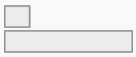
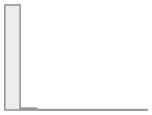
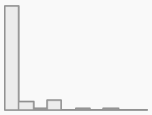
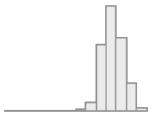
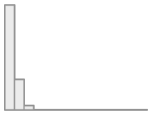
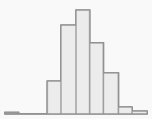
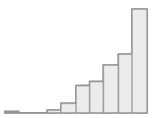
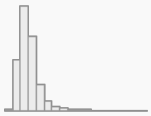
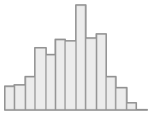
Data Frame Summary

csv_input

Dimensions: 12250 x 33
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	case_id [factor]	1. 0004d251-3f70-4395-b175-c 2. 000d566c-96c7-4f1c-b36e-f 3. 000ead0d-abf5-4606-be04-1 4. 0011a67b-1ba9-4a32-a6b8-7 5. 001887aa-36d0-463f-8bca-d 6. 001944e5-af34-4061-9c09-b 7. 001ad307-4ad3-4f1d-b2fc-e 8. 001cef41-ff86-4d3f-a140-a 9. 001e0309-9c50-42b0-9e38-3 10. 0020317d-d10e-4e75-8fa6-7 [12240 others]	1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 12240 (99.9%)		12250 (100.0%)	0 (0.0%)
2	case_submitter_id [factor]	1. 11LU013 2. 11LU016 3. 11LU022 4. 11LU035 5. AP-6H6G 6. AP-6M60 7. AP-95DK 8. AP-97PG 9. AP-9GTR 10. AP-CL9I [12240 others]	1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 1 (0.0%) 12240 (99.9%)		12250 (100.0%)	0 (0.0%)
3	project_id [factor]	1. CGCI-HTMCP-CC 2. CPTAC-3 3. HCMI-CMDC 4. TCGA-ACC 5. TCGA-BLCA 6. TCGA-BRCA 7. TCGA-CESC 8. TCGA-CHOL 9. TCGA-COAD 10. TCGA-DLBC [27 others]	210 (1.7%) 786 (6.4%) 80 (0.7%) 92 (0.8%) 412 (3.4%) 1097 (9.0%) 307 (2.5%) 48 (0.4%) 459 (3.7%) 48 (0.4%) 8711 (71.1%)		12250 (100.0%)	0 (0.0%)
4	age_at_onset [logical]	All NA's			0 (0.0%)	12250 (100.0%)
5	alcohol_days_per_week [numeric]	1 distinct value	0 : 3 (100.0%)		3 (0.0%)	12247 (100.0%)
6	alcohol_drinks_per_day [logical]	All NA's			0 (0.0%)	12250 (100.0%)
7	alcohol_history [factor]	1. No 2. Yes	646 (41.9%) 897 (58.1%)		1543 (12.6%)	10707 (87.4%)
8	alcohol_intensity [factor]	1. Drinker 2. Lifelong Non-Drinker 3. Non-Drinker 4. Occasional Drinker 5. Unknown	5 (0.6%) 354 (44.4%) 31 (3.9%) 291 (36.5%) 116 (14.6%)		797 (6.5%)	11453 (93.5%)
9	alcohol_type [logical]	All NA's			0 (0.0%)	12250 (100.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
10	asbestos_exposure [factor]	1. No 2. Unknown	1 (16.7%) 5 (83.3%)		6 (0.0%)	12244 (100.0%)
11	bmi [numeric]	Mean (sd) : 28.62 (12.015) min < med < max: 11.735 < 26.839 < 414.894 IQR (CV) : 8.444 (0.42)	2287 distinct values		3507 (28.6%)	8743 (71.4%)
12	cigarettes_per_day [numeric]	Mean (sd) : 4.949 (7.482) min < med < max: 0 < 2.192 < 50 IQR (CV) : 2.74 (1.512)	210 distinct values		1993 (16.3%)	10257 (83.7%)
13	coal_dust_exposure [logical]	All NA's			0 (0.0%)	12250 (100.0%)
14	environmental_tobacco_smoke_exposure [logical]	All NA's			0 (0.0%)	12250 (100.0%)
15	exposure_duration [logical]	All NA's			0 (0.0%)	12250 (100.0%)
16	exposure_type [logical]	All NA's			0 (0.0%)	12250 (100.0%)
17	height [numeric]	Mean (sd) : 167.222 (11.428) min < med < max: 61 < 167 < 200 IQR (CV) : 15 (0.068)	164 distinct values		3523 (28.8%)	8727 (71.2%)
18	marijuana_use_per_week [logical]	All NA's			0 (0.0%)	12250 (100.0%)
19	pack_years_smoked [numeric]	Mean (sd) : 40.283 (33.034) min < med < max: 0.017 < 36 < 730 IQR (CV) : 30 (0.82)	225 distinct values		1961 (16.0%)	10289 (84.0%)
20	radon_exposure [factor]	1. No 2. Unknown	1 (50.0%) 1 (50.0%)		2 (0.0%)	12248 (100.0%)
21	respirable_crystalline_silica_exposure [logical]	All NA's			0 (0.0%)	12250 (100.0%)
22	secondhand_smoke_as_child [logical]	All NA's			0 (0.0%)	12250 (100.0%)
23	smokeless_tobacco_quit_age [logical]	All NA's			0 (0.0%)	12250 (100.0%)
24	smoking_frequency [logical]	All NA's			0 (0.0%)	12250 (100.0%)
25	time_between_waking_and_first_smoke [logical]	All NA's			0 (0.0%)	12250 (100.0%)
26	tobacco_smoking_onset_year [integer]	Mean (sd) : 1976.092 (12.707) min < med < max: 1928 < 1975 < 2015 IQR (CV) : 18.5 (0.006)	56 distinct values		303 (2.5%)	11947 (97.5%)
27	tobacco_smoking_quit_year [integer]	Mean (sd) : 2000.139 (16.458) min < med < max: 1928 < 2004 < 2018 IQR (CV) : 26 (0.008)	53 distinct values		173 (1.4%)	12077 (98.6%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
28	tobacco_smoking_status [integer]	Mean (sd) : 2.079 (1.654) min < med < max: 1 < 1 < 7 IQR (CV) : 2 (0.796)	1 : 598 (56.0%) 2 : 180 (16.9%) 3 : 114 (10.7%) 4 : 97 (9.1%) 5 : 8 (0.7%) 7 : 70 (6.6%)		1067 (8.7%)	11183 (91.3%)
29	tobacco_use_per_day [logical]	All NA's			0 (0.0%)	12250 (100.0%)
30	type_of_smoke_exposure [logical]	All NA's			0 (0.0%)	12250 (100.0%)
31	type_of_tobacco_used [logical]	All NA's			0 (0.0%)	12250 (100.0%)
32	weight [numeric]	Mean (sd) : 83.085 (31.596) min < med < max: 31 < 77 < 371 IQR (CV) : 29 (0.38)	494 distinct values		3598 (29.4%)	8652 (70.6%)
33	years_smoked [numeric]	Mean (sd) : 33.683 (14.089) min < med < max: 0 < 35 < 66 IQR (CV) : 21 (0.418)	67 distinct values		946 (7.7%)	11304 (92.3%)

Generated by summarytools 0.9.9 (R version 4.0.3)
2021-05-22