R script Thesis

Rik Lubbers

# Loading packages

library(tidyverse)  
library(haven)  
library(lattice)  
library(data.table)  
library(knitr)  
library(survey)  
 options(survey.lonely.psu="adjust")  
library(expss)  
library(labelled)  
library(table1)  
library(readxl)  
library(GGally)  
library(kableExtra)  
library(xtable)  
library(modelsummary)  
library(performance)  
library(finalfit)  
library(naniar)  
library(rms)  
library(openxlsx)  
library(rcompanion)  
library(questionr)  
library(flextable)  
library(MASS)  
library(car)  
library(cowplot)  
library(heatmaply)  
library(esquisse)  
library(visreg)  
library(forcats)   
library(mosaic)  
library(DT)  
library(mice)  
library(gtsummary)

# Data handling

#### Loading data and creating subset

DHS\_children\_sub <- read\_sav("D:/Onedrive\_/OneDrive/THESIS RUG/DATA/Uganda Full/DHS Clusters/UGKR7BSV\_Children/UGKR7BFL.SAV") %>% dplyr::select(CASEID,  
 V001, V005,  
 V023, V024,  
 V149, V729,  
 V191, V212,  
 V445, V714,  
 BIDX, B0, B4,  
 B5, BORD,  
 HW5, HW8,  
 HW11, M19,  
 B11, H11,  
 V113, M5,  
 V136,  
 V481, H22,  
 V505, H7)  
  
#Add nearest facility in meters  
  
Travel\_time\_clust = read\_xls("D:/Onedrive\_/OneDrive/THESIS RUG/DATA/Uganda Full/Analyses/Travel\_time.xls", col\_names = T)  
  
DHS\_children\_sub = merge(DHS\_children\_sub, Travel\_time\_clust, by.x = "V001", by.y = "DHSCLUST")  
  
#Convert to KM  
DHS\_children\_sub$NEAR\_DIST = DHS\_children\_sub$NEAR\_DIST/1000

#### Removing unnecessary SPSS metadata

#REMOVING UNNECESSARY SPSS METADATA  
DHS\_children\_sub <- zap\_formats(DHS\_children\_sub)  
DHS\_children\_sub <- zap\_widths(DHS\_children\_sub)  
DHS\_children\_sub <- zap\_labels(DHS\_children\_sub)  
DHS\_children\_sub <- zap\_label(DHS\_children\_sub)

#### Defining missing data

DHS\_children\_sub$HW5[DHS\_children\_sub$HW5==9998] = NA  
DHS\_children\_sub$HW5[DHS\_children\_sub$HW5==9999] = NA  
  
DHS\_children\_sub$HW8[DHS\_children\_sub$HW8==9998] = NA  
DHS\_children\_sub$HW8[DHS\_children\_sub$HW8==9999] = NA  
  
DHS\_children\_sub$HW11[DHS\_children\_sub$HW11==9998] = NA  
DHS\_children\_sub$HW11[DHS\_children\_sub$HW11==9999] = NA  
  
DHS\_children\_sub$V149[DHS\_children\_sub$V149==9] = NA  
  
DHS\_children\_sub$V149[DHS\_children\_sub$V149==9] = NA  
  
DHS\_children\_sub$V729[DHS\_children\_sub$V729==9] = NA  
  
DHS\_children\_sub$V445[DHS\_children\_sub$V445==9998] = NA  
DHS\_children\_sub$V445[DHS\_children\_sub$V445==9999] = NA  
  
DHS\_children\_sub$V714[DHS\_children\_sub$V714==9] = NA  
  
DHS\_children\_sub$M19[DHS\_children\_sub$M19==9996] = NA  
DHS\_children\_sub$M19[DHS\_children\_sub$M19==9998] = NA  
DHS\_children\_sub$M19[DHS\_children\_sub$M19==9999] = NA  
  
DHS\_children\_sub$H11[DHS\_children\_sub$H11==8] = NA  
DHS\_children\_sub$H11[DHS\_children\_sub$H11==9] = NA  
  
DHS\_children\_sub$M5[DHS\_children\_sub$M5==93] = NA  
DHS\_children\_sub$M5[DHS\_children\_sub$M5==94] = NA  
DHS\_children\_sub$M5[DHS\_children\_sub$M5==97] = NA  
DHS\_children\_sub$M5[DHS\_children\_sub$M5==98] = NA  
DHS\_children\_sub$M5[DHS\_children\_sub$M5==99] = NA  
  
