Final Report

due November 16, 2021 by 11:59 PM

G.I. Joe: Isa Rundell and Grace Vo

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

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

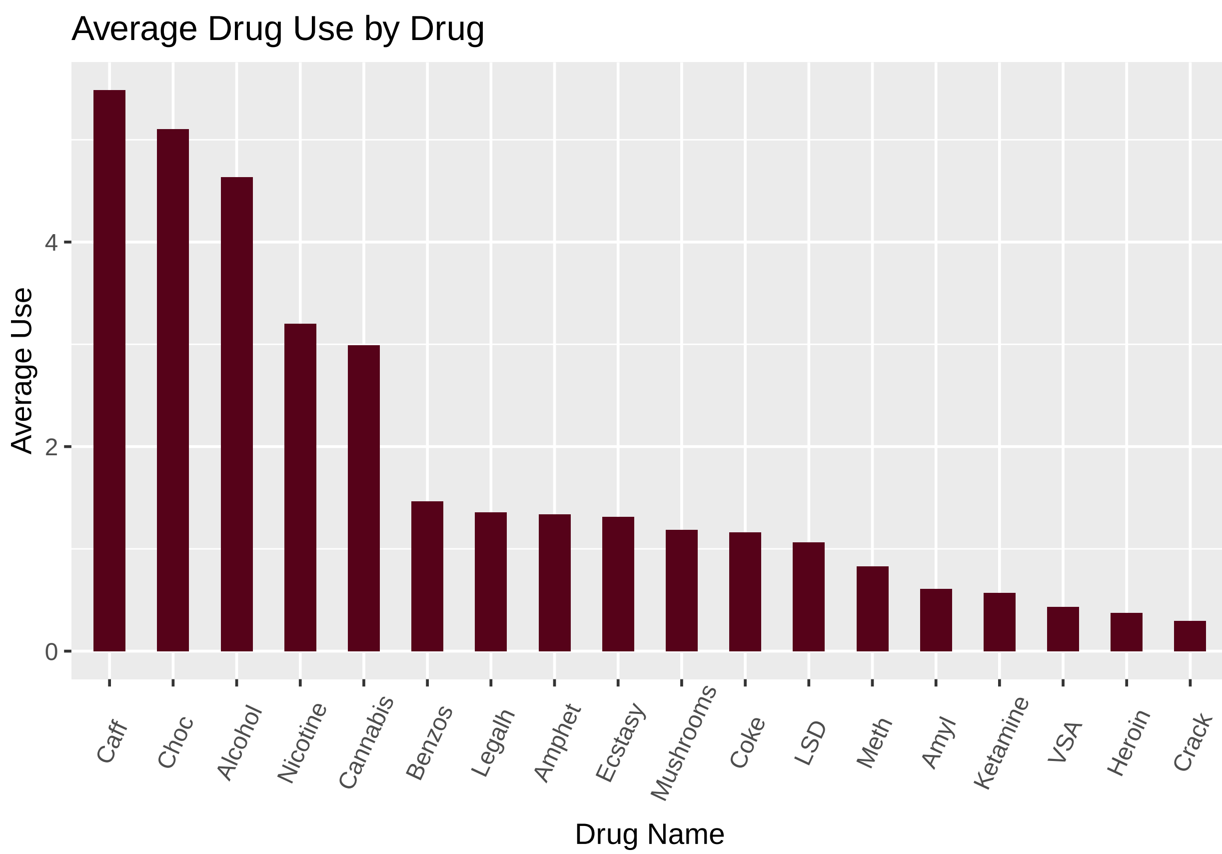
Abstract:

Background and Significance:

Methods: a) Data Collection and Variables

b) Exploratory Data Analysis

```r
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))
```




```r
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)))

head(numdrug)
```

```
## # A tibble: 6 x 32
##      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 22 more variables: Cscore <dbl>, Impulsive <dbl>, SS <dbl>,
## #   Alcohol <dbl>, Amphet <dbl>, Amyl <dbl>, Benzos <dbl>, Caff <dbl>,
## #   Cannabis <dbl>, Choc <dbl>, Coke <dbl>, Crack <dbl>, Ecstasy <dbl>,
## #   Heroin <dbl>, Ketamine <dbl>, Legalh <dbl>, LSD <dbl>, Meth <dbl>,
## #   Mushrooms <dbl>, Nicotine <dbl>, Semer <dbl>, VSA <dbl>
```

```r
correlation_matrix <- round(cor(numdrug),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
```

