MA415 Project
Worldwide Ocean Temperature: Subcontinent East 2001-2016

Submitted to:
Prof. Haviland Wright

Prepared By:
Enas Alahmadi
Nikita Pandya
Vala Rahmani
Wjdan Alharthi

Introduction:

The outcome of this project are datasets that are ready for analysis and this report which discusses the sources for the data and how it has been handled during preparation for analysis.

We are focusing on maritime temperature data – air temperature and sea surface temperature. We are collecting and cleaning data from the Subcontinent East region. Two sources of data have been identified – the NOAA buoy system http://www.ndbc.noaa.gov/ that reports a set of local weather readings on an hourly basis. The NOAA coverage is comprehensive in the US and includes coverage in the Atlantic and Pacific Oceans. However, the NOAA coverage is not worldwide. In order to find data in our region: Subcontinent East, we are using data from the Voluntary Observing Ships (VOS) program http://www.vos.noaa.gov .

How to Run:

Manual:
1- Download the data text files from the source above. Save the data in directory data
2- Create the directory to save the cleaned data, call it cleaned_data
3- Open cleaning_script.R and run the script

Script (bash only):
1- Open the Terminal
2- Navigate to the cloned repo directory
3- Copy the command ./run START END and hit Enter. Replace START and END with the year numbers you want to clean (it will clean the years from START to END inclusive). If you want to clean the data for one year, just the same value for both.

After running the script, the cleaned data will be saved in the cleaned_data directory as Rdata files.

Data Importing: - bash scripts

As mentioned above you can manually download all of the data text files from the source above or you can use the bash scripts to both download and clean the data files. List of sources/URLs of all the data files we imported and cleaned can be found in the appendix. The bash scripts for all 12 months for each of the sixteen years and saves the raw text files in the data directory.

Data Cleaning:

Since we are using ship data instead of buoy data, our data was not collected every hour, or even every day. Our data was dependent on the fact that ships needed to be passing through our region. Despite this, we still collected a fair amount of data. The data on the VOS websites is worldwide and is divided per year and then per month from 2001 to 2016. Below is a sample (2 readings) of what the raw data looks like:

200103311800 1390 4290 040073 1 1VTXK 02403 87 93 101307 100 270 243 99A AAA 165 25432992114 5 1F15111111AA11A1A1AA1 276 2 4 366 99 0 BBXX01040120012000 VTXK 31183 99139 10429 42/93 92417 10270 20243 40130 57010 89/// 22273

200104010000 1490 4200 040073 1 1VTXK 03203 67 98 101284 00 290 221 000 900 165 25442992114 5 1F11111111AA11A1A1AA1

We used the script provided by Scarlet as a starting point to parse the raw data. We then heavily modified and added to this script to further parse and clean the data. The cleaning_data.R script cleans the data for a specific year. We start by looping through ever month of the corresponding year. We read the files (which we have imported in the step above) and create a temporary data frame for the current month. Before adding rows to the data frame, we conduct three main checks. Firstly, that the latitude is within our range (6-20), the longitude is also within our range (80-100). Next, we make sure that the hour in which this data

was measured is within a six-hour range of noon. If the data was recorded out of this six-hour range of noon, we note this record with an “A”. This allows our data to fit into our preferred region while remaining consistent throughout the years.

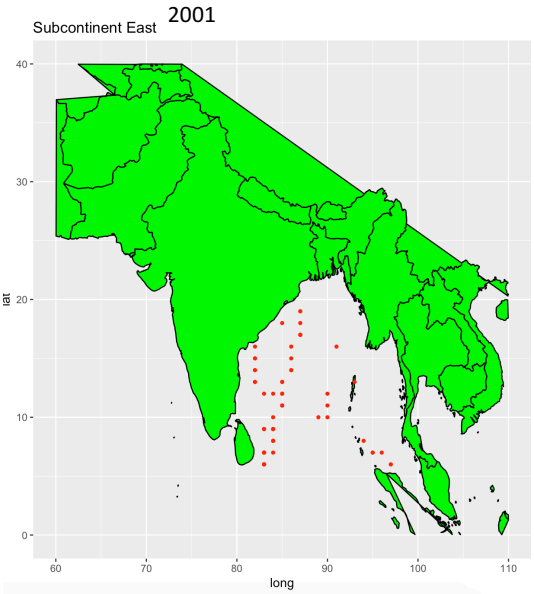
Next, we omit all rows in which any column has the value “NA”. This allows the data to stay meaningful and ensures that all data left in the table is available easily for analysis. Further, we fix the ranges for the following columns: latitude, air temperature (AT) and sea surface temperature (SST). The raw data in its string format, excludes the decimal point which makes the data appear out of its desired range of -99 and 99. Additionally, both the air and sea surface temperature are presented in Celsius.

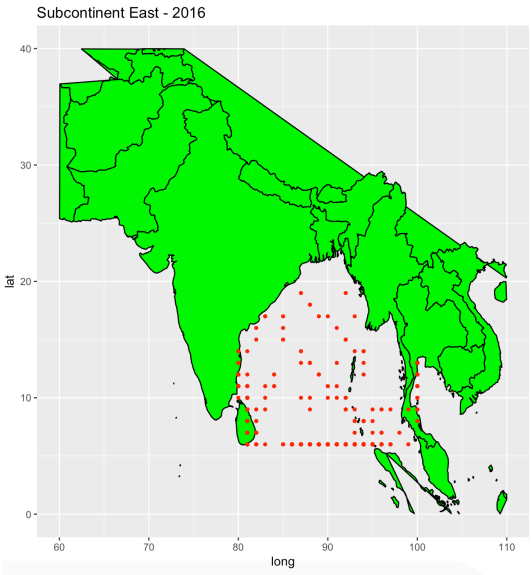
Next, to remove outliers in the temperature data, we performed a quantiles check. To keep the data consistent, accurate and easy to analyze, it is best to remove any extremes in measurement that might indicate severe variability or an experimental error. We removed all data that did not fall in the following percentile: $0.01 < x < 0.99$.

Lastly, to keep our data consistent with the rest of the groups, we formatted each data set into seven columns of following type: group number, reading type (ship), time difference from noon, local time, latitude, longitude, sea temperature and air temperature.

Maps:

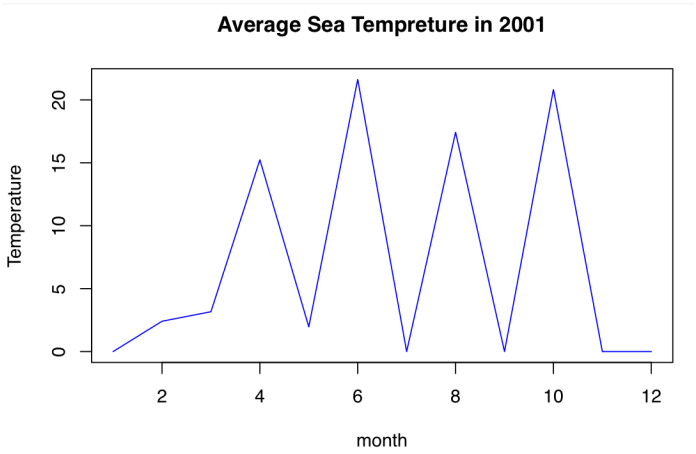
We also created maps per year which show where in the Subcontinent East region the data was collected. We created maps simultaneously as we cleaned the data through the cleaning_data script. Thus, there is one map per year. We can observe distinct shipping lanes throughout the region. Additionally, as the years go on we observe the shipping industry in the region growing as new lanes and locations are introduced and more data is measured throughout the region.

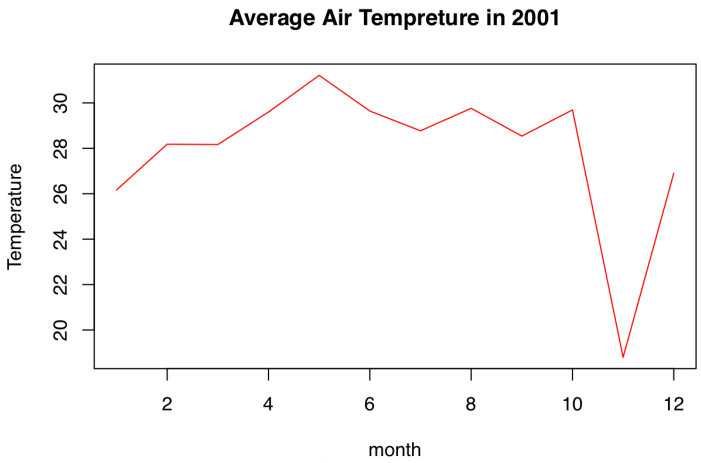


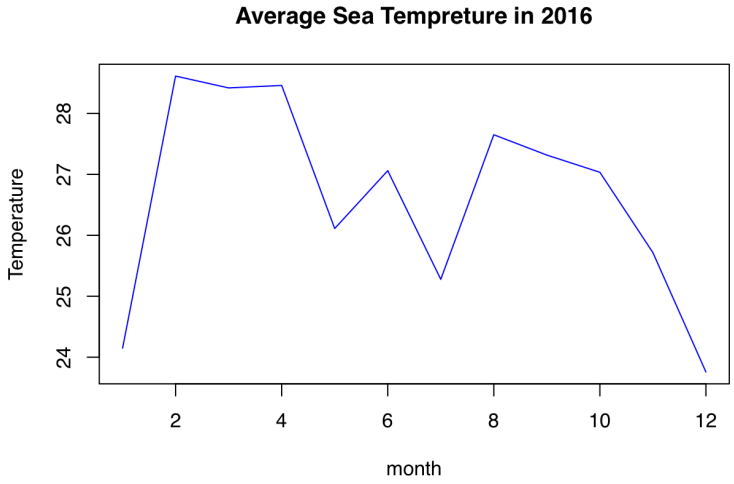


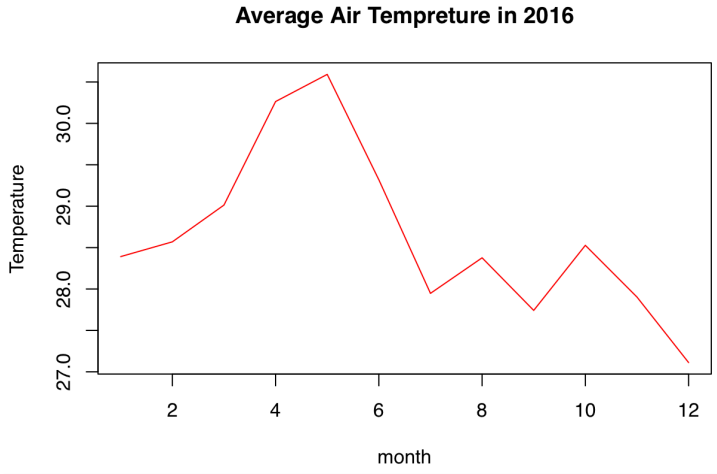
*Maps for every year from 2001-2016 can be found in the appendix at the end of the report.

EDA







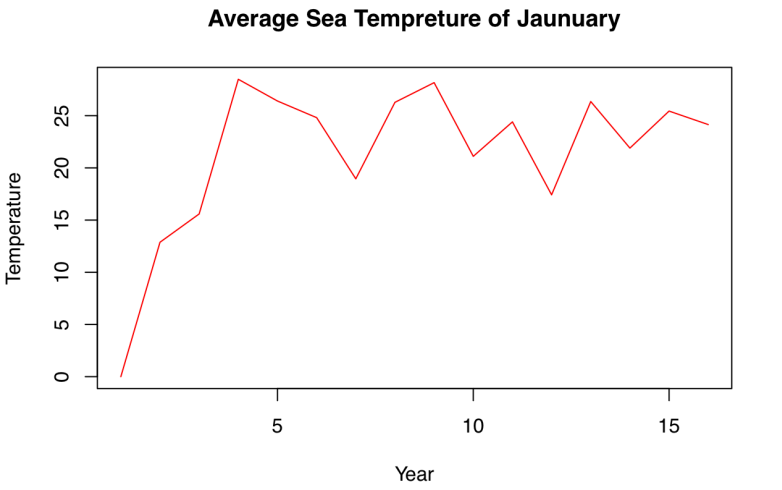


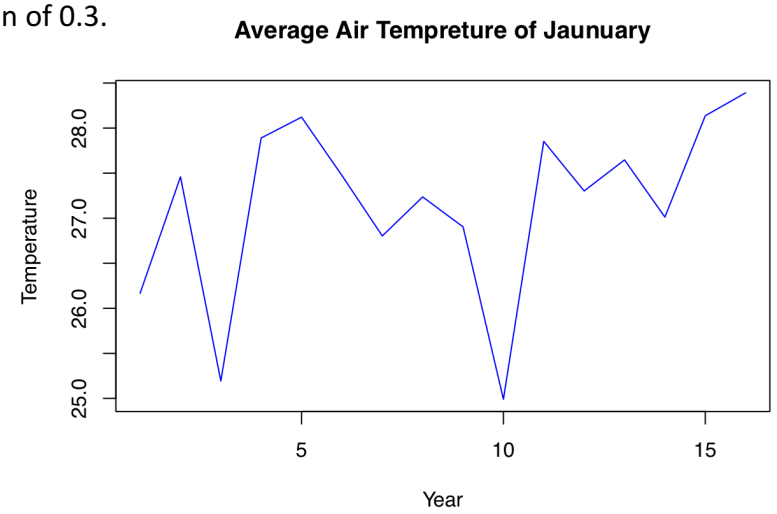
*Graphs for Air and Sea Temperature for every year from 2001-2016 can be found in the appendix at the end of the report.

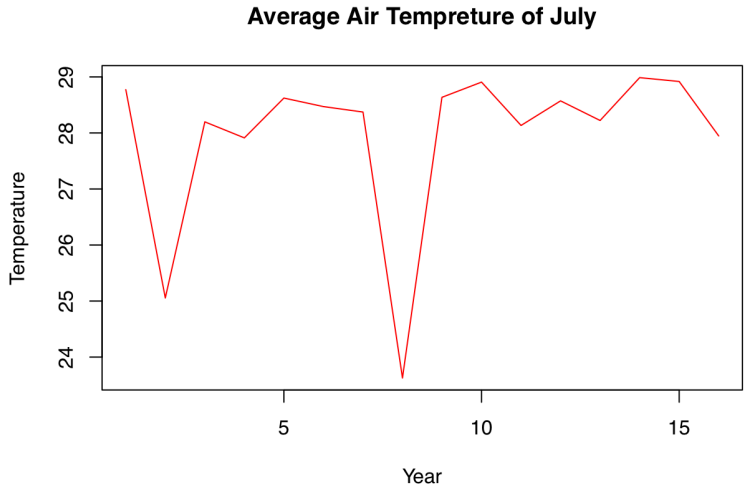
In this section, we create separate plots for each year both for the sea temperature and the air temperature. The plots are structured so that average temperature for each month is calculated and each data point on the plot shows the average temperature in each month. The lowest temperatures at sea are usually observed between December and January and for the rest of the months on average we are looking at slightly higher temperature. Since Subcontinental East is close to the equator and the sunshine is overhead the range of the temperatures is pretty condensed and small. Between 2001 and 2016 this range has been

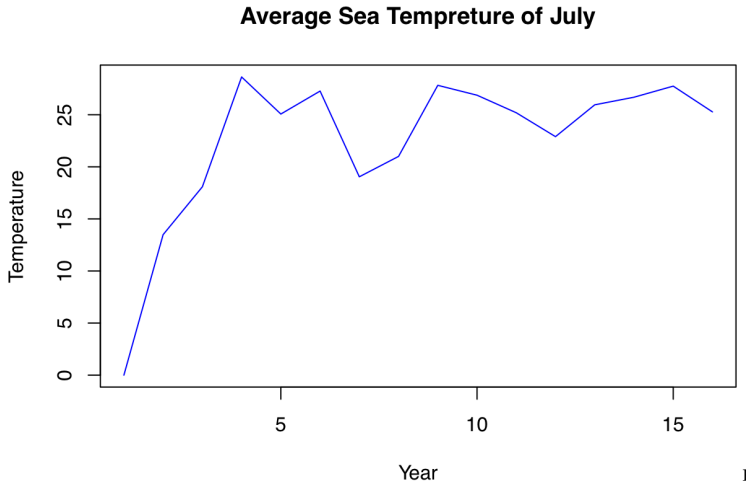
around 15 degrees Celsius for the sea temperature. However, for the air temperature we observe that the range is even more limited around 10 degrees all year round.

