Uncovering the Scientific Impact of Women at the Smithsonian Using Machine Learning

Dr. Rebecca B. Dikow
Dr. Mirian Tsuchiya

OCIO Data Science Lab
Smithsonian Institution

Twitter: @SIDataScience
Web: https://datascience.si.edu

Photo of Vicki Funk by Mauricio Diazgranados

AWHI Science Women Spreadsheet

File Edit View Insert Format Data Tools Add-ons Help Last edit was or

100% $ % .0 .00 123 Calibri 14 B I

fx SI UNIT

	A	B	C	D	
1	SI UNIT	WIKIPEDIA PAGE	WIKIPEDIA PAGE NEEDS IMPROVEMENT?	NAME	WIKIPEDIA ST
22	NMNH	y		Archambault, JoAllyn	Done
23	NMNH	y	y	Piperno, Dolores	Done
24	NMNH	y	y	Bruwelheide, Kari	Done
25	NMNH	y	y	Zeder, Melinda	**Melina Zeder** is a
26	NMNH	n		Báez, Gabriela Pérez	**Gabriela Pérez Bá**
27	NMNH	n		Sholts, Sabrina	**Sabrina Sholts** is
28	NMNH	n		Arnoldi, Mary Jo	**Mary Jo Arnoldi** i
29	NMNH	y	y	Greene, Candace	Done
30	NMNH	n		Isaac, Gwyneira	**Gwyneira Isaac** is
31	NMNH	y		Kaeppler, Adrienne	Done
32	NMNH	n		Scherer, Joanna	**Joanna Scherer** is
33	NMNH	n		Stirling, Marion	**Marion Stirling** w
34	NMNH	y	y	Walsh, Jane	Done
35	NMNH	n		Watanabe, Barbara	**Barbara Watanab**
36	NMNH	y		Stevenson, Matilda	Done
37					

Sheet1

The Funk List

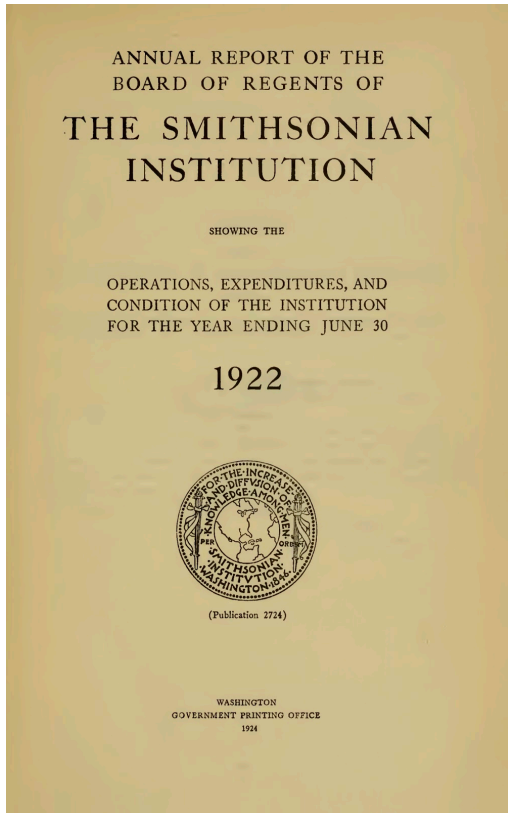
The Funk List includes more than 400 current and past Smithsonian women in science

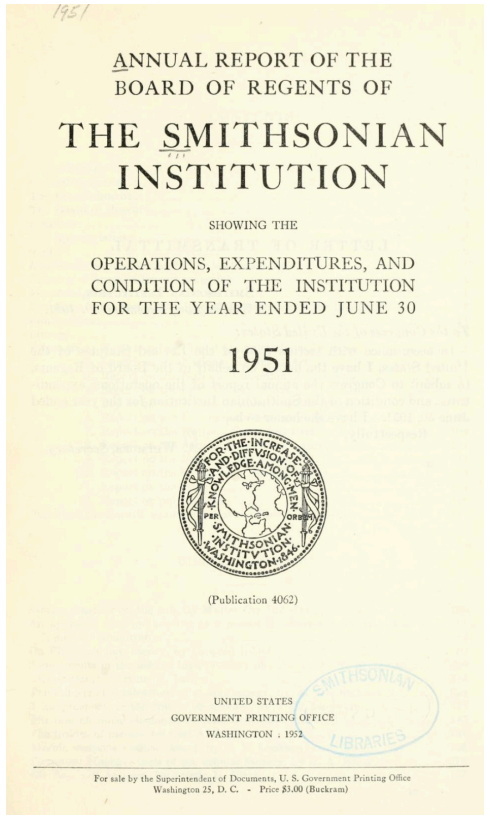
From left to right: Vicki Funk, Sophie Lutterlough, and Jessie Cohen

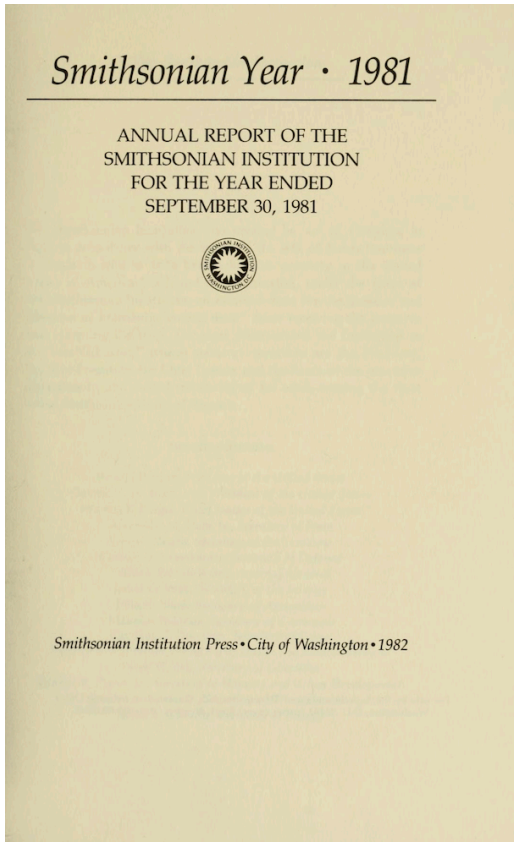
How do we measure scientific impact?

- Publications
- Service
- Collections
- Public outreach

Machine learning tools can help us connect women on the Funk List to Smithsonian archives and collections data to help us better understand their scientific impact.


ANNUAL REPORT OF THE
BOARD OF REGENTS OF
THE SMITHSONIAN
INSTITUTION
SHOWING THE
OPERATIONS, EXPENDITURES, AND
CONDITION OF THE INSTITUTION
FOR THE YEAR ENDING JUNE 30
1922
(Publication 2724)
WASHINGTON
GOVERNMENT PRINTING OFFICE
1924

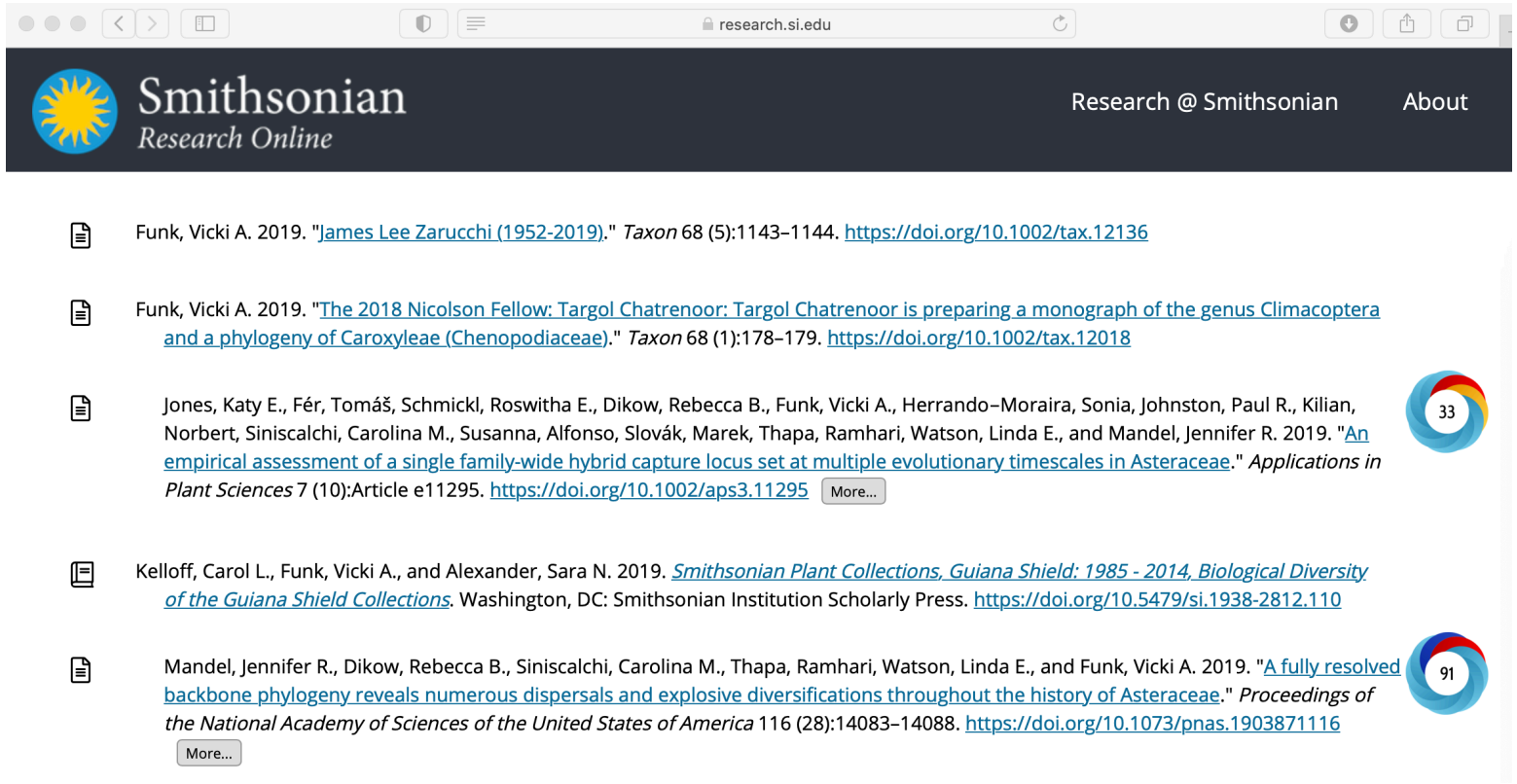

1951
ANNUAL REPORT OF THE
BOARD OF REGENTS OF
THE SMITHSONIAN
INSTITUTION
SHOWING THE
OPERATIONS, EXPENDITURES, AND
CONDITION OF THE INSTITUTION
FOR THE YEAR ENDED JUNE 30
1951
(Publication 4062)
UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1952
For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price $3.00 (Buckram)


