

## Data Summary

I've been writing functions to get a better understanding of the Talisman Test Suite data. There are 97 patients.

emphysema		fibrosis		ground_glass		healthy		micronodules	
Patient	Count	Patient	Count	Patient	Count	Patient	Count	Patient	Count
-1_11	81	-1_10	45	108	2	-1_1	97	-1_12	647
-1_3	11	-1_8	32	11	1	-1_10	27	-1_9	186
-1_5	228	1	35	113	6	-1_11	380	102	1904
-1_7	450	100	28	114	109	-1_2	409	103	14
109	63	105	22	117	10	-1_3	34	110	23
24	27	107	68	118	4	-1_4	1006	111	31
50	174	109	284	119	5	-1_5	43	123	1
51	131	125	17	120	29	-1_6	523	129	20
57	12	134	37	121	1	104	234	130	91
		135	291	122	2	26	179	14	235
		136	161	123	60	28	872	16	349
		18	214	124	205	30	50	31	1971
		24	11	128	93	48	890	32	95
		30	2	19	16	64	255	85	601
		44	83	21	9	99	531	86	498
		45	132	25	23			88	5
		46	91	26	36			97	133
		49	79	27	81			98	17
		50	4	37	311				
		53	64	39	134				
		57	74	44	7				
		58	67	47	399				
		59	47	52	213				
		60	40	69	5				
		61	606	77	323				
		62	10	78	2				
		63	120	80	16				
		64	76	82	58				
		65	40	84	1				
		66	19	91	2				
		67	4	95	12				
		69	3	96	51				
		71	11						
		72	88						
		8	44						
		80	12						
		81	3						
		89	75						
Total	1177	Total	3039	Total	2226	Total	5530	Total	6821
Extreme Values		HU Range Issue		Grand Total: 18793					

- The extreme group count imbalances cause issues with group k-fold cv.