removing outliers

2022-07-26

library(readxl)  
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(writexl)  
library(extrafont)

## Registering fonts with R

library(flextable)

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

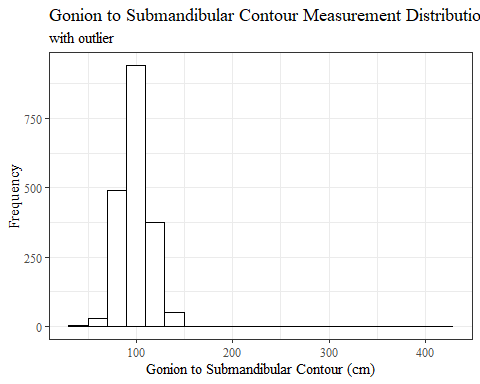
#times new roman tables  
my\_ft\_theme <- function(ft, ...) {  
 # Remove vertical cell padding  
 ft <- padding(ft, padding.top = 0, padding.bottom = 0, part = "all")  
   
 # Change font to TNR 11  
 ft <- font(ft, fontname = "Times New Roman", part = "all")  
 ft <- fontsize(ft, part = "all", size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

headscan\_full<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\headscan\_full.xlsx")  
  
headscan\_full$AA\_C <- headscan\_full$AA\_C \*10  
headscan\_full$BGl\_C <- headscan\_full$BGl\_C \* 10  
headscan\_full$BiW\_C <- headscan\_full$BiW\_C \*10  
headscan\_full$BiW\_L <- headscan\_full$BiW\_L \*10  
headscan\_full$ChCh\_C <- headscan\_full$ChCh\_C \*10  
headscan\_full$GoSub\_C <- headscan\_full$GoSub\_C \*10  
headscan\_full$NRB\_L <- headscan\_full$NRB\_L \*10  
headscan\_full$ProA\_L <- headscan\_full$ProA\_L \*10  
headscan\_full$ProA\_C <- headscan\_full$ProA\_C \*10  
headscan\_full$ProS\_C <- headscan\_full$ProS\_C \*10  
headscan\_full$ProS\_L <- headscan\_full$ProS\_L \*10  
headscan\_full$SelP\_C <- headscan\_full$SelP\_C \*10  
headscan\_full$SelP\_L <- headscan\_full$SelP\_L \*10  
headscan\_full$SelDH\_C <- headscan\_full$SelDH\_C \*10  
headscan\_full$SelM\_L <- headscan\_full$SelM\_L \*10  
headscan\_full$SnasM\_C <- headscan\_full$SnasM\_C \*10  
headscan\_full$SmanM\_C <- headscan\_full$SmanM\_C \*10  
headscan\_full$SmanM\_L <- headscan\_full$SmanM\_L \*10  
headscan\_full$SnasM\_L <- headscan\_full$SnasM\_L \*10  
headscan\_full$TrHO\_C <- headscan\_full$TrHO\_C \*10  
headscan\_full$TrEJ\_C <- headscan\_full$TrEJ\_C \*10  
headscan\_full$TrGo\_C <- headscan\_full$TrGo\_C \*10  
headscan\_full$TrSel\_C <- headscan\_full$TrSel\_C \*10  
headscan\_full$TrSman\_C <- headscan\_full$TrSman\_C \*10  
headscan\_full$TrSnas\_C <- headscan\_full$TrSnas\_C \*10  
headscan\_full$TrTr\_C <- headscan\_full$TrTr\_C \*10  
headscan\_full$TrTr\_L <- headscan\_full$TrTr\_L \*10  
  
  
str(headscan\_full)