Smithsonian Year · 1981
ANNUAL REPORT OF THE
SMITHSONIAN INSTITUTION
FOR THE YEAR ENDED
SEPTEMBER 30, 1981
Smithsonian Institution Press • City of Washington • 1982

Biodiversity Heritage Library

Smithsonian Annual Reports

BECAUSE OF HER STORY
Smithsonian



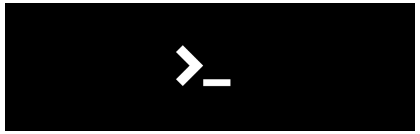
Sample search for publications by Vicki Funk

SRO – Smithsonian Research Online

Smithsonian Libraries

Methods

- Used a combination of Natural Language Processing in spaCy and shell scripting to:
  - Extract and count mentions of women on the Funk List in Annual Reports
  - Count publications for women on the Funk List from Smithsonian Research Online
  - Extract and count occurrences first names and words related to science in the Annual Reports



William J. Bennett PERSON , Secretary of Education ORG

John S. Herrington PERSON , Secretary of Energy Board of Regents ORG

Warren E. Burger PERSON , Chief Justice of the United States GPE ,

ex officio , Chancellor

George H. W. Bush PERSON , Vice President of the United States GPE , ex

officio

Edwin J. PERSON (Jake) Garn PERSON , Senator from Utah GPE

Barry Goldwater PERSON , Senator from Arizona GPE

James R. Sasser PERSON , Senator from Tennessee GPE

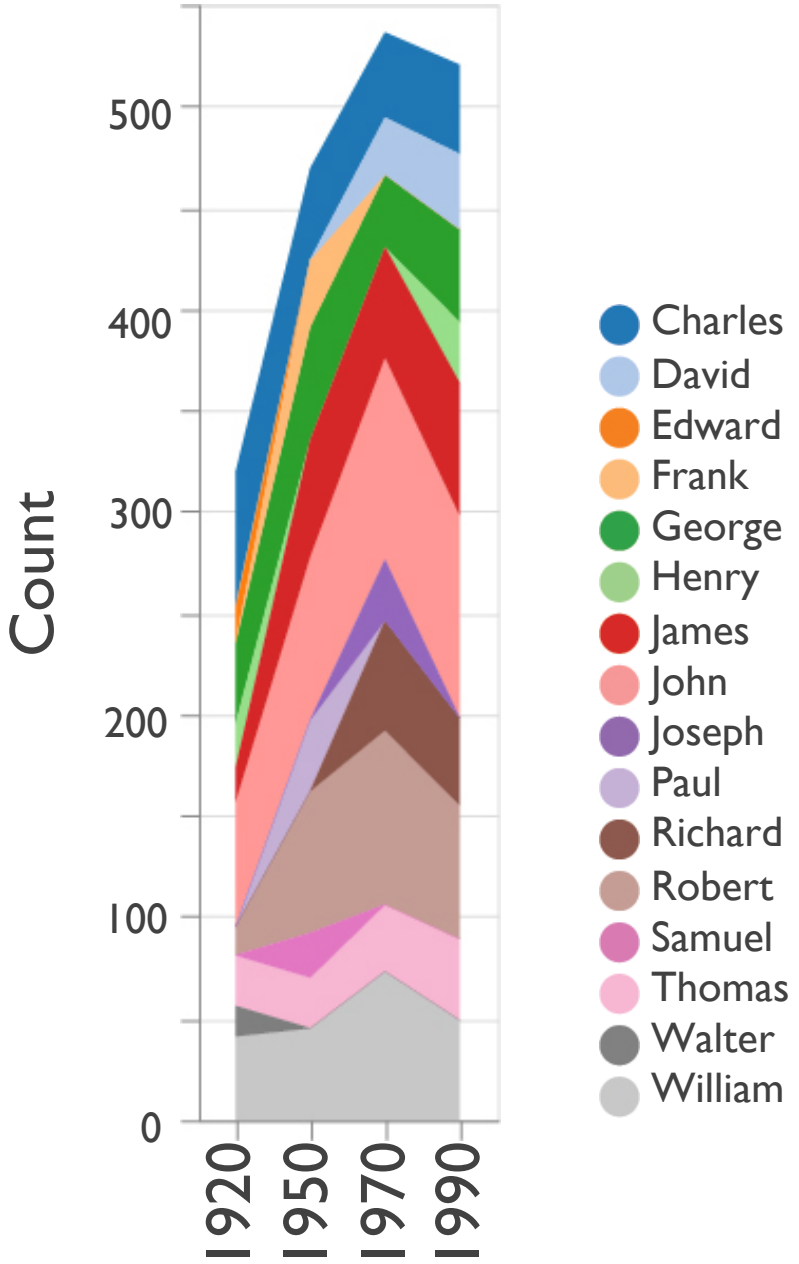
A portion of the 1985 Annual Report - this section lists the members of the Board of Regents

NER – Named Entity Recognition

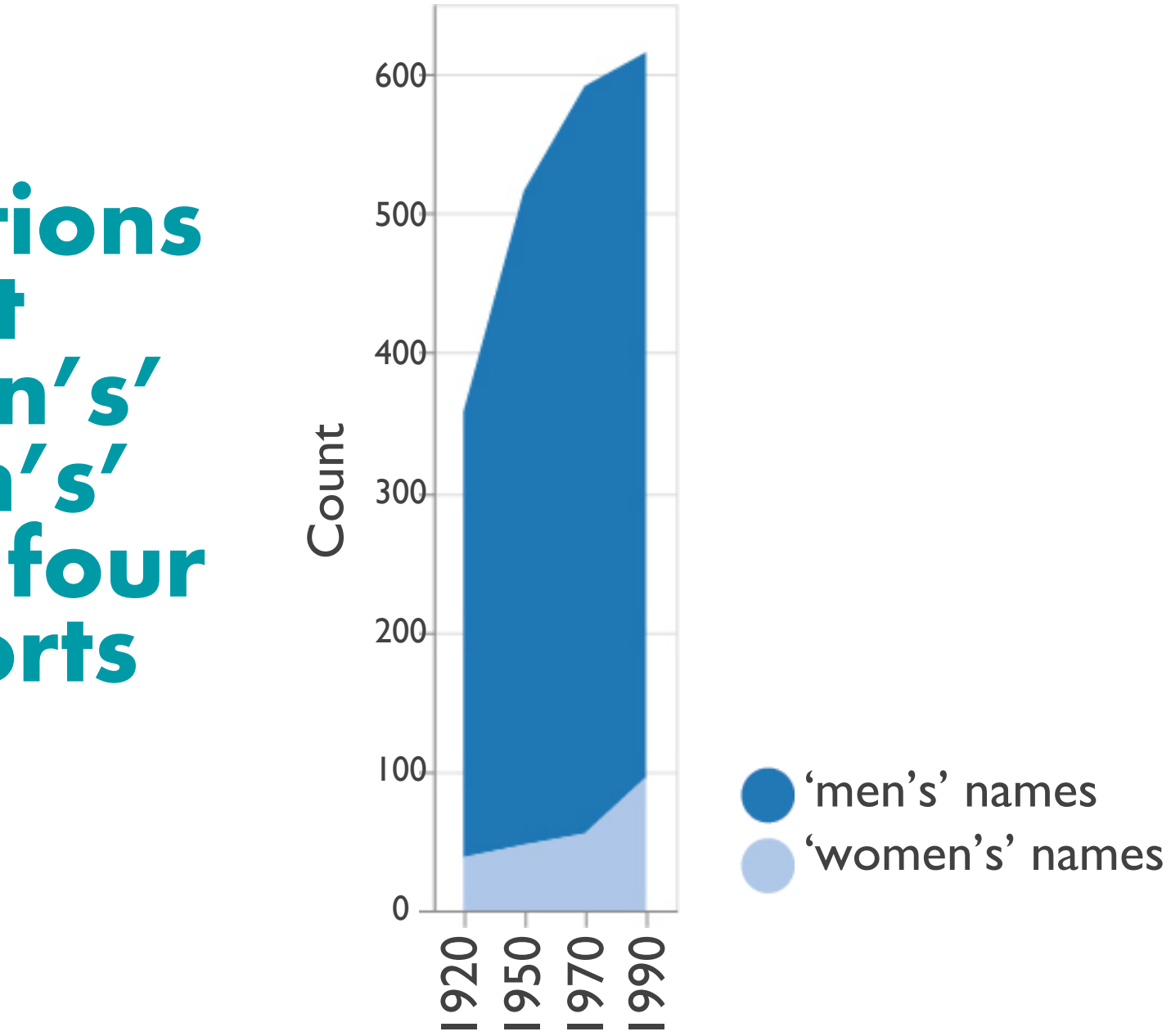
Methods

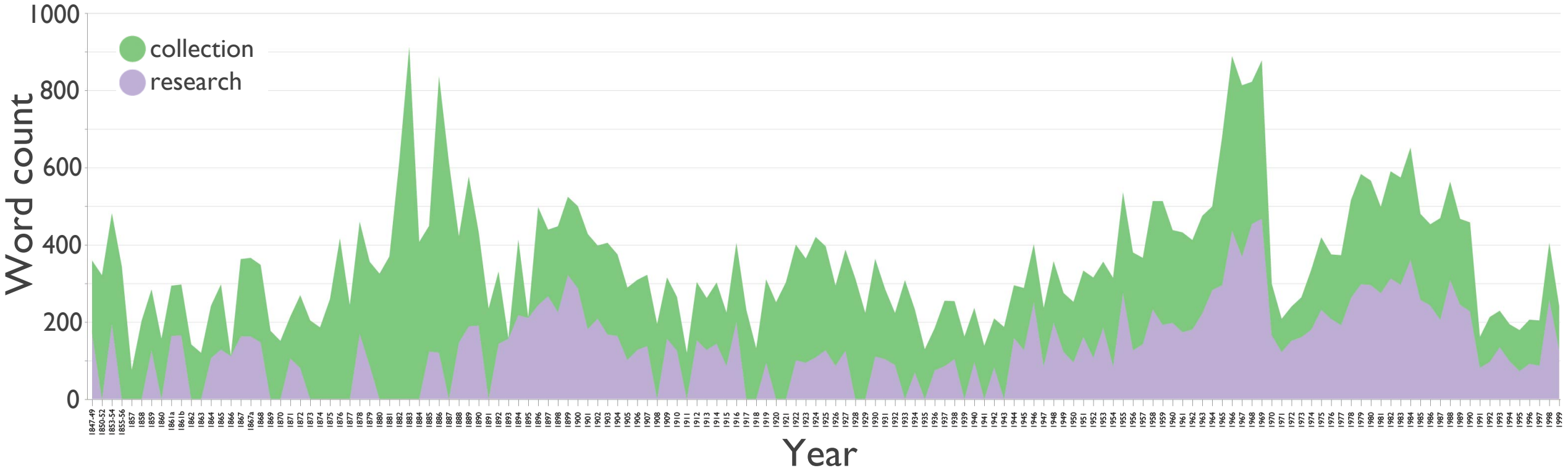
- Included all women from the Funk List no longer at the Smithsonian – 127 total
- Analyzed Annual Reports from 1846-1999
- Downloaded all citations from SRO