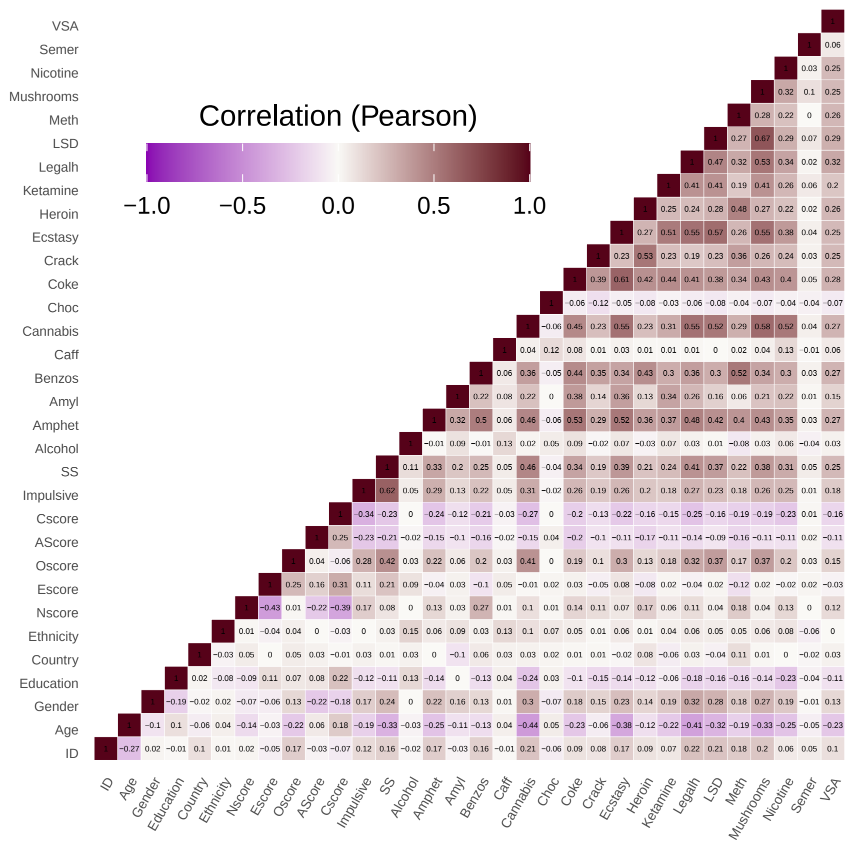
```
## Age       -0.27  1.00  -0.10      0.10   -0.06      0.04  -0.14  -0.03   -0.22
## Gender     0.02 -0.10   1.00     -0.19   -0.02      0.02  -0.07  -0.06    0.13
## Education -0.01  0.10  -0.19      1.00    0.02     -0.08  -0.09   0.11    0.07
## Country    0.10 -0.06  -0.02      0.02    1.00     -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 Alcohol Amphet  Amyl Benzos  Caff
## ID         -0.03  -0.07      0.12  0.16   -0.02   0.17 -0.03   0.16 -0.01
## Age         0.06   0.18     -0.19 -0.33   -0.03  -0.25 -0.11  -0.13  0.04
## Gender     -0.22  -0.18      0.17  0.24    0.00   0.22  0.16   0.13  0.01
## Education   0.08   0.22     -0.12 -0.11    0.13  -0.14  0.00  -0.13  0.04
## Country     0.03  -0.01      0.03  0.01    0.03   0.00 -0.10   0.06  0.03
## Ethnicity   0.00  -0.03      0.00  0.03    0.15   0.06  0.09   0.03  0.13
##           Cannabis  Choc  Coke Crack Ecstasy Heroin Ketamine Legalh   LSD  Meth
## ID            0.21 -0.06  0.09  0.08    0.17   0.09     0.07   0.22  0.21  0.18
## Age          -0.44  0.05 -0.23 -0.06   -0.38  -0.12    -0.22  -0.41 -0.32 -0.19
## Gender        0.30 -0.07  0.18  0.15    0.23   0.14     0.19   0.32  0.28  0.18
## Education    -0.24  0.03 -0.10 -0.15   -0.14  -0.12    -0.06  -0.18 -0.16 -0.16
## Country       0.03  0.02  0.01  0.01   -0.02   0.08    -0.06   0.03 -0.04  0.11
## Ethnicity     0.10  0.07  0.05  0.01    0.06   0.01     0.04   0.06  0.05  0.05
##           Mushrooms Nicotine Semer   VSA
## ID             0.20     0.06  0.05  0.10
## Age           -0.33    -0.25 -0.05 -0.23
## Gender         0.27     0.19 -0.01  0.13
## Education     -0.14    -0.23 -0.04 -0.11
## Country        0.01     0.00 -0.02  0.03
## Ethnicity      0.06     0.08 -0.06  0.00
```

```r
  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
```

```r
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(),
```

```r
  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))
```




```r
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")))
```

