



Gender Bias on Wikipedia

Data Analytics Bootcamp – Final Presentation

Franziska Ulmer | March 13, 2023

Top 10 by Occupation (Level 1)

Ranked by Notability Index

notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
6	Male	Michael Jackson	dancer	Culture	1958	US
8	Male	William Shakespeare	playwright	Culture	1564	United Kingdom
10	Male	Elvis Presley	actor	Culture	1935	US
13	Male	Vincent van Gogh	painter	Culture	1853	Netherlands
14	Male	Wolfgang Amadeus Mozart	composer	Culture	1756	Holy Roman Empire
17	Male	Ludwig van Beethoven	composer	Culture	1770	Germany
18	Male	Pablo Picasso	painter	Culture	1881	Spain
20	Male	Ronald Reagan	actor	Culture	1911	US
23	Male	Charlie Chaplin	film	Culture	1889	United Kingdom
27	Male	John Lennon	producer	Culture	1940	United Kingdom

notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
	1	Male	Barack Obama	politician	Leadership	1961
	2	Male	Donald Trump	magnate	Leadership	1946
	4	Male	Adolf Hitler	politician	Leadership	1889
	7	Male	Winston Churchill	politician	Leadership	1874
	9	Male	Napoleon	politician	Leadership	1769
	11	Male	Mahatma Gandhi	politician	Leadership	1869
	15	Male	Nelson Mandela	politician	Leadership	1918
	18	Male	Abraham Lincoln	politician	Leadership	1809
	24	Male	Che Guevara	politician	Leadership	1928
	25	Female	Hillary Clinton	politician	Leadership	1947

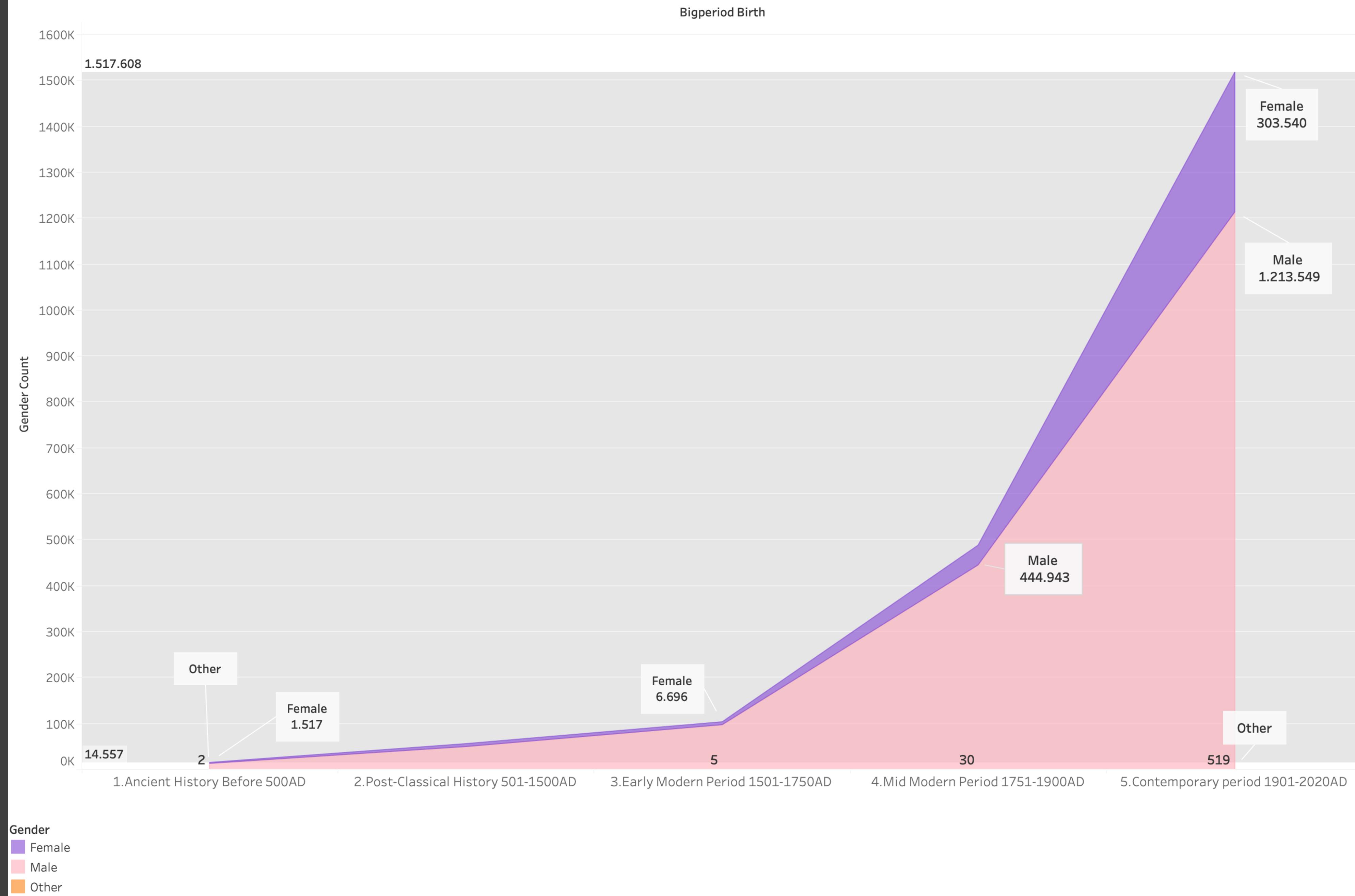
notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
	3	Male	Leonardo da Vinci	polymath	Discovery/Science	1452
	5	Male	Albert Einstein	physicist	Discovery/Science	1879
	12	Male	Charles Darwin	biologist	Discovery/Science	1809
	16	Male	Karl Marx	economist	Discovery/Science	1818
	21	Male	Sigmund Freud	psychoanalyst	Discovery/Science	1856
	22	Male	Martin Luther	linguist	Discovery/Science	1483
	26	Male	Aristotle	biologist	Discovery/Science	-383
	28	Male	Galileo Galilei	astronomer	Discovery/Science	1564
	36	Male	Plato	philosopher	Discovery/Science	-427
	37	Male	Friedrich Nietzsche	philosopher	Discovery/Science	1844

notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
61	Male	Muhammad Ali	boxer	Sports/Games	1942	US
88	Male	Lionel Messi	football	Sports/Games	1987	Argentina
136	Male	David Beckham	football	Sports/Games	1975	United Kingdom
165	Male	Michael Jordan	basket	Sports/Games	1963	US
182	Male	Cristiano Ronaldo	football	Sports/Games	1985	Portugal
200	Male	Pelé	football	Sports/Games	1940	Brazil
223	Male	Diego Maradona	football	Sports/Games	1960	Argentina
256	Male	Zlatan Ibrahimović	football	Sports/Games	1981	Sweden
256	Male	Zinedine Zidane	football	Sports/Games	1972	France
277	Male	Neymar	football	Sports/Games	1992	Brazil

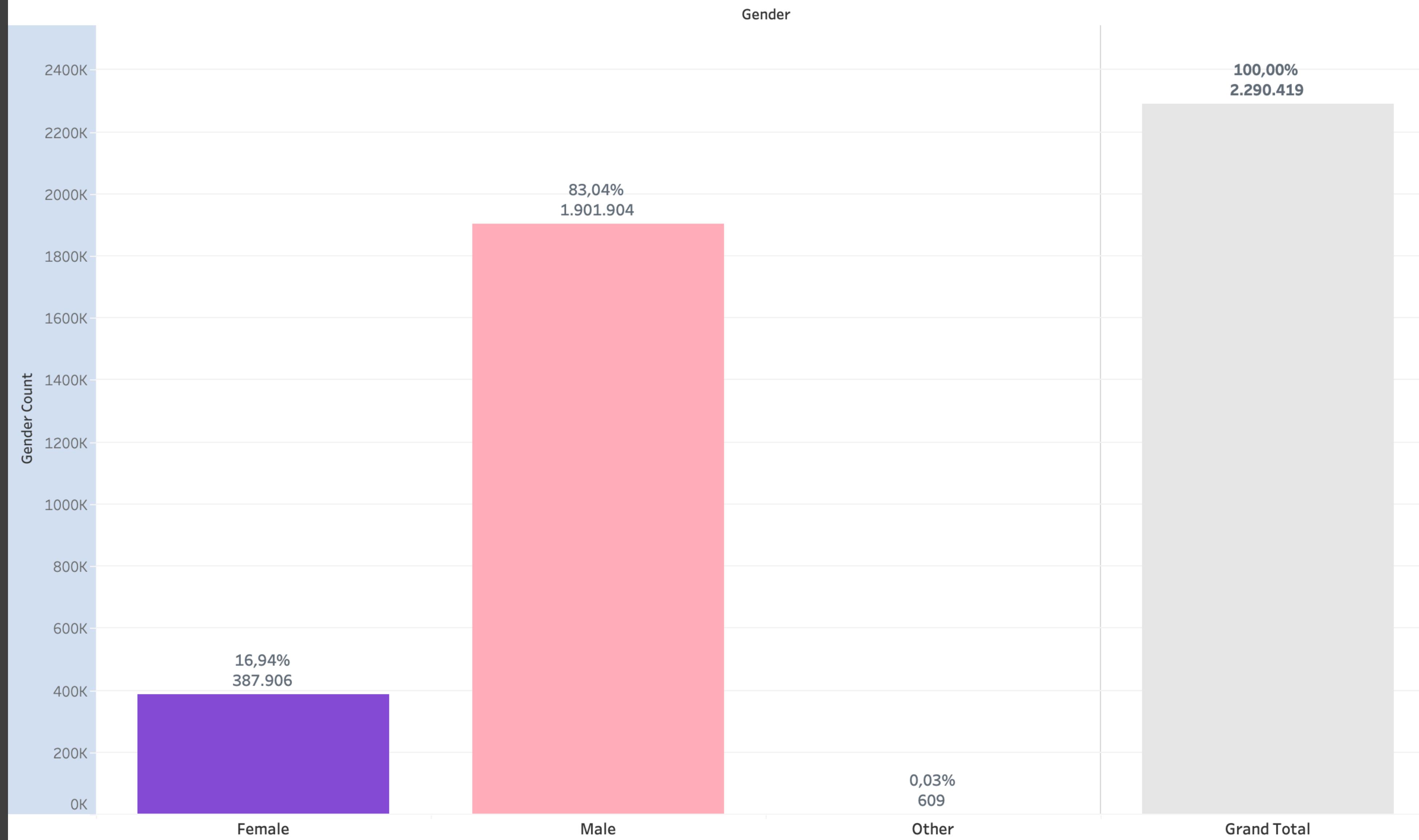
notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
	436	Male	Osama bin Laden	terrorista	Other	1957
	537	Male	Edward Snowden	employee	Other	1983
	1333	Female	Lucrezia Borgia	daughter	Other	1480
	1515	Male	Al Capone	gangster	Other	1899
	1793	Male	Simon Wiesenthal	survivor	Other	1908
	1804	Male	Charles Manson	criminal	Other	1934
	2450	Male	Lee Harvey Oswald	killer	Other	1939
	2956	Male	Germanicus	son	Other	-15
	3074	Female	Abigail Adams	mother_of	Other	1744
	3134	Male	Billy the Kid	outlaw	Other	1859

notability_index_ranking	gender	name	occup_I3	occup_I1	birth	citizenship
	65516	Female	Maria Sabina	NaN	Missing	1894
	124229	Male	Liu Zhiyuan	NaN	Missing	895
	127972	Male	Dahui Zonggao	NaN	Missing	1089
	175622	Male	Jassa Singh Ahluwalia	NaN	Missing	1718
	201966	Male	Jeremiah Brandreth	NaN	Missing	1790
	205964	Male	Shuja-ud-Daula	NaN	Missing	1732
	209974	Male	Marco Feingold	NaN	Missing	1913
	214232	Male	Tāmati Wāka Nene	NaN	Missing	1785
	214579	Male	Vajiranana Javakoski	NaN	Missing	1860
	226172	Male	Thomas Hobson	NaN	Missing	1544

Distribution of Articles across Genders by Time of Living



Wikipedia Articles of Notable People (3500 BC - 2018 AD) – Gender Distribution



Gender Count for each Gender. Color shows details about Gender. The marks are labeled by % of Total Gender Count and Gender Count. The view is filtered on Gender, which keeps Female, Male and Other.

Wikipedia Articles of Notable People (3500 BC - 2018 AD) – Mapped Gender Distribution



Map based on Longitude (generally)

Map based on Longitude (generated) and Latitude (generated). Color shows details about Gender. Size shows No. of Notable People Wikipedia. Details are shown for Map Iso3. The view is filtered on Gender, Latitude (generated) and Longitude (generated). The Gender filter keeps Female, Male and Other. The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

**Where does this “Wikipedia gender gap”
in the content come from?**

**90% of contributors
are male**

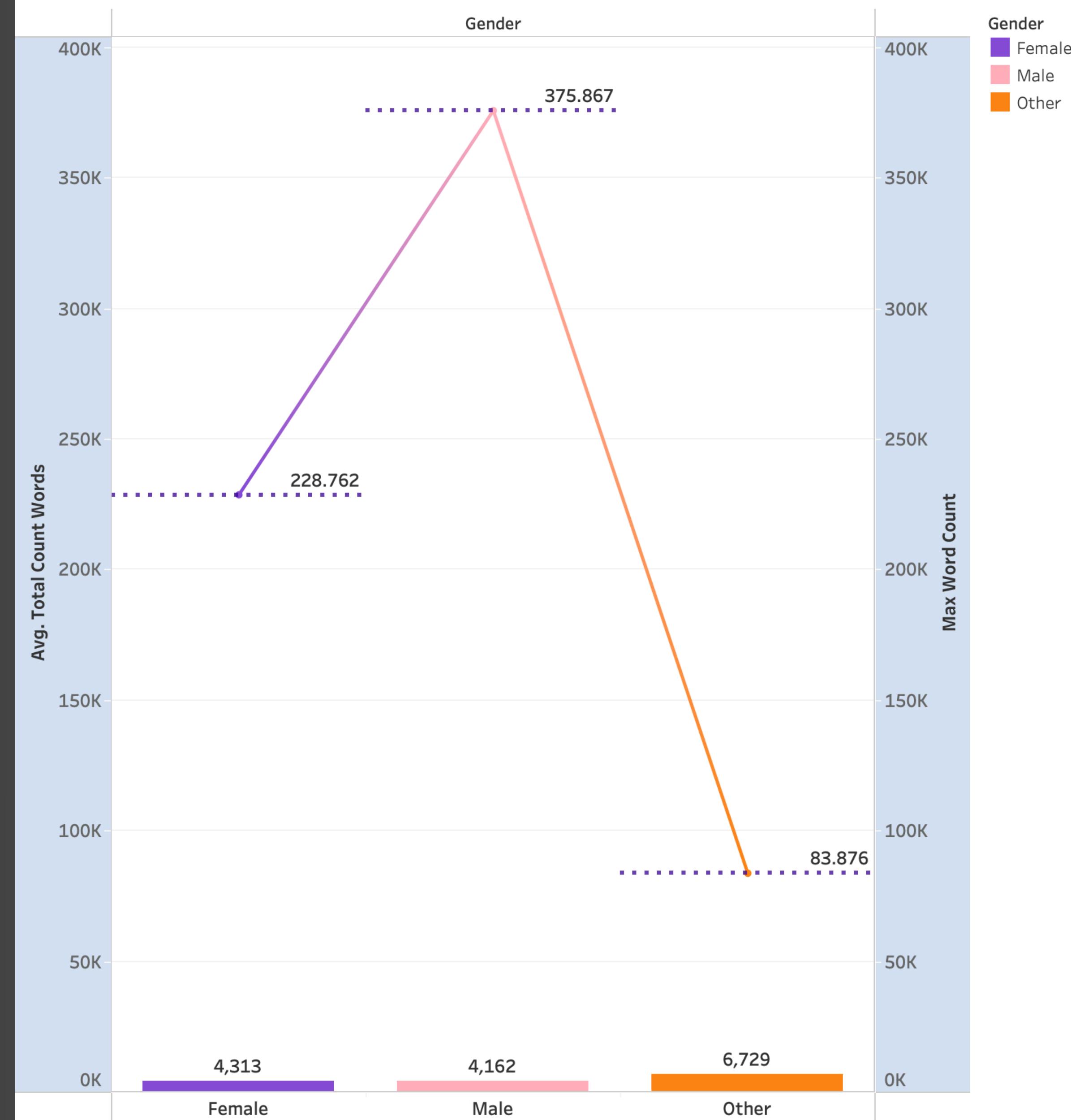
...out of 3,734 respondents of a 2018 survey covering 12 Wikipedia language editions

