Seasonality indices: Environmental applications

ENS-215

7-Mar-2022

Seasonality indices

Many environmental processes/conditions (e.g., precipitation, temperature, vegetation abundance,...) experience distinct seasonal cycles. Understanding the characteristics/dynamics of these seasonal cycles can shed light on the underlying processes that control these cycles as well as provide useful information for understanding their environmental and societal implications. For example, seasonal variations in precipitation can have many important implications ranging from controlling the timing of floods and droughts, influencing the seasonal growth of vegetation/crops, affecting the transmission of waterborne diseases, and impacting the emergence of insect populations and associated diseases (e.g., Malaria).

In many of our classes thus far, we've explored data that have distinct seasonal dynamics and we've seen a few approaches to help visualize these cycles. In today's class were are going to take this a step further and learn some quantitative approaches to characterizing seasonal cycles.

First let's take a look at monthly precipitation for Boston and San Francisco to highlight some features of seasonal cycles.

Mean monthly precipitation (2011–2021)

Data source: NOAA GSOD
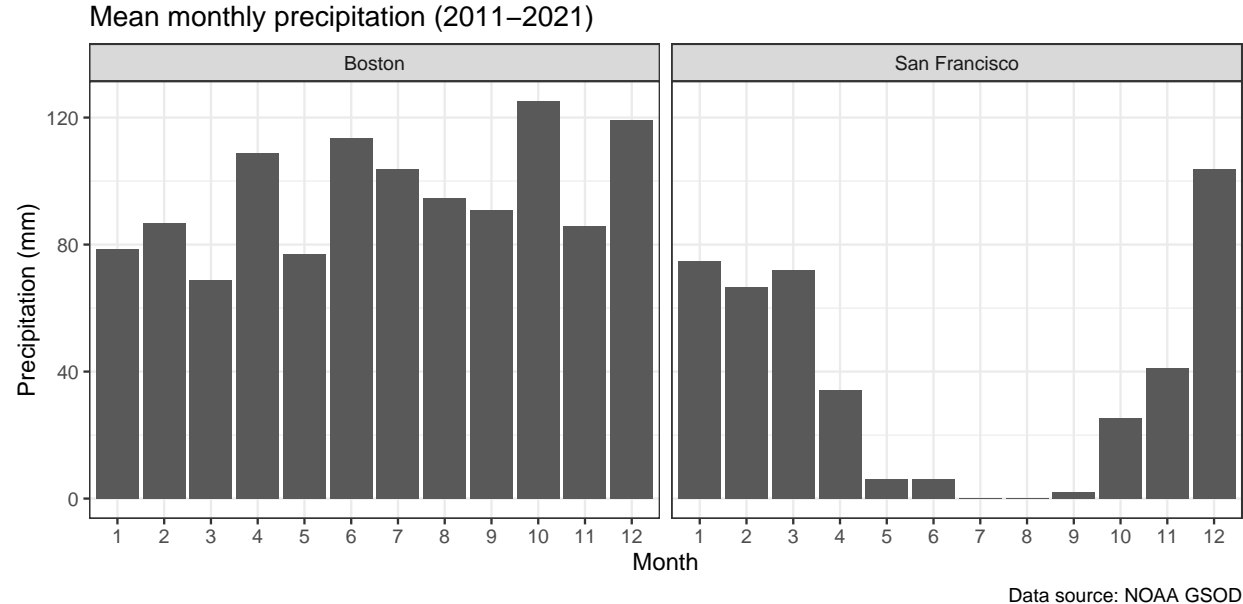
What can we say about the seasonal cycles of precipitation in Boston and In San Francisco? How would you describe these cycles? What features stand out? How do the cycles between the two locations compare and contrast? Can you think of objective metrics that would quantitatively capture the features you observe?

To better understand both the drivers of seasonal cycles and their implications it is important to develop objective, quantitative metrics that summarize seasonal cycles. While there are many possible metrics that can be used - each having different advantages/disadvantages - we will explore two particularly useful metrics

that capture the seasonal timing and seasonal concentration of time-series data.

The metrics we will use were first described in a paper by Markham (1970). These metrics have since been applied to many subsequent studies. Furthermore, many other studies have developed very similar metrics that are also widely used.

The Seasonality Index (SI) describes how uniformly a given variable is spread across the months. The SI ranges from 0 to 1, with a value of 0 indicating that the variable is evenly spread across the months and a value of 1 indicating that the variable is concentration within a single month.

The Seasonal Concentration (SC) describes the time of year that the values are most concentrated. The SC is represented by an angular direction ranging from 0 to 360, thus a value of 0 would indicate the the variable is most focused in the month of January while a value of 180 would indicate the variable being most focused in June. A bit further down I will introduce the mathematical approach to calculating SI and SC, though first let's continue to look at the Boston and San Francisco data to get a better conceptual understanding of these metrics.

Below we are plotting the monthly precipitation where each month's precipitation is represented by a vector. the length of the vector is proportional to the precipitation in each month. The angle of the vector is determined by the month (e.g., picture a clock where the hours represent months - thus January/December would point upwards and then we would move clockwise month by month).

We can see in the figure below that Boston has precipitation that is relatively evenly spread among the 12 months, whereas San Francisco has most of its precipitation in the winter months (long vectors pointing upwards) and very little precipitation in the other months (very short vectors for those months).
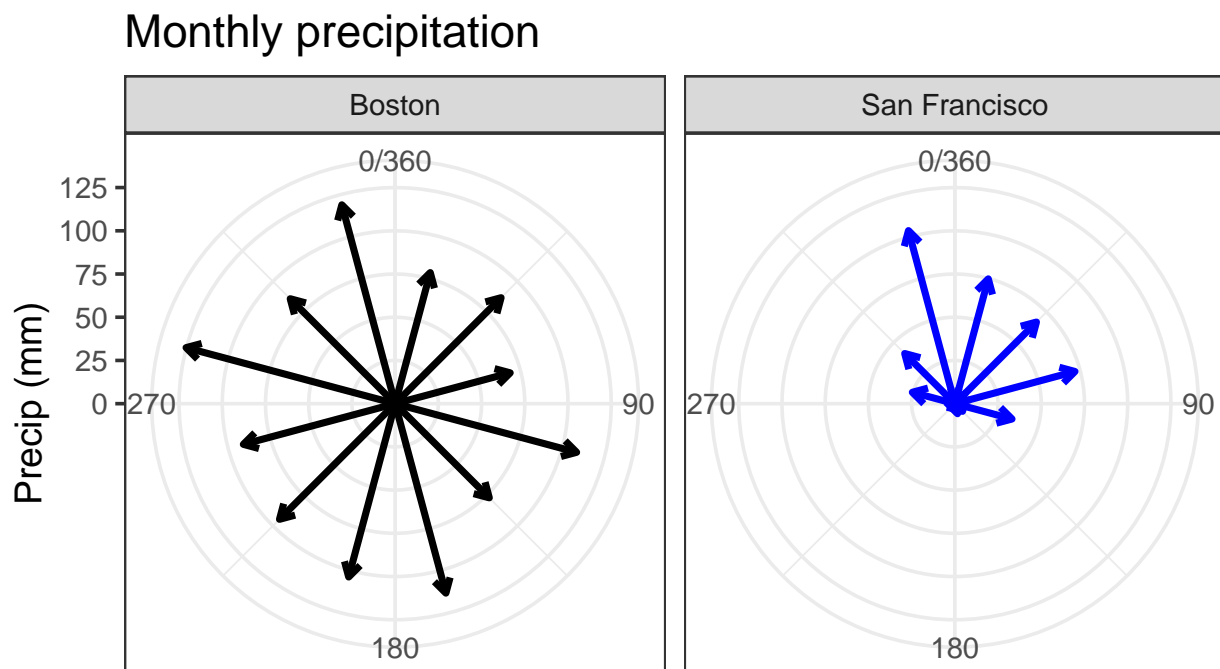
Now let's take those vectors from the figure above and put them end to end (making sure to keep the orientation of each monthly vector the same as in the figure above). When we do that we get a vector (red vector) whose angle represents the seasonal concentration of precipitation (i.e., which season precipitation is

most focused in).
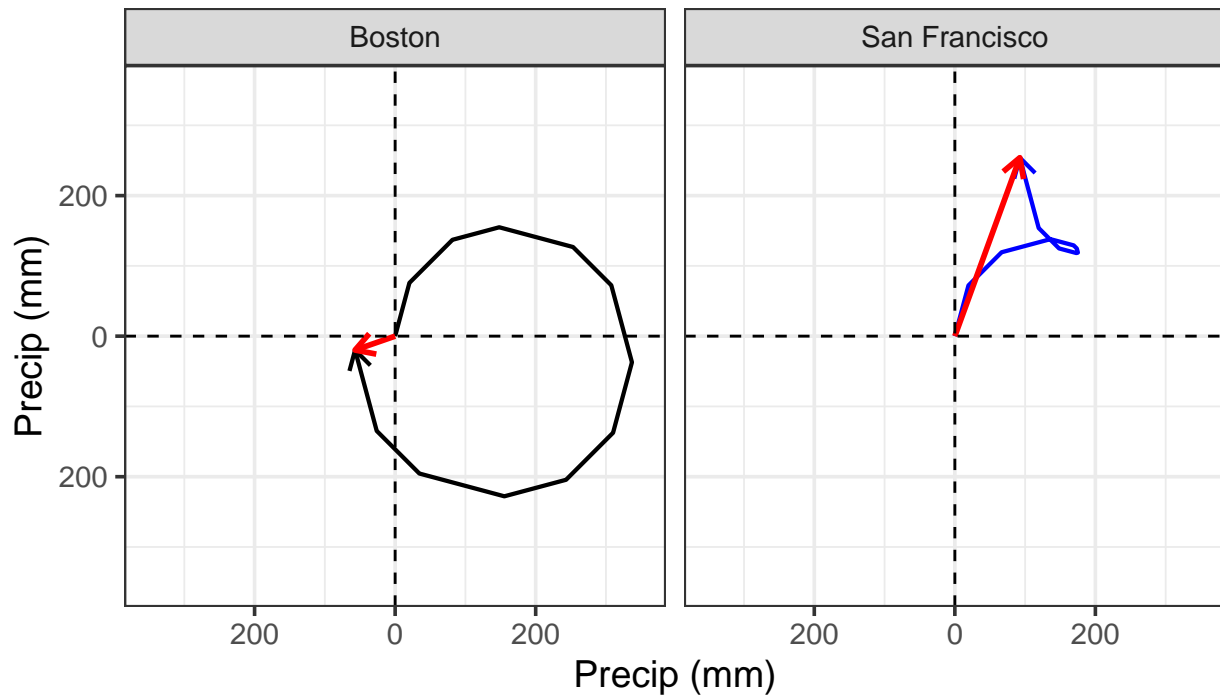
The length of the red vector when compared against the total annual precipitation (i.e., numeric sum of all 12 months precipitation) indicates how uniform or non-uniformly distributed precipitation. The bar chart below shows the length of the vector sum (i.e., red vector from above) to the total annual precipitation. In the bar chart the red bar represents the vector sum and the grey bar represents the total annual precipitation.

You can imagine that if a location had all of its precipitation fall in a single month of the year, then the red vector would be identical in length to the total annual precipitation. Conversely if you had a location with all months having exactly equal precipitation, then the red vector would have a length of zero and thus would be tiny compared to the total annual precipitation.

Based on the figure below it is clear that for Boston the red bar is only a small proportion of the grey bar, thus indicating that precipitation in Boston is relatively evenly distributed throughout the year. Whereas in San Francisco the red bar is similar in magnitude to the total annual precipitation (grey bar) indicating that precipitation is not uniformly distributed (i.e., occurs in a narrow window of months).
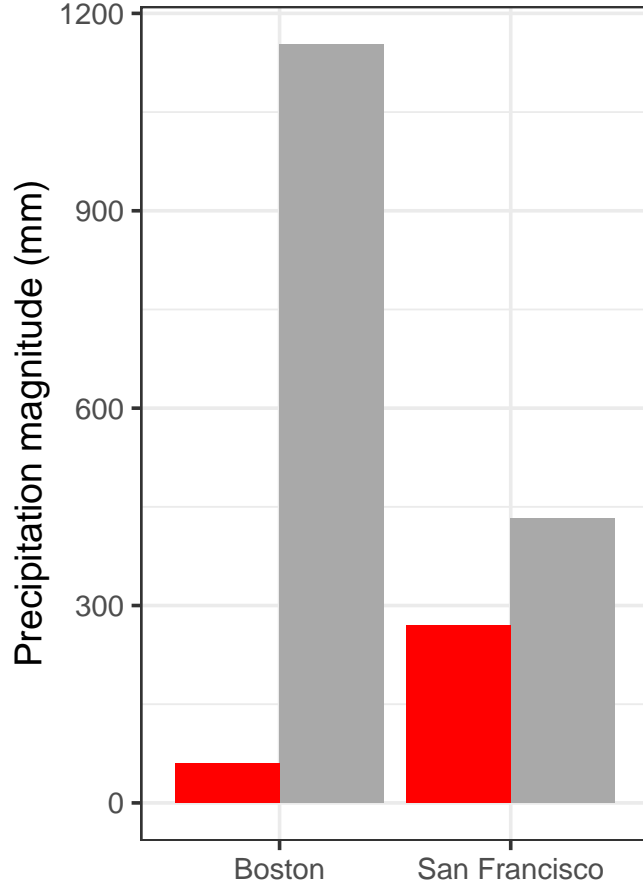
I am now going to introduce the mathematical implementation of the seasonality metrics from Markham (1970): Seasonality Index (SI) and Seasonal Concentration (SC).

$$SI = \frac{(V_x^2 + V_y^2)^2}{V_{tot}}$$

$$SC = tan^{-1}(\frac{V_y}{V_x}) * \frac{180}{\pi} + a$$

$$a = \begin{cases} 180, & \text{when } V_x \leq 0 \\ 0, & \text{when } V_x > 0 \text{ and } V_y > 0 \\ 360, & \text{when } V_x > 0 \text{ and } V_y \leq 0 \end{cases}$$

Note that V_x represents the x directional component of monthly precipitation, where $V_x = \sum V_i * cos(\theta_i)$; V_y is the y directional component of monthly precipitation, where $V_y = \sum V_i * cos(\theta_i)$; V_{tot} is the sum of monthly precipitation, where $V_{tot} = \sum V_i$; V_i is the monthly precipitation for month i and θ_i is the monthly angle for month i (i.e., Jan = 15 degrees, Feb = 45 degrees,...).
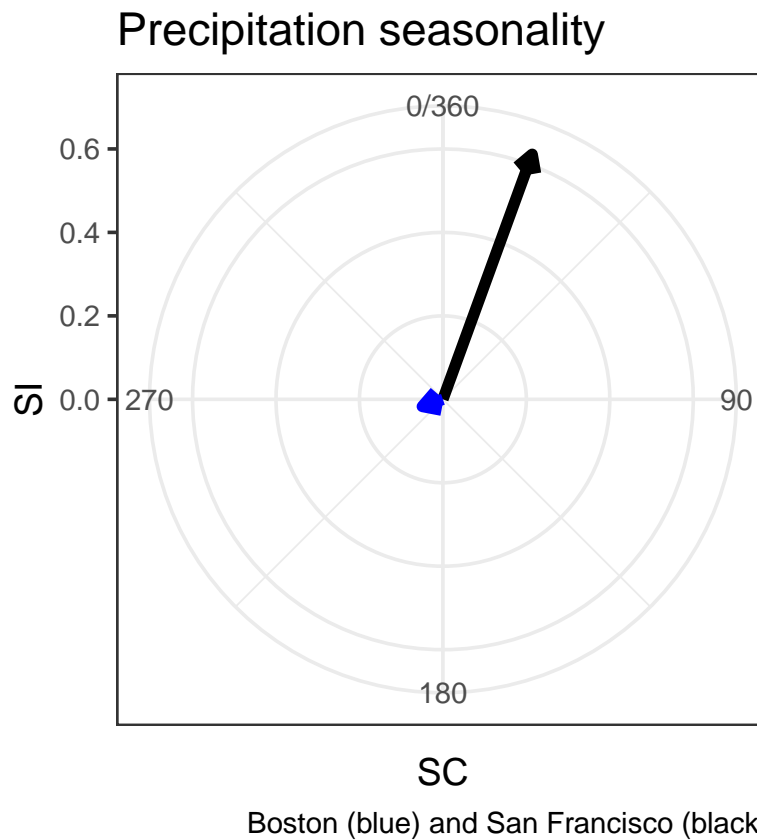
I've computed the seasonality indices (SI and SC) for the Boston and San Francisco precipitation data

We can see that San Francisco has a relatively large Seasonality Index (SI) - indicating that precipitation is highly focused in a few months of the year. Boston, however, has an SI close to zero, indicating that precipitation there is relatively uniformly distributed across all months.

Location	SI	SC
San Francisco	0.62	20
Boston	0.05	251

The Seasonal Concentration (SC) indicates the time of year that the precipitation is most focused. For San Francisco we see that $SC = 20$, indicating precipitation is focused around January. Boston has a $SC = 251$ indicating precipitation most focused around August. However, since Boston has its precipitation uniformly distributed throughout the year (i.e., SI is low), then the SC value is not as informative as it is for San Francisco (which has a large SI value).

Below I've presented the results of the two seasonality indices graphically. Note that the angle of the arrow represents SC and the length of the arrow represents SI. Boston is shown in blue and San Francisco in black.

Precipitation seasonality

Boston (blue) and San Francisco (black)

Precipitation seasonality across the US

Now we are going to compute seasonality indices (SI and SC) for each US state using the method introduced above. First let's load in monthly precipitation data for all of the states in the US (excluding AK and HI). You'll recall working with this data, which comes from NOAA, in some prior class exercises.

```
precip_data <- read_csv("https://stahlm.github.io/ENS_215/Data/NOAA_State_Precip_LabData.csv")
```

Take a look at the dataset to refresh your memory of what it contains. You'll see that it has monthly values for precipitation for each state and each year from 1895 through 2017.

Using just the data from 1980 onward, let's compute the average precipitation in each month for each state.

```
df_monthly_normals <- precip_data %>%
  filter(Year >= 1980) %>%

  group_by(state_cd, Month) %>%
  summarize(observation_value = mean(Precip_inches)) %>%

  rename(group_name = state_cd)
```

Take a look at the data frame you just created, which contains monthly precipitation normals for each state. Make sure you understand what is in the data frame and that everything looks sensible.

Since we want to compute the seasonality indices SI and SC on this data we are going to need some code to do that. I've written a function in R to compute these indices. Let's go ahead a load in this function from our class github site.

```
source("https://raw.githubusercontent.com/stahlm/stahlm.github.io/master/ENS_215/Winter_2022/Lectures/c
```

You'll now see in your environment pane that you have a function called `compute_seasonality_indices()`. This is the function we will use to compute the SI and SC indices. If you click on the function in your environment pane the source code will pop up. Go ahead and do this. Make sure to read the comments within the function file as these comments explain what the function does, the inputs its requires, and the outputs it generates.

FYI, if you want you can also download the R file from my github site.

Now let's use the `compute_seasonality_indices()` function to compute the indices.

```
df_seasonality_precip <- compute_seasonality_indices(df_monthly_normals)
```

The `compute_seasonality_indices()` function returns a list containing two data frames. You will just need the 2nd data frame. Let's use the code below to select just the 2nd data frame which has the results we will need.

```
df_seasonality_precip <- df_seasonality_precip[[2]]
```

Take a look at the `df_seasonality_precip` data frame to make sure you understand what it contains.

You can see that we now have a data frame that has precipitation seasonality indices SI and SC for each US state! Let's create a figure to examine these results.

```
df_seasonality_precip %>%
  ggplot(aes(y = reorder(group_name, SI), x = SI)) +

  geom_point(size = 2) +
  geom_segment(aes(x = 0, xend = SI, yend = group_name)) +

  theme_bw(base_size = 12) +
  labs(x = "SI", y = "State")
```
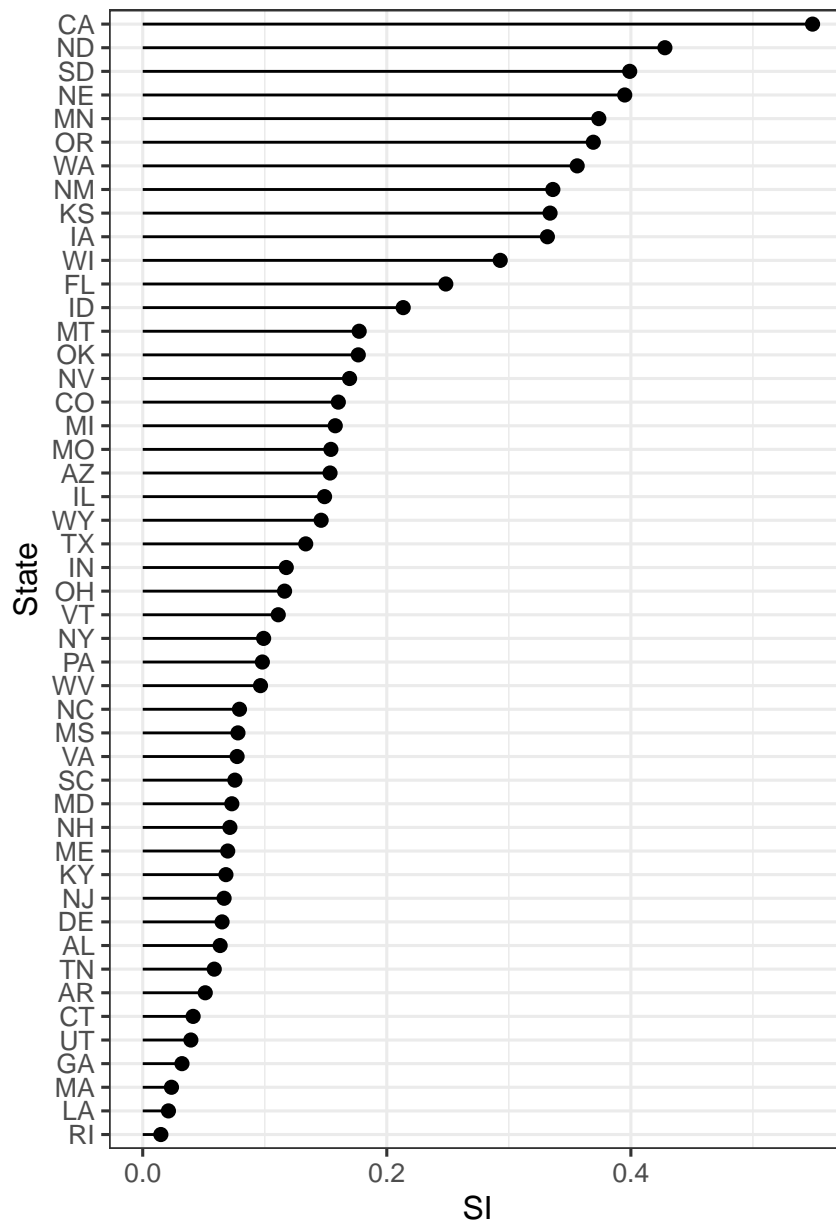
Take some time to look over these results.

- What key patterns/features do you observe?

- What might explain these features (e.g., what do you think is responsible for the wide range of SI values between states)?

- Discuss with your neighbors and think about some additional analysis you could do to help answer any hypotheses you've come up with.

To get a better sense of what the monthly precipitation looks like for different SI values, let's examine the monthly precipitation for California (high SI value), Idaho (intermediate SI value), and Rhode Island (low SI value).
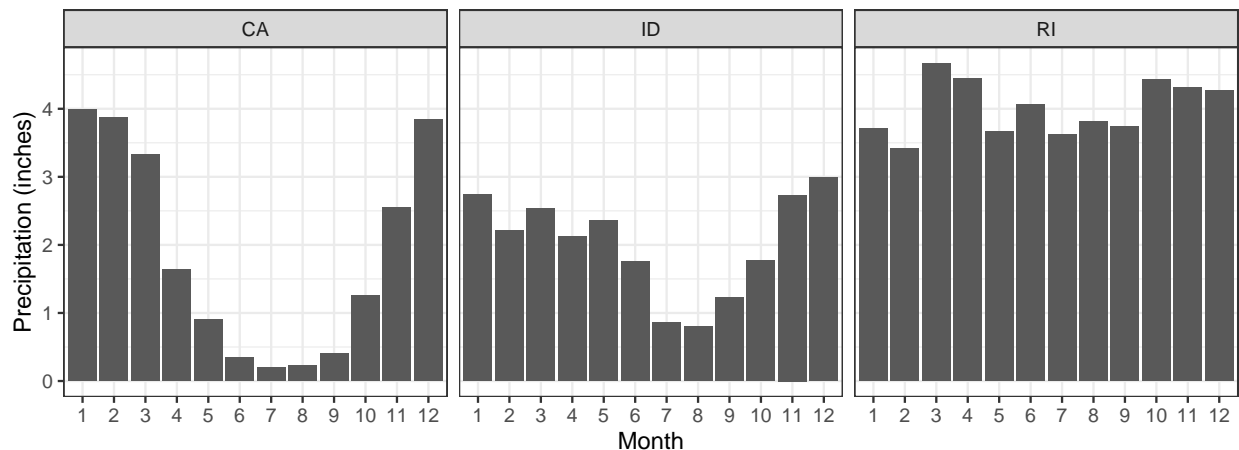
```
df_monthly_normals %>%
  filter(group_name %in% c("RI", "CA", "ID")) %>%

  ggplot() +
  geom_col(aes(x = factor(Month), y = observation_value)) +

  facet_wrap(~ group_name) +

  theme_bw() +

  labs(x = "Month", y = "Precipitation (inches)")
```


Since we've calculated seasonality metrics for precipitation for each US state, let's go ahead and make a map to help us identify if there is any spatial structure/patterns to the data.

We are going to need to join our `df_seasonality_precip` data frame to a sf data frame containing borders for each state. First let's add a column with state names to our `df_seasonality_precip` data frame.

```
state_dictionary <- tibble(NAME = state.name, state_cd = state.abb)
```

```
df_seasonality_precip <- df_seasonality_precip %>%
  rename(state_cd = group_name) %>%
  left_join(state_dictionary)
```

```
## Joining, by = "state_cd"
```

Now let's load in US state borders and then join this data to our `df_seasonality_precip` data frame.

```
library(sf)
library(tmap)
```

```
sf_states <- spData::us_states
```

```
sf_seasonality_precip <- sf_states %>%
  right_join(df_seasonality_precip)
```

```
## Joining, by = "NAME"
```

Now that our seasonality metrics data frame has the state borders associated with it, we can go ahead and make a map. Let's make a map to see how uniformly or non-uniformly precipitation is distributed across months in each US state. To do this we will color each state according to its SI value.

```
sf_seasonality_precip %>%
  tm_shape(bbox = st_bbox(sf_states), projection = 5070) +
  tm_polygons(col = "SI", style = "cont", palette = "viridis")+

  tm_shape(sf_states) +
  tm_borders()
```
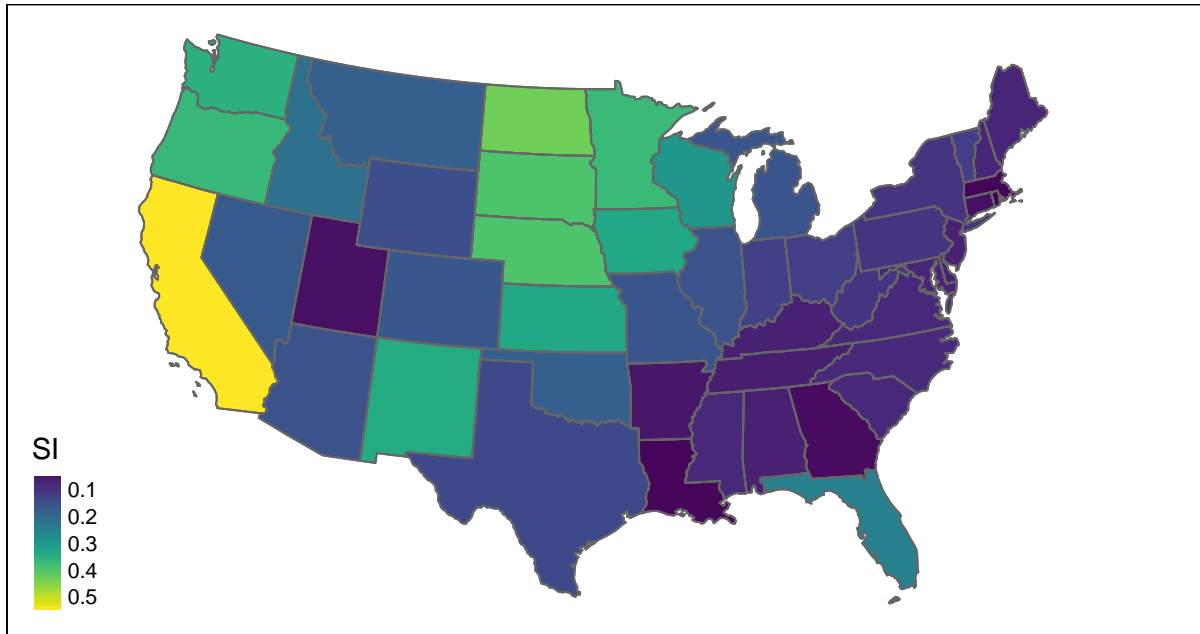

- Do you notice any interesting spatial patterns in the SI data?

- With your neighbors discuss what you think might influence the seasonal distribution of precipitation.

Precipitation and streamflow seasonality: Mohawk Watershed

Now that we've gone over the theory behind seasonality indices and applied these indices to monthly precipitation for each US state, let's delve further into the topic. In particular, let's examine seasonality of precipitation and streamflow for the Mohawk River watershed - this will help us to better characterize and understand the hydrologic processes that influence streamflow in our local watershed.

As you are aware, the water that ultimately makes its way into the Mohawk River originally must have fallen as precipitation within the watershed. However, precipitation that falls does not necessarily make its way to a river/stream immediately - for instance precipitation that falls as snow in the winter may sit on the ground for weeks to months before melting and contributing to streamflow. Furthermore due to seasonal differences in evapotranspiration, a unit of precipitation that comes down in the middle of the summer may generate much less streamflow than a unit of precipitation that comes down in late fall. Thus, the seasonal cycle of precipitation might not be the same as the seasonal cycle of streamflow and by comparing and contrasting these two cycles we can learn more about the underlying hydrologic/environmental processes.

Let's go ahead and download streamflow data for the Mohawk River at Freeman's Bridge in Schenectady (USGS 01354500). To do so we'll use the USGS `dataRetrieval` package, which we've used in past classes.

```r
library(dataRetrieval)
library(lubridate)
```

```r
df_stream_data <- readNWISdv(siteNumbers = "01354500",
                             parameterCd = "00060",
                             statCd = "00003") %>%
  renameNWISColumns()
```

```r
# Add columns with year, month, and day information

df_stream_data <- df_stream_data %>%
  mutate(Year = year(Date),
         Month = month(Date),
         Day = day(Date)
         )
```

Let's select data for 2012-2021 and then compute the mean flow by month (i.e., Jan mean flow, Feb mean flow,..., Dec mean flow) as our function that computes the seasonality indices requires monthly values as inputs. Once you've computed these means take a look at the resulting data frame to confirm that everything makes sense to you and looks reasonable.

```r
df_stream_monthly_normals <- df_stream_data %>%
  filter(Year > 2011 & Year < 2022) %>%

  group_by(Month, site_no) %>%
  summarize(observation_value = mean(Flow, na.rm = T)) %>%

  rename(group_name = site_no)
```

Now that we have the streamflow data, let's use our `compute_seasonality_indices()` function to calculate the seasonality indices.

```r
df_seasonality_flow <- compute_seasonality_indices(df_stream_monthly_normals)

df_seasonality_flow <- df_seasonality_flow[[2]]
```

- Take a look at the seasonality indices for the Mohawk River. Think about these values, what they mean/imply, and discuss with your neighbors.

Now that we have calculated the SI and SC seasonality indices on the Mohawk River streamflow data, let's do the same for local precipitation data.

We are going to obtain precipitation data from Albany International Airport (accessed through NOAA's Global Summary of the Day) using the `GSODR` package that we have used in prior classes.

```r
library(GSODR)
```

```r
# Download data for 2012-2021

met_df <- get_GSOD(years = 2012:2021, station = "725180-14735")
```

You'll see in the `met_df` data frame that we have daily meteorological data for Albany International Airport for 2012-2021. Let's go ahead and compute the monthly means so that we can use our `compute_seasonality_indices()` function to calculate SI and SC on the precipitation data.

obs_type	SI	SC
Precip	0.16	214
Flow	0.24	65

```r
df_alb_precip_monthly_normals <- met_df %>%
  group_by(YEAR, MONTH, STNID) %>%

  summarize(observation_value = sum(PRCP, na.rm = T)) %>%

  ungroup() %>%
  group_by(MONTH, STNID) %>%
  summarise(observation_value = mean(observation_value)) %>%

  rename(group_name = STNID, Month = MONTH)
```

```r
df_seasonality_alb_precip <- compute_seasonality_indices(df_alb_precip_monthly_normals)

df_seasonality_alb_precip <- df_seasonality_alb_precip[[2]]
```

Now that we have calculated SI and SC on the Mohawk River flow data and on the Albany precipitation data, let's combine the two data frames with this info into a single data frame.

```r
df_seasonality_alb_precip$obs_type <- "Precip"

df_seasonality_flow$obs_type <- "Flow"

df_seasonality <- bind_rows(df_seasonality_alb_precip, df_seasonality_flow)
```

Your results should look like this.

```r
df_seasonality %>%
  mutate(group_name = if_else(obs_type == "Flow", "Mohawk River flow", "Precipitation")) %>%
  mutate(SI = round(SI, 2),
         SC = round(SC,0)
         ) %>%
  select(obs_type, SI, SC) %>%

  kableExtra::kable() %>%
  kableExtra::  kable_styling(bootstrap_options = c("striped", "hover", "condensed"))
```

Now let's make a single plot that presents the SI and SC data for the Mohawk River streamflow and the local precipitation.

```r
df_seasonality %>%
  ggplot(aes(x = SC, y = SI, color = obs_type)) +

  geom_segment(aes(y = 0, xend = SC, yend = SI),
               size = 2, arrow = arrow(length = unit(0.03, "npc")) ) +

  scale_color_manual(values = c("Black","Blue")) +

  scale_x_continuous(breaks = c(0,90,180,270,360), limits = c(0,360)) +
  coord_polar() +

  theme_bw(base_size = 14) +
```
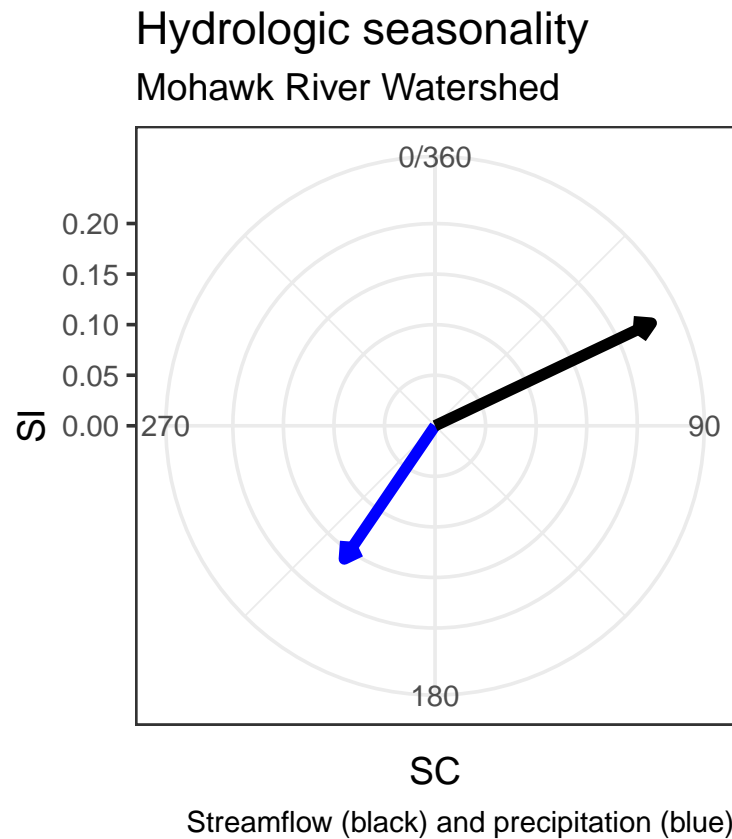
```
theme(legend.position = "none") +

labs(caption = "Streamflow (black) and precipitation (blue)",
     title = "Hydrologic seasonality",
     subtitle = "Mohawk River Watershed"
     )
```

Hydrologic seasonality
Mohawk River Watershed

Streamflow (black) and precipitation (blue)

In the above figure the SC is represented by the angle of the arrow and the SI is represented by the length of the arrow.

- Take a look at these results. How does the SI and SC of the streamflow compare to that of the precipitation?

- What might explain any of the differences and/or similarities?

- What are the environmental, ecological, engineering implications of these differences?