exploring-demo-postimpute

2022-09-13

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

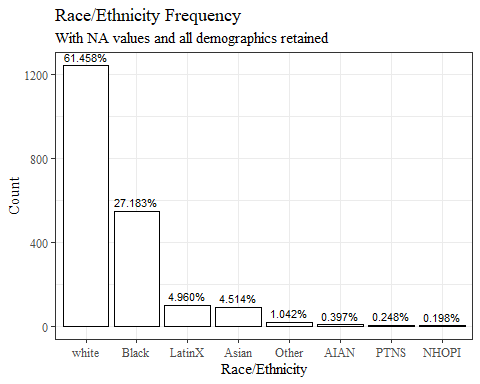
library(forcats)

#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  
}

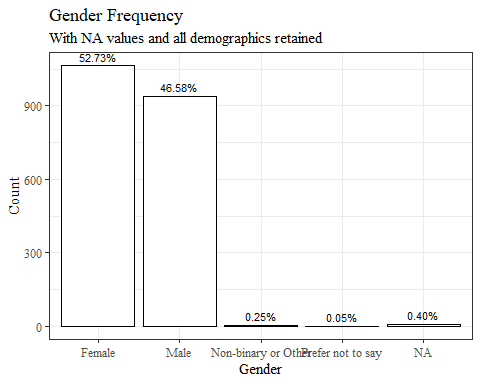
chosen\_withna<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\chosen\_withna.xlsx")  
  
chosen\_withna$gender<-as.factor(chosen\_withna$gender)  
chosen\_withna$race\_eth<-as.factor(chosen\_withna$race\_eth)  
chosen\_withna$age\_group<-as.factor(chosen\_withna$age\_group)  
  
  
str(chosen\_withna)

## tibble [2,016 × 16] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:2016] "400-20210129-009" "400-20201123-009" "400-20210203-001" "400-20210216-006" ...  
## $ AA\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ BiW\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ BiW\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ GoSub\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ NRB\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ ProS\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ SelP\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ SelM\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ SnasM\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ TrSman\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ TrTr\_C : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ TrTr\_L : num [1:2016] NA NA NA NA NA NA NA NA NA NA ...  
## $ gender : Factor w/ 4 levels "Female","Male",..: 2 2 2 2 1 1 2 1 2 1 ...  
## $ race\_eth : Factor w/ 8 levels "AIAN","Asian",..: 8 8 8 8 3 2 8 8 3 8 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 2 1 1 2 2 2 3 3 2 2 ...

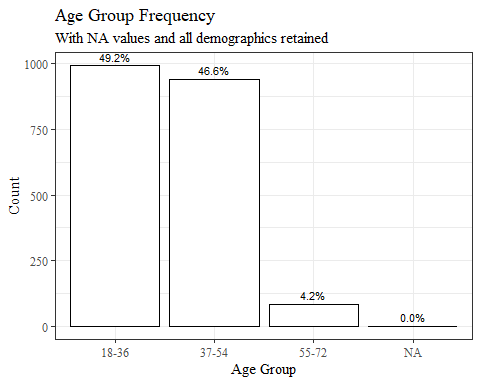
chosen\_withna %>%   
 ggplot(aes(x=fct\_infreq(race\_eth), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Race/Ethnicity Frequency",  
 subtitle="With NA values and all demographics retained",  
 y="Count",  
 x="Race/Ethnicity")



chosen\_withna %>%   
 ggplot(aes(x=fct\_infreq(gender), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Gender Frequency",  
 subtitle="With NA values and all demographics retained",  
 y="Count",  
 x="Gender")



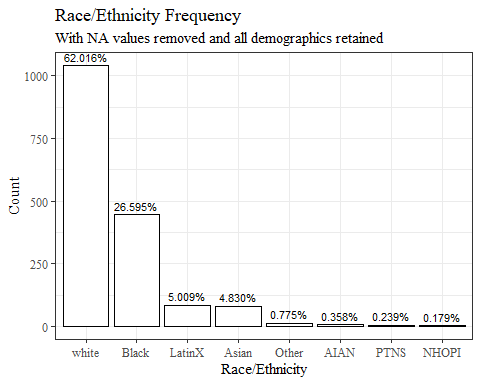
chosen\_withna %>%   
 ggplot(aes(x=fct\_infreq(age\_group), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Age Group Frequency",  
 subtitle="With NA values and all demographics retained",  
 y="Count",  
 x="Age Group")



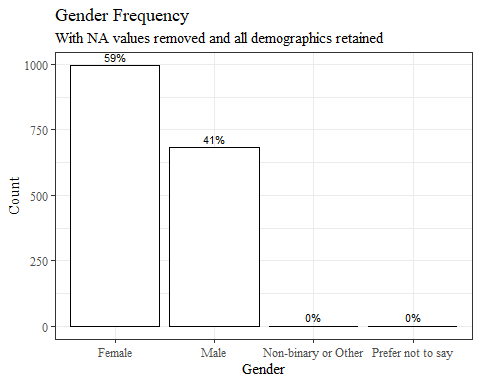
chosen\_nona<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\chosen\_nona.xlsx")  
  
chosen\_nona$gender<-as.factor(chosen\_nona$gender)  
chosen\_nona$race\_eth<-as.factor(chosen\_nona$race\_eth)  
chosen\_nona$age\_group<-as.factor(chosen\_nona$age\_group)  
  
  
str(chosen\_nona)

## tibble [1,677 × 16] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:1677] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:1677] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BiW\_C : num [1:1677] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:1677] 115 108 121 109 104 106 109 111 113 116 ...  
## $ GoSub\_C : num [1:1677] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:1677] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProS\_L : num [1:1677] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_L : num [1:1677] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelM\_L : num [1:1677] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:1677] 82 55 84 74 73 80 78 76 64 75 ...  
## $ TrSman\_C : num [1:1677] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrTr\_C : num [1:1677] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:1677] 155 141 156 149 146 146 147 151 157 144 ...  
## $ gender : Factor w/ 4 levels "Female","Male",..: 2 1 2 2 2 2 2 2 2 2 ...  
## $ race\_eth : Factor w/ 8 levels "AIAN","Asian",..: 3 8 8 8 8 8 3 3 8 8 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 1 2 2 1 2 3 1 1 1 1 ...

