

USBR_EMP_Data_WY2024

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE)
```

```
library(here)
library(tidyverse)
library(kableExtra)
```

Read in raw and clean data

```
dailydata_formatted <- readRDS(here("data/data_clean/usbr_contdata_wy2024.rds"))

missing_flag <-
readRDS(here("data/data_clean/usbr_contdata_wy2024_missingflag.rds"))

ec_raw <- readRDS(here("data/data_clean/ec_WY2024_flagged.rds")) %>%
  select(station,
         station_d1641,
         ec,
         datetime, month, flagged) %>%
  mutate(datetime = ymd_hms(datetime),
         month = month(datetime),
         date = date(datetime)) %>%
  filter(date > ymd("2023-09-30"))

wt_raw <- readRDS("data/data_clean/wt_WY2024_flagged.rds")%>%
  select(station ,
         station_d1641,
         wt,
         datetime, month, flagged, range) %>%
  mutate(datetime = ymd_hms(datetime),
```

```
month = month(datetime),  
date = date(datetime)) %>%  
filter(date > ymd("2023-09-30"))
```

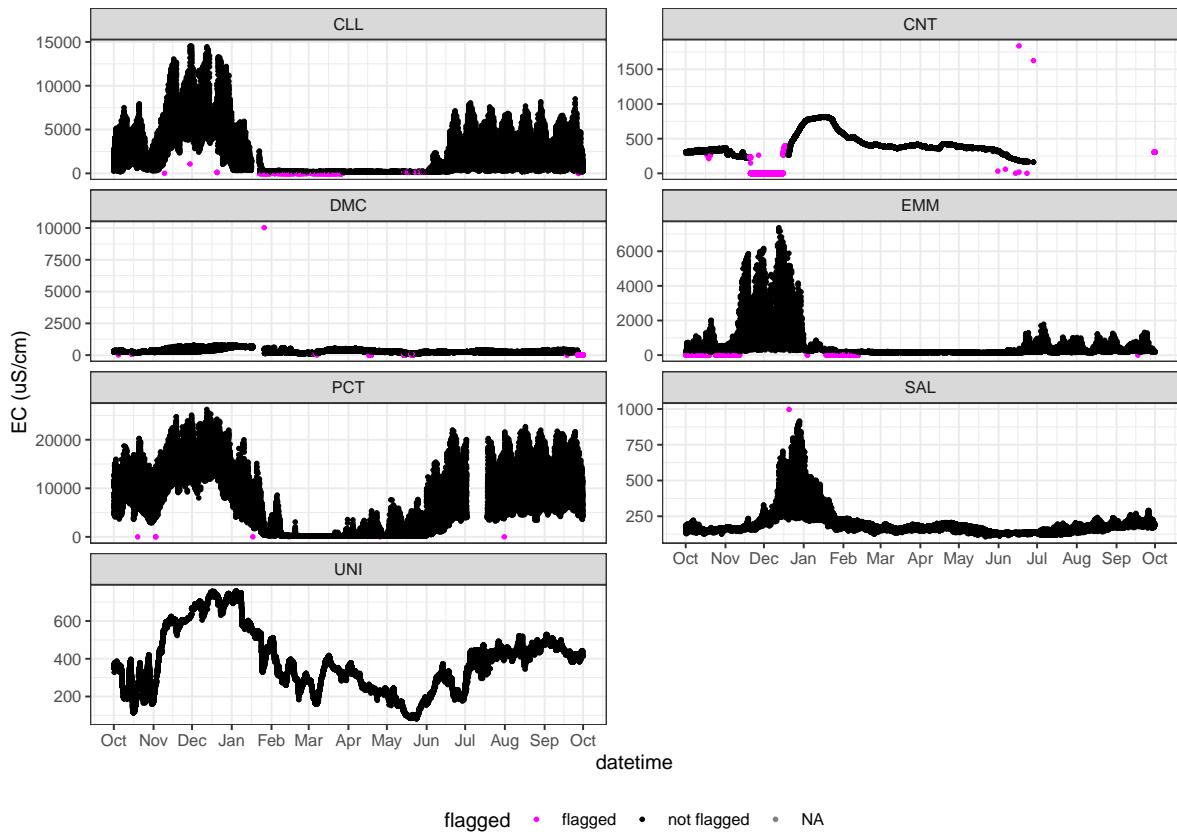
Look at initial data

Removal of data occurred if:

- Value was out of range (less than or equal to zero, aside from CLL)
- Data visually appeared as outlier

EC flagged data

