Birthweight reduced data set

This dataset contains information on new born babies and their parents. It contains mostly continuous variables (although some have only a few values e.g. number of cigarettes smoked per day) and is most useful for correlation and regression. The birthweights of the babies who mothers smoked have been adjusted slightly to exaggerate the differences between mothers who smoked and didn't smoke so students can see the difference more clearly in a scatterplot with gestational age and scatter colour coded by smoking status.

Main dependent variable = Birthweight (kg)

Name	Variable	Data type
ID	Baby number	
length	Length of baby (cm)	Scale
Birthweight	Weight of baby (kg)	Scale
headcirumference	Head Circumference	Scale
Gestation	Gestation (weeks)	Scale
smoker	Mother smokes 1 = smoker 0 = non-smoker	Binary
motherage	Maternal age	Scale
mnocig	Number of cigarettes smoked per day by mother	Scale
mheight	Mothers height (cm)	Scale
mppwt	Mothers pre-pregnancy weight (kg)	Scale
fage	Father's age	Scale
fedyrs	Father's years in education	Scale
fnocig	Number of cigarettes smoked per day by father	Scale
fheight	Father's height (cm)	Scale
lowbwt	Low birth weight, 0 = No and 1 = yes	Binary
mage35	Mother over 35, 0 = No and 1 = yes	Binary

Data contributed by Ellen Marshall, University of Sheffield

Possible research questions

Technique	Question		
1. **Independent t-test**	Do smokers have lighter babies?	Do women over 35 have lighter babies?	
2. **Pearson's Correlation**	Relationship between maternal height and baby length.	Relationship between gestation and baby weight.	Relationship between mother's pre=pregnancy weight and baby weight.
3. **Simple regression**	Can mother's height predict baby length?	Can gestational age predict baby weight?	
4. **Multiple regression**	Any combination of variables to predict baby length	After controlling for gestational age, does mothers pre-pregnancy weight have an effect on birthweight	
5. **Logistic regression**	Predicting probability of low birth weight (binary low < 6lbs) using any independent variables		
6. **Chi-squared**	Is there a relationship between smoking and low baby weight? (use binary variables for both)	Is there a relationship between over 35's and low baby weight? (use binary variables for both)	
7. **Checking normality**	Baby weight is normally distributed. Number of cigarettes smoked per day is not.		
8. **Cluster analysis**	Cluster analysis on the variables shows a clear split on variables relating to the babies and the parents		