

## Data Frame Summary

tobacco

Number of rows: 1000

No	Variable	Properties	Stats / Values	Freqs (% of Valid)	Valid / NA
1	gender	type:integer class:factor	1. F 2. M	1: 489 (50.0%) 2: 489 (50.0%)	Val: 978 (97.8%) NA : 22 ( 2.2%)
2	age	type:double class:numeric	mean (sd) = 49.6 (18.29) min < med < max = 18 < 50 < 80 IQR (CV) = 32 (0.37)	64 distinct values	Val: 975 (97.5%) NA : 25 ( 2.5%)
3	age.gr	type:integer class:factor	1. 18-34 2. 35-50 3. 51-70 4. 71 +	1: 258 (26.5%) 2: 241 (24.7%) 3: 317 (32.5%) 4: 159 (16.3%)	Val: 975 (97.5%) NA : 25 ( 2.5%)
4	BMI	type:double class:numeric	mean (sd) = 25.73 (4.49) min < med < max = 8.83 < 25.62 < 39.44 IQR (CV) = 5.72 (0.17)	975 distinct values	Val: 974 (97.4%) NA : 26 ( 2.6%)
5	smoker	type:integer class:factor	1. Yes 2. No	1: 298 (29.8%) 2: 702 (70.2%)	Val: 1000 (100%) NA : 0 ( 0%)
6	cigs.per.day	type:double class:numeric	mean (sd) = 6.78 (11.88) min < med < max = 0 < 0 < 40 IQR (CV) = 11 (1.75)	38 distinct values	Val: 965 (96.5%) NA : 35 ( 3.5%)
7	diseased	type:integer class:factor	1. Yes 2. No	1: 224 (22.4%) 2: 776 (77.6%)	Val: 1000 (100%) NA : 0 ( 0%)
8	disease	type:character class:character	Most frequent: 1. "Hypertension" 2. "Cancer" 3. "Cholesterol" 4. "Heart" 5. "Pulmonary" 6. "Musculoskeletal" 7. "Diabetes" 8. "Hearing" 9. "Digestive" 10. "Hypotension" ... (3 other values)	- - - - - 1: 36 (16.2%) 2: 34 (15.3%) 3: 21 ( 9.5%) 4: 20 ( 9.0%) 5: 20 ( 9.0%) 6: 19 ( 8.6%) 7: 14 ( 6.3%) 8: 14 ( 6.3%) 9: 12 ( 5.4%) 10: 11 ( 5.0%) Others: 21 (9.5%)	Val: 222 (22.2%) NA : 778 (77.8%)
9	samp.wgts	type:double class:numeric	mean (sd) = 1 (0.08) min < med < max = 0.86 < 1.04 < 1.06 IQR (CV) = 0.19 (0.08)	0.86: 267 (26.7%) 1.04: 249 (24.9%) 1.05: 324 (32.4%) 1.06: 160 (16.0%)	Val: 1000 (100%) NA : 0 ( 0%)