```
ggplot(ec_raw %>% filter(!is.na(datetime))) +  
  geom_point(aes(datetime, ec, color = flagged), size = 0.75) +  
  facet_wrap(~station, scales = "free_y", ncol = 2) +  
  labs(y = "EC (uS/cm)") +  
  scale_x_datetime(date_breaks = "1 month", date_labels = "%b") +  
  scale_color_manual(values = c("magenta", "black")) +  
  theme_bw() +  
  theme(legend.position = "bottom")
```



```

ec_removed <- ec_raw %>%
  group_by(station) %>%
  summarize(removed = sum(flagged == "flagged", na.rm = TRUE),
            values = n(),
            nonNA_vals = sum(!is.na(ec)),
            pct_removed = if_else(nonNA_vals>0,
                                  removed/nonNA_vals*100,
                                  0))

kableExtra::kable(ec_removed) %>%
  kable_styling(c("striped", "condensed"),
                full_width = F,
                latex_options = "striped")

```

Table 1: Removed EC values

station	removed	values	nonNA_vals	pct_removed
CLL	1856	34770	33296	5.5742432
CNT	2955	34770	23898	12.3650515
DMC	771	34770	33581	2.2959412
EMM	3263	34770	34481	9.4631826
PCT	7	34770	32987	0.0212205
SAL	1	34770	34504	0.0028982
UNI	0	34770	34492	0.0000000

Removal of data occurred if:

- Value was out of range (less than or equal to 32, greater than or equal to 90)
- Data visually appeared as outlier

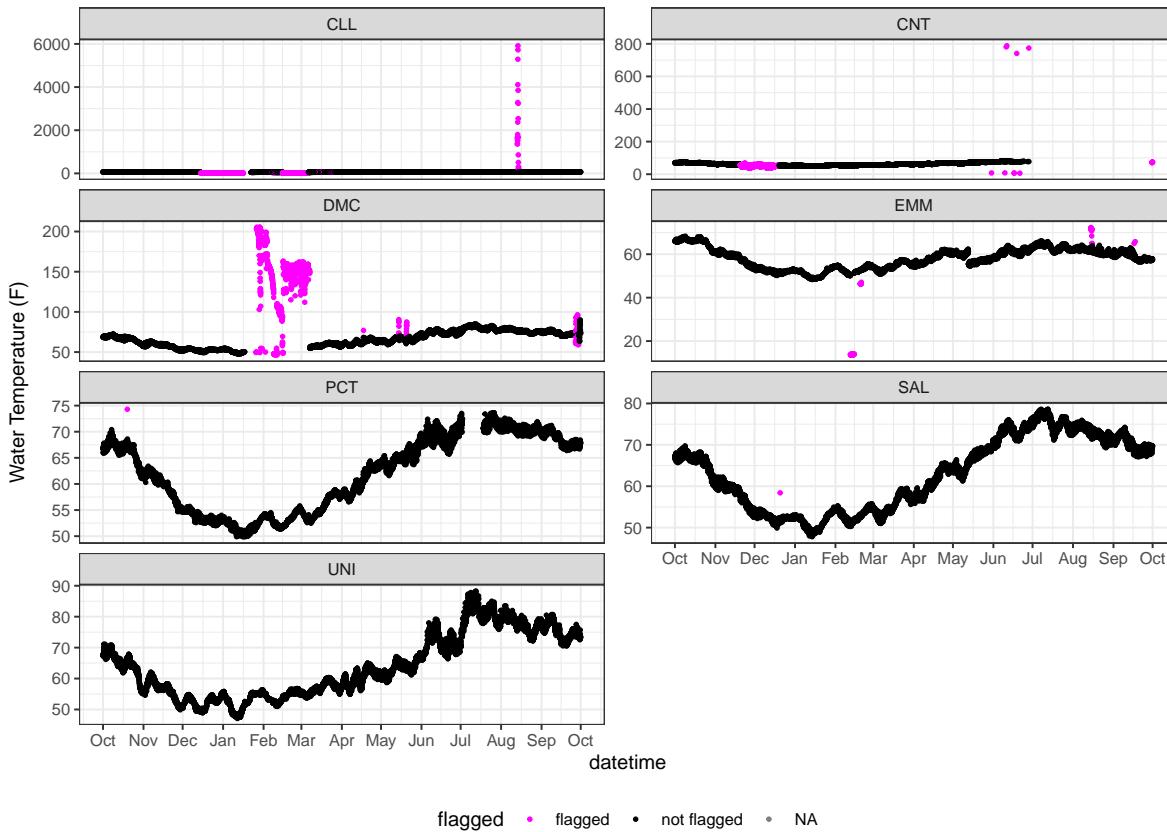
Water temperature flagged data

