

uk_biobank_codebook_2023-Oct-05

Autogenerated data summary from dataMaid

2023-10-05 16:25:25.91789

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	502356
Number of variables	59

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

Codebook summary table

Label	Variable	Class	# unique values	Missing	Description
Participant ID	f.eid	integer	502356	0.00 %	
Sex	sex	ordered	3	0.00 %	
Year of birth	birth_year	integer	38	0.00 %	
Month of birth	birth_month	ordered	13	0.00 %	
Age, baseline	age_0	integer	38	0.00 %	
Birth cohort	birth_cohort	factor	6	0.00 %	
Age group, baseline	age_group_0	factor	7	0.00 %	
Age group, baseline	age_group_retirement_0	factor	6	14.70 %	
Ethnicity	ethnicity	ordered	23	0.18 %	
Ethnicity group	ethnic_group	factor	7	0.55 %	
White or Black and Minority Ethnic	white_BAME	factor	2	0.00 %	

Label	Variable	Class	# unique values	Missing	Description
Age completed full-time education, baseline	age_compl_edu_0	numeric	33	1.30 %	
Highest educational qualification	education	factor	5	1.80 %	
Occupational class (NS-SEC analytical code)	nssec_analytical	factor	9	35.32 %	
Occupational class (NS-SEC analytical code, 5 category)	nssec_5class	factor	6	0.00 %	
Occupational class (NS-SEC analytical code, 5 category)	nssec_5class_v2	factor	6	35.32 %	
Occupation status, wave 0	job_status_0	ordered	8	1.14 %	
Housing tenure, baseline	housing_tenure_0	factor	6	2.41 %	
Type of accomodation, baseline	type_of_accom_0	factor	6	0.52 %	
Number in household, baseline	no_in_household_0	factor	55	0.91 %	
Townsend deprivation index, baseline	town_dep_index	character	6	0.12 %	
IMD for England, baseline	imd_england	factor	6	13.89 %	
IMD for Scotland, baseline	imd_scotland	factor	6	92.86 %	
IMD for Wales, baseline	imd_wales	factor	6	95.78 %	
IMD quantiles, baseline	imd_quantile	factor	6	2.53 %	
Reason lost to follow-up, baseline	rsn_lost_follow_up	factor	4	99.74 %	
Date lost to follow-up	date_lost_follow_up	Date	608	99.74 %	
Part of a multiple birth	part_of_multip_birth	factor	3	1.82 %	
Birth weight (kg)	birth_weight	numeric	367	44.23 %	
Comparative body size at age 10	body_size10yr	factor	4	2.20 %	
Comparative height at age 10	height_10yr	factor	4	2.07 %	
Breastfed as a baby	breastfed	factor	3	23.64 %	
Maternal smoking around birth	maternal_smoking	factor	3	13.85 %	
Above moderate/vigorous recommendation (MET)	abv_modr_vig_recommend	factor	3	19.92 %	

Label	Variable	Class	# unique values	Missing	Description
Above moderate/vigorous walking recommendation (MET)	abv_modr_vig_wlk_recommend	factor	3	19.95 %	
IPAQ activity group	ipaq_active_grp	factor	4	19.92 %	
MET minutes per week for moderate activity	met_mpw_moderate_activ	numeric	312	19.92 %	
MET minutes per week for vigorous activity	met_mpw_vigorous_activ	numeric	270	19.92 %	
MET minutes per week for walking	met_mpw_walking	numeric	319	19.92 %	
Summed minutes activity	summed_minutes_activity	numeric	400	19.92 %	
Summed days activity	summed_days_activity	numeric	23	17.60 %	
Days of mod_act 10+ minutes, wave 0	days_wks_mod_act_0	numeric	9	5.43 %	
Days of vig_act 10+ minutes, wave 0	days_wks_vig_act_0	numeric	9	5.43 %	
Computer use (hours/day)	computer_use_hours_days_0	integer	24	22.44 %	
TV viewing (hours/day)	tv_watching_hours_days_0	integer	25	5.73 %	
Sedentary behaviour (hours/day) categories	sedentary_hours_day_0	factor	7	2.46 %	
Occupational activity, baseline	occupational_activity_0	factor	7	42.93 %	
Smoking status, wave 0	smoking_status_0	factor	4	0.59 %	
Frequency of drinking alcohol, wave 0	alcohol_freq_0	factor	7	0.30 %	
Weight, baseline	weight_0	numeric	1381	0.55 %	
Height, baseline	height_0	numeric	558	0.51 %	
Waist circumference, baseline	waist_circum_0	numeric	739	0.43 %	
Hip circumference, baseline	hip_circum_0	numeric	610	0.44 %	
BMI, baseline	bmi_0	numeric	46965	0.62 %	
Body fat percentage, baseline	body_fat_perc_0	numeric	584	2.07 %	
Comorbiddties, baseline	healthcat_0	factor	4	0.00 %	
Max grip strength (kg), baseline	handgrip_max_0	numeric	90	0.49 %	
Reason for missing handgrip strength assessment	rsn_grip_missing_cat	character	6	99.87 %	

