

Homework 2 – CTN-0094 Demographics Analysis

AUTHOR
Your Name

Table of contents

| 1 Methods

2 Results

1 Methods

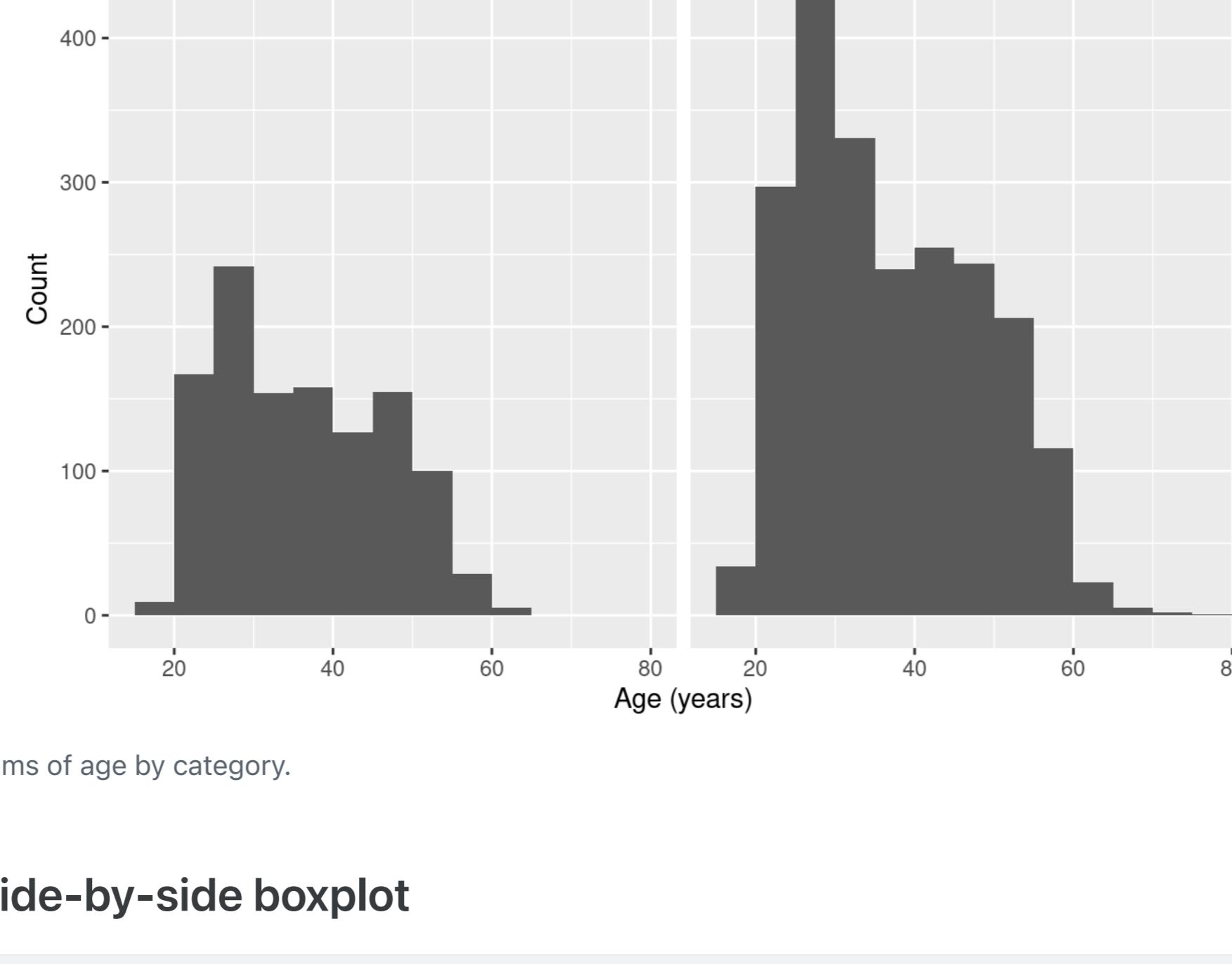
People with missing key analysis variables were excluded. Specifically:

Of the 3560 people with demographics data, 208 were removed because they were missing age (N = 208) and/or the categorical variable is_male (N = 4). The final analytic sample was N = 3352.

2 Results

2.1 Histogram of age (faceted by {CAT})

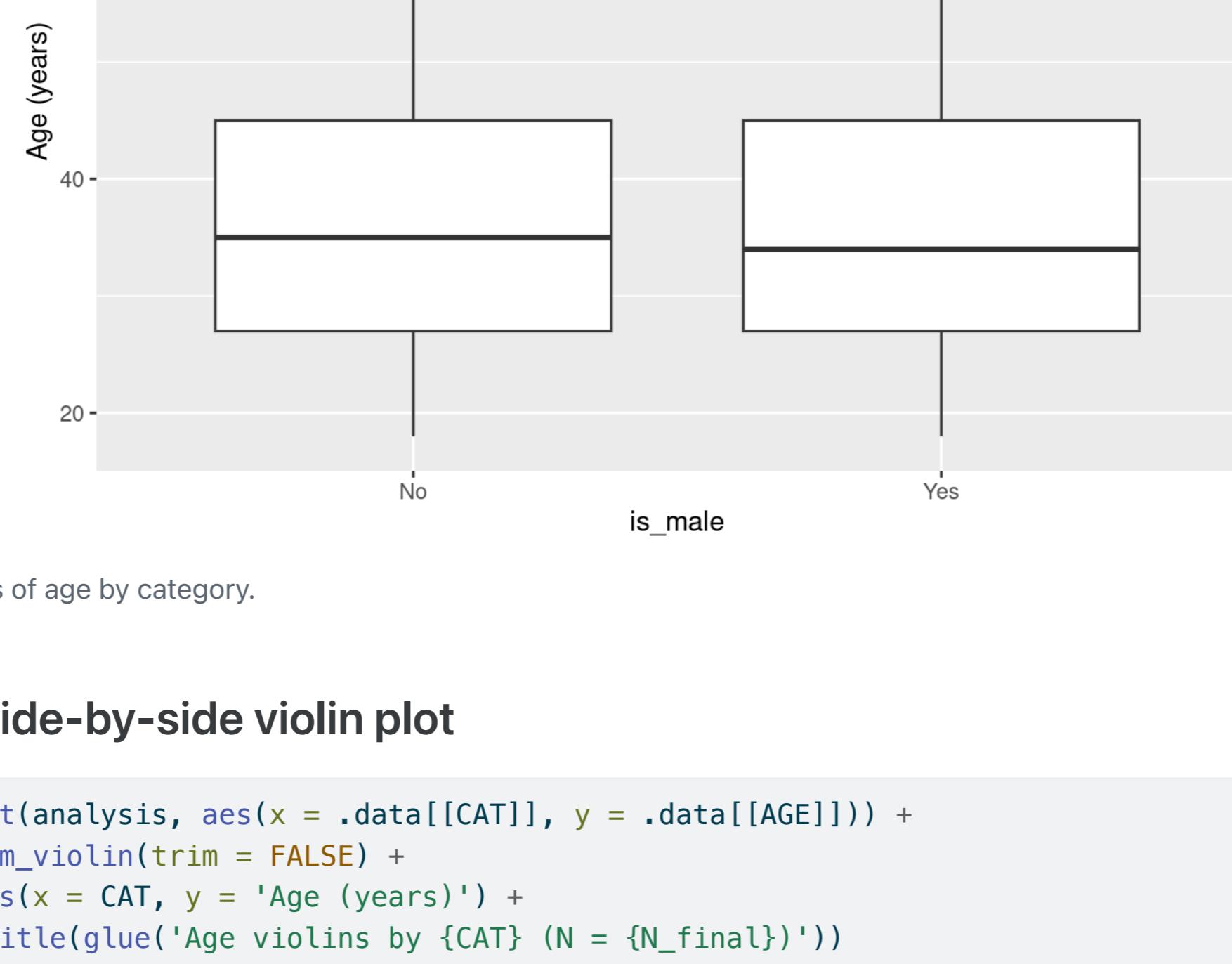
```
ggplot(analysis, aes(x = .data[[AGE]])) +  
  geom_histogram(binwidth = 5, boundary = 0, closed = 'left') +  
  facet_wrap(as.formula(paste('~', CAT))) +  
  labs(x = 'Age (years)', y = 'Count') +  
  ggtitle(glue('Age histograms by {CAT} (N = {N_final})'))
```



Histograms of age by category.

