Modelling

Try:

- Different Modelling Approach: lm, glm … interaction, collinearity (poly)
- Outcome Type / Response Variable: Total Contacts
- Discrete number of predictors, and all & stepwise

Note:

- Use a representative sample of respondents based on population age–group & sex; additionally constrained by an age–group intersection sex cut–off due to survey data limitations. (Starting from 30 – 34)
- The model can decide which predictors are or are not important
- Don't forget to present your homework in R markdown
- A modelling approach that adjusts for all other variables.

Possible hypotheses:

- total contacts is significantly dependent on occupation
- total contacts is significantly dependent on age group
- household_size
- occupation
- interactions, e.g., occupation & age group
- rural/urban

Limitations

- Discrete number of predictors: are we ignoring confounding factors due to how the data was collected?

- all & stepwise