Stream temperatures

Charlotte Wickham

```
library(tidyverse)
library(lubridate)
knitr::opts_chunk$set(fig.height = 3)
```

Explore stream temperatures for the sites:

- GSWS01 Andrews Lookout Creek Gaging Station
- GSMACK Andrews Mack Creek Gaging Station
- GSWS01 Andrews Watershed 1 Gaging Station

This is where we finished last time, but reorganized to group the operations. I've also added spots for the breakout room activities.

Functions

```
# Calculate useful data coverage summaries
coverage_summary <- function(data){
  data %>%
  summarise(
    n_obs = n(),
    n_days = n_distinct(date),
    n_missing = sum(is.na(watertemp_mean_day))
  )
}
```

```
# Plot watertemp_day time series
plot_temperature <- function(data, title){
  data %>%
    ggplot(aes(date, watertemp_mean_day)) +
    geom_line(aes(color = watertemp_method)) +
    scale_color_brewer("Method", palette = "Set2") +
    labs(x = "", y = "Water temperature (C)",
      title = title)
}
```

Data Import

```
streams <- read_csv("data/HT00441_v8.csv")
names(streams) <- tolower(names(streams))
```

Filter to time period of interest, and set observations with anything other than "Accepted" quality to missing:

```
streams <- streams %>%
  filter(year(date) > 2000, year(date) < 2020) %>%
  mutate(
    watertemp_mean_day = ifelse(watertemp_mean_flag != "A",
      NA, watertemp_mean_day)) %>%
  select(sitecode, date, watertemp_mean_day, watertemp_method)
```

Site Subsets

```
gslook <- streams %>%
  filter(sitecode == "GSLOOK")
gsmack <- streams %>%
  filter(sitecode == "GSMACK")
gsws01 <- streams %>%
  filter(sitecode == "GSWS01")
```

Coverage Summaries

```
gslook %>% coverage_summary()
```

```
## # A tibble: 1 x 3
##   n_obs n_days n_missing
##   <int>  <int>     <int>
## 1  8120   6841       543
```

```
gsmack %>% coverage_summary()
```

```
## # A tibble: 1 x 3
##   n_obs n_days n_missing
##   <int>  <int>     <int>
## 1  6847   6847       428
```

```
gsws01 %>% coverage_summary()
```

```
## # A tibble: 1 x 3
##   n_obs n_days n_missing
##   <int>  <int>     <int>
## 1  6847   6847       223
```

Breakout Room Activity #1

Discuss the following code, e.g. in words what does it do?:

```
streams %>%
  filter(sitecode %in% c("GSLOOK", "GSMACK","GSWS01")) %>%
  group_by(sitecode) %>%
  coverage_summary()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

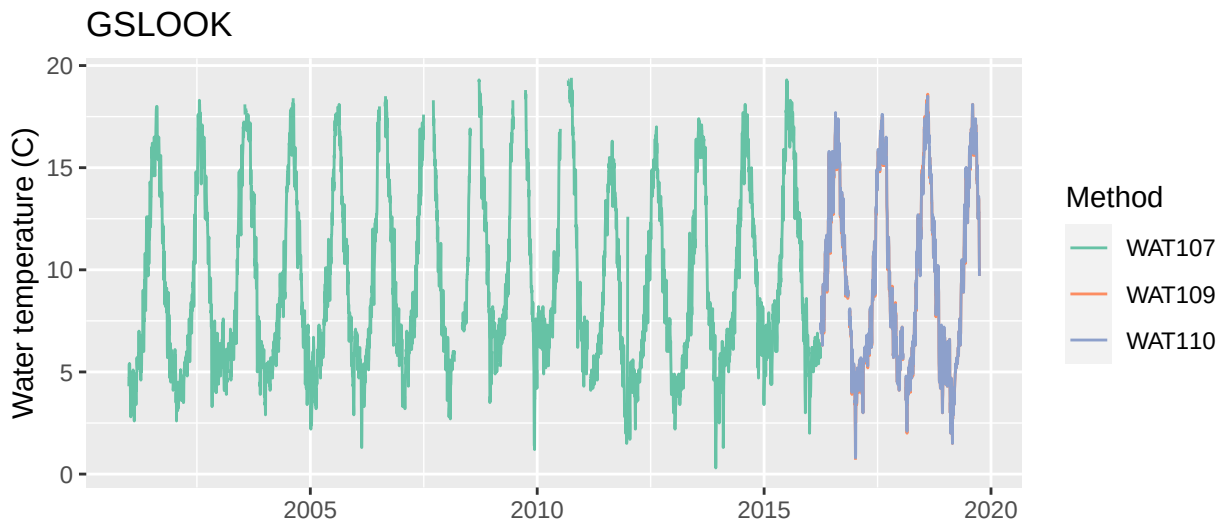
```
## # A tibble: 3 x 4
##   sitecode n_obs n_days n_missing
##   <chr>    <int>  <int>     <int>
## 1 GSLOOK    8120   6841       543
## 2 GSMACK    6847   6847       428
## 3 GSWS01    6847   6847       223
```

Does it work? Run it and see.

Plots

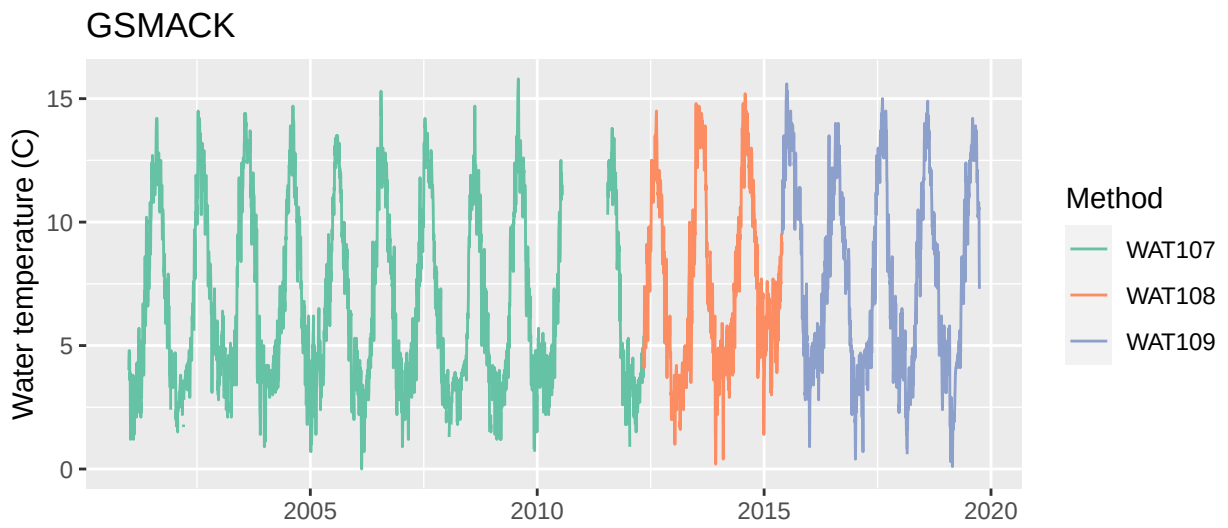
```
gslook %>% plot_temperature("GSLOOK")
```

