Requirements for project report:

* Provide background information on the data set. It should be presented as a summary.
* Add a reference list at the back of the project document when you are referencing information either from an article or from an URL. Use the Harvard or APA method for referencing purposes in your document and your reference list.
* The report should be word processed and with 1.5 line spacing. Font type should be Arial and font size should be "12".
* Create brief, well-organised appendices for each problem where details detract from the report body's simplicity to read.
* For data analysis documented in your report, RAW AND UNINTERPRETED COMPUTER OUTPUT IS UNACCEPTABLE. You should have a caption by every ﬁgure and table that describes it and tells the reader brieﬂy what you see.
* Organise the sections to tell the story you uncovered, not the circuitous path you may have taken to get there. Remember that even that best data analysis is worthless if your reader cannot understand it.

Project instructions:

Use the data provided to build an appropriate regression model to analyse how strongly the measures given above correlate with one another and with the response variable. Do a correlation analysis and principal component analysis as well. Use the backward selection method to select which variables to keep in the regression model and then analyse the effect of each backward selection step of the new model and the significance on parameters. Write out the best full statistical model (using notation you define) and state the model assumptions.

For the model written in the previous part (that is, eﬀects in ﬁtted model should be as in the model speciﬁcation above), assess and address deviations from model assumptions. This may be an iterative process.

 (Note: If model assumptions are not met, try to address that. If you cannot address unsatisﬁed model assumptions, mention this and continue as though the model assumptions are met.)

For the selected model ﬁt provide evidence/justified reasons for the decisions made to arrive at your model (e.g. one hot encoding categorical variables), consider moving intermediate model ﬁt details to the appendix.

State and conduct statistical tests for the parameters and interpret the test results. How many degrees-of-freedom are allocated to each source of variation in the ﬁnal model? Interpret only the significant parameter estimates.