Final Project

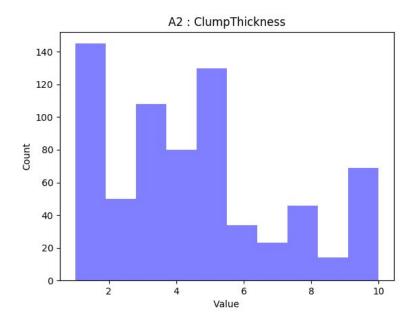
Phase 1 - Statistical Analysis	2
Summary Table	2
Histograms	
Phase 2 - k-means Clustering	7
Program Output	7

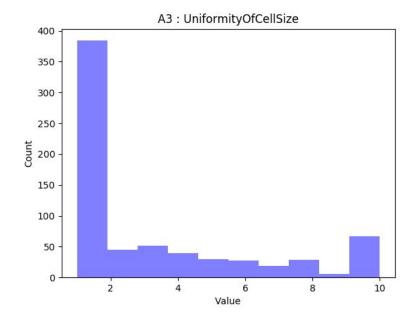
Phase 1 - Statistical Analysis

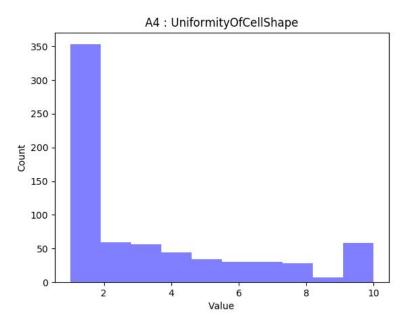
Summary Table

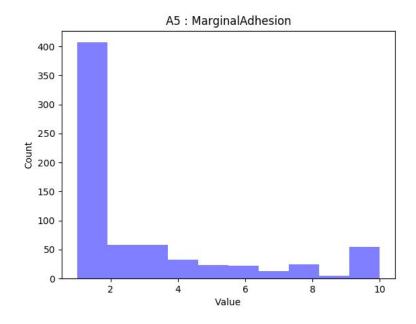
<u>Attı</u>	ribute	Min	Max	Mean	Median	StdDev	<u>Variance</u>
A2	ClumpThickness	1.000	10.000	4.418	4.000	2.816	7.928
А3	UniformityOfCellSize	1.000	10.000	3.134	1.000	3.051	9.311
Α4	UniformityOfCellShape	1.000	10.000	3.207	1.000	2.972	8.832
Α5	MarginalAdhesion	1.000	10.000	2.807	1.000	2.855	8.153
A6	SingleEpithelialCellSize	1.000	10.000	3.216	2.000	2.214	4.903
Α7	BareNuclei	1.000	10.000	3.486	1.000	3.622	13.118
A8	BlandChromatin	1.000	10.000	3.438	3.000	2.438	5.946
Α9	NormalNucleoli	1.000	10.000	2.867	1.000	3.054	9.325
A10	Mitoses	1.000	10.000	1.589	1.000	1.715	2.941

