

SWMP data and retrieval

Marcus W. Beck¹ Todd D. O'Brien²

¹ORISE, USEPA NHEERL Gulf Ecology Division

Email: beck.marcus@epa.gov

²NOAA/NMFS COPEPOD Project

Email: todd.obrien@noaa.gov

Objectives and agenda

- Objectives

- ▶ What are the various ways data are obtained from SWMP?
- ▶ What are some issues that need to be addressed before importing into a statistical program to conduct a time series analysis?

Objectives and agenda

- Objectives

- ▶ What are the various ways data are obtained from SWMP?
- ▶ What are some issues that need to be addressed before importing into a statistical program to conduct a time series analysis?

- Agenda

- ▶ Brief overview of SWMP network and available data
- ▶ Format and potential issues with output data
- ▶ Retrieving and importing the data

Interactive portion

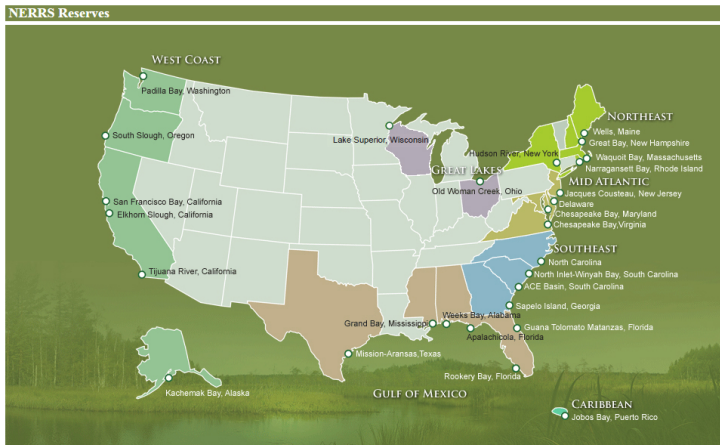
You can follow along later in this module:

- dataset1
- script1

Interactive!

Overview of SWMP and available data

SWMP - System Wide Monitoring Program, initiated in 1995 to provide continuous monitoring data at over 300 stations in 28 US estuaries



<http://nerrs.noaa.gov/ReservesMap.aspx>



Overview of SWMP and available data

CDMO is your one-stop shop for retrieving SWMP data

Home	About CDMO	About Data	Get Data	Web Services	Contact CDMO
					
View / Download Data		Real Time Monitoring Data		CDMO News	
 <p>Requested Citation Format</p>		<p>Choose Reserve... ▾</p> <p>GTMPCMET 10/08/14 09:45 AM GTMPCVQ 10/08/14 09:45 AM</p>  <p>Air Temperature: 27.8 °C (82 °F) Wind Speed: 1.1 m/Sec (02 mph) Water Temperature: 22.7 °C (73 °F) Salinity: 7.1 PPT Dissolved Oxygen: 4.7 mg/L</p>		<p>The CDMO is excited to announce the launch of our new SWMP Mobile application. Near real-time SWMP data is now available on your smartphone or tablet at: www.nerrsdata.org/mobile</p> <p>Our Data Export System has been updated and now has enhanced graphing capabilities! Want to easily export or graph data? If so, check out our Data Export System!</p>	

Overview of SWMP and available data

Data can be exported from CDMO **several** ways:



Data Export System

The DES was developed to provide the majority of users with quick and easy access to SWMP data. The DES utilizes a map-based interface and offers single station exports, yearly authenticated file downloads (these may include non-standard nutrient parameters), charting, and a current conditions display for real-time stations.

[To launch the Data Export System, click here.](#)



Advanced Query System

The AQS was developed to specifically address the data delivery needs of those end-users looking for large amounts of data exported in a format that can be easily imported and manipulated for data analysis. The AQS offers three different query options allowing for mass downloads of annual files, customized queries for specific parameters and multiple stations in the same file, and an option to merge water quality, meteorological and nutrient datasets.

