Project Proposal

due October 11, 2021 by 11:59 PM

G.I. Joe - Isabella Rundell and Grace Vo

October 11, 2021

Load Packages

```
library(tidyverse)
library(dplyr)
```

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, drug use on the y-axis, is dodged by gender, and faced by drug type would be helpful to see the relationship between all of these variables. For certain aspects of analyzing the data, it will be useful to modify the data of the drug usage ratings which currently span from 0 to 6 corresponding to never used, used over a decade ago, used in last decade, used in last year, used in last month, used in last week, and used in last day, respectively. A more useful way of grouping the drug usage would be to only use a scale of 0 to 2 that corresponds to never used, used a long time ago, and used in the past month, respectively. Finally, a t-test is a statistical method that will prove very helpful in answering the proposed research question. A t-test can help determine if there is sufficient evidence that lower personality scores lead to higher levels of drug use. The greater the difference between personality scores and drug usage, the greater the support for this hypothesis.