Workflow: Physical Activity and Well-Being

Sofia Gil-Clavel

- Open the Data

```
#### Opening the Data ####
DATA<-read.table("Replicability/DATA/cchs2012_mh.tab",sep="\t",header = TRUE)
```

- From the variable "PHSGAPA" (Moderate/vigorous physical activity -average/hours) drop those individuals whose content is "NOT STATED".

```
# The variable of interest is "PHSGAPA" 99.99 means "NOT STATED"
DATA<-DATA%>%filter(PHSGAPA<99)
```

- Transform the variable sex (DHH_SEX) into factor

```
# Sex as Factor
DATA$DHH_SEX<-factor(DATA$DHH_SEX,
                     levels = c(1,2),
                     labels = c("MALE","FEMALE"))
```

- Create the variable "GROUP" by transforming "PHSGAPA" into the next categories:
  - "NOT" if they do 0 hours of exercise
  - "0%-25%" for those in the first quantile without considering those that do 0 hours of exercise.
  - "25-50" for those in the second quantile without considering those that do 0 hours of exercise.
  - "50-75" for those in the third quantile without considering those that do 0 hours of exercise.
  - "75-100" for those in the fourth quantile without considering those that do 0 hours of exercise.

```
#### Those that do not exercise
NOTex<-DATA%>%filter(PHSGAPA==0)

NOTex$GROUPS<-"NOT"

#### Those that DO exercise
DOex<-DATA%>%filter(PHSGAPA>0)
# hist(DOex$PHSGAPA,freq = FALSE)

DOex$GROUPS <- with(DOex, cut(PHSGAPA,
                      breaks = qu <- quantile(PHSGAPA),
                      labels = paste0(names(qu)[-5],"-",names(qu)[-1]),
                      include.lowest=TRUE))

#### Join
ALL<-rbind(NOTex,DOex)
```

- Create a table that shows the percentage of population in each "GROUP" by age-groups breaking down by sex.

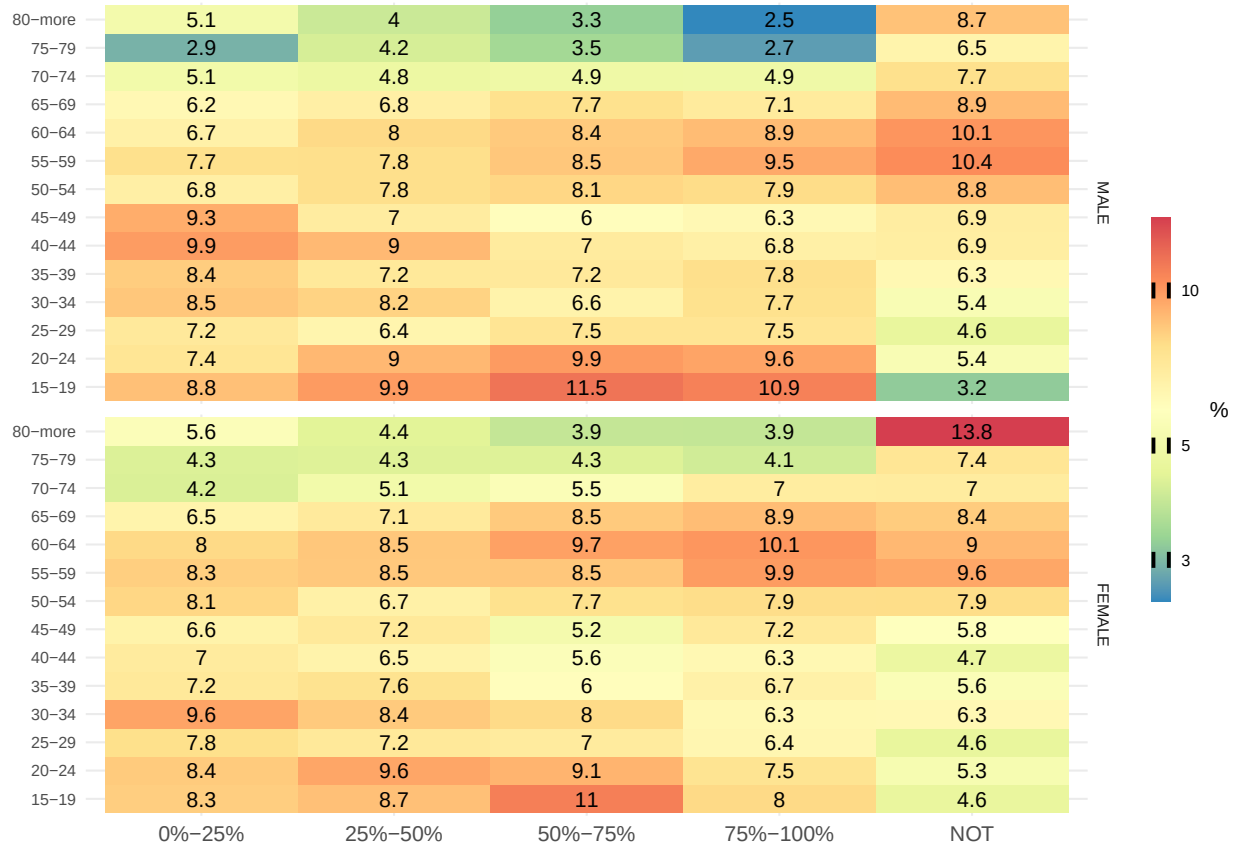
```
#### Some Basic Demogs ####
# DHH_SEX: MALE(1) & FEMALE(2)
# DHHGAGE: Age - grouped variable
```

```r
A1<-seq(15,80,5)
A2<-A1+4; A2[length(A2)]="more"
Alabels<-paste0(A1,"-",A2)

ALL$DHHGAGE<-factor(ALL$DHHGAGE,
                    levels = 1:14,
                    labels = Alabels)

TABLE=ALL%>%
  group_by(GROUPS,DHHGAGE,DHH_SEX)%>%
  count(DHHGAGE,DHH_SEX)%>%
  ungroup()%>%
  group_by(GROUPS,DHH_SEX)%>%
  mutate(perc=100*n/sum(n))# %>%
# select(-n)%>%
# spread(GROUPS,perc)
```

```r
(TABLE%>%
    ggplot() +
    geom_tile(aes(GROUPS,DHHGAGE,fill= perc))+
    geom_text(aes(GROUPS,DHHGAGE,label = round(perc, 1)),size=2.8) +
    facet_grid(DHH_SEX~.)+
    theme_minimal(base_size = 8)+
    theme(axis.text.x = element_text(hjust = 0.5,size = 8),
          axis.title = element_blank(),
          legend.position= "right",
          plot.caption = element_text(hjust =1,size = 8))+
    scale_fill_distiller(palette = "Spectral",trans = 'log10')+
    guides(fill=guide_colorbar(title = "%",
                               title.position = "right",
                               title.hjust = 0.5,
                               barwidth = 0.5, barheight = 10,
                               ticks.colour = "black",
                               ticks.linewidth = 8,ticks = TRUE)))
```



- Transform to NA all the values in "GEN_02A2","GEN_04","GEN_07", and "GEN_10" that correspond to responds: "DON'T KNOW", "REFUSAL", and "NOT STATED".

```
VARIABLES<-c("GEN_02A2","GEN_04","GEN_07","GEN_10")

DATA2<-DATA%>%
  select(c("PHSGAPA","DHH_SEX","DHHGAGE",
           "GEN_02A2","GEN_04","GEN_07","GEN_10","GENDHDI"))

#### Transform "NOT STATED" into NA ####
# GEN_02A2: Satisfaction with life in general - 11 point scale GEN 11
DATA2<-DATA2%>%mutate(GEN_02A2=ifelse(GEN_02A2<97,GEN_02A2,NA))
# GEN_04: Frequency - trouble sleeping GEN 10
DATA2<-DATA2%>%mutate(GEN_04=ifelse(GEN_04<7,GEN_04,NA))
# GEN_07: Perceived life stress GEN 12
DATA2<-DATA2%>%mutate(GEN_07=ifelse(GEN_07<7,GEN_07,NA))
# GEN_10: Sense of belonging to local community GEN 14
DATA2<-DATA2%>%mutate(GEN_10=ifelse(GEN_10<7,GEN_10,NA))
# GENDHDI: Perceived health - (D) GEN 14
DATA2<-DATA2%>%mutate(GENDHDI=ifelse(GENDHDI<9,GENDHDI,NA))
```

- Run a linear model where the outcome is "PHSGAPA" and the independent variables are "DHH_SEX","DHHGAGE", "GEN_02A2","GEN_04","GEN_07","GEN_10", and "GENDHDI".

```
#### Running simple linear model ####
fit<-lm(PHSGAPA~.,data = DATA2)
```

```
SUM<-summary(fit)
```

- Report the results as a table (Table 1) with the coefficients and their 95% confidence intervals, standard errors, p-values, and the standard p-value< '*' coding. Also rename the variables into more descriptive names:
  - DHH_SEXFEMALE: "Sex:Female"
  - DHHGAGE: "Age"
  - GEN_02A2: "Satif. Life"
  - GEN_04: "Trob. Sleep."
  - GEN_07: "Life Stress"
  - GEN_10: "Sense Belong."
  - GENDHDI: "Health"

```
Coef<-coef(fit)

# Confidence Interval
conf<-confint(fit)
CI1<-data.frame(lower=round(conf[,1],4),
                mean=round(Coef,4),
                upper=round(conf[,2],4))
# SE
CI1$SE<-SUM$coefficients[,2]
# p-values
p <- SUM$coefficients[,4]
CI1$`p-value`=round(p,digits = 4)
CI1$NAMES=row.names(CI1)

CI<-rbind(CI1)
CI$`p-value<`<-c("***","**","*",".","  ")[findInterval(CI$`p-value` ,
                                        sort(c(1, 0.1, 0.05, 0.01, 0.001, 0))) ]

CI<-CI[-1,]

CI$NAMES<-factor(CI$NAMES,
                 levels = rev(CI$NAMES),
                 labels = rev(c("Sex:Female","Age","Satif. Life","Trob. Sleep.",
                           "Life Stress","Sense Belong.","Health")))

CI<-CI[,c(6,1:5,7)]

row.names(CI)<-NULL
```

```
kableExtra::kbl(CI, booktabs = T, caption="Results Liner Model") #%>%
```

```
  # kableExtra::kable_styling(position = "center")
```

- Plot the results

```
#### Plotting the result ####
ggplot(CI, aes(x=NAMES, y=mean,label=`p-value<`)) +
   geom_errorbar(aes(ymin=lower, ymax=upper), width=.5,position = position_dodge(1)) +
   geom_point(size=2,position = position_dodge(1))+
   coord_flip()+
   scale_x_discrete(labels = levels(CI$NAMES), breaks=levels(CI$NAMES),
                    expand = c(0.1,0.1))+
```

Table 1: Results Liner Model

NAMES	lower	mean	upper	SE	p-value	p-value<
Sex:Female	-1.3546	-1.2545	-1.1544	0.0510510	0.0000	***
Age	-0.0941	-0.0808	-0.0675	0.0067900	0.0000	***
Satif. Life	0.0432	0.0770	0.1109	0.0172717	0.0000	***
Trob. Sleep.	-0.0220	0.0216	0.0651	0.0222102	0.3314	
Life Stress	-0.0938	-0.0395	0.0147	0.0276776	0.1531	
Sense Belong.	-0.2799	-0.2213	-0.1626	0.0299329	0.0000	***
Health	0.5083	0.5639	0.6194	0.0283423	0.0000	***

```
geom_hline(yintercept = 0, linetype="dashed",color="red")+
theme(panel.spacing = unit(1, "lines"),
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      legend.position = "top",text = element_text(size = 10),
      panel.background = element_rect(fill = "white"),
      panel.grid.major.y = element_line(size = 0.25,
                                        linetype = 'dashed',
                                        colour = "gray"),
      panel.grid.major.x = element_line(size = 0.25,
                                        linetype = 'dashed',
                                        colour = "gray"),
      strip.background = element_rect(color = NA,
                                      fill = NA, size = 1),
      axis.line = element_line(color = 'black'),
      plot.caption = element_text(hjust = 0.5,size = 10))+
geom_text(hjust=0, vjust=0,size=8,position = position_dodge(1))
```