Ellerbee Creek Case Study

Harper Schwab

10/13/2020

Introduction

Water quality can be measured in many different ways, usually resulting in a multi-variable analysis of the water. Examples of variables that can be analyzed include but are not limited too: Water temperature, water pressure, stream depth (mainly used to see if the depth is the cause for changes in conditions), the speed of the flow (used similarly), dissolved oxygen, pH level, NO3 content, turbidity, conductivity, and NH4 content. For a temperate climate like North Carolina, aquatic life likes to be in the range of 25-27 degrees Celsius. the stream depth, as I mentioned can help a analyst understand weather or not, it is the depth affecting different variables and not other things. For instance, deeper water makes for colder temperatures which can result in a differing pH level. The speed of the flow of the water can affect things like dissolved oxygen and turbidity. Places where streams are fast, we then, can disregard to some degree the dissolved oxygen and turbidity levels. Dissolved oxygen is measured in mg/L. 0-4 mg/L cannot support any fish, 4-6.5 can support very few, 6.5-9.5 can support some big fish but not many small fish, and 9.5-12 can support all fish (Dissolved Oxygen (DO)). pH is the acidity basic scale that we measure liquids. The pH scale is 0-14 with 0 being acidic, while 14 is basic. Healthy water stays around 7, aka neutrality. Fish like a pH that doesn't deviate more than roughly .5 away for neutral. pH is affected by temperature in a inverse relationship. Factors such as dissolved CO2 affect pH as well. NO3 is the amount of nitrogen present in water. The federal standard for nitrates in drinking water are 45 mg/L of NO3 (Wagenet).Turbidity is the measure of clarity of the water. it is a visual characteristic to see the amount of dissolved solids in the liquid. Factors such as rain can affect turbidity because of soil runoff. Other factors are dumping/ pollution from humans. Conductivity can also show how 'dirty' the water is because pure water is not a good conductor. A high conductivity can showcase dirty water in the creek. NH4 or amonium is harmful to fish and can cause pH to increase, which is another harmful factor towards wildlife in creeks and bodies of water. High NH4 content can be a sign of runoff form fertilizers and human tended crops.

Data Cleanup

To clean the data I first Started to replace all missing values (N/As) with the Mean of the column.

```
data$CFS <- ifelse(is.na(data$CFS), mean(data$CFS, na.rm=TRUE), data$CFS)
data$DO <- ifelse(is.na(data$DO), mean(data$DO, na.rm=TRUE), data$DO)
data$saturation <- ifelse(is.na(data$saturation), mean(data$saturation, na.rm=TRUE), data$saturati
on)
#
```

I then realized, after looking tat the data table, that there were two extremely high poutliers in the column for NH4. So I voided those values and replaced them with the resultng mean of the column.

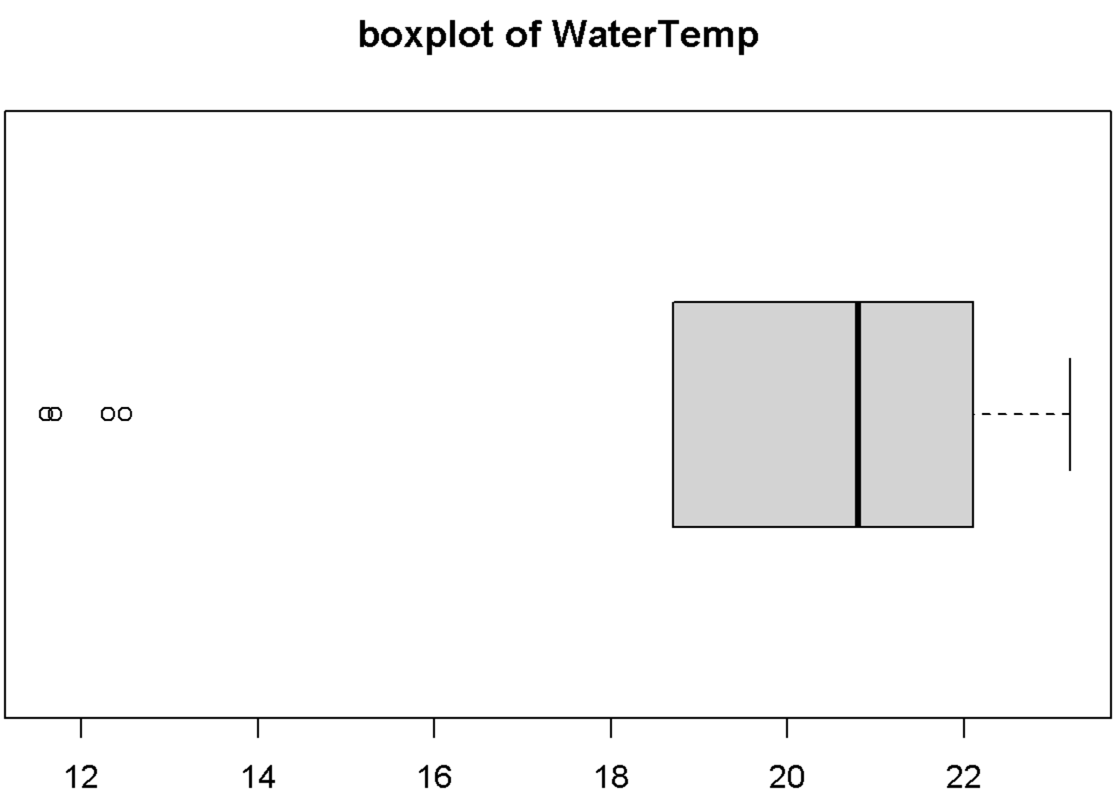
```
#the 100 and 600 seem like false reporting, I will replace them with the NAs then replace N/As wit
h the mean.
data[1,10] <- NA
data[2,10] <- NA
data$NH4 <- ifelse(is.na(data$NH4), mean(data$NH4, na.rm=TRUE), data$NH4)
```