**8.8% reported their gender as female,
1% as other**

Articles about women are less likely to be...

- * included [in Wikipedia]
- * expanded
- * neutral [sexist, loaded, gendered language]
- * detailed

Wikipedia Articles of Notable People (3500 BC - 2018 AD) – Total Word Count Comparison



Wrapping up... & Next Steps in Analysis

Data Set

- * Include more specific gender information apart from male/female
→ “Other” is too broad
- * Enrich data set with most current data (from 2019, esp. incl. COVID)

Analysis

- * German edition
- * Detailing the 5 components of the Notability Index

Tech Stack

Python/Pandas | Jupyter Notebook | Plotly Express/Seaborn | Tableau

References

- * Data Set:
LAOUENAN, Morgane; BHARGAVA, Palaash; EYMOUD, Jean-Benoît; GERGAUD, Olivier; PLIQUE, Guillaume;
WASMER, Etienne, 2022, "A Brief History of Human Time - Cross-verified Dataset", <https://doi.org/10.21410/7E4/RDAG3O>, data.sciencespo, V2
- * Ecosia Image Search:
"Barack Obama", "Hillary Clinton" – March 10, 2023
- * Wikipedia logo:
https://en.wikipedia.org/wiki/Main_Page
- * Gender Bias Definition:
<https://dictionary.apa.org/gender-bias>
- * Article "Gender bias on Wikipedia":
https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia
- * Memes:
<https://www.facebook.com/spoiltmodernwoman>, <https://www.facebook.com/hashtag/genderequality>