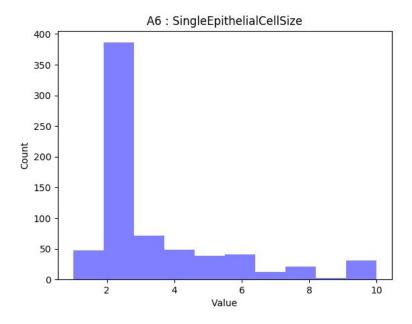
Histograms

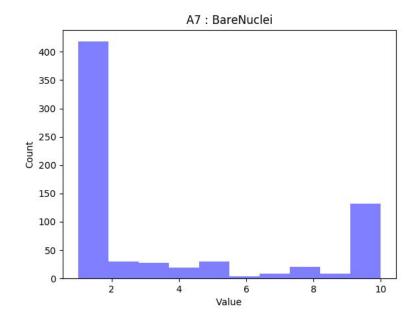


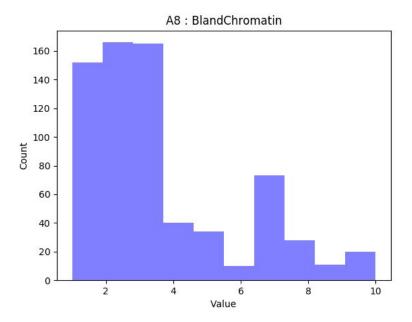


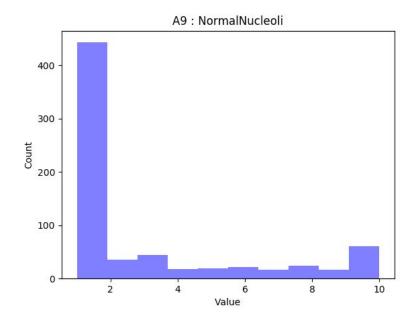


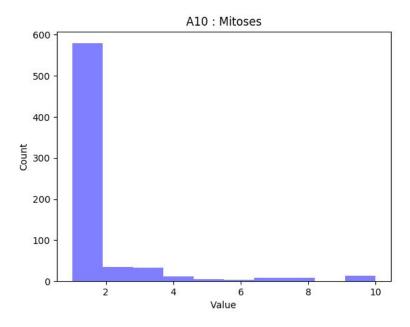












Phase 2 - k-means Clustering

Program Output

```
$ time ./main.py
Please enter file name [default: Breast-Cancer-Wisconsin.csv]:
Generating Histograms...
Statistic Summary Table
Attribute
                              Min
                                      Max
                                            Mean
                                                  Median
                                                          StdDev Variance
A2 ClumpThickness
                            1.000
                                   10.000
                                            4.418
                                                   4.000
                                                           2.816
                                                                  7.928
A3 UniformityOfCellSize
                            1.000
                                   10.000
                                           3.134
                                                   1.000
                                                           3.051
                                                                  9.311
A4 UniformityOfCellShape
                            1.000
                                   10.000
                                           3.207
                                                   1.000
                                                           2.972
                                                                  8.832
                                                           2.855
A5 MarginalAdhesion
                            1.000
                                   10.000
                                           2.807
                                                   1.000
                                                                  8.153
A6 SingleEpithelialCellSize
                            1.000
                                   10.000
                                            3.216
                                                   2.000
                                                           2.214
                                                                  4.903
A7 BareNuclei
                            1.000
                                   10.000
                                           3.486
                                                   1.000
                                                           3.622
                                                                 13.118
A8 BlandChromatin
                                   10.000
                                            3.438
                                                           2.438
                                                                  5.946
                            1.000
                                                   3.000
A9 NormalNucleoli
                            1.000
                                   10.000
                                                           3.054
                                                                  9.325
                                            2.867
                                                   1.000
A10 Mitoses
                            1.000
                                   10.000
                                           1.589
                                                   1.000
                                                           1.715
                                                                  2.941
Running k-means Classification with 1500 iterations (please be patient)
Initial Means (all columns)
u2: Scn=673637.00000 A2=3.00000 A3=1.00000 A4=1.00000 A5=1.00000 A6=2.00000 A7=5.00000 A8=5.00000
A9=1.00000 A10=1.00000 CLASS=2.00000
u4: Scn=1115293.00000 A2=1.00000 A3=1.00000 A4=1.00000 A5=1.00000 A6=2.00000 A7=1.00000 A8=1.00000
A9=1.00000 A10=1.00000 CLASS=2.00000
......
......
```

Final Means

u2: A2=7.14957 A3=6.77778 A4=6.71368 A5=5.72650 A6=5.45726 A7=7.83761 A8=6.08974 A9=6.07692 A10=2.54274

......

......

u4: A2=3.04301 A3=1.30108 A4=1.44301 A5=1.33763 A6=2.08817 A7=1.29677 A8=2.10323 A9=1.25161 A10=1.10968

uster Assignment: u2-Benign						
N I	Index	I	d	Class	Predicte	d
0	1	100294	5	2		2
1	3	101627	7	2		2
2	5	101712	2	4		2
3	14	104457	2	4		2
4	18	105067	0	4		2
5	20	105459	0	4		2
6	21	105459	3	4		2
7	25	106572	6	4		2
8	32	107217	9	4		2
9	36	108018	5	4		2
10	38	108458	4	4		2
11	39	109126	2	4		2
12	40	109680	0	2		2
13	41	109951	0	4		2
14	42	110052	4	4		2
15	43	110257	3	4		2
16	44	110360	8	4		2
17	46	110525	7	4		2
18	49	110682	9	4		2
19	52	111010	2	4		2

Cluster Assignment: u4-Malignant

	Harremanic	Simeries ar	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	ascc.
Predicted	Class	Id	Index	N
4	2	1000025	0	0
4	2	1015425	2	1
4	2	1017023	4	2
4	2	1018099	6	3
4	2	1018561	7	4
4	2	1033078	8	5
4	2	1033078	9	6
4	2	1035283	10	7
4	2	1036172	11	8
4	4	1041801	12	9
4	2	1043999	13	10
4	4	1047630	15	11
4	2	1048672	16	12
4	2	1049815	17	13
4	2	1050718	19	14
4	2	1056784	22	15
4	4	1057013	23	16
4	2	1059552	24	17
4	2	1066373	26	18
4	2	1066979	27	19

real 43m27.623s user 43m21.896s sys 0m3.906s