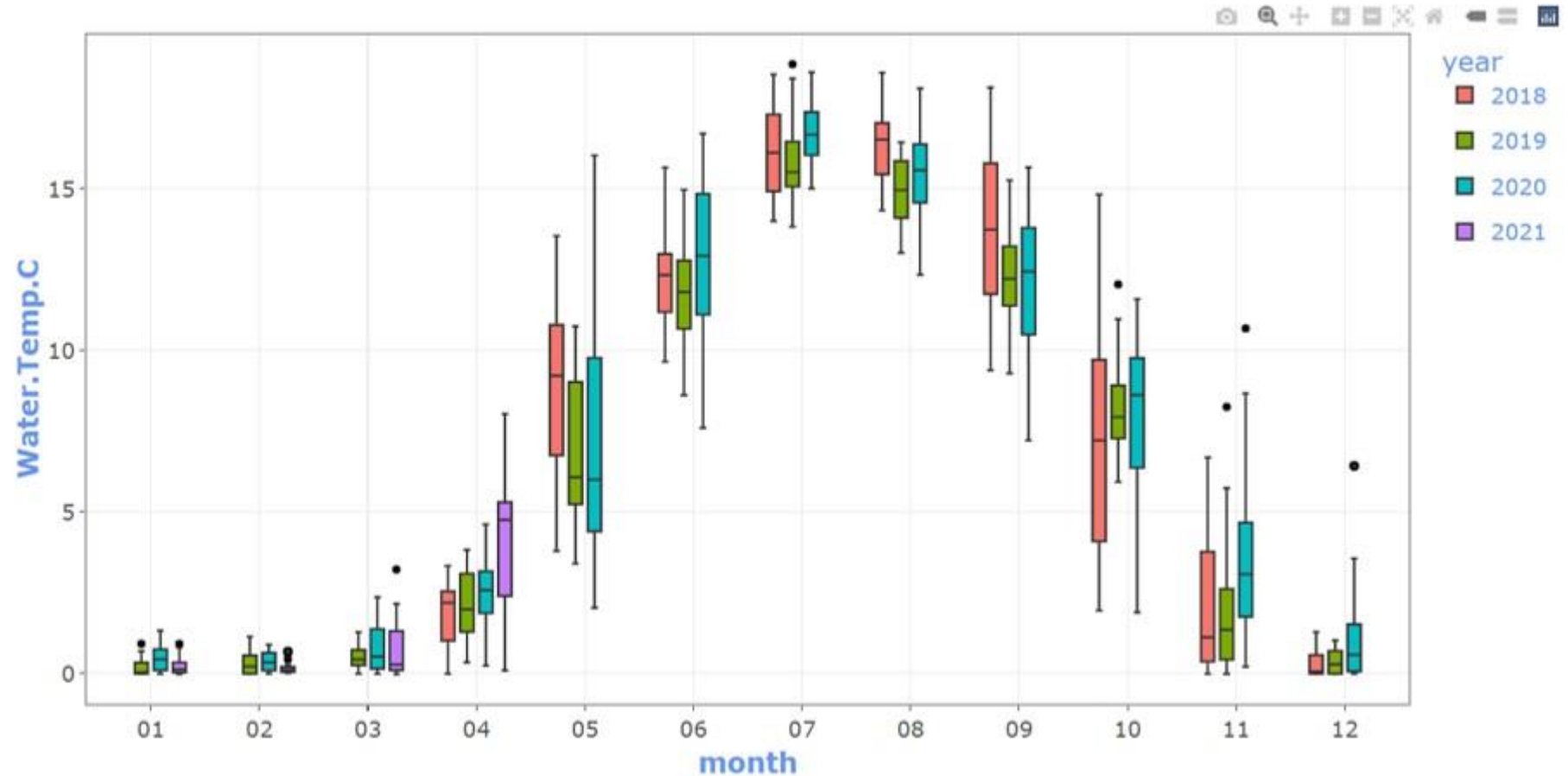
Select variable name  
Water.Temp.C

Select daily statistics metrics  
mean

Group by  
 month  
 month(years side by side)  
 year  
 season  
 season(years side by side)

