

# Class 9: Unsupervised Learning Analysis of Human Breast Cancer Cells

Clarissa Savko (PID:A69028482)

```
wisc.df <- read.csv("WisconsinCancer.csv", row.names= 1)
head(wisc.df)
```

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean
842302	M	17.99	10.38	122.80	1001.0
842517	M	20.57	17.77	132.90	1326.0
84300903	M	19.69	21.25	130.00	1203.0
84348301	M	11.42	20.38	77.58	386.1
84358402	M	20.29	14.34	135.10	1297.0
843786	M	12.45	15.70	82.57	477.1

  

	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean
842302	0.11840	0.27760	0.3001	0.14710
842517	0.08474	0.07864	0.0869	0.07017
84300903	0.10960	0.15990	0.1974	0.12790
84348301	0.14250	0.28390	0.2414	0.10520
84358402	0.10030	0.13280	0.1980	0.10430
843786	0.12780	0.17000	0.1578	0.08089

  

	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se
842302	0.2419	0.07871	1.0950	0.9053	8.589
842517	0.1812	0.05667	0.5435	0.7339	3.398
84300903	0.2069	0.05999	0.7456	0.7869	4.585
84348301	0.2597	0.09744	0.4956	1.1560	3.445
84358402	0.1809	0.05883	0.7572	0.7813	5.438
843786	0.2087	0.07613	0.3345	0.8902	2.217

  

	area_se	smoothness_se	compactness_se	concavity_se	concave.points_se
842302	153.40	0.006399	0.04904	0.05373	0.01587
842517	74.08	0.005225	0.01308	0.01860	0.01340
84300903	94.03	0.006150	0.04006	0.03832	0.02058
84348301	27.23	0.009110	0.07458	0.05661	0.01867

84358402	94.44	0.011490	0.02461	0.05688	0.01885
843786	27.19	0.007510	0.03345	0.03672	0.01137
	symmetry_se	fractal_dimension_se	radius_worst	texture_worst	
842302	0.03003	0.006193	25.38	17.33	
842517	0.01389	0.003532	24.99	23.41	
84300903	0.02250	0.004571	23.57	25.53	
84348301	0.05963	0.009208	14.91	26.50	
84358402	0.01756	0.005115	22.54	16.67	
843786	0.02165	0.005082	15.47	23.75	
	perimeter_worst	area_worst	smoothness_worst	compactness_worst	
842302	184.60	2019.0	0.1622	0.6656	
842517	158.80	1956.0	0.1238	0.1866	
84300903	152.50	1709.0	0.1444	0.4245	
84348301	98.87	567.7	0.2098	0.8663	
84358402	152.20	1575.0	0.1374	0.2050	
843786	103.40	741.6	0.1791	0.5249	
	concavity_worst	concave.points_worst	symmetry_worst		
842302	0.7119	0.2654	0.4601		
842517	0.2416	0.1860	0.2750		
84300903	0.4504	0.2430	0.3613		
84348301	0.6869	0.2575	0.6638		
84358402	0.4000	0.1625	0.2364		
843786	0.5355	0.1741	0.3985		
	fractal_dimension_worst				
842302	0.11890				
842517	0.08902				
84300903	0.08758				
84348301	0.17300				
84358402	0.07678				
843786	0.12440				

```

diagnosis <- as.factor(wisc.df$diagnosis)
wisc.data <- wisc.df[,-1]
head(wisc.data)

```

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
842302	17.99	10.38	122.80	1001.0	0.11840
842517	20.57	17.77	132.90	1326.0	0.08474
84300903	19.69	21.25	130.00	1203.0	0.10960
84348301	11.42	20.38	77.58	386.1	0.14250
84358402	20.29	14.34	135.10	1297.0	0.10030
843786	12.45	15.70	82.57	477.1	0.12780

