

Exploration of Methods for Geographic Visualization

Tova Ibbotson

Static Visualization

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Loading required package: maps

## Loading required package: viridis

## Loading required package: viridisLite

##
## Attaching package: 'viridis'

## The following object is masked from 'package:maps':
##
##   unemp
```

Over 98% of the data in the latitude-longitude variable are missing, so geographic visualization of the data must be done at the state-level. We have state-level location information for roughly 60% of the isolates, so geographic analysis will be restricted to the 1060 observations collected between 2015 and 2022 with state-level location data. Figure 1 shows the number of isolates found in each state over the time period.

Figure 1
Number of Isolates by State

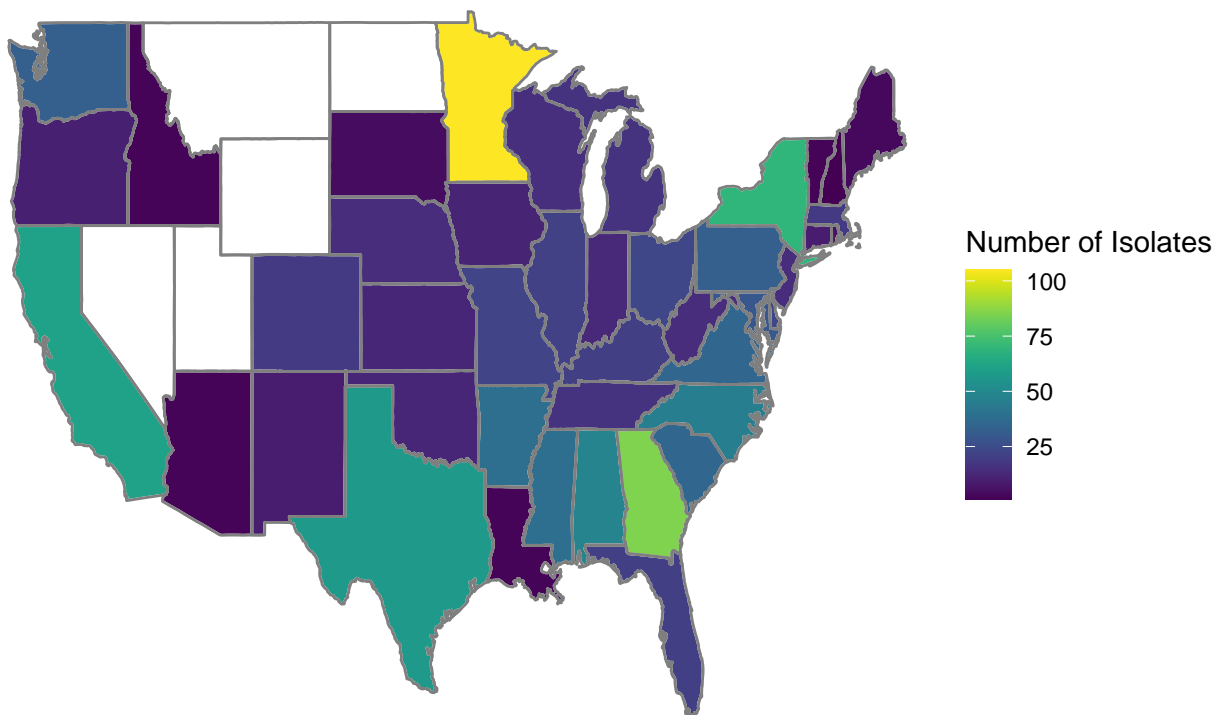


Figure 2
Number of Isolates by Microbiological Monitoring Region

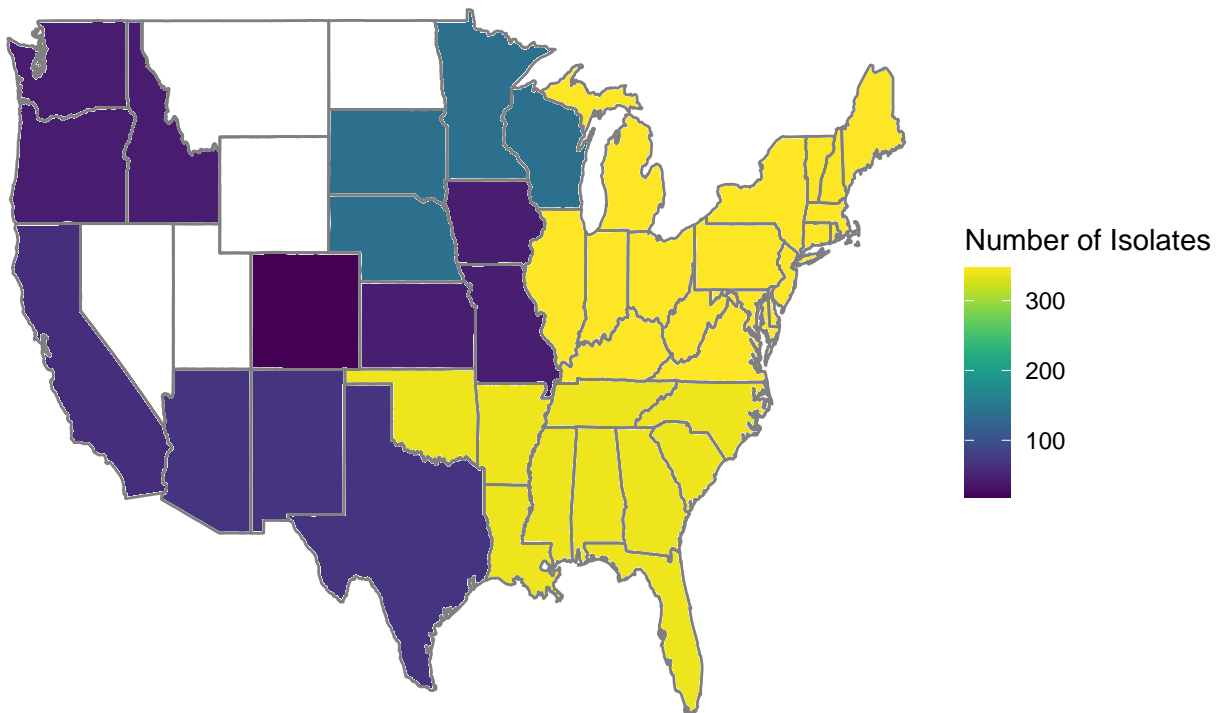


Figure 3
Map of CDC Microbiological Monitoring Regions

