Report 1
Descriptive Statistics
Hunter Stevens

**Stats on Variables**

|	Age	FEV	Hgt	Sex	Smoke
count	654.00000	654.0000	654.00000	654.0000	654.0000
mean	9.9311930	2.636780	61.143578	0.513761	0.099388
std	2.9539350	0.867059	5.703513	0.500193	0.299412
min	3.0000000	0.791000	46.000000	0.000000	0.000000
25%	8.0000000	1.981000	57.000000	0.000000	0.000000
50%	10.000000	2.547500	61.500000	1.000000	0.000000
75%	12.000000	3.118500	65.500000	1.000000	0.000000
max	19.000000	5.793000	74.000000	1.000000	1.000000

Histograms of the variables
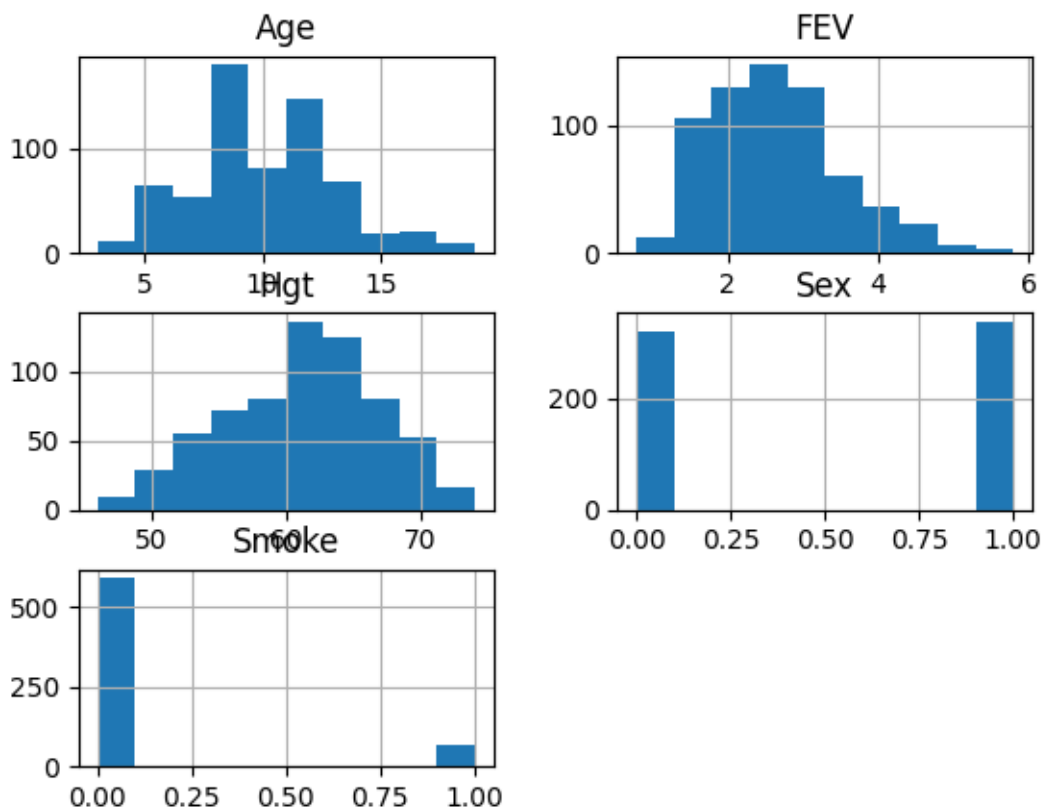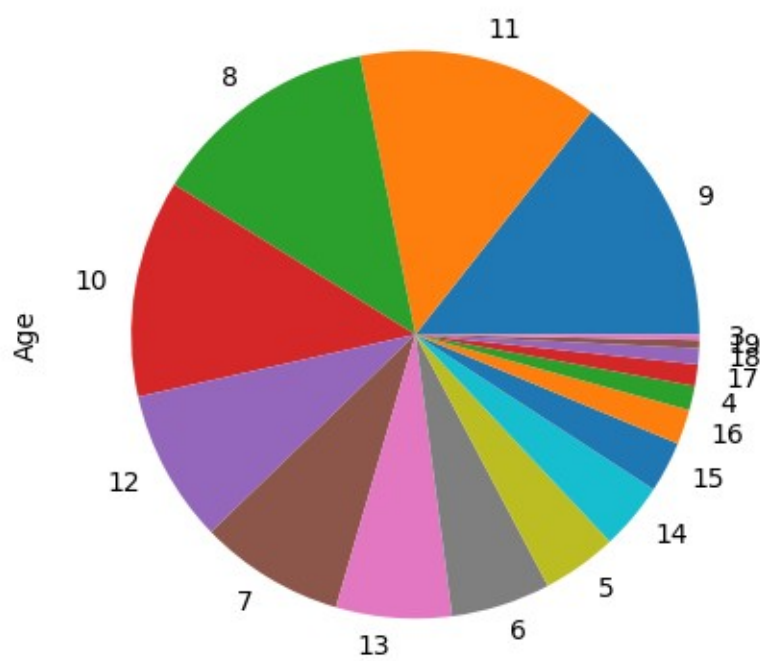
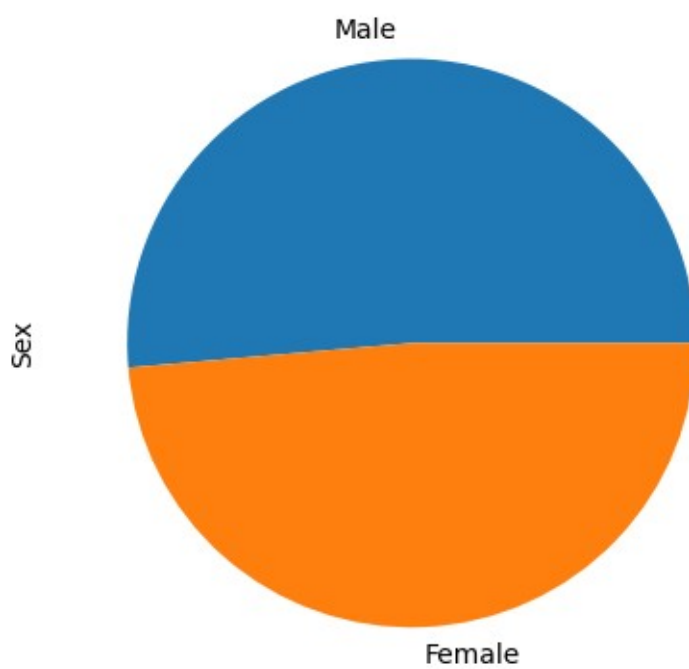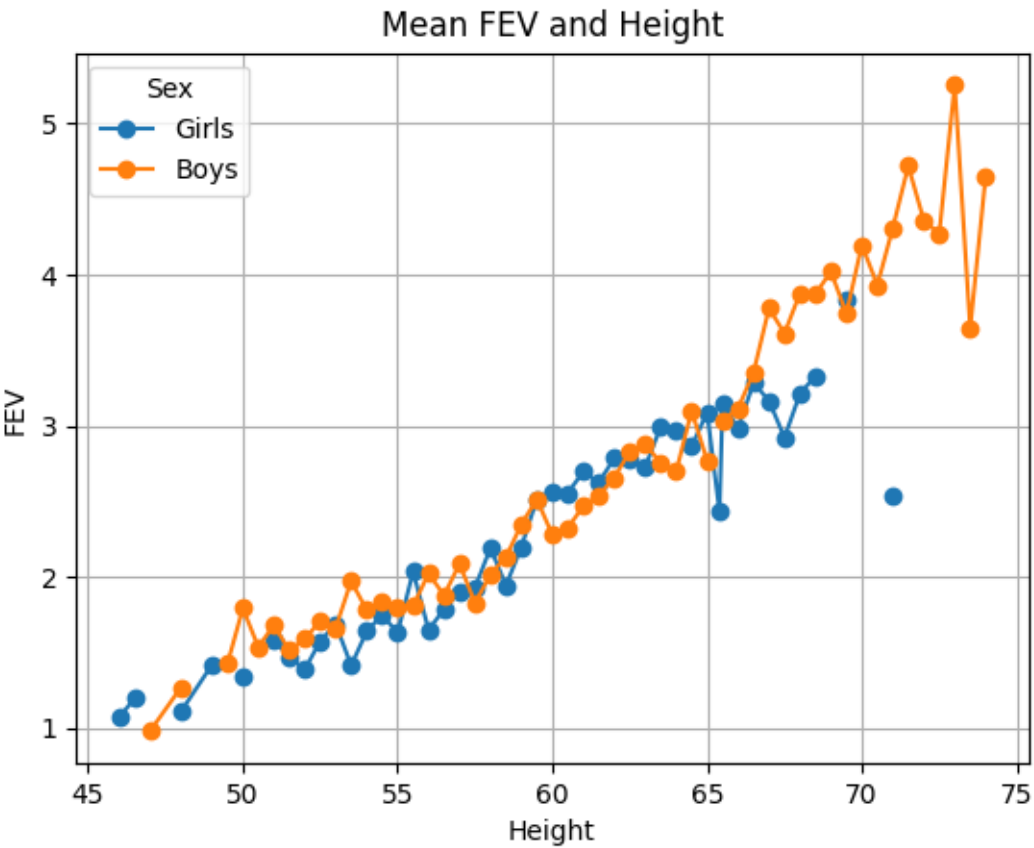Boxplot of Age, boys and girls
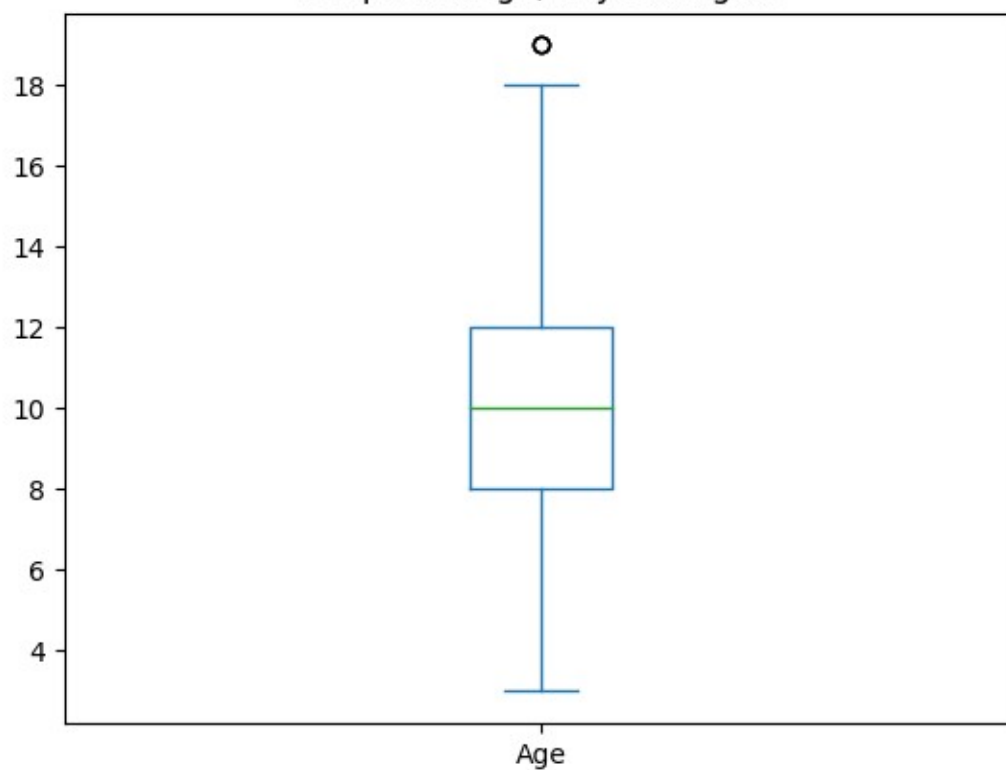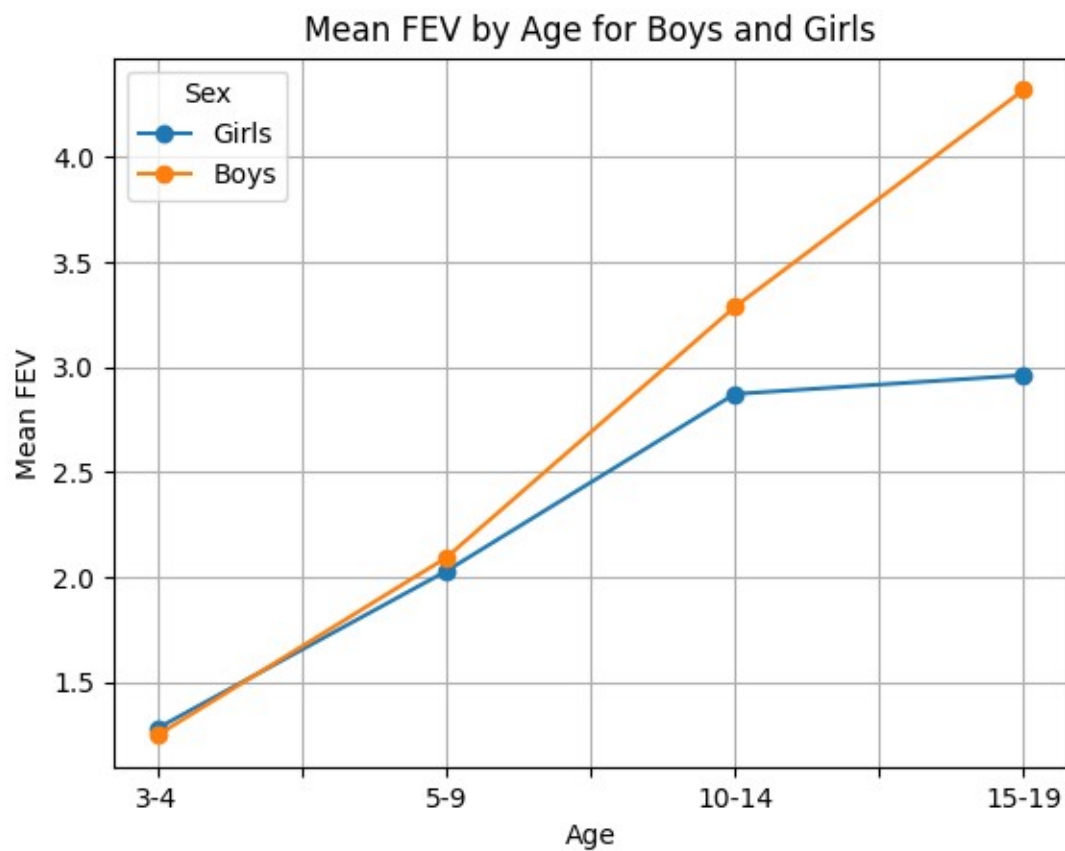

Heights of boys and girls

FEV of boys and girls
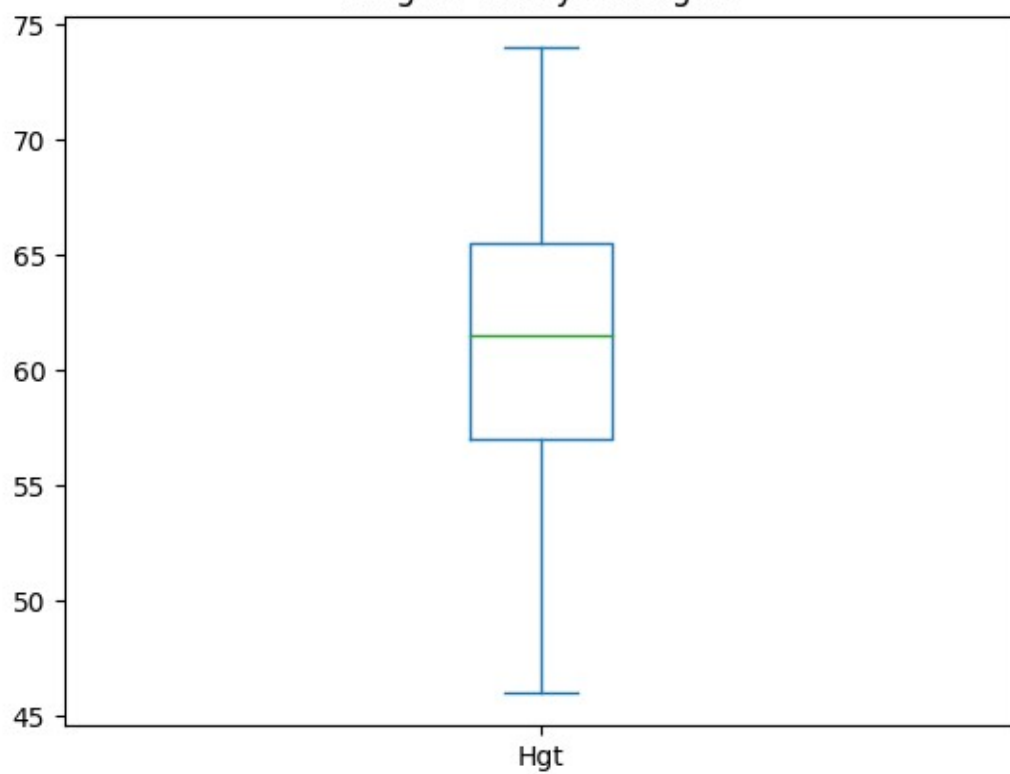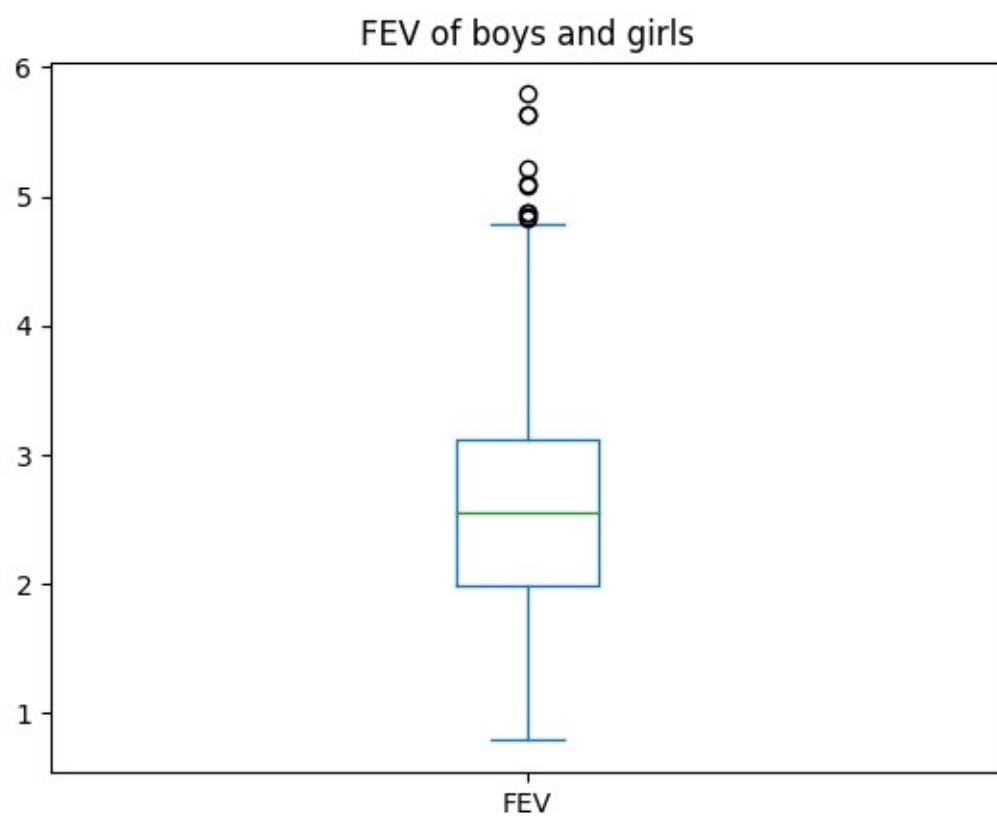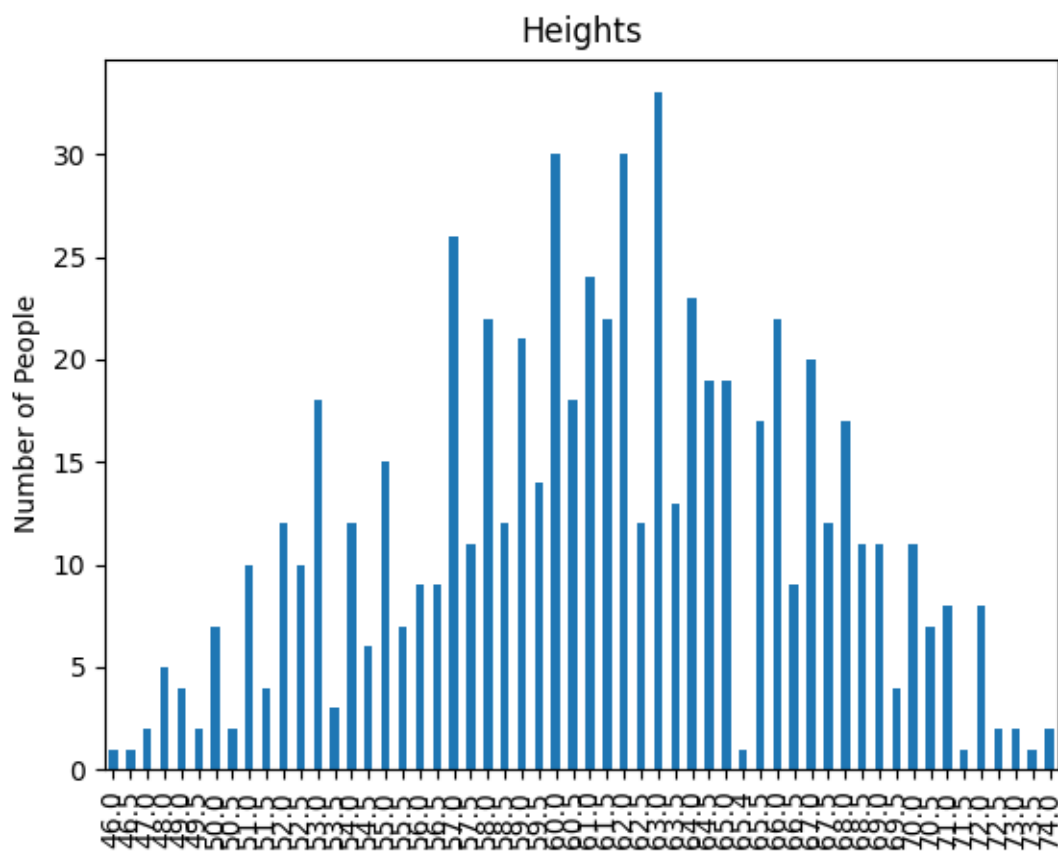
Heights

Ages of boys and girls

Number of Males and Females

FEV and Height:

Mean FEV and Height

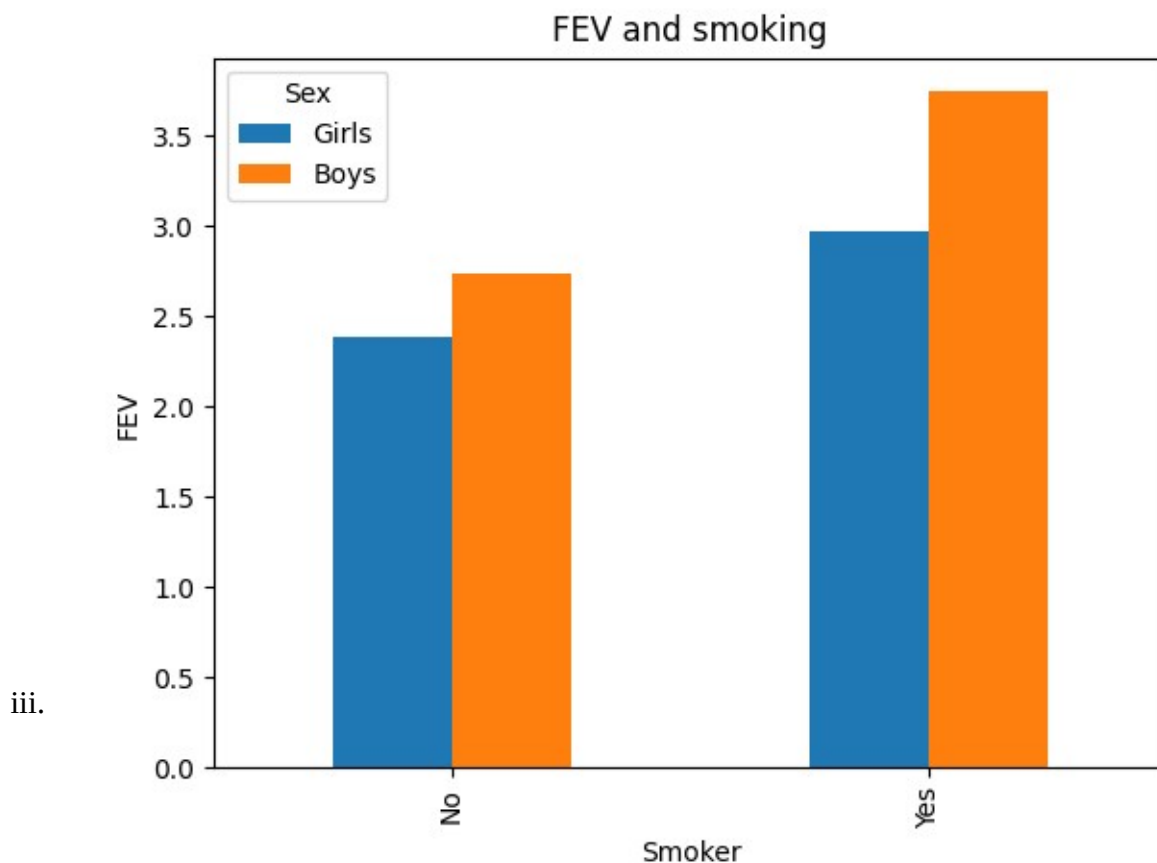**FEV and Age:**

Mean FEV by Age for Boys and Girls

FEV and Smoking

FEV and smoking

iii.

According to the charts above FEV seems to be directly correlated with age, height and smoking status. As you can see in 'FEV and Smoking' among boys and girls FEV is higher among smokers with a bigger jump between boys and girls who smoke vs boys and girls who don't.

Height also seems to play a role in higher FEV levels. In the 'Mean FEV and Height' chart we can see a steady increase in FEV as height increases.

In 'Mean FEV by age for boys and girls' FEV increases with each age group, with the biggest jump being between the 5-9 age group and the 10-14 age group. However, this could also be an indicator of smoking affecting FEV considering the average age of smokers among the dataset is 13. See below.

Stats on smokers:

	Age	**FEV**	**Hgt**	**Sex**
count	65.000000	65.000000	65.000000	65.00000
mean	**13.523077**	3.276862	65.953846	0.40000
std	2.339255	0.749986	3.192671	0.49371
min	9.000000	1.694000	58.000000	0.00000
25%	12.000000	2.795000	63.500000	0.00000
50%	**13.000000**	3.169000	66.000000	0.00000
75%	15.000000	3.751000	68.000000	1.00000
max	19.000000	4.872000	72.000000	1.00000

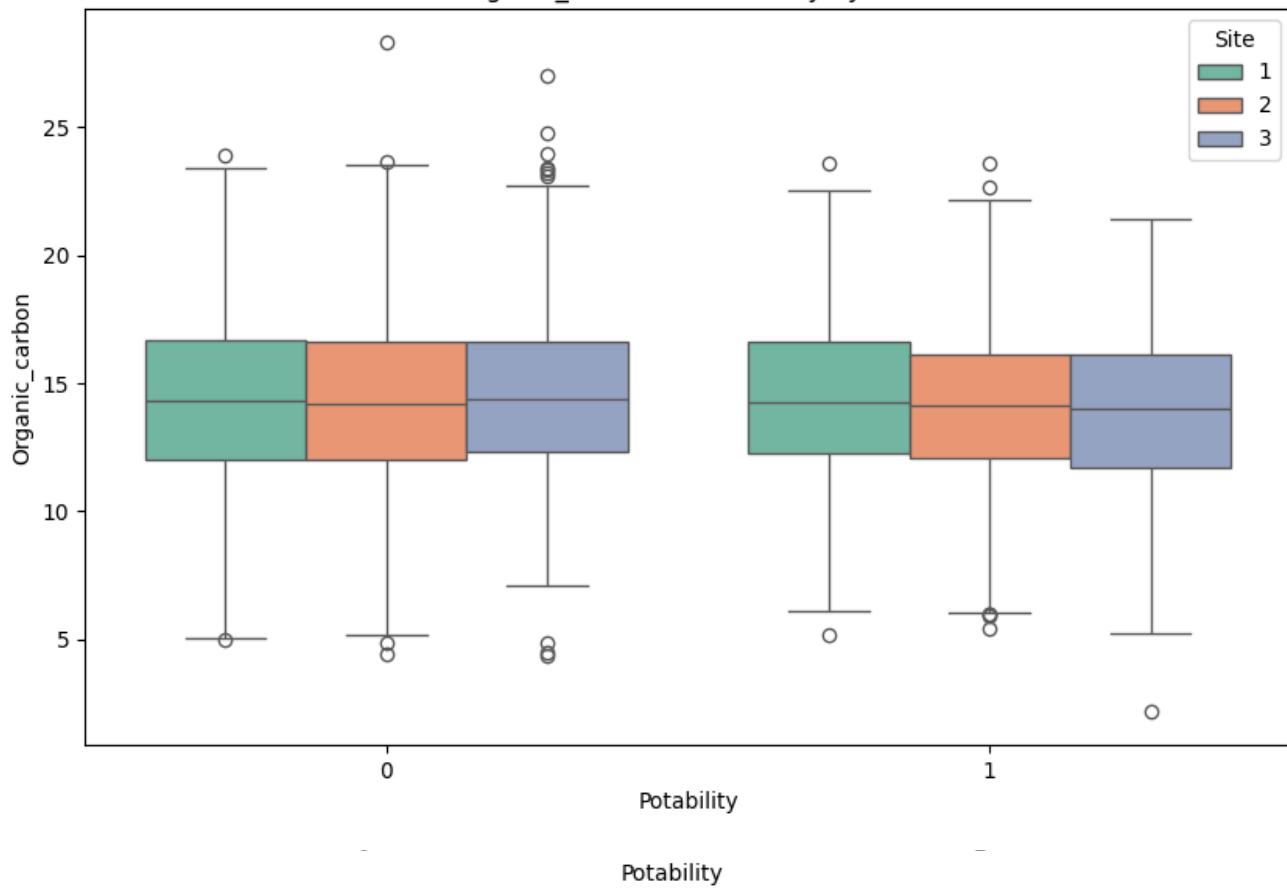# Part 2: Water Potability
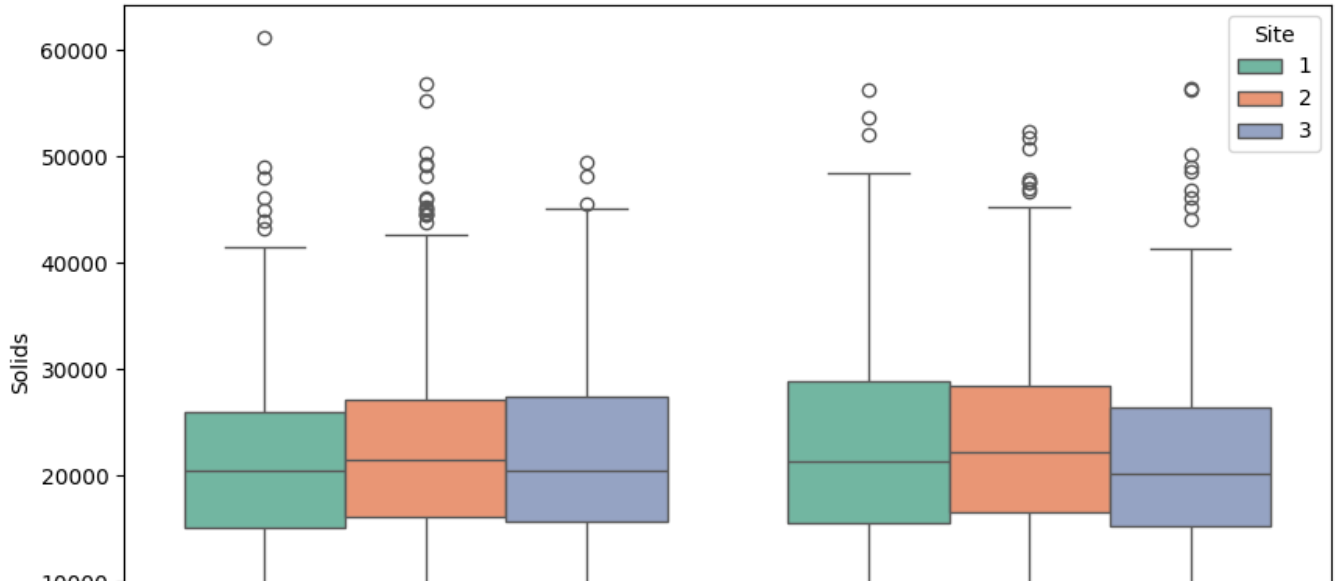
ph vs Potability by Site
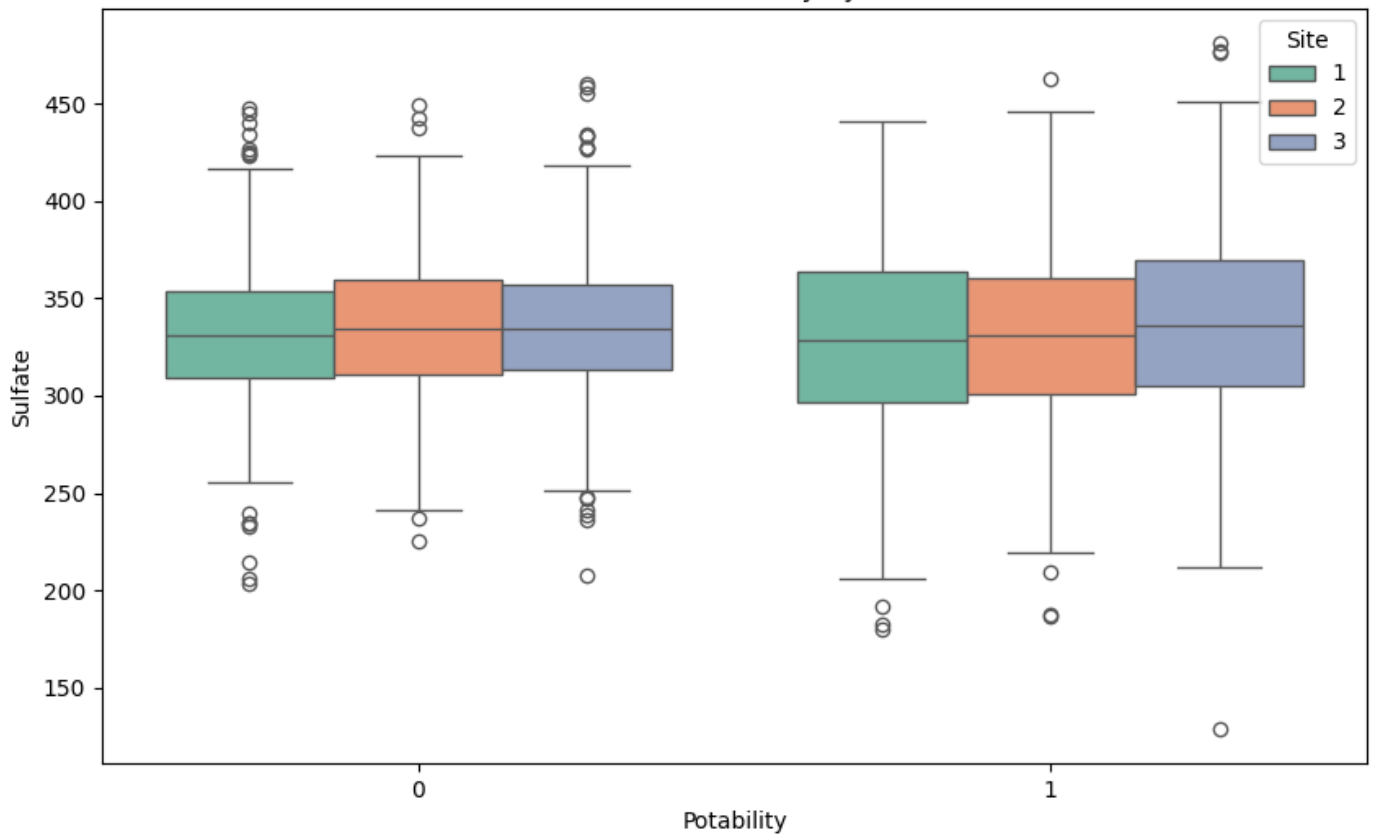
Conductivity vs Potability by Site
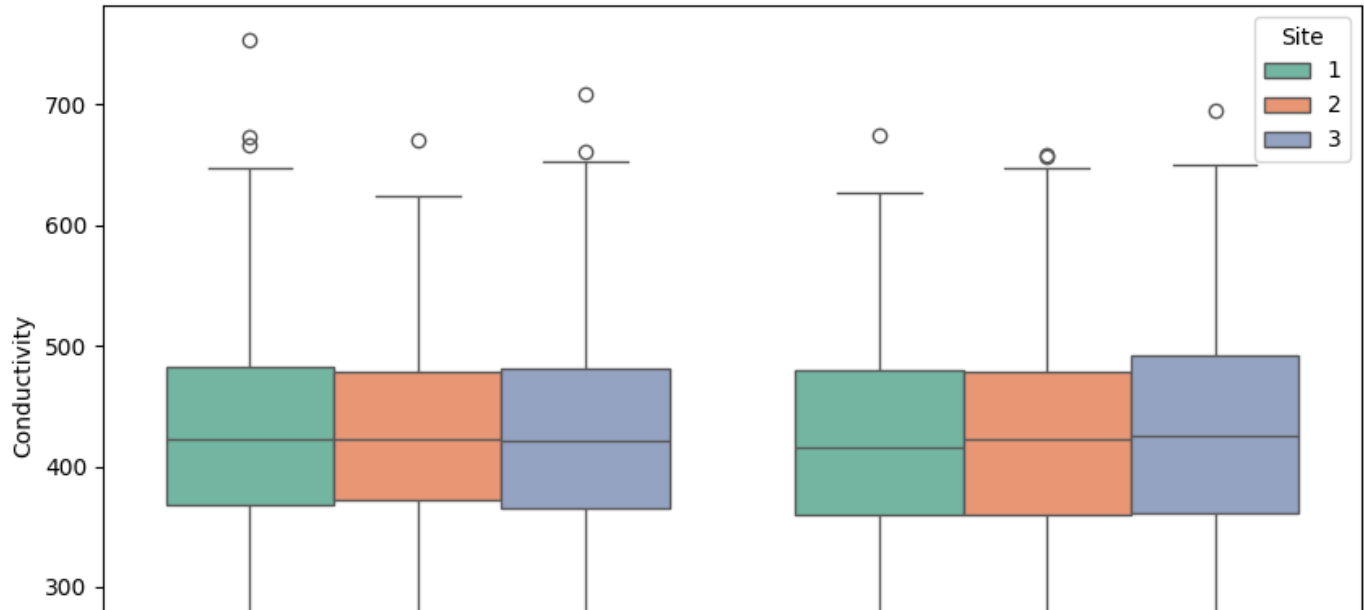
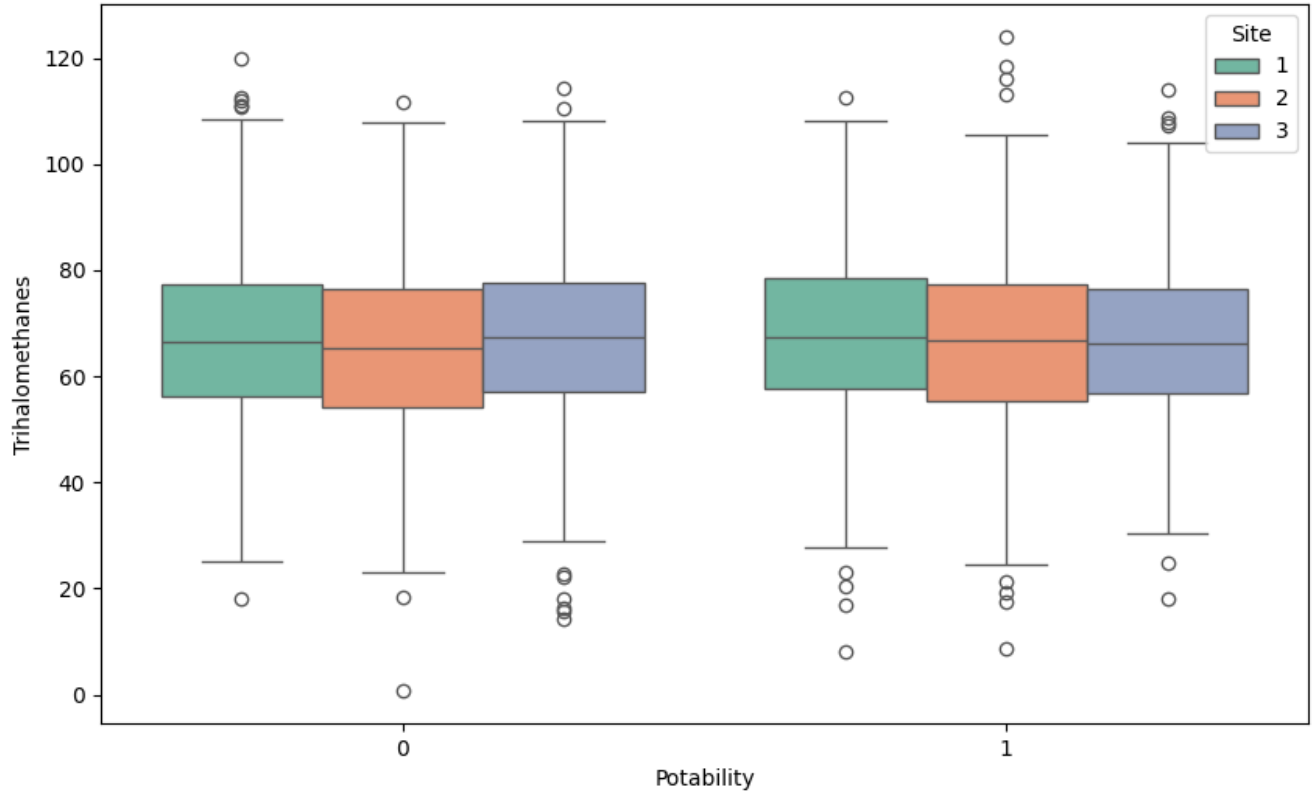Organic_carbon vs Potability by Site

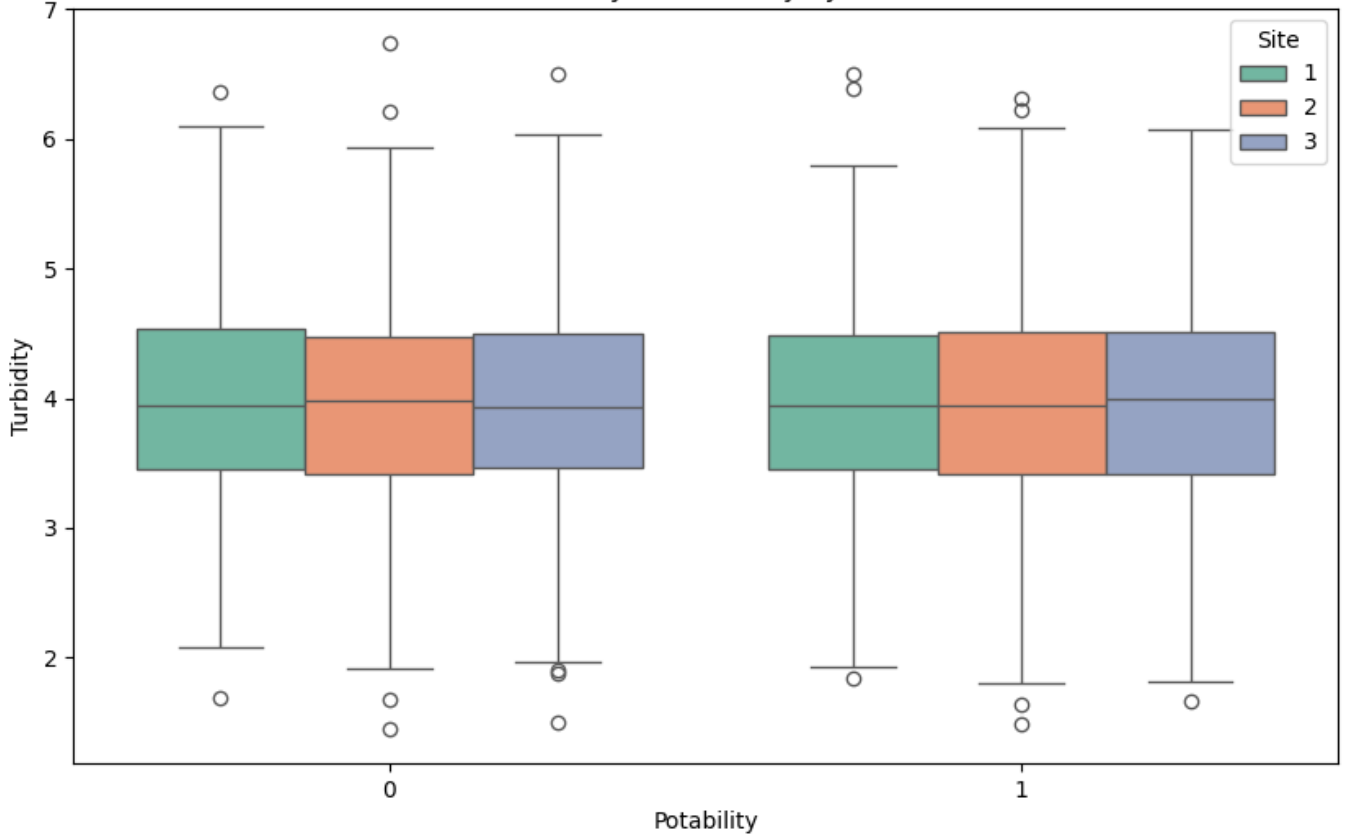Solids vs Potability by Site

Sulfate vs Potability by Site

Trihalomethanes vs Potability by Site


Turbidity vs Potability by Site

Water potability:

Descriptive statistics of entire data set

|	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability	site
count	2785	3276	3276	3276	2495	3276	3276	3114	3276	3276	3276
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.28497	66.396293	3.966786	0.39011	1.999389
std	1.59432	32.879761	8768.570828	1.583085	41.41684	80.824064	3.308162	16.175008	0.780382	0.487849	0.825546
min	0	47.432	320.942611	0.352	129	181.483754	2.2	0.738	1.45	0	1
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0	1
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0	2
75%	8.062066	216.667456	27332.762127	8.114887	359.95017	481.792304	16.557652	77.337473	4.50032	1	3
max	14	323.124	61227.196008	13.127	481.030642	753.34262	28.3	124	6.739	1	3

PH levels

PH
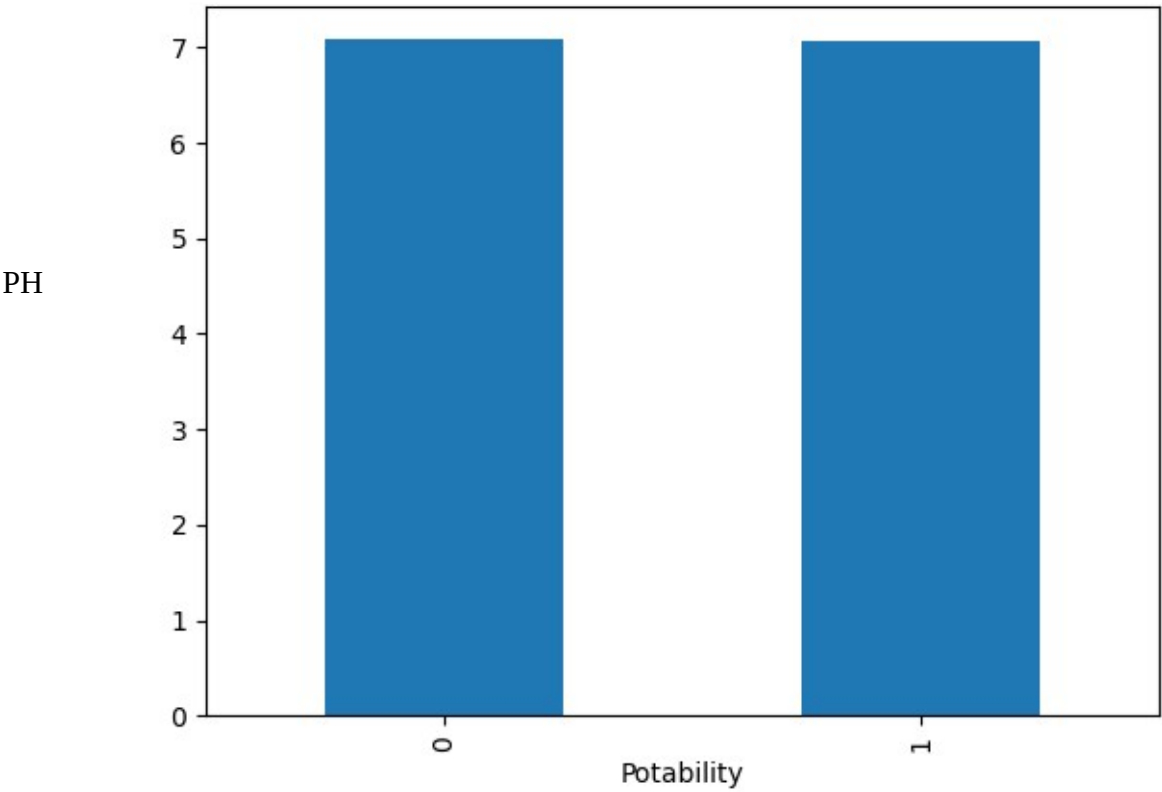
levels were on average identical among potable and non-potable water, however potable water was inside a tighter range. Potable water was not completely base or acidic in this study, while non-potable water had ph levels of 0 and 14.

Potable water
```
            ph
mean     7.073783
min      0.227499
25%      6.179312
50%      7.036752
75%      7.933068
max     13.175402
```

Non-potable water
```
            ph
mean     7.085378
std      1.683499
min      0.000000
25%      6.037723
50%      7.035456
75%      8.155510
max     14.000000
```

As you can see while the average is nearly identical, non-potable water had more entries at the extreme ends of 14 and 0. You can also see in the 3 site ph boxplot image above that the potable water in each site was grouped more tightly together around the mean, which is about 7.

Hardness

Hardness of Potable vs Non-Potable Water
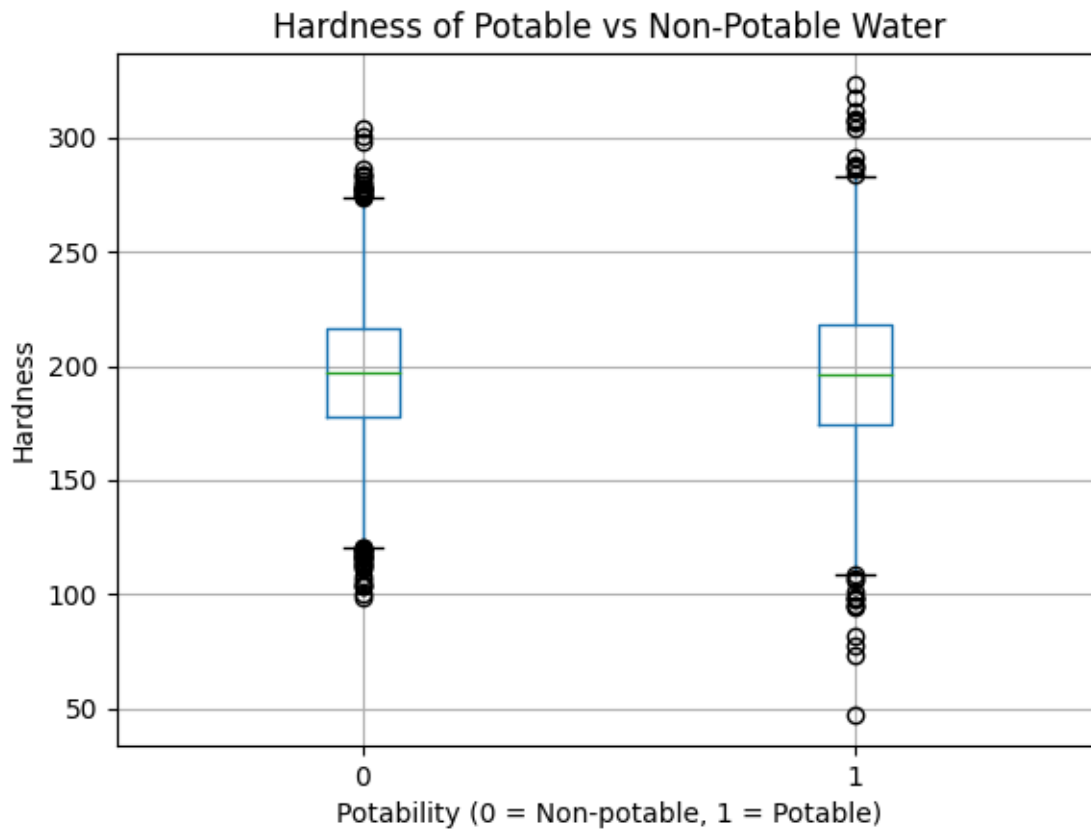
Potable water had a bigger range of hardness levels as you can see in the chart above. The average hardness level for potable and non-potable water is nearly identical. However looking at the above chart and the 3 site comparison chart on hardness levels above we can see that hardness levels had a greater range among potable water than they did among non-potable water.
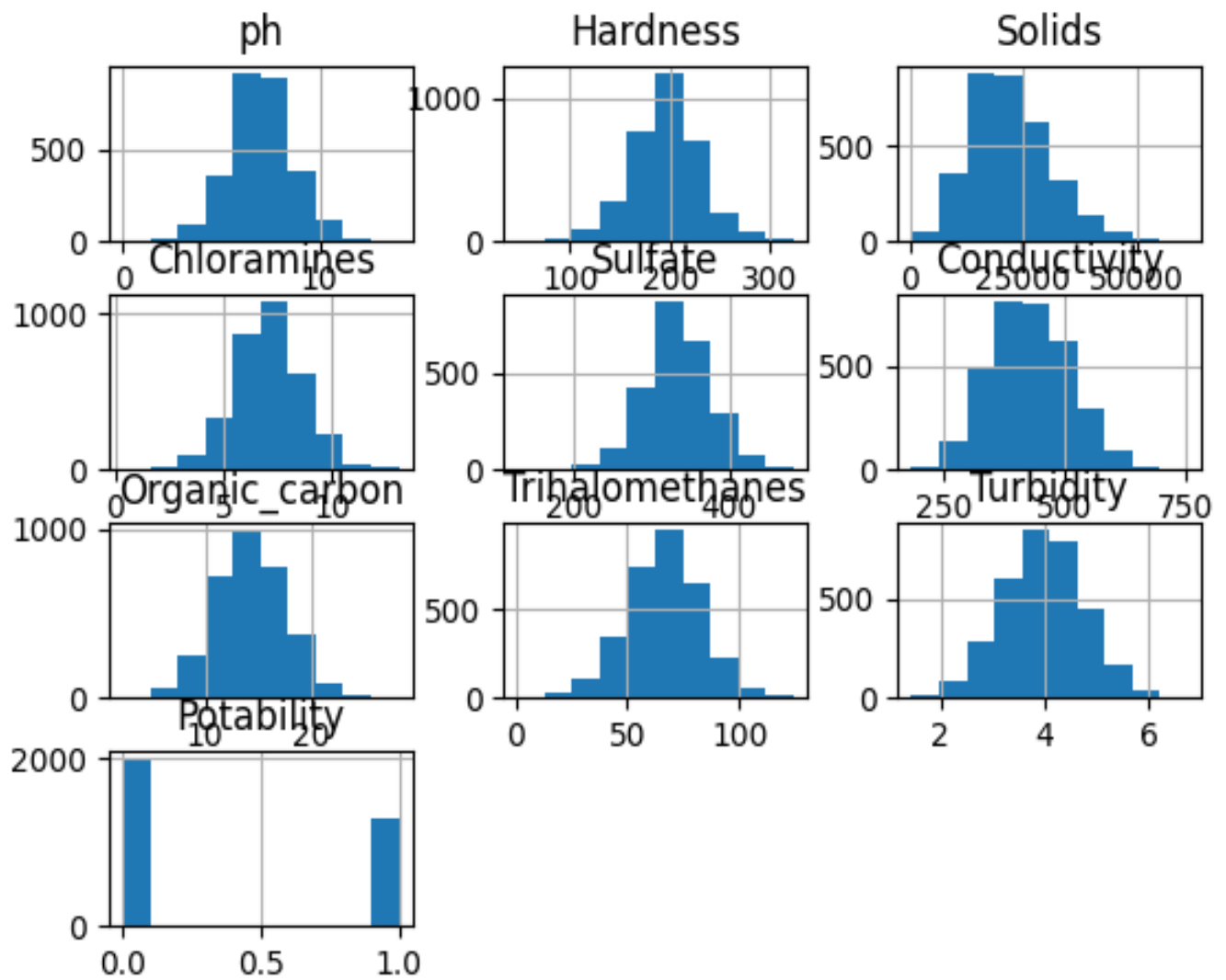
In conclusion, potability has been quite hard to determine based upon these features in the data set. I couldn't seem to see the connections between these variables. The patterns between ph and hardness seemed to be the most obvious connections to potability to me. Everything else seems to be very similar between potable and non-potable water.

Organic Carbon Levels by pH for Potable and Non-Potable Water

Hardness Levels by ph for Potable and Non-Potable Water