

# Laboratory with R

Enrique Akira Takiguchi and Itziar Salazar

2025-12-31

**Problem statement:** Europe is affected by climate change and its effects are not only perceived on earth. European water bodies (lakes, rivers and oceans and seas of the continent) are also affected. Given that there is more water than land on the planet's surface, it is not surprising that the warming of the oceans has accounted for about 93 % of the global warming since the 1950s. This warming occurs as a result of the increase in greenhouse gas emissions, especially carbon dioxide, which in turn trap more and more solar energy within the atmosphere. Most of this trapped heat is eventually stored in the oceans, which affects the temperature and the circulation of water. Sea surface temperatures on European coasts are rising more rapidly than those in the world oceans. Water temperature represents one of the most important regulatory elements of marine life. Temperature increases are already causing major changes under the surface of the water

---

## Exercise 1

### 1.a,b) Data importation, variable specification, and metadata definition

The data set `sea_data` is **directly sourced** from IDESCAT, with the variable names set as per the original dataset. Technical **metadata**, including measurement units and descriptions of variables such as years, months, and depth, has been **added** using the `LABEL` function of the `HMISC` package to tag variables.

The data set shows the temperature of the sea in degrees Celsius at depths of 0, -20, -50, and -80 meters between 2000 and 2017. These measurements were taken at L'Estartit observation point on the Costa Brava, which is one mile east of the Medes Islands (Girona). The coordinates of this location are 42°03' N and 3°15' E.

### 1.c) Data structure

The data frame `sea.deep` is checked to make sure it is complete and correct. The internal organization of the data is inspected to **verify** that **data types** and the previously assigned **attributes** in *1.a,b* are correctly integrated. We make a **summary** of the data so we can see at a glance what the main values are, and how far they range. We also check that there are no **NAs** in the dataset.

```
## Dimension df_yearly
## Rows (obs.): 936
## Columns (var.): 4
## 'data.frame':   936 obs. of  4 variables:
## $ mes          : chr  "January" "January" "January" "January" ...
## $ fondaria     : num  0 -20 -50 -80 0 -20 -50 -80 0 -20 ...
## $ any          : int  2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...
## $ temperatura: num  12.9 12.9 12.8 12.7 12.7 12.7 12.6 12.4 12.8 12.7 ...
## NULL
```

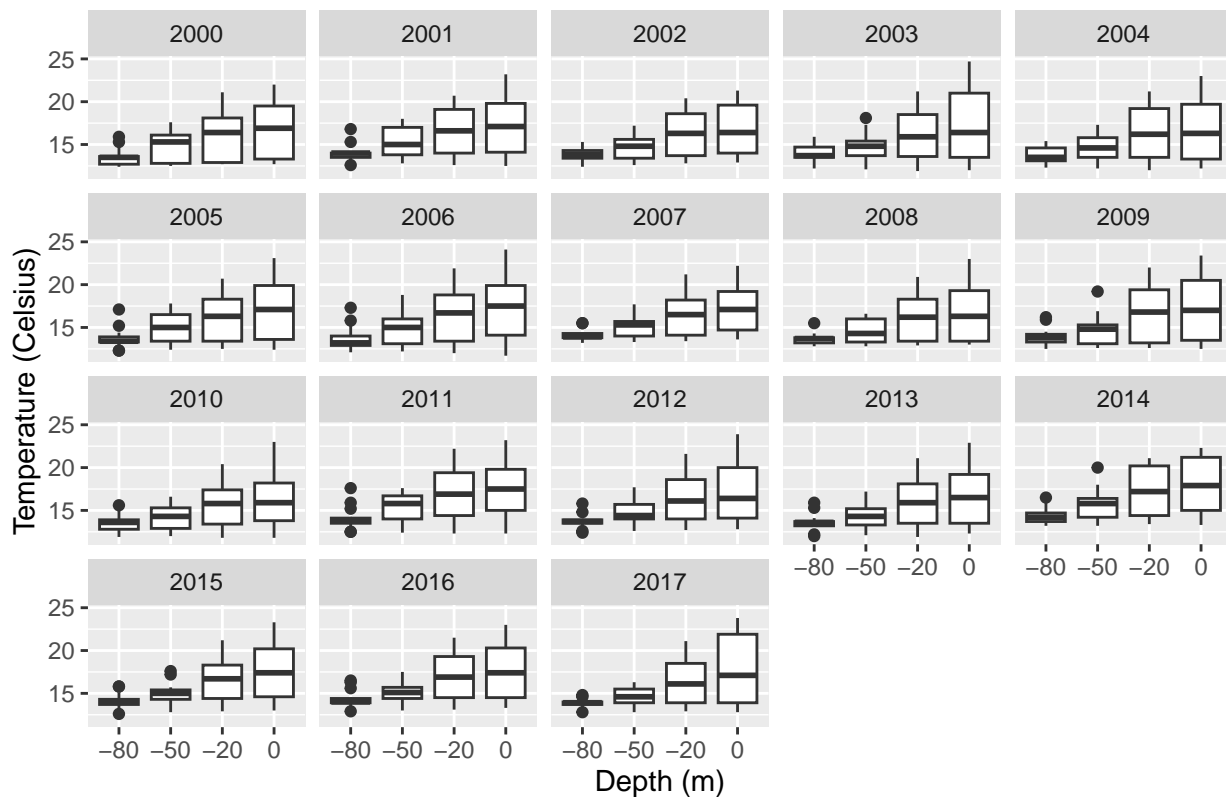
```
##      mes      fondaria      any      temperatura
## Length:936      Min.    :-80.0      Min.    :2000      Min.    :11.70
## Class :character 1st Qu.: -57.5      1st Qu.:2004      1st Qu.:13.50
## Mode  :character Median  :-35.0      Median :2008      Median :14.80
##                Mean   :-37.5      Mean   :2008      Mean   :15.63
##                3rd Qu.: -15.0      3rd Qu.:2013      3rd Qu.:17.10
##                Max.    :  0.0      Max.    :2017      Max.    :24.70

## [1] 0
```