## tibble [2,016 × 33] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:2016] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:2016] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BGl\_C : num [1:2016] 315 289 293 313 288 306 320 NA 300 277 ...  
## $ BiW\_C : num [1:2016] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:2016] 115 108 121 109 104 106 109 111 113 116 ...  
## $ ChCh\_C : num [1:2016] 62 64 68 70 70 70 67 69 67 63 ...  
## $ GoSub\_C : num [1:2016] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:2016] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProA\_L : num [1:2016] 28 25 31 23 28 28 26 27 32 28 ...  
## $ ProA\_C : num [1:2016] 31 27 33 27 31 29 27 29 34 31 ...  
## $ ProS\_C : num [1:2016] 18 20 14 13 22 22 19 14 26 24 ...  
## $ ProS\_L : num [1:2016] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_C : num [1:2016] 42 41 51 45 47 48 46 42 47 44 ...  
## $ SelP\_L : num [1:2016] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelDH\_C : num [1:2016] 15 9 9 11 13 15 9 9 12 14 ...  
## $ SelM\_L : num [1:2016] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:2016] 82 55 84 74 73 80 78 76 64 75 ...  
## $ SmanM\_C : num [1:2016] 59 51 45 43 33 34 55 37 61 41 ...  
## $ SmanM\_L : num [1:2016] 55 50 45 42 33 34 50 36 59 40 ...  
## $ SnasM\_L : num [1:2016] 75 53 78 69 67 76 69 71 62 69 ...  
## $ TrHO\_C : num [1:2016] 179 163 169 166 159 162 169 NA 167 166 ...  
## $ TrEJ\_C : num [1:2016] 40 32 39 29 46 42 29 32 29 33 ...  
## $ TrGo\_C : num [1:2016] 84 57 70 61 68 70 75 61 67 64 ...  
## $ TrSel\_C : num [1:2016] 149 138 150 133 140 151 140 138 156 143 ...  
## $ TrSman\_C : num [1:2016] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrSnas\_C : num [1:2016] 163 142 167 145 152 157 148 149 157 NA ...  
## $ TrTr\_C : num [1:2016] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:2016] 155 141 156 149 146 146 147 151 157 144 ...  
## $ coder : chr [1:2016] "Kayna" "Kayna" "Kayna" "Kayna" ...  
## $ age : num [1:2016] 31 49 49 34 49 55 26 18 25 27 ...  
## $ gender : chr [1:2016] "Male" "Female" "Male" "Male" ...  
## $ race\_eth : chr [1:2016] "Black" "white" "white" "white" ...  
## $ age\_group: chr [1:2016] "18-36" "37-54" "37-54" "18-36" ...

#histogram of all GoSub\_C values w/ outlier  
ggplot(data=headscan\_full, aes(x=GoSub\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Gonion to Submandibular Contour Measurement Distribution",  
 subtitle= "with outlier",  
 y="Frequency",  
 x="Gonion to Submandibular Contour (cm)")

## Warning: Removed 127 rows containing non-finite values (stat\_bin).

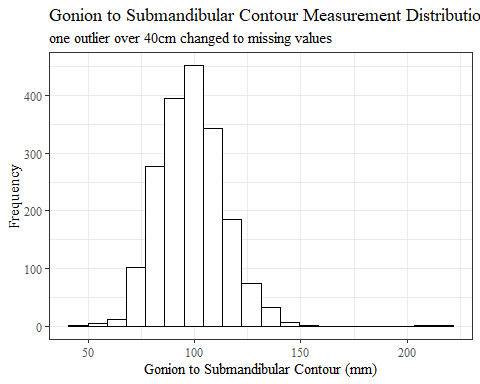


#REMOVING OUTLIER GoSubC  
  
max(headscan\_full$GoSub\_C, na.rm = TRUE)

## [1] 424

headscan\_full1 <- headscan\_full %>% mutate(GoSub\_C = replace(GoSub\_C, GoSub\_C>400, NA))  
  
#histogram of all GoSub\_C values  
ggplot(data=headscan\_full1, aes(x=GoSub\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Gonion to Submandibular Contour Measurement Distribution",  
 subtitle = "one outlier over 40cm changed to missing values",  
 y="Frequency",  
 x="Gonion to Submandibular Contour (mm)")

## Warning: Removed 128 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$GoSub\_C))

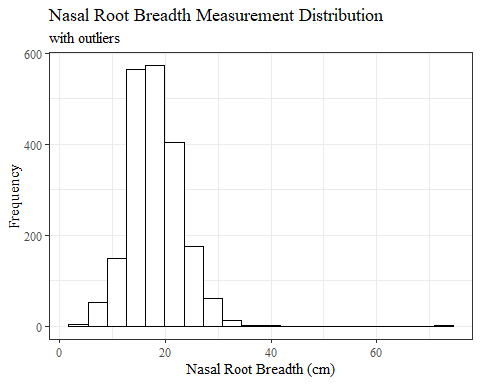
## [1] 127

sum(is.na(headscan\_full1$GoSub\_C))

