comparing-outliers

2022-09-02

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(readxl)  
library(flextable)

##   
## Attaching package: 'flextable'  
##   
## The following object is masked from 'package:purrr':  
##   
## compose

library(extrafont)

## Registering fonts with R

library(fauxnaif)

#times new roman tables  
my\_ft\_theme <- function(ft, ...) {  
 # Remove vertical cell padding  
 ft <- padding(ft, padding.top = 0, padding.bottom = 0, part = "all")  
   
 # Change font to TNR 11  
 ft <- font(ft, fontname = "Times New Roman", part = "all")  
 ft <- fontsize(ft, part = "all", size = 12)  
 ft  
}

vis\_outliers <- read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\all\_vis\_out1.xlsx")  
stat\_outliers <- read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\combin\_out1.xlsx")

filtered\_visoutl <- filter(vis\_outliers, vis\_out %in% c("GoSub\_C",   
 "NRB\_L",   
 "SnasM\_C",   
 "TrSman\_C"))

all\_outliers <- full\_join(filtered\_visoutl, stat\_outliers, by="ID")

all the visual outliers that were changed to NA based on visual assessment of measurement sumstats are also stat outliers (univariate AND multivariate)

all\_outliers1 <- all\_outliers  
  
all\_outliers1$truee <- "1"  
  
all\_outliers1 <- all\_outliers1 %>% pivot\_wider(names\_from = offending, values\_from = truee)  
  
str(all\_outliers1)

## tibble [25 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:25] "400-20210129-009" "400-20201123-009" "400-20210203-001" "400-20210216-006" ...  
## $ vis\_out : chr [1:25] "GoSub\_C" "NRB\_L" "SnasM\_C" "TrSman\_C" ...  
## $ multv\_outlier : chr [1:25] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ extreme\_overall : chr [1:25] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ extreme\_race\_eth : chr [1:25] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ extreme\_gender : chr [1:25] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ extreme\_age : chr [1:25] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ GoSub\_C univar : chr [1:25] "1" NA NA "1" ...  
## $ TrSman\_C univar : chr [1:25] "1" NA NA "1" ...  
## $ NRB\_L univar : chr [1:25] NA "1" NA NA ...  
## $ SnasM\_C univar : chr [1:25] NA NA "1" NA ...  
## $ ProS\_L univar : chr [1:25] NA NA NA NA ...  
## $ Not Univar Outlier: chr [1:25] NA NA NA NA ...  
## $ SelP\_L univar : chr [1:25] NA NA NA NA ...  
## $ SelM\_L univar : chr [1:25] NA NA NA NA ...  
## $ AA\_C univar : chr [1:25] NA NA NA NA ...  
## $ BiW\_C univar : chr [1:25] NA NA NA NA ...

all\_outliers1[is.na(all\_outliers1)]="0"  
  
all\_outliers1$`GoSub\_C univar` <- as.numeric(all\_outliers1$`GoSub\_C univar`)  
all\_outliers1$`TrSman\_C univar` <- as.numeric(all\_outliers1$`TrSman\_C univar`)  
all\_outliers1$`NRB\_L univar` <- as.numeric(all\_outliers1$`NRB\_L univar`)  
all\_outliers1$`SnasM\_C univar`<- as.numeric(all\_outliers1$`SnasM\_C univar`)  
all\_outliers1$`ProS\_L univar` <- as.numeric(all\_outliers1$`ProS\_L univar`)  
all\_outliers1$`SelP\_L univar` <- as.numeric(all\_outliers1$`SelP\_L univar`)  
all\_outliers1$`SelM\_L univar` <- as.numeric(all\_outliers1$`SelM\_L univar`)  
all\_outliers1$`AA\_C univar` <- as.numeric(all\_outliers1$`AA\_C univar`)  
all\_outliers1$`BiW\_C univar` <- as.numeric(all\_outliers1$`BiW\_C univar`)  
  
  
all\_outliers1$univar\_sum <- all\_outliers1$`GoSub\_C univar` +   
 all\_outliers1$`TrSman\_C univar` +   
 all\_outliers1$`NRB\_L univar` +  
 all\_outliers1$`SnasM\_C univar` +   
 all\_outliers1$`ProS\_L univar` +   
 all\_outliers1$`SelP\_L univar` +  
 all\_outliers1$`SelM\_L univar` +   
 all\_outliers1$`AA\_C univar` +   
 all\_outliers1$`BiW\_C univar`

all\_outliers1$`GoSub\_C univar` <- as.character(all\_outliers1$`GoSub\_C univar`)  
all\_outliers1$`TrSman\_C univar` <- as.character(all\_outliers1$`TrSman\_C univar`)  
all\_outliers1$`NRB\_L univar` <- as.character(all\_outliers1$`NRB\_L univar`)  
all\_outliers1$`SnasM\_C univar`<- as.character(all\_outliers1$`SnasM\_C univar`)  
all\_outliers1$`ProS\_L univar` <- as.character(all\_outliers1$`ProS\_L univar`)  
all\_outliers1$`SelP\_L univar` <- as.character(all\_outliers1$`SelP\_L univar`)  
all\_outliers1$`SelM\_L univar` <- as.character(all\_outliers1$`SelM\_L univar`)  
all\_outliers1$`AA\_C univar` <- as.character(all\_outliers1$`AA\_C univar`)  
all\_outliers1$`BiW\_C univar` <- as.character(all\_outliers1$`BiW\_C univar`)  
  
