Popular conceptions in serial publications

© Joris van Eijnatten

NIAS

KB Koninklijke Bibliotheek
Nationale bibliotheek van Nederland

Universiteit Utrecht

Faculty of Humanities

In praise of frequency

Generating and visualising the collective mentalities of the past

Iteration and Collective Mentality

- repetition and ‘mentality’
  - persuasion & repetition: propaganda & advertising
  - collective mentality: religion, teaching, media
- thick description & thin description (Geertz)
  - dense, meaningful contexts vs scarcity of meaning
- imagined community & collective mentality (Anderson)
  - synchronic (re)production vs diachronic repetition

- cultural effects
  - new content enters collective mentalities through iteration

In Praise of Frequency

claim:

- mundane and repetitive source material offers entry into a collective mentality

aim:

- integrate DH into historical discipline, through
  - teaching + research published on mainstream platforms

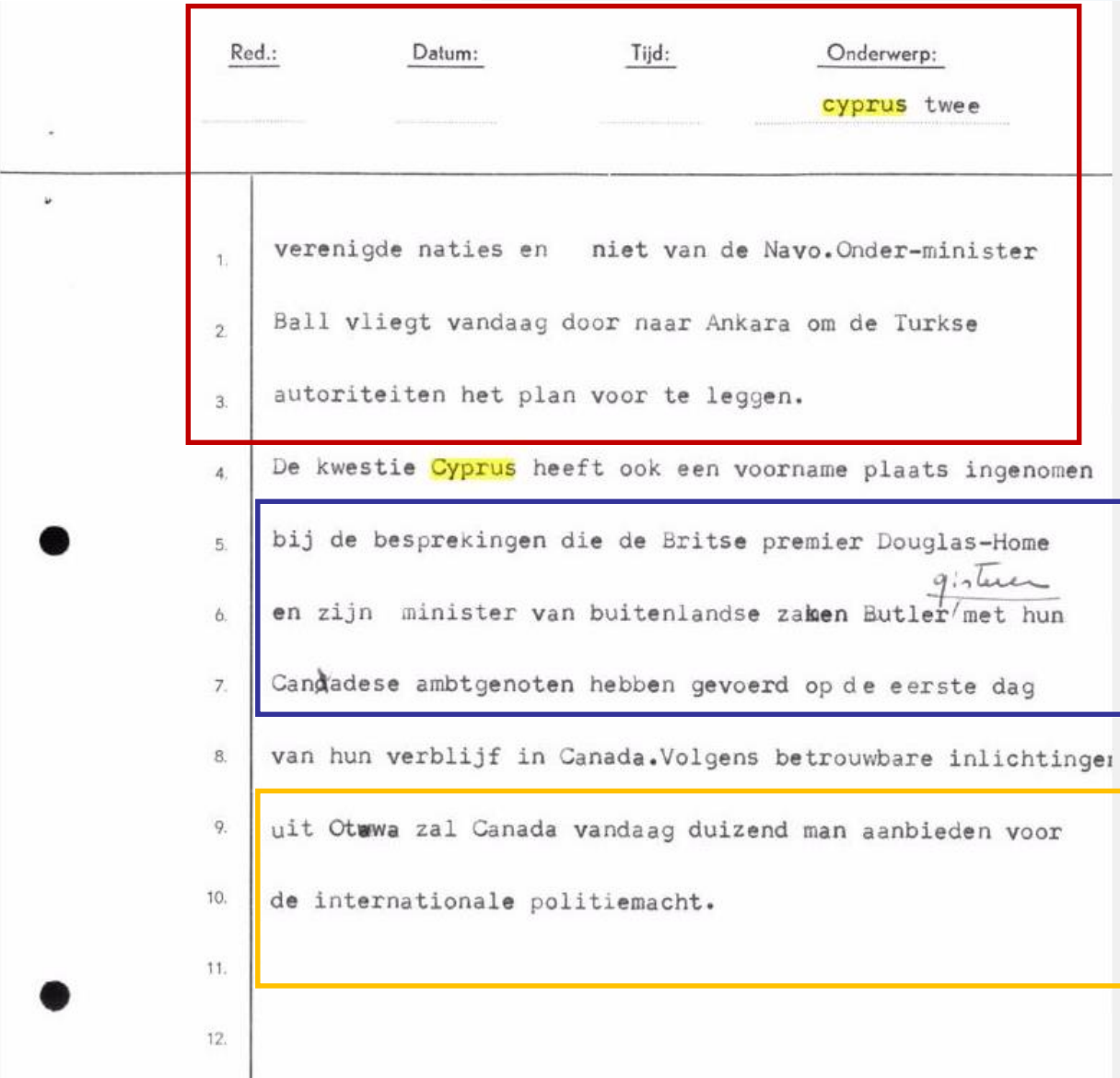
need:

1. highly flexible and malleable infrastructure
2. intuitively understandable DH methods ('radical simplicity': word counts, topic modelling, more complicated things like word embeddings...)
3. easy to access serial data + simple visualisations

Easy to access dataset

- Dutch Press Agency (ANP)
- National Library of the Netherlands (KB)
- metadata
  - identifier (e.g. anp_1964_02_11_6_ocr.xml)
  - content (radio message)
- 1,475,004 broadcasts
- 1937-1984
- brief

11 February 1964 (issue 6)



Red.: Datum: Tijd: Onderwerp:

cyprus twee

verenigde naties en niet van de Navo.Onder-minister
Ball vliegt vandaag door naar Ankara om de Turkse
autoriteiten het plan voor te leggen.
De kwestie Cyprus heeft ook een voorname plaats ingenomen
bij de besprekingen die de Britse premier Douglas-Home
en zijn minister van buitenlandse zaken Butler met hun
Canadese ambtgenoten hebben gevoerd op de eerste dag
van hun verblijf in Canada.Volgens betrouwbare inlichtinge
uit Otwwa zal Canada vandaag duizend man aanbieden voor
de internationale politiemacht.

Red. Datum: Tijd: Onderwerp: Cyprus twee 3. 10 11 12 13 14 15 verenigde naties en niet van de Navo. Onder-minister Ball vliegt vandaag door naar Ankara om de Turkse autoriteiten het plan voor te leggen. (…)

Britse premier Douglas-Home er/met en zijn minister van buitenlandse **zahen Butler' hun Can& adese** ambtgenoten (…)

uit Otwwa zal Canada vandaag duizend man aanbieden voor de internationale politiemacht.

Counting stuff: two techniques

1. Frequency counts of related words that co-occur in documents
   - e.g. ‘victim’ (*slachtoffer*) / ‘wounded’ (*gewond*) / ‘dead’ (*dode[n]*) + ‘accident’ (*ongeluk*) / ‘disaster’ (*ramp*)
   - different frequency counts of (a selection of) words
     1. the mere occurrence (i.e. is the presence of a word in a news item ‘true’ or ‘false’?)
     2. all occurrences
     3. al possible combinations
        - for five words in theory 31 combinations, in practice 26, such as: <victim + accident> or <victim + disaster>

Counting stuff: two techniques

2. Ngrams (1, 2, 3, 4, 5)
   - e.g. 4-gram ‘Dear Ladies and Gentlemen’ (*Geachte dames en heren*)

Visualising frequencies

- Three sets of options for graphs, concerning:
  1. the kind of graph;
  2. the elements in the graph;
  3. the kind of data.

Visualising frequencies

1. Kind of graph
   - timeline: points and lines
   - options
     - simple markers (i.e. scatter plot)
     - lines connecting markers (i.e. line graph)
     - smoothed line ('locally weighted scatterplot smoothing')
     - polynomial
     - univariate interpolation
     - combination of any or all of these (markers, normal lines, smoothed lines, polynomials, etc)
   - all variables underlying the various calculations to be set by the user.

Visualising frequencies

2. Elements in the graph
   - colours, font size and appearance of the x and y axes
   - grid
   - basic timeline: based on month or years
   - zoom function: slices from the timeline
   - vertical bars highlighting missing parts of dataset
   - headings and legends

Visualising frequencies

3. Kind of data
   - four options
   1. unique occurrences
   2. all occurrences of words
   3. occurrences of combinations of words
   4. occurrences of combinations of words as well as unique occurrences of words (dual-scale graph)

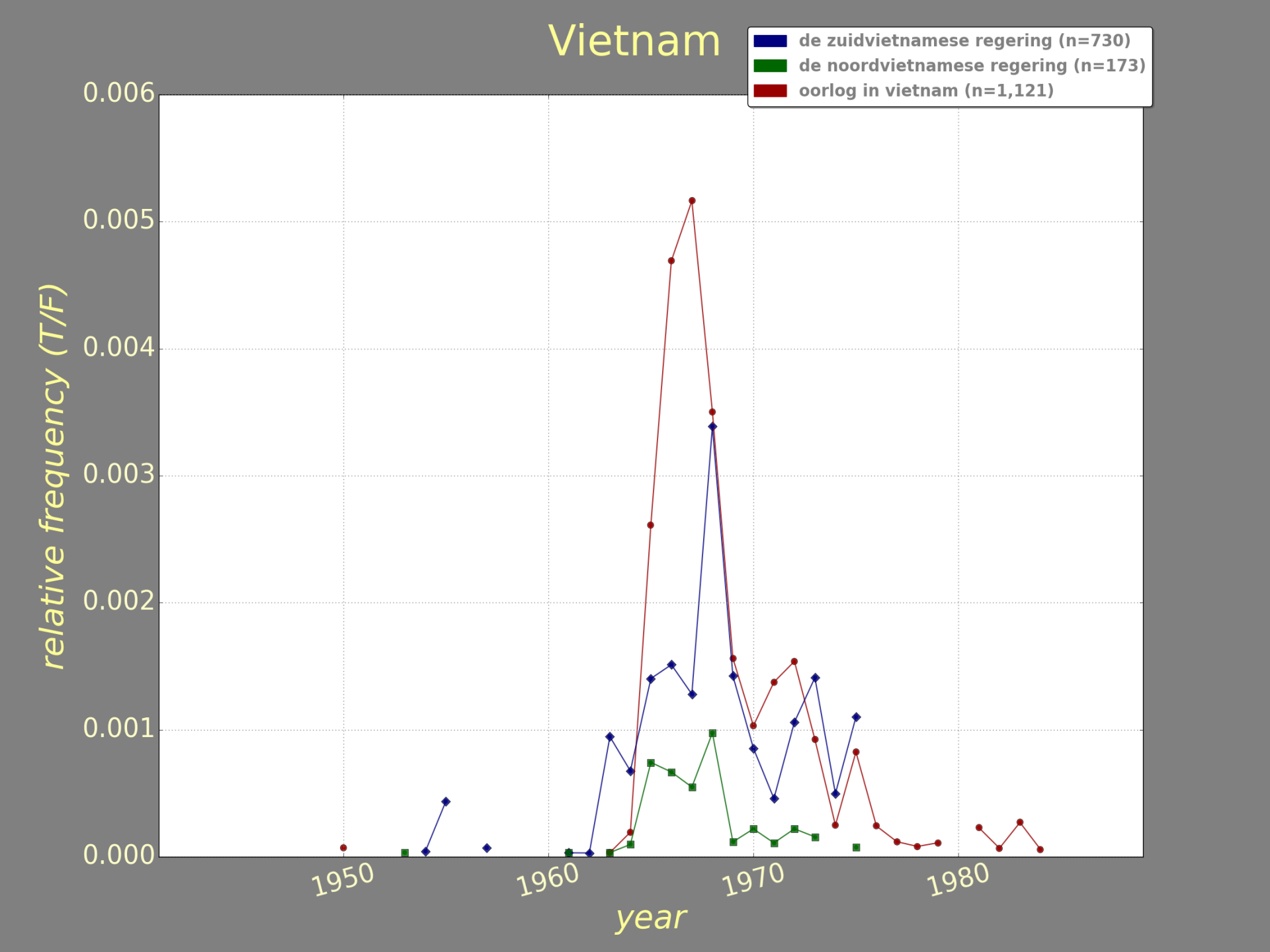
Case: 'Vietnam'

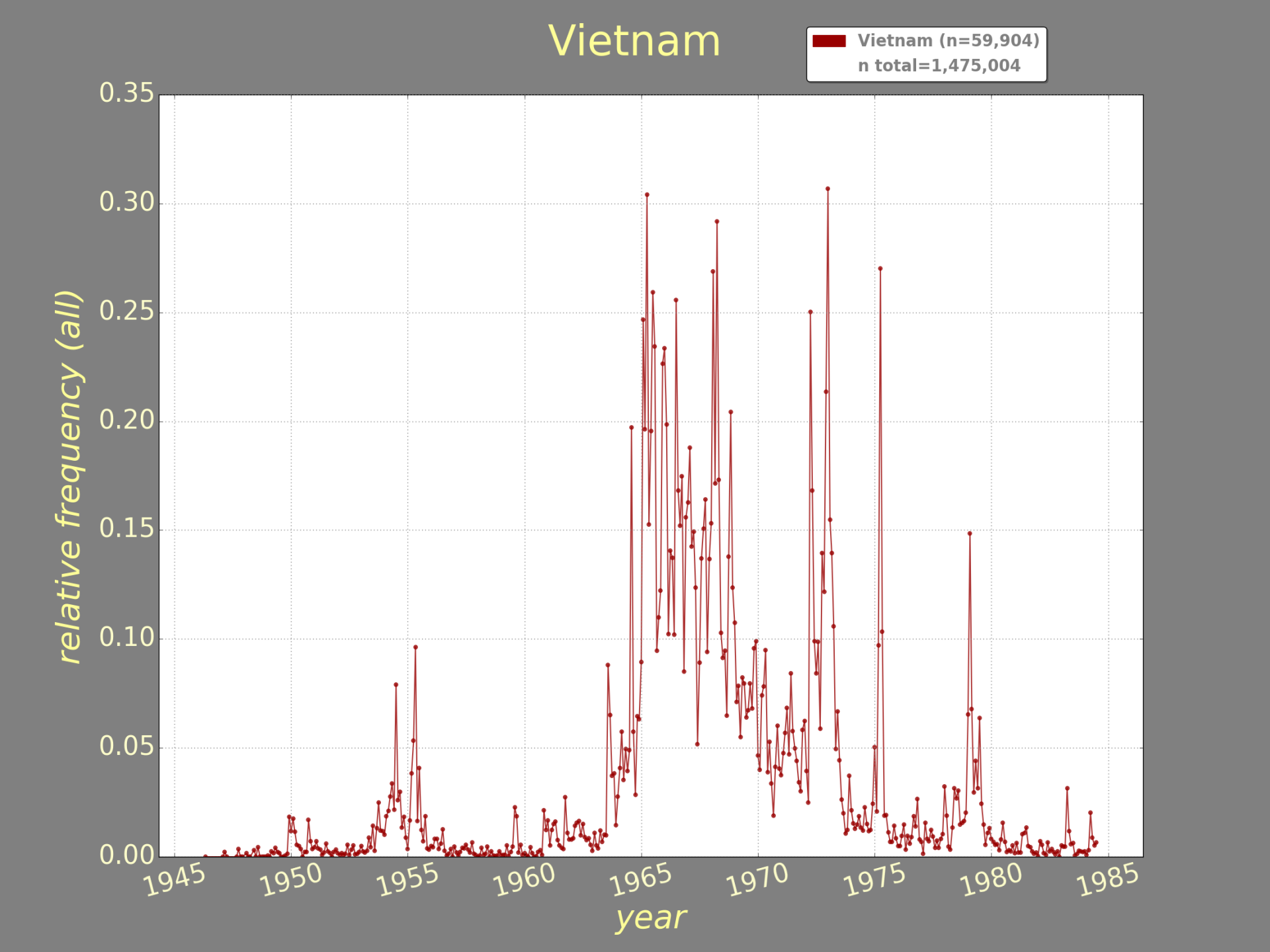
ANP radio news, January 1,1971

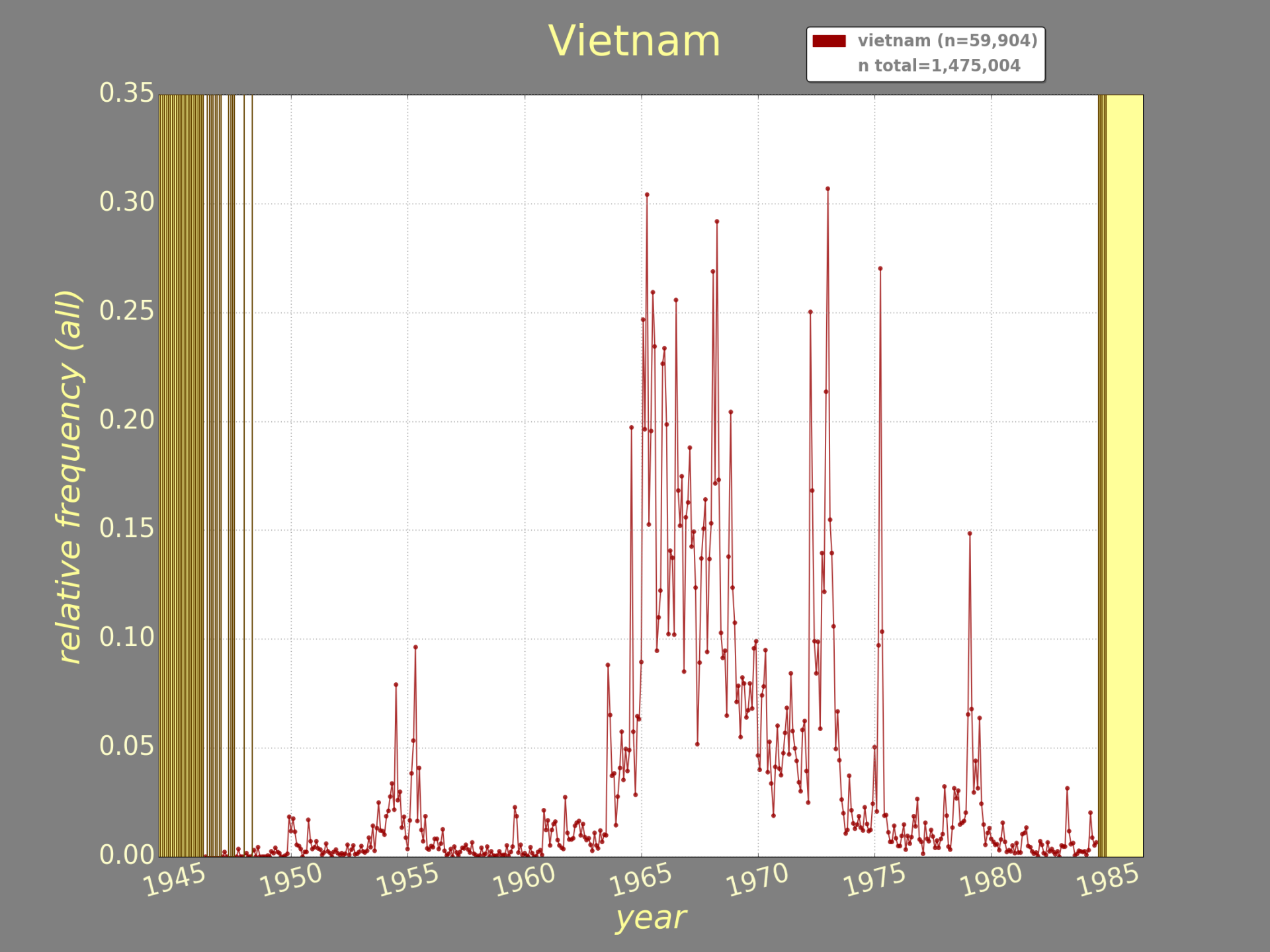
"The South Vietnamese president Van Thieu said at a new years' reception in Saigon that any peace settlement with the Vietcong will be without value as long as the communistist movement in South Vietnam has not been destroyed."

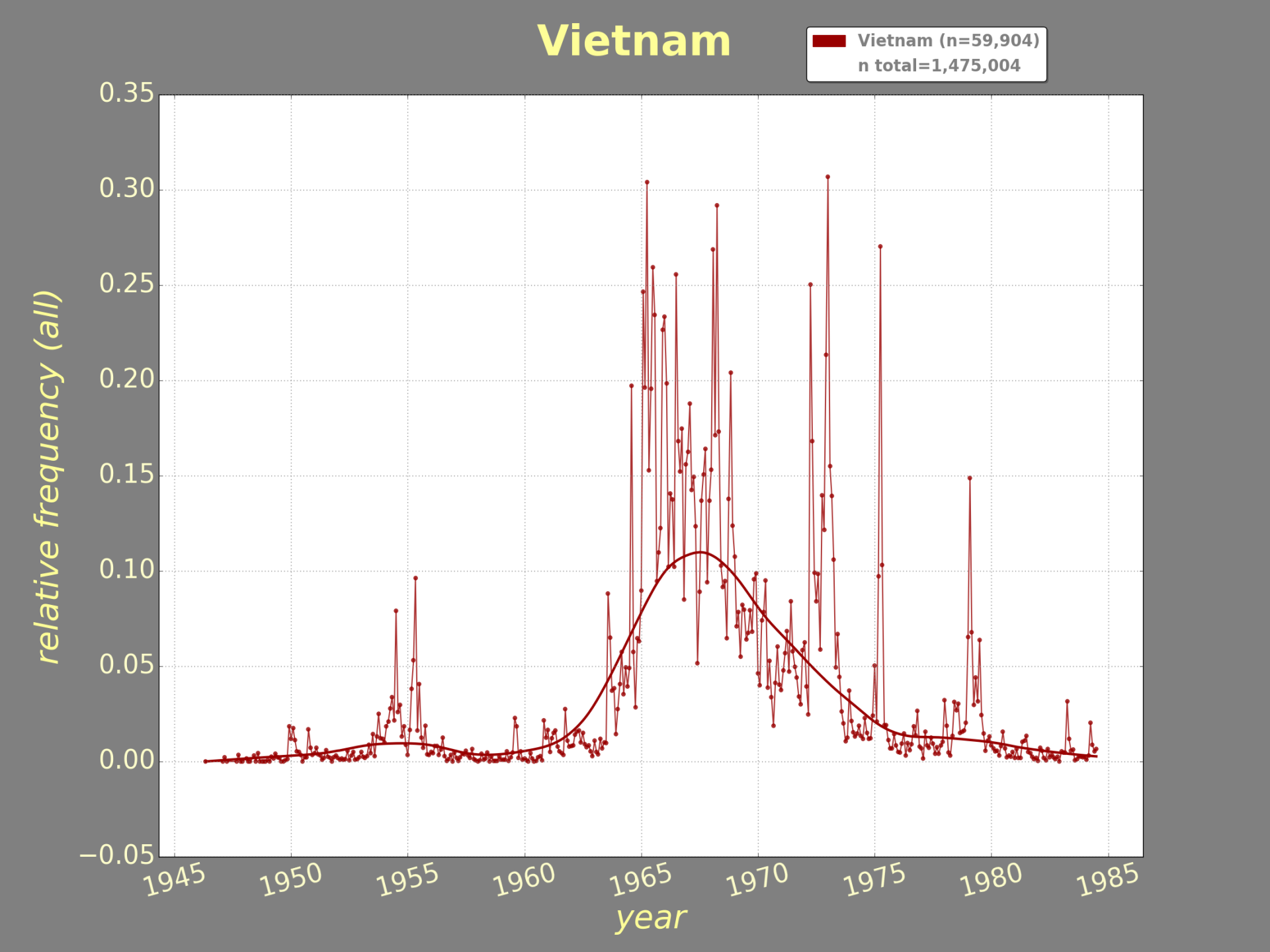
trigrams including ‘vietnam’

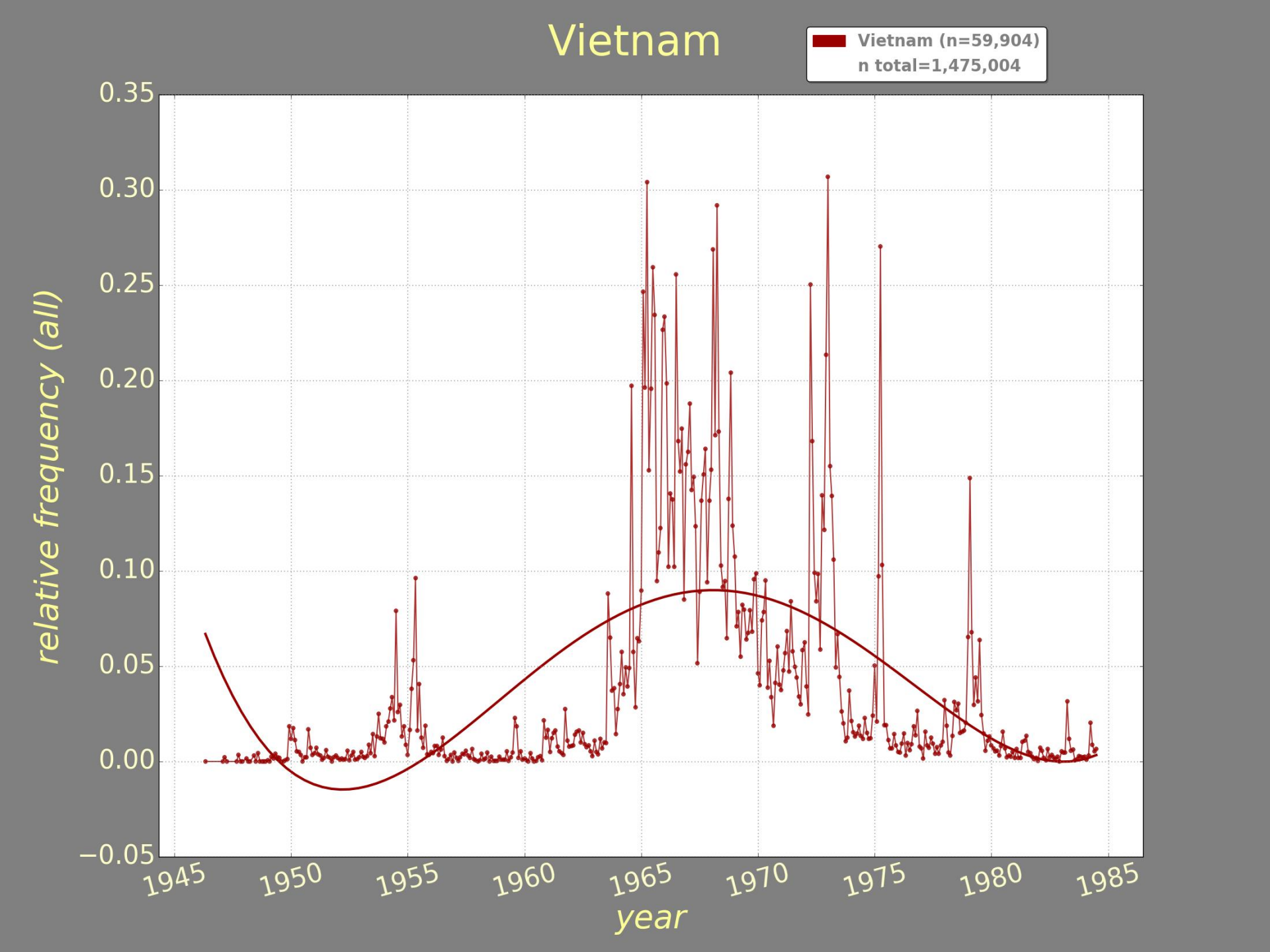
	ngram3	1967	1968	1969	1970	1971	1972	1973
12453471	**oorlog in vietnam**	226	151	67	47	63	77	53
16842873	**vietnam 10 11**	224	159	81	45	38	170	66
15952682	**van de noordvietnamese**	91	122	63	39	24	133	37
8747780	**in de vietnamese**	76	39	21	11	4	12	7
8916758	**in vietnam de**	76	48	18	15	14	23	32
11821686	**noord en zuidvietnam**	76	35	20	2	18	29	33
256156	**14 15 vietnam**	68	61	41	9	13	38	22
17271852	**vrede in vietnam**	62	65	20	5	8	62	44
4266423	**de noordvietnamese hoofdstad**	61	15	9	5	5	66	20
15982933	**van de zuidvietnamese**	61	199	81	41	63	129	60
4624731	**de zuidvietnamese regering**	56	146	61	39	21	53	81

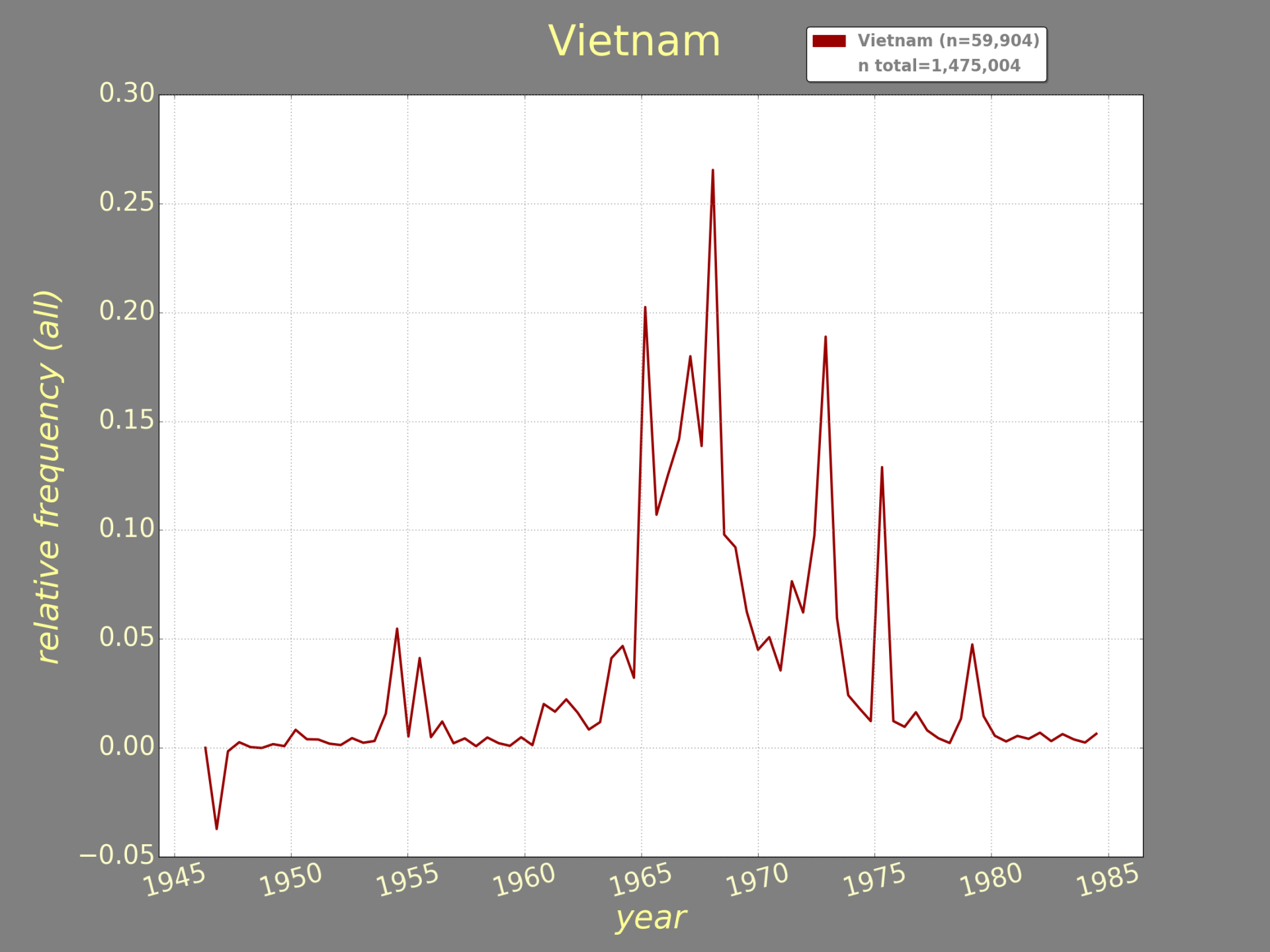

Vietnam
de zuidvietnamese regering (n=730)
de noordvietnamese regering (n=173)
oorlog in vietnam (n=1,121)
0.006
0.005
0.004
0.003
0.002
0.001
0.000
relative frequency (T/F)
1950
1960
1970
1980
year

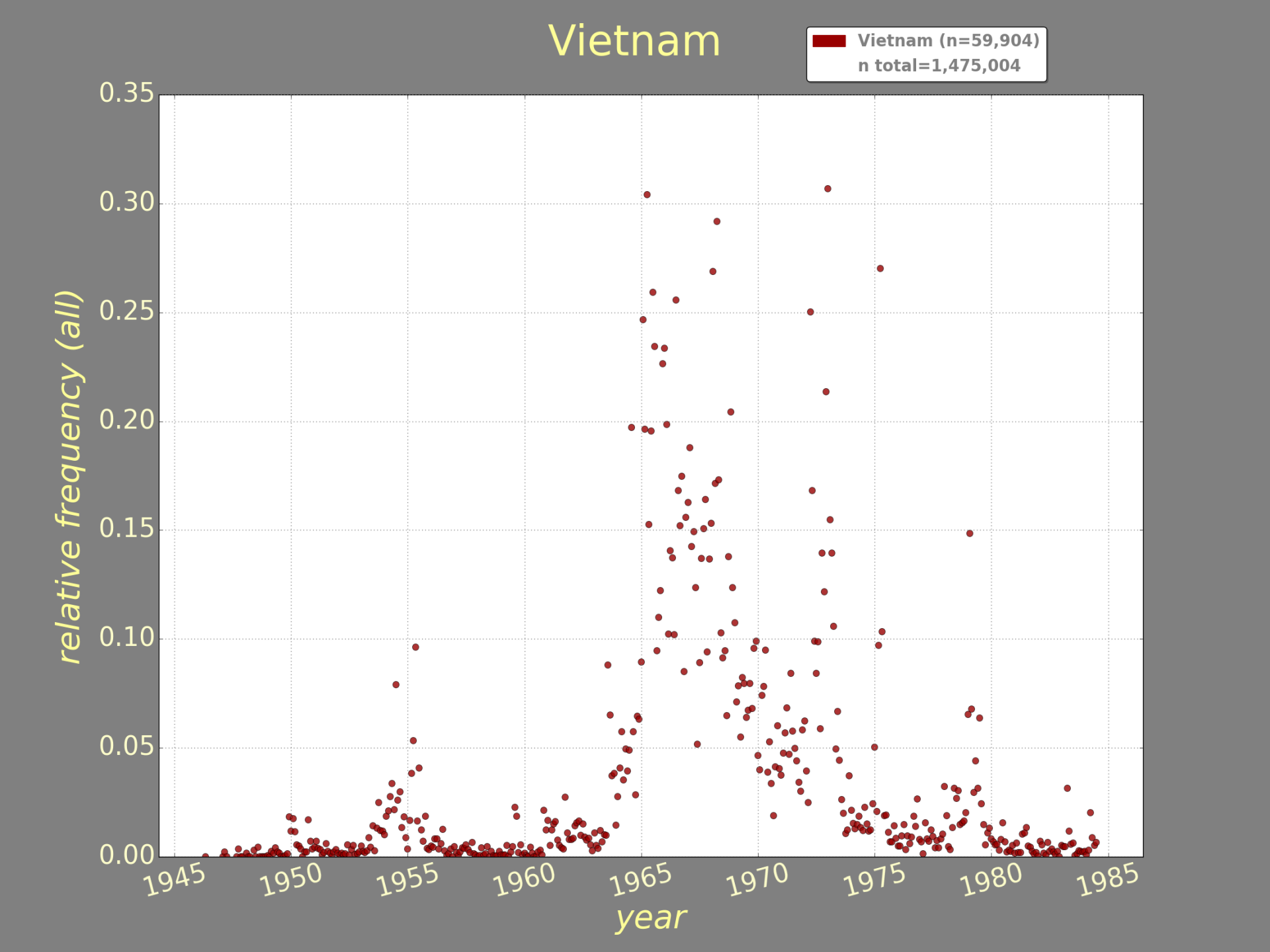

Vietnam
Vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1945
1950
1955
1960
1965
1970
1975
1980
1985
year

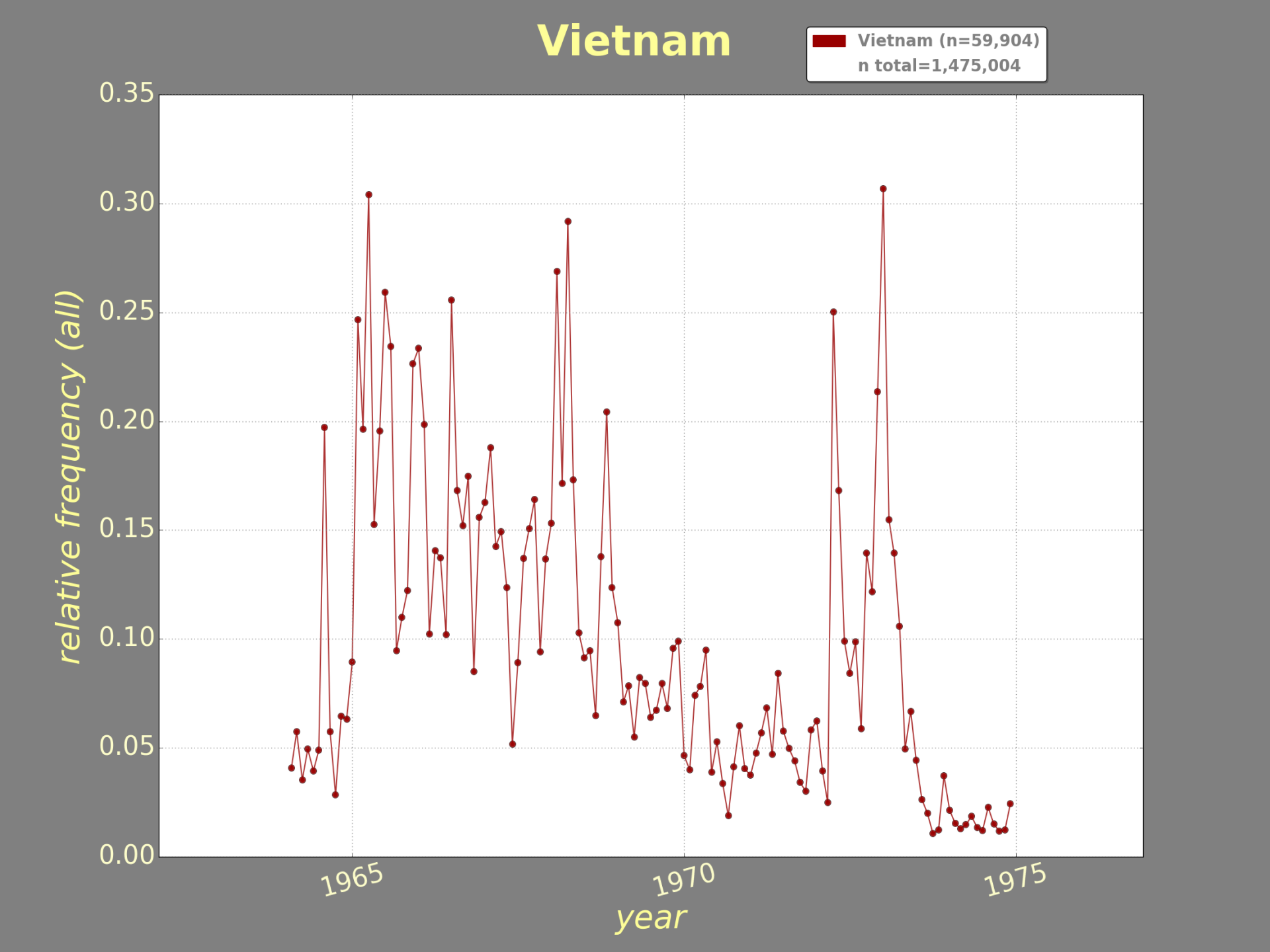

Vietnam
vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1945
1950
1955
1960
1965
1970
1975
1980
1985
year

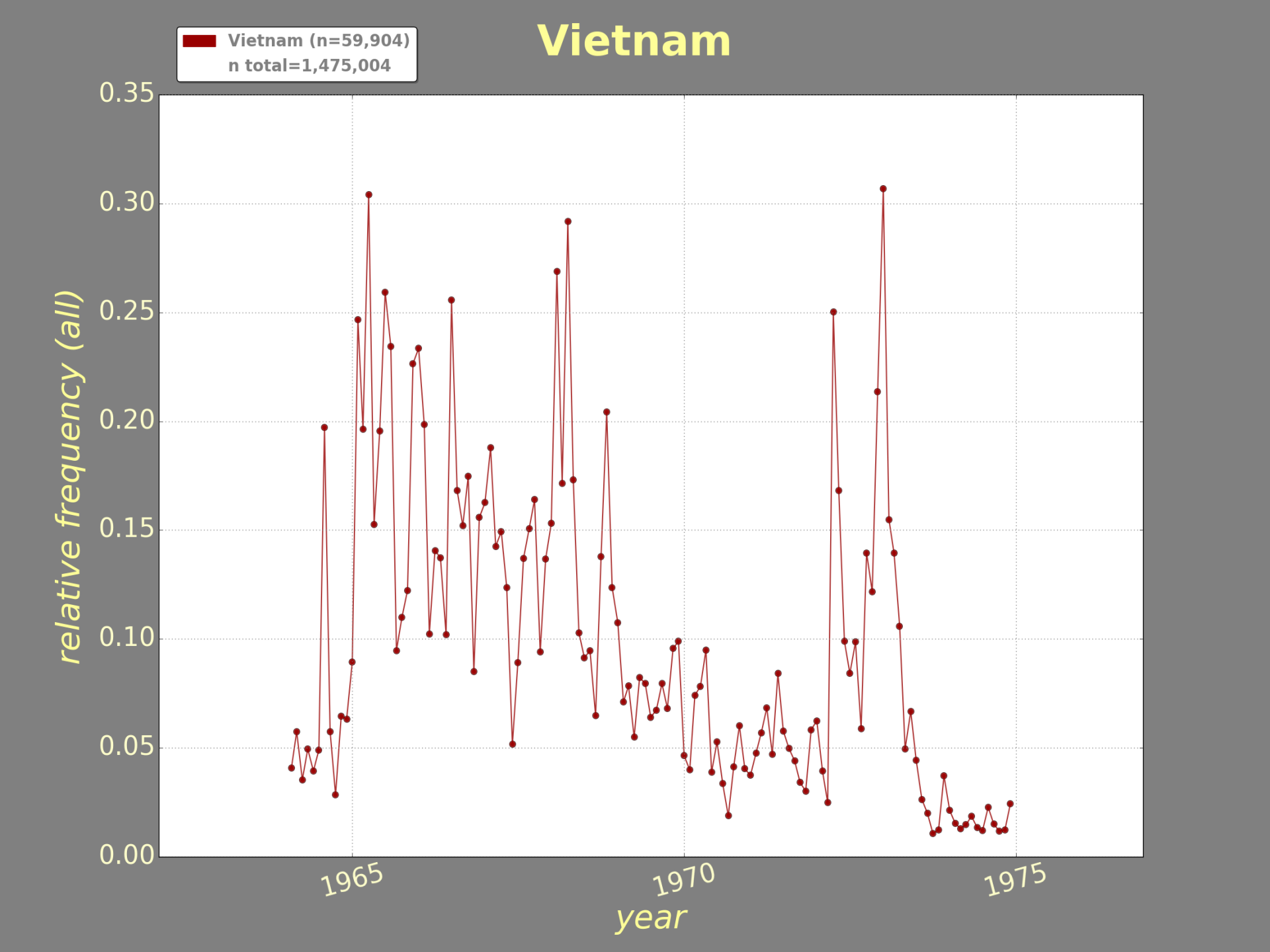

Vietnam
Vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
1945
1950
1955
1960
1965
1970
1975
1980
1985
year

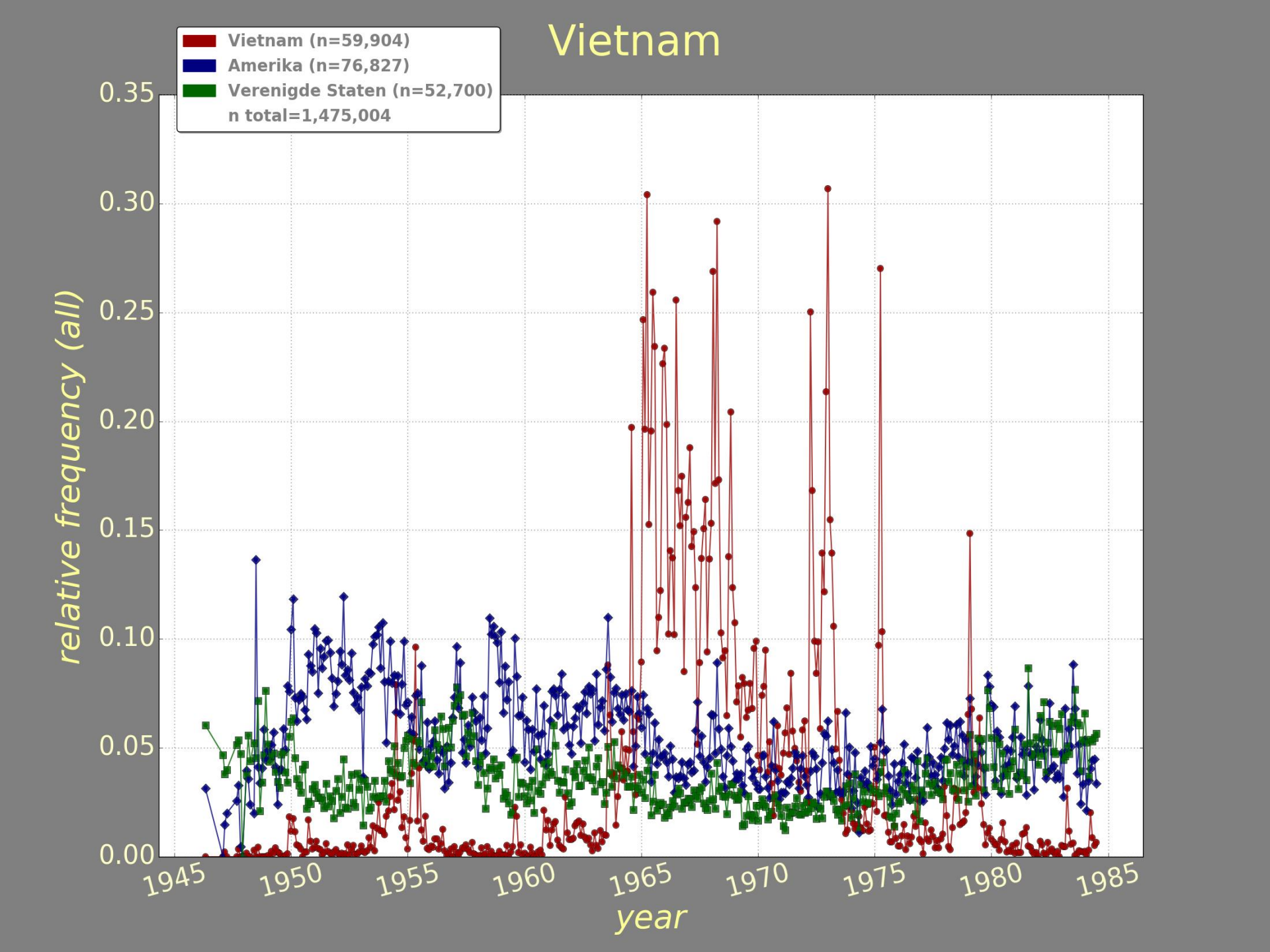

Vietnam
Vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
year
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
1945
1950
1955
1960
1965
1970
1975
1980
1985

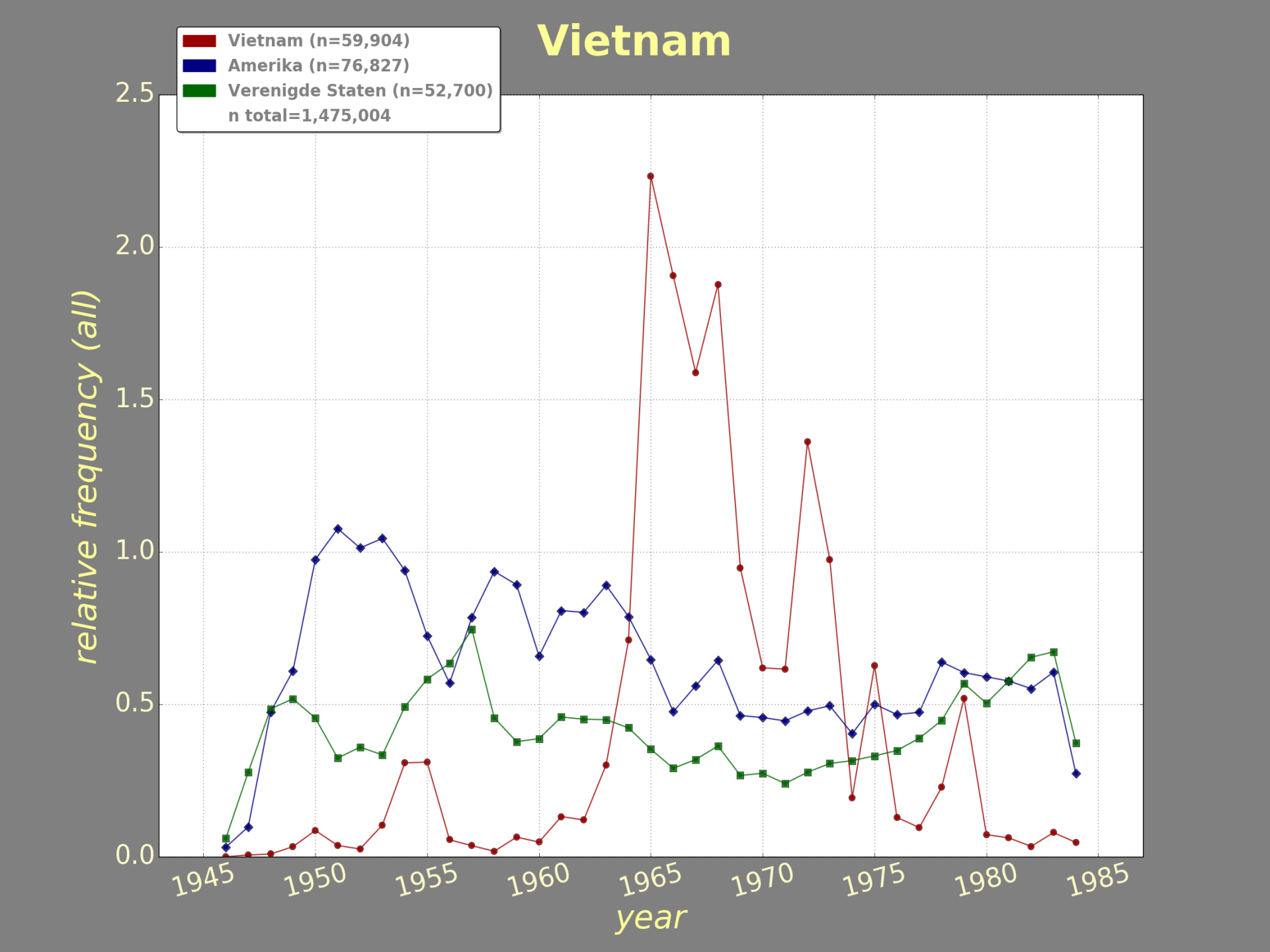

Vietnam
Vietnam (n=59,904)
n total=1,475,004
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
relative frequency (all)
1945
1950
1955
1960
1965
1970
1975
1980
1985
year

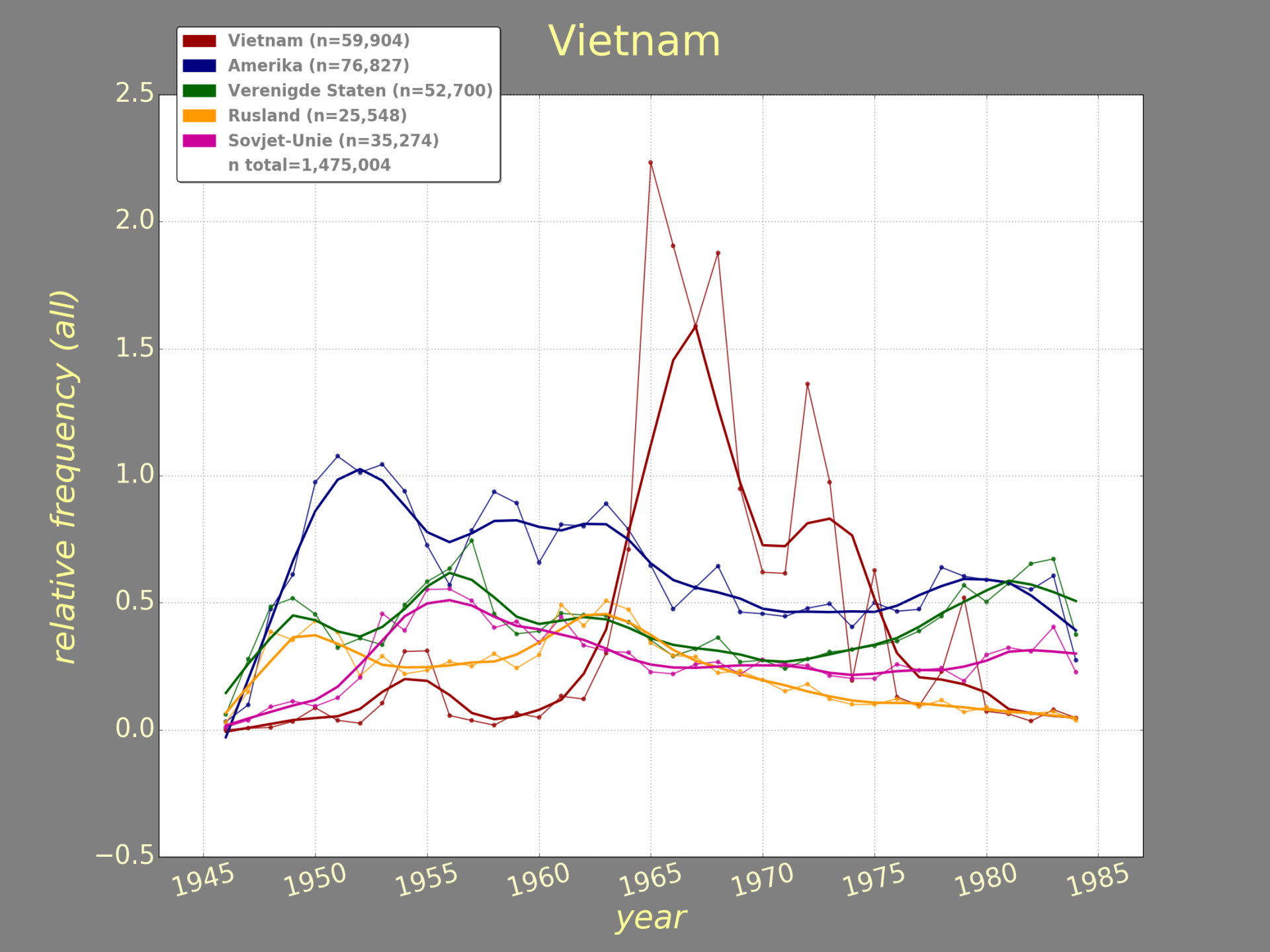

Vietnam
Vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1945
1950
1955
1960
1965
1970
1975
1980
1985
year

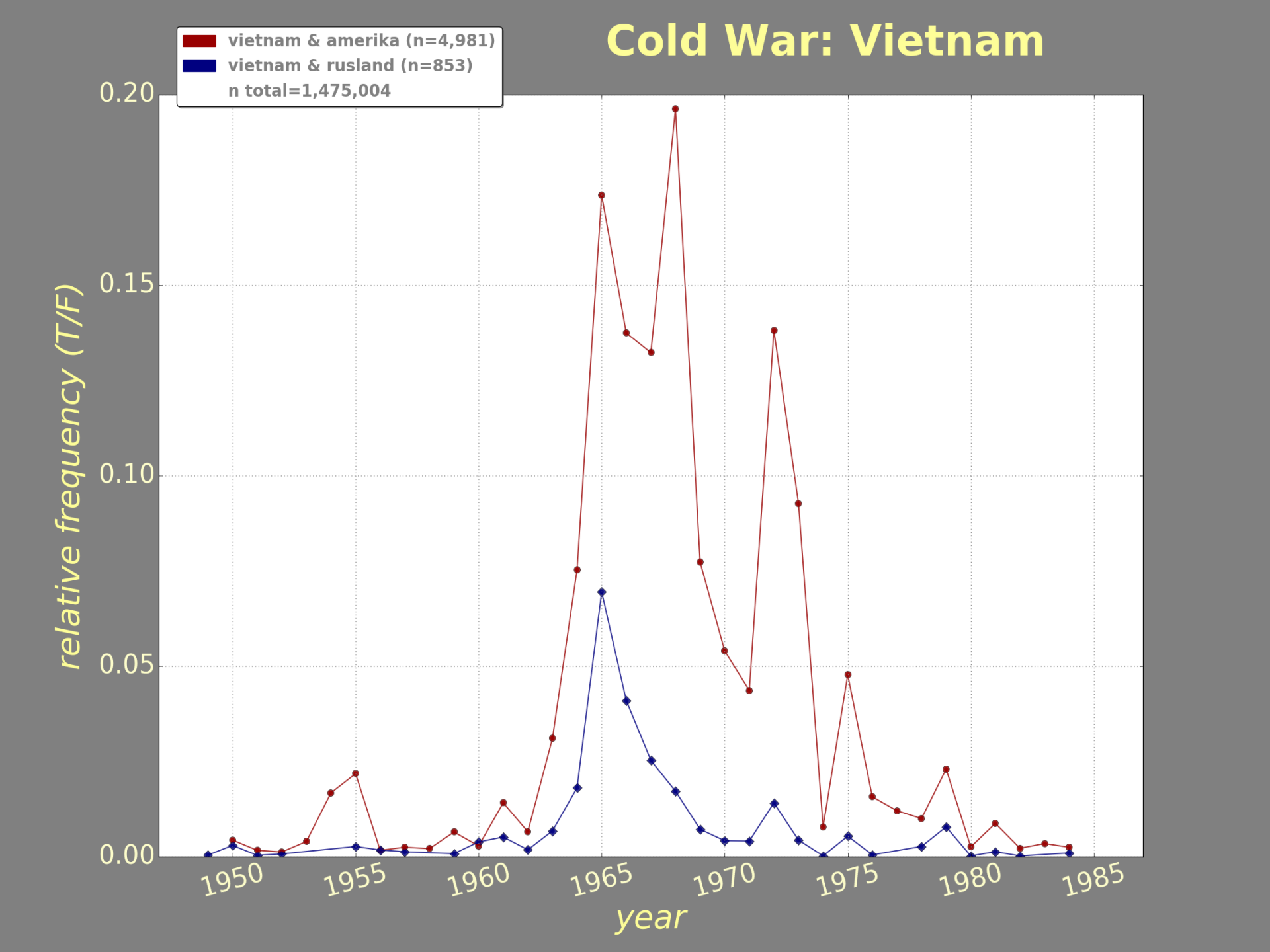

Vietnam
Vietnam (n=59,904)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1965
1970
1975
year

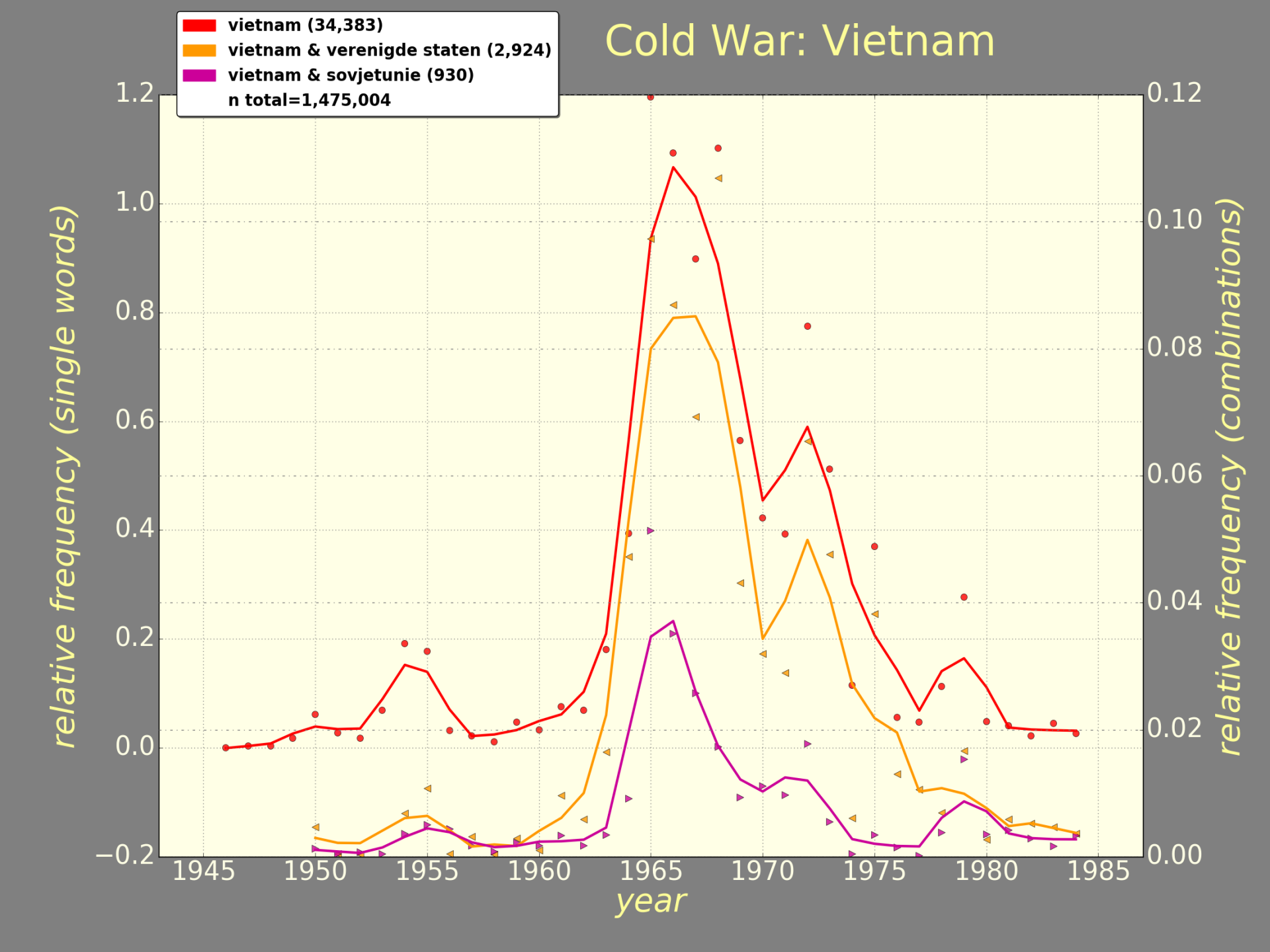

Vietnam (n=59,904)
n total=1,475,004
Vietnam
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1965
1970
1975
year


Vietnam
Vietnam (n=59,904)
Amerika (n=76,827)
Verenigde Staten (n=52,700)
n total=1,475,004
relative frequency (all)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1945
1950
1955
1960
1965
1970
1975
1980
1985
year


Vietnam
Vietnam (n=59,904)
Amerika (n=76,827)
Verenigde Staten (n=52,700)
n total=1,475,004
relative frequency (all)
year
0.0
0.5
1.0
1.5
2.0
2.5
1945
1950
1955
1960
1965
1970
1975
1980
1985


Vietnam
Vietnam (n=59,904)
Amerika (n=76,827)
Verenigde Staten (n=52,700)
Rusland (n=25,548)
Sovjet-Unie (n=35,274)
n total=1,475,004
relative frequency (all)
year
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
1945
1950
1955
1960
1965
1970
1975
1980
1985


Cold War: Vietnam
vietnam & amerika (n=4,981)
vietnam & rusland (n=853)
n total=1,475,004
relative frequency (T/F)
0.20
0.15
0.10
0.05
0.00
1950
1955
1960
1965
1970
1975
1980
1985
year


vietnam (34,383)
vietnam & verenigde staten (2,924)
vietnam & sovjetunie (930)
n total=1,475,004
Cold War: Vietnam
relative frequency (single words)
relative frequency (combinations)
year
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1945
1950
1955
1960
1965
1970
1975
1980
1985