

Concentration de CO2 dans l'atmosphère depuis 1958

Pierre LACOSTE

13/12/2021

1) Importation et verification des données

Les données ont été importées le 13/12/2021.

```
path = "./weekly_in_situ_co2_mlo.csv"
data_url = "https://scrippsco2.ucsd.edu/assets/data/atmospheric/stations/in_situ_co2/weekly/weekly_in_s"
```

On vérifie si les données ont été téléchargées, si ce n'est pas le cas on les télécharge

```
if (file.exists(path)==F) {
  download.file(data_url,path)
}
```

On importe les données, en regardant le jeu de données, on se rend compte que les 44 premières lignes expliquent le tableau, il faut donc les enlever lors de l'importation.

```
data = read.csv(path, skip = 44, header = F, dec = ".", sep = ",")
```

Le tableau contient 2 colonnes:

- La date
- La concentration en CO₂ en micromol/mol

Les colonnes n'ont pas de titres donc on en ajoute un pour plus de clarté dans les analyses.

```
names(data)=c("Date", "C_ppm")
names(data)
```

```
## [1] "Date" "C_ppm"
```

On vérifie que le tableau va bien de 1958 à 2021.

```
head(data)
```

```
##           Date  C_ppm
## 1 1958-03-29 316.19
## 2 1958-04-05 317.31
## 3 1958-04-12 317.69
## 4 1958-04-19 317.58
## 5 1958-04-26 316.48
## 6 1958-05-03 316.95
```

```
tail(data)
```

```
##           Date  C_ppm
## 3237 2021-08-21 414.42
## 3238 2021-08-28 412.68
## 3239 2021-09-04 412.58
## 3240 2021-09-11 413.15
## 3241 2021-09-18 413.09
## 3242 2021-09-25 413.05
```

Les mesures vont bien de Mars 1958 à septembre 2021.

Maintenant il faut vérifier si des données sont manquantes.

```
na_records = apply(data, 1, function (x) any(is.na(x)))
data[na_records,]
```

```
## [1] Date  C_ppm
## <0 lignes> (ou 'row.names' de longueur nulle)
```

Il n'y en a pas, bonne nouvelle pour notre analyse.

Il faut maintenant vérifier la classe de chaque colonne.

```
str(data)
```

```
## 'data.frame': 3242 obs. of 2 variables:
## $ Date : chr "1958-03-29" "1958-04-05" "1958-04-12" "1958-04-19" ...
## $ C_ppm: num 316 317 318 318 316 ...
```

La colonne `date` est un vecteur chaîne de caractère et la colonne concentration (`C_ppm`) est un vecteur de nombre.

Pour pouvoir travailler sur les données il faut transformer les chaînes de caractères en date.

```
library(parsedate)
```

```
## Warning: le package 'parsedate' a été compilé avec la version R 4.1.2
```

```
convert_week = function(iso) {
  as.character(parse_iso_8601(iso))
}

data$Date = as.Date(convert_week(data$Date))
```

```
class(data$Date)
```

```
## [1] "Date"
```

Les points sont dans l'ordre chronologique donc pas besoin de les trier.

On vérifie que les semaines font bien 7 jours.

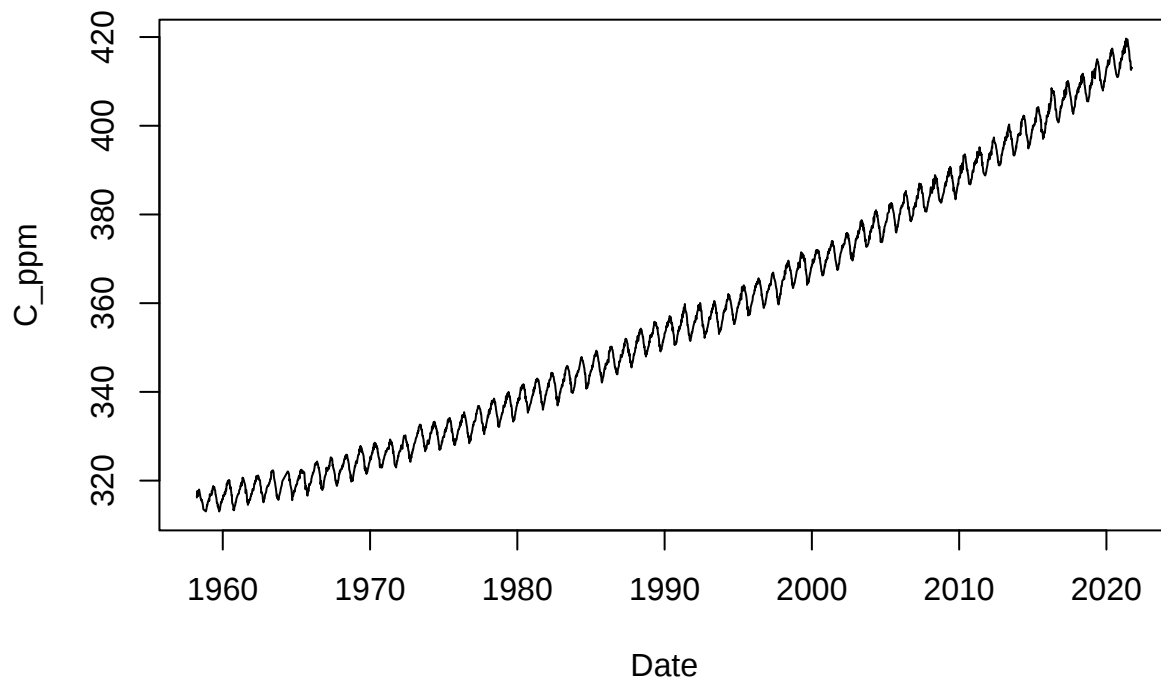
```
all(diff(data$date) == 7)
```

```
## [1] TRUE
```

Représentation graphique:

On fait une première représentation graphique pour visualiser les données.

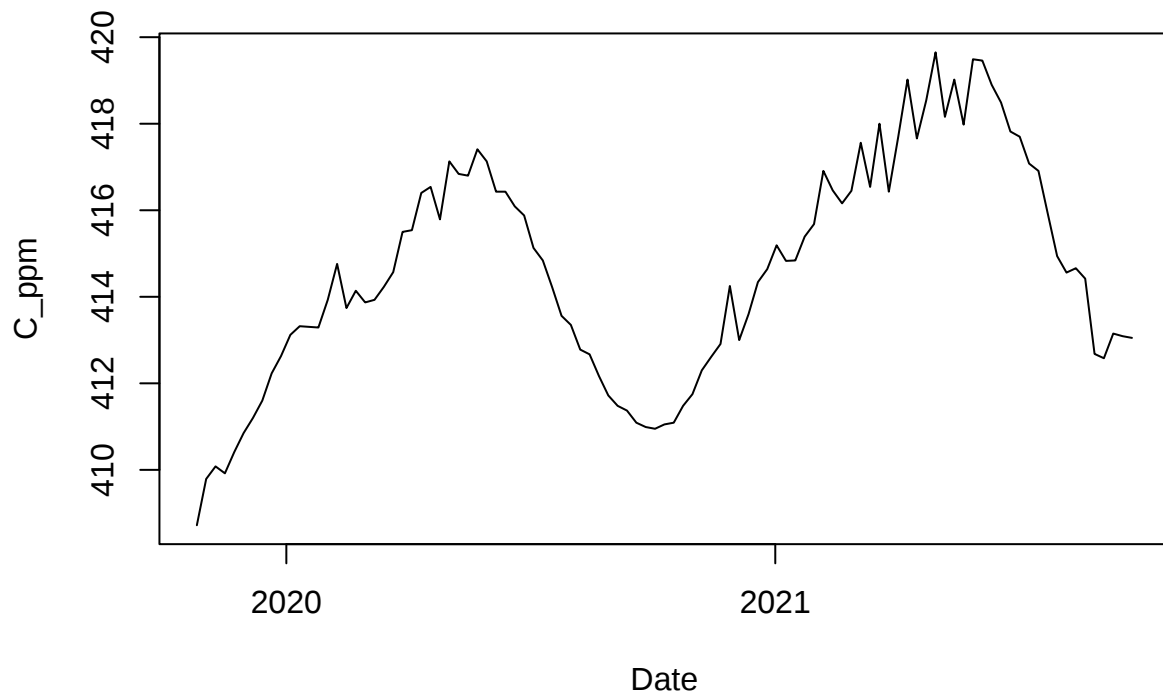
```
plot(data, type = "l")
```



On observe une tendance positive avec une oscillation.

Zoom sur 2ans

```
plot(tail(data,100), type = "l")
```



Pour visualiser l'évolution globale on va se concentrer sur les mesure annuelles.

Les oscillation sont annuelles et le pic semble au printemps. Pour éviter de démarrer une année dans le pic on va commencer l'année au 1er août.

```
pic_annuel = function(annee) {
  debut = paste0(annee-1, "-08-01")
  fin = paste0(annee, "-08-01")
  semaines = data$Date > debut & data$Date <= fin
  return(c(debut, fin, mean(data$C_ppm[semaines]), sd(data$C_ppm[semaines])))
}

annees = 1959:2021

annee = NULL
dates = NULL
mean_C02 = NULL
sd_C02 = NULL

for (i in annees) {
  annee = c(annee, rep(i, 2))
  C02 = pic_annuel(i)
  dates = c(dates, C02[1], C02[2])
  mean_C02 = c(mean_C02, rep(C02[3], 2))
  sd_C02 = c(sd_C02, rep(C02[4], 2))
}

inc_annuelle = data.frame(annee = annee, dates, mean_C02, sd_C02)
```

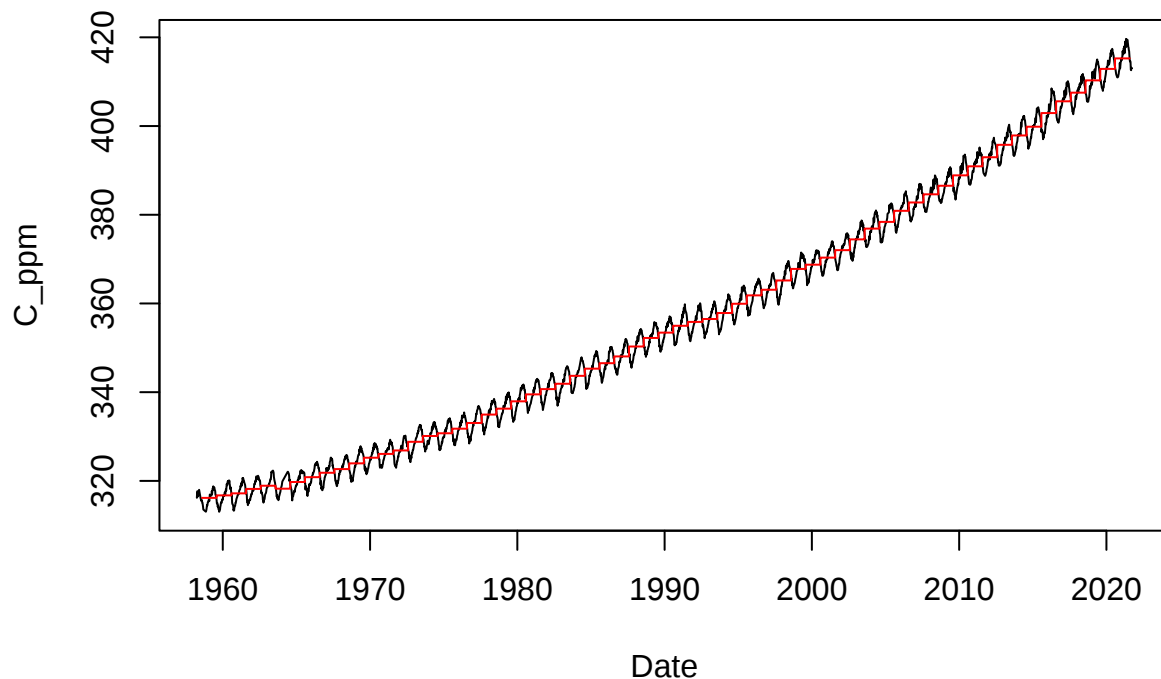
```
str(inc_annuelle)
```

```
## 'data.frame': 126 obs. of 4 variables:
## $ annee : int 1959 1959 1960 1960 1961 1961 1962 1962 1963 1963 ...
## $ dates : chr "1958-08-01" "1959-08-01" "1959-08-01" "1960-08-01" ...
## $ mean_CO2: chr "316.171707317073" "316.171707317073" "316.73568627451" "316.73568627451" ...
## $ sd_CO2 : chr "1.59830989241608" "1.59830989241608" "2.1911186685362" "2.1911186685362" ...
```

```
inc_annuelle$dates = as.Date(convert_week(inc_annuelle$dates))
inc_annuelle$mean_CO2 = as.numeric(inc_annuelle$mean_CO2)
inc_annuelle$sd_CO2 = as.numeric(inc_annuelle$sd_CO2)
```

On refait le graphique en ajoutant par dessus la tendance.

```
plot(data, type = "l")
points(inc_annuelle$mean_CO2~inc_annuelle$dates, col = "red", type = "l")
```



```
#points((inc_annuelle$mean_CO2-inc_annuelle$sd_CO2)~inc_annuelle$dates, col = "blue", type = "l")
#points((inc_annuelle$mean_CO2+inc_annuelle$sd_CO2)~inc_annuelle$dates, col = "blue", type = "l")
```

Etude de la tendance

Pour estimer la tendance on peut utiliser une régression linéaire.

Il faut d'abord créer le modèle.

```
m = lm(inc_annuelle$mean_CO2~inc_annuelle$dates)
```

```
summary(m)
```

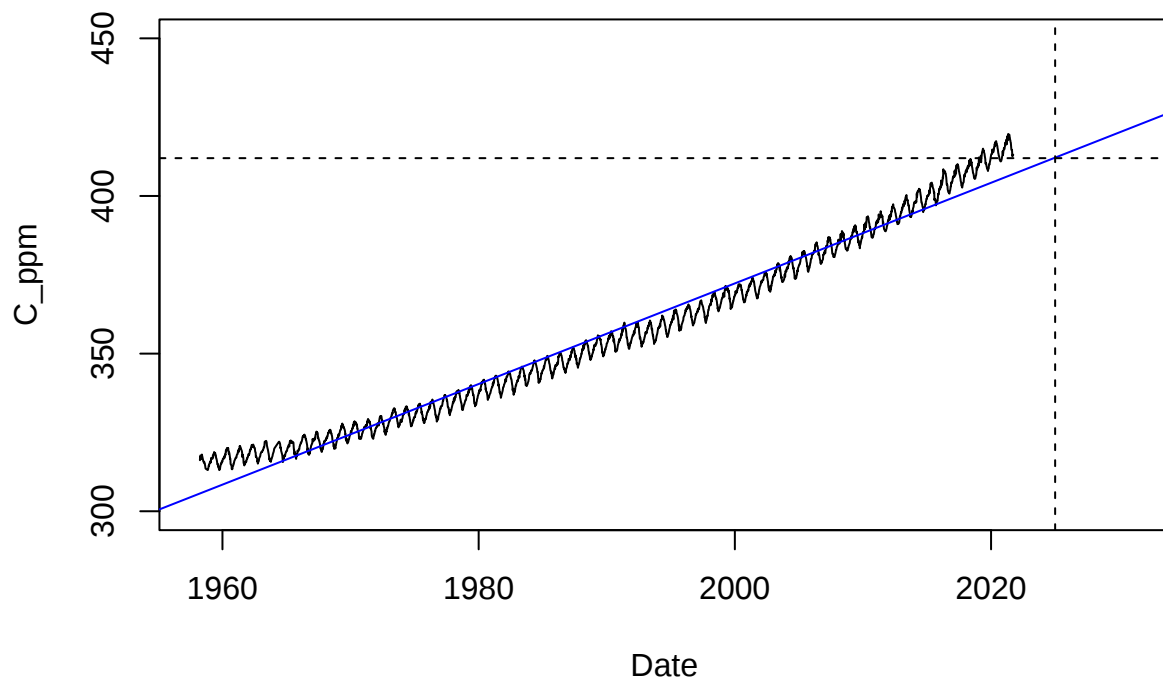
```
##
## Call:
## lm(formula = inc_annuelle$mean_CO2 ~ inc_annuelle$dates)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.793 -3.228 -1.467  2.369 10.110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.244e+02  5.450e-01  595.27  <2e-16 ***
## inc_annuelle$dates 4.369e-03  5.507e-05   79.34  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.107 on 124 degrees of freedom
## Multiple R-squared:  0.9807, Adjusted R-squared:  0.9805
## F-statistic: 6295 on 1 and 124 DF, p-value: < 2.2e-16
```

Notre modèle est bon, il explique 98% des données ($R^2=0.98$). La droite a pour intercept 3.244e+02 et pour pente 4.369e-3

```
plot(data, type = "l", xlim = c(as.Date("1958-01-01"),as.Date("2030-12-31")), ylim = c(300,450))

abline(a = 3.244e+02, b = 4.369e-03, col = "blue")

abline(v=as.Date("2025-01-01"), lty=2)
abline (412, 0, lty = 2)
```



Maintenant on vas tenter d'extrapoler avec notre modèle à l'horizon 2025.

On peut extrapoler grâce au graphique qu'à l'horizon 2025 la quantité de CO₂ sera d'environ 412 mircomol par mol. On cette valeur est déjà dépassé en 2021 ce qui montre que notre modèle devrait être affiné pour ameliorer nos predictions.