Psychoanalytic Data Ontology Project

DRAFT

7/7/2023

Cameron More

1. Introduction

The purpose of this project is to demonstrate my skills as an ontology engineer. Over the course of this project, I primarily used the programming language Python and various libraries, including Pandas (for data analysis), Seaborn (for visualization), and RDFLib (for Owl manipulation). I also used the Web Ontology Language (OWL), in a TTL serialization, the Shapes Constraint Language (SHACL), SPARQL Protocol and RDF Query Language (SPARQL).

I chose to generate a dataset about psychoanalytic patients to analyze because of my subject-matter knowledge about psychoanalysis. Although thousands of people have undergone psychoanalytic analysis, there is no central quantitative data initiative. There are plenty of articles, workshops, and instructional texts that are written about specific cases in qualitative terms, and generalizations are often made by experienced therapists who have dealt with a number of cases. So for the purpose of this project, I supposed that anonymized data did exist that I could analyze.

2. Method

To generate the data, I asked OpenAI's ChatGPT to create a csv file of 150 patients which recorded basic demographic information–age, sex, and a unique ID number–and other factors like the frequency of dreams and nightmares (per week), number of verbal slips (per therapy session), number and types of symptoms, self-described level of addictive tendencies. ChatGPT did not provide me with the entire dataset, no matter how I prompted it, so I tried Google's Bard, which did eventually generate 150 lines of data (though I had to ask for it in 20 line intervals).

After collecting two .csv files, I merged the two together in order to create a file with 16 columns which include the information listed above. Then, I explored the data as a supposed psychoanalytic researcher interested in patients who experienced frequent nightmares. After analyzing the data, I created a number of visualizations which can be found in the appendix.

3. Analysis

My hypothetical framing for this exploratory data analysis was that I was a researcher particularly interested in patients who experience nightmares frequently. After plotting a t-box of the data, I found that there was a group of patients who experienced nightmares 3 times per week, more than average. I began my analysis by creating a subset of patients who experienced more than 2 nightmares per week.

First, I wanted to know if nightmares correlated with neurosis, one of the most fundamental psychoanalytic complexes. I separated out patients who had a high level of neurosis (over 8) and discovered that there was no correlation between a high level of neurosis and a high nightmare frequency.[1] Overall, nightmare patients suffered from a slightly below average level of neurosis (6.67) than average (7.04).

Next, I looked for a correlation of nightmare patients and the types of symptoms that patients reported experiencing. Out of the five symptoms reported (nausea, headache, back ache, stomachache, dizziness), nightmare patients experience headaches and stomachaches the most. Although they experience these two symptoms at slightly higher rates than the total set of patients, all five symptoms did appear in the nightmare patient data.

My next step was to compare the nightmare patients' primary defense mechanism against the total patients. I discovered that almost all of the nightmare patients exhibit denial as their primary defense mechanism (except one). In psychoanalytic theory, the use of a defense mechanism is the cause of other symptoms, like the five listed above and a high frequency of nightmares, so I shifted the focus of my analysis from patients who experience frequent nightmares to patients who exhibit denial as their primary defense mechanism.

I counted the frequency of results in each column for both the denial patients and the entire set to see if there were any other obvious overlaps, and I discovered that the denial patients form a very specific patient profile. While most patients evenly distributed between the measurements (of gender, types of symptoms, whether they were single, married, divorced, or widowed), the denial patients conformed to the following profile: male, widowed, with an avoidant attachment style, and above average level of neurosis (against other criteria, the denial patients did not stand out). The most obvious commonality amongst the denial patients was that they were overwhelmingly widows. Out of the total 150 patients, 28 are widowed, and 27 of them fall into the denial profile.

If this data were real, I would suggest an area of future research be the bifurcation of the denial patients into having either a level 4 or level 8 of perceived death drive.[2] I did not find an obvious explanation with the given data that explains the divergence (**Table 6** in the appendix displays the two clumps).

4. Ontology Curation

The clear starting point for modeling this data is the Mental Functioning Ontology, with its extension Mental Disease Ontology. Although there is a combined MFO-MD file available on the MFO GitHub, I decided to manually merge the two together to keep my ontology as focused and controlled as possible.

[1] Note, this score of 1-10 is hypothetical, and real neurosis scales do exist, which can be read about here: https://pubmed.ncbi.nlm.nih.gov/6844463/.

[2] 'Death drive' itself is a contentious term in psychoanalytic theory, and it is not accepted completely by all practitioners or theorists.

From mf:mental_process, I added md:pathological_mental_prcess, which goes on to define defense mechanisms (our category of interest). However, this was not an obvious place to stitch these ontologies together.

One of the more complicated philosophical issues I dealt with while crafting the ontology was the difference between a regular mental process and a pathological mental process. According to Barry Smith, **mental diseases** are dispositions that are realized in **pathological mental processes**.[3] But, Smith follows a distinction that was made in the Ontology of General Medical Science (OGMS) that not all *instances* of **pathological processes** are *instances* of **mental diseases**. In the OGMS, **diseases** require **disease courses**, defined as "the totality of all processes through which a given disease instance is realized."[4] Disease courses require multiple pathological processes to repetitively and continuously occur in order for us to say a disease has manifested.

Sigmund Freud's theory on the difference between normal and pathological mental processes changed over the course of his career. Through the mid-point of his career, Freud thought that mental processes became pathological at a certain quantitative level. He often called his theory of mind 'economic' because of its quantitative component–just as economics deals with quantity of goods and services to be exchanged, the mind is constantly regulating the quantity of libido (pleasure) that is channeled by certain neuronal paths.[5] At a certain point late in his career, Freud abandoned the idea that there are any non-pathological normal mental processes at all, normality is an "ideal fiction."[6]

To capture the types of symptoms that patients experienced, I imported the relevant terms from the Symptom Ontology (so). Although it is very underdeveloped, I did not want to reinvent the wheel. I made so:symptom a subclass of mfo:mental_quality, which is a specifically dependent continuant (SDC). A symptom, according to the Symptom Ontology, is "a perceived change in function, sensation, loss, disturbance or appearance reported by a patient indicative of a disease." There are a number of things wrong with this definition, including, but not limited to, the fact that according to this definition, only patients can report symptoms, and that only things which are indicative of diseases can be symptoms. As we established above, instances of pathological processes can occur which are not manifestations of mental diseases. Instances of pathological processes, like repression, are not *bona fide* symptoms.

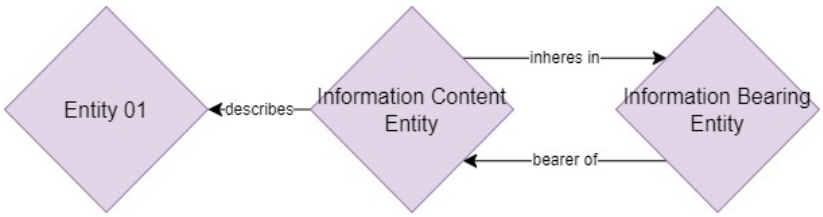
I imported so:symptom as a mental quality because the object of this survey was not whether these symptoms were real in a *measurable* sense but real in a *perspectival* sense; it doesn't matter whether

---

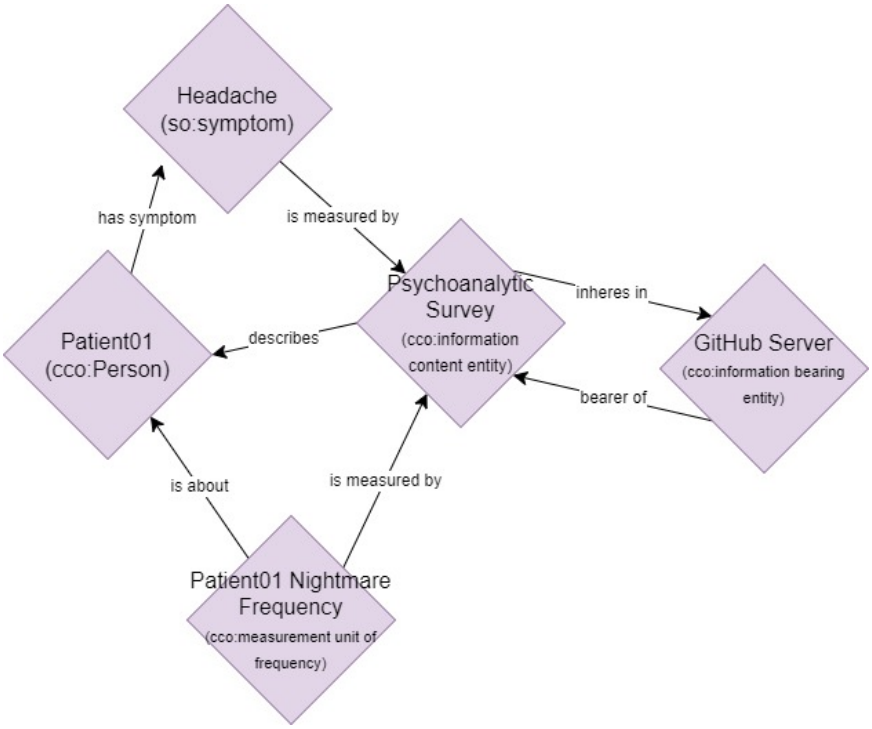
[3] Barry Smith and Werner Ceusters, "Foundations for a realist ontology of mental disease," *Journal of Biomedical Semantics* 1, no. 10 (2010).
[4] Albert Goldfain, Barry Smith, and Lindsay Cowell, "Dispositions and the Infectious Disease Ontology," available at: http://ontology.buffalo.edu/ido/Dispositions_and_IDO.pdf#:~:text=Disease%20Course%3Dde%20f%20The%20totality%20of%20all,processes%20through%20which%20a%20given%20diseaseinstance%20is%20realized.
[5] For more on Freud's neurological background and the quantitative theory of mental stimulus in Austria at the time, see Peter Amacher, *Freud's Neurological Education and Its Influence on Psychoanalytic Theory* (NY: International Universities Press, Inc., 1965).
[6] Freud, *Analysis: Terminable and Interminable*.

someone's stomach has some problem, only that they experience some stomach pain. There are other strategies one may take when modeling symptoms as subjective qualities of experience, but for the purposes of this project, I am taking full advantage of the re-usability of terms. In the future, I may attempt a new strategy.

The next step in modeling the data was attaching measurements to patients and to the survey generally. I followed the strategy that the Common Core Ontologies lays out in the Information Entity Ontology,[7] where an **information content entity** is a bfo:generically_dependent_continuant describes some bfo:entity, and inheres in some **information bearing entity**.[8]



Generally speaking, my design pattern is this:

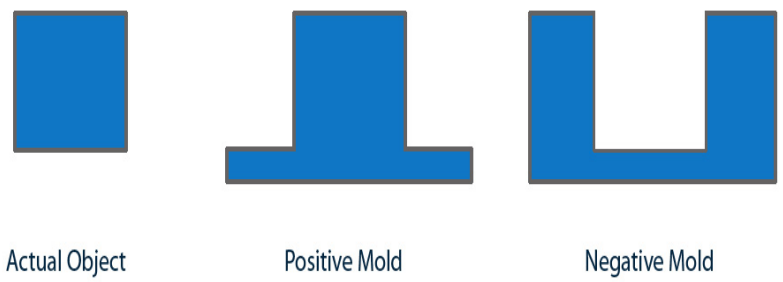


[7] CUBRC, "Modeling Information with the Common Core Ontologies," (2017) available at https://www.nist.gov/system/files/documents/2021/10/14/nist-ai-rfi-cubrc_inc_003.pdf.
[8] For more on use cases, see J. Neil Otte, John Beverley, and Alan Ruttenberg, "Basic Formal Ontology: Case Studies," available at: https://philarchive.org/archive/OTTBBF.

While importing the Information Entity Ontology, I realized that many of the columns in my dataset can be captured as *measurements*, rather than the entities they measure. Take sibling relations, for example: even if I modeled the complex relationship between siblings (let alone someone who doesn't have siblings), I still would need to attach a measurement of the number of siblings to the sibling model; the same goes for marital status. Adhering to the method of ontological realism,[9] this ontology is modeling data about patients, and it is not actually modeling the complex entities that the measurements measure.[10] A benefit of this strategy is that we can capture the survey data and leave the door open for people to model the complex familiar relations to which we can then attach measurements.

5. SHACL Constraints

SHACL is a language designed to validate data, specifically to make sure instance-level data conforms to the axioms of a given ontology. Metaphorically, SHACL shapes take a negative mold of an ontology and ensure that only data that conforms to that shape can enter the ontology (be validated). In this diagram, the actual object is the world, the positive mold is the ontology, and the negative mold is the set of SHACL shapes.



SHACL is often used to quality control instance-level data that enters a database. For a variety of reasons, it is best to validate data with SHACL before ingesting it. Aside from the practical problem that an ontology might be too big to run a reasoner over in a reasonable timeframe, reasoners like Hermit (used by Protege) make the data fit, rather than raising errors. Well-written SHACL constraints can validate very complicated paths, or 'chains' of rdfs.

For this project, I wrote a few shapes with the goal of making sure that only patients with *incomplete information* are not validated. Each patient needs a unique id number, a gender, a level of neurosis, and so on.

[9] Barry Smith and Werner Ceusters, "Ontological realism: A methodology for coordinated evolution of scientific ontologies," *Applied Ontology* 15, no. 5: 139-188.
[10] It is generally not a good practice to model measurements of entities which do not exist in your ontology, but for my purposes here I have simply decided to capture the measurements since my use case is simply to capture the data in a hypothetical database.

APPENDIX

Table 1: Denial Patient Measurements v. Total Patient Measurements

Column	Denial Patients (30 total)	Total Patients (150 total)
Gender	Male: 26 Female: 4	Male 86 Female 64
Marital Status	Widowed: 27 Divorced: 2 Married: 1	Single 61 Married 31 Divorced 30 Widowed 28
Number of Siblings	4: 15 5: 12 3: 2 2: 1	1: 32 2: 31 3: 30 4: 28 0: 17 5: 12
Parental Divorce	No: 28 Yes: 2	No: 120 Yes: 30
Age at Oedipal Trauma	10: 14 11: 13 12: 2 13: 1	11: 26 16: 17 12: 16 18: 15 14: 15 15: 15 13: 15 10: 14 19: 14 17: 3
Dream Frequency	4: 29 2: 1	1: 59 2: 31 3: 30 4: 30
Nightmare Frequency	3: 29 1: 1	0: 59 1: 31 2: 30 3: 30
Attachment Style	Avoidant: 29	Secure: 61

	Anxious: 1	Avoidant: 58 Anxious: 31
Defense Mechanism	Denial: 30	Displacement: 30 Projection: 30 Denial: 30 Intellectualization: 30 Reaction Formation: 29 Repression: 1
Level of Neurosis	7: 11 6: 10 5: 4 8: 3 9: 2	7: 33 8: 32 5: 31 9: 30 6: 24
Number of Slips Per Analytic Session	2: 11 3: 10 5: 4 1: 3 4: 2	2: 33 1: 32 5: 31 4: 30 3: 24
Level of Perceived Death Drive	4: 10 8: 10 1: 4 6: 2 3: 2 2: 2	4: 32 1: 31 2: 30 8: 24 6: 23 3: 9 5: 1
Types of Physical Symptoms	Headache: 12 Stomachache: 10 Nausea: 4 Back pain: 2 Dizziness: 2	Headache: 41 Nausea: 31 Dizziness: 30 Back pain: 24 Stomachache: 24
Number of Symptoms	3: 11 1: 10 4: 4 2: 3 5: 2	3: 33 2: 32 4: 31 5: 30 1: 24