*notice: The sharp points that indicate extremely cold temperatures indicate months that have missing data that the average temperature could not be calculated. Overall 16-year average for sea surface temperature is 22.92 with a standard deviation of 5.10 and the average for air temperature is 28.08 with a standard deviation of 0.3.









For the second set of plots we chose January and July, and got the average temperature for all of the years for these specific months. January and July were chosen so our plots could be better illustrator of the coldest and the warmest months of the year. Again, here we observe patterns that show Air temperature has a smaller range than that of Sea. However, in the plots we don't see any patterned movement that indicates increase or decrease of temperature over these 16 years.

Appendix:

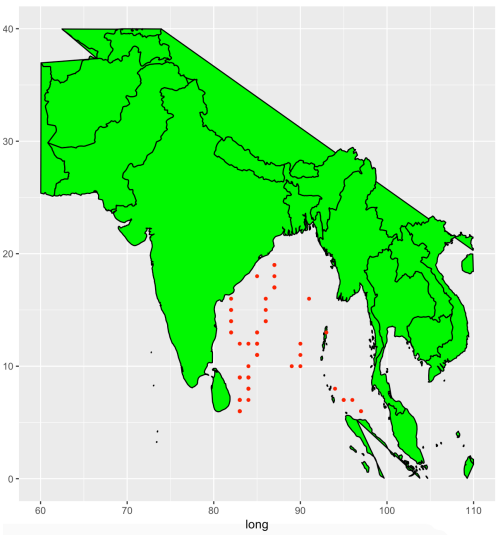
The VOS Clim GTS raw data text files:

2001 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2001/
2002 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2002/
2003 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2003/
2004 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2004/
2005 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2005/
2006 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2006/
2007 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2007/
2008 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2008/
2009 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2009/
2010 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2010/
2011 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2011/
2012 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2012/
2013 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2013/
2014 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2014/
2015 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2015/
2016 (Jan-Dec): https://www1.ncdc.noaa.gov/pub/data/vosclim/2016/

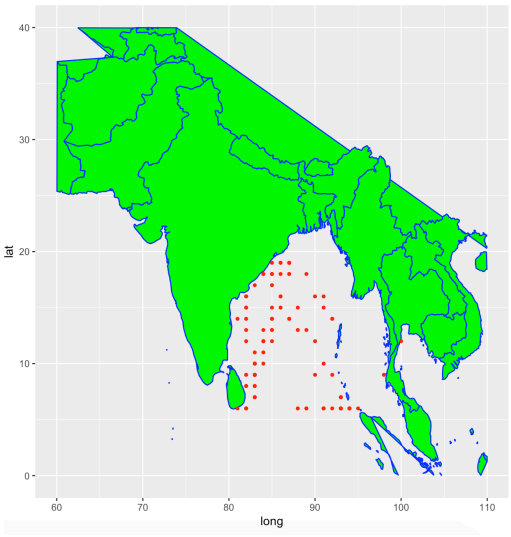
Maps

of the Subcontinent east region year 2001-2016, with data points indicating ships where our data was collected.

