

# MINING the MEDIATIZED MEMORY

#DHH16

Group 2: National broadcasting

Anu Koivunen  
Emily Ohman  
Julia Poutanen  
Matti Lassila  
Salla-Maaria Laaksonen  
Sezin Yaman

# MITÄ MISSÄ MILLOIN KANSALAISEN VUOSIKIRJA 1957

33107 L5502000187876452A000004152B00000D0F02334; SANA VIIKONVAIHEEKSI.; SANA VIIKONVAIHEEKSI.;:::::  
33108 L5502000187876452A000004152B00000D0F03922; PELASTA SIGYN.; OPERAATIO SIGYN.; PELASTA SIGYN; OPERAATI  
33109 L5502000187876452A000004152B00000D0F02614; KAIKKI MUKAAN.; KAIKKI MUKAAN.;:::::1977-01-15;;T  
33110 L5502000187876452A000004152B00000D0F02176; UUTISET JA SÄÄ.; UUTISET JA SÄÄ.;:::::1977-01-15;;T  
33111 L5502000187876452A000004152B00000D0F05474; TYÖPAIKKALIIKUNTA.; TYÖPAIKKALIIKUNTA.;:::::1977-01-15;;T  
33112 L5502000187876452A000004152B00000D0F03894; SÄHKEUUTISET.; SÄHKEUUTISET.;:::::1977-01-15;;T  
33113 L5502000187876452A000004152B00000D0F03155; TV-NYTT (1); TV-NYTT (1);:::::1977-01-15;;T  
33114 L5502000187876452A000004152B00000D0F01872; JAG UNDRAR SA FLUNDRA 3.; JAG UNDRAR SA FLUNDRA 3.;:::::  
33115 L5502000187876452A000004152B00000D0F05246; MINIRULETTI.; MINIRULETTI.;:::::1977-01-15;;TV  
33116 L5502000187876452A000004152B00000D0F02723; KAAKKOIS-SUOMEN VARTTI.; KAAKKOIS-SUOMEN VARTTI.;:::::  
33117 L5502000187876452A000004152B00000D0F01878; ILON JA SURUN RYPYT.; ILON JA SURUN RYPYT.;:::::1977-01-16;  
33118 L5502000187876452A000004152B00000D0F421740; PIENI TALO PREERIALLA. LITTLE HOUSE ON THE PRAIRIE.;P  
33119 L5502000187876452A000004152B00000D0F02179; UUTISET JA SÄÄ.; UUTISET JA SÄÄ.;:::::1977-01-16;  
33120 L5502000187876452A000004152B00000D0F02981; URHEILURUUUTU.; URHEILURUUUTU.;:::::1977-01-16;;T  
33121 L5502000187876452A000004152B00000D0F01855; PYHÄPÄIVÄN HARTAUS.; PYHÄPÄIVÄN HARTAUS.;:::::1977-01-16;  
33122 L5502000187876452A000004152B00000D0F01584; GUDSTJÄNST MED FRALSNINGSARMEN.; GUDSTJÄNST MED FRALSN  
33123 L5502000187876452A000004152B00000D0F01149; KYSSY POIS - KUESI JULGESTI 2.; KYSSY POIS - KUESI JUL  
33124 L5502000187876452A000004152B00000D0F03129; TV-NYTT (2); TV-NYTT (2);:::::1977-01-16;;Om  
33125 L5502000187876452A000004152B00000D0F003814; TULIKO LAMA JÄÄDÄKSEEN?; TULIKO LAMA JÄÄDÄKSEEN?;:::::  
33126 L5502000187876452A000004152B00000D0F01207; LENTOPALLOA/NAISTEN SM-SARJA.; LENTOPALLOA/NAISTEN SM-S  
33127 L5502000187876452A000004152B00000D0F01450; NAISIA MERKEISSÄ 1. MINÄ EN KUULE. NAISIA MERKEISSÄ 1.  
33128 L5502000187876452A000004152B00000D0F00383; SÄHKEUUTISET.; SÄHKEUUTISET.;:::::1977-01-17;;T  
33129 L5502000187876452A000004152B00000D0F02182; UUTISET JA SÄÄ.; UUTISET JA SÄÄ.;:::::1977-01-17;;T  
33130 L5502000187876452A000004152B00000D0F003831; TV-NYTT (1); TV-NYTT (1);:::::1977-01-17;;T  
33131 L5502000187876452A000004152B00000D0F00384; KOKONAIN TOINEN JUTTU 2.; KESTI.; KOKONAIN TOINEN JUT  
33132 L5502000187876452A000004152B00000D0F01323; MATII SAA POLKUPYÖRÄN.; MATTI SAA POLKUPYÖRÄN.;:::::  
33133 L5502000187876452A000004152B00000D0F00323; ILLASATU: HIIRI MORSIAMENA.; ILLASATU: HIIRI MORSIAM  
33134 L5502000187876452A000004152B00000D0F116451; ASIA ON PIHVIT.; ASIA ON PIHVIT.;:::::1977-01-18;;T  
33135 L5502000187876452A000004152B00000D0F005250; ILLASATU: HIIRI MORSIAMENA.; ILLASATU: HIIRI MORSIAM  
33136 L5502000187876452A000004152B00000D0F001055; KAMARIMUSIIKKIA KUHMOSSA.; KAMARIMUSIIKKIA KUHMOSSA.;  
33137 L5502000187876452A000004152B00000D0F002244; MITÄ MINUSSA TAPAHTUU 2.; MITÄ MINUSSA TAPAHTUU 2.;  
33138 L5502000187876452A000004152B00000D0F001232; MENE JA TIEDÄ 77/1.; MENE JA TIEDÄ 77/1.;:::::1977-01-18;  
33139 L5502000187876452A000004152B00000D0F001296; NOPPA.; NOPPA.;:::::1977-01-18;;T; Oma tuotanto;  
33140 L5502000187876452A000004152B00000D0F03386; TASAVALLASSA TAPAHTUU.; TASAVALLASSA TAPAHTUU.;:::::  
33141 L5502000187876452A000004152B00000D0F02924; SEX /6 DAGAR I BENGT ANDERSSONS LIV 2. KUUSI PÄIVÄÄ  
33142 L5502000187876452A000004152B00000D0F03095; SÄHKEUUTISET.; SÄHKEUUTISET.;:::::1977-01-18;;T  
33143 L5502000187876452A000004152B00000D0F01319; ASIA ON PIHVIT.; ASIA ON PIHVIT.;:::::1977-01-18;;T  
33144 L5502000187876452A000004152B00000D0F03146; TRAFF 2/1.; TRAFF 2/1.;:::::1977-01-18;;T  
33145 L5502000187876452A000004152B00000D0F03926; PIKKU KAKKONEN.; PIKKU KAKKONEN.;:::::1977-01-18;;T  
33146 L5502000187876452A000004152B00000D0F005251; PIKKU KAKKONEN.; PIKKU KAKKONEN.;:::::1977-01-18;;T  
33147 L5502000187876452A000004152B00000D0F002141; UUTISET JA SÄÄ.; UUTISET JA SÄÄ.;:::::1977-01-18;;T  
33148 L5502000187876452A000004152B00000D0F003032; TV-NYTT (2); TV-NYTT (2);:::::1977-01-18;;T

AN INDEX OF THE  
NATION?

# Research Frame

**Mapping the “mediatized centre” and the agenda  
setting/indexing power of YLE:**

- What are the major issues covered in the data over time?
- The politics of the TV flow (representation of persons, parties, genders, topics)?

# Yle Data

- **YLE Metro TV-archive metadata** collection from 31.12.1990 backwards to 1.1.1957
- Total 96 838 rows of records, organized in xml-files (Tietomalli Metro 9.4.2013)

```
<Object type="default" mdclass="PROGRAMME">  
<GUID dnname="">20090915052309217202715502000187876452A0000004152B00000D0F128071</GUID>  
<Meta name="FILM_ARC_TIME" format="string" rate="" />  
<Meta name="TVAR_PUBLICATION_READY" format="string" rate="" />  
<Meta name="MEDIA_DELETION_CANDIDATE" format="string" rate="" />  
<Meta name="PRODUCING_DEP_NAME" format="string" rate="" />  
<Meta name="PRODUCTION_CODE" format="string" rate="">44891305000</Meta>  
<Meta name="FIRSTRUN_TIME" format="string" rate="" />  
<Meta name="DELETION_TRANSFERRED_TO_PLASMA" format="string" rate="" />  
<Meta name="KEYWORDS" format="string" rate="" />  
<Meta name="RIGHTS_LIST3" format="string" rate="" />  
<Meta name="TRANSMISSION_READY" format="string" rate="" />  
<Meta name="SE_TITLE" format="string" rate="" />  
<Meta name="RIGHTS_INDICATOR" format="string" rate="" />  
<Meta name="TRANSFERRED_TO_KAVI" format="string" rate="" />  
<Meta name="TRANSFERRED_MEDIA_TO_PLASMA" format="string" rate="" />  
<Meta name="RIGHTS_DESCRIPTION" format="string" rate="">Tv-oikeudet. </Meta>
```

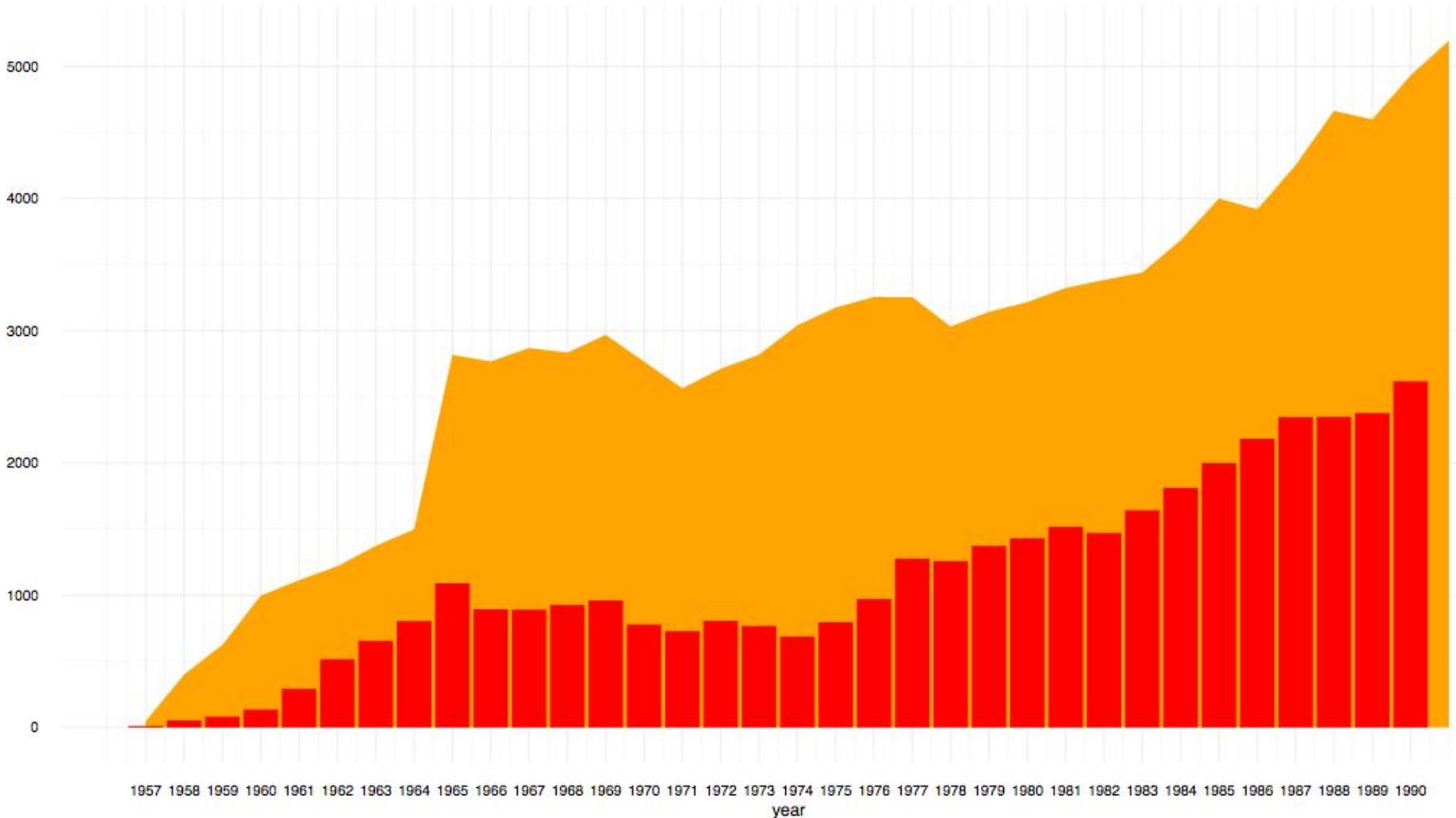
# How did we approach the data?

- **Data enrichment:** categorization (**79 670 out of 96 839**)
  - news (UUTI), actualities (AJAN), factual programming (ASIA), feature films (ELOK), domestic and foreign fiction (KFIK/UFIK), children's programmes (LAST), sports (URHE), lifestyle (LIFE), reality (REAL), and other.
- Creating subsets with **Open Refine** (topics, themes, etc.)
- **Lemmatization** of the text
- **Structural Topic Modelling**
- **Named Entity Recognition**

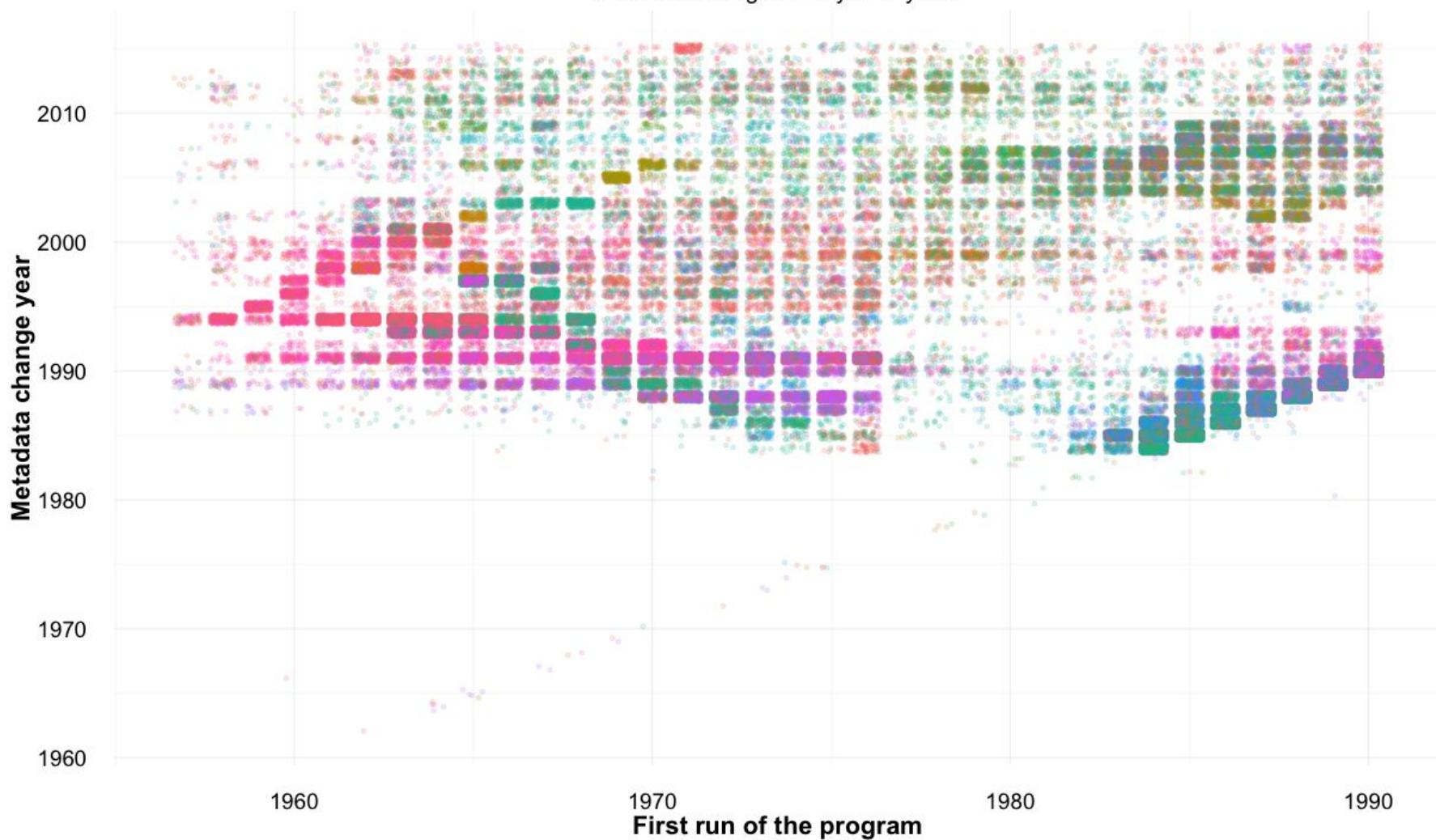
# What's in the data?

---

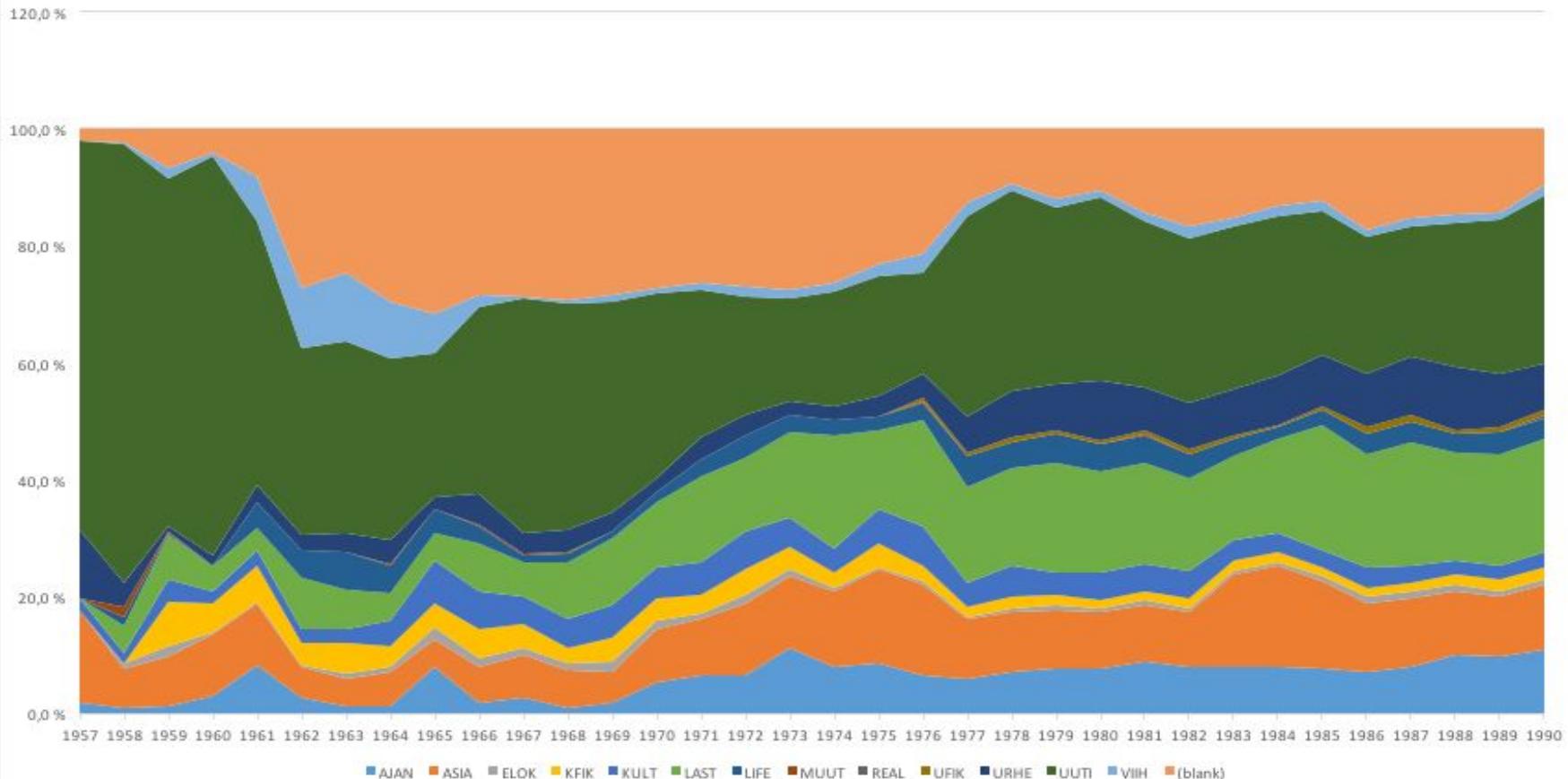
Total program time and total airtime in minutes from 1957–1990



Metadata change activity over years



# Programme categories per year



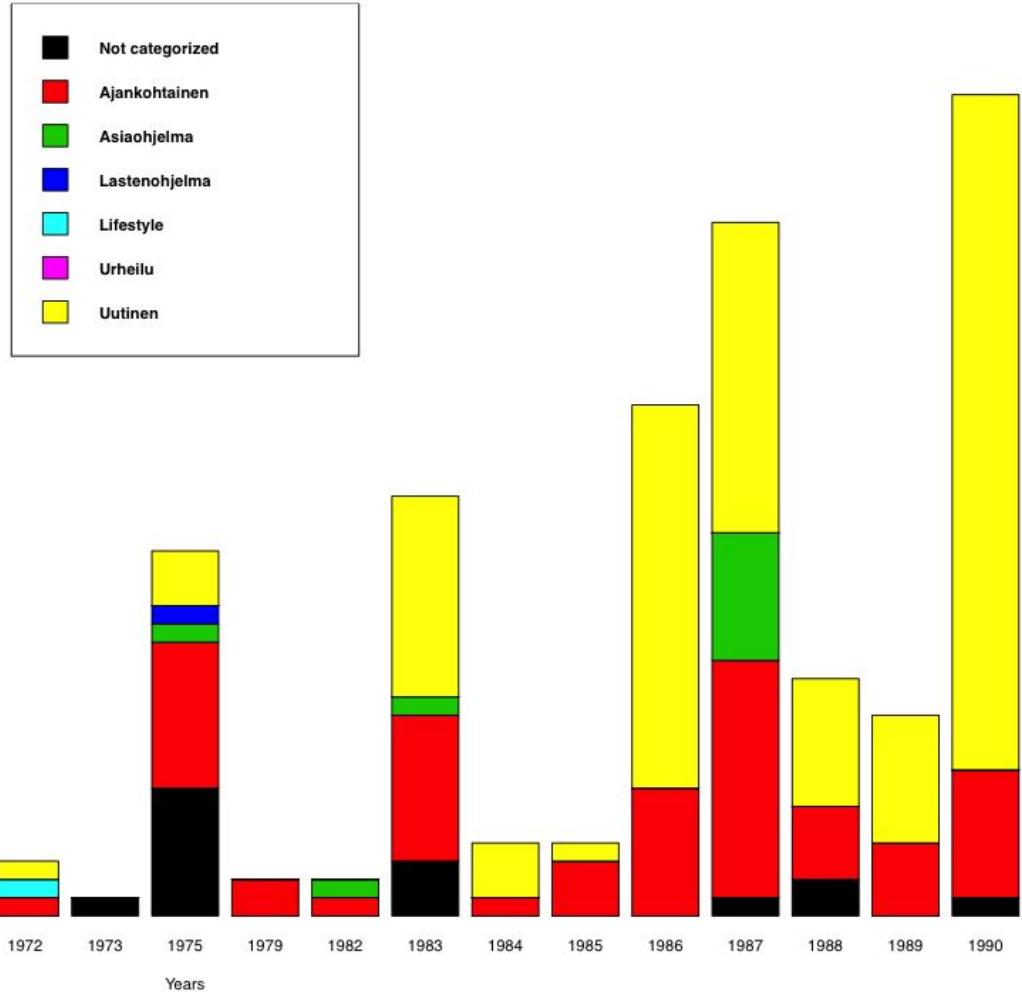
Number	Column Labels															
Row Labels	AJAN	ASIA	ELOK	KFIK	KULT	LAST	LIFE	MUUT	REAL	UFIK	URHE	UTI	VIIH	(blank)	Grand Total	
1957		1	8			1					6	34		1	51	
1958		2	15	2		4	10	3	4		9	164	1	5	219	
1959		4	29	6	25	13	25	1	1		3	199	5	23	334	
1960		19	64	2	30	14	26			1	10	419	2	26	613	
1961		77	95	2	55	25	35	41			26	411	68	75	910	
1962		43	81	7	60	37	137	69	1		40	502	159	424	1560	
1963		23	84	11	96	41	115	114	3		56	572	208	430	1753	
1964		27	122	23	72	96	94	103	2	1	91	655	203	626	2115	
1965		216	128	55	111	195	127	109	2		55	663	184	855	2700	
1966		43	131	32	109	139	180	63	1	2	113	694	47	614	2168	
1967		68	170	27	91	113	137	30	1	1	83	939	12	669	2341	
1968		24	158	25	67	120	240	37		6	92	945	11	718	2443	
1969		42	127	42	103	132	272	34			76	867	28	680	2403	
1970		100	167	23	73	94	206	29		45	580	15	501	1833		
1971		119	172	16	62	100	264	53		68	454	21	475	1804		
1972		130	240	27	85	125	247	78	1	67	394	37	526	1957		
1973		212	229	24	67	95	276	52	1	42	332	24	518	1872		
1974		143	226	9	50	74	341	44		2	42	344	28	467	1770	
1975		163	301	9	77	107	255	46	1	65	386	35	439	1884		
1976		149	347	9	63	145	409	66	6	10	90	387	71	480	2232	
1977		209	369	12	53	151	579	183	1	26	219	1214	77	450	3543	
1978		243	356	24	69	178	577	156		23	279	1173	47	322	3447	
1979		293	381	34	63	143	706	194		21	301	1136	62	451	3785	
1980		282	351	18	62	167	631	163	3	26	370	1136	47	385	3641	
1981		367	398	32	56	197	708	192	1	24	309	1160	52	593	4089	
1982		321	364	27	70	182	622	161	1	43	307	1113	77	661	3949	
1983		338	670	32	77	149	599	128		32	332	1181	67	642	4247	
1984		346	745	24	73	143	687	91		17	355	1176	73	569	4299	
1985		359	704	42	65	145	1005	122		23	415	1163	80	577	4700	
1986		366	608	48	81	179	990	177	5	56	479	1191	58	894	5132	
1987		436	618	65	84	148	1137	181	1	73	526	1194	70	823	5356	
1988		531	570	60	84	126	966	164	1	37	562	1285	80	770	5236	
1989		575	590	53	108	135	1117	208		11	47	521	1531	75	829	5800
1990		727	741	61	138	166	1300	235		22	65	535	1891	121	650	6652
(blank)																
<b>Grand Total</b>		<b>6998</b>	<b>10359</b>	<b>883</b>	<b>2379</b>	<b>3879</b>	<b>15020</b>	<b>3327</b>	<b>37</b>	<b>33</b>	<b>536</b>	<b>6589</b>	<b>27485</b>	<b>2145</b>	<b>17168</b>	<b>96838</b>

# Limitations of the dataset

- Only selected programs were archived until 1984
  - Non-coordination of archival criteria until 1977
  - “The idiosyncratic archivist” (e.g. content descriptions inconsistent, genre classification partial)
  - Loss of original programmes (e.g. 50% of tapes 1977–1984) a problem for retrospective completion
  - Context: to assess the overall significance of topic analysis, we would need comparative datasets (e.g. newspapers, MTV3) which are not accessible
-

# Program that include “eduskuntavaal\*” in metadata, by years

231 entries



Sylvi esittää Helsinki  
Urho nostaa uusia ihmisiä  
auto puheenjohtaja toimittaja  
toivottaa presidentti lippu  
toraissa mua lentokenttä  
kertoo Suomi Lounais-Suomi

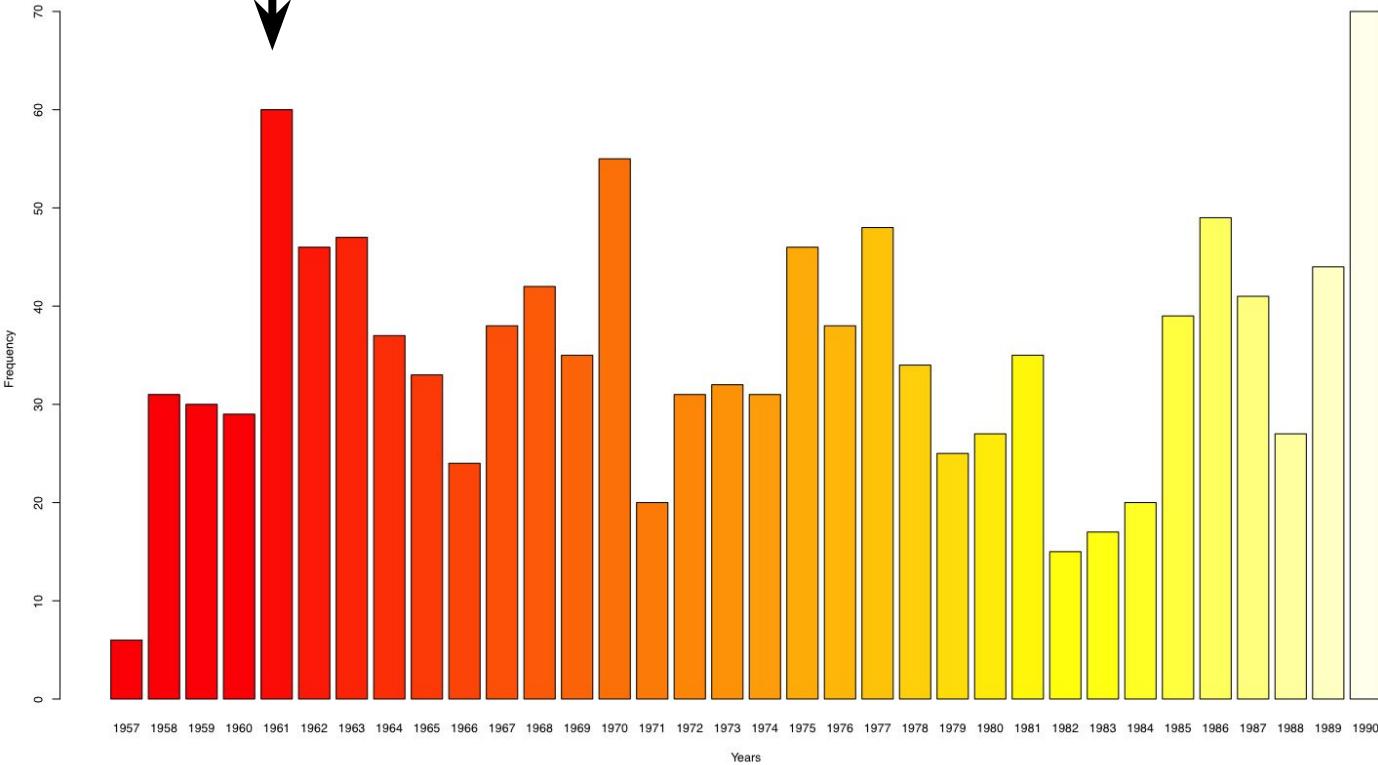
# Case 1: Kekkonen

**Note Crisis**, a political crisis in Soviet–Finnish relations in 1961.

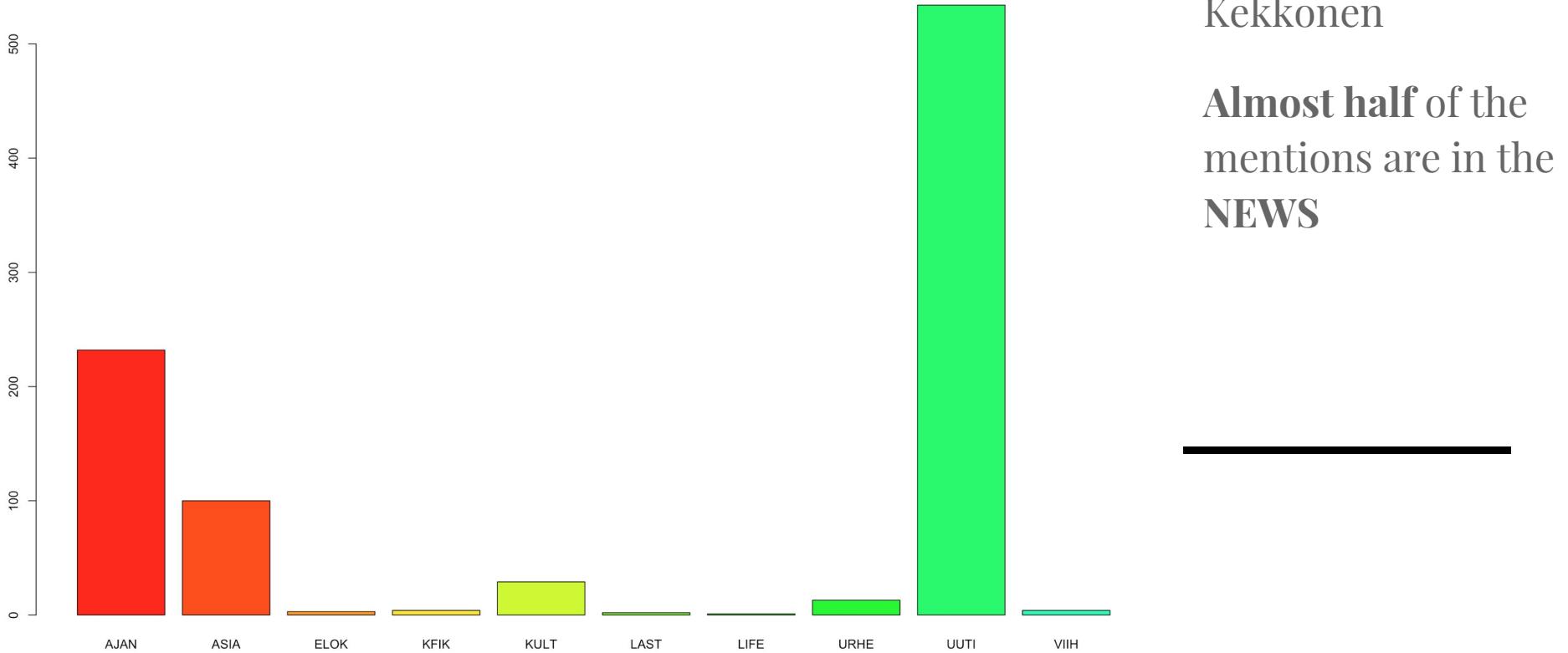
## YEARS OF KEKKONEN

Kekkonen.  
Kekkonen.  
Kekkonen.  
Kekkonen

1202 entries where Urho Kekkonen mentioned in the content description



# PROGRAM TYPES WITH MENTION OF KEKKONEN

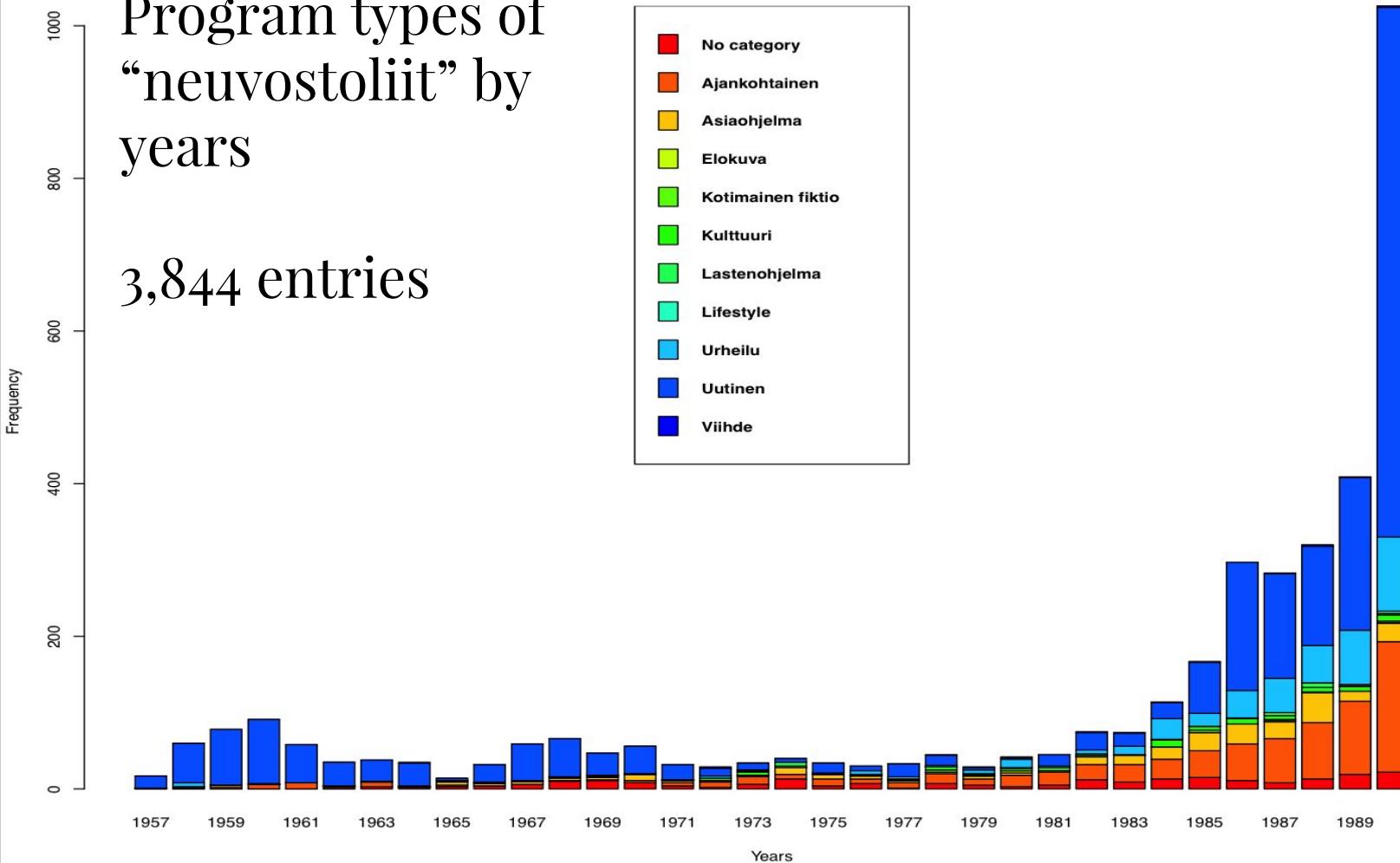


# Case2: Cold War

---

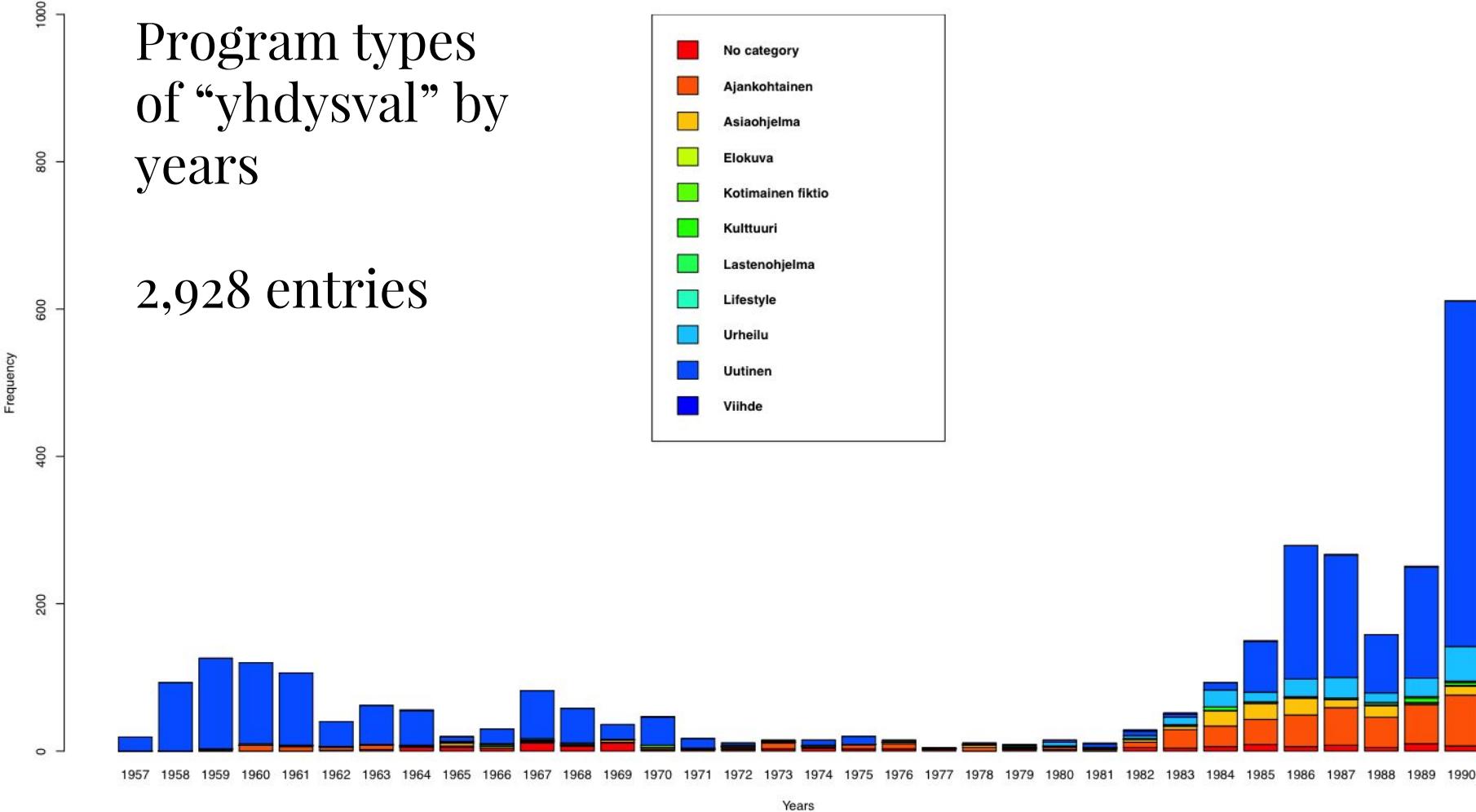
# Program types of “neuvostoliit” by years

3,844 entries



# Program types of “yhdysval” by years

2,928 entries



# Topic Modelling

---

Topic 1: lapsi, kertoa, metsä, talo, tie, puu, kylä

Topic 2: sää, tv-uutiset, sähkeuutiset, suomi, kokous, helsinki, helsing

Topic 3: yhdysvallat, usa, aasi, iso-britannia, afrikka, suomi, saksa

Topic 4: puhua, wtn, toivoa, uosukainen, kokous, lentokone, tiedotustilaisuus

Topic 5: haastateltava, lind, jatkua, ydinvoimala, helsinki, pankki, jakso

Topic 6: tehdas, norja, telakka, työntekijä, laiva, wärtsilä, ruotsi

Topic 7: uutinen, suomi, jatkua, hallitus, uusi, usan, kiina

Topic 8: tv-nytt, lisä, ruotsi, maali, olaf, neuvosto, hypätä

Topic 9: katsaus, kotimaa, tiainen, haastateltava, juonto, jyvänenkylä, eila

Topic 10: voittaa, vuosi, suomi, yhdysvallat, kamerakierto, järjestää, usa

Topic 11: uutiset, myöhäisutiset, varhaisutiset, helsinki, alkaa, helsing, tekstitvn

Topic 12: mies, nainen, ihminen, lähi, katu, lapsi, istua

Topic 13: mykkä, lentokone, saapua, puhua, yleisö, lentokenttä, auto

Topic 14: puheenjohtaja, kansanedustaja, puhua, eduskunta, sorsa, kalla, paavo

Topic 15: sotilas, iraki, ykn, israeli, katu, armeija, libanon

Topic 16: presidentti, koivisto, gorbatshev, neuvostoliitto, puhua, mihail, mauno

Topic 17: kertoa, mäkelä, juhani, studio, esittää, ohjelma, lehdistö

Topic 18: kirkko, paavi, kuvata, ihminen, johannes, pappi, italia

Topic 19: kuvaerittely, aihe, tekstitys, puhua, neuvostoliitto, viro, liettua

Topic 20: haastattelu, toimittaja, kela, sairaala, helsinki, potilas, helsing

Topic 21: juonto, kakkonen, tallella, elokuva, risto, erkki, ote

Topic 22: ebu, wtn, ihminen, katu, sotilas, poliisi, puhua

Topic 23: luetteloida, neuvostoliitto, suomi, maa, hallitus, presidentti, yhdysvallat

Topic 24: studio, suomi, jatkua, välijuonto, tunnus, juonto, aihe

Topic 25: ajankohtainen, studiojuonto, juonto, laukkanen, suomi, markku, saari

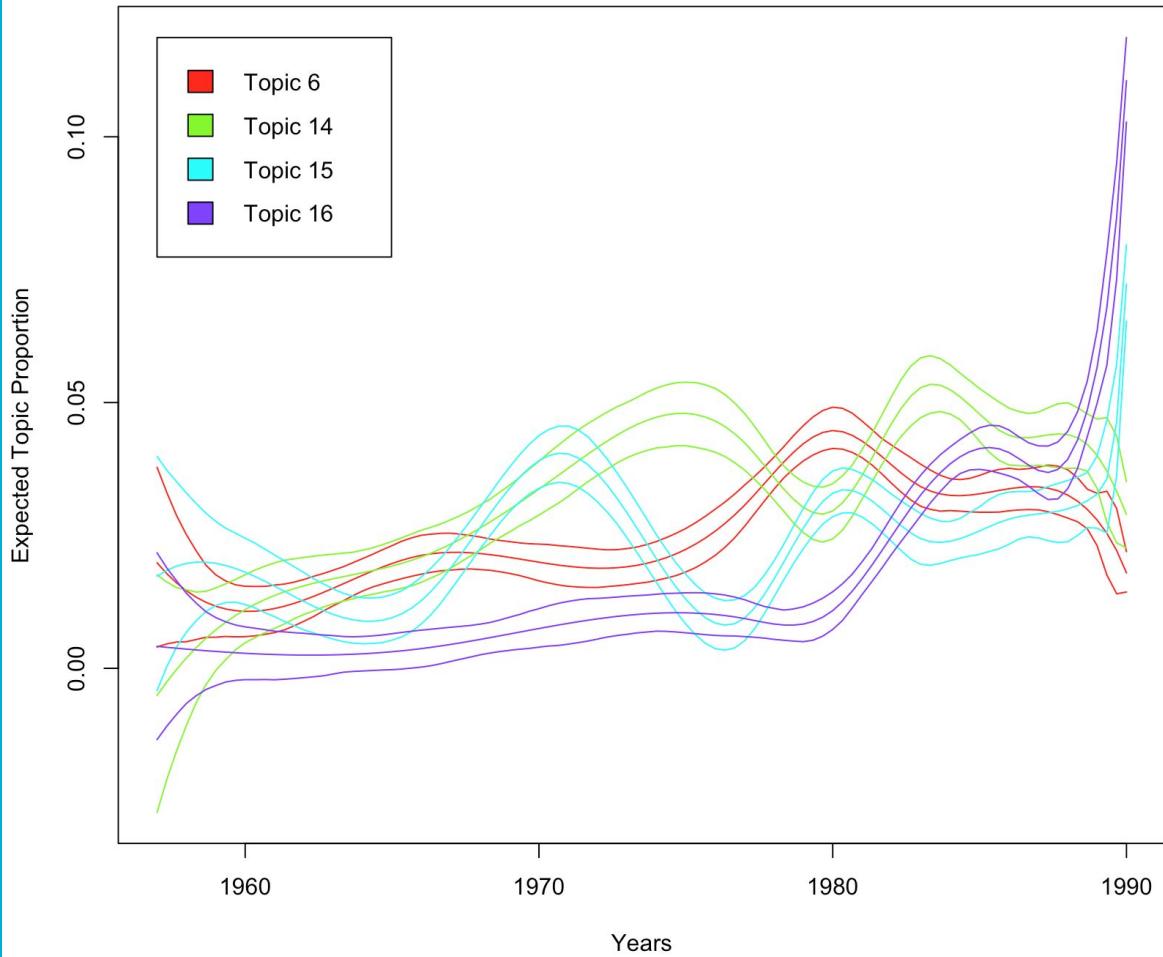
Topic 26: kekkonen, presidentti, urho, tieto, kuvalähde, neuvostoliitto, puhua

Topic 27: puhua, katu, dag, espanja, iso-britannia, reagan, ihminen

Topic 28: toimittaja, haastateltava, välijuonto, ohjelma, jatkua, esittää, tunnus

Topic 29: ääni, teh, väri, kop, musta, mua, puhua

Topic 30: junta, pelto, rautatieasema, lehmä, asema, traktori, vilja



**Topic 3: Foreign politics:**  
yhdysvallat, usa, aasi, isobritannia,  
afrikka, suomi, saksa

**Topic 6: Shipbuilding industry**  
tehdas, norja, telakka, työntekijä,  
laiva, wärtsilä, ruotsi

**Topic 14: Parliament**  
puheenjohtaja, kansanedustaja,  
puhua, eduskunta, sorsa, kalla, paavo

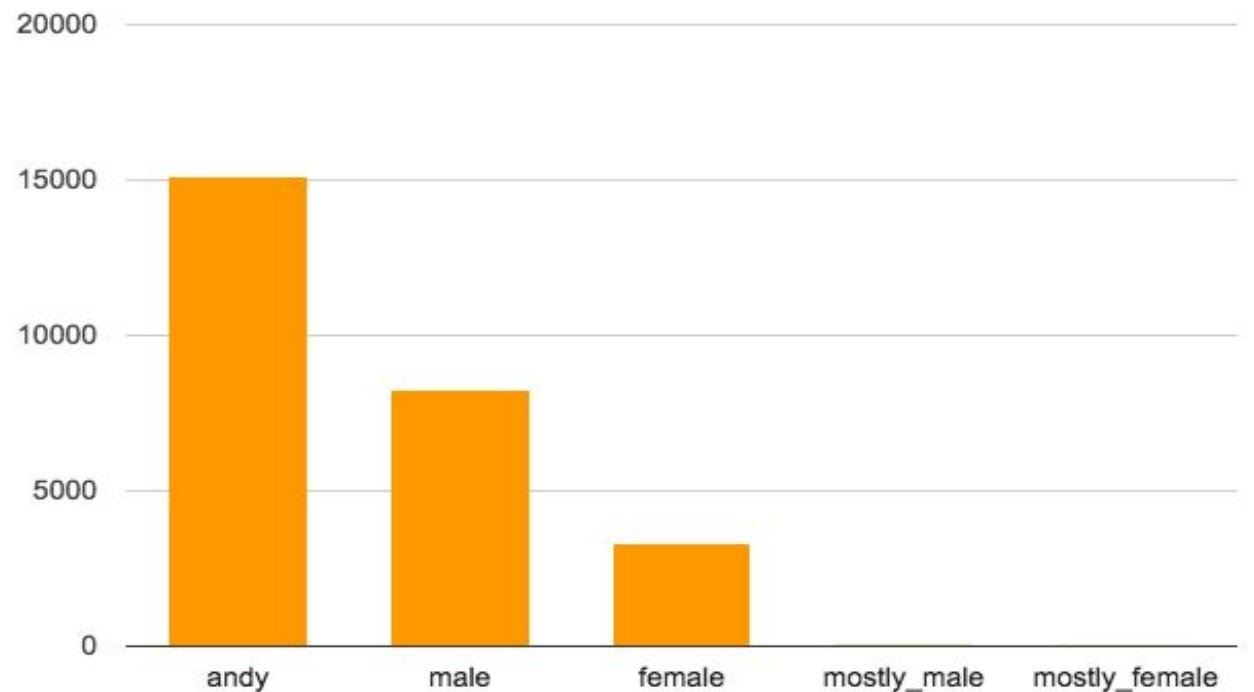
**Topic 15: Middle-Eastern wars**  
sotilas, irak, ykn, israeli, katu,  
armeija, libanon

**Topic 16: Soviet relations in the 80s**  
presidentti, koivisto, gorbatshov,  
neuvostoliitto, puhua, mihail, mauno

# Named Entity / Gender

---

An Sandman Kiviniemi  
Engström Heikkilä  
Franck Salminen  
Kela Björn Taikuri  
Abendgruesse Puska  
Seppälä Tiaisen  
Uotila Korhonen  
Poltilan Lehdistö  
Koiviston Härkönen  
Oksanen Paasilinna  
Panula Gartz  
Tawaststjerna Sirkuspelle  
Jokinen Ahonen  
Sariola Ahvonens  
Forss Liikanen  
Föhr Nalle Hietanen  
Saarela Lehtinen  
Luppakorva Aalto  
Oravisto Loiri  
Maija  
Pohjola Heiskanen  
Kasvi Lindström  
Tiainen Nousiainen  
Johannes Niemelä  
Hätkönen Tikkalan  
Silvasti Nyman  
Pastori Nainen  
Lindblom Piskonen  
Halonen Kurkelan  
Leppänen Linnainmaa  
Sinisalo Nurmi  
Rutanen Jansson  
Haavisto Pyykkö<sup>1</sup>  
Numminen Kukkonen  
Väyrynen  
Holmlund Mustonen Mäkelän Järvelehto Ziliacus Nysten  
Paasonen Bergström  
Lohntander Tuominen Suhonen  
Prost Andersson Rinne  
Vuorinen Karjalainen  
Pellonpää Luukas  
Laine Kallio  
Hann Poika  
Kekkonen



# Lessons Learned

- **Learning a new language:** Working in a team with people from different backgrounds met for the first time
- Working with a big dataset
- Many tools and methods used for the first time
- Dealing with the problems together
- Collaborating and asking for help

46 HOURS.

6 HUMANS.

6 LAPTOPS.

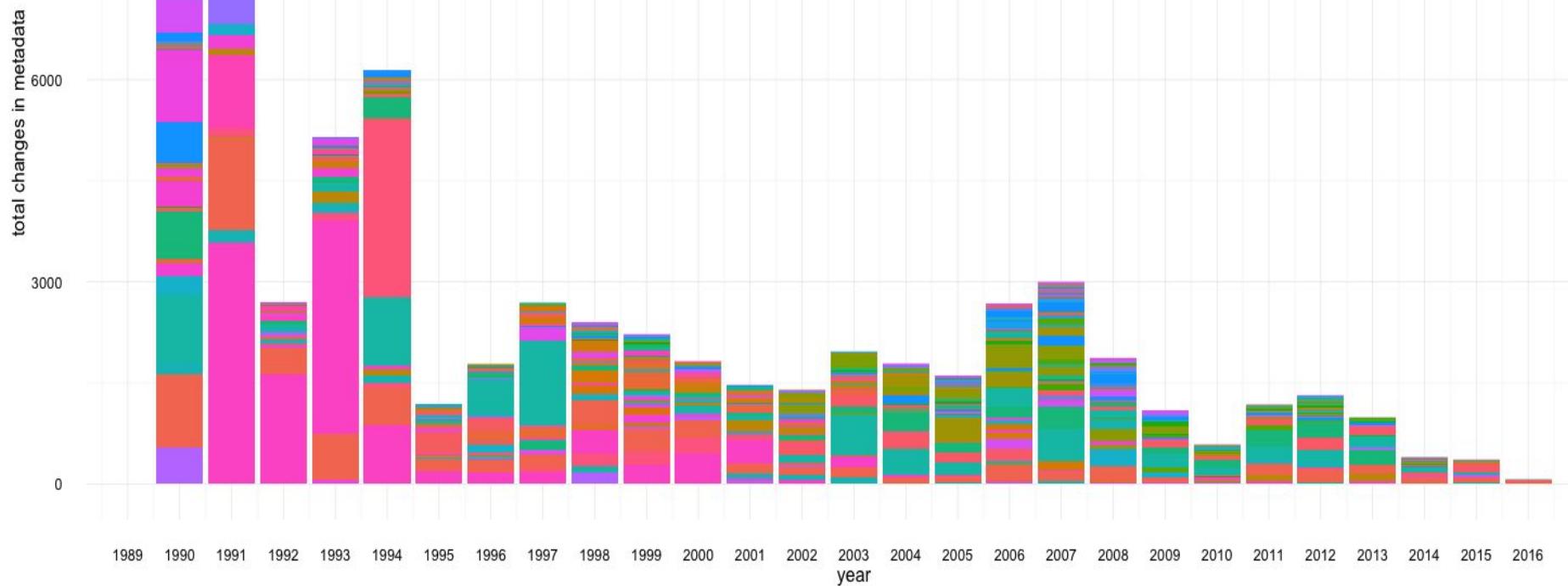
4 MAC CRASHES.

29 CUPS OF COFFEE.

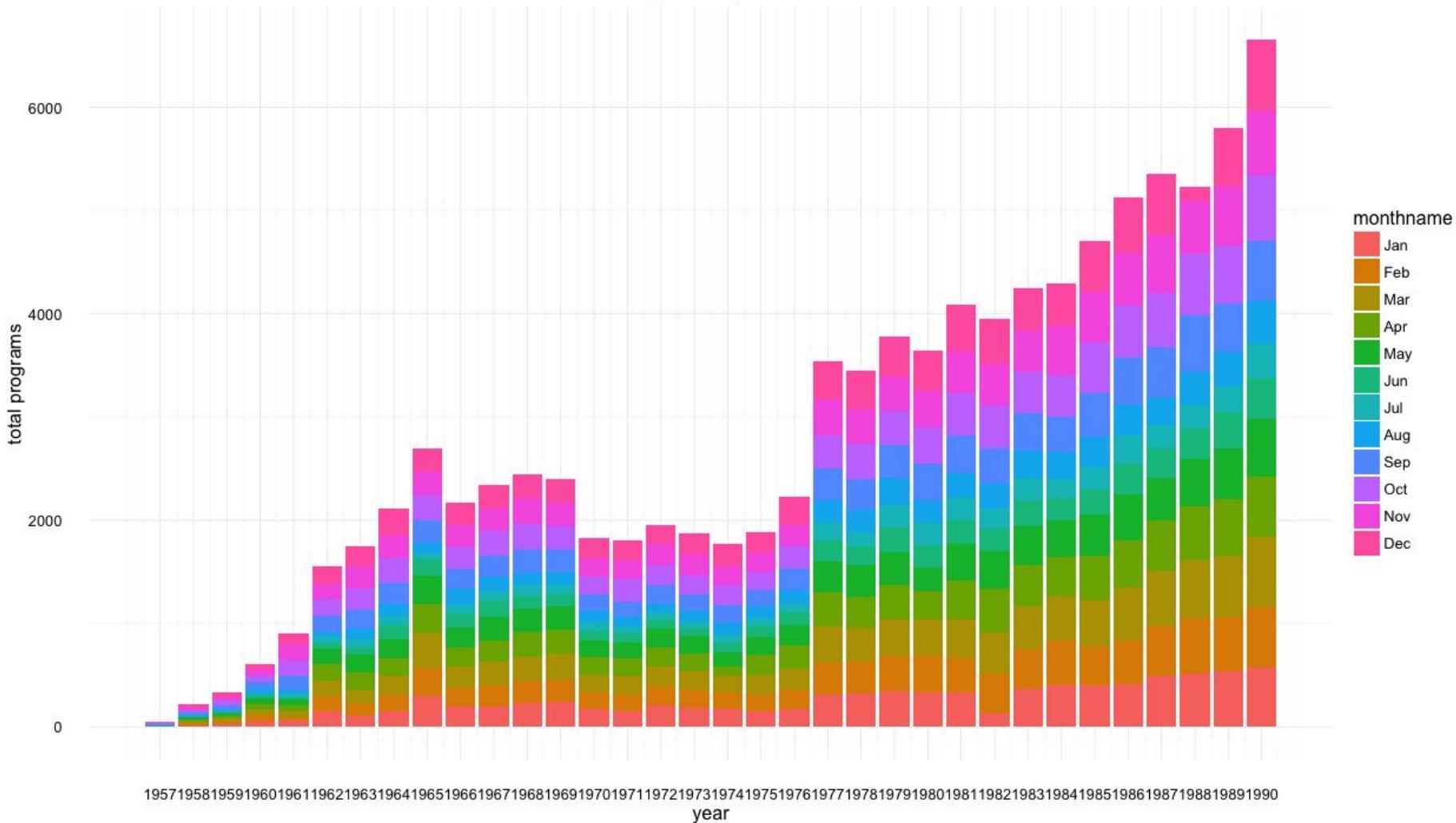
587 ValueError: UTF-8.

Thanks!

## Activity by year and archivist



## Programs by month



### String length of content and description metadata

