# UV Absorbance characteristics in Northern Lakes

https://github.com/rachelbash/absorbance-data-project

*Rachel Bash*

**Abstract**

Experimental overview. This section should be no longer than 250 words. What contributes to absorbance values in the NTL_LTER Carbon data set (will consider things such as DIC, DOC, depth, and water pressure). Also, is there a significant change in absorbance values in lakes over time?

# Contents

# List of Tables

# List of Figures

# 1 Research Question and Rationale

Absorbance is a unitless measurement that describes how much a substance absorbs light over a certain range of wavelenght. The absorbance values of water samples from lakes can provide details regarding its physical characteristics and the health of the lake. The amount of light entering a lake is an component that drives photosynthesis and lake metabolism. Additionally, lake temperature and its absorbance characteristics are deeply intertwined. With the right equipment, absorbance is fairly easy to measure. Therefore, measuring absorbance in lakes can give researchers insight into other processes happening that depend on sunlight.

This research project intends to answer two main questions: What contributes to absorbance values in five lakes located in Michigan's Upper Penninsula? Do absorbance values in these five study lakes change over time? The data that answer these questions come from the North Temperate Lakes Project, which seeks to measure data on carbon and other related variables in lakes. My analysis of the data provides a model that shows the variables that best predict absorbance values and also takes a closer look at how absorbance values have changed over time in different lakes. Time variations in absorbance have implications that other physical characteristics are changing, which may damage biota in the lakes or bring about significant changes in the greater ecosystem that surrounds the lake.

# 2 Dataset Information

The dataset was collected from 1984 to 2016 by researchers working for the Cascade Project and Northern Temperate Lakes at a total of 14 sites. Samples of water were collected, and then were measured. Measurements included dissolved organic and inorganic carbon, particulate organic matter, partial pressure of carbon dioxide, and absorbance. Absorbance was measured using a spectrophotometer at a wavelength of 440nm.

For some variables, a water depth sample was taken that was measured in meters, while in others, samples were taken to reflect a depth that was proportional across all lakes. Therefore, Hypolimnion, Epilimnion, Metalimnion, and pooled mixed layer (PML) are also included as depth values. All water samples were taken with a syringe and then filtered through a mesh filter in order to remove any large debris or zooplankton.

| Data Summary | Relevant Information |
|---|---|
| Date range | 1984-06-03 to 2016-08-17 |
| Structure | 15 variables with 13,557 observations |
| Column names | lakeid, lakename, year4, daynum, sampledate, depth, depth_id, tpc, tpn, DIC_mg, DIC_uM, air_pco2, water_pco2, doc, absorbance |
| Lakes sampled | Crampton Lake, East Long Lake, Hummingbird Lake, Long Lake, Morris Lake, North Gate Bog, Paul Lake, Peter Lake, Reddington Lake, Roach Lake, Tender Bog, Tuesday Lake, Ward Lake, West Long Lake |

