Stat 470/670 Homework 5

Due: Monday, February 20 at 5pm

Submit exactly two files: (i) a PDF/HTML file with your write-up and graphs and (ii) a .r or .Rmd file with code to reproduce your results.

Use the sea ice dataset at
http://jfukuyama.github.io/teaching/stat670/assignments/N-seaice-extent-daily.csv

1. Using `group_by` and `summarise`, create a summarized dataset that contains the monthly average temperature instead of the daily average temperature. Your dataset should now have 533 rows instead of 14,527.
2. Read about banking to 45° in Cleveland (pp. 88-90 and the experiment described on p. 218). Based on your summarized dataset, make a plot with mean sea ice extent on the vertical axis, year on the horizontal axis, and faceted out by month. Add a loess smoother. (This is a variant on a cycle plot, see p. 164 in Cleveland for an example and more information.) Set the aspect ratio of the plot so that the segments corresponding to each month are on average banked to 45°.

   Note 1: Although there are automatic proposals for setting the aspect ratio so as to bank curves to 45°, you can/should do it by eye.

   Note 2: If you are using an Rmd, for your final document preparation, you can set the aspect ratio of the plot by specifying in the chunk heading `fig.height = x` and `fig.width = y` for whatever values you want the height and width of the figure to be.

   Note 3: In this particular example, you might try using `facet_grid` instead of `facet_wrap` to get a row of plots instead of a grid.
3. Make the same plot as in the previous part, but using the unsummarized dataset.
4. Comment about the differences between the two plots, including: (a) Are there any different kinds of conclusions you can draw from them? and (b) Does the smoother seem to be more useful in one than in the other?