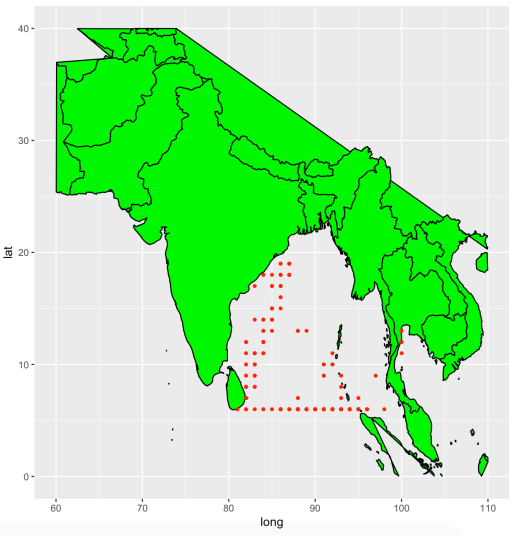
Subcontinent East



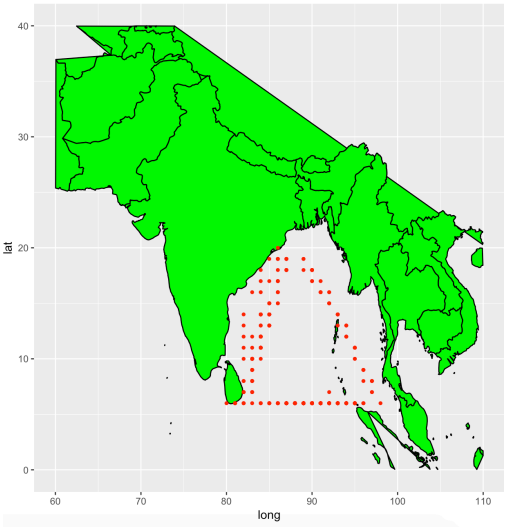
Subcontinent East - 2002



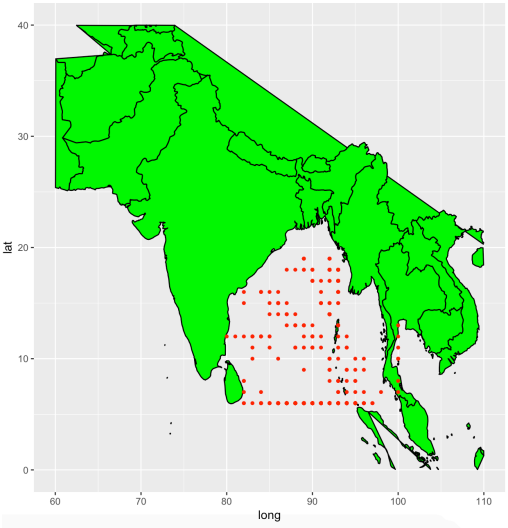
Subcontinent East - 2003



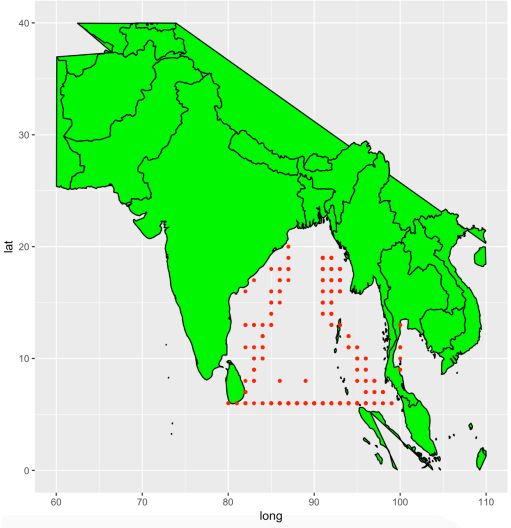
Subcontinent East - 2004



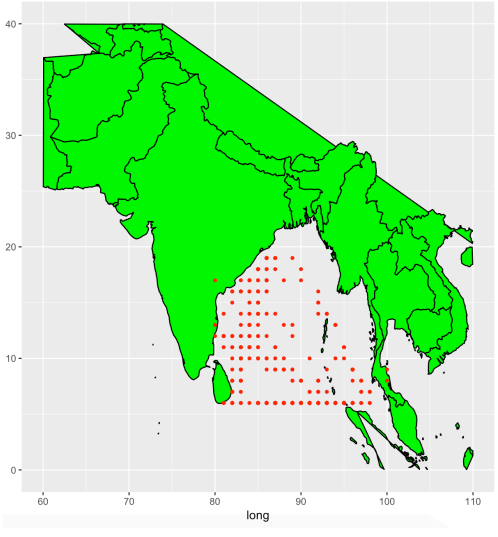
Subcontinent East - 2005



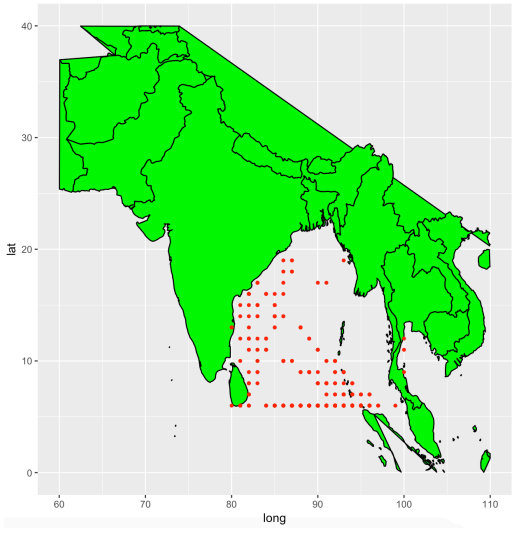
Subcontinent East - 2006



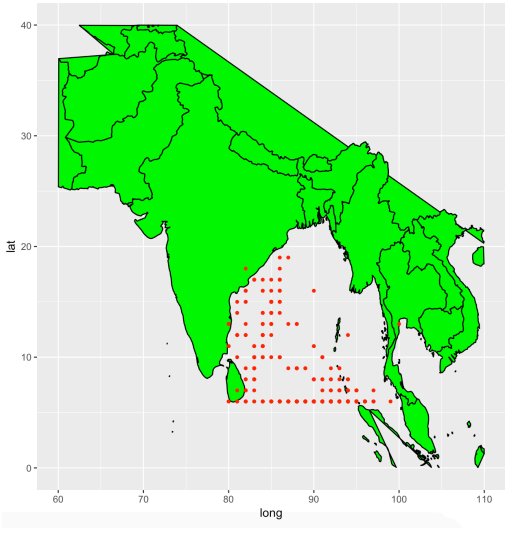
Subcontinent East - 2007

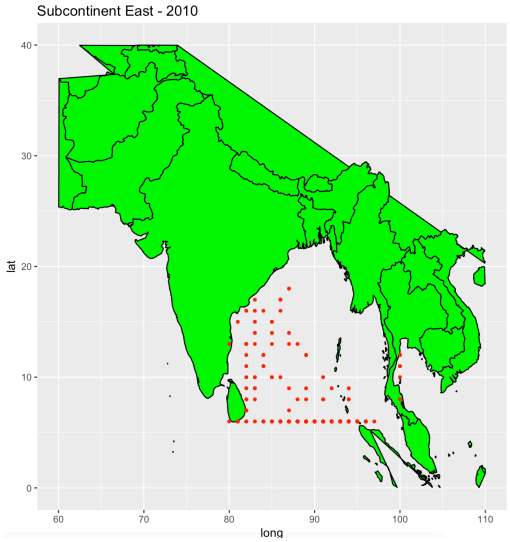


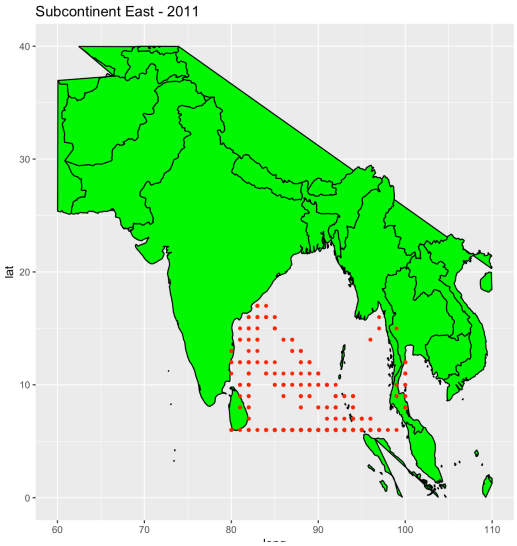
Subcontinent East - 2008

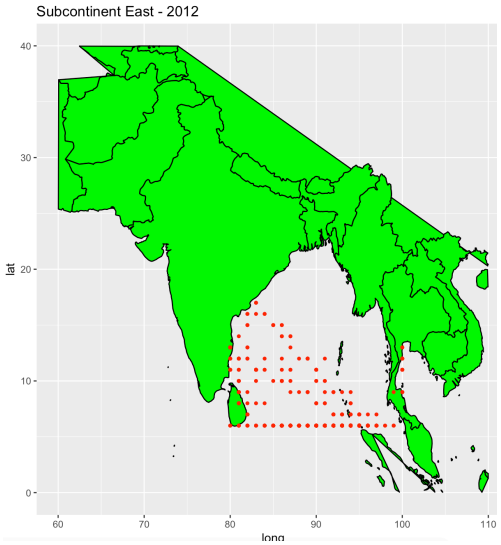


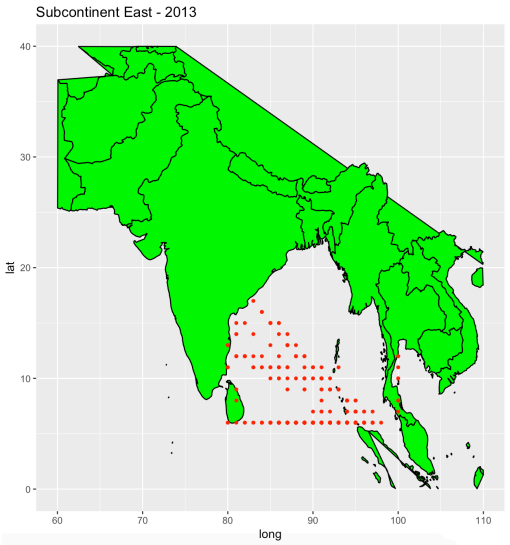
Subcontinent East - 2009

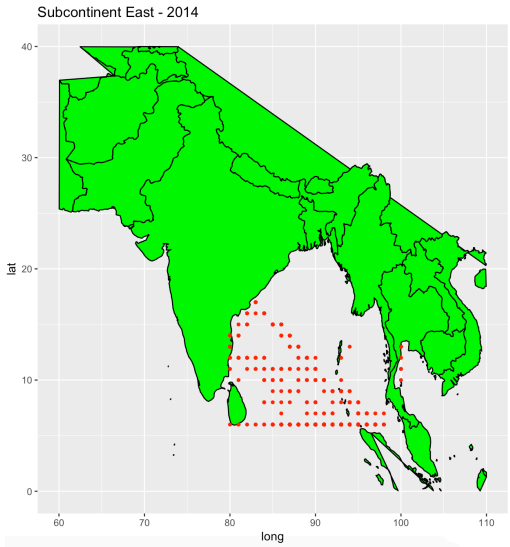


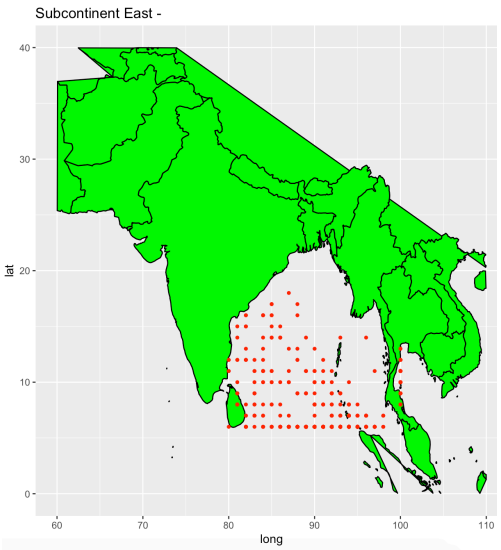

Subcontinent East - 2010
lat
long
40
30
20
10
0
60
70
80
90
100
110

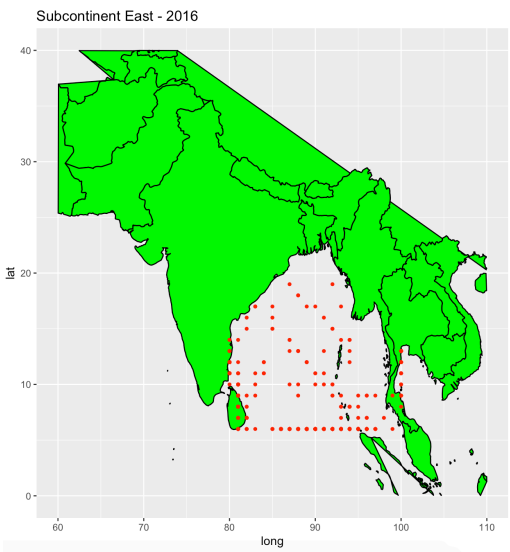

Subcontinent East - 2011
lat
40
30
20
10
0
60
70
80
90
100
110


Subcontinent East - 2012
lat
long
40
30
20
10
0
60
70
80
90
100
110


Subcontinent East - 2013
lat
long
40
30
20
10
0
60
70
80
90
100
110

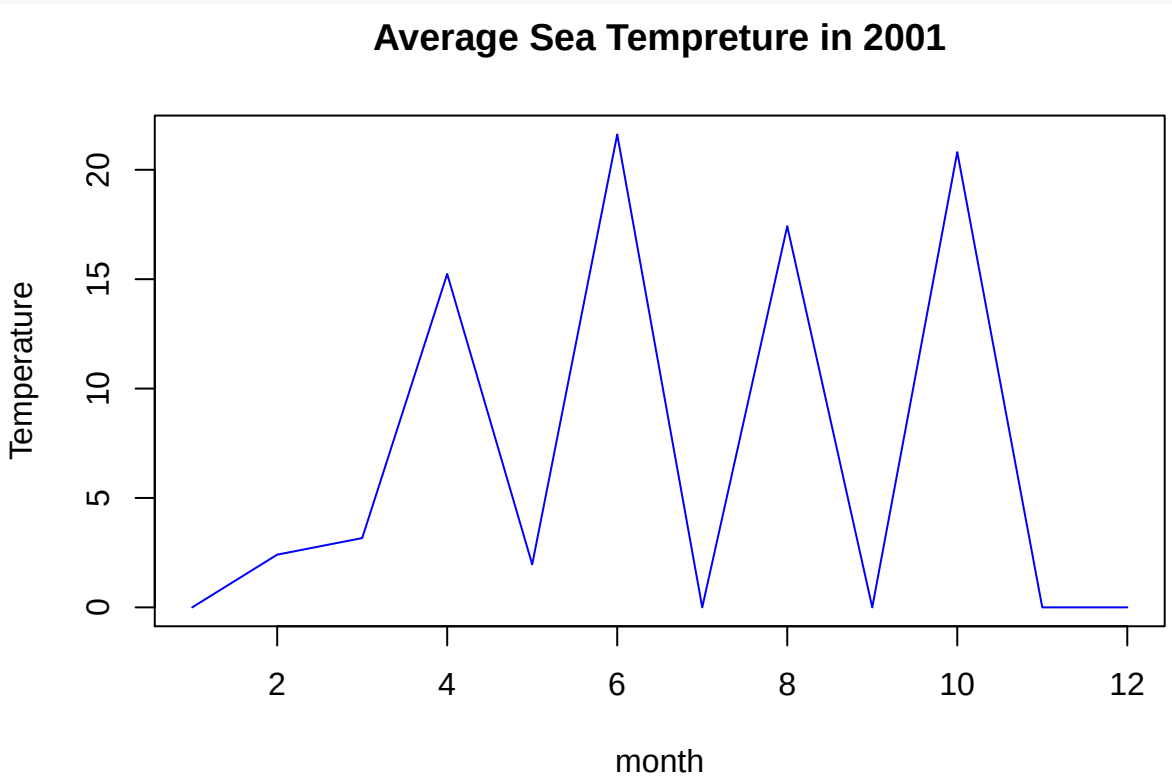

Subcontinent East - 2014
lat
long
40
30
20
10
0
60
70
80
90
100
110

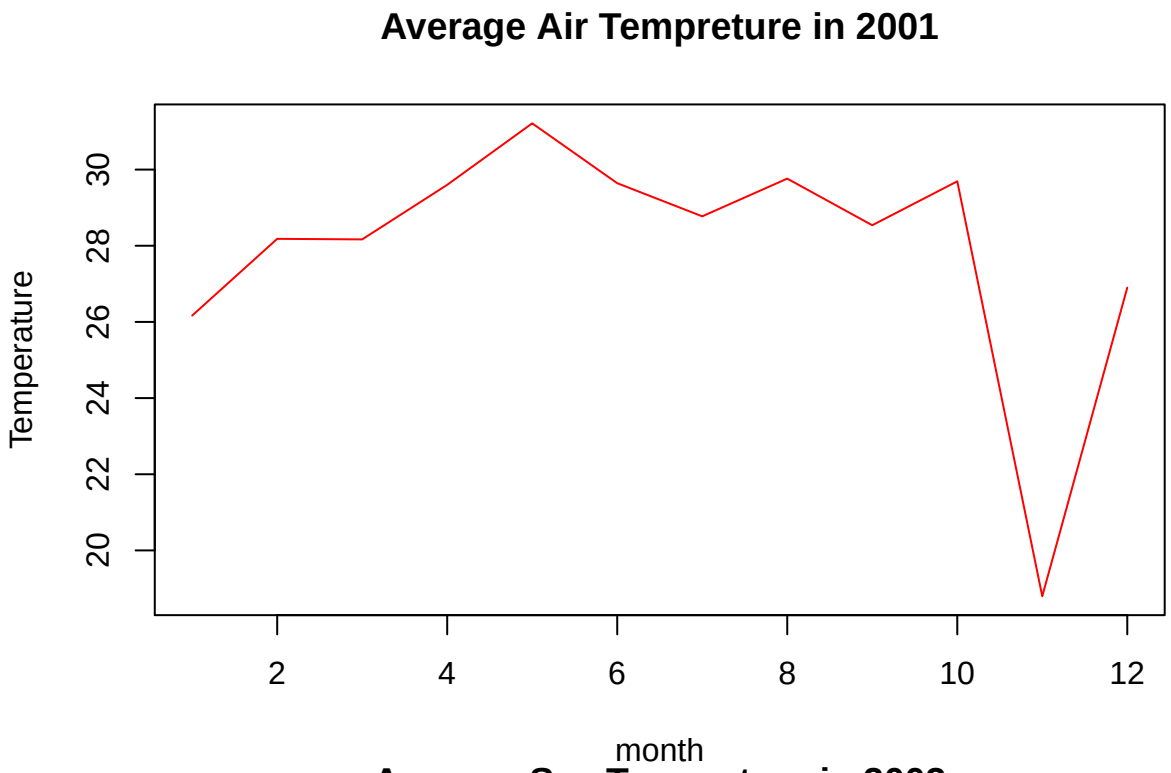

Subcontinent East -
lat
long
40
30
20
10
0
60
70
80
90
100
110

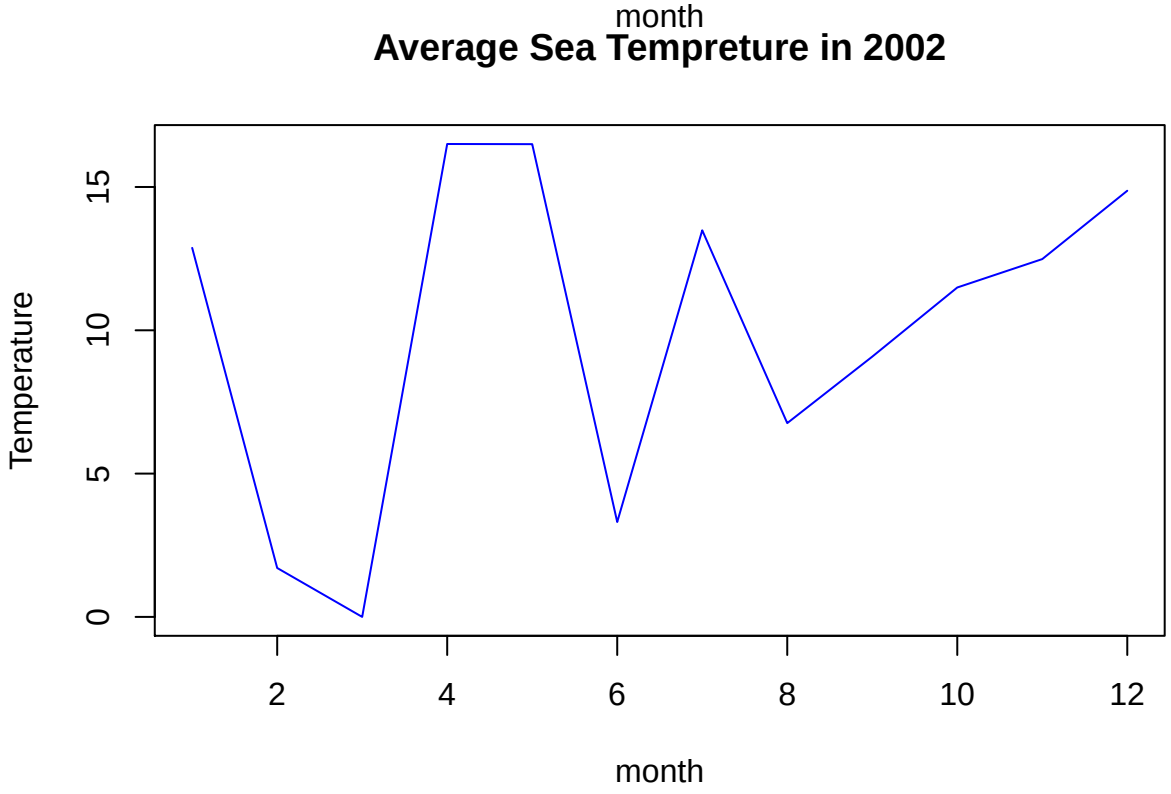

Subcontinent East - 2016
lat
long
40
30
20
10
0
60
70
80
90
100
110

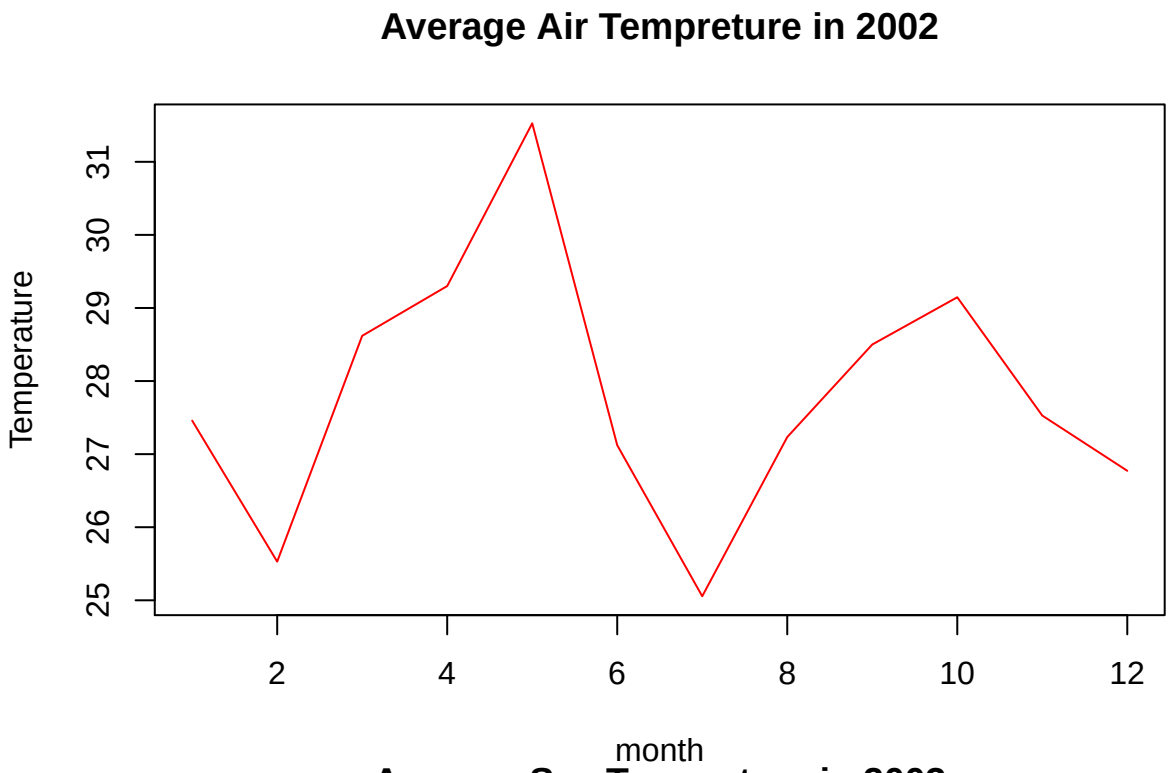
EDA

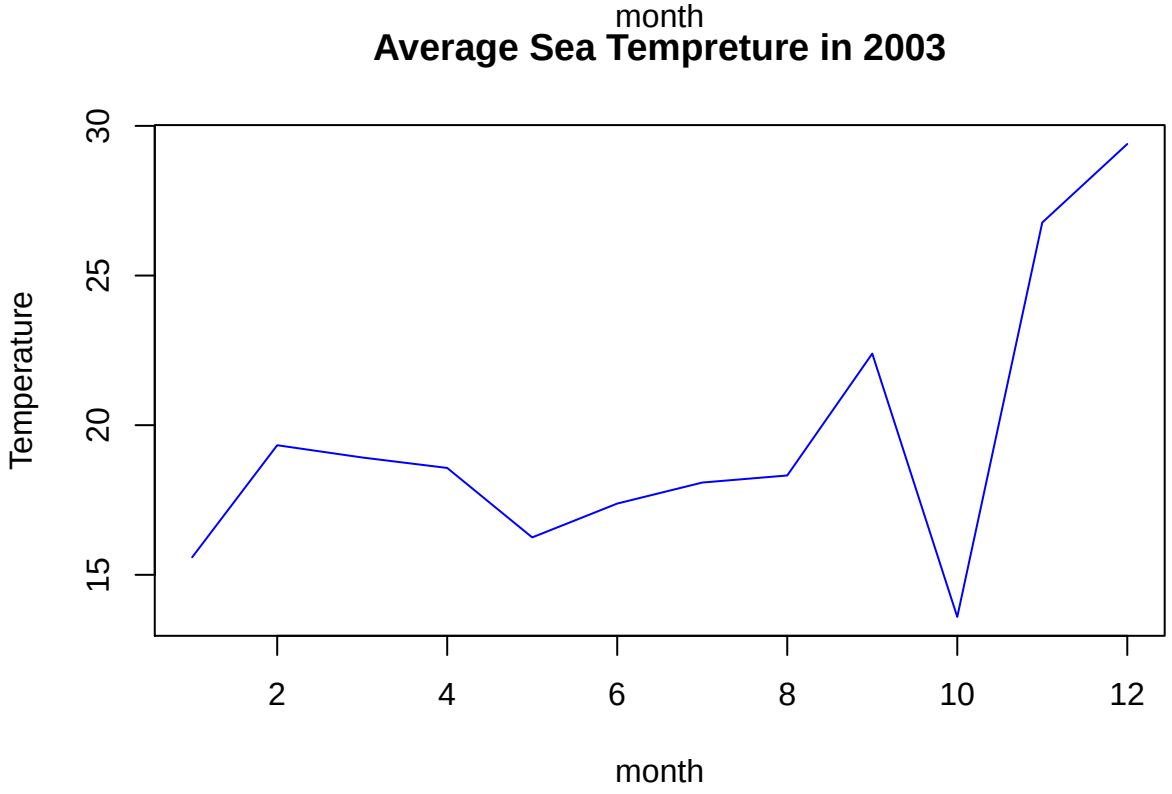
```
matrix.jan.temp = NULL
matrix.jul.temp  = NULL

#### the following loop will generate the graphs
##### and it will build two matrices to track the average temprtures of two months through out the 16 y
for (k in 2001:2016) {
  YEAR = k

  load(paste0("./cleaned_data/ave_temp_",YEAR,".Rdata"))
  plot(EDA.year[,2], type = "l",  main = paste("Average Sea Tempreture in", YEAR, sep = " ") , xlab = "
  plot(EDA.year[,3], type = "l",  main = paste("Average Air Tempreture in", YEAR, sep = " ") , xlab = "
  jan.temp = EDA.year[1,2:3 ]
  jul.temp = EDA.year[7,2:3]
  #jan.temp = cbind(YEAR, jan.temp)
  matrix.jul.temp = rbind(matrix.jul.temp , jul.temp)
  matrix.jan.temp = rbind(matrix.jan.temp , jan.temp)
}
```

