Final Report

due November 16, 2021 by 11:59 PM

G.I. Joe: Isa Rundell and Grace Vo

```
install.packages("taRifx")
install.packages("fastDummies")
library(tidyverse)
library(dplyr)
library(taRifx)
library(fastDummies)
library(infer)
library(parsnip)
```

```
drug <- readr::read_csv("Drug_Consumption.csv")
```

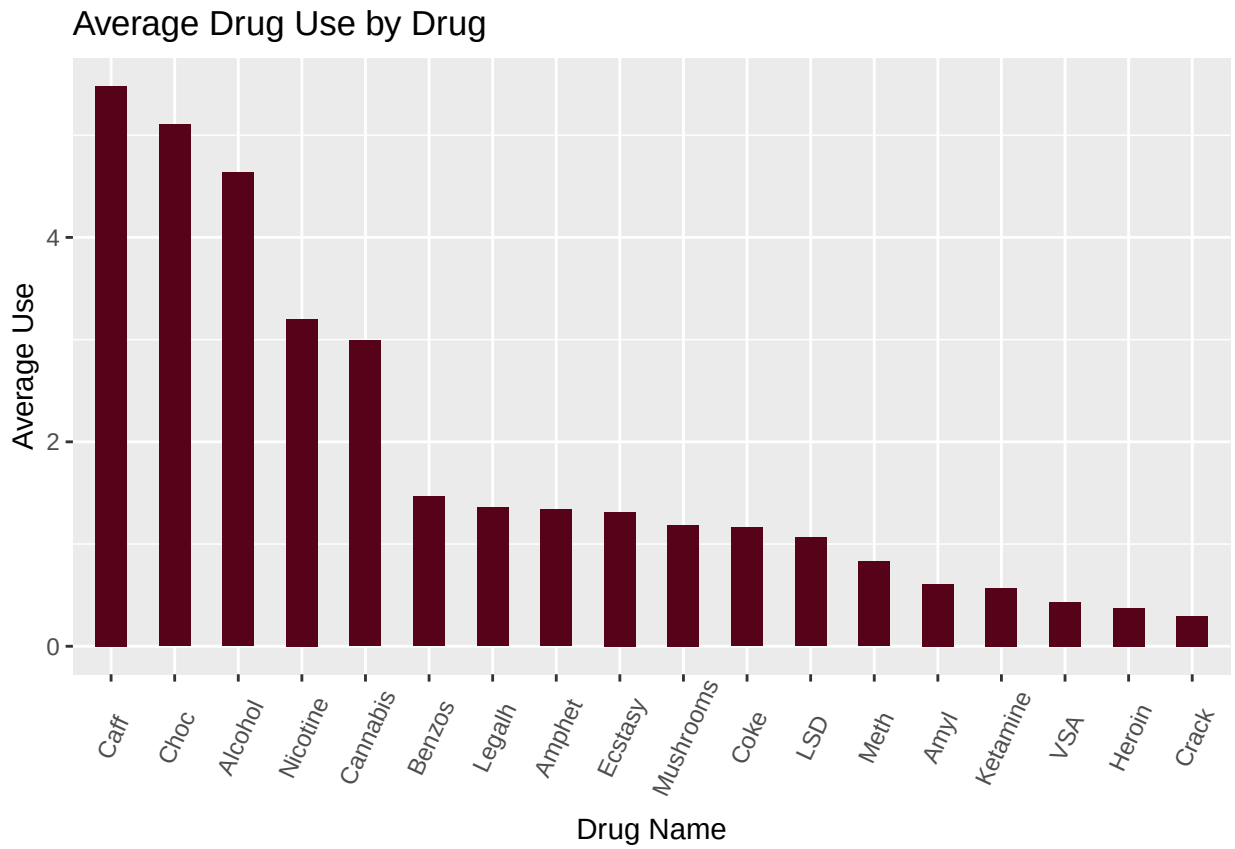
Abstract:

Background and Significance:

Methods: a) Data Collection and Variables

b) Exploratory Data Analysis

```
drug1 <- drug %>%
  mutate(across(Alcohol:VSA,destring))

drug1[,14:32] <- sapply(drug1[,14:32],as.numeric)
drugmeans <- colMeans(drug1[ , 14:32])

drug_name <- c('Alcohol', 'Amphet', 'Amyl', 'Benzos', 'Caff', 'Cannabis', 'Choc', 'Coke', 'Crack', 'Ecs
average_use <- c(4.63481953, 1.34023355, 0.60721868, 1.46496815, 5.48354565, 2.99097665, 5.10668790, 1.

drug_averages <- data.frame(drug_name, average_use)

drug_averages$drug_name <- factor(drug_averages$drug_name,
                  levels = drug_averages$drug_name[order(drug_averages$average_use, decreasing = TRUE)]
ggplot(drug_averages, aes(x=drug_name, y=average_use)) +
  geom_bar(stat="identity", width=.5, fill="#560219") +
  labs(title="Average Drug Use by Drug",
       x = "Drug Name",
       y = "Average Use") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```




```
numdrug <- drug1 %>%
  mutate(Age = replace(Age, Age == "18-24", 0), Age = replace(Age, Age == "25-34", 1), Age = replace(Ag

numdrug <- mutate_all(numdrug, function(x) as.numeric(as.character(x)))

numdrug2 = select(numdrug, -Alcohol, -Amphet, -Amyl, -Benzos, -Caff, -Cannabis, -Choc, -Coke, -Crack, -

head(numdrug2)
```

```
## # A tibble: 6 x 14
##      ID   Age Gender Education Country Ethnicity Nscore Escore  Oscore AScore
##   <dbl> <dbl>  <dbl>     <dbl>   <dbl>     <dbl>  <dbl>  <dbl>   <dbl>  <dbl>
## 1     2     1      1         8       5         6 -0.678  1.94   1.44    0.761
## 2     3     2      1         5       5         6 -0.467  0.805 -0.847  -1.62
## 3     4     0      0         7       5         6 -0.149 -0.806 -0.0193  0.590
## 4     5     2      0         8       5         6  0.735 -1.63  -0.452  -0.302
## 5     6     5      0         3       1         6 -0.678 -0.300 -1.56    2.04
## 6     7     3      1         7       6         6 -0.467 -1.09  -0.452  -0.302
## # ... with 4 more variables: Cscore <dbl>, Impulsive <dbl>, SS <dbl>,
## #   Mushrooms <dbl>
```

```
correlation_matrix <- round(cor(numdrug2),2)
head(correlation_matrix)
```

```
##              ID   Age Gender Education Country Ethnicity Nscore Escore Oscore
## ID         1.00 -0.27   0.02     -0.01    0.10      0.01   0.02  -0.05   0.17
## Age       -0.27  1.00  -0.10      0.10   -0.06      0.04  -0.14  -0.03  -0.22
```

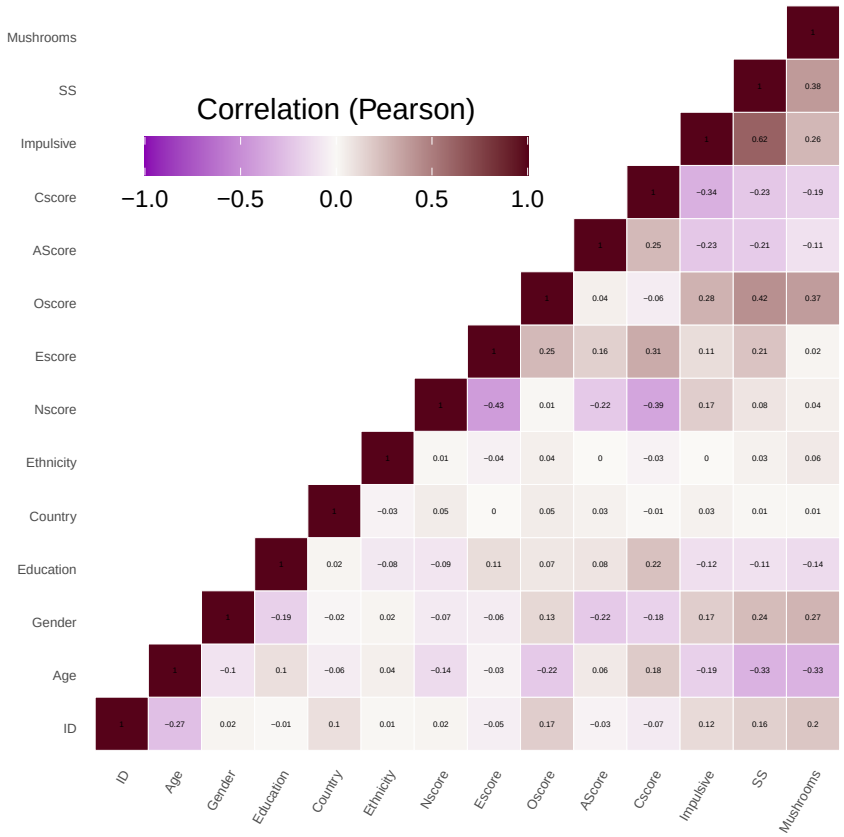
```
## Gender     0.02 -0.10   1.00     -0.19   -0.02      0.02  -0.07  -0.06   0.13
## Education -0.01  0.10  -0.19      1.00    0.02     -0.08  -0.09   0.11   0.07
## Country    0.10 -0.06  -0.02      0.02    1.00     -0.03   0.05   0.00   0.05
## Ethnicity  0.01  0.04   0.02     -0.08   -0.03      1.00   0.01  -0.04   0.04
##           AScore Cscore Impulsive    SS Mushrooms
## ID         -0.03  -0.07      0.12  0.16      0.20
## Age         0.06   0.18     -0.19 -0.33     -0.33
## Gender     -0.22  -0.18      0.17  0.24      0.27
## Education   0.08   0.22     -0.12 -0.11     -0.14
## Country     0.03  -0.01      0.03  0.01      0.01
## Ethnicity   0.00  -0.03      0.00  0.03      0.06
```