## [1] 128

#histogram of all NRB\_L values w/ outlier  
ggplot(data=headscan\_full, aes(x=NRB\_L))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Nasal Root Breadth Measurement Distribution",  
 subtitle= "with outliers",  
 y="Frequency",  
 x="Nasal Root Breadth (cm)")

## Warning: Removed 15 rows containing non-finite values (stat\_bin).

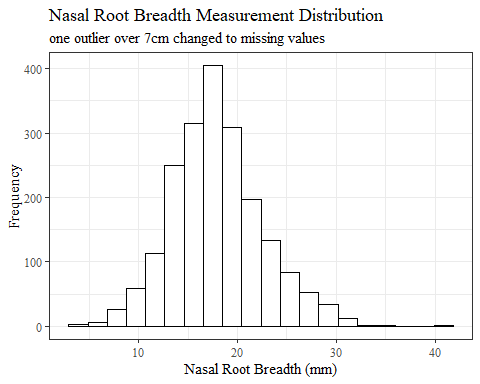


#REMOVING OUTLIER NRB\_L  
  
max(headscan\_full$NRB\_L, na.rm = TRUE)

## [1] 72

headscan\_full1 <- headscan\_full1 %>% mutate(NRB\_L = replace(NRB\_L, NRB\_L>70, NA))  
  
#histogram of all NRB\_L values  
ggplot(data=headscan\_full1, aes(x=NRB\_L))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Nasal Root Breadth Measurement Distribution",  
 subtitle = "one outlier over 7cm changed to missing values",  
 y="Frequency",  
 x="Nasal Root Breadth (mm)")

## Warning: Removed 16 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$NRB\_L))

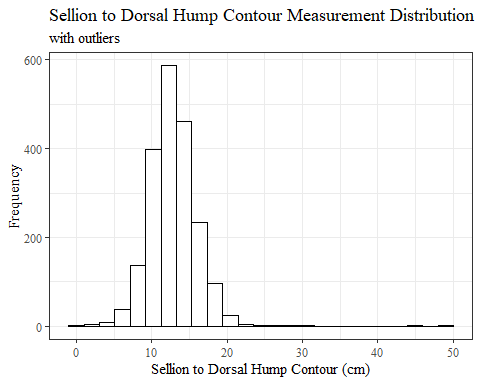
## [1] 15

sum(is.na(headscan\_full1$NRB\_L))

## [1] 16

#histogram of all SelDH\_C values w/ outlier  
ggplot(data=headscan\_full, aes(x=SelDH\_C ))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Sellion to Dorsal Hump Contour Measurement Distribution",  
 subtitle = "with outliers",  
 y="Frequency",  
 x="Sellion to Dorsal Hump Contour (cm)")

## Warning: Removed 14 rows containing non-finite values (stat\_bin).

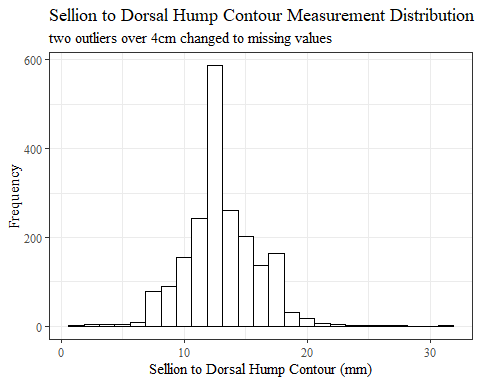


#REMOVING OUTLIER SelDH\_C  
  
max(headscan\_full$SelDH\_C, na.rm = TRUE)

## [1] 50

headscan\_full1 <- headscan\_full1 %>% mutate(SelDH\_C = replace(SelDH\_C, SelDH\_C>40, NA))  
  
#histogram of all SelDH\_C values  
ggplot(data=headscan\_full1, aes(x=SelDH\_C))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Sellion to Dorsal Hump Contour Measurement Distribution",  
 subtitle = "two outliers over 4cm changed to missing values",  
 y="Frequency",  
 x="Sellion to Dorsal Hump Contour (mm)")