[To launch the Advanced Query System, click here.](#)



Real Time Data Application

The Real Time Application allows users to view near real time data, real time gauges, and 24 hour graphs with multiple parameters. You may use a bookmarked link to directly access the station of interest, or browse and select your station. The display will update automatically with the latest information as it comes in.

[To launch the Real Time Data Application, click here.](#)



GIS Application

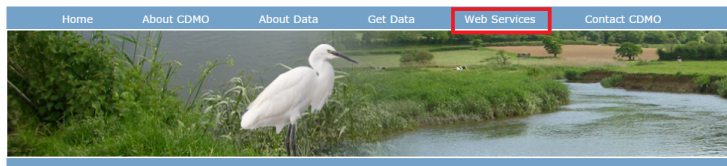
The GIS Application gives users access to Reserve boundary, watershed boundary, and high resolution reserve habitat maps. In addition, Google Earth KML files are available for the Reserve boundaries, watershed boundaries, and monitoring station locations.

[To launch the GIS Application, click here.](#)

Overview of SWMP and available data

You can also use the **SWMP** package to retrieve data

Data retrieval functions connect to the CDMO **web services**, more about this later



Web Services

Web Services

In an effort to increase distribution and use of the data collected by the NERRS, the Centralized Data Management Office has created several web service products for this purpose. These services can be used to pull real-time data from our databases for use by other individuals and organizations.

You must contact cdmowebmaster@belle.baruch.sc.edu for authorization before pulling data from the CDMO.

Transmission Time Schedule

- Realtime satellite transmissions are received hourly based upon the diagram shown below. Decoding occurs at 11 minutes past the hour and 41 minutes past the hour. This data becomes available on the CDMO website at 15 and 45 minutes past the hour. Each transmission will have 4 15-minute records in it.
- What does all this mean? Let's take Block #5 as an example. Three stations are transmitted in that block from 47:00 - 47:30 past the hour. The data from those stations will be decoded at 11 minutes past the hour and will be available on this website at 15 minutes past the hour.

Overview of SWMP and available data

A wide range of data can be requested... a few records for one site to all records for multiple sites

Requests can return a lot of data so make sure you have clear objectives

Check the [available data](#) before making a request!


- station names
- data types
- date ranges
- parameters

Overview of SWMP and available data


Available data: <http://cdmo.baruch.sc.edu/data/availableOne.cfm>

[Home](#) [About CDMO](#) [About Data](#) [Get Data](#) [Web Services](#) [Contact CDMO](#)

[Available Data](#)
[CDMO Manual \(PDF\)](#)
[Data Policy](#)
[Data Citation](#)
[Data QA/QC](#)
[Parameters](#)
[Metadata](#)




View / Download Data



[Requested Citation Format](#)

Real Time Monitoring Data

Choose Reserve...
GRBGLMET 10/08/14 10:45 AM
GRBGBWQ 09/05/14 09:45 AM



Air Temperature: 20.6 °C (69 °F)
Wind Speed: 3.1 m/Sec (07 mph)
Water Temperature: 22.5 °C (73 °F)
Salinity: 0.1 PPT
Dissolved Oxygen: 9.1 mg/L

CDMO News

The CDMO is excited to announce the launch of our new **SWMP Mobile application**. Near real-time SWMP data is now available on your smartphone or tablet at: www.nerrsdata.org/mobile

Our **Data Export System** has been updated and now has enhanced graphing capabilities! Want to easily export or graph data? If so, check out our [Data Export System!](#)

Overview of SWMP and available data

How to view available data:

- Trial-and-error (not recommended)
- View online: <http://cdmo.baruch.sc.edu/data/availableOne.cfm>
- View after request: 'sampling stations.csv'
- View after request: year and station specific .doc file
- Retrieve from within R (will cover later)

Overview of SWMP and available data

How to view available data:

- Trial-and-error (not recommended)
- View online: <http://cdmo.baruch.sc.edu/data/availableOne.cfm>
- View after request: 'sampling stations.csv'
- View after request: year and station specific .doc file
- Retrieve from within R (will cover later)

Now that you have the data, what do they look like?