Count of mentions of 10 most common first names in four Annual Reports



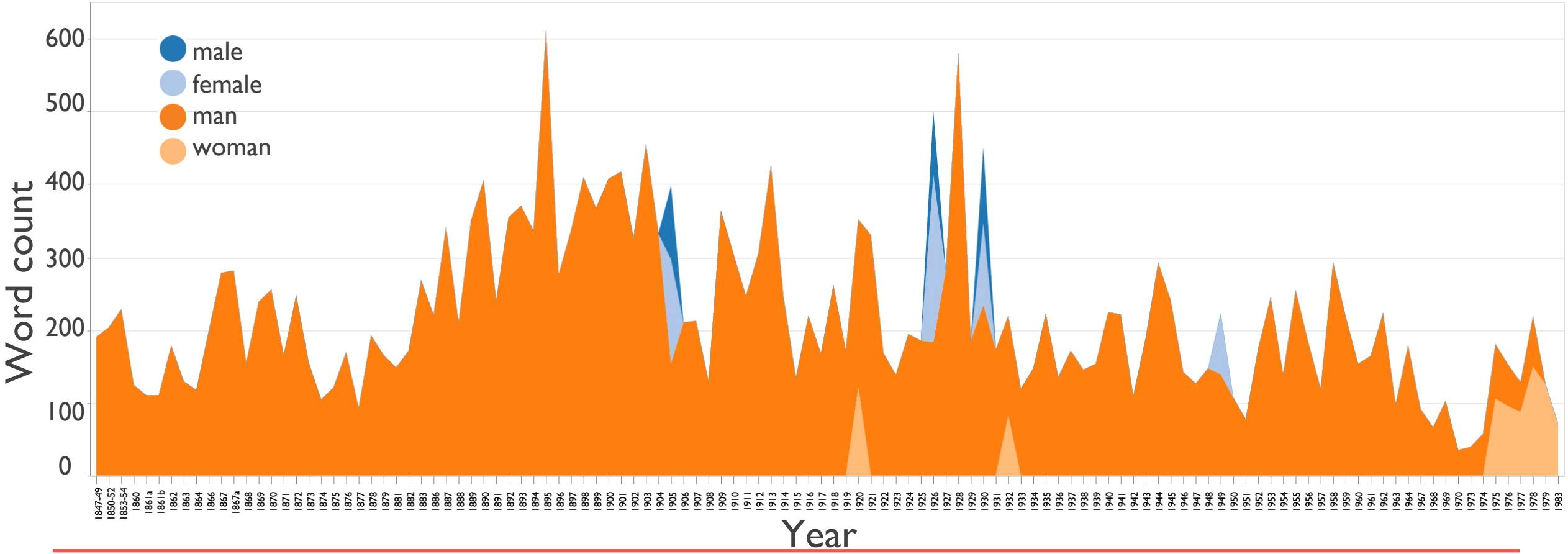
Count of mentions of 10 most common ‘men’s’ and ‘women’s’ first names in four Annual Reports





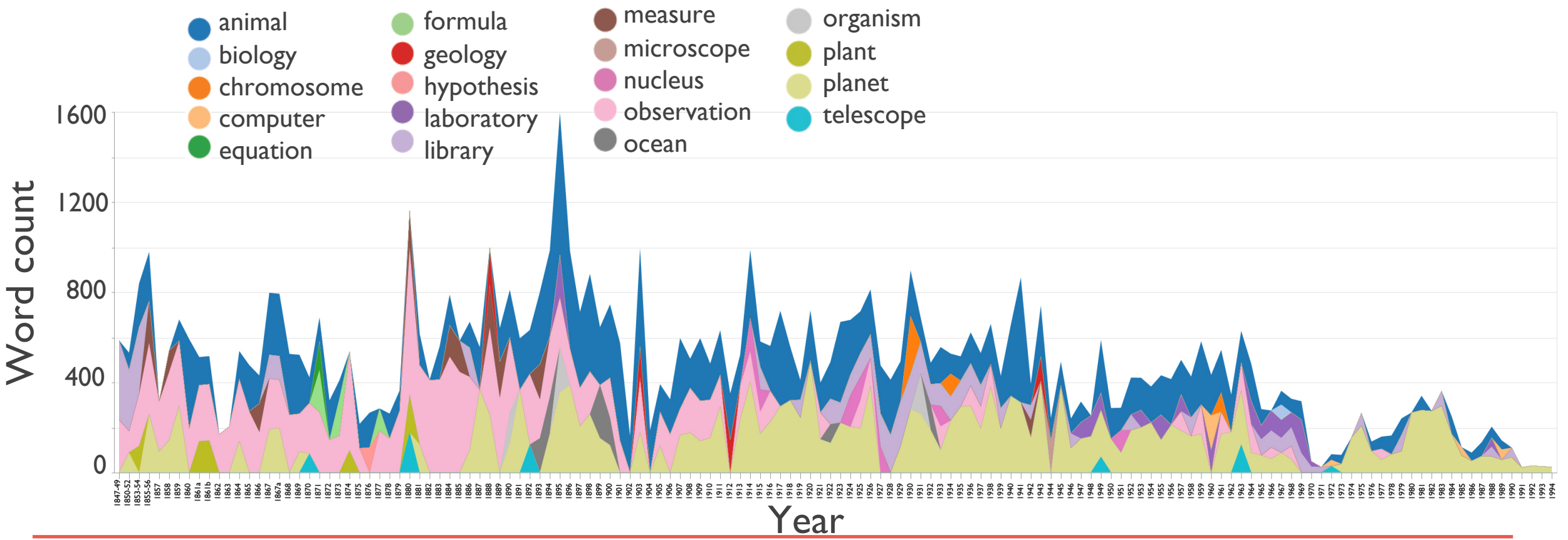
Annual Report word counts through time

Mentions of the words:
research and collection



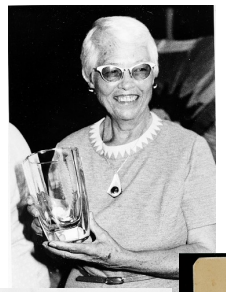
Annual Report word counts through time

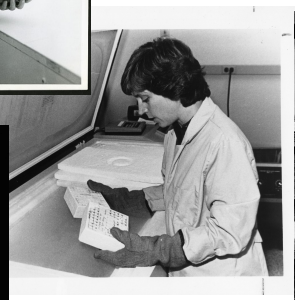
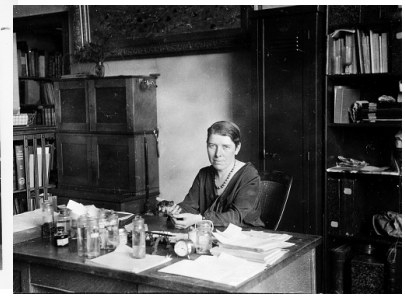
Mentions of the words: man, male, women, and female