```

ggplot(wt_raw %>% filter(!is.na(datetime))) +
  geom_point(aes(datetime,wt, color = flagged), size = 0.75) +
  facet_wrap(~station, scales = "free_y", ncol = 2) +
  labs(y = "Water Temperature (F)") +
  scale_x_datetime(date_breaks = "1 month", date_labels = "%b") +
  scale_color_manual(values = c("magenta", "black"))

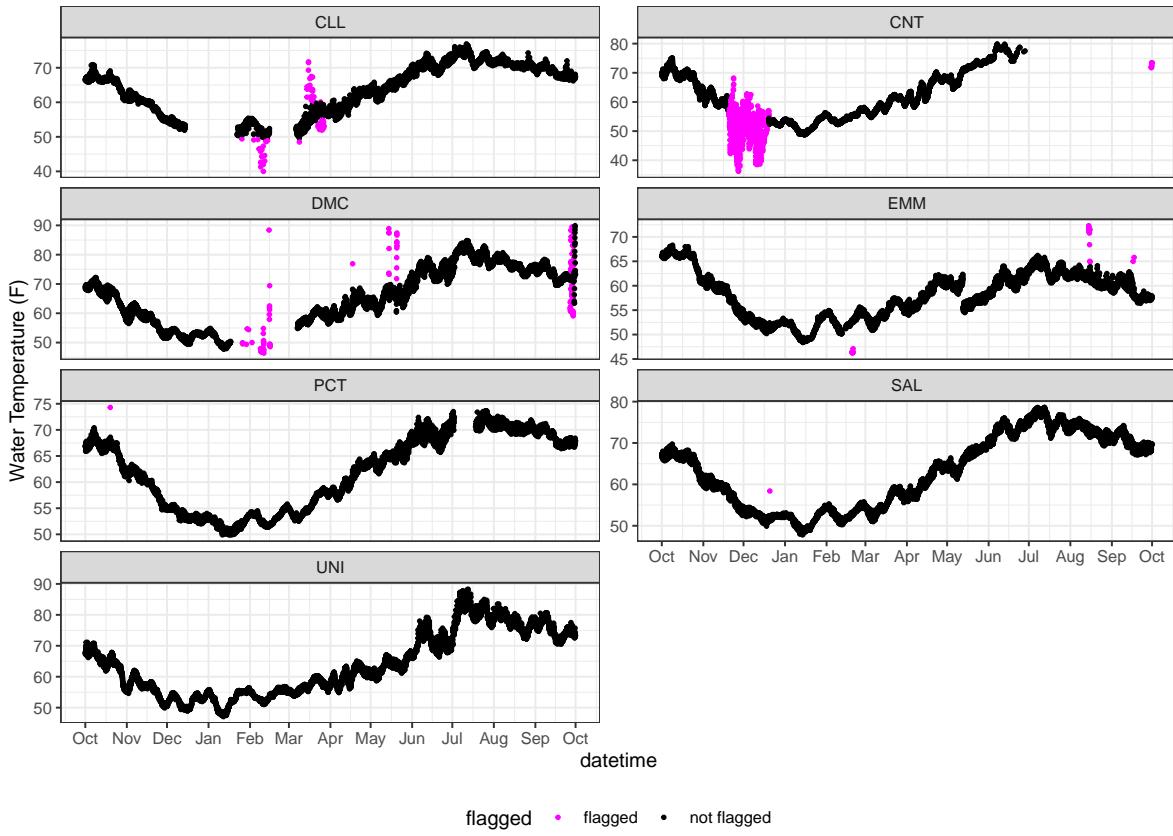
```

```
theme_bw() +
theme(legend.position = "bottom")
```



Water temperature flagged data (removing out of range data)

```
ggplot(wt_raw %>% filter(!is.na(datetime), range==0L)) +
  geom_point(aes(datetime,wt, color = flagged), size = 0.75) +
  facet_wrap(~station, scales = "free_y", ncol = 2) +
  labs(y = "Water Temperature (F)") +
  scale_x_datetime(date_breaks = "1 month", date_labels = "%b") +
  scale_color_manual(values = c("magenta", "black")) +
  theme_bw() +
  theme(legend.position = "bottom")
```



```

wt_removed <- wt_raw %>%
  group_by(station) %>%
  summarize(removed = sum(flagged == "flagged", na.rm = TRUE),
            values = n(),
            nonNA_vals = sum(!is.na(wt)),
            pct_removed = if_else(nonNA_vals>0,
                                  removed/nonNA_vals*100,
                                  0))

kableExtra::kable(wt_removed) %>%
  kable_styling(c("striped", "condensed"),
                full_width = F,
                latex_options = "striped")

```

Table 2: Removed WT Values

station	removed	values	nonNA_vals	pct_removed
CLL	1175	8418	8066	14.5673196
CNT	722	8418	5801	12.4461300
DMC	1030	8418	8130	12.6691267
EMM	103	8418	8349	1.2336807
PCT	1	8418	7988	0.0125188
SAL	1	8418	8353	0.0119717
UNI	0	8418	8352	0.0000000

Look at final data

- While EC was collected at the 15-minute level, SpC required water temperature, which was collected at the hourly level.
- At this stage we converted data to NA if >30% of daily data was missing

```

vals_removed <- missing_flag %>%
  group_by(Site, Station, Analyte) %>%
  summarize(removed = sum(missing_flag == 1L, na.rm = TRUE),
            values = n(),
            nonNA_vals = sum(!is.na(Value)),
            pct_removed = if_else(nonNA_vals>0,
                                  removed/nonNA_vals*100,
                                  0))

kableExtra::kable(vals_removed) %>%
  kable_styling(c("striped", "condensed"),
                full_width = F,
                latex_options = "striped")

```

Table 3: Percent of data removed due to daily missing values

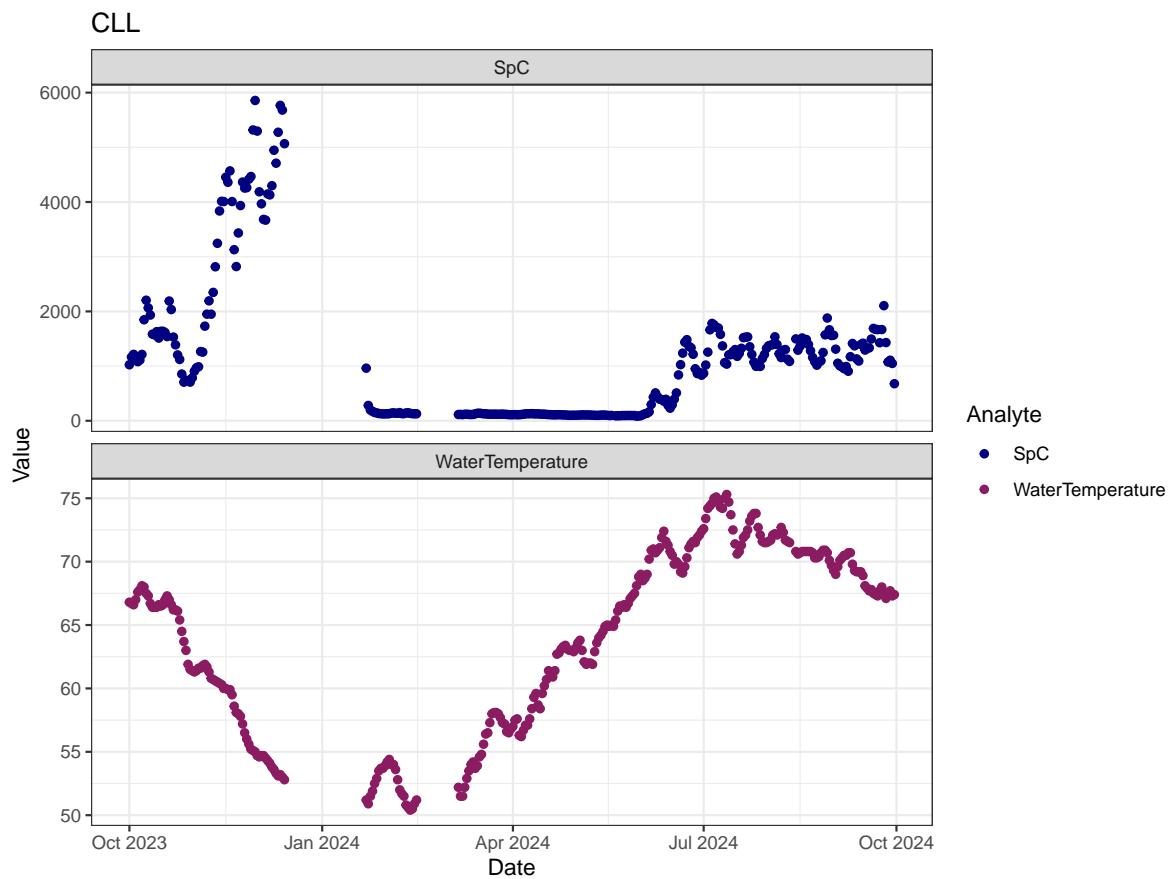
Site	Station	Analyte	removed	values	nonNA_vals	pct_removed
CLL	C2	spc	54	360	307	17.5895765
CLL	C2	wt	53	360	307	17.2638436
CNT	C5	spc	7	240	238	2.9411765
CNT	C5	wt	6	240	240	2.5000000
DMC	DMC1	spc	45	357	313	14.3769968
DMC	DMC1	wt	43	357	316	13.6075949
EMM	D22	spc	24	366	361	6.6481994
EMM	D22	wt	2	366	364	0.5494505
PCT	C14	spc	0	351	351	0.0000000
PCT	C14	wt	0	351	351	0.0000000
SAL	C4	spc	0	366	366	0.0000000
SAL	C4	wt	0	366	366	0.0000000
UNI	C8	spc	0	366	366	0.0000000
UNI	C8	wt	0	366	366	0.0000000

```

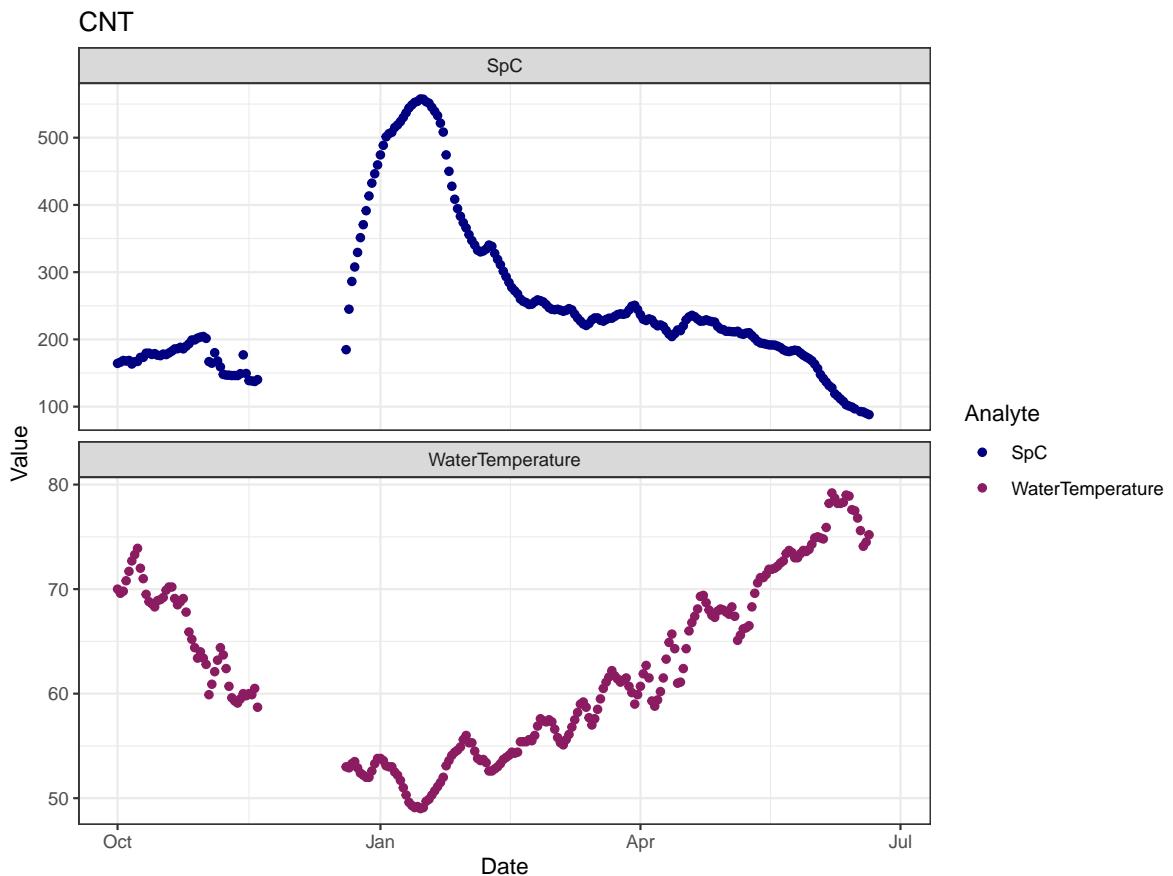
plot_vals <- function(CDEC_sta) {
  ggplot(dailydata_formatted %>% filter(Site == CDEC_sta))+ 
    geom_point(aes(Date, Value, color = Analyte)) + 
    facet_wrap(~Analyte, scales = "free_y", nrow = 2) + 
    labs(title = CDEC_sta)+ 
    scale_color_manual(values = c("navy", "maroon4"))+ 
    theme_bw()
}

```

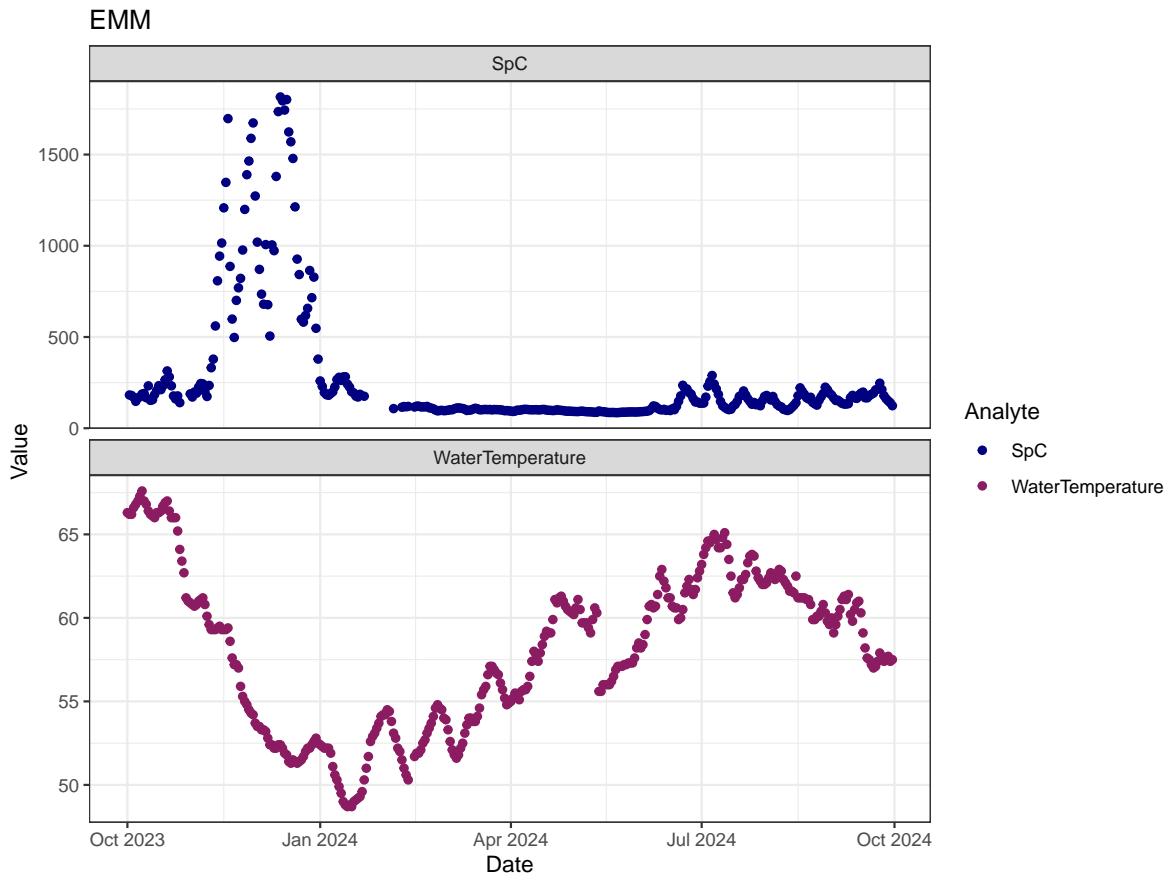
```
plot_vals("CLL")
```



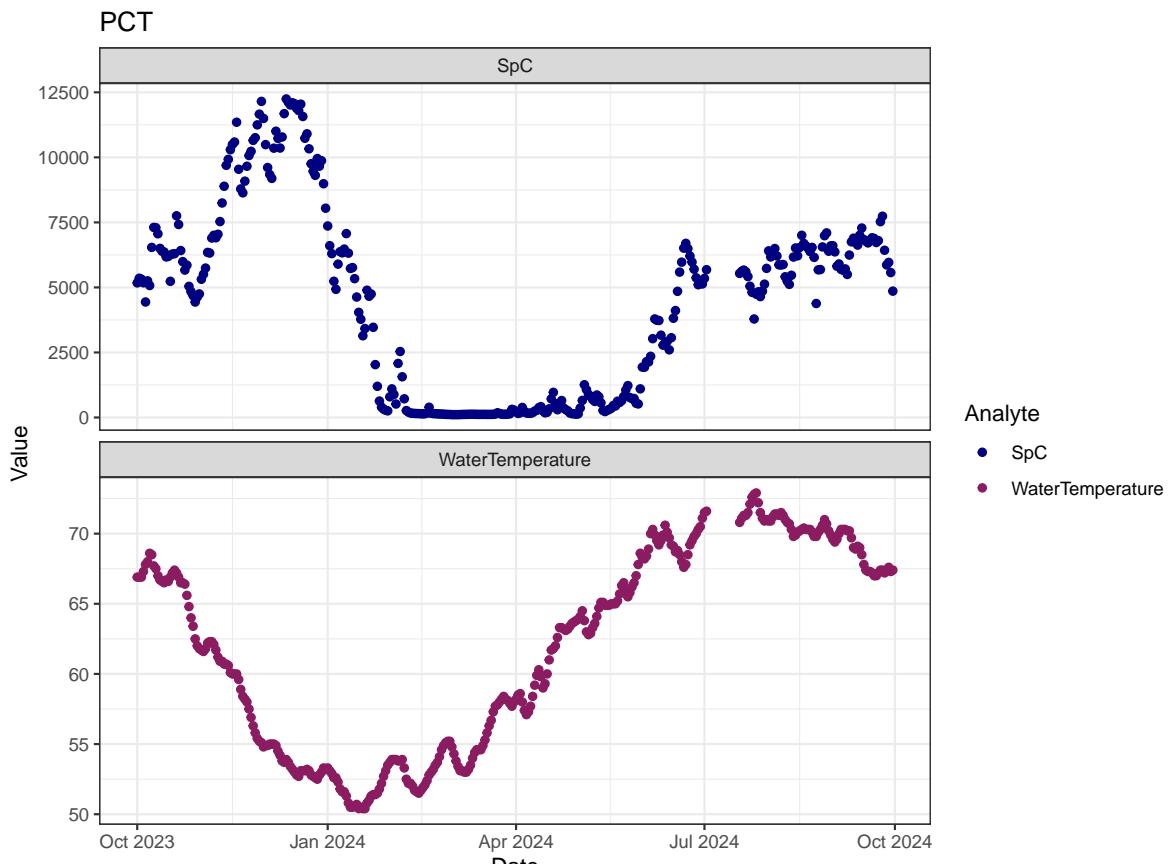
```
plot_vals("CNT")
```



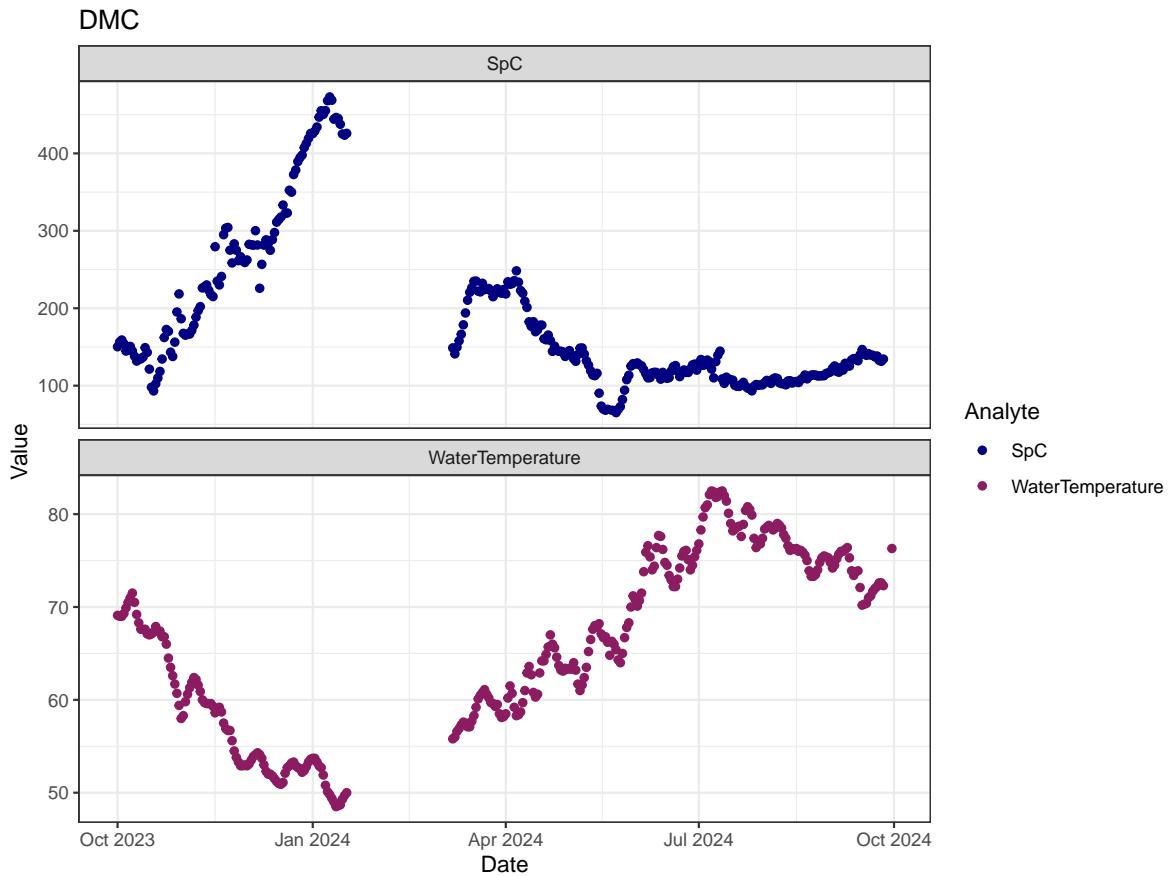
