04\_visualise\_sf12\_relationships

Jon Minton; Martin Taulbut

# Introduction

The aim of this notebook is to better understand how the categories of economic activity and inactivity change by age, sex and wave in the UKHLS sample.

This will help us with the following:

* Think about the right level of disaggregation of economic (in)activity status
* Think about the right age cutoffs
* Better understand differences by sex in the above

## Load prerequisites

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(here)

here() starts at /Users/JonMinton/repos/economic\_inactivity

devtools::load\_all(here('R'))

ℹ Loading economic\_inactivity

base\_dir\_location <- "big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls"  
indresp\_files <- dir(here(base\_dir\_location), pattern = "[a-z]\_indresp.dta", full.names = TRUE)  
  
varnames <- c(  
 "jbstat", "dvage", "sex"  
 )  
  
extract\_what <- c(  
 "labels", "values", "labels"  
 )  
  
overall\_start\_time = Sys.time()  
  
long\_slimmed\_datalist <- lapply(indresp\_files,   
 read\_and\_slim\_data,   
 varnames = varnames,   
 extract\_what = extract\_what,   
 verbose = TRUE  
)

[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/a\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.283735036849976 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.112229108810425 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/b\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.35404896736145 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.0993218421936035 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/c\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.554827928543091 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.099329948425293 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/d\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.410830020904541 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.0884819030761719 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/e\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.446614980697632 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.0881099700927734 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/f\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.412620067596436 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.154217004776001 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/g\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.483757972717285 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.080146074295044 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/h\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.371891021728516 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.138508081436157 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/i\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.473696947097778 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.071552038192749 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/j\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.351300001144409 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.0742568969726562 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/k\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.446302175521851 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.073218822479248 seconds"  
[1] "extracting file: /Users/JonMinton/repos/economic\_inactivity/big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls/l\_indresp.dta"  
[1] "Attempting to find 3 variables"  
[1] "All variables requested found"  
[1] "read file in 0.343468904495239 seconds"  
[1] "slimming file..."  
[1] "extracted 3 variables in 0.0643901824951172 seconds"

long\_slimmed\_data <- bind\_rows(long\_slimmed\_datalist)  
  
overall\_end\_time = Sys.time()  
  
print(paste(  
 "Overall process took",  
 difftime(overall\_end\_time, overall\_start\_time, units = "mins"),  
 "minutes"  
))

[1] "Overall process took 0.103559903303782 minutes"

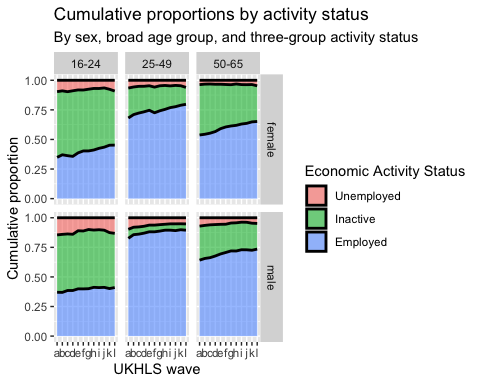
rm(long\_slimmed\_datalist)  
long\_slimmed\_data

# A tibble: 1,516,371 × 4  
 pidp wave variable value   
 <dbl> <chr> <chr> <chr>   
 1 68001367 a sex male   
 2 68004087 a sex male   
 3 68006127 a sex female  
 4 68006135 a sex female  
 5 68006807 a sex female  
 6 68007487 a sex female  
 7 68007491 a sex male   
 8 68007495 a sex male   
 9 68007499 a sex male   
10 68008167 a sex female  
# ℹ 1,516,361 more rows

econ\_act\_groups <- readxl::read\_excel(path = here("data/economic\_activities\_categories.xlsx"), sheet = 'categories') %>%   
 janitor::clean\_names()