Plots and Data

Now that the data was clean, I started looking discecting plots, and looking at each column indavidually. Through simple statistical boxplots I was able to look at each I was able to discect each variable and see the averages and where the data is skewed.Based on the boxplot of water temperature we can see on average the temperature of the water is at 21 degrees Celsius. There are four outliers between 11-13 degrees. As for water pressure the common range is between 755-765 mmHG, however averages more towards 757 mmHG.The stream is mostly between 4-5.5 feet with the shallow ends reaching roughly 2 feet and the deepest parts going about 7 feet deep. The creek has a average flow rate of 10 cubic flow/second. Dissolved oxygen levels are usually between 6-9.5 mg/L. Which is in the good-excellent range according to Washington state’s standards. It more often then not, below the mean. pH is relatively natural only straying roughly .5 away from the mean. N03 averages 1.5 with more of the data above the mean than below. Turbitity is averaged around 18 NTU which is in the in the 2A, cold water and fishery, all recreation category. in the 10-25 turbity range. However some parts of the creek are in the 2A and 2C categories. The conductance of the water has a wide range of data, looking at the box plot. While the average just above 200 most of the data is spread out up untill 700. NH4 levels average around 1-1.5 and dont go above 5, except for an outlier that is > 20. And because so much of the Saturation data was missing the outliers are the real measurable thing. the data does not exceed .3 of the mean. So it is all relatively together.

For these analysis I used simple box plots like the example below.

```
boxplot(data$WaterTemp, main = 'boxplot of WaterTemp', horizontal = T)
```

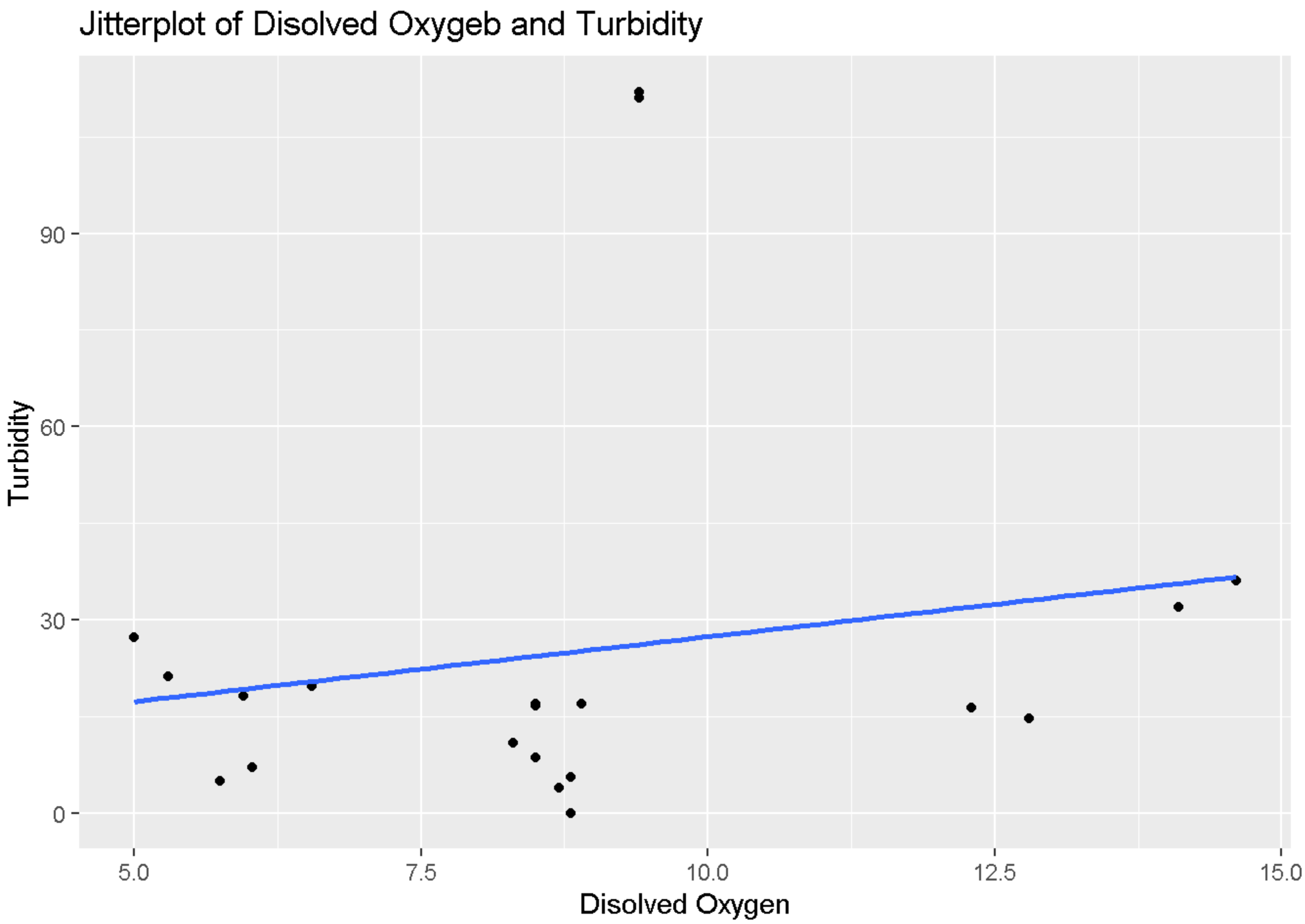


ggplot

The first plot I created was a jitterplot of Disolved Oxygen and Turbidity.

```
jp <- ggplot(data, aes(DO,Turb))
jp <- jp+ geom_point()+geom_smooth(method='lm', se=F)+
  labs(title="Jitterplot of Disolved Oxygeb and Turbidity", y='Turbidity', x='Disolved Oxygen')
jp
```

```
## `geom_smooth()` using formula 'y ~ x'
```



We can see that the is a positive relationship between disolved oxygen and turbidity, except for the points where turbidity has outliers. Maybe the water was so turbid that it was more of a sludge, or perhaps they measured wrong or it is a mistake in the dataset.