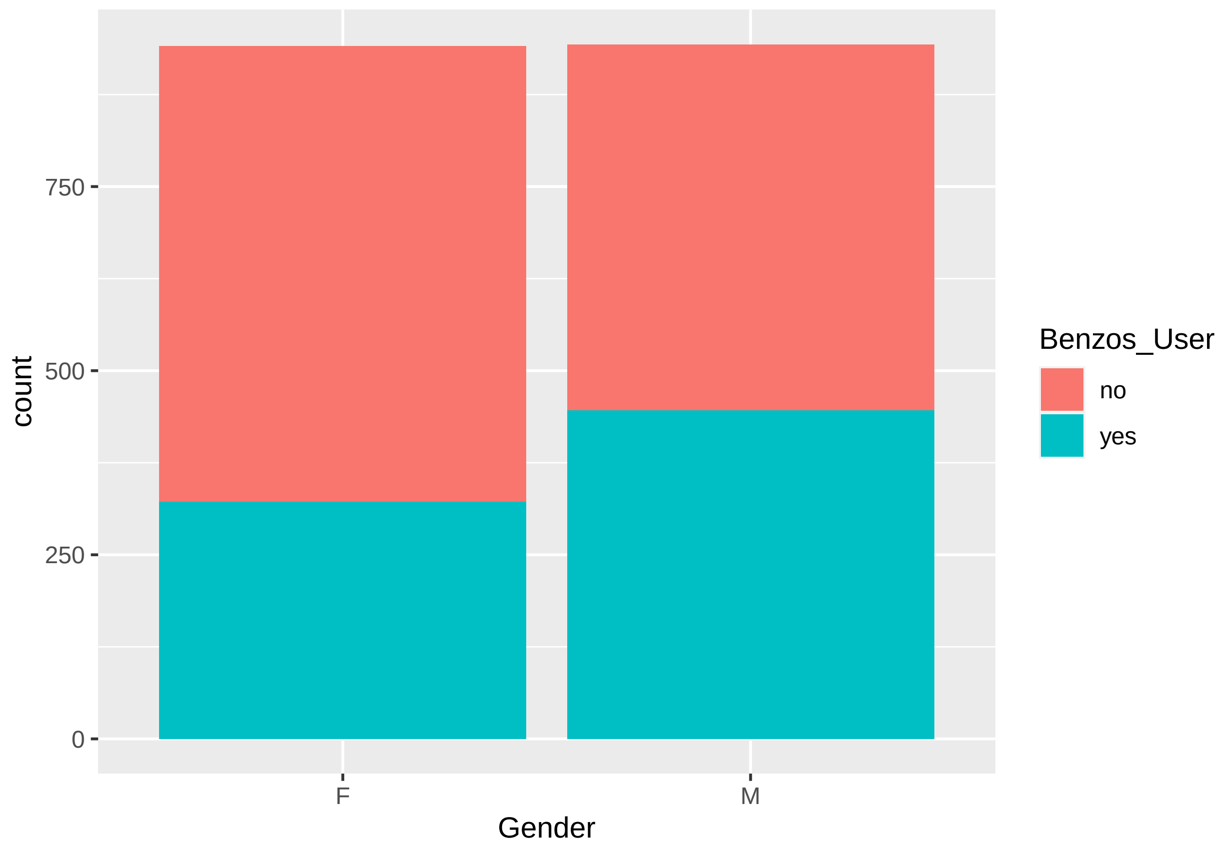
```
drug_byuse <- numdrug %>%
  mutate(Alcohol_User = ifelse(Alcohol > 1, 1, 0),
        Amphetamine_User = ifelse(Amphet > 1, 1, 0),
        AmylNitrite_User = ifelse(Amyl > 1, 1, 0),
        Benzos_User = ifelse(Benzos > 1, 1, 0),
        Caffeine_User = ifelse(Caff > 1, 1, 0),
        Cannabis_User = 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)
```

```r
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

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



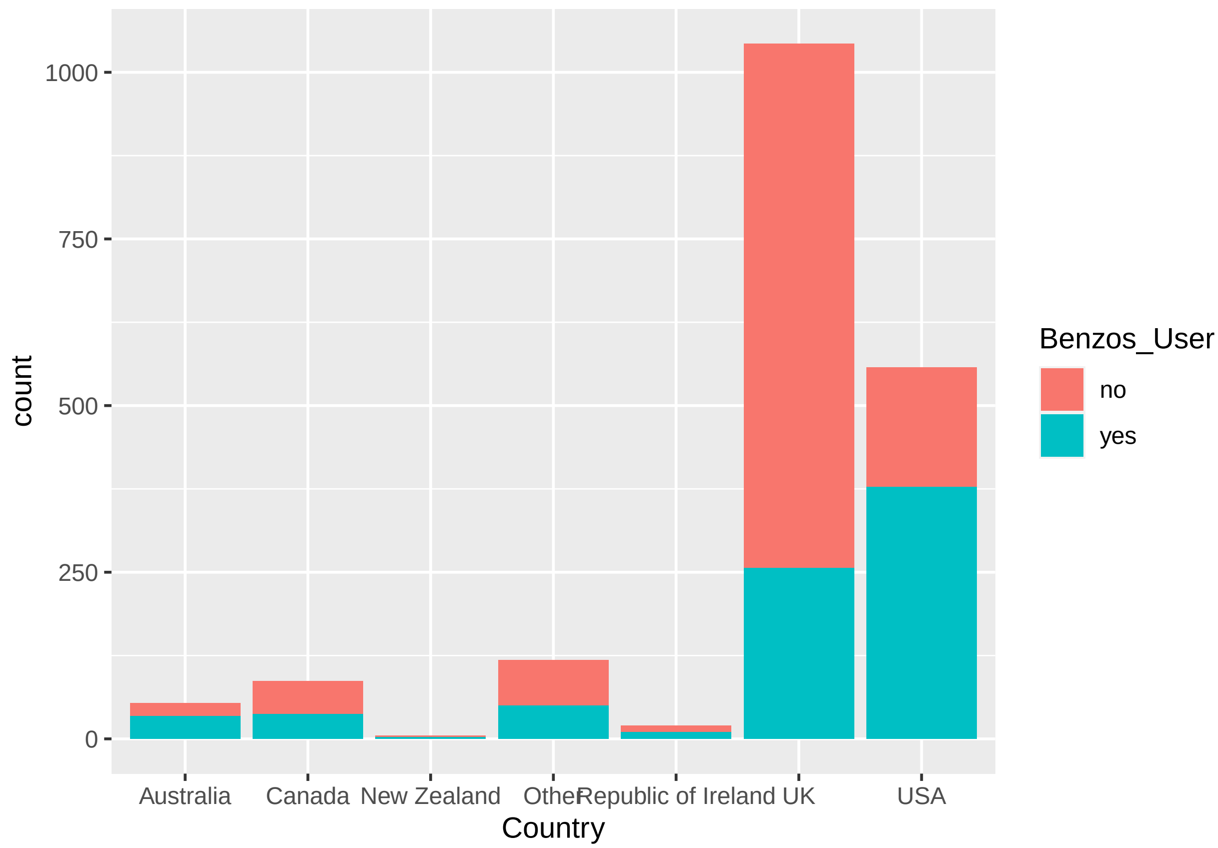