Now to pivot this

long\_slimmed\_data %>%   
 pivot\_wider(names\_from = "variable", values\_from = "value") %>%   
 mutate(age = as.numeric(dvage)) %>%   
 mutate(age = ifelse(age < 0 , NA, age)) %>%   
 mutate(age\_group = case\_when(  
 between(age, 16, 24) ~ "16-24",  
 between(age, 25, 49) ~ "25-49",  
 between(age, 50, 65) ~ "50-65"  
 )) %>%   
 filter(!is.na(age\_group)) %>%   
 filter(sex %in% c("male", "female")) %>%   
 left\_join(  
 econ\_act\_groups %>% mutate(recoded = level\_2\_meso),   
 by = c("jbstat" = "original")  
 ) %>%   
 select(pidp, wave, sex, age, age\_group, econ\_act\_status = recoded) %>%   
 filter(econ\_act\_status != "Missing") %>%   
 group\_by(age\_group, sex, wave, econ\_act\_status) %>%   
 count() %>%   
 group\_by(age\_group, sex, wave) %>%   
 mutate(N = sum(n)) %>%   
 mutate(proportion = n / N) %>%   
 ggplot(aes(x=wave, y=proportion,   
 group = forcats::fct\_rev(econ\_act\_status),   
 fill = forcats::fct\_rev(econ\_act\_status)  
 )  
 ) +   
 geom\_area(alpha=0.6 , linewidth = 1, colour="black") +  
 facet\_grid(sex~age\_group) +  
 labs(  
 x = "UKHLS wave", y = "Cumulative proportion",  
 title = "Cumulative proportions by activity status",  
 subtitle = "By sex, broad age group, and three-group activity status",  
 fill = "Economic Activity Status"  
 )

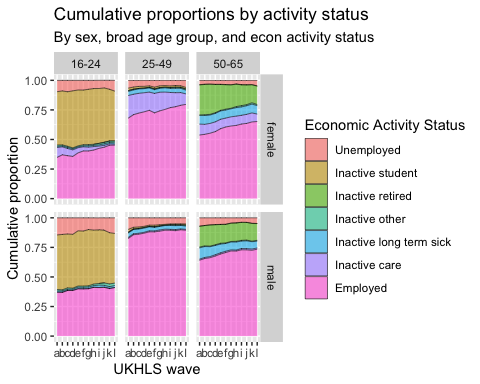


Some observations:

* Rising employment rates for women
* Rising employment rates for prime age men (but from high baseline)
* Inactivity greatest in young adult groups (Students?)
* Rates of decrease in inactivity appear to have slowed down in older working age men
* Not much increase in employment rates in young adult men, compared with steady falls over time for young adult women

We’ll now do the same but disaggregating the economic inactivity categories a bit more…

long\_slimmed\_data %>%   
 pivot\_wider(names\_from = "variable", values\_from = "value") %>%   
 mutate(age = as.numeric(dvage)) %>%   
 mutate(age = ifelse(age < 0 , NA, age)) %>%   
 mutate(age\_group = case\_when(  
 between(age, 16, 24) ~ "16-24",  
 between(age, 25, 49) ~ "25-49",  
 between(age, 50, 65) ~ "50-65"  
 )) %>%   
 filter(!is.na(age\_group)) %>%   
 filter(sex %in% c("male", "female")) %>%   
 left\_join(  
 econ\_act\_groups %>% mutate(recoded = level\_3),   
 by = c("jbstat" = "original")  
 ) %>%   
 select(pidp, wave, sex, age, age\_group, econ\_act\_status = recoded) %>%   
 filter(econ\_act\_status != "Missing") %>%   
 group\_by(age\_group, sex, wave, econ\_act\_status) %>%   
 count() %>%   
 group\_by(age\_group, sex, wave) %>%   
 mutate(N = sum(n)) %>%   
 mutate(proportion = n / N) %>%   
 ggplot(aes(x=wave, y=proportion,   
 group = forcats::fct\_rev(econ\_act\_status),   
 fill = forcats::fct\_rev(econ\_act\_status)  
 )  
 ) +   
 geom\_area(alpha=0.6 , linewidth = 0.2, colour="black") +  
 facet\_grid(sex~age\_group) +  
 labs(  
 x = "UKHLS wave", y = "Cumulative proportion",  
 title = "Cumulative proportions by activity status",  
 subtitle = "By sex, broad age group, and econ activity status",  
 fill = "Economic Activity Status"  
 )



