

Experiment Title

https://github.com/cwatson1013/Env_Data_Analysis_Final_Proj.git

Caroline Watson

Abstract

Experimental overview. This section should be no longer than 250 words. put abstract here

Contents

1	Research Question and Rationale	5
2	Dataset Information	6
3	Exploratory Data Analysis and Wrangling	7
4	Analysis	11
5	Summary and Conclusions	12

Contents

List of Tables

List of Figures

1	Dissolved Organic Carbon (DOC) over time	9
2	Dissolved Inorganic Carbon in each Lake	10

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

The rationale for this analysis is because there typically is a relationship between dissolved organic carbon and depth. There is also typically a relationship between land area surrounding lakes that have high amounts of organic soils usually deposit large amounts of dissolved organic carbon into lakes. Dissolved inorganic carbon is an important part of the carbon cycle and supplies nutrients for some organisms. Most DOC is natural, but high amounts can indicate human influence, such as land surrounding the lake that is high in organic amount.

I want to find out whether there is a relationship between dissolved organic carbon (DOC) and depth. If there is a relationship between these two variables, I want to see if this relationship varies seasonally. I am using a dataset that contains various parameter measurements for different lakes in the North Temperate Region in Wisconsin, USA. Parameters measured include temperature, depth, dissolved organic carbon, dissolved inorganic carbon, particulate organic matter and others.

2 Dataset Information

```
#reading in data file
carbon.data <- read.csv("./Data/Raw/NTL-LTER_Lake_Carbon_Raw.csv")

#structure of data frame
carbon.data_summary <- summary(carbon.data)

#summary of data structure
kable(carbon.data_summary) %>%
  kable_styling()
```

	lakeid	lakename	year4	daynum	sampledate	depth
	R :3887	Peter Lake :3887	Min. :1984	Min. : 82.0	5/24/99: 18	0 :1719
	L :3852	Paul Lake :3852	1st Qu.:1993	1st Qu.:166.0	5/25/99: 18	Metalimnion:1
	T :1818	Tuesday Lake :1818	Median :1999	Median :192.0	5/26/99: 18	Hypolimnion:1
	W :1571	West Long Lake:1571	Mean :2000	Mean :192.4	5/31/99: 18	PML : 876
	E :1435	East Long Lake:1435	3rd Qu.:2007	3rd Qu.:218.0	6/1/99 : 18	Epilimnion : 5
	M : 456	Crampton Lake : 456	Max. :2016	Max. :310.0	6/14/99: 18	(Other) :7918
	(Other): 538	(Other) : 538	NA	NA	(Other):13449	NA's : 157

3 Exploratory Data Analysis and Wrangling

```
#class of sampleddate column
class(carbon.data$sampledate)

## [1] "factor"

#converting sampleddate to a date in R
carbon.data$sampledate <- as.Date(carbon.data$sampledate, format = "%m/%d/%y")

#checking class of sampleddate
class(carbon.data$sampledate)

## [1] "Date"

#summary of the dataset
head(carbon.data)

##   lakeid  lakename year4 daynum sampleddate depth depth_id tpc tpn DIC_mg
## 1      L   Paul Lake  1984   155 1984-06-03     0         1  NA  NA   1.45
## 2      L   Paul Lake  1984   155 1984-06-03     1         2  NA  NA   1.82
## 3      L   Paul Lake  1984   155 1984-06-03     2         3  NA  NA   1.51
## 4      L   Paul Lake  1984   155 1984-06-03    3.5         4  NA  NA   1.47
## 5      L   Paul Lake  1984   155 1984-06-03    5.5         5  NA  NA   2.69
## 6      R Peter Lake  1984   156 1984-06-04     0         1  NA  NA   2.85
##   DIC_uM air_pco2 water_pco2 doc absorbance
## 1 120.8333      NA          NA  NA         NA
## 2 151.6667      NA          NA  NA         NA
## 3 125.8333      NA          NA  NA         NA
## 4 122.5000      NA          NA  NA         NA
## 5 224.1667      NA          NA  NA         NA
## 6 237.5000      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
##   sampleddate      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
```

```
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"
```