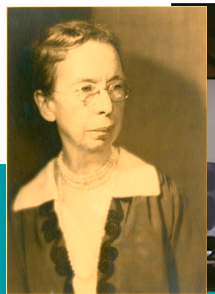


Annual Report word counts through time

Mentions of some common science words

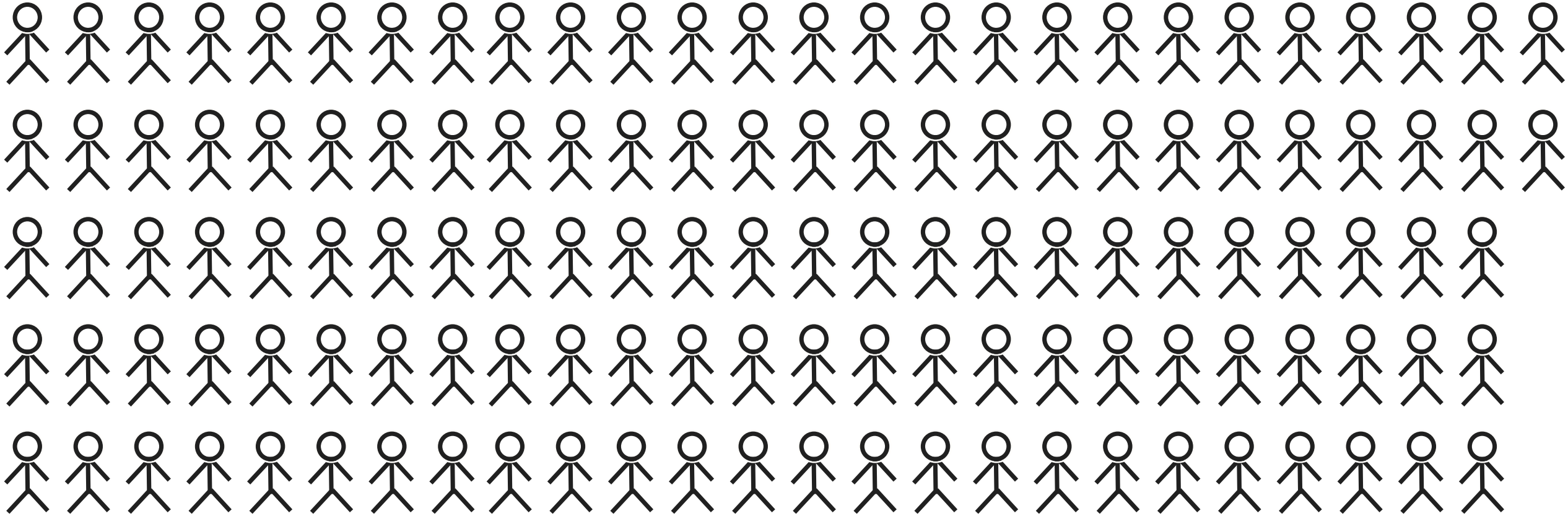


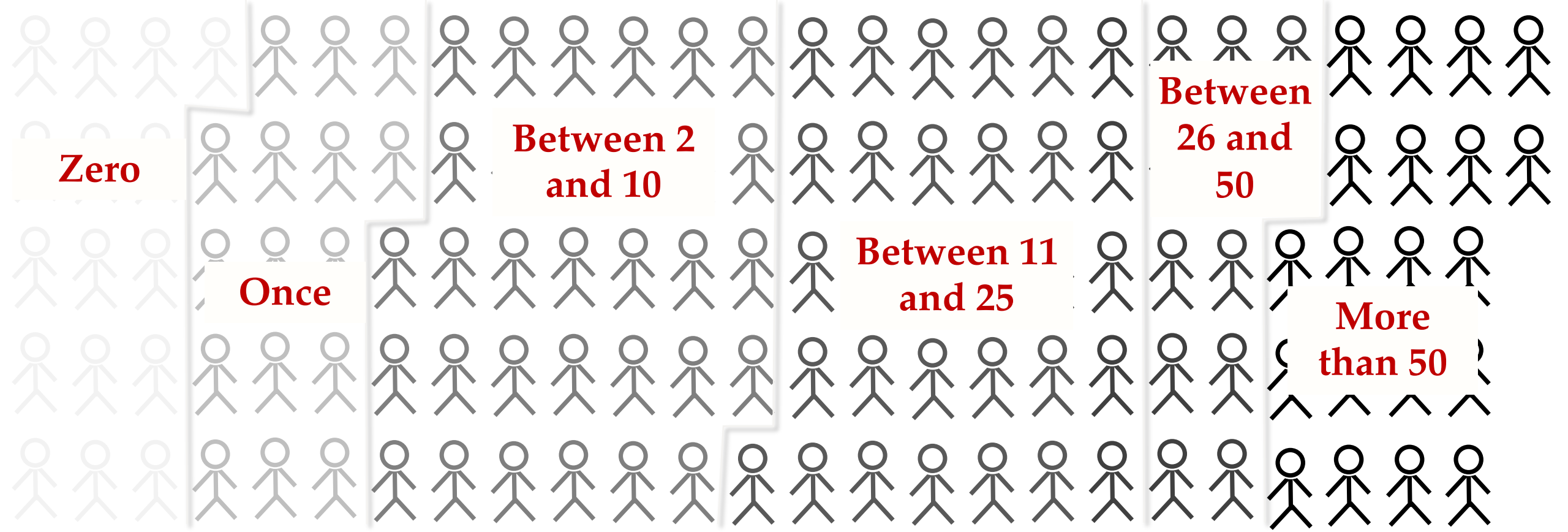


P.5
#4

BECAUSE OF HER STORY
Smithsonian

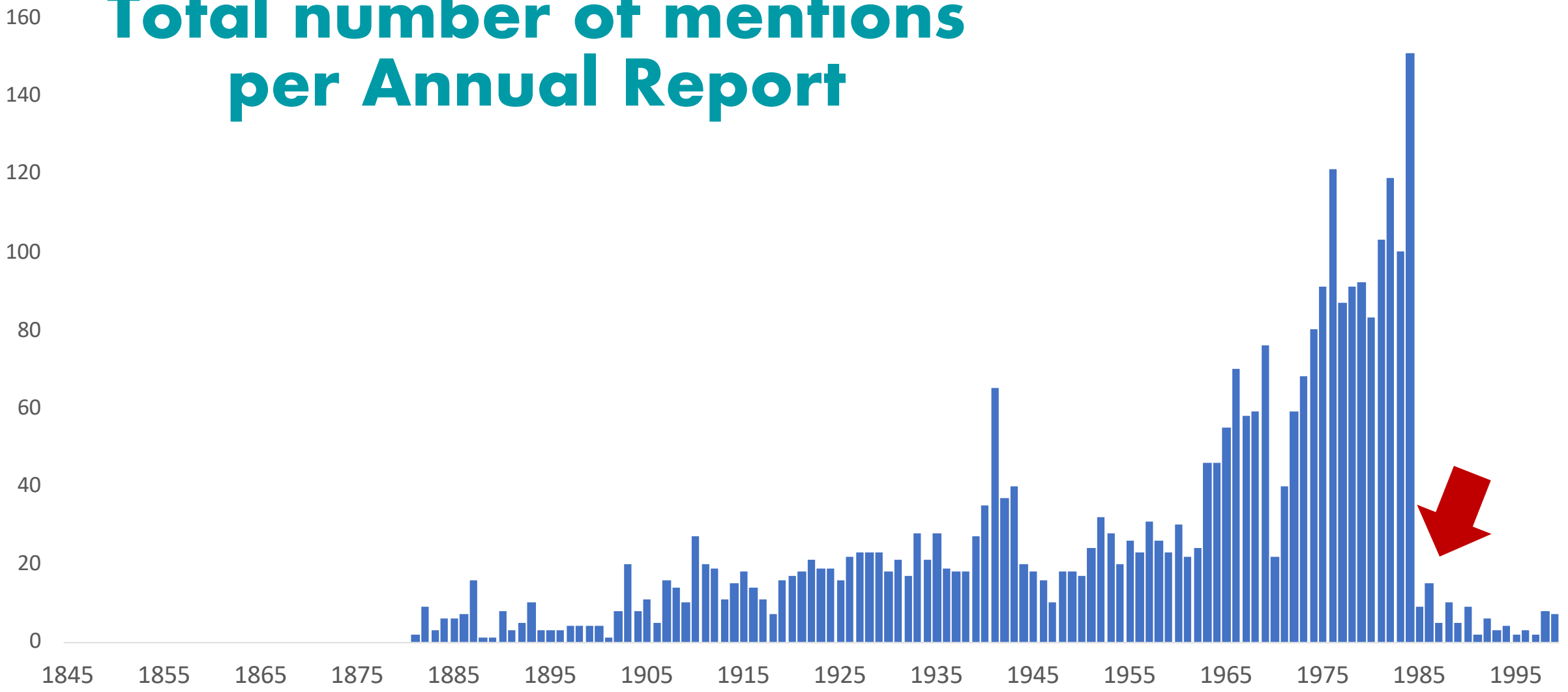


How many times was each woman mentioned?

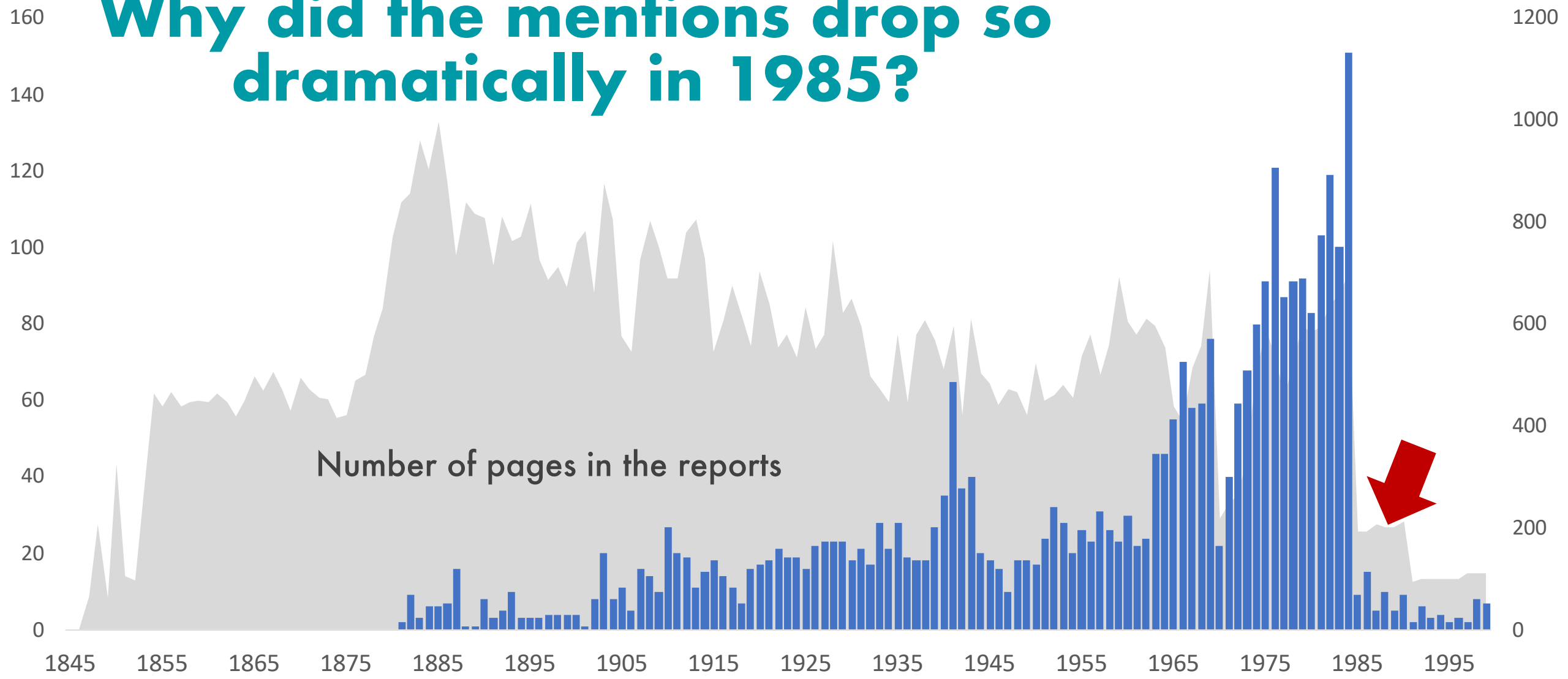


How many times was each woman mentioned?

