Dataset students.xlxs

Variable description

This dataset collect information on students enrolled in a school:

sex	student's sex (binary: 'F' - female or 'M' - male)
age	student's age (numeric: from 15 to 22)
address	student's home address type (binary: 'U' - urban or 'R' - rural)
famsize	family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
Pstatus	parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
Medu	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 “ 5th to 9th grade, 3 “ secondary education or 4 “ higher education)
Fedu	father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 “ 5th to 9th grade, 3 “ secondary education or 4 â€“ higher education)
Mjob	mother's job (nominal: 'teacher' 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
Fjob	father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
reason	reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
guardian	student's guardian (nominal: 'mother', 'father' or 'other')
traveltime	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
studytime	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
failures	number of past class failures (numeric: n if 1<=n<3, else 4)
schoolsup	extra educational support (binary: yes or no)
famsup	family educational support (binary: yes or no)
paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
internet	Internet access at home (binary: yes or no)

romantic	with a romantic relationship (binary: yes or no)
famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
freetime	free time after school (numeric: from 1 - very low to 5 - very high)
goout	going out with friends (numeric: from 1 - very low to 5 - very high)
Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
health	current health status (numeric: from 1 - very bad to 5 - very good)
absences	number of school absences (numeric: from 0 to 93)

Import this dataset within R.studio software and recall it as “stud”

Then, reply to these questions:

1. How many individuals are included in this data? How many variables?
2. How many students study at least 5 weekly hours?
3. What is it the main reason for choosing this school?
4. Plot the rel.freq. distribution of free time after school and provide a comment on this distribution
5. Compute summary statistics of the number of school absences. Could you represent graphically this distribution with a histogram? How is the shape of this distribution?
6. Compute a box plot of the variable absences. Can you detect the presence of any outliers? if yes, Identify the n. of outliers and compute a new boxplot without outliers.
7. Compute a box plot of the variable absences by workday alcohol consumption. Provide a comment of this graphical representation.
8. Compute variability indicators of the variable absence
9. Compute the mean, standard deviation and coefficient of variation of number of school absences by workday alcohol consumption (provide a comment on these statistics).