```
plot_vals("EMM")
```



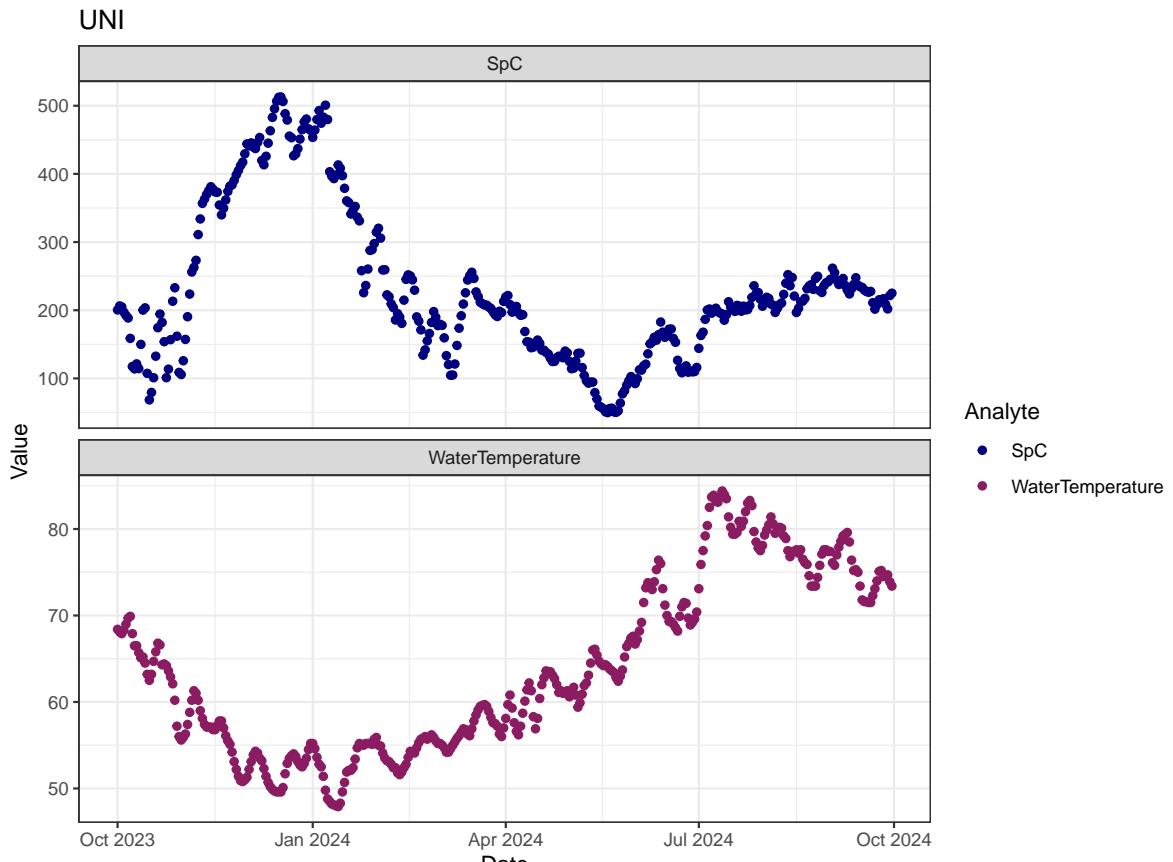
```
plot_vals("PCT")
```



```
plot_vals("DMC")
```



```
plot_vals("UNI")
```



```
plot_vals("SAL")
```