## Warning: Removed 16 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$SelDH\_C))

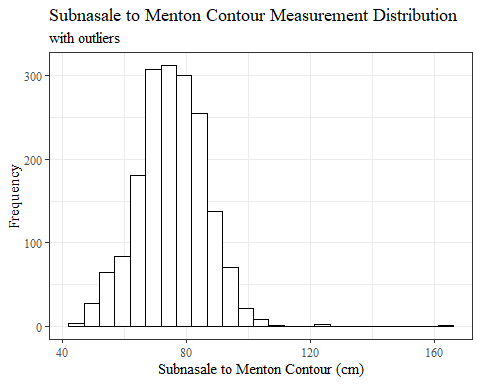
## [1] 14

sum(is.na(headscan\_full1$SelDH\_C))

## [1] 16

#histogram of all SnasM\_C values w/ outlier  
ggplot(data=headscan\_full, aes(x=SnasM\_C))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Subnasale to Menton Contour Measurement Distribution",  
 subtitle = "with outliers",  
 y="Frequency",  
 x="Subnasale to Menton Contour (cm)")

## Warning: Removed 236 rows containing non-finite values (stat\_bin).

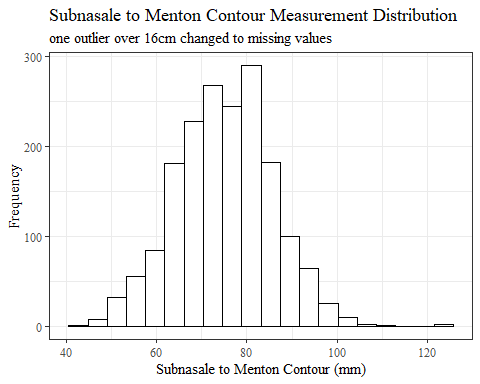


#REMOVING OUTLIER SnasM\_C  
  
max(headscan\_full$SnasM\_C, na.rm = TRUE)

## [1] 163

headscan\_full1 <- headscan\_full1 %>% mutate(SnasM\_C = replace(SnasM\_C, SnasM\_C>160, NA))  
  
#histogram of all SnasM\_C values  
ggplot(data=headscan\_full1, aes(x=SnasM\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Subnasale to Menton Contour Measurement Distribution",  
 subtitle = "one outlier over 16cm changed to missing values",  
 y="Frequency",  
 x="Subnasale to Menton Contour (mm)")

## Warning: Removed 237 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$SnasM\_C))

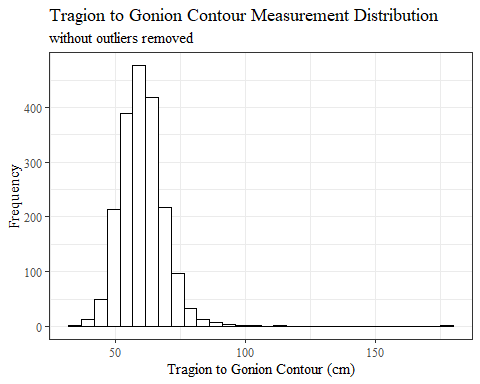
## [1] 236

sum(is.na(headscan\_full1$SnasM\_C))

## [1] 237

#histogram of all TrGo\_C values  
ggplot(data=headscan\_full, aes(x=TrGo\_C))+  
 geom\_bar(stat="bin", bins=30, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Gonion Contour Measurement Distribution",  
 subtitle = "without outliers removed",  
 y="Frequency",  
 x="Tragion to Gonion Contour (cm)")

## Warning: Removed 80 rows containing non-finite values (stat\_bin).

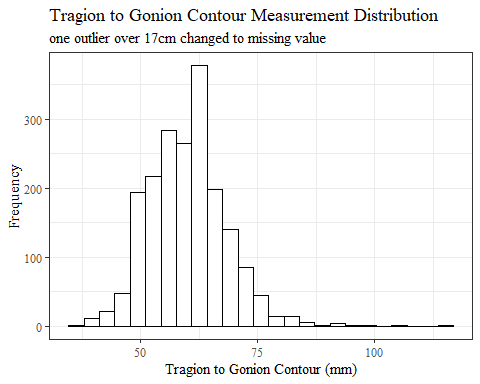


#REMOVING OUTLIER TrGo\_C  
  
max(headscan\_full$TrGo\_C, na.rm = TRUE)