The second plot was a scatterplot of Water tempurature and pH levels. I also had stream depth and turbidity as factors that effected the visual of the plot.

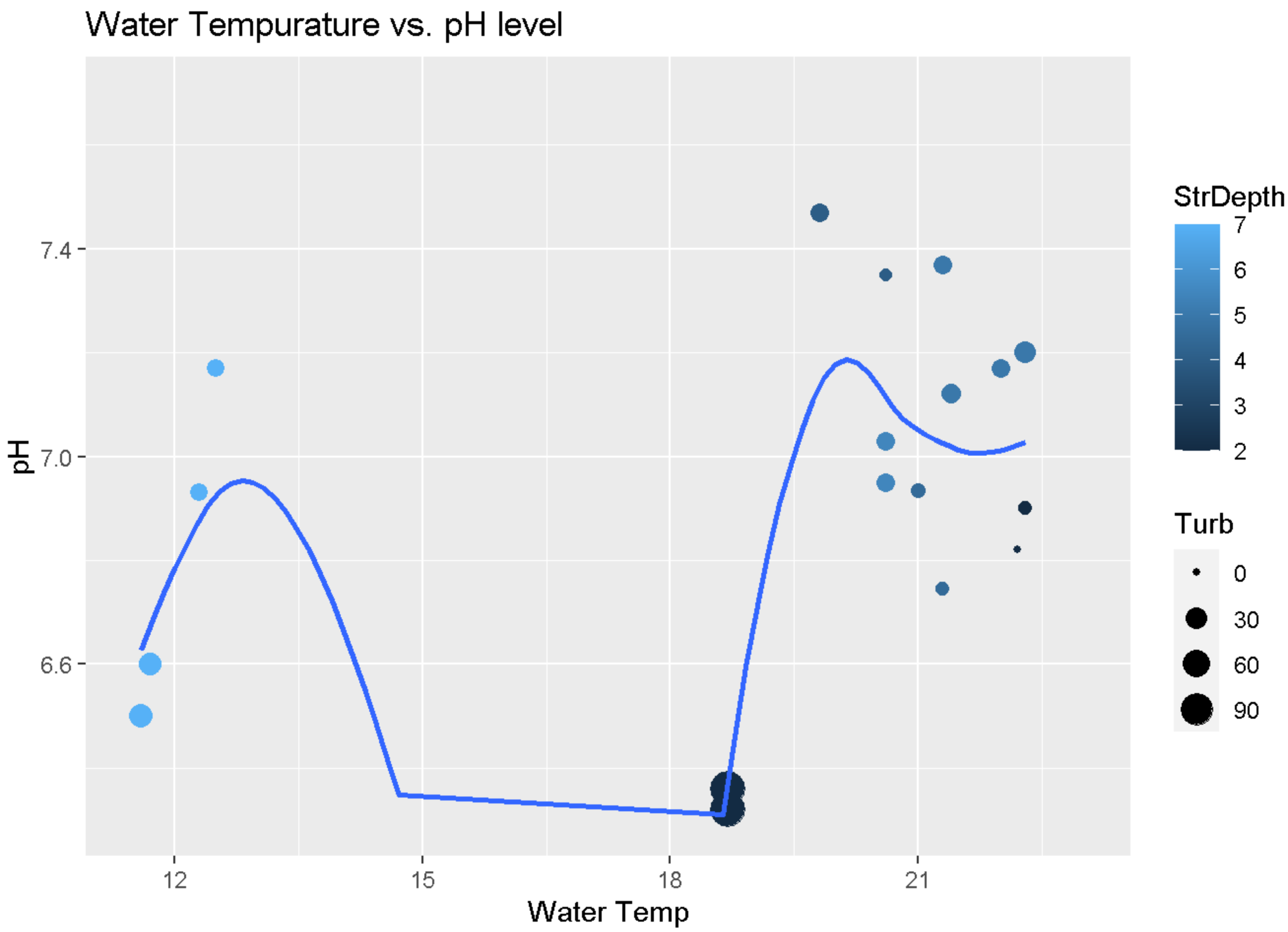
```
sp <- ggplot(data, aes(x=WaterTemp, y=pH)) +
  geom_point(aes(col=StrDepth, size=Turb)) +
  geom_smooth(method='loess', se =F)+
  xlim(c(11.5,23))+
  ylim(c(6.3,7.7))+
  labs(title='Water Tempurature vs. pH level', y = 'pH', x = 'Water Temp')
sp
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 28 rows containing missing values (geom_smooth).
```