	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	
842302	0.27760	0.3001	0.14710	0.2419	
842517	0.07864	0.0869	0.07017	0.1812	
84300903	0.15990	0.1974	0.12790	0.2069	
84348301	0.28390	0.2414	0.10520	0.2597	
84358402	0.13280	0.1980	0.10430	0.1809	
843786	0.17000	0.1578	0.08089	0.2087	
	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se
842302	0.07871	1.0950	0.9053	8.589	153.40
842517	0.05667	0.5435	0.7339	3.398	74.08
84300903	0.05999	0.7456	0.7869	4.585	94.03
84348301	0.09744	0.4956	1.1560	3.445	27.23
84358402	0.05883	0.7572	0.7813	5.438	94.44
843786	0.07613	0.3345	0.8902	2.217	27.19
	smoothness_se	compactness_se	concavity_se	concave.points_se	
842302	0.006399	0.04904	0.05373	0.01587	
842517	0.005225	0.01308	0.01860	0.01340	
84300903	0.006150	0.04006	0.03832	0.02058	
84348301	0.009110	0.07458	0.05661	0.01867	
84358402	0.011490	0.02461	0.05688	0.01885	
843786	0.007510	0.03345	0.03672	0.01137	
	symmetry_se	fractal_dimension_se	radius_worst	texture_worst	
842302	0.03003	0.006193	25.38	17.33	
842517	0.01389	0.003532	24.99	23.41	
84300903	0.02250	0.004571	23.57	25.53	
84348301	0.05963	0.009208	14.91	26.50	
84358402	0.01756	0.005115	22.54	16.67	
843786	0.02165	0.005082	15.47	23.75	
	perimeter_worst	area_worst	smoothness_worst	compactness_worst	
842302	184.60	2019.0	0.1622	0.6656	
842517	158.80	1956.0	0.1238	0.1866	
84300903	152.50	1709.0	0.1444	0.4245	
84348301	98.87	567.7	0.2098	0.8663	
84358402	152.20	1575.0	0.1374	0.2050	
843786	103.40	741.6	0.1791	0.5249	
	concavity_worst	concave.points_worst	symmetry_worst		
842302	0.7119	0.2654	0.4601		
842517	0.2416	0.1860	0.2750		
84300903	0.4504	0.2430	0.3613		
84348301	0.6869	0.2575	0.6638		
84358402	0.4000	0.1625	0.2364		
843786	0.5355	0.1741	0.3985		
	fractal_dimension_worst				

842302	0.11890
842517	0.08902
84300903	0.08758
84348301	0.17300
84358402	0.07678
843786	0.12440

```
skimr::skim(wisc.df)
```

Table 1: Data summary

Name	wisc.df
Number of rows	569
Number of columns	31
Column type frequency:	
character	1
numeric	30
Group variables	None

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
diagnosis	0	1	1	1	0	2	0

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
radius_mean	0	1	14.13	3.52	6.98	11.70	13.37	15.78	28.11	
texture_mean	0	1	19.29	4.30	9.71	16.17	18.84	21.80	39.28	
perimeter_mean	0	1	91.97	24.30	43.79	75.17	86.24	104.10	188.50	
area_mean	0	1	654.89	351.91	143.50	420.30	551.10	782.70	2501.00	
smoothness_mean	0	1	0.10	0.01	0.05	0.09	0.10	0.11	0.16	
compactness_mean	0	1	0.10	0.05	0.02	0.06	0.09	0.13	0.35	
concavity_mean	0	1	0.09	0.08	0.00	0.03	0.06	0.13	0.43	
concave.points_mean	0	1	0.05	0.04	0.00	0.02	0.03	0.07	0.20	
symmetry_mean	0	1	0.18	0.03	0.11	0.16	0.18	0.20	0.30	
fractal_dimension_mean	0	1	0.06	0.01	0.05	0.06	0.06	0.07	0.10	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
radius_se	0	1	0.41	0.28	0.11	0.23	0.32	0.48	2.87	
texture_se	0	1	1.22	0.55	0.36	0.83	1.11	1.47	4.88	
perimeter_se	0	1	2.87	2.02	0.76	1.61	2.29	3.36	21.98	
area_se	0	1	40.34	45.49	6.80	17.85	24.53	45.19	542.20	
smoothness_se	0	1	0.01	0.00	0.00	0.01	0.01	0.01	0.03	
compactness_se	0	1	0.03	0.02	0.00	0.01	0.02	0.03	0.14	
concavity_se	0	1	0.03	0.03	0.00	0.02	0.03	0.04	0.40	
concave.points_se	0	1	0.01	0.01	0.00	0.01	0.01	0.01	0.05	
symmetry_se	0	1	0.02	0.01	0.01	0.02	0.02	0.02	0.08	
fractal_dimension_se	0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.03	
radius_worst	0	1	16.27	4.83	7.93	13.01	14.97	18.79	36.04	
texture_worst	0	1	25.68	6.15	12.02	21.08	25.41	29.72	49.54	
perimeter_worst	0	1	107.26	33.60	50.41	84.11	97.66	125.40	251.20	
area_worst	0	1	880.58	569.36	185.20	515.30	686.50	1084.00	4254.00	
smoothness_worst	0	1	0.13	0.02	0.07	0.12	0.13	0.15	0.22	
compactness_worst	0	1	0.25	0.16	0.03	0.15	0.21	0.34	1.06	
concavity_worst	0	1	0.27	0.21	0.00	0.11	0.23	0.38	1.25	
concave.points_worst	0	1	0.11	0.07	0.00	0.06	0.10	0.16	0.29	
symmetry_worst	0	1	0.29	0.06	0.16	0.25	0.28	0.32	0.66	
fractal_dimension_worst	0	1	0.08	0.02	0.06	0.07	0.08	0.09	0.21	

Q1. How many observations are in this dataset?

There are 569 observations. Q2. How many of the observations have a malignant diagnosis?

Q3. How many variables/features in the data are suffixed with `_mean`?