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

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

```
#renaming columns
```

```
colnames(carbon.data)[1:5] <- c("Lake.ID", "Lake.Name", "Year", "Day.Number", "Date")
```

```
#graph looking at DOC over time
```

```
ggplot(carbon.data) +
  geom_point(aes(x = carbon.data$Year,
y = carbon.data$doc))
```

The following graphs explore the Carbon dataset. Figure 1 shows dissolved organic carbon over time. Figure 1 was created to determine if there was a pattern of dissolved organic carbon in lakes over the years. Figure 2 is a frequency polygon graph looking at the dissolved inorganic carbon (DIC) in each lake. This graph was created to see if there are some lakes with higher DIC than others, which could influence further analysis.

```
#frequency polygon graph looking at DIC in each lake
```

```
ggplot(carbon.data) +
  geom_freqpoly(aes(x = carbon.data$DIC_mg,
```

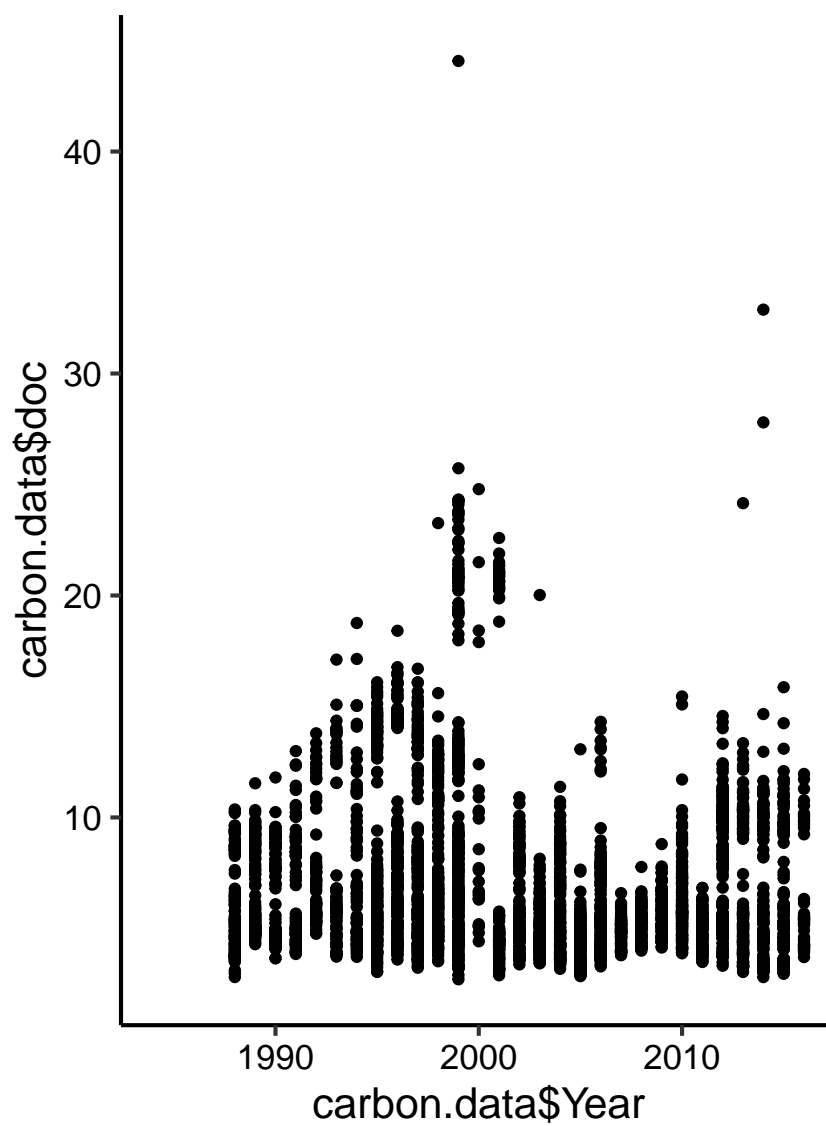



Figure 1: Dissolved Organic Carbon (DOC) over time

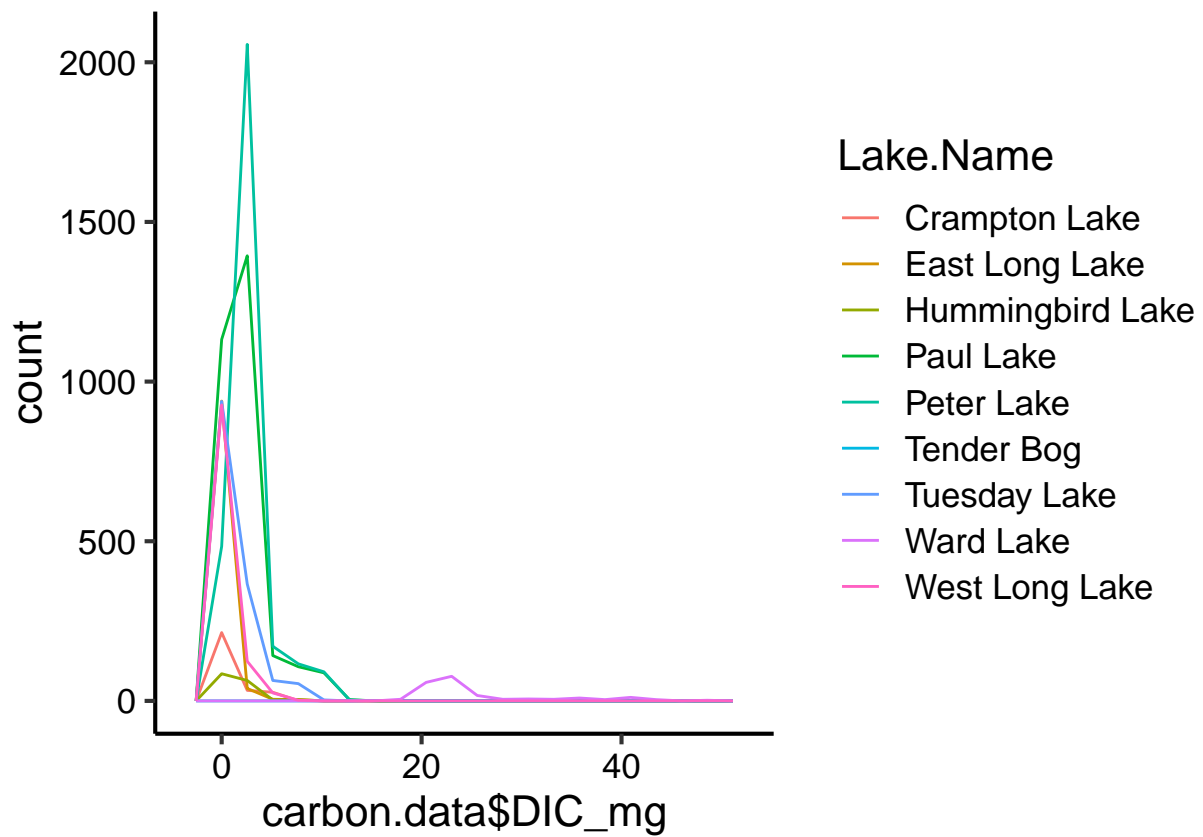


Figure 2: Dissolved Inorganic Carbon in each Lake

```
color = Lake.Name), bins = 20)
```

```
#
```

```
#selecting date, DIC, DOC, Lake Name
```

4 Analysis

5 Summary and Conclusions