  
  
all\_outliers1$`GoSub\_C univar`<-   
 recode\_factor(all\_outliers1$`GoSub\_C univar`, '0'= "no",'1' = "GoSub\_C")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`GoSub\_C univar` = na\_if\_in(`GoSub\_C univar`, "no"))  
  
  
all\_outliers1$`TrSman\_C univar`<-   
 recode\_factor(all\_outliers1$`TrSman\_C univar`, '0'= "no",'1' = "TrSman\_C")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`TrSman\_C univar` = na\_if\_in(`TrSman\_C univar`, "no"))  
  
  
all\_outliers1$`NRB\_L univar` <-   
 recode\_factor(all\_outliers1$`NRB\_L univar`, '0'= "no",'1' = "NRB\_L")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`NRB\_L univar`= na\_if\_in(`NRB\_L univar`, "no"))  
  
  
all\_outliers1$`SnasM\_C univar` <-   
 recode\_factor(all\_outliers1$`SnasM\_C univar`, '0'= "no",'1' = "SnasM\_C")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`SnasM\_C univar`= na\_if\_in(`SnasM\_C univar`, "no"))  
  
  
all\_outliers1$`ProS\_L univar` <-   
 recode\_factor(all\_outliers1$`ProS\_L univar`, '0'= "no",'1' = "ProS\_L")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`ProS\_L univar`= na\_if\_in(`ProS\_L univar`, "no"))  
  
  
all\_outliers1$`SelP\_L univar` <-   
 recode\_factor(all\_outliers1$`SelP\_L univar`, '0'= "no",'1' = "SelP\_L")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`SelP\_L univar`= na\_if\_in(`SelP\_L univar`, "no"))  
  
  
all\_outliers1$`SelM\_L univar` <-   
 recode\_factor(all\_outliers1$`SelM\_L univar`, '0'= "no",'1' = "SelM\_L")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`SelM\_L univar`= na\_if\_in(`SelM\_L univar`, "no"))  
  
  
all\_outliers1$`AA\_C univar` <-   
 recode\_factor(all\_outliers1$`AA\_C univar`, '0'= "no",'1' = "AA\_C")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`AA\_C univar`= na\_if\_in(`AA\_C univar`, "no"))  
  
  
all\_outliers1$`BiW\_C univar` <-   
 recode\_factor(all\_outliers1$`BiW\_C univar`, '0'= "no",'1' = "BiW\_C")  
#all\_outliers1 <- all\_outliers1 %>%   
 #mutate(`BiW\_C univar`= na\_if\_in(`BiW\_C univar`, "no"))

cols <- c("GoSub\_C univar",  
 "TrSman\_C univar",  
 "NRB\_L univar",  
 "SnasM\_C univar",  
 "ProS\_L univar",  
 "SelP\_L univar",  
 "SelM\_L univar",  
 "AA\_C univar",  
 "BiW\_C univar")  
  
  
  
all\_outliers1$univar\_outliers <- apply( all\_outliers1[ , cols], 1, paste, collapse="&")  
  
all\_outliers1$univar\_outliers <- gsub("no","",as.character(all\_outliers1$univar\_outliers))  
  
all\_outliers1$univar\_outliers <- gsub("&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&&&&&","",as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&&&&&&&&&","",as.character(all\_outliers1$univar\_outliers))  
  
  
   
   
all\_outliers1$univar\_outliers <- gsub("GoSub\_C&TrSman\_C&", "GoSub\_C & TrSman\_C", as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&SnasM\_C&", "SnasM\_C", as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&SelP\_L&SelM\_L", "SelP\_L & SelM\_L", as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&TrSman\_C&", "TrSman\_C", as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&SelP\_L&", "SelP\_L", as.character(all\_outliers1$univar\_outliers))  
all\_outliers1$univar\_outliers <- gsub("&AA\_C&", "AA\_C", as.character(all\_outliers1$univar\_outliers))  
  
  
all\_outliers1 <- all\_outliers1 %>%   
 mutate(univar\_outliers= na\_if\_in(univar\_outliers, ""))

all\_outliers1$vis\_out <- gsub("0", "Not Visual Outlier", as.character(all\_outliers1$vis\_out))  
all\_outliers1$multv\_outlier <- gsub("TRUE", "Multivariate Outlier", as.character(all\_outliers1$multv\_outlier))  
all\_outliers1$multv\_outlier <- gsub("Not Outlier", "Not Multivariate Outlier", as.character(all\_outliers1$multv\_outlier))  
  
all\_outliers1$univar\_outliers <- all\_outliers1$univar\_outliers %>%   
 replace\_na("Not Univariate Outlier")

all\_outliers1 <- all\_outliers1[c(1,2,3,19)]