## 1.d) Visual distribution

The way the **seawater temperature changes over time** can be seen in the box-plots, which are divided into categories based on depth and year. This picture helps us to spot changes in temperature, patterns over time, and anything unusual across depths from 2000 to 2017.

Boxplot of temperature by depth (2000–2017)



## 1.e) Quantitative summary

A **quantitative characterization** is done by calculating the mean, median, standard deviation, and interquartile range for each group. This analysis is **supported by other indicators**, such as the **absolute range** and the **coefficient of variation**, which provide a strong assessment of both dispersion and relative variability.

The statistical summary is done in two stages. First, it looks at how the temperature changes with depth and year. Then, it puts these average temperatures together across the whole period (2000–2017). These estimators help us to compare the **central tendency** and the **stability of the vertical temperature distributions** at the specific depths recorded.

```
## # A tibble: 72 x 10
```

```
##      fondaria  any Mitjana Mediana  SD  IQR  Min  Max Range  CV
##      <dbl> <int>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1      -80  2000    13.6    13.5 1.12  0.9   12.4  15.9   3.5 0.0823
##  2      -80  2001    14.0     14  1.05  0.5   12.6  16.8   4.2 0.0752
##  3      -80  2002    13.8    13.8 0.783 0.9   12.4  15.3   2.9 0.0566
##  4      -80  2003    13.9    13.7 1.09  1.2   12.2  15.9   3.7 0.0785
##  5      -80  2004    13.7    13.5 1.03  1.5   12.3  15.4   3.1 0.0750
##  6      -80  2005    13.7    13.4 1.28  0.600 12.3  17.1   4.8 0.0930
##  7      -80  2006    13.8    13.2 1.53  1.1   12.1  17.3   5.2 0.111
##  8      -80  2007    14.1    13.9 0.716 0.5   13.2  15.5   2.3 0.0509
##  9      -80  2008    13.7    13.7 0.707 0.600 12.8  15.5   2.7 0.0518
## 10      -80  2009    13.9    13.8 1.12  0.900 12.5  16.2   3.7 0.0804
## # i 62 more rows

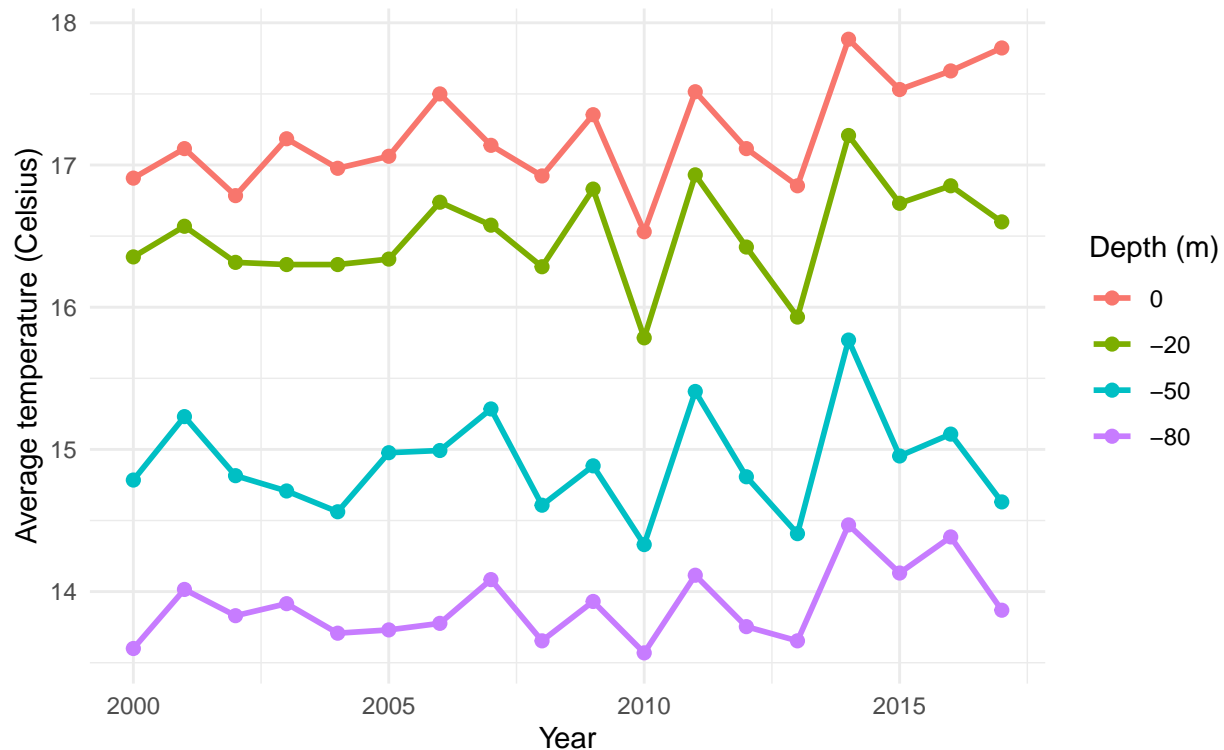
## # A tibble: 72 x 3
##      fondaria  any temp_mitja
##      <dbl> <int>    <dbl>
##  1      -80  2000     13.6
##  2      -80  2001     14.0
##  3      -80  2002     13.8
##  4      -80  2003     13.9
##  5      -80  2004     13.7
##  6      -80  2005     13.7
##  7      -80  2006     13.8
##  8      -80  2007     14.1
##  9      -80  2008     13.7
## 10      -80  2009     13.9
## # i 62 more rows

## # A tibble: 4 x 9
##      fondaria Mitjana Mediana  SD  IQR  Min  Max Range  CV
##      <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1      -80    13.9    13.8 0.258 0.354 13.6  14.5  0.9  0.0186
##  2      -50    14.9    14.8 0.363 0.429 14.3  15.8  1.44 0.0243
##  3      -20    16.5    16.5 0.352 0.433 15.8  17.2  1.42 0.0213
##  4         0    17.2    17.1 0.374 0.575 16.5  17.9  1.35 0.0217
```

## 1.f) Annual thermal trends

To represent the data to be able to see the annual variations of the average temperature in the total depths and years studied, a **multi-line plot** was generated using **GGPLOT2** . By mapping years on the x-axis and average temperatures on the y-axis, and using different colors representing different depths, the graph shows how temperatures have changed over time and how stable the temperature distributions are. It effectively shows **synchronized inter-annual fluctuations across depths**, as well as a **general warming trend** observed during 2000-2017 time period.

Annual average temperature variation by depth  
(2000–2017) Mediterranean sea, at: 42 03' N, 3 15' E



```
## <theme> List of 1
## $ legend.position: chr "bottom"
## @ complete: logi FALSE
## @ validate: logi TRUE
```

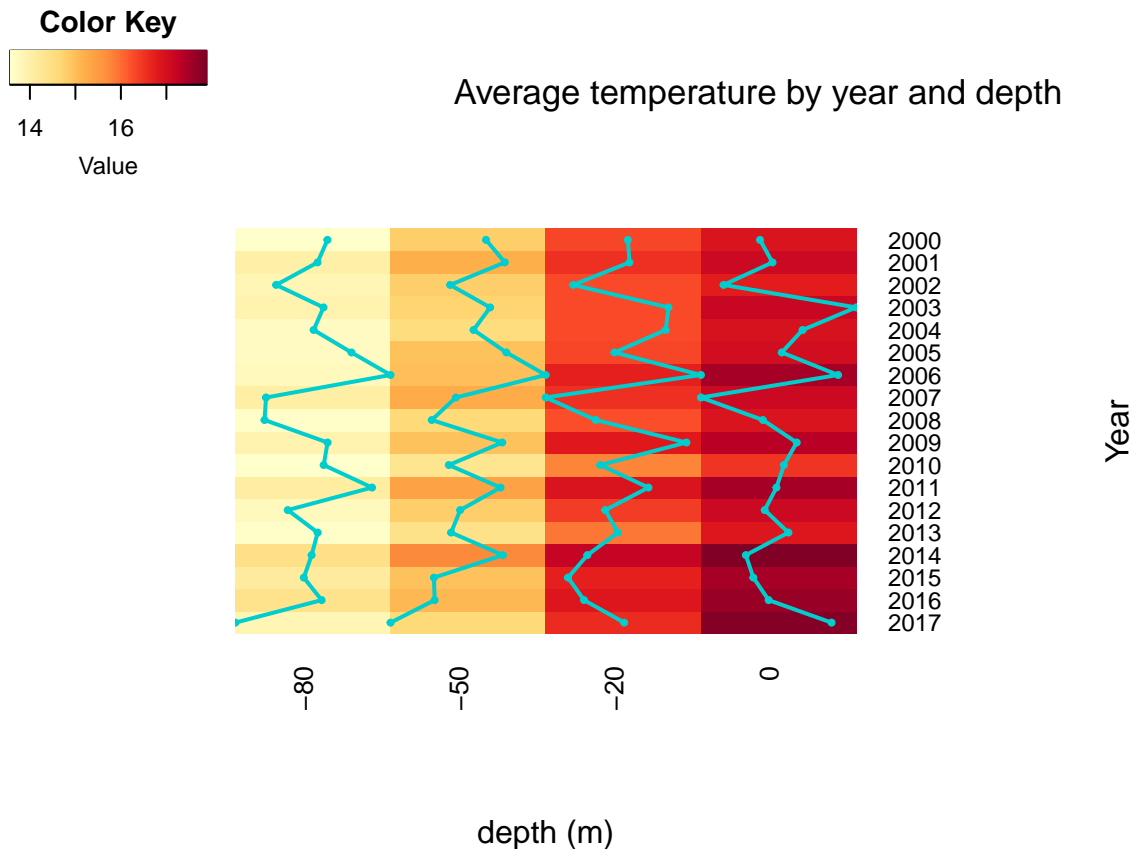
## 1.g) Data export to Excel

The **information** that has been processed is **sent to an outside .xlsx file** using the **OPENXLSX** package. This process makes sure that the data is stored and can be used for external reporting or further analysis. The Excel file has two sheets, including **stats\_per\_any** and **stats\_global**.

## Exercise 2

### 2.a) Heat map with “Gplots”

Following the instructions we generate a heat map using **GPLOTS** package, showing annual mean water temperature changes across time and depths.



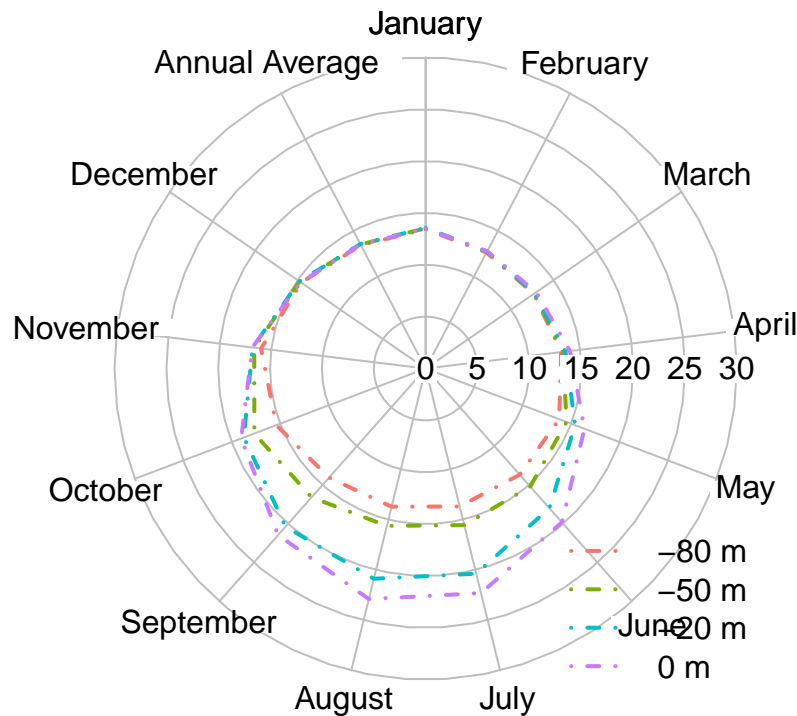
The heat map shows how the annual mean water temperature has changed at different depths (-80, -50, -20 and 0 meters) between 2000 and 2017. The color distribution shows that **surface waters** (0 and -20 meters) are consistently **the warmest** and show the **greatest inter-annual variability**. This variability is evident in the blue lines representing the standard deviation, which are imposed on the heat map. The **deeper layers** (-50 and -80 meters) display a **colder and more stable thermal profile**, much less variation, suggesting less impact from seasonal cycles.

The **years between 2014 and 2016** were a time of particularly **intense warming**. This was seen at the surface, and it is thought that it may have been caused by changes in the ocean or the climate.

## 2.b) Radial plot with “plotrix”

We used the `PLOTrix` package to make a radial plot. This shows the average monthly temperature changes across time and depths.

## Average Monthly Temperature by Depth (All Years)



The radial plot shows how average temperatures recorded at different depths (-80, -50, -20 and 0 meters) changed across each month of the year from 2000 to 2017. The lines are almost perfectly overlapped **between December and March**, showing that **temperature values remain almost identical regardless of depth**. However, there is a **change** that starts **in April**, and becomes more obvious during the summer months. While surface temperatures at 0 and -20 meters show a sharp increase towards higher values, the deeper measurements at -50 and -80 meters remain closer to the center of the graph. As the year progresses into **November and December**, these **temperature values contract and converge again**.

Furthermore, the plot shows that the **surface** has a **lot of** monthly temperature **variations**, while the **deeper layers** have a much **more consistent** and narrow temperature range throughout the year.

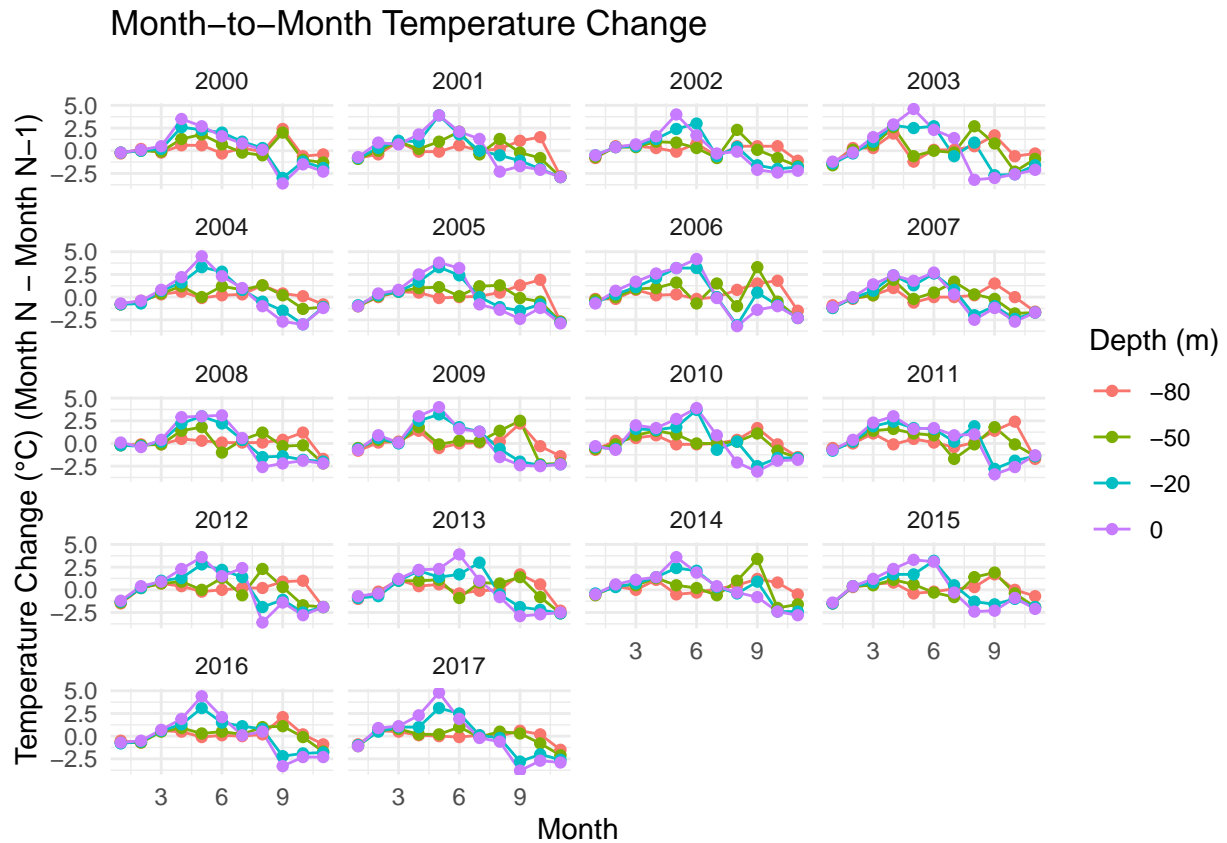
### Exercise 3

A function is scheduled to do the following:

- a. Calculate the temperature **difference between one month and the following month** for each year and depth, offering a graph for each specified year and the average of all years studied.
- b. Calculate the **difference in temperature between each month and the same month of the previous 30 years**, offering a graph for each year and the average for all years studied.

#### 3.a) Monthly rate of temperature change

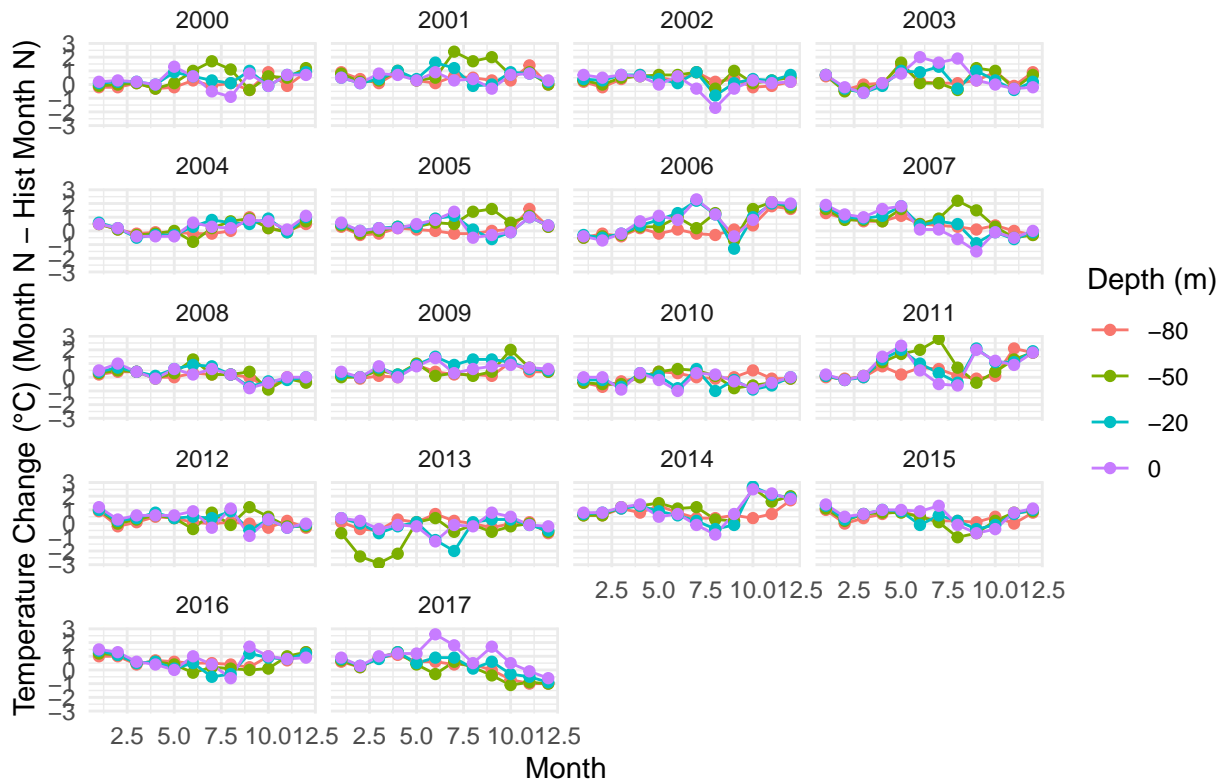
This analysis represents **inter-monthly temperature variation**. This is very useful to observe seasonal warming and cooling rates at different depths.



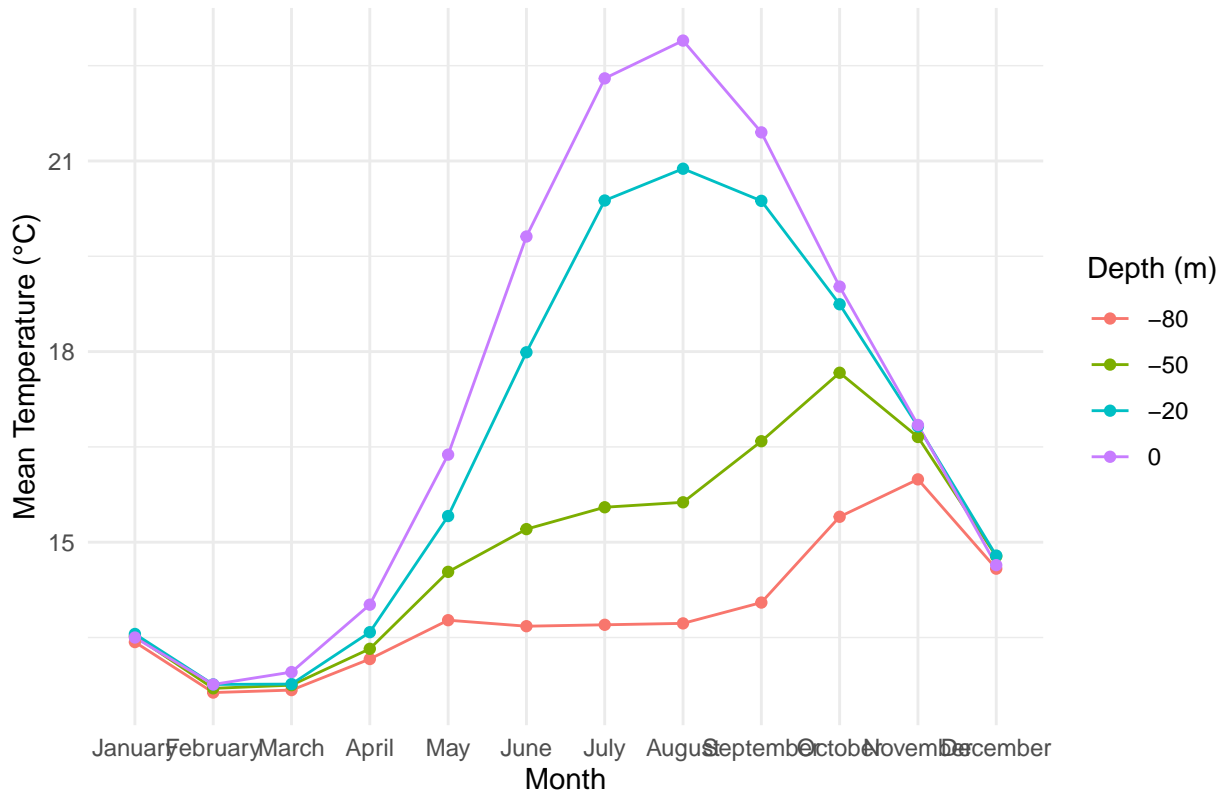
### 3.b) Historical comparison

This analysis looks at **decade temperature anomalies**. The resulting graphs show if a specific month was warmer or cooler than usual, providing insight into ongoing warming trends or unusual years.

## Current year to previous 30 years



## Average monthly changes

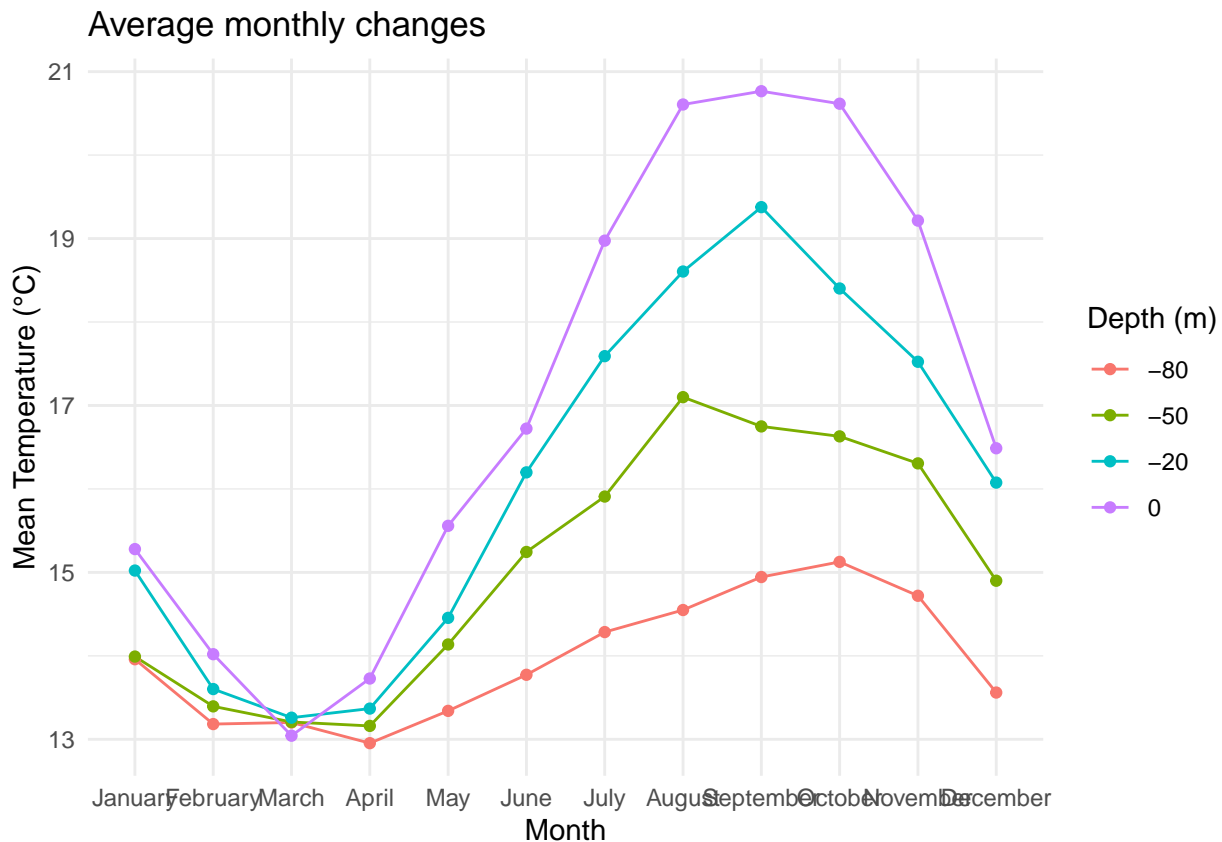




### 3.c) Function validation and testing

To make sure the function works properly and everything runs smoothly, we generate some simulated temperatures. For this, we use a noisy sine function with a period of 12 months.

```
plots = ex3(sim.deep, sim.pred)
print(plots$b2)
```



