**Decisions I am making.**

* I think we want the final data to be entirely raw values merged together
  + Ie. we don’t create % private in the raw data, just present the data as it comes
* I think we want it to be panel data with long form years but and long form variables
* Where APIs are not available from Gov – we will publish the raw ODS, XL, CSV files on our github and clean it all directly from there.
  + UPDATE: struggling to read xlsx off github cos of my own bad tempfile skills.
    - But this is enough to convince me it isn’t worthwhile and we should publish all as CSVs and then clean from there.
    - Ultimately then there should be code to download and convert files to CSVs so that we can reproduce this process via code.
      * Even if no one in their right mind would reproduce it – and dockerise versions might struggle
* The final sheet will work by pulling different functions so that the reproducible file looks like eg.
  + Spend\_Data <- clean\_expenditure\_data()
  + Children\_Data <- clean\_SSDA\_data()
  + Housing\_Data <- clean\_homeless\_data()
  + Full\_Data <- merge\_all\_data(Spend\_Data, Children\_Data, Housing\_Data)
    - Magic
* We should include all data available at LA level.
  + Eg. education outcomes by LA
* I think we should keep only variables that are ongoing to this day
  + Ie. if there is a measure of educational attainment at key stage 1 that is used 2010-2013, we get rid of this.
    - There are a whole bunch of these and we will need to make a call about what is worthwhile and what isn’t – for the sake of usability.
      * Eg. one year they report grades by gender… is it worthwhile keeping this one measure? Not in my opinion.