2.2 Side-by-side boxplot

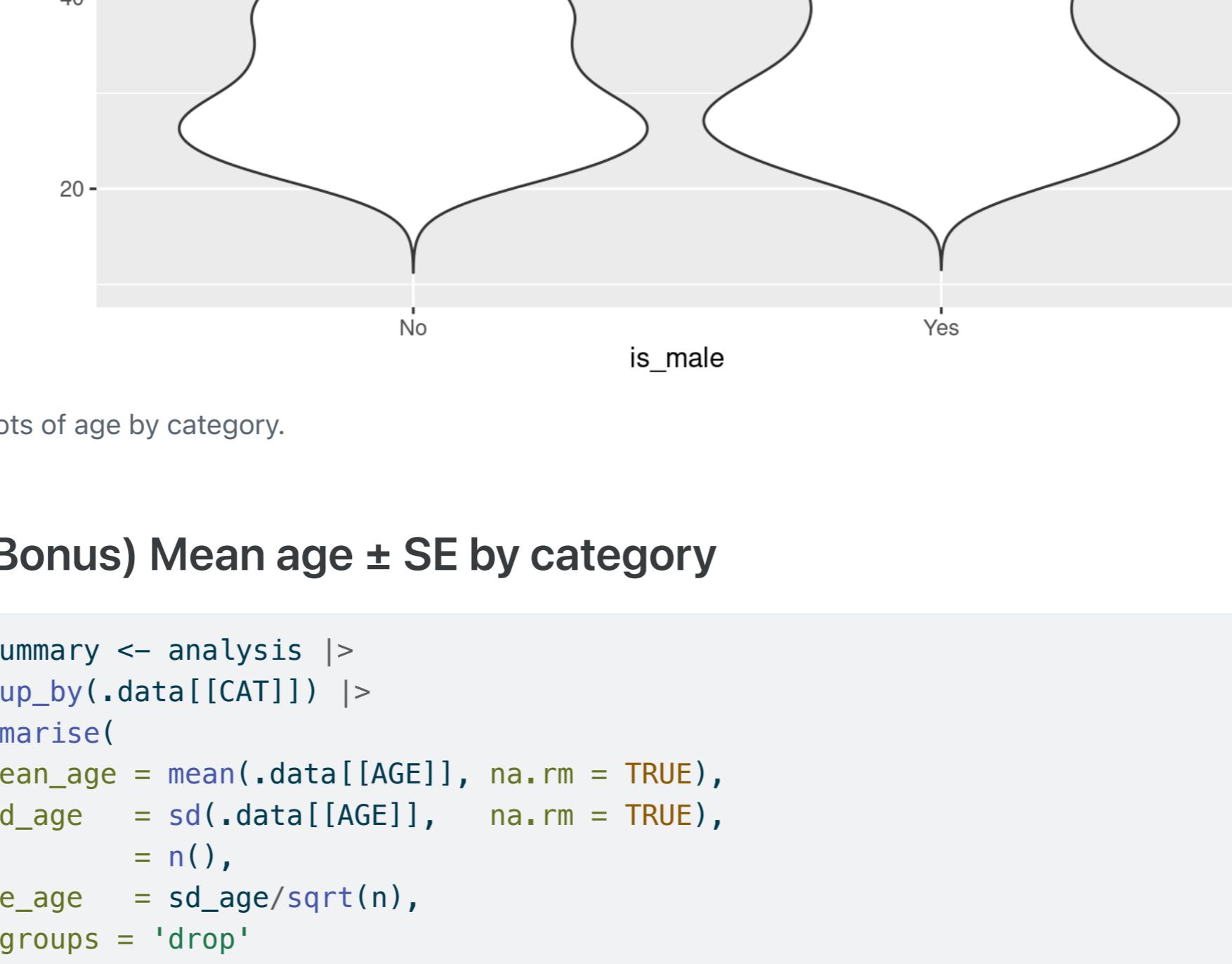
```
ggplot(analysis, aes(x = .data[[CAT]], y = .data[[AGE]])) +  
  geom_boxplot() +  
  labs(x = CAT, y = 'Age (years)') +  
  ggtitle(glue('Age boxplots by {CAT} (N = {N_final})'))
```



Boxplots of age by category.

2.3 Side-by-side violin plot

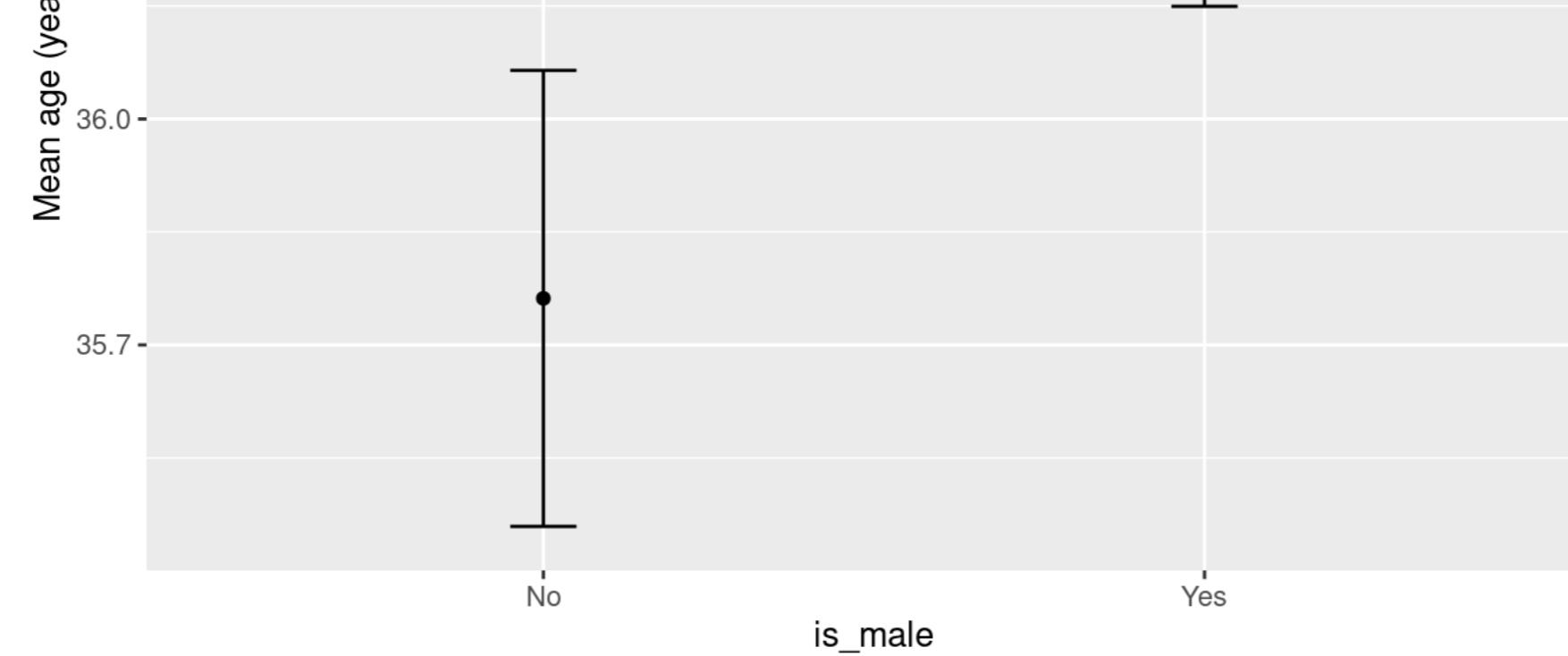
```
ggplot(analysis, aes(x = .data[[CAT]], y = .data[[AGE]])) +  
  geom_violin(trim = FALSE) +  
  labs(x = CAT, y = 'Age (years)') +  
  ggtitle(glue('Age violins by {CAT} (N = {N_final})'))
```



Violin plots of age by category.

2.4 (Bonus) Mean age ± SE by category

```
age_summary <- analysis |>  
  group_by(.data[[CAT]]) |>  
  summarise(  
    mean_age = mean(.data[[AGE]], na.rm = TRUE),  
    sd_age = sd(.data[[AGE]], na.rm = TRUE),  
    n = n(),  
    se_age = sd_age/sqrt(n),  
    .groups = 'drop'  
  )  
  ggplot(age_summary, aes(x = .data[[CAT]], y = mean_age)) +  
  geom_point() +  
  geom_errorbar(aes(ymin = mean_age - se_age, ymax = mean_age + se_age), width = 0.1) +  
  labs(x = CAT, y = 'Mean age (years)') +  
  ggtitle(glue('Mean age ± SE by {CAT} (N = {N_final})'))
```



Mean age and SE by category.