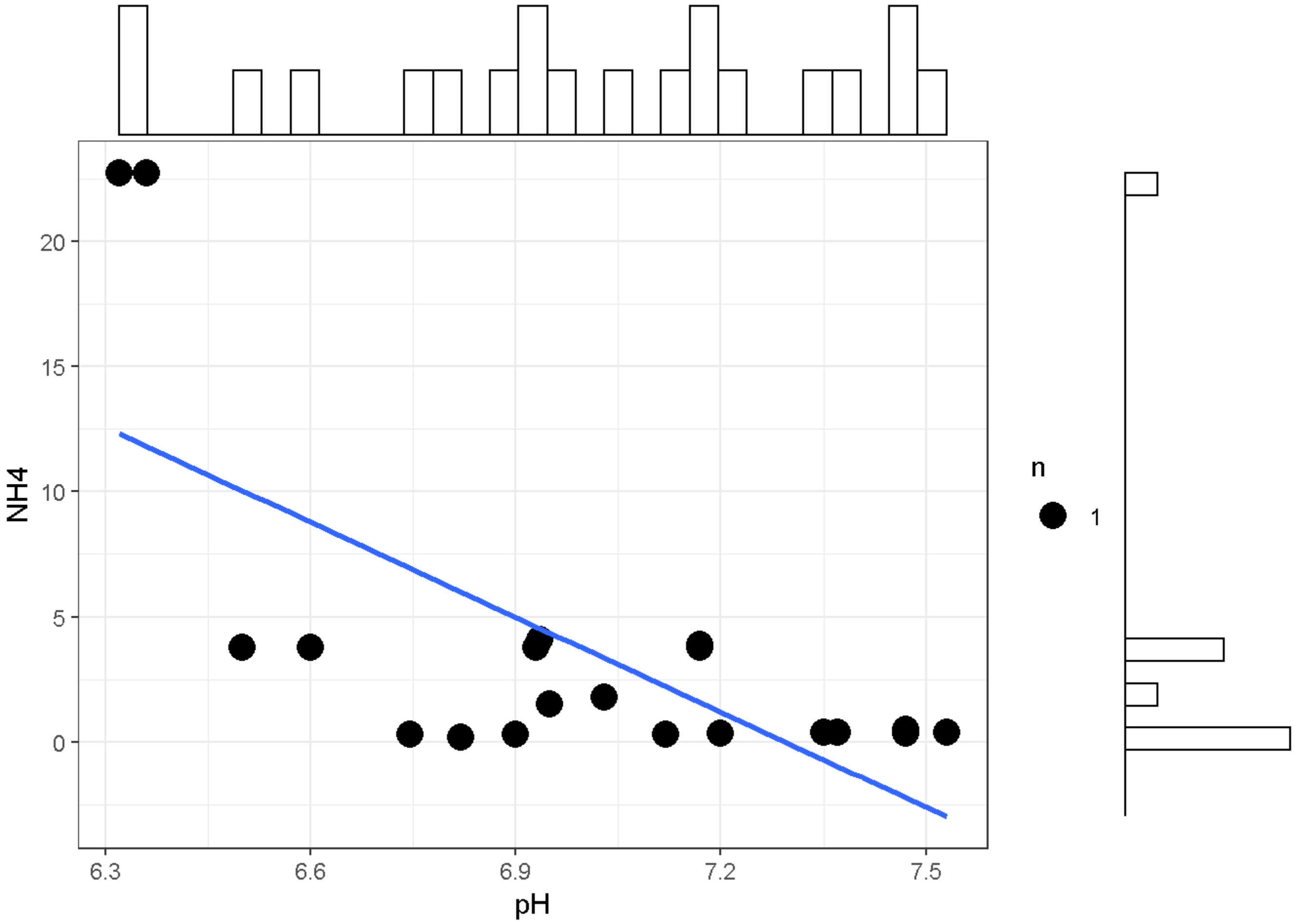
In the plot we are able to see if there is any relationship between temperature and pH, as well as stream depth and turbidity. The thing that first jumps out is the range form x=13-16 that there are no data points. maybe because of the size of the creek, only the 'extremes' are found and not the middle ground. With he extremes meaning roughly 12 degrees or between 18-22 degrees. One relationship that is visible is the two lowest points on the pH scale and the fact that they are the most turbid and in the highest ;pcations of the creek. The coldest areas of the creek have the largest Depth.

Next, because Having excessive NH4 in the water can be harmful to aquatic life I plotted a scatter chart and density histogram on the same graphic for NH4 and pH.

```
theme_set(theme_bw())
HnS <- ggplot(data, aes(pH, NH4))+
  geom_count()+
  geom_smooth(method='lm', se=F)

ggMarginal(HnS, type='histogram',fill='transparent')
```

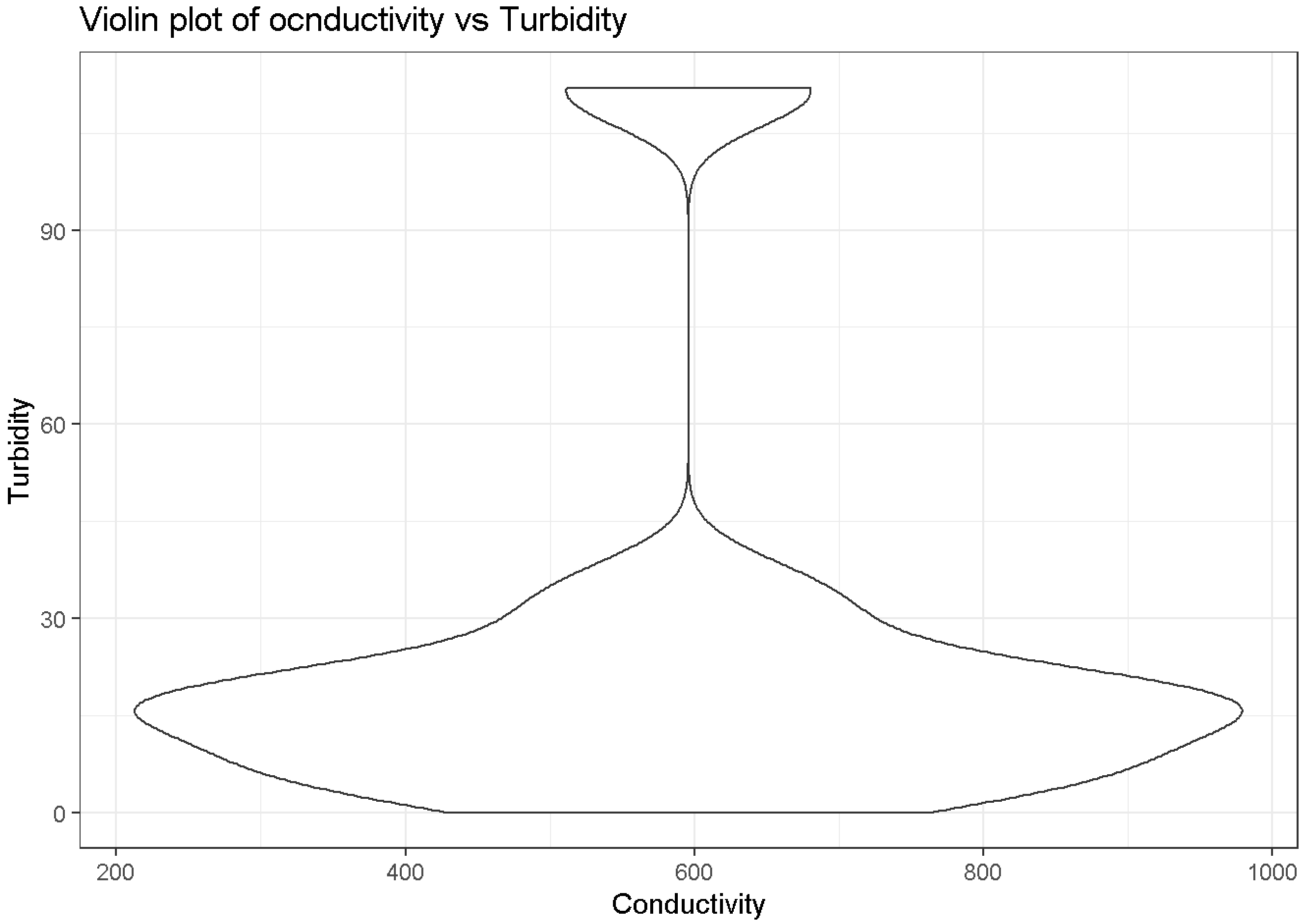
```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



If we exclude the two outlines on the NH4 scale, we can see that the even distribution of pH matches up with the constant levels of NH4.

For my fourth plot I decided to doa volin plot of Conductivity and Turbidity.

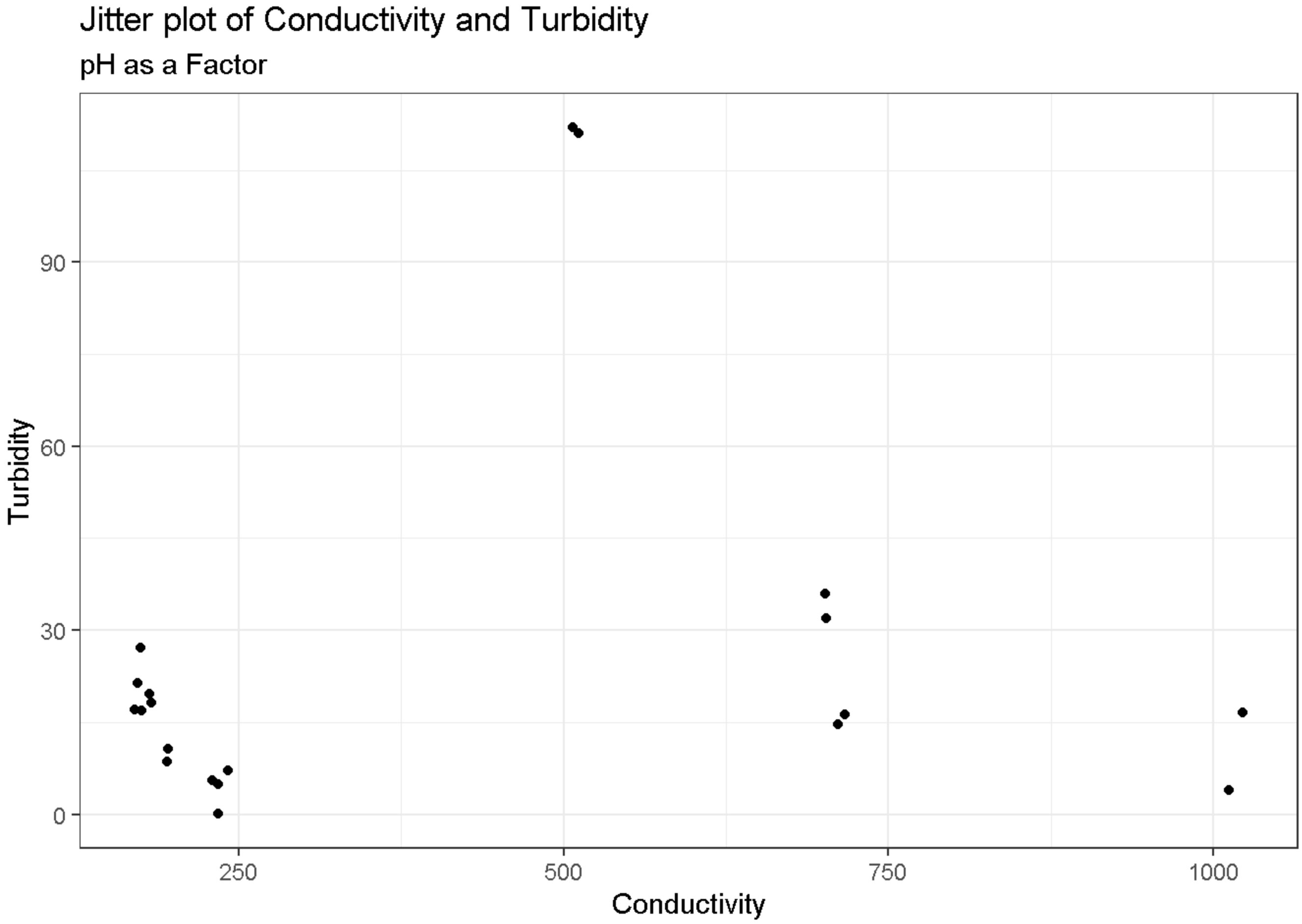
```
vp <- ggplot(data, aes(Conduct, Turb))+ geom_violin()
vp <- vp+labs(title = 'Violin plot of ocnductivity vs Turbidity', x = 'Conductivity', y = 'Turbidi
ty')
vp
```



It seems that conductivity likes more extreme turbidity which is strange. I at first believed it wouldo have a relatively positive correlation. To double check this I will do a scatter plot.

My fifth plot is a jitter plot of Turboidity and Conductivity

