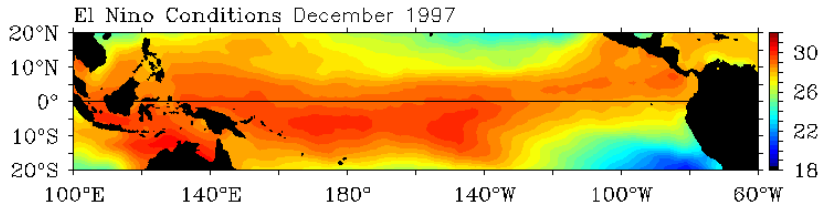
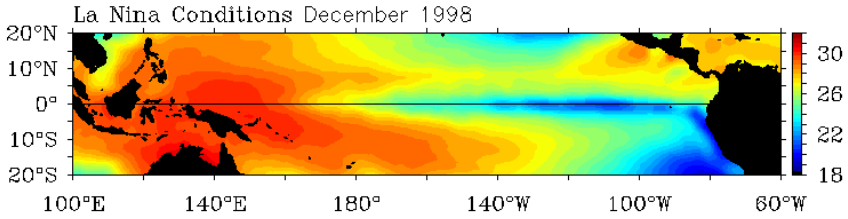
Exercise 1: Effect of El Niño / La Niña in Pacific tuna catches (Total: 25 marks)

Background:

El Niño (https://www.pmel.noaa.gov/elnino/la-nina-pacific) is characterized by unusually warm ocean temperatures (red and orange) all along the equator in the Pacific Ocean (below).



La Niña is characterized by unusually cold ocean temperatures (blue) along the equator in the eastern and central equatorial Pacific Ocean (below).



We want to explore the effects of these climatic conditions in yearly tuna catches of the Pacific tuna stocks.

Data sources

*** Sea Surface Temperatures:**
We have retrieved the data for Sea Surface Temperatures (SST) from NOAA/Pacific Marine Environmental Laboratory: https://www.pmel.noaa.gov/tao/drupal/disdel/
We have downloaded SST data for two geographical localities: (Nauru: 0ºN 165 E) and (Galapagos: 0ºN 110 W). The SST data are 5-day averages from 1980/1986 to 2020 (some data are missing). These data are stored in two different files:
Exercise_1_SST_0N165E_5days.txt
Exercise_1_SST_0N110W_5days.txt

*** Tuna catches:**
We have downloaded tuna catches for all tuna stocks (1980 to 2010) from the FAO website: http://www.fao.org/fishery/statistics/tuna-catches/query/en
These data are stored in Exercise_1_tuna_catches_by_stock_1980_2010.csv

All data files are available from the Canvas webpage of the course and also from https://drive.google.com/drive/folders/11zcrVYQs1y2LMUf7ofMpA8cAKddWNOMk?usp=sharing

Data treatment to do:

We are going to calculate the differences in monthly average temperature between the sampled localities in Nauru (Western Pacific) and Galapagos (Eastern Pacific). The bigger this difference, the higher values for La Niña conditions. We are going to call this variable La_Nina_index.

Then we are going to calculate the yearly average of La_Nina_index for all the years we have enough data, and we will produce a summary table of yearly average La_Nina_index per year.

On the other hand, we are going to select the tuna stocks from the Pacific Ocean. Then we are going to add yearly catches of all types of fishing_gear.
Finally, we will produce three plots:
- Boxplot of La_Nina_index vs year.
- Area plot of Pacific tuna catches vs year
- Scatter plot of the values of total tuna catches in the Pacific versus yearly average values of La_Nina_index, faceted by stock.

Step-by-step analyses:

- Table Nauru-SST:
1. Load the data for Nauru temperatures from **Exercise_1_SST_0N165E_5days.txt**. Remember to remove the first two lines before the column names (commentary lines). You can do that automatically using the option skip=2 in the read_table() or read_table2() functions. Check that the resulting table has four columns (YYYY_MM_DD, HHMM, SST, and Q). **(1 mark)**
2. Remove the missing data (which are codified as negative values (-9.99), as explained in the comment lines of the file). **(1 mark)**
3. Using the function separate(), expand the information from the YYYY_MM_DD column, to create a column for year, another one for month and another one for day. **(1 mark)**
4. Group the data by Year-Month and create a new summary table with the summaries of monthly averages for SST. Remember to keep track of the total number of observations for every month (n). Remove all monthly observations with n < 3 values **(2 mark)**

- Table Galapagos-SST:
5. (Repeat what you did for the Nauru table for the Galapagos table): Load the data for Galapagos temperatures from **Exercise_1_SST_0N180W_5days.txt**. Remember to remove the first two lines before the column names (which are commentary lines). **(1 mark)**
6. Remove the missing data (codified as -9.99, as explained in the comments). **(1 mark)**
7. Expand the information from the YYYY_MM_DD column, to create a column for year, another one for month and another one for day. **(1 mark)**
8. Group the data by Year-Month and create a new table with the summaries of monthly averages for SST. Remember to keep track of the total number of observations for every month (n). Remove all monthly observations with n < 3 values **(2 mark)**

- Table La_Nina_index
9. Join both monthly average SST tables by year and month using an inner_join and create a new column with the differences between monthly average SST in Nauru minus monthly average SST in Galapagos. Call this column La_Nina_Index. **(2 mark)**
10. Grouping by year, create a new summary table with yearly averages of La_Nina_Index, keeping track of the number of monthly observations for every year. Keep just those years when we have enough data (n > 8). **(2 mark)**

- Table Tuna catches
11. Load the table from **Exercise_1_tuna_catches_by_stock_1980_2010.csv (1 mark)**
12. Select those lines which contain data from Pacific stocks (hint: they all contain the word "Pacific" in the stock column, and you can use the function str_detect() to select those observations). **(1 mark)**
13. Reformat the table, so that we have the year in tidy format. The names should go to a new column called "year" and the values will go to a new column called "catch". **(2 mark)**

14. Group by stock and generate a summary table with the total catch by year for each stock. **(2 mark)**

- Joining the final tables and generating the three final plots

15. Join the final table of yearly total catch for every species of tuna with the table of yearly average La_Nina_Index. Keep only those years with available data for the two variables. **(2 mark)**
16. Generate a Boxplot of La_Nina_index vs year. **(1 mark)**
17. Generate an Area plot of Pacific tuna catches vs year. **(1 mark)**
18. Generate a scatter plot of total tuna catches in the Pacific vs. yearly La_Nina_index average, faceted by stock. Use scales = "free_y". **(1 mark)**

Write the R script you have used to perform all calculations for the step-by-step analyses.