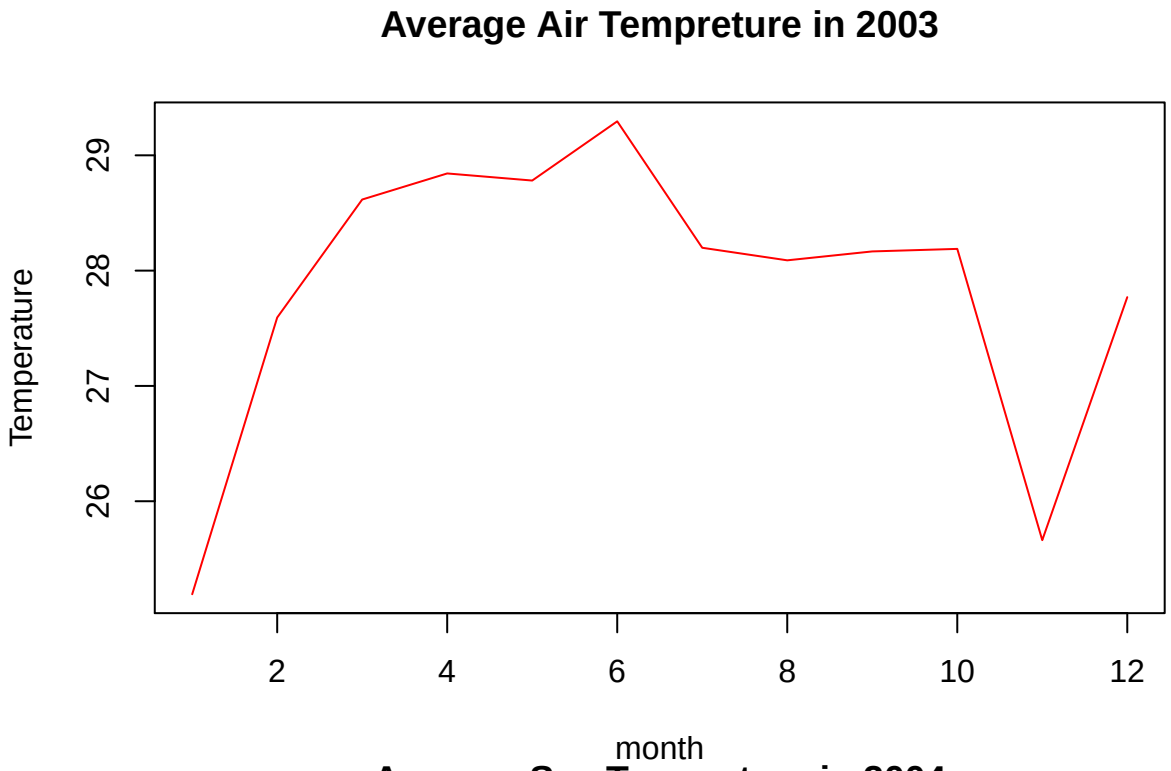


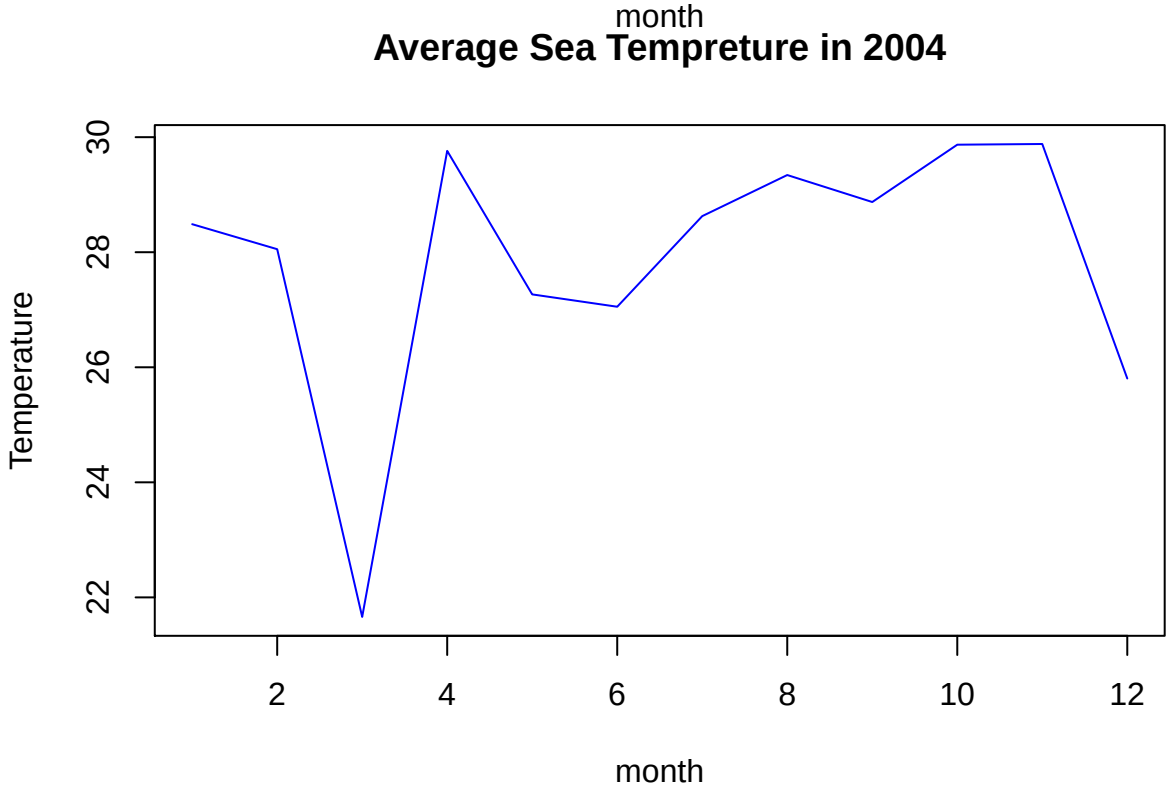

Average Air Tempreture in 2001
Temperature
20
22
24
26
28
30
2
4
6
8
10
12
month

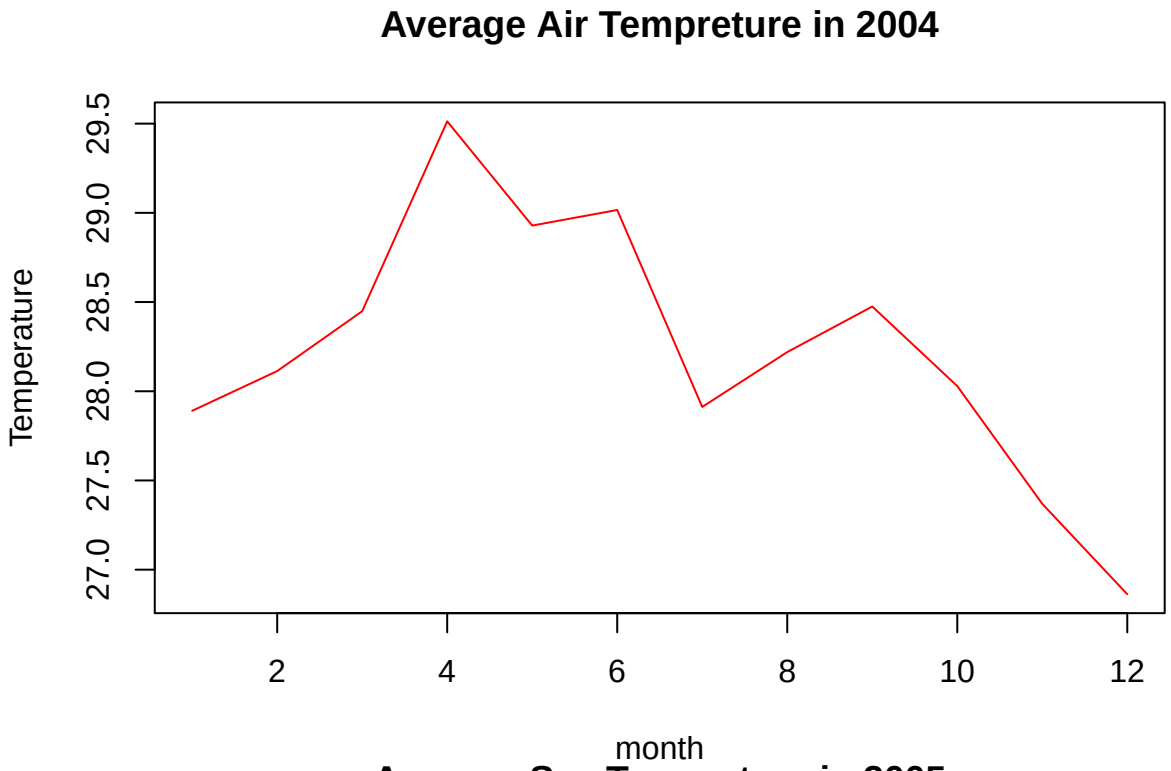

Average Sea Tempreture in 2002
Temperature
0
5
10
15
2
4
6
8
10
12
month

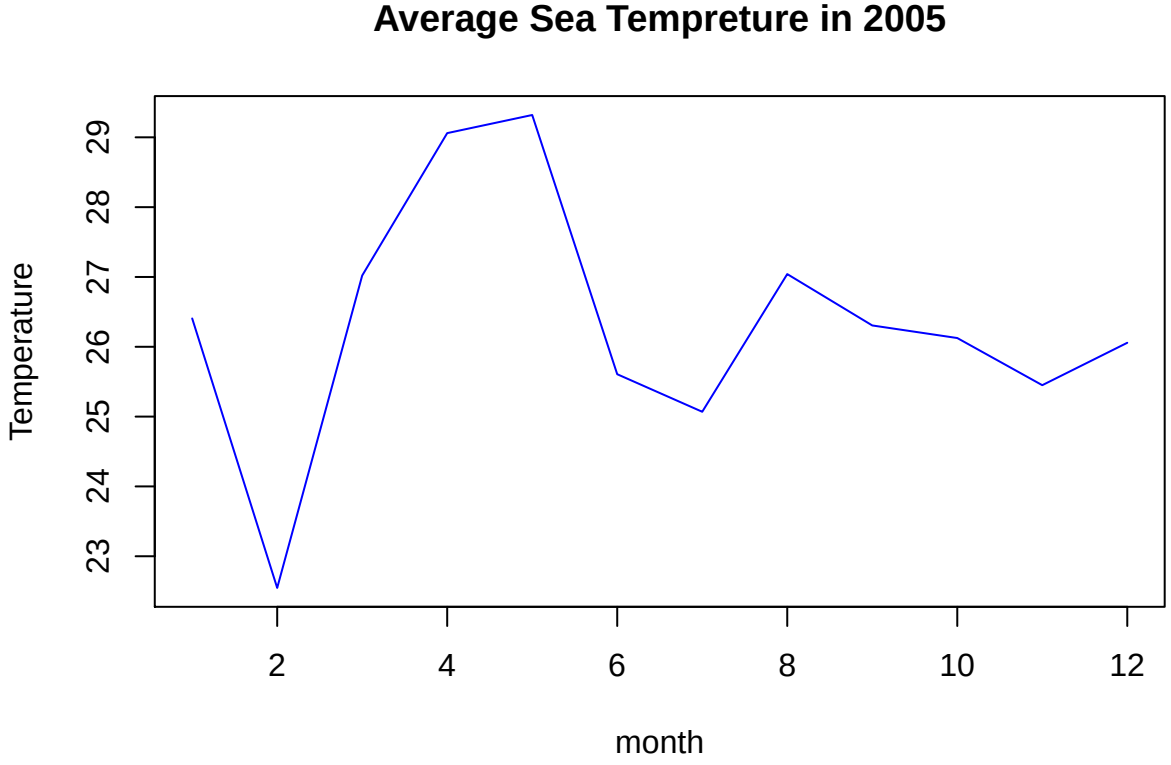

Average Air Tempreture in 2002
Temperature
25
26
27
28
29
30
31
2
4
6
8
10
12
month

