

Gender-responsive budgeting in Australia

- Text Analysis of Women's Budget Statement 1984-2025 & International Indices -

Abstract

The study investigates the longitudinal evolution and policy impact of Australian Women's Budget Statement (WBS) from 1984 to 2025, the world's longest-running example of gender-responsive budgeting (GRB). Using computational text analysis (topic modeling and word frequency) and regression modeling, the research explores how policy priorities have changed across different political administrations and examines correlations between WBS content and gender equity indices including the GII and HDI. The findings reveal consistent thematic emphasis on social services. However, regression analysis shows limited and selective alignment between policy discourse and national gender development indicators. The study contributes to a deeper understanding of how GRB has been shaped and institutionalized in the Australian federal context.

Xi Tang Lin (Coordinator)

XiTang.Lin@UTDallas.edu

Zhangfan Huang

Zhangfan.Huang@UTDallas.edu

Solbee Kim

Solbee.Kim@UTDallas.edu

Xinyu Yan

Xinyu.Yan@UTDallas.edu

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Contents

I. Introduction

II. Literature Review

1. History of GRB

2. Australian Gender-responsive Budgeting

3. Framework

III. Research Design

1. Research Questions

2. Method

3. Data

IV. Result

V. Conclusion

References

Appendix

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

I. Introduction

The incorporation of social equity principles into public budgeting and government finance is essential for addressing and modifying unequal outcomes. Public budgeting provides a key mechanism for advancing social equity and delivering public values as a field of public administration (Bracci et al. 2021; McDonald and McCandless 2022; McDonald III et al. 2024; McDonald III and McCandless, 2025). Among various approaches, gender-responsive budgeting (GRB) is a widely applied tool for analyzing social equity in the budgeting process and its functions, and many countries have implemented GRB in different forms (Galizzi et al. 2021; McDonald III et al. 2024; Rubin and Bartle 2023). While some developed countries have adapted GRB to their own contexts, implementation in several developing countries has been encouraged by multiple international organizations, including the United Nations (UN), International Monetary Fund (IMF), Organization for Economic Cooperation and Development (OECD), and the World Bank (Chakraborty 2016; Downes et al. 2017; OECD 2023; Ostry et al. 2018).

Among the numerous applications, the most distinctive case is Australia's GRB, which began in 1984, a decade before the 1995 Beijing Conference, widely recognized as a milestone in women's rights and as establishing critical areas of concern for governments (Rubin and Bartle 2005; 2023; UN WOMEN 1995a; 1995b). As a pioneer, the Australian federal government sought to address widespread gender disparities in the country by integrating gender analysis into various policy processes, including the budgeting process, a strategy known as "gender mainstreaming" (Rubin and Bartle 2023; Sawer 2002; Walby 2005b; 2005a). In 1984, Australia began publishing an annual Women's Budget Program (later called the Women's Budget Statement, WBS) (Office of the Status of Women and Department of the Prime Minister and

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Cabinet 1984). However, the program was quietly abandoned in 2013, and in 2021 Australia reintroduced new GRB initiatives and published the WBS (Australian Government et al. 2021; Baird et al. 2014; Sharp and Broomhill 2014).

This trajectory raises critical questions about the political, institutional, and social factors that shaped Australia's shifting commitment to GRB. This time frame covers the full historical development of Australian GRB, so the content of these documents may reflect changes in political leadership (ruling parties), historical context, gender discourse, and policy trends. Not many countries have such longitudinal records, so it seems that this is a valuable opportunity to study long-term changes.

II. Literature Review

1. History of Gender-responsive Budgeting

Gender-responsive budgeting (GRB) originated with Australia's Women's Budget Program in 1984 and spread to dozens of countries, influenced by global commitments such as CEDAW (1979), the Beijing Platform for Action (1995), which advocated for integrating a gender perspective into budgetary stages to advance gender equality (Rubin and Bartle 2023b; Sharp and Broomhill, 2002). GRB initiatives have moved from a contested, feminism idea to a mainstream policy tool supported by international organizations such as the United Nations (UN), International Monetary Fund (IMF), Organization for Economic Cooperation and Development (OECD), World Bank, and European Commission (Chakraborty 2016; Downes et al. 2017; 2017; Nolte et al. 2021; O'Hagan and Klatzer, 2018; Ostry et al. 2018; Polzer et al. 2023). The essential concept is the integration of gender perspectives into all budgetary

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

processes to reduce gender gaps, increase accountability, and reframe economic policy as inclusive of care and social priorities, not just growth and efficiency (Nolte et al. 2021; OECD 2023; O'Hagan and Klatzer 2018).

European cases are often led by European Union (EU) with Austria, Belgium, Iceland, Sweden, The Netherlands, Norway, Finland, Germany, Turkey, Scotland, and the UK (Bojičić-Dželić and Hozić 2020; Campbell and Gillespie 2016; Elomäki and Ylöstalo 2021; Gunluk-Senesen 2021; Lodovici 2020; O'Hagan and Klatzer 2018; Polzer and Seiwald 2021). They have demonstrated various notable outcomes at different levels of governments. Asia-Pacific GRB varies widely, with pioneering and legally mandated efforts in the Philippines and South Korea, strong civil society roles in India, and growing government involvement in countries like Malaysia, Indonesia, and Mongolia (Budlender et al. 2011; Budlender and Hewitt 2002). Africa has seen dynamic development of GRB, especially in countries like South Africa, Tanzania, and Uganda, where strong coalitions between civil society and government have driven sustained initiatives over decades (Budlender et al. 2011; Judd 2002). The African experience highlights the importance of coalition-building such as among NGOs, activists, and politicians, iterative capacity-building, and action-oriented research, often focused on health, education, and agriculture sectors. Latin American countries emphasize participatory budgeting, municipal-level innovations, and strong feminist advocacy including Mexico, Ecuador, Peru, Bolivia, and Brazil (Budlender et al. 2011; Budlender and Hewitt 2002; Martínez Guzmán 2024). On the other hand, Middle East and North Africa (MENA) has limited but rising GRB, with Morocco and Egypt leading on national strategies, donor UN initiatives, and pilot integration into key ministries (Nolte et al. 2021; Stotsky 2016).

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Previous studies have demonstrated the importance of contextual diversity, evolving based on international aid and regional differences. Unlike the cases above, Australia has the unique characteristic of being both a country and a continent even though it is classified within Asia-Pacific regions (Coleman 2016). Therefore, Australia is expected to exhibit region-specific characteristics, a long-standing history of GRB, and social changes shaped by its relatively isolated context.

2. Australian Gender-responsive Budgeting

Australia is one of the British Commonwealth of Nations that the government and parliament has closed relationships due to parliamentary cabinet system. Therefore, the leadership of Prime Ministers and its ruling party have strong effects on public policies. Australia has two major political parties: Australian Labor Party which has social democracy and Liberal Party of Australia with National Party of Australia (alias Coalition Party) which have rooted in conservatism, liberalism, and agrarianism (McAllister 1991).

Australian feminists who were called Liberationist, notably activated to improve equal pay and to protest anti-war due to Vietnam War in the 1960s (Grieve and Burns 1994). Their activities in the 1980s centered on the institutionalization of feminism and technical mechanisms in feminist bureaucrats (alias Femocrats) to mainstream gender into federal policies (Grieve and Burns 1994; Sawyer 2002; Sharp and Broomhill 2002; 2014). The then Prime Minister Hawke, who was Labor Party leader, sympathized with Femocrats, so Hawke government introduced the first Women's Budget Program in 1984 (Office of the Status of Women and Department of the Prime Minister and Cabinet 1984). This progressive approach had been maintained for three decades until 2014 through the two different ruling parties.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

While political parties attempted to reduce gender disparities applying GRB, the largest public administration journal in Australia, Australian Journal of Public Administration (AJPA), was less interested in gender. Gender received very limited attention in AJPA, only 0.8% of articles from 1970–1995 and about 1.2% from 1996–2014 addressed gender issues, with a total of 19 articles across 45 years, often clustered in a few special issues rather than consistent coverage (Althaus 2008; 2015). Over time, research shows the initial momentum was challenged by shifts toward neoliberalism, public sector reform, decentralization, and a growing focus on “gender mainstreaming” within public services less linked to budgets directly (Colley and Williamson 2019; Sawyer 2002; Walby 2005a; 2005b). Despite formal GRB legacies in Australia, contemporary journals note declining analytical robustness, limited civil society engagement, and an overemphasis on targeted programs over mainstream fiscal policy (Colley and Williamson 2019; Sawyer 2002; Sharp and Broomhill 2002).

Liberal/National Party won the September 2013 federal election, and new Prime Minister Abbott and his government quietly abandoned Women’s Budget Statements (Baird et al. 2014). Australian GRB returned in 2021 while the Liberal government and then Prime Minister Morrison faced multiple pressures from politics, the press, and society due to sexual scandals (Kim 2025). However, the Labor party grabbed power in 2022 federal election, and the current Prime Minister Albanese and his government has maintained Women’s Budget Statement the present (Australia. Department of Finance et al. 2022; 2023; 2024; Australian Government et al. 2025). Therefore, Australian GRB was a model for other “inside government” approaches globally, underscoring the importance of policy machinery location and political support (Sharp and Broomhill 2002; 2014). Nonetheless, impact on actual resource reallocation and deep policy

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

change was limited and most changes were in accountability, transparency, and data, not in substantive spending priorities (Colley and Williamson 2019; Sharp and Broomhill 2014).

Australian GRB literature reveals an early, innovative experiment that advanced institutional learning internationally, but also demonstrates the fragility of gender equity advances in the face of shifting political priorities and administrative reforms. Contemporary Australian scholarship continues to critique the limits of mainstreaming and calls for renewed, integrated approaches to prevent further erosion of substantive gender equity outcomes (Colley and Williamson 2019; Rubin and Bartle 2005; 2023b; Sawer 2002; Sharp and Broomhill 2002; 2014).

GRB has reflected effort of persistent gender inequality embedded in government budgeting systems. GRB helps governments allocate resources in ways that promote gender equity and social inclusion, making budgeting a tool for advancing both democratic accountability and social justice. Despite its relevance, GRB remains unevenly understood and implemented across countries, and continued academic research is essential for evaluating its effectiveness, improving methodologies, and informing policy innovation.

The Women's Budget Statements represent an unparalleled source of longitudinal data that reflects shifting policy priorities, ideological influences, and institutional engagement with gender equity. The annual documents have received limited systematic academic attention even though Australia has historically the world longest GRB initiatives. Studying the longitudinal changes in Australian Women's Budget Statements from 1984 to 2025 offers a valuable case for understanding the long-term evolution, institutionalization, and political dynamics of GRB. As one of the first countries to adopt GRB in 1984, Australia provides a critical foundation for

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

analyzing trends and turning points of gender-focused programs. This historical and document-based analysis will serve as foundational work that can guide and inspire other governments to design or reform their own GRB systems, tailored to their unique political and institutional contexts.

III. Methods

1. Research Questions

Our research questions aim to trace the longitudinal development of Australian Women's Budget Statements (WBS) over more than four decades. Since the WBS serves as a key policy document reflecting government priorities for advancing gender equity, understanding how its themes, focus areas, and targeted programs have changed over time provides critical insight into the evolution of gender-responsive budgeting (GRB) in Australia. Analyzing this timeline reveals not only shifts in gender-based policy trends but also broader political, economic, and social transformations that have shaped into WBS.

Research Question 1. How have policy targets, program areas, and thematic priorities in the Women's Budget Statements evolved between 1984 and 2025?

Sub-research questions 1.1 How have the two major political parties (Australian Labor party & Liberal party of Australia) designed the Women's Budget Statements differently?

In addition, we seek to assess whether and how shifts in WBS content correlate with measurable changes in gender equity outcomes. By linking thematic emphases in budget statements to international indicators such as the Gender Inequality Index (GII), Human

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Development Index (HDI), and Social Institutions and Gender Index (SIGI), the study aims to bridge discourse and outcome. This relationship is vital for evaluating the real-world impact of gender budgeting beyond rhetorical commitments, helping to validate or critique the effectiveness of the WBS as a policy tool.

Research Question 2. What is the relationship between changes in the Women's Budget Statements' topic focus and changes in the GII, HDI, and SIGI in Australia?

Sub-research questions 2.1 What kind of changes have the most effective impact on index changes?

The last integrative question broadens the scope of analysis to consider how national-level discourse and policy alignment in WBS corresponds with macro-level trends in gender equality. By examining convergence or divergence between WBS content and global indicators, the research contributes to understanding the coherence between policy rhetoric, domestic implementation, and international commitments. It also positions Australia within the global GRB landscape, offering comparative insights for global policy learning and norm diffusion.

Research Question 3. To what extent do shifts in the content and focus of the Women's Budget Statements correspond with broader trends in gender equity, as measured by international indices (e.g., GII & HDI)?

2. Method

The study employs a multi methods approach combining computational text analysis and regression modeling to examine the evolution and impact of Australia's Women's Budget Statements (WBS) from 1984 to 2025.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

2.1 Textual analysis

The study will employ two computational text analysis techniques, including topic modeling and word frequency analysis to examine the evolution of Australia's WBS between 1984 and 2025. Together, these approaches will capture both the latent thematic structures and the explicit linguistic patterns that characterize how GRB has been conceptualized and framed under different political, institutional, and social contexts (Grimmer and Stewart 2013).

2.1.1 Topic modeling

A topic modeling approach using Latent Dirichlet Allocation (LDA) will be used to uncover underlying themes within the corpus of WBS documents. LDA assumes that each document is a mixture of topics, and each topic is represented by a distribution of words (Blei, Ng and Jordan 2003). This unsupervised method enables the identification of dominant policy areas and issue clusters without imposing predefined categories. The number of topics will be determined through coherence scores and iterative model testing to balance interpretability and statistical fit. The resulting topics are expected to align with major sectors relevant to gender policy, such as health, labor market participation, education, social protection, and family and caregiving. Temporal trend analysis will then be conducted by aggregating topic proportions by year to identify changes in emphasis across time. For instance, an increase in the salience of "economic participation" topics under certain administrations may indicate a policy shift toward market-oriented gender framing, while higher prominence of "social protection" themes may signal redistributive or welfare-oriented approaches.

2.1.2 Word frequency analysis

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

To complement the topic modeling, word frequency analysis will provide a direct measure of the visibility and prioritization of gender-related discourse in each year's WBS. Frequency counts of key gender and policy terms (e.g., equity, mainstreaming, empowerment, childcare, leadership) will be calculated using term frequency-inverse document frequency (TF-IDF) weighting to adjust for differences in document length and overall term usage. Tracking these frequencies longitudinally will reveal how the lexical prominence of specific concepts has evolved over the years.

2.2 Regression

2.2.1 Purpose of the Regression

The regression analysis in this project aims to quantitatively evaluate the relationship between the content and thematic priorities of the Women's Budget Statements (WBS) and the allocation of government budgets in Australia from 1984 to 2013 and 2021 to 2025. This approach allows us to assess whether shifts in policy emphasis within the WBS are associated with changes in gender-related budget outcomes over time. In addition, the analysis will examine whether these relationships differ under governments led by the Australian Labor Party and the Liberal Party of Australia.

By applying Ordinary Least Squares (OLS) regression, we aim to establish empirical links between policy narratives (captured through textual analysis) and budgetary decisions, offering evidence on how gender-oriented discourse translates into fiscal outcomes.

2.2.2 Model specification

The baseline regression model can be expressed as follows:

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

$$Budget_t = a + \beta_1 Content_t + \beta_2 Party_t + \beta_3 X_t + \varepsilon_t$$

where:

- $Budget_t$ represents women-related or gender-focused government budget allocations in year t.
- $Content_t$ denotes the quantified thematic or linguistic content extracted from the WBS in year t, based on content analysis results.
- $Party_t$ is a dummy variable for the ruling political party (1 = Labor, 0 = Liberal).
- X_t includes control variables such as macroeconomic indicators (GDP growth, unemployment rate, inflation) and year fixed effects to capture temporal variation.
- ε_t is the error term.

To further explore differences between political parties, an interaction term ($Content_t \times Party_t$) may be included to assess whether the effect of WBS content varies depending on the governing party's ideology.

2.2.3 Variable Construction

Dependent Variable: The dependent variable will be the government budget allocation directed toward gender equality or women-focused initiatives. Data will be obtained from the official Women's Budget Statements and corresponding federal budget papers. Specific subcategories—such as allocations to employment, education, health, or safety—may be analyzed separately for greater specificity.

Independent Variable: The main independent variable will be derived from a content analysis of the WBS documents. Using computational text analysis methods (e.g., keyword

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

frequency or topic modeling), we will quantify the proportion of content dedicated to specific thematic areas (e.g., economic participation, leadership, safety, health). These quantitative measures of thematic emphasis will serve as proxies for policy focus in each year.

Control Variables: To account for external influences, we will include:

- Political Party (1 = Labor, 0 = Liberal)
- Macroeconomic Indicators (GDP growth, unemployment, inflation)
- Year Fixed Effects to capture long-term temporal trends and structural changes.

3. Data

3.1 Women's Budget Statement

Gender-responsive budgeting in Australia has a long, with annual publications released under various titles. This paper focuses on the Women's Budget Statement and related documents, noting that some political leaders have changed the title of these budget papers over time. The data mainly includes documents published under the Women's Budget Program and the Women's Budget Statement, along with several alternative titles such as More Choice for Women, Our Commitment to Women, Maintaining Our Commitment to Women, Delivering on Our Commitment for Women, and Strengthening Our Commitment to Women.

Table 1. Name changes Over Time

Published Year	Name
1984 ~ 1986	Women's Budget Program
1987 ~ 1995	Women's Budget Statement
1996	More Choice for Women
1997	Our Commitment to Women
1998	Maintaining Our Commitment to Women
1999	Delivering on Our Commitments for Women
2000	Strengthening Our Commitment to Women

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

2001 ~ 2003	Women's Budget Statement
2004 & 2005	What the Australian government is doing for women
2006	Budget Information - Women 06
2007	Missing / Unidentified
2008 ~ 2011	Women's Budget Statement
2012	Women's Statement - Achievements and Budget Measures
2013	Women's Budget Highlights
2014 ~ 2020	Abandoned
2021 ~ 2025	Women's Budget Statement

Women's Budget documents are usually accessible via two major archival platforms: *TROVE*, a national archive maintained by the National Library of Australia (National Library of Australia, n.d.), and the *Budget Archive* maintained by the Australian Government (Australian Government, n.d.). However, some documents are not available through these platforms. Australian librarians at the National Library were active and supportive in assisting with the document retrieval process.

Unfortunately, the 2004 document What the Australian government is doing for women is only available in hard copy and cannot be accessed online. In addition, the 2007 document appears to be missing, and the librarian was unable to identify the original women's budget paper for that year. Therefore, while nearly all WBSs were collected, the documents from 2004 and 2007 remain unavailable. Moreover, there was a period of discontinuation from 2014 to 2020 during which no Women's Budget papers were published.

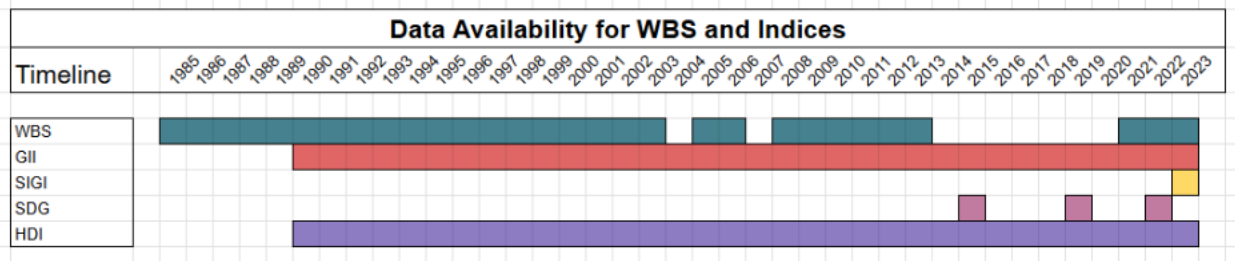
The Women's Budget Statement began in 1984 and was discontinued from 2014 to 2020. The data availability as shown in Figure 1 is below.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Figure 1. Data Availability



3.2 International Gender Indices

The Gender Inequality Index (GII), Social Institutions and Gender Index (SIGI), and Sustainable Development Goals (SDG) Gender Index are international indices that are tools for measuring and analyzing gender inequality. Developed by the United Nations Development Programme (UNDP), The Organization for Economic Cooperation and Development (OCED), and the Sustainable Development Goals (SDG) Gender Index respectively, focus on various aspects contributing to gender inequality. The GII has three major dimensions, reproductive health, empowerment and the labor market involvement, with an emphasis on the differences in outcomes between men and women. With data starting as early as 1990 it will provide a robust foundation for exploring whether there exists a correlation between the topics of Australia's Women's Budget Statements and their progress towards gender inequality.

On the other hand, the SIGI approaches measurement from a policy perspective using the lenses of discrimination in the family, physical integrity, access to productive and financial resources, and civil liberties. Beginning in 2009, it provides important data to help us understand Australia's efforts towards improving gender equity. The SDG Gender Index examines the global progress towards 2030 sustainable development goals from a gender equality perspective. With yearly reports from 2015 to 2024, it is a valuable contribution to gaining insight into Australia's recent progress towards gender inequality. However, SIGI overlaps with the WBS dataset for

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

only one year (2023), and the SDG Gender Index overlaps for just three years, making these data points too limited to conduct meaningful analysis in conjunction with WBS trends.

IV. Result

1. Textual analysis results

1.1 Topic modeling results

The Latent Dirichlet Allocation (LDA) model extracted five stable topics from the Women's Budget Statements (WBS) spanning 1997–2023. Although the specific keywords composing each topic vary slightly over time, the core thematic structure remains consistent, allowing for longitudinal interpretation of Australia's gender-responsive budgeting priorities. Table 1 and table 2 present the evolution of topic keywords by year. Below, we summarize each topic and identify its temporal patterns.

Throughout the entire period, topics consistently reflect the government's emphasis on women's access to social services—including care, health, and national support programs—indicating that foundational welfare-oriented language has remained a core component of WBS discourse. Topics center on gender-based violence, women's safety, and health-related protection, with its salience increasing during the late 2000s and again after 2021. Furthermore, some topics capture the administrative and programmatic dimensions of GRB—employment initiatives, departmental programs, and community assistance—highlighting an enduring bureaucratic focus on workforce participation and program delivery.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Table 2. Topic Modeling Results

Year	Topic 1 - Keywords	Topic 2 - Keywords	Topic 3 - Keywords
1997	women, government, care, australian, million, services, support, year, national, health	women, government, support, violence, per, health, gender, cent, work, australia	women, program, programs, services, government, employment, department, community, assistance, education
1998	women, government, care, australian, million, services, support, year, national, health	women, program, programs, services, government, employment, department, community, assistance, education	women, government, support, violence, per, health, gender, cent, work, australia,
2000	women, government, care, australian, million, services, support, year, national, health	women, community, communities, government, budget, national, families, australia, deg, morph	women, government, support, violence, per, health, gender, cent, work, australia,
2001	women, government, care, australian, million, services, support, year, national, health	women, government, support, violence, per, health, gender, cent, work, australia	women, program, programs, services, government, employment, department, community, assistance, education
2002	women, government, care, australian, million, services, support, year, national, health	women, budget, national, australia, page, la, library, obj, statement, impact	women, government, support, violence, per, health, gender, cent, work, australia,
2003	women, government, care, australian, million, services, support, year, national, health	women, program, programs, services, government, employment, department, community, assistance, education	women, government, support, violence, per, health, gender, cent, work, australia,
2005	women, government, care, australian, million, services, support, year, national, health	women, program, programs, services, government, employment, department, community, assistance, education	women, government, support, violence, per, health, gender, cent, work, australia,
2008	women, government, care, australian, million, services, support, year, national, health	women, government, support, violence, per, health, gender, cent, work, australia	women, budget, national, australia, page, la, library, obj, statement, impact
2009	women, government, support, violence, per, health, gender, cent, work, australia	women, government, care, australian, million, services, support, years, national, health	women, program, programs, services, government, employment, department, community, assistance, education
2010	women, government, care, australian, million, services, support, year, national, health	women, government, support, violence, per, health, gender, cent, work, australia	women, program, programs, services, government, employment, department, community, assistance, education
2011	women, government, support, violence, per, health, gender, cent, work, australia	women, government, care, australian, million, services, support, years, national, health	women, program, programs, services, government, employment, department, community, assistance, education
2012	women, government, care, australian, million, services, support, year, national, health	women, government, support, violence, per, health, gender, cent, work, australia	women, program, programs, services, government, employment, department, community, assistance, education
2013	women, government, support, violence, per, health, gender, cent, work, australia	women, government, care, australian, million, services, support, years, national, health	women, program, programs, services, government, employment, department, community, assistance, education
2021	women, government, support, violence, per, health, gender, cent, work, australia	women, government, care, australian, million, services, support, years, national, health	women, program, programs, services, government, employment, department, community, assistance, education
2022. March	women, government, support, violence, per, health, gender, cent, work, australia	women, government, care, australian, million, services, support, years, national, health	women, program, programs, services, government, employment, department, community, assistance, education
2022. October	women, government, support, violence, per, health, gender, cent, work, australia	women, budget, national, australia, page, la, library, obj, statement, impact	women, program, programs, services, government, employment, department, community, assistance, education
2023	women, government, support, violence, per, health, gender, cent, work, australia	women, budget, national, australia, page, la, library, obj, statement, impact	women, program, programs, services, government, employment, department, community, assistance, education

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Table 3. Topic Modeling Results (cont.)

Year	Topic 4 - Keywords	Topic 4 - Keywords
1997	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
1998	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2000	women, program, programs, services, government, employment, department, community, assistance, education	women, budget, national, australia, page, la, obj, statement, impact
2001	women, community, communities, government, budget, national, families, australia, deg, morph	women, budget, national, australia, page, la, obj, statement, impact
2002	women, program, programs, services, government, employment, department, community, assistance, education	women, community, communities, government, budget, national, families, australia, deg, morph
2003	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2005	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2008	women, program, programs, services, government, employment, department, community, assistance, education	women, community, communities, government, budget, national, families, australia, deg, morph
2009	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2010	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2011	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2012	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2013	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2021	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2022. March	women, budget, national, australia, page, la, obj, statement, impact	women, community, communities, government, budget, national, families, australia, deg, morph
2022. October	women, government, care, australian, million, services, support, years, national, health	women, community, communities, government, budget, national, families, australia, deg, morph
2023	women, government, care, australian, million, services, support, years, national, health	women, community, communities, government, budget, national, families, australia, deg, morph

Additionally, some topics represent the fiscal and institutional framing of GRB itself, featuring recurrent references to national budgets, policy statements, and governmental priorities; this topic peaks during years when the WBS format or communication style was revised, such as the early 2000s and post-2021 reinstatement. Finally, the topic encompasses family- and community-oriented initiatives, linking gender policy to broader social resources and national budgeting for households and communities. Taken together, these topics reveal a long-term continuity in Australia's gender budgeting narrative, with periodic shifts in emphasis that correspond to political transitions, evolving social contexts, and the temporary discontinuation and later revival of GRB. The post-2021 period shows a clear reintensification of discourse

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

around women's safety and social services, suggesting that the reinstated WBS responds directly to contemporary political pressures and public concern about gender inequality.

1.2 World frequency analysis results

This study conducts world frequency analysis by abstracting the top 10 most frequently used meaningful words from each document. The word-frequency analysis provides a clear, quantitative view of how the Women's Budget Statements (WBS) prioritized specific concepts across time, revealing both stable framing devices and notable shifts in emphasis between 1997 and 2023.

Across all years, the term "women" is consistently the most frequent content word, ranging from roughly 1.6% to above 3% of all tokens per document. This persistent lexical prominence reflects the identity-anchored structure of the WBS, reinforcing that these documents remain explicitly targeted toward women's issues rather than gender-neutral or gender-mainstreamed framing. The rising peaks in 2021–2023 (above 2.8–3.0%) coincide with the reinstatement of GRB and the national reckoning around women's safety—indicating a renewed rhetorical centering of women in policy language.

A second stable pattern is the continuous presence of "government," "care," and "support." These terms consistently appear among the top words (typically between ~0.5–1.5% per year), suggesting that the WBS has historically framed gender equality through the lens of state-led service provision. Words such as "care," "services," "support," "health," "community," and "families" appear repeatedly in the top 10 across most years prior to 2013, signaling that early and mid-period WBS documents primarily emphasized the government's responsibility in addressing social needs and care infrastructures. The strong appearance of "care" and "health" in

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

almost every year demonstrates that care-economy and wellbeing themes have long been central pillars of Australian GRB, irrespective of political administration.

Notably, the frequency of “violence” displays a distinct temporal pattern. It is present sporadically in the late 1990s and early 2000s but increases substantially from 2008 onward, and especially after 2021. The resurgence of “violence” as a top word—from 2021 through the 2022 and 2023 statements—marked by national inquiries, public protests, and heightened visibility of gendered violence cases. This reinforces the finding from the topic modeling results that violence prevention re-emerges as a major thematic priority in the latest era of WBS documents.

The early 2000s also show increases in terms like “community,” “families,” “children,” and “rural.” These frequencies suggest that the Howard-era WBS adopted a more community- and family-oriented framing of gender equity, consistent with conservative policy agendas emphasizing household-level supports, regional targeting, and localized service delivery rather than system-level feminist reforms. Terms such as “pension,” “indigenous,” and “rural” appearing in the top ranks (e.g., 1997, 2000, 2001, 2008) indicate selective focus on specific population groups within the broader umbrella of women’s programs.

Another noticeable trend is the declining visibility of words associated with employment and paid work in the top 10 most frequent tokens. Terms like “work,” “men,” “paid,” and “employment” appear occasionally (especially around 2009–2013) but do not maintain stable prominence across all years. This suggests that while women’s labor market participation is an important policy domain, it is not the dominant rhetorical anchor of WBS documents related to care, health, safety, and community themes. These findings parallel prior scholarship noting

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Australia's tendency to approach gender inequality primarily through welfare and social policy channels rather than structural economic reforms.

Finally, the dataset shows an abrupt break in the time series between 2014 and 2020, when no WBS was issued. The immediate reappearance of "support," and "violence" as high-frequency terms in 2021 suggests that the revival of GRB intentionally re-established the discursive foundations of earlier periods while also responding to modern gender-equality debates. The 2021–2023 frequency patterns reflect a renewed governmental commitment emphasizing safety, national accountability, and measurable fiscal commitments, with vocabulary indicating a more institutionalized and metric-driven approach than earlier decades.

Overall, the word-frequency patterns reveal long-term continuity in placing women, government support, care, and social services at the center of WBS language, while shifts in emphasis on violence, community, and equality mirror key political, social, and administrative turning points in Australia's GRB trajectory.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Table 4. Top 10 Most Frequently Used Words Over Years

Year	1	2	3	4	5	6	7	8	9	10
1997	women (1.64%)	government (0.99%)	care (0.69%)	service (0.45%)	pension (0.44%)	health (0.42%)	rural (0.38%)	work (0.37%)	violence (0.37%)	child (0.33%)
1998	women (2.46%)	government (1.01%)	service (0.57%)	australia (0.47%)	violence (0.45%)	rural (0.43%)	care (0.41%)	support (0.39%)	commonwealth (0.36%)	community (0.36%)
2000	women (1.37%)	government (1.17%)	community (0.91%)	families (0.78%)	family (0.71%)	communities (0.66%)	care (0.62%)	work (0.46%)	support (0.46%)	programmer (0.41%)
2001	women (1.89%)	government (0.85%)	community (0.55%)	care (0.53%)	indigenous (0.47%)	violence (0.46%)	programmer (0.44%)	support (0.44%)	services (0.42%)	health (0.41%)
2002	women (2.65%)	care (1.40%)	government (1.03%)	aged (0.94%)	funding (0.63%)	support (0.54%)	budget (0.47%)	community (0.40%)	older (0.40%)	provide (0.37%)
2003	women (1.93%)	care (1.12%)	service (0.98%)	funding (0.93%)	government (0.87%)	health (0.69%)	support (0.66%)	family (0.60%)	rural (0.59%)	community (0.53%)
2005	women (2.27%)	government (1.17%)	australian (0.98%)	care (0.74%)	family (0.65%)	support (0.63%)	million (0.56%)	services (0.47%)	child (0.43%)	health (0.43%)
2008	women (2.43%)	government (1.74%)	health (1.24%)	australian (0.85%)	child (0.80%)	budget (0.69%)	children (0.80%)	working (0.50%)	families (0.44%)	measures (0.44%)
2009	women (2.24%)	government (0.90%)	support (0.63%)	million (0.60%)	australian (0.59%)	budget (0.56%)	work (0.56%)	australia (0.56%)	paid (0.54%)	provide (0.51%)
2010	women (2.42%)	government (1.26%)	australian (1.03%)	national (0.73%)	care (0.64%)	health (0.57%)	australia (0.56%)	services (0.53%)	men (0.49%)	work (0.43%)
2011	women (1.86%)	australian (0.90%)	support (0.84%)	government (0.80%)	health (0.70%)	families (0.64%)	service (0.62%)	australia (0.57%)	men (0.54%)	children (0.50%)
2012	women (2.73%)	australian (0.78%)	government (0.76%)	australia (0.67%)	work (0.60%)	care (0.52%)	gender (0.47%)	violence (0.45%)	families (0.43%)	support (0.42%)
2013	women (2.04%)	disability (1.00%)	government (0.98%)	australia (0.81%)	australian (0.75%)	support (0.74%)	care (0.67%)	budget (0.48%)	education (0.46%)	services (0.44%)
2021	women (3.07%)	government (0.93%)	support (0.64%)	health (0.57%)	violence (0.55%)	million (0.51%)	care (0.48%)	work (0.45%)	budget (0.44%)	services (0.40%)
2022. March	women (2.84%)	government (0.87%)	million (0.84%)	violence (0.83%)	support (0.83%)	national (0.62%)	health (0.62%)	budget (0.55%)	australia (0.50%)	family (0.45%)
2022. October	women (2.69%)	gender (1.20%)	violence (0.89%)	equality (0.89%)	government (0.65%)	health (0.62%)	australia (0.48%)	children (0.48%)	work (0.46%)	men (0.45%)
2023	women (2.75%)	gender (0.92%)	government (0.83%)	health (0.69%)	support (0.69%)	violence (0.63%)	budget (0.58%)	work (0.55%)	equality (0.41%)	australia (40%)

* Note: The number in parentheses indicates the proportion of that word relative to the total word count.

2. Regression analysis results

2.1 Regression analysis

Prior to conducting the regression analysis, we examined the frequency distributions of the key words. As shown in Table 5, a substantial number of missing values were observed in the frequency data. In the first step, we selected the most frequently occurring key words, specifically the top eight across all observations: *health*, *violence*, *work*, *family*, *rural*, *children*, *men*, and *care*. In the second step, all missing frequency values were recorded as zero. Finally, we implemented Ordinary Least Squares (OLS) regression to establish the empirical relationships between keyword frequencies derived from textual analysis and the Gender Inequality Index (GII) and Human Development Index (HDI) scores. To ensure robustness, we estimated the models using two alternative specifications of the keyword variables: (1) a binary indicator capturing the presence or absence of each keyword, and (2) the original continuous frequency measures. This approach provides empirical evidence on how gender-oriented discourse translates into international development indicators.

Figure 2. Frequency of Top 20 keywords in Women's Budget Statement

Year	gender	disability	aged	health	violence	equality	work	pension	family	community	indigenous	rural	child & children	parents	men	cares	leave	education	economic, tax, income	sexual
1997				42	37		37	44				38	33			69				
1998					72					58		70	50			66				
2000							26													
2001				43	48		40				49					55				
2002				57	15															
2003			43	72								62	39							
2004																				
2005				106					162	91	94		107			183				
2006																				
2007																				
2008				45			21		16				48						12	
2009							35										47			
2010				50	37								33		43	56				
2011				94	50				86				67		67					
2012	56				54	43	115		51	42					44	62				
2013		98					39									66		45		
2021				151	146		121		101							129				
2022	330			172	245	187	127						133		123	123			103	
2023	253			191	173	112	152						95		91	113			103	

* Note: Blue-colored cells represent the Liberal Party, and red-colored cells represent the Labor Party.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

2.2 Regression result: keyword frequency as binary variables.

The binary regression models do not reveal statistically meaningful associations between the presence of key gender-related terms in Australia's Women's Budget Statements and shifts in either the GII or HDI over time. Across both models, none of the keyword variables reach conventional significance thresholds, and the coefficients fluctuate in sign and size, indicating instability rather than systematic directional influence.

This suggests that simply mentioning a policy theme (e.g., health, violence, family, rural, women, men, care) — without considering intensity, depth of discussion, or budgetary scale — is insufficient to predict measurable movement in national gender or human development outcomes. Put differently, rhetorical reference alone does not align closely with macro-level equity patterns.

Figure 3. Regression Analysis of Keywords (coded as binary) and GII scores

Source	SS	df	MS	Number of obs =	20
Model	.010531851	9	.001170206	F(9, 10) =	1.08
Residual	.010849349	10	.001084935	Prob > F =	0.4502
Total	.0213812	19	.001125326	R-squared =	0.4926
				Adj R-squared =	0.0359
				Root MSE =	.03294

GIIScore	Coefficient	Std. err.	t	P> t	[95% conf. interval]
health	-.0229896	.0266385	-0.86	0.408	-.0823438 .03636
violence	.0069289	.0232397	0.30	0.772	-.0448523 .05871
work	-.0138829	.0186286	-0.75	0.473	-.05539 .02762
family	.0102313	.017926	0.57	0.581	-.0297102 .05017
rural	.0302919	.0356951	0.85	0.416	-.0492417 .10982
children	.0062761	.0345587	0.18	0.860	-.0707256 .08327
men	-.0321295	.0361123	-0.89	0.395	-.1125927 .04833
cares	-.0048426	.0201967	-0.24	0.815	-.0498437 .04015
party_liberal_1	-.0050875	.027321	-0.19	0.856	-.0659624 .05578
_cons	.1490826	.025267	5.90	0.000	.0927841 .20538

Figure 4. Regression Analysis of Keywords (coded as binary) and HDI scores

Source	SS	df	MS	Number of obs =	20
Model	.005552118	9	.000616902	F(9, 10) =	1.94
Residual	.003180082	10	.000318008	Prob > F =	0.1582
Total	.0087322	19	.000459589	R-squared =	0.6358
				Adj R-squared =	0.3081
				Root MSE =	.01783

HDIScore	Coefficient	Std. err.	t	P> t	[95% conf. interval]
health	.0110207	.0144221	0.76	0.462	-.0211137 .0431
violence	-.0063016	.0125819	-0.50	0.627	-.0343359 .02173
work	.0078572	.0100855	0.78	0.454	-.0146147 .03032
family	.0002442	.0097051	0.03	0.980	-.0213801 .02186
rural	-.0132786	.0193253	-0.69	0.508	-.056338 .02978
children	-.0073992	.0187101	-0.40	0.701	-.0490879 .03428
men	.021279	.0195511	1.09	0.302	-.0222837 .06484
cares	.0031335	.0109345	0.29	0.780	-.0212301 .0274
party_liberal_1	-.0093163	.0147915	-0.63	0.543	-.0422739 .02364
_cons	.9168066	.0136795	67.02	0.000	.8863267 .94728

Taken together, these findings imply that discursive presence is not the same as policy impact, and that the translation from language to social outcome is likely mediated by implementation capacity, institutional priorities, program scale, and external social force factors not captured in binary text indicators. Future work may detect clearer relationships using

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

continuous measures of emphasis, lag structures, alternative indicators, or causal estimators, but under binary specification, the connection between WBS topic focus and national development scores appears weak or nonexistent.

2.3 Regression result: keyword frequency as continuous variable

The continuous regression models suggest that only selective aspects of Women's Budget Statement discourse exhibit measurable association with gender equality outcomes, and even these relationships are modest.

First, the frequency of the term "health" displays a statistically significant relationship with GII scores ($p = 0.046$), implying that as health discourse becomes more prominent in national gender budgeting language, gender inequality metrics shift correspondingly. This may reflect that public commitment to women's health topics aligns with moments when gender inequality issues become more visible or politically salient.

Figure 5. Regression Analysis of Keywords (coded as continuous) and GII scores

Source	SS	df	MS	Number of obs	=	20
Model	.018528697	9	.002058744	F(9, 10)	=	7.22
Residual	.002852503	10	.00028525	Prob > F	=	0.0024
Total	.0213812	19	.001125326	R-squared	=	0.8666
				Adj R-squared	=	0.7465
				Root MSE	=	.01689

GIIScore	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
health	-.0002535	.0001113	-2.28	0.046	-.0005016	-5.45e-06
violence	-.0001714	.0001631	-1.05	0.318	-.0005349	.000192
work	-.0002383	.0001721	-1.39	0.196	-.0006217	.0001451
family	.0000309	.0001182	0.26	0.799	-.0002324	.0002942
rural	.0004212	.0002429	1.73	0.114	-.0001201	.0009625
children	-.0000794	.0002171	-0.37	0.722	-.000563	.0004043
men	.0001214	.0002996	0.41	0.694	-.0005461	.0007889
cares	.0001185	.0001525	0.78	0.455	-.0002212	.0004583
party_liberal_1	-.0021788	.0120518	-0.18	0.860	-.0290318	.0246742
_cons	.153476	.0103179	14.87	0.000	.1304863	.1764656

Second, the variable capturing party change (party_liberal_1) shows a marginally significant association with HDI ($p \approx 0.10$), indicating that government turnover may relate

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

indirectly to shifts in human development framing within the WBS. This suggests institutional or ideological differences influencing policy focus or reporting structure, even if substantive outcome changes may lag behind rhetorical shifts.

Figure 6. Regression Analysis of Keywords (coded as continuous) and HDI scores

Source	SS	df	MS	Number of obs	=	20
Model	.006754795	9	.000750533	F(9, 10)	=	3.80
Residual	.001977405	10	.00019774	Prob > F	=	0.0246
				R-squared	=	0.7736
				Adj R-squared	=	0.5697
Total	.0087322	19	.000459589	Root MSE	=	.01406

HDIScore	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
health	.0001145	.0000927	1.23	0.245	-.0000921	.000321
violence	1.36e-06	.0001358	0.01	0.992	-.0003013	.000304
work	.0001488	.0001433	1.04	0.323	-.0001704	.000468
family	.0000234	.0000984	0.24	0.817	-.0001958	.0002426
rural	-.0001617	.0002023	-0.80	0.443	-.0006124	.000289
children	-.0000763	.0001807	-0.42	0.682	-.000479	.0003264
men	.0000426	.0002494	0.17	0.868	-.0005132	.0005984
cares	.0000133	.0001269	0.11	0.918	-.0002695	.0002962
party_liberal_1	-.0154697	.0100343	-1.54	0.154	-.0378274	.006888
_cons	.9180259	.0085906	106.86	0.000	.8988848	.937167

However, the vast majority of keywords exhibit no statistically significant relationship with either GII or HDI, meaning that increasing linguistic emphasis on themes such as violence, childcare, employment, men, or community does not systematically track with national development or gender equity trends. This implies that only certain thematic areas, particularly health, align meaningfully with broader gender outcomes, and that rhetorical presence alone does not reliably translate into development movement.

Overall, the findings indicate that WBS topic shifts show weak but detectable alignment with gender inequality measurements, yet only in limited domains. Australia's GRB discourse appears responsive to political change and health-related equity priorities, but the link between rhetoric and measured social progress remains partial and highly mediated, suggesting the need for deeper investigation into policy implementation, timing effects, and budget allocation mechanisms.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

2.4 Regression conclusion

To binary models that we regressed, there is no significant association between the presence of keywords in WBS and changes in GII or HDI. Simply mentioning themes such as health, violence, family, or care does not predict national gender or human development outcomes.

Based on the continuous models that we coded, we find overall linkages between keyword intensity and national outcomes are weak and selective. “Health” frequency is the only keyword significantly associated with GII ($p = 0.046$), suggesting that health-focused discourse aligns more closely with shifts in gender inequality. Party change shows a marginal association with HDI ($p \approx 0.10$), implying that institutional or ideological differences may shape reporting emphasis and framing.

Most thematic language in WBS does not correspond with measurable progress in gender or human development indicators. Rhetoric alone does not translate into social outcomes — budget implementation, policy scale, institutional capacity, and external pressures likely mediate these effects. The WBS reflects evolving policy narratives but only weakly maps onto national gender equity trends. Specific themes (e.g., health) show alignment, but broader discourse shifts do not reliably predict national progress, highlighting gaps between communication, resource delivery, and social outcomes.

V. Conclusion

The study finds that Australia’s WBSs have maintained consistent emphasis on care, health, and social support, with increased attention to gender-based violence in recent years.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Differences between Labor and Liberal governments are present but often subtle, with each prioritizing different aspects of gender-responsive policy. However, regression analysis reveals that these rhetorical shifts in WBS content show only weak associations with measurable improvements in gender equality indicators like GII and HDI. Notably, only “health” discourse showed a significant correlation with GII. These findings suggest that while WBS reflects evolving political priorities and public concerns, rhetorical emphasis alone is insufficient to drive national gender outcomes.

Research Question 1. How have policy targets, program areas, and thematic priorities in the Women’s Budget Statements evolved between 1984 and 2025?

Textual analysis reveals long-term continuity in themes such as care, health, and government support, with growing attention to violence and women’s safety post-2021. Thematic shifts often aligned with political changes or public concern.

Sub-research questions 1.1 How have the two major political parties (Australian Labor party & Liberal party of Australia) designed the Women's Budget Statements differently?

Under Liberal governments, WBS emphasized family, community, and localized support, while Labor governments lean more toward redistributive and public service narratives. However, these differences are sometimes subtle and overlapping.

Research Question 2. What is the relationship between changes in the Women’s Budget Statements’ topic focus and changes in the GII, HDI, and SIGI in Australia?

Sub-research questions 2.1 What kind of changes have the most effective impact on index changes?

Regression analysis shows limited statistical association between keyword presence in the WBS and changes in GII or HDI. The word “health” was the only term significantly

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

correlated with GII. This suggests that rhetorical emphasis alone does not strongly influence national-level gender outcomes.

Research Question 3. To what extent do shifts in the content and focus of the Women's Budget Statements correspond with broader trends in gender equity, as measured by international indices (e.g., GII & HDI)?

While WBS discourse aligns with moments of political or social shifts (e.g., post-2021), the connection to actual improvements in international gender indices remains weak. This implies that WBS reflects political narratives more than measurable development outcomes.

The results are closely related to limitations due to data. Data set has missing data for 2004 and 2007 WBS, and international indicators also have limited temporal, which may not fully capture the scope of national gender policy progress. Furthermore, because the WBS has produced at the federal level, it may overlook the complexities of implementation, regional variation, and community-level impacts.

Future research should therefore explore how GRB is operationalized across different levels of government, how budget commitments translate into concrete outcomes, and how institutional mechanisms support (or hinder) policy follow-through. Integrating qualitative insights, such as stakeholder interviews or program evaluations, could also enrich understanding of the pathways from rhetoric to results.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

References

- Althaus, Catherine. 2008. "What Do We Talk About? Publications in AJPA 1970–95." *Australian Journal of Public Administration* 56 (1): 141–146. <https://doi.org/10.1111/j.1467-8500.1997.tb01253.x>
- Althaus, Catherine. 2015. "What Do We Talk About Now? Reflecting on Publications in AJPA 1970–2015." *Australian Journal of Public Administration* 74 (2): 227–238. <https://doi.org/10.1111/1467-8500.12150>
- Australian Government. n.d. Budget Archive [Government]. Archive of Budgets. <https://archive.budget.gov.au/index.htm> (accessed April 11, 2025)
- Australian Government, Australia. Department of Foreign Affairs, Australia. Office for Women, Payne, M., Australia. Department of Families and Social Services, Australia. Women's Safety Council, Ruston, A., Australia. Office for Superannuation, Financial Services and the Digital Economy, Australia. Office for Women's economic equality, and Hume, J. 2021. *Women's Budget Statement 2021-22*. Commonwealth of Australia.
- Baird, M., Darlene Cox, Mary Crawford, Frances Davies, Helen Hodgson, Kathy MacDermott, Ruth Medd, Amanda Robbins, Joan Ross, Sue Salthouse, and Jozefa Sobski. 2014. *Gender Lens on the Budget 2014-2015*. National Foundation for Australian Women.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3 (Jan): 993-1022.
- Bracci, Enrico., Iris Saliterer, Mariafrancesca Sicilia, and Ileana Steccolini. 2021. "Accounting for (public) value(s): Reconsidering publicness in accounting research and practice." *Accounting, Auditing & Accountability Journal* 34 (7): 1513–1526.
- Budlender, Debbie. Diane Elson. Guy Hewitt. and Tanni Mukhopadhyay. 2011. *Gender Budgets Make Cents: Understanding Gender Responsive Budgets*. The Commonwealth Secretariat.
- Budlender, Debbie. And Guy Hewitt. 2002. *Gender Budgets Make More Cents: Country Studies and Good Practice*. The Commonwealth Secretariat. <https://www.oecd-ilibrary.org/content/publication/9781848597983-en>
- Campbell, Jim, and Morag Gillespie. 2016. *Feminist Economics and Public Policy: Reflections on the Work and Impact of Ailsa McKay*. Routledge, Taylor & Francis Group.
- Chakraborty, Lekha. 2016. "Asia: A Survey of Gender Budgeting Efforts." Working Paper No. 2016/150. International Monetary Fund.
- Diesner, Jana, and Kathleen M. Carley. 2005. "Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis." *Causal Mapping for Information Systems and Technology Research*: 81–108.
- Downes, Ronnie, Lisa von Trapp, and Scherie Nicol. 2017. "Gender budgeting in OECD countries." *OECD Journal on Budgeting* 16 (3): 71–107.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Elomäki, Anna, and Hanna Ylöstalo. 2021. "Gender budgeting in the crossroad of gender policy and public financial management: The Finnish case." *Public Money & Management* 41 (7): 516–526. <https://doi.org/10.1080/09540962.2021.1927528>

Equal Measures 2030. 2024. "2024 SDG Gender Index—Equal Measures 2030." Equal Measures 2030.

Galizzi, Giovanna., Elina Meliou, and Ileana Steccolini. 2021. "Theme: Experiences and challenges with gender budgeting and accounting. Moving towards gender-responsive forms of accountability?" *Public Money & Management* 41 (7): 499–501.

Grieve, Norma, and Ailsa Burns. 1994. *Australian Women: Contemporary Feminist Thought*. Oxford University Press.

Grimmer J, Stewart BM. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*. 2013;21(3):267-297.
[doi:10.1093/pan/mps028](https://doi.org/10.1093/pan/mps028)

Gunluk-Senesen, Gulay. (2021). "Wellbeing gender budgeting to localize the UN SDGs: Examples from Turkey." *Public Money & Management* 41 (7): 554–560.
<https://doi.org/10.1080/09540962.2021.1965402>

Kim, Solbee. 2025. "Why do countries adopt (or abandon) gender-responsive budget policies?: The case of Australia, 2014–2022." University of Delaware."

Lodovici, Manuela Samek. 2020. "EU Funding for Gender Equality: The Need for a Gender Responsive Budget." *Rivista Internazionale Di Scienze Sociali* 128 (1): 67–84.

Martínez Guzmán, Juan Pablo. 2024. "Can gender-responsive budgeting change how governments budget?: Lessons from the case of Ecuador." *Public Administration* 102 (2): 388–404. <https://doi.org/10.1111/padm.12926>

McDonald, Bruce D., and Sean McCandless. 2022. "11. Incorporating Social Equity." In *Teaching public budgeting and finance: A practical guide*, eds. Meagan M. Jordan and Bruce D. McDonald. Abingdon, Oxon: Routledge.

McDonald, Bruce D., and Sean McCandless. 2025. "Budgeting for Social Equity: Exploring the (Nearly) Unknown." *Journal of Social Equity and Public Administration* 3 (1): 9–27.
<https://doi.org/10.24926/jsepa.v3i1.4661>

National Library of Australia. n.d. Trove. <https://trove.nla.gov.au/> (accessed April 11, 2025)

OECD. n.d. "Social Institutions and Gender Index (SIGI)." <https://www.oecd.org/en/about/programmes/social-institutions-and-gender-index-sigi.html> (accessed October 1, 2025)

OECD. 2023. "OECD Best Practices for Gender Budgeting." *OECD Journal on Budgeting* 23 (1): 1–18.

Office of the Status of Women, and Department of the Prime Minister and Cabinet. 1984. "Women's Budget Program 1984–85." Commonwealth of Australia.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

- O'Hagan, Angela, and Elisabeth Klatzer. 2018. *Gender Budgeting in Europe: Developments and Challenges*. Springer. <http://link.springer.com/10.1007/978-3-319-64891-0>
- Ostry, Jonathan D., Jorge Alvarez, Raphael A. Espinoza, and Chris Papageorgiou. 2018. *Economic Gains From Gender Inclusion: New Mechanisms, New Evidence*. International Monetary Fund.
- Polzer, Tobias, Isabella M. Nolte, and Johann Seiwald. 2023. "Gender budgeting in public financial management: A literature review and research agenda." *International Review of Administrative Sciences* 89 (2): 450–466. <https://doi.org/10.1177/00208523211031796>
- Polzer, Tobias, and Johann Seiwald. 2021. "Gender-responsive budgeting in Austria: The narrow line between implementation and confirmation." *Public Money & Management* 41 (7): 527–538. <https://doi.org/10.1080/09540962.2021.1927516>
- Rubin, Marilyn M., and John R. Bartle. 2005. "Integrating Gender into Government Budgets: A New Perspective." *Public Administration Review* 65 (3): 259–272.
- Rubin, Marilyn M., and John R. Bartle. 2023. "Gender-responsive budgeting: A budget reform to address gender inequity." *Public Administration* 101 (2): 391–405.
- Sawer, Marian. 2002. "Australia: The mandarin approach to gender budgets." In *Gender Budgets Make More Cents: Country Studies and Good Practice*, eds. Debbie Budlender and Tina Johnson. The Commonwealth Secretariat.
- Sharp, Rhonda., and Ray Broomhill. 2014. "A case study of gender responsive budgeting in Australia." Research Report. Commonwealth Secretariat.
- Stotsky, Janet Gale. 2016. *Gender budgeting: Fiscal context and current outcomes*. International Monetary Fund.
- UN WOMEN. 1995a. "Beijing Declaration and Platform for Action, Beijing +5 Political Declaration and Outcome." United Nations.
- UN WOMEN. 1995b. "Fourth World Conference on Women, Beijing 1995." <https://www.un.org/womenwatch/daw/beijing/platform/plat2.htm> (accessed December 11, 2022)
- United Nations Development Fund for Women (UNIFEM). 2002. *Gender Budget Initiatives: Strategies. Concepts and Experiences*. The United Nations Development Fund for Women.
- United Nations Human Development (UNDP). n.d.-a. "Gender Inequality Index (GII)." <https://hdr.undp.org/data-center/thematic-composite-indices/gender-inequality-index> (accessed November 30, 2024)
- United Nations Human Development (UNDP). n.d.-b. "Human Development Index (HDI)." <https://hdr.undp.org/data-center/human-development-index> (accessed October 1, 2025)
- Walby, Sylvia. 2005a. "Gender Mainstreaming: Productive Tensions in Theory and Practice." *Social Politics: International Studies in Gender, State & Society* 12 (3): 321–343.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

<https://doi.org/10.1093/sp/jxi018>

Walby, Sylvia. 2005b. "Introduction: Comparative gender mainstreaming in a global era."
International Feminist Journal of Politics 7 (4): 453–470.

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Appendix A

Data Cleaning and Tokenization Code (R Studio)

title: "EPPS 6302 Final Project" author: "Tan Lin" date: "2025-11-24"

```
knitr::opts_chunk$set(echo = TRUE)
```

Setting Up

```
#Loading Libraries
```

```
library(dplyr)
```

```
library(readtext)
```

```
library(quantda)
```

```
library(hunspell)
```

```
library(ggplot2)
```

```
# Load PDFs
```

```
corpus <- readtext(file = "WBS PDFs/**") %>% as.data.frame(stringsAsFactors = FALSE) %>% corpus()
```

Initial Text Processing

```
# Extracting Raw Tokens and Doing Simple Cleaning
```

```
raw_toks = tokens(corpus,  
                  remove_punct = TRUE,  
                  remove_symbols = TRUE,  
                  remove_numbers = TRUE,  
                  remove_url = TRUE)
```

```
# Getting Unique Tokens
```

```
toks = tokens_tolower(raw_toks) %>% # Converting Tokens to Lower Case  
tokens_remove(, pattern = stopwords("en")) # Removing Stopwords  
unique_toks = unique(unlist(toks)) # Unlist and Finding Unique Tokens
```

```
# Finding Misspelled Unique Tokens
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
raw_misspells = hunspell_find(unique_toks) # Getting a list of misspellings

# Selecting and Merging Misspellings
misspells = sapply(raw_misspells[lengths(raw_misspells) > 0], paste, collapse = "")

# Generating Suggestions Using Hunspell (Australian English)
suggests = hunspell_suggest(misspells, dict = "en_AU")

# Accepting the First Suggestion
first_suggests = unlist(lapply(suggests, function(x) x[1]))

# Counting Tokens
my_dfm = dfm(toks)
tok_counts <- colSums(my_dfm) %>% data.frame(word = names(.), count = .)

#Combining Data into a Dataframe
misspells_suggests = data.frame(misspells,first_suggests)
colnames(misspells_suggests) = c("word","suggestion")
spellcheck_df = left_join(misspells_suggests,tok_counts) %>%
  distinct(word, .keep_all = TRUE) %>% #Keep Distinct
  arrange(desc(count)) #Sort
```

Manual Corrections

```
# Filling in Acronyms
acronyms = read.csv("ACRONYMS.csv") %>% as.data.frame()
acronyms$word = sapply(acronyms$word,tolower)

spellcheck_df2 = left_join(spellcheck_df,acronyms) %>%
  mutate(suggestion = ifelse(!is.na(full_name),full_name, suggestion))
spellcheck_df2[lengths(spellcheck_df2$suggestion) > 1,]
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
# If it exists, use the manual corrections csv to update the spellcheck_df, otherwise
create the csv
```

```
if(!file.exists("spellcheck_update")){
#write.csv(spellcheck_df2[601:800,c("word","suggestion")],"spellcheck_update.csv") }
spellcheck_update = read.csv("spellcheck_update.csv") %>% as.data.frame()
```

```
# Apply the manual corrections to the df. Apply the manual corrections to the tokens
list.
```

```
colnames(spellcheck_update) = c("index","word","update")
spellcheck_df3 = left_join(spellcheck_df2[,1:2],spellcheck_update) %>%
  mutate(suggestion = ifelse(!is.na(update),update, suggestion))
spellcheck = spellcheck_df3[,c("word","suggestion")] %>%
  distinct(word, .keep_all = TRUE) %>%
  mutate(suggestion = ifelse(is.na(suggestion), "", suggestion))
spellchecked_toks = tokens_replace(toks,spellcheck$word,spellcheck$suggestion) %>%
tokens_remove("") %>% tokens_wordstem()
```

Tokens into CSV

```
# Reconstruct text
cleaned_text = sapply(spellchecked_toks, paste, collapse = " ")
cleaned_df = data.frame(docnames(corpus),cleaned_text)
write.csv(cleaned_df, "corpus_cleaned.csv")
```

Some Fun Visualizations

```
raw_counts = dfm(raw_toks) %>% rowSums()
raw_counts_df = data.frame(document = names(raw_counts),words = raw_counts)
row.names(raw_counts_df) = NULL
raw_counts_df$document = substring(raw_counts_df$document,0,11)
ggplot(raw_counts_df,aes(document,words)) +
geom_bar(stat="identity",fill="lightblue") + theme(axis.text.x = element_text(angle =
90, vjust = 0.5, hjust=1),legend.position = "none") + xlab("WBS Year") + ylab("Number
of Words") + ggtitle("Word Count for WBS Documents")
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

Appendix B

Textual analysis (Python)

title: "EPPS 6302 Final Project" author: "Xinyu Yan" date: "2025-11-26"

```
{
  "cells": [
    {
      "cell_type": "code",
      "execution_count": 30,
      "id": "562f4e83-415f-441e-aaa7-eb73024778d4",
      "metadata": {},
      "outputs": [],
      "source": [
        "import pandas as pd\n",
        "\n",
        "# Load the CSV file\n",
        "df = pd.read_csv(r"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget\n",
        "Project\\WBP.csv")\n",
        "df.columns = [\"year\", \"content\"]\n",
        "df.iloc[0, 0] = \"1984\"\n",
        "df.iloc[1, 0] = \"1985\"\n",
        "df.iloc[2, 0] = \"1986\"\n",
        "df.iloc[3, 0] = \"1987\"\n",
        "df.iloc[4, 0] = \"1988\"\n",
        "df.iloc[5, 0] = \"1989\"\n",
        "df.iloc[6, 0] = \"1990\"\n",
        "df.iloc[7, 0] = \"1991\"\n",
        "df.iloc[8, 0] = \"1992\"\n",
        "df.iloc[9, 0] = \"1993\"\n",
        "df.iloc[10, 0] = \"1994\"\n",
        "df.iloc[11, 0] = \"1995\"\n",
        "df.iloc[12, 0] = \"1997\"\n",
        "df.iloc[13, 0] = \"1998\"
      ]
    }
  ]
}
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"df.iloc[14, 0] = \"1999\\n\",
"df.iloc[15, 0] = \"2000\\n\",
"df.iloc[16, 0] = \"2001\\n\",
"df.iloc[17, 0] = \"2002\\n\",
"df.iloc[18, 0] = \"2003\\n\",
"df.iloc[19, 0] = \"2005\\n\",
"df.iloc[20, 0] = \"2008\\n\",
"df.iloc[21, 0] = \"2009\\n\",
"df.iloc[22, 0] = \"2010\\n\",
"df.iloc[23, 0] = \"2011\\n\",
"df.iloc[24, 0] = \"2012\\n\",
"df.iloc[25, 0] = \"2013\\n\",
"df.iloc[26, 0] = \"2021\\n\",
"df.iloc[27, 0] = \"2022October\\n\",
"df.iloc[28, 0] = \"2022March\\n\",
"df.iloc[29, 0] = \"2023\\n\",
"\\n"
]
},
{
  "cell_type": "code",
  "execution_count": 31,
  "id": "f589fe37-6c07-47cb-9493-1e3afae062c6",
  "metadata": {},
  "outputs": [],
  "source": [
    "df[\"total_words\"] = df[\"content\"].astype(str).apply(lambda x: len(x.split()))"
  ]
},
{
  "cell_type": "code",
  "execution_count": 32,
  "id": "9b449237-4fd9-48cf-92cc-fa8206c745a6",
```


Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling ``frame.insert`` many times, which has poor performance. Consider joining all
columns at once using `pd.concat(axis=1)` instead. To get a de-fragmented frame, use
``newframe = frame.copy()``"

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0"
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
    " df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n",
```

```
"C:\\Users\\xxy230002\\AppData\\Local\\Temp\\ipykernel_2636\\194985112.py:26:
PerformanceWarning: DataFrame is highly fragmented. This is usually the result of
calling `frame.insert` many times, which has poor performance. Consider joining all
columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use
`newframe = frame.copy()`\n",
```

```
" df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"] if
df.at[idx, "total_words"] > 0 else 0\n"
```

```
]
```

```
},
```

```
{
```

```
"name": "stdout",
```

```
"output_type": "stream",
```

```
"text": [
```

```
"          year          content \\n",
```

```
"0      1984  deparment op he senate paper no date c o aug a...  \n",
```

```
"1      1985  o p a department of the prime minister and ani...  \n",
```

```
"2      1986  department of the prime minister and cabinet o...  \n",
```

```
"3      1987  budget related paper no pp no of date am http ...  \n",
```

```
"4      1988  budget related paper no pp no of date am http ...  \n",
```

```
"5      1989  budget related paper no pp no of date am http ...  \n",
```

```
"6      1990  budget related paper no pp no of date am http ...  \n",
```

```
"7      1991  budget related paper no pp no of date am http ...  \n",
```

```
"8      1992  budget related paper no pp no of date am http ...  \n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"9	1993	budget related paper no pp no of date am http ...	\n",
"10	1994	budget related paper no pp no of date am http ...	\n",
"11	1995	budget related paper no pp no of date am http ...	\n",
"12	1997	our commitment to women statement by senator t...	\n",
"13	1998	maintaining our commitment to women statement ...	\n",
"14	1999	pp no of date pm http la gov au tar kine la ob...	\n",
"15	2000	strengthening our commitment to women statemen...	\n",
"16	2001	rqwhqwv women s budget statement iii major ini...	\n",
"17	2002	women s budget statement statement by senator ...	\n",
"18	2003	women s budget statement statement by senator ...	\n",
"19	2005	what the australian government is doing for wo...	\n",
"20	2008	women a message from the prime minister and th...	\n",
"21	2009	women the hon tanya berserk mp minister for th...	\n",
"22	2010	women the hon tanya berserk mp minister for th...	\n",
"23	2011	women s budget statement the hon take ellis mp...	\n",
"24	2012	women s statement achievements and budget meas...	\n",
"25	2013	women s budget highlights a strong and fair so...	\n",
"26	2021	women s budget statement by senator the honora...	\n",
"27	2022October	women s budget statement october circulated by...	\n",
"28	2022March	women s budget statement by senator the honora...	\n",
"29	2023	women s budget statement statement by senator ...	\n",
"\n",			
"	total_words	top1 top1_count top2 top2_count top3 top3_count	
\\n",			
"0	24902	the 1377 of 1120 to 823	\n",
"1	90431	the 6158 of 4242 and 3437	\n",
"2	102244	the 6626 of 4449 and 4060	\n",
"3	6664	of 939 the 627 budget 625	\n",
"4	5908	of 831 the 555 budget 553	\n",
"5	5236	of 735 the 491 budget 489	\n",
"6	5404	of 759 the 507 budget 505	\n",
"7	9016	of 1275 the 851 budget 849	\n",
"8	7042	of 993 the 663 budget 661	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"9	7378	of	1041	the	695	budget	693	\n",
"10	868	of	111	the	75	budget	73	\n",
"11	826	of	105	the	71	budget	69	\n",
"12	9958	the	601	to	374	and	366	\n",
"13	16103	the	1074	and	681	of	664	\n",
"14	767	la	53	library	53	of	52	\n",
"15	5633	the	284	and	266	to	195	\n",
"16	10404	the	547	and	494	to	453	\n",
"17	4272	the	255	of	185	and	161	\n",
"18	10447	the	587	and	407	of	378	\n",
"19	24844	the	1464	and	1060	to	900	\n",
"20	3628	the	204	and	168	to	127	\n",
"21	6301	the	337	and	264	to	217	\n",
"22	8793	the	539	and	347	to	302	\n",
"23	13401	the	599	and	594	to	480	\n",
"24	11900	and	565	the	556	to	424	\n",
"25	9810	the	493	and	410	to	336	\n",
"26	26680	the	1371	and	1236	to	992	\n",
"27	27537	and	1314	the	1292	to	937	\n",
"28	23918	the	1189	and	1127	to	959	\n",
"29	27577	the	1295	and	1251	to	950	\n",

"\n",

"	top4	...	top31_proportion	top32_proportion	top33_proportion	\\\n",
"0	and	...	0.004176	0.003976	0.003895	\n",
"1	to	...	0.003516	0.003340	0.003207	\n",
"2	to	...	0.003276	0.003218	0.003218	\n",
"3	women	...	0.000150	0.000150	0.000150	\n",
"4	women	...	0.000169	0.000169	0.000169	\n",
"5	women	...	0.000191	0.000191	0.000191	\n",
"6	women	...	0.000185	0.000185	0.000185	\n",
"7	women	...	0.000111	0.000111	0.000111	\n",
"8	women	...	0.000142	0.000142	0.000142	\n",
"9	women	...	0.000136	0.000136	0.000136	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"10	women	...	0.001152	0.001152	0.001152	\n",
"11	women	...	0.001211	0.001211	0.001211	\n",
"12	of	...	0.003916	0.003816	0.003716	\n",
"13	to	...	0.003726	0.003726	0.003664	\n",
"14	for	...	0.001304	0.001304	0.001304	\n",
"15	of	...	0.004438	0.004438	0.004438	\n",
"16	of	...	0.004229	0.004133	0.004133	\n",
"17	to	...	0.003979	0.003979	0.003979	\n",
"18	to	...	0.004307	0.004307	0.004212	\n",
"19	of	...	0.004267	0.004065	0.003985	\n",
"20	of	...	0.004961	0.004961	0.004686	\n",
"21	of	...	0.004285	0.004285	0.004126	\n",
"22	of	...	0.004322	0.004322	0.004322	\n",
"23	of	...	0.004477	0.004328	0.004253	\n",
"24	of	...	0.004286	0.004202	0.004202	\n",
"25	of	...	0.004587	0.004587	0.004383	\n",
"26	women	...	0.004423	0.004273	0.004123	\n",
"27	of	...	0.004467	0.004467	0.004140	\n",
"28	of	...	0.004390	0.004348	0.003888	\n",
"29	of	...	0.004098	0.004061	0.004025	\n",
"\n",						
"	top34_proportion	top35_proportion	top36_proportion	top37_proportion	\\\n",	
"0	0.003775	0.003775	0.003735	0.003694	\n",	
"1	0.003207	0.003163	0.003129	0.003085	\n",	
"2	0.003188	0.003159	0.003130	0.003003	\n",	
"3	0.000150	0.000150	0.000150	0.000150	\n",	
"4	0.000169	0.000169	0.000169	0.000169	\n",	
"5	0.000191	0.000191	0.000191	0.000191	\n",	
"6	0.000185	0.000185	0.000185	0.000185	\n",	
"7	0.000111	0.000111	0.000111	0.000111	\n",	
"8	0.000142	0.000142	0.000142	0.000142	\n",	
"9	0.000136	0.000136	0.000136	0.000136	\n",	
"10	0.001152	0.001152	0.001152	0.001152	\n",	

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"11	0.001211	0.001211	0.001211	0.001211	\n",
"12	0.003716	0.003615	0.003615	0.003515	\n",
"13	0.003602	0.003602	0.003291	0.003229	\n",
"14	0.001304	0.001304	0.001304	0.001304	\n",
"15	0.004438	0.004261	0.004261	0.004083	\n",
"16	0.004133	0.004037	0.004037	0.003941	\n",
"17	0.003979	0.003979	0.003979	0.003745	\n",
"18	0.004212	0.004116	0.004116	0.003925	\n",
"19	0.003824	0.003784	0.003784	0.003743	\n",
"20	0.004410	0.004410	0.003859	0.003859	\n",
"21	0.004126	0.003968	0.003968	0.003809	\n",
"22	0.004208	0.004208	0.004208	0.003980	\n",
"23	0.004179	0.004104	0.003806	0.003731	\n",
"24	0.003950	0.003697	0.003697	0.003697	\n",
"25	0.004383	0.004383	0.004383	0.004077	\n",
"26	0.004048	0.004048	0.003861	0.003861	\n",
"27	0.004140	0.004031	0.003813	0.003777	\n",
"28	0.003888	0.003888	0.003846	0.003846	\n",
"29	0.003771	0.003735	0.003699	0.003445	\n",
"\n",					
"	top38_proportion	top39_proportion	top40_proportion		\n",
"0	0.003534	0.003534	0.003494		\n",
"1	0.003063	0.002930	0.002698		\n",
"2	0.002983	0.002973	0.002944		\n",
"3	0.000150	0.000150	0.000150		\n",
"4	0.000169	0.000169	0.000169		\n",
"5	0.000191	0.000191	0.000191		\n",
"6	0.000185	0.000185	0.000185		\n",
"7	0.000111	0.000111	0.000111		\n",
"8	0.000142	0.000142	0.000142		\n",
"9	0.000136	0.000136	0.000136		\n",
"10	0.001152	0.001152	0.001152		\n",
"11	0.001211	0.001211	0.001211		\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"12      0.003414      0.003314      0.003113  \n",
"13      0.003105      0.003105      0.002981  \n",
"14      0.001304      0.001304      0.001304  \n",
"15      0.004083      0.004083      0.003906  \n",
"16      0.003941      0.003941      0.003845  \n",
"17      0.003745      0.003745      0.003511  \n",
"18      0.003733      0.003637      0.003637  \n",
"19      0.003703      0.003663      0.003663  \n",
"20      0.003583      0.003308      0.003308  \n",
"21      0.003809      0.003492      0.003333  \n",
"22      0.003980      0.003867      0.003753  \n",
"23      0.003731      0.003656      0.003656  \n",
"24      0.003613      0.003613      0.003529  \n",
"25      0.003976      0.003976      0.003874  \n",
"26      0.003861      0.003786      0.003673  \n",
"27      0.003777      0.003740      0.003740  \n",
"28      0.003763      0.003679      0.003637  \n",
"29      0.003409      0.003300      0.003300  \n",
"\n",
"[30 rows x 123 columns]\n"
]
}
],
"source": [
"from collections import Counter\n",
"import re\n",
"\n",
"def get_top_words(text):\n",
"    # Clean & extract words\n",
"    text = str(text).lower()\n",
"    words = re.findall(r"\\b\\w+\\b", text)\n",
"    \n",
"    freq = Counter(words).most_common(40) # top 20
```


Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"    \n",
"    # Pad to exactly 40 entries\n",
"    freq += [(\n", 0)] * (40 - len(freq))\n",
"    return freq\n",
"\n",
"# Create 80 new columns\n",
"for i in range(40):\n",
"    df[f"top{i+1}"] = \n"\n",
"    df[f"top{i+1}_count"] = 0\n",
"\n",
"# Fill columns\n",
"for idx, text in df["content"].items():\n",
"    top_words = get_top_words(text)\n",
"    for i, (word, count) in enumerate(top_words):\n",
"        df.at[idx, f"top{i+1}"] = word\n",
"        df.at[idx, f"top{i+1}_count"] = count\n",
"        df.at[idx, f"top{i+1}_proportion"] = count / df.at[idx, "total_words"]
if df.at[idx, "total_words"] > 0 else 0\n",
"\n",
"print (df)"
]
},
{
    "cell_type": "code",
    "execution_count": 33,
    "id": "b851fa93-14bc-4989-8fbb-5d1934ff38e6",
    "metadata": {},
    "outputs": [],
    "source": [
        "df.to_csv(r"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget
Project\\WBP_frenquency.csv", index=False)"
    ]
},
{
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"cell_type": "code",
"execution_count": 23,
"id": "fceef94e-f57e-4ebe-a227-8c4152716c95",
"metadata": {},
"outputs": [
  {
    "name": "stdout",
    "output_type": "stream",
    "text": [
      "
      year  total_words  top1  top1_count  top2  top2_count  top3
\\n",
      "0      1984      24902  the      1377      of      1120      to
\\n",
      "1      1985      90431  the      6158      of      4242      and
\\n",
      "2      1986      102244 the      6626      of      4449      and
\\n",
      "3      1987      6664   of      939      the      627  budget
\\n",
      "4      1988      5908   of      831      the      555  budget
\\n",
      "5      1989      5236   of      735      the      491  budget
\\n",
      "6      1990      5404   of      759      the      507  budget
\\n",
      "7      1991      9016   of      1275     the      851  budget
\\n",
      "8      1992      7042   of      993      the      663  budget
\\n",
      "9      1993      7378   of      1041     the      695  budget
\\n",
      "10     1994      868    of      111      the      75   budget
\\n",
      "11     1995      826    of      105      the      71   budget
\\n",
      "12     1997      9958  the      601      to      374   and
\\n",
      "13     1998      16103 the      1074     and      681   of
\\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
\n",
    "14      1999      767  la      53  library      53      of
\n",
    "15      2000      5633 the      284      and      266      to
\n",
    "16      2001      10404 the      547      and      494      to
\n",
    "17      2002      4272 the      255      of      185      and
\n",
    "18      2003      10447 the      587      and      407      of
\n",
    "19      2005      24844 the      1464      and      1060      to
\n",
    "20      2008      3628 the      204      and      168      to
\n",
    "21      2009      6301 the      337      and      264      to
\n",
    "22      2010      8793 the      539      and      347      to
\n",
    "23      2011      13401 the      599      and      594      to
\n",
    "24      2012      11900 and      565      the      556      to
\n",
    "25      2013      9810 the      493      and      410      to
\n",
    "26      2021      26680 the      1371      and      1236      to
\n",
    "27  2022October      27537 and      1314      the      1292      to
\n",
    "28  2022March      23918 the      1189      and      1127      to
\n",
    "29      2023      27577 the      1295      and      1251      to
\n",
    "\n",
    "    top3_count  top4  top4_count  ... top31_proportion  top32_proportion
\\n",
    "0      823  and      800  ...      0.004176      0.003976  \n",
    "1      3437  to      2740  ...      0.003516      0.003340  \n",
    "2      4060  to      3089  ...      0.003276      0.003218  \n",
    "3      625  women      624  ...      0.000150      0.000150  \n",
    "4      553  women      552  ...      0.000169      0.000169  \n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"5	489	women	488	...	0.000191	0.000191	\n",
"6	505	women	504	...	0.000185	0.000185	\n",
"7	849	women	848	...	0.000111	0.000111	\n",
"8	661	women	660	...	0.000142	0.000142	\n",
"9	693	women	692	...	0.000136	0.000136	\n",
"10	73	women	72	...	0.001152	0.001152	\n",
"11	69	women	68	...	0.001211	0.001211	\n",
"12	366	of	361	...	0.003916	0.003816	\n",
"13	664	to	566	...	0.003726	0.003726	\n",
"14	52	for	52	...	0.001304	0.001304	\n",
"15	195	of	189	...	0.004438	0.004438	\n",
"16	453	of	362	...	0.004229	0.004133	\n",
"17	161	to	146	...	0.003979	0.003979	\n",
"18	378	to	359	...	0.004307	0.004307	\n",
"19	900	of	785	...	0.004267	0.004065	\n",
"20	127	of	103	...	0.004961	0.004961	\n",
"21	217	of	185	...	0.004285	0.004285	\n",
"22	302	of	276	...	0.004322	0.004322	\n",
"23	480	of	413	...	0.004477	0.004328	\n",
"24	424	of	369	...	0.004286	0.004202	\n",
"25	336	of	271	...	0.004587	0.004587	\n",
"26	992	women	818	...	0.004423	0.004273	\n",
"27	937	of	894	...	0.004467	0.004467	\n",
"28	959	of	681	...	0.004390	0.004348	\n",
"29	950	of	822	...	0.004098	0.004061	\n",
"\n",							
"	top33_proportion	top34_proportion	top35_proportion	top36_proportion	\\n",		
"0	0.003895	0.003775	0.003775	0.003735	\n",		
"1	0.003207	0.003207	0.003163	0.003129	\n",		
"2	0.003218	0.003188	0.003159	0.003130	\n",		
"3	0.000150	0.000150	0.000150	0.000150	\n",		
"4	0.000169	0.000169	0.000169	0.000169	\n",		
"5	0.000191	0.000191	0.000191	0.000191	\n",		

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"6	0.000185	0.000185	0.000185	0.000185	\n",
"7	0.000111	0.000111	0.000111	0.000111	\n",
"8	0.000142	0.000142	0.000142	0.000142	\n",
"9	0.000136	0.000136	0.000136	0.000136	\n",
"10	0.001152	0.001152	0.001152	0.001152	\n",
"11	0.001211	0.001211	0.001211	0.001211	\n",
"12	0.003716	0.003716	0.003615	0.003615	\n",
"13	0.003664	0.003602	0.003602	0.003291	\n",
"14	0.001304	0.001304	0.001304	0.001304	\n",
"15	0.004438	0.004438	0.004261	0.004261	\n",
"16	0.004133	0.004133	0.004037	0.004037	\n",
"17	0.003979	0.003979	0.003979	0.003979	\n",
"18	0.004212	0.004212	0.004116	0.004116	\n",
"19	0.003985	0.003824	0.003784	0.003784	\n",
"20	0.004686	0.004410	0.004410	0.003859	\n",
"21	0.004126	0.004126	0.003968	0.003968	\n",
"22	0.004322	0.004208	0.004208	0.004208	\n",
"23	0.004253	0.004179	0.004104	0.003806	\n",
"24	0.004202	0.003950	0.003697	0.003697	\n",
"25	0.004383	0.004383	0.004383	0.004383	\n",
"26	0.004123	0.004048	0.004048	0.003861	\n",
"27	0.004140	0.004140	0.004031	0.003813	\n",
"28	0.003888	0.003888	0.003888	0.003846	\n",
"29	0.004025	0.003771	0.003735	0.003699	\n",
"\n",					
"	top37_proportion	top38_proportion	top39_proportion	top40_proportion	\n",
"0	0.003694	0.003534	0.003534	0.003494	\n",
"1	0.003085	0.003063	0.002930	0.002698	\n",
"2	0.003003	0.002983	0.002973	0.002944	\n",
"3	0.000150	0.000150	0.000150	0.000150	\n",
"4	0.000169	0.000169	0.000169	0.000169	\n",
"5	0.000191	0.000191	0.000191	0.000191	\n",
"6	0.000185	0.000185	0.000185	0.000185	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"7          0.000111          0.000111          0.000111          0.000111  \n",
"8          0.000142          0.000142          0.000142          0.000142  \n",
"9          0.000136          0.000136          0.000136          0.000136  \n",
"10         0.001152          0.001152          0.001152          0.001152  \n",
"11         0.001211          0.001211          0.001211          0.001211  \n",
"12         0.003515          0.003414          0.003314          0.003113  \n",
"13         0.003229          0.003105          0.003105          0.002981  \n",
"14         0.001304          0.001304          0.001304          0.001304  \n",
"15         0.004083          0.004083          0.004083          0.003906  \n",
"16         0.003941          0.003941          0.003941          0.003845  \n",
"17         0.003745          0.003745          0.003745          0.003511  \n",
"18         0.003925          0.003733          0.003637          0.003637  \n",
"19         0.003743          0.003703          0.003663          0.003663  \n",
"20         0.003859          0.003583          0.003308          0.003308  \n",
"21         0.003809          0.003809          0.003492          0.003333  \n",
"22         0.003980          0.003980          0.003867          0.003753  \n",
"23         0.003731          0.003731          0.003656          0.003656  \n",
"24         0.003697          0.003613          0.003613          0.003529  \n",
"25         0.004077          0.003976          0.003976          0.003874  \n",
"26         0.003861          0.003861          0.003786          0.003673  \n",
"27         0.003777          0.003777          0.003740          0.003740  \n",
"28         0.003846          0.003763          0.003679          0.003637  \n",
"29         0.003445          0.003409          0.003300          0.003300  \n",
"\n",
"[30 rows x 122 columns]\n"
]
}
],
"source": [
    "df = pd.read_csv(r\"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget\n",
    Project\\WBP_frenquency.csv\\")\n",
    "df = df.drop(df.columns[1], axis=1)\n",
    "print (df)\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"df.to_csv(r"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget
Project\\Simple_WBP_frenquency.csv", index=False)"

]
},
{
  "cell_type": "code",
  "execution_count": 14,
  "id": "726c7ca8-3cdd-4a26-85cb-29530b206b0c",
  "metadata": {},
  "outputs": [
    {
      "name": "stdout",
      "output_type": "stream",
      "text": [
        "Collecting gensim\n",
        "  Downloading gensim-4.4.0-cp310-cp310-win_amd64.whl.metadata (8.6 kB)\n",
        "Requirement already satisfied: pandas in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages
(2.2.3)\n",
        "Requirement already satisfied: nltk in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages
(3.9.1)\n",
        "Requirement already satisfied: numpy>=1.18.5 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
gensim) (2.2.5)\n",
        "Collecting scipy>=1.7.0 (from gensim)\n",
        "  Downloading scipy-1.15.3-cp310-cp310-win_amd64.whl.metadata (60 kB)\n",
        "Collecting smart_open>=1.8.1 (from gensim)\n",
        "  Downloading smart_open-7.5.0-py3-none-any.whl.metadata (24 kB)\n",
        "Requirement already satisfied: python-dateutil>=2.8.2 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
pandas) (2.9.0.post0)\n",
        "Requirement already satisfied: pytz>=2020.1 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
pandas) (2025.2)\n",
        "Requirement already satisfied: tzdata>=2022.7 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
pandas) (2025.2)\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"Requirement already satisfied: click in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
nlTK) (8.1.8)\\n",

"Requirement already satisfied: joblib in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
nlTK) (1.4.2)\\n",

"Requirement already satisfied: regex>=2021.8.3 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
nlTK) (2024.11.6)\\n",

"Requirement already satisfied: tqdm in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
nlTK) (4.67.1)\\n",

"Requirement already satisfied: six>=1.5 in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
python-dateutil>=2.8.2->pandas) (1.17.0)\\n",

"Collecting wrapt (from smart_open>=1.8.1->gensim)\\n",

"  Downloading wrapt-2.0.1-cp310-cp310-win_amd64.whl.metadata (9.2 kB)\\n",

"Requirement already satisfied: colorama in
c:\\users\\xxy230002\\appdata\\local\\miniconda3\\envs\\myenv\\lib\\site-packages (from
click->nlTK) (0.4.6)\\n",

"Downloading gensim-4.4.0-cp310-cp310-win_amd64.whl (24.4 MB)\\n",
"  ----- 0.0/24.4 MB ? eta -:--:--\\n",
"  ----- - 23.6/24.4 MB 114.9 MB/s eta
0:00:01\\n",
"  ----- 24.4/24.4 MB 96.5 MB/s eta
0:00:00\\n",

"Downloading scipy-1.15.3-cp310-cp310-win_amd64.whl (41.3 MB)\\n",
"  ----- 0.0/41.3 MB ? eta -:--:--\\n",
"  ----- 23.9/41.3 MB 116.2 MB/s eta
0:00:01\\n",
"  ----- 41.2/41.3 MB 113.9 MB/s eta
0:00:01\\n",
"  ----- 41.3/41.3 MB 97.1 MB/s eta
0:00:00\\n",

"Downloading smart_open-7.5.0-py3-none-any.whl (63 kB)\\n",

"Downloading wrapt-2.0.1-cp310-cp310-win_amd64.whl (60 kB)\\n",

"Installing collected packages: wrapt, scipy, smart_open, gensim\\n",

"\\n",

"  ----- 0/4 [wrapt]\\n",
"  ----- 1/4 [scipy]\\n",
```


Date: 12. 05. 2025

57

Date: 12. 05. 2025

58

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
" ----- 3/4 [gensim]\n",
" ----- 3/4 [gensim]\n",
" ----- 3/4 [gensim]\n",
" ----- 4/4 [gensim]\n",
"\n",
"Successfully installed gensim-4.4.0 scipy-1.15.3 smart_open-7.5.0 wrapt-
2.0.1\n",
"Note: you may need to restart the kernel to use updated packages.\n"
]
}
],
"source": [
"pip install gensim pandas nltk"
]
},
{
"cell_type": "code",
"execution_count": 36,
"id": "27853109-0ed3-4e00-818e-8312dc4673ae",
"metadata": {},
"outputs": [
{
"name": "stderr",
"output_type": "stream",
"text": [
"[nltk_data] Downloading package stopwords to\n",
"[nltk_data] C:\\Users\\xxy230002\\AppData\\Roaming\\nltk_data...\n",
"[nltk_data] Package stopwords is already up-to-date!\n"
]
},
{
"name": "stdout",
"output_type": "stream",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"text": [  
  "  
    year                                     content  \\n",  
  "0      1984  deparment op he senate paper no date c o aug a...  \\n",  
  "1      1985  o p a department of the prime minister and ani...  \\n",  
  "2      1986  department of the prime minister and cabinet o...  \\n",  
  "3      1987  budget related paper no pp no of date am http ...  \\n",  
  "4      1988  budget related paper no pp no of date am http ...  \\n",  
  "5      1989  budget related paper no pp no of date am http ...  \\n",  
  "6      1990  budget related paper no pp no of date am http ...  \\n",  
  "7      1991  budget related paper no pp no of date am http ...  \\n",  
  "8      1992  budget related paper no pp no of date am http ...  \\n",  
  "9      1993  budget related paper no pp no of date am http ...  \\n",  
  "10     1994  budget related paper no pp no of date am http ...  \\n",  
  "11     1995  budget related paper no pp no of date am http ...  \\n",  
  "12     1997  our commitment to women statement by senator t...  \\n",  
  "13     1998  maintaining our commitment to women statement ...  \\n",  
  "14     1999  pp no of date pm http la gov au tar kine la ob...  \\n",  
  "15     2000  strengthening our commitment to women statemen...  \\n",  
  "16     2001  rqwhqwv women s budget statement iii major ini...  \\n",  
  "17     2002  women s budget statement statement by senator ...  \\n",  
  "18     2003  women s budget statement statement by senator ...  \\n",  
  "19     2005  what the australian government is doing for wo...  \\n",  
  "20     2008  women a message from the prime minister and th...  \\n",  
  "21     2009  women the hon tanya berserk mp minister for th...  \\n",  
  "22     2010  women the hon tanya berserk mp minister for th...  \\n",  
  "23     2011  women s budget statement the hon take ellis mp...  \\n",  
  "24     2012  women s statement achievements and budget meas...  \\n",  
  "25     2013  women s budget highlights a strong and fair so...  \\n",  
  "26     2021  women s budget statement by senator the honora...  \\n",  
  "27  2022October  women s budget statement october circulated by...  \\n",  
  "28  2022March   women s budget statement by senator the honora...  \\n",  
  "29     2023  women s budget statement statement by senator ...  \\n",  
  "\\n",  
]
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"	total_words	top1	top1_count	top2	top2_count	top3	top3_count	
\\n",								
"0	24902	the	1377	of	1120	to	823	\n",
"1	90431	the	6158	of	4242	and	3437	\n",
"2	102244	the	6626	of	4449	and	4060	\n",
"3	6664	of	939	the	627	budget	625	\n",
"4	5908	of	831	the	555	budget	553	\n",
"5	5236	of	735	the	491	budget	489	\n",
"6	5404	of	759	the	507	budget	505	\n",
"7	9016	of	1275	the	851	budget	849	\n",
"8	7042	of	993	the	663	budget	661	\n",
"9	7378	of	1041	the	695	budget	693	\n",
"10	868	of	111	the	75	budget	73	\n",
"11	826	of	105	the	71	budget	69	\n",
"12	9958	the	601	to	374	and	366	\n",
"13	16103	the	1074	and	681	of	664	\n",
"14	767	la	53	library	53	of	52	\n",
"15	5633	the	284	and	266	to	195	\n",
"16	10404	the	547	and	494	to	453	\n",
"17	4272	the	255	of	185	and	161	\n",
"18	10447	the	587	and	407	of	378	\n",
"19	24844	the	1464	and	1060	to	900	\n",
"20	3628	the	204	and	168	to	127	\n",
"21	6301	the	337	and	264	to	217	\n",
"22	8793	the	539	and	347	to	302	\n",
"23	13401	the	599	and	594	to	480	\n",
"24	11900	and	565	the	556	to	424	\n",
"25	9810	the	493	and	410	to	336	\n",
"26	26680	the	1371	and	1236	to	992	\n",
"27	27537	and	1314	the	1292	to	937	\n",
"28	23918	the	1189	and	1127	to	959	\n",
"29	27577	the	1295	and	1251	to	950	\n",
"\n",								

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

	top4	...	top35_proportion	top36_proportion	top37_proportion	\\\n",
"0	and	...	0.003775	0.003735	0.003694	\n",
"1	to	...	0.003163	0.003129	0.003085	\n",
"2	to	...	0.003159	0.003130	0.003003	\n",
"3	women	...	0.000150	0.000150	0.000150	\n",
"4	women	...	0.000169	0.000169	0.000169	\n",
"5	women	...	0.000191	0.000191	0.000191	\n",
"6	women	...	0.000185	0.000185	0.000185	\n",
"7	women	...	0.000111	0.000111	0.000111	\n",
"8	women	...	0.000142	0.000142	0.000142	\n",
"9	women	...	0.000136	0.000136	0.000136	\n",
"10	women	...	0.001152	0.001152	0.001152	\n",
"11	women	...	0.001211	0.001211	0.001211	\n",
"12	of	...	0.003615	0.003615	0.003515	\n",
"13	to	...	0.003602	0.003291	0.003229	\n",
"14	for	...	0.001304	0.001304	0.001304	\n",
"15	of	...	0.004261	0.004261	0.004083	\n",
"16	of	...	0.004037	0.004037	0.003941	\n",
"17	to	...	0.003979	0.003979	0.003745	\n",
"18	to	...	0.004116	0.004116	0.003925	\n",
"19	of	...	0.003784	0.003784	0.003743	\n",
"20	of	...	0.004410	0.003859	0.003859	\n",
"21	of	...	0.003968	0.003968	0.003809	\n",
"22	of	...	0.004208	0.004208	0.003980	\n",
"23	of	...	0.004104	0.003806	0.003731	\n",
"24	of	...	0.003697	0.003697	0.003697	\n",
"25	of	...	0.004383	0.004383	0.004077	\n",
"26	women	...	0.004048	0.003861	0.003861	\n",
"27	of	...	0.004031	0.003813	0.003777	\n",
"28	of	...	0.003888	0.003846	0.003846	\n",
"29	of	...	0.003735	0.003699	0.003445	\n",
"\n",						
"	top38_proportion	top39_proportion	top40_proportion	\\\n",		

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"0	0.003534	0.003534	0.003494	\n",
"1	0.003063	0.002930	0.002698	\n",
"2	0.002983	0.002973	0.002944	\n",
"3	0.000150	0.000150	0.000150	\n",
"4	0.000169	0.000169	0.000169	\n",
"5	0.000191	0.000191	0.000191	\n",
"6	0.000185	0.000185	0.000185	\n",
"7	0.000111	0.000111	0.000111	\n",
"8	0.000142	0.000142	0.000142	\n",
"9	0.000136	0.000136	0.000136	\n",
"10	0.001152	0.001152	0.001152	\n",
"11	0.001211	0.001211	0.001211	\n",
"12	0.003414	0.003314	0.003113	\n",
"13	0.003105	0.003105	0.002981	\n",
"14	0.001304	0.001304	0.001304	\n",
"15	0.004083	0.004083	0.003906	\n",
"16	0.003941	0.003941	0.003845	\n",
"17	0.003745	0.003745	0.003511	\n",
"18	0.003733	0.003637	0.003637	\n",
"19	0.003703	0.003663	0.003663	\n",
"20	0.003583	0.003308	0.003308	\n",
"21	0.003809	0.003492	0.003333	\n",
"22	0.003980	0.003867	0.003753	\n",
"23	0.003731	0.003656	0.003656	\n",
"24	0.003613	0.003613	0.003529	\n",
"25	0.003976	0.003976	0.003874	\n",
"26	0.003861	0.003786	0.003673	\n",
"27	0.003777	0.003740	0.003740	\n",
"28	0.003763	0.003679	0.003637	\n",
"29	0.003409	0.003300	0.003300	\n",
"\n",				
"			tokens dominant_topic	\\\n",
"0	[deparment, op, senate, paper, date, c, aug, a...		1	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"1  [p, department, prime, minister, anisette, off...      1  \n",
"2  [department, prime, minister, cabinet, office,...      1  \n",
"3  [budget, related, paper, pp, date, http, la, g...      3  \n",
"4  [budget, related, paper, pp, date, http, la, g...      3  \n",
"5  [budget, related, paper, pp, date, http, la, g...      3  \n",
"6  [budget, related, paper, pp, date, http, la, g...      3  \n",
"7  [budget, related, paper, pp, date, http, la, g...      3  \n",
"8  [budget, related, paper, pp, date, http, la, g...      3  \n",
"9  [budget, related, paper, pp, date, http, la, g...      3  \n",
"10 [budget, related, paper, pp, date, http, la, g...      3  \n",
"11 [budget, related, paper, pp, date, http, la, g...      3  \n",
"12 [commitment, women, statement, senator, honora...      4  \n",
"13 [maintaining, commitment, women, statement, ho...      4  \n",
"14 [pp, date, pm, http, la, gov, au, tar, kine, l...      3  \n",
"15 [strengthening, commitment, women, statement, ...      4  \n",
"16 [rqwhqwv, women, budget, statement, iii, major...      4  \n",
"17 [women, budget, statement, statement, senator,...      4  \n",
"18 [women, budget, statement, statement, senator,...      4  \n",
"19 [australian, government, women, minister, mess...      4  \n",
"20 [women, message, prime, minister, minister, st...      4  \n",
"21 [women, hon, tanya, berserk, mp, minister, sta...      0  \n",
"22 [women, hon, tanya, berserk, mp, minister, sta...      4  \n",
"23 [women, budget, statement, hon, take, ellis, m...      0  \n",
"24 [women, statement, achievements, budget, measu...      4  \n",
"25 [women, budget, highlights, strong, fair, soci...      0  \n",
"26 [women, budget, statement, senator, honorable,...      0  \n",
"27 [women, budget, statement, october, circulated...      0  \n",
"28 [women, budget, statement, senator, honorable,...      0  \n",
"29 [women, budget, statement, statement, senator,...      0  \n",
"\n",
"      topic_score                                topic_keywords \n",
"0      0.994356  women, program, programs, services, government... \n",
"1      0.999925  women, program, programs, services, government... \n"
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"2      0.999209  women, program, programs, services, government...  \n",
"3      0.999806  women, budget, national, australia, page, la, ...  \n",
"4      0.999781  women, budget, national, australia, page, la, ...  \n",
"5      0.999754  women, budget, national, australia, page, la, ...  \n",
"6      0.999761  women, budget, national, australia, page, la, ...  \n",
"7      0.999857  women, budget, national, australia, page, la, ...  \n",
"8      0.999817  women, budget, national, australia, page, la, ...  \n",
"9      0.999825  women, budget, national, australia, page, la, ...  \n",
"10     0.998535  women, budget, national, australia, page, la, ...  \n",
"11     0.998462  women, budget, national, australia, page, la, ...  \n",
"12     0.594843  women, government, care, australian, million, ...  \n",
"13     0.834976  women, government, care, australian, million, ...  \n",
"14     0.998503  women, budget, national, australia, page, la, ...  \n",
"15     0.721227  women, government, care, australian, million, ...  \n",
"16     0.873899  women, government, care, australian, million, ...  \n",
"17     0.999685  women, government, care, australian, million, ...  \n",
"18     0.994287  women, government, care, australian, million, ...  \n",
"19     0.928612  women, government, care, australian, million, ...  \n",
"20     0.809581  women, government, care, australian, million, ...  \n",
"21     0.750866  women, government, support, violence, per, hea...  \n",
"22     0.682557  women, government, care, australian, million, ...  \n",
"23     0.673844  women, government, support, violence, per, hea...  \n",
"24     0.545533  women, government, care, australian, million, ...  \n",
"25     0.756768  women, government, support, violence, per, hea...  \n",
"26     0.997614  women, government, support, violence, per, hea...  \n",
"27     0.998894  women, government, support, violence, per, hea...  \n",
"28     0.999578  women, government, support, violence, per, hea...  \n",
"29     0.997497  women, government, support, violence, per, hea...  \n",
"\n",
"[30 rows x 127 columns]\n"
]
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"source": [  
    "from gensim import corpora, models\n",  
    "import gensim\n",  
    "import re\n",  
    "import nltk\n",  
    "from nltk.corpus import stopwords\n",  
    "nltk.download('stopwords')\n",  
    "df= pd.read_csv(r\"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget  
Project\\WBP_frenquency.csv\")\n",  
    "\n",  
    "# ----- CLEAN & TOKENIZE ----- \n",  
    "stop_words = set(stopwords.words('english'))\n",  
    "\n",  
    "def tokenize(text):\n",  
    "    text = str(text).lower()\n",  
    "    words = re.findall(r'\\b\\w+\\b', text)\n",  
    "    words = [w for w in words if w not in stop_words]\n",  
    "    return words\n",  
    "\n",  
    "df['tokens'] = df['content'].apply(tokenize)\n",  
    "\n",  
    "# ----- DICTIONARY & CORPUS ----- \n",  
    "from gensim import corpora\n",  
    "\n",  
    "tokens = df['tokens'].tolist()\n",  
    "\n",  
    "dictionary = corpora.Dictionary(tokens)\n",  
    "corpus = [dictionary.doc2bow(text) for text in tokens]\n",  
    "\n",  
    "# ----- TRAIN TOPIC MODEL ----- \n",  
    "from gensim.models import LdaModel\n",  
    "NUM_TOPICS = 5 # you can adjust\n",  
    "\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
lda_model = LdaModel(\n",
    corpus=corpus,\n",
    id2word=dictionary,\n",
    num_topics=NUM_TOPICS,\n",
    passes=10\n",
)\n",
\n",
"# ----- PER-ROW TOPIC EXTRACTION -----\n",
dominant_topic = []\n",
topic_score = []\n",
topic_keywords = []\n",
\n",
for bow in corpus:\n",
    topic_probs = lda_model.get_document_topics(bow)\n",
    topic_probs = sorted(topic_probs, key=lambda x: x[1], reverse=True)\n",
\n",
    top_topic, top_prob = topic_probs[0]\n",
\n",
    dominant_topic.append(top_topic)\n",
    topic_score.append(top_prob)\n",
    topic_keywords.append(\n",
        \", \".join(w for w, p in lda_model.show_topic(top_topic, topn=10))\n",
    )\n",
"# ----- ADD RESULTS TO DATAFRAME -----\n",
df["dominant_topic"] = dominant_topic\n",
df["topic_score"] = topic_score\n",
df["topic_keywords"] = topic_keywords\n",
\n",
print(df)
]
```

}

{

"cell_type": "code",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"execution_count": 38,
"id": "144f6d1e-4abe-40de-8ec9-8314d0ff4f49",
"metadata": {},
"outputs": [
  {
    "name": "stdout",
    "output_type": "stream",
    "text": [
      "          year                                content  \\n",
      "0          1984  deparment op he senate paper no date c o aug a...  \\n",
      "1          1985  o p a department of the prime minister and ani...  \\n",
      "2          1986  department of the prime minister and cabinet o...  \\n",
      "3          1987  budget related paper no pp no of date am http ...  \\n",
      "4          1988  budget related paper no pp no of date am http ...  \\n",
      "5          1989  budget related paper no pp no of date am http ...  \\n",
      "6          1990  budget related paper no pp no of date am http ...  \\n",
      "7          1991  budget related paper no pp no of date am http ...  \\n",
      "8          1992  budget related paper no pp no of date am http ...  \\n",
      "9          1993  budget related paper no pp no of date am http ...  \\n",
      "10         1994  budget related paper no pp no of date am http ...  \\n",
      "11         1995  budget related paper no pp no of date am http ...  \\n",
      "12         1997  our commitment to women statement by senator t...  \\n",
      "13         1998  maintaining our commitment to women statement ...  \\n",
      "14         1999  pp no of date pm http la gov au tar kine la ob...  \\n",
      "15         2000  strengthening our commitment to women statemen...  \\n",
      "16         2001  rqwhqwv women s budget statement iii major ini...  \\n",
      "17         2002  women s budget statement statement by senator ...  \\n",
      "18         2003  women s budget statement statement by senator ...  \\n",
      "19         2005  what the australian government is doing for wo...  \\n",
      "20         2008  women a message from the prime minister and th...  \\n",
      "21         2009  women the hon tanya berserk mp minister for th...  \\n",
      "22         2010  women the hon tanya berserk mp minister for th...  \\n",
      "23         2011  women s budget statement the hon take ellis mp...  \\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"24 2012 women s statement achievements and budget meas... \n",
"25 2013 women s budget highlights a strong and fair so... \n",
"26 2021 women s budget statement by senator the honora... \n",
"27 2022October women s budget statement october circulated by... \n",
"28 2022March women s budget statement by senator the honora... \n",
"29 2023 women s budget statement statement by senator ... \n",
"\n",