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), **station**, and **parameter type** are identified by a 7 or 8 character name

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), **station**, and **parameter type** are identified by a 7 or 8 character name

E.g., elkcwmet

Format and potential issues with output data

To orient yourself, understand the NERRS/SWMP naming convention

Site (reserve), **station**, and **parameter type** are identified by a 7 or 8 character name

E.g., elkcwmet

- elk: site, Elkhorn Slough
- cw: station, Caspian Weather Station
- met: parameter type (weather)

Format and potential issues with output data

The fundamental unit of data is the 'station' defined by a parameter type

The parameters for a station are specific to the parameter type

Format and potential issues with output data

The fundamental unit of data is the 'station' defined by a parameter type

The parameters for a station are specific to the parameter type

Nutrients

po4f, chla_n, no3f,
no2f, nh4f, no23f,
ke_n, urea

Water quality

temp, spcond, sal,
do_pct, do_mgl,
depth, cdepth, level,
clevel, ph, turb,
chlfluor

Meteorology

atemp, rh, bp, wspd,
maxwspd, wdir,
sdwdir, totpar,
totprcp, cumprcp,
totsorad

Format and potential issues with output data

Each parameter will also have a QAQC column, with the prefix 'f_'

E.g., 'atemp' and 'f_atemp'

Values in these columns describe whether the data passed automated QAQC checks

- 5 Outside high sensor range
- 4 Outside low sensor range
- 3 Data rejected due to QAQC
- 2 Missing data
- 1 Optional SWMP supported parameter
- 0 Passed initial QAQC checks
- 1 Suspect data
- 2 Open - reserved for later flag
- 3 Calculated data: non-vented depth/level sensor correction for changes in barometric pressure
- 4 Historical data: Pre-auto QAQC
- 5 Corrected data

Format and potential issues with output data

You will have to decide how to handle QAQC...

	A	B	C	D	E	F	G	H	I	J	K	L
1	StationCo	isSWMP	DateTimeStamp	Historical	Provision	CollMethc	REP	F_Record	PO4F	F_PO4F	NH4F	F_NH4F
2	apacpnut	P	1/10/2012 10:20	0	1	1	1		0.003	<-4> [SBL]	0.03	<0>
3	apacpnut	P	2/7/2012 11:41	0	1	1	1		0.005	<0>	0.019	<0>
4	apacpnut	P	3/5/2012 11:51	0	1	1	1		0.003	<-4> [SBL]	0.041	<0>
5	apacpnut	P	4/4/2012 10:30	0	1	1	1		0.003	<-4> [SBL]	0.043	<0>
6	apacpnut	P	5/9/2012 10:12	0	1	1	1		0.003	<0>	0.053	<0>
7	apacpnut	P	5/9/2012 10:15	0	1	1	2		0.003	<-4> [SBL]	0.022	<0>
8	apacpnut	P	5/9/2012 10:20	0	1	1	3		0.003	<0>	0.016	<0>
9	apacpnut	P	6/5/2012 8:30	0	1	1	1		0.003	<-4> [SBL]	0.04	<0>
10	apacpnut	P	7/3/2012 9:58	0	1	1	1 {CSM}		0.004	<0>	0.094	<0>
11	apacpnut	P	7/3/2012 9:59	0	1	1	2 {CSM}		0.004	<0>	0.066	<0>
12	apacpnut	P	7/3/2012 10:01	0	1	1	3 {CSM}		0.005	<0>	0.069	<0>
13	apacpnut	P	8/7/2012 9:53	0	1	1	1 {CSM}		0.003	<-4> [SBL]	0.05	<0>
14	apacpnut	P	9/5/2012 10:56	0	1	1	1		0.003	<-4> [SBL]	0.026	<0>
15	apacpnut	P	10/2/2012 9:22	0	1	1	1		0.003	<-4> [SBL]	0.042	<0>
16	apacpnut	P	10/2/2012 9:27	0	1	1	2		0.003	<-4> [SBL]	0.024	<0>
17	apacpnut	P	10/2/2012 9:32	0	1	1	3		0.003	<0>	0.042	<0>
18	apacpnut	P	11/6/2012 10:30	0	1	1	1		0.003	<-4> [SBL]	0.07	<0>
19	apacpnut	P	11/26/2012 11:39	0	1	1	1		0.003	<-4> [SBL]	0.041	<0>

CDMO has limited export options for dealing with bad QAQC flags

We will learn how to handle QAQC flags in R

Format and potential issues with output data

The final piece of of the puzzle is the DateTimeStamp

Format is month/day/year hours:minutes, based on UTC offset and no daylight savings!