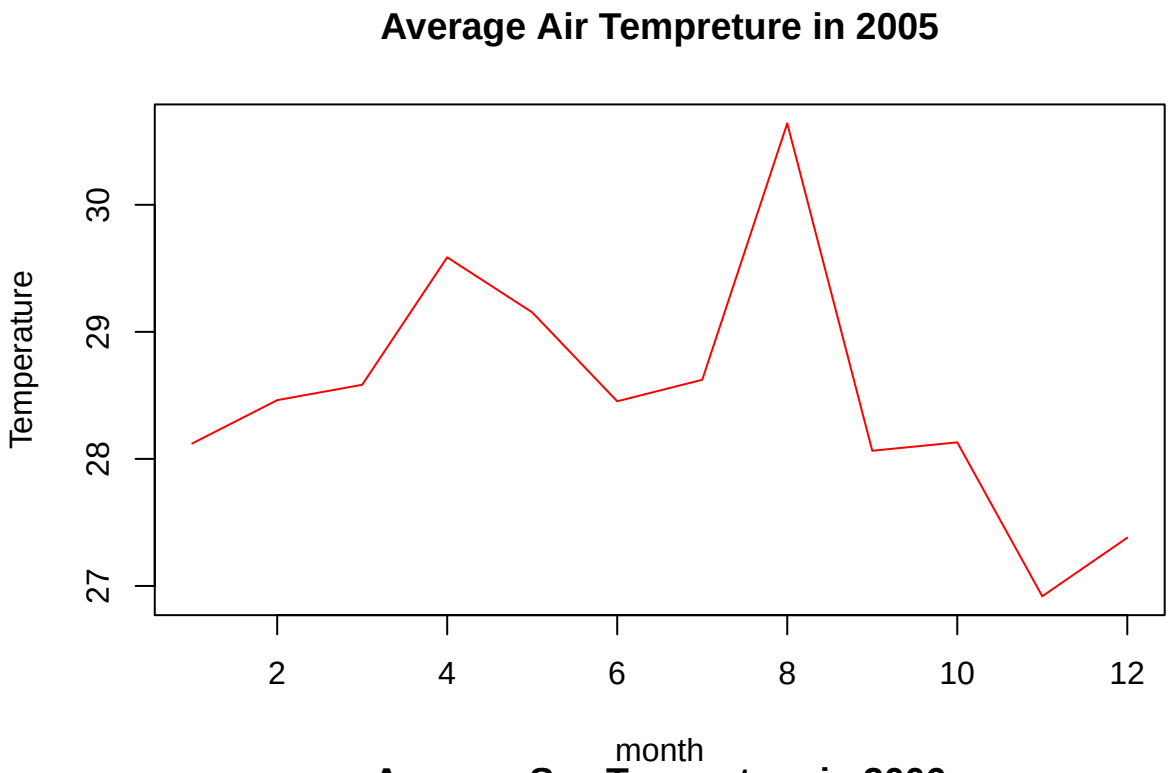

Average Sea Tempreture in 2003
Temperature
15
20
25
30
2
4
6
8
10
12
month

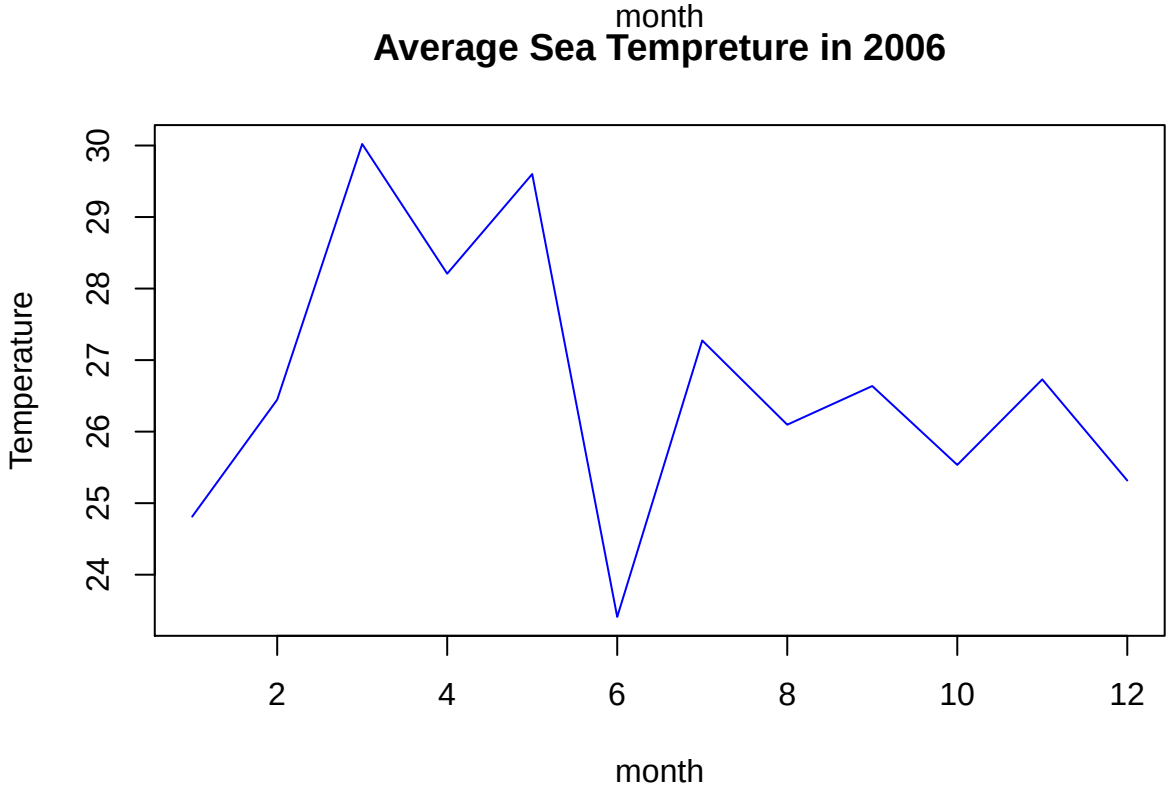

Average Air Tempreture in 2003
Temperature
26
27
28
29
2
4
6
8
10
12
month

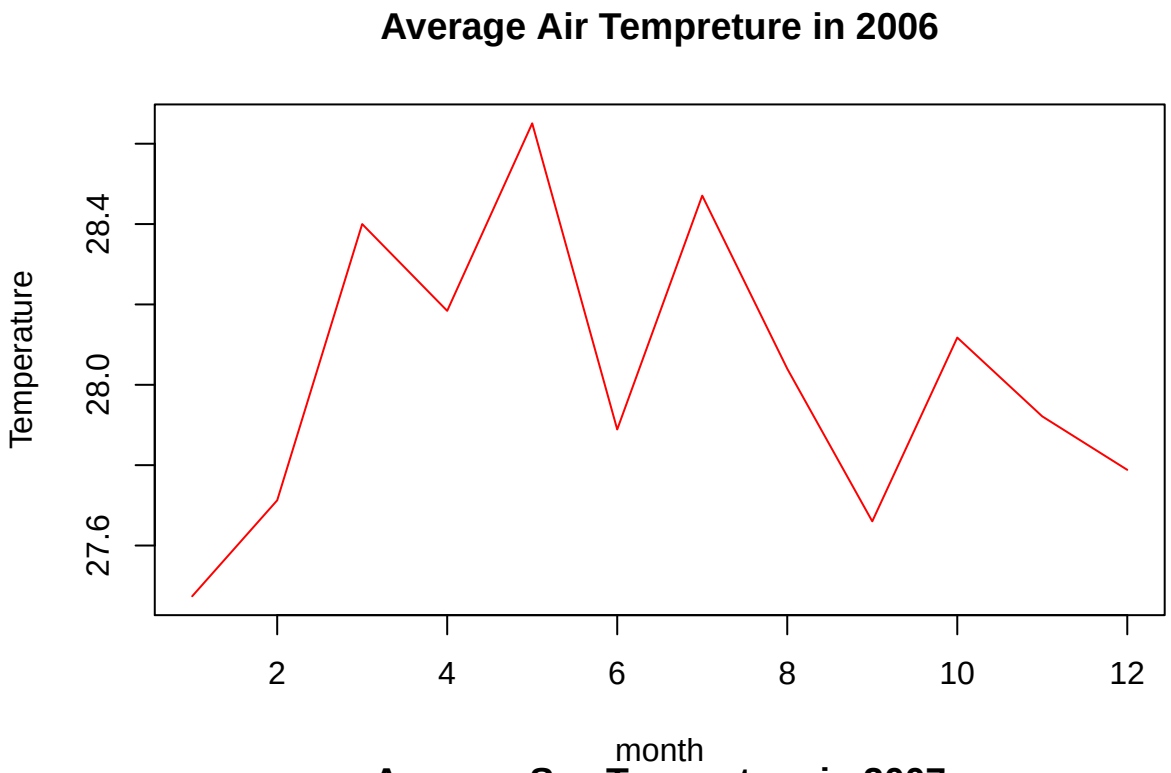

Average Sea Tempreture in 2004
Temperature
22
24
26
28
30
2
4
6
8
10
12
month

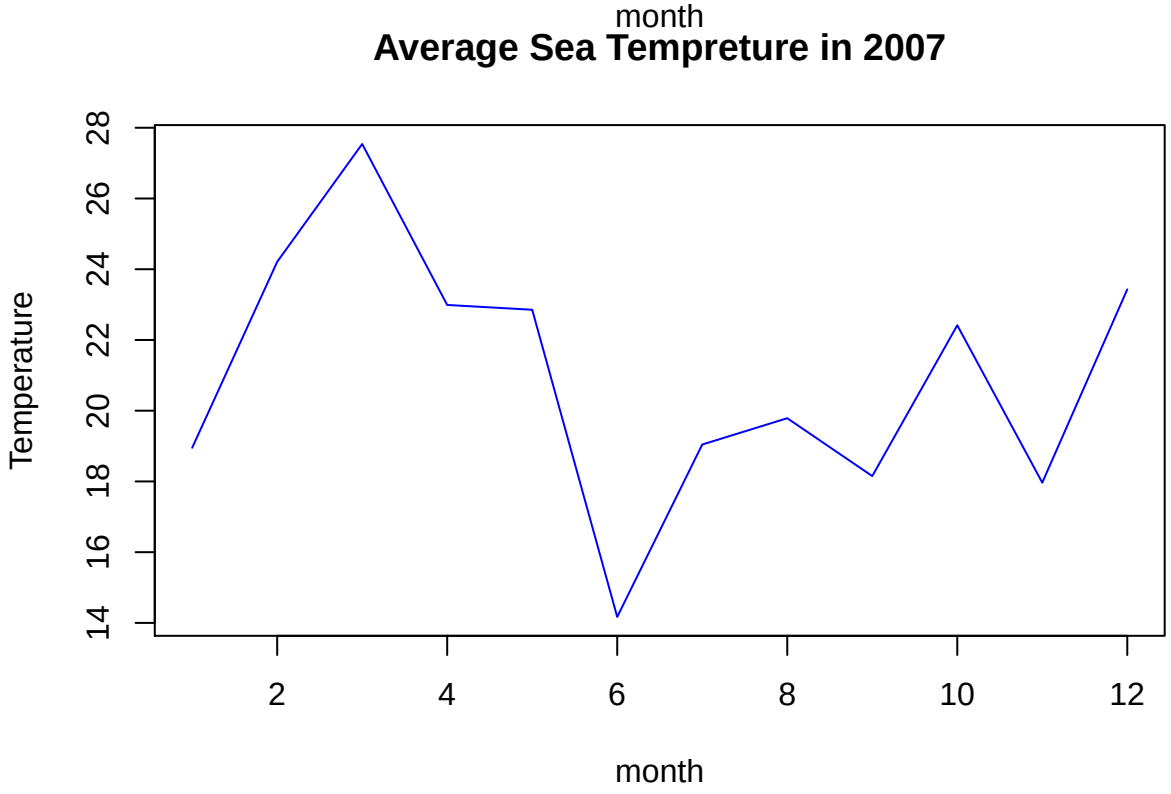

Average Air Tempreture in 2004
Temperature
27.0
27.5
28.0
28.5
29.0
29.5
2
4
6
8
10
12
month

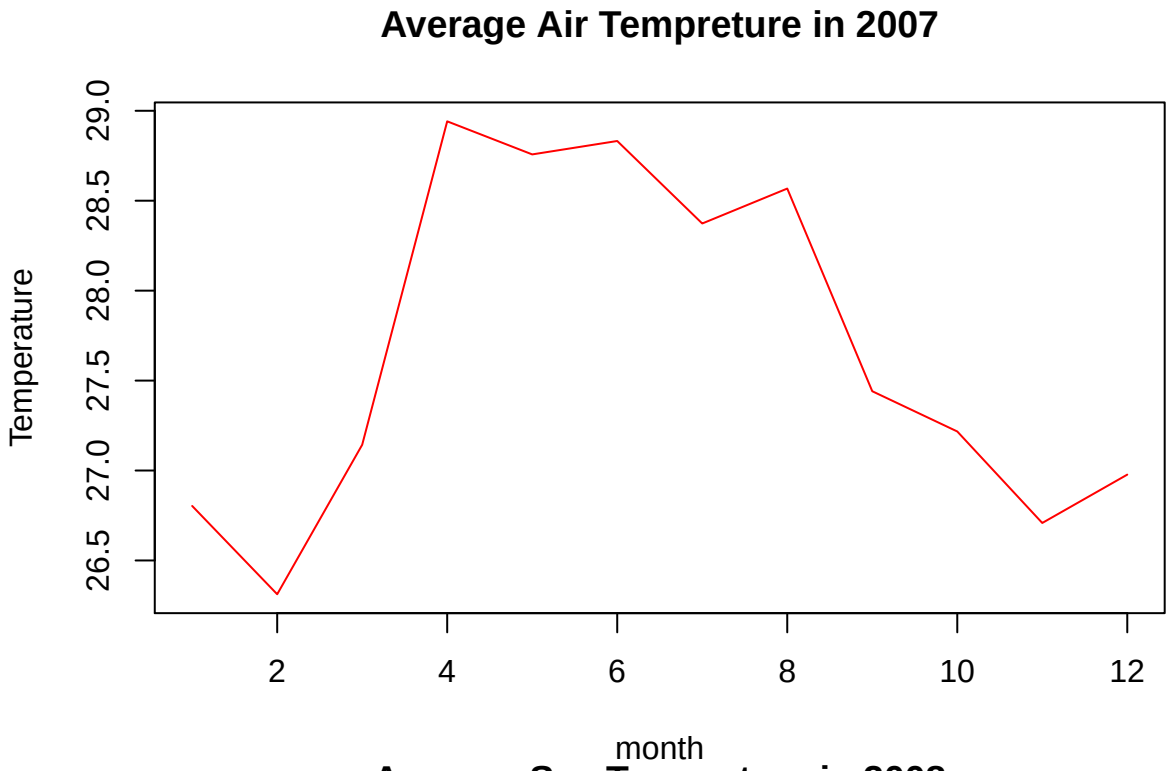

Average Sea Tempreture in 2005
Temperature
23
24
25
26
27
28
29
2
4
6
8
10
12
month

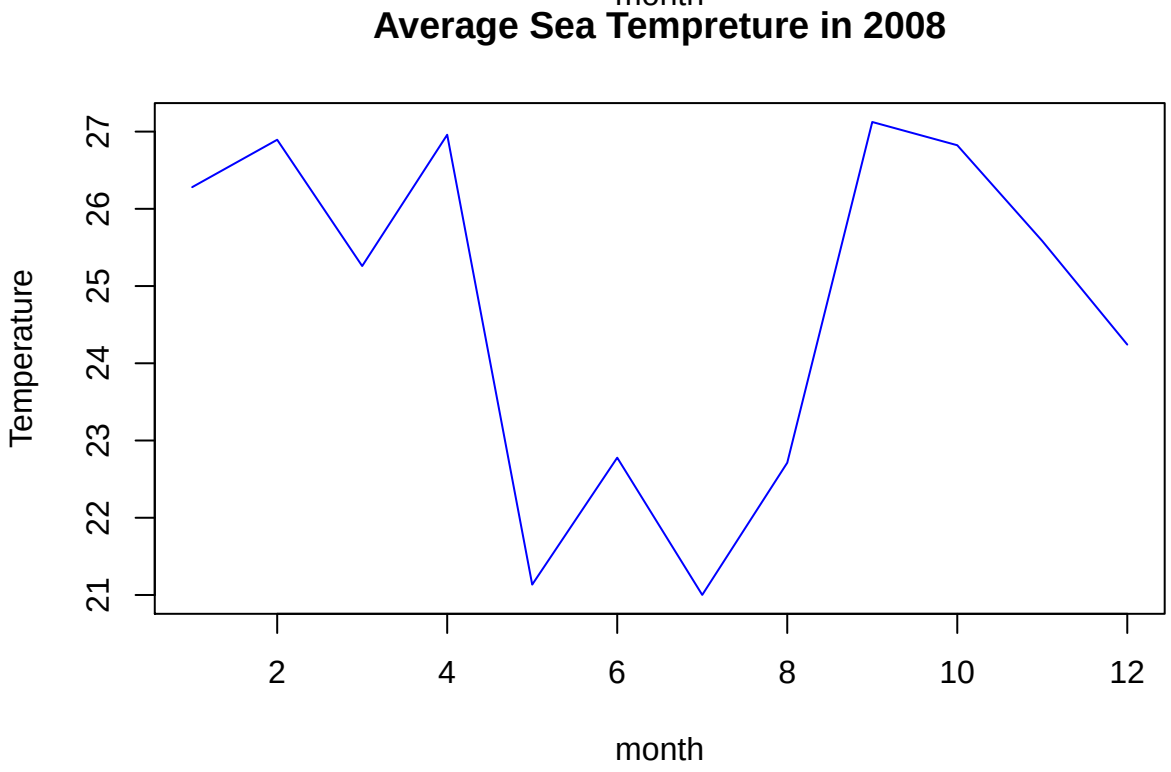

Average Air Tempreture in 2005
Temperature
27
28
29
30
2
4
6
8
10
12
month

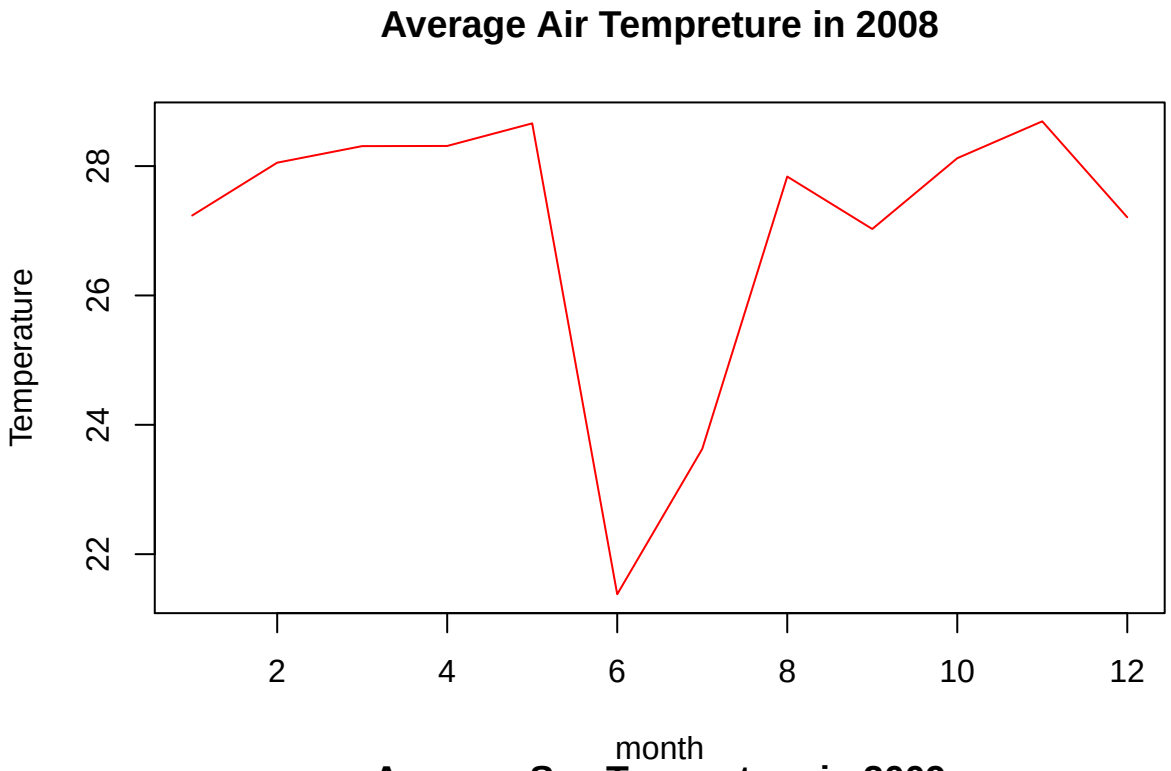

Average Sea Tempreture in 2006
Temperature
24
25
26
27
28
29
30
2
4
6
8
10
12
month

