

Process the Raw Data

```
here::i_am("data/process_raw_data.qmd")

library(conflicted)
library(here)
library(readxl)
library(rnaturalearth)
library(rnaturalearthdata)
library(sf)
library(tidyverse)
conflicts_prefer(dplyr::filter, dplyr::lag)

site_data <- read_xlsx(
  here("data/raw/USFWS_Bat_Acoustic_Data/NightlyPassCounts.xlsx")
)
site_metadata <- read_csv(
  here("data/raw/Bat_Detector_Metadatas/acoustic_sites.csv")
)

theme_set(theme_bw())
```

The following objects contain information on state, province, and lake boundaries and will be used when making maps.

```
usa <- ne_states(country = "United States of America", returnclass = "sf")
ontario <- ne_states(country = "Canada", returnclass = "sf") %>%
  filter(name == "Ontario")
great_lakes <- ne_download(
  scale = 50, type = "lakes", category = "physical", returnclass = "sf"
) %>%
  filter(
    name %in% c(
```

```

    "Lake Superior", "Lake Michigan", "Lake Huron", "Lake Erie",
    "Lake Ontario"
  )
)

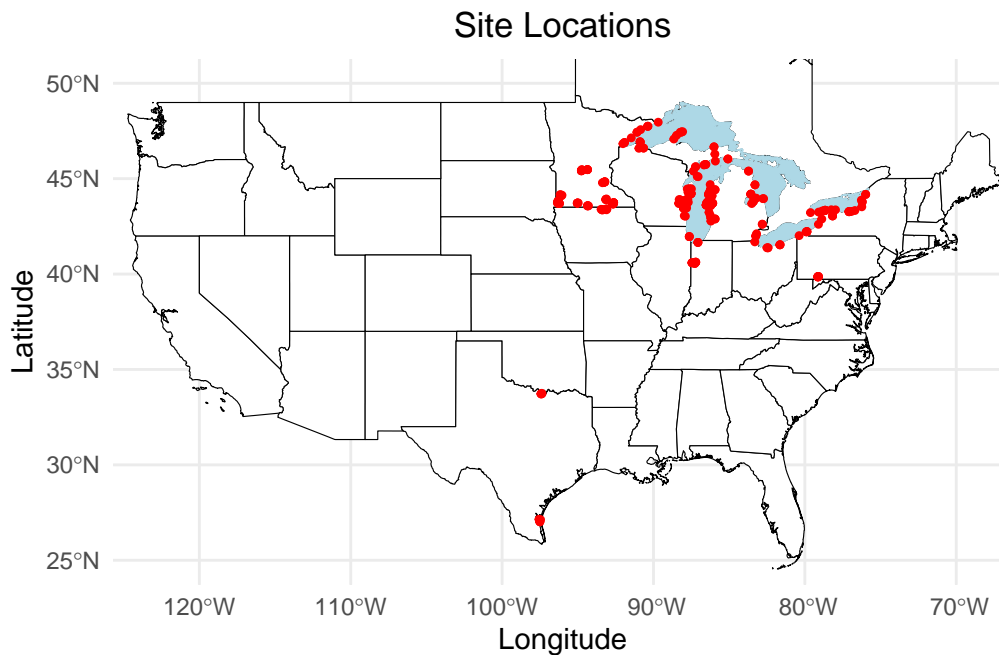
```

We're interested in bat populations living near the Great Lakes. The map below reveals that a number of sites are far away from the Great Lakes. Those sites are in southern Minnesota, northern Iowa, central Indiana, southern Pennsylvania, and Texas.

```

ggplot() +
  geom_sf(data = usa, color = "black", fill = "white") +
  geom_sf(data = ontario, color = "black", fill = "white") +
  geom_sf(data = great_lakes, color = "lightblue", fill = "lightblue") +
  geom_point(aes(Long, Lat), site_metadata, color = "red", shape = "bullet") +
  coord_sf(xlim = c(-123, -69), ylim = c(25, 50)) +
  labs(x = "Longitude", y = "Latitude", title = "Site Locations") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