Varies by station and parameter type

	A	B	C
1	StationCo	isSWMP	DateTimeStamp
2	apacpnut	P	1/10/2012 10:20
3	apacpnut	P	2/7/2012 11:41
4	apacpnut	P	3/5/2012 11:51
5	apacpnut	P	4/4/2012 10:30
6	apacpnut	P	5/9/2012 10:12
7	apacpnut	P	5/9/2012 10:15
8	apacpnut	P	5/9/2012 10:20
9	apacpnut	P	6/5/2012 8:30
10	apacpnut	P	7/3/2012 9:58

	A	B	C
1	StationCo	isSWMP	DateTimeStamp
2	apacpwq	P	1/1/2012 0:00
3	apacpwq	P	1/1/2012 0:15
4	apacpwq	P	1/1/2012 0:30
5	apacpwq	P	1/1/2012 0:45
6	apacpwq	P	1/1/2012 1:00
7	apacpwq	P	1/1/2012 1:15
8	apacpwq	P	1/1/2012 1:30
9	apacpwq	P	1/1/2012 1:45
10	apacpwq	P	1/1/2012 2:00

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)
- Dealing with QAQC columns and removing 'bad' observations

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)
- Dealing with QAQC columns and removing 'bad' observations
- Data we don't want... extra columns or irrelevant parameters

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)
- Dealing with QAQC columns and removing 'bad' observations
- Data we don't want... extra columns or irrelevant parameters
- Combining data for comparison

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)
- Dealing with QAQC columns and removing 'bad' observations
- Data we don't want... extra columns or irrelevant parameters
- Combining data for comparison
- Issues inherent with time series, e.g., missing data

Format and potential issues with output data

What are the challenges for evaluating SWMP data??

- Knowing what we want (I can't help with this)
- Dealing with QAQC columns and removing 'bad' observations
- Data we don't want... extra columns or irrelevant parameters
- Combining data for comparison
- Issues inherent with time series, e.g., missing data

We will learn how to handle most of these challenges!

Overview of the SWMP_r package

What: An R package for retrieving, organizing and analyzing SWMP data

Overview of the SWMP_r package

What: An R package for retrieving, organizing and analyzing SWMP data

Why: There are many challenges for working with SWMP data... a toolkit for addressing these challenges using an open-source format will be useful (I hope!)

Overview of the SWMP_r package

What: An R package for retrieving, organizing and analyzing SWMP data

Why: There are many challenges for working with SWMP data... a toolkit for addressing these challenges using an open-source format will be useful (I hope!)

How:

- Install R/RStudio on your computer (done already?)
- Install the SWMP_r package
- Use the SWMP_r functions to **retrieve**, **organize**, and **analyze** SWMP data

Overview of the SWMPPr package

This is where SWMPPr lives - <https://github.com/fawda123/SWMPPr>

The screenshot shows the GitHub repository page for **fawda123 / SWMPPr**. The repository is described as "r package for accessing, processing, and evaluating data from SWMP of NERRS". It has 21 commits, 1 branch, 0 releases, and 1 contributor. The current branch is **master**. A recent commit by **fawda123** is shown, titled "added method to na.approx for swmpr objects", made 23 minutes ago. The commit message lists several files that were added or modified, including `R`, `README_files`, `data/zip_ex`, `man`, `.Rbuildignore`, `.Rprofile`, `.gitignore`, and `DESCRIPTION`. The right sidebar shows the repository's navigation menu, including **Code**, **Issues**, **Pull Requests**, **Wiki**, **Pulse**, **Graphs**, and **Settings**. The **HTTPS clone URL** is provided as `https://github.com/fawda123/SWMPPr`.

