EDA

Allison Shi

'r Sys.Date()

```
library(tidyverse)
library(knitr)
library(broom)
library(nnet) # for multinomial logistic regression
library(patchwork)
```

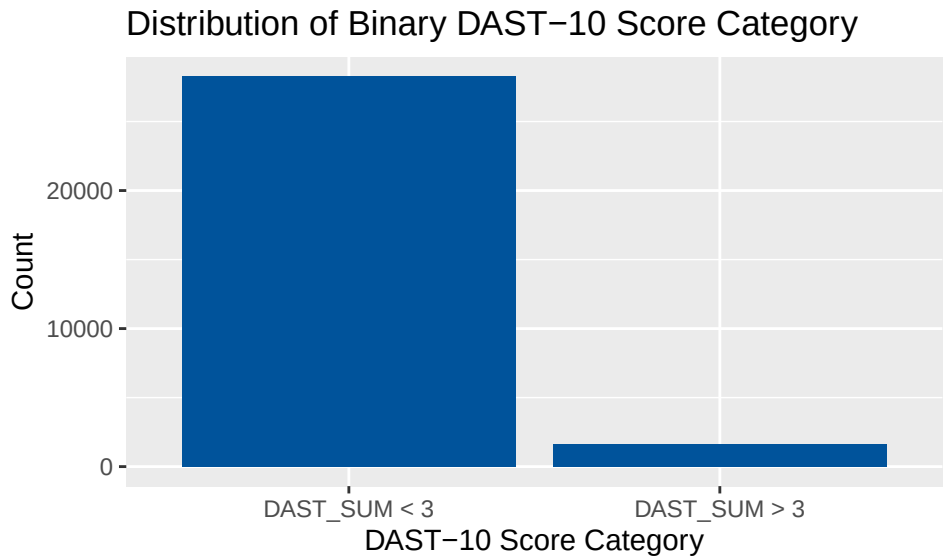
```
us_19 <- read_csv('~/df_data/US/us_19.csv') %>%
  mutate(DAST_binary = if_else(DAST_SUM < 3, 0, 1))
#glimpse(us_19)
```

```
col_names <- colnames(us_19)
us_19 <- lapply(us_19, as.factor)
us_19 <- data.frame(matrix(unlist(us_19), nrow=length(us_19), byrow=TRUE))
us_19 <- data.frame(t(us_19))
colnames(us_19) <- col_names
```

EDA

Distribution of Binary Response Variable

```
ggplot(data = us_19, aes(x = DAST_binary))+
  geom_bar(fill = "#00539B") +
  labs(title = "Distribution of Binary DAST-10 Score Category",
       x = "DAST-10 Score Category", y = "Count") +
  scale_x_discrete(labels=c("0" = "DAST_SUM < 3", "1" = "DAST_SUM > 3"))
```



```
  theme_bw()
```

Distribution of demographic predictor variables

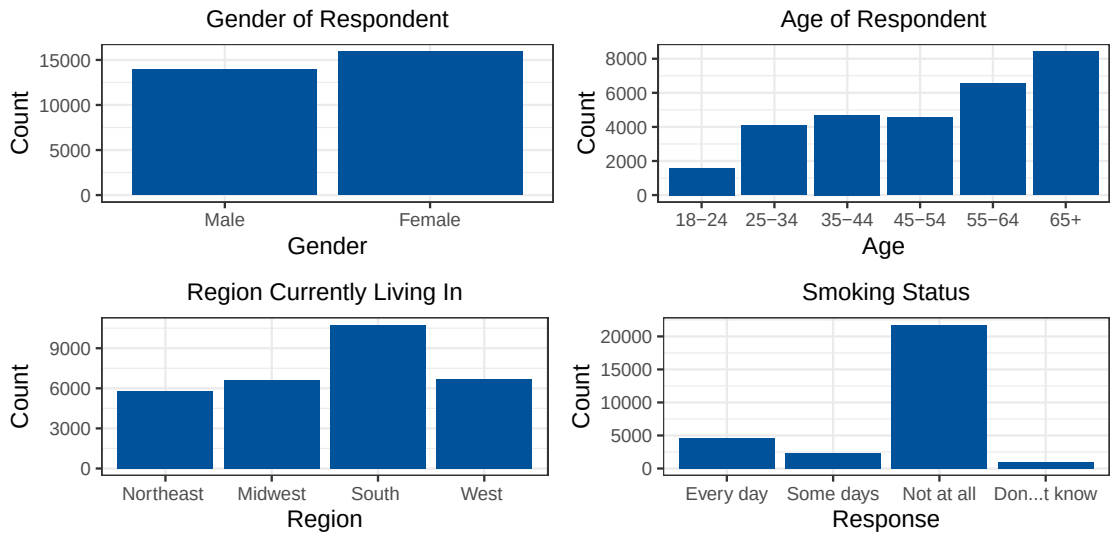
```
gender <- ggplot(data = us_19, aes(x = DEM_GENDER))+
  geom_bar(fill = "#00539B") +
   xlab("Gender") + ylab("Count") +
   ggtitle("Gender of Respondent") +
   scale_x_discrete(labels=c("1" = "Male", "2" = "Female")) +
   theme_bw(base_size = 9) +
   theme(plot.title = element_text(size = 9, hjust = 0.5))

age <- ggplot(data = us_19, aes(x = DEM_AGE10))+
  geom_bar(fill = "#00539B") +
   xlab("Age") + ylab("Count") +
   ggtitle("Age of Respondent") +
   scale_x_discrete(labels=c("1" = "18-24", "2" = "25-34 ", "3" = "35-44",
                             "4" = "45-54", "5" = "55-64", "6" = "65+")) +
   theme_bw(base_size = 9) +
   theme(plot.title = element_text(size = 9, hjust = 0.5))

region <- ggplot(data = us_19, aes(x = DEM_REGION))+
  geom_bar(fill = "#00539B") +
   xlab("Region") + ylab("Count") +
   ggtitle("Region Currently Living In") +
   scale_x_discrete(labels=c("1" = "Northeast ", "2" = "Midwest", "3" = "South ",
                             "4" = "West ")) +
   theme_bw(base_size = 9) +
   theme(plot.title = element_text(size = 9, hjust = 0.5))

income <- ggplot(data = us_19, aes(x = DEM_INCOME))+
  geom_bar(fill = "#00539B") +
   xlab("Income (in $)") + ylab("Count") +
   ggtitle("Combined Household Income in the Last 12 Months") +
   scale_x_discrete(labels=c("1" = "< 25,000", "2" = "25,000-49,999", "3" = "50,000-74,999",
                             "4" = "75,000-99,999", "5" = "> 100,000")) +
   theme_bw(base_size = 9) +
   theme(plot.title = element_text(size = 9, hjust = 0.5))

smoking <- ggplot(data = us_19, aes(x = TOB_LIFE))+
  geom_bar(fill = "#00539B") +
   xlab("Response") + ylab("Count") +
   ggtitle("Smoking Status") +
   scale_x_discrete(labels=c("1" = "Every day", "2" = "Some days", "3" = "Not at all",
                             "4" = "Don't know")) +
   theme_bw(base_size = 9) +
   theme(plot.title = element_text(size = 9, hjust = 0.5))

(gender | age ) /
  (region | smoking)
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don't know' in 'mbcsToSbcs': dot substituted for <e2>
```

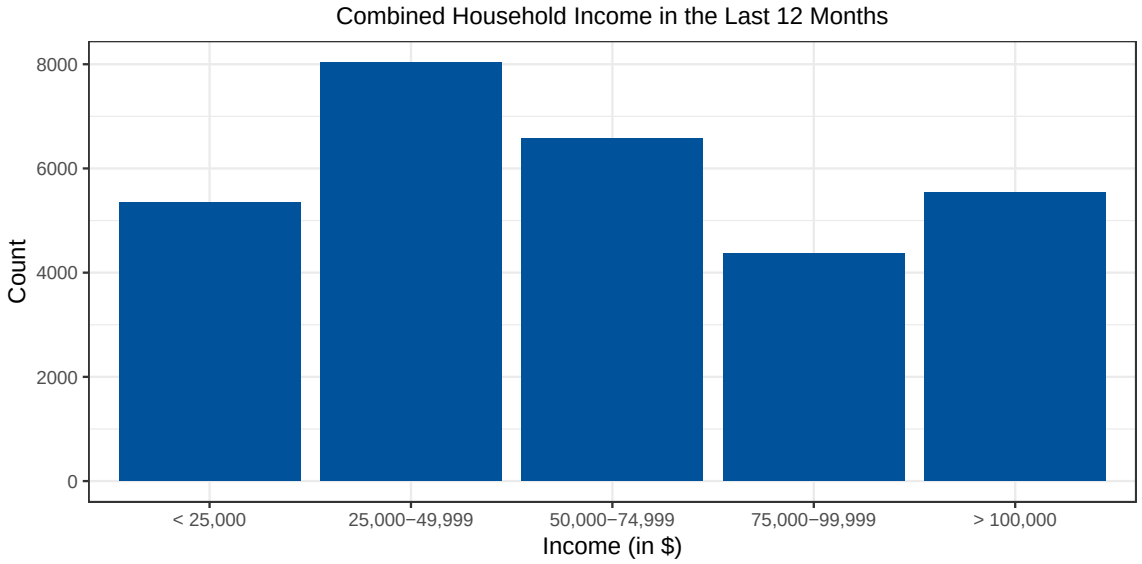
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

```
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Don’t know' in 'mbcsToSbcs': dot substituted for <99>
```



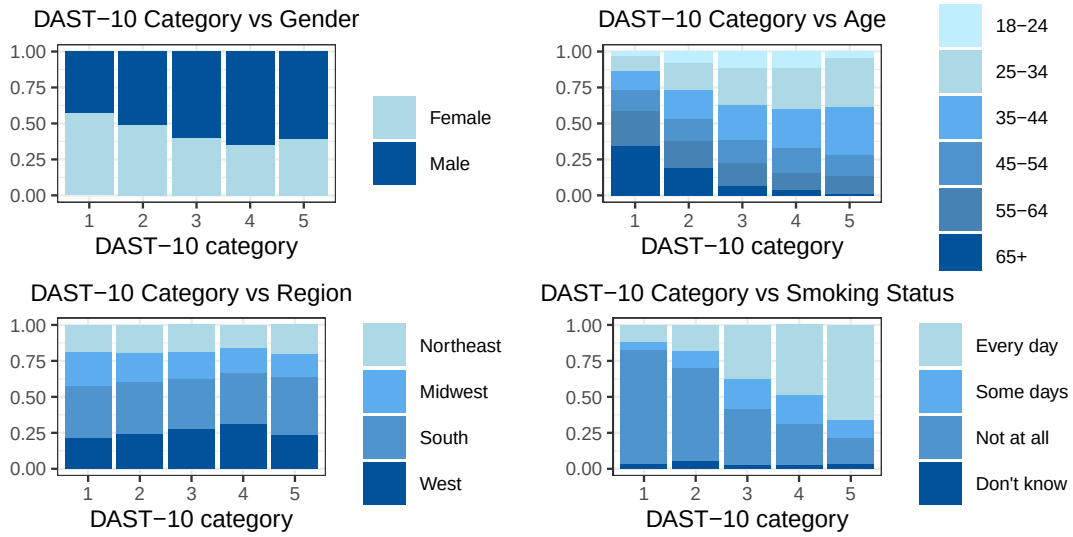

```
income
```



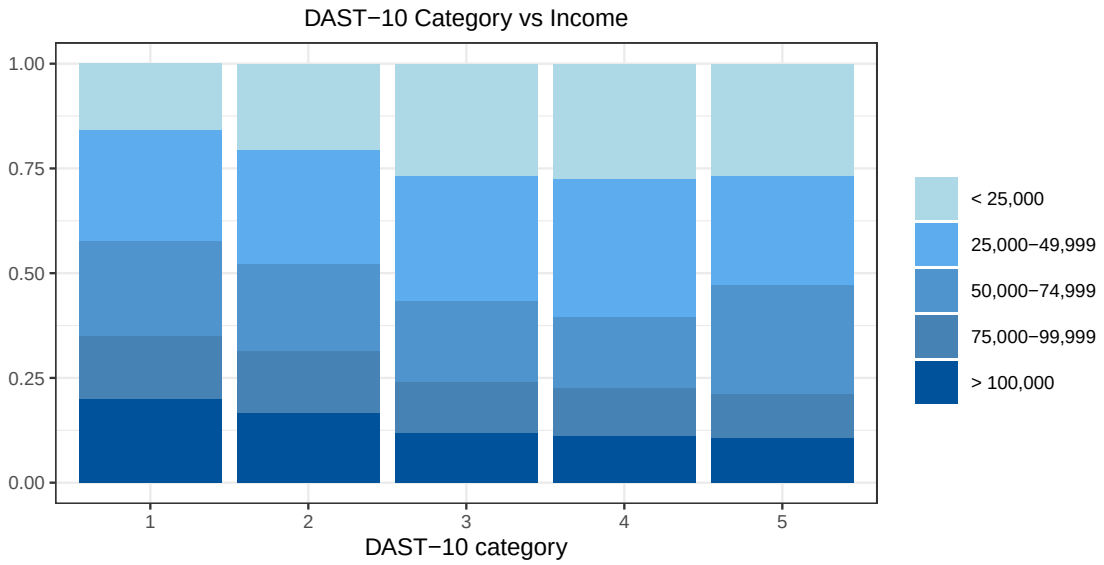
Relationship between demographic predictor variables and response variable

```
gender_dast_10 <- ggplot(data = us_19, mapping = aes(x = DAST_CAT, fill = DEM_GENDER)) +
  geom_bar(position = "fill") +
  labs(title = "DAST-10 Category vs Gender",
       x = "DAST-10 category", y = "") +
  theme_bw(base_size = 9) +
  theme(plot.title = element_text(size = 9, hjust = 0.5)) +
  scale_fill_manual(values=c("#00539B", "lightblue"), name = "", labels = c("Male", "Female"), guide = g

age_dast_10 <- ggplot(data = us_19, mapping = aes(x = DAST_CAT, fill = DEM_AGE10)) +
  geom_bar(position = "fill") +
  labs(title = "DAST-10 Category vs Age",
       x = "DAST-10 category", y = "") +
  theme_bw(base_size = 9) +
  theme(plot.title = element_text(size = 9, hjust = 0.5)) +
  scale_fill_manual(values=c("lightblue1", "lightblue", "steelblue2", "steelblue3", "steelblue", "#0053
```

```
region_dast_10 <- ggplot(data = us_19, mapping = aes(x = DAST_CAT, fill = DEM_REGION)) +
  geom_bar(position = "fill") +
  labs(title = "DAST-10 Category vs Region",
       x = "DAST-10 category", y = "") +
  theme_bw(base_size = 9) +
  theme(plot.title = element_text(size = 9, hjust = 0.5)) +
  scale_fill_manual(values=c("lightblue", "steelblue2", "steelblue3", "#00539B"), name = "", labels = c

income_dast_10 <- ggplot(data = us_19, mapping = aes(x = DAST_CAT, fill = DEM_INCOME)) +
  geom_bar(position = "fill") +
  labs(title = "DAST-10 Category vs Income",
       x = "DAST-10 category", y = "") +
  theme_bw(base_size = 9) +
  theme(plot.title = element_text(size = 9, hjust = 0.5)) +
  scale_fill_manual(values=c("lightblue", "steelblue2", "steelblue3", "steelblue", "#00539B"), name = "

smoking_dast_10 <- ggplot(data = us_19, mapping = aes(x = DAST_CAT, fill = TOB_LIFE)) +
  geom_bar(position = "fill") +
  labs(title = "DAST-10 Category vs Smoking Status",
       x = "DAST-10 category", y = "") +
  theme_bw(base_size = 9) +
  theme(plot.title = element_text(size = 9, hjust = 0.5)) +
  scale_fill_manual(values=c("lightblue", "steelblue2", "steelblue3", "#00539B"), name = "", labels = c

(gender_dast_10 | age_dast_10 ) /
  (region_dast_10 | smoking_dast_10)
```




```
income_dast_10
```


DAST−10 Category vs Income
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
DAST−10 category
< 25,000
25,000−49,999
50,000−74,999
75,000−99,999
> 100,000