실록통계청 · 라인 기획안 v2.3

작성 2026-09-25. v1(2026-08-23)을 1호·2호 원고와 시즌2~8 기획서에 맞춰 고친 것이다.
주 1회 발행 체제를 유지한다.

v2.1 (2026-10-17): 3호를 「신하를 믿지 못한 영조」로 바꾸고 이후 호수를 하나씩 밀었다.
틀 F(왕대 비교)를 더하고, 틀 F 편의 분량 예외와 공용 도구 yeongjo_census.py를 적었다.
3호에서 넘긴 재료(게장과 탕약)를 6장 003 아래에 모았다.

v2.2 (2026-10-22): 2호를 「송시열 통설 검증」 축으로 다시 썼다. 그 과정에서 틀 G(통설 검증)가
생겼다. 다음 카페 스크랩 「조선왕조 기네스북」(2014)에서 실록 데이터로 검증할 수 있는 항목을
골라 9장에 모으고, 그 가운데 관직 기록을 014로 세웠다.

v2.3 (2026-10-22): **분량 상한(2,000~2,500자)과 "해석 절 네 개" 규칙을 없앴다.** 2호 실측 2,083자에
맞춰 정한 값이었는데, 다른 라인의 실제 길이와 맞지 않았다. 서지수 졸기는 공백 포함 본문 33,405자이고
졸기 대부분이 18,000자를 넘는다. 통계청도 같은 잣대(공백 포함)로 재고, 길이는 소재가 정한다.

---

0. v1에서 바뀐 것

항목	v1	v2	근거
첫 편	이름 순위	**분모 해설 편(1호)** 을 앞에 둠	1호가 이미 발행됨 (wp_id 602)
호수	통계청01	001, 002 … 세 자리	1호·2호 파일명, 졸기 세 자리 규칙
분모	편마다 판단	**1호가 정한 네 범위** 중 하나를 명시	1호 VI절

항목	v1	v2	근거
순위 편 구성	순위 하나	**순위 + 두 번째 지표**	2호의 제목 등장률
분량	1,200~1,800자	~~2,000~2,500자(공백 제외)~~ → **v2.3: 상한 없음, 공백 포함으로 잰다**	졸기 18,000자 이상, 서지수 졸기 33,405자
4편·8편	기사 길이, 史臣曰 비율 따로	**한 편으로 합침**	시즌4 기획서가 이미 측정하고 통계청으로 넘김
시즌 연동	시즌1·시즌6	시즌4·6·9, 졸기	시즌 번호 재편
도구	census.py 7개 명령 신설	term만 완성. **series는 count_word.py로 대체**	실제 도구 현황
슬러그	census-01-names	**번호를 넣지 않음**	시즌2와 같은 이유

---

1. 이 라인은 무엇인가

하나의 숫자를 제시하고, 그 숫자가 무엇을 말하고 무엇을 말하지 않는지를 밝힌다.
v1의 정의를 그대로 둔다. 졸기는 인물 한 명을 깊게 읽고, 시즌은 편집의 흔적을 추적하며,
통계청은 분포를 보여 준다.

시즌과의 분업

같은 데이터를 두 라인이 쓸 때는 역할을 나눈다. **통계청은 500년 전체의 분포를 보여 주고,**
시즌은 그 분포 안의 개별 기록을 읽는다. 같은 표를 두 곳에 싣지 않는다. 시즌 원고는 통계청 편을 링크로 부르고, 통계청 편은 시즌 원고를 "이 숫자 안의 기록 하나"로 링크한다.

어조

"~입니다" 체를 쓰되 졸기처럼 말을 걸지 않는다. 숫자를 놓은 뒤 물러선다.
감탄을 유도하는 문장을 넣지 않는다.

2. 분모 규칙 — 1호가 정한 것

1호 「41만 건에는 무엇이 들어 있나」가 이 라인 전체의 기준 문서다. 모든 편은 아래 범위 가운데
하나를 골라 **「헤아리는 방법」 절에 건수와 함께 밝힌다.** 그리고 그 문장에서 1호를 링크한다.

범위	건수	쓰는 경우	census.py
전체	413,577	실록 기록의 총량을 물을 때	옵션 없음
정본 계열	345,133	**기본값.** 순위, 분포	--edition original
왕대 비교용	344,325	왕대별 두께·밀도를 견줄 때. 정본에서 세종·세조실록 부록 808건을 뺀다	(수동 보정)
판본 짝	다섯 쌍	판본 차이를 볼 때. 두 판본을 모두 두고 짝으로 견준다	--edition revised · draft
중복을 뺀 전체 (v2.3)	377,930	실록 전체에서 "가장 많은" 것을 물을 때. 다시 쓴 판본(revised·draft)만 빼고 고종·순종실록은 넣는다	rank_dedup.py (기본값)

추가 규칙은 넷이다.

- **"실록 전체에서 가장 많은 것"을 묻는 편은 중복을 뺀 전체를 쓴다.** 정본 계열은 고종·순종실록을 빼므로
  그 시대에 나오는 사람·사물이 덜 잡힌다. 고종·순종실록은 그 시대의 유일한 실록이라 중복이 생기지 않는다.
  2호에서 정했다(송시열 정본 1,993건 → 중복을 뺀 전체 2,042건, 둘 다 7위). 왕대끼리 견주는 편은 여전히
  정본 계열(또는 왕대 비교용)을 쓴다
- **일제강점기 편찬분(32,797건)은 정본 계열에 들어 있지 않다.** 시계열 편에서 1864년 이후를
  그려야 하면 포함하되 막대나 선의 색을 달리하고, 캡션에 편찬 사정을 적는다.
- **건수만 쓰지 않는다.** 왕대·연대를 견주는 편은 밀도(해당 기사 ÷ 그 구간 기사 수)를 기본값으로 쓰고 건수는 괄호에 병기한다.
- **344,325건은 계산값이다.** 첫 사용 전에 부록 808건이 edition 값으로 구분되는지, 아니면 날짜 없는 기사로만 구분되는지 DB에서 확인한다.

3. 원고 틀

절 구성

```
[도입 2~3문장]              질문과 답을 함께 놓는다. AI 검색 요약이 여기를 인용한다
## I. 헤아리는 방법         범위와 건수. 1호 링크. 건너뛰지 않는다
[차트 1]                    인라인 SVG
## II~. 해석                각 절 첫 문장에 그 절의 결론. 절 수는 소재가 정한다
[차트 2]                    두 번째 지표(순위 편은 필수)
[이 숫자가 말하지 않는 것]  고정 상자
[출처 상자]                 집계 기준과 기준일 한 줄
[꼬리말]                    「실록통계청」 N호 · 목차 링크
```

고정 요소의 결정

닫는 상자는 「이 숫자가 말하지 않는 것」 하나로 둔다. 「편집장의 메모」는 넣지 않는다. 1호는 라인 전체의 기준 편이라 편집장의 메모로 닫았으나, 2호부터는 이 상자가 라인의
정체성이다. 세 라인이 각기 다른 상자로 닫혀야 독자가 라인을 구별한다.

순위 편(틀 A)에는 두 번째 지표를 붙인다. 2호에서 제목 등장률이 1위 윤필상의 정체를 드러냈다. 순위 하나만 놓으면 숫자가 과장된다. 두 번째 지표의 후보는 다음과 같다.

- 제목 등장률(--title-rate): 주인공인가, 명단에 있었을 뿐인가
- 기준을 바꾼 순위(--by ref): 표기가 흩어진 항목이 합쳐지면 순위가 바뀌는가
- 한 실록 집중도: 등장 기사의 몇 퍼센트가 한 왕대에 몰렸는가

H2는 로마 숫자로 쓴다. SEO 지침과 1호를 따른다. 2호는 발행 전에 맞춘다.

차트 색

색	뜻
`#2c3e50`	기본
`#9b6b52` (갈색)	그대로 견주기 어려운 항목(묘호, 부록 포함 실록, 일제강점기 편찬분, 중초본)

색	뜻
`#b0b0b0` (회색)	대조 불가 항목(인물 정보표 미등재 등)

갈색의 뜻을 1호와 2호가 이미 이렇게 썼다. 매 편 캡션에 한 번 풀어 준다.

분량 (v2.3 개정)

상한을 두지 않는다. 길이는 소재가 정한다. v2의 "본문 2,000~2,500자(공백 제외)"와 "해석 절이 네 개를
넘으면 둘로 가른다"는 규칙은 없앤다. 2호 첫 원고(2,083자)에 맞춰 정한 값이었고, 다른 라인과 견주면
지나치게 짧았다.

라인	공백 포함 본문
졸기 (대부분)	18,000자 이상
졸기 · 서지수	33,405자
통계청 2호 (10-22 개정)	약 8,900자 (표 포함)

재는 법: 다른 라인과 같은 잣대로 **공백 포함, HTML 태그·SVG 제외** 글자 수를 잰다. 표는 본문에 넣어 잰다.
v2까지 적던 "공백 제외" 값은 더 쓰지 않는다.

대신 지킬 것 두 가지.