```
jp2 <- ggplot(data, aes(Conduct, Turb)) + geom_jitter(width = .5) + labs(title = 'Jitter plot of C
onductivity and Turbidity', subtitle = 'pH as a Factor', x ='Conductivity', y = 'Turbidity')
jp2
```

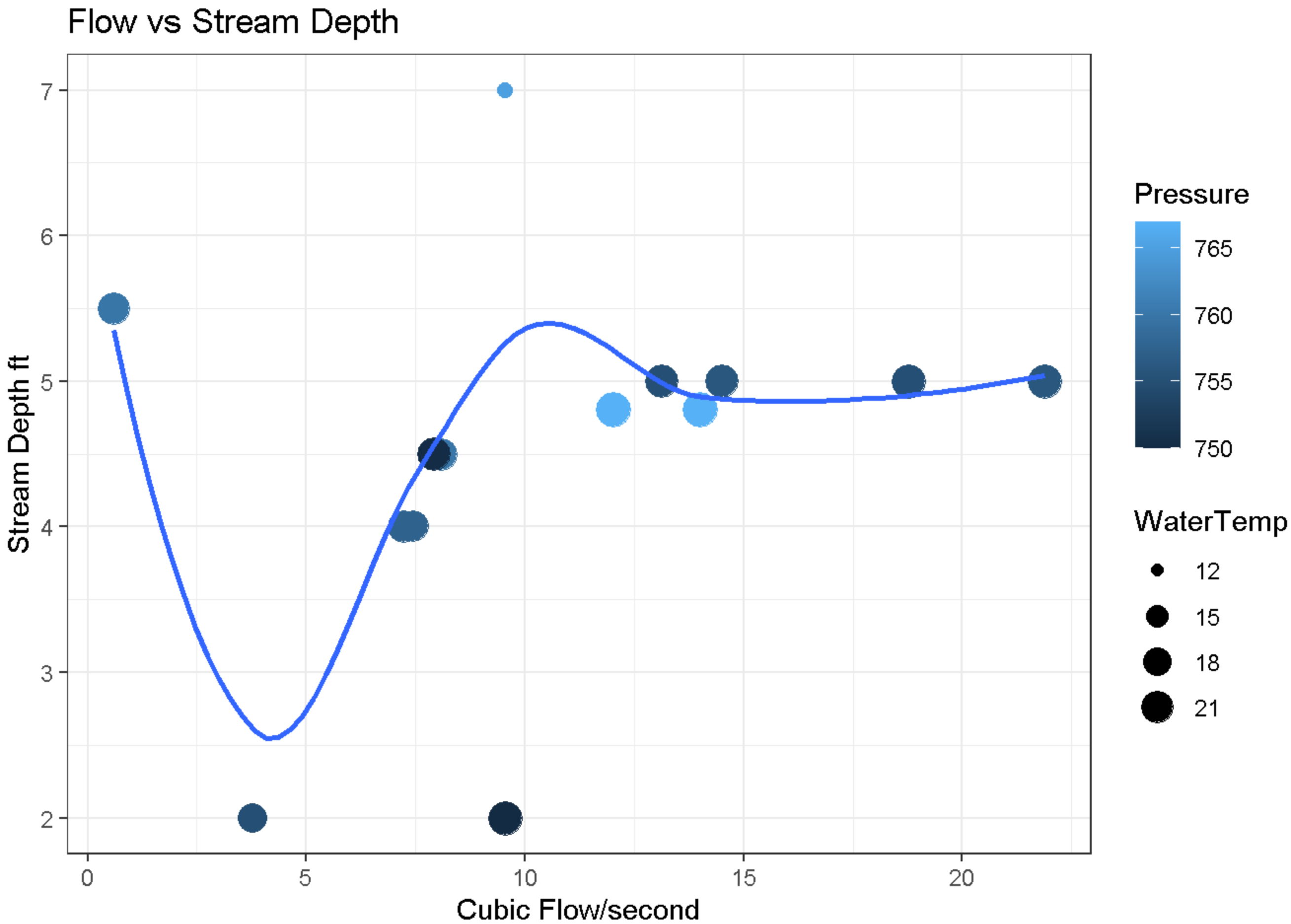


With the outlier it seems that there isn't a high correlation. However, we can see lower conductivity happens with lower turbidity which confirms my belief from before.

Then I did a cubic flow vs stream depth plot to see correlations between that.

