

# DA3 Exercise1

Peter Kaiser

2022 01 23

```
# setup and import
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.5      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

df <- read_csv('https://osf.io/4ay9x/download')

## New names:
## * ' ' -> ...1

## Rows: 149316 Columns: 23

## -- Column specification -----
## Delimiter: ","
## chr  (9): intmonth, stfips, prcitshp, state, ind02, class, unionmme, unionco...
## dbl  (14): ...1, hhid, weight, earnwke, uhours, grade92, race, ethnic, age, s...

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

#str(df)
#summary(df)

# select columns

df <- df %>% select(occ2012,earnwke,uhours,grade92,race,age,sex,state,unionmme)

na_count <- sum(is.na(df))

df <- df %>% filter(occ2012==9120 | occ2012==9130 | occ2012==9140)
```