Plot title

Display



# ContDataSumViz - CDFs

Summary tables

Time series plots

Time series - Annual overlays

Box plots

CDFs

Raster graphs

Select variable name

Water.Temp.C

Add shading with

- 25th & 75th percentiles
- minimum & maximum
- newData

Select year

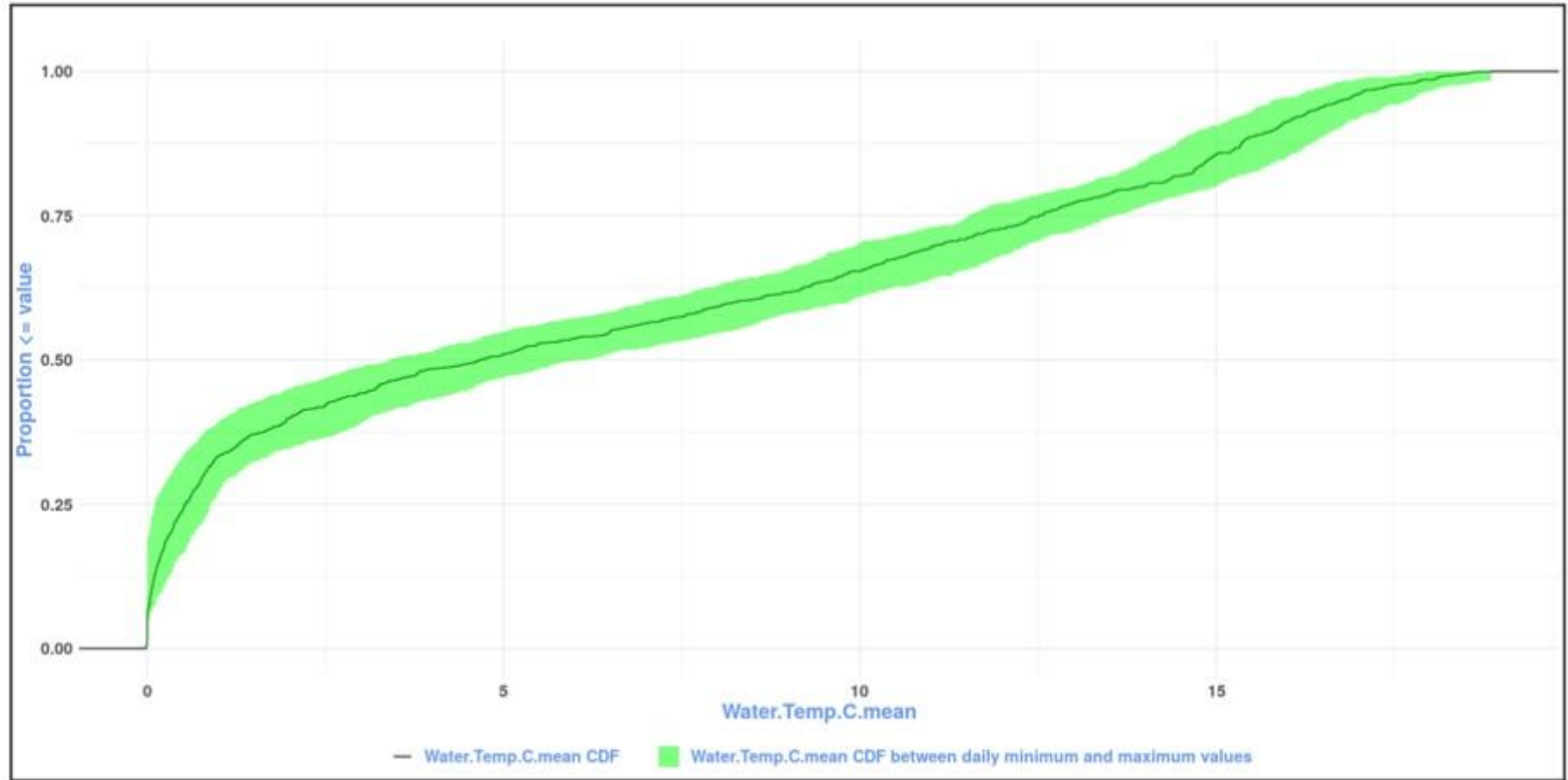
All

Select season

All

Plot title

Run and display



# ContDataSumViz – raster graphs

Summary tables

Time series plots

Time series - Annual overlays

Box plots

CDFs

Raster graphs

Select variable name

Water.Temp.C

Select daily statistics metrics

mean

Plot title

Adjust plot aspect ratio

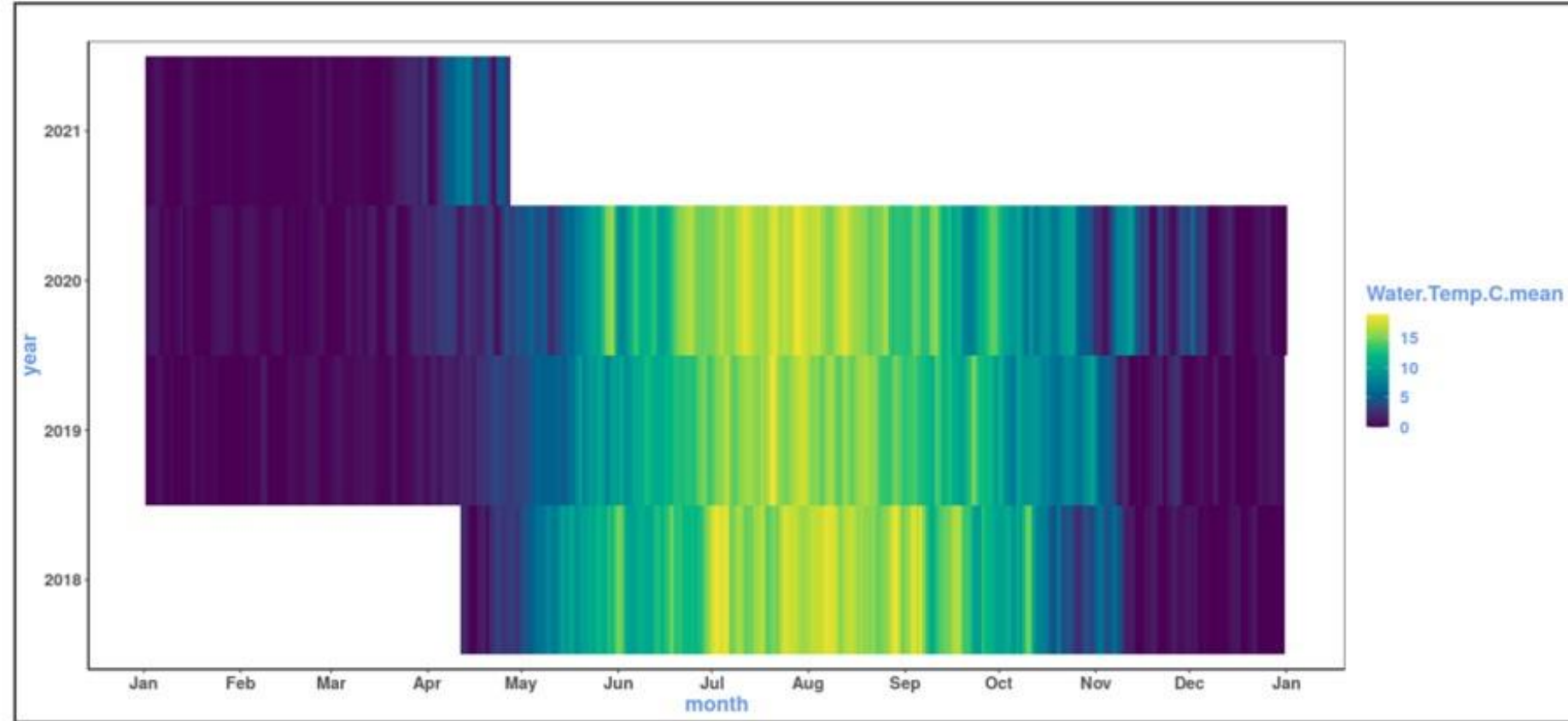
0.5

Color palette options

hcl

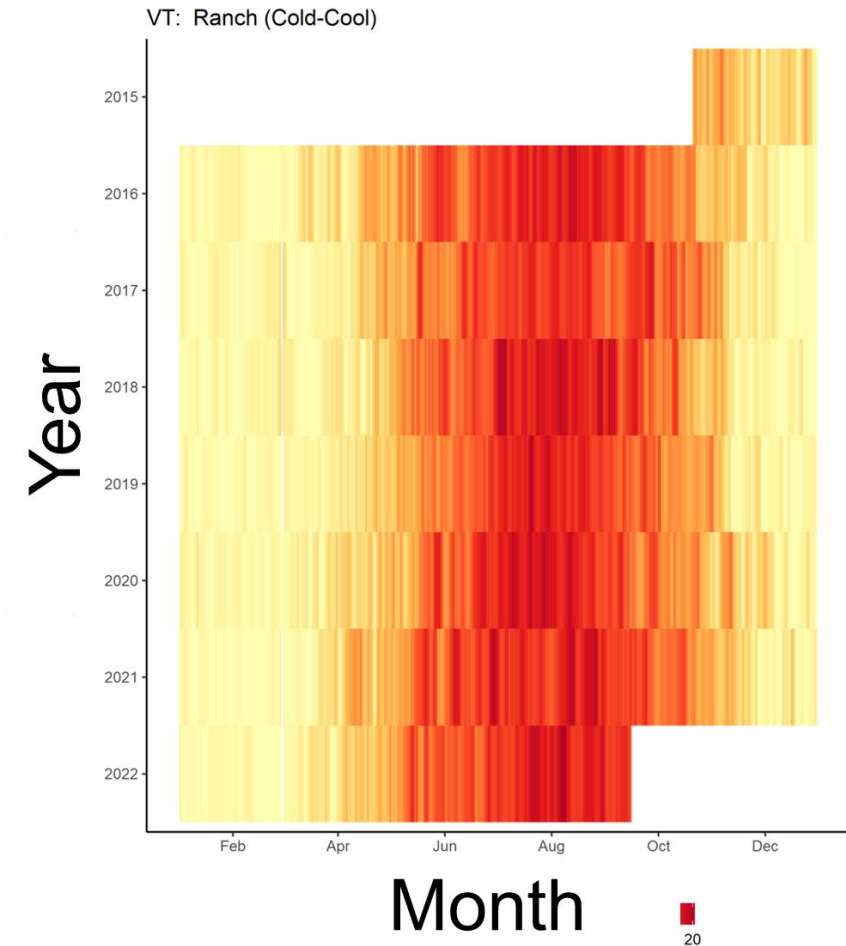
rainbow

heat

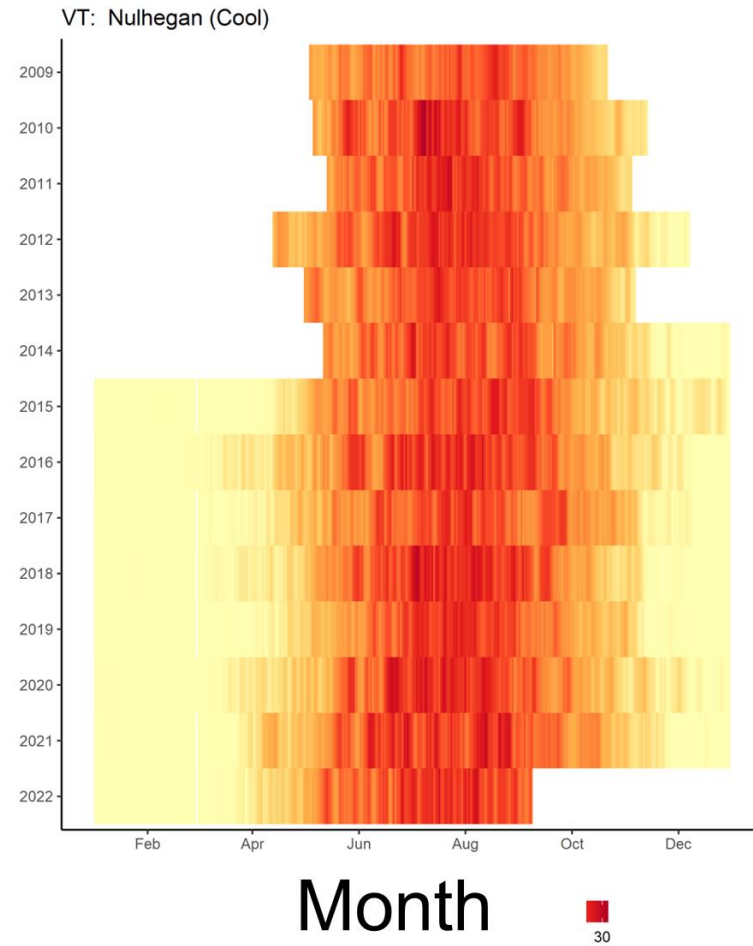




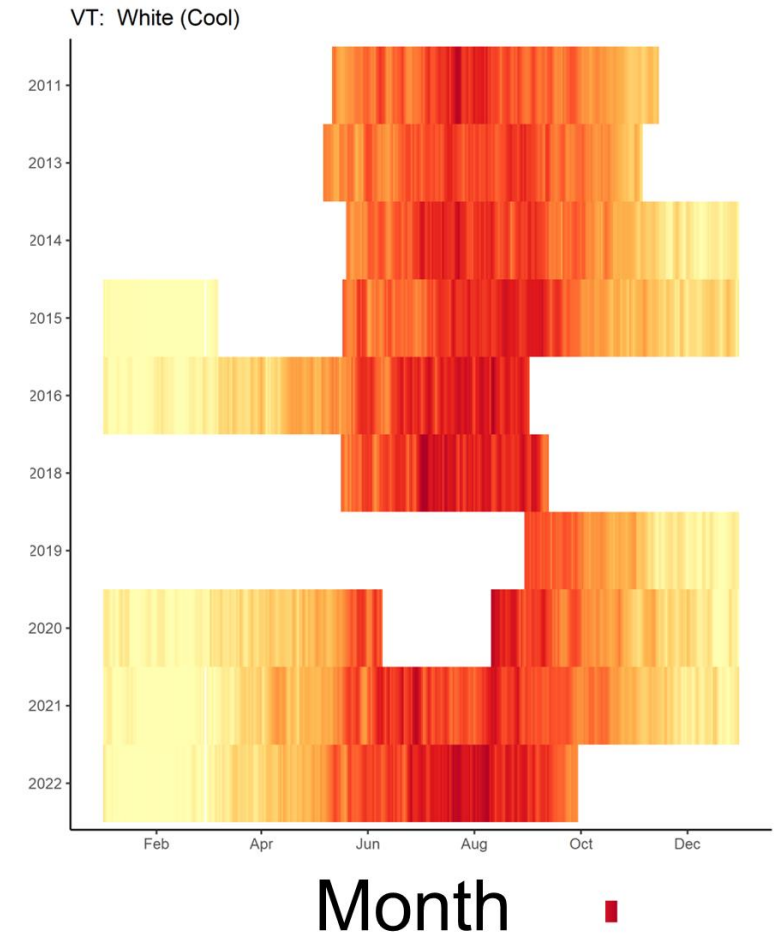
# Visualize changes in duration, timing, magnitude



- Year-round
- No gaps



- Seasonal, then year-round
- No gaps



- Seasonal, then year-round
- Gaps

# Acknowledgments

- ContDataSumViz version 2 –
  - Tom Faber (EPA Region 1), Leah Ettema & Lou Reynolds (EPA Region 3) and other workgroup members
- RMN partners –
  - Providing data, helping with testing and development
  - Special thanks to Shane & Kayla Bowe (Red Lake), Tim Martin (MN DNR), David Smith (KY Division of Water), Chris McArthur (EPA R4) and more...