```
CFSplot <- ggplot(data, aes(x=CFS, y=StrDepth))+ geom_point(aes(col=Pressure, size=WaterTemp)) + g
eom_smooth(method='loess',se=F)+
  labs(title="Flow vs Stream Depth", y='Stream Depth ft', x = 'Cubic Flow/second')
CFSplot
```

```
## `geom_smooth()` using formula 'y ~ x'
```

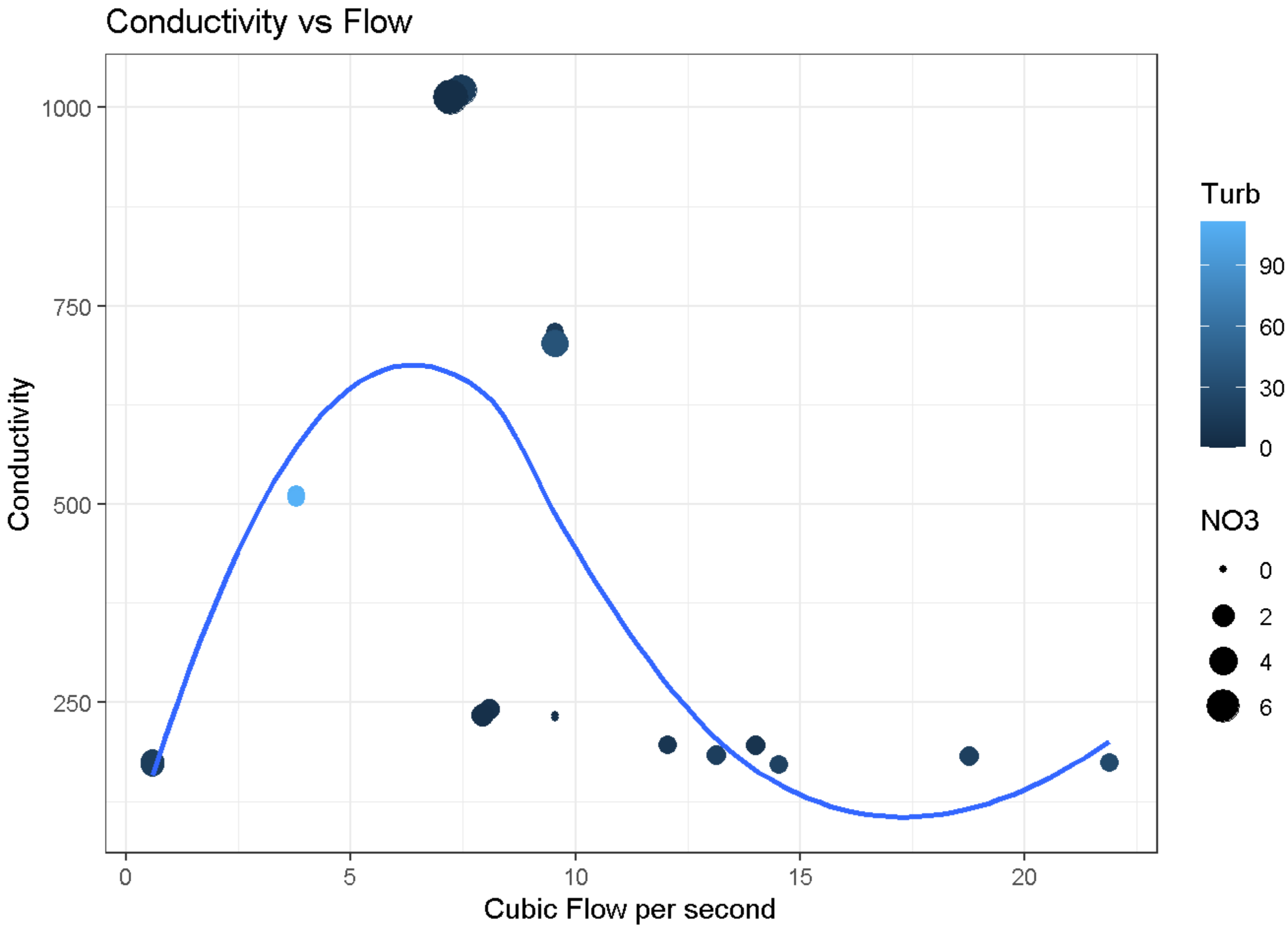


In this plot We can see the relationship of Cubic flow of water and stream depth. For the most part the depth of the river is constant. We can also see again that the depest part of the creek, is the coldest. The pressure of the water also seems to increase as the stream becomes shallower. The pressure of the water doesn't seem to have much affect.

A Flow vs Conductivity plot follows:

```
FlowconPlot <- ggplot(data, aes(x=CFS, y=Conduct))+ geom_point(aes(col=Turb, size=NO3)) + geom_smo
oth(method='loess',se=F)+
  labs(title="Conductivity vs Flow", y='Conductivity', x = 'Cubic Flow per second')
FlowconPlot
```