## [1] 178

headscan\_full1 <- headscan\_full1 %>% mutate(TrGo\_C = replace(TrGo\_C, TrGo\_C>170, NA))  
  
#histogram of all TrGo\_C values  
ggplot(data=headscan\_full1, aes(x=TrGo\_C))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Gonion Contour Measurement Distribution",  
 subtitle = "one outlier over 17cm changed to missing value",  
 y="Frequency",  
 x="Tragion to Gonion Contour (mm)")

## Warning: Removed 81 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$TrGo\_C))

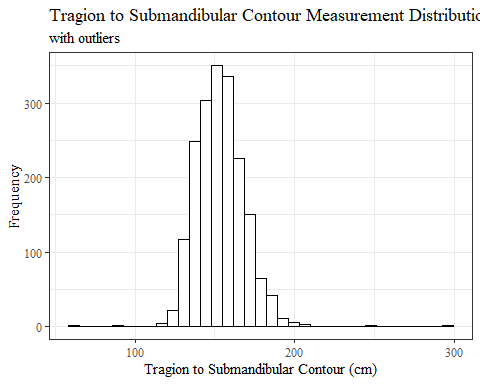
## [1] 80

sum(is.na(headscan\_full1$TrGo\_C))

## [1] 81

#histogram of all TrSman\_C values  
ggplot(data=headscan\_full, aes(x=TrSman\_C))+  
 geom\_bar(stat="bin", bins=35, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Submandibular Contour Measurement Distribution",  
 subtitle = "with outliers",  
 y="Frequency",  
 x="Tragion to Submandibular Contour (cm)")

## Warning: Removed 132 rows containing non-finite values (stat\_bin).

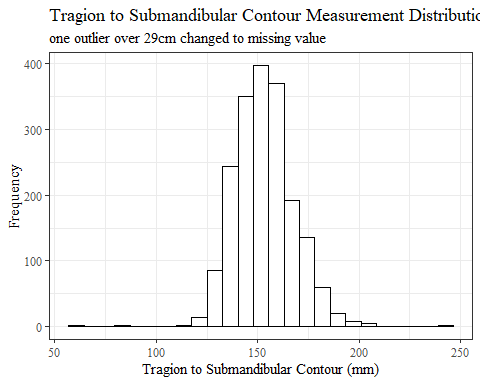


#REMOVING OUTLIER TrSman\_C  
  
max(headscan\_full$TrSman\_C, na.rm = TRUE)

## [1] 298

headscan\_full1 <- headscan\_full1 %>% mutate(TrSman\_C = replace(TrSman\_C, TrSman\_C>290, NA))  
  
#histogram of all TrSman\_C values  
ggplot(data=headscan\_full1, aes(x=TrSman\_C))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Submandibular Contour Measurement Distribution",  
 subtitle = "one outlier over 29cm changed to missing value",  
 y="Frequency",  
 x="Tragion to Submandibular Contour (mm)")

## Warning: Removed 133 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$TrSman\_C))

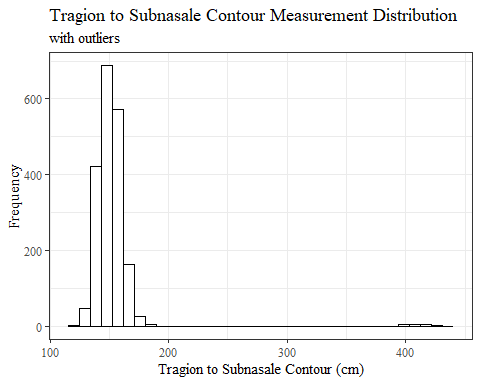
## [1] 132

sum(is.na(headscan\_full1$TrSman\_C))

## [1] 133

#histogram of all TrSnas\_C values  
ggplot(data=headscan\_full, aes(x=TrSnas\_C))+  
 geom\_bar(stat="bin", bins=35, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Subnasale Contour Measurement Distribution",  
 subtitle = "with outliers",  
 y="Frequency",  
 x="Tragion to Subnasale Contour (cm)")