```



We are going to identify the sites that are far away from the Great Lakes. It will help to have the state/province for each site; we find the sites' states and provinces below.

```

site_metadata_sf <- st_as_sf(
  select(site_metadata, c(AcousticSite, Long, Lat)),
  coords = c("Long", "Lat"),
  crs = 4326 # Need to use WGS84 CRS
)
site_states <- site_metadata_sf %>%
  st_join(rbind(usa, ontario)["name"]) %>%
  st_drop_geometry()
site_metadata <- site_metadata %>%
  inner_join(site_states, by = "AcousticSite") %>%
  relocate(state_province = name, .after = AcousticSite)

```

For a small fraction of the sites, no state or province was found.

```

site_metadata %>%
  filter(is.na(state_province)) %>%
  select(AcousticSite:Lat) %>%
  print(n = Inf)

```

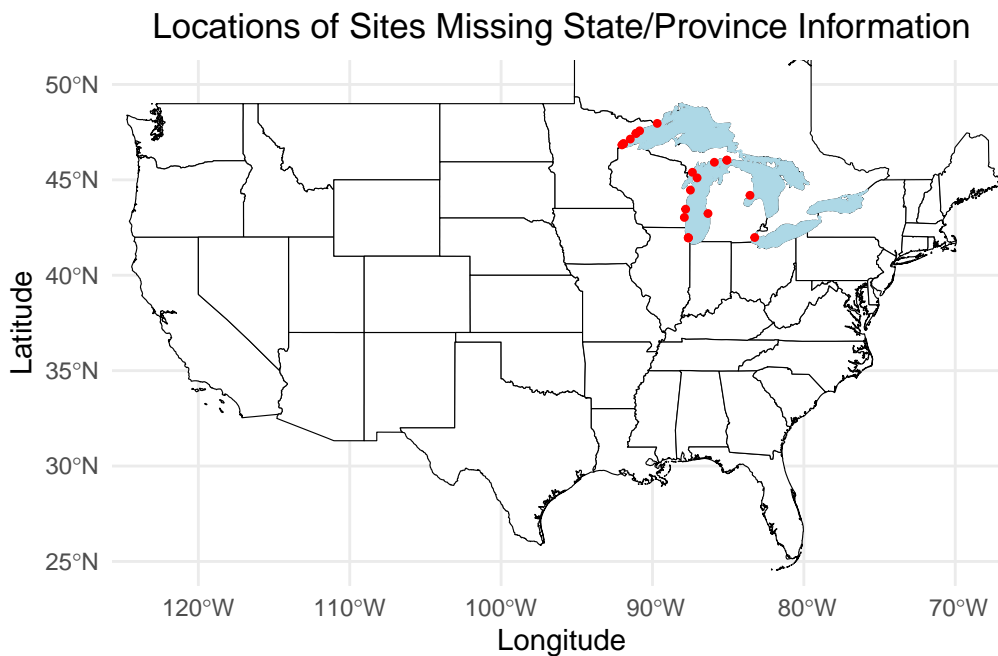
A tibble: 22 x 4

	AcousticSite	state_province	Long	Lat
	<chr>	<chr>	<dbl>	<dbl>
1	BAILEY	<NA>	-87.1	45.1
2	FOREST	<NA>	-87.8	43.5
3	GOOSE	<NA>	-91.5	47.1
4	HWKRDG	<NA>	-92.0	46.8
5	Kewaunee	<NA>	-87.5	44.5
6	LSSP	<NA>	-87.9	43.0
7	MUSK	<NA>	-86.3	43.2
8	NPS-GP	<NA>	-89.7	48.0
9	RadarF14_Batman-Manitou1	<NA>	-91.1	47.4
10	RadarF14_Batman-Manitou2	<NA>	-91.1	47.4
11	RadarF18_Robin-DetS2	<NA>	-83.2	42.0
12	RadarS15_Robin-Silica1	<NA>	-85.1	46.0
13	RadarS15_Robin-Silica2	<NA>	-85.1	46.0
14	RadarS17_Robin-Montrose1	<NA>	-87.6	42.0
15	RadarS17_Robin-Montrose2	<NA>	-87.6	42.0
16	RadarS17_Robin-Montrose3	<NA>	-87.6	42.0
17	RadarS18A_Robin-Trappist1	<NA>	-91.9	46.9
18	RadarS18A_Robin-Trappist2	<NA>	-91.9	46.9
19	SEUL-CHOIX	<NA>	-85.9	45.9