# 3 Exploratory Data Analysis and Wrangling

#### 3.0.0.1 Importing raw data and identifying its attributes

```
#exploratory code to see the full dataset and its attributes
colnames(carbon.data)
```

```
##  [1] "lakeid"     "lakename"   "year4"      "daynum"     "sampledate"
##  [6] "depth"      "depth_id"   "tpc"        "tpn"        "DIC_mg"
## [11] "DIC_uM"     "air_pco2"   "water_pco2" "doc"        "absorbance"
```

```
str(carbon.data)
```

```
## 'data.frame':    13557 obs. of  15 variables:
##  $ lakeid    : Factor w/ 14 levels "E","H","L","Long",..: 3 3 3 3 3 8 8 8 8 8 ...
##  $ lakename  : Factor w/ 14 levels "Crampton Lake",..: 7 7 7 7 7 8 8 8 8 8 ...
##  $ year4     : int  1984 1984 1984 1984 1984 1984 1984 1984 1984 1984 ...
##  $ daynum    : int  155 155 155 155 155 156 156 156 156 156 ...
##  $ sampledate: Date, format: "1984-06-03" "1984-06-03" ...
##  $ depth     : Factor w/ 231 levels "0","0.1","0.15",..: 1 62 102 140 180 1 62 102 14
##  $ depth_id  : int  1 2 3 4 5 1 2 3 4 5 ...
##  $ tpc       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ tpn       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ DIC_mg    : num  1.45 1.82 1.51 1.47 2.69 2.85 2.84 3.27 2.98 7.26 ...
##  $ DIC_uM    : num  121 152 126 122 224 ...
##  $ air_pco2  : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ water_pco2: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ doc       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ absorbance: num  NA NA NA NA NA NA NA NA NA NA ...
```

```
summary(carbon.data)
```

```
##      lakeid               lakename        year4          daynum
##  R      :3887   Peter Lake    :3887   Min.   :1984   Min.   : 82.0
##  L      :3852   Paul Lake     :3852   1st Qu.:1993   1st Qu.:166.0
##  T      :1818   Tuesday Lake  :1818   Median :1999   Median :192.0
##  W      :1571   West Long Lake:1571   Mean   :2000   Mean   :192.4
##  E      :1435   East Long Lake:1435   3rd Qu.:2007   3rd Qu.:218.0
##  M      : 456   Crampton Lake : 456   Max.   :2016   Max.   :310.0
##  (Other): 538   (Other)       : 538
##    sampledate                 depth         depth_id           tpc
##  Min.   :1984-06-03   0          :1719   Min.   :-2.000   Min.   : 0.100
##  1st Qu.:1993-06-16   Metalimnion:1297   1st Qu.: 1.000   1st Qu.: 0.580
##  Median :1999-07-06   Hypolimnion:1020   Median : 3.000   Median : 0.890
##  Mean   :2000-07-14   PML        : 876   Mean   : 2.775   Mean   : 1.110
##  3rd Qu.:2007-08-28   Epilimnion : 570   3rd Qu.: 5.000   3rd Qu.: 1.305
##  Max.   :2016-08-17   (Other)    :7918   Max.   : 7.000   Max.   :11.860
```

```
##                       NA's       : 157   NA's   :170      NA's   :11410
##       tpn              DIC_mg            DIC_uM            air_pco2
##  Min.   :0.000   Min.   : 0.023   Min.   :   1.917   Min.   :197.7
##  1st Qu.:0.070   1st Qu.: 0.812   1st Qu.:  67.625   1st Qu.:343.4
##  Median :0.103   Median : 1.322   Median : 110.167   Median :362.9
##  Mean   :0.149   Mean   : 2.310   Mean   : 192.487   Mean   :360.4
##  3rd Qu.:0.180   3rd Qu.: 1.968   3rd Qu.: 164.000   3rd Qu.:379.0
##  Max.   :2.170   Max.   :48.599   Max.   :4049.883   Max.   :608.1
##  NA's   :11409   NA's   :3642     NA's   :3642       NA's   :12411
##    water_pco2          doc            absorbance
##  Min.   :   0.0   Min.   : 2.710   Min.   :0.011
##  1st Qu.: 478.0   1st Qu.: 4.570   1st Qu.:0.060
##  Median : 838.5   Median : 5.603   Median :0.146
##  Mean   :1012.3   Mean   : 6.932   Mean   :0.194
##  3rd Qu.:1175.6   3rd Qu.: 8.370   3rd Qu.:0.265
##  Max.   :9348.2   Max.   :44.080   Max.   :1.213
##  NA's   :12411    NA's   :9993     NA's   :10658
```

```
dim(carbon.data)
```

```
## [1] 13557    15
```

```
summary(carbon.data$absorbance)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.011   0.060   0.146   0.194   0.265   1.213   10658
```

```
class(carbon.data$depth)
```

```
## [1] "factor"
```

```
head(carbon.data$depth, 10)
```

```
##  [1] 0   1   2   3.5 5.5 0   1   2   3.5 7
## 231 Levels: 0 0.1 0.15 0.17 0.18 0.19 0.2 0.21 0.22 0.23 0.25 0.28 ... surface
```

These exploratory commands above function as helpful tools that help me see what kind of shape my data are in. It shows me how many NA's I have, what variables I am working with, the classes of my variables, and basic summary statistics. An important thing I discovered while doing the initial exploratory data analysis is that the depth variable has both numeric and factor-level observations, which is why its class is listed as `factor`. In other words, depth was measured in both numeric terms (1 meter, 13 meters, etc), but also in thermally stratified terms, such as Hypolimnion, Metalimnion, and Epilimnion. This was an important discovery that led to further data wrangling and filtering of this specific variable.

#### 3.0.0.2 Visualizing the data

As seen by Figure 1, Absorbance values are not normally distributed. This is expected, as we

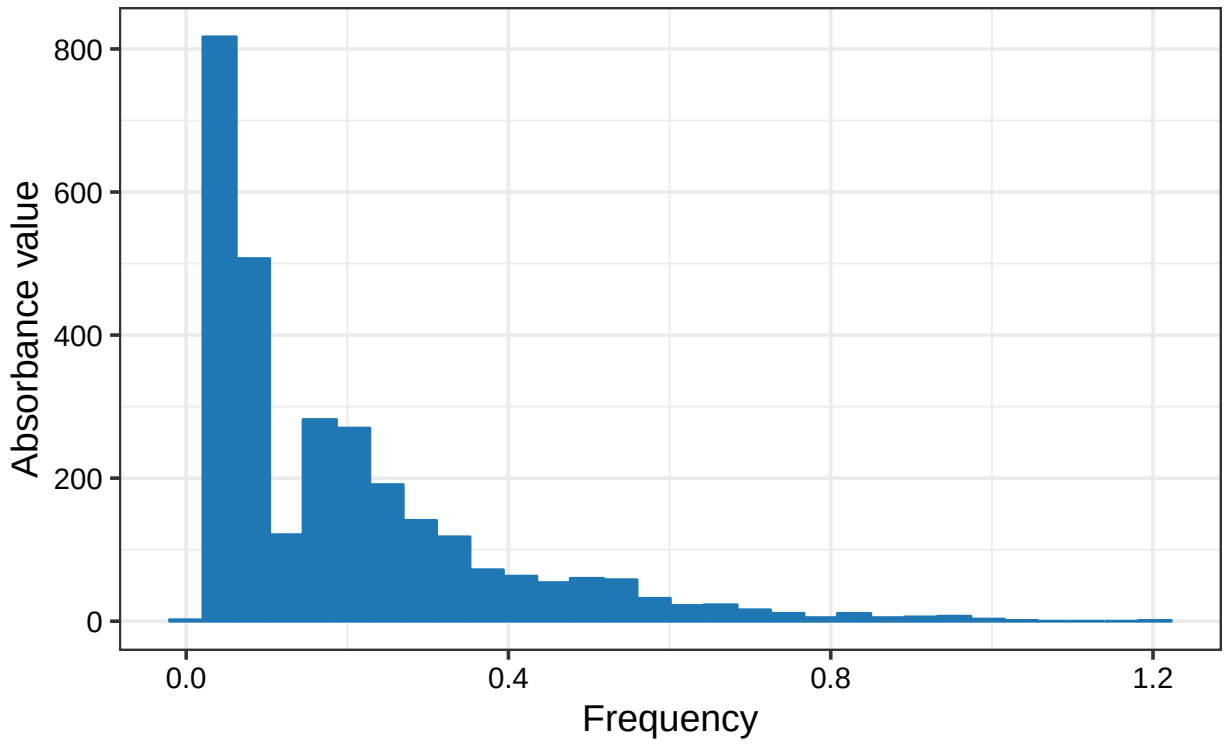


Figure 1: Absorbance frequency

are dealing with ecological data.

Relatedly, Figure 2 shows that different levels of depth (factor) had difference absorbance freqency values. It was helpful to create this graph to show that absorbance was measured at multiple different water depth levels.

