Quizzes > Solutions

Missing **data**

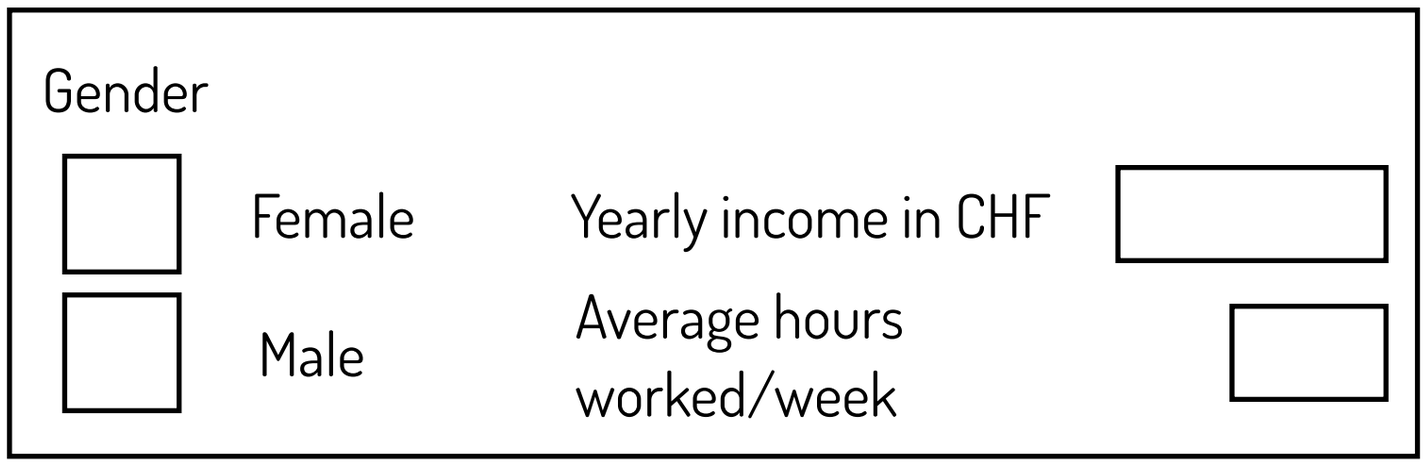
by Sina Rüeger

Notes & Material

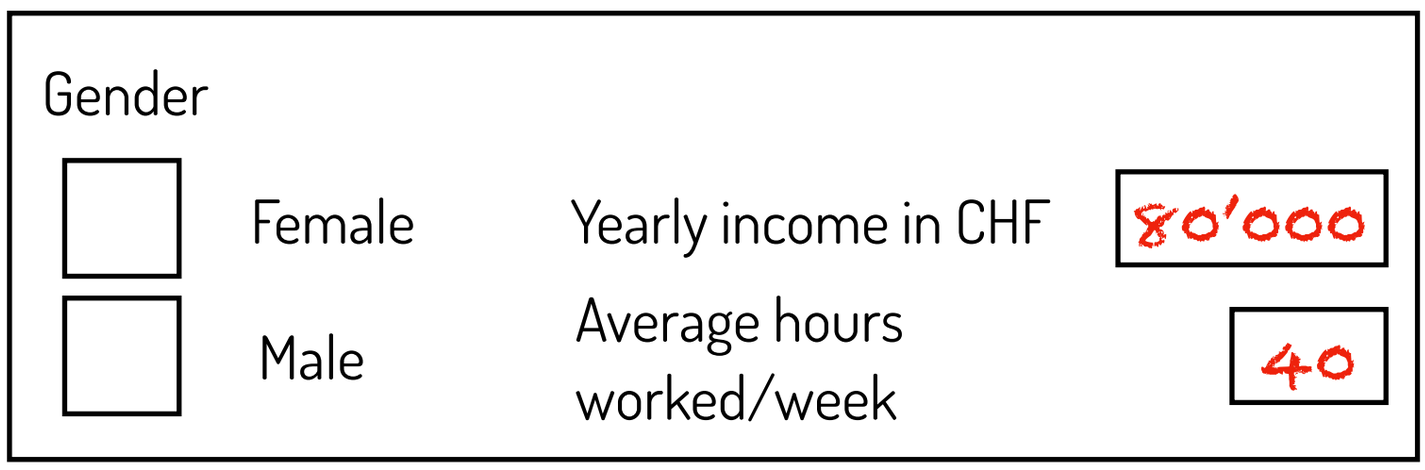
⬇

https://github.com/sinarueeger/teaching/tree/master/missing_data

The following excerpt was part of a larger questionnaire that was sent to a number of individuals.
The goal of the questionnaire was to assess income and number of hours worked.



In several cases, no gender was reported, but a yearly income and the average hours worked per week provided.



Questions

1. What could be the reason for the missing data in “gender”?
2. For the subsequent analysis of income and hours worked per week, observations with missing gender were removed. What is the danger of doing an analysis with removed observations that had missing values?

1. “Male” and “female” are not the only gender. Therefore, if somebody does identify not as male/female, both boxes will remain empty.
   Some people might simply not want to report their gender.
2. Two main consequences:
   - There could be bias introduced. For example, over- or underestimation of the average income or hours worked.
   - Ultimately, removing samples leads to lower sample size, therefore the estimations are less precise.