Diabetes-Eye Exam Analysis

Mridula (Mally) Shan - Summer 2024

Introduction & Code Set-Up

```
library(tidyverse)
library(dplyr)
library(infer)
library(readr)
library(broom)
library(gglm)
library(MASS)
library(AICcmodavg)
library(finalfit)
library(knitr)
library(lme4)
library(tinytex)

# (For Reference) Code to Reformat the Data for Analysis
# Collected data from the following link:
#https://www.cdc.gov/nchs/nhis/2022nhis.htm
# surveying 27,651 people throughout the United States

NHIS_Diabetes_Eye <- read_csv("adult22.csv")

vars_names <- colnames(NHIS_Diabetes_Eye)

#checking if variables of interest is actually in the dataset!
#"REGION" %in% vars_names

#if seeing the following error: Error in select(., REGION) : unused argument (REGION)
#then you need to specify the dplyr package in front of the command
```

Descriptive Statistics

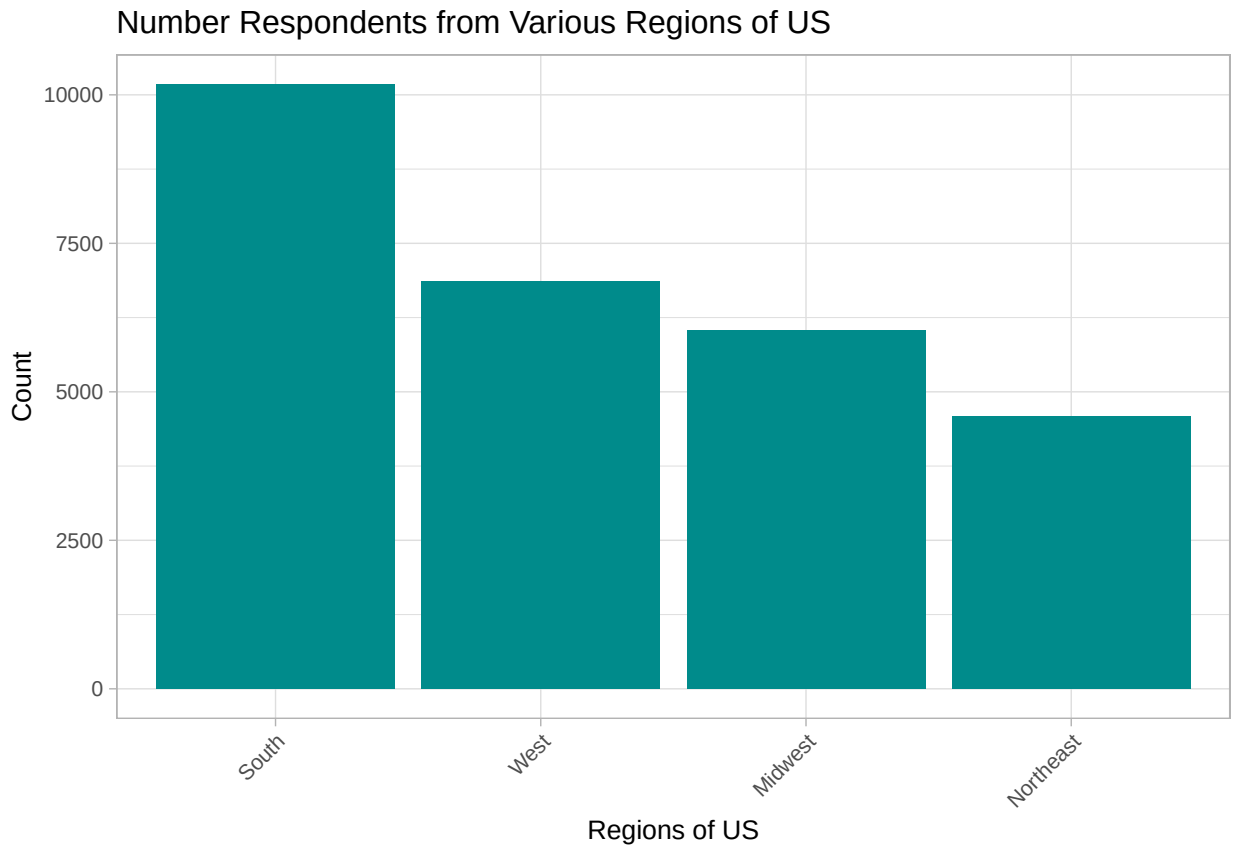
1) Number Respondents from Various Regions of US

```
geographic_summary <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(REGION = case_when(REGION == 1 ~ "Northeast",
                                   REGION == 2 ~ "Midwest",
                                   REGION == 3 ~ "South",
                                   REGION == 4 ~ "West")) %>%
  dplyr::count(REGION) %>%
  dplyr::mutate(perc_region = (n/sum(n)*100)) %>%
  arrange(desc(perc_region))
```

```r
kable(geographic_summary)
```

REGION	n	perc_region
South	10171	36.78348
West	6857	24.79838
Midwest	6033	21.81838
Northeast	4590	16.59976

```r
geographic_summary %>%
  ggplot(aes(x = reorder(REGION, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Number Respondents from Various Regions of US ",
       x ="Regions of US",
       y = "Count")
```



2) Number Respondents from Various County Types in US

The 2013 NCHS Urban-Rural Classification scheme defines:

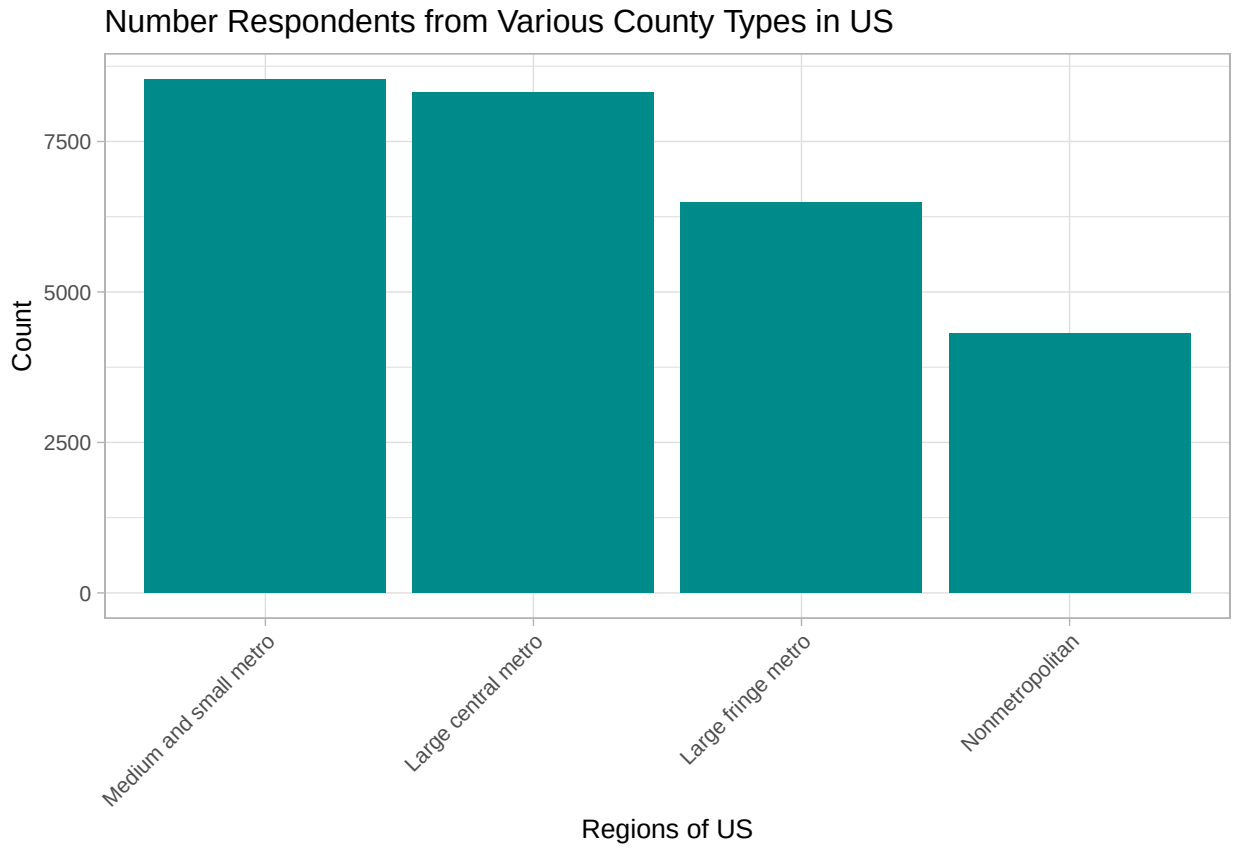
- **Large central metro** as counties (or county equivalents) in metropolitan statistical areas (MSA) of 1 million or more population that 1) contain the entire population of the largest principal city of the MSA, or 2) are completely contained in the largest principal city of the MSA, or 3) contain at least 250,000 residents of any principal city of the MSA.
- **Large fringe metro** are counties (or county equivalents) in MSAs of 1 million or more population that do not qualify as large central.
- **Medium and small metro** are counties (or county equivalents) in MSAs of 250,000 to 999,999 population or in MSAs of less than 250,000 population.
- **Nonmetropolitan** are counties (or county equivalents) in micropolitan statistical areas and non-core counties.

```
urban_summary <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(URBRRL = case_when(URBRRL == 1 ~ "Large central metro",
                                   URBRRL == 2 ~ "Large fringe metro",
                                   URBRRL == 3 ~ "Medium and small metro",
                                   URBRRL == 4 ~ "Nonmetropolitan")) %>%
  dplyr::count(URBRRL) %>%
  dplyr::mutate(perc_urban = (n/sum(n)*100)) %>%
  arrange(desc(perc_urban))

kable(urban_summary)
```

URBRRL	n	perc_urban
Medium and small metro	8539	30.88134
Large central metro	8321	30.09294
Large fringe metro	6480	23.43496
Nonmetropolitan	4311	15.59076

```
urban_summary %>%
  ggplot(aes(x = reorder(URBRRL, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Number Respondents from Various County Types in US ",
       x ="Regions of US",
       y = "Count")
```



3) Distribution of Respondents' Sex

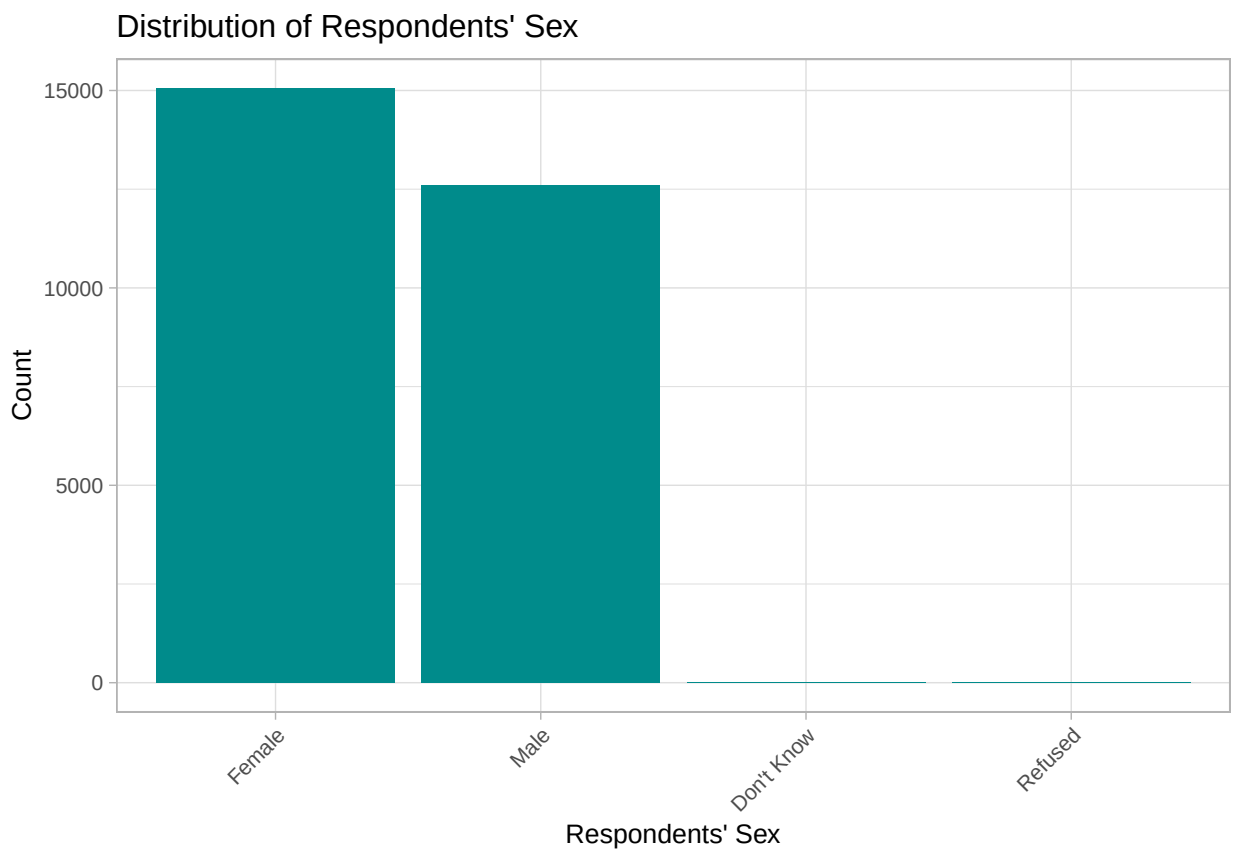
```
dist_sex <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(SEX_A = case_when(SEX_A == 1 ~ "Male",
                                  SEX_A == 2 ~ "Female",
                                  SEX_A == 7 ~ "Refused",
                                  SEX_A == 8 ~ "Not Ascertained",
                                  SEX_A == 9 ~ "Don't Know")) %>%
  dplyr::count(SEX_A) %>%
  dplyr::mutate(perc_sex = (n/sum(n)*100)) %>%
  arrange(desc(perc_sex))

kable(dist_sex)
```

SEX_A	n	perc_sex
Female	15050	54.4284113
Male	12598	45.5607392
Don't Know	2	0.0072330
Refused	1	0.0036165

```
dist_sex %>%
  ggplot(aes(x = reorder(SEX_A, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
  theme(text = element_text(size = 8),
```

```
              axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
              axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title= "Distribution of Respondents' Sex",
       x ="Respondents' Sex",
       y = "Count")
```



4) Distribution of Respondents' Race

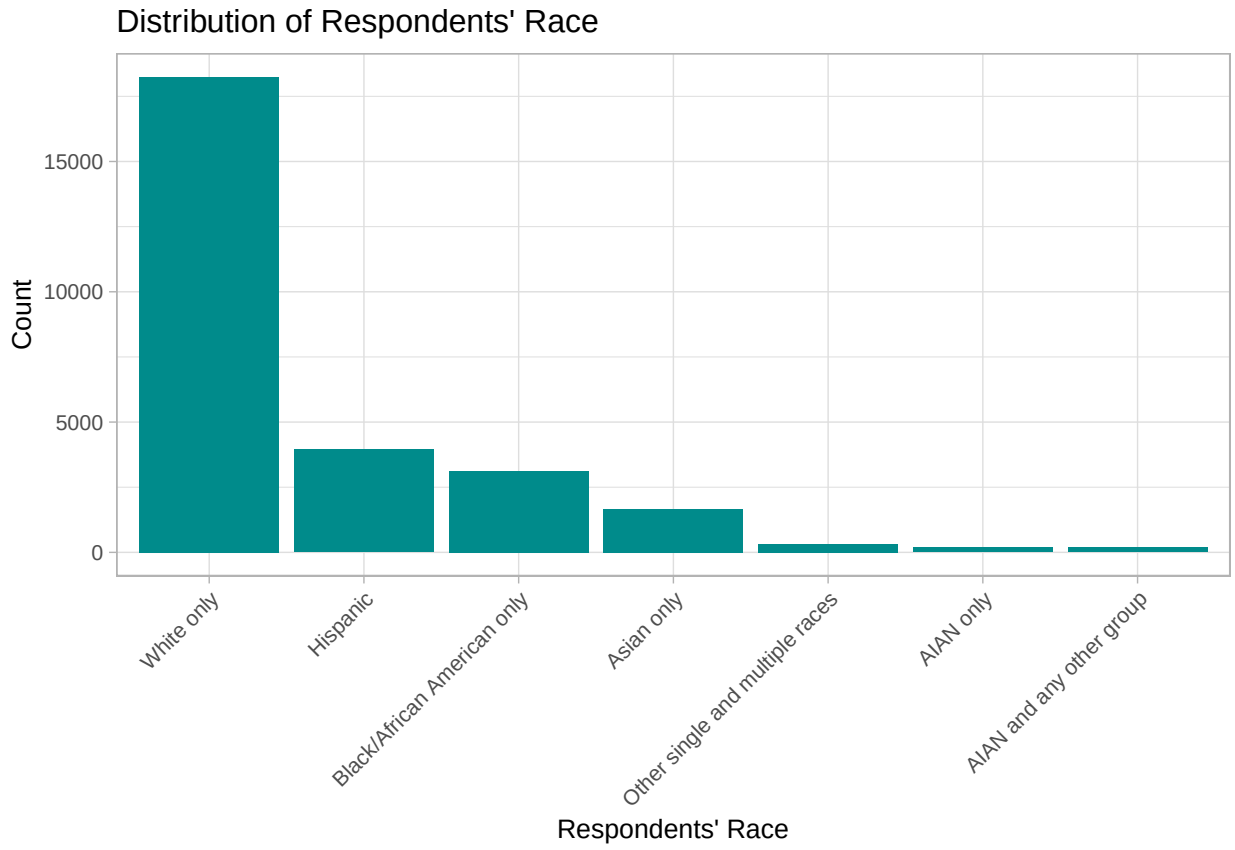
```
dist_race <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(RACEALLP_A = case_when(HISPALLP_A == 1 ~ "Hispanic",
                                       HISPALLP_A == 2 ~ "White only",
                                       HISPALLP_A == 3 ~ "Black/African American only",
                                       HISPALLP_A == 4 ~ "Asian only",
                                       HISPALLP_A == 5 ~ "AIAN only",
                                       HISPALLP_A == 6 ~ "AIAN and any other group",
                                       HISPALLP_A == 7 ~ "Other single and multiple races",
                                       HISPALLP_A == 97 ~ "Refused",
                                       HISPALLP_A == 98 ~ "Not Ascertained",
                                       HISPALLP_A == 98 ~ "Don't know")) %>%
  dplyr::count(RACEALLP_A) %>%
  dplyr::mutate(perc_race = (n/sum(n)*100)) %>%
```

```
  arrange(desc(perc_race))

kable(dist_race)
```

RACEALLP_A	n	perc_race
White only	18242	65.9722976
Hispanic	3943	14.2598821
Black/African American only	3112	11.2545658
Asian only	1663	6.0142490
Other single and multiple races	319	1.1536653
AIAN only	187	0.6762866
AIAN and any other group	185	0.6690536

```
dist_race %>%
  ggplot(aes(x = reorder(RACEALLP_A, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Distribution of Respondents' Race ",
       x ="Respondents' Race",
       y = "Count")
```



5) Number of Respondents that had an Eye Exam

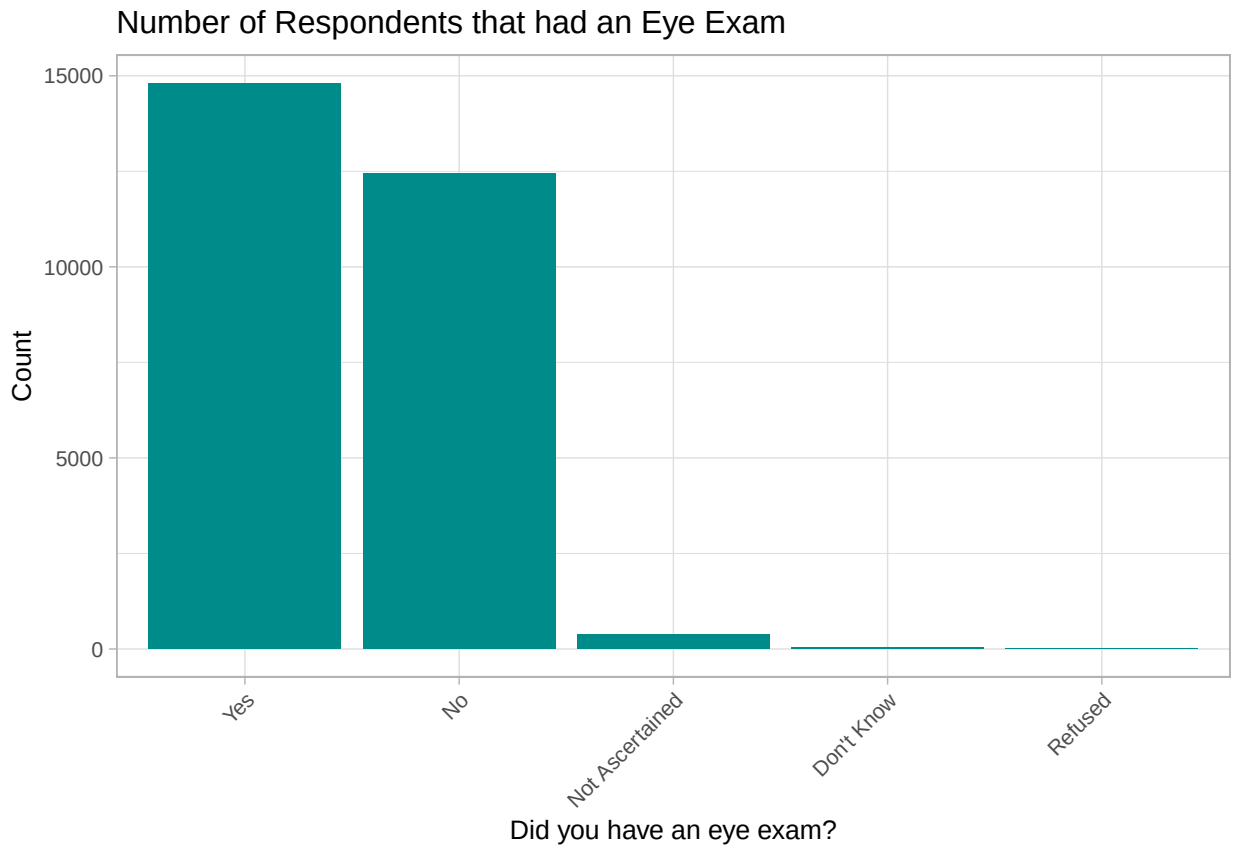
```
num_eye_exam <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::count(EYEEX12M_A) %>%
  dplyr::mutate(perc_eye_exam = (n/sum(n)*100)) %>%
  arrange(desc(perc_eye_exam))

kable(num_eye_exam)
```

EYEEX12M_A	n	perc_eye_exam
Yes	14812	53.5676829
No	12434	44.9676323
Not Ascertained	370	1.3381071
Don't Know	32	0.1157282
Refused	3	0.0108495

```
num_eye_exam %>%
  ggplot(aes(x = reorder(EYEEX12M_A, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
```

```
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Number of Respondents that had an Eye Exam",
       x ="Did you have an eye exam?",
       y = "Count")
```



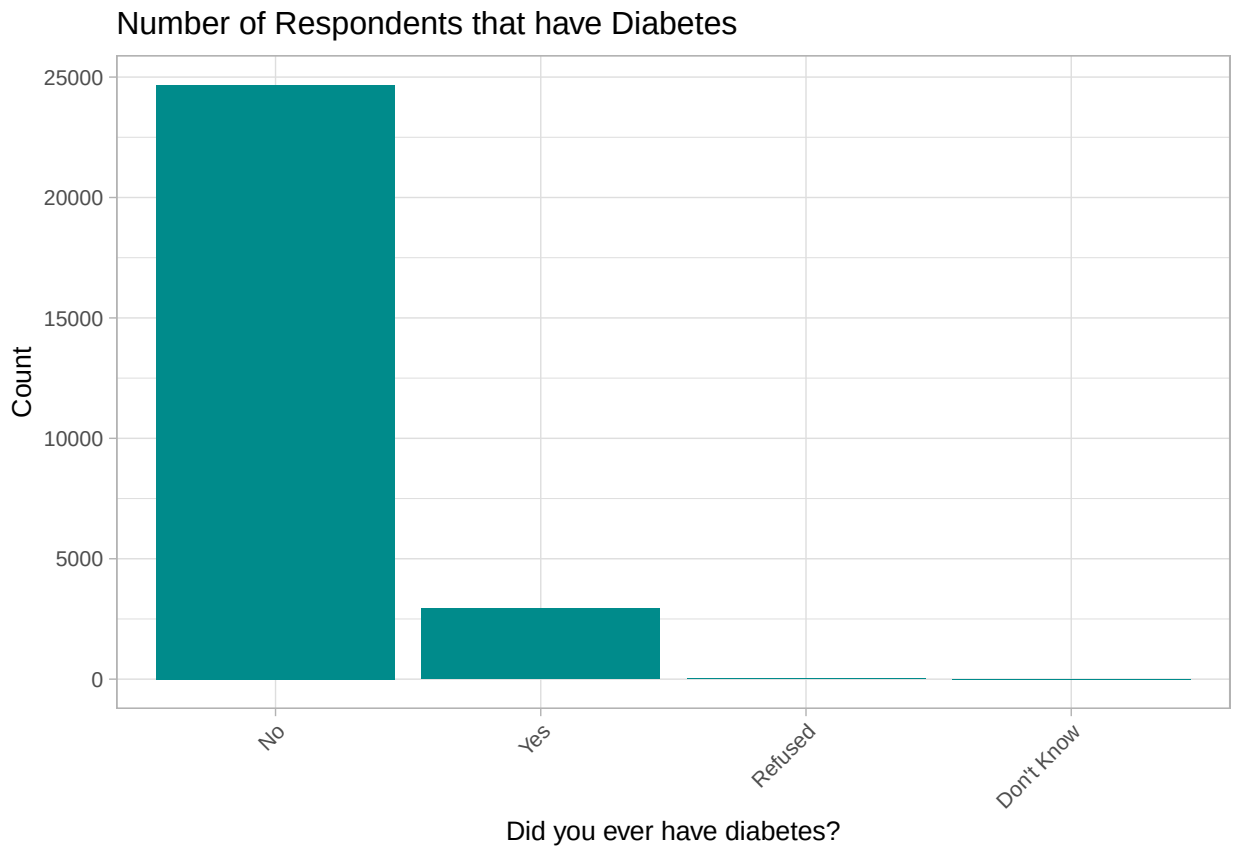
6) Number of Respondents that have Diabetes

```
num_diabetes <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(DIBEV_A = case_when(DIBEV_A == 1 ~ "Yes",
                                    DIBEV_A == 2 ~ "No",
                                    DIBEV_A == 7 ~ "Refused",
                                    DIBEV_A == 8 ~ "Not Ascertained",
                                    DIBEV_A == 9 ~ "Don't Know")) %>%
  dplyr::count(DIBEV_A) %>%
  dplyr::mutate(perc_diabetes = (n/sum(n)*100)) %>%
  arrange(desc(perc_diabetes))

kable(num_diabetes)
```

DIBEV_A	n	perc_diabetes
No	24673	89.2300459
Yes	2946	10.6542259
Refused	23	0.0831796
Don't Know	9	0.0325486

```
num_diabetes %>%
  ggplot(aes(x = reorder(DIBEV_A, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Number of Respondents that have Diabetes",
       x ="Did you ever have diabetes?",
       y = "Count")
```



Exploratory Analysis

1) Number of Respondents with Diabetes who got an Eye Exam

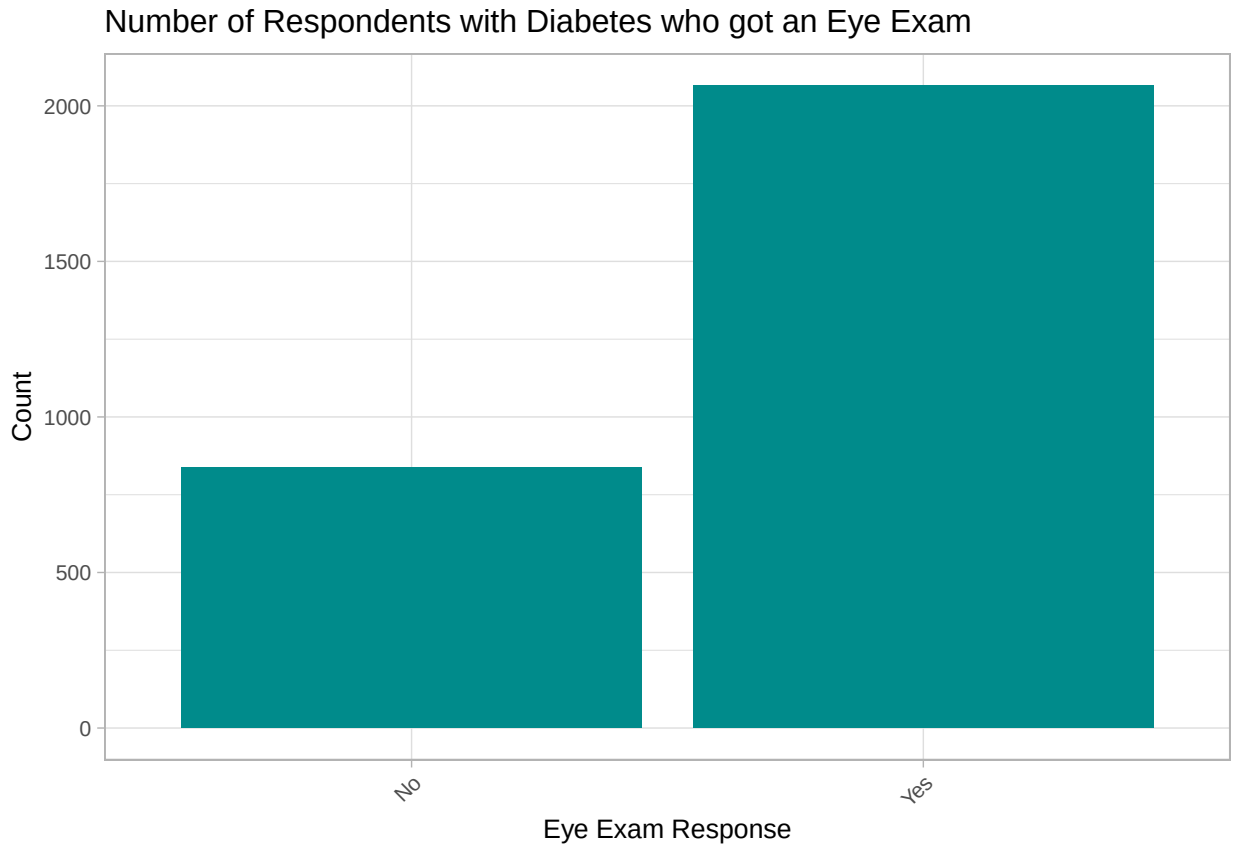
```
has_diabetes_tbl <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(DIBEV_A = case_when(DIBEV_A == 1 ~ "Yes",
                                       DIBEV_A == 2 ~ "No",
                                       DIBEV_A == 7 ~ "Refused",
                                       DIBEV_A == 8 ~ "Not Ascertained",
                                       DIBEV_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"), DIBEV_A %in% c("Yes", "No")) %>%
  dplyr::group_by(DIBEV_A) %>%
  dplyr::count(EYEEX12M_A) %>%
  dplyr::mutate(perc_eye_exam = (n/sum(n)*100)) %>%
  rename(Eye_Exam = EYEEX12M_A,
         Has_Diabetes = DIBEV_A)


kable(has_diabetes_tbl)
```

Has_Diabetes	Eye_Exam	n	perc_eye_exam
No	No	11584	47.64529
No	Yes	12729	52.35471
Yes	No	838	28.86669
Yes	Yes	2065	71.13331

```
NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(DIBEV_A = case_when(DIBEV_A == 1 ~ "Yes",
                                       DIBEV_A == 2 ~ "No",
                                       DIBEV_A == 7 ~ "Refused",
                                       DIBEV_A == 8 ~ "Not Ascertained",
                                       DIBEV_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"), DIBEV_A %in% c("Yes")) %>%
  ggplot() +
  geom_bar(mapping = aes(x = EYEEX12M_A),
           fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
```

```
  labs(title="Number of Respondents with Diabetes who got an Eye Exam",
       x ="Eye Exam Response",
       y = "Count")
```



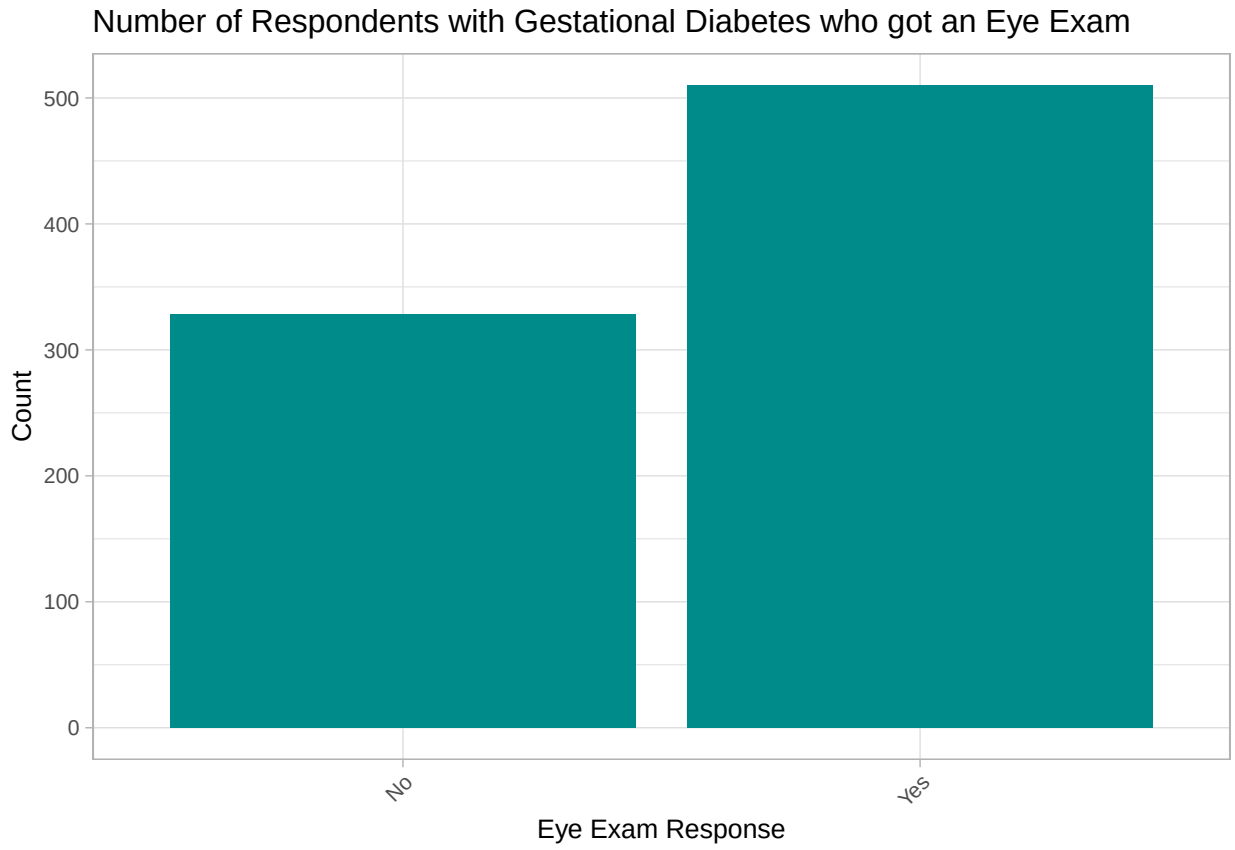
2) Number of Respondents with Gestational Diabetes who got an Eye Exam

```
#Gestational diabetes
gestational_type_tbl <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(GESDIB_A = case_when(GESDIB_A == 1 ~ "Yes",
                                     GESDIB_A == 2 ~ "No",
                                     GESDIB_A == 7 ~ "Refused",
                                     GESDIB_A == 8 ~ "Not Ascertained",
                                     GESDIB_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"), GESDIB_A %in% c("Yes")) %>%
  dplyr::group_by(GESDIB_A) %>%
  dplyr::count(EYEEX12M_A) %>%
  dplyr::mutate(perc_eye_exam = (n/sum(n)*100)) %>%
  rename(Eye_Exam = EYEEX12M_A,
         Gestational_Diabetes = GESDIB_A)
```

```
kable(gestational_type_tbl)
```

Gestational_Diabetes	Eye_Exam	n	perc_eye_exam
Yes	No	328	39.14081
Yes	Yes	510	60.85919

```
NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(GESDIB_A = case_when(GESDIB_A == 1 ~ "Yes",
                                     GESDIB_A == 2 ~ "No",
                                     GESDIB_A == 7 ~ "Refused",
                                     GESDIB_A == 8 ~ "Not Ascertained",
                                     GESDIB_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"), GESDIB_A %in% c("Yes")) %>%
  ggplot(mapping = aes(x = EYEEX12M_A)) +
  geom_bar(fill = "darkcyan") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title = "Number of Respondents with Gestational Diabetes who got an Eye Exam",
       x ="Eye Exam Response",
       y = "Count")
```



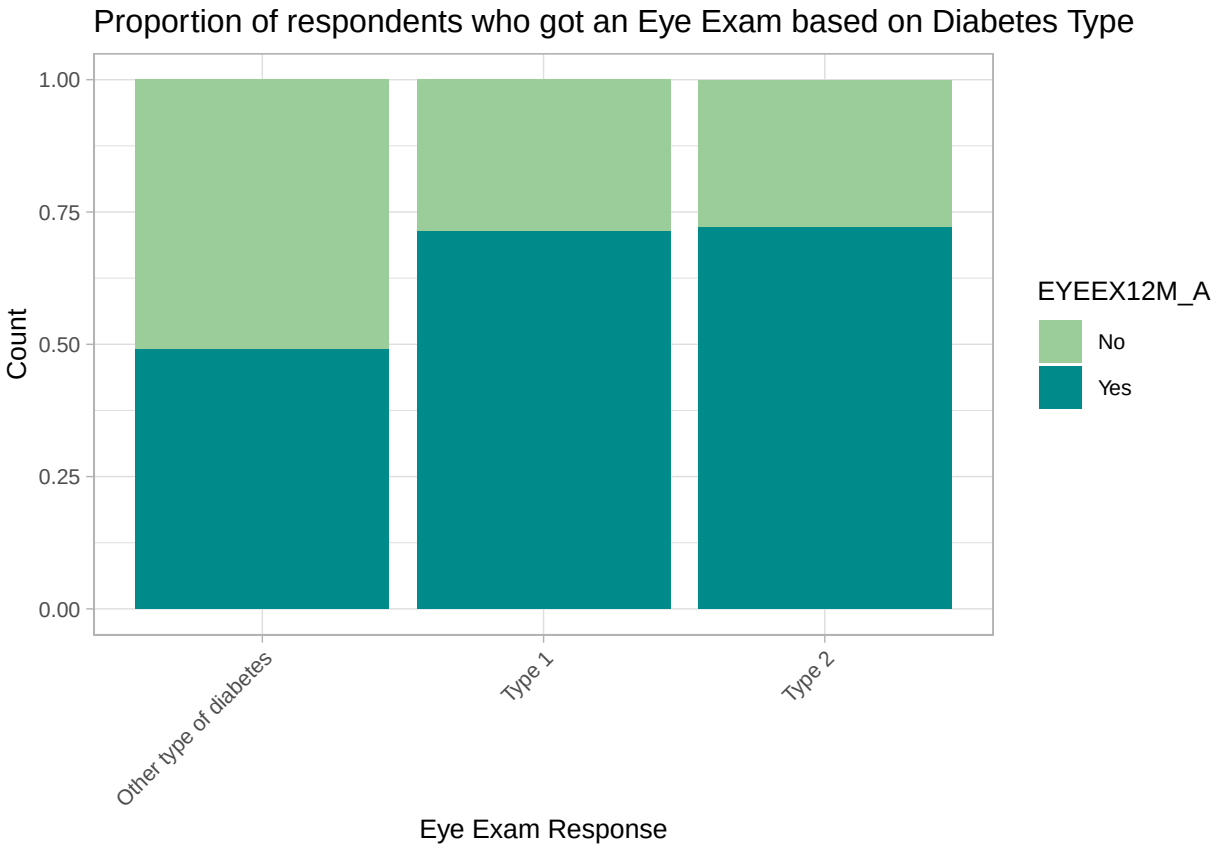
3) Proportion of respondents who got an Eye Exam based on Diabetes type

```
diabetes_type_tbl <- NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(DIBTYPE_A = case_when(DIBTYPE_A == 1 ~ "Type 1",
                                      DIBTYPE_A == 2 ~ "Type 2",
                                      DIBTYPE_A == 3 ~ "Other type of diabetes",
                                      DIBTYPE_A == 7 ~ "Refused",
                                      DIBTYPE_A == 8 ~ "Not Ascertained",
                                      DIBTYPE_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"),
         DIBTYPE_A %in% c("Type 1", "Type 2", "Other type of diabetes")) %>%
  dplyr::group_by(DIBTYPE_A) %>%
  dplyr::count(EYEEX12M_A) %>%
  dplyr::mutate(perc_eye_exam = (n/sum(n)*100)) %>%
  rename(Diabetes_Type = DIBTYPE_A,
         Eye_Exam = EYEEX12M_A)

kable(diabetes_type_tbl)
```

Diabetes_Type	Eye_Exam	n	perc_eye_exam
Other type of diabetes	No	29	50.87719
Other type of diabetes	Yes	28	49.12281
Type 1	No	68	28.57143
Type 1	Yes	170	71.42857
Type 2	No	686	27.89752
Type 2	Yes	1773	72.10248

```
NHIS_Diabetes_Eye %>%
  dplyr::mutate(EYEEX12M_A = case_when(EYEEX12M_A == 1 ~ "Yes",
                                       EYEEX12M_A == 2 ~ "No",
                                       EYEEX12M_A == 7 ~ "Refused",
                                       EYEEX12M_A == 8 ~ "Not Ascertained",
                                       EYEEX12M_A == 9 ~ "Don't Know")) %>%
  dplyr::mutate(DIBTYPE_A = case_when(DIBTYPE_A == 1 ~ "Type 1",
                                      DIBTYPE_A == 2 ~ "Type 2",
                                      DIBTYPE_A == 3 ~ "Other type of diabetes",
                                      DIBTYPE_A == 7 ~ "Refused",
                                      DIBTYPE_A == 8 ~ "Not Ascertained",
                                      DIBTYPE_A == 9 ~ "Don't Know")) %>%
  filter(EYEEX12M_A %in% c("Yes", "No"),
         DIBTYPE_A %in% c("Type 1", "Type 2", "Other type of diabetes")) %>%
  ggplot(mapping = aes(x = DIBTYPE_A, fill = EYEEX12M_A)) +
  geom_bar(position = "fill") +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme(text = element_text(size = 8),
                axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  theme_light() +
  scale_fill_manual(values = c("darkseagreen3",
                               "darkcyan")) +
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title="Proportion of respondents who got an Eye Exam based on Diabetes Type",
       x ="Eye Exam Response",
       y = "Count")
```


Proportion of respondents who got an Eye Exam based on Diabetes Type
1.00
0.75
0.50
0.25
0.00
Count
Other type of diabetes
Type 1
Type 2
Eye Exam Response
EYEEX12M_A
No
Yes