#Size 12 Table TNR  
flextable(all\_outliers1) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Univariate, Multivariate, and Visual Outlier Status") %>%   
 autofit

**Table** : Univariate, Multivariate, and Visual Outlier Status

| **ID** | **vis\_out** | **multv\_outlier** | **univar\_outliers** |
| --- | --- | --- | --- |
| 400-20210129-009 | GoSub\_C | Multivariate Outlier | GoSub\_C & TrSman\_C |
| 400-20201123-009 | NRB\_L | Multivariate Outlier | NRB\_L |
| 400-20210203-001 | SnasM\_C | Multivariate Outlier | SnasM\_C |
| 400-20210216-006 | TrSman\_C | Multivariate Outlier | GoSub\_C & TrSman\_C |
| 400-20201013-003 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20201013-003 | Not Visual Outlier | Multivariate Outlier | SnasM\_C |
| 400-20210210-009 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210317-009 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20210317-009 | Not Visual Outlier | Multivariate Outlier | SnasM\_C |
| 400-20201209-011 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201214-021 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201215-009 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201202-019 | Not Visual Outlier | Multivariate Outlier | SelP\_L & SelM\_L |
| 400-20201203-004 | Not Visual Outlier | Multivariate Outlier | GoSub\_C |
| 400-20201204-018 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201208-024 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210120-002 | Not Visual Outlier | Multivariate Outlier | TrSman\_C |
| 400-20210122-007 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20210128-015 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210201-002 | Not Visual Outlier | Multivariate Outlier | SelP\_L |
| 400-20210203-009 | Not Visual Outlier | Multivariate Outlier | TrSman\_C |
| 400-20201012-012 | Not Visual Outlier | Not Multivariate Outlier | AA\_C |
| 400-20210108-006 | Not Visual Outlier | Not Multivariate Outlier | BiW\_C |
| 400-20201216-014 | Not Visual Outlier | Not Multivariate Outlier | SelP\_L |
| 400-20210126-012 | Not Visual Outlier | Not Multivariate Outlier | SelP\_L |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(all\_outliers1) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Univariate, Multivariate, and Visual Outlier Status") %>%   
 autofit() %>%   
 fit\_to\_width(7.5)

**Table** : Univariate, Multivariate, and Visual Outlier Status

| **ID** | **vis\_out** | **multv\_outlier** | **univar\_outliers** |
| --- | --- | --- | --- |
| 400-20210129-009 | GoSub\_C | Multivariate Outlier | GoSub\_C & TrSman\_C |
| 400-20201123-009 | NRB\_L | Multivariate Outlier | NRB\_L |
| 400-20210203-001 | SnasM\_C | Multivariate Outlier | SnasM\_C |
| 400-20210216-006 | TrSman\_C | Multivariate Outlier | GoSub\_C & TrSman\_C |
| 400-20201013-003 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20201013-003 | Not Visual Outlier | Multivariate Outlier | SnasM\_C |
| 400-20210210-009 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210317-009 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20210317-009 | Not Visual Outlier | Multivariate Outlier | SnasM\_C |
| 400-20201209-011 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201214-021 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201215-009 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201202-019 | Not Visual Outlier | Multivariate Outlier | SelP\_L & SelM\_L |
| 400-20201203-004 | Not Visual Outlier | Multivariate Outlier | GoSub\_C |
| 400-20201204-018 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20201208-024 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210120-002 | Not Visual Outlier | Multivariate Outlier | TrSman\_C |
| 400-20210122-007 | Not Visual Outlier | Multivariate Outlier | ProS\_L |
| 400-20210128-015 | Not Visual Outlier | Multivariate Outlier | Not Univariate Outlier |
| 400-20210201-002 | Not Visual Outlier | Multivariate Outlier | SelP\_L |
| 400-20210203-009 | Not Visual Outlier | Multivariate Outlier | TrSman\_C |
| 400-20201012-012 | Not Visual Outlier | Not Multivariate Outlier | AA\_C |
| 400-20210108-006 | Not Visual Outlier | Not Multivariate Outlier | BiW\_C |
| 400-20201216-014 | Not Visual Outlier | Not Multivariate Outlier | SelP\_L |
| 400-20210126-012 | Not Visual Outlier | Not Multivariate Outlier | SelP\_L |