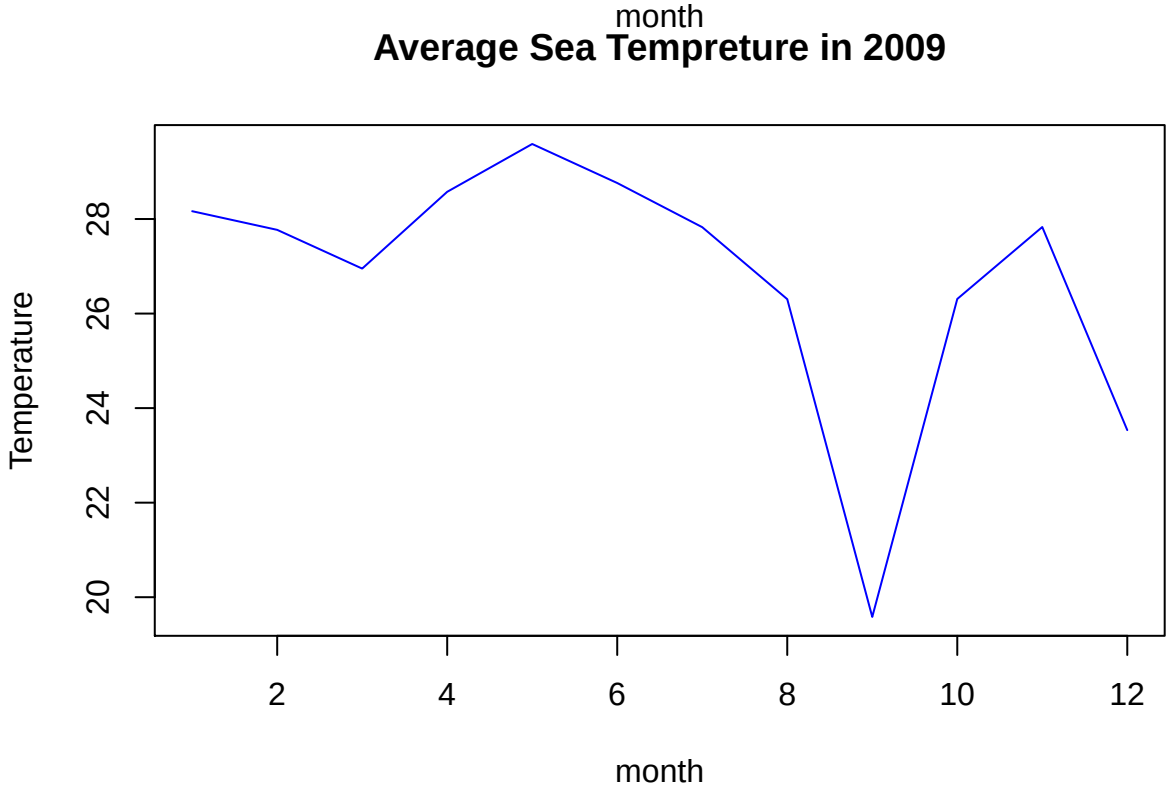

Average Air Tempreture in 2006
Temperature
27.6
28.0
28.4
2
4
6
8
10
12
month

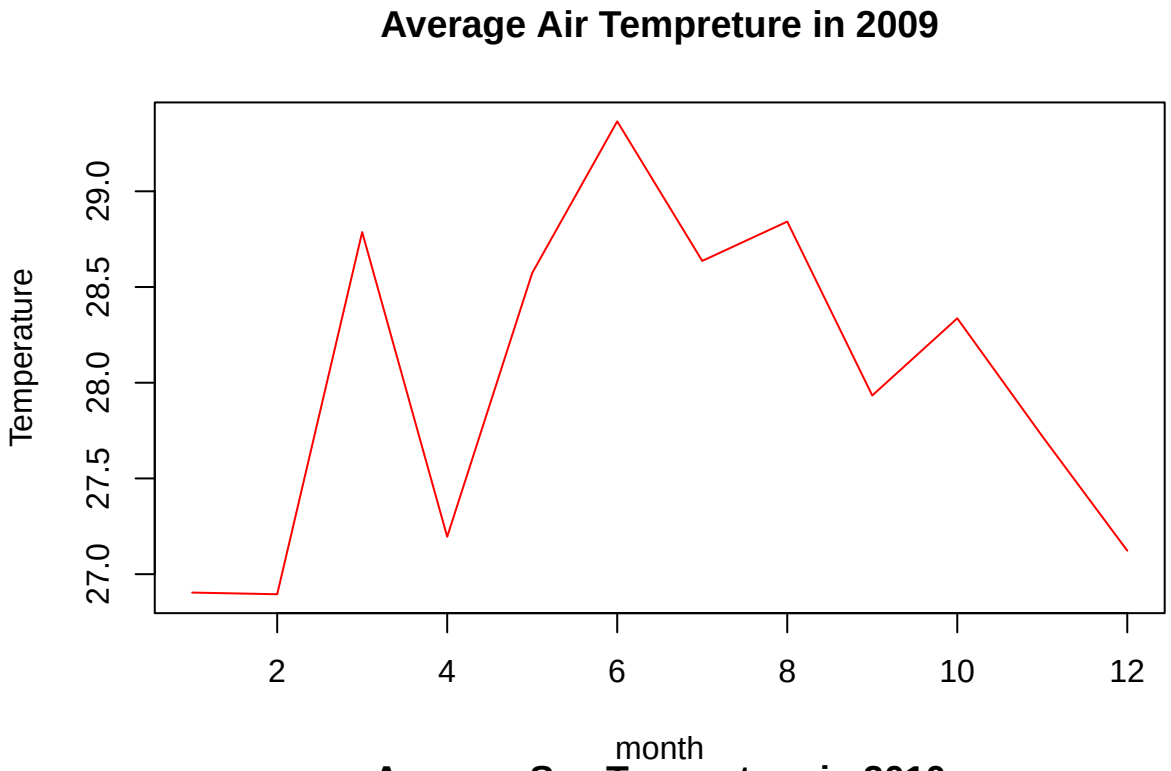

Average Sea Tempreture in 2007
Temperature
14
16
18
20
22
24
26
28
2
4
6
8
10
12
month

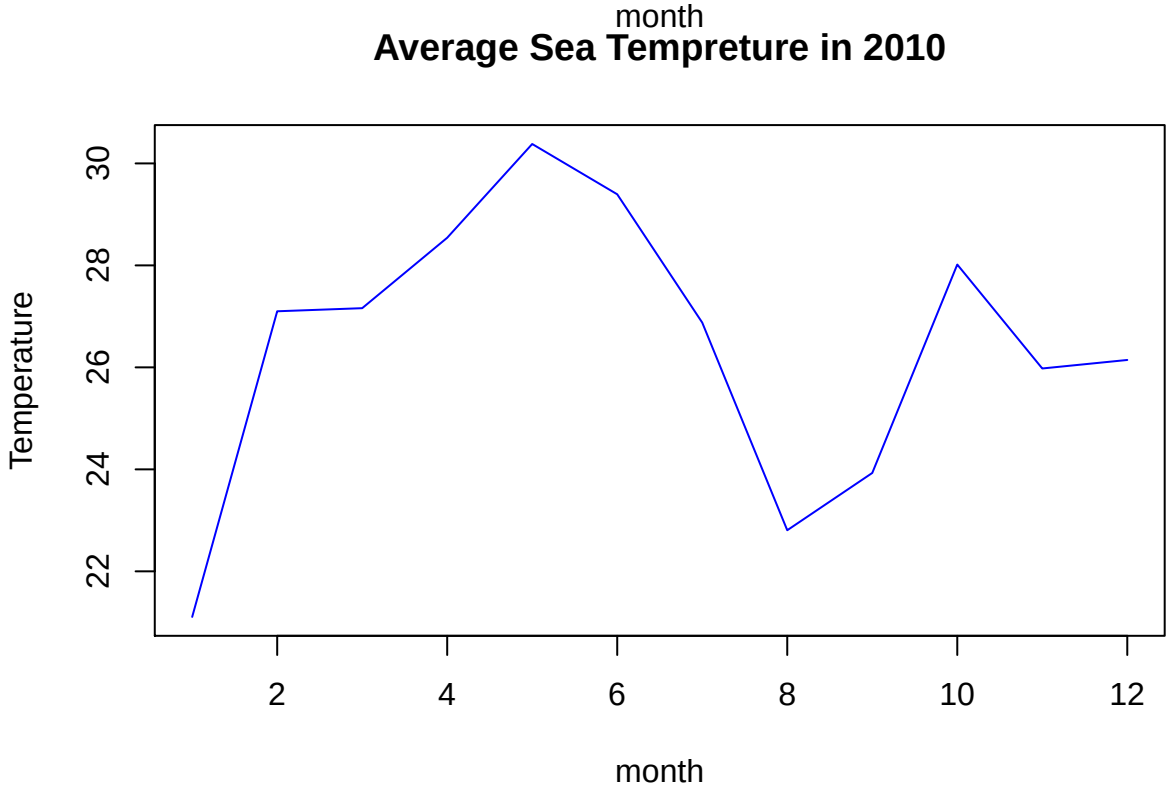

Average Air Tempreture in 2007
Temperature
26.5
27.0
27.5
28.0
28.5
29.0
2
4
6
8
10
12
month

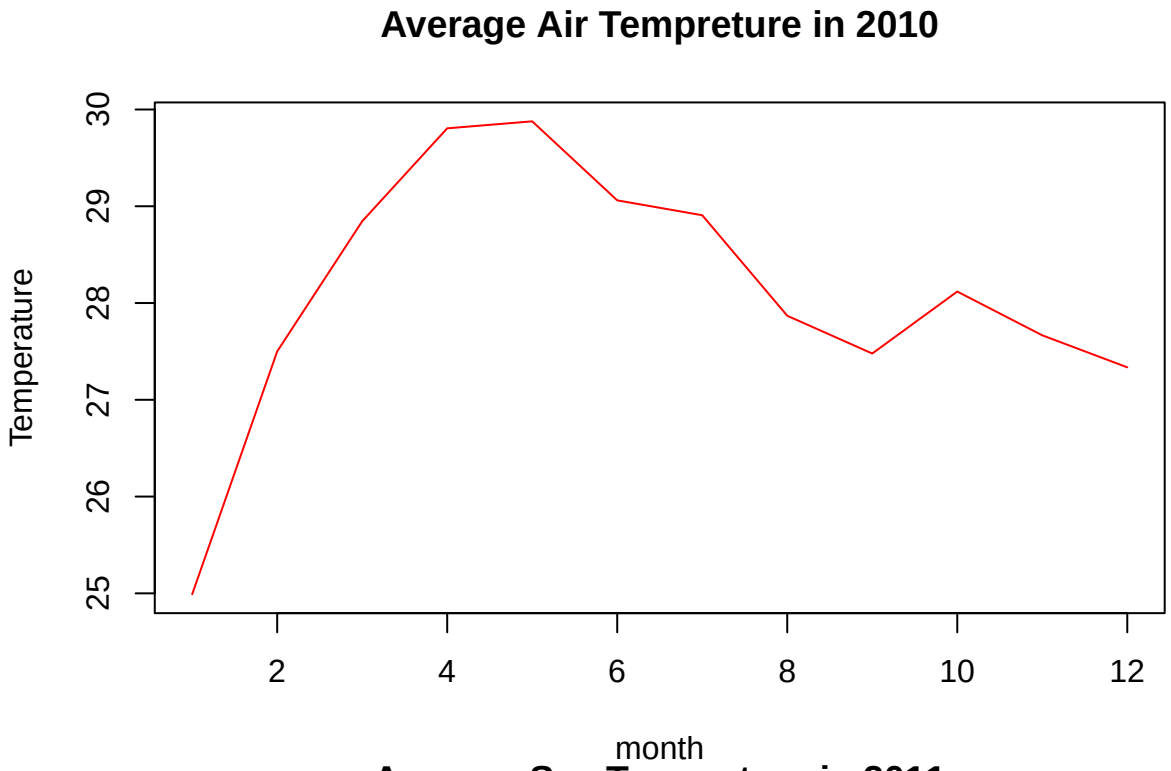

Average Sea Tempreture in 2008
Temperature
21
22
23
24
25
26
27
2
4
6
8
10
12
month