Table 2: Boxplot of Nightmare Frequency of All Patients

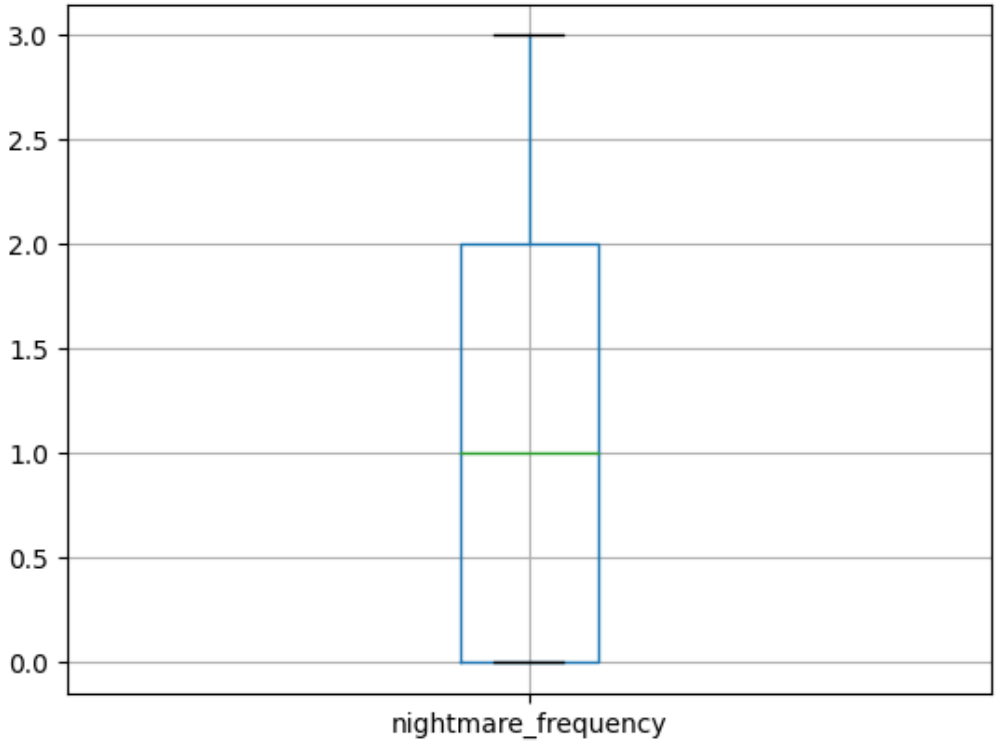


Table 3: Strip Plot of Age and Level of Neurosis for All Patients (with defense mechanism highlighted)

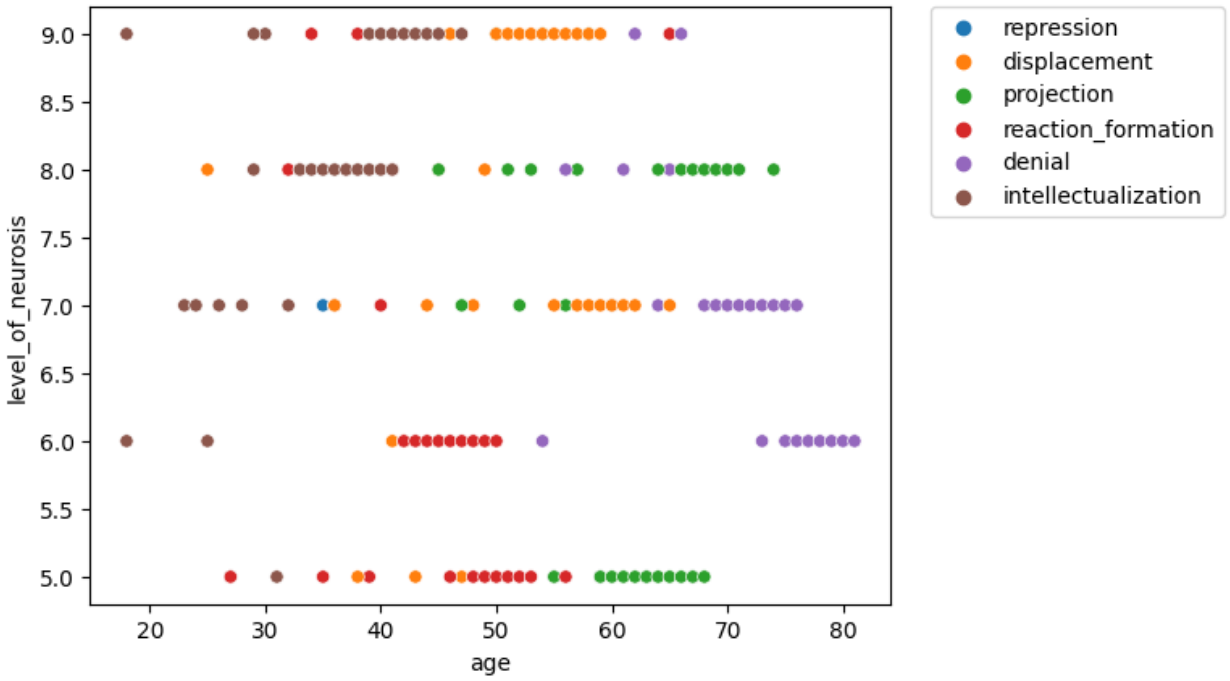


Table 4: Scatterplot of Age and Level of Neurosis for Patients Who Exhibit Denial (highlighted by gender)

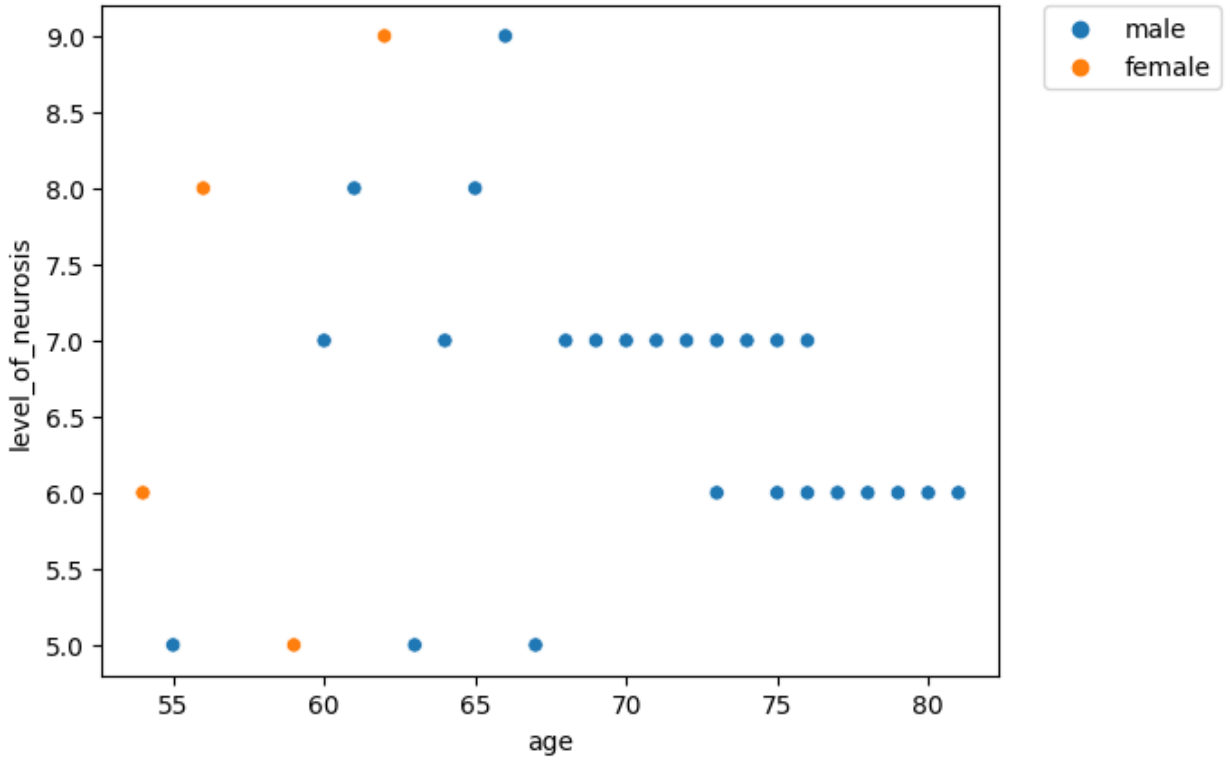


Table 5: Strip Plot of Number of Freudian Slips Per Therapy Session and Nightmare Frequency (highlighted by defense mechanism)

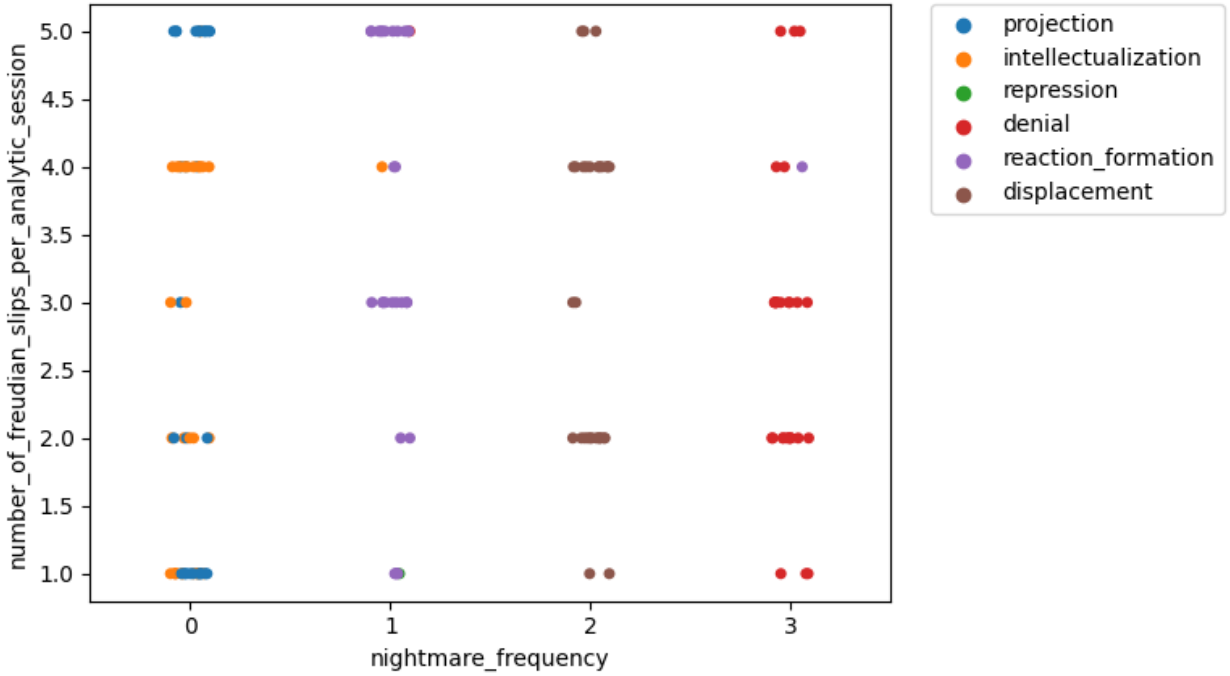


Table 6: Violin Plot of Level of Perceived Death Drive