

# 大規模データ処理法 課題1

氏名: 金城 信哉

学籍番号: 71942674

学部・学年: 環境情報学部 4年

## 1. 解析の目的

降水量の合計とCO2にどの程度相関があるか調べる。

教室内の生徒数に応じてCO2濃度は上昇すると考えられるため、登校する生徒数に影響を与えるであろう降水量はCO2濃度と相関があるのではないかと考えた。

## 2. 解析の手順

気象庁サイト(<https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>)を利用し、以下の条件でデータを取得する。

- 地点: 辻堂
- 項目: 降水量の日合計
- 期間: 2021/10/20 - 2022/05/01

nkfコマンドを使用して、取得したデータの文字コードをshift-jisに変換する。

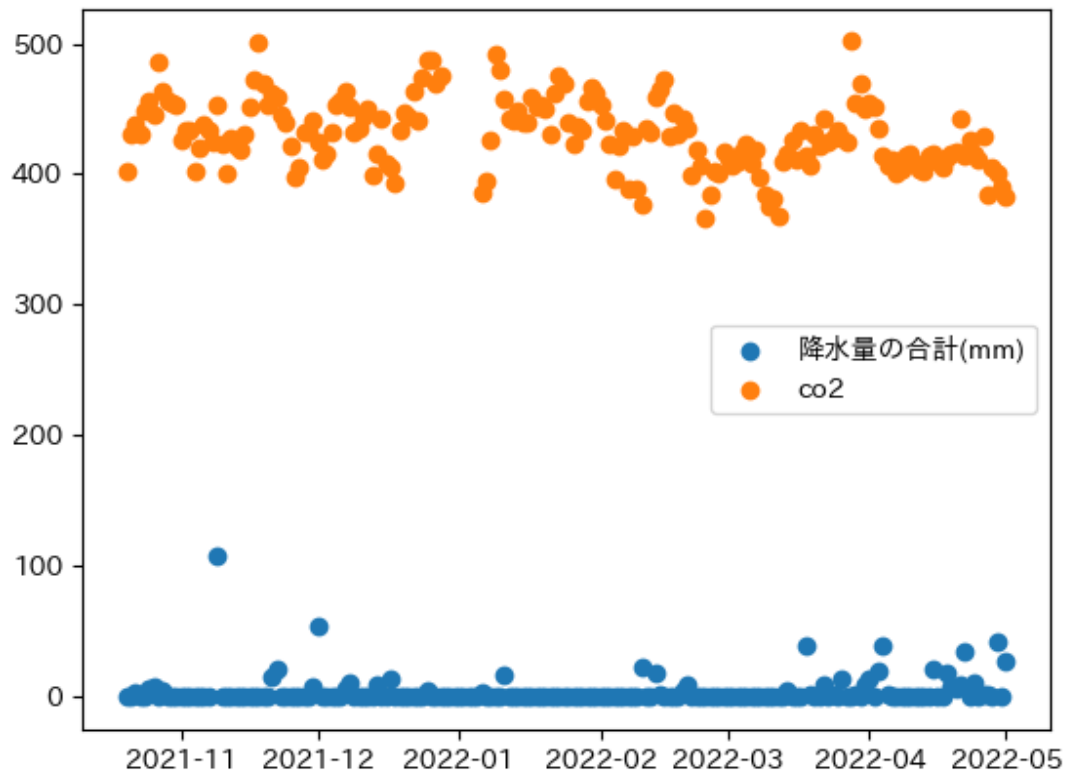
SFCのデータは、θ館のCO2を利用し、ファイル名をtheta.csvとする。θ館を利用する理由は、生徒の収容数が最も多いため降水量により変化する量が大きいと考えるためである。

theta.csvのtimestampをdatetime型に変換し、インデックスに設定する。

co2データについて日毎に平均を取る。

「気象庁から入手した降水量のファイル」と「theta.csvのco2データについて日毎に平均を取ったファイル」を突合させたファイルをcombined.csvとする。

以下の図は、降水量の合計とCO2濃度を散布図に表示させたものである。



また、降水量の合計とCO2濃度の相関係数は、-0.019292であった。

### 3. 解析の結果わかったこと

残念ながら、当初想定していた降水量の合計とCO2濃度に相関は見られなかった。また、今回の結果をもとに、以下2つの可能性があるのではないかと私は考える。1つ目は「学生数はCO2濃度に関係ない」という点である。学生の数そのもののデータが存在しないため、学生数とCO2濃度に関係があるかどうかは議論の余地があるだろう。2つ目は「降水量の合計値は登校する学生数に関係ない」という点である。仮に最初に立てた「学生数が増えるとCO2濃度も上昇する」という仮説が正しかった場合、降水量が増えても登校する学生数はほぼ変化しないことが考えられる。

### 4. 感想

仮説を立ててデータの分析をし、得られた結果をもとに考察をする、という一連のデータ分析の流れを実感できました。また、今回のように想定していた結果にならなかったケースであっても「なぜうまく行かなかったのか」「他にはどのような可能性があるだろうか」といったことが考えられ、大変勉強になりました。