```
  get_upper_tri<-function(correlation_matrix){
    correlation_matrix[lower.tri(correlation_matrix)] <- NA
    return(correlation_matrix)
  }
  upper_tri <- get_upper_tri(correlation_matrix)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
melted_cormat <- melt(upper_tri, na.rm = TRUE)
library(ggplot2)
ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
 geom_tile(color = "white")+
 scale_fill_gradient2(low = "#8a02b2", high = "#560219", mid = "#FAF9F6",
   midpoint = 0, limit = c(-1,1), space = "Lab",
   name="Correlation (Pearson)") +
  theme_minimal()+
 theme(axis.text.x = element_text(angle = 60, vjust = 1,
    size = 5, hjust = 1), axis.text.y = element_text(vjust = 1, size = 5, hjust = 1))+
 coord_fixed() +
geom_text(aes(Var2, Var1, label = value), color = "black", size = 1) +
theme(
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  panel.grid.major = element_blank(),
  panel.border = element_blank(),
  panel.background = element_blank(),
  axis.ticks = element_blank(),
  legend.justification = c(1, 0),
  legend.position = c(0.6, 0.7),
  legend.direction = "horizontal",
  legend.key.size = unit(0.5, 'cm'))+
  guides(fill = guide_colorbar(barwidth = 10, barheight = 1,
                title.position = "top", title.hjust = 0.5))
```




```
drug_clean <- numdrug %>%
  mutate(Alcohol_User = as.factor(ifelse(Alcohol > 1, "yes", "no")),
         Amphetamine_User = as.factor(ifelse(Amphet > 1, "yes", "no")),
         AmylNitrite_User = as.factor(ifelse(Amyl > 1, "yes", "no")),
         Benzos_User = as.factor(ifelse(Benzos > 1, "yes", "no")),
         Caffeine_User = as.factor(ifelse(Caff > 1, "yes", "no")),
         Cannabis_User = as.factor(ifelse(Cannabis > 1, "yes", "no")),
         Chocolate_User = as.factor(ifelse(Choc > 1, "yes", "no")),
         Cocaine_User = as.factor(ifelse(Coke > 1, "yes", "no")),
         Crack_User = as.factor(ifelse(Crack > 1, "yes", "no")),
         Ecstasy_User = as.factor(ifelse(Ecstasy > 1, "yes", "no")),
         Heroine_User = as.factor(ifelse(Heroin > 1, "yes", "no")),
         Ketamine_User = as.factor(ifelse(Ketamine > 1, "yes", "no")),
         LegalHighs_User = as.factor(ifelse(Legalh > 1, "yes", "no")),
         LSD_User = as.factor(ifelse(LSD > 1, "yes", "no")),
         Meth_User = as.factor(ifelse(Meth > 1, "yes", "no")),
         Mushrooms_User = as.factor(ifelse(Mushrooms > 1, "yes", "no")),
         Nicotine_User = as.factor(ifelse(Nicotine > 1, "yes", "no")),
         Semeron_User = as.factor(ifelse(Semer > 1, "yes", "no")),
         VSA_User = as.factor(ifelse(VSA > 1, "yes", "no")))

drug_clean_2 <- drug1 %>%
    mutate(Alcohol_User = as.factor(ifelse(Alcohol > 1, "yes", "no")),
         Amphetamine_User = as.factor(ifelse(Amphet > 1, "yes", "no")),
         AmylNitrite_User = as.factor(ifelse(Amyl > 1, "yes", "no")),
         Benzos_User = as.factor(ifelse(Benzos > 1, "yes", "no")),
```

```
         Caffeine_User = as.factor(ifelse(Caff > 1, "yes", "no")),
         Cannabis_User = as.factor(ifelse(Cannabis > 1, "yes", "no")),
         Chocolate_User = as.factor(ifelse(Choc > 1, "yes", "no")),
         Cocaine_User = as.factor(ifelse(Coke > 1, "yes", "no")),
         Crack_User = as.factor(ifelse(Crack > 1, "yes", "no")),
         Ecstasy_User = as.factor(ifelse(Ecstasy > 1, "yes", "no")),
         Heroine_User = as.factor(ifelse(Heroin > 1, "yes", "no")),
         Ketamine_User = as.factor(ifelse(Ketamine > 1, "yes", "no")),
         LegalHighs_User = as.factor(ifelse(Legalh > 1, "yes", "no")),
         LSD_User = as.factor(ifelse(LSD > 1, "yes", "no")),
         Meth_User = as.factor(ifelse(Meth > 1, "yes", "no")),
         Mushrooms_User = as.factor(ifelse(Mushrooms > 1, "yes", "no")),
         Nicotine_User = as.factor(ifelse(Nicotine > 1, "yes", "no")),
         Semeron_User = as.factor(ifelse(Semer > 1, "yes", "no")),
         VSA_User = as.factor(ifelse(VSA > 1, "yes", "no")))
```

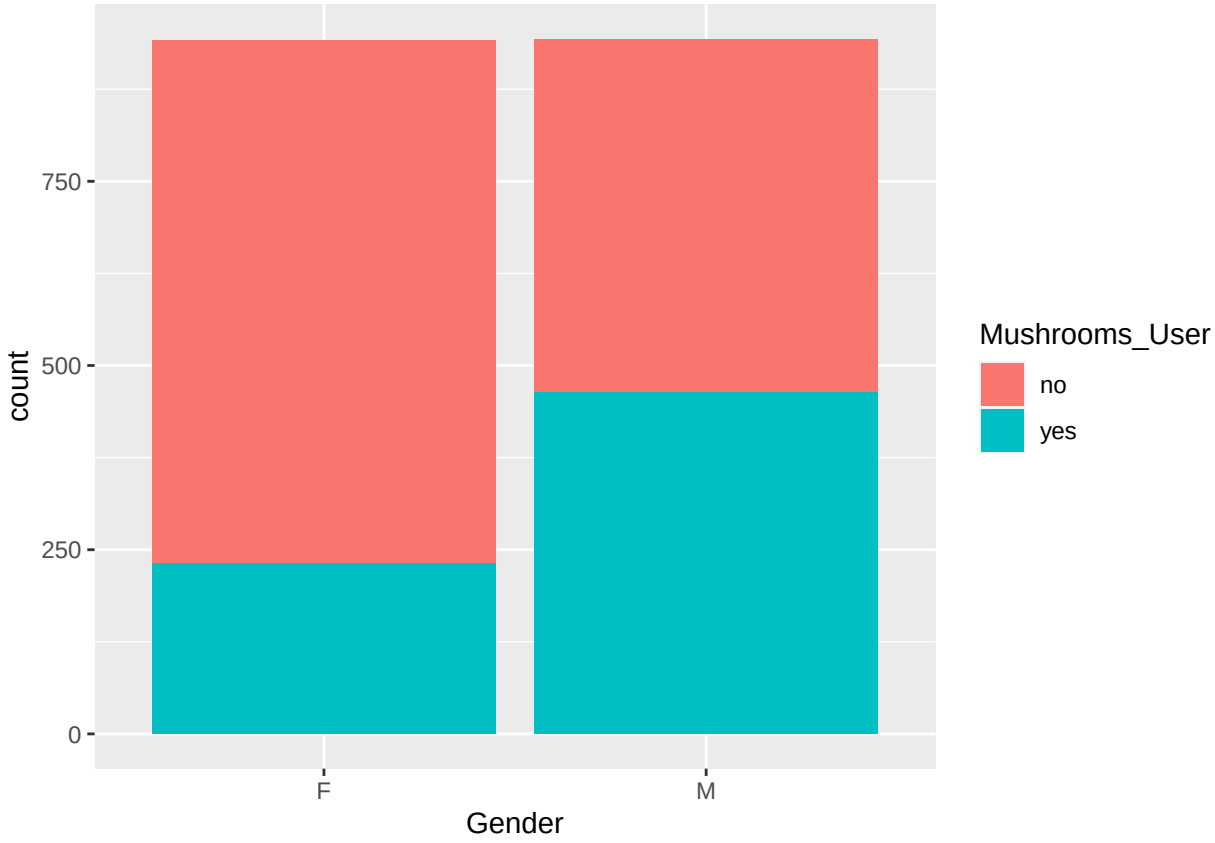
```
drug_byuse <- numdrug %>%
  mutate(Alcohol_User = ifelse(Alcohol > 1, 1, 0),
         Amphetamine_User = ifelse(Amphet > 1, 1, 0),
         AmylNitrite_User = ifelse(Amyl > 1, 1, 0),
         Benzos_User = ifelse(Benzos > 1, 1, 0),
         Caffeine_User = ifelse(Caff > 1, 1, 0),
         Cannabis_User = ifelse(Cannabis > 1, 1, 0),
         Chocolate_User = ifelse(Choc > 1, 1, 0),
         Cocaine_User = ifelse(Coke > 1, 1, 0),
         Crack_User = ifelse(Crack > 1, 1, 0),
         Ecstasy_User = ifelse(Ecstasy > 1, 1, 0),
         Heroine_User = ifelse(Heroin > 1, 1, 0),
         Ketamine_User = ifelse(Ketamine > 1, 1, 0),
         LegalHighs_User = ifelse(Legalh > 1, 1, 0),
         LSD_User = ifelse(LSD > 1, 1, 0),
         Meth_User = ifelse(Meth > 1, 1, 0),
         Mushrooms_User = ifelse(Mushrooms > 1, 1, 0),
         Nicotine_User = ifelse(Nicotine > 1, 1, 0),
         Semeron_User = ifelse(Semer > 1, 1, 0),
         VSA_User = ifelse(VSA > 1, 1, 0)) %>%
  dplyr::select(Alcohol_User, Amphetamine_User, AmylNitrite_User, Benzos_User,
                Caffeine_User, Cannabis_User, Chocolate_User, Cocaine_User,
                Crack_User, Ecstasy_User, Heroine_User, Ketamine_User,
                LegalHighs_User, LSD_User, Meth_User, Mushrooms_User,
                Nicotine_User, Semeron_User, VSA_User)


data.frame(yes = colSums(drug_byuse), no = 1885 - colSums(drug_byuse),
           drug = colnames(drug_byuse)) %>%
  summarise(diff = abs(yes - no), drug = drug) %>%
  arrange(diff) %>%
  head()
```

```
##   diff             drug
## 1  349      Benzos_User
## 2  361  LegalHighs_User
## 3  383     Ecstasy_User
## 4  497   Mushrooms_User
```

```
## 5  511     Cocaine_User
## 6  529 Amphetamine_User
```

c) Analytical Methods

```
drug_clean_2 %>%
  ggplot(aes(x = Gender,
             fill = Mushrooms_User)) +
  geom_bar()
```



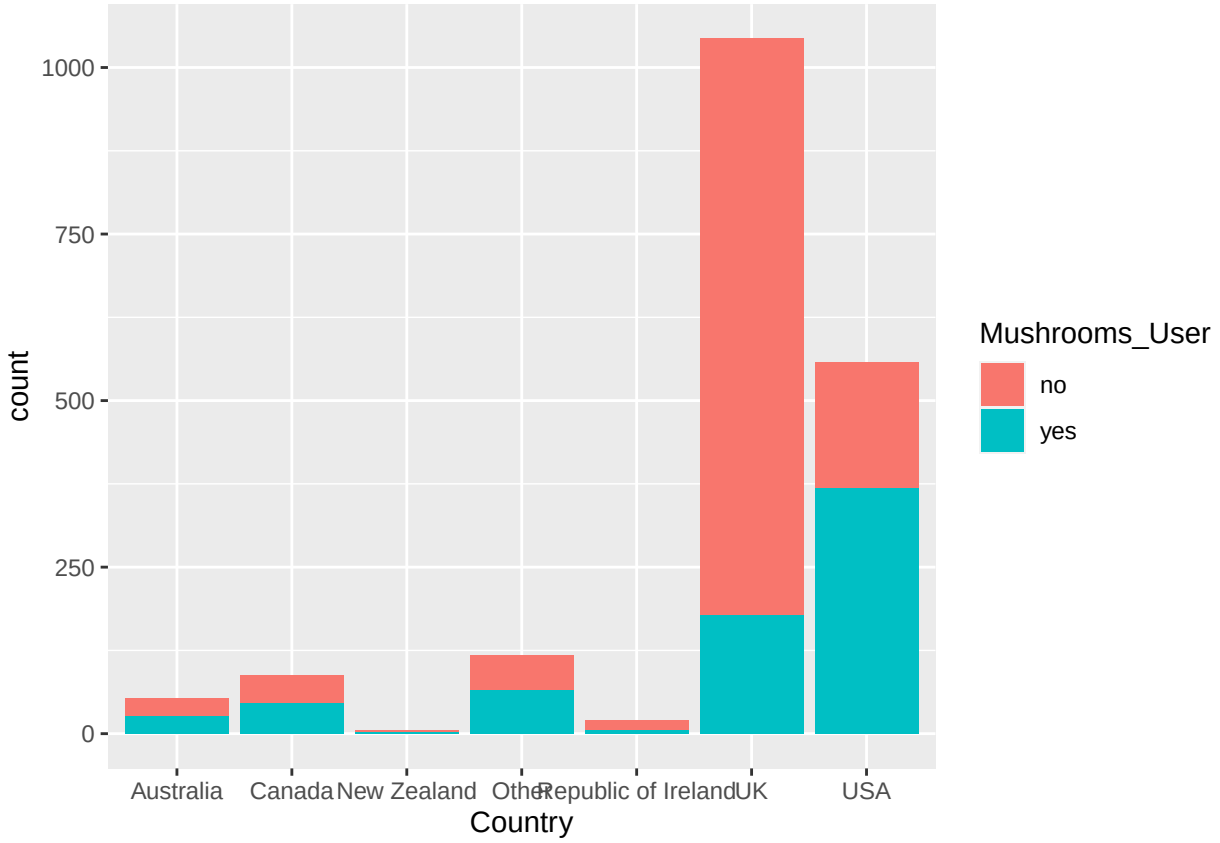

```
fisher.test(drug_clean_2$Gender, drug_clean_2$Mushrooms_User)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  drug_clean_2$Gender and drug_clean_2$Mushrooms_User
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  2.426426 3.624594
## sample estimates:
## odds ratio
##   2.962888
```

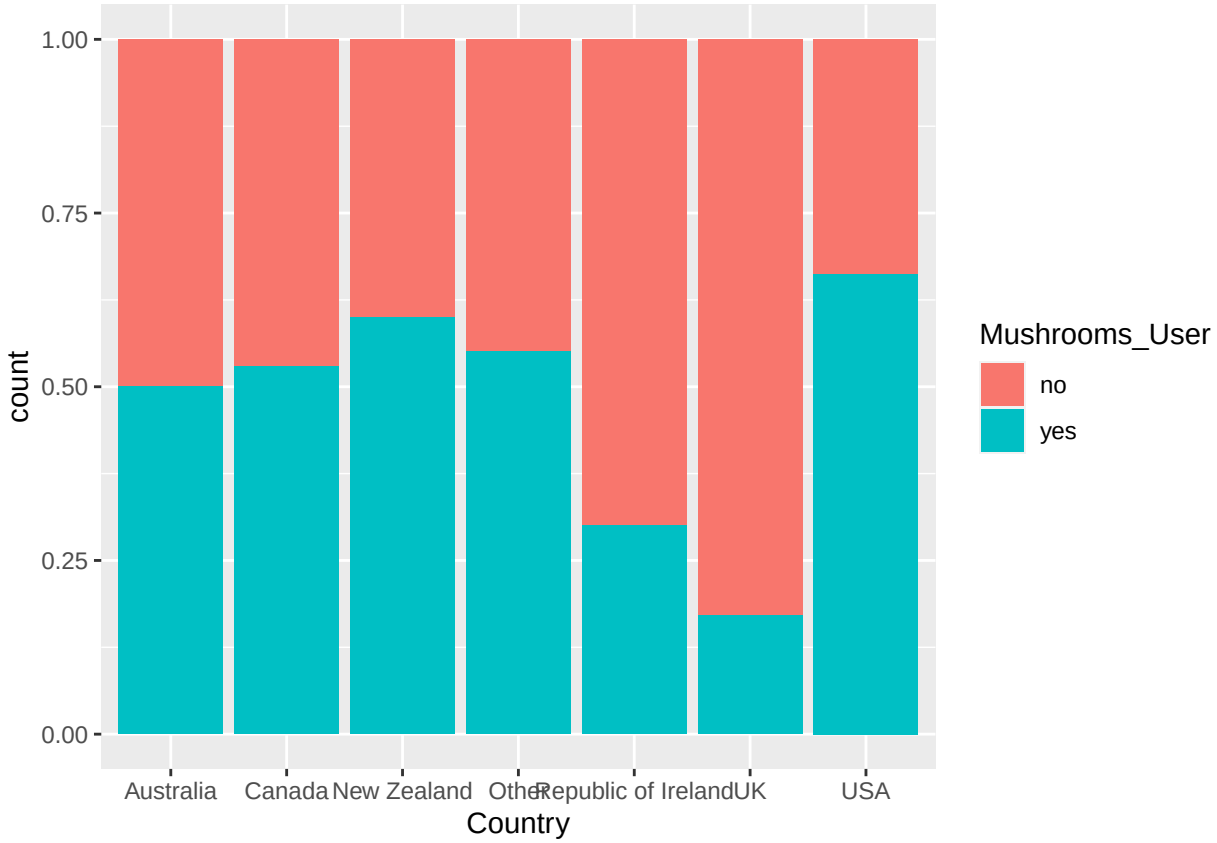
-Since the p-value is less than the significance level, gender is statistically significant.

```
drug_clean_2 %>%
  ggplot(aes(x = Country,
             fill = Mushrooms_User)) +
```

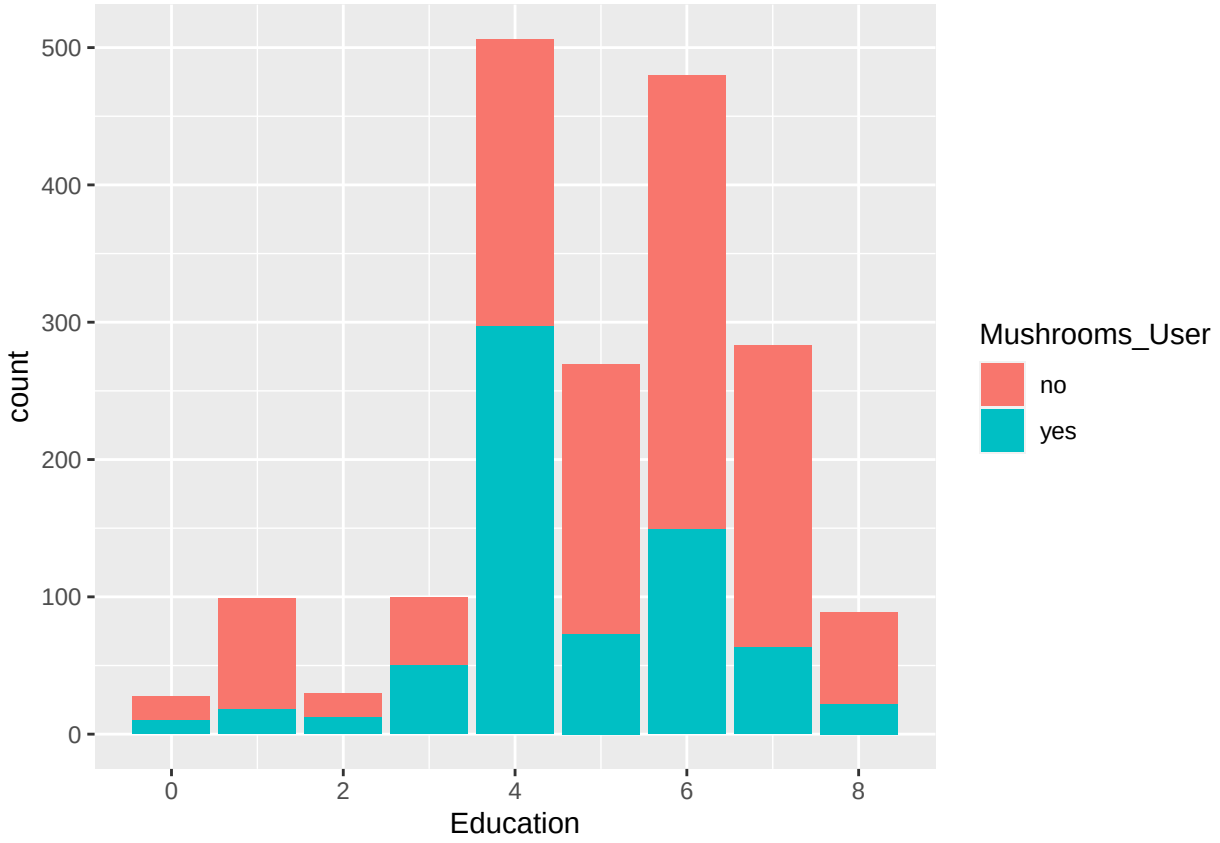
```
  geom_bar()
```



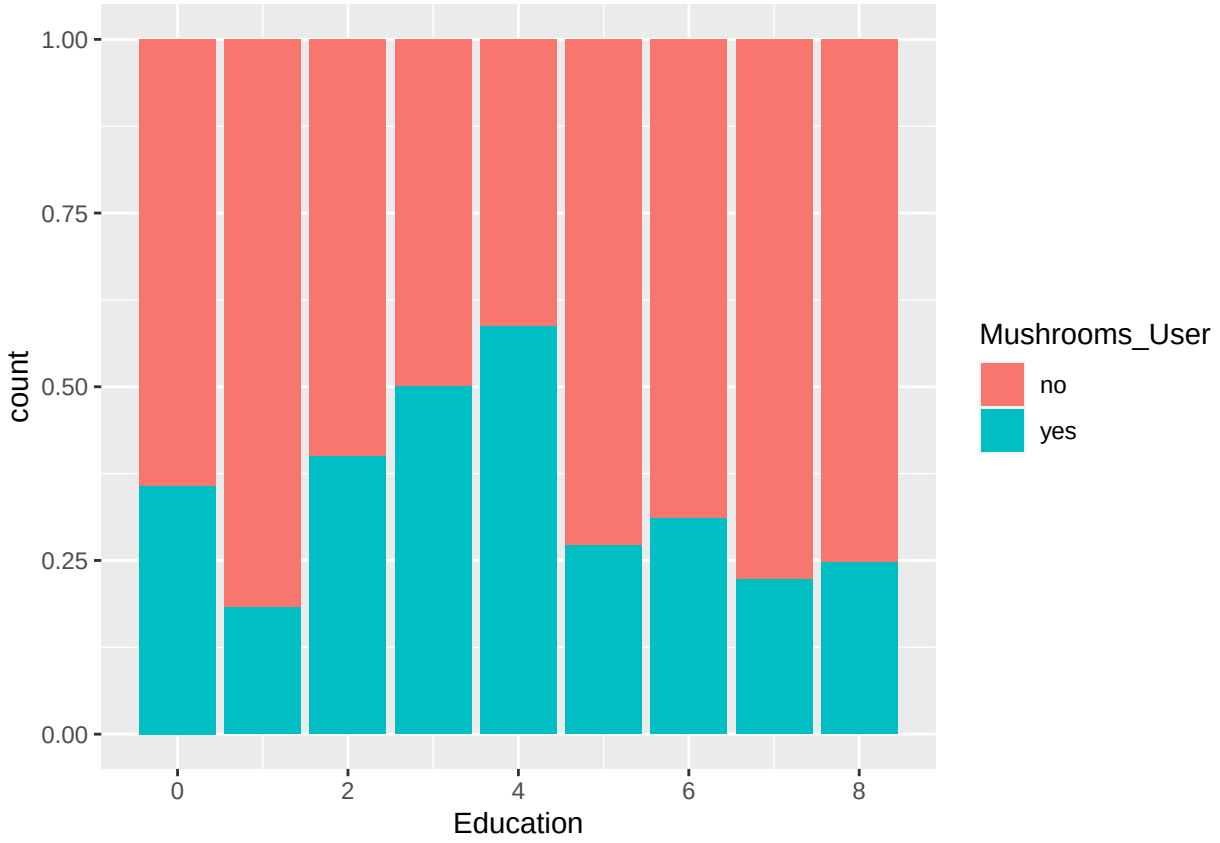

```
drug_clean_2 %>%
  ggplot(aes(x = Country,
             fill = Mushrooms_User)) +
  geom_bar(position = "fill")
```



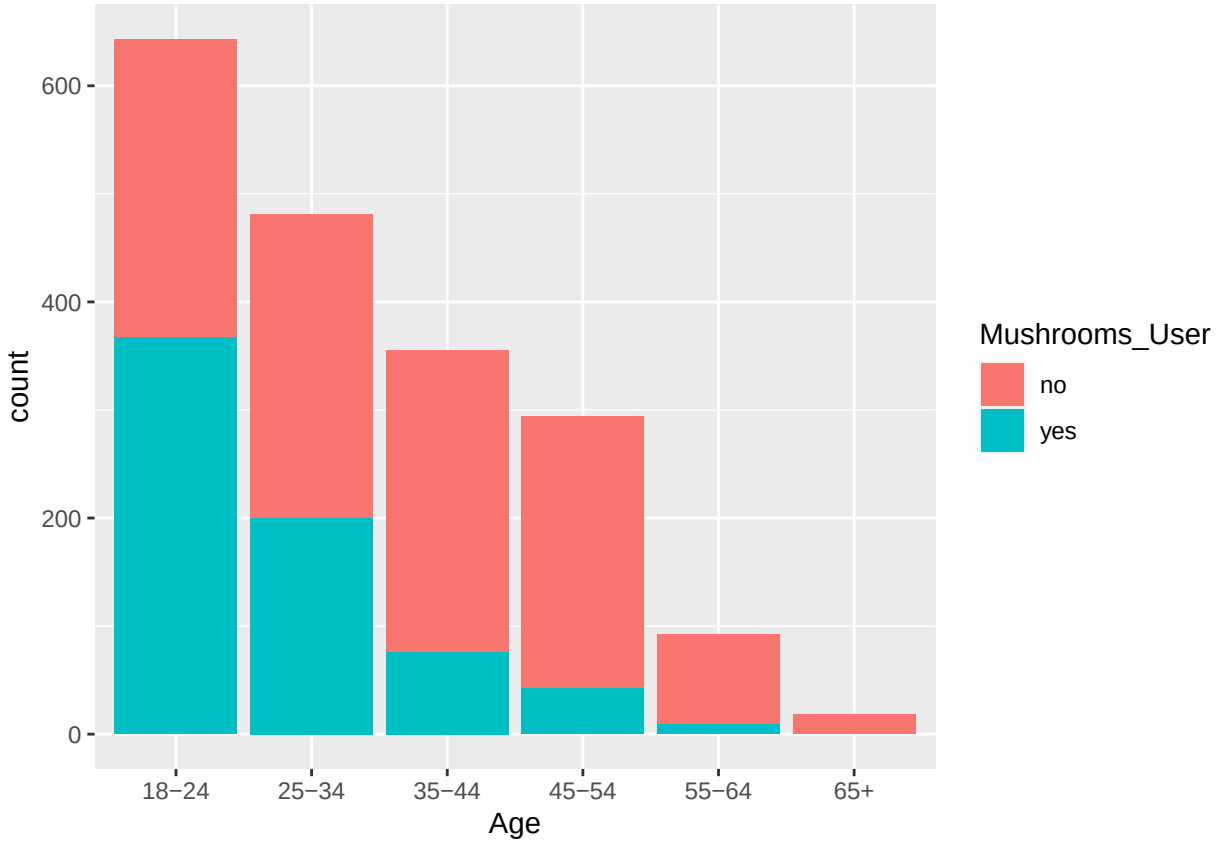