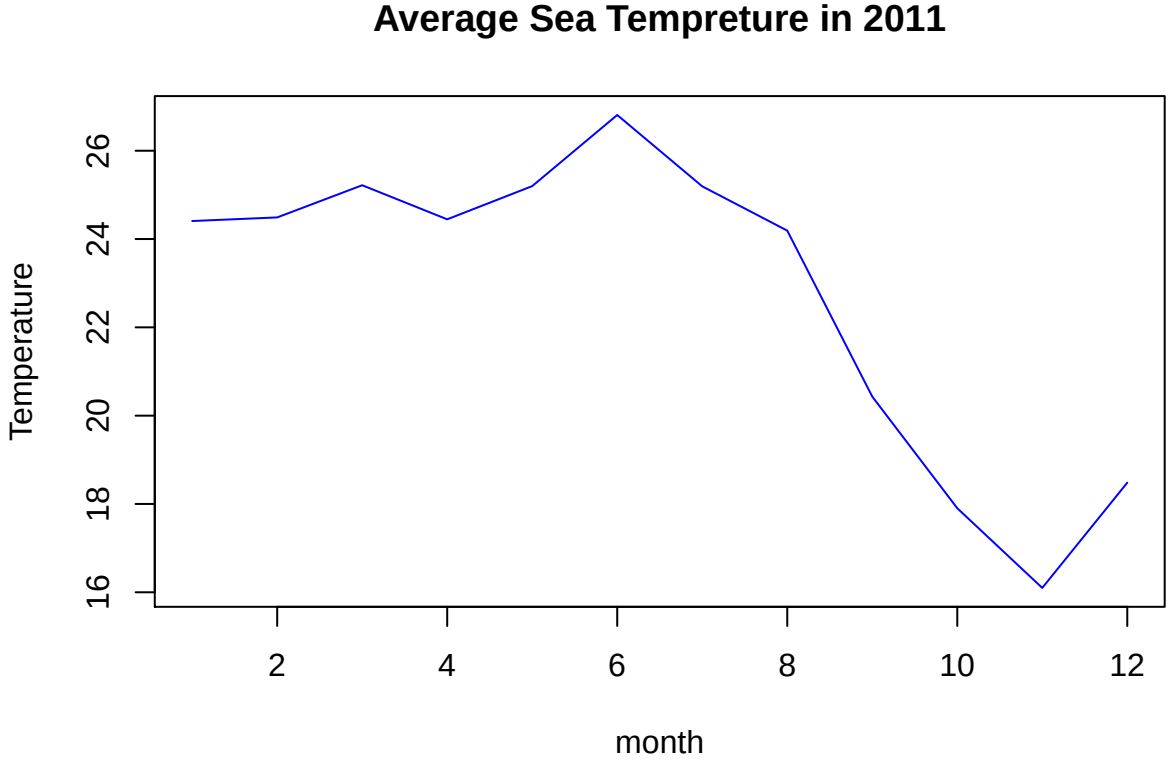

Average Air Tempreture in 2008
Temperature
22
24
26
28
2
4
6
8
10
12
month

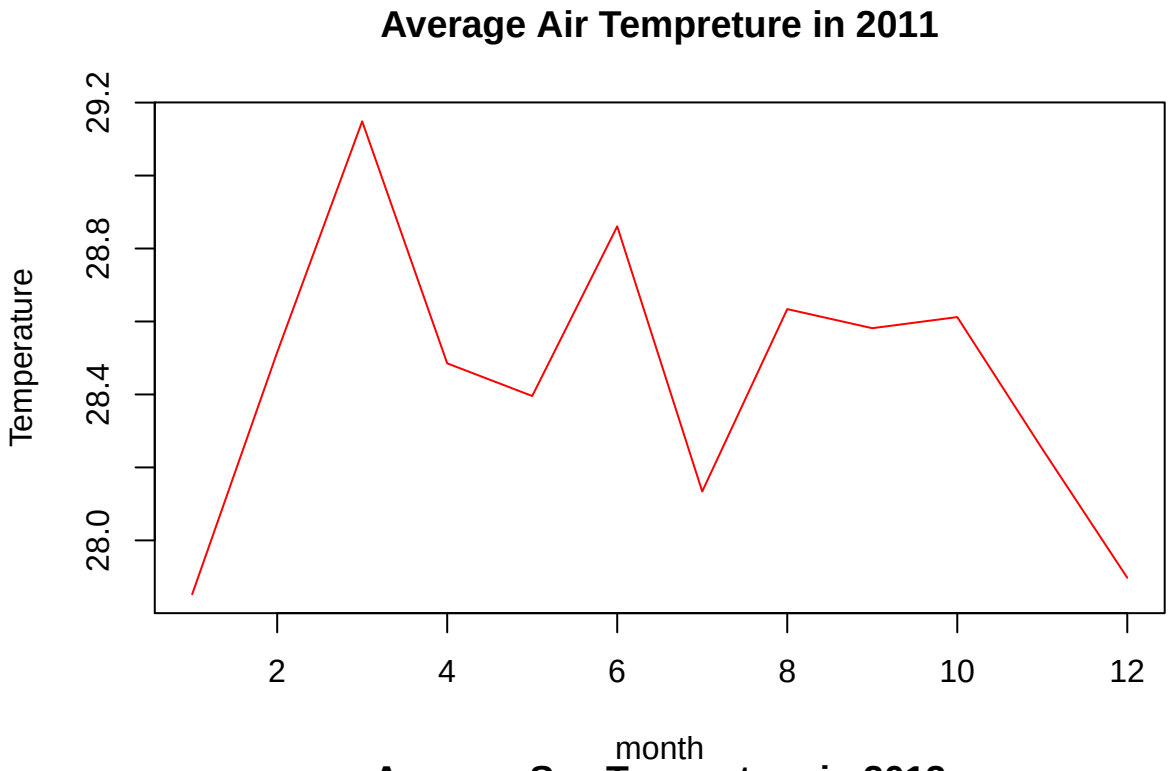

Average Sea Tempreture in 2009
Temperature
20
22
24
26
28
2
4
6
8
10
12
month

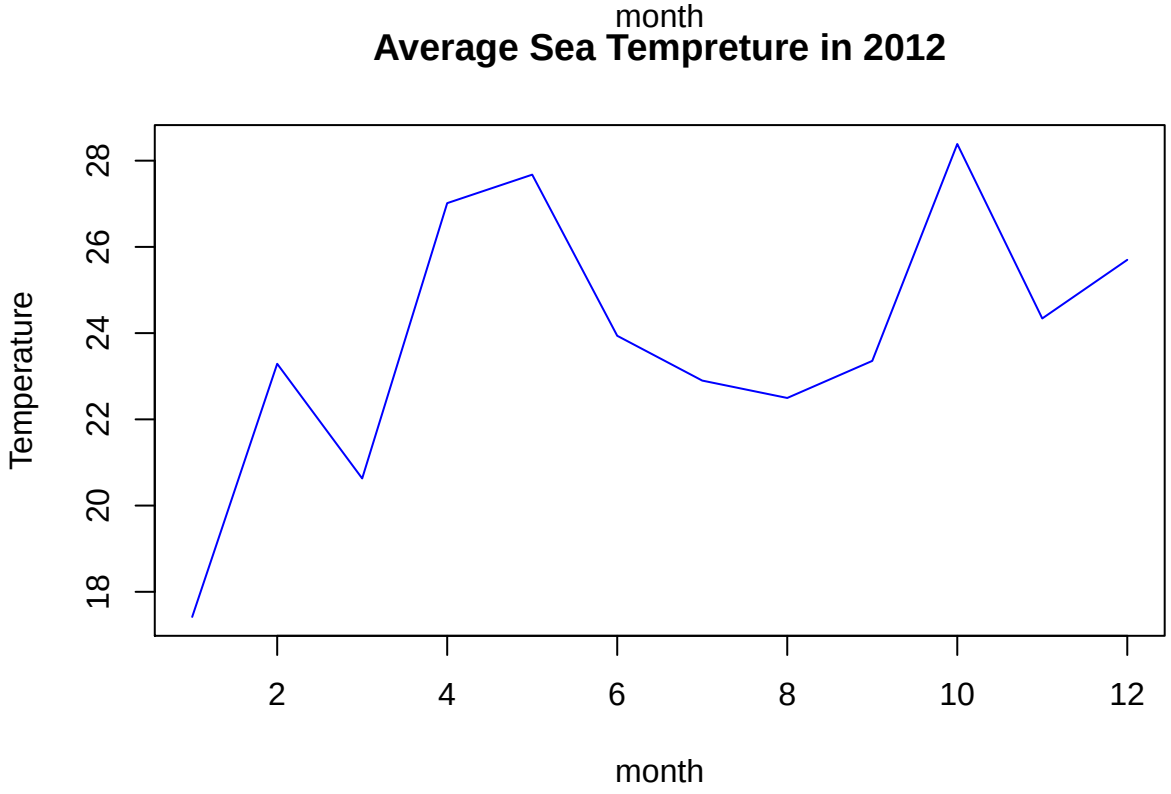

Average Air Tempreture in 2009
Temperature
27.0
27.5
28.0
28.5
29.0
2
4
6
8
10
12
month

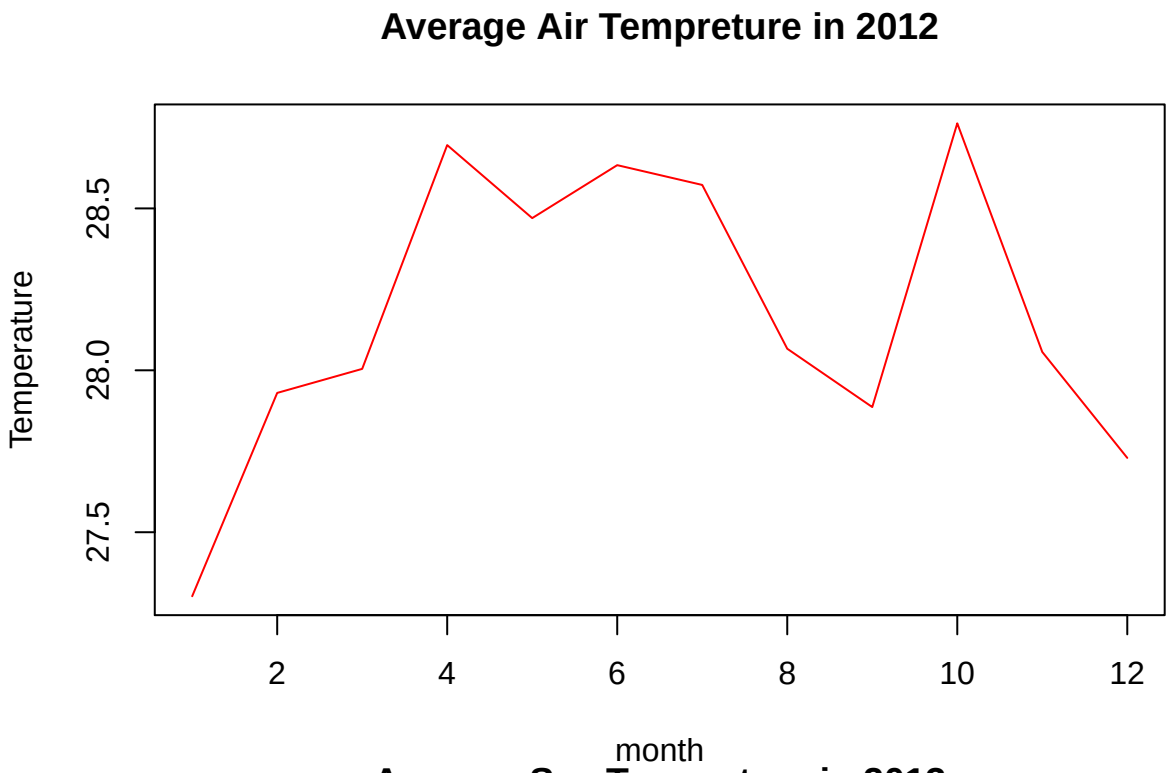

Average Sea Tempreture in 2010
Temperature
22
24
26
28
30
2
4
6
8
10
12
month

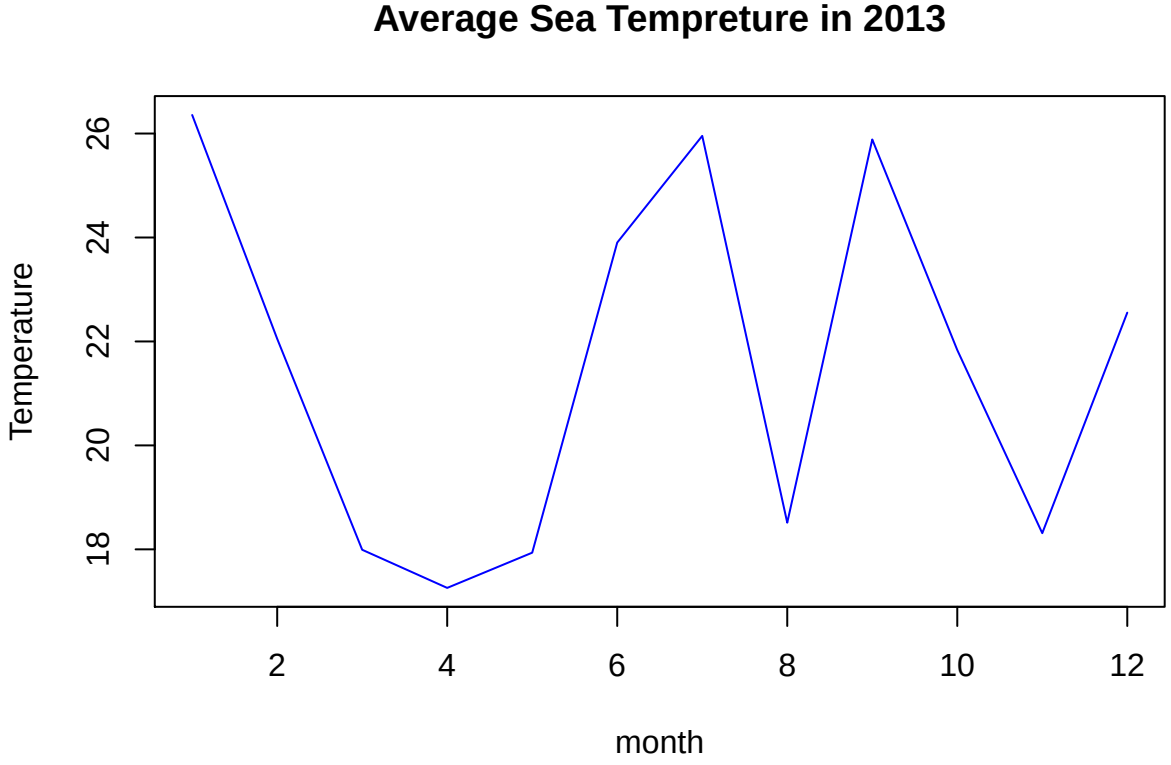

Average Air Tempreture in 2010
Temperature
25
26
27
28
29
30
2
4
6
8
10
12
month

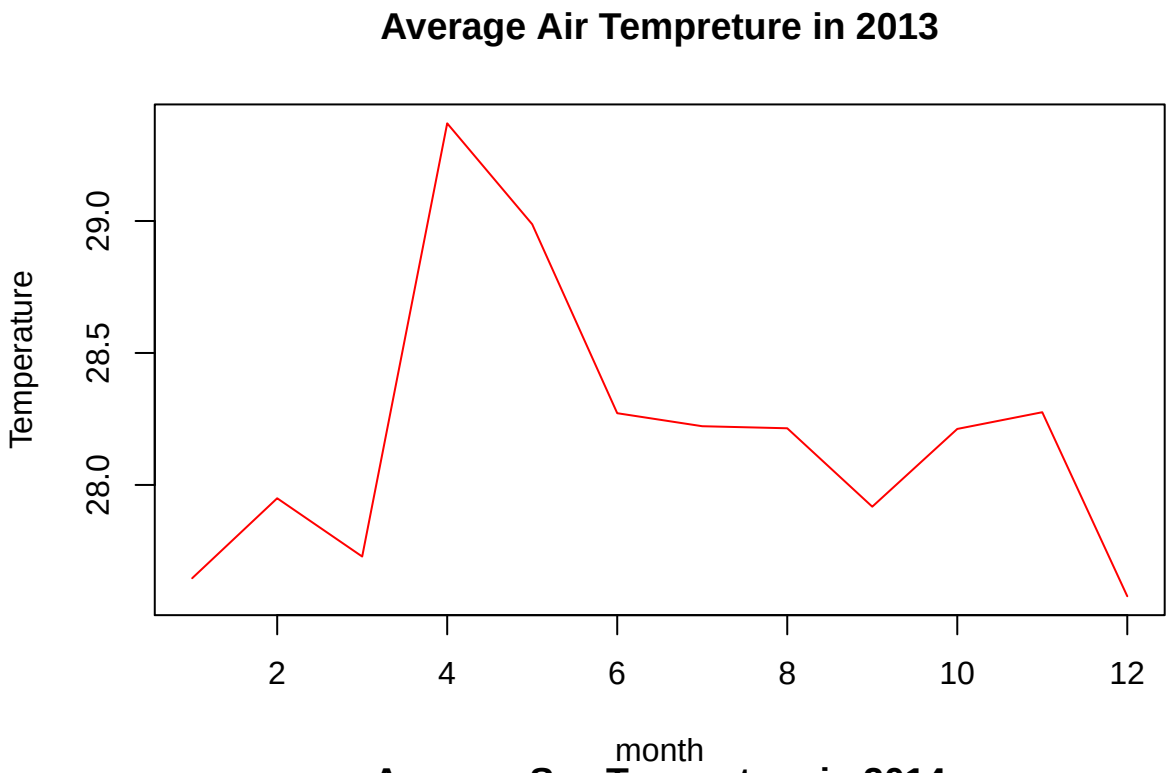

Average Sea Tempreture in 2011
Temperature
16
18
20
22
24
26
2
4
6
8
10
12
month

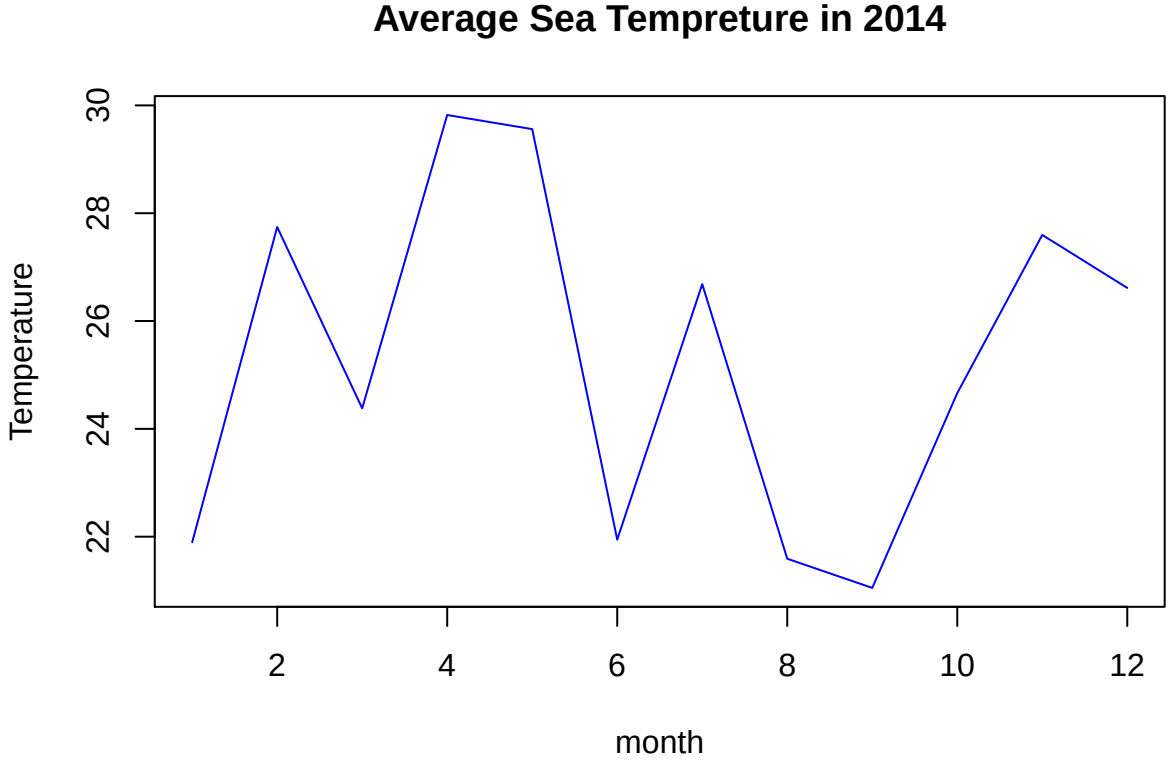

Average Air Tempreture in 2011
Temperature
28.0
28.4
28.8
29.2
2
4
6
8
10
12
month

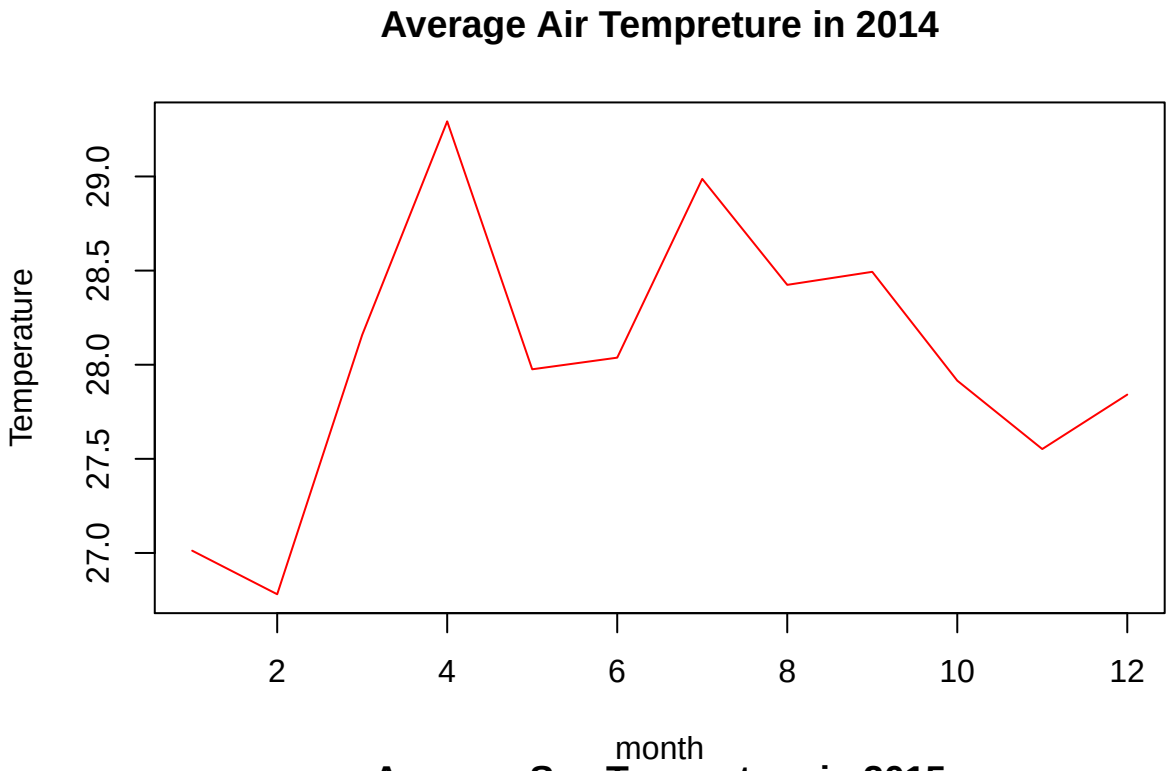

Average Sea Tempreture in 2012
Temperature
18
20
22
24
26
28
2
4
6
8
10
12
month

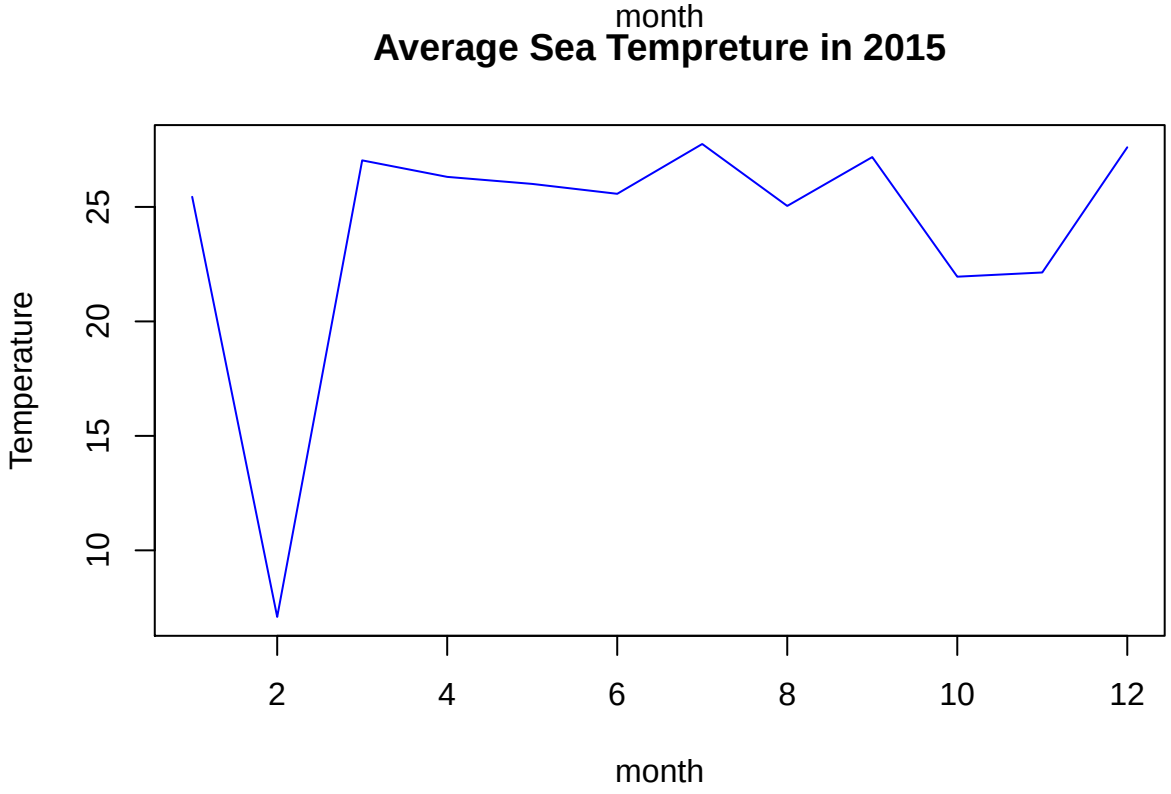

Average Air Tempreture in 2012
Temperature
27.5
28.0
28.5
2
4
6
8
10
12
month

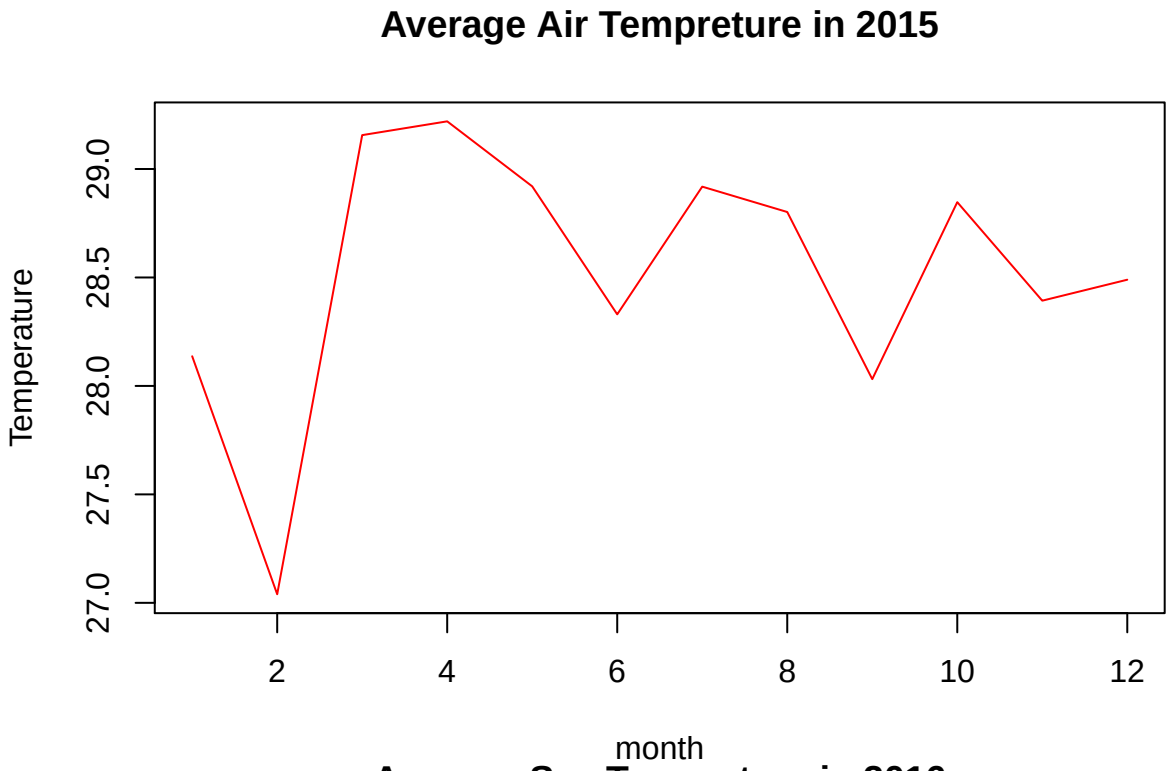

Average Sea Tempreture in 2013
Temperature
18
20
22
24
26
2
4
6
8
10
12
month

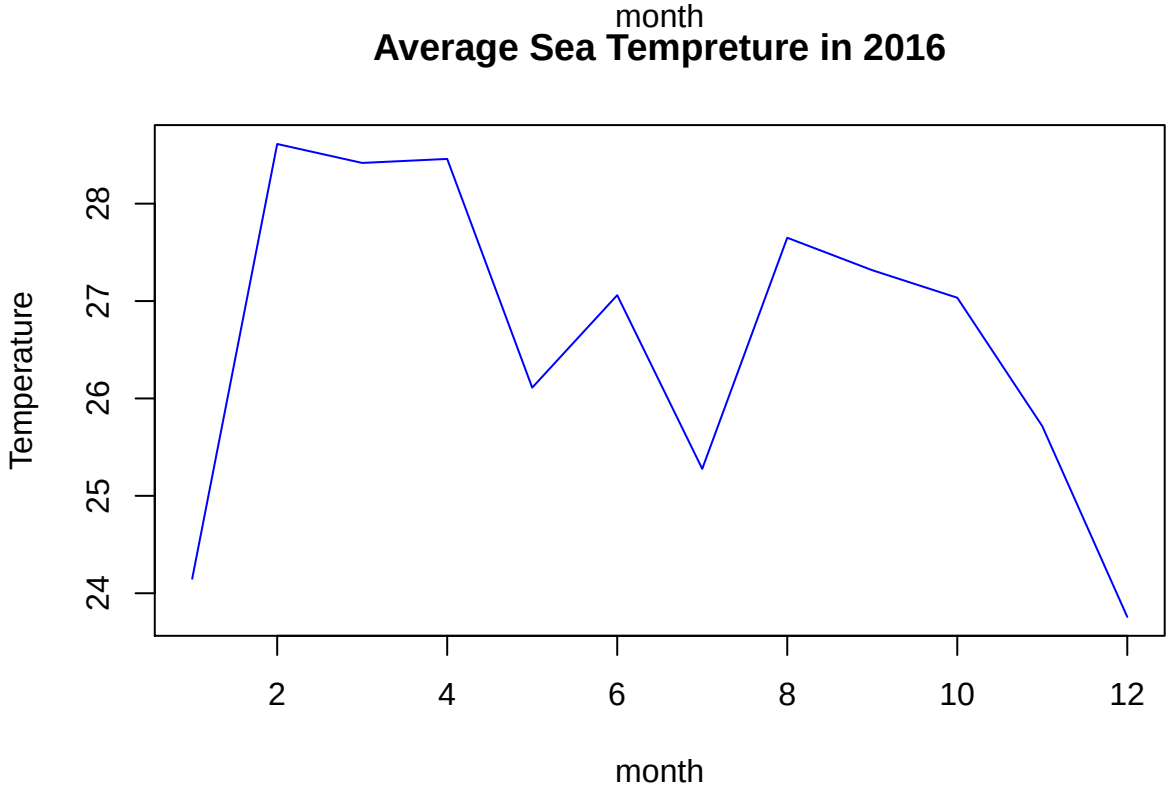

Average Air Tempreture in 2013
Temperature
28.0
28.5
29.0
2
4
6
8
10
12
month

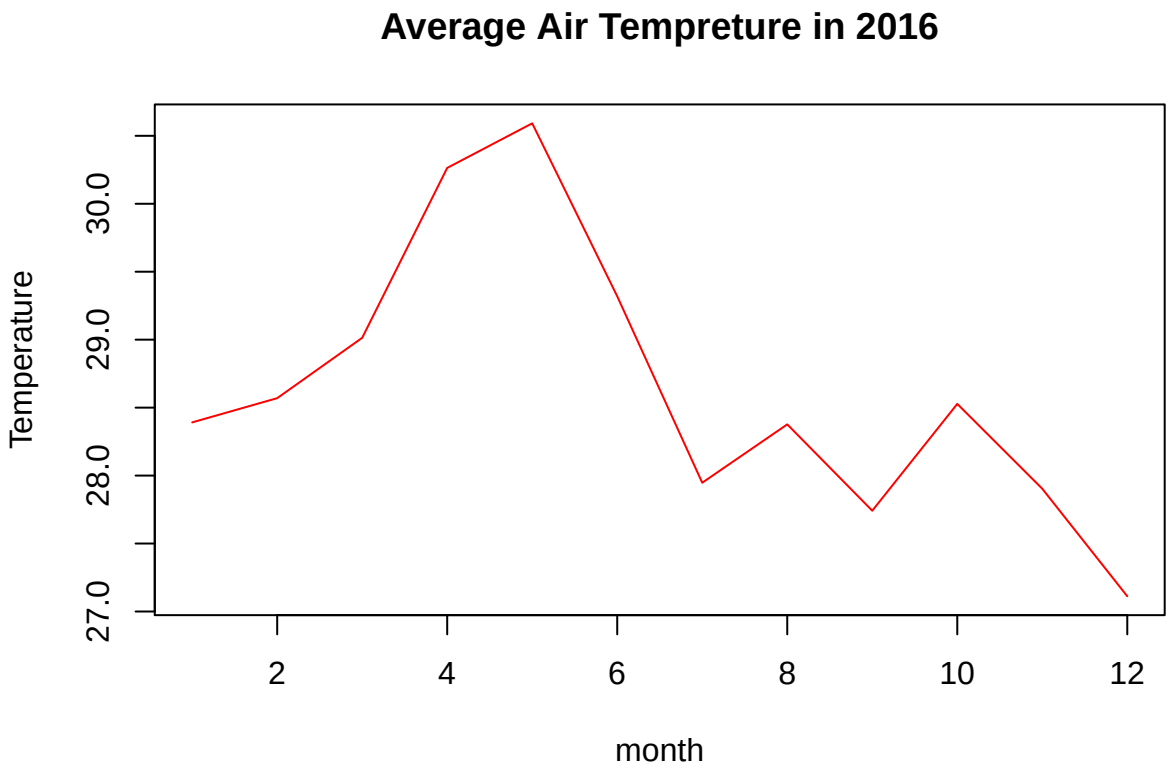

Average Sea Tempreture in 2014
Temperature
22
24
26
28
30
2
4
6
8
10
12
month


Average Air Tempreture in 2014
Temperature
27.0
27.5
28.0
28.5
29.0
2
4
6
8
10
12
month


Average Sea Tempreture in 2015
Temperature
10
15
20
25
2
4
6
8
10
12
month


Average Air Tempreture in 2015
Temperature
27.0
27.5
28.0
28.5
29.0
2
4
6
8
10
12
month


Average Sea Tempreture in 2016
Temperature
24
25
26
27
28
2
4
6
8
10
12
month




```
colnames(matrix.jan.temp) = c("SEA" , "AIR")
rownames(matrix.jan.temp) = c("2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","20

colnames(matrix.jul.temp) = c("SEA" , "AIR")
rownames(matrix.jul.temp) = c("2001","2002","2003","2004","2005","2006","2007","2008","2009","2010","20
```

Average Month Plots

In this section, we create separate plots for each year both for the sea temperature and the air temperature. The plots are structured so that average temperature for each month is calculated and each data point on the plot shows the average temperature in each month. The lowest temperatures at sea are usually observed between December and January and for the rest of the months on average we are looking at slightly higher temperature. Since Subcontinental East is close to the equator and the sunshine is overhead the range of the temperatures is pretty condensed and small. Between 2001 and 2016 this range has been around 15 degrees Celsius for the sea temperature. However, for the air temperature we observe that the range is even more limited around 10 degrees all year round. *notice: The sharp points that indicate extremely cold temperatures indicate months that have missing data that the average temperature could not be calculated.