Overview of the SWMP_r package

Scroll down the page to view the [README](#) file, all instructions here...

The screenshot shows a web browser window with the GitHub repository page for `fawda123/SWMP_r`. The browser's address bar shows the URL `https://github.com/fawda123/SWMP_r`. The page lists recent commits:

- `test_analyze.R`: added method to na.approx for swmp_r objects (26 minutes ago)
- `test_organize.R`: modification of station attribute if subsetting removes data type, ad... (5 days ago)
- `test_retrieval.R`: Finished/tested subset function, started comb function, added test da... (11 days ago)

Below the commits, the `README.md` file is selected. The README content is as follows:

SWMP_r package for estuarine monitoring data

This repository contains materials to retrieve, organize, and analyze estuarine monitoring data from the System Wide Monitoring Program (SWMP) implemented by the National Estuarine Research Reserve System (NERRS). SWMP was initiated in 1995 to provide continuous monitoring data at over 300 stations in 28 estuaries across the United States. SWMP data are maintained online by the Centralized Data Management Office (CDMO). This R package will provide several functions to retrieve, organize, and analyze SWMP data from the CDMO. Information on the CDMO web services are available [here](#). Your computer's IP address must be registered with the CDMO website to use most of the data retrieval functions, see contact info in the link. All other functions can be used after obtaining data from the CDMO, as described below.

The package has many dependencies, the most important being the SSOAP package for retrieving data from the CDMO using a SOAP client interface. The SSOAP package is not required to use

Overview of the SWMP_r package

Installation instructions are in the [README](#)

Run these four lines to install the package

```
install.packages('devtools')  
library(devtools)  
install_github('fawda123/SWMPr')  
library(SWMPr)
```

What is it doing?

Packages can be installed from Github using the 'install_github' function from the devtools package

Overview of the SWMP_r package

Your R console should look like this...

```
> install.packages("devtools")
Installing package into 'C:/Users/mbeck/R/library'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/devtools_1.6.1.zip'
Content type 'application/zip' length 284413 bytes (277 kb)
opened URL
downloaded 277 kb

package 'devtools' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\mbeck\AppData\Local\Temp\1\Rtmpupes08\downloaded_packages
> library(devtools)
> install_github('fawda123/SWMPr')
Downloading github repo fawda123/SWMPr@master
Installing SWMPr
"C:/PROGRA~1/R/R-31~1.1/bin/x64/R" --vanilla CMD INSTALL \
"C:/Users/mbeck/AppData\Local\Temp\1\Rtmpupes08\devtools1a304b
64545\fawda123-SWMPr-552d945" \
--library="C:/Users/mbeck/R/library" --install-tests

* installing *source* package 'SWMPr' ...
** R
** data
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
*** arch - i386
*** arch - x64
* DONE (SWMPr)
> library(SWMPr)
> |
```

Overview of the SWMP_r package

What is provided in the SWMP_r package?

Retrieve

```
all_params  
all_params_dtrng  
single_param  
import_local
```

Organize

```
qaqc.swmpr  
subset.swmpr  
setstep.swmpr  
comb.swmpr
```

Analyze

```
aggregate.swmpr  
smoother.swmpr  
na.approx.swmpr  
plot.swmpr  
hist.swmpr  
lines.swmpr  
decomp.swmpr
```

Built around the concept of ***object-oriented programming*** - retrieval functions return a data type with specific methods to organize and analyze

Overview of the SWMP_r package

To view the help file for any function (including examples):

```
?all_params
```

all_params {SWMP_r}

R Documentation

Import current station records from the CDMO

Description

Import current station records from the CDMO starting with the most current date, CDMO equivalent of `exportAllParamsXMLNew`

Usage

```
all_params(station_code, Max = 100)
```

Arguments

`station_code` chr string of station, 7 or 8 characters

`Max` numeric value for number of records to obtain from the current date, maximum of 100

Value

Returns a `swmpr` object, all available parameters including QAQC columns

Overview of the SWMP_r package

Let's get some data into R!

