

# 1 여성리더십 텍스트마이닝 7차분석 리포트 \*

2 **채충일**

3 **조연주, 유지은, 이수미**

4 **Acknowledgement to WKU BA Lab Students: Allen, Owen, V, Claudia, Jason, Felix, Jenny,**

5 **Arya**

---

6 이 연구는 한국 여성 리더십에 대한 기존 10년 동안 수행된 여덟 개의 질적 조사를 통합하고 컴퓨터 지원  
7 텍스트 마이닝 기술을 활용하여 숨겨진 연구 주제와 통찰력을 발견하려는 목적을 가지고 있습니다. 200  
8 개의 인터뷰 데이터를 기반으로 한 이 연구는 주제 모델링 및 계층적 클러스터링 기술을 사용하여 연구  
9 결과를 분석하고 타당성을 검증합니다. 이론적 기반과 텍스트 마이닝을 결합하여 여성 리더십에 대한  
10 심층적인 이해를 제공하며 학문적 및 실용적 기여를 동시에 추구하는것을 목적으로 합니다.

11 *Keywords:* 분석보고서, 텍스트마이닝, 토픽모델링

---

## 12 Contents

13	<b>노트</b>	<b>2</b>
14	작업내용 개요 . . . . .	2
15	제안 및 향후단계 . . . . .	2
16	<b>서론</b>	<b>3</b>
17	<b>방법론적 프로시저</b>	<b>4</b>
18	<b>결과</b>	<b>5</b>
19	데이터 . . . . .	5
20	데이터 . . . . .	5
21	어근추출 데이터 . . . . .	5
22	스탑워드 . . . . .	5
23	기본분석 . . . . .	5
24	최빈단어 분석 . . . . .	5
25	워드클라우드 . . . . .	7
26	각 메타데이터(변수)에 따른 최빈단어 비교 . . . . .	8
27	결혼상태 . . . . .	8
28	아이유무 . . . . .	9
29	직종별 . . . . .	10
30	KWIC(Key Word In Context) 분석 . . . . .	10
31	주요단어와 연관성이 높은 단어들 . . . . .	12
32	Keyword Network . . . . .	13
33	토픽모델링 . . . . .	14
34	토픽모델 : 하한 단어갯수 3, 3389단어, 30개 토픽 . . . . .	15
35	모델핏 . . . . .	16

---

\*Replication files are available on the author's Github account (<http://github.com/chadchae>). | Corresponding author: [chadchae@gmail.com](mailto:chadchae@gmail.com) | Submission to : TBA

36	토픽결과 . . . . .	17
37	수정 레이블 . . . . .	22
38	대안 레이블 . . . . .	23

### 39 노트

#### 40 작업내용 개요

#### 41 제안 및 향후단계

- 42 • 다음 미팅시 까지 레이블링 확정
- 43 • 레이블 확정시 레이블의 의미에 대해 토론 및 정리 필요
- 44 • Nelson (2020) 및 논문들을 읽고 연구 결과 및 프로시저를 어떻게 일치시킬것인가에 대한 논의도
- 45 필요함.
- 46 • 레이블 확정이후 추가분석(토픽 출현확률 상대적인 양에 대한 제시된 변수들에 따른 차이비교)
- 47 • 글쓰기 시작 전 전반적인 파인딩들에 대해 토론과 글쓰기에서 어떻게 반영할것인지에 대한 논의
- 48 필요

## 49 서론

- 50 • 이 연구의 제목은 “한국 여성 리더십에 대한 통찰력 종합 분석: 컴퓨터 지원 텍스트 마이닝 기술을  
51 통한 체계적인 연구” 로써 지난 10년간 한국의 여성 리더십에 대한 8편의 질적연구에서 얻어진  
52 200개의 인터뷰 데이터에 컴퓨터 지원 텍스트 마이닝 기술기반으로 체계적인 분석을 제시하려  
53 합니다. 이 연구는 기존 연구 결과를 통합하고 고급 분석 방법을 사용하며 질적연구와 컴퓨터이셔  
54 널 그라운드드 피어리 절차에 따른 방법을 통해 기존 연구들의 종합분석 및 다량의 텍스트분석에  
55 대한 인지의 한계를 넘어 이전에 간과된 잠재적인 연구 주제와 통찰력을 발굴하는 것을 목적으로  
56 합니다.
- 57 • 이 연구는 한국 여성 리더십 연구의 개요, 사용된 방법 및 데이터 분석 가능성을 포함한 종합적인  
58 개요에 대한 필요성을 다룹니다. 연구 동기는 200개의 인터뷰 대본을 축적한 결과에서 비롯되며,  
59 이는 풍부한 질적 데이터의 원천입니다. 분석 방법을 향상시키기 위해 연구팀은 주제 모델링 기술  
60 을 적용하여 이전에 간과된 연구 주제와 테마적 요소를 식별하고자 합니다.
- 61 • 이 연구는 Nelson (2020) 가 제시한 텍스트 마이닝, 주제 모델링 및 계층적 클러스터링 기술을 등  
62 을 포함하는 컴퓨터이셔널 퀄리티에티브 분석방식을 도입하여 200개의 인터뷰 대본으로 구성된  
63 데이터셋에 대한 분석을 시도합니다.
- 64 • 데이터셋은 익명화 및 메타데이터 생성을 거칩니다. 이 메타데이터에는 인터뷰 대상자 식별자, 산  
65 업 소속, 직위, 근무 기간, 가족 구성원 및 지원 구조가 포함되어 있어 인구 특성에 기반한 패턴  
66 분석이 가능합니다.
- 67 • 연구팀은 먼저 주제 모델링 기술을 대본의 대표적인 하위 집합에 적용하여 데이터와 호환성을 확  
68 인합니다. 메타데이터의 불일치 사항은 데이터 전체에 주제 모델링을 적용하기 전에 해결됩니다.  
69 질적 연구 결과와 비교를 통해 주제 모델링 결과의 타당성과 신뢰성을 평가하기 위해 다학제적  
70 연구자 협의가 진행됩니다.
- 71 • 데이터 분석 프로토콜은 데이터 통합 및 재구성, 데이터 정제, 전처리 및 모델링 네 가지 주요 단  
72 계로 구분됩니다. 텍스트 대본은 텍스트 파일로 변환되어 엑셀 환경 내에서 데이터 프레임으로  
73 구조화됩니다. 두 번째 단계는 데이터 유형의 일치 여부를 확인하고 데이터의 공백 여부를 확인하  
74 는 데이터 정제를 포함합니다. 전처리 단계는 텍스트를 분석 가능한 말뭉치로 변환하는 과정으로,  
75 n-gram 생성, 불용어 필터링 및 품사 태깅과 같은 기술을 사용하여 토큰화된 단어 목록을 작성  
76 합니다. 마지막으로 Latent Dirichlet Allocation (LDA)를 기반으로 한 구조화된 주제 모델링이  
77 적용되어 주제와 관련된 확률을 제공합니다.
- 78 • 이론적/개념적 프레임워크: 연구는 관련 주제, 이론 및 모델 논문을 통합하여 분석에 대한 이론적  
79 기반을 제공합니다. 종합적인 시각에서의 분석에서 나온 연구 결과를 기존 출판된 연구 결과와  
80 비교함으로써, 연구는 지난 10년간 학술 논문에서 발표된 문학을 체계적으로 통합합니다. 텍스트  
81 마이닝 방법은 질적 연구의 어휘 패턴과 테마적 트렌드를 깊이 있게 조사함으로써 한국의 여성  
82 리더십에 대한 우리의 이해를 높이는 데 기여하려합니다.
- 83 • 이 연구는 풍부한 질적 인터뷰 대본 데이터에 고급 텍스트 마이닝 기술을 적용하여 한국의 여성  
84 리더십에 대한 통찰력을 종합적으로 제시합니다. 계산 방법과 해석 분석을 결합하는 다학제적 접  
85 근은 편향되지 않고 견고하며 재현 가능한 연구 주제를 도출할 것으로 기대됩니다. 이 연구 결과는  
86 학문적 기여를 뿐만 아니라 한국의 여성 리더십 분야에서의 실용적 적용에도 기여할 것입니다.

## 87 방법론적 프로시저

## 88 • 전처리

- 89 - 데이터 합치기
- 90 - 무결성 확인
- 91 - 코드북 생성 및 메타데이터 확인
- 92 - 익명화관련 코드 및 데이터셋 생성
- 93 - 한글문서에 대한 전처리

- 94 \* 클리닝
- 95 \* 토큰나이징
- 96 \* 스탑워드

## 97 • 디스크립티브 분석

- 98 - 최빈단어 뽑기
- 99 - [추가분석시] TF-IDF 단어스코어
- 100 - [추가분석시] Log-Odd 스코어
- 101 - 워드클라우드
- 102 - 각 메타데이터(변수)에 따른 최빈단어 비교

- 103 \* 결혼상태
- 104 \* 아이유무
- 105 \* 직종별

## 106 - KWIC(Key Word In Context) 분석

- 107 \* 성공
- 108 \* 경력
- 109 \* 육아
- 110 \* 리더십 여성

## 111 - Keyword Network

## 112 • 토픽모델링 (STM, Structural Topic Modeling)

## 113 - Model 2:

- 114 \* 하한 단어갯수 3, 3389단어
- 115 \* 30개 토픽

## 116 결과

## 117 데이터

## 118 데이터

- 119 • 총 199개 문서에서 612개의 문단으로 나누어진 데이터셋 확보.
- 120 • 문서레벨에서 199개 분석
- 121 • 단어 및 토픽레벨에서 분석에 612개 observation 사용
- 122 • 총 20개의 변수
- 123 • 텍스트와 토큰나이징된 텍스트 포함
- 124 • 각 변수별 상세 사항은 코드북 참조

## 125 어근추출 데이터

- 126 • 파이썬 KoNLP와 ko-Mecab라이브러리를 활용하여 명사들만 추출
- 127 • 추출된 명사들은 각 문서(문단)에 맞게 재배치

## 128 스타트업워드

- 129 • 2023년 11월 24일 합의된 1004개의 스타트업워드
- 130 • 인터뷰 응답자의 이름 및 일터들의 이름을 추출
- 131 • 총 1322개의 스타트업워드를 포함한 데이터셋을 활용 - 생각, 사람, 회사, 들이, 우리, 저희, 때문, 얘기, 부분, 정도, 사실, 그것, 해서, 말씀...
- 132

## 133 기본분석

## 134 최빈단어 분석

- 135 • 여자 (2974회): 여성 리더십에 대한 논의는 주로 여성의 역할과 위치에 초점을 맞춥니다.
- 136 • 여성 (2213회): '여자'와 유사하게, 여성 리더십의 논의는 여성의 사회적, 전문적 역할에 집중합니
- 137 다.
- 138 • 교수 (2199회): 여성이 학계에서의 리더십 역할, 특히 교수직에서의 위치를 차지하는 빈도가 높음
- 139 을 시사합니다.
- 140 • 남편 (1949회): 여성 리더십과 개인적인 삶, 특히 결혼과 가정 생활이 어떻게 상호작용하는지에
- 141 대한 관심을 반영합니다.
- 142 • 교육 (1856회): 여성 리더십과 교육의 중요성을 강조합니다. 교육은 여성이 리더십 역할을 수행하
- 143 는 데 중요한 기반이 됩니다.
- 144 • 남자 (1804회): 여성 리더십에 대한 논의는 종종 남성 리더십과의 비교를 통해 이루어집니다.
- 145 • 공부 (1709회): 학문적 노력과 지속적인 학습이 여성 리더십 발달에 중요함을 나타냅니다.
- 146 • 시간 (1685회): 여성 리더로서의 역할과 시간 관리, 워크-라이프 밸런스의 중요성을 시사합니다.
- 147 • 그렇 (1674회): 이 단어는 추가분석시 스타트업워드로 분류하는것이 맞을것 같습니다
- 148 • 학교 (1669회): 교육 기관에서의 여성 리더십의 중요성과 영향력을 나타냅니다.
- 149 • 아이 (1553회): 여성 리더십과 육아의 병행, 그리고 이로 인한 도전과 기회를 시사합니다.
- 150 • 직원 (1509회): 조직 내에서의 여성 리더십과 직장 내 여성의 역할을 나타냅니다.
- 151 • 성공 (1485회): 여성 리더십과 성공 사이의 연관성에 대한 탐구를 반영합니다.
- 152 • 엄마 (1381회): 여성 리더십과 가정 생활, 특히 모성 역할과의 균형에 대한 논의를 나타냅니다.
- 153 • 박사 (1235회): 학문적 성취와 여성 리더십 간의 관계를 강조합니다.
- 154 • 힘들 (1229회): 여성 리더십의 경로와 그 과정에서 겪는 어려움을 시사합니다.
- 155 • 대학 (1177회): 고등 교육 기관에서의 여성 리더십과 관련된 논의를 반영합니다.

- 156 • 중요 (1176회): 여성 리더십의 중요성과 그 영향력을 강조합니다.
- 157 • 조직 (1171회): 조직 내에서의 여성의 역할과 리더십의 중요성을 나타냅니다.
- 158 • 필요 (1144회): 여성 리더십 발전과 그 필요성에 대한 인식을 나타냅니다.

159 이러한 키워드들은 여성 리더십에 대한 다양한 측면과 그 중요성, 도전과 기회들을 포괄적으로 반영  
 160 하고 있습니다. 여성이 사회적, 전문적 영역에서 중요한 위치를 차지하며, 그들의 리더십이 교육, 가정,  
 161 직장 등 다양한 분야에서 중요한 역할을 하고 있음을 보여줍니다. 단어의 빈도는 그 자체로도 중요한  
 162 의미를 지니고 있지만 때로는 의미를 편향되게 이해하게 될 위험성도 존재합니다. 때문에 추후 자세한  
 163 분석들에 적용에서 이런 이런 단어들이 중요하게 자주 빈출되었다 정도로 이해하면 좋을듯 싶습니다.

### 상위 20개 최빈 단어

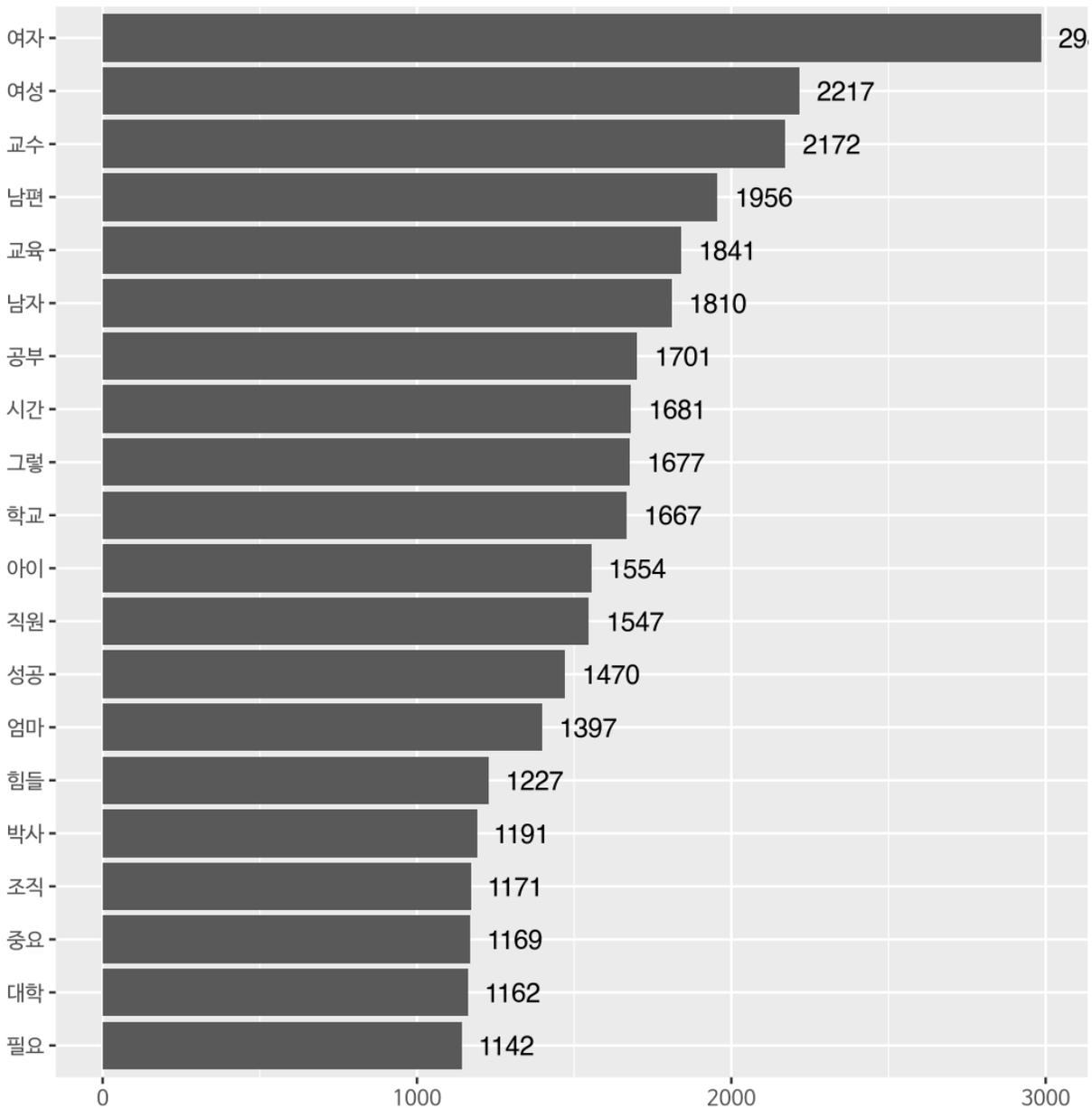


Figure 1: 20개 최빈단어



167 각 메타데이터(변수)에 따른 최빈단어 비교

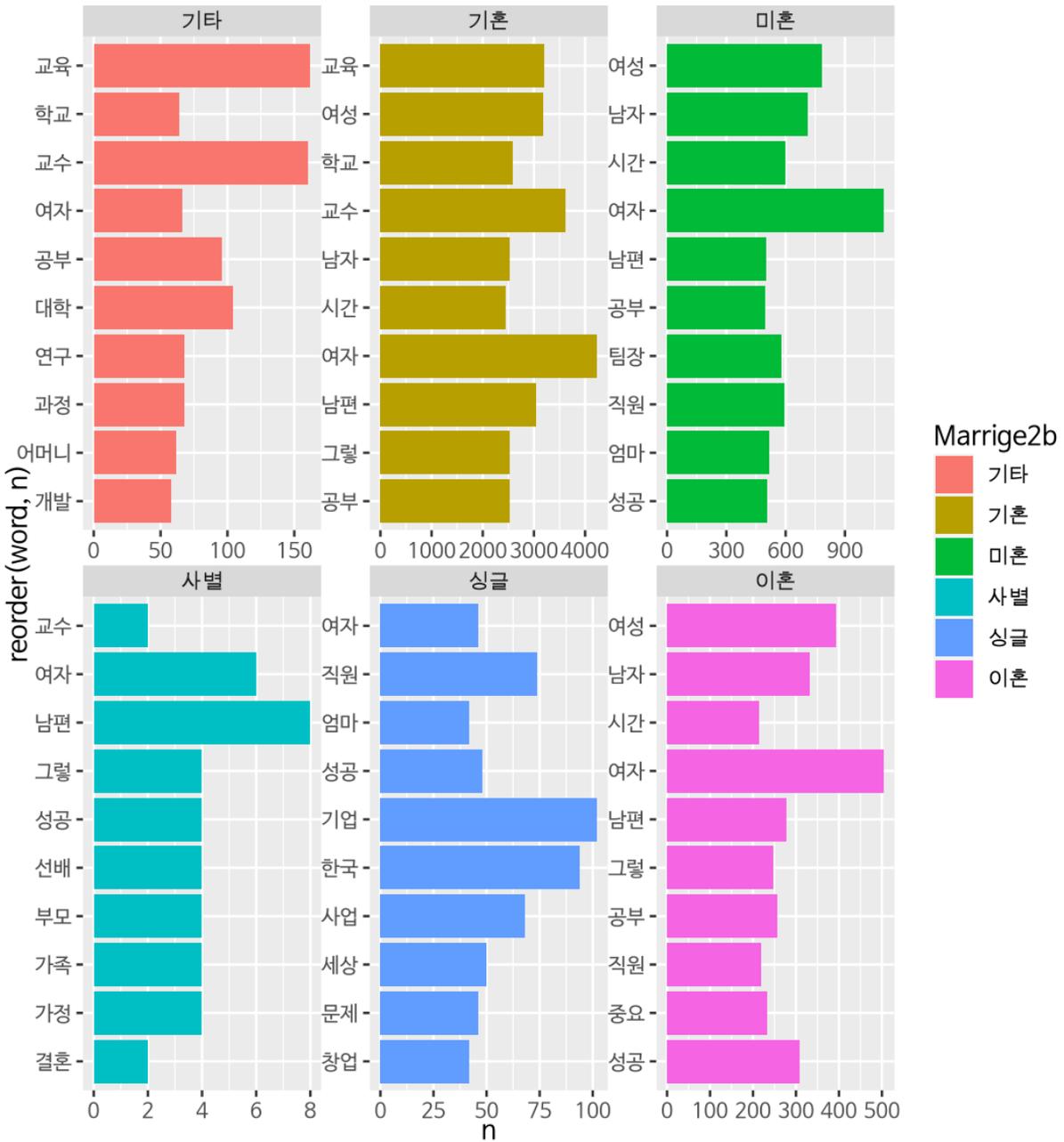


Figure 3: 변수에 따른 주요단어확인: 결혼상태

168 결혼상태

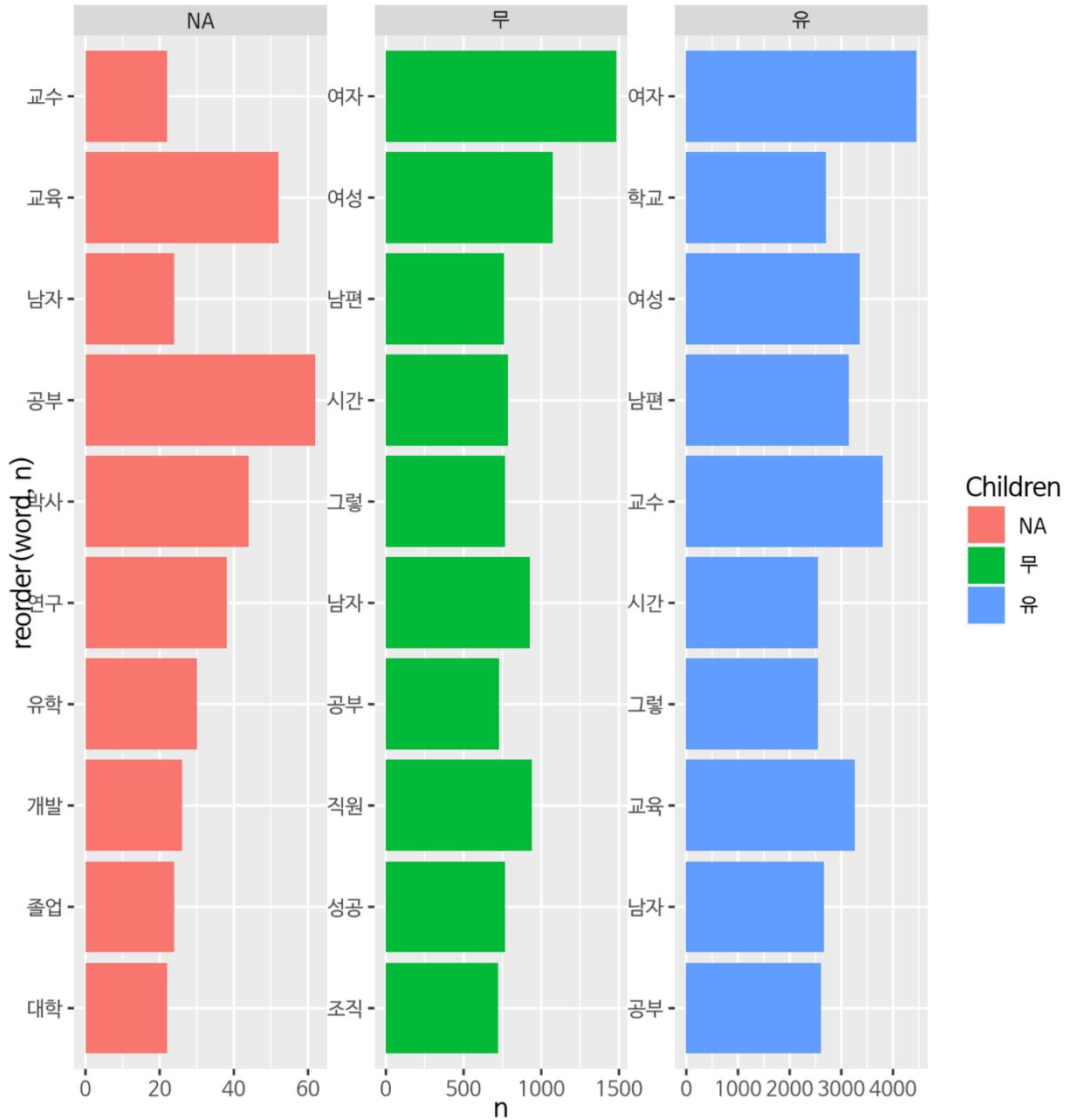


Figure 4: 변수에 따른 주요단어확인: 아이유무

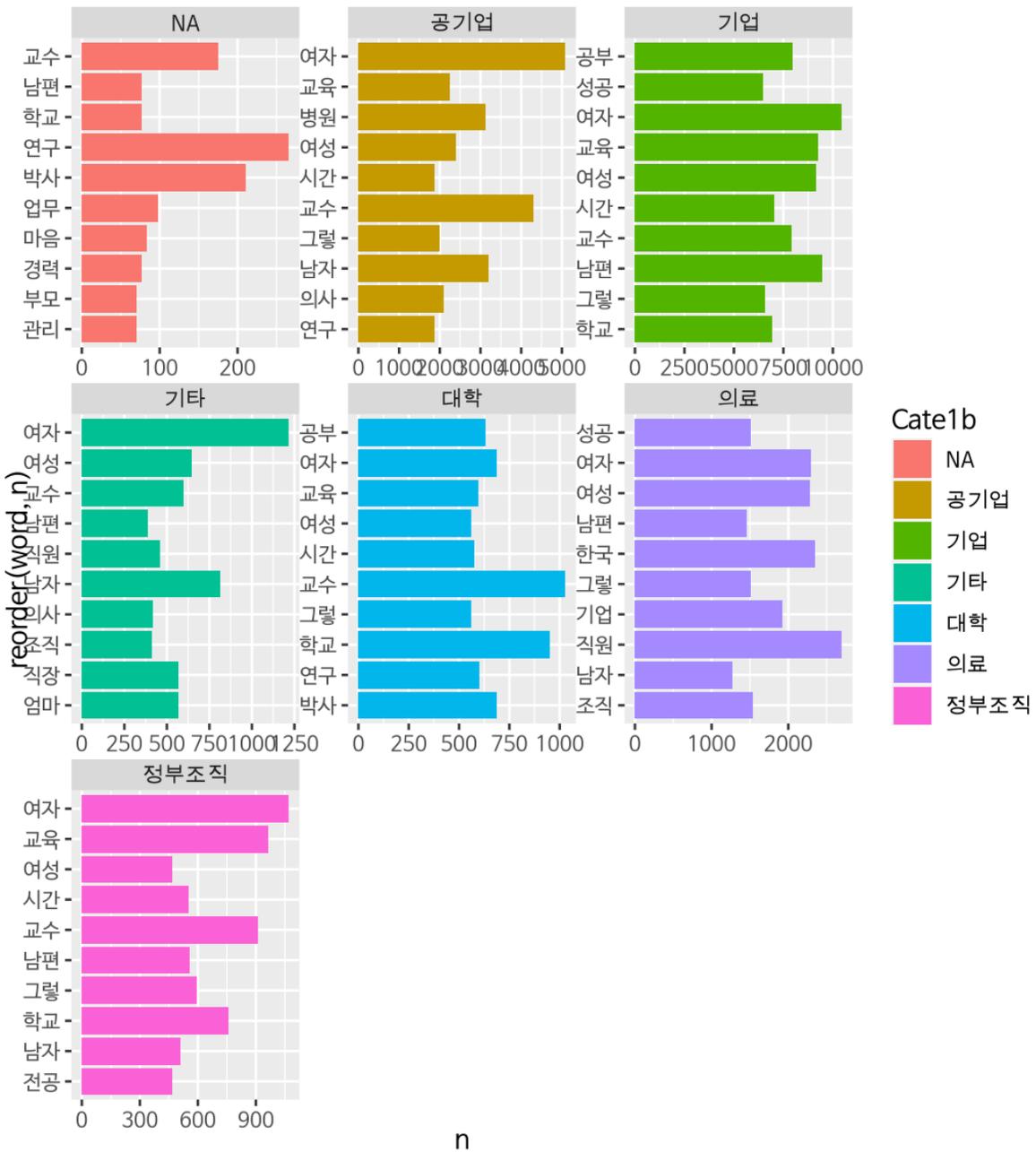


Figure 5: 변수에 따른 주요단어확인: 직종

170 직종별

171 KWIC(Key Word In Context) 분석

172 KWIC분석은 Key Word In Context의 약자로서 제시된 단어들이 어떤 맥락에서 사용되었는지를 보여  
 173 줍니다. 이를 통해 단어의 사용과 그 맥락에 대한 이해에 도움이 됩니다.

174 ## Keyword-in-context with 5 matches.

175 ## [text1, 809] 엄마 엄마 전문직 여성 편안 | 성공 | 사회 상태 사회 사회 장애물

```

176 ## [text1, 819] 장애물 극복 사회 극복 성취 | 성공 | 성공 성공 열매 주관 성공
177 ## [text1, 820]   극복 사회 극복 성취 성공 | 성공 | 성공 열매 주관 성공 성공
178 ## [text1, 821]   사회 극복 성취 성공 성공 | 성공 | 열매 주관 성공 성공 박사
179 ## [text1, 824]   성공 성공 성공 열매 주관 | 성공 | 성공 박사 고속 승격 성공

180 ## Keyword-in-context with 5 matches.
181 ## [text2, 634]   입주 보모 워킹 컨설팅 준비 | 경력 |
182 ## [text2, 879]   시간 성공 개인 경영 관리 | 경력 |
183 ## [text13, 128]  고객 프로필 인맥 달인 달인 | 경력 |
184 ## [text24, 549]  연구원 기술 일반직 기술 백그라운드 | 경력 |
185 ## [text24, 1949] 대학교 대학교 리더십 중요 여성 | 경력 |
186 ##
187 ## 힘들 공수 월화 수목 금금
188 ## 지원 단축 근무 업무 심장
189 ## 착각 기업 외환 사업 종합
190 ## 향후 진로 고민 연구 개발
191 ## 관리 필요 남자 마투 네트

192 ## Keyword-in-context with 5 matches.
193 ## [text2, 232]   제도 시행 이유 아이 어린이 | 육아 |
194 ## [text2, 434]   아이 이모 퇴근 시간 퇴근 | 육아 |
195 ## [text2, 542]   아이 루마니아 문제 아이 아이 | 육아 |
196 ## [text2, 683]   나이 아이 덜하 승진 역량 | 육아 |
197 ## [text2, 871]   팀장 팀장 단계 리더십 기회 | 육아 |
198 ##
199 ## 아이 아이 양해 퇴근 정규
200 ## 독립군 어머니 시댁 신랑 약속
201 ## 정석 과정 이모 숙식 조건
202 ## 업무 컨설팅 지원 배려 제도
203 ## 힘들 퇴근 시간 성공 개인

204 ## Keyword-in-context with 5 matches.
205 ## [text2, 777]   지식 필요 리더 빠르 승진 | 리더십 | 후배 연습 기회 핸드폰 사업
206 ## [text2, 851]   기미 부정 생활 즐겁 업무 | 리더십 | 연습 절호 기회 슬프 구성원
207 ## [text2, 869]   신입 서무 팀장 팀장 단계 | 리더십 | 기회 육아 힘들 퇴근 시간
208 ## [text7, 5]    극복 인색 시간 여자 | 리더십 |
209 ## [text10, 4]   이론 직원 공포 | 리더십 | 기준 정체 부인 판단 시기

210 ## Keyword-in-context with 5 matches.
211 ## [text1, 67]    아이 시간 어떻 박사 특별 | 여성 |
212 ## [text1, 232]  화랑 선제공격 위치 터프 업적 | 여성 |
213 ## [text1, 497]  퍼포먼스 남녀 조직 관리 남성 | 여성 |
214 ## [text1, 501]  남성 여성 사고방식 남성 사냥 | 여성 |
215 ## [text1, 807]  대학 시스템 엄마 엄마 전문직 | 여성 |
216 ##
217 ## 직속 노트 유동 시간 근무
218 ## 굴레 남녀 전공 물리학 연극
219 ## 사고방식 남성 사냥 여성 채집
220 ## 채집 남성 사냥 걱정 집형
221 ## 편안 성공 사회 상태 사회

```

222 주요단어와 연관성이 높은 단어들

- 223 • 단어들 간의 연관성을 확인하기 위해 6개의 주요 단어가 제시되었고 각 단어들과 연관성이 높은  
 224 단어 10씩을 제시하였습니다.

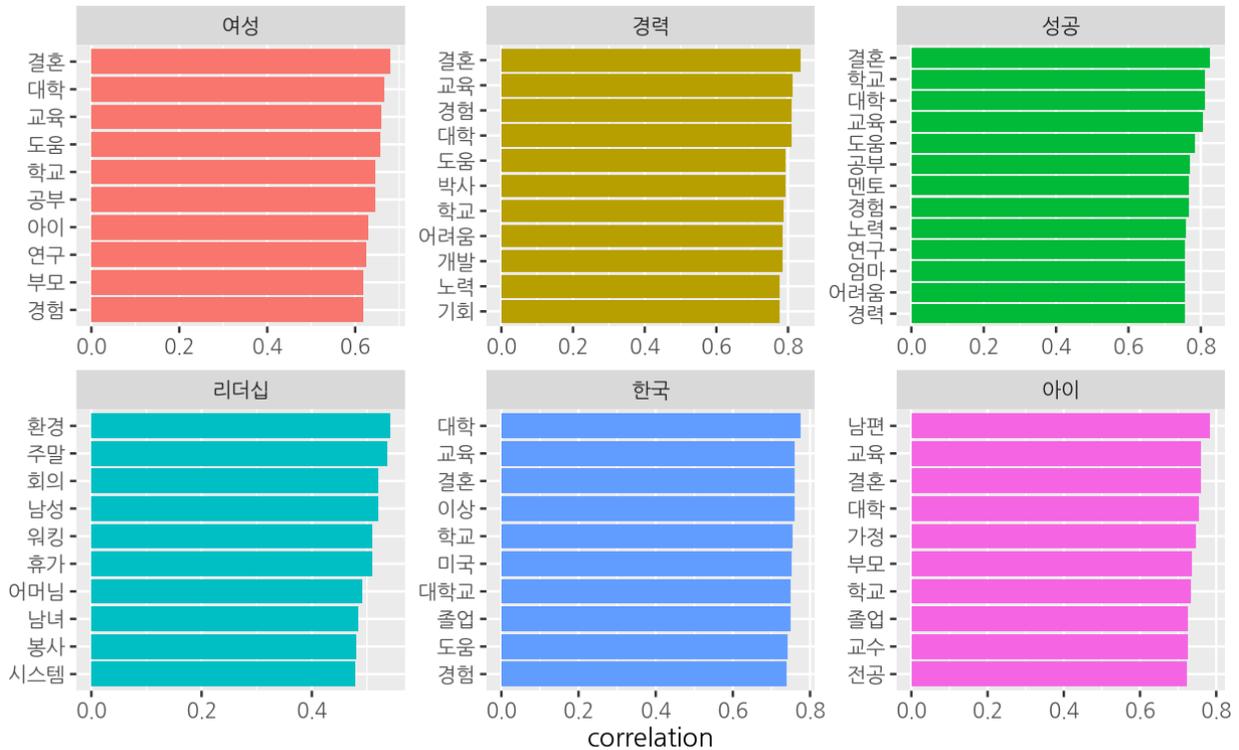


Figure 6: 주요단어 연관성

- 225 • 여성: 이 카테고리는 '여성'이라는 단어와 높은 상관관계를 가지는 다른 단어들 보여줍니다. 여성  
 226 과 가장 연관성이 높은 단어들은 여성의 사회적 역할, 가족 내 위치, 그리고 개인적 특성과 관련된  
 227 것으로 추정됩니다. 결혼, 공부, 대학, 과정 등의 단어들과 상관관계가 있습니다. 이는 여성의 삶  
 228 에서 교육과 가정생활이 중요한 역할을 차지함을 보여주며, 여성의 사회적 진출에 있어 이러한  
 229 요소들이 큰 영향을 미칠 수 있음을 시사합니다.
- 230 • 경력: '경력'과 관련하여 높은 상관관계를 가지는 단어들은 전문성, 교육, 그리고 일과 관련된 요소  
 231 들로 보입니다. 이는 경력 개발과 관련된 중요한 요소들을 반영합니다. 결혼, 경험, 대학, 도움 등의  
 232 단어들과 높은 상관계수를 가지며, 이는 경력 발전에 있어 개인의 학력, 경험, 결혼 여부 및 사회적  
 233 지원이 중요한 요소로 작용할 수 있음을 나타냅니다.
- 234 • 성공: '성공'이라는 단어와 관련성이 높은 단어들은 아마도 성공을 측정하거나 이를 달성하기 위한  
 235 조건들과 관련이 있을 것입니다. 대학, 학교, 결혼, 도움, 교육 등과 높은 상관관계를 보이는데, 이는  
 236 성공이 교육 수준과 사회적 지원, 가정 환경과 밀접한 연관이 있음을 나타냅니다.
- 237 • 리더십: '리더십'과 연관된 단어들은 리더의 특성, 리더십의 영향, 그리고 리더십이 발휘되는 맥락  
 238 과 관련된 것으로 보입니다. 신입, 휴가, 회의, 환경 등과의 상관계수가 상대적으로 낮습니다. 이는  
 239 리더십이 다양한 요소들과 복잡한 관계를 맺고 있음을 보여줄 수 있으며, 리더십은 단일 요인보다  
 240 는 여러 조건의 조합에 의해 영향을 받을 수 있음을 시사합니다.
- 241 • 한국: '한국'과 높은 상관관계를 가지는 단어들은 한국의 사회적, 문화적 맥락과 관련이 있을 수

242 있습니다. 이는 한국에서의 여성 리더십에 영향을 미치는 특정한 요소들을 나타낼 수 있습니다. 대  
 243 학, 박사, 이상, 관계, 외국 등과의 상관관계수가 있습니다. 이는 한국 사회에서 학력이 중요한 역할을  
 244 하며, 국제적인 관계나 외국과의 교류가 경력 및 개인적 성장에 영향을 미칠 수 있음을 나타냅니다.  
 245 • 아이: '아이'와 상관관계가 높은 단어들은 주로 가정, 육아, 그리고 아이들과 관련된 교육적 측면  
 246 을 나타낼 것입니다. 남편, 교육, 결혼, 교수 등과의 상관관계수가 높습니다. 이는 아이들의 성장과  
 247 발달에 가정 환경과 부모의 교육적 배경이 크게 영향을 미치는 것을 의미할 수 있습니다.

248 이러한 상관관계를 분석하는 것은 여성 리더십이 각각의 맥락에서 어떤 요소들과 강하게 연결되어  
 249 있는지 이해하는 데 도움이 됩니다. 예를 들어, '여성'과 '남편', '아이'와의 높은 상관관계는 가정과 직  
 250 장 사이의 역할 분담과 밸런스를 논의할 때 중요할 수 있습니다. 반면에, '경력'과 '교육', '전문성' 간의  
 251 연결은 직업 성장과 개발에 중점을 둘 때 유용합니다.

252 *Keyword Network*

- 253 • 단어의 상관관계를 사용해 네트워크를 그림.
- 254 • 우선 100개의 상관관계가 높은 단어들을 우선 도출
- 255 • 이중 상관관계가 0.8 이상인 단어쌍들의 네트워크를 도식화
- 256 • 시각화의 간명화를 위해 전체 단어쌍의 네트워크가 아닌 (172,360쌍) 간명화된 네트워크임에 주의
- 257 할 것

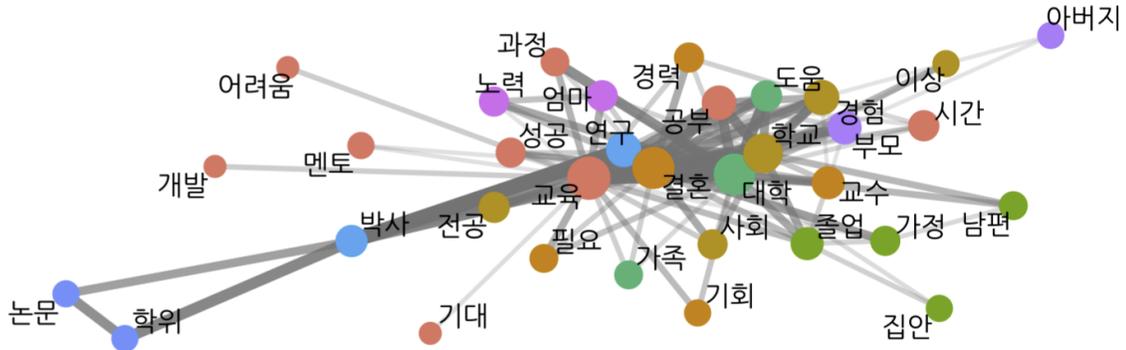


Figure 7: 주요단어 네트워크

- 258 • 교육적 배경과 직업 경로: “초등”과 “경험”, “전공”과 “대학교”, “아버지”와 “전공” 등의 쌍들은  
 259 교육과 초기 경험, 그리고 가족 (특히 아버지의 역할)이 여성의 직업 선택과 전문성 발달에 영향을  
 260 미칠 수 있음을 나타냅니다.
- 261 • 사회적 지원과 학문적 성취: “도움”과 “대학교”, “멘토”와 “가족”, “관심”과 “멘토” 등은 학문적  
 262 성취에 있어서 멘토링, 가족의 지원, 그리고 사회적 관심이 중요한 역할을 함을 시사합니다.

- 263 • 직업적 성공과 개인적 노력: “노력”과 “연구”, “논문”과 “가정”, “집안”과 “직업” 등의 쌍들은  
264 직업적 성공과 학문적 성취가 개인적인 노력과 가정환경, 가족의 지원과 밀접한 관계가 있음을  
265 보여줍니다.
- 266 • 경력 발달과 개인적 동기: “동기”와 “기회”, “노력”과 “시간” 등은 경력 발달에 있어서 개인의  
267 동기부여와 시간 투자가 기회를 창출하는 데 중요함을 강조합니다.

268 주요단어들간의 상관관계와 이를 이용한 네트워크 분석은 여성 리더십에 영향을 미치는 다양한 요소  
269 들을 조명하며, 이들 간의 상호작용이 어떻게 개인의 경력 발달과 리더십 형성에 영향을 미칠 수 있는지에  
270 대한 인사이트를 제공. 예를 들어, 교육적 배경과 가족의 지원은 여성이 리더십 역할을 추구하는 데 있  
271 어 기초를 형성할 수 있습니다. 또한, 개인적인 노력과 멘토링의 중요성은 목표 달성과 성취에 필수적인  
272 요소로 나타납니다.

### 273 토픽모델링

#### 274 • 모델갯수와 관련한 지표

- 275 - 독점성(독점): 일반적으로 값이 높을수록 주제가 서로 더 뚜렷하다는 것을 의미하므로 더 나  
276 은 모델을 나타냅니다.
- 277
- 278 - 의미적 일관성(semcoh): 값이 높을수록(음수가 적을수록) 한 토픽 내의 단어가 의미적으로  
279 더 연관성이 있음을 나타내므로 더 나은 모델입니다.
- 280
- 281 - 홀드아웃 난해성(난해성): 값이 낮을수록 모델이 보이지 않는 데이터를 더 잘 예측한다는 의  
282 미이므로 더 좋습니다.
- 283
- 284 - 잔여 및 바운드(잔여, 바운드, 바운드): 이러한 메트릭은 개별적으로 해석하기는 쉽지 않지만  
285 일반적으로 (잔여 및 바운드의 경우) 값이 낮을수록 모델 적합도가 더 높다는 것을 나타낼 수  
286 있습니다. 그러나 다른 측정값과 함께 사용할 때 더 유용합니다.

#### 287 • 토픽내 키워드들의 순위와 관련한 지표들

- 288 - Highest Probability (Highest Prob): 이 메트릭은 특정 주제 내에서 가장 가능성이 높은  
289 단어를 식별합니다. 해당 주제에 할당된 문서에서 가장 자주 등장하는 단어입니다. 그러나  
290 이러한 단어는 해당 주제에만 국한된 것이 아니라 여러 주제 또는 전체 말뭉치에서 공통적으  
291 로 나타날 수 있습니다.
- 292 - FREX (Frequency and Exclusivity): FREX는 단어 빈도와 독점성 사이의 균형을 맞추는  
293 기능입니다. 한 주제에서 빈도가 높으면서도 다른 주제에서는 그다지 자주 나타나지 않는,  
294 즉 해당 주제에만 한정된 단어를 강조 표시합니다. 이 메트릭은 토픽의 특징적인 단어이지만  
295 가장 가능성이 높은 단어는 아닐 수 있는 단어를 식별하는 데 특히 유용합니다. FREX는 한  
296 토픽과 다른 토픽을 구분하는 미묘한 차이에 대한 보다 세밀한 시각을 제공할 수 있습니다.
- 297 - Lift: 상승도는 단어가 말뭉치에서 기준 빈도와 비교하여 주제에 얼마나 더 많이 나타나는지  
298 를 측정합니다. 상승도 값이 높을수록 해당 단어가 토픽과 더 관련이 있다는 뜻입니다. 이는  
299 여러 토픽에서 공통적으로 사용되는 단어가 아닌, 한 토픽에서 빈번하게 사용될 뿐만 아니라  
300 그 토픽과 뚜렷하게 연관된 단어를 찾는 데 좋은 방법입니다.
- 301 - Score: 점수 메트릭은 종종 빈도, 독점성 및 기타 요소를 결합하여 주제에 대한 단어의 중요  
302 도에 대한 단일하고 전반적인 척도를 제공합니다. 이 점수의 정확한 성격은 구체적인 구현  
303 방식이나 연구자의 선호도에 따라 달라질 수 있습니다. 각 주제에서 가장 중요한 단어를 전  
304 체적으로 빠르게 파악할 수 있도록 설계되었습니다.

305 • 주의사항

306 - 도출된 모든 토픽을 사용해야 하는가?

307 \* 아닙니다. 질적자료내에 잠재적으로 (겸쳐서 또는 내재적으로, latent) 존재하는 토픽들  
308 을 도출하는 것이기 때문에 알고리즘적으로 도출되었으나 해석에 곤란하거나, 이해할  
309 수 없는 토픽이 나오기도 합니다. 이런 경우 맥락에 많은 의존하게 됩니다.

310 - 토픽갯수를 산정하는 절대적인 산식이나 합의가 있는가?

311 \* 없습니다. 각각의 지표들은 토픽갯수를 정하는데 가이드라인이 되기는 하지만 정확히  
312 몇개를 지칭하는 단일의 지표는 없습니다. 때문에 많은 연구들에서 각각의 측면을 고  
313 려한 지표들에 대한 연구들이 진행되어왔고 이런 지표들을 고려하여 범위를 좁혀가며  
314 최대한 합리적으로 토픽의 갯수를 결정하게 됩니다

315 - 토픽내 모든 키워드들을 해석에 사용해야 하는가?

316 \* 아닙니다. ~의 연구에 따르면 클리닝이나 고차맥락에 의해 때론 이해할수 없는 키워드  
317 들이 뽑히기도 하며 이를 intruder라고 부르고 있고 해석과 레이블링에서 다른 지표들  
318 과의 비교와 디스크립티브 분석에서의 이해와 연구자의 데이터에 대한 이해를 바탕으  
319 로 제외하도록 권고되고 있습니다.

320 • 이 분석에서 알아두어야 할 사항

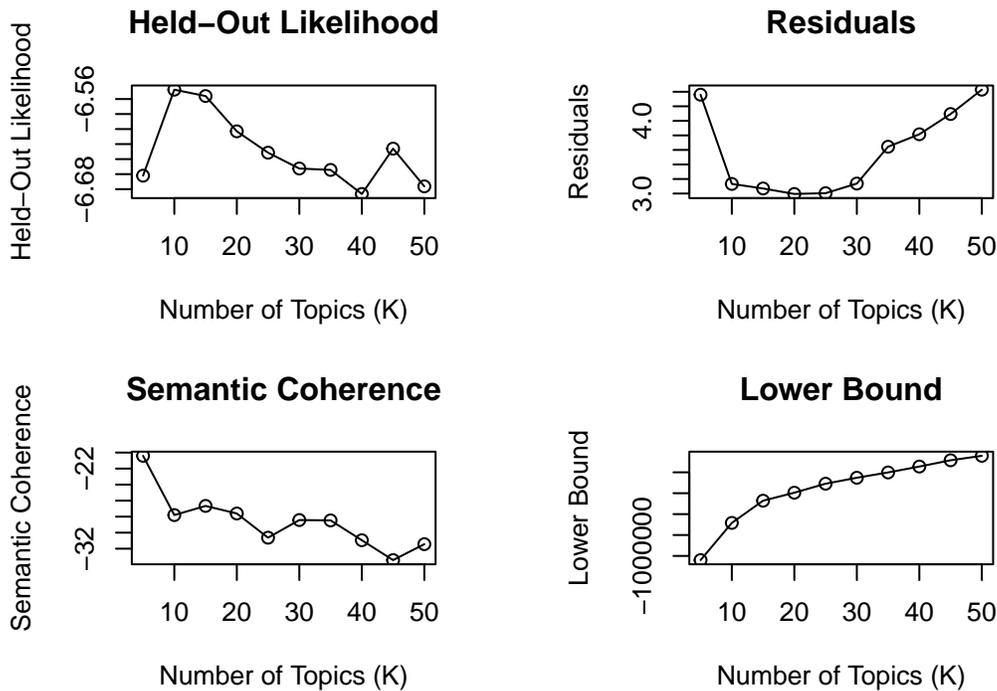
321 - 단어의 분포가 매우 극단적으로 치우친(적게 출현한 단어들이 많고 아주 많이 출현한 단어  
322 들은 적음) 형태이므로 단어의 클리닝 이후, 단어의 하한(몇번 출현한것이상)에 따라 토픽이  
323 매우 다르게 나타납니다.

324 - 일반적으로 많은 단어가 사용된 경우, 그리고 토픽의 갯수가 많은 경우 디테일함을 보여주는  
325 경향이 있으며 단어의 사용이 적거나 토픽의 갯수가 적은 경우 좀 더 일반화된 토픽을 뽑는  
326 경향이 있습니다

327 토픽모델: 하한 단어갯수 3, 3389단어, 30개 토픽

- 328 • Held-out likelihood는 10에서 최대 이후로 즉 감소세였다가 25에서 소폭반등
- 329 • semantic coherence는 30이전까지 등락을 반복하다가 30에서 최대 이후로 떨어지다가 45에서  
330 반등
- 331 • Residuals는 2-30 까지 비슷한 양상을 보이다가 45에서 최대 40에서 최저
- 332 • Lower bound는 완만하게 계속 상승
- 333 • Perplexity는 계속 하강하다가 25개에서 한번 저점 이후로 반등다시 35개에서 하강 상승 그리고  
334 50개에서 최저
- 335 • 10개에서 30개 사이의 토픽이 적당해 보임 20개 토픽에서 퍼플렉시티도 꽤 저점이고 시멘틱 코히  
336 런스는 첫번째 상승 고점 그리고 리더쥬얼은 비슷한 상태로 저점유지 로옴바운드는 중간 때문에  
337 절충하여 20개 토픽으로 결정

### Diagnostic Values by Number of Topics



338 모델핏

##	K	exclus	semcoh	heldout	residual	bound	lbound	perplexity	
340	## 1	5	9.377788	-20.4039	-6.662055	4.360042	-1004016	-1004011	1318.505
341	## 2	10	9.390588	-27.80497	-6.547531	3.131972	-968411.7	-968396.6	1251.515
342	## 3	15	8.864709	-26.66005	-6.556033	3.069624	-947194	-947166.1	1225.269
343	## 4	20	8.846668	-27.60497	-6.603015	2.994263	-939549.7	-939507.3	1213.160
344	## 5	25	8.999873	-30.62945	-6.631424	3.004979	-930963.1	-930905.1	1210.506
345	## 6	30	9.056996	-28.42308	-6.652538	3.137362	-925276.3	-925201.6	1219.142
346	## 7	35	9.125991	-28.46755	-6.654362	3.643879	-920232.2	-920140.1	1213.063
347	## 8	40	9.1839	-30.9445	-6.686386	3.815404	-914596.6	-914486.3	1202.886
348	## 9	45	9.216255	-33.42794	-6.626042	4.094104	-908615.3	-908486.2	1203.155
349	## 10	50	9.271124	-31.43572	-6.676295	4.429038	-904372.9	-904224.4	1201.685

- 350 • Held-out likelihood는 10에서 최대 이후로 하강하다가 45에서 다시 반등하나 50에서 다시하강
- 351 • semantic coherence는 전반적으로 하강하는 추세에서 중간중간 15, 30 에서 반등하는 모습
- 352 • Residuals는 10에서 급격하게 떨어져서 20에서 최저 이후로는 상승
- 353 • Lower bound는 완만하게 계속 상승
- 354 • Perplexity는 20에서 한번 저점 이후에 상승 하강 반복
- 355 • 10개에서 40개 사이의 토픽이 적당해 보임

356 각 메트릭이 어떻게 토픽 모델의 품질을 반영하는지를 고려해야 합니다. 그리고 각각의 변화 추세를  
357 바탕으로 가장 균형 잡힌 토픽 수를 선택하는 것이 중요합니다.

358 Perplexity: 25에서 최저점을 기록한 후 다시 상승하고, 30개 이후에 50개까지 지속적으로 하강하는  
359 경향을 보입니다. 여기서 25개의 토픽에서 perplexity가 처음으로 저점을 찍는 것을 볼 수 있습니다. 이  
360 는 모델이 해당 토픽 수에서 데이터를 잘 예측하고 있다는 신호일 수 있습니다. 그 후 30개 이후부터 50  
361 개까지 다시 하강하는 경향을 보이는데, 이는 추가적인 토픽이 모델의 예측 능력을 향상시키고 있음을

362 의미합니다. 이러한 트렌드는 30개의 토픽이 적절한 복잡도와 예측 능력 사이의 균형을 제공한다는 것을  
363 시사합니다.

364 **Held-out likelihood:** 10개에서 최대값을 기록한 후 28까지 급격히 하락하다가, 30-35 사이에서 완  
365 만하게 하강하며, 40개에서 다시 급강하합니다. 이는 토픽의 개수가 적을 때 모델이 과적합을 경험할 수  
366 있음을 나타내며, 30개의 토픽에서 모델이 비교적 안정적인 성능을 보인다는 것을 나타냅니다. 이는 30  
367 개의 토픽이 데이터에 대해 합리적인 해석을 제공하면서도 과적합을 피하는 좋은 지점이 될 수 있음을  
368 시사합니다.

369 **Residuals:** 10개의 토픽까지 급격하게 하강하고, 20-30개 사이에서 완만한 상승을 보인 후 50개까지  
370 다시 급상승한다는 것은, 20-30개의 토픽 구간에서 모델이 데이터에 포함된 노이즈에 대처하는 데 있어서  
371 적절한 타협점을 찾고 있음을 의미합니다. 이는 30개의 토픽이 데이터의 중요한 구조를 잘 포착하면서도  
372 과대적합을 피하는 적절한 수준을 제공한다고 볼 수 있습니다.

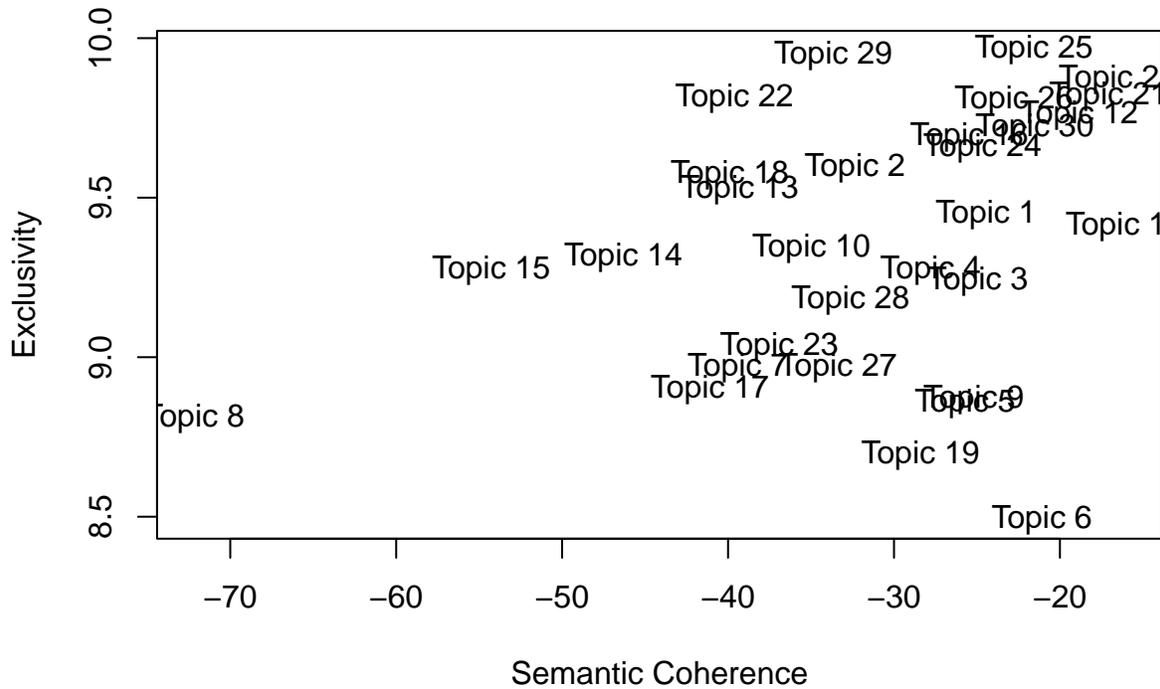
373 **Lower Bound:** Lower Bound가 15개까지 급상승한 후 완만하게 50까지 상승한다는 것은, 이 지점  
374 이후에 모델이 더 이상 크게 개선되지 않음을 나타낼 수 있습니다. 즉, 30개의 토픽에서 이미 데이터의  
375 주요 패턴을 충분히 포착하고 있음을 시사합니다

376 **Semantic Coherence:** 초기 최고값에서 10개까지 급격히 하락한 후 15개에서 첫 고점을 기록하고,  
377 25까지 하락한 다음 30개에서 두 번째 고점을 기록하며 하락합니다. 이는 토픽들이 어느 정도 의미론적  
378 일관성을 가지고 있으며, 30개의 토픽에서 이 일관성이 잘 유지되고 있음을 나타냅니다.

379 위의 분석을 바탕으로, 토픽 개수를 30개로 설정하는 것은 여러 메트릭에서 균형 잡힌 성능을 보여주  
380 기 때문입니다. Perplexity와 Held-out likelihood 모두 이 지점에서 상대적으로 안정된 경향을 보이며,  
381 Residuals는 급격한 변화가 없고, Lower Bound는 지속적으로 상승합니다. 또한 Semantic Coherence  
382 는 30개의 토픽에서 두 번째로 높은 값을 보여주어, 토픽들이 의미론적으로도 일관성이 있음을 보여줍니  
383 다. 이러한 여러 메트릭의 조화는 토픽의 개수를 30개로 설정하는 것이 모델의 품질을 최적화하는 모델의  
384 갯수로 해석하였습니다

### 385 토픽결과

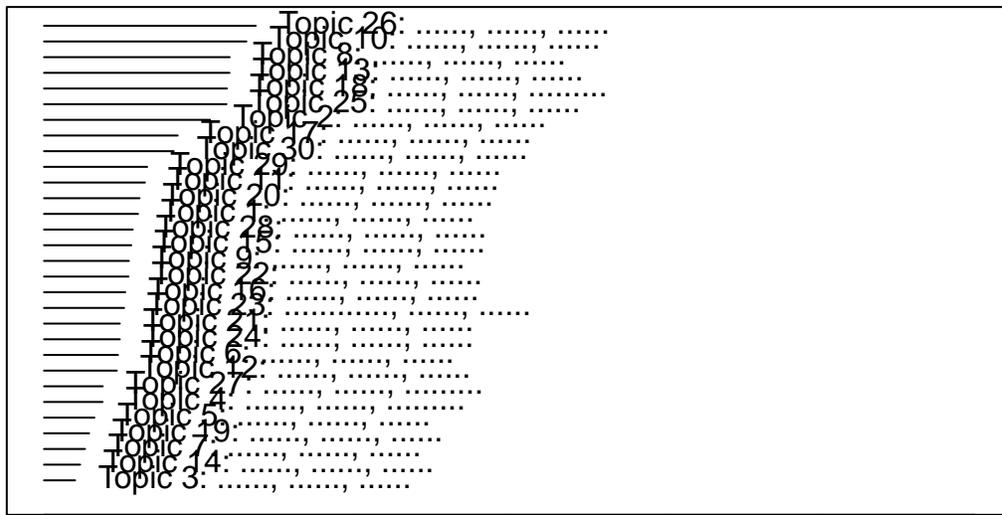
```
386 ## [1] -24.45019 -32.35187 -24.95220 -27.78681 -25.70807 -21.08110 -39.39862
387 ## [8] -72.16872 -25.18917 -34.97596 -16.09075 -18.84093 -39.31211 -46.30967
388 ## [15] -54.27163 -25.46350 -41.08829 -39.88471 -28.40016 -16.50017 -17.06237
389 ## [22] -39.61342 -36.91644 -24.65313 -21.56133 -22.77238 -33.37082 -32.60928
390 ## [29] -33.65343 -21.51034
391 ## [1] 9.446479 9.592248 9.239827 9.273401 8.854341 8.490021 8.967001 8.808211
392 ## [9] 8.868868 9.341397 9.408449 9.758667 9.523646 9.314079 9.271962 9.689968
393 ## [17] 8.898007 9.570456 8.694444 9.870958 9.818032 9.811191 9.031309 9.656556
394 ## [25] 9.963895 9.804758 8.968347 9.180259 9.946028 9.719824
```



395

396 ## A topic model with 30 topics, 612 documents and a 3364 word dictionary.

### Top Topics



0.00 0.05 0.10 0.15 0.20 0.25 0.30

### Expected Topic Proportions

397

398 ## Topic 1 Top Words:

399 ## Highest Prob: 사업, 기술, 기업, 개발, 매출, 직원, 창업, 조직, 남편, 지원

400 ## FREX: 기술, 사업, 매출, 창업, 기상, 모바일, 자금, 사업가, 경영, 손익

401 ## Lift: 기상, 클라우드, 외산, 기업가, 이직률, 모바일, 혁명, 로드맵, 상승, 담보

402 ## Score: 매출, 창업, 사업, 기술, 기업, 보안, 사업가, 자금, 고객, 기상  
403 ## Topic 2 Top Words:  
404 ## Highest Prob: 업무, 시간, 육아, 힘들, 주말, 필요, 근무, 휴직, 휴가, 출산  
405 ## FREX: 업무, 야근, 신랑, 휴가, 휴직, 육아, 직급, 배려, 출산, 근무  
406 ## Lift: 명회, 아모레, 조직원, 신랑, 불가피, 대다수, 단축, 관계자, 자재, 마사지  
407 ## Score: 업무, 아모레, 야근, 신랑, 휴직, 직급, 육아, 자재, 퍼시픽, 호성  
408 ## Topic 3 Top Words:  
409 ## Highest Prob: 교수, 여자, 환자, 의사, 남자, 병원, 힘들, 학교, 기회, 자리  
410 ## FREX: 환자, 내과, 티칭, 리서치, 소화기, 부교수, 연수, 교수, 클리닉, 치료  
411 ## Lift: 클리닉, 개수, 소화기, 보드, 통찰, 개원, 란스, 티칭, 남대, 세요  
412 ## Score: 환자, 내과, 교수, 소화기, 클리닉, 전공의, 티칭, 부교수, 란스, 의과  
413 ## Topic 4 Top Words:  
414 ## Highest Prob: 기업, 경험, 컨설팅, 성공, 힘들, 조직, 외국, 직원, 친구, 관심  
415 ## FREX: 컨설팅, 홍콩, 클라이언트, 기업, 비서, 미팅, 통역, 호텔, 커뮤니케이션, 퍼포  
416 ## 먼스  
417 ## Lift: 잠재력, 스폰서, 동시통역, 홍콩, 차관, 종속, 랭귀지, 통역, 재충전, 이타  
418 ## Score: 기업, 클라이언트, 컨설팅, 컬처, 홍콩, 밸류, 미팅, 고객, 통역, 퍼포먼스  
419 ## Topic 5 Top Words:  
420 ## Highest Prob: 공부, 졸업, 교회, 힘들, 과정, 친구, 논문, 결혼, 학교, 박사  
421 ## FREX: 교회, 기도, 신앙, 국문학, 방통대, 이성호, 하나님, 한국어, 국어, 학당  
422 ## Lift: 방통대, 학당, 이성호, 목사, 국문학, 시누, 한국어, 교무, 선교사, 양심  
423 ## Score: 학당, 이성호, 신앙, 방통대, 국문학, 하나님, 국어, 논문, 시누, 교회  
424 ## Topic 6 Top Words:  
425 ## Highest Prob: 창업, 직원, 기업, 아버지, 공부, 개발, 기술, 사업, 성공, 지원  
426 ## FREX: 창업, 보안, 사업가, 적자, 장치, 손익, 자금, 이익, 분기점, 상공  
427 ## Lift: 창업자, 화성시, 상공, 회의소, 이율, 적자, 일주, 도산, 화합, 댄스  
428 ## Score: 창업, 보안, 사업가, 적자, 매출, 자금, 화성시, 기업인, 분기점, 장치  
429 ## Topic 7 Top Words:  
430 ## Highest Prob: 의사, 의학, 교수, 병원, 학교, 공부, 시간, 어머니, 진료, 여성  
431 ## FREX: 의학, 재활, 진료, 의과, 의원, 의사, 내과, 병원장, 학술, 위원장  
432 ## Lift: 의학, 재활, 복지부, 은사, 의원, 병원장, 여의사, 비례, 위상, 생색  
433 ## Score: 의학, 의과, 진료, 재활, 내과, 환자, 병원장, 응급, 전공의, 여의사  
434 ## Topic 8 Top Words:  
435 ## Highest Prob: 기자, 직원, 후배, 구청, 엄마, 인터넷, 자리, 시험, 업무, 팀장  
436 ## FREX: 구청, 기자, 점포, 책임자, 인천, 행원, 손님, 일반직, 건축, 인터넷  
437 ## Lift: 구청, 부시장, 지면, 서초, 정치부, 서초구, 점포, 행원, 교통부, 수협  
438 ## Score: 구청, 기자, 책임자, 지면, 본점, 점포, 인천, 부시장, 행원, 교통부  
439 ## Topic 9 Top Words:  
440 ## Highest Prob: 한국, 미국, 외국, 공부, 영어, 나이, 대학, 고객, 조직, 중요  
441 ## FREX: 한국, 미국, 텔레, 항공사, 파티, 외교관, 독일, 그램, 로컬, 오피스  
442 ## Lift: 외교관, 텔레, 항공사, 맥킨지, 파이널, 그램, 가교, 보유, 스텝, 시카고  
443 ## Score: 한국, 항공사, 텔레, 미국, 고객, 교포, 그램, 독일, 파이널, 스텝  
444 ## Topic 10 Top Words:  
445 ## Highest Prob: 팀장, 조직, 리더, 팀원, 여자, 임원, 고민, 개발, 고객, 사업  
446 ## FREX: 팀장, 팀원, 구성원, 회계, 리더, 건설, 성과, 전사, 추진력, 합병  
447 ## Lift: 부류, 레코드, 실로, 과업, 팀원, 팀장, 추진력, 구성원, 회계학, 이만큼  
448 ## Score: 팀장, 팀원, 고객, 조직, 임원, 구성원, 회계, 리더, 자금, 결산  
449 ## Topic 11 Top Words:

450 ## Highest Prob: 아이, 여성, 육아, 엄마, 시간, 생활, 직장, 힘들, 필요, 사회  
 451 ## FREX: 아이, 육아, 양립, 어린이집, 휴직, 출산, 양육, 여성, 보육, 인화  
 452 ## Lift: 목동, 일수, 수유, 배분, 리그, 위탁, 안심, 아이, 보육, 강박  
 453 ## Score: 아이, 여성, 육아, 인화, 어린이집, 엄마, 보안, 휴직, 양립, 수유  
 454 ## Topic 12 Top Words:  
 455 ## Highest Prob: 부모, 과정, 전공, 개발, 경력, 분야, 기회, 직업, 졸업, 경험  
 456 ## FREX: 직업, 전공, 부모, 임용, 기대, 동생, 분야, 석사, 단절, 과정  
 457 ## Lift: 일차, 천안, 문헌, 폭력, 자매, 일부분, 이중, 양대, 소진, 바이올린  
 458 ## Score: 직업, 전공, 임용, 천안, 경력, 석사, 부모, 개발, 교직원, 과정  
 459 ## Topic 13 Top Words:  
 460 ## Highest Prob: 직원, 고객, 친구, 힘들, 공부, 마음, 욕심, 영어, 언니, 업무  
 461 ## FREX: 직원, 관리자, 욕심, 부서장, 한전, 게임, 고객, 협상, 부하, 국세청  
 462 ## Lift: 전력, 시중, 한전, 초급, 조급, 저학년, 격주, 나이브, 전세, 급급  
 463 ## Score: 직원, 고객, 한전, 부서장, 본점, 보안, 국세청, 산원, 시중, 전력  
 464 ## Topic 14 Top Words:  
 465 ## Highest Prob: 공부, 교육, 정책, 유학, 분석, 교육부, 고시, 결과, 가족, 결정  
 466 ## FREX: 교육부, 의약품, 고시, 이혼, 행시, 광주, 연구실, 교육청, 보고서, 공직  
 467 ## Lift: 의약품, 교육청, 공직, 결단, 실태, 디펜스, 심층, 양분, 스테이트, 방침  
 468 ## Score: 의약품, 교육부, 광주, 교육청, 행시, 고시, 사무관, 결단, 공직, 주문  
 469 ## Topic 15 Top Words:  
 470 ## Highest Prob: 병원, 여자, 의사, 남자, 환자, 교수, 힘들, 외과, 출신, 업무  
 471 ## FREX: 외과, 병원, 외래, 환자, 전공의, 진료, 의사, 당직, 정형, 출신  
 472 ## Lift: 성형외과, 수가, 외래, 외과, 당직, 의국, 여선생, 유방, 병동, 마취  
 473 ## Score: 환자, 외과, 전공의, 외래, 병원, 진료, 의국, 내과, 수가, 당직  
 474 ## Topic 16 Top Words:  
 475 ## Highest Prob: 가정, 도움, 전문, 가족, 시간, 필요, 균형, 경력, 사회, 멘토  
 476 ## FREX: 가정, 균형, 전문, 한계, 성취감, 도움, 쇼핑, 기준, 사회생활, 전문가  
 477 ## Lift: 명상, 방편, 영위, 타인, 지주, 균형, 성가대, 쇼핑, 유대, 부지런  
 478 ## Score: 가정, 균형, 전문, 도움, 성취감, 방편, 그러, 영위, 약대, 명상  
 479 ## Topic 17 Top Words:  
 480 ## Highest Prob: 조직, 문화, 직원, 친구, 영업부, 중요, 능력, 여자, 그렇, 여성  
 481 ## FREX: 영업부, 딜러, 디렉터, 내일, 소통, 업계, 직군, 용품, 사물, 문화  
 482 ## Lift: 테팔, 딜러, 라운드, 기념일, 사물, 지론, 영업부, 제주, 디렉터, 용품  
 483 ## Score: 딜러, 영업부, 테팔, 디렉터, 용품, 매출, 라운드, 챌린지, 조직, 현지  
 484 ## Topic 18 Top Words:  
 485 ## Highest Prob: 여성, 남성, 리더십, 직원, 기업, 성장, 환경, 필요, 코칭, 관리  
 486 ## FREX: 남성, 여성, 코칭, 지속, 리더십, 환경, 수익, 생산, 특별, 평등  
 487 ## Lift: 카네기, 폼하, 걸레, 체인, 주목, 볼링, 주경야독, 소심, 남성, 바지  
 488 ## Score: 여성, 남성, 코칭, 창업, 기업, 직원, 리더십, 임원, 생태, 걸레  
 489 ## Topic 19 Top Words:  
 490 ## Highest Prob: 그러, 학교, 대학, 한국, 어머니, 아버지, 부모, 시간, 생활, 마음  
 491 ## FREX: 그러, 교육열, 유아, 이북, 부산대, 프레임, 오인택, 연구비, 영화, 계약직  
 492 ## Lift: 노선, 터널, 방송국, 그러, 부산대, 급박, 유학생, 교육열, 쓸모없, 비롯  
 493 ## Score: 그러, 유아, 교육열, 오인택, 부산대, 개론, 연구비, 이북, 사회학, 프레임  
 494 ## Topic 20 Top Words:  
 495 ## Highest Prob: 교육, 학교, 남편, 전공, 공부, 미국, 과정, 아이, 관심, 한국  
 496 ## FREX: 교육, 학교, 전공, 남편, 교사, 심리, 자녀, 관심, 대학원, 노인  
 497 ## Lift: 교육, 노인, 학교, 홈페이지, 한양대, 의뢰, 교사, 학부모, 액티브, 지방대

498 ## Score: 교육, 학교, 남편, 교사, 전공, 노인, 학부모, 아이, 심리, 미국  
 499 ## Topic 21 Top Words:  
 500 ## Highest Prob: 교수, 연구, 대학, 논문, 경력, 사회, 박사, 개발, 평가, 관계  
 501 ## FREX: 연구, 교수, 논문, 지도, 서울대, 대학, 평가, 고등교육, 발표, 경력  
 502 ## Lift: 연구, 아시, 고등교육, 소수자, 비정규직, 교육자, 교수, 연구회, 백인, 비판  
 503 ## Score: 교수, 연구, 논문, 지도, 고등교육, 경력, 박사, 사회학, 연구자, 직업  
 504 ## Topic 22 Top Words:  
 505 ## Highest Prob: 친구, 워킹, 네트, 관계, 웃음, 멘토, 인력, 그렇, 후배, 분위기  
 506 ## FREX: 워킹, 웃음, 네트, 인력, 친구, 관계, 독일, 동료, 성숙, 제한  
 507 ## Lift: 수반, 웃음, 워킹, 가르침, 묘하, 원고, 중도, 대수, 네트, 중성  
 508 ## Score: 웃음, 워킹, 네트, 친구, 인력, 독일, 입찰, 철학, 학부모, 관계  
 509 ## Topic 23 Top Words:  
 510 ## Highest Prob: 프로젝트, 고객, 임원, 기업, 조직, 자리, 고민, 결정, 갈등, 필요  
 511 ## FREX: 프로젝트, 컨설턴트, 갈등, 이스트, 텔레콤, 여사원, 체인지, 보스턴, 본질, 고  
 512 객  
 513 ## Lift: 체인지, 작정, 여사원, 서번트, 보스턴, 컨설턴트, 홀릭, 도망, 존재감, 워커  
 514 ## Score: 고객, 프로젝트, 컨설턴트, 임원, 텔레콤, 여사원, 효성, 이스트, 보스턴, 팀  
 515 장  
 516 ## Topic 24 Top Words:  
 517 ## Highest Prob: 박사, 학위, 공부, 과정, 지원, 석사, 대학, 경력, 논문, 유학  
 518 ## FREX: 학위, 박사, 취득, 연세대, 취업, 석사, 연구자, 유학, 병행, 토플  
 519 ## Lift: 졸업식, 한국교육개발원, 취득, 학위, 교육사, 구직, 석박, 토플, 박사, 외대  
 520 ## Score: 박사, 학위, 취득, 연세대, 논문, 취업, 연구자, 석사, 구직, 강의  
 521 ## Topic 25 Top Words:  
 522 ## Highest Prob: 성공, 결혼, 노력, 중요, 행복, 멘토, 생활, 준비, 자리, 목표  
 523 ## FREX: 성공, 노력, 행복, 결혼, 강의, 목표, 중요, 준비, 조언, 자신  
 524 ## Lift: 성공, 노력, 행복, 조언, 결혼, 후임자, 강의, 의지, 전부, 마인드  
 525 ## Score: 성공, 노력, 강의, 결혼, 행복, 목표, 조언, 멘토, 후배, 준비  
 526 ## Topic 26 Top Words:  
 527 ## Highest Prob: 여자, 남자, 조직, 리더, 직장, 필요, 사회, 시간, 리더십, 중요  
 528 ## FREX: 여자, 남자, 리더, 조직, 뒷사람, 차별, 군대, 인정, 회식, 회의  
 529 ## Lift: 충고, 여자, 경제력, 네이버, 남자, 단기간, 뒷사람, 진솔, 각광, 정원  
 530 ## Score: 여자, 남자, 조직, 리더, 뒷사람, 리더십, 사회, 직장, 육아, 회식  
 531 ## Topic 27 Top Words:  
 532 ## Highest Prob: 아이, 남편, 어머니, 한국, 시간, 어머니, 결정, 필요, 외국, 학교  
 533 ## FREX: 캐나다, 어머니, 엘지, 아버님, 주방, 스텐, 가전, 시댁, 제사, 바이어  
 534 ## Lift: 저력, 주방, 계란, 엘지, 국민학교, 캐나다, 정보원, 노동력, 바이어, 단지  
 535 ## Score: 아이, 주방, 엘지, 캐나다, 스텐, 가전, 어머니, 바이어, 지망, 매장  
 536 ## Topic 28 Top Words:  
 537 ## Highest Prob: 임원, 인사, 승진, 사업, 시스템, 공부, 기업, 관리, 경영, 기회  
 538 ## FREX: 임원, 인사, 라이프, 채용, 비서실, 효성, 컴퍼니, 분가, 발탁, 밸런스  
 539 ## Lift: 엘리베이터, 태반, 비서실, 업데이트, 저변, 오퍼레이션, 실리, 핸디캡, 파일, 디  
 540 스플레이  
 541 ## Score: 임원, 채용, 효성, 인사, 비서실, 컴퍼니, 라이프, 업데이트, 코어, 인화  
 542 ## Topic 29 Top Words:  
 543 ## Highest Prob: 그렇, 문제, 시간, 나이, 재미있, 해결, 과정, 의사, 집안, 훌륭  
 544 ## FREX: 그렇, 문제, 해결, 재미있, 나이, 훌륭, 배려, 시간, 집안, 체계  
 545 ## Lift: 문제, 그렇, 축구, 야구, 훌륭, 재미있, 해결, 나이, 중공업, 코어

546 ## Score: 그렇, 문제, 재미있, 나이, 해결, 시간, 코어, 야구, 훌륭한, 배려  
 547 ## Topic 30 Top Words:  
 548 ## Highest Prob: 엄마, 남편, 아들, 아버지, 결혼, 힘들, 아빠, 대학, 마음, 부모  
 549 ## FREX: 엄마, 아들, 아빠, 할머니, 남편, 사랑, 아버지, 대화, 담임, 감사  
 550 ## Lift: 목록, 외아들, 원수, 죽음, 용돈, 자격지심, 담임, 비디오, 할머니, 엄마  
 551 ## Score: 엄마, 남편, 아들, 죽음, 할머니, 담임, 아빠, 교사, 광주, 용돈

## 552 수정 레이블

- 553 1. 여성 창업가와 기술 기반 사업
- 554 2. 직장과 가정 생활의 균형
- 555 3. 여성 의료 전문가와 환자 관리
- 556 4. 국제 기업 컨설팅과 여성 리더십
- 557 5. 여성의 교육과 신앙 관점
- 558 6. 여성 기업가의 창업과 성장
- 559 7. 여성 의료 교육자와 의학 연구
- 560 8. 공공 서비스에서의 여성 기자 리더십
- 561 9. 글로벌 인식과 해외 경험
- 562 10. 여성 리더십과 팀 관리
- 563 11. 가정생활과 여성의 역할
- 564 12. 학문적 발전과 직업 선택
- 565 13. 조직 내 여성 리더의 직원 관리
- 566 14. 여성의 학문적 기여와 교육 정책
- 567 15. 의료 분야의 여성 리더십
- 568 16. 가정과 전문직 사이의 균형
- 569 17. 영업 분야에서의 여성 리더십
- 570 18. 조직 관리와 여성 리더십
- 571 19. 사회 문제 해결과 여성의 시각
- 572 20. 여성의 감정적 측면과 가정생활
- 573 21. 여성 교수의 학문적 연구와 경력
- 574 22. 인간 관계와 네트워킹의 중요성
- 575 23. 기업 프로젝트 관리와 여성 임원
- 576 24. 교육 경력과 학문적 성취
- 577 25. 멘토십의 중요성과 개인적 성공
- 578 26. 조직 내 여성의 역할과 차별
- 579 27. 다문화 경험과 해외 생활
- 580 28. 인사전략과 임원 경력 관리
- 581 29. 일상적 문제 해결과 여성의 관점
- 582 30. 가족 관계와 여성의 감정 처리

### 583 • 여성 기업가와 창업:

- 584 - 속한 토픽들: 1, 6, 14, 19
- 585 - 클러스터 설명: 이 클러스터는 여성 기업가들의 창업 경험, 사업 개발, 기술적 혁신, 그리고
- 586 기업 경영에 관한 주제를 다룹니다. 여성 창업가들이 직면한 도전과 성공 사례, 그리고 사업
- 587 운영과 기술 개발에 대한 내용이 주를 이룹니다.

### 588 • 여성의 직장 생활과 가정 생활 균형:

- 589           - 속한 토픽들: 2, 11, 16, 30  
590           - 클러스터 설명: 이 클러스터는 여성이 직장 생활과 가정 생활을 어떻게 균형있게 관리하는  
591            지에 대해 탐구합니다. 육아, 가정 내 역할, 직장에서의 책임, 그리고 개인적인 시간 관리에  
592            대한 주제가 포함됩니다.
- 593       • 여성의 학문적 성취와 교육:
- 594           - 속한 토픽들: 5, 12, 20, 24, 28  
595           - 클러스터 설명: 이 클러스터는 학문적 성취에 집중하는 여성들의 경험을 다룹니다. 대학 교  
596            육, 박사 학위 취득, 학문적 연구, 그리고 교육 분야에서의 경력 개발에 대한 내용을 포함합  
597            니다.
- 598       • 여성의 의료 분야 전문성과 리더십:
- 599           - 속한 토픽들: 3, 7, 15  
600           - 클러스터 설명: 이 클러스터는 여성이 의료 분야에서 어떻게 전문성을 발휘하고 리더십을 키  
601            워나가는지에 대해 탐구합니다. 의사, 교수, 병원 관리 등 의료 분야에서 여성이 겪는 도전과  
602            기회에 대한 내용이 중심을 이룹니다.
- 603       • 여성 리더의 조직 관리와 네트워킹:
- 604           - 속한 토픽들: 4, 9, 17, 23, 26, 29  
605           - 클러스터 설명: 이 클러스터는 조직 내에서 여성 리더가 어떻게 영향력을 행사하고, 네트워  
606            킹을 통해 직장 내외에서 관계를 구축하는지에 대해 다룹니다. 컨설팅, 영업, 네트워크 구축,  
607            팀 관리 등이 주요 주제입니다.
- 608       대안 레이블 [https://docs.google.com/spreadsheets/d/1oadR8LgtB0euvq6YYcNw7CoaxWZojQx8Aczb-](https://docs.google.com/spreadsheets/d/1oadR8LgtB0euvq6YYcNw7CoaxWZojQx8Aczb-u6wJXc/edit#gid=506785807)  
609       u6wJXc/edit#gid=506785807

	Chae	Cho	Lee	Yoo
1	여성 창업가와 기술 기반 사업	여성 창업가와 사업 성장	여성 창업가와 기술 개발	여성창업가의 기술개발 및 경영
2	직장과 가정 생활의 균형	직장과 가정 생활의 균형	일과 삶의 균형	직장과 가정생활의 균형
3	여성 의료 전문가와 환자 관리	연관성 없음	여성 의사와 환자 관리	여자의사의 조직 내 어려움
4	국제 기업 컨설팅과 여성 리더십	연관성 없음	다국적 기업의 리더와 조직 관리	(빼는걸로)
5	여성의 교육과 신앙 관점	연관성 없음	여성과 신앙 생활	(빼는걸로)
6	여성 기업가의 창업과 성장			
7	여성 의료 교육자와 의학 연구	여성 의료 교육과 의학 연구	여성 의사의 의학 연구	여자의사의 경력관리
8	공공 서비스에서의 여성 기자 리더	언론분야 여성 리더십	여성 리더십과 대외 업무	(빼는걸로)
9	글로벌 인식과 해외 경험	해외 경험과 일가정 양립 인식	여성의 해외 공부 도전	다국적기업 조직문화
10	여성 리더십과 팀 관리	팀 관리와 여성 리더십	여성리더십과 팀 관리	팀 관리와 여성 리더십
11	가정생활과 여성의 역할	가정 생활과 여성의 역할	여성의 가정에서의 역할	일과 자녀 양육의 양립 어려움
12	학문적 발전과 직업 선택	직업 선택과 학문적 발전	여성의 직업과 학문적 발전	직업선택 및 경력개발
13	조직 내 여성 리더의 직원 관리	직원 관리와 조직 내 여성의 역할	조직 내 여성의 직원 관리	직원 관리와 조직 내 여성의 역할
14	여성의 학문적 기여와 교육 정책	교육 정책과 여성의 경력개발	여성학자의 학문적 기여	(빼는걸로)
15	의료 분야의 여성 리더십	의료 분야에서 여성 리더십	여자의사의 의료 리더십	의료분야에서의 여성리더십
16	가정과 전문직 사이의 균형	가정과 일의 균형	전문직 여성의 일과 삶의 균형	가정과 일에서의 균형
17	영업 분야에서의 여성 리더십	영업 분야 여성 리더십과 문화	영업분야의 여성리더십	여성리더십과 조직 문화
18	조직 관리와 여성 리더십	여성 리더십과 조직 관리	여성리더십과 조직 관리	여성리더십과 조직 관리
19	사회 문제 해결과 여성의 시각	연관성 없음	여성의 사회적 시각과 문제 해결	(빼는걸로)
20	여성의 감정적 측면과 가정생활	가정생활과 여성의 감정적 측면	여성의 가정 생활	자녀양육 및 경력개발
21	여성 교수의 학문적 연구와 경력	학문적 연구와 여성 교수의 경력	여성 교수의 경력관리와 학문적 연	여자박사의 연구 및 경력 개발
22	인간 관계와 네트워킹의 중요성	네트워킹과 인간 관계의 중요성	네트워킹의 중요성	여성리더의 네트워킹
23	기업 프로젝트 관리와 여성 임원	기업 프로젝트 관리와 여성 임원	여성 임원의 기업 프로젝트와 관리	여성리더의 어려움 및 고민
24	교육 경력과 학문적 성취	연관성 없음	여성의 학문적 성취	여자박사의 학위취득 및 경력
25	멘토십의 중요성과 개인적 성공	개인적 성공과 멘토십의 중요성	성공과 멘토십의 중요성	개인적 성공과 멘토십의 중요성
26	조직 내 여성의 역할과 차별	조직 내 여성의 역할과 차별	조직내 여성의 역할과 차별	조직 내 여성의 역할과 차별
27	다문화 경험과 해외 생활	해외 생활과 다문화 경험	해외 생활과 다문화 경험	해외경험 및 가족관계
28	인사전략과 임원 경력 관리	연관성 없음	여성 임원의 경력관리와 인사 전략	(빼는걸로)
29	일상적 문제 해결과 여성의 관점	일상적 문제 해결과 여성의 관점	여성의 삶의 문제와 해결	(빼는걸로)
30	가족 관계와 여성의 감정 처리	가족 관계와 여성의 감정 처리	여성의 가족관계와 어려움	여성의 가족관계와 어려움

610

611

- 이상의 보고서의 내용을 참조하여 최종레이블 선정을 위한 준비를 부탁드립니다.

612

1. 보고서의 내용을 다시 검토

613

2. 구글 드라이브의 레이블비교 탭의 레이블을 업데이트

614

3. 미팅 중 최정 레이블을 선정 및 해석에 포함 불포함 토픽을 선정 (해석 불포함 레이블의 경우 수정 레이블에서 제시된 레이블로만 레이블을 기입하고 추가적인 분석이나 해석에서는 제외하는것을 의미)

615

616

617 **References**

- 618 Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological*  
619 *Methods & Research*, 49(1):3–42.