```
drug_clean %>%
  ggplot(aes(x = Education,
             fill = Mushrooms_User)) +
  geom_bar()
```



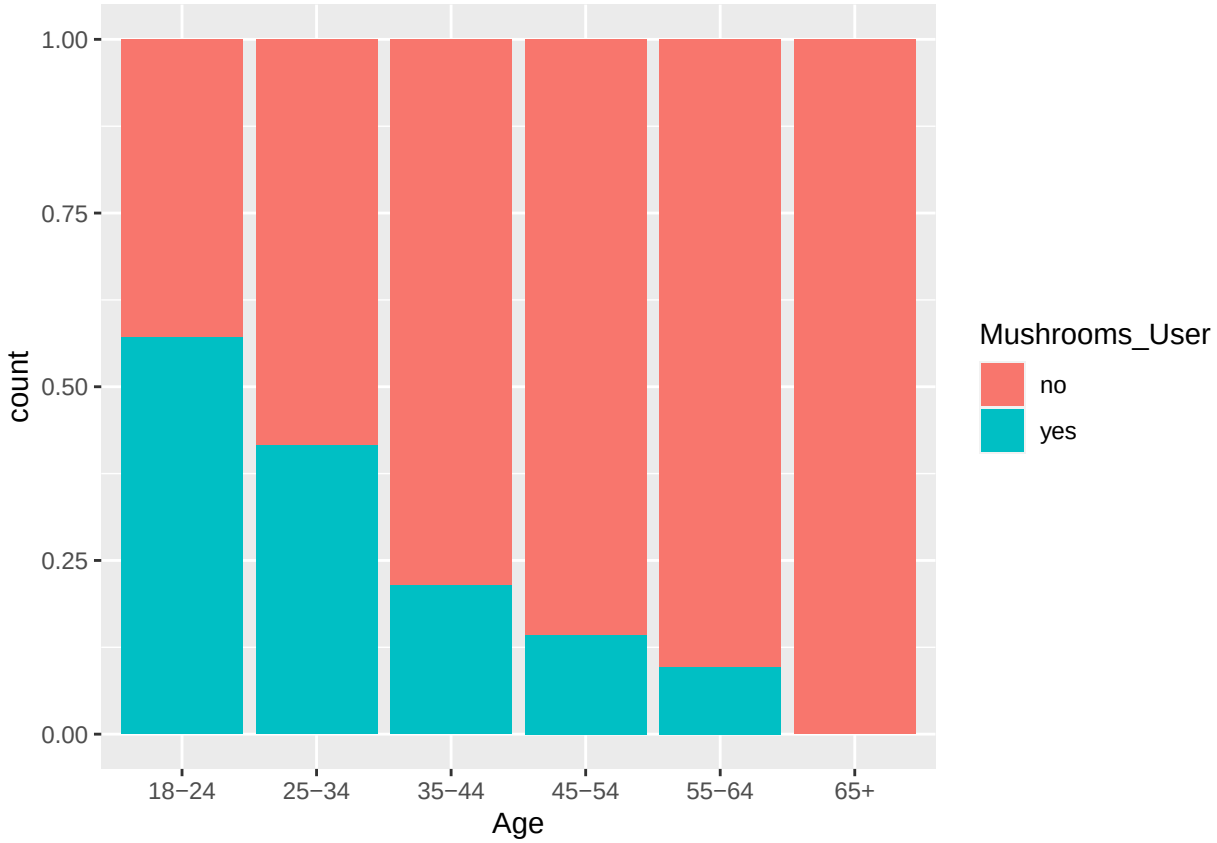

```
drug_clean %>%
  ggplot(aes(x = Education,
             fill = Mushrooms_User)) +
  geom_bar(position = "fill")
```



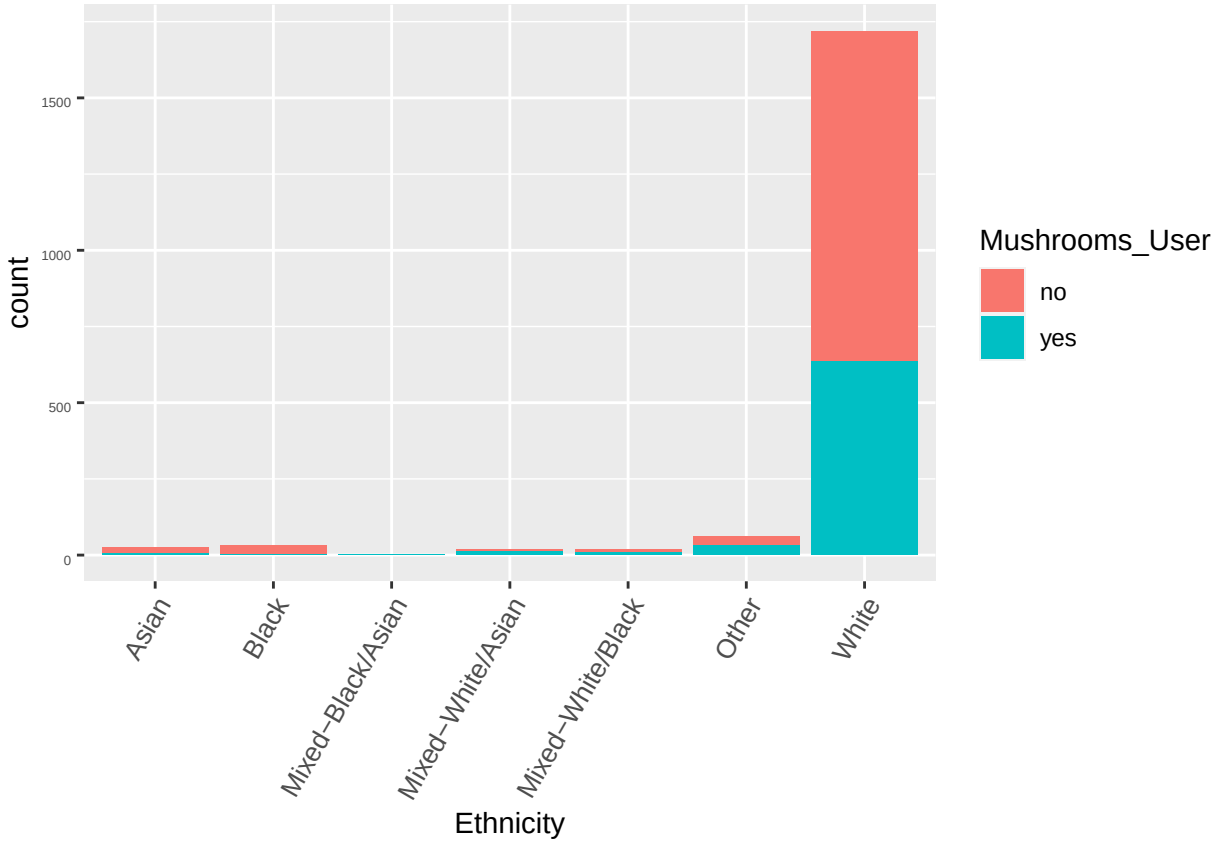

```
drug_clean_2 %>%
  ggplot(aes(x = Age,
             fill = Mushrooms_User)) +
  geom_bar()
```



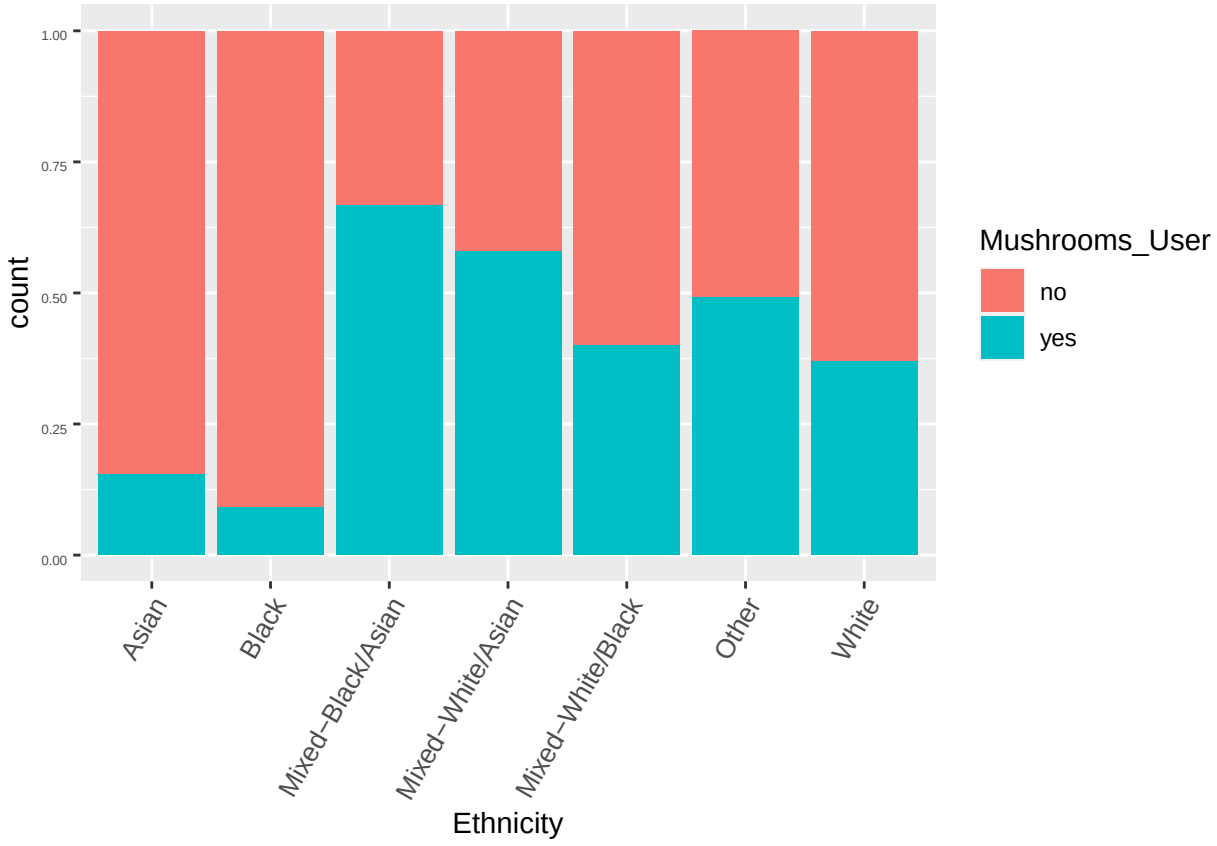

```
drug_clean_2 %>%
  ggplot(aes(x = Age,
             fill = Mushrooms_User)) +
  geom_bar(position = "fill")
```



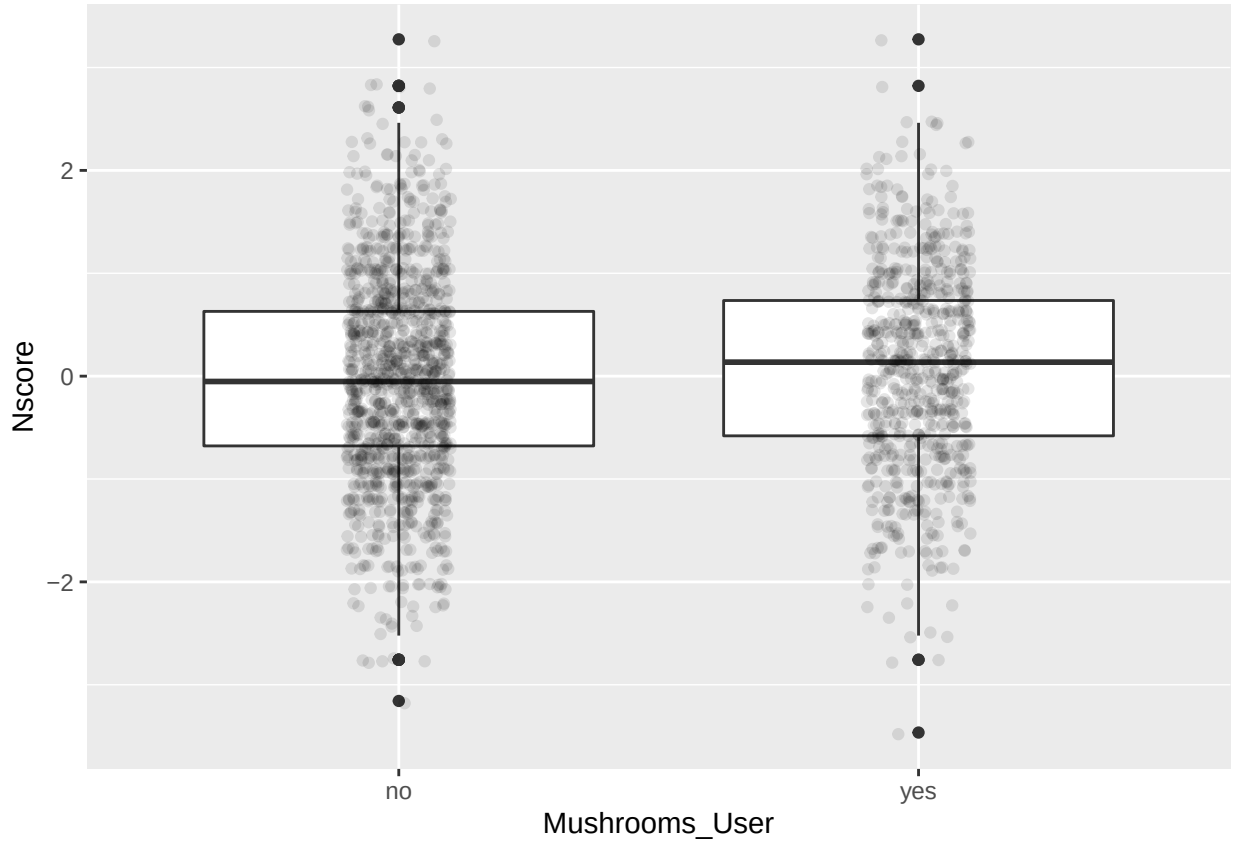

```
drug_clean_2 %>%
  ggplot(aes(x = Ethnicity,
             fill = Mushrooms_User)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 60, vjust = 1,
    size = 10, hjust = 1), axis.text.y = element_text(vjust = 1, size = 5, hjust = 1))
```



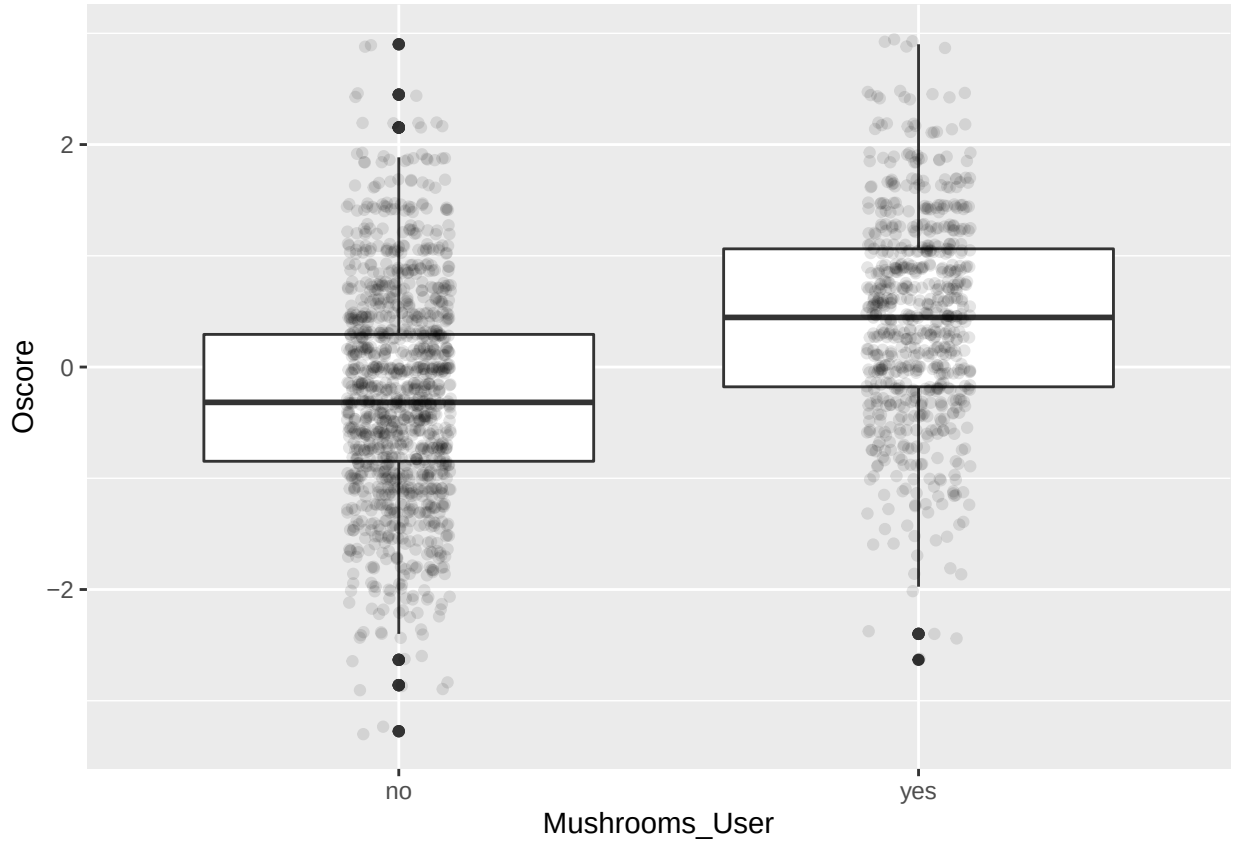

```
drug_clean_2 %>%
  ggplot(aes(x = Ethnicity,
             fill = Mushrooms_User)) +
  geom_bar(position = "fill") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1,
    size = 10, hjust = 1), axis.text.y = element_text(vjust = 1, size = 5, hjust = 1))
```



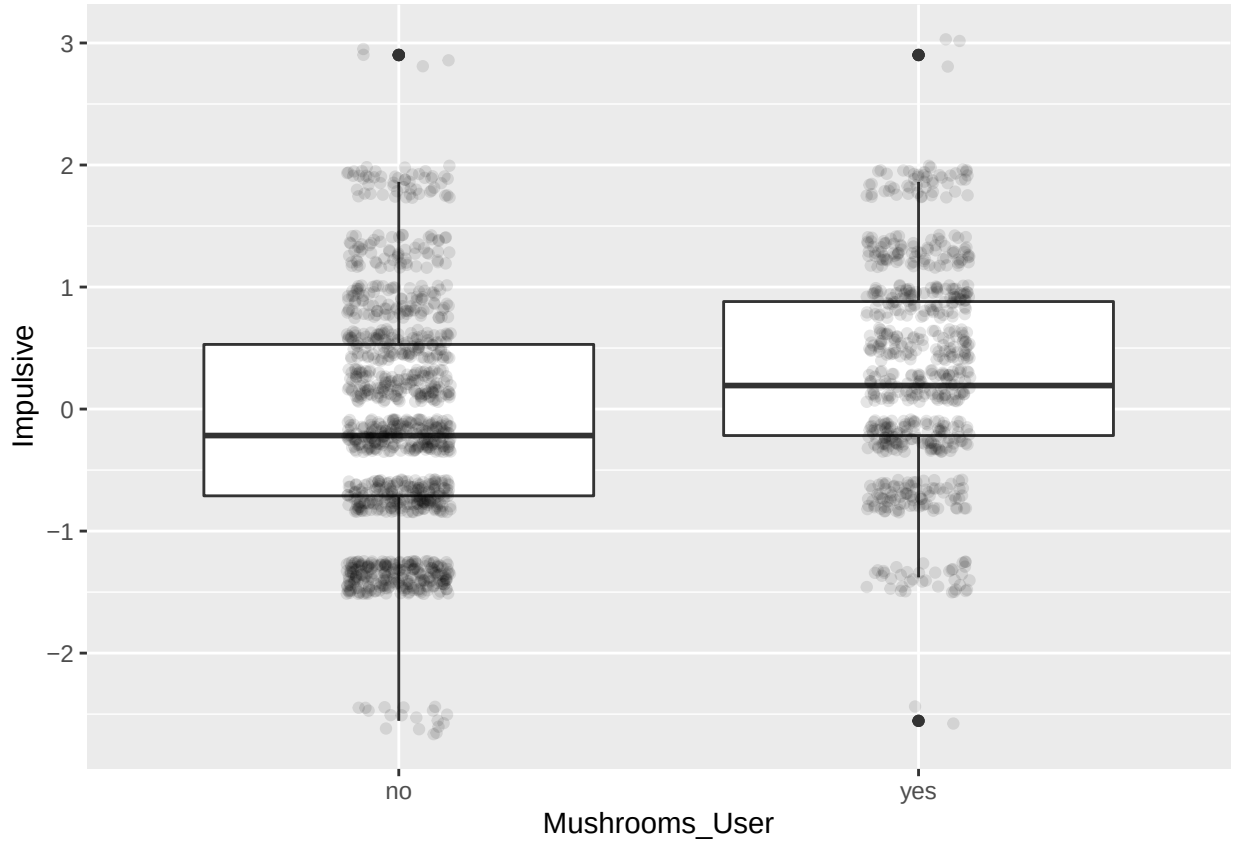

```
drug_clean %>%
  ggplot(aes(Mushrooms_User, Nscore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```



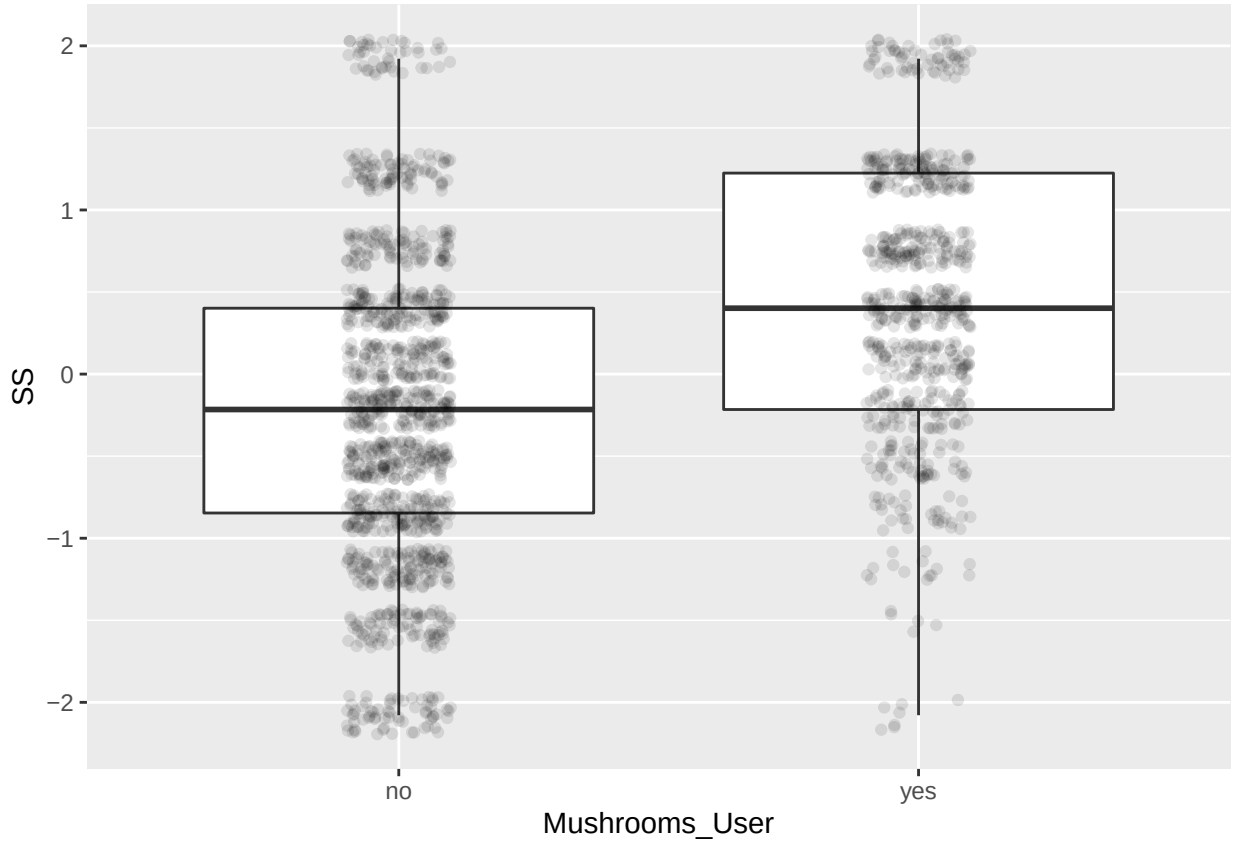

```
drug_clean %>%
  ggplot(aes(Mushrooms_User, Oscore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```



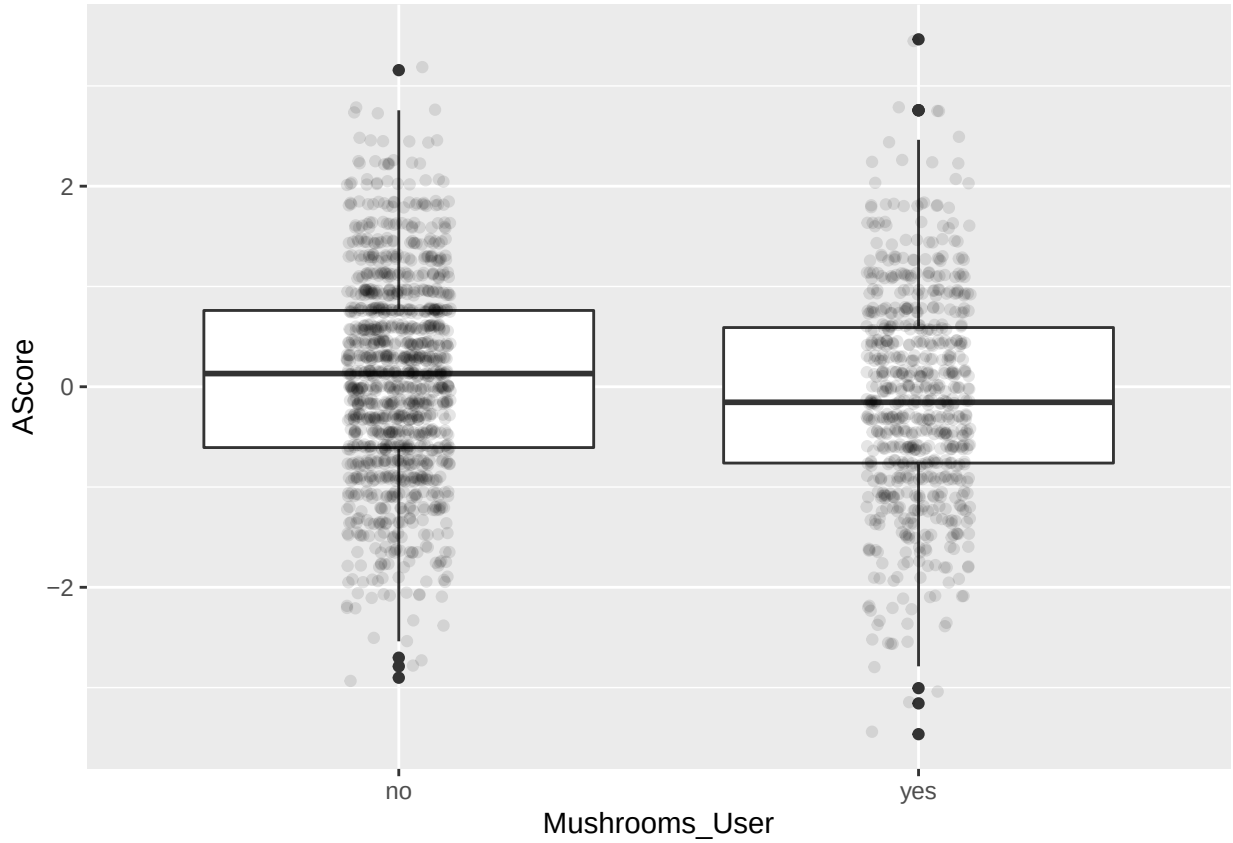

```
drug_clean %>%
  ggplot(aes(Mushrooms_User, Impulsive)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```