20	USFS-TFTE	<NA>	-90.9	47.6
21	USG2	<NA>	-83.6	44.2
22	WELLS	<NA>	-87.4	45.4

All of these sites appear to be coastal sites.

```
ggplot() +
  geom_sf(data = usa, color = "black", fill = "white") +
  geom_sf(data = ontario, color = "black", fill = "white") +
  geom_sf(data = great_lakes, color = "lightblue", fill = "lightblue") +
  geom_point(
    aes(Long, Lat),
    filter(site_metadata, is.na(state_province)),
    color = "red", shape = "bullet"
  ) +
  coord_sf(xlim = c(-123, -69), ylim = c(25, 50)) +
  labs(
    x = "Longitude", y = "Latitude",
    title = "Locations of Sites Missing State/Province Information"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



We use Google Maps to identify the states and provinces for those sites.

```

site_metadata <- mutate(
  site_metadata,
  state_province = case_when(
    # https://www.google.com/maps/place/45°06'15.8%22N+87°03'10.1%22W
    AcousticSite == "BAILEY" ~ "Wisconsin",
    # https://www.google.com/maps/place/43°27'43.5%22N+87°48'38.8%22W
    AcousticSite == "FOREST" ~ "Wisconsin",
    # https://www.google.com/maps/place/47°08'15.9%22N+91°28'05.5%22W
    AcousticSite == "GOOSE" ~ "Minnesota",
    # https://www.google.com/maps/place/46°49'57.4%22N+92°01'08.3%22W
    AcousticSite == "HWKRDG" ~ "Minnesota",
    # https://www.google.com/maps/place/44°27'51.7%22N+87°29'42.2%22W
    AcousticSite == "Kewaunee" ~ "Wisconsin",
    # https://www.google.com/maps/place/43°01'52.7%22N+87°53'42.4%22W
    AcousticSite == "LSSP" ~ "Wisconsin",
    # https://www.google.com/maps/place/43°14'08.9%22N+86°19'57.4%22W
    AcousticSite == "MUSK" ~ "Michigan",
    # https://www.google.com/maps/place/47°57'31.8%22N+89°41'16.5%22W
    AcousticSite == "NPS-GP" ~ "Minnesota",
    # https://www.google.com/maps/place/47°26'11.7%22N+91°05'14.7%22W
    AcousticSite == "RadarF14_Batman-Manitou1" ~ "Minnesota",
    # https://www.google.com/maps/place/47°26'02.8%22N+91°05'16.0%22W
    AcousticSite == "RadarF14_Batman-Manitou2" ~ "Minnesota",
    # https://www.google.com/maps/place/41°58'47.2%22N+83°14'52.4%22W
    AcousticSite == "RadarF18_Robin-DetS2" ~ "Michigan",
    # https://www.google.com/maps/place/46°01'55.6%22N+85°04'59.0%22W
    AcousticSite == "RadarS15_Robin-Silica1" ~ "Michigan",
    # https://www.google.com/maps/place/46°02'00.9%22N+85°05'06.3%22W
    AcousticSite == "RadarS15_Robin-Silica2" ~ "Michigan",
    # https://www.google.com/maps/place/41°57'47.7%22N+87°38'24.5%22W
    AcousticSite == "RadarS17_Robin-Montrose1" ~ "Illinois",
    # https://www.google.com/maps/place/41°57'45.9%22N+87°37'59.2%22W
    AcousticSite == "RadarS17_Robin-Montrose2" ~ "Illinois",
    # https://www.google.com/maps/place/41°57'47.7%22N+87°37'53.4%22W
    AcousticSite == "RadarS17_Robin-Montrose3" ~ "Illinois",
    # https://www.google.com/maps/place/46°53'49.8%22N+91°54'47.4%22W
    AcousticSite == "RadarS18A_Robin-Trappist1" ~ "Minnesota",
    # https://www.google.com/maps/place/46°53'46.9%22N+91°54'46.8%22W
    AcousticSite == "RadarS18A_Robin-Trappist2" ~ "Minnesota",
    # https://www.google.com/maps/place/45°55'16.0%22N+85°54'42.8%22W
    AcousticSite == "SEUL-CHOIX" ~ "Michigan",
    # https://www.google.com/maps/place/47°34'11.0%22N+90°51'03.2%22W

```

```

    AcousticSite == "USFS-TFTE" ~ "Minnesota",
    # https://www.google.com/maps/place/44°11'25.7%22N+83°33'31.4%22W
    AcousticSite == "USG2" ~ "Michigan",
    # https://www.google.com/maps/place/45°23'48.5%22N+87°21'48.7%22W
    AcousticSite == "WELLS" ~ "Michigan",
    .default = state_province
  )
)

```

Below are the states with sites.

```
count(site_metadata, state_province, sort = TRUE, name = "num_sites")
```

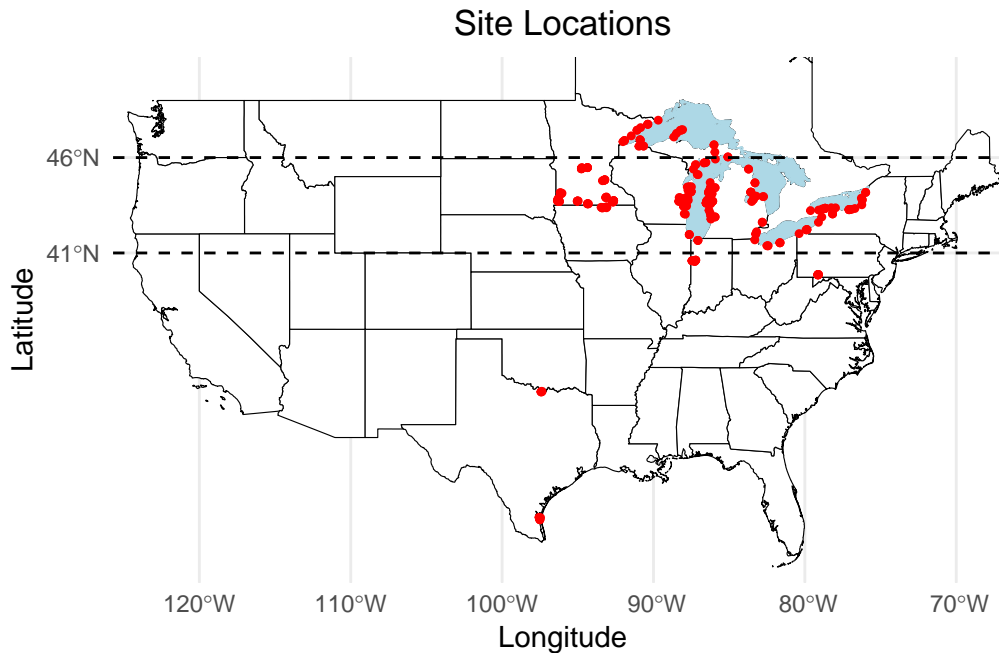
```
# A tibble: 11 x 2
  state_province num_sites
  <chr>          <int>
1 Michigan             74
2 Minnesota            73
3 Wisconsin           45
4 New York            25
5 Indiana             15
6 Pennsylvania        14
7 Texas              11
8 Iowa                7
9 Ohio                7
10 Illinois           3
11 Ontario            2
```

It looks like the sites far away from the Great Lakes are those that are either in Minnesota and below 46°N, or are in Iowa, or are below 41°N.