- **한 편에 축 하나.** 길이가 아니라 축으로 가른다. 절이 늘어나도 모든 절이 같은 물음에 답하면 한 편이다.
  다른 물음에 답하는 절이 생기면 그때 떼어 낸다(예: 2호의 승정원일기 절은 시즌3 2화의 물음에 더 가깝다.
  지금은 2호의 결론을 받치므로 남긴다)
- **도입 2~3문장에 결론을 둔다.** 길어질수록 앞이 중요하다. AI 검색 요약과 네이버 미리보기가 여기를 가져간다

틀 F(3호)에 두었던 분량 예외는 이 개정으로 필요 없어졌다.

---

4. 반복 가능한 틀 일곱 가지

v2에 F와 G를 더했다. 3호가 틀 F의 첫 편이고, 2호(10-22 개정)가 틀 G의 첫 편이다.

틀	형태	차트	필수 요소
A 순위	상위 20	가로 막대	두 번째 지표
B 시계열	500년 추이	꺾은선 또는 세로 막대	밀도 기본, 일제강점기 편찬분 색 구분
C 달력	1년 주기	12칸 막대	하루당 평균 보정값
D 인명부	집단의 통계	가로 막대	인물 정보표에 오른 사람만 센다는 경고
E 대조	두 집단	좌우 대칭 막대	판본 짝의 분모 두 개를 모두 표기
F 왕대비교	재위 20년+ 임금 11명의 지표	가로 막대 + 순위표	총수와 10년당(재위 길이 보정) 두 기준 모두 표기, TOP3 표로 요약
G 통설검증	흔히 알려진 숫자 하나	기준별 비교표	통설을 먼저 놓고, 무엇을 어떻게 세었을 때 그 숫자가 맞는지 밝힌다. 통설의 숫자는 인용하지 않고 직접 센 값만 쓴다

틀 G의 규칙. 통설은 "흔히 ~라고 한다" 수준으로만 쓴다. 출처가 불분명한 글(카페 스크랩, AI 검색 요약)은
이름을 대지 않는다. 통설을 반박하는 데서 멈추지 않고, **통설이 맞아지는 조건**을 찾는다.
2호에서 송시열 1위는 틀린 말이 아니라 판본을 가리지 않고 셀 때만 맞는 말이었다. 이 조건이 그 편의 발견이다.

---

5. 로드맵 001~012

준비가 끝난 편을 앞에, 도구를 새로 만들어야 하는 편을 뒤에 놓았다.

호	제목안	틀	원료	준비	연동	슬러그안
001	조선왕조실록 41만 건에는 무엇이 들어 있나	·	전체 집계	**발행**	전 라인	what-is-joseon-annals (확정)
002	실록 최다 등장 인물은 송시열? 정본만 카운팅하면 7위, 1위는 윤필상	A·G	term 이름, count_word.py, census_by_king.py	**원고 완료 (10-22 개정)**	졸기, 시즌1 4화, 시즌3 2화	sillok-most-named-people
003	신하를 믿지 못한 영조	F	yeongjo_census.py	**2차 원고 진행**	게장·탕약 편(라인 미정)	sillok-yeongjo-trust
004	실록에 가장 많이 오른 지명 20곳	A	term 지명	즉시	시즌 7·8	sillok-most-named-places
005	실록은 무엇을 기록했나: 주제 분류의 분포	A	db.py cls	즉시	시즌5 3화	sillok-subject-classes
006	졸기는 어느 왕대에 몰려 있나	B	db.py jolgi	즉시	졸기	sillok-jolgi-by-reign
007	실록이 평가를	B	db.py style, 시즌4 측정값	절반 측정	시즌 4·9	sillok-historian-

호	제목안	틀	원료	준비	연동	슬러그안
	그만둔 해: 사론과 기사 길이					comments
008	실록의 1년: 기사가 가장 많은 달	C	calendar	**도구 필요**	·	sillok-busiest-month
009	지진·혜성·일식: 재이 기록의 500년	B	count_word.py	즉시	시즌6, 시즌3 3화	sillok-omens-500-years
010	실록 인물 본관 상위 20	D	person	**도구 필요**	졸기	sillok-person-bongwan
011	실록이 쓴 측정 단위: 도량형과 역법	A	count_word.py	표본 필요	시즌6	sillok-units-measures
012	가장 오래 쓰인 관직명	A·B	office	**도구 필요**	시즌9	sillok-office-titles
013	윤달의 기사: 조선이 끼워 넣은 시간	C	calendar	**도구 필요**	시즌6	sillok-leap-months
014	영의정 여덟 번은 사실인가: 관직 기록을 다시 세다	G	person_office	**도구 필요** (012와 같은 office)	012, 003	sillok-office-records

배치 근거

003은 2호 발행 뒤 급히 끼워 넣은 편이다. 원래 3호였던 지명 20곳은 4호로 밀렸다. 라이영이 "변덕스러운 영조"를 축으로 요청하면서 열한 임금의 교차 비교를 하게 되었고,
그 과정에서 틀 F(왕대 비교)가 생겼다. 틀 F는 005·007·009에서도 쓸 수 있도록 공용 도구
`yeongjo_census.py`로 남겨 둔다.