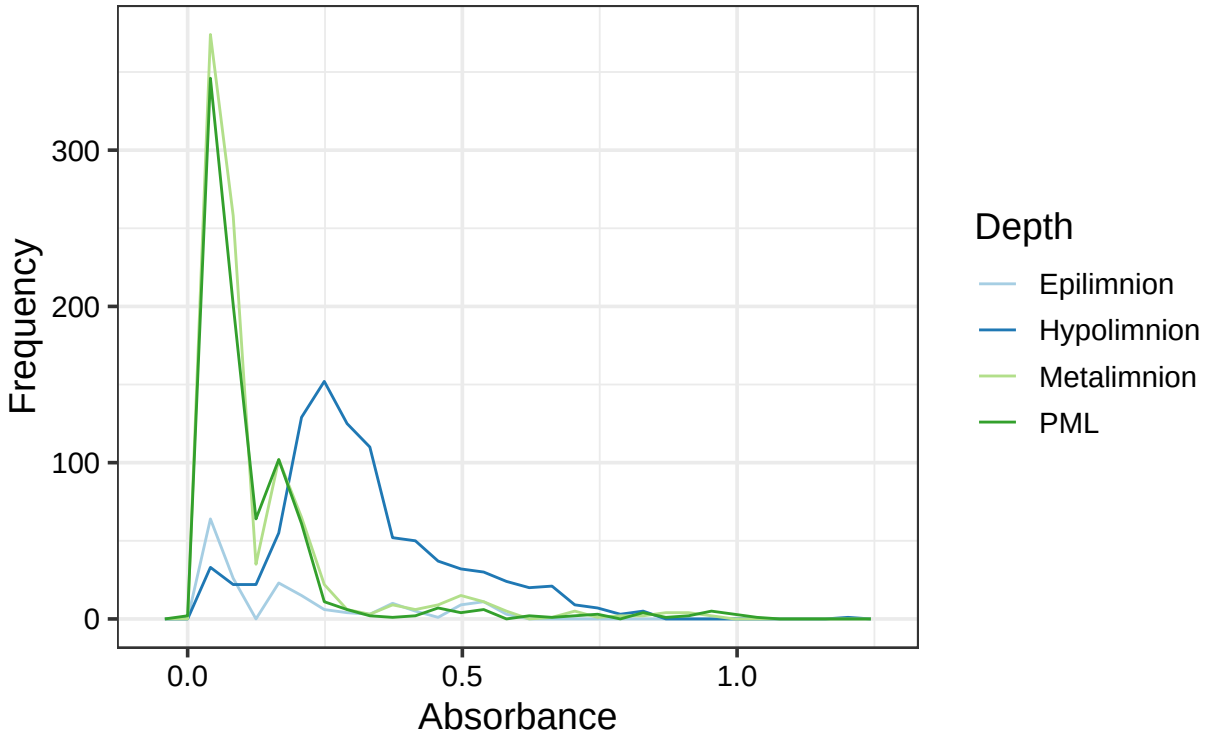


Figure 2: Absorbance frequency by depth category

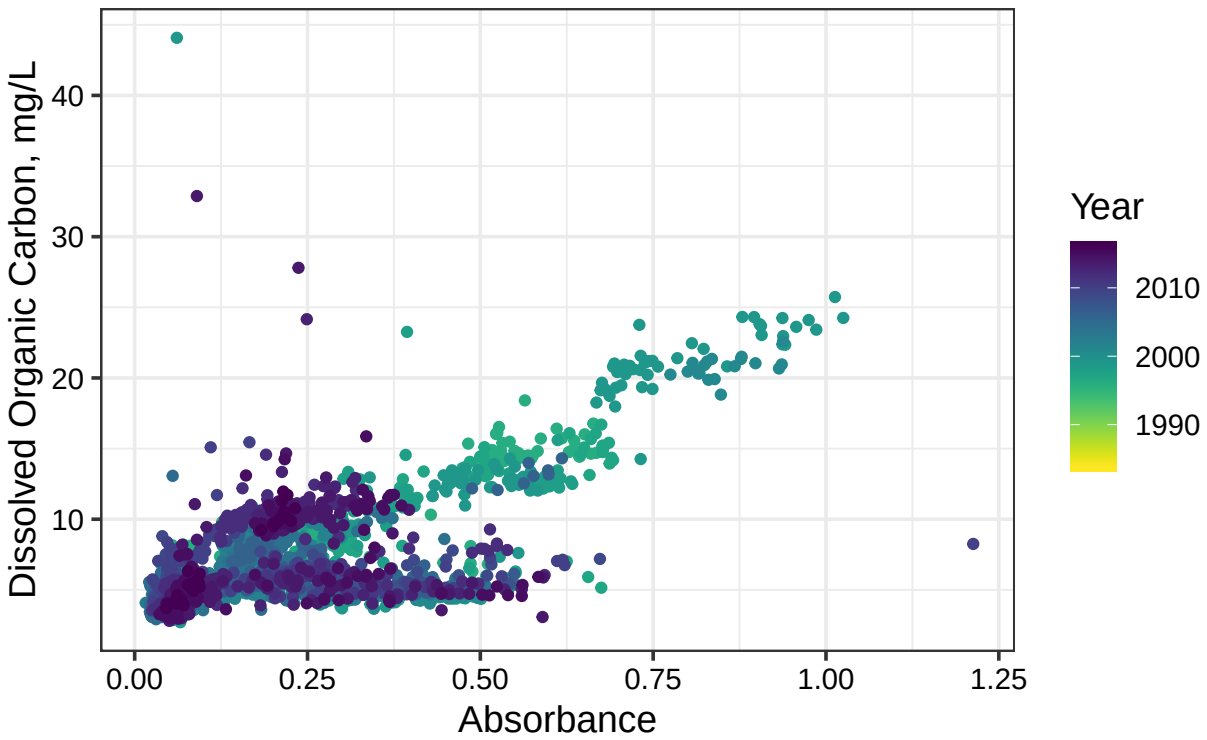


Figure 3: Disolved organic carbon and absorbance relationship by year

# 4 Analysis

# 5 Summary and Conclusions