Application Research Report

Paige X. Phillips

11/27/22

Table of contents

# Where Is Your Data From?

**Application Area**: Neuroscience

**Article**: The Five Factor Model of Personality and Evaluation of Drug Consumption Risk

**Database**: [Drug Consumptions (UCI)](https://www.kaggle.com/datasets/obeykhadija/drug-consumptions-uci)

**DOI**: <https://doi.org/10.1007/978-3-319-55723-6_18>

**Journal**: [SpringerLink](https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18#Sec2)

You can access this page here: [GitHub](https://github.com/paph3285/APR_Drug_Consumption)

**Reference**

|  |
| --- |
| Fehrman, Elaine, et al. “The Five Factor Model of Personality and Evaluation of Drug Consumption Risk.” *SpringerLink*, Springer International Publishing, 5 July 2017, <https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18#Sec2>.  Khadija. “Drug Consumptions (UCI).” *Kaggle*, 26 Sept. 2021, <https://www.kaggle.com/datasets/obeykhadija/drug-consumptions-uci>. |

# Why Should I Care?

Change is inevitable and one of the hot topics in change is the discussion, advancement, and push to legalize different drugs. However, we need to remind ourselves before we get too carried away, drug consumption usage is very much still an on-going global problem. Such risk behavior cannot be isolated alone but constitute from a multitude of factors. These factors are defined as anything that increases the chances or probability of one’s drug consumption(s). Primary this can be predicted and measured through an individual’s attribute, characteristics, and/or life events

## Summary

### Topic:

Assessing one’s risk of drug consumption use and misuse according to their personality trait(s).

### Data Collection:

The study initially received 2051 responses, but only included 1885 respondents. 12 attributes and 18 central nervous system (CNS) psychoactive drugs were included in the study. The 12 attributes can be categorized based on demographic and personality measurements. The 18 CNS psychoactive drugs includes both illicit and licit drugs. Additionally, there was a fake drug (Semeron) included to help identify any over-claimers in the study. Finally, respondents were asked to rate their drug use per drug across seven different categories. Based on their response of drug use, respondents were classified either as users or non-users.

| ID | Age | Gender | Education | Country | Ethnicity | Nscore | Escore | Oscore | AScore | Cscore | Impulsive | SS | Alcohol | Amphet | Amyl | Benzos | Caff | Cannabis | Choc | Coke | Crack | Ecstasy | Heroin | Ketamine | Legalh | LSD | Meth | Mushrooms | Nicotine | Semer | VSA |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 25-34 | M | Doctorate degree | UK | White | -0.67825 | 1.93886 | 1.43533 | 0.76096 | -0.14277 | -0.71126 | -0.21575 | CL5 | CL2 | CL2 | CL0 | CL6 | CL4 | CL6 | CL3 | CL0 | CL4 | CL0 | CL2 | CL0 | CL2 | CL3 | CL0 | CL4 | CL0 | CL0 |
| 3 | 35-44 | M | Professional certificate/ diploma | UK | White | -0.46725 | 0.80523 | -0.84732 | -1.62090 | -1.01450 | -1.37983 | 0.40148 | CL6 | CL0 | CL0 | CL0 | CL6 | CL3 | CL4 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL1 | CL0 | CL0 | CL0 |
| 4 | 18-24 | F | Masters degree | UK | White | -0.14882 | -0.80615 | -0.01928 | 0.59042 | 0.58489 | -1.37983 | -1.18084 | CL4 | CL0 | CL0 | CL3 | CL5 | CL2 | CL4 | CL2 | CL0 | CL0 | CL0 | CL2 | CL0 | CL0 | CL0 | CL0 | CL2 | CL0 | CL0 |
| 5 | 35-44 | F | Doctorate degree | UK | White | 0.73545 | -1.63340 | -0.45174 | -0.30172 | 1.30612 | -0.21712 | -0.21575 | CL4 | CL1 | CL1 | CL0 | CL6 | CL3 | CL6 | CL0 | CL0 | CL1 | CL0 | CL0 | CL1 | CL0 | CL0 | CL2 | CL2 | CL0 | CL0 |
| 6 | 65+ | F | Left school at 18 years | Canada | White | -0.67825 | -0.30033 | -1.55521 | 2.03972 | 1.63088 | -1.37983 | -1.54858 | CL2 | CL0 | CL0 | CL0 | CL6 | CL0 | CL4 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL6 | CL0 | CL0 |
| 7 | 45-54 | M | Masters degree | USA | White | -0.46725 | -1.09207 | -0.45174 | -0.30172 | 0.93949 | -0.21712 | 0.07987 | CL6 | CL0 | CL0 | CL0 | CL6 | CL1 | CL5 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL0 | CL6 | CL0 | CL0 |