```r
fisher.test(drug_clean_2$Gender, drug_clean_2$Benzos_User)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  drug_clean_2$Gender and drug_clean_2$Benzos_User
## p-value = 7.832e-09
```

```
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.426549 2.086420
## sample estimates:
## odds ratio
##   1.724624
```

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

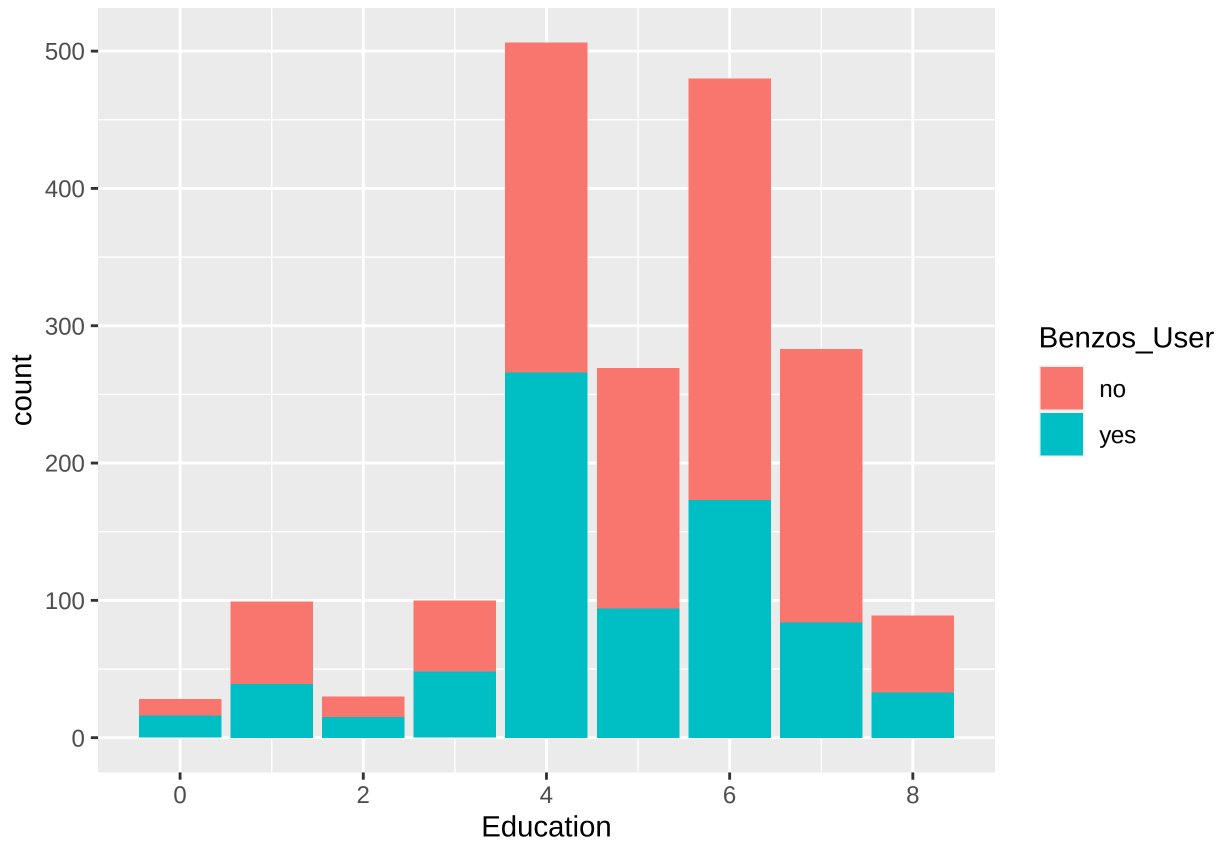
```
drug_clean_2 %>%
  ggplot(aes(x = Country,
             fill = Benzos_User)) +
  geom_bar()
```



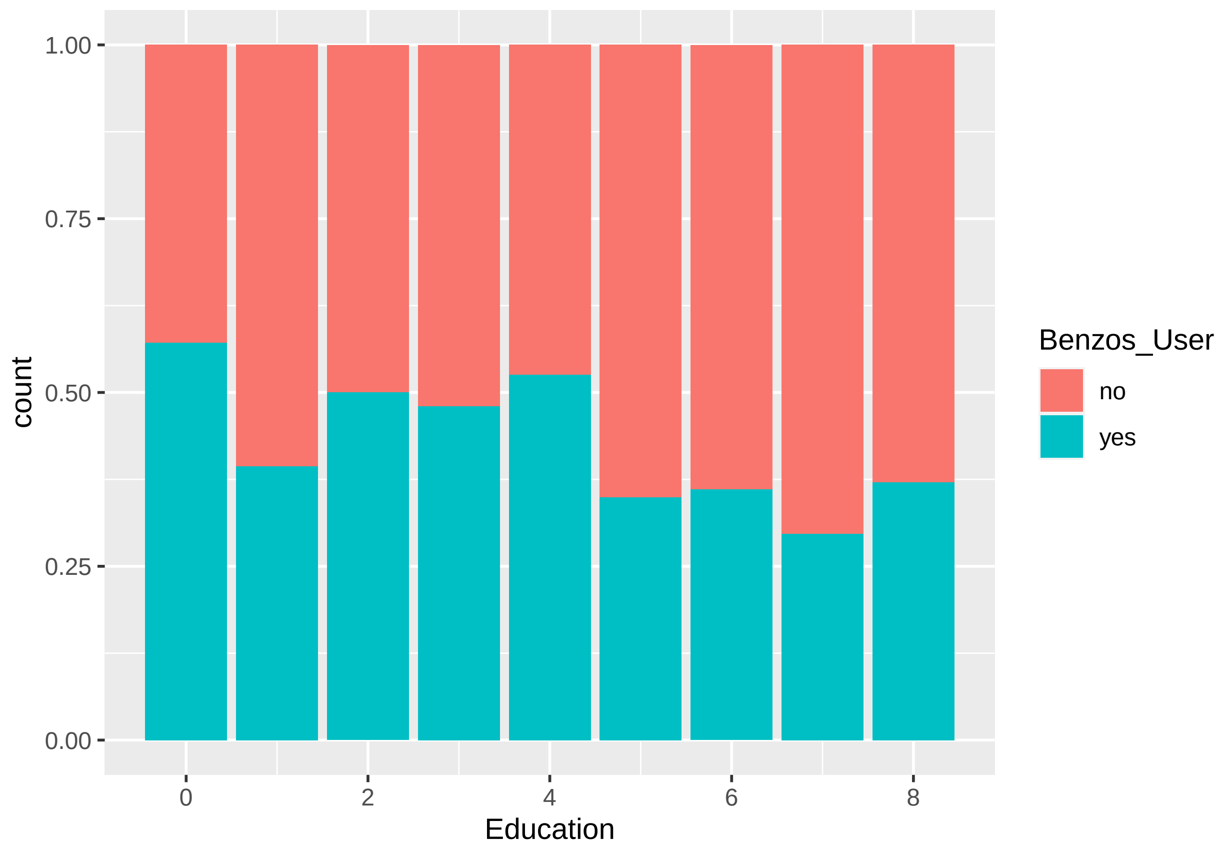