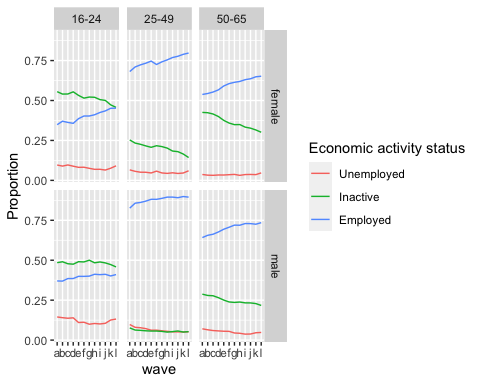
Some observations:

* Inactivity due to caring reduced for young women
* Inactivity due to caring reduced for prime age women, but at a lower rate than for young women
  + I suspect it’s reduced more in absolute terms for prime age women, but in relative terms for young women
* Inactivity due to early retirement also appears to have reduced for older women
* economic activity due to employment has increased for women in all age groups.
* for men it also appears to have increased, but at a slower rate, and perhaps not in the last few waves for prime age men
* there is an apparent increase in unemployment in young men and women in the last three waves
* inactivity due to long term sickness is highest in the oldest age group
  + it appears to be fairly constant in females but possibly falling in males (in the oldest age group)
* rates of inactivity due to long term sickness appear around 3-5% in prime age men and women, and similar in both sexes, as well as not to have changed appreciably over time

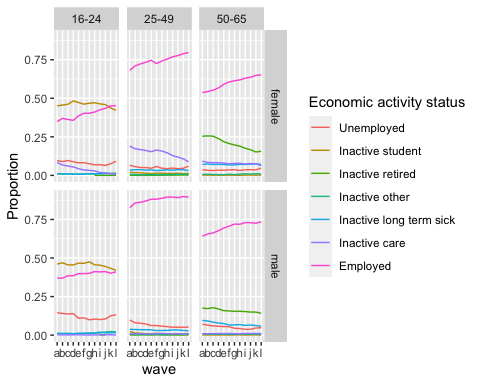
## line charts

We now want to show the same data but with the actual proportions rather than cumulative proportions

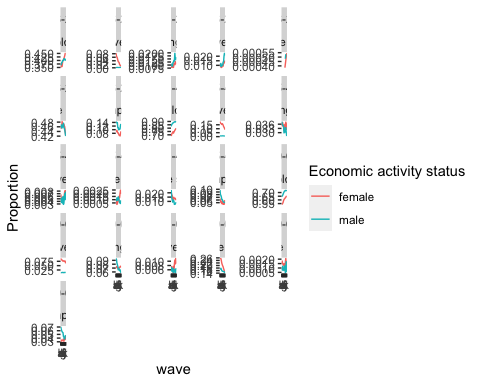
long\_slimmed\_data %>%   
 pivot\_wider(names\_from = "variable", values\_from = "value") %>%   
 mutate(age = as.numeric(dvage)) %>%   
 mutate(age = ifelse(age < 0 , NA, age)) %>%   
 mutate(age\_group = case\_when(  
 between(age, 16, 24) ~ "16-24",  
 between(age, 25, 49) ~ "25-49",  
 between(age, 50, 65) ~ "50-65"  
 )) %>%   
 filter(!is.na(age\_group)) %>%   
 filter(sex %in% c("male", "female")) %>%   
 left\_join(  
 econ\_act\_groups %>% mutate(recoded = level\_2\_meso),   
 by = c("jbstat" = "original")  
 ) %>%   
 select(pidp, wave, sex, age, age\_group, econ\_act\_status = recoded) %>%   
 filter(econ\_act\_status != "Missing") %>%   
 group\_by(age\_group, sex, wave, econ\_act\_status) %>%   
 count() %>%   
 group\_by(age\_group, sex, wave) %>%   
 mutate(N = sum(n)) %>%   
 mutate(proportion = n / N) %>%   
 ggplot(aes(x=wave, y=proportion,   
 group = forcats::fct\_rev(econ\_act\_status),   
 colour = forcats::fct\_rev(econ\_act\_status)  
 )  
 ) +   
 geom\_line() +  
 facet\_grid(sex~age\_group) +  
 labs(  
 x = "wave",  
 y = "Proportion",  
 colour = "Economic activity status"  
 )



long\_slimmed\_data %>%   
 pivot\_wider(names\_from = "variable", values\_from = "value") %>%   
 mutate(age = as.numeric(dvage)) %>%   
 mutate(age = ifelse(age < 0 , NA, age)) %>%   
 mutate(age\_group = case\_when(  
 between(age, 16, 24) ~ "16-24",  
 between(age, 25, 49) ~ "25-49",  
 between(age, 50, 65) ~ "50-65"  
 )) %>%   
 filter(!is.na(age\_group)) %>%   
 filter(sex %in% c("male", "female")) %>%   
 left\_join(  
 econ\_act\_groups %>% mutate(recoded = level\_3),   
 by = c("jbstat" = "original")  
 ) %>%   
 select(pidp, wave, sex, age, age\_group, econ\_act\_status = recoded) %>%   
 filter(econ\_act\_status != "Missing") %>%   
 group\_by(age\_group, sex, wave, econ\_act\_status) %>%   
 count() %>%   
 group\_by(age\_group, sex, wave) %>%   
 mutate(N = sum(n)) %>%   
 mutate(proportion = n / N) %>%   
 ggplot(aes(x=wave, y=proportion,   
 group = forcats::fct\_rev(econ\_act\_status),   
 colour = forcats::fct\_rev(econ\_act\_status)  
 )  
 ) +   
 geom\_line() +  
 facet\_grid(sex~age\_group) +  
 labs(  
 x = "wave",  
 y = "Proportion",  
 colour = "Economic activity status"  
 )



long\_slimmed\_data %>%   
 pivot\_wider(names\_from = "variable", values\_from = "value") %>%   
 mutate(age = as.numeric(dvage)) %>%   
 mutate(age = ifelse(age < 0 , NA, age)) %>%   
 mutate(age\_group = case\_when(  
 between(age, 16, 24) ~ "16-24",  
 between(age, 25, 49) ~ "25-49",  
 between(age, 50, 65) ~ "50-65"  
 )) %>%   
 filter(!is.na(age\_group)) %>%   
 filter(sex %in% c("male", "female")) %>%   
 left\_join(  
 econ\_act\_groups %>% mutate(recoded = level\_3),   
 by = c("jbstat" = "original")  
 ) %>%   
 select(pidp, wave, sex, age, age\_group, econ\_act\_status = recoded) %>%   
 filter(econ\_act\_status != "Missing") %>%   
 group\_by(age\_group, sex, wave, econ\_act\_status) %>%   
 count() %>%   
 group\_by(age\_group, sex, wave) %>%   
 mutate(N = sum(n)) %>%   
 mutate(proportion = n / N) %>%   
 ggplot(aes(x=wave, y=proportion,   
 group = sex,   
 colour = sex  
 )  
 ) +   
 geom\_line() +  
 facet\_wrap(~ age\_group + econ\_act\_status, scales = "free\_y") +  
 labs(  
 x = "wave",  
 y = "Proportion",  
 colour = "Economic activity status"  
 )



Some observations

* young adults (16-24)
  + rates of employment in young adults have continued to increase for young women but stalled for young men after around wave h (2016-2017). Rates of employment in this age group are now higher for women than men
  + rates of inactivity due to caring declined from around 8% to about 1%, whereas rates of inactive care for men in this age group have consistently been around negligible in comparison.
  + rates of inactivity due to long-term sickness in this young adult age group are consistently low for both sexes, but there are indications of an increasing trend, and slightly higher rates for males than females
  + rates of inactivity for other reasons are low in this age group, by have increased from around 1% to 1.5% for women, and from around 1% to around 2% for men.
  + rates of inactivity in young adults due to being a full time student were similar for both sexes, steady at around 46-47%, up until around wave h (2016-17). Afterwards they fell for both sexes year on year, reaching around 42% by wave l.
  + rates of unemployment have in this age group have always been higher for males than females, starting around 14% for males and slightly below 10% for females. Rates of female unemployment continually fell for females up until wave j, reaching around 7%, then increased back to around 9%. For males rates of unemployment are more u-shaped, declining to around 10-11% between waves f and j, then increasing back to around 13% in the most recent wave.
* prime age adults
  + rates of employment are higher for men than women, but the gap has been closing. This is due to rates of employment stalling at around 89-90% for men from about wave h, whereas continuing to increase for women over the whole series of waves, from under 70% in wave a, to around 80% in wave l.
    - there was a possible dip in employment for females around wave e (around 2013-14), but generally rates have continued to increase. (It’s possible this dip may be in part due to welfare reform, including policies/changes like benefits sanctions which peaked around this same period, with many lone parents being moved onto JSA)
  + Rates of inactivity due to caring responsibilities have consistently been low for men, and declining over time for women, starting at around 16% in wave a, and declining to around 9.5% by wave l.
    - Rates for women increased slightly around wave f, i.e. the same wave where rates of prime age female emplotment temporarily fell.
    - It appears for females in this age group that rates of employment and inactivity for caring responsibilities are negatively correlated. (it might be worth exploring whether attempts to move lone parents to JSA temporarily led to increases in caring inactivity rather than employment, i.e. had what might be the opposite effect to that expected?!)
  + inactive long term sick: rates are around 3-3.6% throughout the period. Rates appear slightly more consistent at around 3.5% for women, whereas there are indications that they may have fallen for men, to below 3% in the last period.
  + Rates of inactivity due to retirement are very low for both sexes, at around 0.1-0.2%. It may be that they are increasing for women.
  + unemployment rates have been falling for both sexes, are consistently higher for men except in the last wave. For men they fell from around 10% to around 5%, with the rate of decline decreasing over time. For females they have been more stable at around 5-6%. In the most recent wave rates of unemployment for women were higher than for wave, increasing comparatively sharply in the last wave
    - In wave e-f the rates for women temporarily increased, possibly consistent with other changes in other categories in this age group
* 50-65
  + rates of employment are higher for males than females, and increasing over time. However the rates of increase slowed for both sexes, though more so for men. They changed from around 65% to around 72% for men, and from around 54% to around 65% for women, from the start to the end of the period.
  + rates of inactivity due to caring responsibilities are much higher for women than for men, and have been fairly stable at around 7.5%. There are indications that the general tendency has been for reducing rates for women, especially in the first and last wave.
  + inactive long term sick: rates were higher for men at the start of the period, then higher for women at the end. This has been because rates fell, from around 9% to around 6% for men, whereas they remained stable at around 7% for women.
  + Rates of inactive other appear too small to make reasonable inferences about
  + Rates of inactivity due to retirement fell for both sexes, especially women. For women they fell from around 25% at the start to around 16% at the end. More modest falls were seen for men, from around 17.% at the start of the period to around 14% at the end
  + Rates of inactive due to being a student appear too low to make reasonable inferences about
  + Rates of unemployment are higher for men than women, but the gap reduced markedly over time. This is because they tended to fall, from around 7% at the start to around 4.5% at the end, reaching a nadir of around 4% in wave i-j. Whereas for women they were stable at around 3.5% for most of the period, though possibly increasing slowly throughout the period, before rising much more quickly, to around 4.5%, in the last wave