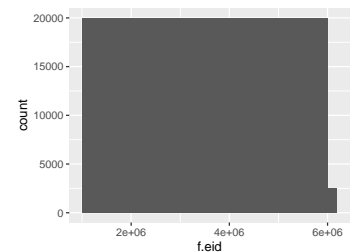
Label	Variable	Class	# unique values	Missing	Description
Maximum handgrip strength (kg) (imputed), baseline	handgrip_max_0_imp	numeric	91	0.43 %	

Variable list

f.eid

Participant ID

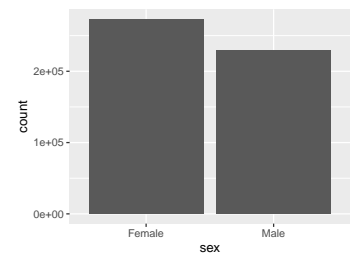
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	502356
Median	3512377
1st and 3rd quartiles	2256180.5; 4768528.25
Min. and max.	1000010; 6024784



sex

Sex

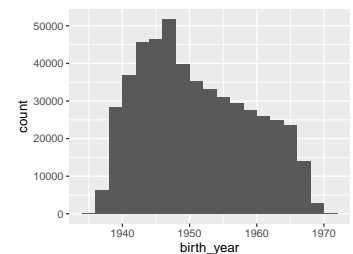
Feature	Result
Variable type	ordered
Number of missing obs.	1 (0 %)
Number of unique values	2
Mode	"Female"
Reference category	Female



birth_year

Year of birth

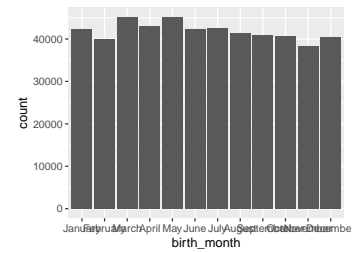
Feature	Result
Variable type	integer
Number of missing obs.	1 (0 %)
Number of unique values	37
Median	1950
1st and 3rd quartiles	1945; 1958
Min. and max.	1934; 1971



birth_month

Month of birth

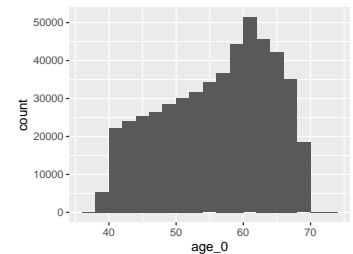
Feature	Result
Variable type	ordered
Number of missing obs.	1 (0 %)
Number of unique values	12
Mode	"May"
Reference category	January



age_0

Age, baseline

Feature	Result
Variable type	integer
Number of missing obs.	1 (0 %)
Number of unique values	37
Median	58
1st and 3rd quartiles	50; 63
Min. and max.	37; 73



birth_cohort

Birth cohort

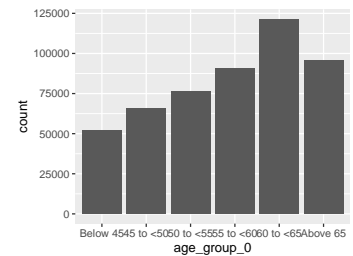
Feature	Result
Variable type	factor
Number of missing obs.	1 (0 %)
Number of unique values	5
Mode	"1940s"
Reference category	1930s



age_group_0

Age group, baseline

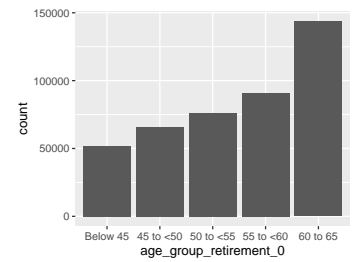
Feature	Result
Variable type	factor
Number of missing obs.	1 (0 %)
Number of unique values	6
Mode	"60 to <65"
Reference category	Below 45



age_group_retirement_0

Age group, baseline

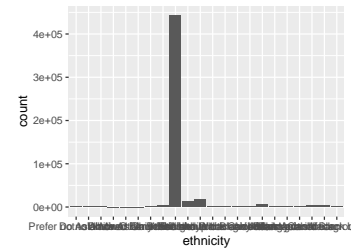
Feature	Result
Variable type	factor
Number of missing obs.	73866 (14.7 %)
Number of unique values	5
Mode	"60 to 65"
Reference category	Below 45



ethnicity

Ethnicity

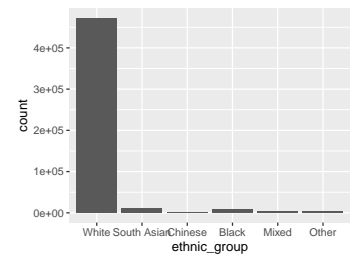
Feature	Result
Variable type	ordered
Number of missing obs.	899 (0.18 %)
Number of unique values	22
Mode	"British"
Reference category	Prefer not to answer



ethnic_group

Ethnicity group

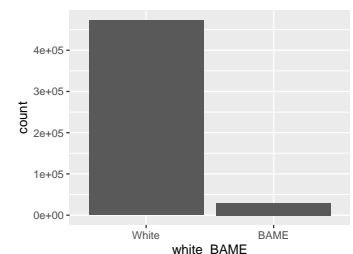
Feature	Result
Variable type	factor
Number of missing obs.	2777 (0.55 %)
Number of unique values	6
Mode	"White"
Reference category	White



white_BAME

White or Black and Minority Ethnic

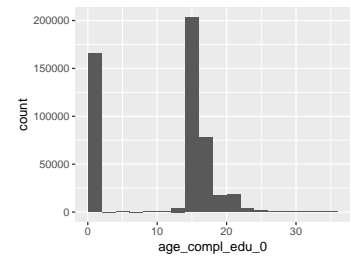
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"White"
Reference category	White



age_compl_edu_0

Age completed full-time education, baseline

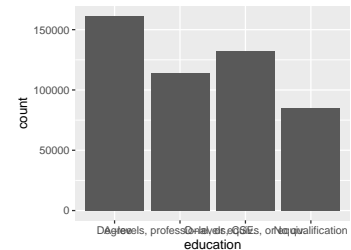
Feature	Result
Variable type	numeric
Number of missing obs.	6531 (1.3 %)
Number of unique values	32
Median	15
1st and 3rd quartiles	0; 16
Min. and max.	0; 35



education

Highest educational qualification

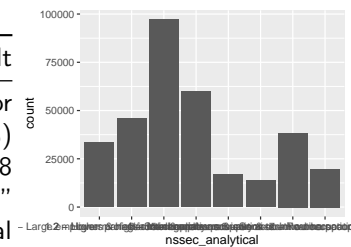
Feature	Result
Variable type	factor
Number of missing obs.	9049 (1.8 %)
Number of unique values	4
Mode	"Degree"
Reference category	Degree



nssec_analytical

Occupational class (NS-SEC analytical code)

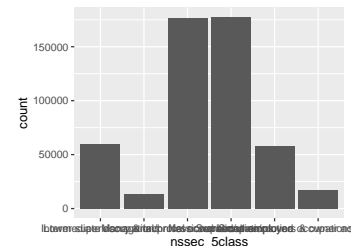
Feature	Result
Variable type	factor
Number of missing obs.	177415 (35.32 %)
Number of unique values	8
Mode	"2 - Lower managl/admin & profes. occupations"
Reference category	1.1 - Large employers & higher managerial



nssec_5class

Occupational class (NS-SEC analytical code, 5 category)

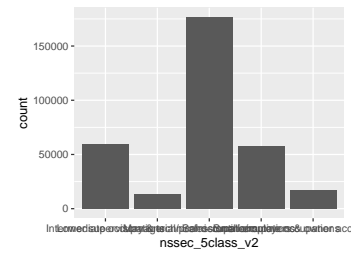
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	6
Mode	"Never worked/unemployed"
Reference category	Intermediate occupations



nssec_5class_v2

Occupational class (NS-SEC analytical code, 5 category)

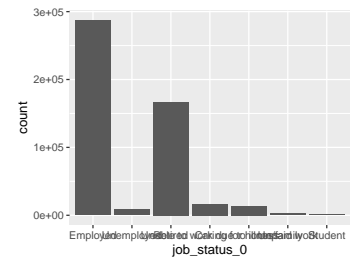
Feature	Result
Variable type	factor
Number of missing obs.	177415 (35.32 %)
Number of unique values	5
Mode	"Managerial/professional occupations"
Reference category	Intermediate occupations



job_status_0

Occupation status, wave 0

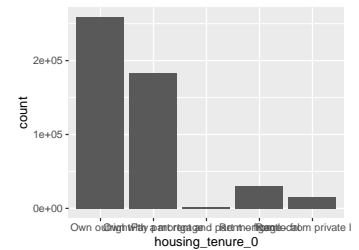
Feature	Result
Variable type	ordered
Number of missing obs.	5749 (1.14 %)
Number of unique values	7
Mode	"Employed"
Reference category	Employed



housing_tenure_0

Housing tenure, baseline

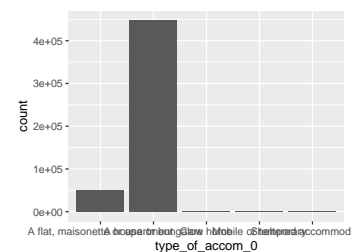
Feature	Result
Variable type	factor
Number of missing obs.	12106 (2.41 %)
Number of unique values	5
Mode	"Own outright"
Reference category	Own outright



type_of_accom_0

Type of accommodation, baseline

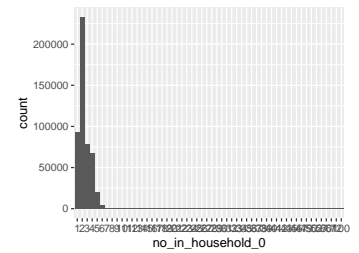
Feature	Result
Variable type	factor
Number of missing obs.	2618 (0.52 %)
Number of unique values	5
Mode	"A house or bungalow"
Reference category	A flat, maisonette or apartment



no_in_household_0

Number in household, baseline

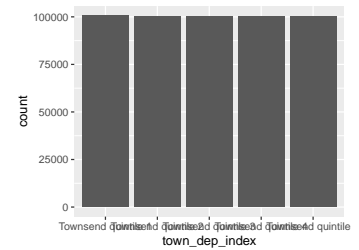
Feature	Result
Variable type	factor
Number of missing obs.	4548 (0.91 %)
Number of unique values	54
Mode	"2"
Reference category	1



town_dep_index

Townsend deprivation index, baseline

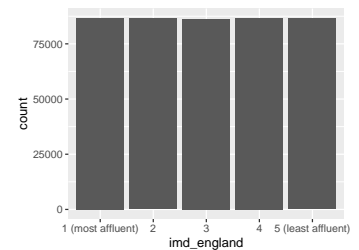
Feature	Result
Variable type	character
Number of missing obs.	624 (0.12 %)
Number of unique values	5
Mode	"Townsend quintile 1"



imd_england

IMD for England, baseline

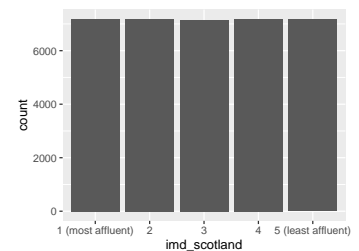
Feature	Result
Variable type	factor
Number of missing obs.	69756 (13.89 %)
Number of unique values	5
Mode	"1 (most affluent)"
Reference category	1 (most affluent)



imd_scotland

IMD for Scotland, baseline

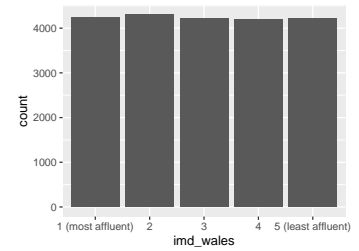
Feature	Result
Variable type	factor
Number of missing obs.	466512 (92.86 %)
Number of unique values	5
Mode	"1 (most affluent)"
Reference category	1 (most affluent)



imd_wales

IMD for Wales, baseline

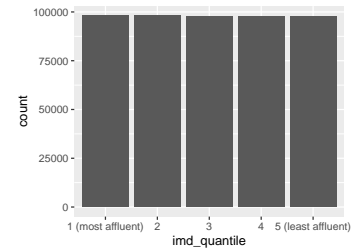
Feature	Result
Variable type	factor
Number of missing obs.	481174 (95.78 %)
Number of unique values	5
Mode	"2"
Reference category	1 (most affluent)



imd_quantile

IMD quantiles, baseline

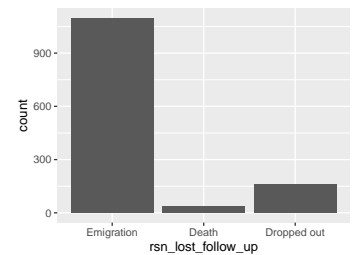
Feature	Result
Variable type	factor
Number of missing obs.	12730 (2.53 %)
Number of unique values	5
Mode	"1 (most affluent)"
Reference category	1 (most affluent)



rsn_lost_follow_up

Reason lost to follow-up, baseline

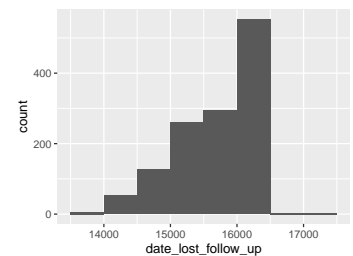
Feature	Result
Variable type	factor
Number of missing obs.	501062 (99.74 %)
Number of unique values	3
Mode	"Emigration"
Reference category	Emigration



date_lost_follow_up

Date lost to follow-up

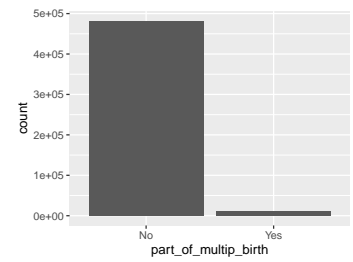
Feature	Result
Variable type	Date
Number of missing obs.	501059 (99.74 %)
Number of unique values	607
Mode	"2014-06-03"
Min. and max.	2007-12-03; 2016-12-06
1st and 3rd quartiles	2012-01-11; 2014-06-03



part_of_multip_birth

Part of a multiple birth

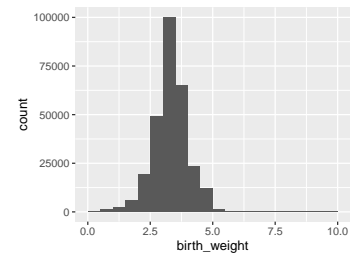
Feature	Result
Variable type	factor
Number of missing obs.	9146 (1.82 %)
Number of unique values	2
Mode	"No"
Reference category	No



birth_weight

Birth weight (kg)

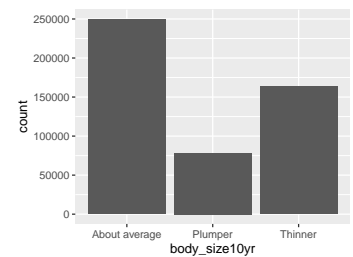
Feature	Result
Variable type	numeric
Number of missing obs.	222192 (44.23 %)
Number of unique values	366
Median	3.32
1st and 3rd quartiles	2.95; 3.66
Min. and max.	0.45; 10



body_size10yr

Comparative body size at age 10

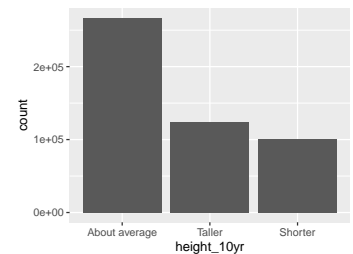
Feature	Result
Variable type	factor
Number of missing obs.	11043 (2.2 %)
Number of unique values	3
Mode	"About average"
Reference category	About average



height_10yr

Comparative height at age 10

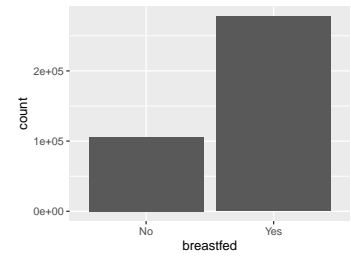
Feature	Result
Variable type	factor
Number of missing obs.	10394 (2.07 %)
Number of unique values	3
Mode	"About average"
Reference category	About average



breastfed

Breastfed as a baby

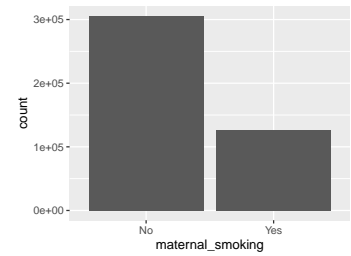
Feature	Result
Variable type	factor
Number of missing obs.	118756 (23.64 %)
Number of unique values	2
Mode	"Yes"
Reference category	No



maternal_smoking

Maternal smoking around birth

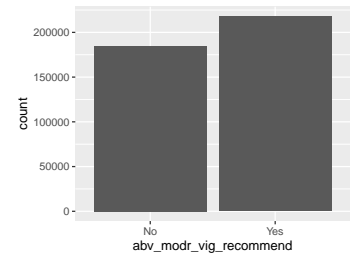
Feature	Result
Variable type	factor
Number of missing obs.	69595 (13.85 %)
Number of unique values	2
Mode	"No"
Reference category	No



abv_modr_vig_recommend

Above moderate/vigorous recommendation (MET)

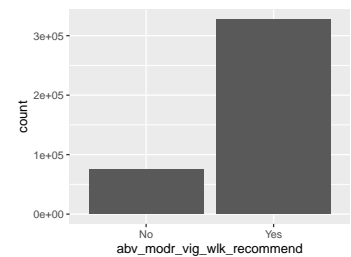
Feature	Result
Variable type	factor
Number of missing obs.	100091 (19.92 %)
Number of unique values	2
Mode	"Yes"
Reference category	No



abv_modr_vig_wlk_recommend

Above moderate/vigorous walking recommendation (MET)