The ***retrieval*** functions do two things:

Import data directly from the CDMO:

```
all_params  
all_params_dtrng  
single_param
```

These functions require [registering your IP address](#) with CDMO, data are also rate-limited
















Import data from a local path:

```
import_local
```

Allows import of data obtained from (and only from) the [zip downloads](#) feature

Overview of the SWMP_r package

After unzipping, data from [zip downloads](#) will have separate .csv files for each station and year

Name	Date modified	Type	Size
 apacpnut2011.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
 apacpnut2012.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
 apacpnut2013.csv	9/19/2014 7:04 AM	Microsoft Excel C...	3 KB
 apacpwq2011.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,481 KB
 apacpwq2012.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,472 KB
 apacpwq2013.csv	9/19/2014 7:06 AM	Microsoft Excel C...	5,567 KB
 apadbnut2011.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
 apadbnut2012.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
 apadbnut2013.csv	9/19/2014 7:06 AM	Microsoft Excel C...	3 KB
 apadbwq2011.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,407 KB
 apadbwq2012.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,483 KB
 apadbwq2013.csv	9/19/2014 7:08 AM	Microsoft Excel C...	5,337 KB
 apaebmet2011.csv	9/19/2014 7:10 AM	Microsoft Excel C...	5,453 KB
 apaebmet2012.csv	9/19/2014 7:10 AM	Microsoft Excel C...	5,401 KB
 apaebmet2013.csv	9/19/2014 7:11 AM	Microsoft Excel C...	5,669 KB

Overview of the SWMP_r package

Use the following to import some data into R... these files are included with the package, otherwise specify the path manually with the commented line for your own data

```
# get data for apacpwq, all years

# location of data
# mypath <- 'C:/data/dataset1'
mypath <- system.file('zip_ex', package = 'SWMPr')

# import and assign to 'dat'
dat <- import_local(mypath, 'apacpwq', trace = T)
```

Overview of the SWMP_r package

Your console should look something like this:

```
# get data for apacpwq, all years

# location of data
# mypath <- 'C:/data/dataset1'
mypath <- system.file('zip_ex', package = 'SWMPr')

# import and assign to 'dat'
dat <- import_local(mypath, 'apacpwq', trace = T)

## Loading files...
##
## apacpwq2011.csv apacpwq2012.csv apacpwq2013.csv
##
## Data imported...
```


Overview of the SWMP_r package

Now we have data in our 'workspace' that we can organize/analyze

```
head(dat)
```

```
##           datetimestamp temp f_temp spcond f_spcond sal f_sal do_pct f_do_pct
## 1 2011-01-01 00:00:00   11  <0>    44    <0>   28  <0>    68    <0>
## 2 2011-01-01 00:15:00   11  <0>    44    <0>   28  <0>    68    <0>
## 3 2011-01-01 00:30:00   11  <0>    44    <0>   28  <0>    68    <0>
## 4 2011-01-01 00:45:00   11  <0>    44    <0>   28  <0>    68    <0>
## 5 2011-01-01 01:00:00   11  <0>    44    <0>   29  <0>    68    <0>
## 6 2011-01-01 01:15:00   11  <0>    44    <0>   29  <0>    67    <0>
##   do_mgl f_do_mgl depth f_depth cdepth f_cdepth level f_level clevel f_clevel
## 1      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
## 2      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
## 3      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
## 4      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
## 5      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
## 6      6    <0>     2    <0>     2    <3>     NA    <-1>     NA     NA
##   ph f_ph turb f_turb chlfluor f_chlfluor
## 1  8 <0>    3  <0>    NA    <-1>
## 2  8 <0>    3  <0>    NA    <-1>
## 3  8 <0>    2  <0>    NA    <-1>
## 4  8 <0>    1  <0>    NA    <-1>
## 5  8 <0>    2  <0>    NA    <-1>
## 6  8 <0>    1  <0>    NA    <-1>
```



NERRS / SWMP

Data Analysis Workshop: *Time Series*

November 17, 2014

Questions??