Project 2 and Project 3

Austine Lotanna Iheji

2024-04-02

Research Questions

Changed Variable: Ocean PH in Hawaii with Ocean heat content (top 700) in the world

Association questions:

1. Is the sea level rise associated with Ocean heat content (top 2000m)?

Explanatory Variable: Ocean heat content(Quantitative)

Response Variable: Sea level Rise(Quantitative)

2. Is the Ocean heat content (top 700m) associated with CO2 concentration?

Explanatory Variable: CO2 Concentration (Quantitative)

Response Variable: Ocean heat content (top 700)

3. Is temperature anomaly associated with sea surface temperatures?

Explanatory Variable: Temperature anomaly(Quantitative)

Response Variable: Sea Surface Temperature(Quantitative)

1. Load data set(s) and libraries

```
setwd("C:\\Users\\user\\OneDrive - Ashesi University\\Statistics\\Statistics
Homework")
load("climate_change")
library(stats)
library(descr)
```

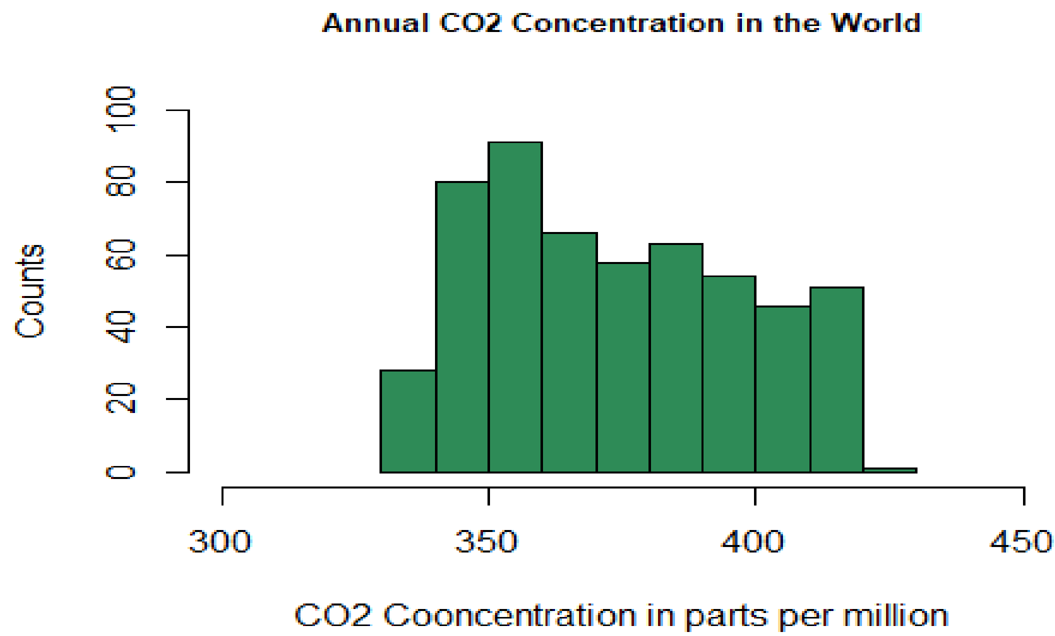
2. Create variable subset

```
#creating the data frame with my needed variables
variables <- c("Entity", "Date","Average", "Sea.surface.temp..lower.bound.",
"Sea.surface.temp..upper.bound.", "Annual.averaged.2", "Temperature.anomaly",
"monthly_sea_surface_temperature_anomaly")

my_data <- climate_change[variables]
save(my_data, file = "my_data.Rdata")
```

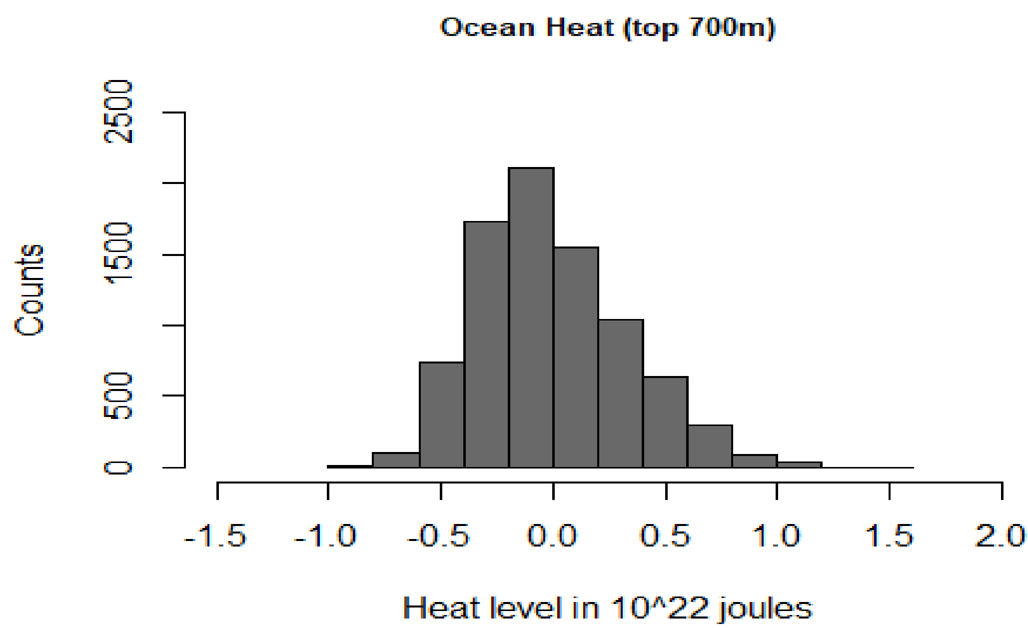
3. Data management I: check for and recode errors and NAs, if the need be.

```
#creating a histogram for each of my quantitaive variables
hist(my_data$Annual.averaged.2, col = "seagreen",ylim = c(0, 100), xlim =
```