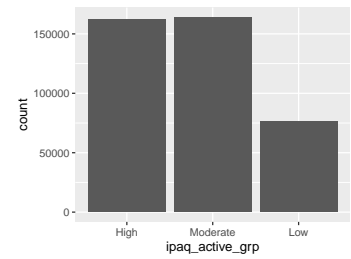
Feature	Result
Variable type	factor
Number of missing obs.	100203 (19.95 %)
Number of unique values	2
Mode	"Yes"
Reference category	No



ipaq_active_grp

IPAQ activity group

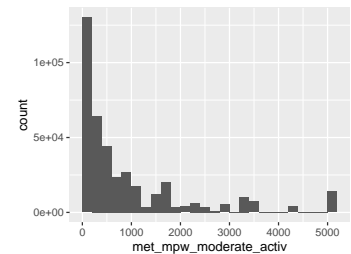
Feature	Result
Variable type	factor
Number of missing obs.	100091 (19.92 %)
Number of unique values	3
Mode	"Moderate"
Reference category	High



met_mpw_moderate_activ

MET minutes per week for moderate activity

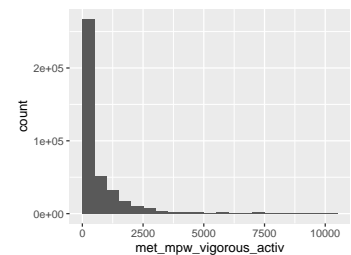
Feature	Result
Variable type	numeric
Number of missing obs.	100091 (19.92 %)
Number of unique values	311
Median	480
1st and 3rd quartiles	120; 1200
Min. and max.	0; 5040



met_mpw_vigorous_activ

MET minutes per week for vigorous activity

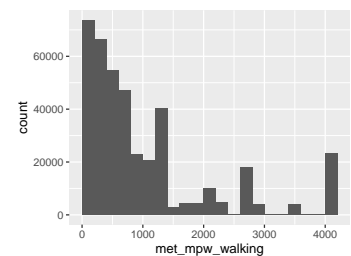
Feature	Result
Variable type	numeric
Number of missing obs.	100091 (19.92 %)
Number of unique values	269
Median	240
1st and 3rd quartiles	0; 960
Min. and max.	0; 10080



met_mpw_walking

MET minutes per week for walking

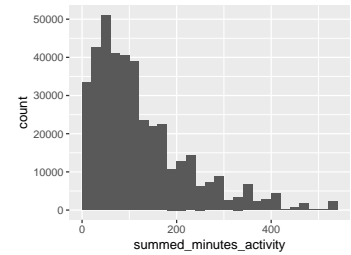
Feature	Result
Variable type	numeric
Number of missing obs.	100091 (19.92 %)
Number of unique values	318
Median	693
1st and 3rd quartiles	297; 1386
Min. and max.	0; 4158



summed_minutes_activity

Summed minutes activity

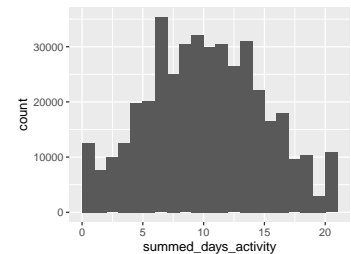
Feature	Result
Variable type	numeric
Number of missing obs.	100091 (19.92 %)
Number of unique values	399
Median	100
1st and 3rd quartiles	55; 180
Min. and max.	0; 540



summed_days_activity

Summed days activity

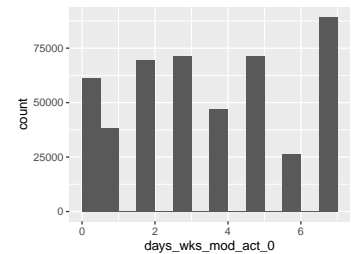
Feature	Result
Variable type	numeric
Number of missing obs.	88438 (17.6 %)
Number of unique values	22
Median	11
1st and 3rd quartiles	7; 14
Min. and max.	0; 21



days_wks_mod_act_0

Days of mod_act 10+ minutes, wave 0

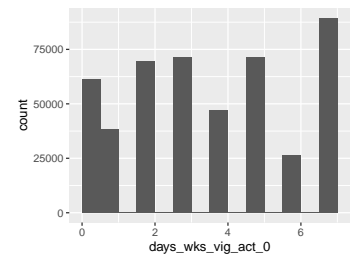
Feature	Result
Variable type	numeric
Number of missing obs.	27263 (5.43 %)
Number of unique values	8
Median	3
1st and 3rd quartiles	2; 5
Min. and max.	0; 7



days_wks_vig_act_0

Days of vig_act 10+ minutes, wave 0

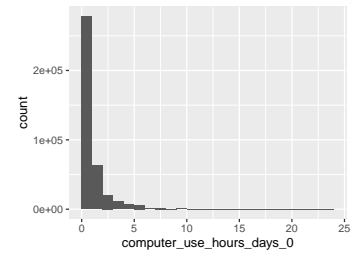
Feature	Result
Variable type	numeric
Number of missing obs.	27263 (5.43 %)
Number of unique values	8
Median	3
1st and 3rd quartiles	2; 5
Min. and max.	0; 7



computer_use_hours_days_0

Computer use (hours/day)