004~006은 명령 한 줄로 원료가 나온다. 004는 2호와 같은 명령에 type만 바꾸면 되므로
2호의 방법(두 번째 지표 포함)을 그대로 되풀이할 수 있다.

007은 v1의 4편과 8편을 합친 것이다. 시즌4 기획서 2-1·2-2절이 기사당 자수와 사론 밀도를
이미 대부분 측정했고, 9장 결정 대기에서 "정조 이후 사론 0 · 헌종 43.3자 · 철종 33.9자"를
실록통계청 한 편으로 빼기로 정했다. 두 곡선을 한 차트에 겹치는 편이 두 편으로 나누는
것보다 강하다. **시즌4 1화 발행 뒤에 놓는다.** 시즌4 1화는 폐위된 왕 셋의 비교이고,
007은 500년 전체의 곡선이다. 007에서 시즌4 1화를 링크한다.

008·010·012·013은 census.py에 명령을 더해야 한다. 그래서 뒤에 몰았다.
008과 013은 같은 calendar 명령을 쓰므로 한 번에 만든다.

009와 011은 count_word.py로 원료가 나온다. v1이 계획한 census.py series 명령은
만들지 않는다. 다만 2장의 밀도 규칙을 지키려면 연도별 분모가 필요하다(6장 참조).

---

6. 호별 설계 메모

002 · 실록 최다 등장 인물 — 10-22 개정

20명 순위 편에서 **틀 G(통설 검증)** 편으로 바꾸었다. 순위(틀 A)는 III·IV절에 남았다.

축: 흔히 송시열이 실록 최다 등장 인물이라고 한다. 판본을 가리지 않고 인물 ID로 세면 맞다(2,699건).
정본만 세면 7위(1,993건)이고 1위는 윤필상(2,612건)이다. 송시열의 차이 706건은 현

종개수실록·
숙종실록보궐정오·경종수정실록과 일제강점기 편찬분에서 온다. 윤필상은 7건이다.

확인된 수치 (2026-10-22, 10-27 범위 추가):

범위	기준	송시열	윤필상
정본	색인어 문자열	1,800 (8위)	2,453 (1위)
정본	인물 ID	1,993 (7위)	2,612 (1위)
전체	원문 문자열 (count_word.py)	2,460	2,473
전체	인물 ID	2,699 (1위)	2,619 (2위)
중복을 뺀 전체	인물 ID	**2,042 (7위)**	**2,612 (1위)**

- 宋時烈 원문 2,460건 가운데 655건이 정본 밖: 현종개수실록 522 · 보궐정오 87 · 고종·순종 35 · 경종수정 11
- 인물 ID 기준 판본별(edition_split.py): 송시열 2,699 = 정본 1,993 + 현종개수 548 + 보궐정오 96 + 경종수정 13 + 일제강점기 49. 윤필상 2,619 = 정본 2,612 + 중초본 6 + 현종개수 1
- **본문의 주 범위를 '중복을 뺀 전체'(377,930건)로 바꿨다(10-27).** 상위 20 차트도 이 범위로 다시 그렸다. 20위 이집(李㙫)은 인물 확인 필요
- **현종개수실록의 宋時烈(522)이 현종실록(437)보다 많다.** 허적 647 대 561, 정태화 601 대 526도 같다
- 인터넷에 도는 원문 검색 개인 집계(윤필상 2,474 · 송시열 2,473)를 count_word.py로 재현했다. 상위 열 명이 많아야 13건 차이
- 실록 25권별 1~3위(인물 ID, 부록 제외), 1위의 제목 등장률, 임금의 이름(묘호), 기준에 따라 1위가 바뀌는
  다섯 권, 승정원일기 대조(인조~철종)도 본문에 넣었다

절 구성 (7개): I 카운팅하는 방법 · II 송시열의 2,699건은 어디서 왔나 · III 정본의 1위, 윤필상 ·
IV 실록마다 자주 언급된 인물 · V 임금의 이름 · VI 기준을 바꾸면 1위가 바뀐다 · VII 다른 기록으로 카운팅하면

발행 전 남은 일:

- 분량: 공백 포함 약 8,900자(표 포함). v2.3 기준으로 문제없다

- 「이 숫자가 말하지 않는 것」 다섯째 항목이 VI절의 太宗 분석과 겹친다. 상자에서는 한 줄로 줄인다
- 시즌1 4화와 상호 링크. 둘 다 미발행이므로 지금은 평문. 4화 III절에 2호를 가리키는 문단을 넣었다
- 1호 링크, 목차 링크, 꼬리말
- 원문 검색 집계의 원 게시글은 찾지 못했다. 본문은 "인터넷에 도는 개인 집계"로만 쓴다

지난 점검 항목의 처리: 제목의 "002_"는 뺐다. H2는 로마 숫자다. 출처 상자를 넣었다. "성종실록 최다 인물 여섯 명" 대목은 개정에서 인물 ID 기준 일곱 명(신하)으로 다시 셌다.
金在魯 "인물표 미등재"는 인물 정보표 한자 이름의 BOM 문자 때문이었다.
fix_bom.py로 1,169건을 고쳤다.

003 · 신하를 믿지 못한 영조

v2의 원래 3호(지명 20곳)에서 바뀐 편이다. 틀 F(왕대 비교)는 이 편에서 처음 쓴다.

축: 재위 20년이 넘는 임금 열한 명을 놓고 일곱 가지 지표를 **총수와 10년당**으로 두 번씩
헤아린다. 두 기준 모두 1위면 그 임금의 특징이고, 총수에서만 1위면 재위 길이가 만든 숫자다.
영조는 영의정 임명과 친국 두 가지에서만 두 기준 모두 1위다. 나머지 다섯 가지는 재위 52년의
몫이거나 다른 임금이 앞선다.

지표: 영의정 임명, 친국, 유배, 거둔 명령, 처형, 약방·주방 관원 처벌, 전위 파동.

제목: 「신하를 믿지 못한 영조」(라이영 확정). 제목이 해석을 단언하므로, 본문은 이 말을
물음으로 받아 숫자와 반대 사례(1755년 5월 20일 "누가 나를 믿고 섬기겠는가")로 시험한다.

원고 구성 (2차):

1. 헤아리는 방법 — 열한 임금 표와 제목 기준의 함정
2. 두 기준 모두 1위 ① 영의정 41번 — 홍봉한 엿새 사례 포함
3. 여든 살 임금의 영의정 — 1773년 봄 한익모의 45일, 1774년 6월 사관의 평

4. 두 기준 모두 1위 ② 친국 186건 — 1755년 신치운 기사의 蟹醬 복선
5. 총수에서만 1위 — 유배와 거둔 명령
6. 영조가 1위가 아닌 것 — 처형(세종), 전위(선조), 약방·주방 처벌(정조)
7. 영조의 순위표

데이터 도구: `yeongjo_census.py` v3.3 (현재 scripts/). 왕대별 지표를 한 번에 뽑고 `--count` 정규식으로 임의의 제목 패턴을 셀 수 있다. 다른 편(졸기 분포 005, 사론 007, 재이 009)의
"임금별 분포"를 낼 때도 쓸 수 있도록 틀 F의 공용 도구로 둔다.

이 숫자가 말하지 않는 것 — 넷째: 친국을 많이 한 것이 신하 전체를 믿지 않아서인지, 자기에
관한 질문만은 직접 듣고 싶어서인지는 숫자 밖의 문제라고 적는다. 결론을 열어 둔다.

발행 전 남은 일:

- 1774년 6월 28일 인용 번역 감수 (申晦의 貪淫縱恣)
- 네이버 블로그 1호 주소 확인 후 1절·꼬리말 링크
- 「편집장의 메모」를 둘지 뺄지 결정 (라인 지침은 뺀다, 프로젝트 지침은 매 편 넣음 — 이 편은 메모가 결론을 한 번 더 비틀어 주므로 두는 쪽을 권함)
- 분량: 다른 라인 평균 25,000자(공백 포함) 기준으로 작성함. v2.3에서 상한을 없앴으므로 예외가 아니다

이어지는 편(게장과 탕약):
3호에서 "다음 편으로 넘긴다"고 적은 재료를 모은 편. 통계청 라인이 아니라 별도 연재 (시즌 미정)
또는 역사다큐로 뺀다.

- 1773년 2월 2일 탕제 사건 (kua_14902002_002) — 영조가 탕약을 땅에 쏟고 "오늘날 여러 신하는
  모두 풍도와 같은 자들이다"라고 한 날
- 1739년 10월 11일 감선 명령과 10월 14일 감선 환수 (kua_11510011_003, kua_11510014_002)
- 1755년 5월 20일·21일 신치운 신문·복주 (kua_13105020_002, kua_13105021_001) — 蟹醬 自服
- 1755년 10월 9일 《천의소감》 찬수자 인견 (kua_13110009_002) — 영조가 경종의 승하에 대해

