data exploration

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ 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(knitr)  
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.2.1

## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output  
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'

##   
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':  
##   
## group\_rows

library(officer)

## Warning: package 'officer' was built under R version 4.2.1

library(flextable)

## Warning: package 'flextable' was built under R version 4.2.1

##   
## Attaching package: 'flextable'

## The following objects are masked from 'package:kableExtra':  
##   
## as\_image, footnote

## 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  
}

#times new roman ggplots  
#library(remotes)  
#remotes::install\_version("Rttf2pt1", version = "1.3.8")  
#library(extrafont)  
#font\_import(paths = NULL, recursive = TRUE, prompt = TRUE,pattern = NULL)  
#loadfonts(device = "win")

load(file="headscan\_full")

age\_sumstats <- headscan\_full %>%   
 summarise(n = n(),  
 min = min(age, na.rm = TRUE),  
 max = max(age, na.rm = TRUE),  
 mean = mean(age, na.rm = TRUE),  
 sd = sd(age, na.rm = TRUE),  
 se = sd/sqrt(n),  
 quant5th = quantile(age, 0.05, na.rm=TRUE),  
 quant25th = quantile(age, 0.25, na.rm=TRUE),  
 quant50th = quantile(age, 0.50, na.rm=TRUE),  
 quant75th = quantile(age, 0.75, na.rm=TRUE),  
 quant95th = quantile(age, 0.95, na.rm=TRUE),  
 na = sum(is.na(age)))  
  
age\_sumstats <- age\_sumstats %>%   
 mutate(across(where(is.numeric), round, 2))  
   
age\_sumstats %>%   
 kbl(caption = "Age SumStats") %>%   
 kable\_styling(bootstrap\_options = c("striped", "hover", "condensed"), full\_width = TRUE)

Age SumStats

n

min

max

mean

sd

se

quant5th

quant25th

quant50th

quant75th

quant95th

na

2017

18

72

36.39

11.51

0.26

19

26

37

46

54

1

flextable(age\_sumstats) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Age SumStats")

**Table** : Age SumStats

| **n** | **min** | **max** | **mean** | **sd** | **se** | **quant5th** | **quant25th** | **quant50th** | **quant75th** | **quant95th** | **na** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2,017 | 18 | 72 | 36.39 | 11.51 | 0.26 | 19 | 26 | 37 | 46 | 54 | 1 |

raceage\_sumstats <- headscan\_full %>%   
 group\_by(race\_eth) %>%   
 summarise(n = n(),  
 min = min(age, na.rm = TRUE),  
 max = max(age, na.rm = TRUE),  
 mean = mean(age, na.rm = TRUE),  
 sd = sd(age, na.rm = TRUE),  
 se = sd/sqrt(n),  
 quant5th = quantile(age, 0.05, na.rm=TRUE),  
 quant25th = quantile(age, 0.25, na.rm=TRUE),  
 quant50th = quantile(age, 0.50, na.rm=TRUE),  
 quant75th = quantile(age, 0.75, na.rm=TRUE),  
 quant95th = quantile(age, 0.95, na.rm=TRUE),  
 na = sum(is.na(age)))  
  
raceage\_sumstats <- raceage\_sumstats %>%   
 mutate(across(where(is.numeric), round, 2))  
   
raceage\_sumstats %>%   
 kbl(caption = "Age SumStats by Race/Ethnicity") %>%   
 kable\_styling(bootstrap\_options = c("striped", "hover", "condensed"), full\_width = TRUE)

Age SumStats by Race/Ethnicity

race\_eth

n

min

max

mean

sd

se

quant5th

quant25th

quant50th

quant75th

quant95th

na

AIAN

8

27

56

43.25

11.78

4.17

27.70

33.50

45.5

53.00

56.00

0

Asian

91

18

56

33.23

11.76

1.23

18.00

21.50

31.0

42.50

54.00

0

Black

548

18

71

37.92

10.79

0.46

21.00

29.00

39.0

47.00

54.00

0

LatinX

100

18

55

34.63

11.93

1.19

19.00

23.00

34.0

44.50

53.10

1

NHOPI

4

19

40

27.00

9.76

4.88

19.15

19.75

24.5

31.75

38.35

0

Other

21

20

72

37.48

14.75

3.22

20.00

25.00

33.0

51.00

55.00

0

PTNS

5

29

40

36.60

4.72

2.11

30.20

35.00

39.0

40.00

40.00

0

white

1240

18

62

36.05

11.64

0.33

19.00

25.00

36.0

46.00

54.00

0

flextable(raceage\_sumstats) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Age SumStats by Race/Ethnicity") %>%   
 set\_header\_labels(values = list(race\_eth = "Race/Ethnicity"))

**Table** : Age SumStats by Race/Ethnicity

| **Race/Ethnicity** | **n** | **min** | **max** | **mean** | **sd** | **se** | **quant5th** | **quant25th** | **quant50th** | **quant75th** | **quant95th** | **na** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AIAN | 8 | 27 | 56 | 43.25 | 11.78 | 4.17 | 27.70 | 33.50 | 45.5 | 53.00 | 56.00 | 0 |
| Asian | 91 | 18 | 56 | 33.23 | 11.76 | 1.23 | 18.00 | 21.50 | 31.0 | 42.50 | 54.00 | 0 |
| Black | 548 | 18 | 71 | 37.92 | 10.79 | 0.46 | 21.00 | 29.00 | 39.0 | 47.00 | 54.00 | 0 |
| LatinX | 100 | 18 | 55 | 34.63 | 11.93 | 1.19 | 19.00 | 23.00 | 34.0 | 44.50 | 53.10 | 1 |
| NHOPI | 4 | 19 | 40 | 27.00 | 9.76 | 4.88 | 19.15 | 19.75 | 24.5 | 31.75 | 38.35 | 0 |
| Other | 21 | 20 | 72 | 37.48 | 14.75 | 3.22 | 20.00 | 25.00 | 33.0 | 51.00 | 55.00 | 0 |
| PTNS | 5 | 29 | 40 | 36.60 | 4.72 | 2.11 | 30.20 | 35.00 | 39.0 | 40.00 | 40.00 | 0 |
| white | 1,240 | 18 | 62 | 36.05 | 11.64 | 0.33 | 19.00 | 25.00 | 36.0 | 46.00 | 54.00 | 0 |

genderage\_sumstats <- headscan\_full %>%   
 group\_by(gender) %>%   
 summarise(n = n(),  
 min = min(age, na.rm = TRUE),  
 max = max(age, na.rm = TRUE),  
 mean = mean(age, na.rm = TRUE),  
 sd = sd(age, na.rm = TRUE),  
 se = sd/sqrt(n),  
 quant5th = quantile(age, 0.05, na.rm=TRUE),  
 quant25th = quantile(age, 0.25, na.rm=TRUE),  
 quant50th = quantile(age, 0.50, na.rm=TRUE),  
 quant75th = quantile(age, 0.75, na.rm=TRUE),  
 quant95th = quantile(age, 0.95, na.rm=TRUE),  
 na = sum(is.na(age)))  
  
genderage\_sumstats <- genderage\_sumstats %>%   
 mutate(across(where(is.numeric), round, 2))  
   
genderage\_sumstats %>%   
 kbl(caption = "Age SumStats by Gender") %>%   
 kable\_styling(bootstrap\_options = c("striped", "hover", "condensed"), full\_width = TRUE)

Age SumStats by Gender

gender

n

min

max

mean

sd

se

quant5th

quant25th

quant50th

quant75th

quant95th

na

Female

1064

18

71

36.75

11.57

0.35

20.00

26.0

38.0

47.00

54.0

1

Male

939

18

72

35.97

11.48

0.37

19.00

26.0

35.0

46.00

54.1

0

Non-binary or Other

5

29

43

34.00

5.48

2.45

29.40

31.0

32.0

35.00

41.4

0

Prefer not to say

1

39

39

39.00

NA

NA

39.00

39.0

39.0

39.00

39.0

0

NA

8

22

49

39.12

9.79

3.46

24.45

33.5

41.5

46.75

49.0

0

flextable(genderage\_sumstats) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Age SumStats by Gender") %>%   
 set\_header\_labels(values = list(gender = "Gender"))

**Table** : Age SumStats by Gender

| **Gender** | **n** | **min** | **max** | **mean** | **sd** | **se** | **quant5th** | **quant25th** | **quant50th** | **quant75th** | **quant95th** | **na** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Female | 1,064 | 18 | 71 | 36.75 | 11.57 | 0.35 | 20.00 | 26.0 | 38.0 | 47.00 | 54.0 | 1 |
| Male | 939 | 18 | 72 | 35.97 | 11.48 | 0.37 | 19.00 | 26.0 | 35.0 | 46.00 | 54.1 | 0 |
| Non-binary or Other | 5 | 29 | 43 | 34.00 | 5.48 | 2.45 | 29.40 | 31.0 | 32.0 | 35.00 | 41.4 | 0 |
| Prefer not to say | 1 | 39 | 39 | 39.00 |  |  | 39.00 | 39.0 | 39.0 | 39.00 | 39.0 | 0 |
|  | 8 | 22 | 49 | 39.12 | 9.79 | 3.46 | 24.45 | 33.5 | 41.5 | 46.75 | 49.0 | 0 |

flextable(genderage\_sumstats) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Age SumStats by Gender") %>%   
 autofit() %>%   
 fit\_to\_width(7.5) %>%   
 set\_header\_labels(values = list(gender = "Gender"))

**Table** : Age SumStats by Gender

| **Gender** | **n** | **min** | **max** | **mean** | **sd** | **se** | **quant5th** | **quant25th** | **quant50th** | **quant75th** | **quant95th** | **na** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Female | 1,064 | 18 | 71 | 36.75 | 11.57 | 0.35 | 20.00 | 26.0 | 38.0 | 47.00 | 54.0 | 1 |
| Male | 939 | 18 | 72 | 35.97 | 11.48 | 0.37 | 19.00 | 26.0 | 35.0 | 46.00 | 54.1 | 0 |
| Non-binary or Other | 5 | 29 | 43 | 34.00 | 5.48 | 2.45 | 29.40 | 31.0 | 32.0 | 35.00 | 41.4 | 0 |
| Prefer not to say | 1 | 39 | 39 | 39.00 |  |  | 39.00 | 39.0 | 39.0 | 39.00 | 39.0 | 0 |
|  | 8 | 22 | 49 | 39.12 | 9.79 | 3.46 | 24.45 | 33.5 | 41.5 | 46.75 | 49.0 | 0 |

agegroup\_sumstats <- headscan\_full %>%   
 group\_by(age\_group) %>%   
 summarise(n = n(),  
 min = min(age, na.rm = TRUE),  
 max = max(age, na.rm = TRUE),  
 mean = mean(age, na.rm = TRUE),  
 sd = sd(age, na.rm = TRUE),  
 se = sd/sqrt(n),  
 quant5th = quantile(age, 0.05, na.rm=TRUE),  
 quant25th = quantile(age, 0.25, na.rm=TRUE),  
 quant50th = quantile(age, 0.50, na.rm=TRUE),  
 quant75th = quantile(age, 0.75, na.rm=TRUE),  
 quant95th = quantile(age, 0.95, na.rm=TRUE),  
 na = sum(is.na(age)))

## Warning in min(age, na.rm = TRUE): no non-missing arguments to min; returning  
## Inf

## Warning in max(age, na.rm = TRUE): no non-missing arguments to max; returning  
## -Inf

agegroup\_sumstats <- agegroup\_sumstats %>%   
 mutate(across(where(is.numeric), round, 2))  
   
agegroup\_sumstats %>%   
 kbl(caption = "Age SumStats by Age Group") %>%   
 kable\_styling(bootstrap\_options = c("striped", "hover", "condensed"), full\_width = TRUE)

Age SumStats by Age Group

age\_group

n

min

max

mean

sd

se

quant5th

quant25th

quant50th

quant75th

quant95th

na

18-36

992

18

36

26.19

5.60

0.18

18.00

21

26

31

35.0

0

37-54

940

37

54

45.41

4.99

0.16

37.95

41

45

50

53.0

0

55-72

84

55

72

55.96

2.94

0.32

55.00

55

55

55

61.4

0

NA

1

Inf

-Inf

NaN

NA

NA

NA

NA

NA

NA

NA

1

flextable(agegroup\_sumstats) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Age SumStats by Age Group")%>%   
 set\_header\_labels(values = list(age\_group = "Age Group"))

**Table** : Age SumStats by Age Group

| **Age Group** | **n** | **min** | **max** | **mean** | **sd** | **se** | **quant5th** | **quant25th** | **quant50th** | **quant75th** | **quant95th** | **na** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 18-36 | 992 | 18 | 36 | 26.19 | 5.60 | 0.18 | 18.00 | 21 | 26 | 31 | 35.0 | 0 |
| 37-54 | 940 | 37 | 54 | 45.41 | 4.99 | 0.16 | 37.95 | 41 | 45 | 50 | 53.0 | 0 |
| 55-72 | 84 | 55 | 72 | 55.96 | 2.94 | 0.32 | 55.00 | 55 | 55 | 55 | 61.4 | 0 |
|  | 1 | Inf | -Inf |  |  |  |  |  |  |  |  | 1 |

#explore race\_eth, frequency of each

#explore gender, frequency of each

PLOTS

ggplot(data=headscan\_full, aes(x= age, fill=gender))+  
 theme(text=element\_text(family= "Times New Roman"))+  
 scale\_fill\_grey(na.value="gray90")+  
 geom\_histogram(binwidth = 3)+  
 labs(title="Headscan Sample Demographics: Age",  
 subtitle="sorted by Race/Ethnicity, color by Gender",  
 y="Count (variable scale per group)",  
 x="Age",  
 fill="Gender")+  
 facet\_wrap(~race\_eth, scales="free\_y")+  
 scale\_y\_continuous(breaks = function(x) unique(floor(pretty(seq(0, (max(x) + 1) \* 1.1)))))+  
 geom\_vline(xintercept = 37, color="red", linetype= "dashed")+  
 geom\_vline(xintercept = 55, color="red", linetype= "dashed")

## Warning: Removed 1 rows containing non-finite values (stat\_bin).

## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): font family not  
## found in Windows font database  
  
## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): font family not  
## found in Windows font database

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database

## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): font family not  
## found in Windows font database

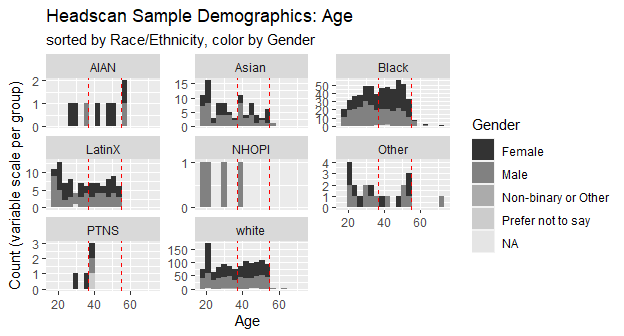
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database

## Warning in grid.Call.graphics(C\_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database



#ggplot(data=headscan\_full)+  
 #geom\_histogram(binwidth=2, aes(x=age, fill=race\_eth))+  
 #scale\_fill\_grey()+  
 #labs(title= "Demographics of Headscan Sample",  
 #subtitle = "Data collected by Human Solutions",  
 #fill= "Ethnicity",  
 #y="Count",  
 #x="Age")

#ggplot(data=headscan\_full)+  
 #geom\_boxplot(aes(x=race\_eth, y=age, fill=gender))+  
 #scale\_fill\_grey()+  
 #theme(axis.text.x = element\_text(angle = 45))