***Demographic***

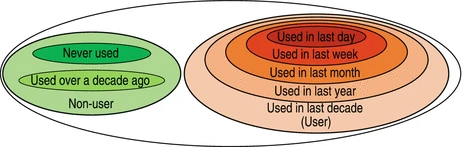
|  |
| --- |
| **Age**: 18-24, 25-34, 35-44, 45-54, 55-64 |
| **Country**: Australia, Canada, Other, USA, UK |
| **Gender**: Female, Male |
| **Ethnicity**: Asian, Black, Mixed-White/Black, Other, White |
| **Level of Education**: Left school at 18 years old, Some college or university with no certificate or degree, Professional certificate/diploma, University degree, Masters degree |

***Personality Measurements***

| Five Factor Model (FFM): |
| --- |
| 1. Agreeableness (A) |
| 1. Conscientiousness (C) |
| 1. Extraversion (E) |
| 1. Neuroticism (N) |
| 1. Openness to Experience (O) |
| Impulsivity (BIS-11) |
| Impulsiveness Sensation Seeking Scale (ImpSS) |

***CNS Psychoactive Drug Use***

|  |  |
| --- | --- |
| Alcohol | Ecstasy |
| Amphetamines | Heroin |
| Amyl nitrite | Ketamine |
| Benzodiazepines | Legal Highs |
| Cannabis | LSD |
| Chocolate | Methadone |
| Cocaine | Magic Mushrooms |
| Caffeine | Nicotine |
| Crack | Volatile Substance Abuse (VSA). |

[](https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18/figures/1)

***Category Group of Drug Users***

### Data Cleaning:

In order to measure and account for the accuracy of each drug measured across the groups, a number of different classification methods were implemented and tested. Only the most effective methods were selected to help identify between users and non-users for each drug.

***Data Mining Classification Methods***

|  |
| --- |
| Decision Tree (DT) |
| Gaussian Mixture |
| K-Nearest Neighbours (KNN) |
| Linear Discriminant Analysis (LDA) |
| Logistic Regression |
| Naïve Bayes |
| Probability Density Function Estimation (PDFE) |
| Random Forest |

# What Does It Tell You?

### Data Analysis:

***Quantifying Categorical Variables***

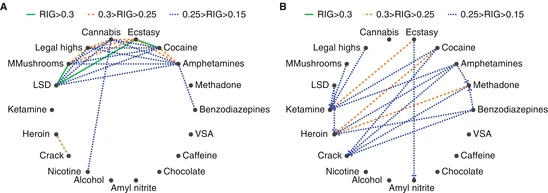
| Ordinal Features Quantification | Nominal Feature Quantification |
| --- | --- |
| Polychoric Correlation (PolC) | Nonlinear Categorical Principal Components Analysis (CatPCA) |

***Rule of Thumb:*** The greater the value of RIG, the stronger is the indicated correlation

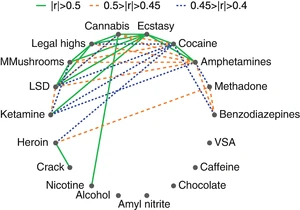
* We can consider any correlation to be considered **weak** **if | *r* | < 0. 4**
* We can consider any correlation to be considered **medium** **if 0. 45 > | *r* | ≥ 0. 4**
* We can consider any correlation to be considered **strong** **if 0. 5 > | *r* | ≥ 0. 45**
* We can consider any correlation to be considered **very strong** **if | *r* | ≥ 0. 5**

### Data Visualizations:

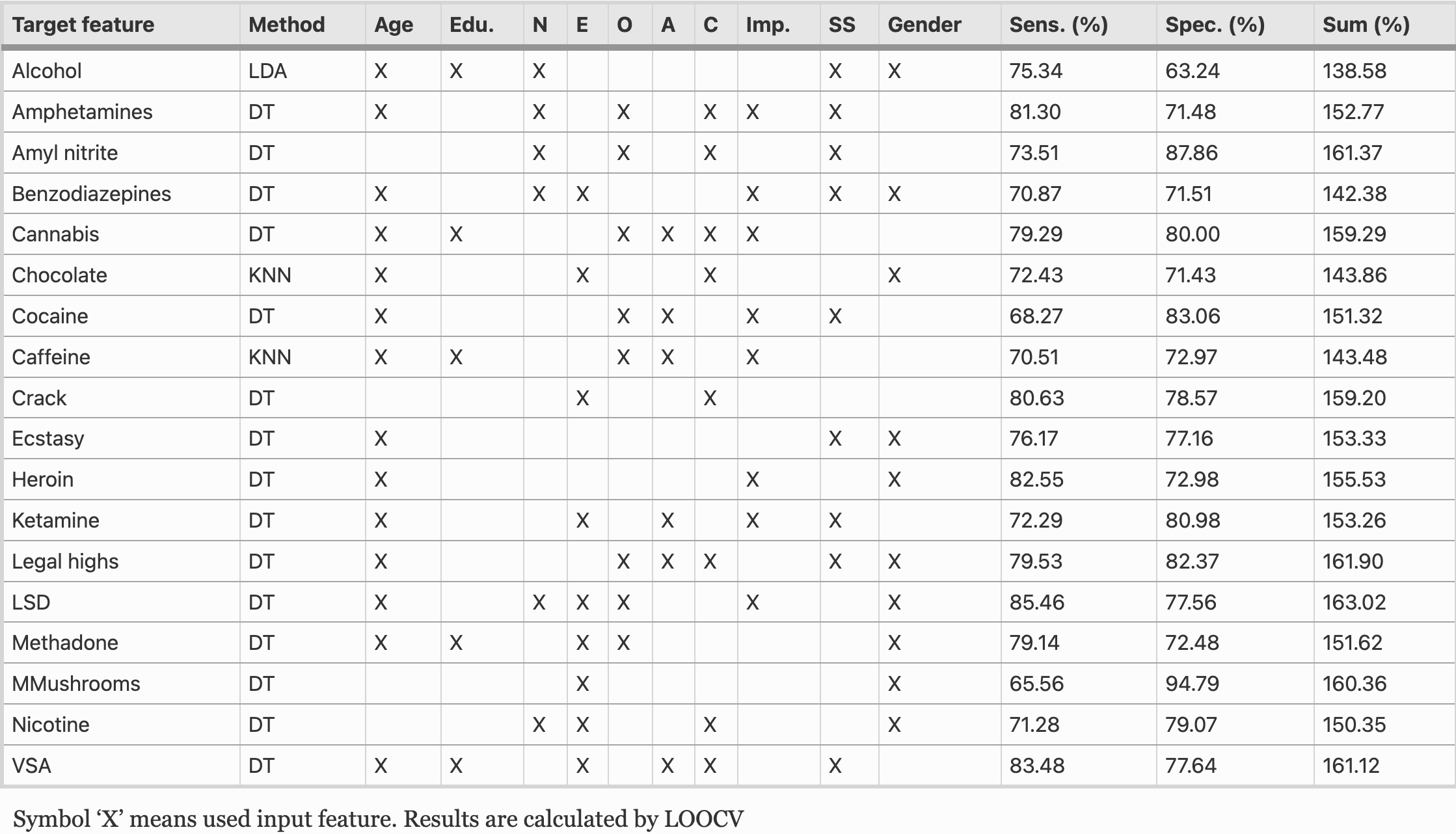
For our specific data we are using, please note that any correlations **above 0.4** are to be considered relatively *strong*; any correlations **between 0.2 and 0.4** are to be considered relatively *moderate*, and for those correlations **below 0.2** are to considered to be relatively *weak*.