```
## `geom_smooth()` using formula 'y ~ x'
```

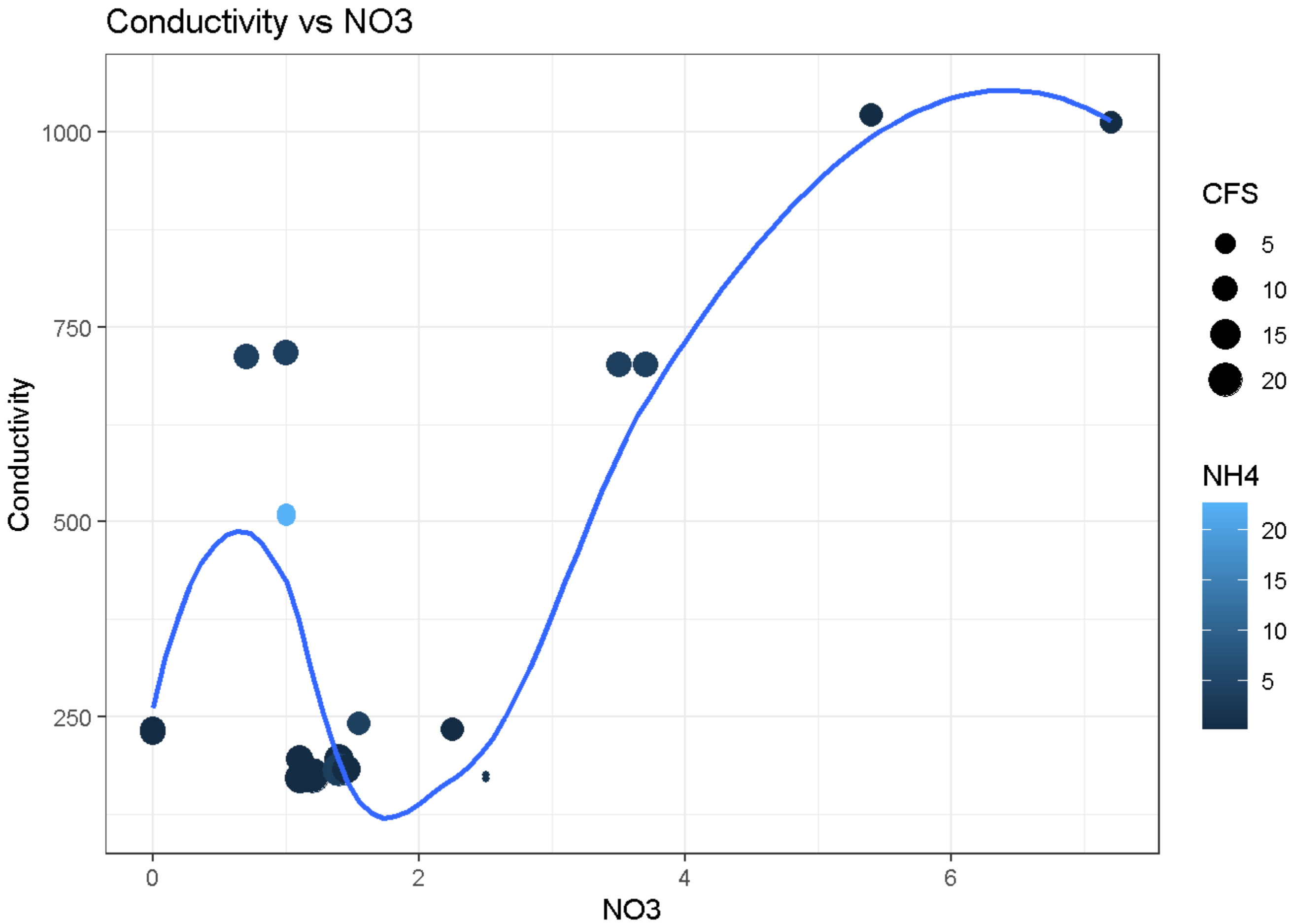


In this plot I notice that the NO3 level is larger with Higher conductivity, That is what I'll look at next. The sweet spot for conductivity is around 5-10 cubic flow per second.

In my final plot I will look at NO3 and conductivity levels

```
NOCon <- ggplot(data, aes(x=NO3, y=Conduct))+ geom_point(aes(col=NH4, size=CFS)) + geom_smooth(met
hod='loess',se=F)+
  labs(title="Conductivity vs NO3", y='Conductivity', x = 'NO3')
NOCon
```

```
## `geom_smooth()` using formula 'y ~ x'
```



I can see a positive correlation between conductivity and NO3. These factors do not seem to affect NH4. However, it seems that the flow is heavier with a lower conductivity and higher NO4 levels.

Closing Statements

Overall, the conditions at Ellerbee creek are relatively constant. The bad thing is, some of the things that are constant are nopt good. Dor instance the average dissolved oxygen levels in the water are in the 6.5-9.5 range, which allow for some big fish but not many small fish. And over 50% of the data collected had dissolved oxygen levels below 9.5, and even reaching below 6. in water, lower temperatures are supposed to have higher dissolved oxygen counts. The water in Ellerbee creek just according to dissolved oxygen count is okay. While it is at a safe level for fish a lot of the time, it at many times trails down into unsafe levels. As for NH4, the safest amount for marine life is between 0.02 and 0.4 ppm. The NH4 levels show the over 50% of the data has extremely unsafe NH4 levels. Those factors alone show how polluted Ellerbee creek is. Turbidity, which is measured in Nephelometric Turbidity Units, in Ellerbee creek has a wide range of values. One thing is for sure, the turbidity levels of Ellerbee creek are to inconsistent for drinking water and are more close to the normal levels for fish. Because the NH4 levels and Dissolved Oxygen levels, I assume that there is intense runoff from the nearby roads, and human infrastructure. Efforts need to take place to make sure that aquatic life can thrive in Ellerbee creek.

Sources

“Dissolved Oxygen (DO).” Government of Northwest Territories, www.enr.gov.nt.ca/sites/enr/files/dissolved_oxygen.pdf.

Wagenet, Robert J. “Fact Sheets :: Nitrate: Health Effects in Drinking Water.” PSEP, Margaret McCasland, Nancy M. Trautmann, and Keith S. Porter Center for Environmental Research, 2020, psep.cce.cornell.edu/facts-slides-self/facts/nit-heef-grw85.aspx.