Extracting Climate Data from Prism

Shannon Carter

February 22, 2019

This is my portion of a collaborative project with a group of 10 scientists around the world (folks in Germany, New Zealand, Phoenix, and Santa Barbara!). My role so far was to find open source global climate data, and extract monthly mean temperature and precipitation for specific geographic sites for which we have biological data. Here, I download climate data for each of the LTER sites from Prism and put it in a long format data frame.

First, load required packages:

```
library(tidyverse)
library(stringr)
library(prism)
library(raster)
library(magrittr)
library(popler)
```

Fetch Climate Data

Prism is an opensource climate database that compiles data from a range of monitoring networks around the world uses sophisticated modeling techniques to interpolate across different spatial and temporal resolutions. Here, I've extracted global monthly mean temperature and precipitation data for from 2000-2005.

```
# First, set a file path where prism data will be stored
options(prism.path = 'C:\\Users\\Shannon\\Documents\\F18 Topics in Ecology\\prism.path_monthly')

# Now, select the type and date range
get_prism_monthlys(type = 'tmean', years = 2000:2016, mo = 1:12, keepZip = F)
get_prism_monthlys(type = 'ppt', years = 2000:2016, mo = 1:12, keepZip = F)
```

Process Climate Data

The data is downloaded as a zip folder of raster files. Luckily, R's 'prism' package has some simple functions to compile the files into a format we can work with. Here, I first stack the raster data and then extract coordinates.

```
# Here, you'll need to specify which dataset to pull if you've downloaded multiple to the path.
mystack <- ls_prism_data() %>%
  prism_stack(.)

# Get project coordinates from raster stack
mycrs <- mystack@crs@projargs
```

Fetch & Process Biological Data

Now I have the whole world's climate data, but I only need the data for specific locations for which we have biological data. These locations are Long Term Ecological Research (LTER) sites. This network of 30 sites across North America supports a huge amount of ecological research. I think this project is really neat becaue we don't collect any data for it– it's all coming from open source data! At each site, there is census data for a huge range of different species- everything from herons to penguins to reef fish. Eventually, this project aims to connect population sizes of these animals to climate. But we're not there yet. At this stage, I'm only pulling the LTER data so I can extract the location coordinates to connect with the climate data. Here, I download all LTER site metadata from a database called "popler". From the data pulled from popler, extract the coordinates for each LTER site, put them in the same coordinate reference system (CRS) as the Prism climate data, and match them.

```
# First, pull the metadata for all LTER sites using pplr_browse
all_studies <- popler::pplr_browse()

# Select just the lat/long and site ID columns and make a df
lter_sites <- all_studies %>%
  dplyr::select(lng_lter,lat_lter, lterid)

lter_sites <- as.data.frame(lter_sites)

# Convert these locations to format that can be matched to prism data
coordinates(lter_sites) <- c('lng_lter', 'lat_lter')
proj4string(lter_sites) <- CRS(mycrs)
```

Join Climate & LTER data

The Prism climate data comes in pretty nasty shape, so here I wrangle that into a manageable format that I can join with the LTER data.

```
# Make a data frame including LTER site locations and the climate data extracted from the raster stack
data <- data.frame(coordinates(lter_sites),
                   lter_sites$lterid,
                   extract(mystack, lter_sites))

# Reshape data. Col 1:3 are lat, long, and site ID. Col 4:ncol are climate data
# Column headers include date and climate type info
data <- data %>%
  gather(date, value, 4:ncol(data))

# The column header includes the date and data type, but also some other metadata that we don't need
# Here, I remove the extra info from the column header
data$date <- gsub('PRISM_', '', data$date) %>%
  gsub('stable_4kmM3_', '', .) %>%
  gsub('stable_4kmM2_', '', .) %>%
  gsub('_bil', '', .)

# Split header into type (precipitation or temperature), year, and month
data <- separate(data, 'date',
                 into = c('type', 'YearMonth'),
                 sep = '_')
data <- separate(data, 'YearMonth',
                 into = c('year', 'month'),
                 sep = 4)

# Reshape data-- make a separate column for temperature and precipitation
data <- unique(data)
data <- data %>%
```

```
  spread(type, value) %>%
  rename(lng = lng_lter, lat = lat_lter, lterid = lter_sites.lterid)

# Order data by LTER site
data <- data[order(data$lterid),]
```

View Data

The final product is a long format dataset with lat/long, LTER site ID code, year and month, precipitation (in mm) and mean temperature (in °C)

```
head(data, 15)
```

```
##         lng   lat lterid year month    ppt  tmean
## 613 -122.26 44.21    AND 2000    01 474.28  1.640
## 614 -122.26 44.21    AND 2000    02 256.22  4.310
## 615 -122.26 44.21    AND 2000    03 171.27  5.440
## 616 -122.26 44.21    AND 2000    04 103.51 10.070
## 617 -122.26 44.21    AND 2000    05 142.10 11.270
## 618 -122.26 44.21    AND 2000    06  82.94 17.200
## 619 -122.26 44.21    AND 2000    07   8.01 18.820
## 620 -122.26 44.21    AND 2000    08   0.80 19.585
## 621 -122.26 44.21    AND 2000    09  42.05 16.050
## 622 -122.26 44.21    AND 2000    10 132.58 10.750
## 623 -122.26 44.21    AND 2000    11 127.39  3.370
## 624 -122.26 44.21    AND 2000    12 244.37  3.190
## 625 -122.26 44.21    AND 2001    01 110.95  3.110
## 626 -122.26 44.21    AND 2001    02  72.73  3.410
## 627 -122.26 44.21    AND 2001    03 171.77  6.715
```

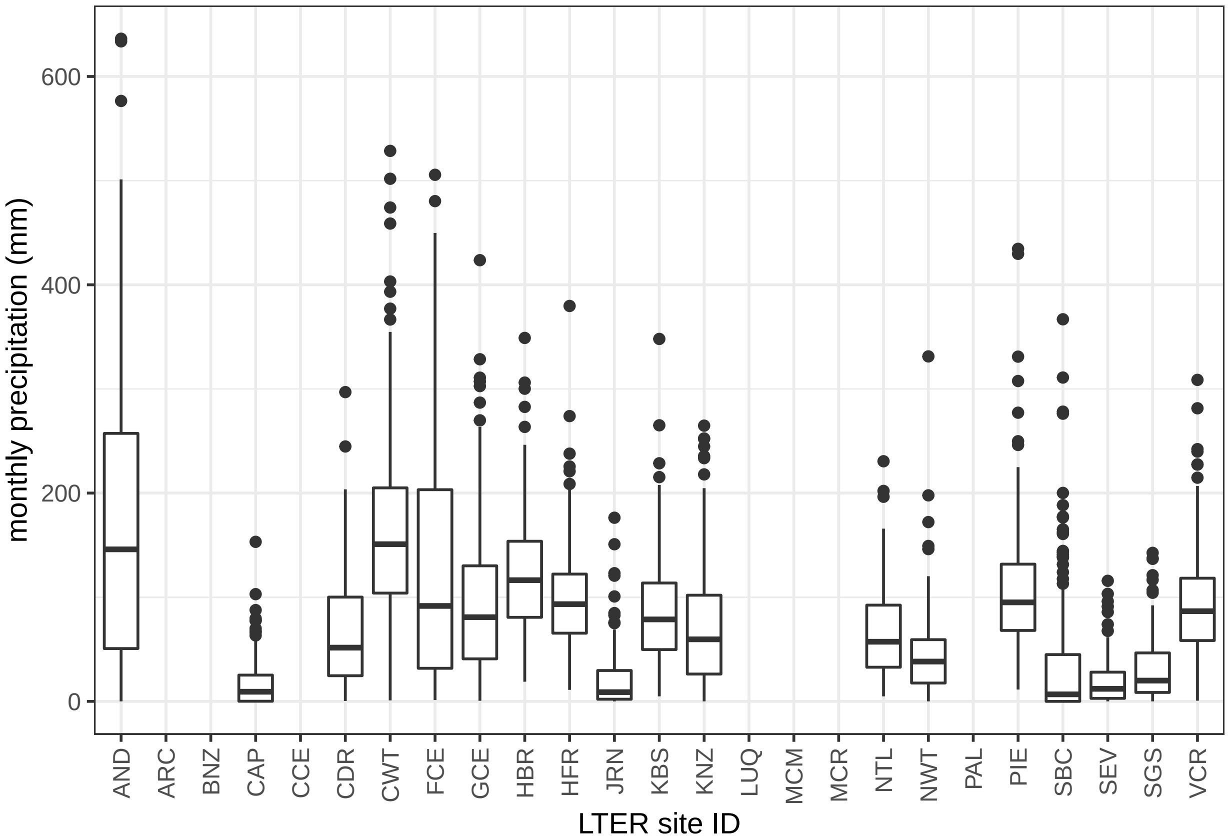
```
str(data)
```

```
## 'data.frame':    5100 obs. of  7 variables:
##  $ lng   : num  -122 -122 -122 -122 -122 ...
##  $ lat   : num  44.2 44.2 44.2 44.2 44.2 ...
##  $ lterid: Factor w/ 25 levels "AND","ARC","BNZ",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year  : chr  "2000" "2000" "2000" "2000" ...
##  $ month : chr  "01" "02" "03" "04" ...
##  $ ppt   : num  474 256 171 104 142 ...
##  $ tmean : num  1.64 4.31 5.44 10.07 11.27 ...
```

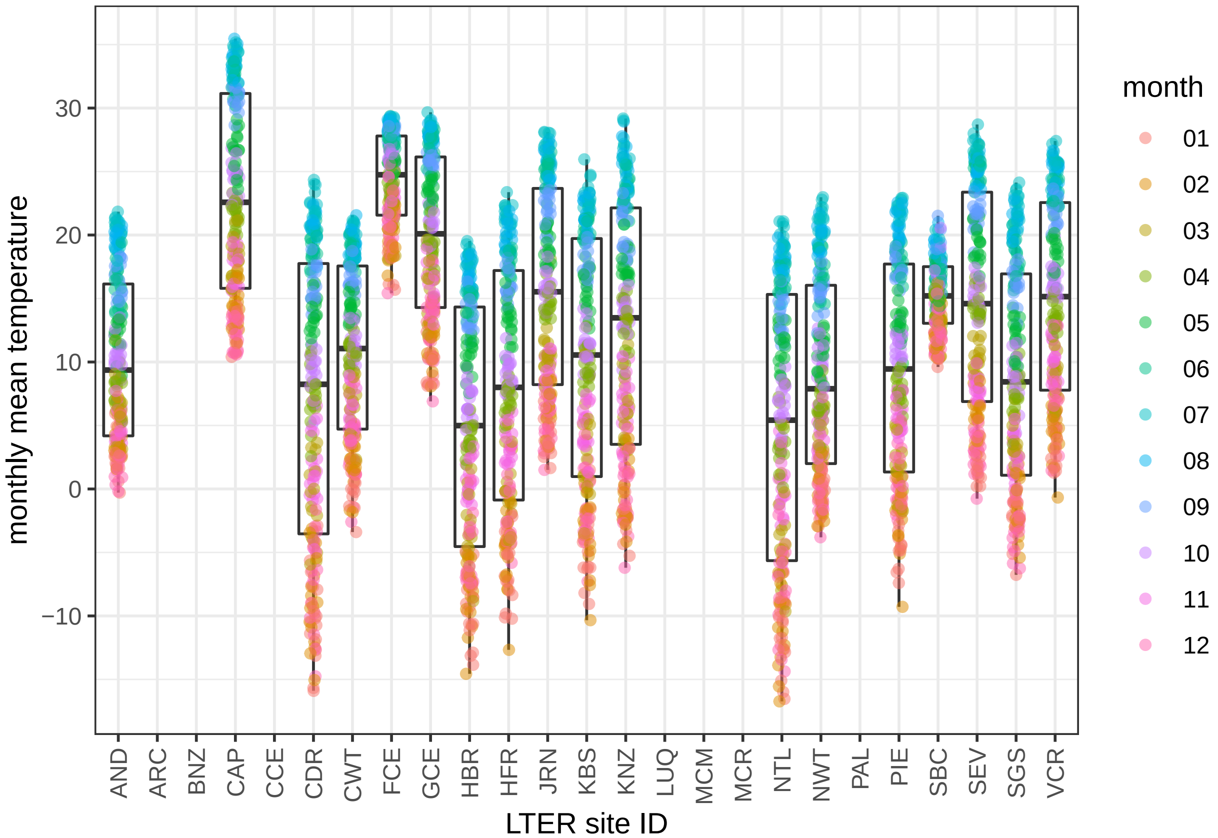
Some Plots

These plots aren't super informative, but are useful for confirming that everything went well. I used these to make sure the climate data was in the range expected and spot check some sites.

First, precipitation. CAP, JRN, and SEV are 3 sites in Arizona/New Mexico, and as expected, are looking quite dry. The site with the most rainfall, AND, is in Portland. So looks like the precipitation data worked well.



Now, temperature. Here, I colored points by month so we can spot check by seasonal trends as well as sites. It's clear that the blue points that represent summer months are warmer than red/orange points that represent winter months. Good. The hottest sites, CAP and FCE, are in Phoenix and the Florida Everglades.



monthly mean temperature
30
20
10
0
−10
AND
ARC
BNZ
CAP
CCE
CDR
CWT
FCE
GCE
HBR
HFR
JRN
KBS
KNZ
LUQ
MCM
MCR
NTL
NWT
PAL
PIE
SBC
SEV
SGS
VCR
LTER site ID
month
01
02
03
04
05
06
07
08
09
10
11
12