### 3.d) Discussion of results

When we look at how **temperatures change from month to month**, we can see a **clear difference between the surface and deeper waters during summer periods**. For instance in August, the surface is about 23°C, but the water at -80 meters is much cooler and more stable. This means there is a risk of different layers of water getting too hot or too cold. This happens because warm water is lighter than cold water creating a “cap” on the surface that acts like a barrier that stops oxygen from moving down and nutrients from moving up.

This process is called **stratification** and can eventually lead to a **collapse of the food chain** and create **dead zones** where there is too little oxygen for animals and plants to survive. Furthermore, if we compare the temperatures of the last few years to the average temperature between 1974 and 1999, we can see that the **temperature has been getting steadily higher**. The **water was much warmer than usual in 2003, 2011, and 2017**.

We used pretend data to show that our function and our calculations are correct. We created a mathematical “sine wave” to copy the natural 12-month cycle of the seasons. This showed that we can spot regular patterns even when there is random “noise” in the data. If we calculate the average of the total, we can make the results more stable and reliable. This helps us to see the bigger picture without getting distracted by small changes that happen every year.

In summary, the results show that **the sea is getting warmer**, which is a big threat to marine biodiversity.

The difference between the temperatures at the surface of the ocean and at depth is getting bigger, and we are seeing more heat spikes than before. These changes are not just part of a natural cycle but they are clear signs of climate change.

### 3.e) Additional data crossing and correlation analysis

We look for other data to better explain the climate change that is taking place. To be more specific, we are looking for information about the salt content of the water at a close location to 42°03' N and 3°15' E and at depths 0, -20, -50 and -80 meters. We get our information from the Copernicus Marine Service.

The **saltness of water** is as important as temperature when diagnosing climate change, but it tells us different things than temperature alone. **Salinity** is a **key indicator** of **changes in the water cycle**, which is important for understanding how water moves around the world. When it gets warmer, more water evaporates and the water becomes saltier. If the salt levels in the ocean suddenly dropped a lot, it could be a sign of extreme weather or melting ice on other parts of the planet.

**Changes in the amount of salt in the water change how thick or thin the water is**, and can make the layers of the water more stable. Alongside temperature, salinity drives the thermohaline circulation of ocean currents, so disruptions in these levels can increase stratification.

```
## [1] -0.122125
##
## Call:
## lm(formula = temperatura ~ salinitat, data = df_combined)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5255 -2.1806 -0.9757  1.6320  9.0387
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 100.7065    23.5505   4.276 2.11e-05 ***
## salinitat   -2.2342     0.6184  -3.613 0.000321 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.875 on 862 degrees of freedom
## Multiple R-squared:  0.01491,    Adjusted R-squared:  0.01377
## F-statistic: 13.05 on 1 and 862 DF,  p-value: 0.0003207
```

By combining these sets of data, we can **better tell the difference between the usual seasonal changes and long-term shifts caused by climate change**, which changes how water moves through the environment. The data shows that temperature and salinity usually go in opposite directions. Although the connection is only weak with a **correlation** of **-0.12**, the very low **p-value** of **0.00032** shows that this link is not just a coincidence.

The **multiple R-squared** of **0.0149** shows that salinity explains only about 1.5% of the variation in temperature. This **result is expected as the temperature of seawater is primarily driven by the sun** and changes with the seasons, rather than the salt content alone. But the data also shows that **when the sea is very salty, it is often very hot**.

In conclusion, while salinity is not the main cause of temperature changes, it is an important tool for understanding the biochemical changes triggered by global warming. Our model clearly shows a connection between these two variables. When we look at how climate change is affecting the sea, measuring salt levels and temperature helps us to see **how the layers of the sea are becoming more separated**. This can help us to see more accurately, the increasing stratification of the water column, **predicting the emergence of hypoxic zones** that threaten marine biodiversity.

### 3.f) Project reproductibility and additional documentation

To make sure the analysis is transparent and can be recreated by anyone, a **public repository** named “**SeaTemperature**” has been set up on **GitHub** <<https://github.com/eaTaki/SeaTemperature>>. This repository is a central open-source storage space for the following components:

- The assignment instructions.
- The presented Rmarkdown document.
- The data used to complete the assignment `undersea_data.RData` . As well as salinity data used to complete exercise 3.e.
- An R library that stores the functions used for the completion of this project, thus being: `CreateDF.R` and `Ex3.R`.