직접 진술한 어제
- 경종수정실록 1724년 8월 20일 (ktb_10408020_001) — 蟹醬과 생시를 함께 든 뒤 밤에 가슴과
  배가 아팠다는 기사. 이것이 1755년 사건의 뿌리
- 같은 1773년 2월 2일 기사의 영조 하교 "오늘날 조선의 양반들은 이미 갈라섰으니, 임금이 어떻게
  임금 노릇을 하겠는가". 영조와 신하의 관계를 집약한 장면
- 1755년 5월 21일 사론: 1724년 8월 20일의 게장은 대비전이 보낸 것이 아니라 어주(御廚)에서
  올린 것이라고 밝힌다. 10월 9일 어제도 같은 내용을 영조가 직접 적는다
- 1755년 5월 20일 기사 끝: 소론 대신 이종성·박문수·이철보가 공초에 거듭 끌려 나왔으나 영조가
  끝내 묻지 않고 "오래 부린 신하를 한 사람의 말로 역적으로 의심하면 누가 나를 믿고 섬기겠는가"
  라고 했다는 대목. 3호에도 쓴다(반대 사례)
- 확인 사실: 약방·주방 관원 처벌 빈도에서 영조는 정조·숙종보다 낮다(3호 6절). 그러므로 이 편은
  "음식과 약 앞에서 늘 의심했다"가 아니라 "게장 이야기가 영조에게 얼마나 무거웠나"를 축으로 삼는다
- 축 후보: "게장과 탕약 — 1724년 경종의 게장에서 1755년 신치운의 자복까지"

004 · 지명 20곳

```
python scripts\census.py term --type 지명 -n 30 --edition original --
out 통계청003_지명_word.tsv
python scripts\census.py term --type 지명 -n 30 --edition original --
by ref --out 통계청003_지명_ref.tsv
python scripts\census.py term --type 지명 --diag
```

예상되는 함정: 같은 곳이 시대마다 다른 이름으로 불린다(한양 계열). 중국 지명이 상위에
섞일 수 있다. 2호의 순(舜)·요(堯)처럼 **조선 밖의 항목이 순위에 들어오면 숨기지 않고 절을
하나 준다.** --title-rate가 지명에도 작동하는지 먼저 확인한다. 작동하지 않으면 두 번째
지표는 한 실록 집중도로 바꾼다.

시즌 연동: 시즌7은 北京·燕京·瀋陽을, 시즌8은 對馬島·鬱陵島·釜山浦를 같은 명령으로 조회할 계획이다. 003은 상위 20만 보여 주고 개별 지명은 시즌에 남긴다.

005 · 주제분류

```
python scripts\db.py cls -n 40
```

v1은 "216종"이라고 적었다. **값을 다시 확인한 뒤 제목에 쓴다.** 한 기사에 여러 분류가 붙으므로
합계가 기사 수를 넘는다는 점, 분류 체계가 현대 편찬자의 관점이라는 점을 「이 숫자가 말하지
않는 것」에 둔다. 시즌5 3화가 "db.py cls 사법"을 먼저 보기로 했으므로, 004에서 사법 계열의
위치를 한 줄 보여 주면 시즌5가 그것을 받는다.

006 · 졸기의 분포

```
python scripts\db.py jolgi --stats
```

명령의 정확한 옵션은 db.py에서 확인한다. v1이 적은 "2,179건"은 도구로 다시 센다. **분모는 왕대 비교용(344,325건)** 이다. 졸기 수를 그대로 늘어놓으면 기사가 많은 왕대가 이긴다.
졸기가 0건인 인물(유자광 편)과 두 판본에 모두 졸기가 있는 인물(이순신 편)을 예외 사례로
링크할 수 있다.

007 · 실록이 평가를 그만둔 해

```
python scripts\db.py style --tsv > 통계청006_기사길이.tsv
```

사론 곡선은 시즌4 2-2절의 표를 쓴다. **史臣曰(하한)과 史臣(상한)을 둘 다 밝힌다.**
차이가 4.3%라 곡선의 모양은 같다는 시즌4의 판단을 그대로 옮긴다. 謹按과 論曰은 지표에서
뺀 이유를 한 문장으로 둔다.

세우지 않는 것: 사론이 사라진 이유. 편찬 방식의 변화인지 정치 환경인지는 해석이며, 이 편은 "그만둔 해"까지만 쓴다. 이유는 시즌9 사관열전의 몫이다.

남은 조회: 인조 이후 모든 실록의 기사당 자수. 시즌4 표는 효종에서 끝난다.
세종실록은 부록 750건을 뺀 값으로 다시 잰다.

008 · 실록의 1년

calendar 명령이 필요하다. **하루당 평균을 함께 낸다.** 작은달(29일)과 큰달(30일),
윤달이 낀 해가 섞여 있다. 정월에 기사가 몰릴 것이라는 짐작은 세기 전에는 쓰지 않는다.