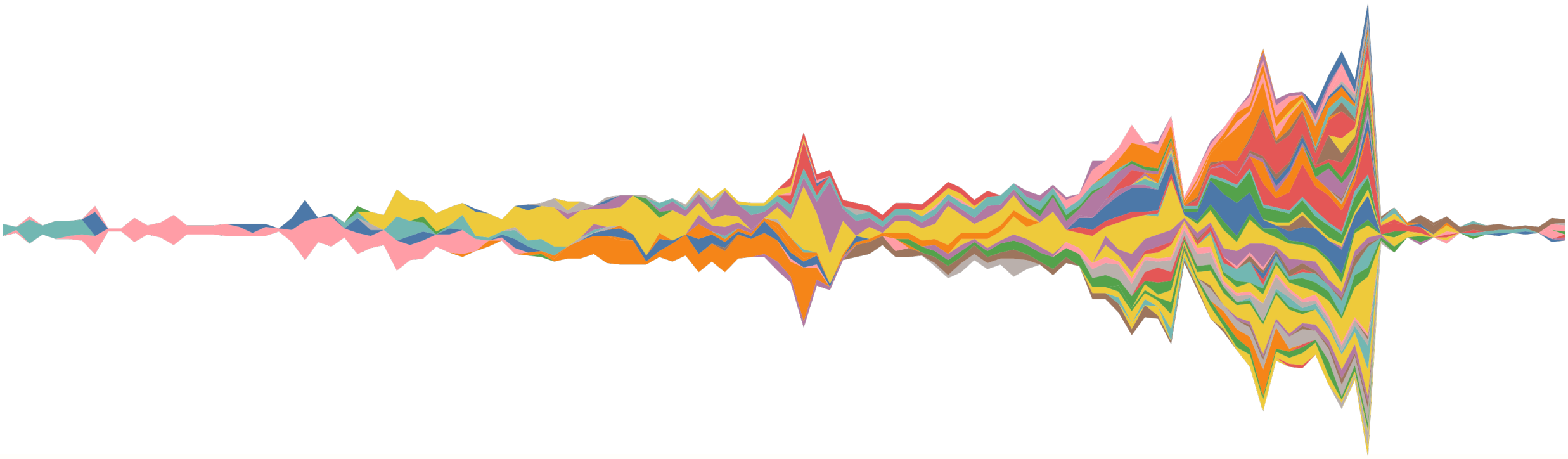
Total number of mentions per Annual Report



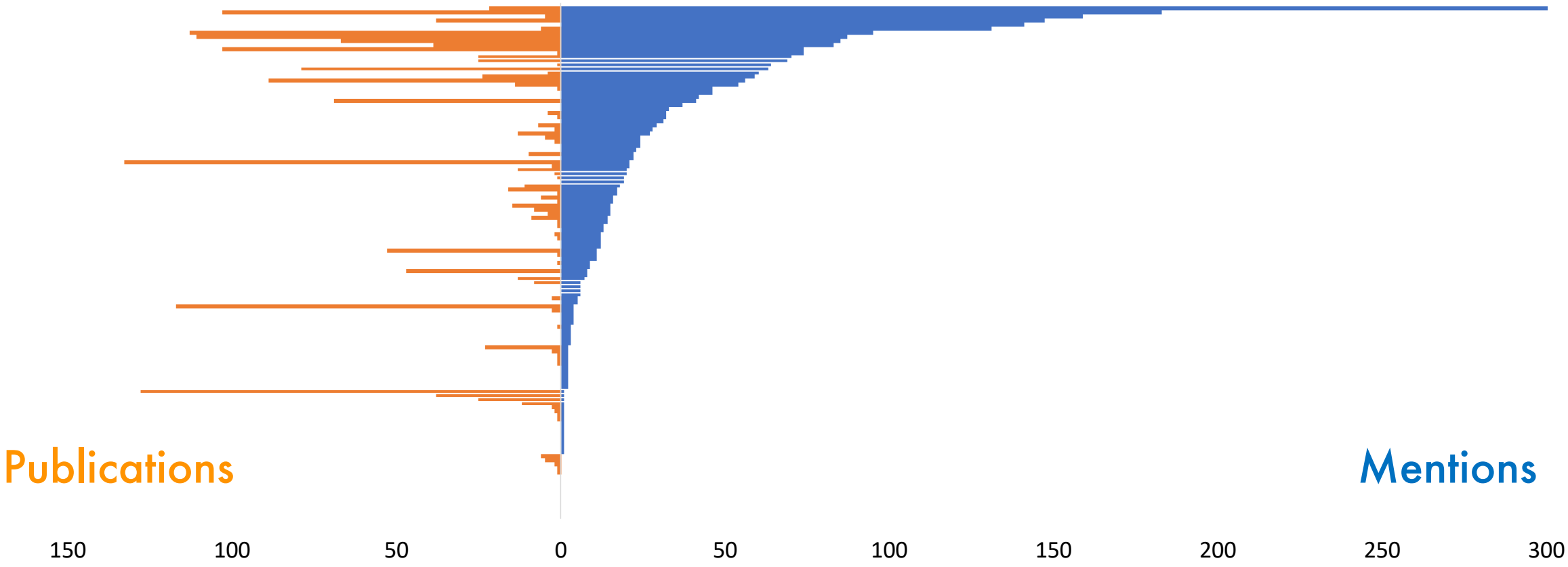
Why did the mentions drop so dramatically in 1985?



How many women are represented in each year?



Number of mentions in the report do not correspond to scientific publications



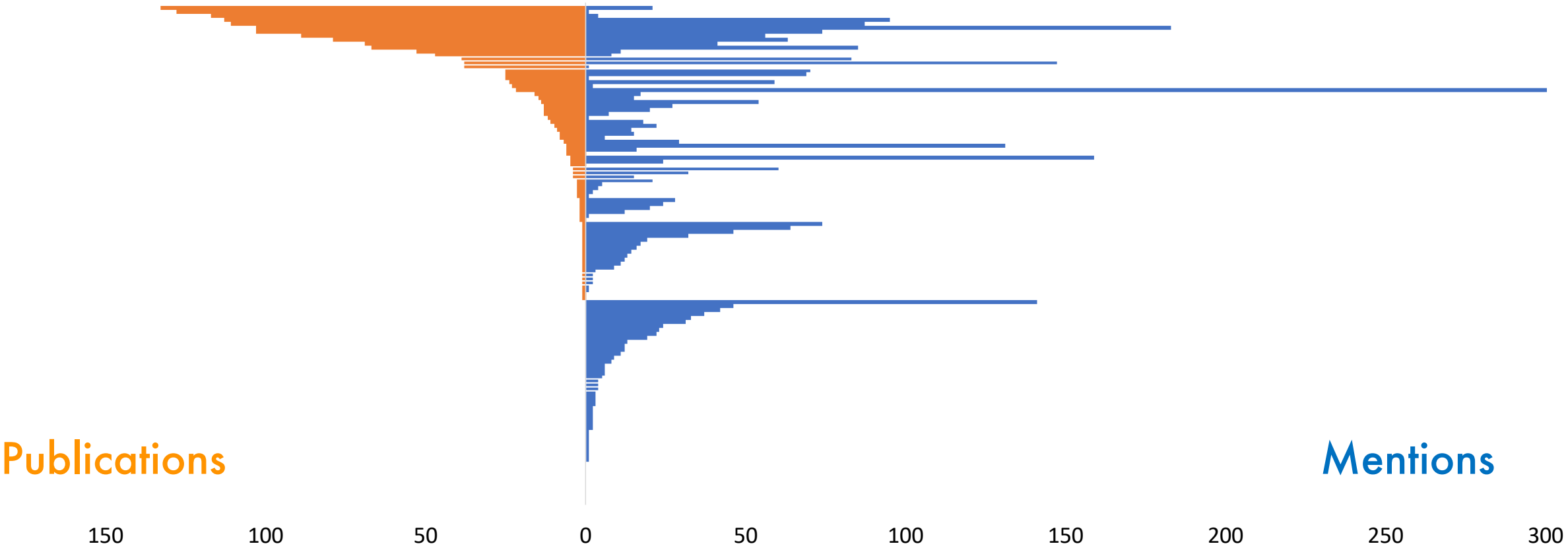
Number of mentions in the report do not correspond to scientific publications

Betty Jane Meggers
NMNH
Tenure: 1954-2012
Mentions: 183
Publications: 103

Frances Theresa Densmore
NMNH
Tenure: 1907-1957
Mentions: 330
Publications: 22

Matilda C. (Evans) Stevenson
NMNH
Tenure: 1889-1915
Mentions: 159
Publications: 5

Number of mentions in the report do not correspond to scientific publications



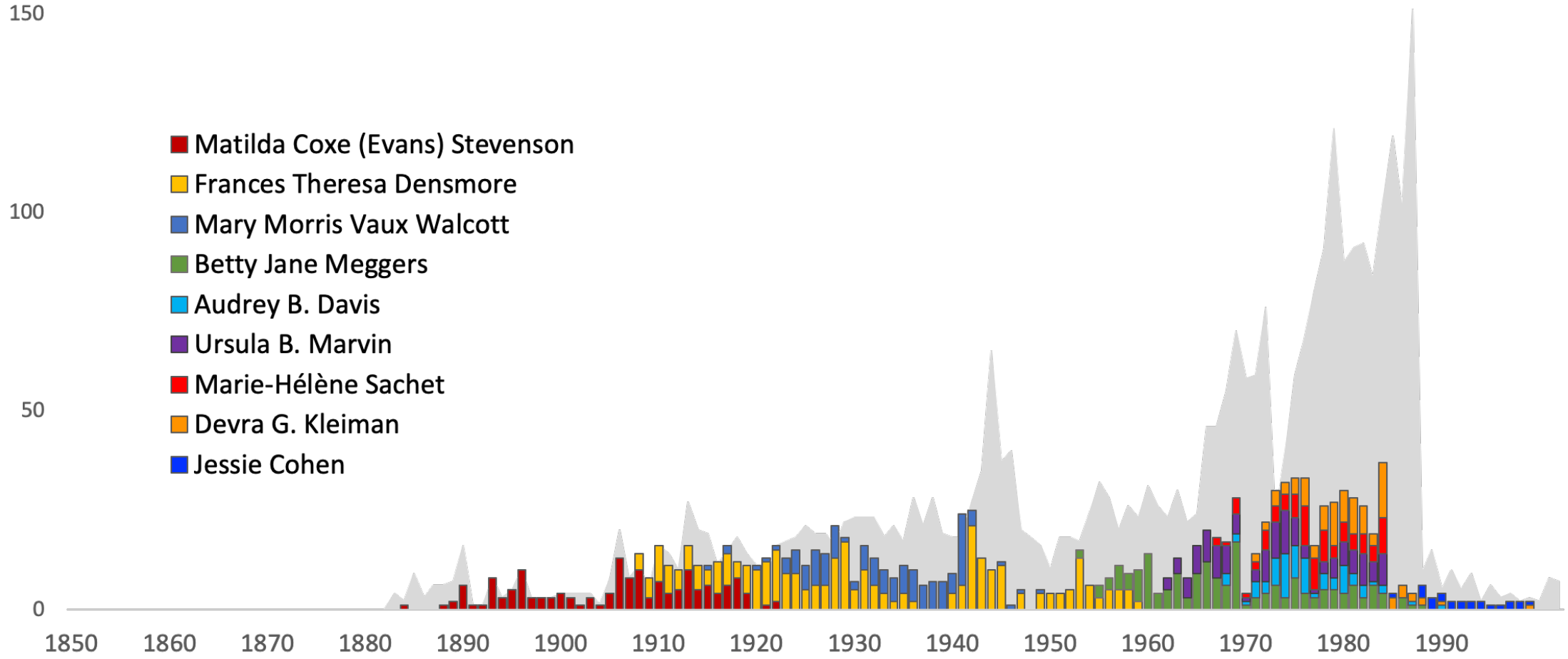
Number of mentions in the report do not correspond to scientific publications

JoGayle Howard
NZP/SCBI
Tenure: 1994-2011
Mentions: 4
Publications (until 1999): 117

Vicki A. Funk
NMNH
Tenure: 1981-2019
Mentions: 21
Publications (until 1999): 133

Pamela B. Vandiver
MCI
Tenure: 1985-2003
Mentions: 1
Publications: 128

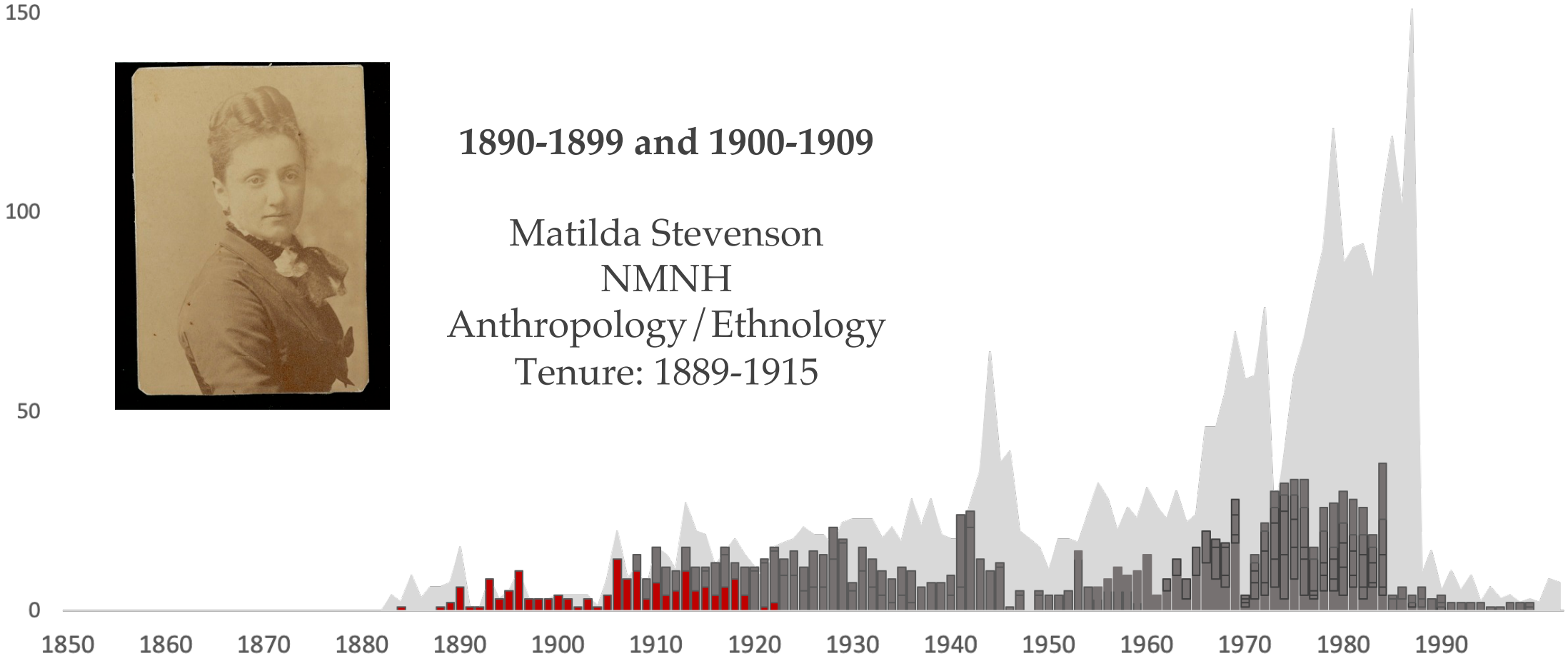
Who are the most-mentioned women in each decade?



Who are the most-mentioned women in each decade?

1890-1899 and 1900-1909

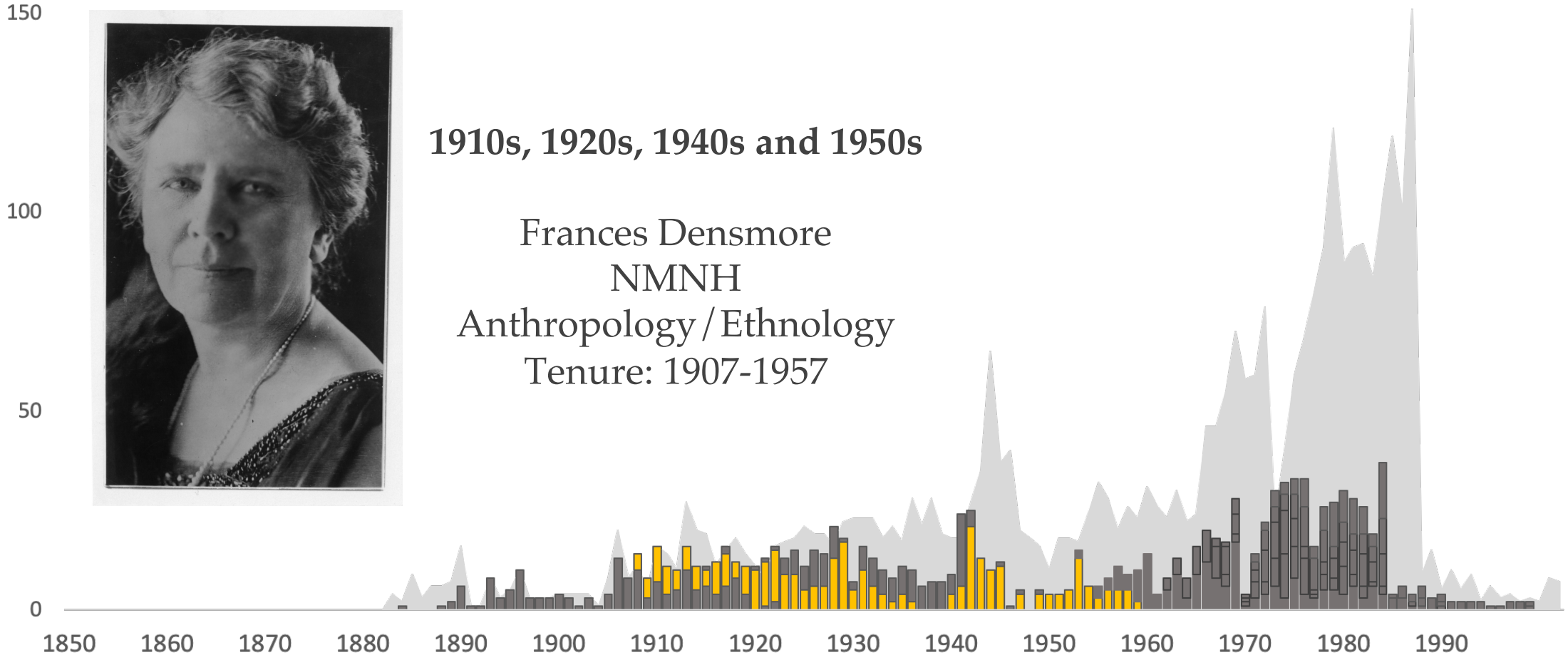
Matilda Stevenson
NMNH
Anthropology / Ethnology
Tenure: 1889-1915



Who are the most-mentioned women in each decade?

1910s, 1920s, 1940s and 1950s

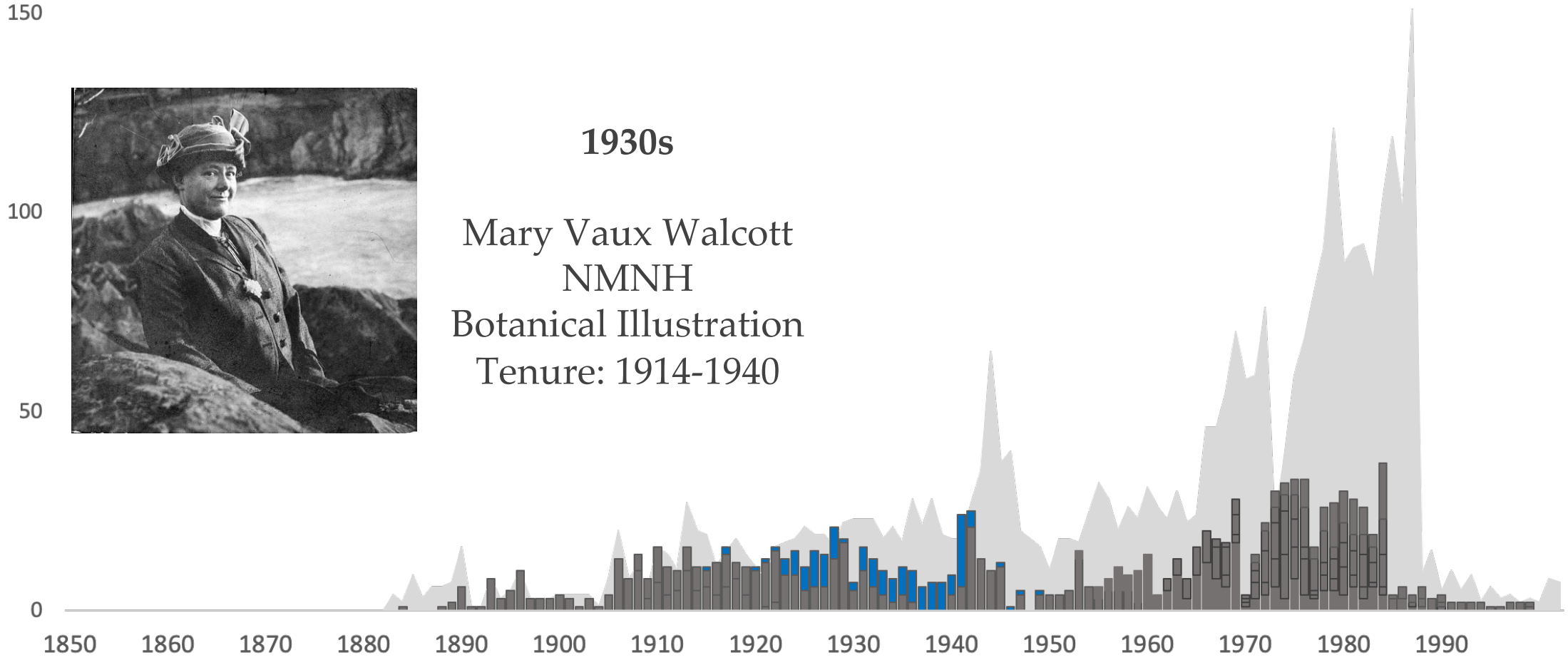
Frances Densmore
NMNH
Anthropology/Ethnology
Tenure: 1907-1957



Who are the most-mentioned women in each decade?

1930s

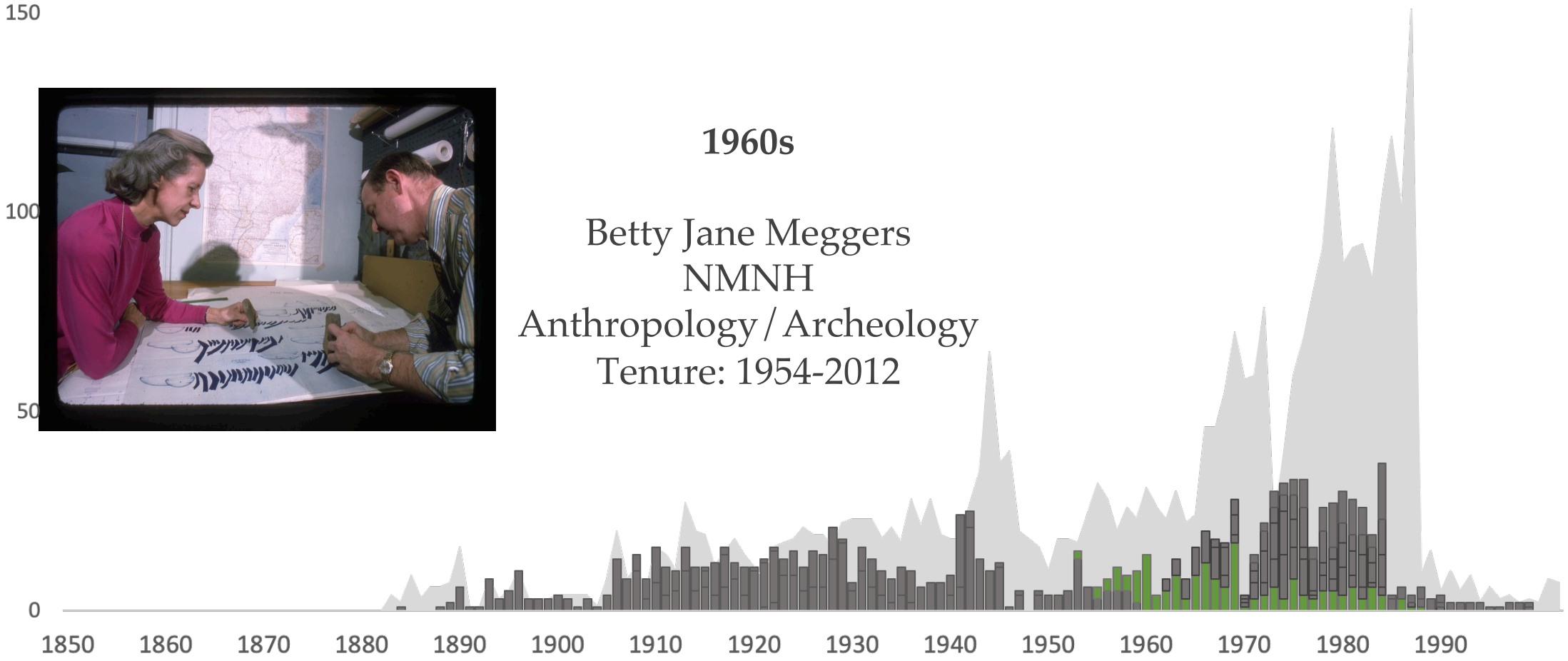
Mary Vaux Walcott
NMNH
Botanical Illustration
Tenure: 1914-1940



Who are the most-mentioned women in each decade?

1960s

Betty Jane Meggers
NMNH
Anthropology / Archeology
Tenure: 1954-2012

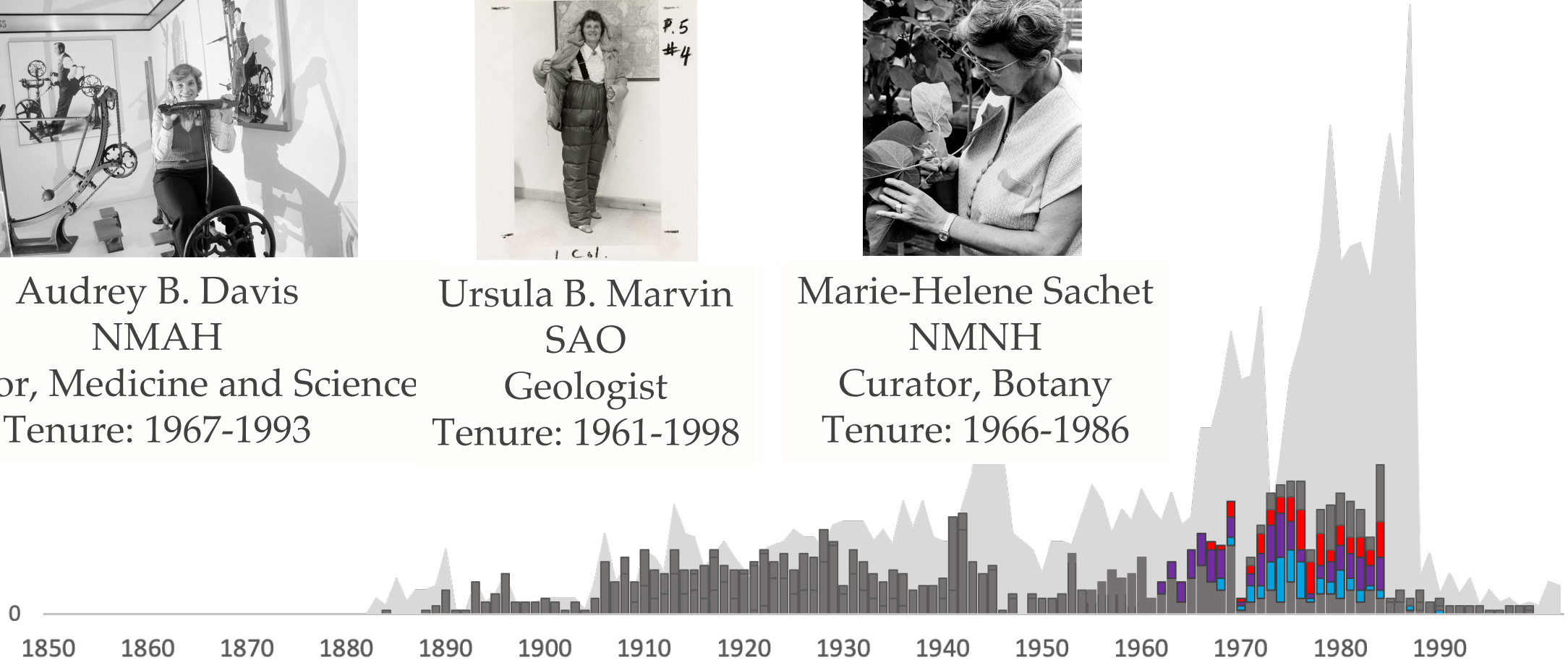


Who are the most-mentioned women in each decade?

Audrey B. Davis
NMAH
Curator, Medicine and Science
Tenure: 1967-1993

Ursula B. Marvin
SAO
Geologist
Tenure: 1961-1998

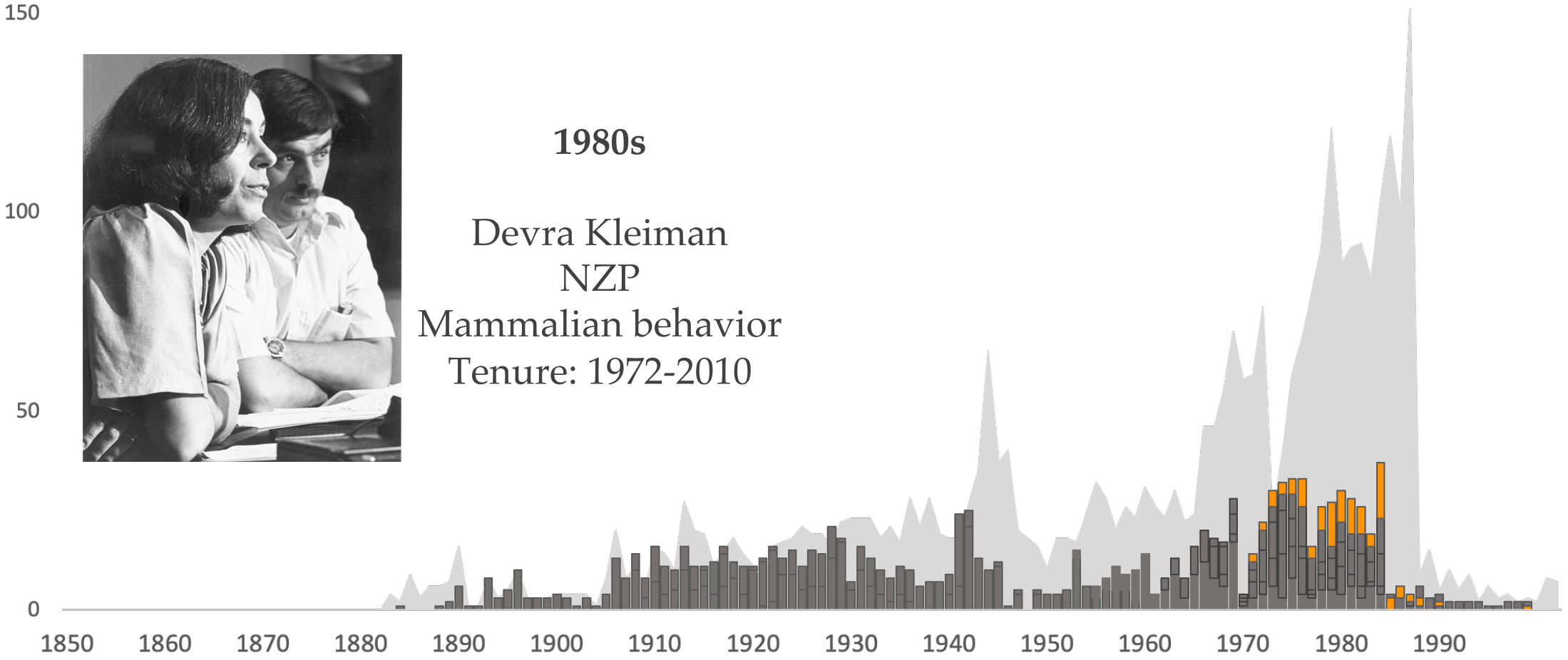
Marie-Helene Sachet
NMNH
Curator, Botany
Tenure: 1966-1986



Who are the most-mentioned women in each decade?

1980s

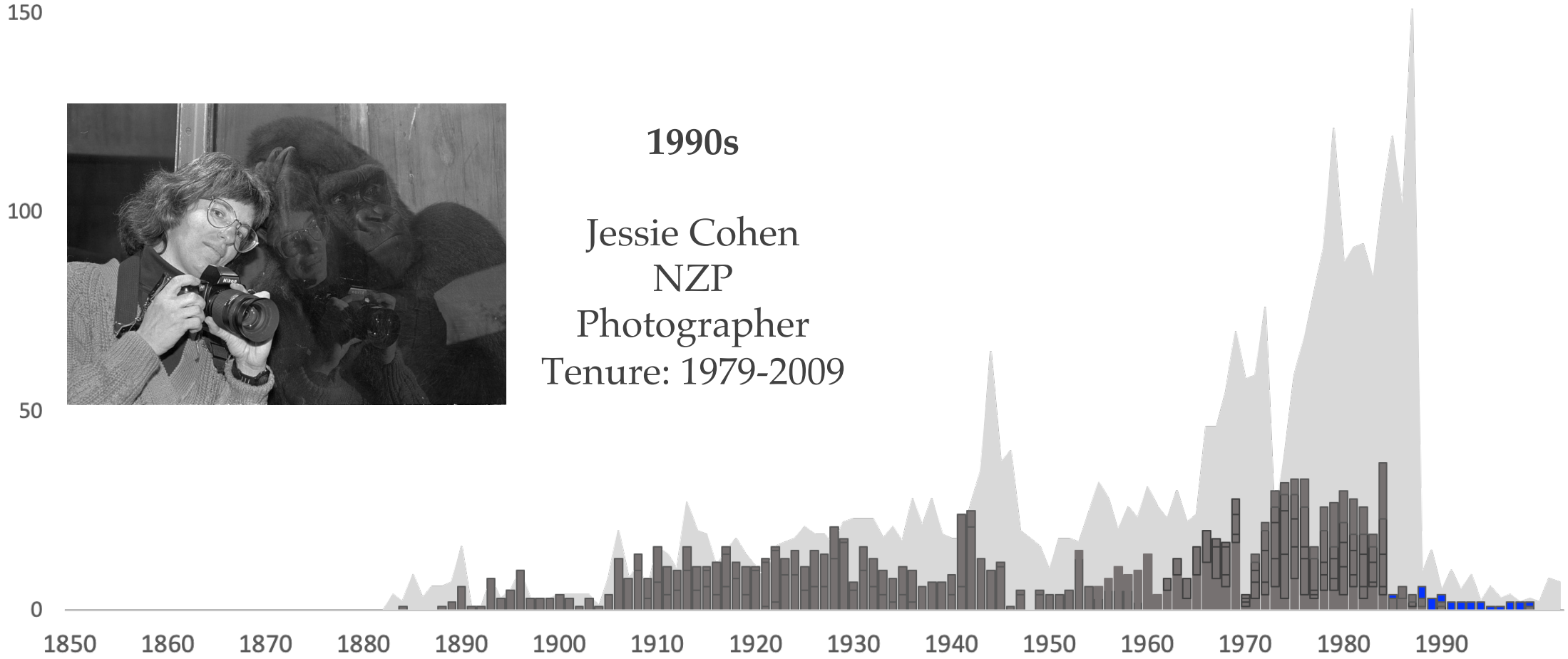
Devra Kleiman
NZP
Mammalian behavior
Tenure: 1972-2010



Who are the most-mentioned women in each decade?

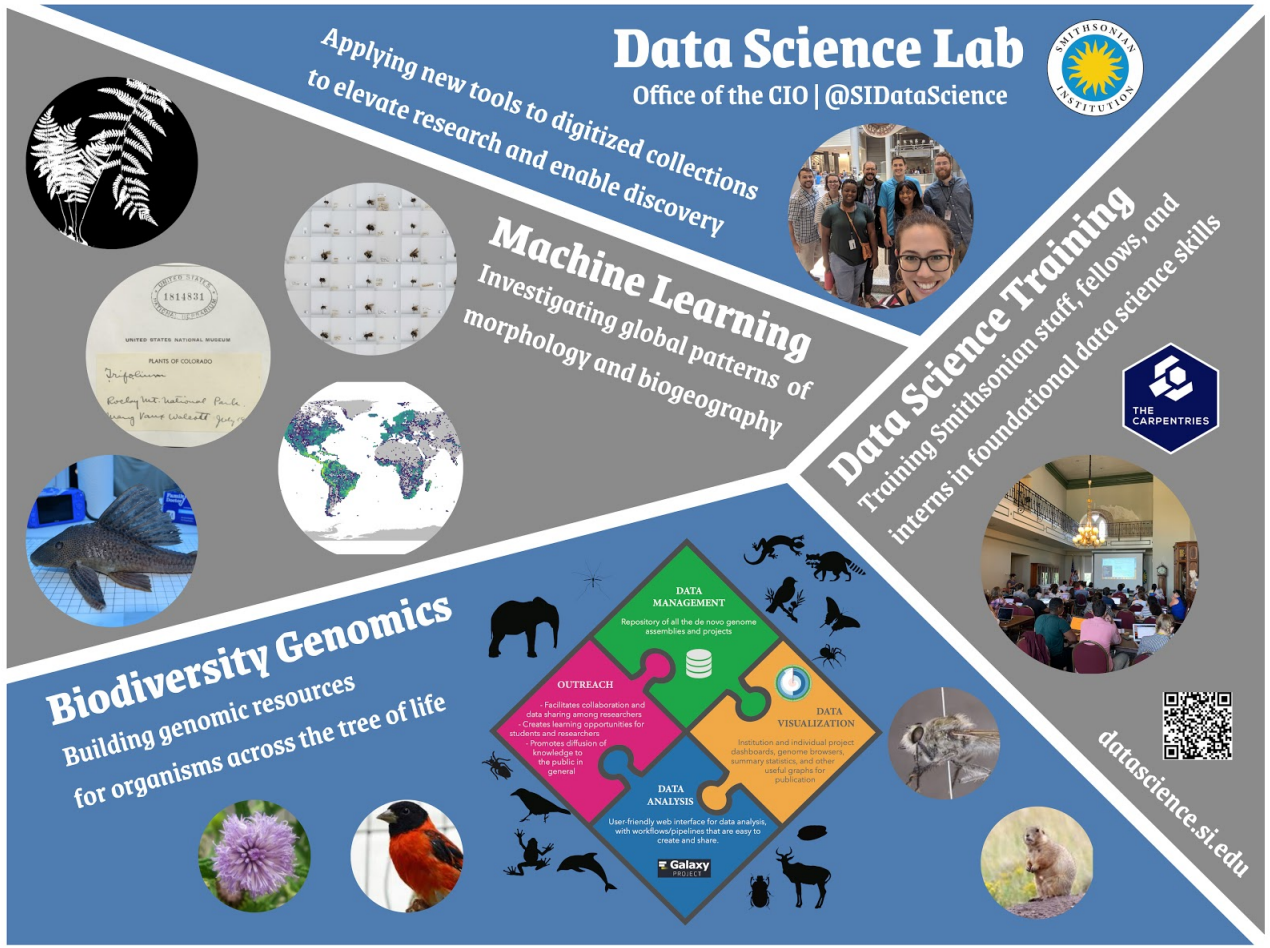
1990s

Jessie Cohen
NZP
Photographer
Tenure: 1979-2009



find more online:
datascience.si.edu

Acknowledgments

- Smithsonian Women's Committee

- Tiana Curry
- Megan Glenn
- Liz Harmon
- Effie Kapsalis
- Katrina Lohan
- Grace May
- Richard Naples
- Jenn Schneider
- Keri Thompson
- Mike Trizna