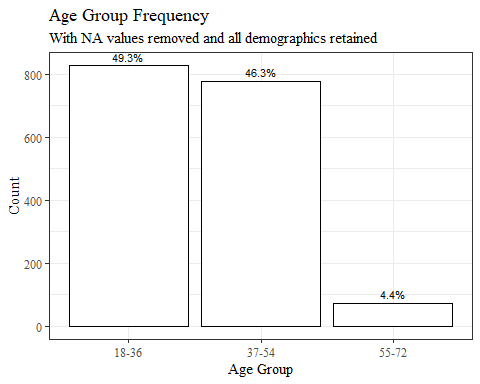
chosen\_nona %>%   
 ggplot(aes(x=fct\_infreq(race\_eth), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Race/Ethnicity Frequency",  
 subtitle="With NA values removed and all demographics retained",  
 y="Count",  
 x="Race/Ethnicity")



chosen\_nona %>%   
 ggplot(aes(x=fct\_infreq(gender), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Gender Frequency",  
 subtitle="With NA values removed and all demographics retained",  
 y="Count",  
 x="Gender")



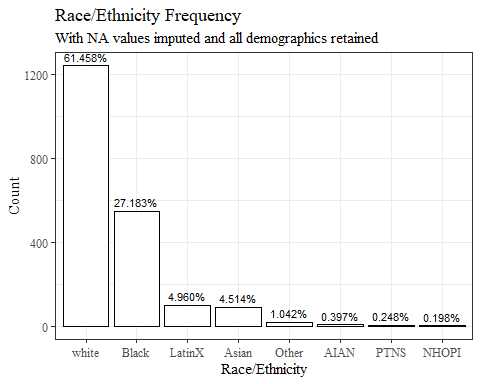
chosen\_nona %>%   
 ggplot(aes(x=fct\_infreq(age\_group), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Age Group Frequency",  
 subtitle="With NA values removed and all demographics retained",  
 y="Count",  
 x="Age Group")



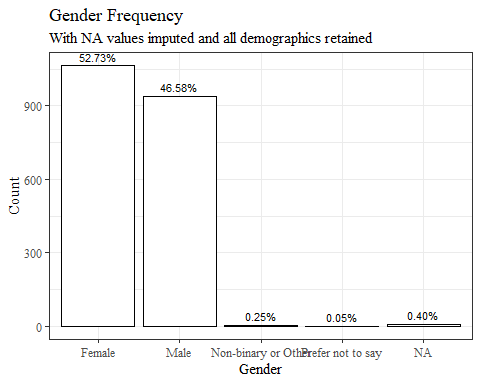
chosen\_imputed<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\chosen\_imputed.xlsx")  
  
chosen\_imputed$gender<-as.factor(chosen\_imputed$gender)  
chosen\_imputed$race\_eth<-as.factor(chosen\_imputed$race\_eth)  
chosen\_imputed$age\_group<-as.factor(chosen\_imputed$age\_group)  
  
  
str(chosen\_imputed)

## tibble [2,016 × 16] (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 ...  
## $ 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 ...  
## $ 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 ...  
## $ ProS\_L : num [1:2016] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_L : num [1:2016] 42 41 51 44 47 48 46 41 46 44 ...  
## $ 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 ...  
## $ TrSman\_C : num [1:2016] 177 145 178 147 157 164 149 159 151 160 ...  
## $ 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 ...  
## $ gender : Factor w/ 4 levels "Female","Male",..: 2 1 2 2 2 2 2 2 2 2 ...  
## $ race\_eth : Factor w/ 8 levels "AIAN","Asian",..: 3 8 8 8 8 8 3 3 8 8 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 1 2 2 1 2 3 1 1 1 1 ...

chosen\_imputed %>%   
 ggplot(aes(x=fct\_infreq(race\_eth), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Race/Ethnicity Frequency",  
 subtitle="With NA values imputed and all demographics retained",  
 y="Count",  
 x="Race/Ethnicity")



chosen\_imputed %>%   
 ggplot(aes(x=fct\_infreq(gender), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Gender Frequency",  
 subtitle="With NA values imputed and all demographics retained",  
 y="Count",  
 x="Gender")



chosen\_imputed %>%   
 ggplot(aes(x=fct\_infreq(age\_group), label=scales::percent(prop.table(stat(count)))))+  
 geom\_bar(stat="count", color= "black", fill = "white")+  
 geom\_text(stat="count",  
 position= position\_dodge(0.9),  
 vjust = -0.5,   
 size = 3)+  
 theme\_bw()+theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="Age Group Frequency",  
 subtitle="With NA values imputed and all demographics retained",  
 y="Count",  
 x="Age Group")