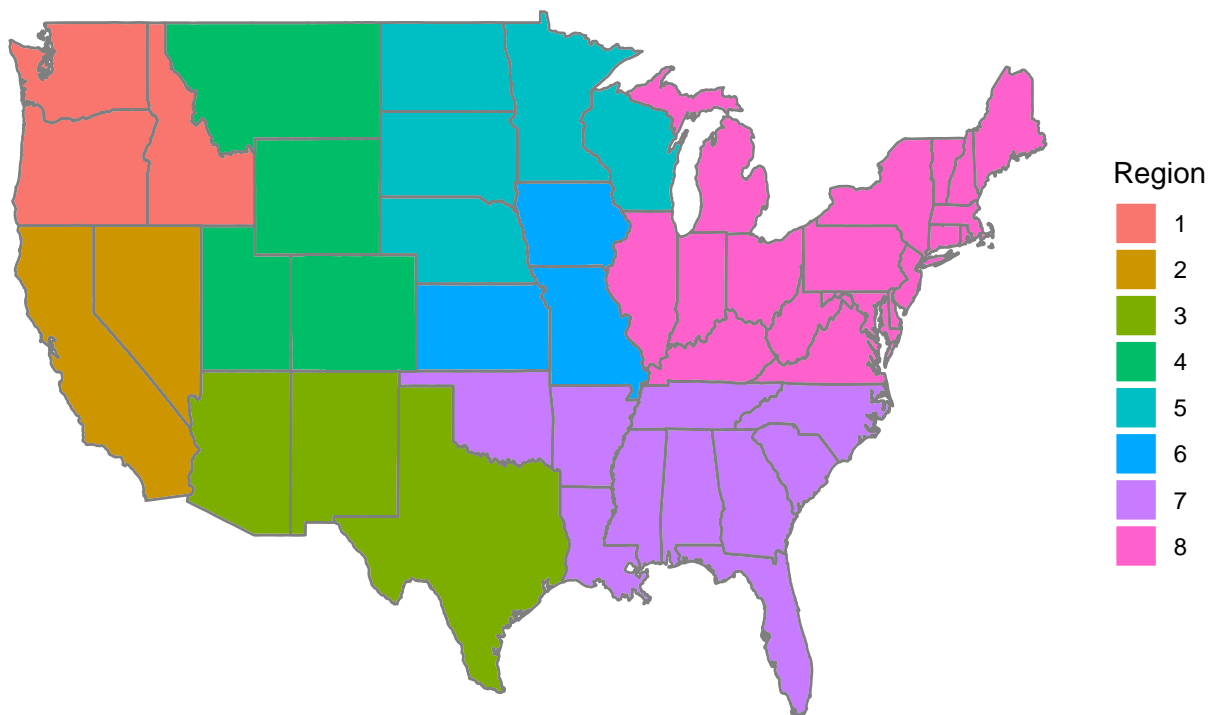


Figure 4
Number of Isolates Over Time

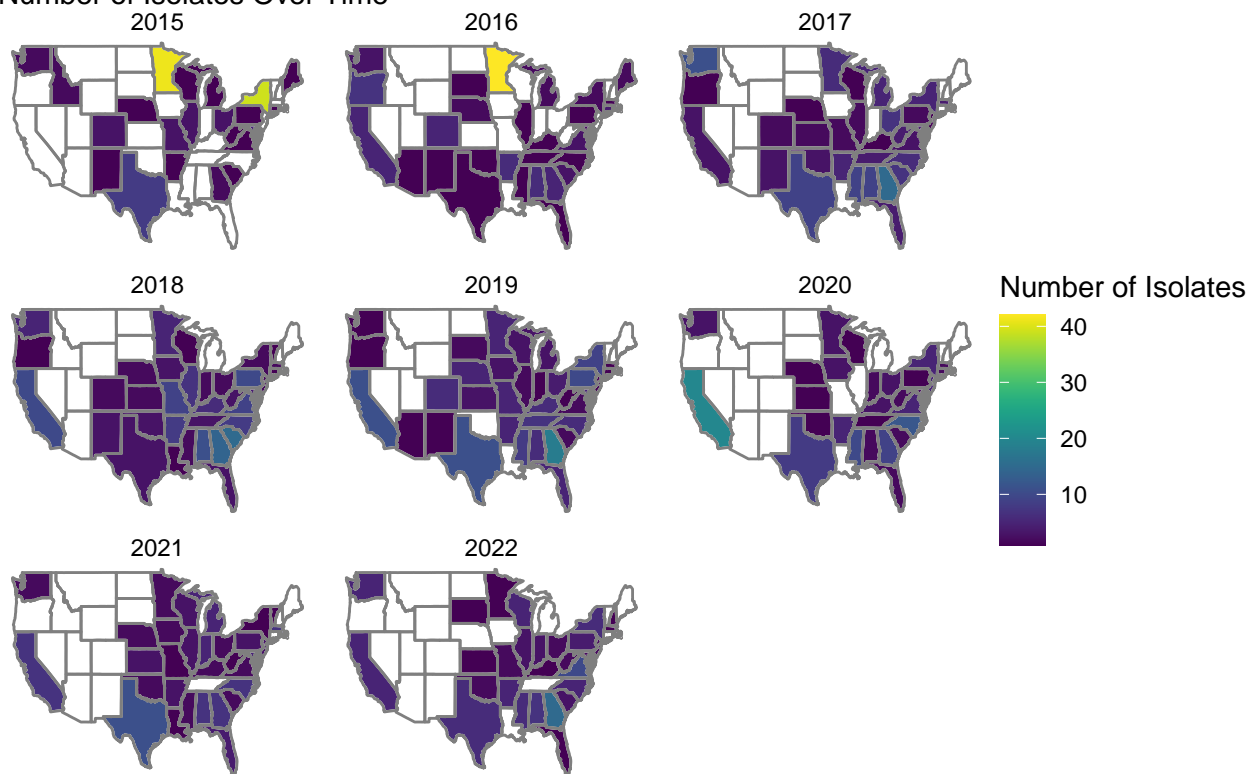
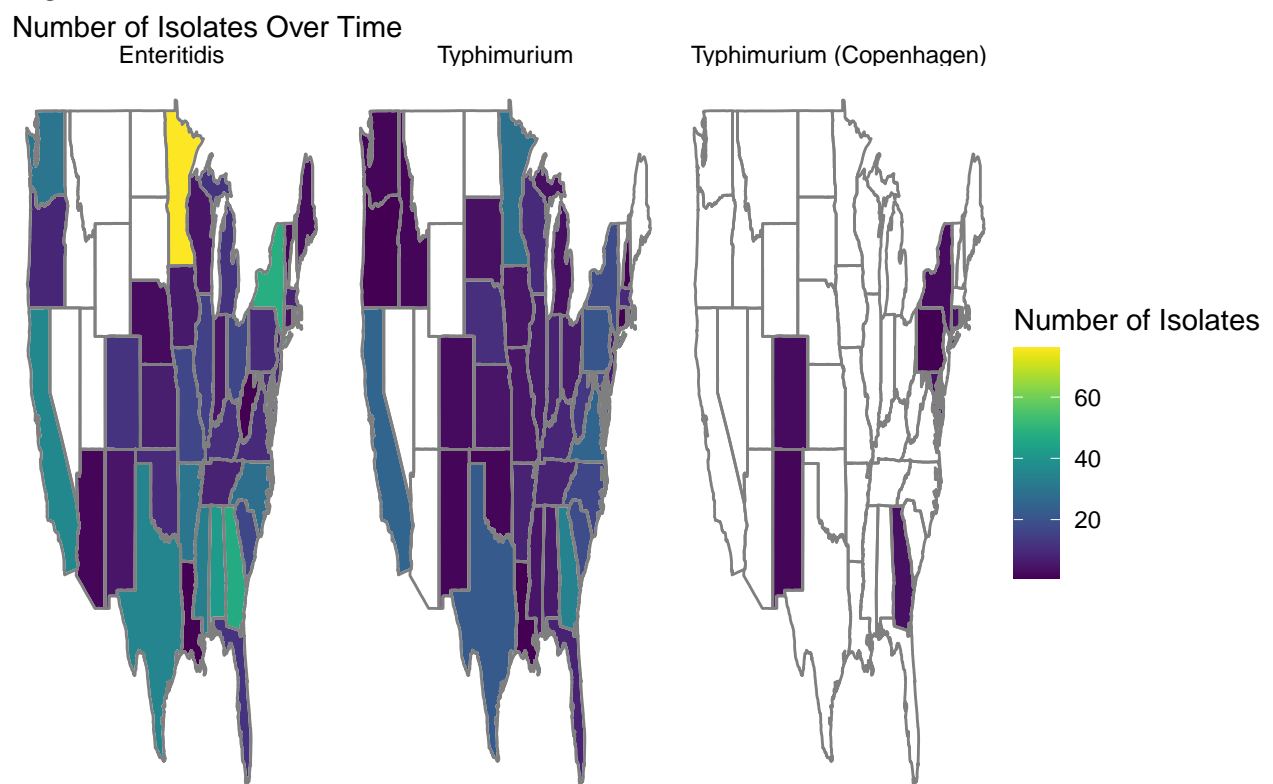


Figure 4

Number of Isolates Over Time



Interactive Visualization

shiny with above facet wrap as filter, if time allows try to add hover summary box

References

Appendix: All code for this report