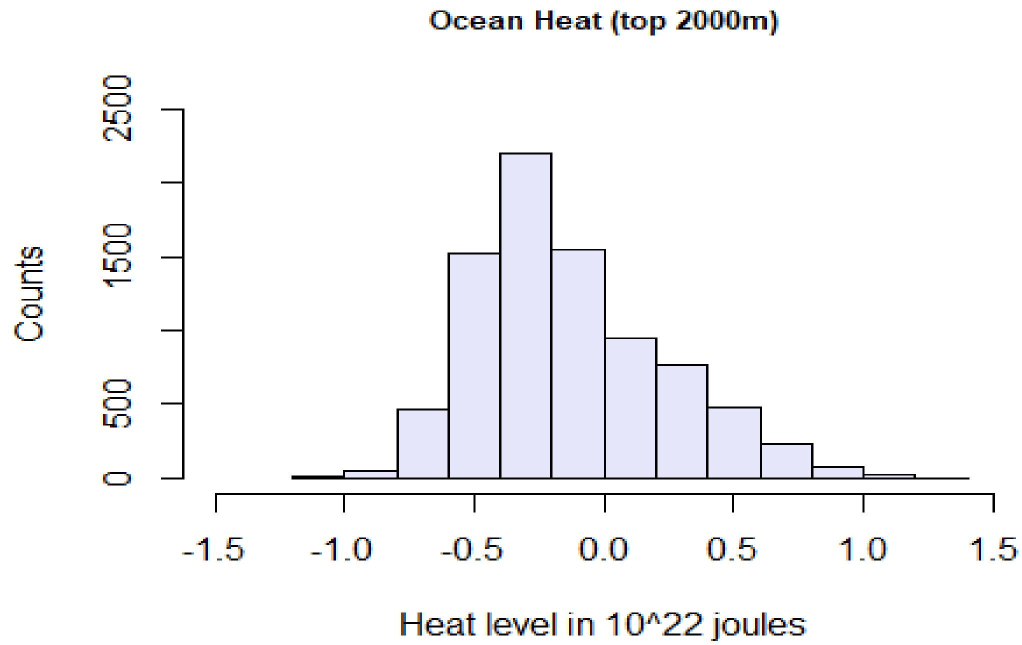
```
c(300, 450), main="Annual CO2 Concentration in the World", xlab = "CO2 Cooncentration in parts per million", ylab = "Counts", cex.main = 0.8)
```



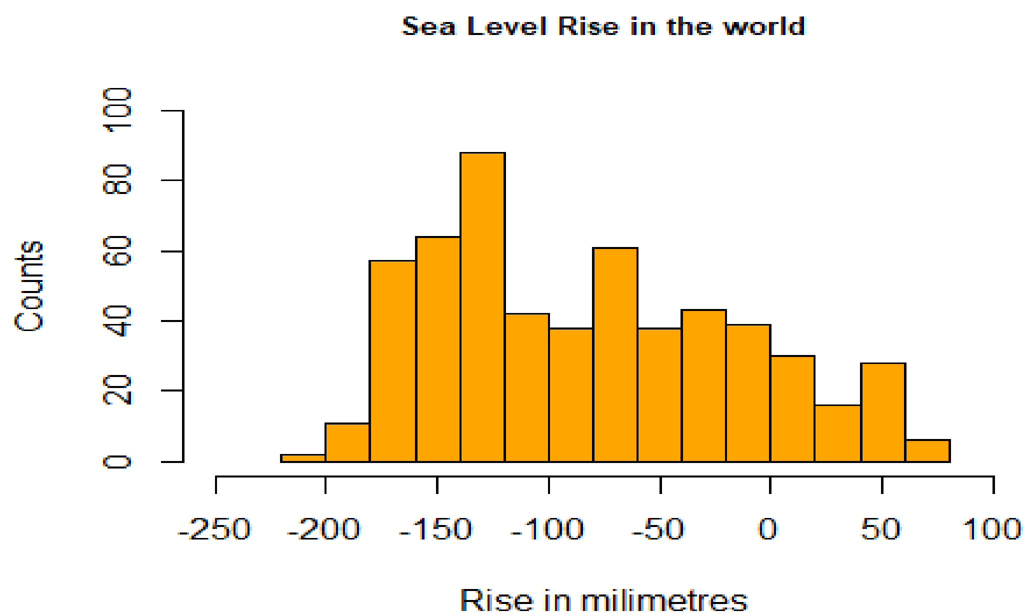

```
hist(my_data$Sea.surface.temp..upper.bound., col = "#696969",ylim = c(0, 2500), xlim = c(-1.5, 2) ,main="Ocean Heat (top 700m)", xlab = "Heat level in 10^22 joules", ylab = "Counts", cex.main = 0.8)
```