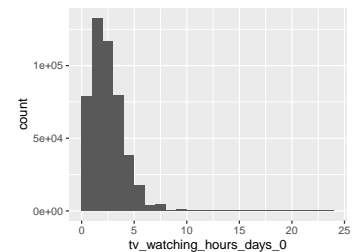
Feature	Result
Variable type	integer
Number of missing obs.	112716 (22.44 %)
Number of unique values	23
Median	1
1st and 3rd quartiles	0; 2
Min. and max.	0; 24



tv_watching_hours_days_0

TV viewing (hours/day)

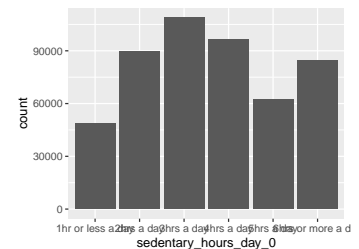
Feature	Result
Variable type	integer
Number of missing obs.	28772 (5.73 %)
Number of unique values	24
Median	3
1st and 3rd quartiles	2; 4
Min. and max.	0; 24



sedentary_hours_day_0

Sedentary behaviour (hours/day) categories

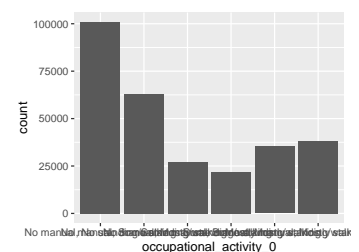
Feature	Result
Variable type	factor
Number of missing obs.	12360 (2.46 %)
Number of unique values	6
Mode	"3hrs a day"
Reference category	1hr or less a day



occupational_activity_0

Occupational activity, baseline

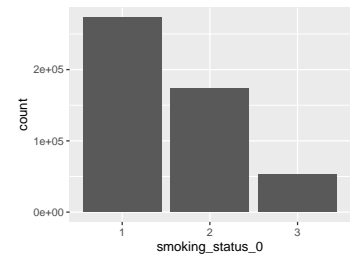
Feature	Result
Variable type	factor
Number of missing obs.	215684 (42.93 %)
Number of unique values	6
Mode	"No manual, No standing/walking"
Reference category	No manual, No standing/walking



smoking_status_0

Smoking status, wave 0

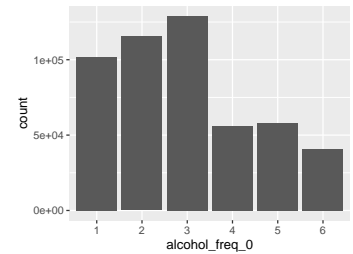
Feature	Result
Variable type	factor
Number of missing obs.	2949 (0.59 %)
Number of unique values	3
Mode	"1"
Reference category	1



alcohol_freq_0

Frequency of drinking alcohol, wave 0

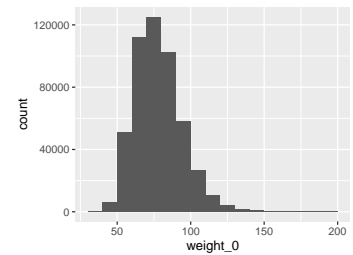
Feature	Result
Variable type	factor
Number of missing obs.	1501 (0.3 %)
Number of unique values	6
Mode	"3"
Reference category	1



weight_0

Weight, baseline

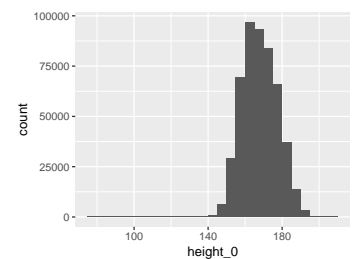
Feature	Result
Variable type	numeric
Number of missing obs.	2775 (0.55 %)
Number of unique values	1380
Median	76.4
1st and 3rd quartiles	66.5; 87.5
Min. and max.	30; 197.7



height_0

Height, baseline

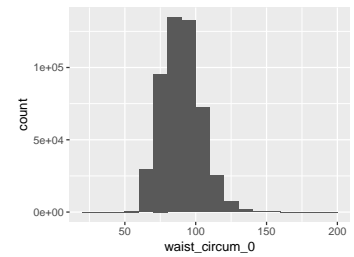
Feature	Result
Variable type	numeric
Number of missing obs.	2539 (0.51 %)
Number of unique values	557
Median	168
1st and 3rd quartiles	161.8; 175
Min. and max.	75; 209



waist_circum_0

Waist circumference, baseline

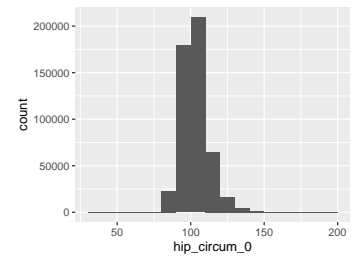
Feature	Result
Variable type	numeric
Number of missing obs.	2161 (0.43 %)
Number of unique values	738
Median	90
1st and 3rd quartiles	80; 99
Min. and max.	20; 197



hip_circum_0

Hip circumference, baseline

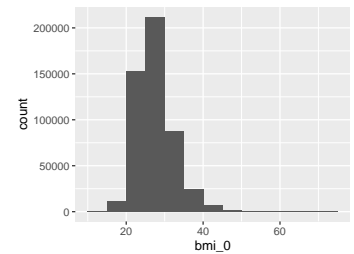
Feature	Result
Variable type	numeric
Number of missing obs.	2220 (0.44 %)
Number of unique values	609
Median	102
1st and 3rd quartiles	97; 108
Min. and max.	30; 195



bmi_0

BMI, baseline

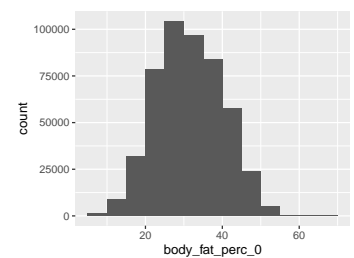
Feature	Result
Variable type	numeric
Number of missing obs.	3105 (0.62 %)
Number of unique values	46964
Median	26.74
1st and 3rd quartiles	24.14; 29.91
Min. and max.	12.12; 74.68



body_fat_perc_0

Body fat percentage, baseline

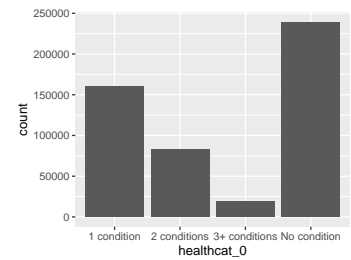
Feature	Result
Variable type	numeric
Number of missing obs.	10405 (2.07 %)
Number of unique values	583
Median	31.1
1st and 3rd quartiles	25.2; 37.7
Min. and max.	5; 69.8



healthcat_0

Comorbidities, baseline

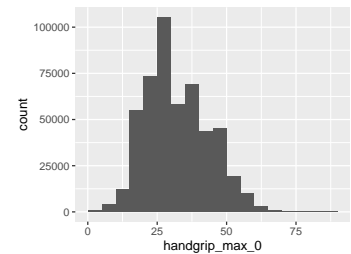
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"No condition"
Reference category	1 condition



handgrip_max_0

Max grip strength (kg), baseline

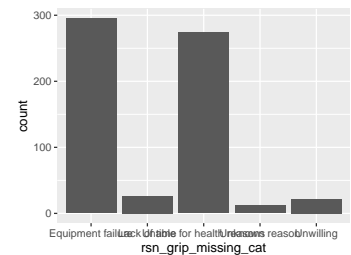
Feature	Result
Variable type	numeric
Number of missing obs.	2458 (0.49 %)
Number of unique values	89
Median	30
1st and 3rd quartiles	24; 40
Min. and max.	1; 90



rsn_grip_missing_cat

Reason for missing handgrip strength assessment

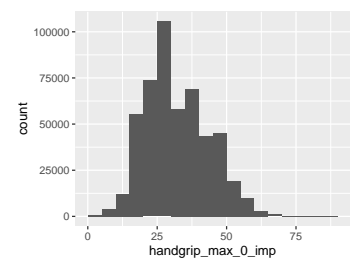
Feature	Result
Variable type	character
Number of missing obs.	501727 (99.87 %)
Number of unique values	5
Mode	"Equipment failure"



handgrip_max_0_imp

Maximum handgrip strength (kg) (imputed), baseline

Feature	Result
Variable type	numeric
Number of missing obs.	2184 (0.43 %)
Number of unique values	90
Median	30
1st and 3rd quartiles	24; 40
Min. and max.	1; 90



Report generation information:

- Created by: Mohamed Yusuf (username: mohamedyusuf).
- Report creation time: Thu Oct 05 2023 16:25:55
- Report was run from directory: /Users/mohamedyusuf/Library/CloudStorage/OneDrive-LondonSchoolofHygieneandTrop

- dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.3.0)]
- R version 4.3.1 (2023-06-16).
- Platform: aarch64-apple-darwin20 (64-bit)(Africa/Nairobi).
- Function call: `dataMaid::makeDataReport(data = ukb_final, mode = c("summarize", "visualize"),
file = "01_data_cleaning/1c_cleaned_data/uk_biobank_codebook.Rmd", replace = TRUE,
codebook = TRUE, reportTitle = paste0("uk_biobank_codebook_", format(Sys.time(),
"%Y-%b-%d")))`