

Project Proposal

due October 11, 2021 by 11:59 PM

G.I. Joe - Isabella Rundell and Grace Vo

October 11, 2021

Load Packages

```
install.packages("taRifx")  
library(tidyverse)  
library(dplyr)  
library(taRifx)
```

Load Data

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

Introduction and Data, including Research Questions

The goal of this research is to determine whether or not lower personality scores correlate with more frequent abuse of illegal drugs. Further, how do these trends differ across gender and age lines? A study conducted by Turiano, Nicholas A et al., “Personality and Substance Use in Midlife: Conscientiousness as a Moderator and the Effects of Trait Change,” highlights the cruciality of examining the links between personality and substance abuse, for the former is a prime predictor of the latter across stages of life (Turiano et al., 2012). This dataset amasses figures pertaining to the drug consumption and personality scores of 1885 participants hailing from predominantly white, English speaking countries. The data include observations on both legal and illegal drugs: alcohol, amphetamines, amyl nitrite, benzodiazepine, cannabis, chocolate, cocaine, caffeine, crack, ecstasy, heroin, ketamine, legal highs, LSD, methadone, mushrooms, nicotine, and a class of volatile substance abuse. The various personality traits, neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness, were quantified using the NEO-FFI-R measurement, and impulsivity and sensation seeking attributes were measured using BIS-11 and ImpSS, respectively. The dataset also contains the binary gender identity, age category, ethnicity, country of residence, and educational background of all of the participants. For the purposes of this research project, ethnicity, country of residence, and educational background are likely to be unimportant or unhelpful given that the vast majority are white and the data on education are not readily quantifiable. Further, the primary focus will be on the use, or lack thereof, of illegal drugs and will not qualify the legal drugs as “drug usage.”

Glimpse

```
glimpse(drug, width = getOption("width"))
```

```
## Rows: 1,884  
## Columns: 32
```

```
## $ ID      <dbl> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, ~
## $ Age     <chr> "25-34", "35-44", "18-24", "35-44", "65+", "45-54", "35-44", ~
## $ Gender  <chr> "M", "M", "F", "F", "F", "M", "M", "F", "M", "F", "M", "F", ~
## $ Education <chr> "Doctorate degree", "Professional certificate/ diploma", "Ma~
## $ Country <chr> "UK", "UK", "UK", "UK", "Canada", "USA", "UK", "Canada", "UK~
## $ Ethnicity <chr> "White", "White", "White", "White", "White", "White", "White~
## $ Nscore  <dbl> -0.67825, -0.46725, -0.14882, 0.73545, -0.67825, -0.46725, --
## $ Escore  <dbl> 1.93886, 0.80523, -0.80615, -1.63340, -0.30033, -1.09207, 1.~
## $ Oscore  <dbl> 1.43533, -0.84732, -0.01928, -0.45174, -1.55521, -0.45174, --
## $ AScore  <dbl> 0.76096, -1.62090, 0.59042, -0.30172, 2.03972, -0.30172, -0.~
## $ Cscore  <dbl> -0.14277, -1.01450, 0.58489, 1.30612, 1.63088, 0.93949, 1.63~
## $ Impulsive <dbl> -0.71126, -1.37983, -1.37983, -0.21712, -1.37983, -0.21712, ~
## $ SS      <dbl> -0.21575, 0.40148, -1.18084, -0.21575, -1.54858, 0.07987, -0~
## $ Alcohol <chr> "CL5", "CL6", "CL4", "CL4", "CL2", "CL6", "CL5", "CL4", "CL6~
## $ Amphet  <chr> "CL2", "CL0", "CL0", "CL1", "CL0", "CL0", "CL0", "CL0", "CL1~
## $ Amyl    <chr> "CL2", "CL0", "CL0", "CL1", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Benzos  <chr> "CL0", "CL0", "CL3", "CL0", "CL0", "CL0", "CL0", "CL0", "CL1~
## $ Caff    <chr> "CL6", "CL6", "CL5", "CL6", "CL6", "CL6", "CL6", "CL6", "CL6~
## $ Cannabis <chr> "CL4", "CL3", "CL2", "CL3", "CL0", "CL1", "CL0", "CL0", "CL1~
## $ Choc    <chr> "CL6", "CL4", "CL4", "CL6", "CL4", "CL5", "CL4", "CL6", "CL6~
## $ Coke    <chr> "CL3", "CL0", "CL2", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Crack   <chr> "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Ecstasy <chr> "CL4", "CL0", "CL0", "CL1", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Heroin  <chr> "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Ketamine <chr> "CL2", "CL0", "CL2", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Legalh  <chr> "CL0", "CL0", "CL0", "CL1", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ LSD     <chr> "CL2", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Meth    <chr> "CL3", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Mushrooms <chr> "CL0", "CL1", "CL0", "CL2", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ Nicotine <chr> "CL4", "CL0", "CL2", "CL2", "CL6", "CL6", "CL0", "CL6", "CL6~
## $ Semer   <chr> "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
## $ VSA     <chr> "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0", "CL0~
```

Data Analysis Plan

In order to analyze these data, the drug usage of the various illegal drugs will be considered the outcome variable while the personality scores will be the explanatory variable. Both gender and age categories are additional data that constitute comparison groups to facilitate the answering of the overarching research question and provide compelling juxtapositions. To visualize these trends, a bar plot that has age on the x-axis, average drug use on the y-axis, is dodged by gender, and faceted by drug type would be helpful to see the relationship between all of these variables. Finally, an ANOVA model is a statistical method that will prove very helpful in answering the proposed research question. An ANOVA model can help determine if there is sufficient evidence that lower personality scores lead to more frequent illegal drug usage. “Drug usage” is quantified by the average use of each drug based on the rating. In other words, a drug with an average rating of 5 is considered more frequently used than a drug with an average rating of 2.

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

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

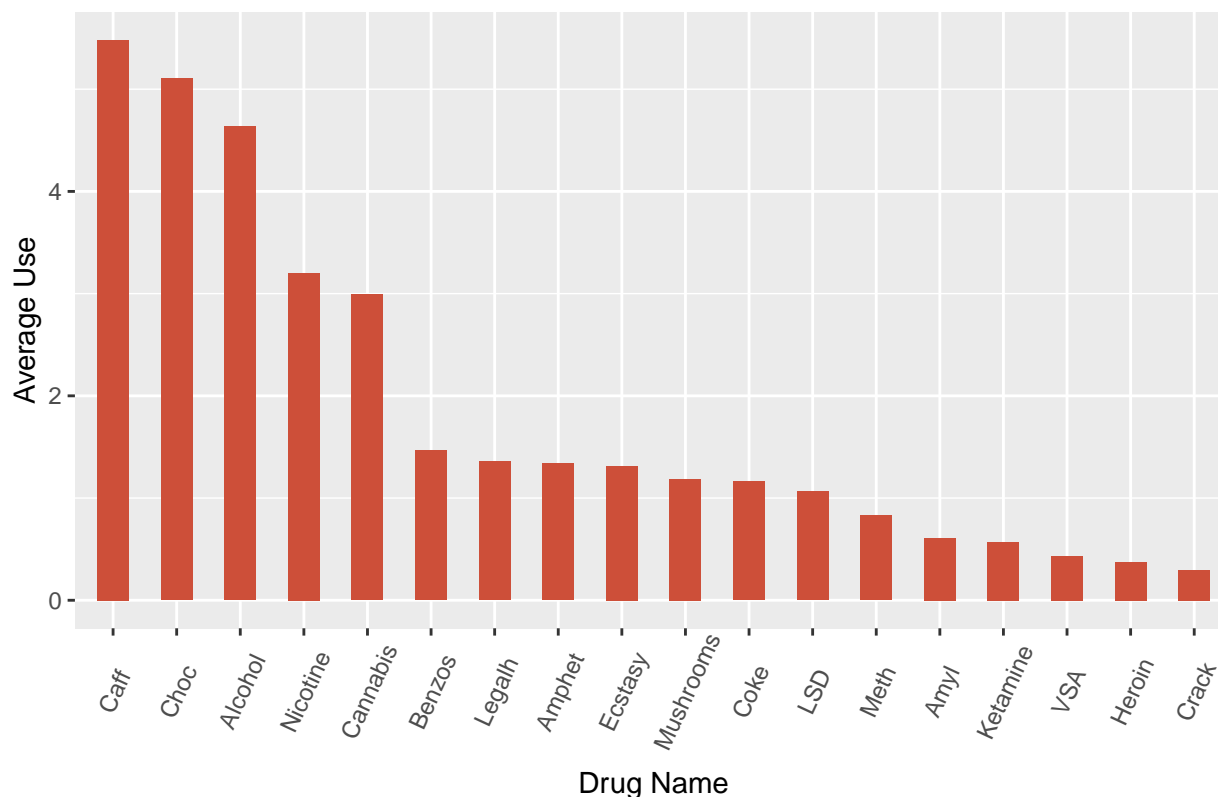
```
drug_name <- c('Alcohol', 'Amphet', 'Amyl', 'Benzos', 'Caff', 'Cannabis', 'Choc', 'Coke', 'Crack', 'Ecs
average_use <- c(4.63481953, 1.34023355, 0.60721868, 1.46496815, 5.48354565, 2.99097665, 5.10668790, 1.~
```

```
drug_averages <- data.frame(drug_name, average_use) %>%
print
```

```
##      drug_name average_use
## 1      Alcohol  4.6348195
## 2       Amphet  1.3402335
## 3        Amyl   0.6072187
## 4       Benzos  1.4649682
## 5        Caff  5.4835456
## 6     Cannabis  2.9909766
## 7        Choc  5.1066879
## 8         Coke  1.1618896
## 9        Crack  0.2977707
## 10    Ecstasy  1.3147558
## 11     Heroin  0.3742038
## 12   Ketamine  0.5695329
## 13    Legalh   1.3566879
## 14        LSD  1.0621019
## 15        Meth  0.8269639
## 16 Mushrooms  1.1878981
## 17  Nicotine  3.2011677
## 18         VSA  0.4336518
```

```
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="tomato3") +
  labs(title="Average Drug Use by Drug",
       x = "Drug Name",
       y = "Average Use") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```

Average Drug Use by Drug



```
num_drug <- drug1 %>%
  mutate(Age = replace(Age, Age == "18-24", 0), Age = replace(Age, Age == "25-34", 1), Age = replace(Age, Age == "35-44", 2), Age = replace(Age, Age == "45-54", 3), Age = replace(Age, Age == "55-64", 4), Age = replace(Age, Age == "65+", 5))

numdrug <- mutate_all(num_drug, 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>
```

```
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
```

```
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))

```

