01\_extract\_pid\_and\_econ

Jon Minton; Martin

## Aim

The aim of this document is to extract the PID, year, and economic activity status from the UKHLS.  
Later iterations will also incorporate predictor variables.

## Preparation

First we load some packages, including the economic\_inactivity package created as part of this project.

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.2 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
✔ purrr 1.0.1   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(haven)  
#library(economic\_inactivity)

Now we specify the base folder location containing all the UKHLS files we will want to access and iterate over.

base\_dir\_location <- "big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls"

Now we want to declare the columns that we want to extract.

Initially this will just be economic activity status and PID. But soon we will expand this to cover additional explanatory variables too.