009 · 재이 기록의 500년

```
python scripts\count_word.py 地震 彗星 日食 --by-year
```

두 글자 낱말이지만 count_word.py는 원문 문자열을 직접 세므로 db.py find의 bigram 오탐이
없다. 다만 **地震은 "땅이 흔들렸다"가 아닌 문맥**(비유, 인용)이 섞일 수 있으니 표본을
읽는다. 시즌3 3화(날씨)가 시즌6으로 넘어갈 경우 008이 그 편의 앞자리가 된다.

010 · 본관 상위 20

인물 정보표에 오른 사람만 센다. 2호에서 인물 정보표에 없는 항목(순)이 나왔다. (金在魯는 미등재가 아니라 BOM 문자 오류였다. 10-22 수정)
표에 오른 사람과 실록에 이름이 오른 사람은 다른 집단이다. 이 경고가 「이 숫자가 말하지
않는 것」의 첫 항목이다.

011 · 측정 단위

尺·斗·升 같은 한 글자 단위는 오탐이 감당되지 않는다. **두 글자 이상 표현으로만 센다.**
표본 확인이 가장 많이 드는 편이다. 밀리면 예비 소재와 바꾼다.

012 · 관직명

person_office에서 최초·최종 등장 연도를 뽑는다. 연속해서 쓰였는지, 끊겼다 다시 나왔는지를
구분한다. 시즌9 사관열전의 예고편 역할을 한다(사관 관직명의 수명).

013 · 윤달

007과 같은 명령을 쓴다. 윤달이 든 해와 들지 않은 해의 기사 수를 견준다.
시즌6에서 역법을 다룰 때 이 편을 링크한다.

---

7. 이름 규칙

파일명

```
001_실록이란무엇인가.md
002_실록에_이름이_가장_많이_오른_사람_20명.md
003_신하를_믿지_못한_영조.md
004_지명20곳.md
```

세 자리 호수로 시작한다. 볼트 폴더는 30*월간*실록통계청/20_원고를 유지한다.
publish_note.py의 DRAFT_DIRS가 이 경로를 보고 있으므로, 발행 주기와 맞지 않더라도
폴더명은 바꾸지 않는다.

제목

핵심 키워드를 앞에 두고 호수는 넣지 않는다. 호수는 꼬리말에만 적는다
("「실록통계청」 3호"). 시즌3 기획서도 같은 방식을 쓰기로 했다. 제목 전체는 사이트명을
포함해 60자 안팎으로 둔다.

1호의 현재 제목 "실록통계청 1호 · 41만 건에는 무엇이 들어 있나"는 이 규칙과 다르다.
"조선왕조실록 41만 건에는 무엇이 들어 있나"로 바꾸기를 권한다. focus_keyword가 제목
맨 앞에 오고, 제목 변경은 노트에서 고쳐 다시 발행하면 되므로 부담이 없다. 슬러그는
그대로 둔다.

슬러그

번호를 넣지 않는다. 시즌2가 화수 재배치 때 슬러그를 지키려고 번호를 뺀 것과 같은
이유다. 예비 소재가 끼어들면 호수가 밀린다. 모든 슬러그를 sillok-로 시작해 라인 안에서
찾기 쉽게 한다. 1호의 what-is-joseon-annals는 이미 발행되어 그대로 둔다.

카테고리

실록통계청(census)은 1호 발행과 함께 자리를 잡았다. v1의 "새로 만들 것" 표시는 지운다.

---

8. 도구 현황과 남은 일

명령	상태	쓰는 편
census.py term	**완성** (--by word·ref, --edition, --title-rate, --word, --diag, --sample, --out)	002, 004
yeongjo_census.py	**완성 (v3.3)**. 재위 20년+ 왕대 교차 비교. --count 정규식, --title, --years, --episodes, --forms, --post, --list, --sample	003, 틀 F 공용
census.py cls	**만들지 않는다.** db.py cls로 충분하다	005
census.py calendar	필요. 일수 보정값과 윤달 분리	008, 013
census.py person	필요. --by bongwan · born	010
census.py office	필요. 최초·최종 등장 연도, 관직별 임명 횟수·재임 기간·임명 나이	012, 014
census.py series	**만들지 않는다.** count_word.py --by-year가 대신한다	009, 011
census.py saron	**만들지 않는다.** 시즌4 측정값과 db.py style로 충분하다	007

확인할 것 하나: count_word.py --by-year가 연도별 전체 기사 수(분모)를 함께 찍는지.
찍지 않으면 밀도를 낼 수 없다. 그때는 count_word.py에 --density 옵션을 더하는 쪽이
새 명령을 만드는 쪽보다 가볍다.

공통 규칙(v1 유지): 모든 명령은 --out으로 UTF-8 파일을 남기고, 출력 끝에 분모를 찍는다.

---

9. 예비 소재

로드맵의 편이 데이터 문제로 밀리면 여기서 끌어온다.