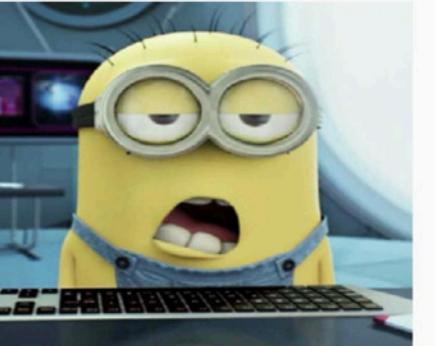
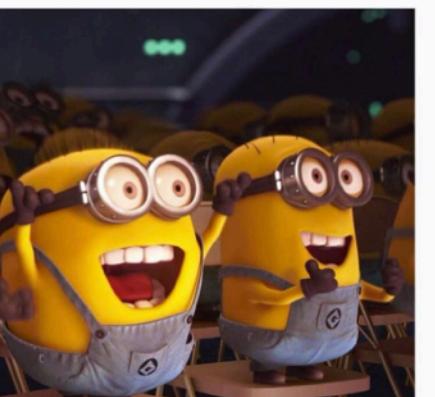
 **myisland** 
@opinionated75

Please pray for my son who had to empty the dish washer when “he just did this yesterday and he’s tired”.

Society noticeboard

We are celebrating Women's day this weekend!!

Event includes:
skin care session ,
bakery workshop ,
fashion show 

Thank You ;-)



 **Mr. Red Pill**
@JonesGfashaw

Do women still cook?

4:43 AM · 12 Jun 19 · Twitter for iPhone

 **CVS Rewards Member**
@mdeeeeeee_

Do men still hunt and gather?

 **Raza** 
@TrademarkRaza

if my wife makes 6 figures my place is in the kitchen, it's called cooKING not cookQUEEN for a reason

08:56 · 03/10/21 · Twitter for iPhone

Appendix

What is Gender Bias?

**“any one of a variety of stereotypical beliefs
about individuals on the basis of their sex
[...]”**

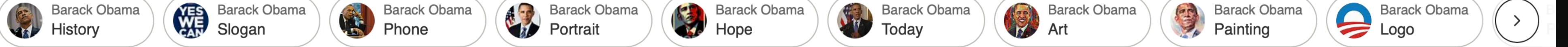
— APA Dictionary of Psychology

ECOSIA barack obama 

All Images News Videos Maps More Settings

Search region: Germany All sizes All colors All types Any time All layouts All usage rights

Barack Obama History Barack Obama Slogan Barack Obama Phone Barack Obama Portrait Barack Obama Hope Barack Obama Today Barack Obama Art Barack Obama Painting Barack Obama Logo >



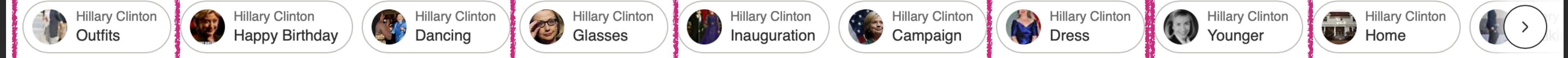
"Absolute Chaotic Disaster": In a Campaign ... Interview With Barac... What books former President Barack Obam... Barack Obama Announces ... Two Nigerians featured in Ba... Barack Obama I The White ...
www.vanityfair.com www.popsugar.com www.cnbc.com www.everythingzoomer.com www.adomonline.com www.whitehouse.gov

ECOSIA hillary clinton 

All Images News Videos Maps More Settings

Search region: Germany All sizes All colors All types Any time All layouts All usage rights

Hillary Clinton Outfits Hillary Clinton Happy Birthday Hillary Clinton Dancing Hillary Clinton Glasses Hillary Clinton Inauguration Hillary Clinton Campaign Hillary Clinton Dress Hillary Clinton Younger Hillary Clinton Home >



Secretary Hillary Clinton t... Hillary Clinton C... Hillary Clinton appears in 'SNL' skit - ... Why Hillary Clinton is the world's greate... Hillary Clinton withholding donations f... Hillary Clinton DEFENDS emails and c...
www.laughingplace.com time.com www.bostonglobe.com nypost.com www.salon.com altnewscoin.com

Data Set & Tech Stack

Data

A cross-verified database of notable people, 3500BC-2018AD

- data extracted from 7 different editions of Wikipedia + Wikidata
- includes 2.29 million individuals

Tech Stack

Python/Pandas | Jupyter Notebook | Plotly Express/Seaborn | Tableau

Top 25 Overall & Top 10 by Gender Ranked by Notability Index

Top 25 Wikipedia Articles of Notable People by Notability Index			
Name	Occup L3	Gender	
		Female	Male
Barack Obama	politician		1
Donald Trump	magnate		2
Leonardo da Vinci	polymath		3
Adolf Hitler	politician		4
Albert Einstein	physicist		5
Michael Jackson	dancer		6
Winston Churchill	politician		7
William Shakespeare	playwright		8
Napoleon	politician		9
Elvis Presley	actor		10
Mahatma Gandhi	politician		11
Charles Darwin	biologist		12
Vincent van Gogh	painter		13
Wolfgang Amadeus Mozart	composer		14
Nelson Mandela	politician		15
Karl Marx	economist		16
Ludwig van Beethoven	composer		17
Abraham Lincoln	politician		18,5
Pablo Picasso	painter		18,5
Ronald Reagan	actor		20
Sigmund Freud	psychoanalyst		21
Martin Luther	linguist		22
Charlie Chaplin	film		23
Che Guevara	politician		24
Hillary Clinton	politician	25	

Sum of Notability Index Ranking broken down by Gender vs. Name and Occup L3. Color shows sum of Notability Index Ranking. The marks are labeled by sum of Notability Index Ranking. The data is filtered on Wikidata Code Set, which keeps 25 members.

notability_index_ranking	gender	name	occup_l3	occup_l1	birth	citizenship
25	Female	Hillary Clinton	politician	Leadership	1947	US
32	Female	Madonna	singer	Culture	1958	US
34	Female	Marilyn Monroe	actor	Culture	1926	US
58	Female	Lady Gaga	singer	Culture	1986	US
59	Female	Britney Spears	singer	Culture	1981	US
66	Female	Marie Curie	physicist	Discovery/Science	1867	Poland
68	Female	Angelina Jolie	actor	Culture	1975	US
71	Female	Elizabeth I of England	queen	Leadership	1533	United Kingdom
72	Female	Queen Victoria	queen	Leadership	1819	United Kingdom
76	Female	Rihanna	singer	Culture	1988	Barbados

notability_index_ranking	gender	name	occup_l3	occup_l1	birth	citizenship
1923	Other	Caitlyn Jenner	olympic	Sports/Games	1949	US
2145	Other	Steven Tyler	singer	Culture	1948	US
3001	Other	Sam Smith	singer	Culture	1992	United Kingdom
8650	Other	Lili Elbe	painter	Culture	1882	Denmark
8721	Other	Dana International	singer	Culture	1969	Israel
9293	Other	Alexis Arquette	actor	Culture	1969	US
9528	Other	Wendy Carlos	composer	Culture	1939	US
9873	Other	Jake Zyrus	singer	Culture	1992	Philippines
10863	Other	Richard O'Brien	actor	Culture	1942	United Kingdom
11452	Other	Amandla Stenberg	actor	Culture	1998	US

Where does this “Wikipedia gender gap” come from?

Wikipedia depends on secondary sources,
which are similarly biased.

Why is contribution of non-males so low?

- * Dominance of white, high-skilled, cis-gendered males
- * Editors are more likely to write about their own gender
- * Unappealing atmosphere within the contributors community
 - * includes difficult and high-conflict people, even misogynists
 - * intimidating, contentious content discussions
- * Lack of confidence editing other contributors' work
- * Educational & (technical) skills gap
- * Anonymity often perceived as male
- * Rejection

How to contribute to Wikipedia?

Requirements

- * Wikipedia has a free to edit policy: “anyone can edit in good faith”
- * Follow “Wikipedia’s policies, including being verifiable by published sources”
- * “Experienced editors watch and patrol bad edits”

“Wikipedia has tested the wisdom of the crowd since 2001
and found that it succeeds.”