

Data Cleaning Checklist

- ☐ What are the final clean datasets that should be produced?
- ☐ What file format should each clean dataset be in?
 - What did we say in our Data Management Plan?
- ☐ What are the team standards for what constitutes a clean dataset?
 - Are these standards laid out in documentation?
- ☐ Who is in charge of cleaning data?
 - When will data cleaning plans be created and reviewed?
 - When will data cleaning occur?
 - Is there a timeline for when clean data is needed?
- ☐ What tools will be used for data cleaning?
 - Are there any costs associated with our data cleaning tools?
- ☐ What structure should clean datasets be in?
 - Long vs wide
 - Merged vs separate (across time, across forms)
- ☐ How will data and code be versioned?
 - Manually?
 - Is this plan written out in a Style Guide?
 - Programmatically?
 - What program will be used? (ex: Git/GitHub, SharePoint, Box)