Lab 05

Adam Guerra

2022-10-26

```
#Step #1 Set global options and load packages

knitr::opts_chunk$set(echo = TRUE)                                    #set global options

library(tidyverse)                                                    #load tidyverse
library(here)                                                         #load here
library(janitor)                                                      #load janitor
library(readxl)                                                       #load readxl
library(sf)                                                           #load sf
library(USAboundaries)                                                #load USAboundaries
```

```
#Step 2: Import water use data (estimated, county-level, 2015, USGS; Clean data)

water_use <- read_excel(here("data/usco2015v2.0.xlsx"), sheet = 1, skip = 1) |> #load data
             clean_names() |>                                         #clean names
             mutate_at(vars(6:141), as.numeric, na.rm = TRUE)         #convert relevant data

data_dictionary <- read_excel(here("data/usco2015v2.0.xlsx"),         #load dictionary
                              sheet = 2) |>
                   clean_names() |>                                   #clean names
                   pivot_wider(names_from = "column_tag",             #pivot dictionary wide
                               values_from = "attribute") |>
                   clean_names() |>                                   #clean wide names
                   select(contains(c("to"))) |>                       #select data with "to"
                   pivot_longer(cols = 1:60, names_to = "column_tag",  #pivot back longer
                                values_to = "attributes")

centroids <- readRDS(here("data/state_centroids.rds"))                #load centroid data
```

```
#Step 3: Organize data for goal 1: water use by sector by state, highest to lowest withdrawals
fresh_sectors <- select(water_use, state, county, contains(c("w_fr_to"))) |>    #select fresh data
                 mutate(domestic = ps_w_fr_to + do_w_fr_to,                     #add domestic column
                        industrial = in_w_fr_to,                                #add industrial column
                        agricultural = ir_w_fr_to + li_w_fr_to + aq_w_fr_to,    #add agricultural colum
                        mining = mi_w_fr_to,                                    #add mining column
                        thermoelectric = pt_w_fr_to,                            #add thermoelectric col
                        total = to_w_fr_to,                                     #add total column
                        total_check = domestic + industrial + agricultural +    #add total_check column
                         mining + thermoelectric) |>
                 select(state, county, domestic, industrial, agricultural,      #select new columns
                        mining, thermoelectric) |>
                 group_by(state) |>                                             #group by state
                 summarize_at(2:6, sum) |>                                      #summarize
```

```
                ungroup() |>                                                    #ungroup
                filter(!state %in% c("DC", "VI", "PR")) |>                      #filter out non-states
                pivot_longer(2:6, names_to = "sector",                          #pivot longer
                             values_to = "withdrawals") |>
                inner_join(centroids, by = "state")                             #join centroids data
```

```
#Step 4: Plot goal 1

ggplot() +                                                                      #make ggplot
  geom_col(data = fresh_sectors, aes(x = reorder(state, -withdrawals),          #bar plot, set x/y/fill
                                     y = withdrawals,
                                     fill = sector),
           position = "stack", width = 0.5) +                                   #stack bars
  labs(x = "State", y = "Withdrawals (Mgal/day)", caption = "Figure 1: State withdrawals
       by sector. Data from USGS (2015). Created by Adam Guerra ",              #set labels
       fill = "") +
  scale_y_continuous(limits = c(0, 30000),                                      #set scale continuous
 expand = c(0,0)) +                                                             #expand
theme_bw() +                                                                    #add theme
theme(axis.text = element_text(color = "black", size = 6),                      #edit axis text color
 axis.title.x = element_text(color = "black", size = 10, face = "bold"),        #change x axis title co
 axis.title.y = element_text(color = "black", size = 10, face = "bold"),        #change y axis text col
 plot.caption = element_text(hjust = 0, color = "black", face = "bold"),        #change plot caption te
 legend.title = element_text(size = 10),                                        #change legend title te
 legend.position = "top") +                                                     #change legend position
  scale_fill_manual(values = c("green", "blue", "red", "grey", "orange"),       #change fill colors
                    labels = c("Agriculture", "Domestic", "Industrial",         #capitalize legend labe
                               "Mining", "Thermoelectric"))
```

