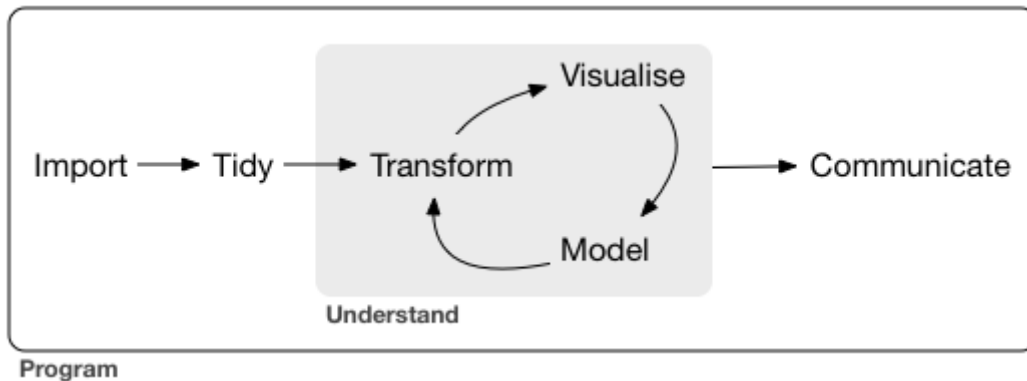


# Lab 1: Your Data in R

One of the best ways to learn a new computer language is to use it for a project that *matters to you*. Here at the beginning of the semester, you'll select a data analysis that you've done in the past and set it as a goal for something that you will be able to re-implement in Python by the end of the semester. The main guideline is that your analysis touches on as many of the steps of the data science cycle as possible.



Some further guidelines:

1. This analysis could come from a course project, a particularly challenging lab, a research project, or a fun hobby project that you've worked on. If you don't have any candidates from your own work that seem suited to the task, you can use an analysis that you've found elsewhere; Rpubs is a good place to look. Be sure to cite any work that is not your own.
2. The source file can either be an R Markdown document or a well-commented R script. The file that you will be submitting to gradescope will need to be a pdf.
3. In terms of length, 1 - 3 pages of R code (on a pdf) and potentially 1 - 5 pages of output would be appropriate, but your mileage may vary. It's possible you were able to do a rich data analysis in only a half page of code.
4. Before you compile to pdf, add very visible section labels (through comments or markdown) to demarcate which step of the data science cycle that span of code represents.
5. In terms of sophistication of the data analysis, you should be shooting for moderate. It could be fine to use an analysis that fits a very sophisticated statistical model, but know that that may be a component that you may need to leave in R (and pass objects to Python using reticulate).

This data analysis will be due as a pdf to gradescope by Sunday at 8 pm.