# Lab 1.2 – Processing a Healthcare Survey

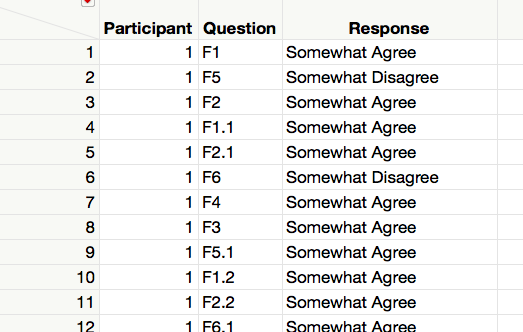
Dr. Bergen, Director of the Statistical Consulting Center at WSU, needs you to prepare the attached data for analysis. Note that **health\_survey.csv** contains the responses to a series of health-related questions that we want to recode on a numeric scale and then aggregate. Dr. Bergen had a follow-up meeting with his client, and it turns out that some of the columns need a reverse coding, see the *Needs Reverse Coding?* column in **ReverseCodingItems.csv**. Please perform the following steps to prepare the required csv file.

|  |  |  |
| --- | --- | --- |
| Old Label | New Coded Value | Reverse Coding |
| “Strongly Disagree” | 1 | 5 |
| “Somewhat Disagree” | 2 | 4 |
| “Neither Agree nor Disagree” | 3 | 3 |
| “Somewhat Agree” | 4 | 2 |
| “Strongly Agree” | 5 | 1 |

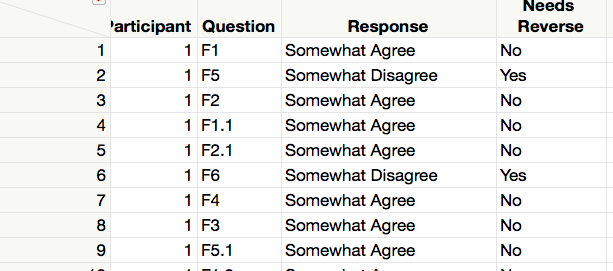
1. Look at the questions that need reverse coding and explain why it makes sense to reverse the coding on these items.

The reverse coding in this case makes sense as the intent of the survey questions is ‘opposite’ of the others.

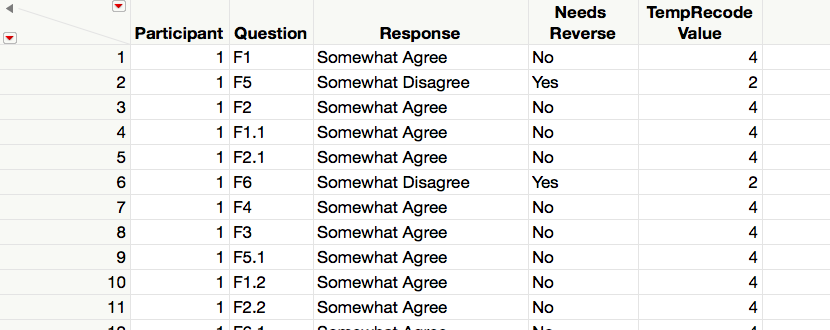
1. Create a data repository for this project. Make sure that notebook is in the root of the repository and the data files are stored in a data folder.
2. You will need to redo the file construction, but now need to take the reverse coding into account.
   1. *Stack* the columns.



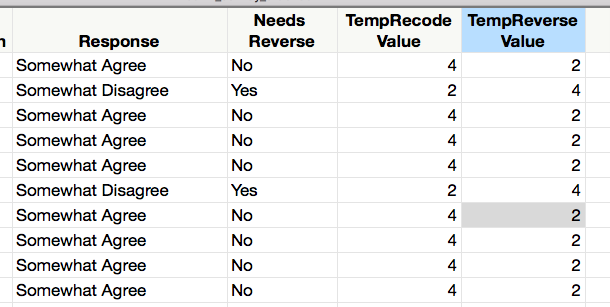
* 1. Make a new column called *Needs Reverse* by *Recoding* the Question Identifier to “Yes” or “No” per the *Needs Reverse Coding?* column in **ReverseCodingItems.csv.**



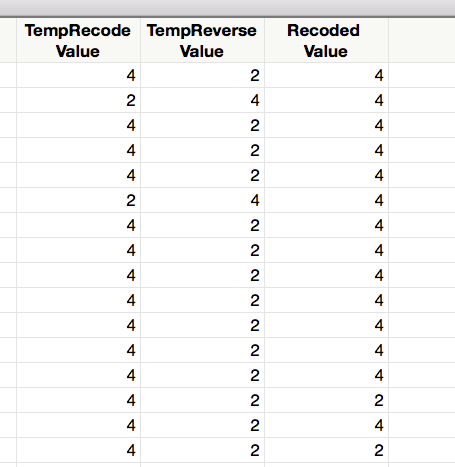
* 1. Make a new column called *Temp Coded Value* by *Recoding* the Questions Responses to the New Coded Values.



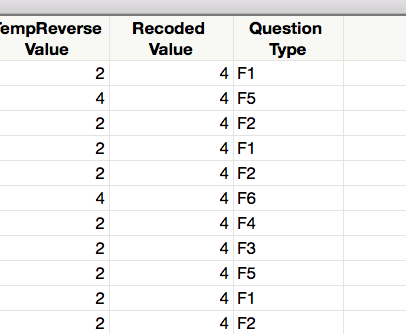
* 1. Make a new column called *Temp Coded Value* by *Recoding* the Questions Responses to the Reversed Coded Values.



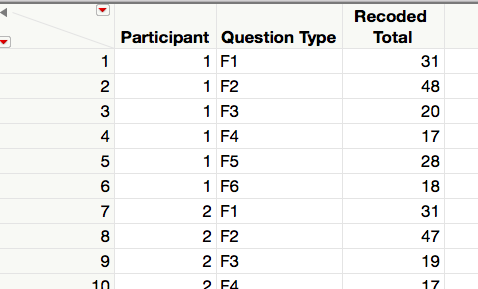
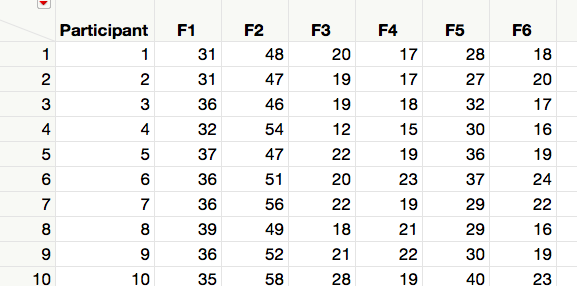
* 1. Make a new column called *Recoded Value* that holds the correct value for each question. You will want to use an case-when statement and use *Needs Reverse* to decide which temporary value to use.



* 1. Make a new column by *Recoding* the Question Types to *F1, F2, …, F6.*

**

* 1. *Aggregate* and *Split.*

1. Write the resulting table to a CSV file named **health\_survey\_summary.csv** in the data folder.

**Deliverables.** Submit this document with your answer to question 1, a link to your data repository, and a csv file with your final table.

<https://github.com/JacBGar/DSCI326/tree/main/Lab1_2>