```
knitr::opts_chunk$set(echo = FALSE)
library(knitr)
library(ggplot2)
library(dplyr)
require(maps)
require(viridis)

theme_set(theme_void())

isolates <- read.csv("isolates2.csv")

# Restrict to isolates that have a state identified
isolate_states <- isolates %>%
  filter(!state == "Not Specified") %>%
  rename(cdc_region = region) %>%
  rename(region = state) %>%
  filter(!(region == "Other"|region == "PR")) %>%
  filter(Collection.year==c(2015:2022))

isolate_states$region <- tolower(state.name[match(isolate_states$region,state.abb)])

sum(is.na(isolates$Lat.Lon))/nrow(isolates)
sum(!isolates$state=="Not Specified")/nrow(isolates)

states_map <- map_data("state")

print_map_states <-
  isolate_states %>%
  group_by(region) %>%
  summarize(n_isolates = n()) %>%
  left_join(states_map, by = "region") %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = n_isolates), color = "white") +
  scale_fill_viridis(option = "D", name = "Number of Isolates") +
  borders("state") +
  labs(title = "Figure 1", subtitle = "Number of Isolates by State")

print_map_states

# Create table of count of isolates by state
a <- isolate_states %>%
  group_by(region) %>%
  summarize(n_isolates = n())

# Create table of count of isolates by state
b <- isolate_states %>%
  group_by(cdc_region) %>%
  summarize(n_isolates = n())
```

```

print_map_cdc <-
  left_join(b, isolate_states, by = "cdc_region") %>%
  left_join(states_map, by = "region") %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = n_isolates), color = "white") +
  scale_fill_viridis(option = "D", name = "Number of Isolates") +
  borders("state") +
  labs(title = "Figure 2", subtitle = "Number of Isolates by Microbiological Monitoring Region")

print_map_cdc

bifscsco_states <- isolate_states %>%
  distinct(state, region) %>%
  filter(!(state == "Not Specified" | state == "Other" |
           state == "HI" | state == "AK" | state == "PR")) %>%
  rename(cdc_region = region) %>%
  rename(region = state)

bifscsco_states$region <- tolower(state.name[match(bifscsco_states$region, state.abb)])

print_bifscsco_map <- bifscsco_states %>%
  left_join(states_map, by = "region") %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = factor(cdc_region)), color = "white") +
  scale_fill_discrete(name = "Region") +
  borders("state") +
  labs(title = "Figure 3", subtitle = "Map of CDC Microbiological Monitoring Regions")

print_bifscsco_map

# By Collection.year
c <- isolate_states %>%
  group_by(region, Collection.year) %>%
  summarize(n_isolates = n())

print_map_year <- c %>%
  left_join(states_map, by = "region") %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = n_isolates), color = "white") +
  scale_fill_viridis(option = "D", name = "Number of Isolates") +
  borders("state") +
  labs(title = "Figure 4", subtitle = "Number of Isolates Over Time") +
  facet_wrap(vars(Collection.year))

print_map_year

# By Serovar
e <- isolate_states %>%
  group_by(region, Serovar) %>%
  summarize(n_isolates = n())

print_map_serovar <- e %>%

```



```
left_join(states_map, by = "region") %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = n_isolates), color = "white") +
  scale_fill_viridis(option = "D", name = "Number of Isolates") +
  borders("state") +
  labs(title = "Figure 4", subtitle = "Number of Isolates Over Time") +
  facet_wrap(vars(Serovar))

print_map_serovar
```