PCA demographics

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

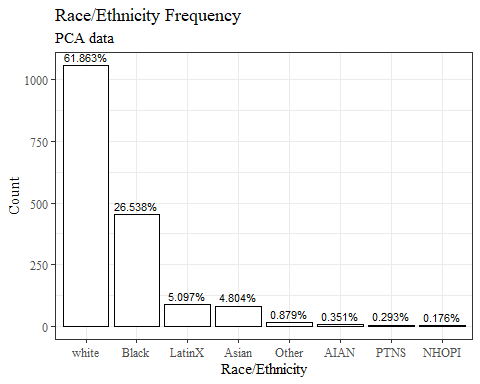
library(forcats)

#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  
}

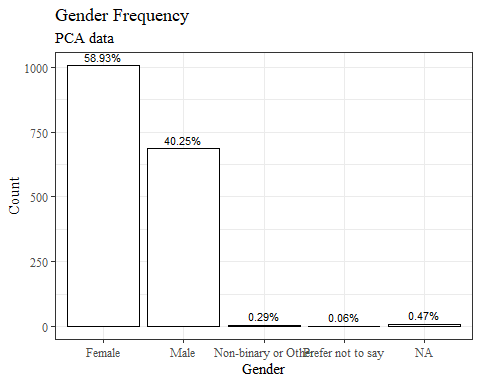
PCAdata\_full<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\PCAdata\_full.xlsx")  
  
PCAdata\_full$gender<-as.factor(PCAdata\_full$gender)  
PCAdata\_full$race\_eth<-as.factor(PCAdata\_full$race\_eth)  
PCAdata\_full$age\_group<-as.factor(PCAdata\_full$age\_group)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

## tibble [1,707 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ AA\_C : num [1:1707] 650 550 700 580 670 600 590 590 650 650 ...  
## $ BiW\_C : num [1:1707] 1300 1270 1430 1400 1370 1300 1410 1380 1430 1500 ...  
## $ BiW\_L : num [1:1707] 1150 1080 1210 1090 1040 1060 1090 1110 1130 1160 ...  
## $ GoSub\_C : num [1:1707] 930 930 1150 930 1030 1000 790 1060 850 1020 ...  
## $ NRB\_L : num [1:1707] 170 180 190 210 190 140 170 180 160 170 ...  
## $ ProS\_L : num [1:1707] 170 180 140 130 200 200 180 120 240 220 ...  
## $ SelP\_L : num [1:1707] 420 410 510 440 470 480 460 410 460 440 ...  
## $ SelM\_L : num [1:1707] 1220 990 1300 1150 1190 1260 1170 1120 1170 1170 ...  
## $ SnasM\_C : num [1:1707] 820 550 840 740 730 800 780 760 640 750 ...  
## $ TrSman\_C : num [1:1707] 1770 1450 1780 1470 1570 1640 1490 1590 1510 1600 ...  
## $ TrTr\_C : num [1:1707] 2960 2760 2920 2730 2790 3000 2830 2750 3070 2860 ...  
## $ TrTr\_L : num [1:1707] 1550 1410 1560 1490 1460 1460 1470 1510 1570 1440 ...  
## $ 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 ...

PCAdata\_full %>%   
 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="PCA data",  
 y="Count",  
 x="Race/Ethnicity")



PCAdata\_full %>%   
 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="PCA data",  
 y="Count",  
 x="Gender")



PCAdata\_full %>%   
 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="PCA data",  
 y="Count",  
 x="Age Group")