DHS\_children\_sub$V113[DHS\_children\_sub$V113==99] = NA  
DHS\_children\_sub$V113[DHS\_children\_sub$V113==97] = NA  
DHS\_children\_sub$V113[DHS\_children\_sub$V113==96] = NA  
  
DHS\_children\_sub$H7[DHS\_children\_sub$H7==8] = NA  
DHS\_children\_sub$H7[DHS\_children\_sub$H7==9] = NA  
  
DHS\_children\_sub$V481[DHS\_children\_sub$V481==9] = NA  
  
DHS\_children\_sub$H22[DHS\_children\_sub$H22==8] = NA  
DHS\_children\_sub$H22[DHS\_children\_sub$H22==9] = NA  
  
DHS\_children\_sub$V505[DHS\_children\_sub$V505==98] = NA  
DHS\_children\_sub$V505[DHS\_children\_sub$V505==99] = NA

#### Recoding variables

#Birth interval from continuous in dichotomous with more or less than 24 months  
DHS\_children\_sub$B11 = ifelse(DHS\_children\_sub$B11 >=24, 0, 1)  
  
#Combined diarrhea last 24 hours and 2 weeks in 1 category  
DHS\_children\_sub$H11 = ifelse(DHS\_children\_sub$H11 == 0, 0, 1)  
  
#Recode Source of Drinking water in simplified categories  
DHS\_children\_sub$V113 = expss::recode(DHS\_children\_sub$V113, 11 %thru% 14 ~ 1, 21 ~ 2, 31 %thru% 32 ~ 3, 41 %thru% 43 ~ 4, 51 ~ 5, 63 ~ 6, 71 ~ 6)  
  
#Make factor of both  
DHS\_children\_sub$B11 = factor(DHS\_children\_sub$B11,  
 levels = c(0,1),  
 labels = c("Birth interval more than 24 months",  
 "Birth interval less than 24 months"))  
  
DHS\_children\_sub$H11 = factor(DHS\_children\_sub$H11,  
 levels = c(0,1),  
 labels = c("No diarrhea in the last 2 weeks",   
 "Had diarrhea in the last 2 weeks"))  
  
#Recode sex of child   
DHS\_children\_sub$B4 = ifelse(DHS\_children\_sub$B4==1, 0, 1) %>% factor(levels = c(0, 1), labels = c("Male","Female"))  
  
#Make factor of twin  
DHS\_children\_sub$B0 = factor(DHS\_children\_sub$B0, levels = c(0:3),  
 labels = c("Single Birth", "1st of multiple", "2nd of multiple", "3rd of multiple"))  
  
#Recode DPT 3 into yes and no answers and make factor out of it  
DHS\_children\_sub$H7 = ifelse(DHS\_children\_sub$H7 == 0, 0, 1)  
  
DHS\_children\_sub$H7 = factor(DHS\_children\_sub$H7,   
 levels = c(0,1),   
 labels = c("No have not received",   
 "Yes, received DPT 3"))  
  
  
#Set other variables as factors  
DHS\_children\_sub$V149 = factor(DHS\_children\_sub$V149,   
 levels = c(0,1,2,3,4,5),   
 labels = c("No education",   
 "Incomplete Primary",   
 "Complete primary",   
 "Incomplete secondary",   
 "Complete secondary",   
 "Higher"))  
  
DHS\_children\_sub$V729 = factor(DHS\_children\_sub$V729,   
 levels = c(0,1,2,3,4,5),   
 labels = c("No education",   
 "Incomplete Primary",   
 "Complete primary",   
 "Incomplete secondary",   
 "Complete secondary",   
 "Higher"))  
  
DHS\_children\_sub$V714 = factor(DHS\_children\_sub$V714,   
 levels = c(0,1),   
 labels = c("Not working",  
 "Currently working"))  
  
DHS\_children\_sub$V113 = factor(DHS\_children\_sub$V113,   
 levels = c(1, 2, 3, 4, 5, 6),   
 labels = c("Piped Water",  
 "Tube Well Water",  
 "Dug Well (Open/Protected",  
 "Surface from Spring",  
 "Rainwater",  
 "Other"))  
  
DHS\_children\_sub$V481 = factor(DHS\_children\_sub$V481,   
 levels = c(0,1),   
 labels = c("No",  
 "Yes"))  
  
DHS\_children\_sub$H22 = factor(DHS\_children\_sub$H22,   
 levels = c(0,1),   
 labels = c("No",  
 "Yes"))  
  
#Changing stunting, wasting, underweight, BMI mother and wealth index variables to interpretative numbers  
DHS\_children\_sub$HW5 = DHS\_children\_sub$HW5/100  
DHS\_children\_sub$HW8 = DHS\_children\_sub$HW8/100  
DHS\_children\_sub$HW11 = DHS\_children\_sub$HW11/100  
  
DHS\_children\_sub$V445 = DHS\_children\_sub$V445/100  
DHS\_children\_sub$V191 = DHS\_children\_sub$V191/1000  
  
#Creating categorical variables for stunting, wasting, underweight, severe acute malnutrition and region,  
DHS\_children\_sub$HW5\_cat = ifelse(DHS\_children\_sub$HW5 >=-2, 0, 1) %>%   
 factor(levels = c(0,1),   
 labels = c("Not stunted",  
 "Stunted"))  
  
  
DHS\_children\_sub$HW8\_cat = ifelse(DHS\_children\_sub$HW8 >=-2, 0, 1) %>%   
 factor(levels = c(0,1),   
 labels = c("Not underweight",  
 "Underweight"))  
  
DHS\_children\_sub$HW11\_cat = ifelse(DHS\_children\_sub$HW11 >=-2, 0, 1) %>%   
 factor(levels = c(0,1),   
 labels = c("Not wasted",  
 "Wasted"))  
  
  
DHS\_children\_sub$SAM = ifelse(DHS\_children\_sub$HW11 >=-3, 0, 1) %>%   
 factor(levels = c(0,1),   
 labels = c("Not SAM",  
 "SAM"))  
  
DHS\_children\_sub$V024 = factor(DHS\_children\_sub$V024,  
 levels = c(0:14),  
 labels = c("Kampala",   
 "South Buganda",   
 "North Buganda",  
 "Busoga",  
 "Bukedi",  
 "Bugisu",  
 "Teso",  
 "Karamoja",  
 "Lango",  
 "Acholi",  
 "West Nile",  
 "Bunyoro",  
 "Tooro",  
 "Ankole",  
 "Kigezi"))  
  
  
#Renaming Travel time   
DHS\_children\_sub = dplyr::rename(DHS\_children\_sub, travel\_time = raster\_travel\_time\_Lower\_travel\_time\_V1)  
  
#Categorical variable for travel time  
DHS\_children\_sub$travel\_time\_cat = expss::recode(DHS\_children\_sub$travel\_time, 0 %thru% 30 ~ 1, 31 %thru% 60 ~ 2, 61 %thru% 120 ~ 3, 121 %thru% 999 ~ 4) %>%   
 factor(levels = c(1,2,3,4),  
 labels = c("30 min. or less", "30 - 60 min.", "1 -2 hours", "More than 2 hours"))  
  
Travel\_time\_clust$travel\_time\_cat = expss::recode(Travel\_time\_clust$raster\_travel\_time\_Lower\_travel\_time\_V1, 0 %thru% 30 ~ 1, 31 %thru% 60 ~ 2, 61 %thru% 120 ~ 3, 121 %thru% 999 ~ 4) %>%   
 factor(levels = c(1,2,3,4),  
 labels = c("30 min. or less", "30 - 60 min.", "1 -2 hours", "More than 2 hours"))

#### Providing labels

var\_lab(DHS\_children\_sub$V001) = "Cluster number"  
var\_lab(DHS\_children\_sub$V005) = "Women's individual sample weight"  
var\_lab(DHS\_children\_sub$V023) = "Stratification used in sample design"  
var\_lab(DHS\_children\_sub$BORD) = "Birth order"  
var\_lab(DHS\_children\_sub$V149) = "Educational attainment of the mother"  
var\_lab(DHS\_children\_sub$V729) = "Educational attainment of the father"   
var\_lab(DHS\_children\_sub$V191) = "Wealth index"  
var\_lab(DHS\_children\_sub$V212) = "Age of respondent at first birth"  
var\_lab(DHS\_children\_sub$V445) = "BMI of mother (kg/m^2)"  
var\_lab(DHS\_children\_sub$V714) = "Work status of mother"  
var\_lab(DHS\_children\_sub$B4) = "Sex of child"  
var\_lab(DHS\_children\_sub$HW5) = "Stunting (continious)"  
var\_lab(DHS\_children\_sub$HW8) = "Underweight (continious)"   
var\_lab(DHS\_children\_sub$HW11) = "Wasting (continious)"  
var\_lab(DHS\_children\_sub$V024) = "Region"  
var\_lab(DHS\_children\_sub$M19) = "Birth weight in grams"  
var\_lab(DHS\_children\_sub$M5) = "Duration of breastfeeding"  
var\_lab(DHS\_children\_sub$V113) = "Source of drinking water"  
var\_lab(DHS\_children\_sub$V136) = "Number of household members"  
var\_lab(DHS\_children\_sub$V481) = "Covered by health insurance"  
var\_lab(DHS\_children\_sub$H22) = "Fever last two weeks"  
var\_lab(DHS\_children\_sub$V505) = "Number of other wives"  
var\_lab(DHS\_children\_sub$V714) = "Mother is currently working"  
var\_lab(DHS\_children\_sub$V729) = "Education of father/partner"  
var\_lab(DHS\_children\_sub$V024) = "Region of residence"  
var\_lab(DHS\_children\_sub$SAM) = "SAM categorical"  
var\_lab(DHS\_children\_sub$V149) = "Education of mother"  
var\_lab(DHS\_children\_sub$B4) = "Sex"  
var\_lab(DHS\_children\_sub$H11) = "Had diarrhea in the last two weeks"  
var\_lab(DHS\_children\_sub$B11) = "Preceding birth interval <=24"  
var\_lab(DHS\_children\_sub$B0) = "Child is twin"  
var\_lab(DHS\_children\_sub$HW11\_cat) = "Wasted categorical"  
var\_lab(DHS\_children\_sub$HW8\_cat) = "Underweight categorical"  
var\_lab(DHS\_children\_sub$HW5\_cat) = "Stunting categorical"  
var\_lab(DHS\_children\_sub$travel\_time) = "Travel Time in min."  
var\_lab(DHS\_children\_sub$URBAN\_RURA) = "Cluster type urban/rural"  
var\_lab(DHS\_children\_sub$NEAR\_DIST) = "Distance to nearest HC facility in km."  
var\_lab(DHS\_children\_sub$H7) = "Child has received DPT3 vaccination"  
var\_lab(DHS\_children\_sub$travel\_time\_cat) = "Travel Time Categorical"  
var\_lab(Travel\_time\_clust$travel\_time\_cat) = "Travel Time Categorical"

#### Removing invalid data

#Dead children   
DHS\_children\_sub = subset(DHS\_children\_sub, B5==1)  
  
#Only keepng children with anthropomorgic measurements  
DHS\_children\_sub = DHS\_children\_sub %>% drop\_na(HW5)

# Creating table 1

#GENERATING TABLE 1  
  
#To calculate group differences  
pvalue <- function(x, ...) {  
 # Construct vectors of data y, and groups (strata) g  
 y <- unlist(x)  
 g <- factor(rep(1:length(x), times=sapply(x, length)))  
 if (is.numeric(y)) {  
 # For numeric variables, perform a standard 2-sample t-test  
 p <- t.test(y ~ g)$p.value  
 } else {  
 # For categorical variables, perform a chi-squared test of independence  
 p <- chisq.test(table(y, g))$p.value  
 }  
 # Format the p-value, using an HTML entity for the less-than sign.  
 # The initial empty string places the output on the line below the variable label.  
 c("", sub("<", "&lt;", format.pval(p, digits=3, eps=0.001)))  
}  
  
  
#Stunting  
#OVERALL  
table1\_stunted = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW5\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra")  
  
#Group comparison  
table1\_stunted = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW5\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra", extra.col=list(`P-value`=pvalue), overall=F)  
  
table1\_stunted  
#t1flex(table1\_stunted) %>%  
 #save\_as\_docx(path = "table1\_stunted.docx")  
  
#Wasted  
#OVERALL  
table1\_wasted = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW11\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra")  
  
#Group differences  
table1\_wasted = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW11\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra", extra.col=list(`P-value`=pvalue), overall=F)  
  
table1\_wasted  
#t1flex(table1\_wasted) %>%  
 #save\_as\_docx(path = "table1\_wasted.docx")  
  
#Underweight  
#OVERALL  
table1\_underweight = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW8\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra")  
  
#Group Differences  
table1\_underweight = table1(~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + H7 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + travel\_time | HW8\_cat, data=DHS\_children\_sub, topclass="Rtable1-zebra", extra.col=list(`P-value`=pvalue), overall=F)  
  
table1\_underweight  
#t1flex(table1\_underweight) %>%  
 #save\_as\_docx(path = "table1\_underweight.docx")

# Missing values

## Investigating missing data

prop\_miss(DHS\_children\_sub)  
n\_miss(DHS\_children\_sub)  
miss\_var\_summary(DHS\_children\_sub)  
miss\_case\_summary(DHS\_children\_sub)  
miss\_case\_table(DHS\_children\_sub)  
vis\_miss(DHS\_children\_sub) + theme(axis.text.x = element\_text(angle=80))  
gg\_miss\_upset(DHS\_children\_sub)  
  
dependent1 = "HW5"  
dependent2 = "HW8"  
dependent3 = "HW11"  
  
explanatory = c("V149", "V729", "V191", "V212", "V445", "V714", "B4", "BORD", "M19", "B11", "H11", "V113", "V136", "H7", "V481", "H22", "V505", "travel\_time")  
  
missing\_plot(DHS\_children\_sub, dependent1, explanatory)  
missing\_plot(DHS\_children\_sub, dependent2, explanatory)  
missing\_plot(DHS\_children\_sub, dependent3, explanatory)  
  
  
missing = ff\_glimpse(DHS\_children\_sub, dependent1, explanatory)  
 flextable(missing$Continuous) %>%  
 save\_as\_image("missing\_cont.pdf")  
 flextable(missing$Categorical) %>%  
 save\_as\_image("missing\_cat.pdf")  
  
ff\_glimpse(DHS\_children\_sub, dependent2, explanatory)  
ff\_glimpse(DHS\_children\_sub, dependent3, explanatory)  
  
DHS\_children\_sub %>%  
 missing\_pairs(dependent1, explanatory, position = "fill")  
  
DHS\_children\_sub %>%  
 missing\_pairs(dependent2, explanatory, position = "fill")  
  
DHS\_children\_sub %>%  
 missing\_pairs(dependent3, explanatory, position = "fill")

## Imputing missing data

############################  
#USING MICE #  
  
pred\_mat = quickpred(DHS\_children\_sub, mincor = 0.25)  
  
#used 20 imputations as the most up to date version of R does not work with the proposed package "HowManyImputation" for replicate SE. 2- 10 is advised but we are more liberal with 20 for reproducible SE  
DHS\_children\_sub\_multimp <- mice(DHS\_children\_sub, m=20, predictorMatrix = pred\_mat, set.seed = 123, method = "pmm")  
  
#Complete dataset  
DHS\_children\_sub <-mice::complete(DHS\_children\_sub\_multimp, action = "long")  
  
#Checking quality of imputations  
#ALL  
densityplot(DHS\_children\_sub\_multimp)  
  
#Per Variable   
densityplot(DHS\_children\_sub\_multimp, ~H7)  
  
densityplot(DHS\_children\_sub\_multimp, ~H7 | .imp)

# Applying survey weights

#dividing weight variable by 1000000  
  
DHS\_children\_sub$V005 = DHS\_children\_sub$V005/1000000  
  
#Creating survey weights  
DHS\_children\_sub\_weighted <- svydesign(id = DHS\_children\_sub$V001,  
 strata = DHS\_children\_sub$V023,  
 weights = DHS\_children\_sub$V005,  
 nest = TRUE,  
 data = DHS\_children\_sub,  
 variables = NULL)

# Visual exploration of data

## Outcome variables

#########################  
# It is advised for computational reasons to not impute missing data before conducting exploratory analyses  
#########################  
  
### Travel time per cluster in categories ###  
ggplot(Travel\_time\_clust) +  
 aes(x = travel\_time\_cat) +  
 geom\_bar(fill = '#00bfc4') +  
 labs(title = "Travel Time per Household Cluster in Categories", x = "Travel Time Categories") +  
 theme\_light() +  
 geom\_text(stat='count',aes(label=paste0("N = ", ..count..)),vjust=2) +  
 geom\_text(aes(label = scales::percent(..count.. / 677)), stat = "count", vjust = 4)   
  
  
ggplot(Travel\_time\_clust) +  
 aes(x = travel\_time\_cat, fill = URBAN\_RURA) +  
 scale\_fill\_manual(values = c('#00bfc4', '#f9766c'),   
 name="Type of Household Cluster",  
 labels=c("Rural", "Urban")) +  
 geom\_bar(position = 'fill', col = "black") +  
 labs(title = "Travel Time per Type of Household Cluster in Categories",   
 x = "Travel Time Categories",   
 y ='Proportion') +  
 theme\_light()  
  
chisq.test(Travel\_time\_clust$travel\_time\_cat, Travel\_time\_clust$URBAN\_RURA)  
  
#OUTCOME VARIABLES  
#Histogram for stunting  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = HW5) +  
 geom\_histogram(aes(fill = HW5 < -2)) +  
 scale\_fill\_manual(values = c("light green", "red"))) %>%  
 ggplotly()  
  
(ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = HW5\_cat) +   
 geom\_bar(position = 'fill') +  
 scale\_fill\_manual(values = c("light green", "red"))) %>%  
 ggplotly()  
  
svytable(~DHS\_children\_sub$HW5\_cat, DHS\_children\_sub\_weighted)  
  
#Histogram for underweight  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = HW8) +  
 geom\_histogram(aes(fill = HW8 < -2)) +  
 scale\_fill\_manual(values = c("light green", "red"))  
  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = HW8\_cat) +   
 geom\_bar(position = 'fill') +  
 scale\_fill\_manual(values = c("light green", "red"))  
  
svytable(~DHS\_children\_sub$HW8\_cat, DHS\_children\_sub\_weighted)  
  
#Histogram for wasting  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = HW11) +  
 geom\_histogram(aes(fill = HW11 < -2)) +  
 scale\_fill\_manual(values = c("light green", "red"))  
  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = HW11\_cat) +   
 geom\_bar(position = 'fill') +  
 scale\_fill\_manual(values = c("light green", "red"))  
  
svytable(~DHS\_children\_sub$HW11\_cat, DHS\_children\_sub\_weighted)  
  
#Severe Acute Malnutrition  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = SAM) +   
 geom\_bar(position = 'fill') +  
 scale\_fill\_manual(values = c("light green", "red"))  
  
svytable(~DHS\_children\_sub$SAM, DHS\_children\_sub\_weighted)

## Covariates

#########################  
# It is advised for computational reasons to not impute missing data before conducting exploratory analyses  
#########################  
  
#Nearest facility in kilometers  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = NEAR\_DIST) +  
 geom\_histogram(fill = 'light blue', col = "black")  
 ) %>%  
 ggplotly()  
  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = NEAR\_DIST, y = V024) +  
 geom\_boxplot()  
  
#Travel time in minutes  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = travel\_time)+  
 geom\_histogram(fill = 'light blue', col = "black")) %>%  
 ggplotly()  
  
#Region of residence  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = V024) +  
 geom\_bar(fill = "light blue", col = "black")  
  
#Birt Order  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = BORD) +  
 geom\_bar(fill = "light blue", col = "black")  
  
#Wealth Index  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V191)+  
 geom\_histogram(fill = 'light blue', col = "black")  
  
#Age of respondent at first birth  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V212)+  
 geom\_histogram(fill = 'light blue', col = "black")  
  
#Body mass index (BMI) of mother  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V445)+  
 geom\_histogram(fill = 'light blue', col = "black")  
  
#Birth weight in kilograms  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = M19)+  
 geom\_histogram(fill = 'light blue', col = "black")  
  
#Mother is currently working  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = V714)+  
 geom\_bar(position = 'fill')  
  
#Mother's educational attainment  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = V149)+  
 geom\_bar(position = 'fill')+  
 scale\_fill\_brewer(palette = "RdYlGn")  
  
#Father/partners educational attainment  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = V729)+  
 geom\_bar(position = 'fill')+  
 scale\_fill\_brewer(palette = "RdYlGn")  
  
#Sex of child  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = B4)+  
 geom\_bar(position = 'fill')  
  
#Birth interval (more or less than 24 months)  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = '', fill = B11)+  
 geom\_bar(position = 'fill') +  
 scale\_fill\_brewer(palette = "Dark2")  
  
#Diarrhea last two weeks  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes(x = '', fill = H11)+  
 geom\_bar(position = 'fill') +  
 scale\_fill\_brewer(palette = "Dark2")  
  
#Number of other wives  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V505) +  
 geom\_histogram(fill = 'light blue', col = "black")  
  
svytable(~DHS\_children\_sub$V505, DHS\_children\_sub\_weighted)  
  
#Fever last two weeks  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = H22) +  
 geom\_bar(fill = 'light blue', col = "black")  
  
ggsurvey(design = DHS\_children\_sub\_weighted)+  
 aes(x = '', fill = H22)+  
 geom\_bar(position = 'fill')  
  
svytable(~DHS\_children\_sub$H22, DHS\_children\_sub\_weighted)  
  
#Covered by health insurance  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V481) +  
 geom\_bar(fill = 'light blue', col = "black")  
  
svytable(~DHS\_children\_sub$V481, DHS\_children\_sub\_weighted)  
  
#Number of household members  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V136) +  
 geom\_histogram(fill = 'light blue', col = "black")  
  
svytable(~DHS\_children\_sub$V136, DHS\_children\_sub\_weighted)  
  
#Source of drinking water  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = V113) +  
 geom\_bar(fill = 'light blue', col = "black")  
  
svytable(~DHS\_children\_sub$V113, DHS\_children\_sub\_weighted)  
  
#Breastfeeding  
ggsurvey(DHS\_children\_sub\_weighted)+  
 aes(x = M5) +  
 geom\_bar(fill = 'light blue', col = "black")  
  
svytable(~DHS\_children\_sub$M5, DHS\_children\_sub\_weighted)

# Investigating correlation

## Visually

esquisser()  
  
#########################  
# It is advised for computational reasons to not impute missing data before conducting exploratory analyses  
#########################  
  
#Main interested variables  
#Stunting by Euclidian distance to nearest HC facility  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = NEAR\_DIST, y = HW5) +  
 geom\_point() +  
 geom\_smooth(method = loess)  
  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW5\_cat, y = NEAR\_DIST) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
  
#Underweight by Euclidian distance to nearest HC facility  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = NEAR\_DIST, y = HW8) +  
 geom\_point()+  
 geom\_smooth(method = loess)  
  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW8\_cat, y = NEAR\_DIST) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
  
#Wasting by Euclidian distance to nearest HC facility  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = NEAR\_DIST, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess)  
  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW11\_cat, y = NEAR\_DIST) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
  
ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = SAM, y = NEAR\_DIST) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
  
  
#TRAVEL TIME  
#Stunting by traveltime to nearest HC facility  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = travel\_time, y = HW5) +  
 geom\_point() +  
 geom\_smooth(method = loess) +  
 xlab("Travel time in min.")+  
 ylab("Stunting") +  
 ggtitle("Stunting by travel time in min.")  
) %>%   
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW5\_cat, y = travel\_time) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
) %>%  
 ggplotly()  
  
#Underweight by Travel Time to nearest HC facility  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = travel\_time, y = HW8) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 xlab("Travel time in min.")+  
 ylab("Underweight") +  
 ggtitle("Underweight by travel time in min.")  
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW8\_cat, y = travel\_time) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
) %>%  
 ggplotly()  
  
#Wasting by Travel Time to nearest HC facility  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = travel\_time, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 xlab("Travel time in min.")+  
 ylab("Wasting") +  
 ggtitle("Wasting by travel time in min.")  
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = HW11\_cat, y = travel\_time) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
) %>%   
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = SAM, y = travel\_time) +  
 geom\_boxplot(fill = c("Green", "red"), alpha = 0.5)  
) %>%  
 ggplotly()  
  
#Other continuous variables  
#Wealth index  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V191, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V191, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V191, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#age of respondent at first birth  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V212, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V212, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V212, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Bmi of mother  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V445, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V445, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V445, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Birth weight  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = M19, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = M19, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = M19, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Number of household members  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V136, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = loess)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V136, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V136, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = loess) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Number of other wives  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V505, y = HW5) +  
 geom\_point()+  
 geom\_smooth(method = lm)+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V505, y = HW8) +  
 geom\_point()+   
 geom\_smooth(method = lm) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V505, y = HW11) +  
 geom\_point()+  
 geom\_smooth(method = lm) +  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Categorical variables  
#Education of the mother  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V149, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V149, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V149, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Education of the father  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V729, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V729, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V729, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Mother is currently working  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V714, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V714, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V714, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Child is twin  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B0, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B0, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B0, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Sex of the child  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B4, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B4, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B4, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Birth order  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = BORD, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = BORD, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = BORD, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Birth interval  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B11, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B11, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = B11, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Diarrhea last two weeks  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Duration of breastfeeding  
X  
  
#Source of drinking water  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V113, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V113, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V113, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Vaccination status  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H11, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Health insurance  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V481, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V481, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = V481, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Fever last two weeks  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H22, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H22, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = H22, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
#Urban\_rural  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = URBAN\_RURA, y = HW5) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = URBAN\_RURA, y = HW8) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()  
  
(ggsurvey(DHS\_children\_sub\_weighted) +  
 aes (x = URBAN\_RURA, y = HW11) +  
 geom\_boxplot()+  
 scale\_y\_continuous(breaks=seq(-6, 6, by = 0.5))   
) %>%  
 ggplotly()

## Statistically

ggpairs(DHS\_children\_sub, columns = c(6, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 20, 23, 25, 27, 29, 37), cardinality\_threshold = 17)  
  
#Exludes V729, B11, H11, V113, H22  
  
DHS\_children\_sub\_num = select\_if(DHS\_children\_sub, is.numeric)  
   
DHS\_children\_sub\_num = subset(DHS\_children\_sub\_num, select = c(V191, V212, V445, B0, HW5, HW8, HW11, M19, V136, V505, travel\_time))  
  
heatmaply\_cor(x = cor(na.omit(DHS\_children\_sub\_num)), xlab = "Features",  
 ylab = "Features")  
  
  
DHS\_children\_sub\_num = select\_if(DHS\_children\_sub, is.numeric)  
   
DHS\_children\_sub\_num = subset(DHS\_children\_sub\_num, select = c(V191, V212, V445, B0, HW5, HW8, HW11, M19, V136, V505, travel\_time))  
  
heatmaply\_cor(x = cor(na.omit(DHS\_children\_sub\_num)), xlab = "Features",  
 ylab = "Features")

# Statistical testing

## GLM stunting

#Generalized linear model COMPLETE  
  
 ####### STUNTING #######  
glm\_stunted = svyglm(formula = HW5 ~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
#Removed V149, B0 for high VIF  
glm\_stunted = svyglm(formula = HW5 ~ B4 + BORD + M19 + B11 + H11 + H22 + V481 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
glm\_stuned\_null = svyglm(formula = HW5~1,  
 family = gaussian(),  
 design = DHS\_children\_sub\_weighted,  
 rescale = T) # null model  
  
#AIC  
glm\_stunted = stepAIC(glm\_stunted, direction = "both")  
  
#Check model  
modelsummary(glm\_stunted, stars = T, coef\_rename = TRUE)  
  
summary(glm\_stunted)  
check\_model(glm\_stunted)  
  
plot(glm\_stunted)  
  
performance(glm\_stunted)  
vif(glm\_stunted)  
  
visreg(glm\_stunted)  
ggcoef\_model(glm\_stunted)

## GLM underweight

####### UNDERWEIGHT #######  
glm\_underweight = svyglm(formula = HW8 ~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
#Removed V149, B0 for high VIF  
glm\_underweight = svyglm(formula = HW8 ~ B4 + BORD + M19 + B11 + H11 + H22 + V481 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
  
#AIC  
glm\_underweight = stepAIC(glm\_underweight, direction = "both")  
  
#Check model  
modelsummary(glm\_underweight, stars = T, coef\_rename = TRUE)  
summary(glm\_underweight)  
check\_model(glm\_underweight)  
  
plot(glm\_underweight)  
  
performance(glm\_underweight)  
vif(glm\_underweight)  
  
ggcoef\_model(glm\_underweight)

## GLM wasting

####### WASTING #######  
glm\_wasted = svyglm(formula = HW11 ~ B4 + BORD + B0 + M19 + B11 + H11 + H22 + V481 + V149 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
  
#Removed V149, B0 for high VIF  
glm\_wasted = svyglm(formula = HW11 ~ B4 + BORD + M19 + B11 + H11 + H22 + V481 + V729 + V505 + V191 + V212 + V445 + V714 + V113 + V136 + M5 + H7 + travel\_time,  
 design = DHS\_children\_sub\_weighted,  
 family = gaussian(),  
 rescale = T)  
  
  
#AIC  
glm\_wasted = stepAIC(glm\_wasted, direction = "both")  
  
#Check model  
modelsummary(glm\_wasted, stars = T, coef\_rename = TRUE)  
summary(glm\_wasted)  
check\_model(glm\_wasted)  
  
plot(glm\_wasted)  
  
performance(glm\_wasted)  
vif(glm\_wasted)  
  
ggcoef\_model(glm\_wasted)

## Model comparison

####### COMPARING MODELS #######  
models = modelsummary(list(glm\_stunted, glm\_wasted, glm\_underweight), stars = TRUE, coef\_rename = T, col.names=c("", "Stunting", "Wasting", "Underweight"), output = "models.docx")  
  
models  
  
#output = "models.docx"  
statistic = 'conf.int'  
#gof\_map = c("nobs", "rmse")  
  
## as command for modelsummary to get document "output = "test.docx""   
ggcoef\_compare(model = list("Stunted" = glm\_stunted, "Wasted" = glm\_wasted, "Underweight" = glm\_underweight), type = "dodged", conf.int = TRUE, conf.level = 0.95, add\_reference\_rows = TRUE, significance = 0.05, include = c("V729", "V191", "V445", "B4", "BORD", "M19", "B11", "V113", "M5", "V505", "H7", "travel\_time"), no\_reference\_row = c("B11", "H7"))