GDA of *disabData* (from **addhaz**)

Background Information on disability and chronic conditions from the 2013 National Health Survey in Brazil. Only women aged 60 or older are included.

Aims How are chronic conditions and old age related to disability?

Source Instituto Brasileiro de Geografia e Estatistica (IBGE) Pesquisa Nacional de Saude 2013

Structure 6294 observations on 7 variables (1 discrete variable, 5 binary variables, 1 weight variable)

The data are recorded numerically, so the variables have first been recoded to make the labels more understandable. The categories for the two disability variables have additionally been reordered for better display possibilities later on.

```
data(disabData, package="addhaz")

library(dplyr)
disab <- disabData %>% mutate(Disb = factor(dis.bin), Dism = factor(dis.mult),
        Age = factor(age), Diab = factor(diab), Arth = factor(arth), Stro = factor(stro))

disab <- within(disab, levels(Disb) <- c("no", "disability"))
disab <- within(disab, Disb <- factor(Disb, levels = c("disability", "no")))

disab <- within(disab, levels(Dism) <- c("no", "mild", "severe"))
disab <- within(disab, Dism <- factor(Dism, levels = c("severe", "mild", "no")))

disab <- within(disab, levels(Age) <- c("60to79", "over80"))
disab <- within(disab, levels(Diab) <- c("no", "diabetes"))
disab <- within(disab, levels(Arth) <- c("no", "arthritis"))
disab <- within(disab, levels(Stro) <- c("no", "stroke"))
```

The distributions of the two disability variables are shown in Figure 1.

```
p1 <- ggplot(disab, aes(Disb)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
p2 <- ggplot(disab, aes(Dism)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
library(gridExtra)
grid.arrange(p1, p2, nrow=1, widths=c(42,58))
```

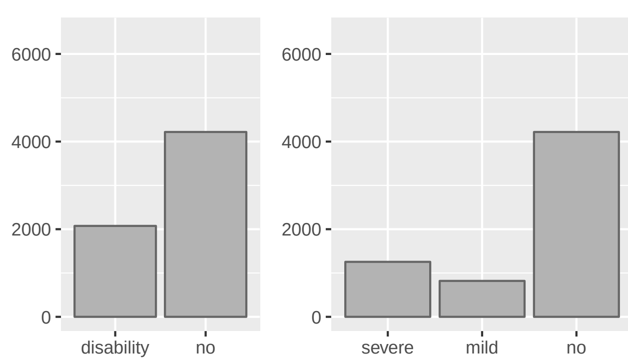


Fig 1: About one third of the survey have a disability. Of those the majority have a severe disability.

The distributions of the other four main variables are shown in Figure 2.

```r
p3 <- ggplot(disab, aes(Age)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
p4 <- ggplot(disab, aes(Diab)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
p5 <- ggplot(disab, aes(Arth)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
p6 <- ggplot(disab, aes(Stro)) + geom_bar() + labs(x=NULL, y=NULL) + ylim(0,6500)
grid.arrange(p4, p5, p6, p3, nrow=1)
```

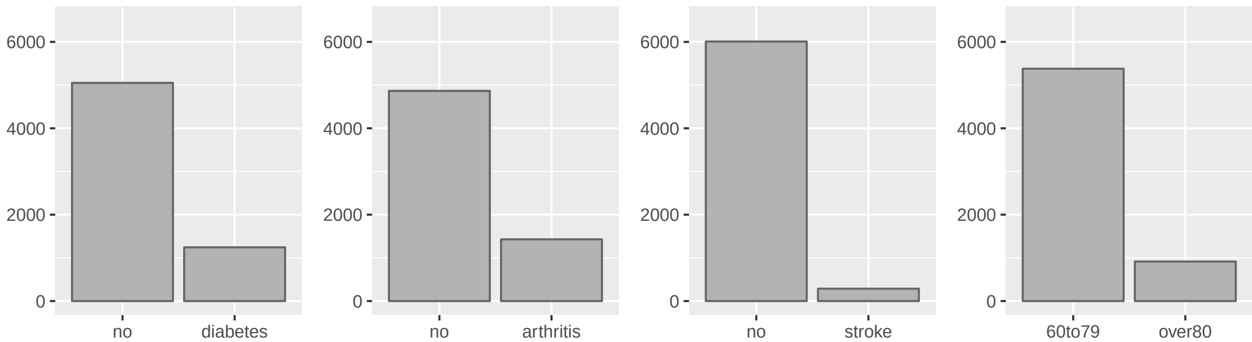
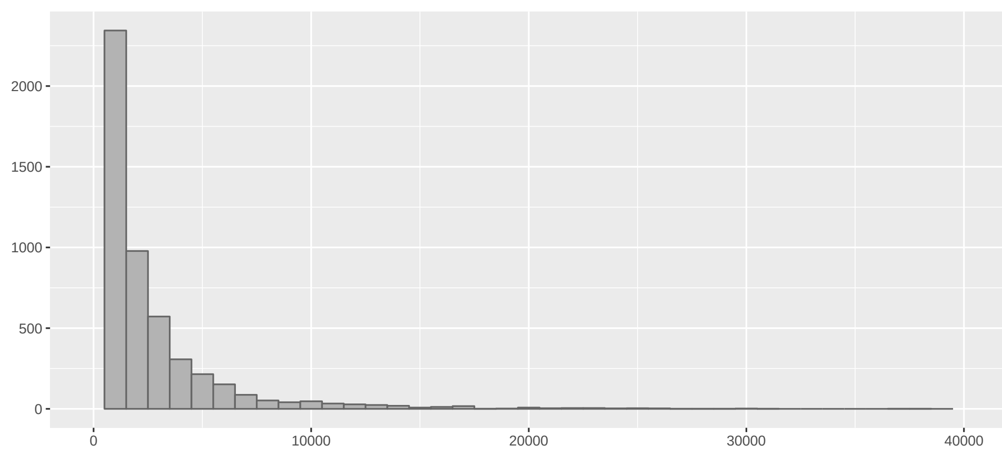


Fig 2: Arthritis and diabetes occur fairly often, evidence of a stroke far less often. Most of the study participants are aged 60 to 79.

Each case has a survey weight. A histogram shows the distribution of weights and a boxplot emphasises the extreme values.

```r
ggplot(disab, aes(wgt)) + geom_histogram(binwidth=1000) + labs(x=NULL, y=NULL) +
  xlim(0, 40000)
```




```r
ggplot(disab, aes("wgt", wgt)) + geom_boxplot()  + labs(x=" ", y=NULL) +
  ylim(0, 40000) + coord_flip()
```

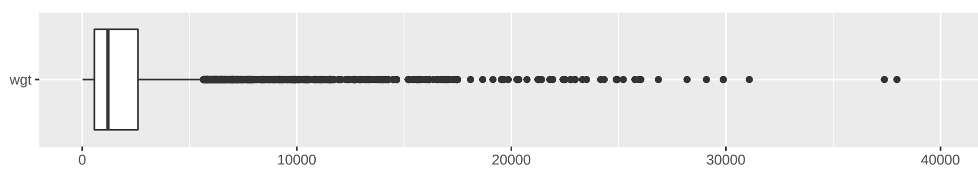


Fig 3: The distribution of survey weights is highly skewed, going from close to 0 (in fact 18.6) to almost 40,000.

Associations between disability and the other variables can be examined using spineplots. A column's width represents the number of cases in the group and the proportion in red represents the disabled cases in that group. Note that the survey weights (wgt) have been used to draw these plots.

```r
plotbar <- function(df, xv, yv) {
  df$xv <- df[, xv]
  df$yv <- df[, yv]
  ggplot(df %>% group_by(xv, yv) %>% summarise(freq = sum(wgt)) %>% group_by(xv) %>%
           mutate(weight=sum(freq))) + geom_bar(aes(xv, freq/weight, fill = yv,
           width = weight/max(weight)), stat = 'identity', position = 'fill') +
           labs(x=NULL, y=NULL, fill="") + theme(legend.position="bottom")
}
s1 <- plotbar(disab, "Diab", "Disb")
s2 <- plotbar(disab, "Arth", "Disb")
s3 <- plotbar(disab, "Stro", "Disb")
s4 <- plotbar(disab, "Age", "Disb")
grid.arrange(s1, s2, s3, s4, nrow=1)
```

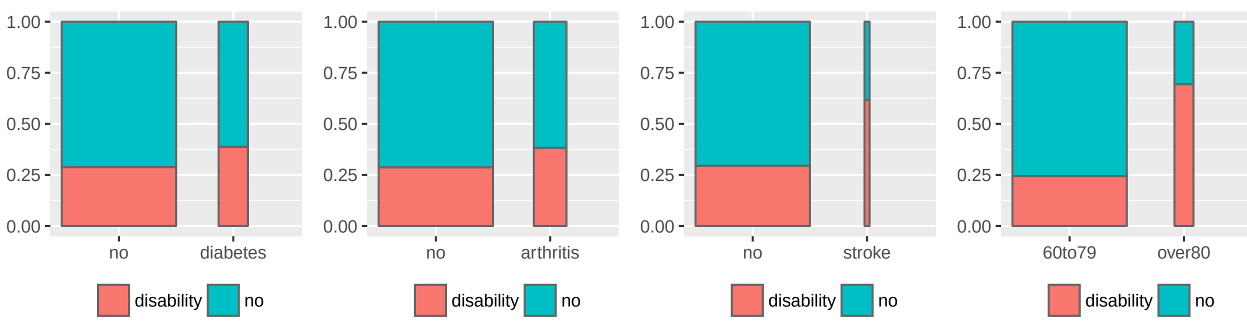


Fig 4: The biggest differences in the groups are for age and disability and for stroke and disability.

Before looking at multivariate effects it is worth checking how big the various groups are.

```r
disab <- within(disab, disb <- factor(Disb, levels = c("no", "disability")))
disab <- within(disab, levels(disb) <- c("no", "disab"))
ggplot(disab, aes(disb)) + geom_bar() + facet_grid(.~Stro+Age+Diab+Arth) + labs(x=NULL, y=NULL)
```

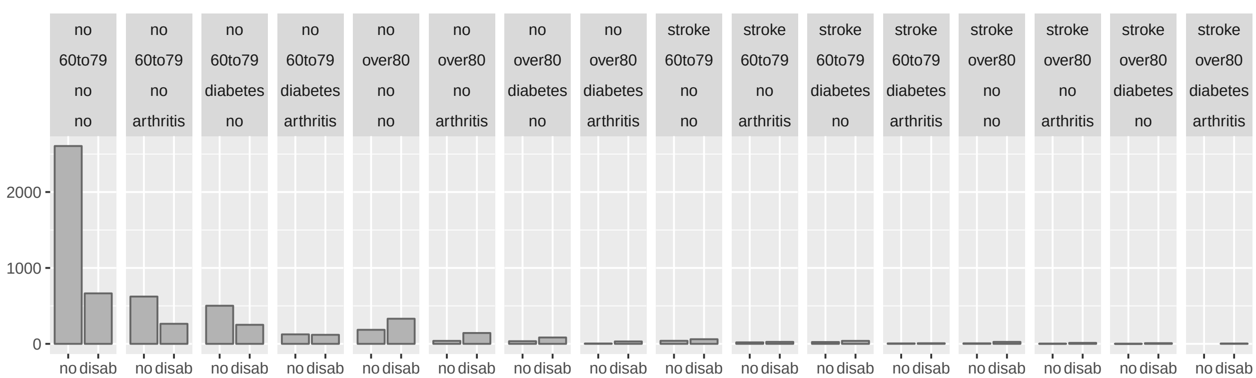


Fig 5: The largest group by far (about 40% of the cases) is the younger women with none of the health issues in question. In the first four groups, younger women with arthritis and/or diabetes or neither, there are fewer disabled than without a disability. In all the other groups there are more disabled.

Looking at more than one variable together with disability requires a choice of variables and a choice of variable ordering. Figure 6 excludes strokes (because there are not many cases with stroke) and uses an ordering of age, arthritis, and diabetes (because age has the strongest association on its own).

```
mvspine <- function(df, x1, x2, x3, y) {
  df$x1 <- df[, x1]
  df$x2 <- df[, x2]
  df$x3 <- df[, x3]
  df$y <- df[, y]
  ggplot(df %>% group_by(x1, x2, x3, y) %>% summarise(freq = sum(wgt)) %>% group_by(x2) %>%
           mutate(weight=sum(freq))) + geom_bar(aes(x2, freq/weight, fill = y,
           width = weight/max(weight)), stat = 'identity', position = 'fill') +
    facet_grid(.~x1 + x3) + labs(x=NULL, y=NULL, fill="") + theme(legend.position="bottom")
}
mvspine(disab, "Age", "Arth", "Diab", "Disb")
```

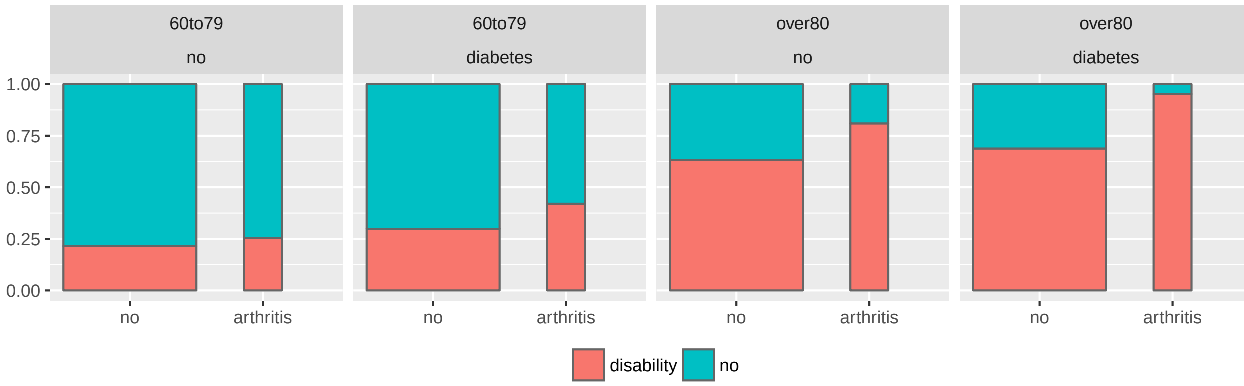


Fig 6: Rates of disability are higher for arthritis cases than cases with no arthritis in every age group, for both diabetics and non-diabetics. The rates for each group of the patients over 80 are higher than for every group of the younger patients. Younger diabetic women are more likely to be disabled whether they have arthritis or not, whereas this is not quite true for the older patients. (Survey weights have been used.)

A display distinguishing between mild and severe disability gives similar results.

```
mvspine(disab, "Age", "Arth", "Diab", "Dism")
```

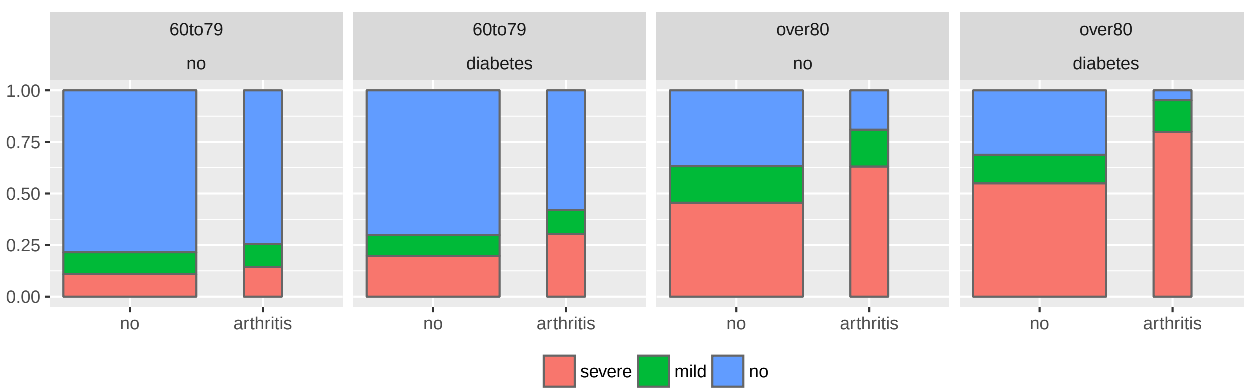


Fig 7: Conclusions are very similar to those for Figure 5.

The **addhaz** package help pages include some modelling of these data using binomial and multinomial additive hazard models.