# Using R to analyze, understand, and communicate tidal wetland data

Kim Cressman, Grand Bay National Estuarine Research Reserve Kimberly.Cressman@dmr.ms.gov

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

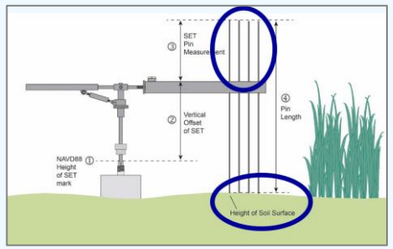
Suzanne Shull, Padilla Bay NERR; Margo Posten, Grand Bay NERR; Kristin Evans, Mission Aransas NERR; Jenni Schmitt, South Slough NERR; Kari St. Laurent, Delaware NERR; Megan Tyrrell, Waquoit Bay NERR; Brook Russell, Clemson University

more info

back to cover slide

**The Problem**: same type of data; many different formats

**The Data:**
Surface Elevation Tables, aka **SETs**




**Question**: is the marsh surface keeping up with sea level change?

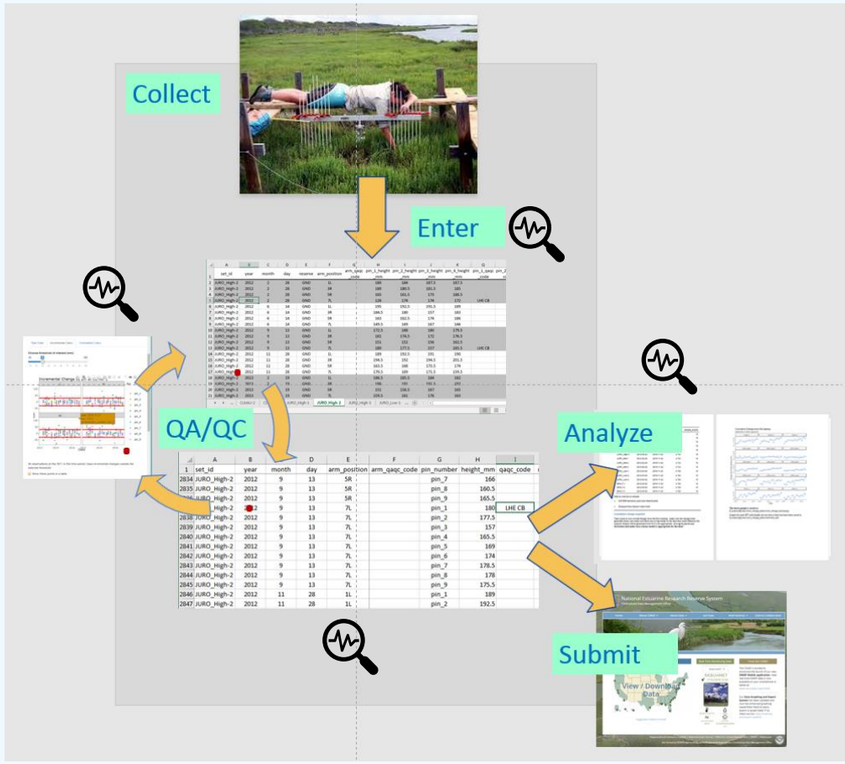
Pin heights (36 per date, top blue circle) are a proxy for shape and height of the marsh surface (bottom blue circle).

**Over time, we can use these data to see how the marsh surface is changing.**
Marshes provide valuable services, such as flood protection and nursery habitat for seafood species. We want to see them keeping up with sea level rise!

**See and download other NERR data:**
http://nerrsdata.org

**This Project:**

- tidied SET data from 15 National Estuarine Research Reserves (NERRs)
- generated a ***workflow*** that can be used to *keep* things tidy***:***




**Think about the users!**
This project's end users don't have much, if any, programming experience.

**file setup**: self-contained (.Rproj); descriptive folder and file names

**inputs**: Excel files

**outputs**: interactive (shiny, leaflet) and static (Word, from rmarkdown)

- help with QA/QC
- perform basic analyses
- generate documents that are useful communication tools

**instructions:** everywhere

**installation help**: scripts to install and check for required packages

*This work is sponsored by the National Estuarine Research Reserve System Science Collaborative, which supports collaborative research that addresses coastal management problems important to the reserves. The Science Collaborative is funded by the National Oceanic and Atmospheric Administration and managed by the University of Michigan Water Center (NAI4NOS4190145).* More info: http://nerrssciencecollaborative.org/

more info

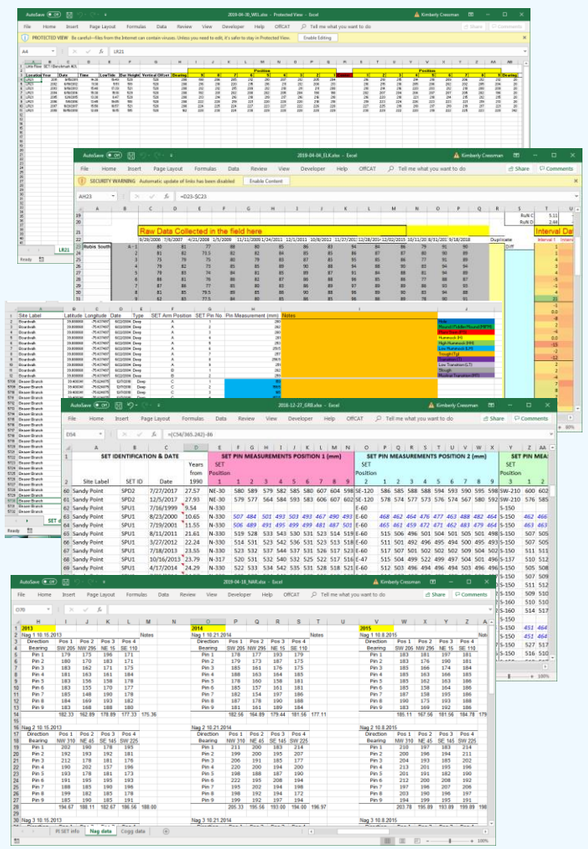
# Using R to analyze, understand, and communicate tidal wetland data