```

ggplot() +
  geom_sf(data = usa, color = "black", fill = "white") +
  geom_sf(data = ontario, color = "black", fill = "white") +
  geom_sf(data = great_lakes, color = "lightblue", fill = "lightblue") +
  geom_point(aes(Long, Lat), site_metadata, color = "red", shape = "bullet") +
  geom_hline(yintercept = c(41, 46), linetype = "dashed") +
  scale_y_continuous(breaks = c(41, 46)) +
  coord_sf(xlim = c(-123, -69), ylim = c(25, 50)) +
  labs(x = "Longitude", y = "Latitude", title = "Site Locations") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

```



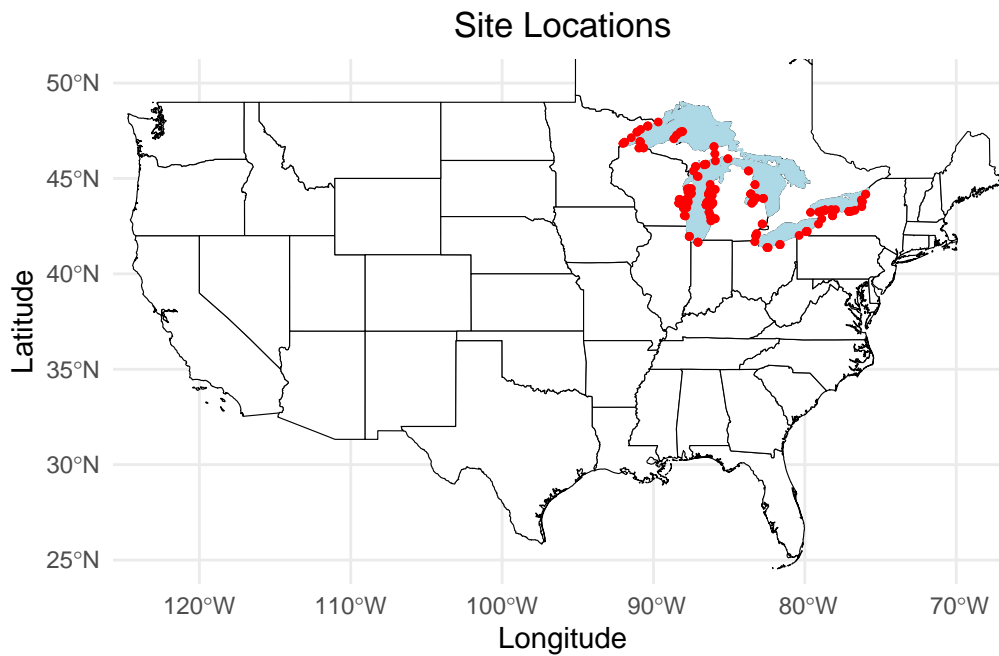
We filter out the sites that are far away from the Great Lakes.

```
great_lakes_site_metadata <- filter(
  site_metadata,
  !(
    ((state_province == "Minnesota") & (Lat <= 46)) |
    (state_province == "Iowa") |
    (Lat <= 41)
  )
)
```

The site map now looks like this.

```
ggplot() +
  geom_sf(data = usa, color = "black", fill = "white") +
  geom_sf(data = ontario, color = "black", fill = "white") +
  geom_sf(data = great_lakes, color = "lightblue", fill = "lightblue") +
  geom_point(
    aes(Long, Lat),
    great_lakes_site_metadata,
    color = "red", shape = "bullet"
  ) +
  coord_sf(xlim = c(-123, -69), ylim = c(25, 50)) +
  labs(x = "Longitude", y = "Latitude", title = "Site Locations") +
```

```
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))
```



We filter out the data for the sites that are far away from the Great Lakes.

```
great_lakes_site_data <- filter(
  site_data, AcousticSite %in% great_lakes_site_metadata$AcousticSite
)
```

The percentage reductions in the number of sites and the number of data records are below.

```
c(
  num_sites_reduc = 100 * (
    nrow(great_lakes_site_metadata) / nrow(site_metadata) - 1
  ),
  num_data_recs_reduc = 100 * (
    nrow(great_lakes_site_data) / nrow(site_data) - 1
  )
)
```

```
num_sites_reduc num_data_recs_reduc
-36.23188      -43.11083
```


We save the data and metadata for the Great Lakes sites under shorter names.

```
env <- list2env(  
  list(  
    site_metadata = great_lakes_site_metadata,  
    site_data = great_lakes_site_data  
  )  
)  
save(  
  list = ls(env),  
  file = here("data/processed/site_data_metadata.RData"),  
  envir = env  
)
```