BIO 423 - Lab 9

Benjamin Blonder

Spring 2019

Learning outcomes

Content goals:

- Describe community dynamics using paleoecological datasets (residence times, lag times, disturbance responses)
- Propose hypotheses for drivers of community dynamics
- Estimate climate-driven rates of range expansions

R goals:

- Gain familiarity with the Neotoma database
- Calibrate radiocarbon dates to calendar-year dates

Based on a lab written by Jack Williams at University of Wisconsin:

`https://serc.carleton.edu/neotoma/activities/121251.html`

and demos written by Simon Goring:

`https://github.com/SimonGoring/neotoma_paper/blob/master/Neotoma_paper.md`

This week we are going to work with real paleoecological data in the Neotoma database. The database is named for the genus of packrats, Neotoma. This database contains records of species occurrences over space and time, compiled from many types of proxy records (e.g. middens, pollen cores, macrofossils) and many investigators. It is currently the best and most well-regarded paleoecological database. You can explore it through the web at:

`https://www.neotomadb.org/`

or visually explore the database at:

`http://apps.neotomadb.org/explorer/`

We can also explore the database programmatically via the `neotoma` R package - our goal for today, in a two-part lab.

Part one - community dynamics

We will examine changes in species composition at two contrasting lakes in British Columbia for which pollen core data are available. Each slice in these cores has been carbon-dated and calibrated to calendar-year dates.

In the Neotoma database, we have to follow a somewhat complex workflow to get at the data in a useful form. We first query the database for sites, then get site metadata, then download palynological/age data, then convert these records into standardized taxonomic names, then convert pollen counts into relative abundances - and can finally analyze the results. The below script takes you through this workflow. Don't worry about the details of this unless you want to use similar data in your own research - these are just the necessary steps to access the data, and reflect the complexity of a database that must store information for a range of species and record types across space and time.

First, we will load in data from two sites.

```
library(neotoma)
library(analogue)
library(ggplot2)

marion <- get_site(sitename = 'Marion Lake%')
louise <- get_site(sitename = 'Louise Pond%')

marion.data <- get_dataset(marion)
louise.data <- get_dataset(louise)

louise.data[[1]]$site.data

# get underlying data for each site
marion.dl <- get_download(marion.data)
louise.dl <- get_download(louise.data)

# calculate taxon/count information for each site
# we have to match the counts against a known species list, here the 'P25' list
# see the help for this function for details if curious
marion.taxa <- compile_taxa(marion.dl[[1]], list.name ='P25')
louise.taxa <- compile_taxa(louise.dl[[1]], list.name ='P25')
```

```
# list all the taxa we have
print(louise.taxa$taxon.list)
```

```
##                            taxon.name variable.units variable.element
## 1                   Alnus viridis-type           NISP           pollen
## 2       Asteraceae subf. Cichorioideae           NISP           pollen
## 3                         Cupressaceae           NISP           pollen
## 4                     Dryopteris-type           NISP            spore
## 5                             Ericales           NISP           pollen
## 6                 Gentiana douglasiana           NISP           pollen
## 7                     Huperzia selago           NISP            spore
## 8                             Isoëtes           NISP            spore
## 9                   Lycopodium spike         number          counted
## 10                 Lycopodium tablets  grains/tablet    concentration
## 10.1               Lycopodium tablets  grains/tablet    concentration
## 12                       Other plants           NISP     pollen/spore
## 13                             Picea           NISP           pollen
## 14                             Pinus           NISP           pollen
## 15                           Poaceae           NISP           pollen
## 16   Polypodiophyta (monolete) undiff.           NISP            spore
## 17                         Polypodium           NISP            spore
## 18                    Ranunculus-type           NISP           pollen
## 19                   Rosaceae undiff.           NISP           pollen
## 20                    Sample quantity             ml           volume
## 21            Selaginella selaginoides           NISP            spore
## 22    Spermatophyta undiff. (aquatics)           NISP           pollen
## 23                  Tsuga heterophylla           NISP           pollen
## 24                   Tsuga mertensiana           NISP           pollen
## 25                    Alnus rubra-type           NISP           pollen
## 26                         Botrychium           NISP            spore
## 27                 Caltha leptosepala           NISP           pollen
## 28                         Cyperaceae           NISP           pollen
```

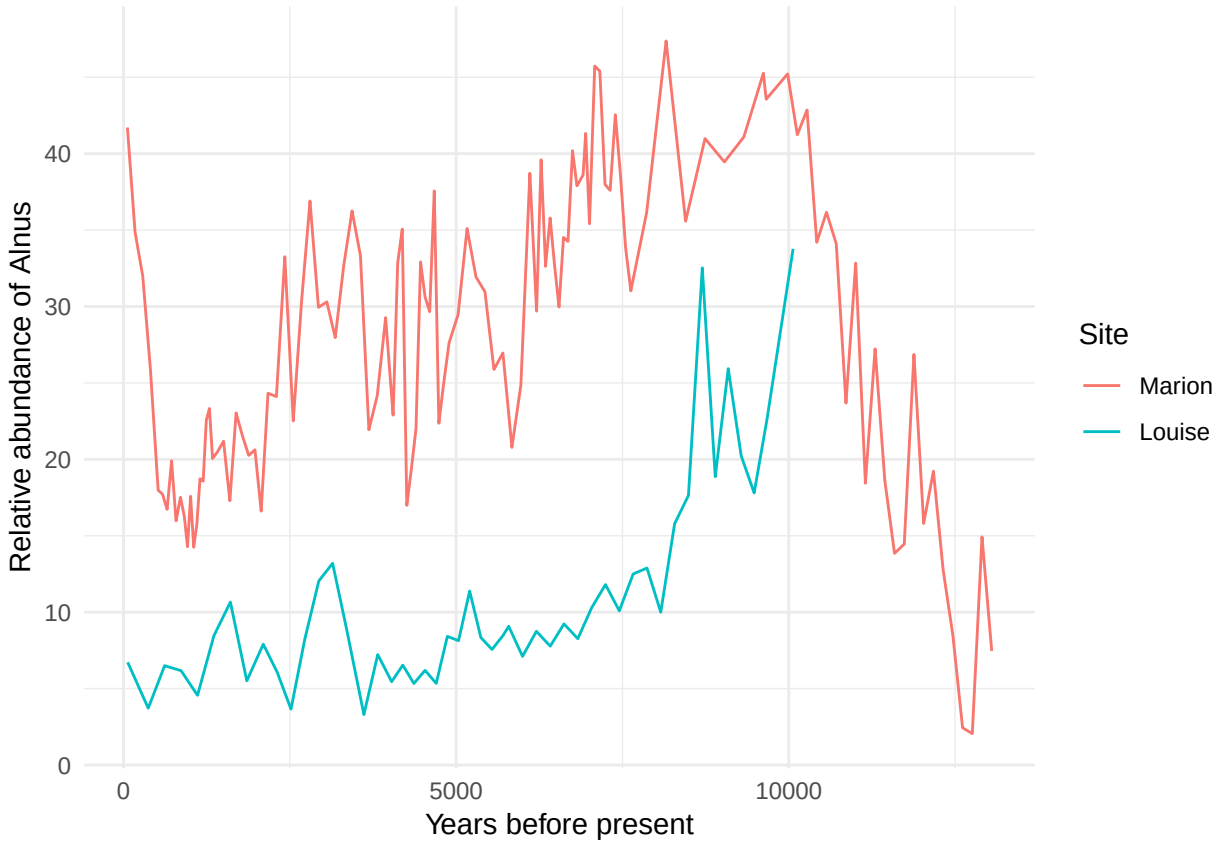
```
## 29                         Sanguisorba           NISP           pollen
## 30          Thelypteris quelpaertensis           NISP            spore
## 31        Asteraceae subf. Asteroideae           NISP           pollen
## 32                Coptis aspleniifolia           NISP           pollen
## 33                           Pteridium           NISP            spore
## 34                    Potamogetonaceae           NISP           pollen
##      variable.context         taxon.group ecological.group
## 1                  NA     Vascular plants             TRSH
## 2                  NA     Vascular plants             UPHE
## 3                  NA     Vascular plants             TRSH
## 4                  NA     Vascular plants             VACR
## 5                  NA     Vascular plants             TRSH
## 6                  NA     Vascular plants             UPHE
## 7                  NA     Vascular plants             VACR
## 8                  NA     Vascular plants             AQVP
## 9                  NA Laboratory analyses             LABO
## 10                 NA Laboratory analyses             LABO
## 10.1               NA Laboratory analyses             LABO
## 12                 NA      Plants undiff.             UNID
## 13                 NA     Vascular plants             TRSH
## 14                 NA     Vascular plants             TRSH
## 15                 NA     Vascular plants             UPHE
## 16                 NA     Vascular plants             VACR
## 17                 NA     Vascular plants             VACR
## 18                 NA     Vascular plants             UPHE
## 19                 NA     Vascular plants             UPHE
## 20                 NA Laboratory analyses             LABO
## 21                 NA     Vascular plants             VACR
## 22                 NA     Vascular plants             AQVP
## 23                 NA     Vascular plants             TRSH
## 24                 NA     Vascular plants             TRSH
## 25                 NA     Vascular plants             TRSH
## 26                 NA     Vascular plants             VACR
## 27                 NA     Vascular plants             UPHE
## 28                 NA     Vascular plants             UPHE
## 29                 NA     Vascular plants             UPHE
## 30                 NA     Vascular plants             VACR
## 31                 NA     Vascular plants             UPHE
## 32                 NA     Vascular plants             UPHE
## 33                 NA     Vascular plants             VACR
## 34                 NA     Vascular plants             AQVP
##                                          alias            compressed
## 1                        Alnus viridis-type                 Alnus
## 2          Asteraceae subf. Cichorioideae          Prairie Forbs
## 3                      Cupressaceae Cupressaceae/Taxaceae
## 4                    Dryopteris-type                 Other
## 5                           Ericales                 Other
## 6                 Gentiana douglasiana                 Other
## 7                      Huperzia selago                 Other
## 8                           Isoëtes                 Other
## 9                    Lycopodium spike                  <NA>
## 10   Lycopodium tablets|concentration|grains/tablet      <NA>
## 10.1 Lycopodium tablets|concentration|grains/tablet      <NA>
## 12                        Other plants                 Other
```

```
## 13                                    Picea                 Picea
## 14                                    Pinus                 Pinus
## 15                                 Poaceae               Poaceae
## 16 Polypodiophyta (monolete) undiff.                  Other
## 17                                Polypodium                 Other
## 18                             Ranunculus-type               Other
## 19                            Rosaceae undiff.               Other
## 20                             Sample quantity                <NA>
## 21                    Selaginella selaginoides               Other
## 22 Spermatophyta undiff. (aquatics)                   Other
## 23                        Tsuga heterophylla               Tsuga
## 24                         Tsuga mertensiana               Tsuga
## 25                          Alnus rubra-type               Alnus
## 26                                Botrychium                 Other
## 27                           Caltha leptosepala               Other
## 28                                Cyperaceae           Cyperaceae
## 29                               Sanguisorba                 Other
## 30                  Thelypteris quelpaertensis               Other
## 31                Asteraceae subf. Asteroideae       Prairie Forbs
## 32                        Coptis aspleniifolia               Other
## 33                                  Pteridium                 Other
## 34                             Potamogetonaceae               Other
```

Next, we will focus on the data for the genus Alnus, which occurs at both sites:

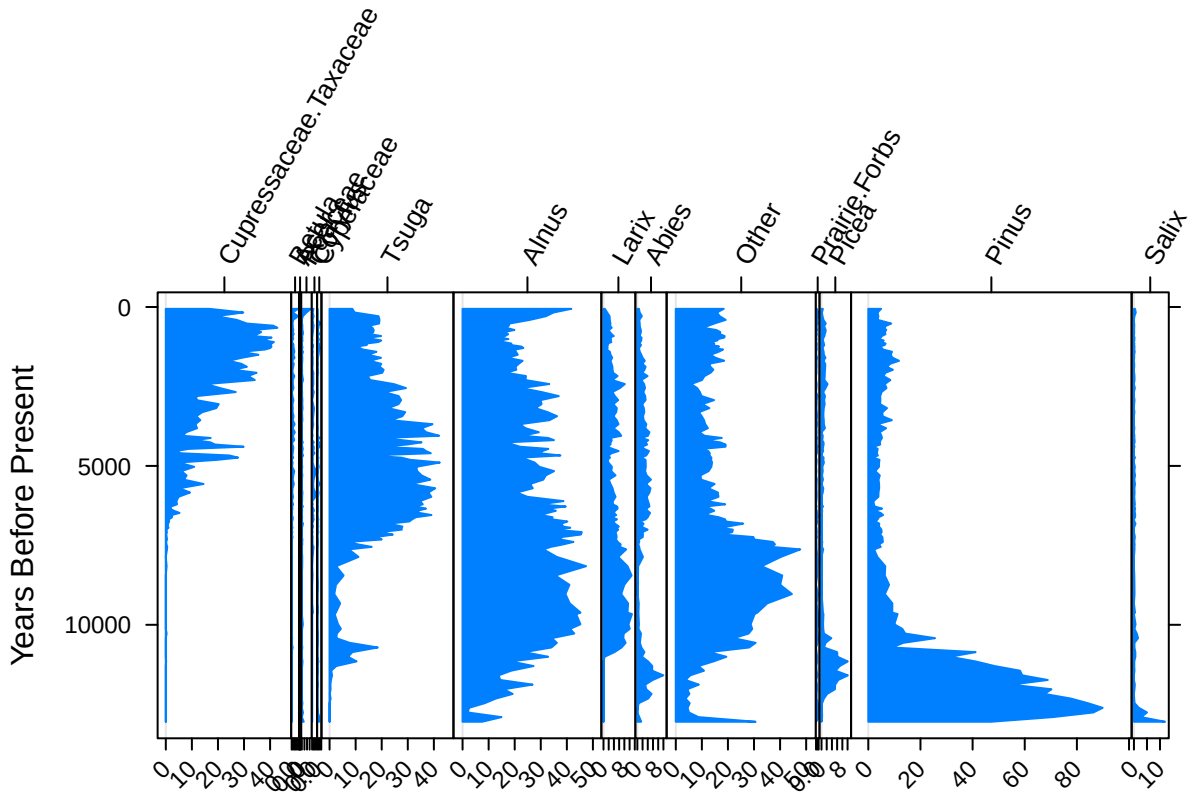
```
# convert pollen counts into relative abundances, assuming that pollen counts
# are proportional to species abundance
marion.alnus <- tran(x = marion.taxa$counts, method ='percent')[,'Alnus']
louise.alnus <- tran(x = louise.taxa$counts, method ='percent')[,'Alnus']

# get age information
marion.age = marion.taxa$sample.meta$age
louise.age = louise.taxa$sample.meta$age

# assemble final dataframe of alnus relative abundance
df_marion.louise = rbind(
  data.frame(Site='Marion',Age=marion.age, Percent.Alnus=marion.alnus),
  data.frame(Site='Louise',Age=louise.age, Percent.Alnus=louise.alnus)
  )
```

```
# make a plot
ggplot(df_marion.louise, aes(x=Age,y=Percent.Alnus,col=Site)) +
  geom_line() +
  theme_minimal() +
  xlab("Years before present") +
  ylab("Relative abundance of Alnus")
```



We can also make a stratigraphic plot of the entire pollen core:

```
# make stratigraphic plot
marion.pct <- data.frame(tran(marion.taxa$counts, method = "percent"))
marion.pct$age <- marion.age
marion.pct.strati <- chooseTaxa(marion.pct)
Stratiplot(age ~ ., marion.pct.strati, sort = 'wa', type = 'poly',
           ylab = "Years Before Present")
```



Questions (part one)

1. Include a Google Map screenshot for each of the two sites, Louise and Marion. Hint: you can get coordinates via: `get_site(louise.data)$long`
2. For approximately how many thousand years have cypresses (Cupressaceae) been present at the Louise site?
3. How long do pines (Pinus) persist at the Marion site at high abundance? What hypotheses might explain their decline in dominancee?
4. Why might Alnus have rapidly increased in the last 300 years at the Marion site but not the Louise site? (Hint: see Mathewes, R., A palynological study of postglacial vegetation changes in the University Research Forest, southwestern British Columbia. Canadian Journal of Botany, 1973, 51(11): 2085-2103, in subheader 'Zone ML-5 (Spectra 4-1, About 500 B.P. to the Present', and think about white settler history in Canada...)

Part two

We can also use Neotoma to do spatial analyses to examine how species distributions have changed over time by using presence data from multiple sites. In this part, we will examine evidence for climate-driven range expansions in Canada after the end of the last ice age. This analysis is based on a study of Macdonald and Cwynar (1991), who used Pinus pollen percentages to map the northward migration of lodgepole pine following the retreat of the Laurentide Ice Sheet. They used a cutoff of 15% Pinus pollen as an indicator of Pinus presence, but Strong and Hills (2013) have since conducted a more robust analysis with more sites and a 5% cutoff. The below (from a Simon Goring exercise) replicates part of this analysis.

The approach will be to get pollen data from many sites in Canada, then determine when in each site Pinus

is first detected. An analysis of time-of-detection vs site latitude then allows estimation of expansion rates of the species. But you will see below that it is not so simple to do this...

```
# load some necessary packages
library("ggmap")
```

```
## Warning: package 'ggmap' was built under R version 3.5.2
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library("ggplot2")
library("reshape2")
library("Bchron")
library("gridExtra")
library("mapproj")
```

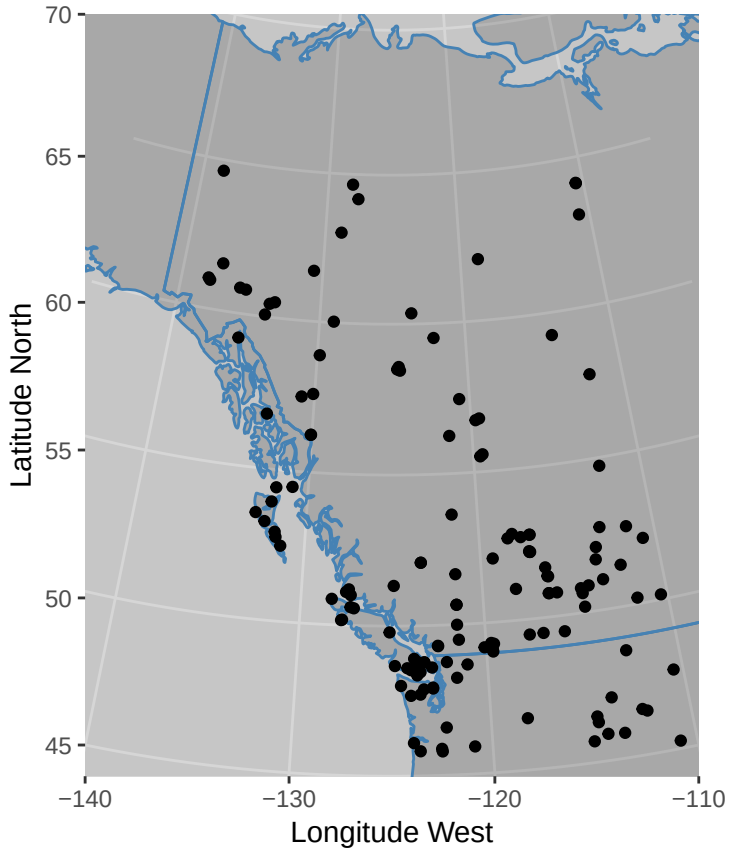
```
## Loading required package: maps
```

```
# find pollen datasets containing Pinus in British Columbia
# note that loc is in (lonW, latS, lonE, latN) format
all.datasets <- get_dataset(loc = c(-140, 45, -110, 65),
                            datasettype = 'pollen',
                            taxonname = 'Pinus%')
```

```
## The API call was successful, you have returned 143 records.
```

```
# extract coordinates from the search results
all.coords <- get_site(all.datasets)
```

```
# make a quick plot of all of these sites
map <- map_data('world')
ggplot(data = data.frame(map), aes(long, lat)) +
  geom_polygon(aes(group=group), color = 'steelblue', alpha = 0.2) +
  geom_point(data = all.coords, aes(x = long, y = lat)) +
  xlab('Longitude West') +
  ylab('Latitude North') +
  coord_map(projection = 'albers', lat0 = 40, lat1 = 65,
            xlim = c(-140, -110), ylim = c(45, 70))
```




```
# download the data for all of these sites
# (this may take a few minutes, ignore warnings, be patient)
all.downloads <- get_download(all.datasets, verbose = FALSE)
```

Next, convert the data into pollen counts with matches species names, and retain only first-occurrences of Pinus at 5 percent abundance in each site.

```
# match species names
compiled.cores <- compile_taxa(all.downloads, 'P25')

# filter the data to find the first occurrence of pinus in the dataset at above 5% relative abundance
# discarding also any cores that span less than 5 kya - present
top.pinus <- function(x) {
  # first convert the pollen counts to proportions
  x.pct <- tran(x$counts, method = "proportion")
  #  Cores must span at least the last 5000 years (and have no missing dates):
  old.enough <- max(x$sample.meta$age) > 5000 & !all(is.na(x$sample.meta$age))
  #  Find the highest row index associated with Pinus presence over 5%
  oldest.row <- ifelse(any(x.pct[, 'Pinus'] > .05 & old.enough),
                       max(which(x.pct[, 'Pinus'] > .05)),
                       0)
  #  return a data.frame with site name & location, and the age and date type
  #  (since some records have ages in radiocarbon years) for the oldest Pinus.
  out <- if (oldest.row > 0) {
    data.frame(site = x$dataset$site.data$site.name,
               lat = x$dataset$site.data$lat,
               long = x$dataset$site.data$long,
```

```
              age = x$sample.meta$age[oldest.row],
              date = x$sample.meta$age.type[oldest.row])
  } else {
    return(NULL)
  }
  return(out)
}

# apply the 'top.pinus' function to each site, then bind results into a dataframe
summary.pinus <- do.call("rbind", lapply(compiled.cores, top.pinus))
```

```
# note that some of the ages are suspect...
# and that the date types are either radiocarbon, or calendar/calibrated radiocarbon dates
summary(summary.pinus)
```

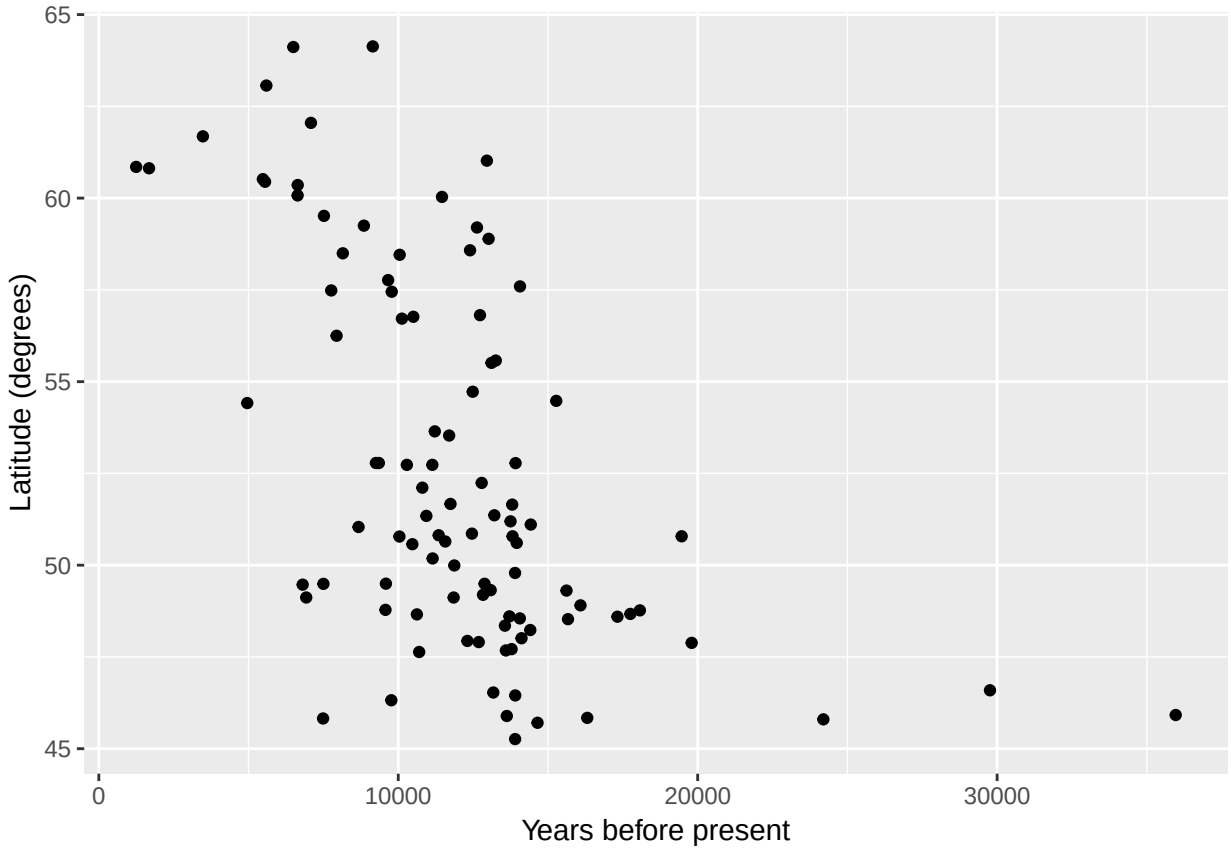
```
##                   site         lat             long             age
##  Lost Lake          : 2   Min.   :45.26   Min.   :-138.4   Min.   :   -45
##  Battle Ground Lake: 1   1st Qu.:48.62   1st Qu.:-125.1   1st Qu.:  8641
##  Boone Lake         : 1   Median :50.95   Median :-121.9   Median : 11241
##  Candelabra Lake    : 1   Mean   :52.66   Mean   :-122.2   Mean   : 12508
##  Carp Lake          : 1   3rd Qu.:57.29   3rd Qu.:-116.9   3rd Qu.: 13303
##  Lac Ciel Blanc     : 1   Max.   :64.63   Max.   :-110.5   Max.   :125223
##  (Other)            :95
##                                date
##  Radiocarbon years BP            :48
##  Calendar years BP               : 1
##  Calibrated radiocarbon years BP:53
##
##
##
##
```

Because of the date issues, we also have to filter the dates and do a calibration between radiocarbon dates and calendar dates, here with the Int13 calibration curve.

```
# keep only sites that have dates within the last 40 kyr
summary.pinus <- summary.pinus[summary.pinus$age < 40000 & summary.pinus$age > 1000,]

# for data that have uncalibrated radiocarbon ages, convert them to calibrated ages
radio.years <- summary.pinus$date %in% 'Radiocarbon years BP'
sryears <- sum(radio.years, na.rm = TRUE)
# use the 'intcal13' calibration
calibrated.dates <- BchronCalibrate(summary.pinus$age[radio.years],
                            ageSds = rep(100, sryears),
                            calCurves = rep('intcal13', sryears))

# define a weighted mean giving more evidence to points on the calibration curve
# with more data
wmean.date <- function(x) sum(x$ageGrid*x$densities / sum(x$densities))

# replace the uncalibrated ages with the calibrated ages
summary.pinus$age[radio.years] <- sapply(calibrated.dates, wmean.date)
```

Finally, we are ready to make the key plot of date vs. latitude.

```
ggplot(summary.pinus, aes(x=age,y=lat)) +
  geom_point() +
  xlab("Years before present") +
  ylab("Latitude (degrees)")
```



Questions (part two)

5. Redraw the map, coloring sites by the age at which Pinus first appears.

Hint: `scale_color_gradient(low="yellow",high="blue")`

6. Since the Last Glacial Maximum (21 Kya), it appears as though Pinus continues migrating northwards. Estimate the migration rate in degrees per kyr. (Hint: remember to subset the data before linear regression).

7. Why is it necessary to use a (say 5%) relative abundance cutoff to determine that Pinus is present at a site? (Hint: think about biases in the data).

8. Why is it necessary to calibrate radiocarbon dates to calendar dates? (Hint, see

`https://en.m.wikipedia.org/wiki/Calibration_of_radiocarbon_dates`

9. Use the Neotoma Explorer website (advanced search) or the `get_dataset()` function to identify all datasets with type pollen core, occurring anywhere on earth. how many sites do you find? given that a core typically costs about 2 person-years and $30,000 to analyze (including carbon dating), what is your estimate for the total time and money cost of this part of the Neotoma database?

Optional questions for graduate students

- Find all mammoth records in the database (search for `mammut*`). Make a plot of # of occurrences vs. time. Is there a bias towards more recent observations? Next, make maps of mammoth occurrences in 5 Kyr intervals and describe what you find. Estimate mammoth range size in each temporal bin using a convex hull around the occurrence points (possibly clipped to the continental extents). (Hint, use `geometry::convhulln(...,options="FA")$vol`). How does mammoth range size change over time? When do mammoths go extinct? Is there a slow or rapid decline in range size before their extinction?

What to hand in

- A single Word Document including:
- written answers (1-2 sentences) and figures for each question above
- A copy of your R script (the contents of your `.R` file pasted into the Word document)
- Author contribution statement