"	total_words	top1	top1_count	top2	top2_count	top3	top3_count	\\\n",
"0	24902	the	1377	of	1120	to	823	\n",
"1	90431	the	6158	of	4242	and	3437	\n",
"2	102244	the	6626	of	4449	and	4060	\n",
"3	6664	of	939	the	627	budget	625	\n",
"4	5908	of	831	the	555	budget	553	\n",
"5	5236	of	735	the	491	budget	489	\n",
"6	5404	of	759	the	507	budget	505	\n",
"7	9016	of	1275	the	851	budget	849	\n",
"8	7042	of	993	the	663	budget	661	\n",
"9	7378	of	1041	the	695	budget	693	\n",
"10	868	of	111	the	75	budget	73	\n",
"11	826	of	105	the	71	budget	69	\n",
"12	9958	the	601	to	374	and	366	\n",
"13	16103	the	1074	and	681	of	664	\n",
"14	767	la	53	library	53	of	52	\n",
"15	5633	the	284	and	266	to	195	\n",
"16	10404	the	547	and	494	to	453	\n",
"17	4272	the	255	of	185	and	161	\n",
"18	10447	the	587	and	407	of	378	\n",
"19	24844	the	1464	and	1060	to	900	\n",
"20	3628	the	204	and	168	to	127	\n",
"21	6301	the	337	and	264	to	217	\n",
"22	8793	the	539	and	347	to	302	\n",
"23	13401	the	599	and	594	to	480	\n",
"24	11900	and	565	the	556	to	424	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"25      9810  the      493      and      410      to      336      \n",
"26      26680 the      1371     and      1236     to      992      \n",
"27      27537 and      1314     the      1292     to      937      \n",
"28      23918 the      1189     and      1127     to      959      \n",
"29      27577 the      1295     and      1251     to      950      \n",
"\n",
"      top4 ...                               topic2_keywords topic3_id  \\\n",
"0      and ... women, government, support, violence, per, hea...      3      \n",
"1      to ... women, government, support, violence, per, hea...      4      \n",
"2      to ... women, government, support, violence, per, hea...      4      \n",
"3  women ... women, government, support, violence, per, hea...      4      \n",
"4  women ... women, government, support, violence, per, hea...      4      \n",
"5  women ... women, government, support, violence, per, hea...      4      \n",
"6  women ... women, government, support, violence, per, hea...      4      \n",
"7  women ... women, government, support, violence, per, hea...      4      \n",
"8  women ... women, government, support, violence, per, hea...      4      \n",
"9  women ... women, government, support, violence, per, hea...      4      \n",
"10 women ... women, government, care, australian, million, ...      0      \n",
"11 women ... women, government, care, australian, million, ...      0      \n",
"12    of ... women, government, support, violence, per, hea...      1      \n",
"13    to ... women, program, programs, services, government...      0      \n",
"14   for ... women, government, care, australian, million, ...      0      \n",
"15    of ... women, community, communities, government, bud...      0      \n",
"16    of ... women, government, support, violence, per, hea...      1      \n",
"17    to ... women, budget, national, australia, page, la, ...      0      \n",
"18    to ... women, program, programs, services, government...      0      \n",
"19    of ... women, program, programs, services, government...      0      \n",
"20    of ... women, government, support, violence, per, hea...      3      \n",
"21    of ... women, government, care, australian, million, ...      1      \n",
"22    of ... women, government, support, violence, per, hea...      1      \n",
"23    of ... women, government, care, australian, million, ...      1      \n",
"24    of ... women, government, support, violence, per, hea...      1      \n",
"25    of ... women, government, care, australian, million, ...      1      \n
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"26 women ... women, government, care, australian, million, ...      1  \n",
"27   of ... women, budget, national, australia, page, la, ...      1  \n",
"28   of ... women, government, care, australian, million, ...      1  \n",
"29   of ... women, budget, national, australia, page, la, ...      1  \n",
"\n",
"   topic3_prob                                topic3_keywords  topic4_id  \\n",
"0    0.000458 women, budget, national, australia, page, la, ...      4  \n",
"1    0.000014 women, government, care, australian, million, ...      3  \n",
"2    0.000214 women, government, care, australian, million, ...      3  \n",
"3    0.000048 women, government, care, australian, million, ...      1  \n",
"4    0.000055 women, government, care, australian, million, ...      1  \n",
"5    0.000062 women, government, care, australian, million, ...      1  \n",
"6    0.000060 women, government, care, australian, million, ...      1  \n",
"7    0.000036 women, government, care, australian, million, ...      1  \n",
"8    0.000046 women, government, care, australian, million, ...      1  \n",
"9    0.000044 women, government, care, australian, million, ...      1  \n",
"10   0.000366 women, government, support, violence, per, hea...      1  \n",
"11   0.000385 women, government, support, violence, per, hea...      1  \n",
"12   0.175172 women, program, programs, services, government...      3  \n",
"13   0.000730 women, government, support, violence, per, hea...      3  \n",
"14   0.000375 women, government, support, violence, per, hea...      1  \n",
"15   0.043232 women, government, support, violence, per, hea...      1  \n",
"16   0.054883 women, program, programs, services, government...      2  \n",
"17   0.000079 women, government, support, violence, per, hea...      1  \n",
"18   0.000333 women, government, support, violence, per, hea...      3  \n",
"19   0.026290 women, government, support, violence, per, hea...      3  \n",
"20   0.000090 women, budget, national, australia, page, la, ...      1  \n",
"21   0.093399 women, program, programs, services, government...      3  \n",
"22   0.000365 women, program, programs, services, government...      3  \n",
"23   0.027505 women, program, programs, services, government...      3  \n",
"24   0.000029 women, program, programs, services, government...      3  \n",
"25   0.009380 women, program, programs, services, government...      3  \n",
"26   0.000216 women, program, programs, services, government...      3  \n",
```


Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"27      0.000012  women, program, programs, services, government...      4  \n",
"28      0.000014  women, program, programs, services, government...      3  \n",
"29      0.000012  women, program, programs, services, government...      4  \n",
"\n",
"      topic4_prob                                topic4_keywords topic5_id  \\\n",
"0      0.000154  women, government, care, australian, million, ...      2  \n",
"1      0.000004  women, budget, national, australia, page, la, ...      2  \n",
"2      0.000003  women, budget, national, australia, page, la, ...      2  \n",
"3      0.000048  women, program, programs, services, government...      2  \n",
"4      0.000055  women, program, programs, services, government...      2  \n",
"5      0.000062  women, program, programs, services, government...      2  \n",
"6      0.000060  women, program, programs, services, government...      2  \n",
"7      0.000036  women, program, programs, services, government...      2  \n",
"8      0.000046  women, program, programs, services, government...      2  \n",
"9      0.000044  women, program, programs, services, government...      2  \n",
"10     0.000366  women, program, programs, services, government...      2  \n",
"11     0.000385  women, program, programs, services, government...      2  \n",
"12     0.000035  women, budget, national, australia, page, la, ...      2  \n",
"13     0.000021  women, budget, national, australia, page, la, ...      2  \n",
"14     0.000374  women, program, programs, services, government...      2  \n",
"15     0.000630  women, program, programs, services, government...      3  \n",
"16     0.000033  women, community, communities, government, bud...      3  \n",
"17     0.000079  women, program, programs, services, government...      2  \n",
"18     0.000031  women, budget, national, australia, page, la, ...      2  \n",
"19     0.000013  women, budget, national, australia, page, la, ...      2  \n",
"20     0.000090  women, program, programs, services, government...      2  \n",
"21     0.000052  women, budget, national, australia, page, la, ...      2  \n",
"22     0.000037  women, budget, national, australia, page, la, ...      2  \n",
"23     0.000024  women, budget, national, australia, page, la, ...      2  \n",
"24     0.000028  women, budget, national, australia, page, la, ...      2  \n",
"25     0.000034  women, budget, national, australia, page, la, ...      2  \n",
"26     0.000012  women, budget, national, australia, page, la, ...      2  \n",
"27     0.000012  women, government, care, australian, million, ...      2  \n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"28    0.000014  women, budget, national, australia, page, la, ...      2    \"n\",
"29    0.000012  women, government, care, australian, million, ...    2    \"n\",
\"n\",
"      topic5_prob                                     topic5_keywords \"n\",
"0      0.000014  women, community, communities, government, bud... \"n\",
"1      0.000004  women, community, communities, government, bud... \"n\",
"2      0.000003  women, community, communities, government, bud... \"n\",
"3      0.000048  women, community, communities, government, bud... \"n\",
"4      0.000055  women, community, communities, government, bud... \"n\",
"5      0.000062  women, community, communities, government, bud... \"n\",
"6      0.000060  women, community, communities, government, bud... \"n\",
"7      0.000036  women, community, communities, government, bud... \"n\",
"8      0.000046  women, community, communities, government, bud... \"n\",
"9      0.000044  women, community, communities, government, bud... \"n\",
"10     0.000366  women, community, communities, government, bud... \"n\",
"11     0.000384  women, community, communities, government, bud... \"n\",
"12     0.000034  women, community, communities, government, bud... \"n\",
"13     0.000021  women, community, communities, government, bud... \"n\",
"14     0.000372  women, community, communities, government, bud... \"n\",
"15     0.000058  women, budget, national, australia, page, la, ... \"n\",
"16     0.000032  women, budget, national, australia, page, la, ... \"n\",
"17     0.000078  women, community, communities, government, bud... \"n\",
"18     0.000031  women, community, communities, government, bud... \"n\",
"19     0.000013  women, community, communities, government, bud... \"n\",
"20     0.000088  women, community, communities, government, bud... \"n\",
"21     0.000051  women, community, communities, government, bud... \"n\",
"22     0.000037  women, community, communities, government, bud... \"n\",
"23     0.000024  women, community, communities, government, bud... \"n\",
"24     0.000028  women, community, communities, government, bud... \"n\",
"25     0.000033  women, community, communities, government, bud... \"n\",
"26     0.000012  women, community, communities, government, bud... \"n\",
"27     0.000012  women, community, communities, government, bud... \"n\",
"28     0.000014  women, community, communities, government, bud... \"n\",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"29      0.000011  women, community, communities, government, bud...  \n",
"\n",
"[30 rows x 142 columns]\n"
]
}
],
"source": [
"for k in range(1, 6): # 1..5\n",
"    df[f\"topic{k}_id\"] = -1\n",
"    df[f\"topic{k}_prob\"] = 0.0\n",
"    df[f\"topic{k}_keywords\"] = \"\"\n",
"\n",
"for idx, bow in enumerate(corpus):\n",
"    # get topic distribution for this row\n",
"    topic_probs = lda_model.get_document_topics(bow, minimum_probability=0)\n",
"\n",
"    # sort by probability, highest first\n",
"    topic_probs = sorted(topic_probs, key=lambda x: x[1], reverse=True)\n",
"\n",
"    # take top 5 (or fewer if you used <5 global topics)\n",
"    top5 = topic_probs[:5]\n",
"\n",
"    for rank, (topic_id, prob) in enumerate(top5, start=1):\n",
"        # topic ID and probability\n",
"        df.at[idx, f\"topic{rank}_id\"] = topic_id\n",
"        df.at[idx, f\"topic{rank}_prob\"] = prob\n",
"\n",
"        # top words for this topic\n",
"        keywords = \"\", \".join(\n",
"            w for w, p in lda_model.show_topic(topic_id, topn=10)\n",
"        )\n",
"        df.at[idx, f\"topic{rank}_keywords\"] = keywords\n",
"\n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"df.to_csv(r"C:\\Users\\xxy230002\\Desktop\\Data updated\\Women Budget
Project\\WBP_frenquency_topic.csv", index=False)\n",
    "print(df)"
]
},
{
    "cell_type": "code",
    "execution_count": 42,
    "id": "2382fe59-8373-41c3-9d72-714f480bd9da",
    "metadata": {},
    "outputs": [
        {
            "name": "stdout",
            "output_type": "stream",
            "text": [
                "      year  total_words top1  top1_count      top2  top2_count      top3  \\\n",
                "0      1984        24902  the        1377      of        1120      to  \n",
                "1      1985        90431  the        6158      of        4242      and  \n",
                "2      1986       102244  the        6626      of        4449      and  \n",
                "3      1987         6664  of          939      the         627  budget  \n",
                "4      1988         5908  of          831      the         555  budget  \n",
                "5      1989         5236  of          735      the         491  budget  \n",
                "6      1990         5404  of          759      the         507  budget  \n",
                "7      1991         9016  of        1275      the         851  budget  \n",
                "8      1992         7042  of          993      the         663  budget  \n",
                "9      1993         7378  of        1041      the         695  budget  \n",
                "10     1994          868  of          111      the          75  budget  \n",
                "11     1995          826  of          105      the          71  budget  \n",
                "12     1997         9958  the         601      to         374      and  \n",
                "13     1998        16103  the        1074      and         681      of  \n",
                "14     1999          767  la           53  library         53      of  \n",
                "15     2000         5633  the         284      and         266      to  \n",
                "16     2001        10404  the         547      and         494      to  \n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"17	2002	4272	the	255	of	185	and	\n",
"18	2003	10447	the	587	and	407	of	\n",
"19	2005	24844	the	1464	and	1060	to	\n",
"20	2008	3628	the	204	and	168	to	\n",
"21	2009	6301	the	337	and	264	to	\n",
"22	2010	8793	the	539	and	347	to	\n",
"23	2011	13401	the	599	and	594	to	\n",
"24	2012	11900	and	565	the	556	to	\n",
"25	2013	9810	the	493	and	410	to	\n",
"26	2021	26680	the	1371	and	1236	to	\n",
"27	2022October	27537	and	1314	the	1292	to	\n",
"28	2022March	23918	the	1189	and	1127	to	\n",
"29	2023	27577	the	1295	and	1251	to	\n",

"\n",

"	top3_count	top4	top4_count	...	\\n",
"0	823	and	800	...	\n",
"1	3437	to	2740	...	\n",
"2	4060	to	3089	...	\n",
"3	625	women	624	...	\n",
"4	553	women	552	...	\n",
"5	489	women	488	...	\n",
"6	505	women	504	...	\n",
"7	849	women	848	...	\n",
"8	661	women	660	...	\n",
"9	693	women	692	...	\n",
"10	73	women	72	...	\n",
"11	69	women	68	...	\n",
"12	366	of	361	...	\n",
"13	664	to	566	...	\n",
"14	52	for	52	...	\n",
"15	195	of	189	...	\n",
"16	453	of	362	...	\n",
"17	161	to	146	...	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"18      378      to      359 ...  \n",
"19      900      of      785 ...  \n",
"20      127      of      103 ...  \n",
"21      217      of      185 ...  \n",
"22      302      of      276 ...  \n",
"23      480      of      413 ...  \n",
"24      424      of      369 ...  \n",
"25      336      of      271 ...  \n",
"26      992  women      818 ...  \n",
"27      937      of      894 ...  \n",
"28      959      of      681 ...  \n",
"29      950      of      822 ...  \n",
"\n",
"
                                topic2_keywords  topic3_id  topic3_prob
\\n",
"0  women, government, support, violence, per, hea...      3  0.000458  \n",
"1  women, government, support, violence, per, hea...      4  0.000014  \n",
"2  women, government, support, violence, per, hea...      4  0.000214  \n",
"3  women, government, support, violence, per, hea...      4  0.000048  \n",
"4  women, government, support, violence, per, hea...      4  0.000055  \n",
"5  women, government, support, violence, per, hea...      4  0.000062  \n",
"6  women, government, support, violence, per, hea...      4  0.000060  \n",
"7  women, government, support, violence, per, hea...      4  0.000036  \n",
"8  women, government, support, violence, per, hea...      4  0.000046  \n",
"9  women, government, support, violence, per, hea...      4  0.000044  \n",
"10 women, government, care, australian, million, ...      0  0.000366  \n",
"11 women, government, care, australian, million, ...      0  0.000385  \n",
"12 women, government, support, violence, per, hea...      1  0.175172  \n",
"13 women, program, programs, services, government...      0  0.000730  \n",
"14 women, government, care, australian, million, ...      0  0.000375  \n",
"15 women, community, communities, government, bud...      0  0.043232  \n",
"16 women, government, support, violence, per, hea...      1  0.054883  \n",
"17 women, budget, national, australia, page, la, ...      0  0.000079  \n",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

"18	women, program, programs, services, government...	0	0.000333	\n",
"19	women, program, programs, services, government...	0	0.026290	\n",
"20	women, government, support, violence, per, hea...	3	0.000090	\n",
"21	women, government, care, australian, million, ...	1	0.093399	\n",
"22	women, government, support, violence, per, hea...	1	0.000365	\n",
"23	women, government, care, australian, million, ...	1	0.027505	\n",
"24	women, government, support, violence, per, hea...	1	0.000029	\n",
"25	women, government, care, australian, million, ...	1	0.009380	\n",
"26	women, government, care, australian, million, ...	1	0.000216	\n",
"27	women, budget, national, australia, page, la, ...	1	0.000012	\n",
"28	women, government, care, australian, million, ...	1	0.000014	\n",
"29	women, budget, national, australia, page, la, ...	1	0.000012	\n",
"\n",				
"	topic3_keywords topic4_id topic4_prob \\n",			
"0	women, budget, national, australia, page, la, ...	4	0.000154	\n",
"1	women, government, care, australian, million, ...	3	0.000004	\n",
"2	women, government, care, australian, million, ...	3	0.000003	\n",
"3	women, government, care, australian, million, ...	1	0.000048	\n",
"4	women, government, care, australian, million, ...	1	0.000055	\n",
"5	women, government, care, australian, million, ...	1	0.000062	\n",
"6	women, government, care, australian, million, ...	1	0.000060	\n",
"7	women, government, care, australian, million, ...	1	0.000036	\n",
"8	women, government, care, australian, million, ...	1	0.000046	\n",
"9	women, government, care, australian, million, ...	1	0.000044	\n",
"10	women, government, support, violence, per, hea...	1	0.000366	\n",
"11	women, government, support, violence, per, hea...	1	0.000385	\n",
"12	women, program, programs, services, government...	3	0.000035	\n",
"13	women, government, support, violence, per, hea...	3	0.000021	\n",
"14	women, government, support, violence, per, hea...	1	0.000374	\n",
"15	women, government, support, violence, per, hea...	1	0.000630	\n",
"16	women, program, programs, services, government...	2	0.000033	\n",
"17	women, government, support, violence, per, hea...	1	0.000079	\n",
"18	women, government, support, violence, per, hea...	3	0.000031	\n",

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"19 women, government, support, violence, per, hea... 3 0.000013 \n",
"20 women, budget, national, australia, page, la, ... 1 0.000090 \n",
"21 women, program, programs, services, government... 3 0.000052 \n",
"22 women, program, programs, services, government... 3 0.000037 \n",
"23 women, program, programs, services, government... 3 0.000024 \n",
"24 women, program, programs, services, government... 3 0.000028 \n",
"25 women, program, programs, services, government... 3 0.000034 \n",
"26 women, program, programs, services, government... 3 0.000012 \n",
"27 women, program, programs, services, government... 4 0.000012 \n",
"28 women, program, programs, services, government... 3 0.000014 \n",
"29 women, program, programs, services, government... 4 0.000012 \n",
"\n",
"
                                topic4_keywords topic5_id topic5_prob
\\n",
"0 women, government, care, australian, million, ... 2 0.000014 \n",
"1 women, budget, national, australia, page, la, ... 2 0.000004 \n",
"2 women, budget, national, australia, page, la, ... 2 0.000003 \n",
"3 women, program, programs, services, government... 2 0.000048 \n",
"4 women, program, programs, services, government... 2 0.000055 \n",
"5 women, program, programs, services, government... 2 0.000062 \n",
"6 women, program, programs, services, government... 2 0.000060 \n",
"7 women, program, programs, services, government... 2 0.000036 \n",
"8 women, program, programs, services, government... 2 0.000046 \n",
"9 women, program, programs, services, government... 2 0.000044 \n",
"10 women, program, programs, services, government... 2 0.000366 \n",
"11 women, program, programs, services, government... 2 0.000384 \n",
"12 women, budget, national, australia, page, la, ... 2 0.000034 \n",
"13 women, budget, national, australia, page, la, ... 2 0.000021 \n",
"14 women, program, programs, services, government... 2 0.000372 \n",
"15 women, program, programs, services, government... 3 0.000058 \n",
"16 women, community, communities, government, bud... 3 0.000032 \n",
"17 women, program, programs, services, government... 2 0.000078 \n",
"18 women, budget, national, australia, page, la, ... 2 0.000031 \n",
```


Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"19 women, budget, national, australia, page, la, ...      2      0.000013      \"n\",
"20 women, program, programs, services, government...      2      0.000088      \"n\",
"21 women, budget, national, australia, page, la, ...      2      0.000051      \"n\",
"22 women, budget, national, australia, page, la, ...      2      0.000037      \"n\",
"23 women, budget, national, australia, page, la, ...      2      0.000024      \"n\",
"24 women, budget, national, australia, page, la, ...      2      0.000028      \"n\",
"25 women, budget, national, australia, page, la, ...      2      0.000033      \"n\",
"26 women, budget, national, australia, page, la, ...      2      0.000012      \"n\",
"27 women, government, care, australian, million, ...      2      0.000012      \"n\",
"28 women, budget, national, australia, page, la, ...      2      0.000014      \"n\",
"29 women, government, care, australian, million, ...      2      0.000011      \"n\",
\"n\",
\"
                                topic5_keywords      \"n\",
"0  women, community, communities, government, bud...      \"n\",
"1  women, community, communities, government, bud...      \"n\",
"2  women, community, communities, government, bud...      \"n\",
"3  women, community, communities, government, bud...      \"n\",
"4  women, community, communities, government, bud...      \"n\",
"5  women, community, communities, government, bud...      \"n\",
"6  women, community, communities, government, bud...      \"n\",
"7  women, community, communities, government, bud...      \"n\",
"8  women, community, communities, government, bud...      \"n\",
"9  women, community, communities, government, bud...      \"n\",
"10 women, community, communities, government, bud...      \"n\",
"11 women, community, communities, government, bud...      \"n\",
"12 women, community, communities, government, bud...      \"n\",
"13 women, community, communities, government, bud...      \"n\",
"14 women, community, communities, government, bud...      \"n\",
"15 women, budget, national, australia, page, la, ...      \"n\",
"16 women, budget, national, australia, page, la, ...      \"n\",
"17 women, community, communities, government, bud...      \"n\",
"18 women, community, communities, government, bud...      \"n\",
"19 women, community, communities, government, bud...      \"n\",
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"20 women, community, communities, government, bud... \n",
"21 women, community, communities, government, bud... \n",
"22 women, community, communities, government, bud... \n",
"23 women, community, communities, government, bud... \n",
"24 women, community, communities, government, bud... \n",
"25 women, community, communities, government, bud... \n",
"26 women, community, communities, government, bud... \n",
"27 women, community, communities, government, bud... \n",
"28 women, community, communities, government, bud... \n",
"29 women, community, communities, government, bud... \n",
"\n",
"[30 rows x 140 columns]\n"
]
}
],
"source": [
    "df = pd.read_csv(r\"C:\\\\Users\\\\xxy230002\\\\Desktop\\\\Data updated\\\\Women Budget
Project\\\\WBP_frenquency_topic.csv\")\n",
    "df = df.drop(df.columns[1], axis=1)\n",
    "df = df.drop(columns=[\"tokens\"])\n",
    "print (df)\n",
    "df.to_csv(r\"C:\\\\Users\\\\xxy230002\\\\Desktop\\\\Data updated\\\\Women Budget
Project\\\\Simple_WBP_frenquency_topic.csv\", index=False)"
]
},
{
    "cell_type": "code",
    "execution_count": null,
    "id": "00ae85ee-6edf-41bb-b703-191613c1c3d4",
    "metadata": {},
    "outputs": [],
    "source": []
}
],
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025

```
"metadata": {
  "kernelspec": {
    "display_name": "Python 3 (ipykernel)",
    "language": "python",
    "name": "python3"
  },
  "language_info": {
    "codemirror_mode": {
      "name": "ipython",
      "version": 3
    },
    "file_extension": ".py",
    "mimetype": "text/x-python",
    "name": "python",
    "nbconvert_exporter": "python",
    "pygments_lexer": "ipython3",
    "version": "3.10.16"
  }
},
"nbformat": 4,
"nbformat_minor": 5
```

Subject: [EPPS6302] Methods of Data Collection and Production

Recipient: Dr. Carl Ho

Date: 12. 05. 2025