이미 측정된 값이 있는 것

- **연산군일기 謹按 70건 83회** (시즌4 2-3절). 謹按이 한 왕대에 몰린 이유를 묻는 편
- ~~**커준을 바꾸면 1위가 바뀐다**~~ → **2호(10-22 개정)에 흡수했다.** 송시열 2,699건은 정본이 아니라
  전체 판본 값이었다. 표기·ID 문제는 2호「이 숫자가 말하지 않는 것」에 남겼다
- **왕은 죽은 뒤에 이름을 얻는다** (2호「왕의 이름」 절). 묘호가 어느 실록에서 불리는지를
  27명 전부로 넓힌 편
- **선조수정실록 기사당 239.1자의 이유** (시즌4 미결). 사론 때문이 아니라는 것까지만 확인됨

「조선왕조 기네스북」에서 가져올 것 (v2.2)

다음 카페 스크랩(2014년 3월, 원출처·근거 표기 없음)에 실린 '최다·최장·최연소' 항목 가운데
실록 DB로 다시 셀 수 있는 것만 골랐다. 그 글의 숫자는 인용하지 않는다. 틀 G의 통설로만 쓴다.

통설	검증 재료	판단
영의정 최다 최석정 8회 · 최장기 영의정 황희 18년	person_office, 003의 영의정 임명 집계	**014 핵심**
이조판서 최다 박종덕 18회 · 판서 최다 박종래 15회 · 한성판윤 최다 이가우 10회	person_office	014
최장기 호조판서 정홍순 11년 · 예조판서 신숙주 15년 · 대제학 서거정 23년 · 대사헌 권상하 13년	person_office (재임 기간 계산)	014. 기간은 임명·체직 날짜가 다 있어야 한다
최연소 영의정 이준 28세 · 최연소 병조판서 남이 27세 · 최연소 대제학 이덕형 31세	person.born + person_office	014. born 9999 오류 주의

통설	검증 재료	판단
최장기 관직 정원용 72년 · 김세민 7대 임금 70년 봉직	person_office 첫·끝 연도	014. 정원용은 철종실록 1위(2호)
사직상소 최다 정태화 37회	제목 정규식 (yeongjo_census.py --count)	단독 소재. 정태화는 2호에서 현종개수실록 609건으로 나온 인물
최장수 홍유손·권황 99세	person born·died	표본 확인 후
최다 아들 세종 18명 · 왕비 최다 가문 청주 한씨 6명	선원계보 CSV	적재 여부부터 확인

다루지 않는 것: 과거 최연소·장원(방목 자료가 DB에 없다), 김정희의 호, 이민구의 저술,
최초 개신교·천주교 신자, 족보. 실록 데이터로 셀 수 없다.

주의: 그 글은 윤증의 관직을 한 곳에서는 우의정, 한 곳에서는 정승이라고 적는 등 서술이 일정하지 않다.
숫자가 맞지 않으면 그 글이 틀렸을 수도, 우리 기준이 다를 수도 있다. 2호처럼 **어떻게 세면 그 숫자가 나오는지**를 찾는다.

시즌을 접을 경우 넘어오는 것

각 시즌 기획서가 "실록통계청으로 옮겨도 성립한다"고 적어 둔 편이다. 시즌이 진행되면 여기서
쓰지 않는다.

시즌	편	틀
시즌3	1화 승정원일기란 무엇인가, 6화	· , B
시즌4	1화 폐위된 왕의 실록, 12화	E, B
시즌5	1화 대명률·경국대전 분포, 9화, 12화 사면	B
시즌7	1화 연호 분포, 8화	B
시즌8	1화 倭와 日本	B

시즌7 가설 가(崇禎이 1644년 이후에도 나온다)가 깨지면 연호 편은 시즌에서 빠져 이리로 온다.

10. 운영

주간 작업 순서

v1과 같다. 통계청은 주 초반(월~화)에 쓴다. 데이터가 비어 나와도 소재를 바꿀 여유가 있다.

원료 선확보

003·004·005·008의 명령을 한 번에 돌려 TSV를 받아 둔다. 명령만 돌리면 되는 편이 넷이므로,
calendar·person·office 명령을 만드는 동안에도 발행이 끊기지 않는다.

고칠 문서

- 프로젝트_소스.md의 콘텐츠 구조 표: 통계청이 "월 1회"로 적혀 있다. 주 1회로 고친다
- 프로젝트_소스.md의 카테고리 계층: 옛 시즌 번호 체계다. 시즌2~9 재편에 맞춘다 (시즌4 기획서의 결정 대기 항목과 같은 일)
- v1 8장의 시즌 번호 충돌은 해결되었다. 조선의 과학은 시즌6, 사관열전은 시즌9다