```
#null_distribution_simulated <- drug_clean_2 %>%
  #specify(Country ~ Benzos_User) %>%
  #hypothesize(null = "independence") %>%
  #generate(reps = 5000, type = "permute") %>%
  #calculate(stat = "Chisq")

#null_distribution_simulated %>%
  #visualize(method = "both") +
  #shade_p_value(observed_chisq_statistic,
               #direction = "greater")
```

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

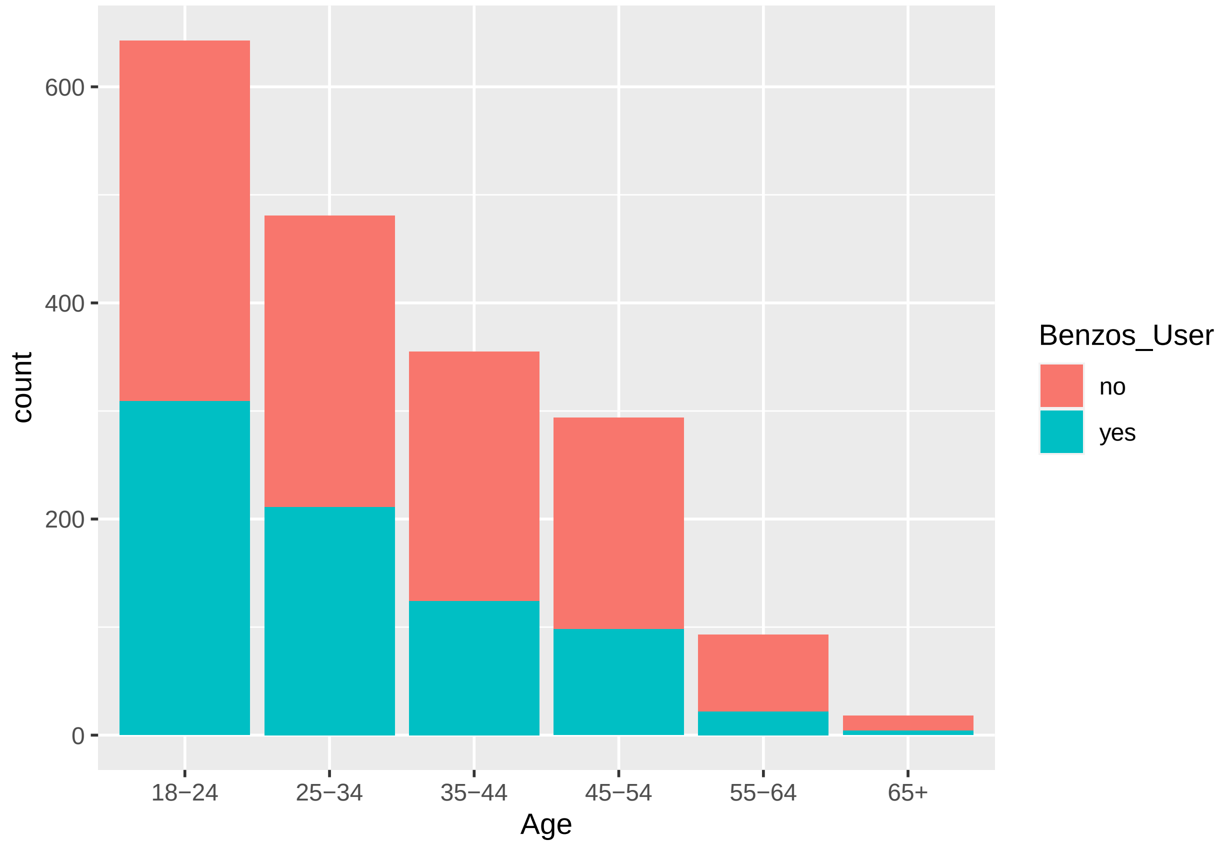
```
  geom_bar()
```



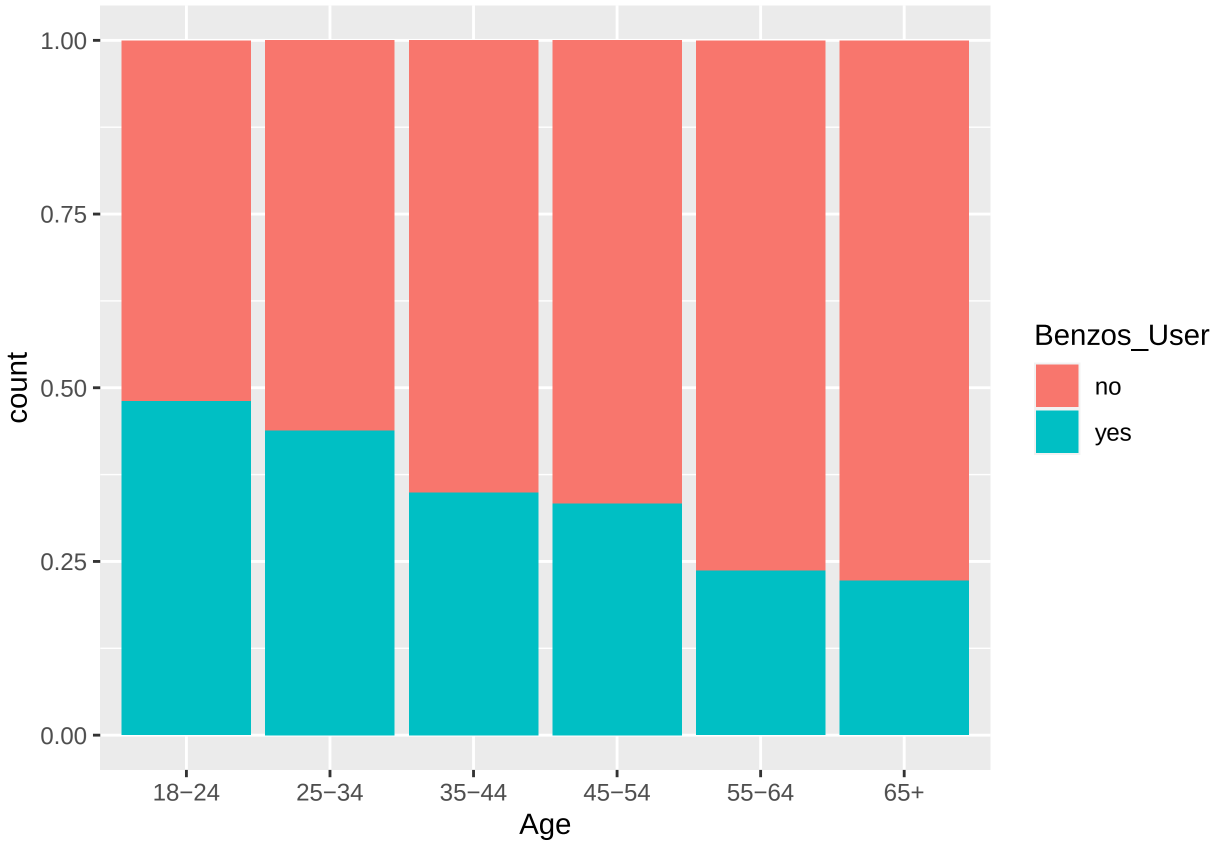

```
drug_clean %>%
  ggplot(aes(x = Education,
             fill = Benzos_User)) +
  geom_bar(position = "fill")
```



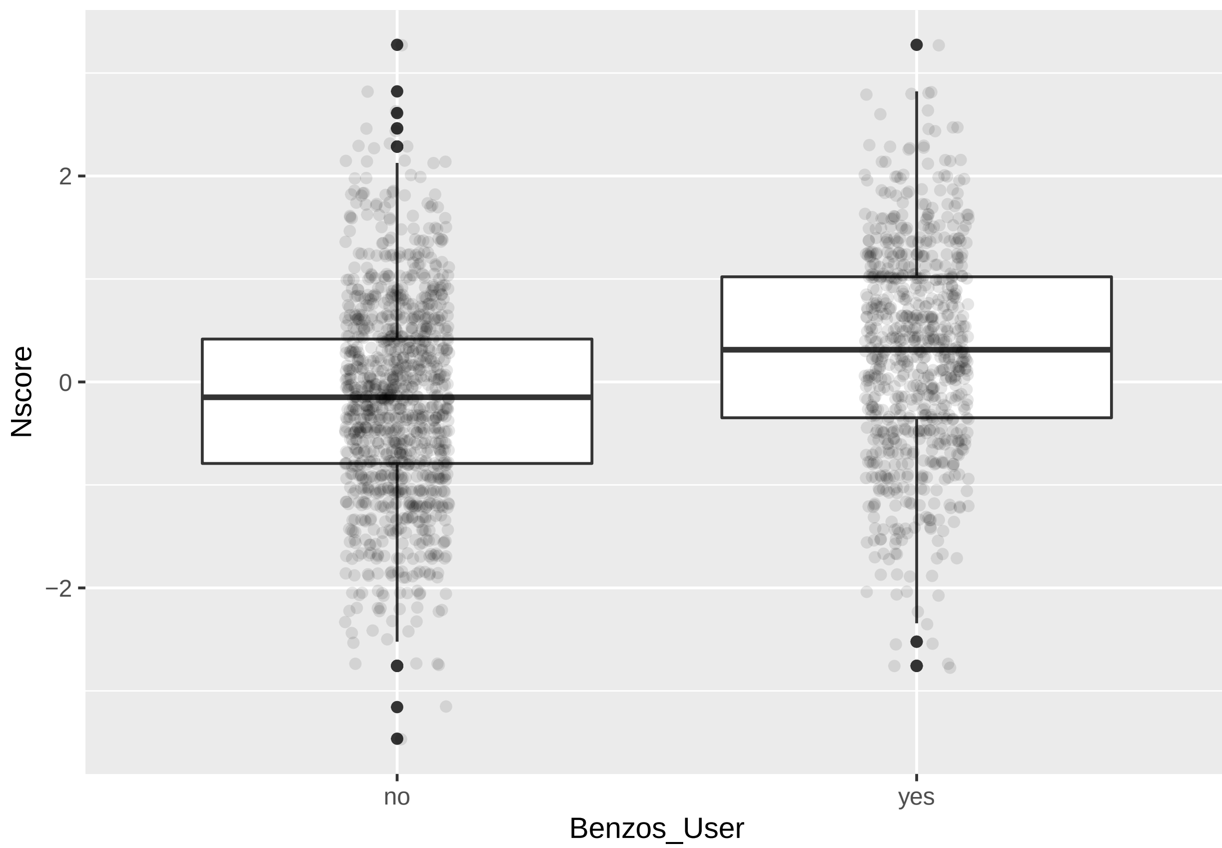

```
drug_clean_2 %>%
  ggplot(aes(x = Age,
             fill = Benzos_User)) +
  geom_bar()
```



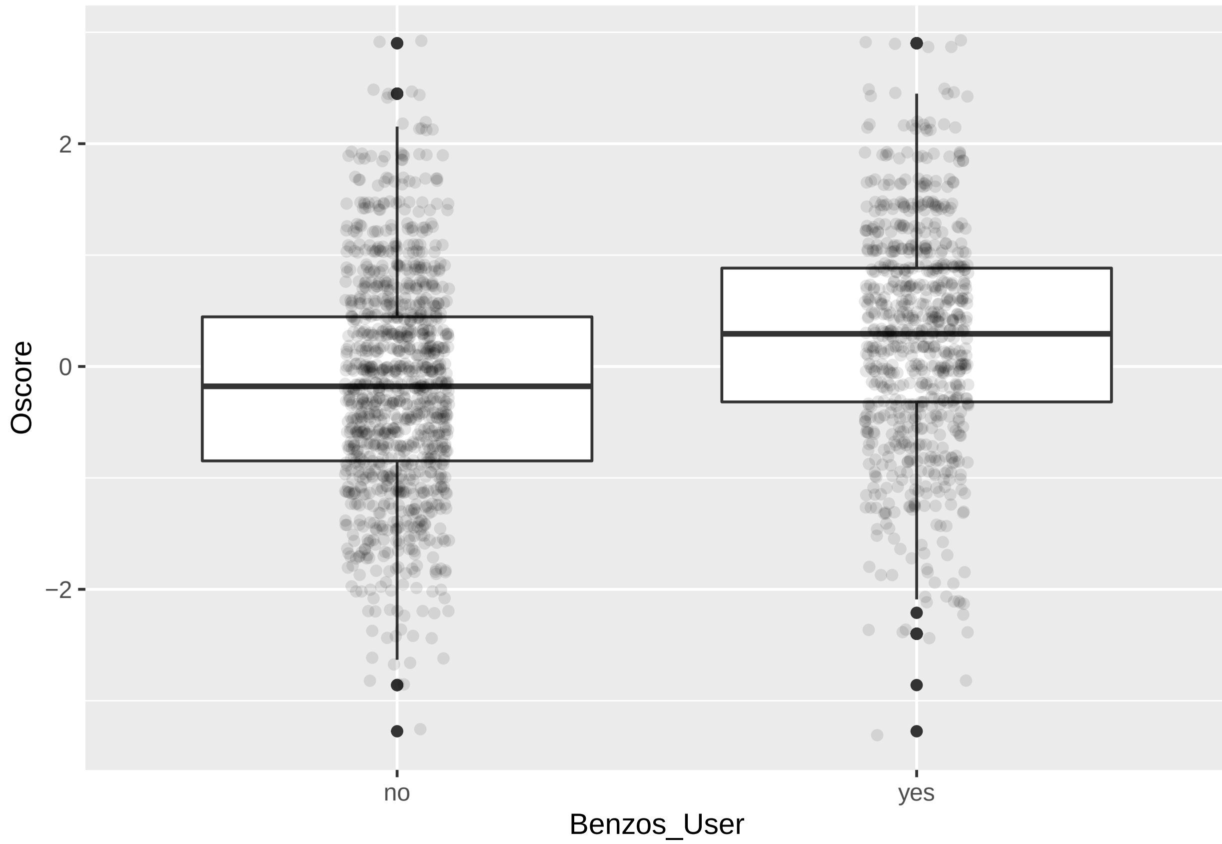

```r
drug_clean_2 %>%
  ggplot(aes(x = Age,
             fill = Benzos_User)) +
  geom_bar(position = "fill")
```



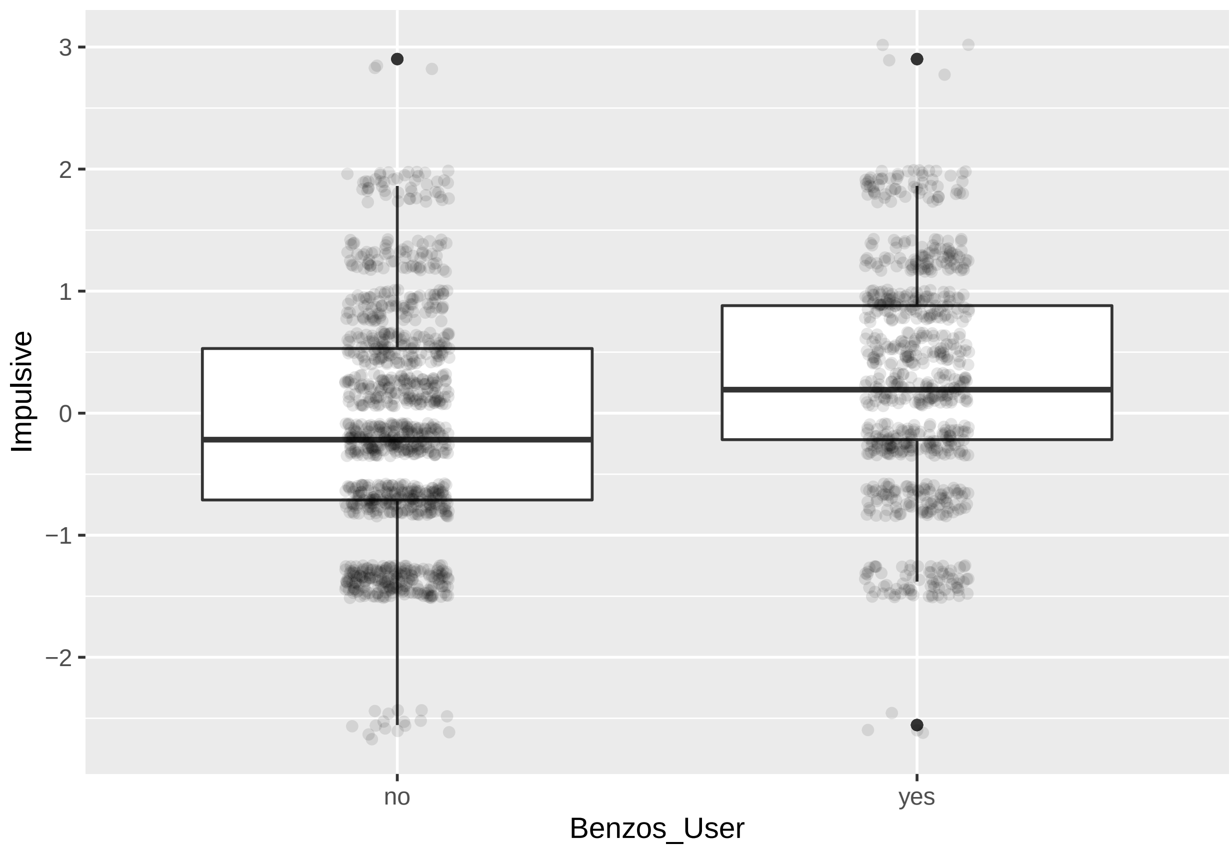

```
drug_clean %>%
  ggplot(aes(Benzos_User, Nscore)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.1)
```




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




```
drug_clean %>%
  ggplot(aes(Benzos_User, Impulsive)) +
  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, data=drug_clean_2, fami
result<-tidy(fit_multi, conf.int=TRUE, exponentiate=TRUE)
print(result, n=20)
```

```
## # A tibble: 19 x 7
##    term                  estimate std.error statistic  p.value conf.low conf.high
##    <chr>                    <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
##  1 (Intercept)              0.468    0.269     -2.82  4.78e- 3    0.274     0.789
##  2 GenderM                  1.49     0.108      3.67  2.38e- 4    1.20      1.84
##  3 Age25-34                 1.45     0.144      2.58  9.91e- 3    1.09      1.92
##  4 Age35-44                 1.14     0.162      0.821 4.12e- 1    0.832     1.57
##  5 Age45-54                 1.16     0.174      0.836 4.03e- 1    0.822     1.63
##  6 Age55-64                 0.697    0.277     -1.30  1.93e- 1    0.398     1.18
##  7 Age65+                   0.718    0.637     -0.520 6.03e- 1    0.179     2.30
##  8 EducationLeft schoo~     1.21     0.330      0.565 5.72e- 1    0.631     2.31
##  9 EducationLeft schoo~     1.23     0.464      0.453 6.51e- 1    0.496     3.09
## 10 EducationLeft schoo~     1.32     0.332      0.828 4.08e- 1    0.689     2.53
## 11 EducationLeft schoo~     3.05     0.493      2.27  2.35e- 2    1.17      8.18
## 12 EducationMasters de~     0.693    0.272     -1.35  1.78e- 1    0.408     1.19
## 13 EducationProfession~     0.884    0.277     -0.445 6.57e- 1    0.516     1.53
## 14 EducationSome colle~     1.28     0.267      0.917 3.59e- 1    0.760     2.17
## 15 EducationUniversity~     0.867    0.258     -0.553 5.80e- 1    0.525     1.45
## 16 Nscore                   1.61     0.0545     8.71  3.08e-18    1.45      1.79
## 17 Oscore                   1.42     0.0590     5.98  2.27e- 9    1.27      1.60
```

```
## 18 Impulsive                  1.10     0.0690     1.39  1.65e- 1    0.962     1.26
## 19 SS                         1.30     0.0745     3.49  4.90e- 4    1.12      1.50
```

Results:

Discussion:

References: