SDS 439 - Final Project

Due May 7, 5:00 pm

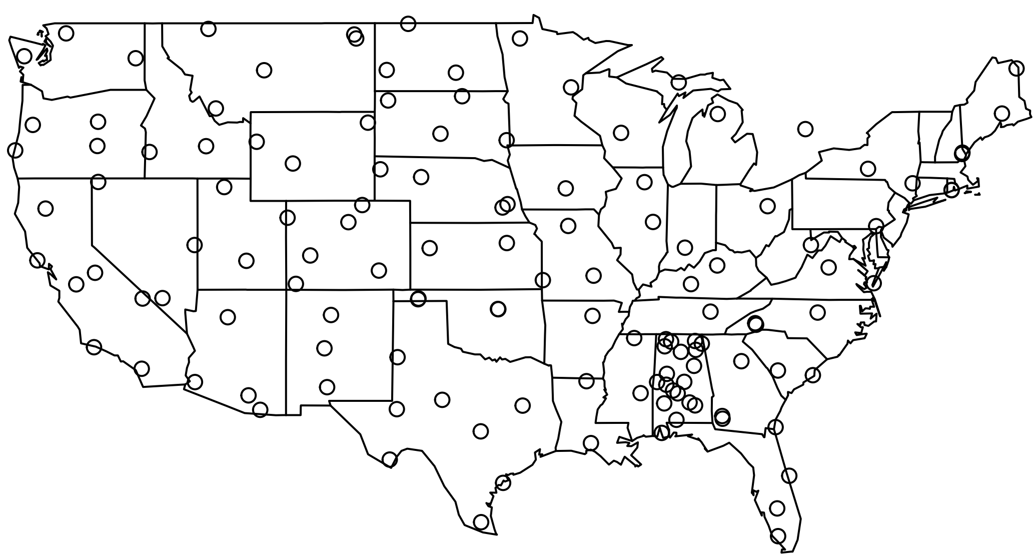
Weather Data

We have analyzed data from the US Climate Reference Network (USCRN) (https://www.ncei.noaa.gov/access/crn/) a few times this semester, looking at data from a handful of stations. Now, we'll do a more complete analysis of data from all of the lower-48 stations.

There is a file in the course github: `datasets/uscrn_daily.RData` which contains daily observations from 132 different sites.

1. As usual, start by plotting the data. I'll help by making one plot that uses the maps package to create a map. Make three more plots showing various features of the data, focusing on the `T_DAILY_AVG` variable, which will be our primary response of interest. One of the plots should explore how temperatures vary over the course of the year.

```
load("../datasets/uscrn_daily.RData")
library("maps")
map("state")
site_locs <- unique( dat[c("LONGITUDE","LATITUDE")] )
points( site_locs$LONGITUDE, site_locs$LATITUDE )
```



2. Fix all of the missing values, and create a `date` column in the dataframe, using the `as.Date` function and the `LST_DATE` column. Also create a `days` column that is equal to the number of days since `2020-01-01`. For reference, see the documentation at https://www.ncei.noaa.gov/pub/data/uscrn/products/daily01/readme.txt

3. Notate the data as follows: Let i refer to the row of the dataset, y_i the daily average, d_i the number of days since `2020-01-01`, and $j(i)$ be the site.

   We will be fitting the following model to the responses:

$$Y_i = b_{0,j(i)} + b_{1,j(i)} d_i + b_{2,j(i)} \sin\left(\frac{2\pi d_i}{365.25}\right) + b_{3,j(i)} \cos\left(\frac{2\pi d_i}{365.25}\right) + b_{4,j(i)} \sin\left(\frac{4\pi d_i}{365.25}\right) + b_{5,j(i)} \cos\left(\frac{4\pi d_i}{365.25}\right) + \varepsilon_i$$

$$\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2_{j(i)})$$

where $b_{3,j(i)}$ just means that each site has a separate regression coefficient.

How do you interpret the intercept $b_{0,j(i)}$? Think carefully because it's actually not possible to set all of the covariates equal to zero at the same time.

How do you interpret $b_{1,j(i)}$?

4. Fit this model separately to each site, and store the results. (this is a short prompt, but it will require some coding).
5. Make a plot summarizing the results of estimating $b_{1,j}$. Your plot should include information about the estimates and their uncertainties, and it should be visually informative. Make sure that outliers are not dominating the plot.
6. From your analysis of $b_{1,j(i)}$, what can you conclude about long term trends in temperature in the continental US? Be specific about the size of the effects and their uncertainties.
7. Comment on any problems with the model, how they might be influencing the results, and how you might address the problems.
8. For each site, calculate the following quantites based on your fitted models

- Amplitude: difference in temperature between warmest and coldest day
- Day number with coldest temperature: (1-365)
- Day number with warmest temperature: (1-365)
- Daily standard deviation: how much actual temperatures tend to differ from the expected temperature.

  You can accomplish this in various ways, but a relatively straightforward way is to create a data frame with a row for each day of the year, and use your fitted model for each site and the `predict` function to make a prediction of temperature on each day in the data frame.

9. Plot the yearly cycles for the following sites, making sure to label your plots:

- the site with the largest amplitude
- the site with the smallest amplitude
- the site with the earliest coldest day in the winter season
- the site with the latest coldest day in the winter season

  For the coldest day, you'll need to deal with the fact that December 31 is earlier in the winter season than January 1. Try to get all of the yearly cycles on the same plot, if it looks ok.

10. Make a map showing the day of year with the warmest temperature. You might try combining the `fields::quilt.plot` function with the `maps::map` function.
11. Make a map showing the daily residual standard deviation. Where in the U.S. do the temperatures deviate most from expected?