```
## Warning: Removed 5 row(s) containing missing values (geom_path).
```



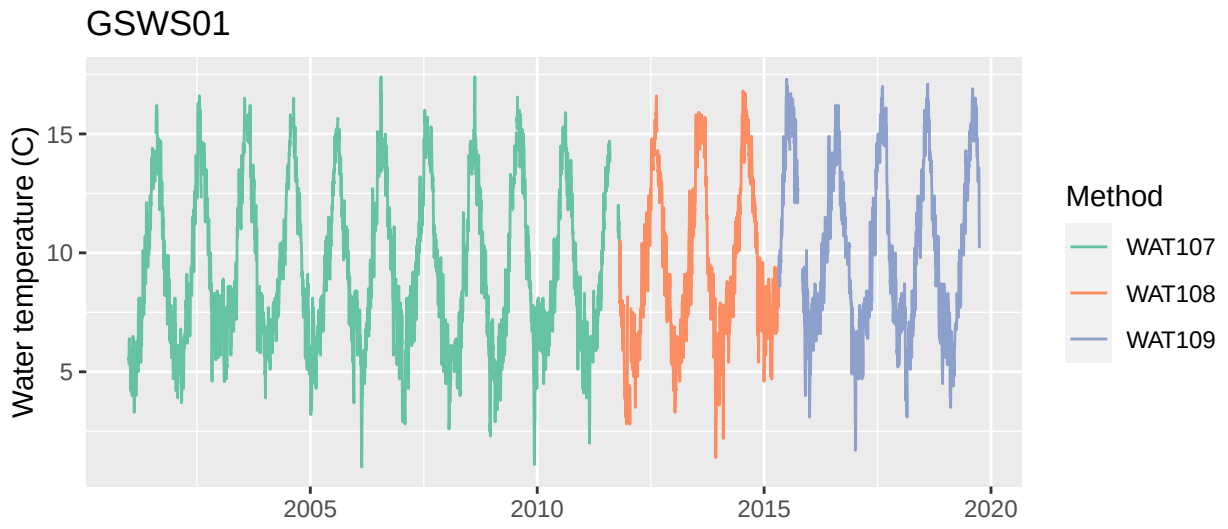

```
gsmack %>% plot_temperature("GSMACK")
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```




```
gsws01 %>% plot_temperature("GSWS01")
```

```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



Take the same approach with plotting:

Breakout Room Activity #2

Run the following code and explore `streams_nested`. You might like to look at it on the Console, or using `View()`.

How does `nest_by()` differ to 'group_by()?

```
streams_nested <- streams %>%
  nest_by(sitecode)
```

Combine with `rowwise()`:

Look at a plot: