PHP 2550 HW #4

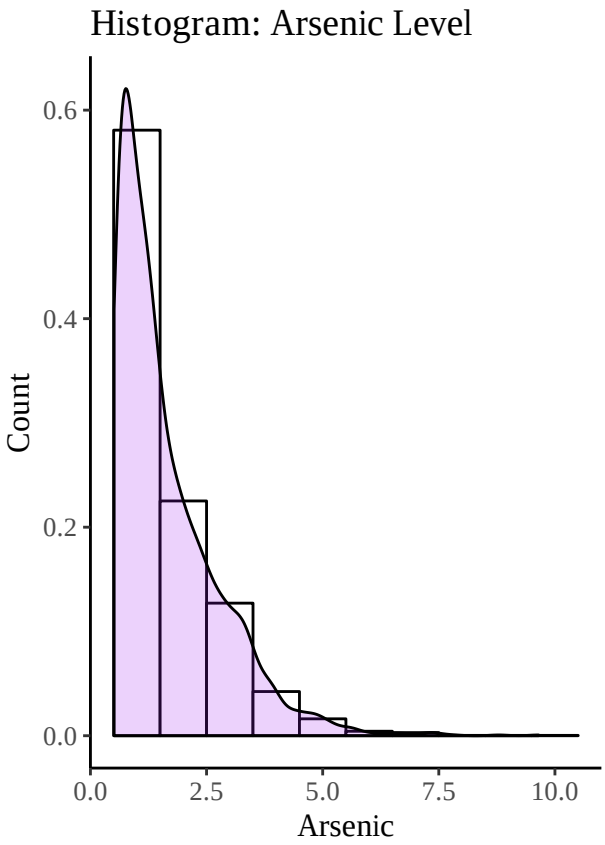
Blain Morin

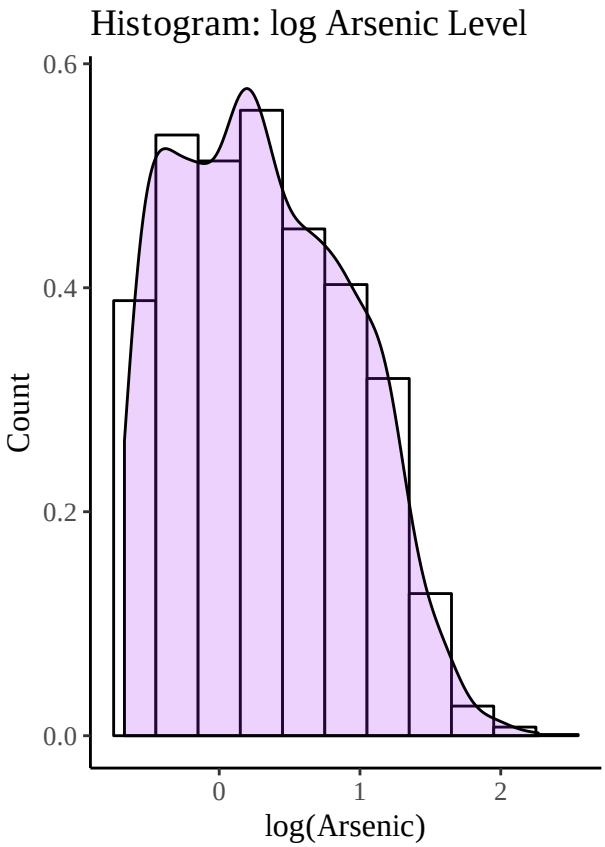
November 9, 2018

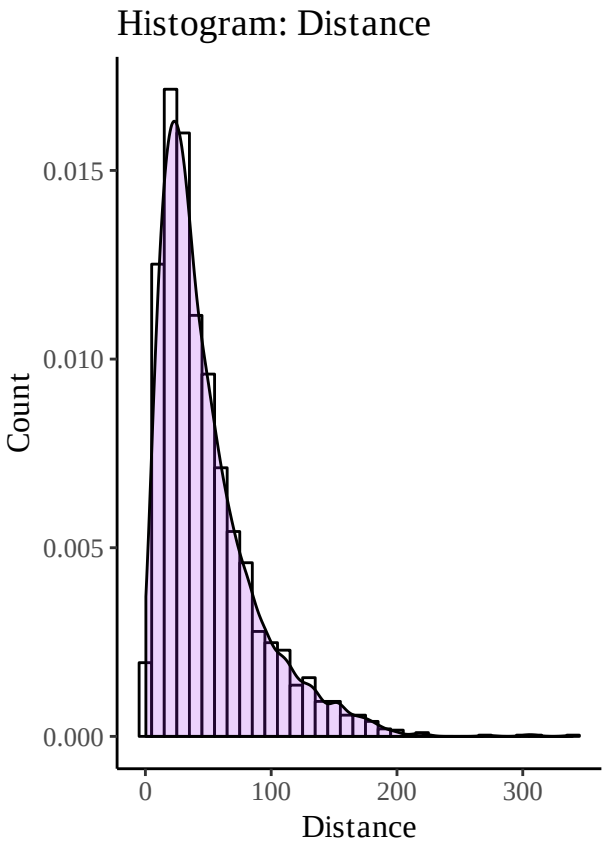
1a. Construct a good logistic regression model predicting the decision to switch wells as a function of the 4 predictors (arsenic, distance, association and education) on the training data. Consider potential transformations of continuous variables and possible interactions.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Switched Wells	3,020	0.575	0.494	0	0	1	1
Arsenic Level	3,020	1.657	1.107	0.510	0.820	2.200	9.650
Distance	3,020	48.332	38.479	0.387	21.117	64.041	339.531
Active in Community = 1	3,020	0.423	0.494	0	0	1	1
Years of Education	3,020	4.828	4.017	0	0	8	17


Histogram: Arsenic Level
Count
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
Arsenic


Histogram: log Arsenic Level
Count
0.6
0.4
0.2
0.0
0
1
2
log(Arsenic)


Histogram: Distance
Count
0.015
0.010