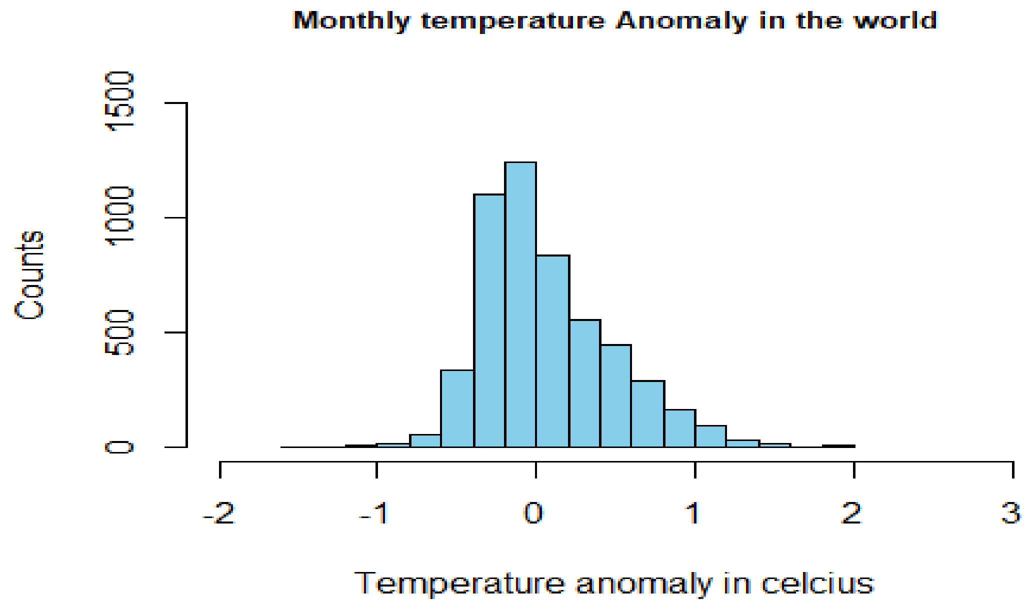
```r
hist(my_data$Sea.surface.temp..lower.bound., col = "lavender",ylim = c(0,
2500), xlim = c(-1.5, 1.5), main="Ocean Heat (top 2000m)", xlab = "Heat level
in 10^22 joules", ylab = "Counts", cex.main = 0.8)
```



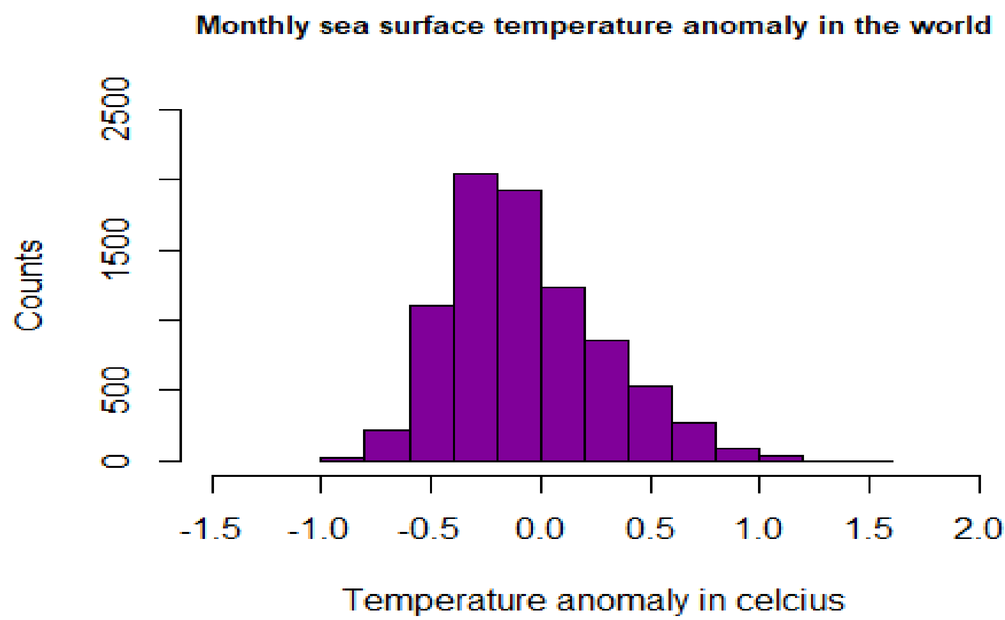

```r
hist(my_data$Average, col = "orange",ylim = c(0, 100), xlim = c(-250, 100),
main="Sea Level Rise in the world", xlab = "Rise in milimetres", ylab =
"Counts", cex.main = 0.8)
```



```
hist(my_data$Temperature.anomaly, col = "skyblue",ylim = c(0, 1500), xlim =
c(-2, 3), main="Monthly temperature Anomaly in the world", xlab =
"Temperature anomaly in celcius", ylab = "Counts", cex.main = 0.8)
```




```
hist(my_data$monthly_sea_surface_temperature_anomaly, col =  "#800099",ylim =
c(0, 2500), xlim = c(-1.5, 2), main="Monthly sea surface temperature anomaly
in the world", xlab = "Temperature anomaly in celsius", ylab = "Counts",
cex.main = 0.8)
```



4. Data management II: further subset and create a secondary variable, if need be.

```
#Creating a subset for the world
World_subset <- subset(my_data, Entity == "World")
```

5. Descriptive statistics (sample means, standard deviations, proportions) and univariate displays

```
#Mean, SD for Sea level Rise
summary(World_subset$Average)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -204.38 -138.41  -87.45  -82.17  -28.78   68.81    2061

mean(World_subset$Average,na.rm = TRUE)

## [1] -82.1743

sd(World_subset$Average,na.rm = TRUE)

## [1] 68.08795

length(na.omit(World_subset$Average))

## [1] 563

#Mean, SD for Ocean Heat Level (top 2000m)
summary(World_subset$Sea.surface.temp..lower.bound.)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -0.8185 -0.3730 -0.2160 -0.1320  0.0534  1.0755     536

mean(World_subset$Sea.surface.temp..lower.bound.,na.rm = TRUE)

## [1] -0.131982

sd(World_subset$Sea.surface.temp..lower.bound.,na.rm = TRUE)

## [1] 0.3465258

length(na.omit(World_subset$Sea.surface.temp..lower.bound.))

## [1] 2088

#Mean, SD for Ocean Heat Level(top 700m)
summary$Sea.surface.temp..upper.bound.)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -0.7035 -0.2460 -0.0712 -0.0117  0.1732  1.1195     536

mean(World_subset$Sea.surface.temp..upper.bound.,na.rm = TRUE)

## [1] -0.01168367
```

```
sd(World_subset$Sea.surface.temp..upper.bound.,na.rm = TRUE)

## [1] 0.3198274

length(na.omit(World_subset$Sea.surface.temp..upper.bound.))

## [1] 2088

#Mean, SD for CO2 concentration in the world
summary(World_subset$Annual.averaged.2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   335.9   353.7   370.3   373.6   392.7   420.2    2086

mean(World_subset$Annual.averaged.2,na.rm = TRUE)

## [1] 373.6143

sd(World_subset$Annual.averaged.2,na.rm = TRUE)

## [1] 24.09929

length(na.omit(World_subset$Annual.averaged.2))

## [1] 538

#Mean, SD for Atmospheric Temperature Anomaly
summary(World_subset$Temperature.anomaly)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
## -0.8100 -0.2200 -0.0300  0.0677  0.2900  1.4800     896

mean(World_subset$Temperature.anomaly,na.rm = TRUE)

## [1] 0.06768519

sd(World_subset$Temperature.anomaly,na.rm = TRUE)

## [1] 0.3926423

length(na.omit(World_subset$Temperature.anomaly))

## [1] 1728

#Mean, SD for Sea surface temperature anomaly
summary(World_subset$monthly_sea_surface_temperature_anomaly)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
## -0.7539 -0.3051 -0.1399 -0.0715  0.1151  1.0977     536

mean(World_subset$monthly_sea_surface_temperature_anomaly,na.rm = TRUE)

## [1] -0.07145969

sd(World_subset$monthly_sea_surface_temperature_anomaly,na.rm = TRUE)
```

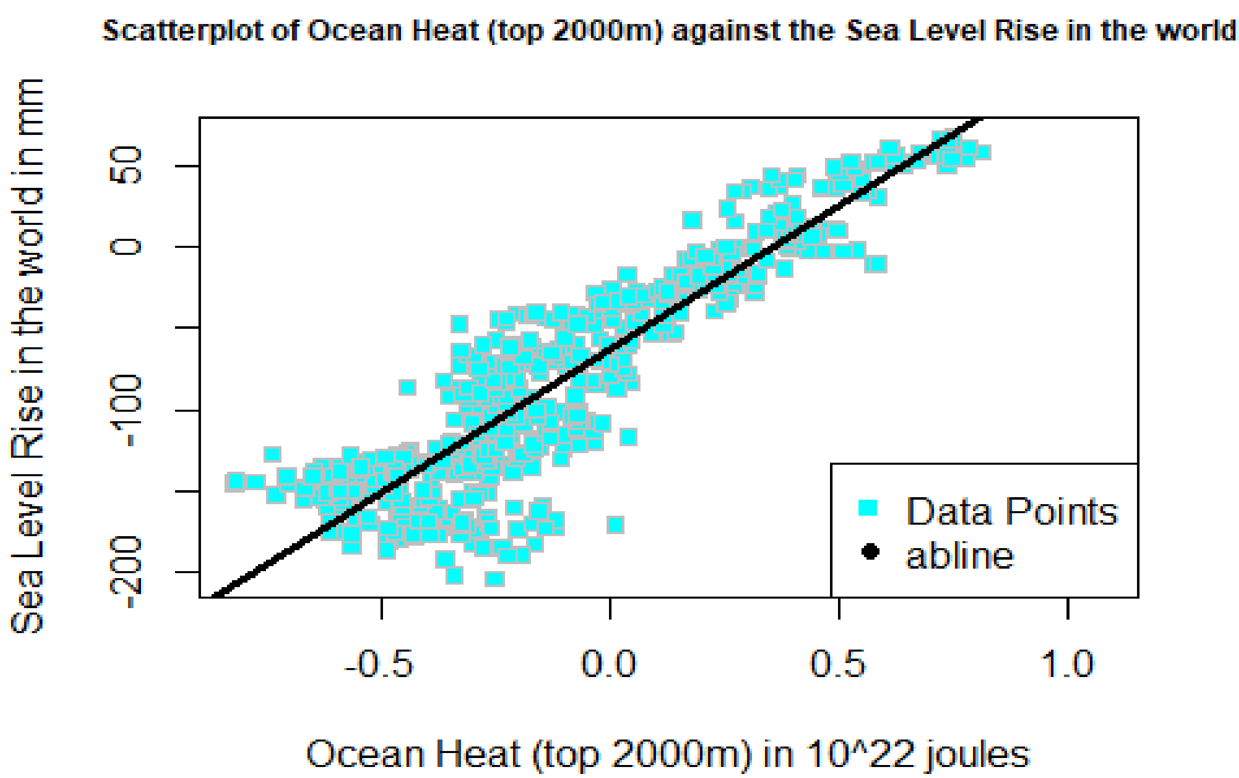
```
## [1] 0.331959
```

```r
length(na.omit(World_subset$monthly_sea_surface_temperature_anomaly))
```

```
## [1] 2088
```

6. Bivariate tables and graphs

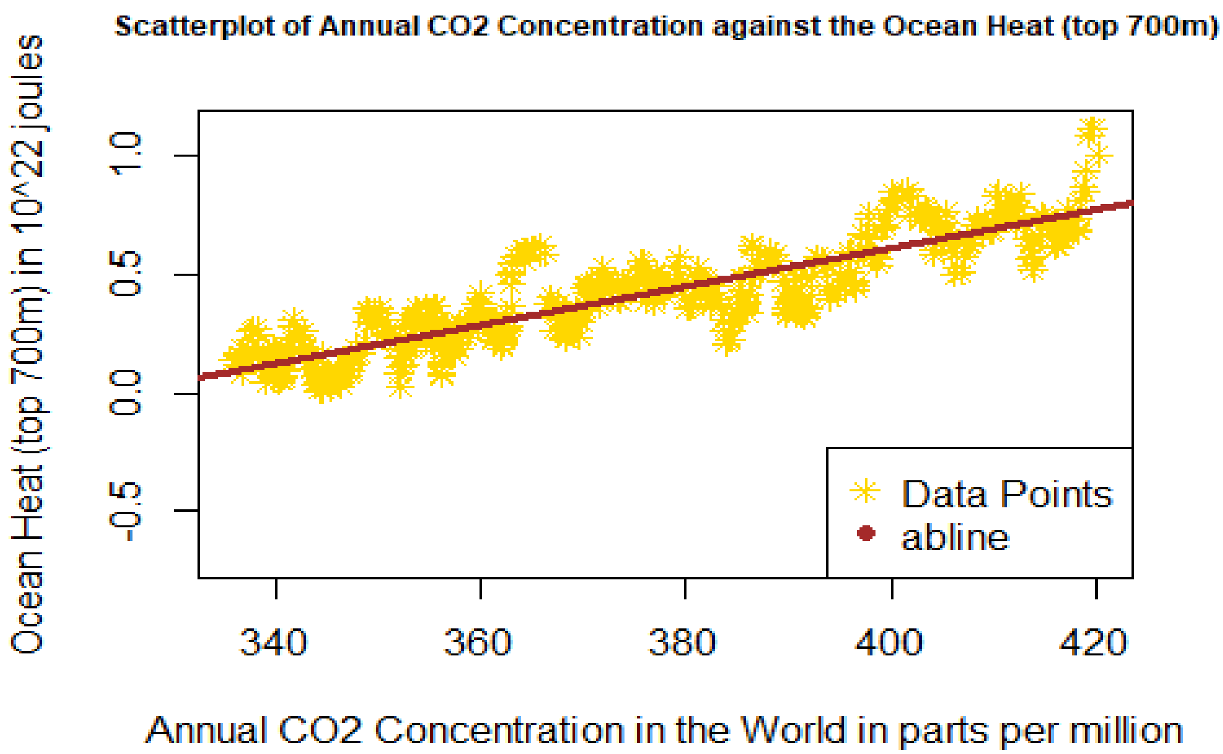
```r
#Explanatory Variable: Ocean heat content(Quantitative)
#Response Variable: Sea level Rise(Quantitative)

plot(Average ~ Sea.surface.temp..lower.bound., data = World_subset, xlab =
"Ocean Heat (top 2000m) in 10^22 joules", ylab = "Sea Level Rise in the world
in mm", col = "grey", bg = "cyan", pch = 22, main = "Scatterplot of Ocean
Heat (top 2000m) against the Sea Level Rise in the world", cex.main = 0.75)
legend("bottomright", legend = c("Data Points", "abline"), col = c("cyan",
"black"), pch = c(15 ,19))
abline(lm(World_subset$Average ~
World_subset$Sea.surface.temp..lower.bound.), col = "black", lwd= 3)
```




```r
cor(World_subset$Average , World_subset$Sea.surface.temp..lower.bound., use =
"complete.obs")
```

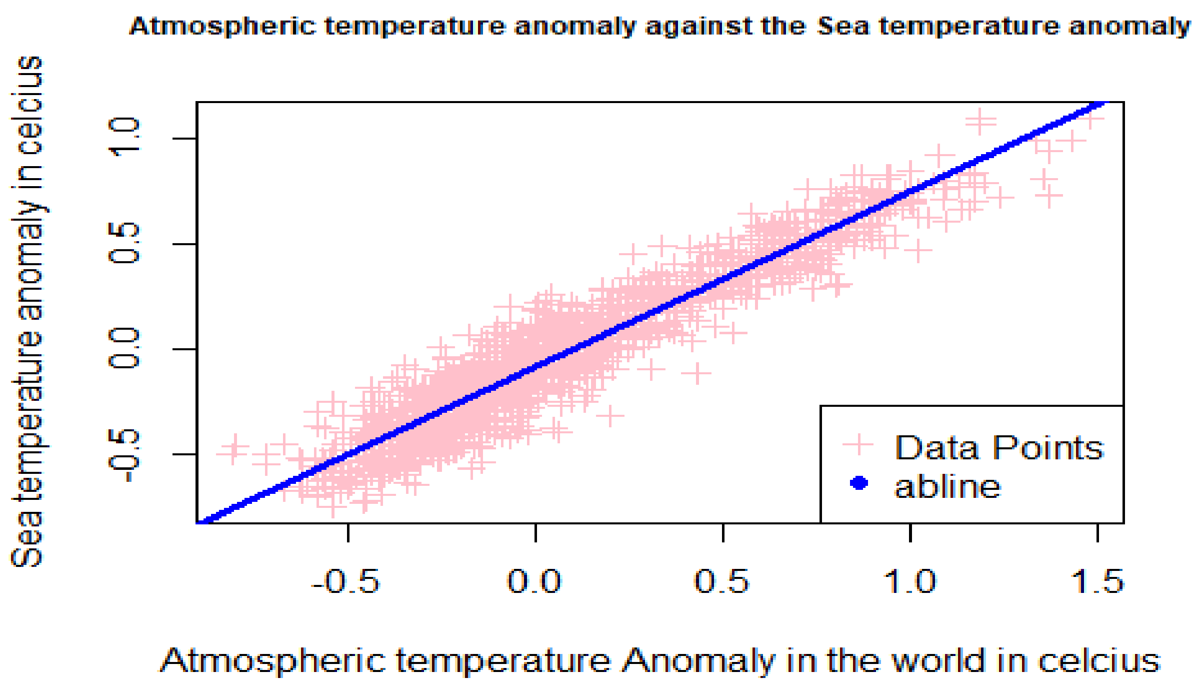
```
## [1] 0.9040855
```

```
#Explanatory Variable: CO2 Concentration (Quantitative)
#Response Variable: Response Variable: Ocean heat content (top 700)

plot(Sea.surface.temp..upper.bound. ~ Annual.averaged.2  , data =
World_subset, xlab = "Annual CO2 Concentration in the World in parts per
million", ylab = "Ocean Heat (top 700m) in 10^22 joules", col = "gold", pch =
8, main = "Scatterplot of Annual CO2 Concentration against the Ocean Heat
(top 700m)", cex.main = 0.75)
legend("bottomright", legend = c("Data Points", "abline"), col = c("gold",
"brown"), pch = c(8,19))
abline(lm(World_subset$Sea.surface.temp..upper.bound. ~
World_subset$Annual.averaged.2), col = "brown", lwd= 3)
```




```
cor(World_subset$Sea.surface.temp..upper.bound.,
World_subset$Annual.averaged.2, use = "complete.obs")

## [1] 0.8854638

#Explanatory Variable: Temperature anomaly(Quantitative)
#Response Variable: Sea Surface Temperature(Quantitative)

plot(monthly_sea_surface_temperature_anomaly ~ Temperature.anomaly, data =
World_subset, xlab = "Atmospheric temperature Anomaly in the world in
celcius", ylab = "Sea temperature anomaly in celcius", col = "pink", bg =
"red", pch = 3, main = "Atmospheric temperature anomaly against the Sea
```

```
temperature anomaly", cex.main = 0.75)
legend("bottomright", legend = c("Data Points", "abline"), col = c("pink", "blue"), pch = c(3,19))

abline(lm(World_subset$monthly_sea_surface_temperature_anomaly ~ World_subset$Temperature.anomaly), col = "blue", lwd= 3)
```




```
cor(World_subset$monthly_sea_surface_temperature_anomaly, World_subset$Temperature.anomaly, use = "complete.obs")

## [1] 0.9467204
```

Summary of Findings

a. Sea Level Rise and Ocean Heat Content (Top 2000 meters)

The analysis revealed a robust correlation between sea level rise and ocean heat content (top 2000 meters), with a correlation coefficient of approximately 0.904. This strong positive relationship suggests that as ocean heat content in the top 700 meters increases, so does sea level rise. Also with a mean of about -82.17 millimeters, it can be said that the sea level was falling in most years, while the ocean heat content remains rather low at about -0.135*10^22 joules.

b. Ocean Heat Content (Top 700 meters) and CO2 Concentration

Investigating the relationship between ocean heat content (top 700 meters) and CO2 concentration revealed a similarly strong positive correlation, with a coefficient of approximately 0.885. This finding indicates that higher CO2 concentrations may lead to increased ocean heat content. Also, the findings reveal the fact that CO2 concentrations are somewhat high at an average of 373.61 parts per million.

c. Temperature Anomaly and Sea Surface Temperatures

The analysis explored the association between temperature anomaly and sea surface temperatures, revealing a very strong positive relationship with a correlation coefficient of approximately 0.947. This suggests that an increase in atmospheric temperature anomaly may contribute to higher sea surface temperatures.

7. Bivariate analysis (hypothesis tests and posthoc tests)

Association questions:

1. **Is the sea level rise associated with Ocean heat content (top 2000m)?**

Step 1: State the claim

Null Hypothesis: Ho: There is no linear relationship between sea level rise and Ocean heat content (top 2000m).

Alternate Hypothesis: Ha: There is a linear relationship between sea level rise and Ocean heat content (top 2000m).

Step 2: Collect and summarize the sample

Mean of sea level rise: -82.17

SD of sea level rise: 68.09

Mean of Ocean heat content (top 2000m): -0.132

SD of Ocean heat content (top 2000m): 0.346

Sample size = 563 (The number of valid pairs in the data set)

The observations are independent.

The histograms for the variables are normally distributed, and even though the sea level rise is a bit skewed, its skew is not very potent. Also, the sample size(563) is more than 30, which indicates that the criteria for using a normal distribution is met.

Step 3: Assess the evidence

```
cor_results1<-cor.test(World_subset$Average,
World_subset$Sea.surface.temp..lower.bound.)
cor_results1

##
##  Pearson's product-moment correlation
##
## data:  World_subset$Average and
```

```
World_subset$Sea.surface.temp..lower.bound.
## t = 50.108, df = 561, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8877754 0.9181279
## sample estimates:
##       cor
## 0.9040855
```

```
cor_results1$estimate^2
```

```
##       cor
## 0.8173705
```

```
my_lm1 = lm(World_subset$Average ~
World_subset$Sea.surface.temp..lower.bound.); my_lm1
```

```
##
## Call:
## lm(formula = World_subset$Average ~
World_subset$Sea.surface.temp..lower.bound.)
##
## Coefficients:
##                                 (Intercept)
##                                      -62.95
## World_subset$Sea.surface.temp..lower.bound.
##                                      177.80
```

```
summary(my_lm1)
```

```
##
## Call:
## lm(formula = World_subset$Average ~
World_subset$Sea.surface.temp..lower.bound.)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -110.264  -16.125    3.194   19.749   74.180
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                  -62.951      1.286  -48.95
## World_subset$Sea.surface.temp..lower.bound.  177.803      3.548   50.11
##                                             Pr(>|t|)
## (Intercept)                                   <2e-16 ***
## World_subset$Sea.surface.temp..lower.bound.   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.12 on 561 degrees of freedom
##   (2061 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.8174, Adjusted R-squared:  0.817 
## F-statistic:  2511 on 1 and 561 DF,  p-value: < 2.2e-16
```

Step 4: Make a conclusion

With a p-value of 2.2e-16, when compared with a significance level of 0.05, the p-value indicates that the results are significant, and we would reject the null Hypothesis.

Conclusion: There is a linear relationship between sea level rise and Ocean heat content (top 2000m)

In this scenario, there's a possibility of committing a type 1 error, wherein we incorrectly reject the null hypothesis despite it being true. To explain further, we might conclude that there's a linear relationship between sea level rise and Ocean heat content when, in reality, no such association exists.

The slope of the regression line: 177.803

2. Is the Ocean heat content (top 700m) associated with CO2 concentration?

Step 1: State the claim

Null Hypothesis: Ho: There is no linear relationship between Ocean heat content (top 700m) and CO2 concentration

Alternate Hypothesis: Ha: There is a linear relationship between Ocean heat content (top 700m) and CO2 concentration

Step 2: Collect and summarize the sample

Mean of Ocean heat content (top 700m): -0.0117

SD of Ocean heat content (top 700m): 0.320

Mean of CO2 concentration: 373.61

SD of CO2 concentration: 24.1

In this case, the observations of CO2 concentrations and Ocean heat are independent.

Sample size = 538 (The number of valid pairs in the data set)

The histograms for the variables are normally distributed, and the sample size(538) is more than 30, which indicates that the criteria for using a normal distribution are met.

Step 3: Assess the evidence

```
cor(World_subset$Sea.surface.temp..upper.bound.,
World_subset$Annual.averaged.2, use = "complete.obs")

## [1] 0.8854638
```

```
cor_results2<- cor.test(World_subset$Sea.surface.temp..upper.bound.,
World_subset$Annual.averaged.2)
cor_results2

##
##  Pearson's product-moment correlation
##
## data:  World_subset$Sea.surface.temp..upper.bound. and
World_subset$Annual.averaged.2
## t = 44.114, df = 536, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8657312 0.9024480
## sample estimates:
##       cor
## 0.8854638

cor_results2$estimate^2

##       cor
## 0.7840461

my_lm2 = lm(World_subset$Sea.surface.temp..upper.bound. ~
World_subset$Annual.averaged.2); my_lm2

##
## Call:
## lm(formula = World_subset$Sea.surface.temp..upper.bound. ~
World_subset$Annual.averaged.2)
##
## Coefficients:
##                    (Intercept)  World_subset$Annual.averaged.2
##                      -2.617499                        0.008072

summary(my_lm2)

##
## Call:
## lm(formula = World_subset$Sea.surface.temp..upper.bound. ~
World_subset$Annual.averaged.2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27140 -0.06999 -0.00162  0.05899  0.35094
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -2.617500   0.068508  -38.21   <2e-16 ***
## World_subset$Annual.averaged.2  0.008072   0.000183   44.11   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1022 on 536 degrees of freedom
##   (2086 observations deleted due to missingness)
## Multiple R-squared:  0.784,  Adjusted R-squared:  0.7836
## F-statistic:  1946 on 1 and 536 DF,  p-value: < 2.2e-16
```

Step 4: Make a conclusion

Similar to the previous test, we have a p-value of 2.2e-16; when compared with a significance level of 0.05, the p-value indicates that the results are significant, and we would reject the null Hypothesis.

Conclusion: There is a linear relationship between Ocean heat content (top 700m) and CO2 concentration

Also, in this scenario, there's a possibility of committing a type 1 error, wherein we incorrectly reject the null hypothesis despite it being true.

The slope of the regression line: 0.008072

3. Is temperature anomaly associated with sea surface temperatures?

Step 1: State the claim

Null Hypothesis: Ho: There is no linear relationship between temperature anomaly and sea surface temperatures

Alternate Hypothesis: Ha: There is a linear relationship between temperature anomaly and sea surface temperatures

Step 2: Collect and summarize the sample

Mean of temperature anomaly: 0.068

SD of temperature anomaly:

Mean of sea surface temperatures: -0.0715

SD of sea surface temperatures: 0.332

Sample size = 1728 (The number of valid pairs in the data set)

The observations of temperature anomaly and sea surface temperatures are independent.

The histograms for the variables are normally distributed, and the sample size(1728) is more than 30, which indicates that the criteria for using a normal distribution are met.

Step 3: Assess the evidence

```
cor(World_subset$monthly_sea_surface_temperature_anomaly,
World_subset$Temperature.anomaly, use = "complete.obs")

## [1] 0.9467204
```

```r
cor_results3<- cor.test(World_subset$monthly_sea_surface_temperature_anomaly,
World_subset$Temperature.anomaly)
cor_results3
```

```
##
##  Pearson's product-moment correlation
##
## data:  World_subset$monthly_sea_surface_temperature_anomaly and
World_subset$Temperature.anomaly
## t = 122.13, df = 1726, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9416008 0.9514023
## sample estimates:
##       cor
## 0.9467204
```

```r
cor_results3$estimate^2
```

```
##       cor
## 0.8962794
```

```r
my_lm3 = lm(World_subset$monthly_sea_surface_temperature_anomaly ~
World_subset$Temperature.anomaly); my_lm3
```

```
##
## Call:
## lm(formula = World_subset$monthly_sea_surface_temperature_anomaly ~
##     World_subset$Temperature.anomaly)
##
## Coefficients:
##                      (Intercept)  World_subset$Temperature.anomaly
##                         -0.08827                           0.83366
```

```r
summary(my_lm3)
```

```
##
## Call:
## lm(formula = World_subset$monthly_sea_surface_temperature_anomaly ~
##     World_subset$Temperature.anomaly)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39745 -0.07473  0.00428  0.07723  0.35074
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -0.088274   0.002719  -32.47   <2e-16 ***
## World_subset$Temperature.anomaly  0.833662   0.006826  122.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1114 on 1726 degrees of freedom
##   (896 observations deleted due to missingness)
## Multiple R-squared:  0.8963, Adjusted R-squared:  0.8962
## F-statistic: 1.491e+04 on 1 and 1726 DF,  p-value: < 2.2e-16
```

Step 4: Make a conclusion

Lastly, we also have a p-value of 2.2e-16, and when compared with a significance level of 0.05, the p-value indicates that the results are significant, and we would reject the null Hypothesis.

Conclusion: There is a linear relationship between temperature anomaly and sea surface temperatures

Similarly, in this scenario, there's a possibility of committing a type 1 error, wherein we incorrectly reject the null hypothesis despite it being true. We might conclude that there's a linear relationship between temperature anomaly and sea surface temperatures where no such association exists.

The slope of the regression line: 0.83366

8. Moderation

9. Save

Summary:

For all associations that were studied here, namely:

a. Sea level rise and Ocean heat content (top 2000m):

There is a linear relationship between each variable, as the null hypothesis was rejected. According to the slope of the regression line (177.803), we can say that for every one-unit increase (10^22 joules) in Ocean heat content, sea levels are expected to rise by approximately 177.803mm, suggesting a positive relationship between these quantities. This positive relation is also strengthened by the correlation coefficient of approximately 0.904.

In a broader context, this statistical conclusion reinforces the notion that as sea temperatures increase, the consequential thermal expansion of seawater contributes significantly to the increase of global sea levels. Therefore, these findings not only elucidate the relationship of oceanic processes but also underscore the urgency of addressing climate change to mitigate the adverse impacts of rising sea levels on coastal communities and ecosystems worldwide.

b. Ocean heat content (top 700m) and CO2 concentration:

In this relationship, we rejected the null hypothesis, leading us to conclude that there is indeed a positive, strong linear relationship between the two variables. Also, according to the slope of the regression line, we can quantify this relationship and say that for every one-unit increase in CO_2 concentration, there is an increase in Ocean heat content (top 700m) by 0.008072 (10^{22} joules). This positive association is also seen in their correlation coefficient of approximately 0.885.

In the broader context, this statistical result confirms the coherence between rising CO_2 levels and the increased accumulation of heat in the ocean, highlighting the significant role of anthropogenic activities in driving oceanic warming. In the context of climate change, anthropogenic activities typically refer to the burning of fossil fuels, deforestation, industrial processes, and agriculture, which release greenhouse gases such as carbon dioxide (CO_2) into the atmosphere, contributing to global warming and increasing the heat content.

c. Temperature anomaly and sea surface temperatures:

Similarly, the null hypothesis was rejected in the association, leaving us with the conclusion that a linear relationship exists between these quantities. In other words, the slope of the regression line tells us that for every 1 Celsius increase in atmospheric temperature anomaly, there is an increase in sea surface anomaly by 0.83366 Celsius. This conclusion underscores the interconnectedness between atmospheric and oceanic conditions, highlighting their influence on each other's temperature variations.