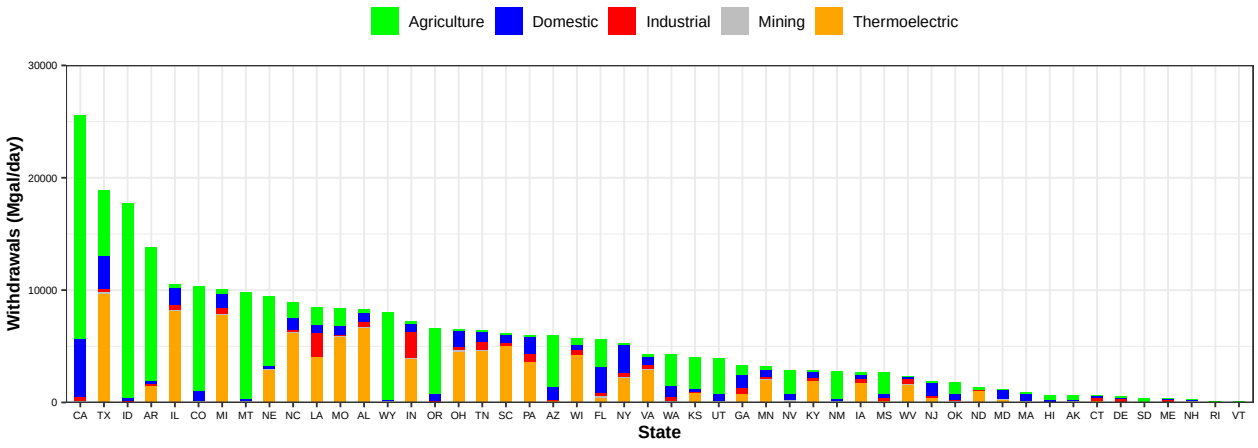


Figure 1: State withdrawals by sector. Data from USGS (2015). Created by Adam Guerra

```
#Step 5: Organize data for goal 2: total sw + gw by state, highest to lowest withdrawal volume

sources <- select(water_use, state, contains(c(groundwater = "to_wgw_fr",       #creates sources from w
                                               surface_water = "to_wsw_fr"))) |>#selecting ground ans s
           group_by(state) |>                                                   #group by state
           summarize_at(1:2, sum) |>                                            #sum columns
           ungroup() |>                                                         #ungroup
           filter(!state %in% c("DC", "VI", "PR")) |>                           #filter out non-states
           pivot_longer(2:3, names_to = "source", values_to = "withdrawals")    #pivot longer
```

```r
#Step 6: Plot goal 2

ggplot() +                                                                    #making ggplot
  geom_col(data = sources, aes(x = reorder(state, withdrawals),               #column plot, setting x
                               y = withdrawals, fill = source),
           position = "stack", width = 0.5) +                                 #adjusting position and
  coord_flip() +                                                              #coord flip
  labs( x = "State", y = "Freshwater Withdrawals (Mgal/day)",                 #label x and y
        caption = "Figure 2: Withdrawals by source. Data from USGS (2015).
        Created by Adam Guerra", fill = "") +                                 #label caption
  scale_y_continuous(limits = c(0, 30000), expand = c(0,0)) +                 #scale continuous
theme_bw() +                                                                  #set theme
theme(axis.text = element_text(color = "black", size = 6),                    #set axis text color
 axis.title.x = element_text(color = "black", size = 10, face = "bold"),      #set x axis text color
 axis.title.y = element_text(color = "black", size = 10, face = "bold"),      #set y axis text color
 plot.caption = element_text(hjust = 0, color = "black", face = "bold"),      #set plot caption text
 legend.title = element_text(size = 10),                                      #set legend text color
 legend.position = "top",                                                     #set legend position
 panel.background = element_blank(),                                          #set panel background c
 panel.grid.major = element_line(size = 0.25, linetype = 'solid',             #set element line size
 colour = "lightgrey"),                                                       #set grid major color
 panel.grid.minor = element_line(size = 0.25, linetype = 'dotted',            #set grid minor size
 colour = "lightgrey"),                                                       #set grid minor color
 axis.ticks.x=element_blank(),                                                #set x ticks blank
 axis.ticks.y=element_blank()) +                                              #set y ticks blank
  scale_fill_manual(values = c("light blue", "dark blue"),                    #set columns color
                    labels = c("Groundwater", "Surface Water"))               #set label with capital
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
```

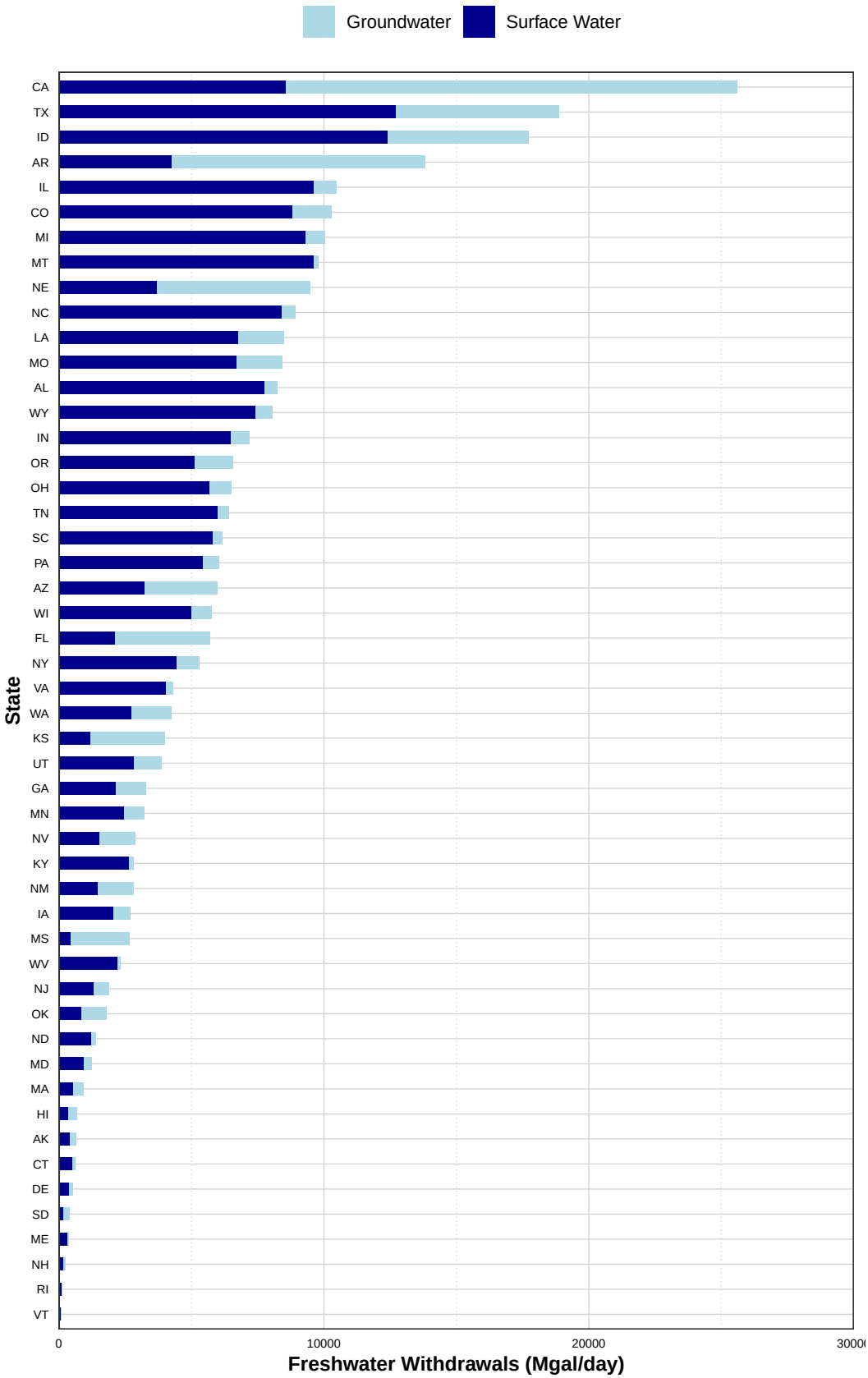


Figure 2: Withdrawals by source. Data from USGS (2015). Created by Adam Guerra

Q1: What is the take home message specific to Figure 1?
That agriculture and thermoelectric are the highest users of water across all states. High population states also have larger domestic uses, but not close to thermoelectric or agriculture. Also the west coast states tend to have higher agriculture and as you move east it transitions to more thermoelectric.

Q2: What is the take home message specific to Figure 2?
Surface water is used by a vast majority of states more heavily than ground water. With the exception of some outlines that have access to large aquifers, mainly on the west side of the United States.

