Final Solution

```
library(dplyr)
library(ggplot2)
library(lubridate)
```

```
member <- read.csv("https://collinn.github.io/data/member.csv")
claims <- read.csv("https://collinn.github.io/data/claims.csv")
```

Inclusion criteria (age)

Here, we inner_join claims with ICD codes O80 and O82, as we only want women who had a delivery. We then determine age with `lubridate::interval()` and filter to include those older than 18 and younger than 40

```
cc <- subset(claims, ICD %in% c("O80", "O82"))
full <- inner_join(member, cc, by = "id")

## Determine age using lubridate
full <- mutate(full, age = interval(DoB, date) / years(1)) %>%
  filter(age >= 18 & age <= 40) %>% arrange(id, year, month)
```

Inclusion criteria (insurance coverage)

Here, we need to be sure that they are covered for 3 months after and 9 months before. We will do this with two functions

```
happen_after <- function(y, m, d, n) {
  m <- m + 12*(y == 2018)
  event <- month(d) + 12*(year(d) == 2018)
  max(m) - max(event) >= n
}

happen_before <- function(y, m, d, n) {
  m <- m + 12*(y == 2018)
  event <- month(d) + 12*(year(d) == 2018)
  min(event) - min(m) >= n
}

full <- group_by(full, id) %>%
  mutate(ha = happen_after(year, month, date, 3),
         hb = happen_before(year, month, date, 9),
         in_interval = ha & hb) %>%
  filter(in_interval) # keep only those where in_interval == TRUE

## Inclusion criteria leaves us with 4240 individuals
length(unique(full$id))
```

```
## [1] 4240
```

Delivery type

We can get this from the data as we have it now using the ICD code. After, let's remove variables we no longer need in preparation for another join

```
full <- mutate(full,
               birth_type = ICD) %>%
  select(-ha, -hb, -in_interval, -date, -CPT, -ICD)
```

Indicator variables

Here we construct indicators for opioid, migraine, and mental health status. We do this with a left join, keeping all of our member data but adding their claims information. Using the ICD code data, we will create boolean for each of our indicators

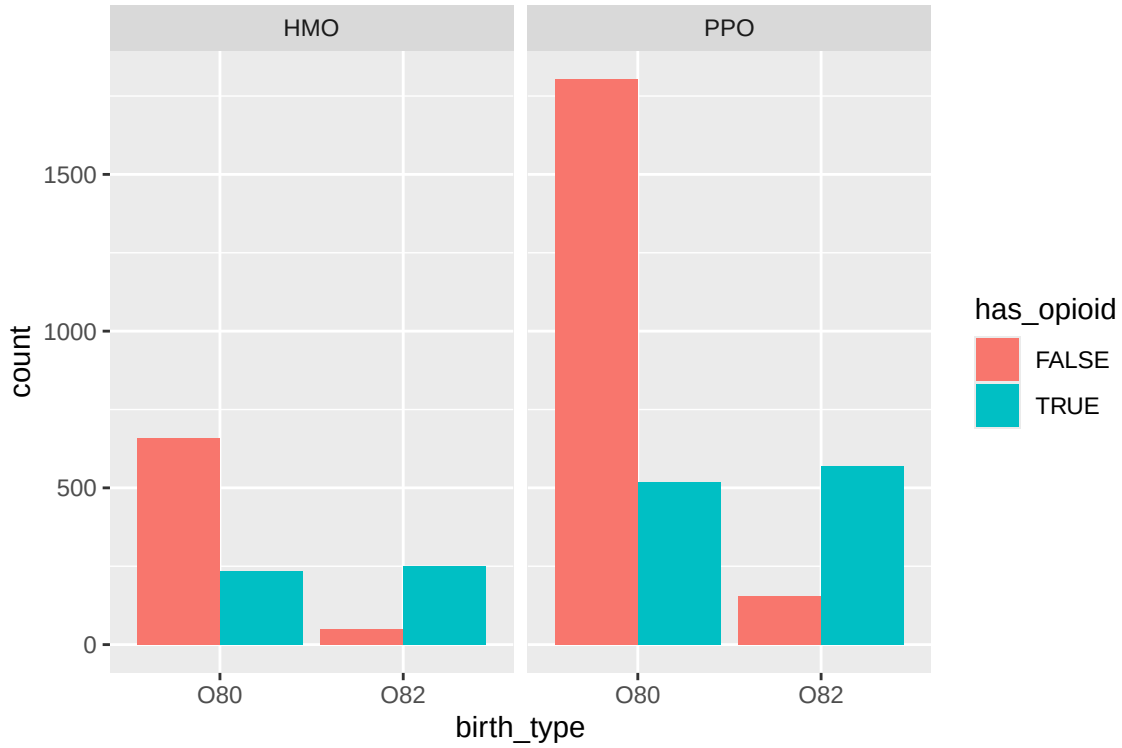
```
# Add ICD data to our members (we will get warning, that's fine)
full <- left_join(full, claims, by = "id")

# Now let's create our variables
full <- group_by(full, id) %>%
  mutate(has_opioid = "J0745" %in% ICD,
         has_migraine = "G43.4" %in% ICD,
         has_anxiety = "F41.8" %in% ICD,
         has_depression = "F32.3" %in% ICD,
         anx_and_dep = has_anxiety & has_depression,
         anx_or_dep = has_anxiety | has_depression) %>%
  select(id, PlanType, age, birth_type, has_opioid, has_anxiety, has_depression,
         anx_and_dep, anx_or_dep, has_migraine) %>% unique()


head(full)
```

```
## # A tibble: 6 x 10
## # Groups:   id [6]
##   id        PlanType   age birth_type has_opioid has_anxiety has_depression
##   <chr>     <chr>    <dbl> <chr>      <lgl>      <lgl>       <lgl>
## 1 PID234671 PPO       32.0 O82        TRUE       FALSE       FALSE
## 2 PID235034 PPO       39.4 O80        TRUE       FALSE       FALSE
## 3 PID235286 HMO       22.9 O82        TRUE       FALSE       FALSE
## 4 PID235455 PPO       29.5 O80        FALSE      FALSE       FALSE
## 5 PID235675 HMO       23.5 O80        FALSE      FALSE       FALSE
## 6 PID235692 HMO       29.8 O80        TRUE       FALSE       FALSE
## # i 3 more variables: anx_and_dep <lgl>, anx_or_dep <lgl>, has_migraine <lgl>
```

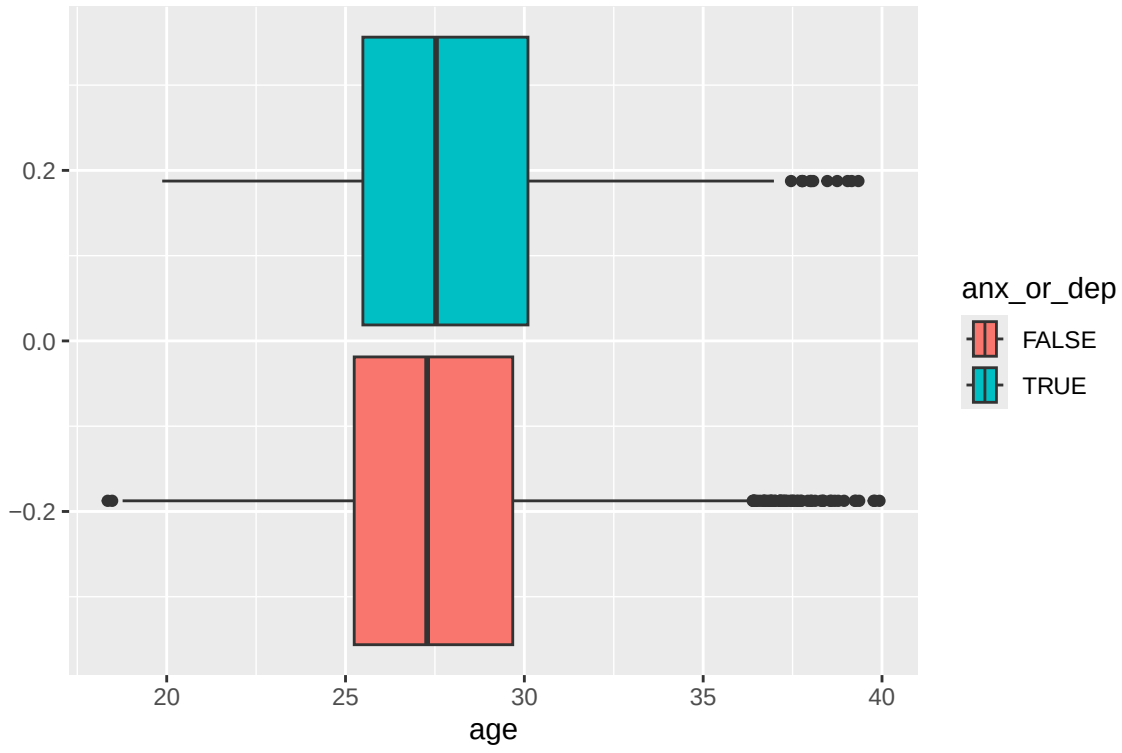
Create exploratory plots

Here, we are just looking for 2-3 plots with some description of what they entail. Here are a few possibilities without description

```
# Plot 1
ggplot(full, aes(x = birth_type, fill = has_opioid)) + geom_bar(position = "dodge") +
  facet_wrap(~PlanType)
```

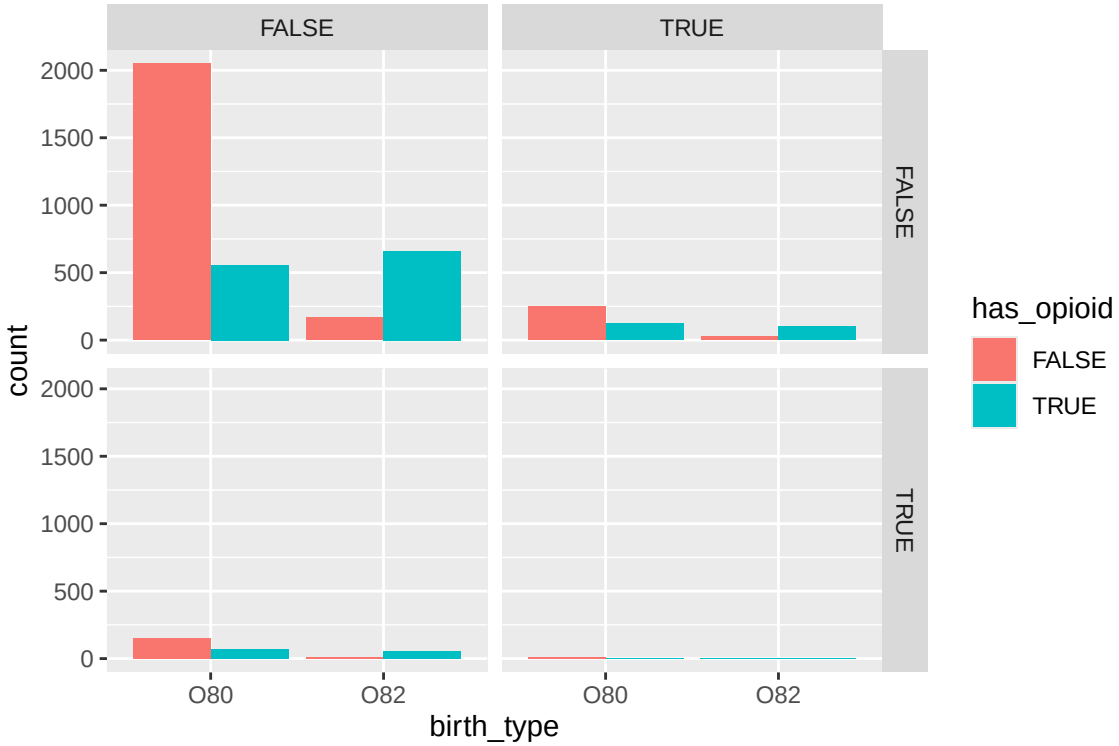



```
# Plot 2
ggplot(full, aes(x = age, fill = anx_or_dep)) + geom_boxplot()
```




```
# Plot 3
ggplot(full, aes(x = birth_type, fill = has_opioid)) + geom_bar(position = "dodge") +
```

```
  facet_grid(has_anxiety~has_depression)
```



Build model

This can really be whatever. They should find that mental health diagnosis increases odds of opioid prescription.

```
fit <- glm(has_opioid ~ age + birth_type + PlanType + anx_or_dep,
           data = full, family = binomial)
exp(coef(fit))
```

```
##    (Intercept)            age  birth_typeO82    PlanTypePPO anx_or_depTRUE
##        0.31020        1.00190       13.40368        0.79665        1.55766
```