10 Considerations for Data Analysis

Read Read data into R	Consider (1) the type of data, (2) how they are stored, and (3) the size of the dataset.
Format Format the data	Long or wide format? Plan the analysis keeping in mind what your needs are. Long: summarising, visualising. Wide: human-readable, tables.
Select Select portions of the dataset	Decide what data you'll need in advance from the dataset: subsets vs. filtered dataset. Consider the pros and cons of the approach.
Modify Create and modify variables	Consider: (1) what variables you need, (2) calculations you will perform, and (3) algorithms needed to create the variables.
Link Conduct data linkage(s)	Assign a key variable(s) that you will link datasets on. Consider the join type you will need: inner, left, right, full, semi, or anti join.
Analyse Analyse the data	Consider grouping data meaningfully in performing calculations and obtaining descriptive statistics. Evaluate outputs and interpret summaries.
Visualise Visualise the data	Use tables or charts, and consider the best way to display information. Often, the simpler the better. What are the data telling you?
Report Report the findings	Communicate the findings to those who need the information created from your work. Consider the audience you are communicating with.
Celebrate	Preferably with cake!
Review Review, revise, and re-do as needed	Often, we can go back and simplify/optimise. This is important especially if the code will be used again.

Introduction to R for Public Health Investigations
Training and Development Unit, PHAC
Winter/Spring, 2021