[](https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18/figures/3)

***Medium to Weak Drug Use Correlation***

[](https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18/figures/2)

***Strong Drug Use Correlation***

[](https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18/figures/3)

***Best Classification Drug User***

# What Do You Conclude?

Based on the findings, there are strong correlations between consumption of certain group of drugs. For instance, amphetamines, cannabis, cocaine, ecstasy, legal highs, LSD, and magic mushrooms can be identified as a group of drugs with strong correlation consumption. The same can be said between the strongly correlated consumption of crack and heroin. In regards to the 12 attributes, attributes will differentiate for different drugs. For example, while age can be used as the best classifier for 14 drugs; gender can be used as the best classifiers for 10 drugs, while sensation seekers (SS) can be used as the best classifiers for 8 drugs. Lastly, the classification accuracy of the drugs are considered to be relatively high. The sensitivity and specificity are greater than 70% across nearly all drug classifications, except for alcohol, cocaine, and magic mushrooms.

# How Could You Be Wrong?

* Classification method testing - This is highly important to critically think and laterally think about how different methods of testing may get different sensitivity and specificity results.
* Sample size - This may not entirely reflect the general public and is bias to what is actually present in the real-world.
* Cultural - This was a global validation study and thus does not necessarily reflect inner cultural differences and identification of specific drug consumption.
* Bias - The word “drug(s)” carries negative connotation behind it. More often than not my mind jumps to 1) illegal, 2) substance use disorder, 3) ruins lives. Again, a lot of these are associated with more illicit drugs like cocaine, meth, heroin, etc., not so much with licit drugs such as alcohol, caffeine, cannabis, chocolate, and nicotine. However, that is not to say that others may not perceive those drugs in the same way as me. How I perceive these drugs is a bias of itself. Again, mainly due to the cultural climate I live in today.

# How Would One Pursue Studies In This Area?

## University of Colorado - Boulder Courses

Below outlines courses taught at CU Boulder that would be applicable to help me pursue my studies in this area.

| Class - Purpose | Requirements (Instructors) |
| --- | --- |
| **Neuropharmacology** **(NRSC 4132)** - Designed to provide a fundamental understanding of the neurobiological and neurochemical mechanisms of drug action within the central nervous system.   * Principles of pharmacology * Brain neurotransmitter systems * Biochemical basis of psychiatric disorders * Pharmacological treatment | Requires a prerequisite course of NRSC 2100/ NRSC 2150 (minimum grade C-)  (**Ryan Bachtell & Heidi Day**) |
| **Introduction to Neuroscience 1** **(NRSC 5100)** - Designed to provide a fundamental understanding of the principles of neuroscience.   * Neuroanatomy * Physiology * Neuropysiology * Neurochemical * Developmental characteristics of the CNS | Restricted to Psychology and Neuroscience (PSYC & NRSC) graduate students and students in the interdepartmental neuroscience program. Email jefferey.greeson@colorado.edu for assistance in enrolling.  (**David Root**) |
| **Data Mining (CSCI 5502**) - Designed to provide a fundamental understanding of data mining concepts and techniques for discovering interesting patterns hidden in large scale data sets.   * Focuses on issues relating to effectiveness and efficiency * Data preprocessing, warehouse, association, classification, clustering * Mining of specific data types | Restricted to graduate students only.  (**Qin Lv & Di Wu**) |
| **Methods in Statistical Learning (STAT 5600)** - Designed to provide a fundamental understanding of different statistical concepts, models, and algorithms of machine learning.   * Supervised learning for regression * Classification and resampling methods * Discriminant analysis * Classification and regression trees * Random forests and associated tuning * Diagnostics and performance evaluation | Requires prerequisite: APPM 3310 or CSCI 2820 or MATH 2130 or 2135 or 3130 or 3135) OR (APPM 3570 or 4570 or CHEN 3010 or CSCI 3022 or CVEN 3227 or ECEN 3810 or ECON 3818 or MATH 3510 or 4510 or MCEN 3047 or STAT 3100 or 4000 or 4520) (min grade B) Grad students only.  (**Ami Gates & David Quigley**) |
| **Information Visualization (INFO 5602)** - Designed to provide a fundamental understanding of visual representations of data.   * Using and building exploratory tools and data narratives * Interactive systems * User-centered and graphic design * Perception, data storytelling and analysis * Insight generation | Requires prerequisite course of STAT 5010 (minimum grade C-). Restricted to MS-DS students.  (**Osita Onyejekwe**) |
| **Machine Learning (CSCI 5622)** - Designed to provide a fundamental, practical, and theoretical understanding different widely used algorithms.   * Algorithms: neural networks, decision trees, support vector machines, Q-learning) * Building computer systems that learn from experience * Connecting data mining and statistical modeling | Restricted to graduate students only.  (**Abel Iyasele**) |

# sessionInfo()

R version 4.2.2 (2022-10-31)  
Platform: x86\_64-apple-darwin17.0 (64-bit)  
Running under: macOS Big Sur ... 10.16  
  
Matrix products: default  
BLAS: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib  
LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib  
  
locale:  
[1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base   
  
other attached packages:  
 [1] latexpdf\_0.1.7 tinytex\_0.42 forcats\_0.5.2 stringr\_1.4.1   
 [5] dplyr\_1.0.10 purrr\_0.3.5 readr\_2.1.3 tidyr\_1.2.1   
 [9] tibble\_3.1.8 ggplot2\_3.4.0 tidyverse\_1.3.2  
  
loaded via a namespace (and not attached):  
 [1] tidyselect\_1.2.0 xfun\_0.34 haven\_2.5.1   
 [4] gargle\_1.2.1 colorspace\_2.0-3 vctrs\_0.5.0   
 [7] generics\_0.1.3 htmltools\_0.5.3 yaml\_2.3.6   
[10] utf8\_1.2.2 rlang\_1.0.6 pillar\_1.8.1   
[13] withr\_2.5.0 glue\_1.6.2 DBI\_1.1.3   
[16] dbplyr\_2.2.1 modelr\_0.1.9 readxl\_1.4.1   
[19] lifecycle\_1.0.3 munsell\_0.5.0 gtable\_0.3.1   
[22] cellranger\_1.1.0 rvest\_1.0.3 evaluate\_0.17   
[25] knitr\_1.40 tzdb\_0.3.0 fastmap\_1.1.0   
[28] fansi\_1.0.3 highr\_0.9 broom\_1.0.1   
[31] backports\_1.4.1 scales\_1.2.1 googlesheets4\_1.0.1  
[34] jsonlite\_1.8.3 fs\_1.5.2 hms\_1.1.2   
[37] digest\_0.6.30 stringi\_1.7.8 grid\_4.2.2   
[40] cli\_3.4.1 tools\_4.2.2 magrittr\_2.0.3   
[43] crayon\_1.5.2 pkgconfig\_2.0.3 ellipsis\_0.3.2   
[46] xml2\_1.3.3 reprex\_2.0.2 googledrive\_2.0.0   
[49] lubridate\_1.9.0 timechange\_0.1.1 assertthat\_0.2.1   
[52] rmarkdown\_2.18 httr\_1.4.4 rstudioapi\_0.14   
[55] R6\_2.5.1 compiler\_4.2.2