Scientometric Analysis
of Psychology and Related Fields

Himadri Mandal (BS2327)
B.Stat II Year, Indian Statistical Institute, Kolkata

Under the guidance of Dr. Garga Chatterjee
Assistant Professor, Psychology Research Unit (PRU)
Indian Statistical Institute, Kolkata

May 2025

Abstract

We present a scalable, data-driven bibliometric analysis of psychology and related fields over 2000–2025, using author-assigned keywords from the top 20 h5-index journals in SCOPUS. By tracking keyword frequency (popularity) and citation counts (impact), and introducing an impact-density metric, we uncover both enduring themes—such as Personality, Depression, and Anxiety—and high-impact yet less frequent topics like Meta-analysis and Positive Psychology. We further quantify divergences between popularity and impact rankings to highlight emerging research areas, and demonstrate a replicable workflow for mapping disciplinary evolution through simple yet powerful computational methods.

1 Introduction

The field of psychology has long been marked by a dynamic interplay between its subfields—ranging from basic cognitive research to applied clinical work—and by its interactions with neighboring disciplines such as neuroscience, sociology, and computer science (Marshall, 2009; Morf, 2018; Schwartz et al., 2016). Over the past two decades, this interdisciplinarity has been further shaped by technological advances and computational methods, which have introduced new ways to study the evolution of psychological inquiry. Particularly, the proliferation of digital publication archives and citation indexing has opened new avenues for understanding the intellectual structure and topical trajectories within the field.

Despite the rich diversity of psychological research, it remains unclear how specific topics gain prominence, fade, or re-emerge over time. What kinds of ideas dominate the discourse in high-impact journals versus lesser-known publications? How do emerging areas like machine learning or neuroinformatics influence traditional psychological themes? These questions are essential to understanding how psychology evolves as both a science and a discipline. Yet, due to the sheer scale of scholarly output, these questions are difficult to answer through manual review alone.

This project adopts a computational bibliometric approach to address these gaps. By analyzing a large-scale dataset of psychological research metadata, obtained through the **SCOPUS** database, we aim to trace the frequency and semantic associations of **Author Keywords** over time. This methodology distinctly differs from previous work [1], which uses STMs on a corpus of abstracts. We believe this crucial change does justice to how Authors self-identify the topics associated to their papers. However, this reduces our corpus allowing meaningful analysis only after the 2000s, as we demonstrate in Figure 1.

This project allows us to approximate the shifting landscape of psychological research from a macro-level perspective. Drawing inspiration from recent applications of natural language processing in science studies, we use keyword ranking, citation-weighted keyword ranking, temporal trend analysis. One can further look into applying clustering based techniques using **BERT** embeddings (SciBERT, PsychBERT, etc.), but rudimentary analysis suggest the cosine similarity doesn't do justice to the concepts.

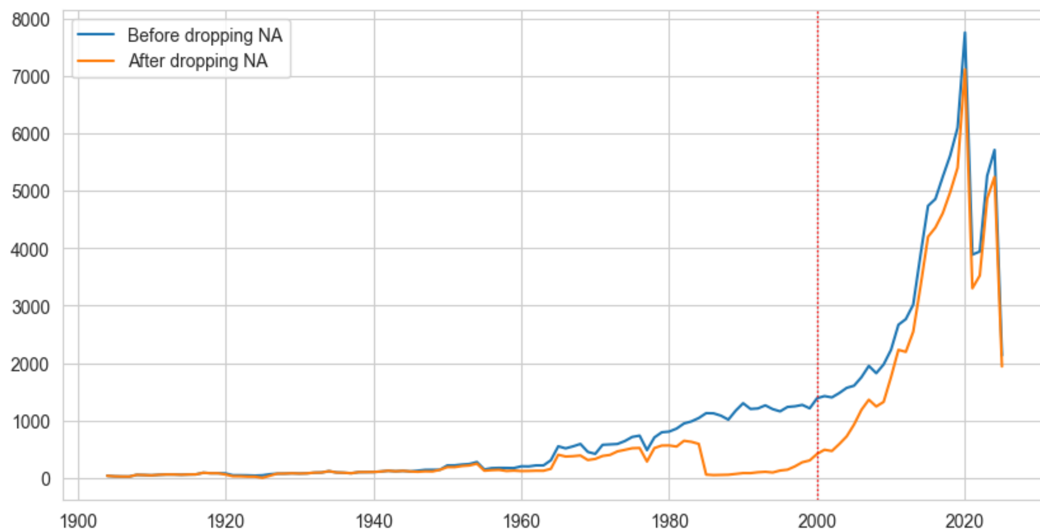


Figure 1: Corpus with and w/o **Author Keywords**

In this paper, we pursue some fundamentally important research questions:

1. What keywords have been the most **frequent/popular** throughout time?
2. What keywords have been the most **impactful** throughout time?
3. What keywords have the highest **impact density**?
4. How far away **popularity rankings** and **impact rankings** are?

By leveraging computational tools, we aim to contribute a replicable and scalable methodology to the study of disciplinary evolution in psychology. Our hope is that such work can complement qualitative and theoretical insights with data-driven perspectives on the transformation of psychological science.

2 Methodology

We use computational methods to study the metadata of thousands of papers in the Psychological literature. We use opensource libraries like **Matplotlib, Pandas, Numpy, etc.**, and would like to extend our gratitude to their dedication to keeping software free and accessible for all.

2.1 Database

We explored various options for choosing the metadata database. **Semantic Scholar**, **OpenAlex**, **Web of Science** were our top choices, however, we couldn't proceed with any of them, because:

1. **Semantic Scholar**: no access to API.
2. **OpenAlex**: cannot access Author Keywords of closed-source articles.

3. **Web Of Science**: institute didn't have access

We ended up choosing **SCOPUS** because of its beautiful UI (and us having access to it.) However, there were issues with SCOPUS as well. SCOPUS can't be used to export more than 20000 entries from one query, which meant creating many different non-contiguous queries, if we were to study all of Psychological Literature.

Instead, we chose to stick to the top 20 Journals ordered by their **h5-indices** (March 2025) (see Figure 2).

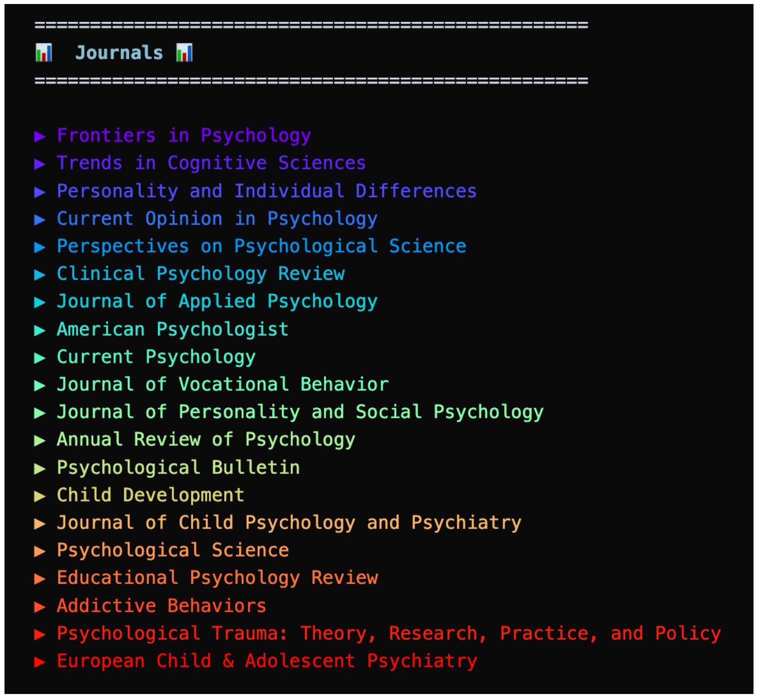


Figure 2: Journals Picked

Then, we exported metadata of papers on these journals throughout time. Of course, this leads to a post hoc analysis of journals that actually manage to be in the Top 20. However, the methodology is generalizable, and one could add a host of journals into the oven, and repeat the process.

2.2 Data Extraction and Cleanup

We extract the Author Keywords using Pandas. Ofcourse, as demonstrated in Figure 1, the number of entries one could work with reduces as many metadata entries do not have the Author Keyword variable filled.

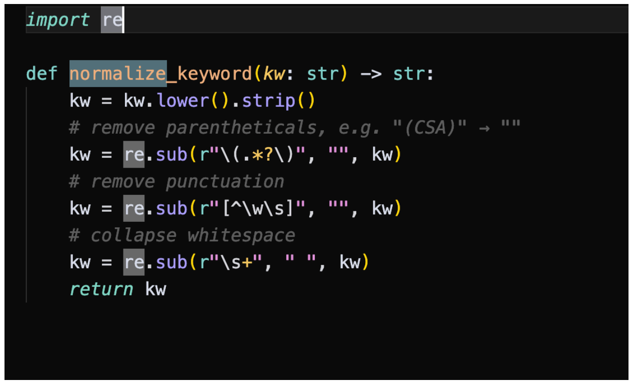
We '**normalize**' the keywords by removing parentheticals, removing punctuations, and collapsing whitespace. Then, we convert every keyword to title case. This step is important to bucket same keywords together. (See Figure 3.)

Size of KeywordsDF before dropping Author Keywords: **124512**
Size of KeywordsDF after dropping Author Keywords: **87352**

We perform the analysis on the timeline **2000-2025**.

3 Analysis

We study two important trends in the Keywords: **Popularity** and **Impact**.



```python
import re

def normalize_keyword(kw: str) -> str:
    kw = kw.lower().strip()
    # remove parentheticals, e.g. "(CSA)" → ""
    kw = re.sub(r"\(.*?\)", "", kw)
    # remove punctuation
    kw = re.sub(r"[^\w\s]", "", kw)
    # collapse whitespace
    kw = re.sub(r"\s+", " ", kw)
    return kw
```

Figure 3: Normalize Keywords

3.1 Popularity and Impact Metrics

We implement **Popularity** trends by counting the number of times a Keyword K has appeared in year Y,

$$N(Y, K) = \#\text{count}(K \text{ in year } Y)$$

and then plot $N(Y, K)$ across years. How we implement **Impact** trends is more interesting. For keyword K, we count the total number of *citations* papers with keyword K has received in year Y, say,

$$c(K, Y) = \#\text{citations(papers with } K \text{ in year } Y).$$

Then, we plot $c(K, Y)$ through time.

3.2 Most Popular Keywords

We find the top 40 most popular keywords (Figure 4).

3.3 Most Impactful Keywords

We find the top 40 most impactful keywords (Figure 5).

3.4 Trends of Top 10 Most Popular Keywords

(Figure 6).

3.5 Trends of Top 10 Most Impactful Keywords

(Figure 7).

3.6 Popularity vs Impact Rankings

(Table of top impactful keywords ranked by popularity.)

Keywords	Impact Ranking	Popularity Rank	Occurrences
Metaanalysis	#1	#5	1,486
Depression	#2	#2	2,498
Personality	#3	#1	2,508
Anxiety	#4	#3	1,713
Motivation	#5	#14	1,037
Emotion	#6	#11	1,100
Adolescence	#7	#6	1,387
Mental Health	#8	#7	1,290
Adolescents	#9	#4	1,635
Stress	#10	#15	949

Table 1: Popularity rankings of Impactful Keywords

3.7 Deviation Between Rankings

(Top 20 keywords with highest deviation.)

We wanted to find keywords that have high impact but rank low in popularity rankings. This meant studying

$$\max_{K \in \mathcal{S}} \left(\text{rank}_{\text{impact}}(K) - \text{rank}_{\text{popularity}}(K) \right) .$$

We fixed $\mathcal{S}$ to be the top 100 most impactful keywords. Then, we find the top 25 keywords which have the highest deviations.

Keyword	Deviation	Impact Rank	Popularity Rank
Review	102.0	44	146
Job Performance	100.0	67	167
Power	75.0	100	175
Positive Psychology	71.0	90	161
Epidemiology	69.0	75	144
Academic Achievement	66.0	71	137
Happiness	54.0	60	114
Risk Factors	53.0	95	148
Leadership	43.0	70	113
Psychotherapy	42.0	82	124
Performance	40.0	39	79
Burnout	38.0	85	123
Intervention	36.0	27	63
Health	36.0	35	71
Emotions	33.0	36	69
Prejudice	32.0	83	115
Subjective Wellbeing	31.0	52	83
Morality	31.0	76	107
Prosocial Behavior	26.0	78	104
Cognitive Control	26.0	91	117

Table 2: Top 20 keywords with highest deviation between impact and popularity rankings

3.8 Impact Density

(See Figure 8). We define

$$\text{Impact\_Density}(K) = \frac{\sum_Y c(K, Y)}{\sum_Y N(K, Y)}$$

Now, we sort the keywords for impact density.

Then, we find the keywords K with the highest impact densities (conditional on K having been quoted in atleast 25 papers).

3.9 Repeated Common Keywords

(See Figure 9). We define *common keywords* as keywords which appear in 75% of the years (2000-2025) **AND** are in top 10 for at least 50% of years.

3.10 Top Keywords Excluding Common

(See Figure 10). We exclude the *common keywords* from the top 10 popularity rankings to understand which keywords became specifically popular in certain years. We create a heatmap to demonstrate this.

4 Conclusion

In this study, we have presented a scalable, data-driven bibliometric analysis of the evolution of psychology and its related fields over the period 2000–2025. By focusing on author-assigned keywords in the metadata of the top 20 h5-index journals from SCOPUS, we traced both the raw popularity (frequency) and scholarly impact (citation counts) of research topics over time. Our findings reveal that while traditional keywords such as *Personality*, *Depression*, and *Anxiety* remain dominant in sheer volume, other terms like *Meta-analysis* and *Positive Psychology* exhibit disproportionately high impact relative to their frequency. The impact-density metric further highlights topics—e.g., *Pharmacotherapy* and *Measurement Invariance*—that, although less common, generate significant scholarly attention per paper.

A key insight is the divergence between popularity and impact rankings: several keywords with moderate use (e.g., *Burnout*, *Intervention*, *Well-being*) show high impact deviations, suggesting emergent or rapidly maturing areas that punch above their weight in citations.

Despite these contributions, our approach has limitations. The top-20 journal selection may bias toward established, high-impact outlets. Future work could incorporate semantic embeddings (e.g., SciBERT, PsychBERT) to capture deeper conceptual relationships, extend the journal set for broader coverage, and delve into interesting questions like which keywords appear together etc.

Overall, this paper demonstrates the power of accessible bibliometric methods to map the shifting landscape of psychological science. By providing an open, replicable workflow and clear visualizations of keyword trends, we offer a foundation for further qualitative and theoretical studies on how ideas emerge, spread, and transform within the discipline.

Acknowledgements

I would like to express my sincere gratitude to **Dr. Garga Chatterjee** and **Dr. Kuntal Ghosh** for their invaluable methodological guidance and birds-eye view of the project. I am also thankful to the open-source community for developing tools such as **Python**, **Pandas**, and **Matplotlib**, which greatly facilitated data analysis and visualization. My thanks go to ISI Kolkata for providing the access to the **SCOPUS** database, and academic freedom necessary to pursue this project. Lastly, I am deeply grateful to my friends and peers for their unwavering support and encouragement—this study would not have been possible without them.

References

[1] Oliver Wieczorek, Saïd Unger, Jan Riebling, Lukas Erhard, Christian Koß, and Raphael Heiberger. Mapping the field of psychology: Trends in research topics 1995–2015. *Scientometrics*, 126:9699–9731, June 2021.

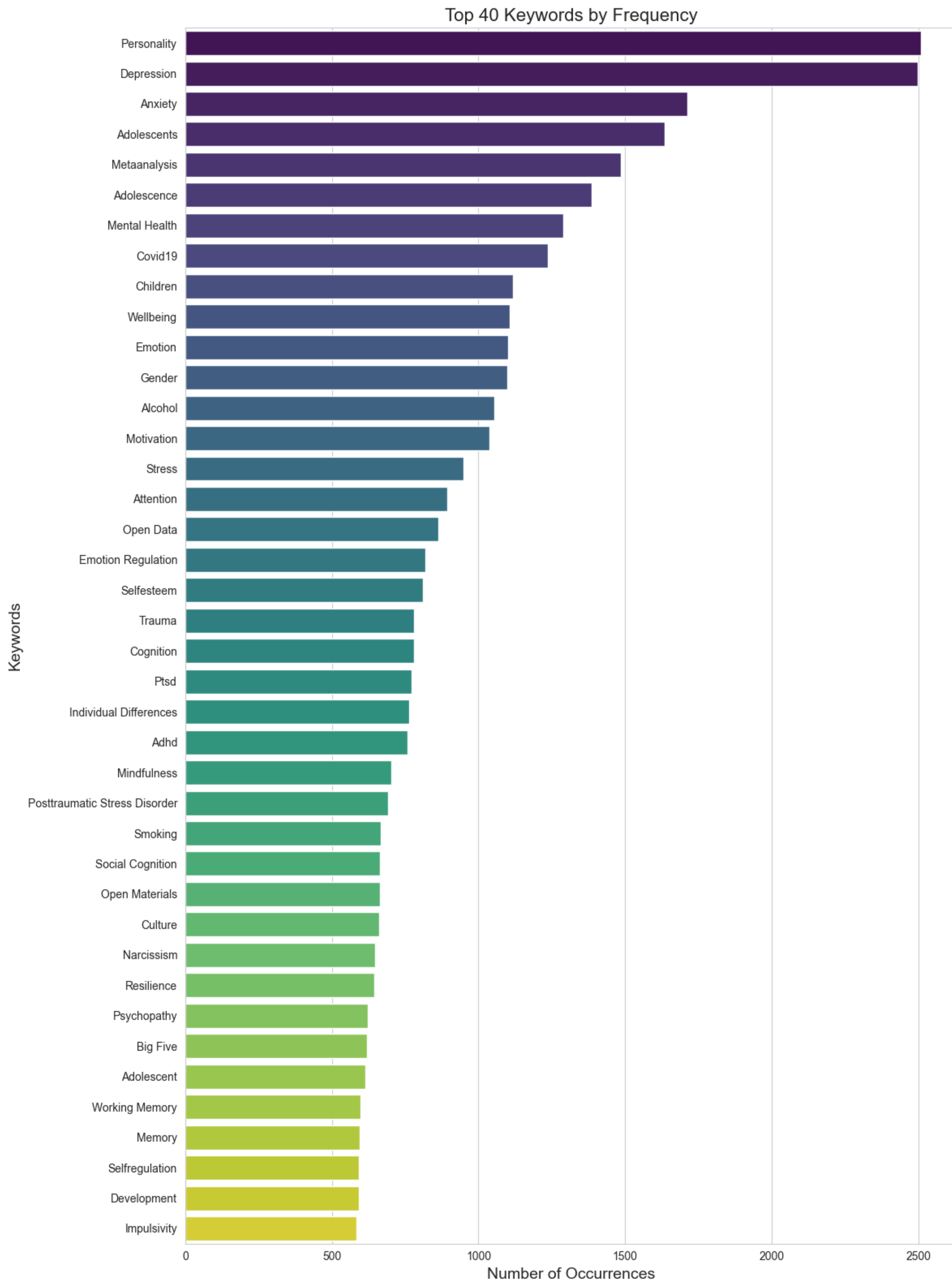


Figure 4: 40 Most Popular Keywords

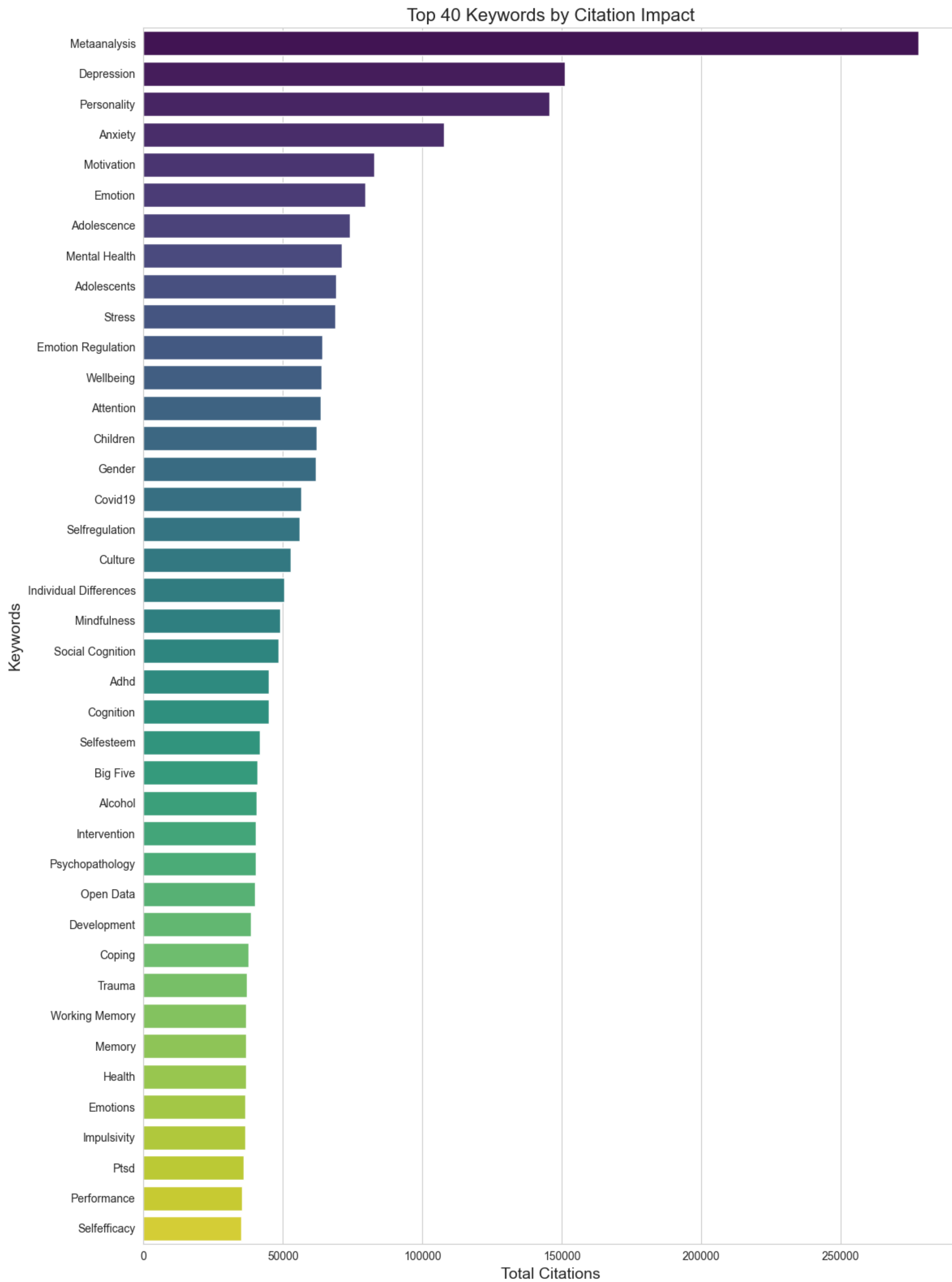


Figure 5: Most Impactful Keywords

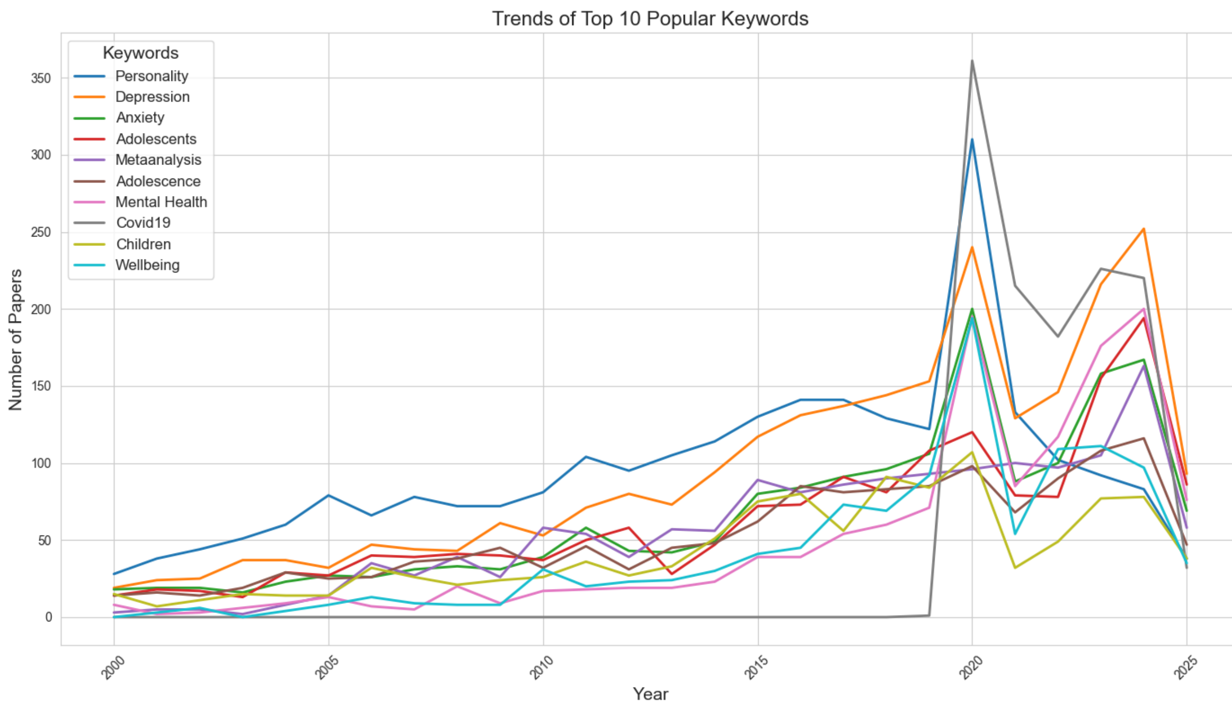


Figure 6: Trends of Most Popular Keywords

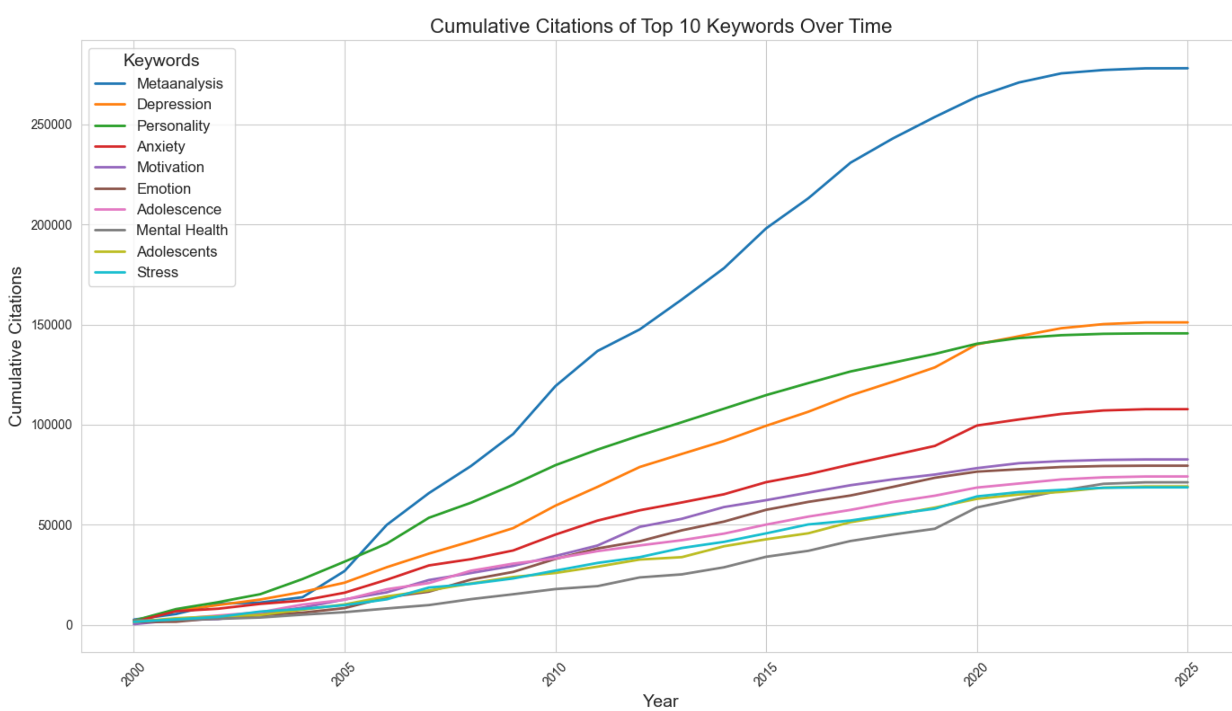


Figure 7: Trends of Most Impactful Keywords

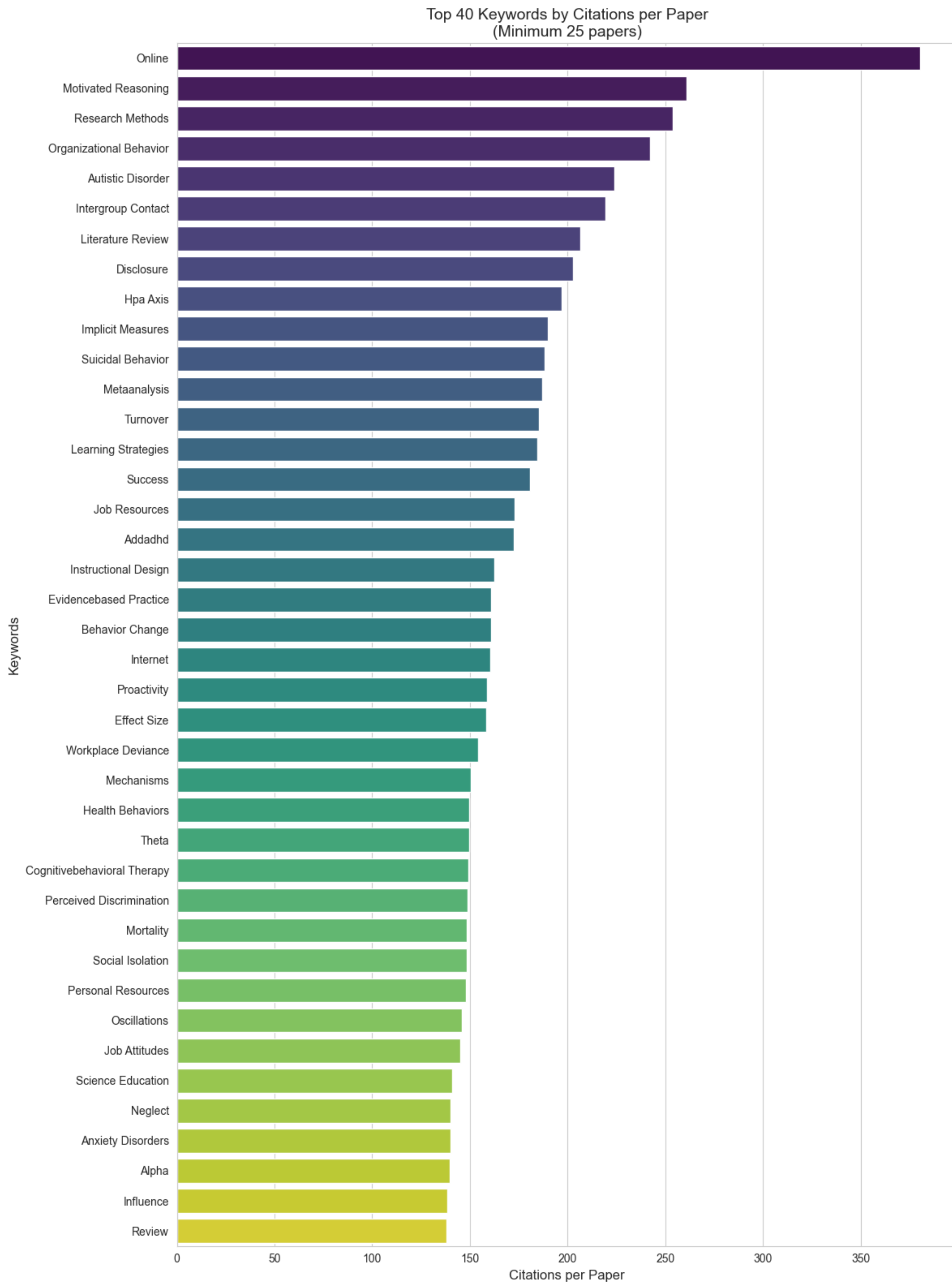


Figure 8: Impact Density Rankings

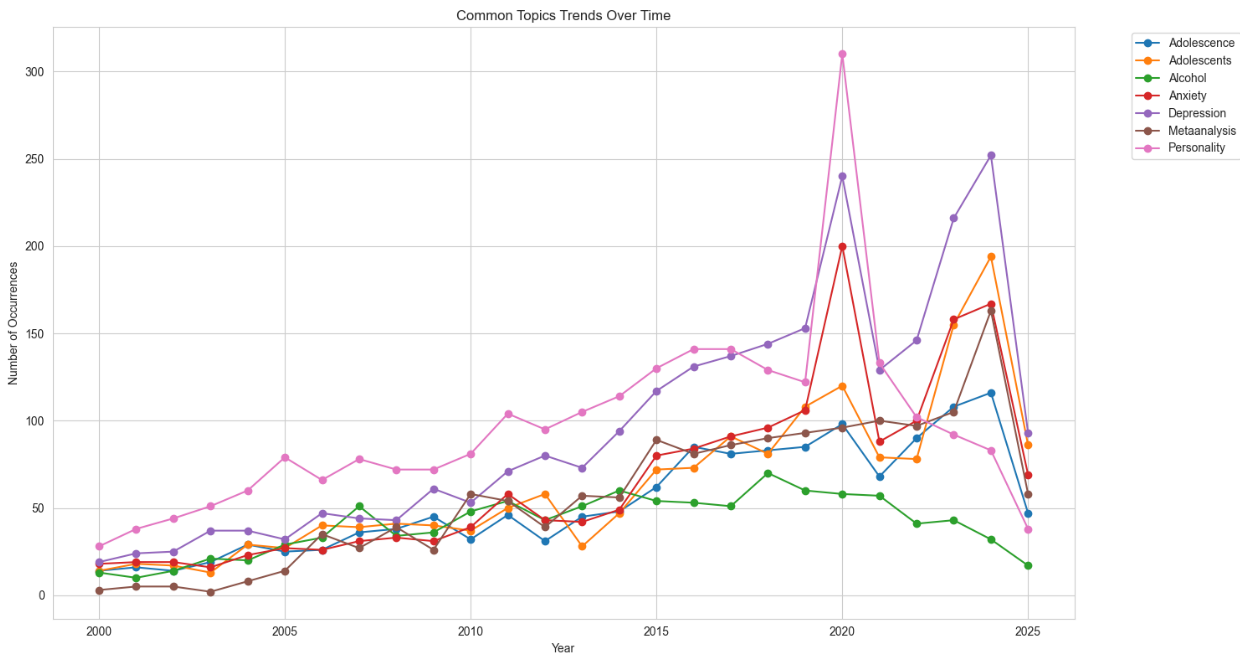


Figure 9: Trends of Repeated Common Keywords

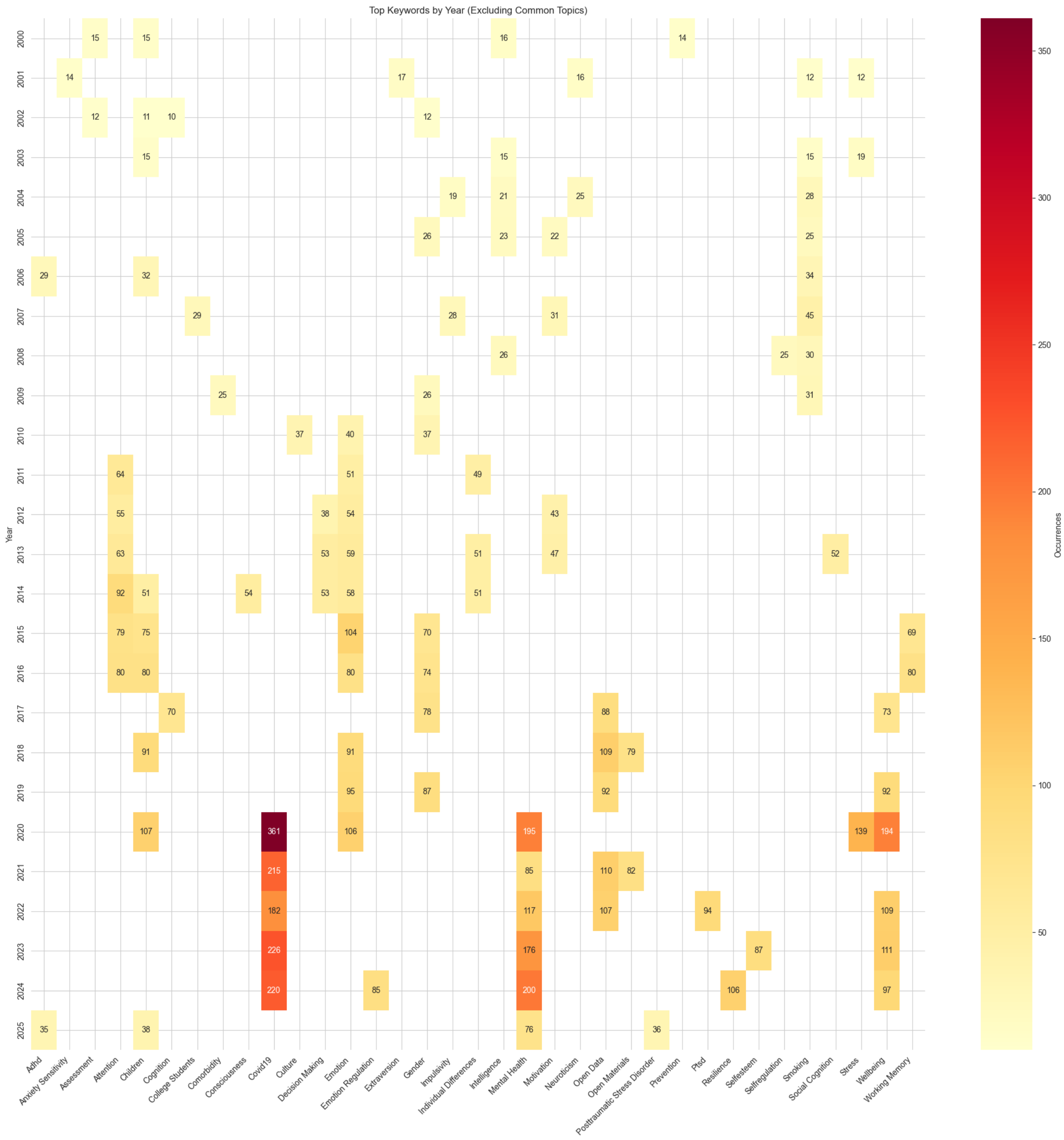


Figure 10: Top Keywords Excluding Common