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

back to cover slide

## Dealing with Data

individual scripts to get each unique data format into new, standardized “raw” format



## Human-friendly data entry

- need to minimize data entry error
- with 36 measurements per date, this was easiest to do by dividing into 4 rows per date – the rows correspond to positions of the measurement device

## Computer-friendly tidy file

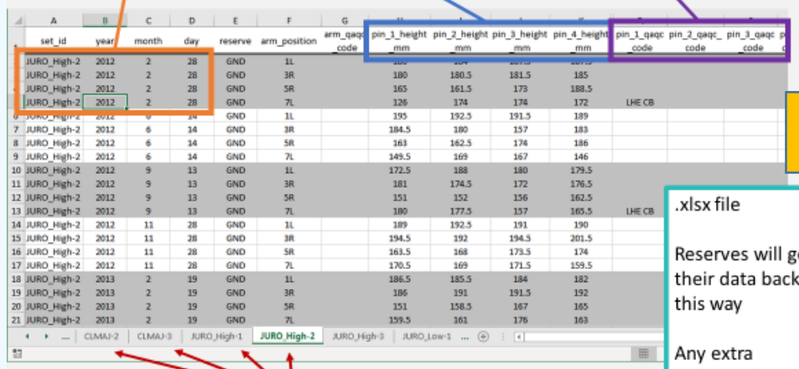
used to generate all other outputs in the workflow

### Output: “Raw” Data Format for future data entry

one row per arm, per date = 4 rows per SET, per date

9 columns for readings (pins 1-9; can be mm or cm)

columns for QC flags/codes off to the right to ease data entry



| set_id | year | month | day | reserve | arm_position | arm_qaqc_code | pin_1_height_mm | pin_2_height_mm | pin_3_height_mm | pin_4_height_mm | pin_1_qaqc_code | pin_2_qaqc_code | pin_3_qaqc_code |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| JURO_High-2 | 2012 | 2 | 28 | GND | 1L | | 195 | 192.5 | 191.5 | 189 | | | |
| JURO_High-2 | 2012 | 2 | 28 | GND | 3R | | 180 | 180.5 | 181.5 | 185 | | | |
| JURO_High-2 | 2012 | 2 | 28 | GND | 5R | | 165 | 161.5 | 173 | 188.5 | | | |
| JURO_High-2 | 2012 | 2 | 28 | GND | 7L | | 126 | 174 | 174 | 172 | LHE CB | | |
| JURO_High-2 | 2012 | 6 | 14 | GND | 1L | | 195 | 192.5 | 191.5 | 189 | | | |
| JURO_High-2 | 2012 | 6 | 14 | GND | 3R | | 184.5 | 180 | 157 | 183 | | | |
| JURO_High-2 | 2012 | 6 | 14 | GND | 5R | | 163 | 162.5 | 174 | 186 | | | |
| JURO_High-2 | 2012 | 6 | 14 | GND | 7L | | 149.5 | 169 | 167 | 146 | | | |
| JURO_High-2 | 2012 | 9 | 13 | GND | 1L | | 172.5 | 188 | 180 | 179.5 | | | |
| JURO_High-2 | 2012 | 9 | 13 | GND | 3R | | 181 | 174.5 | 172 | 176.5 | | | |
| JURO_High-2 | 2012 | 9 | 13 | GND | 5R | | 151 | 152 | 156 | 162.5 | | | |
| JURO_High-2 | 2012 | 9 | 13 | GND | 7L | | 180 | 177.5 | 157 | 165.5 | LHE CB | | |
| JURO_High-2 | 2012 | 11 | 28 | GND | 1L | | 189 | 192.5 | 191 | 190 | | | |
| JURO_High-2 | 2012 | 11 | 28 | GND | 3R | | 194.5 | 192 | 194.5 | 201.5 | | | |
| JURO_High-2 | 2012 | 11 | 28 | GND | 5R | | 163.5 | 168 | 173.5 | 174 | | | |
| JURO_High-2 | 2012 | 11 | 28 | GND | 7L | | 170.5 | 169 | 171.5 | 159.5 | | | |
| JURO_High-2 | 2013 | 2 | 19 | GND | 1L | | 186.5 | 185.5 | 184 | 182 | | | |
| JURO_High-2 | 2013 | 2 | 19 | GND | 3R | | 186 | 191 | 191.5 | 192 | | | |
| JURO_High-2 | 2013 | 2 | 19 | GND | 5R | | 151 | 158.5 | 167 | 165 | | | |
| JURO_High-2 | 2013 | 2 | 19 | GND | 7L | | 159.5 | 161 | 176 | 163 | | | |

CLMAJ-2 | CLMAJ-3 | JURO_High-1 | JURO_High-2 | JURO_High-3 | JURO_Low-1

Each SET on its own worksheet in the same file

.xlsx file

Reserves will get their data back this way

Any extra columns (e.g. Observer) will be preserved

### “Processed” Data: generated from raw file by R script

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | set_id | year | month | day | arm_position | arm_qaqc_code | pin_number | height_mm | qaqc_code | reserve |
| 2834 | JURO_High-2 | 2012 | 9 | 13 | 5R | | pin_7 | 166 | | GND |
| 2835 | JURO_High-2 | 2012 | 9 | 13 | 5R | | pin_8 | 160.5 | | GND |
| 2836 | JURO_High-2 | 2012 | 9 | 13 | 5R | | pin_9 | 165.5 | | GND |
| 2837 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_1 | 180 | LHE CB | GND |
| 2838 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_2 | 177.5 | | GND |
| 2839 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_3 | 157 | | GND |
| 2840 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_4 | | | GND |
| 2841 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_5 | | | GND |
| 2842 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_6 | | | GND |
| 2843 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_7 | | | GND |
| 2844 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_8 | | | GND |
| 2845 | JURO_High-2 | 2012 | 9 | 13 | 7L | | pin_9 | | | GND |
| 2846 | JURO_High-2 | 2012 | 11 | 28 | 1L | | pin_1 | | | GND |
| 2847 | JURO_High-2 | 2012 | 11 | 28 | 1L | | pin_2 | | | GND |

....

all SETs, all readings in one file

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 10252 | SPALT-3 | 2016 | 11 | 23 | 5R | | pin_9 | | | GND |
| 10253 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_1 | | | GND |
| 10254 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_2 | | | GND |
| 10255 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_3 | | | GND |
| 10256 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_4 | | | GND |
| 10257 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_5 | | | GND |
| 10258 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_6 | | | GND |
| 10259 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_7 | 75 | | GND |
| 10260 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_8 | 76.5 | | GND |
| 10261 | SPALT-3 | 2016 | 11 | 23 | 7L | | pin_9 | 79 | | GND |
| 10262 | | | | | | | | | | |

gndset_QC

.csv file - should not be modified by hand

“long” format: one row per pin = 36 rows per SET per date

this file gets used in all subsequent scripts and analyses

This format, even for non-reserve data, would work well with the scripts from this project

more info

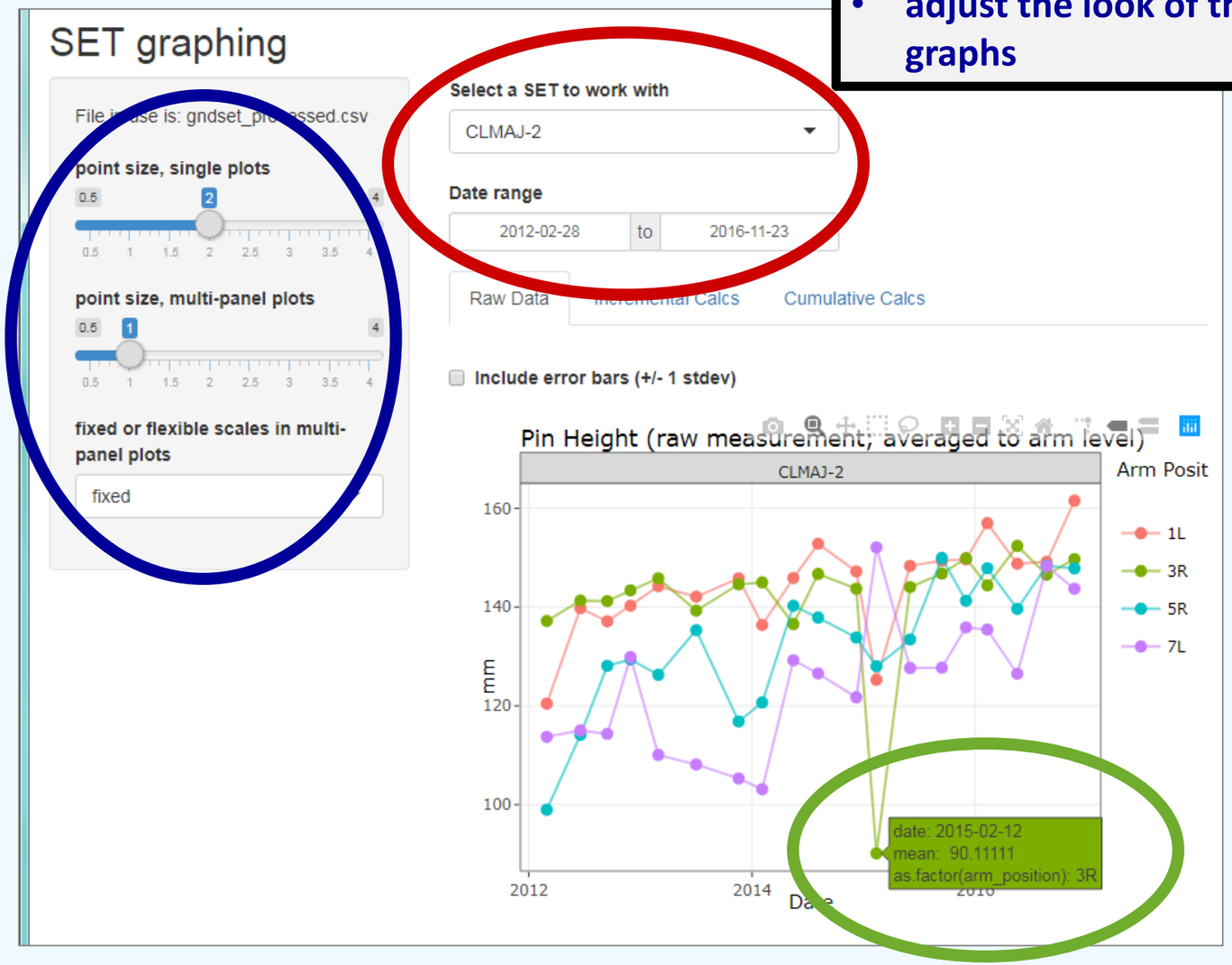
# Using R to analyze, understand, and communicate tidal wetland data

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

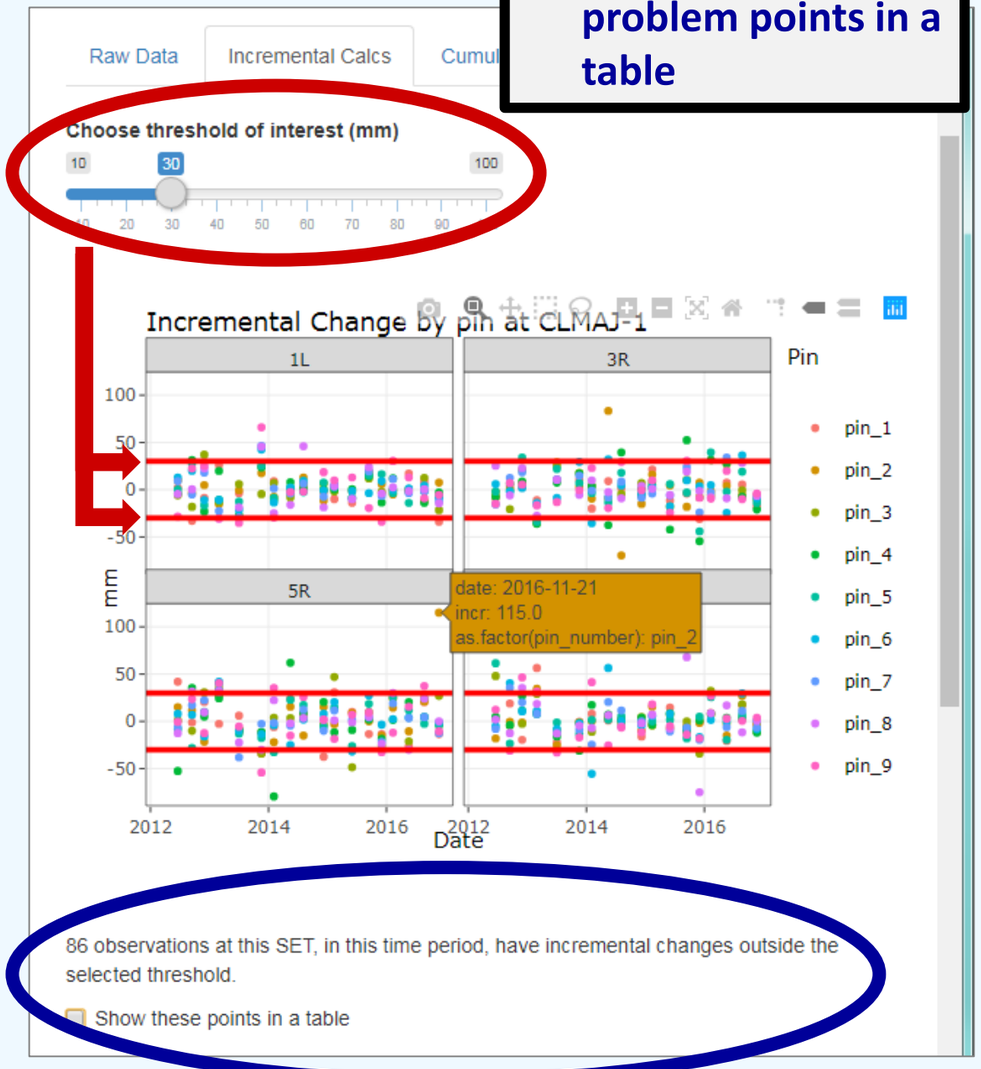
back to cover slide

## shiny app for interactive QA/QC

- Choose which site to work with
- look for outliers; hover over a point for its details
- adjust the look of the graphs




- Check for large differences between readings
- count and show problem points in a table



more info

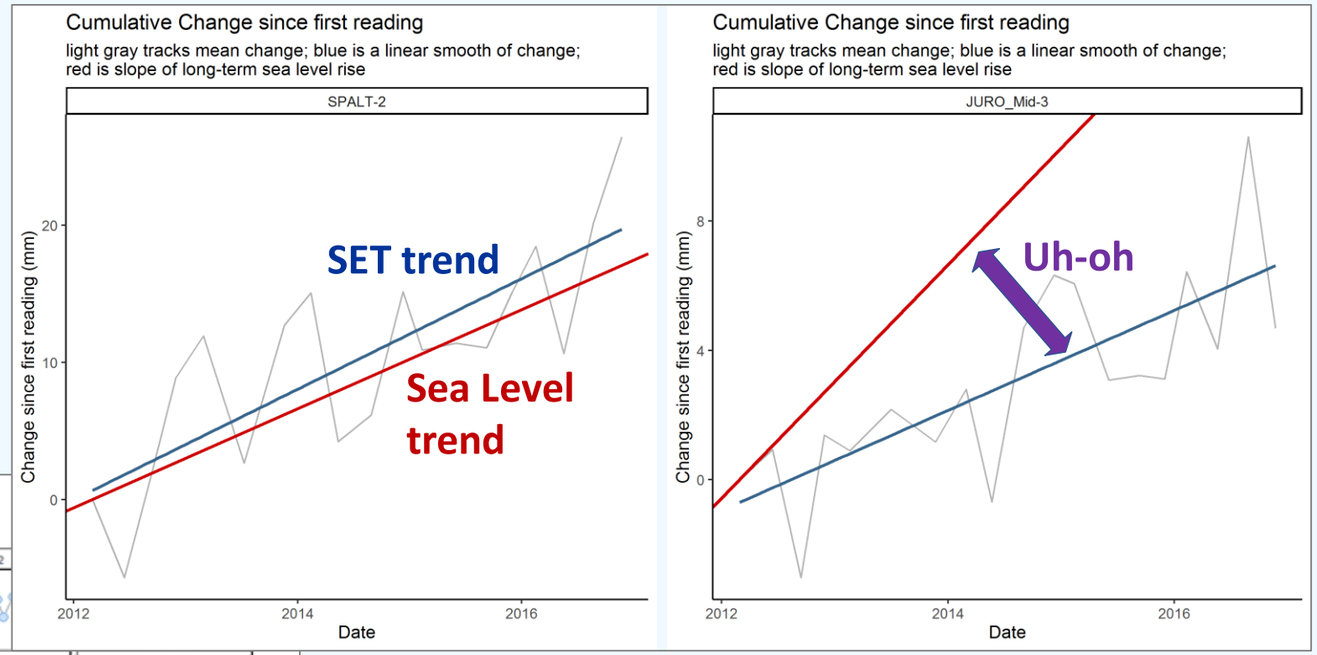
# Using R to analyze, understand, and communicate tidal wetland data

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

back to cover slide

## Analyze and Visualize

- generated reports are in Word because users know how to edit .docx files – can remove figures, modify descriptions, and make other changes as needed
- figures in report also saved as .png files, with file location indicated in document




**sampling information**

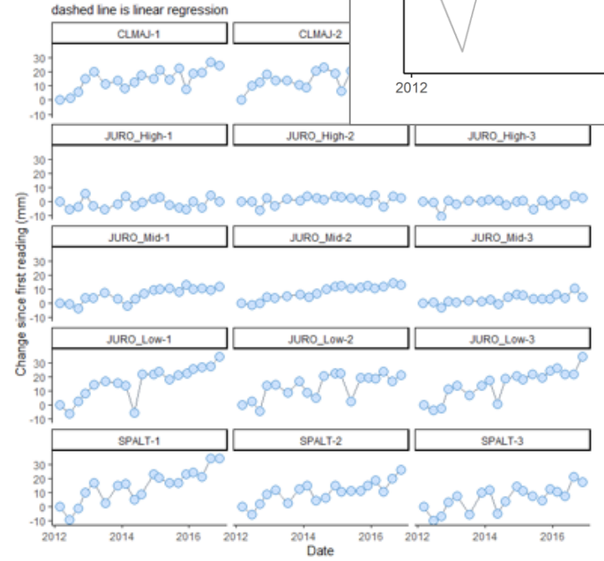
| set_id | first_sampled | last_sampled | years_sampled | sample_events |
|---|---|---|---|---|
| CLMAJ-1 | 2012-02-29 | 2016-11-21 | 4.728 | 19 |
| CLMAJ-2 | 2012-02-29 | 2016-11-21 | 4.728 | 19 |
| CLMAJ-3 | 2012-02-29 | 2016-11-21 | 4.728 | 19 |
| JURO_High-1 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_High-2 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_High-3 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_Mid-1 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_Mid-2 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_Mid-3 | 2012-02-28 | 2016-11-22 | 4.734 | 19 |
| JURO_Low-1 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |
| JURO_Low-2 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |
| JURO_Low-3 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |
| SPALT-1 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |
| SPALT-2 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |
| SPALT-3 | 2012-03-02 | 2016-11-23 | 4.728 | 19 |

Still on wish list to include:

- NAVD88 elevation (and year determined)
- Distance from closest water body

**Cumulative change snapshot**

Take a look at your overall change since the first reading - make sure the change looks generally linear, and make sure there are no big breaks in the data that could influence the outputs. Output will be generated even if it is not appropriate - **it is up to you to use discretion and make sure a linear model is appropriate for the data!**




**The above graph is saved as:**
*R_output/figures/cumu_change_plots/cumu_change_noLine.png*

Graphs for each SET individually are not shown here but have been saved in *R_output/figures/cumu_change_plots/individual_sets*

above left: Sea level change (red line) is similar to change in marsh surface at the SET (blue line). This is good – it means this marsh is not drowning (yet).

above right: Sea level is increasing much faster than the marsh surface – this area could be in particular danger from sea level rise.

left: page spread in the Word output

more info

# Using R to analyze, understand, and communicate tidal wetland data

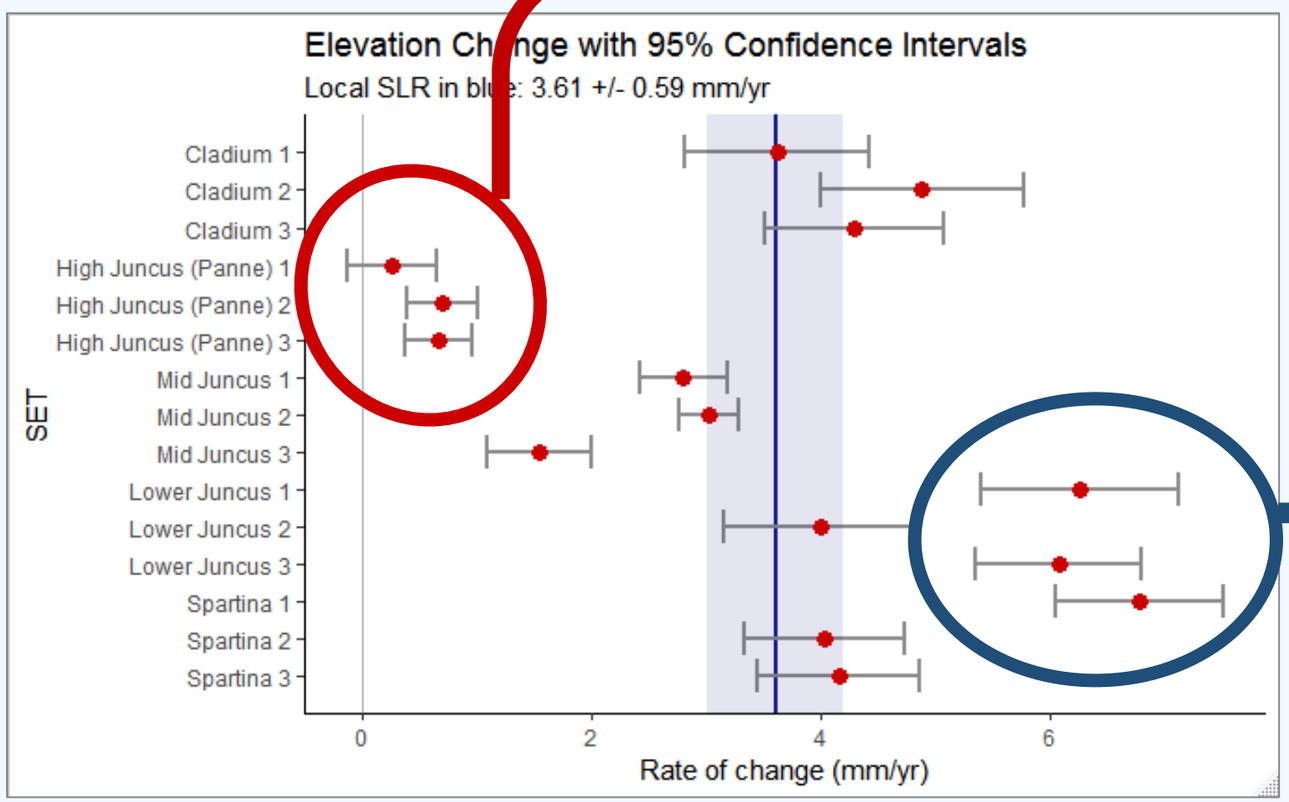
Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

back to cover slide

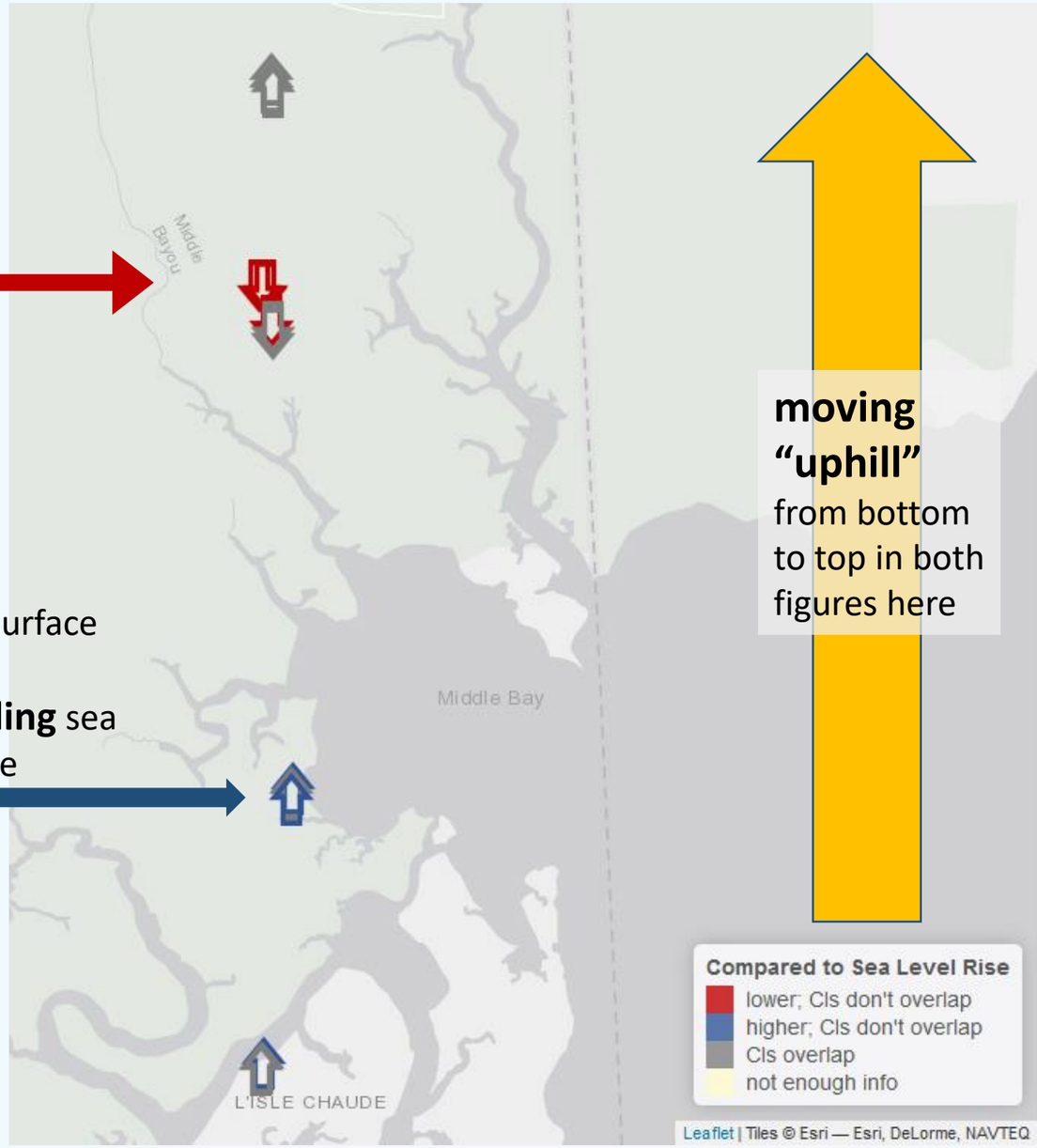
## Analyze and Visualize

summaries for the entire reserve: how does each SET's surface change compare to the rate of sea level rise?

marsh surface **not keeping up** with sea level rise




marsh surface change **exceeding** sea level rise




**moving "uphill"** from bottom to top in both figures here

more info

# Using R to analyze, understand, and communicate tidal wetland data

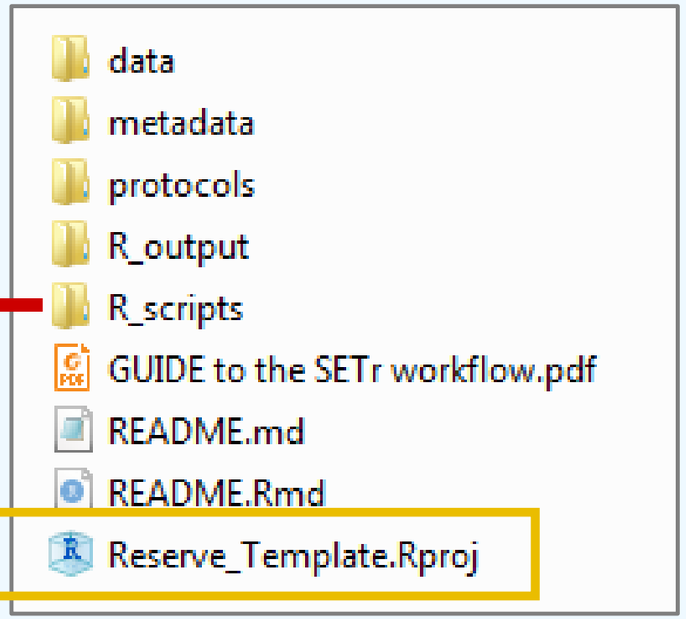
Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

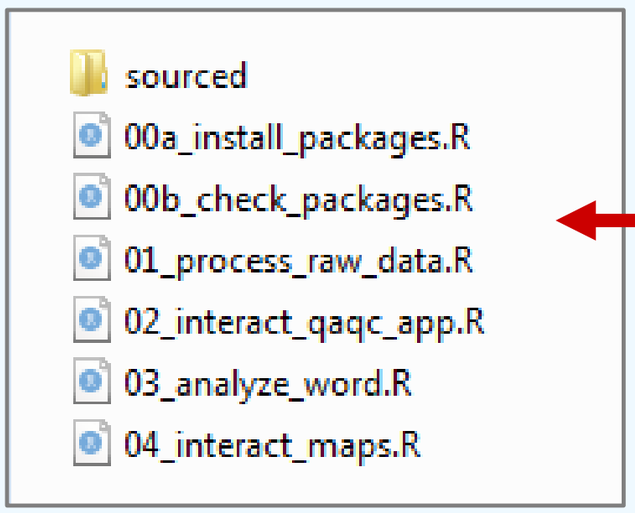
back to cover slide

## Thinking about the USERS: File setup and inputs

### Entire Directory is unified and portable

- Directory is self-contained, thanks to .Rproj and the here package
- Descriptive file names: what do I *do* with this script?
- Instructions in many places




### Excel file for users to define options

| | | |
|---|---|---|
| **If you want to analyze only a subset of dates:** | | |
| When to start, in yyyymmdd (no dashes or slashes) format | | custom_start |
| When to end, in yyyymmdd (no dashes or slashes) format | | custom_end |
| | | |
| **Do you want lat/long included in the informational tables?** (must be either TRUE or FALSE, all caps) | | |
| ...for technical reports | FALSE | coords_tech |
| ...for outreach documents | FALSE | coords_outreach |

photo credits: middle, Sandra Huynh; right, Delaware NERR

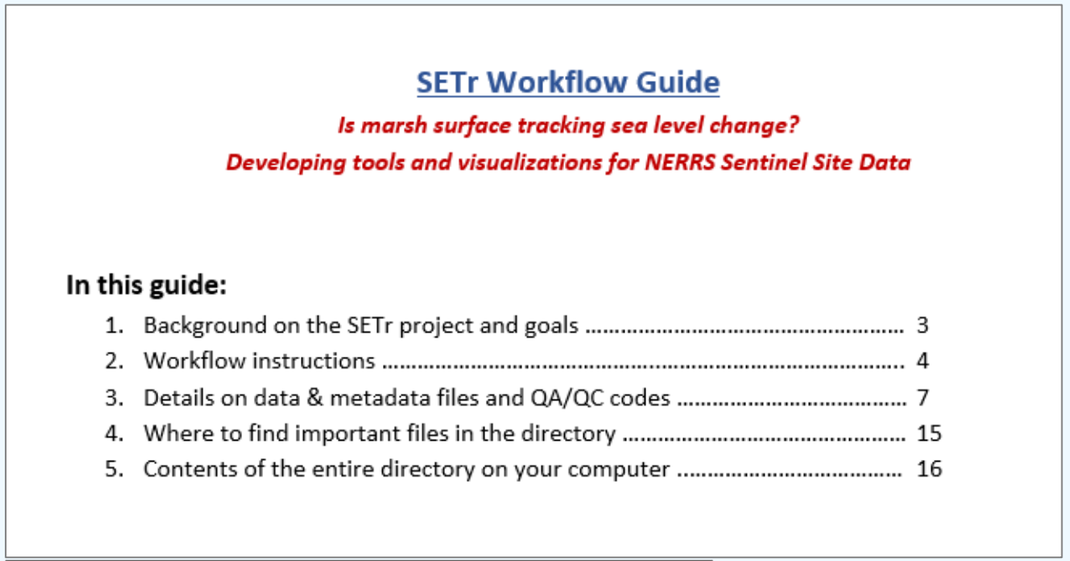
more info

# Using R to analyze, understand, and communicate tidal wetland data

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg

back to cover slide

## Thinking about the USERS: Instructions everywhere



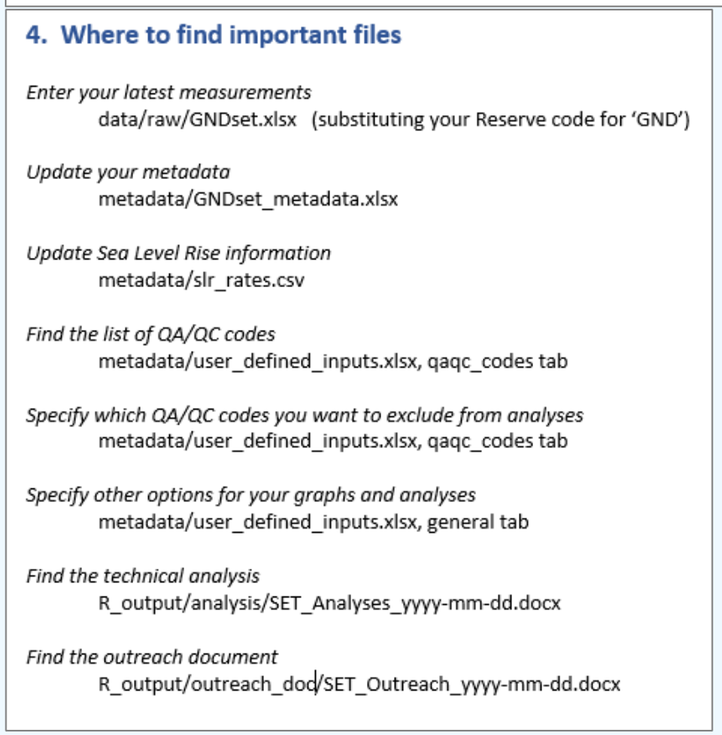
**SETr Workflow Guide**

***Is marsh surface tracking sea level change?***

***Developing tools and visualizations for NERRS Sentinel Site Data***

**In this guide:**





**4. Where to find important files**

*Enter your latest measurements*
data/raw/GNDset.xlsx (substituting your Reserve code for 'GND')

*Update your metadata*
metadata/GNDset_metadata.xlsx

*Update Sea Level Rise information*
metadata/slr_rates.csv

*Find the list of QA/QC codes*
metadata/user_defined_inputs.xlsx, qaqc_codes tab

*Specify which QA/QC codes you want to exclude from analyses*
metadata/user_defined_inputs.xlsx, qaqc_codes tab

*Specify other options for your graphs and analyses*
metadata/user_defined_inputs.xlsx, general tab

*Find the technical analysis*
R_output/analysis/SET_Analyses_yyyy-mm-dd.docx

*Find the outreach document*
R_output/outreach_doc/SET_Outreach_yyyy-mm-dd.docx

## Workflow guide in top level of directory

20 pages = balance of short-enough-to-actually-skim and thorough

## Beginning of every script tells how to use it

top: analysis script
bottom: shiny app

```
#### INSTRUCTIONS ############################################

# 1:
# Re-start your R session to make sure there's no interference:
# From the menu bar, select 'Session', then 'Restart R'
# Windows keyboard shortcut is Ctrl + Shift + F10

# 2:
# Select this entire script.
# Keyboard shortcut is Ctrl + a on windows or Cmd + a on Mac

#3:
# Run it: either using the "Run" button in the upper right corner
# or the keyboard shortcut Ctrl/Cmd + Enter

##############################################################################
```

```
#### INSTRUCTIONS ############################################

# 1:
# Re-start your R session to make sure there's no interference:
# From the menu bar, select 'Session', then 'Restart R'
# Windows keyboard shortcut is Ctrl + Shift + F10

# 2:
# In the upper right-hand corner of this (source) window,
# there is a button that says "Run App".
# Push it.
# Make sure pop-ups are enabled; the app comes up in a browser window.

##############################################################################
```

# Using R to analyze, understand, and communicate tidal wetland data

Working example: https://github.com/swmpkim/SETr_example_reserve_pkg



## Thinking about the USERS: Installation help

The goal isn't to make R programmers out of people, but to **enable them** to use tools built using R.

**Installation is a pain point that can scare people away!**

These scripts aim to minimize that pain by **not** making people pick through a long string of error messages to figure out what didn't work.

## Again: instructions everywhere!

Installation script

```
# Install packages
# THIS ONLY NEEDS TO BE RUN ONCE

################ INSTRUCTIONS ###################################
### Use Ctrl+A (Windows) / Cmd+A (Mac)   to select this entire script
### Then Ctrl+Enter / Cmd+Enter  to run it

### After this completes, run the script '00b_check_packages.R'
### to make sure everything installed properly
################################################################
```

Script that checks installations

```
################ INSTRUCTIONS ###################################
### Use Ctrl+A (Windows) / Cmd+A (Mac)   to select this entire script
### Then Ctrl+Enter / Cmd+Enter  to run it
### Then check the output in your console and see if you
### need to re-install any packages

### If you do, I recommend trying one at a time, using the command
### install.packages("package_name_here")

### Then re-run this script to make sure everything worked

### If you have any problems, contact Kim

################################################################
```

## Script to (helpfully) check package installations

Indicate success

```
All required packages are installed and loading properly!
```

Or provide helpful messages and advice if something is wrong

```
Your version of tidyr needs to be updated to at least 1.0.0;
try running install.packages("tidyr")

AND

You need to install the following packages. You can try again now by running:
install.packages(pkgs_missing)
readr
shiny
leaflet
webshot
mapview
```