0.005
0.000
0
100
200
300
Distance

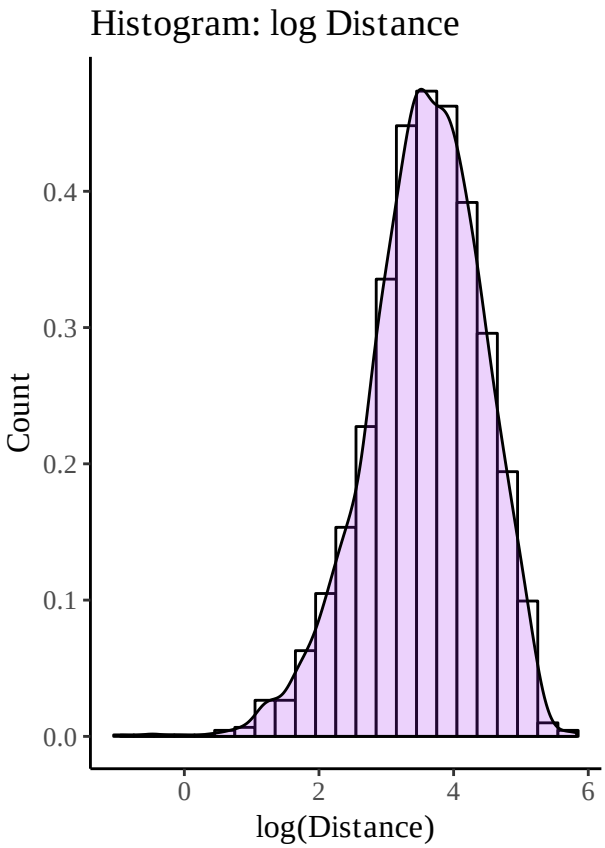

Histogram: log Distance
Count
0.4
0.3
0.2
0.1
0.0
0
2
4
6
log(Distance)

Table 2: Regression Results

	Dependent variable:			
	log odds Switched Well			
	(1)	(2)	(3)	(4)
Arsenic	0.467***		0.463***	0.348***
	(0.042)		(0.041)	(0.059)
Distance	−0.009***		−0.009***	−0.009***
	(0.001)		(0.001)	(0.001)
log(Arsenic)		0.844***		
		(0.068)		
log(Distance)		−0.348***		
		(0.048)		
Active in Community	−0.124	−0.128*	−0.124	−0.123
	(0.077)	(0.077)	(0.077)	(0.077)
Years Education	0.042***	0.041***		−0.0004
	(0.010)	(0.010)		(0.018)
Graduated Highschool = 1			0.608***	
			(0.156)	
Arsenic * Education				0.029***
				(0.011)
Constant	−0.157	1.159***	0.006	0.028
	(0.100)	(0.183)	(0.088)	(0.120)
Observations	3,020	3,020	3,020	3,020
Log Likelihood	−1,953.913	−1,955.532	−1,955.784	−1,950.173
Akaike Inf. Crit.	3,917.826	3,921.063	3,921.569	3,912.346
Note:				*p<0.1; **p<0.05; ***p<0.01