## Warning: Removed 71 rows containing non-finite values (stat\_bin).

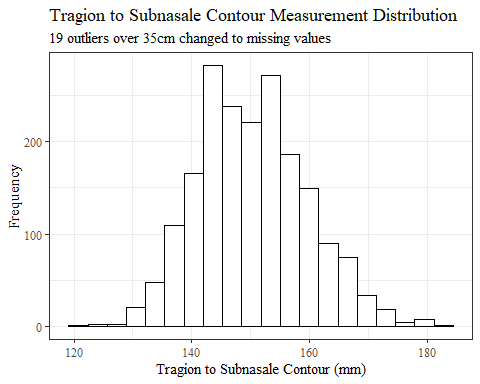


#REMOVING OUTLIER TrSnas\_C  
  
max(headscan\_full$TrSnas\_C, na.rm = TRUE)

## [1] 437

headscan\_full1 <- headscan\_full1 %>% mutate(TrSnas\_C = replace(TrSnas\_C, TrSnas\_C>350, NA))  
  
#histogram of all TrSnas\_C values  
ggplot(data=headscan\_full1, aes(x=TrSnas\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Tragion to Subnasale Contour Measurement Distribution",  
 subtitle = "19 outliers over 35cm changed to missing values",  
 y="Frequency",  
 x="Tragion to Subnasale Contour (mm)")

## Warning: Removed 90 rows containing non-finite values (stat\_bin).



sum(is.na(headscan\_full$TrSnas\_C))

## [1] 71

sum(is.na(headscan\_full1$TrSnas\_C))

## [1] 90

str(headscan\_full1)

## tibble [2,016 × 33] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:2016] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:2016] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BGl\_C : num [1:2016] 315 289 293 313 288 306 320 NA 300 277 ...  
## $ BiW\_C : num [1:2016] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:2016] 115 108 121 109 104 106 109 111 113 116 ...  
## $ ChCh\_C : num [1:2016] 62 64 68 70 70 70 67 69 67 63 ...  
## $ GoSub\_C : num [1:2016] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:2016] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProA\_L : num [1:2016] 28 25 31 23 28 28 26 27 32 28 ...  
## $ ProA\_C : num [1:2016] 31 27 33 27 31 29 27 29 34 31 ...  
## $ ProS\_C : num [1:2016] 18 20 14 13 22 22 19 14 26 24 ...  
## $ ProS\_L : num [1:2016] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_C : num [1:2016] 42 41 51 45 47 48 46 42 47 44 ...  
## $ SelP\_L : num [1:2016] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelDH\_C : num [1:2016] 15 9 9 11 13 15 9 9 12 14 ...  
## $ SelM\_L : num [1:2016] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:2016] 82 55 84 74 73 80 78 76 64 75 ...  
## $ SmanM\_C : num [1:2016] 59 51 45 43 33 34 55 37 61 41 ...  
## $ SmanM\_L : num [1:2016] 55 50 45 42 33 34 50 36 59 40 ...  
## $ SnasM\_L : num [1:2016] 75 53 78 69 67 76 69 71 62 69 ...  
## $ TrHO\_C : num [1:2016] 179 163 169 166 159 162 169 NA 167 166 ...  
## $ TrEJ\_C : num [1:2016] 40 32 39 29 46 42 29 32 29 33 ...  
## $ TrGo\_C : num [1:2016] 84 57 70 61 68 70 75 61 67 64 ...  
## $ TrSel\_C : num [1:2016] 149 138 150 133 140 151 140 138 156 143 ...  
## $ TrSman\_C : num [1:2016] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrSnas\_C : num [1:2016] 163 142 167 145 152 157 148 149 157 NA ...  
## $ TrTr\_C : num [1:2016] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:2016] 155 141 156 149 146 146 147 151 157 144 ...  
## $ coder : chr [1:2016] "Kayna" "Kayna" "Kayna" "Kayna" ...  
## $ age : num [1:2016] 31 49 49 34 49 55 26 18 25 27 ...  
## $ gender : chr [1:2016] "Male" "Female" "Male" "Male" ...  
## $ race\_eth : chr [1:2016] "Black" "white" "white" "white" ...  
## $ age\_group: chr [1:2016] "18-36" "37-54" "37-54" "18-36" ...

outl1 <- subset(headscan\_full, GoSub\_C>400)   
outl1$vis\_out <- "GoSub\_C"  
   
outl2 <- subset(headscan\_full, NRB\_L>70)  
outl2$vis\_out <- "NRB\_L"  
  
outl3 <- subset(headscan\_full, SelDH\_C>40)  
outl3$vis\_out <- "SelDH\_C"  
  
  
outl4 <- subset(headscan\_full, SnasM\_C>160)  
outl4$vis\_out <- "SnasM\_C"  
  
outl5 <- subset(headscan\_full, TrGo\_C>170)  
outl5$vis\_out <- "TrGo\_C"  
  
outl6 <- subset(headscan\_full, TrSman\_C>290)  
outl6$vis\_out <- "TrSman\_C"  
  
outl7 <- subset(headscan\_full, TrSnas\_C>350)  
outl7$vis\_out <- "TrSnas\_C"  
  
  
all\_vis\_out <- rbind(outl1, outl2, outl3, outl4, outl5, outl6, outl7)

sum(is.na(headscan\_full))

## [1] 2801

sum(is.na(headscan\_full1))

## [1] 2827

all\_vis\_out1 <- all\_vis\_out[-c(2:33)]  
   
#Size 12 Table TNR  
flextable(all\_vis\_out1) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Visual Outliers") %>%   
 autofit

**Table** : Visual Outliers

| **ID** | **vis\_out** |
| --- | --- |
| 400-20210129-009 | GoSub\_C |
| 400-20201123-009 | NRB\_L |
| 400-20201202-019 | SelDH\_C |
| 400-20210107-021 | SelDH\_C |
| 400-20210203-001 | SnasM\_C |
| 400-20210120-002 | TrGo\_C |
| 400-20210216-006 | TrSman\_C |
| 400-20201020-001 | TrSnas\_C |
| 400-20201117-012 | TrSnas\_C |
| 400-20201214-013 | TrSnas\_C |
| 400-20201217-017 | TrSnas\_C |
| 400-20201223-003 | TrSnas\_C |
| 400-20201203-004 | TrSnas\_C |
| 400-20201203-013 | TrSnas\_C |
| 400-20201204-009 | TrSnas\_C |
| 400-20201204-020 | TrSnas\_C |
| 400-20201207-007 | TrSnas\_C |
| 400-20201207-011 | TrSnas\_C |
| 400-20201207-016 | TrSnas\_C |
| 400-20210104-013 | TrSnas\_C |
| 400-20210108-017 | TrSnas\_C |
| 400-20210114-017 | TrSnas\_C |
| 400-20210121-021 | TrSnas\_C |
| 400-20210126-002 | TrSnas\_C |
| 400-20210128-007 | TrSnas\_C |
| 400-20210129-013 | TrSnas\_C |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))

#write\_xlsx(headscan\_full1, "C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\headscan\_full1.xlsx")  
  
#write\_xlsx(all\_vis\_out1, "C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\all\_vis\_out1.xlsx")

measureNAsums <- colSums(is.na(headscan\_full1))  
  
measureNAprops <- colMeans(is.na(headscan\_full1))  
  
measureNAprops1 <- as.data.frame(measureNAprops)  
measureNAprops1 <- rownames\_to\_column(measureNAprops1, "measure\_name")  
measureNAprops1 <- measureNAprops1 %>% slice(-c(1, 29:33))  
  
measureNAsums1 <- as.data.frame(measureNAsums)  
measureNAsums1 <- rownames\_to\_column(measureNAsums1, "measure\_name")  
measureNAsums1 <- measureNAsums1 %>% slice(-c(1, 29:33))  
  
measureNAs <- inner\_join(measureNAprops1, measureNAsums1, by = "measure\_name", desc)  
  
measureNAs$measureNAprops<- round(measureNAs$measureNAprops, digits=4)  
  
measureNAs$measure\_name <- fct\_reorder(measureNAs$measure\_name, measureNAs$measureNAsums, .desc=TRUE)  
  
str(measureNAs$measure\_name)

## Factor w/ 27 levels "BGl\_C","TrHO\_C",..: 19 1 20 21 12 9 22 24 23 16 ...

ggplot(data=measureNAs, aes(x=measure\_name, y=measureNAsums))+  
 geom\_bar(stat= "identity", color= "black", fill = "white")+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 theme(axis.text.x = element\_text(angle = 45, vjust=0.7))+  
 labs(title="Missing Values by Measurement Location (Count)",  
 y="Count of Missing Values",  
 x="Measure Abbreviation (see Appendix B)")