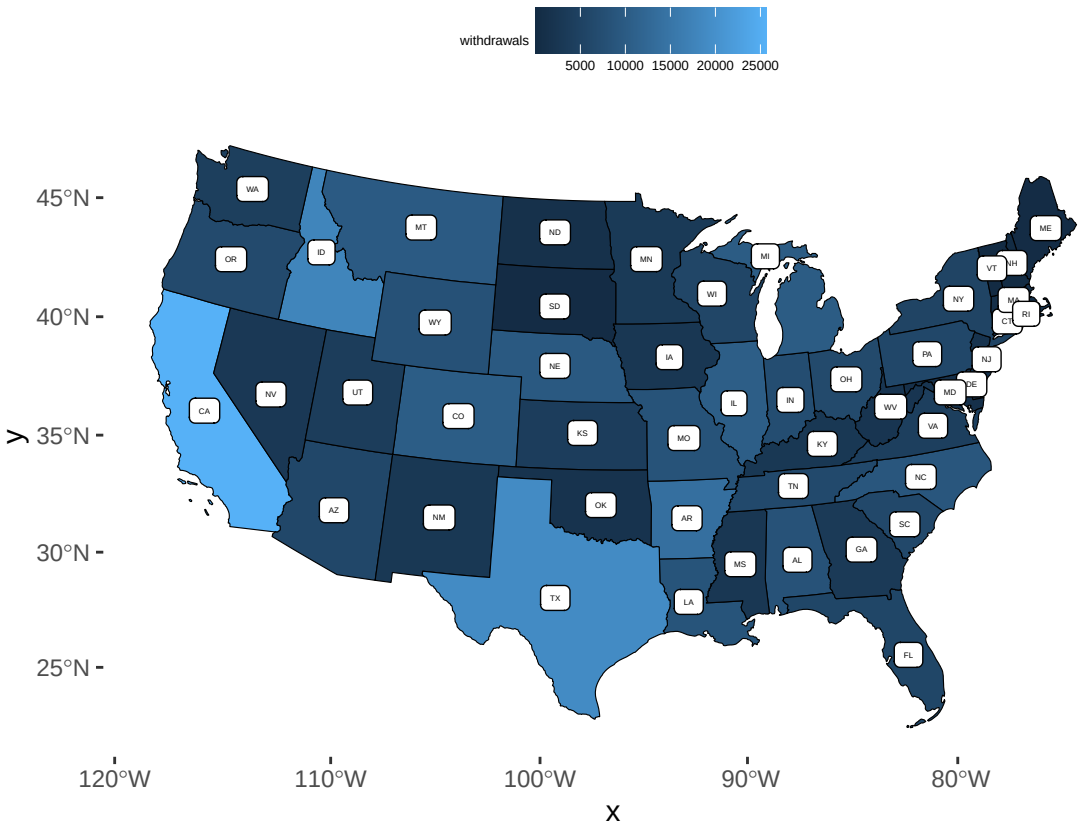
Q3: What is the most surprising outcome to you of Figure 1 or Figure 2?
The most surpising outcome on both figures for me was that Arkansas had such a high water usage. Especially with their higher ground water usage and agriculture uses while be located on the eastern side of the US.

Q4: These water use data (i.e., water_use) can be used to answer many other research questions. Identify at least one question that you would like to answer using these data.
Do states with a higher population tend to use more water per capita than states with a lower population?

Q5: Science communication is a critical part of doing good science. Identify two changes you would make to either plot to promote how the public engages with the materials—that is, identify two changes to make the plot prettier!
I think have some minor grid lines on either plot would help people draw better comparisons across the bars. Also some extra values on the y axis could help people understand the magnitude of the data rather than just seeing two large numbers.

```
#Step 7: Geo spatial Plot data setup

conus <- us_states() |>
         filter(!state_name %in% c("Alaska", "Hawaii", "Puerto Rico")) |>
         st_transform(crs = 5070) |>
         select("state_name", "state_abbr")

centroids_spatial <- centroids |>
                     st_as_sf(coords = c("lng","lat"), crs = 4326)

total_map <- water_use |>
             select(state_abbr = "state", withdrawals = "to_w_fr_to") |>
             filter(!state_abbr %in% c("PR", "VI", "DC", "HI", "AK")) |>
             group_by(state_abbr) |>
             summarize_at(1, sum) |>
             ungroup() |>
             inner_join(conus, by = "state_abbr") |>
             st_as_sf(sf_column_name = "geometry")
```

```
#Step 8: plot geo spatial data

ggplot() +
  geom_sf(data = total_map, aes(fill = withdrawals), col = "black",
          size = 0.1) +
  geom_sf_label(data = total_map, aes(label = state_abbr), size = 1) +
  theme(legend.position = "top", panel.background = element_blank(),
        legend.title = element_text(size = 5), legend.text = element_text(size = 5))
```




```
  scale_fill_gradient()
```

```
## <ScaleContinuous>
##  Range:
##  Limits:    0 --    1
```