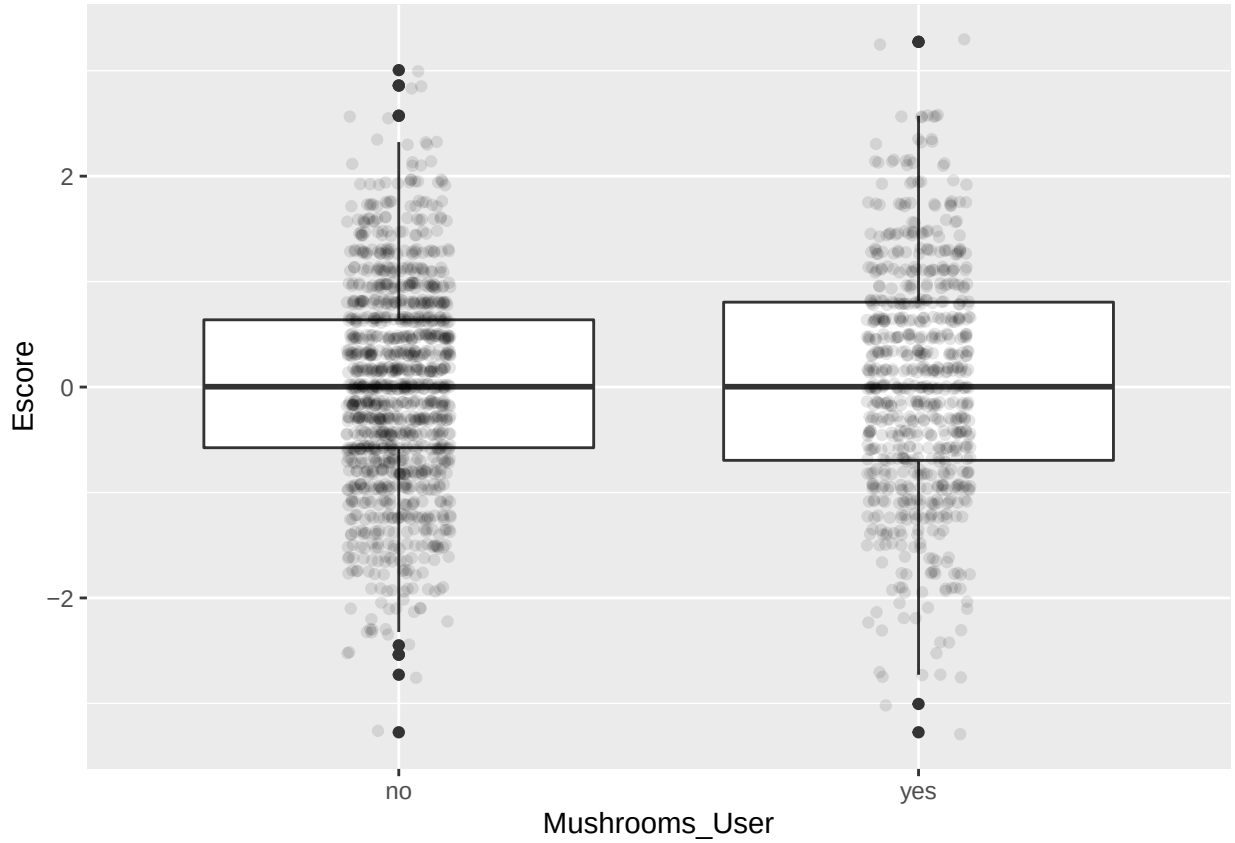

```
drug_clean %>%
  ggplot(aes(Mushrooms_User, SS)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```



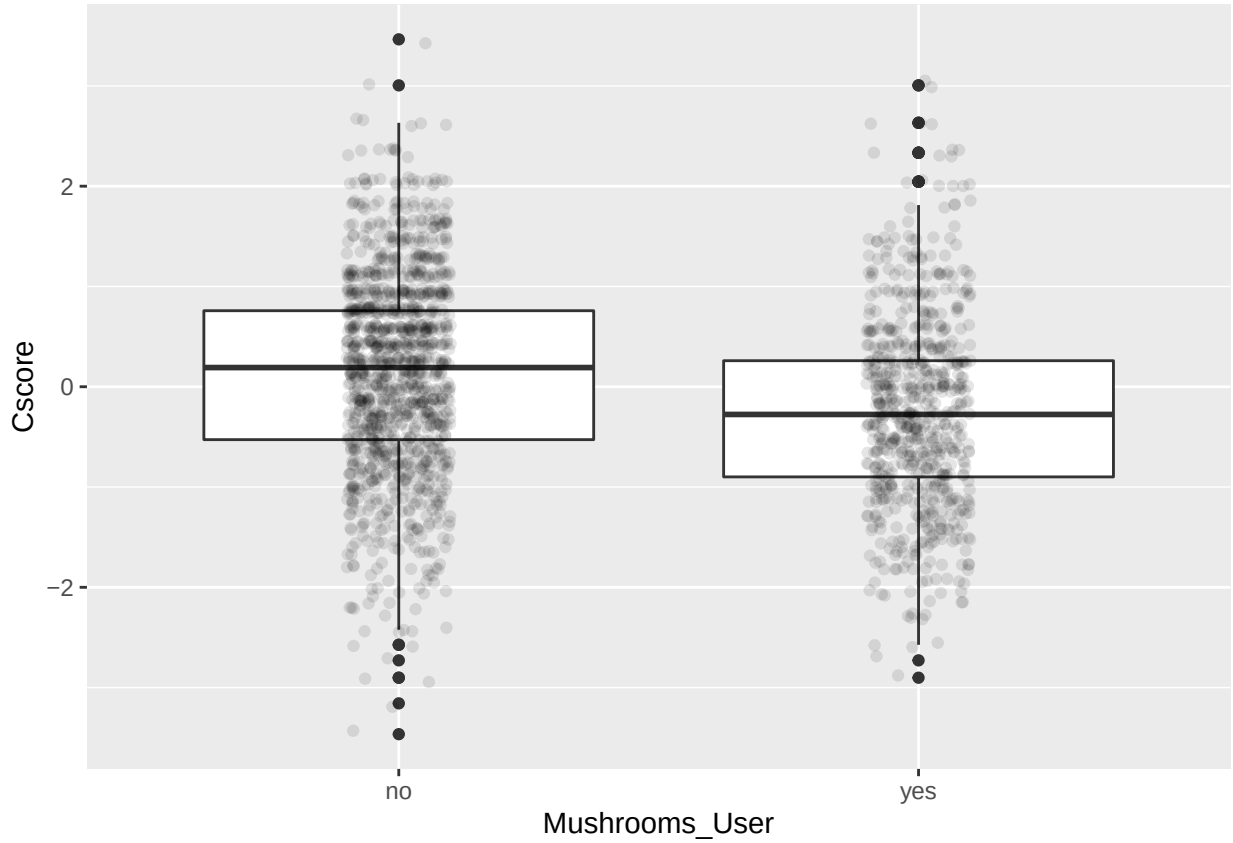

```
drug_clean %>%
  ggplot(aes(Mushrooms_User, AScore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```




```
drug_clean %>%
  ggplot(aes(Mushrooms_User, Escore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```




```
drug_clean %>%
  ggplot(aes(Mushrooms_User, Cscore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```




```
fit_multi <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(Benzos_User ~ Gender + Age + Education + Nscore + Oscore + Impulsive + SS + Cscore + AScore + Esc
result<-tidy(fit_multi, conf.int=TRUE, exponentiate=TRUE)
print(result, n=20)
```

```
## # A tibble: 22 x 7
##    term                 estimate std.error statistic  p.value conf.low conf.high
##    <chr>                   <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
##  1 (Intercept)             0.498    0.271     -2.58  1.00e- 2    0.290     0.842
##  2 GenderM                 1.35     0.112      2.66  7.84e- 3    1.08      1.68
##  3 Age25-34                1.46     0.145      2.62  8.83e- 3    1.10      1.94
##  4 Age35-44                1.13     0.162      0.758 4.48e- 1    0.822     1.56
##  5 Age45-54                1.18     0.175      0.924 3.56e- 1    0.834     1.66
##  6 Age55-64                0.674    0.278     -1.42  1.56e- 1    0.384     1.15
##  7 Age65+                  0.702    0.633     -0.560 5.76e- 1    0.177     2.23
##  8 EducationLeft schoo~    1.19     0.332      0.528 5.98e- 1    0.622     2.29
##  9 EducationLeft schoo~    1.18     0.467      0.348 7.28e- 1    0.470     2.95
## 10 EducationLeft schoo~    1.25     0.334      0.671 5.02e- 1    0.652     2.42
## 11 EducationLeft schoo~    3.06     0.495      2.26  2.40e- 2    1.17      8.22
## 12 EducationMasters de~    0.703    0.273     -1.29  1.97e- 1    0.413     1.21
## 13 EducationProfession~    0.888    0.278     -0.427 6.69e- 1    0.517     1.54
## 14 EducationSome colle~    1.23     0.269      0.755 4.50e- 1    0.726     2.09
## 15 EducationUniversity~    0.859    0.259     -0.588 5.57e- 1    0.519     1.44
## 16 Nscore                  1.46     0.0634     5.97  2.35e- 9    1.29      1.66
## 17 Oscore                  1.49     0.0611     6.49  8.48e-11    1.32      1.68
```

```
## 18 Impulsive              1.07     0.0714     0.989 3.22e- 1    0.933     1.23
## 19 SS                     1.30     0.0761     3.49  4.89e- 4    1.12      1.51
## 20 Cscore                 0.929    0.0621    -1.19  2.35e- 1    0.822     1.05
## # ... with 2 more rows
```

Results:

Discussion:

References: