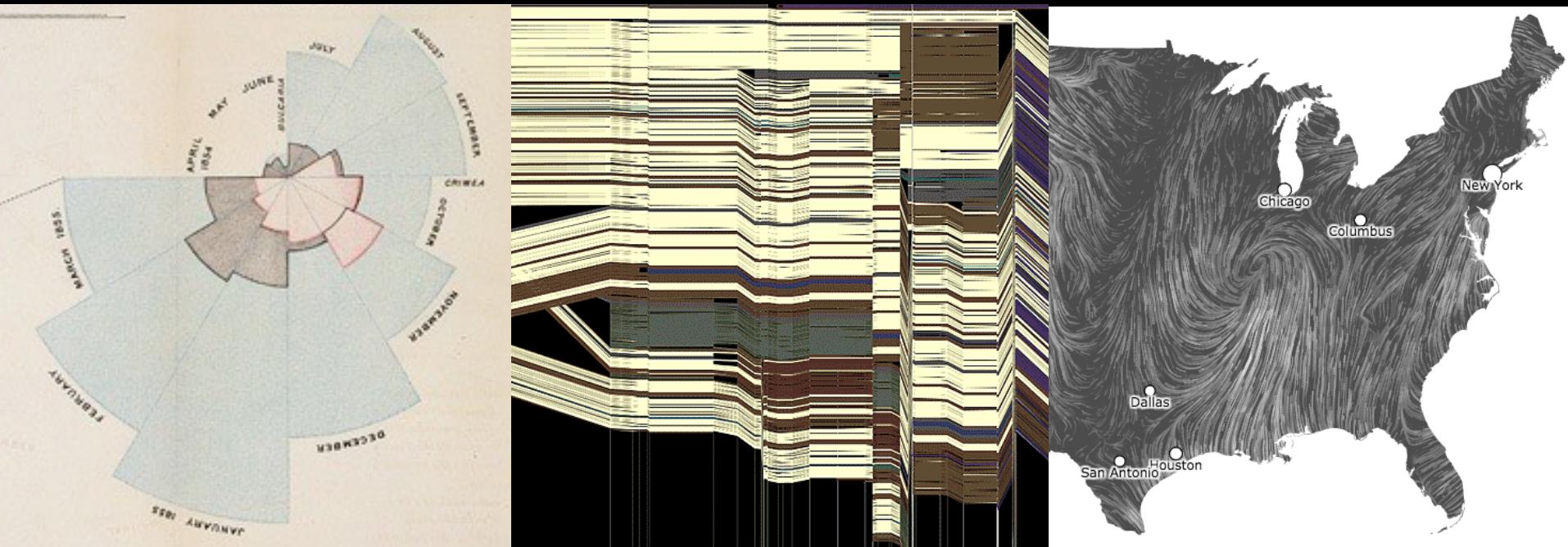


CSE 442 - Data Visualization

A1 Review



Jeffrey Heer University of Washington

Last Time:
Data & Image Models

The Big Picture

task
questions, goals
assumptions

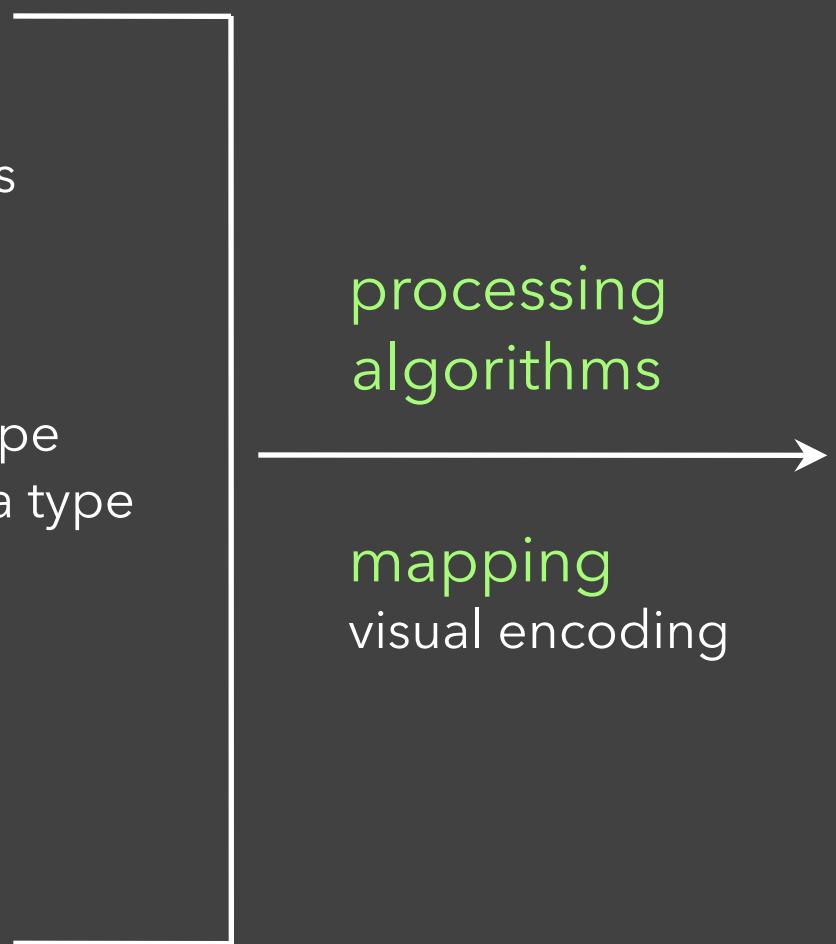
data
physical data type
conceptual data type

domain
metadata
semantics
conventions

processing
algorithms

mapping
visual encoding

image
visual channel
graphical marks



Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Operations: $=, \neq$

O - Ordered

- Operations: $=, \neq, <, >$

Q - Interval (location of zero arbitrary)

- Operations: $=, \neq, <, >, -$
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: $=, \neq, <, >, -, \%$
- Can measure ratios or proportions

Visual Encoding Variables

Position (x 2)

Size

Value

Texture

Color

Orientation

Shape

Others?

LES VARIABLES DE L'IMAGE			
	POINTS	LIGNES	ZONES
XY 2 DIMENSIONS DU PLAN			
Z			
VALEUR			
LES VARIABLES DE SÉPARATION DES IMAGES			
GRAIN			
COULEUR			
ORIENTATION			
FORME			

Bertin's “Levels of Organization”

Position

N	O	Q
---	---	---

Nominal

Size

N	O	Q
---	---	---

Ordinal

Value

N	O	Q
---	---	---

Quantitative

Texture

N	o	
---	---	--

Color

N		
---	--	--

Orientation

N		
---	--	--

Shape

N		
---	--	--

Note: **Q** ⊂ **O** ⊂ **N**

Choosing Visual Encodings

Assume k visual encodings and n data attributes.
We would like to pick the “best” encoding among
a combinatorial set of possibilities of size $(n+1)^k$

Principle of Consistency

The properties of the image (visual variables)
should match the properties of the data.

Principle of Importance Ordering

Encode the most important information in the
most effective way.

Design Criteria [Mackinlay 86]

Expressiveness

A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness

A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Design Criteria *Translated*

Tell the truth and nothing but the truth
(don't lie, and don't lie by omission)

Use encodings that people decode better
(where better = faster and/or more accurate)

Effectiveness Rankings

[Mackinlay 86]

QUANTITATIVE

Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL

Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL

Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume

Effectiveness Rankings

[Mackinlay 86]

QUANTITATIVE

Position

Length

Angle

Slope

Area (Size)

Volume

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Shape

ORDINAL

Position

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Length

Angle

Slope

Area (Size)

Volume

Shape

NOMINAL

Position

Color Hue

Texture

Connection

Containment

Density (Value)

Color Sat

Shape

Length

Angle

Slope

Area

Volume

Effectiveness Rankings

[Mackinlay 86]

QUANTITATIVE

Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL

Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL

Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume

A1 Review

A1 Submission Designs

Measures: Population count, Gender ratio, Growth rate, Age, Difference between years or genders

Transforms: Percentages, Counts, Proportions

Marks: Bar, Line/Area, Dot/Scatter, Pie, Other

Bars: Stacked, Grouped, Opposed Axes

Extra Context: Labor Force, Other Countries

Design Considerations

Title, labels, legend, captions, source!

Expressiveness and Effectiveness

Avoid unexpressive marks (lines? gradients?)

Use perceptually effective encodings

Don't distract: faint gridlines, pastel highlights/fills

The "elimination diet" approach – start minimal

Support comparison and pattern perception

Between elements, to a reference line, or to totals

Use human-friendly units (10M or 10,000,000?)

Design Considerations

Transform data (e.g., invert, log, normalize)

Group / sort data by meaningful dimensions

Reduce cognitive overhead

Minimize visual search, minimize ambiguity

Appropriate size, aspect ratio, legible text

Avoid legend lookups if direct labeling works

Avoid color mappings with indiscernible colors

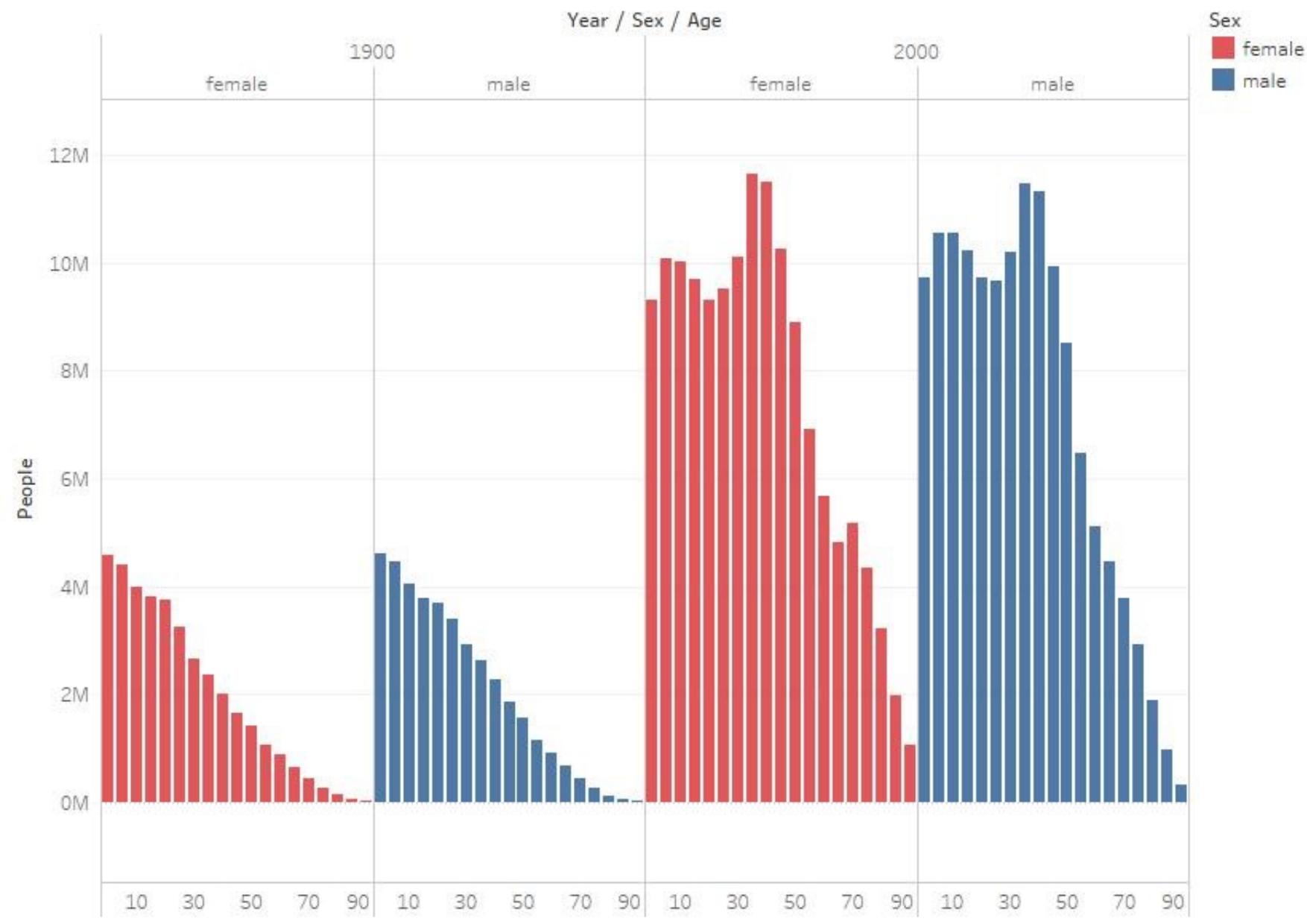
Be consistent! Visual inferences should consistently support data inferences.

Bar Charts

Age, Sex, Year

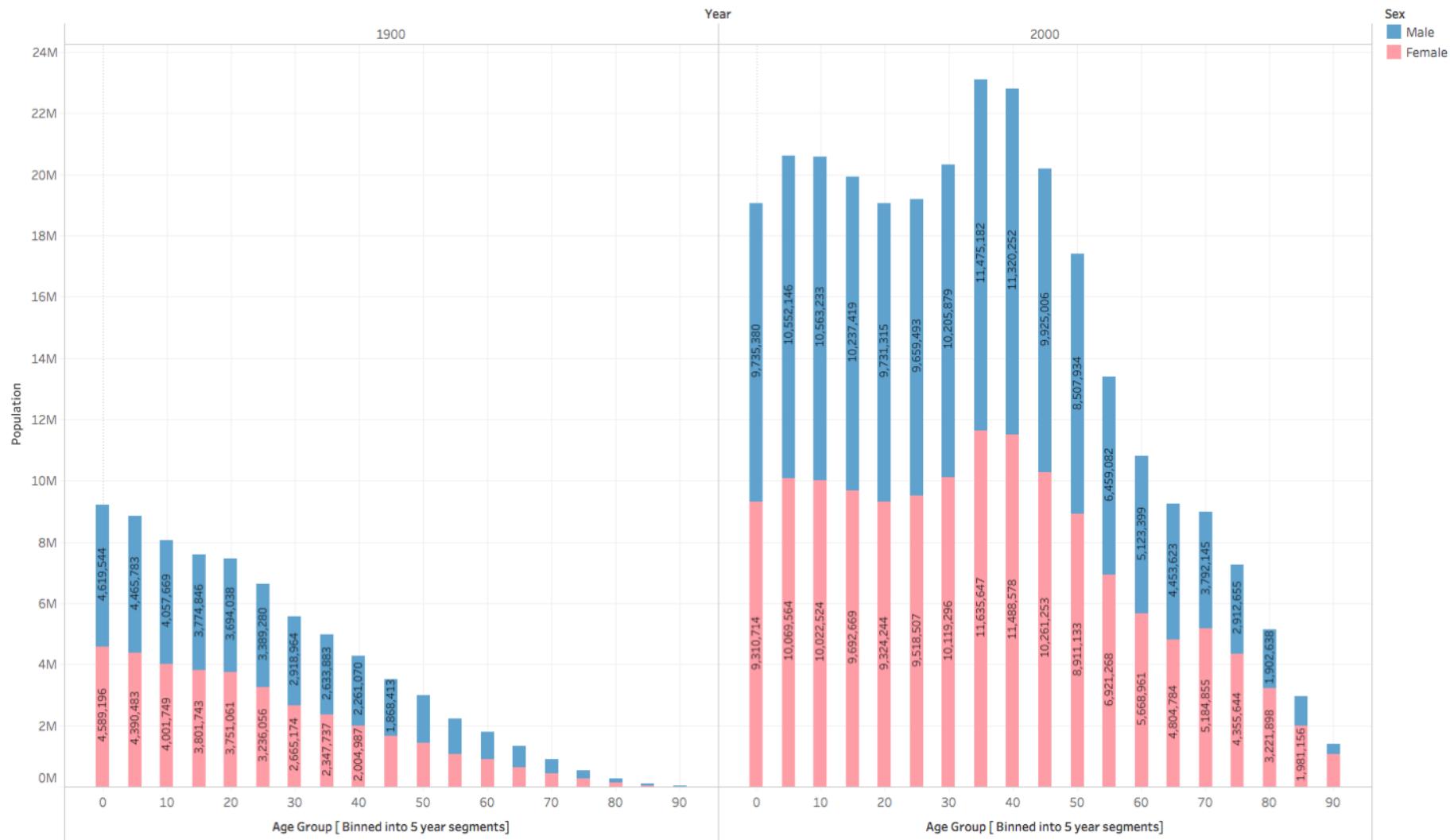
Population Count

Population of different age groups in 1900 and 2000



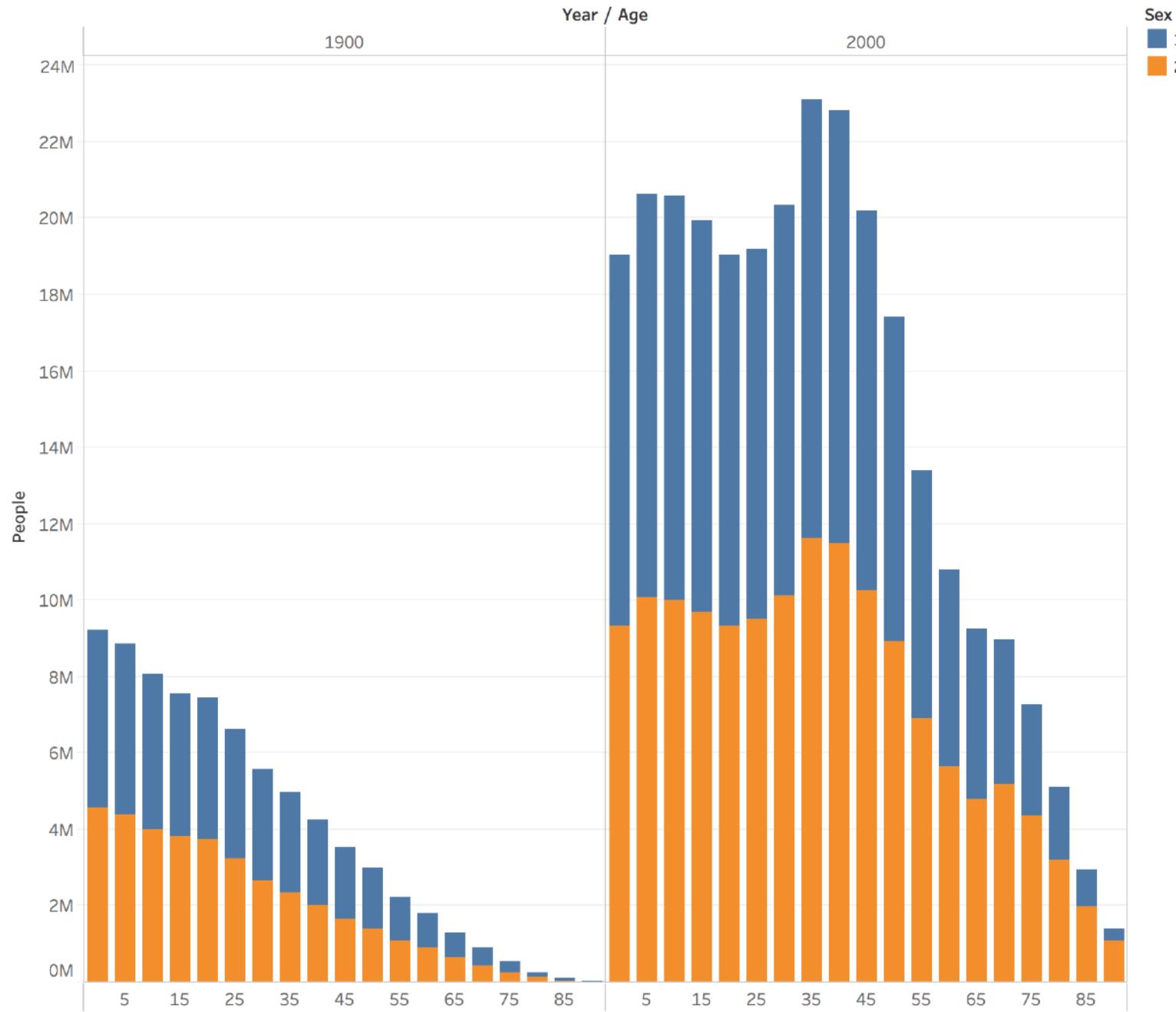
Sum of People for each Age broken down by Year and Sex. Color shows details about Sex.

How different is the US Population of 1900 compared to the population in 2000 by age group and sex?



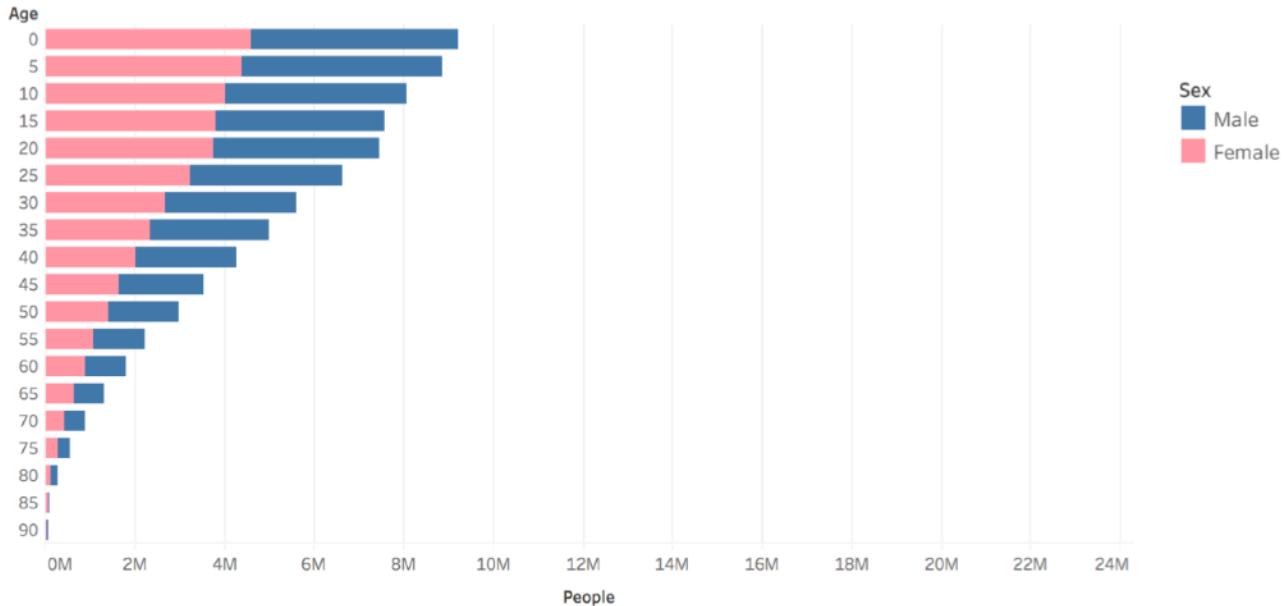
The plot of Population(Sum of People) for Age broken down by Age Group(Year). Color shows details about Sex with Labels denoting Population for that Sex within the Age Group. The data is filtered on Year, which ranges from 1900 to 2000 discretely.

Population of different ages and sexes: 1900 v.s. 2000

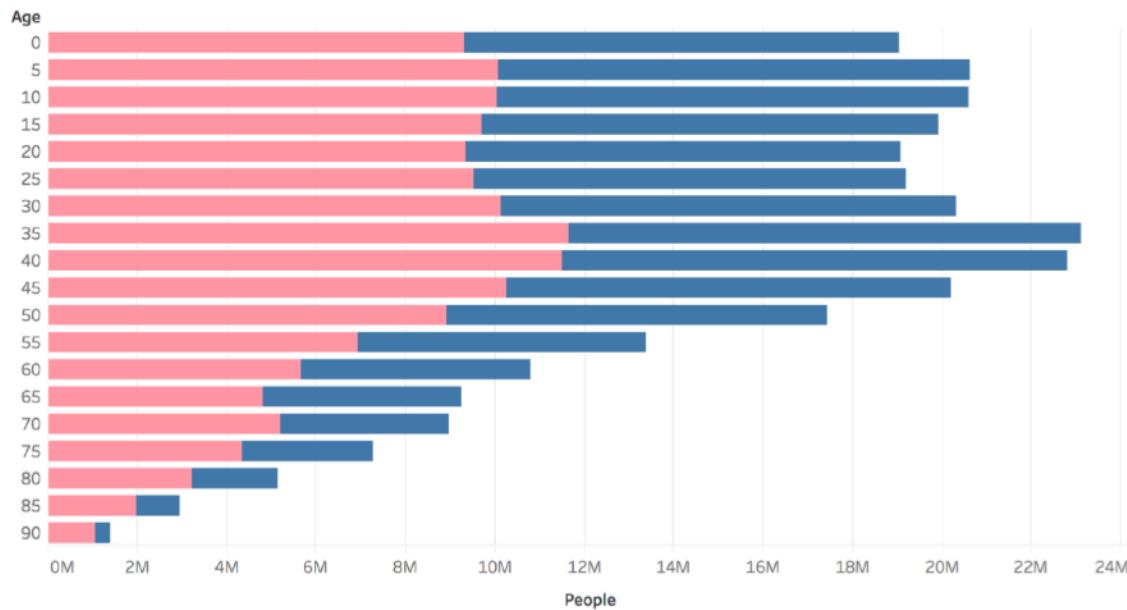


How does the age distribution of the US population vary between 1900 and 2000?

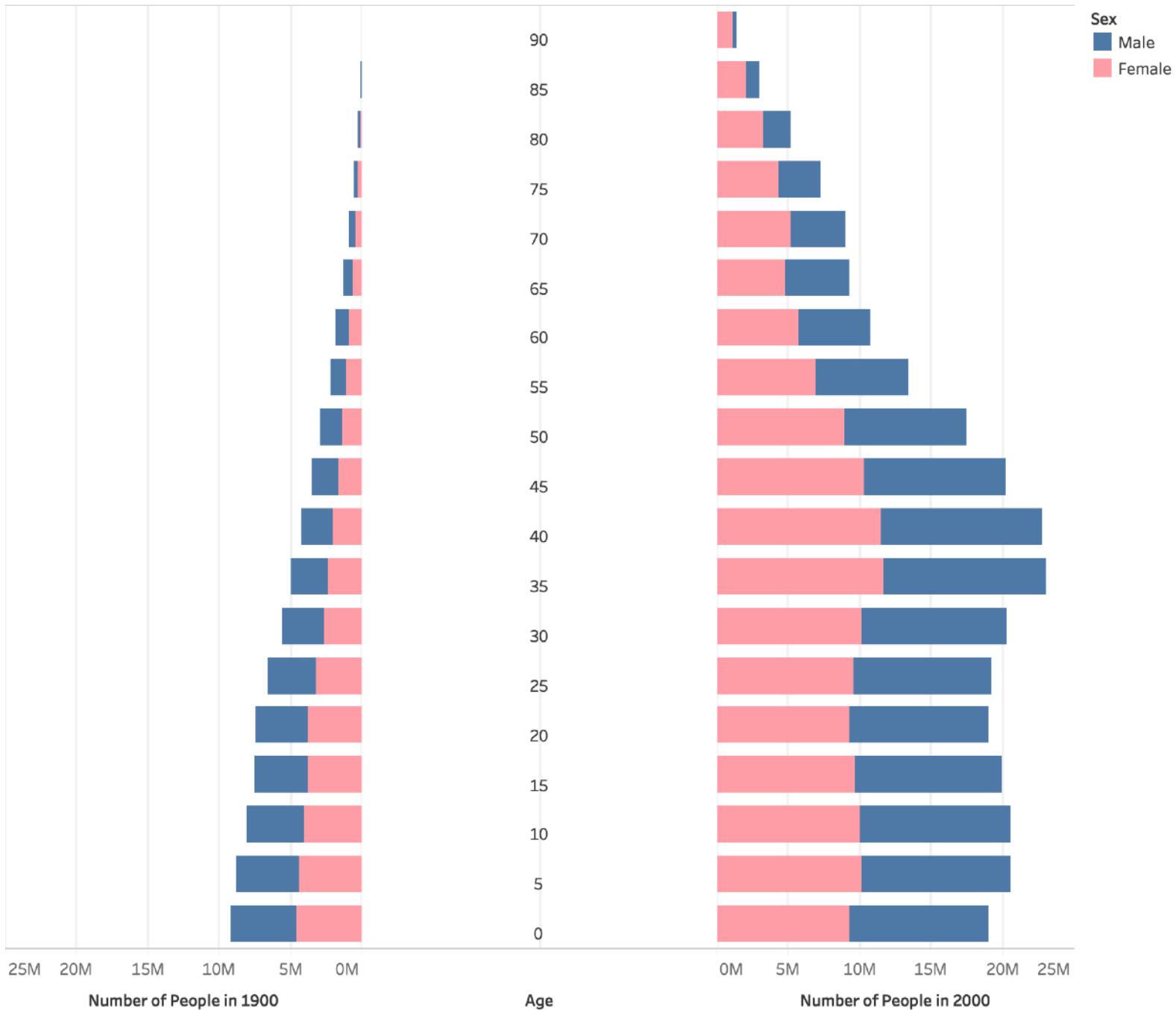
1900



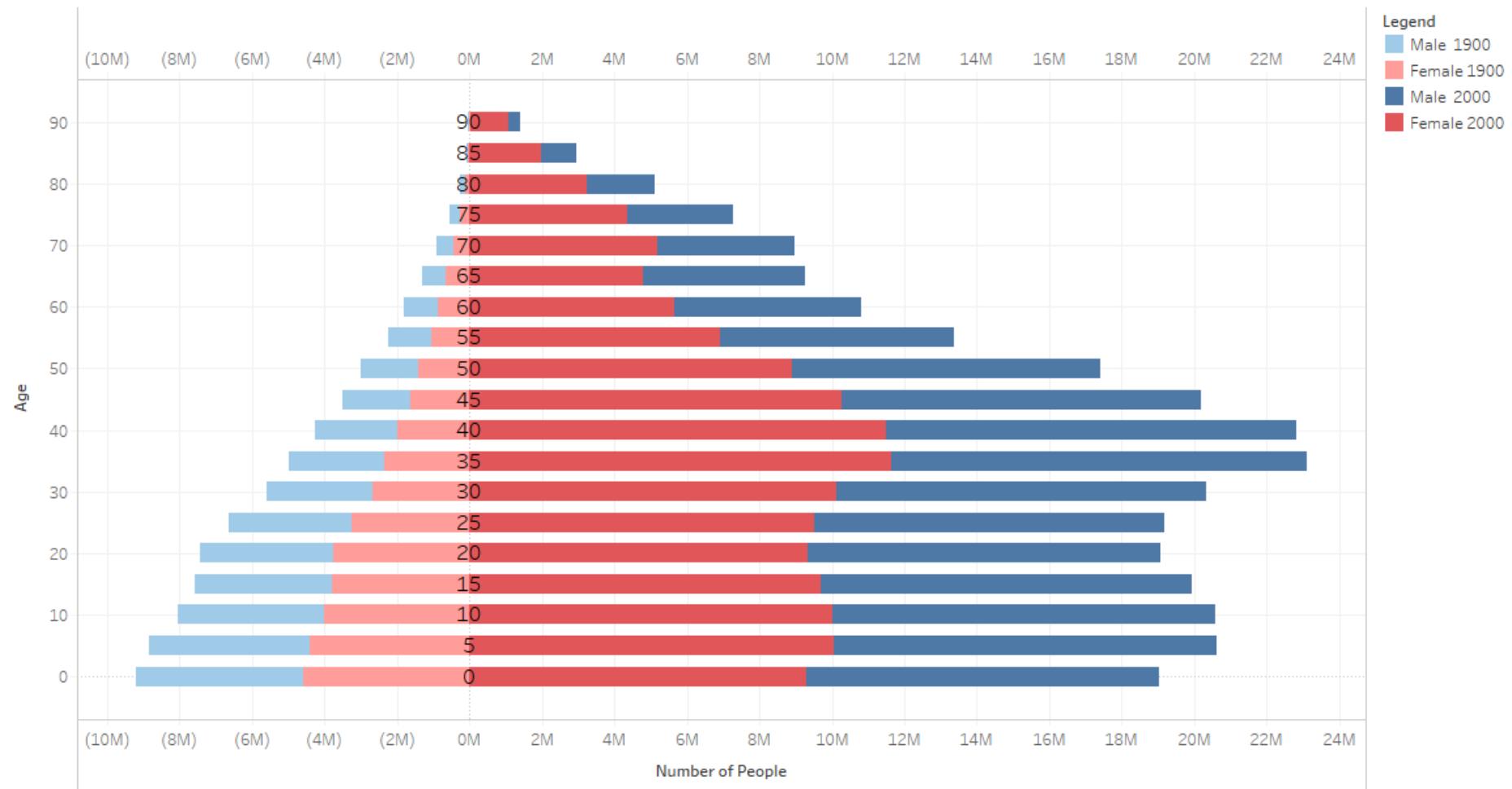
2000



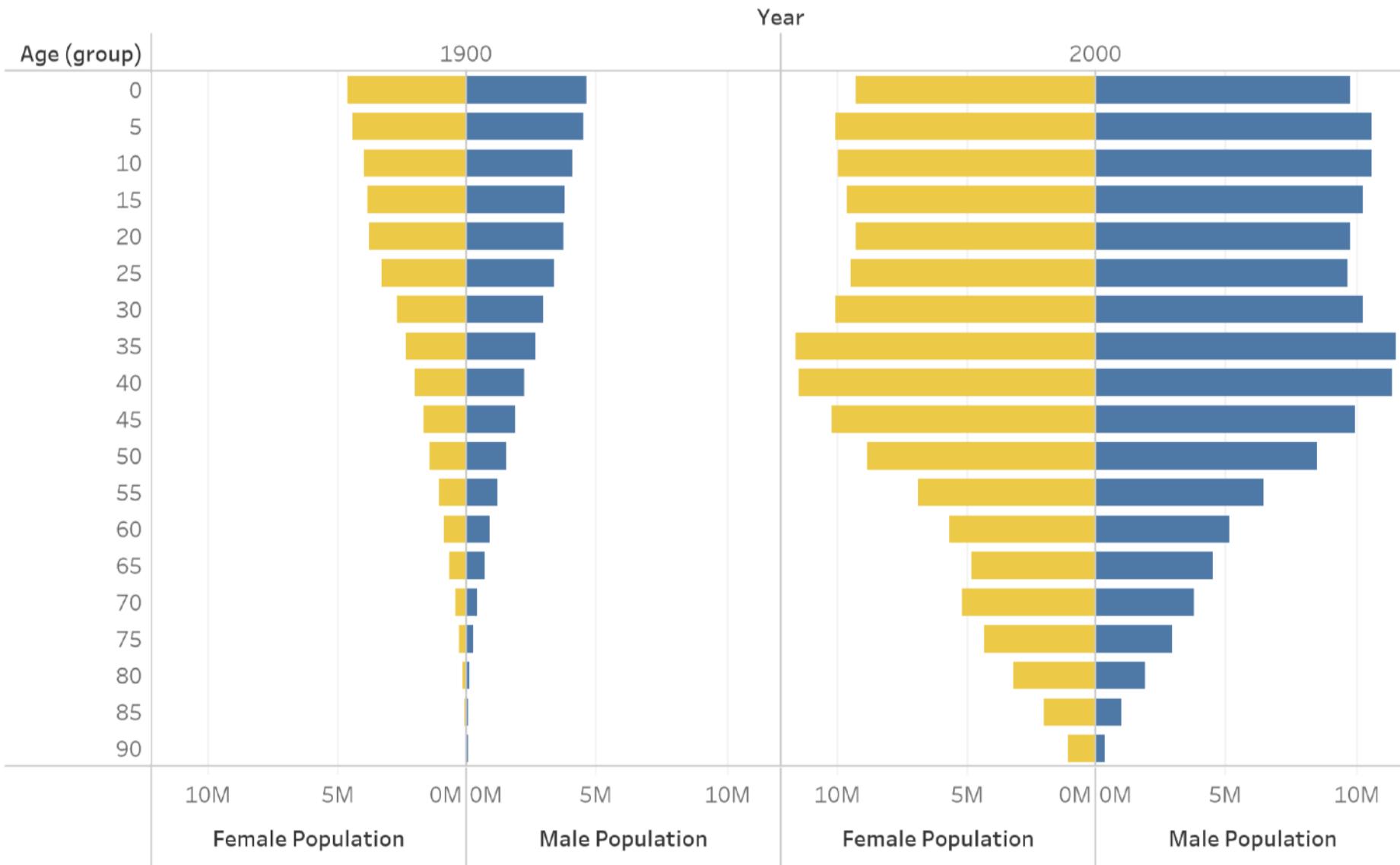
What has U.S population changed in Age and Sex composition from 1900 to 2000?



Does 100 years of societal advances increase lifespan difference between genders?

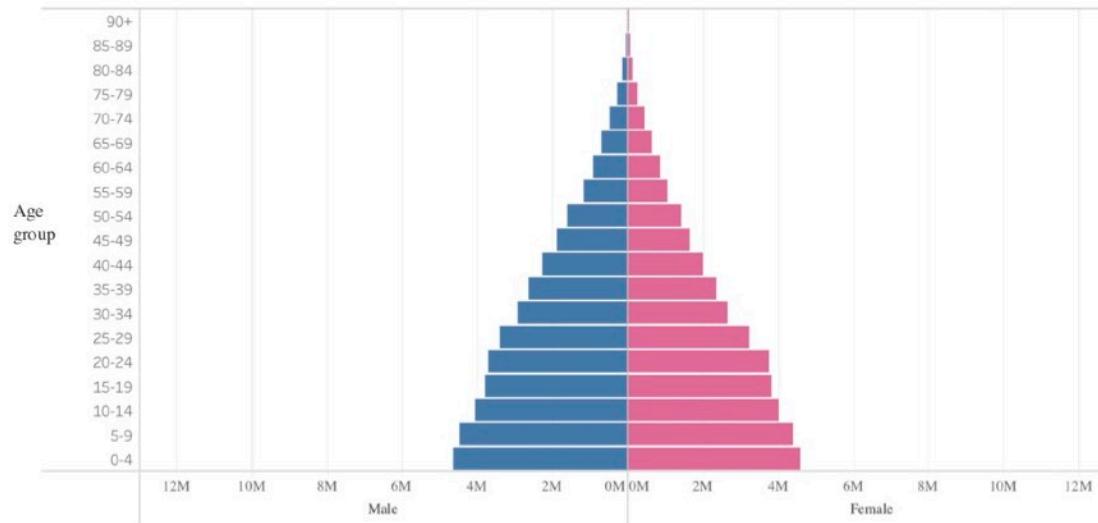


How has the Population Distribution of the US Changed From 1900 to 2000?

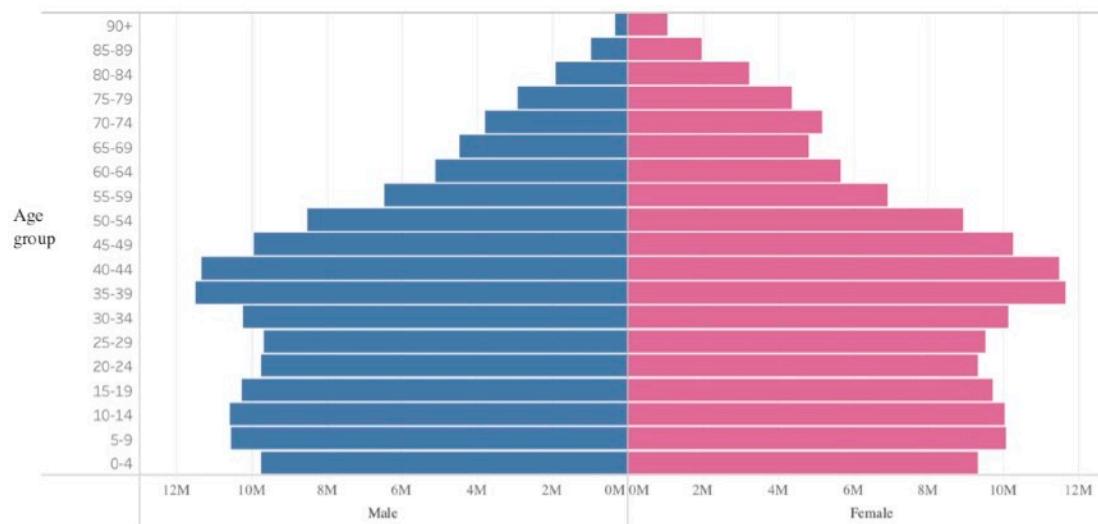


How does the U.S population's general trend change in 1900-2000?

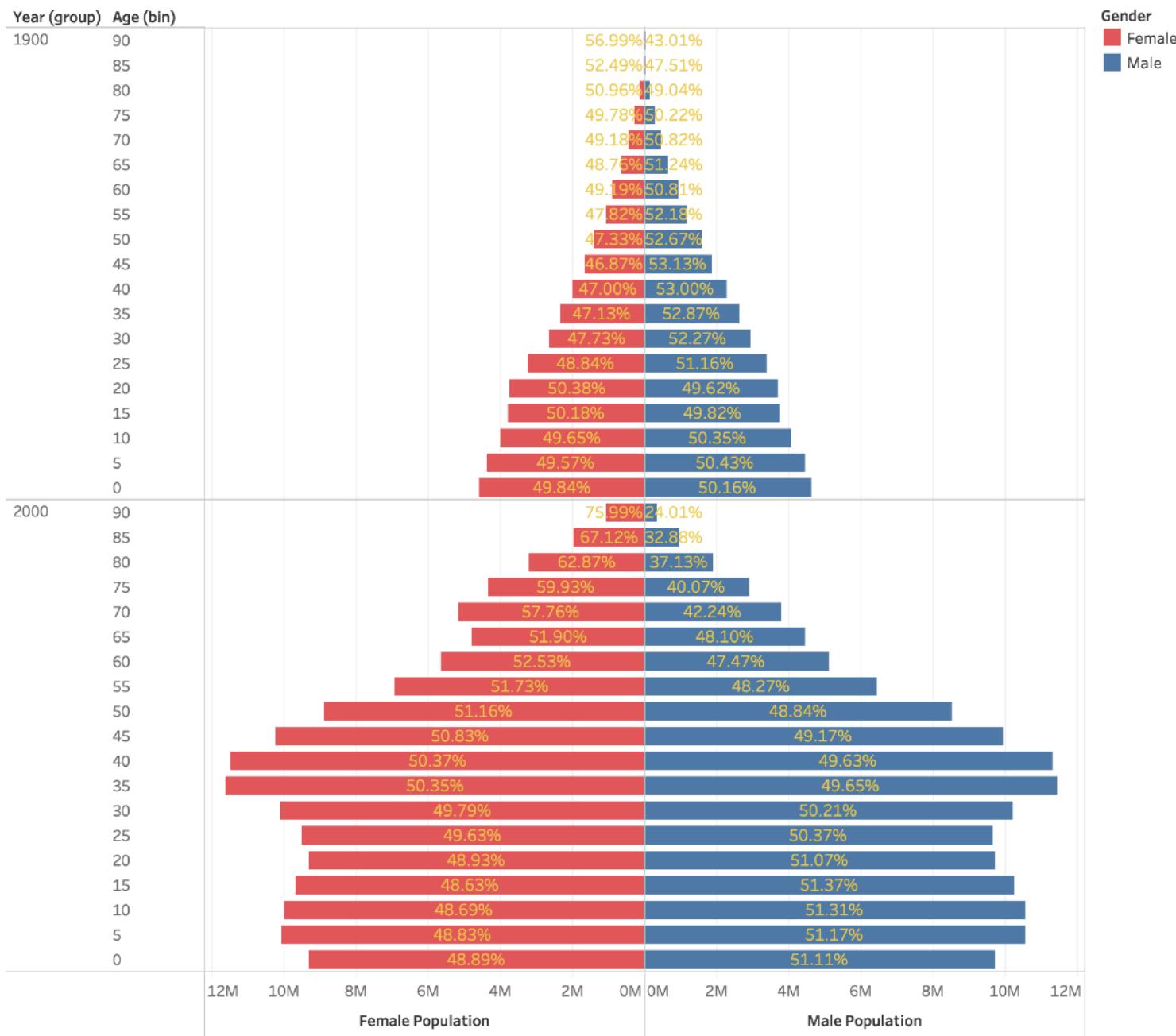
population pyramid - 1900



population pyramid - 2000

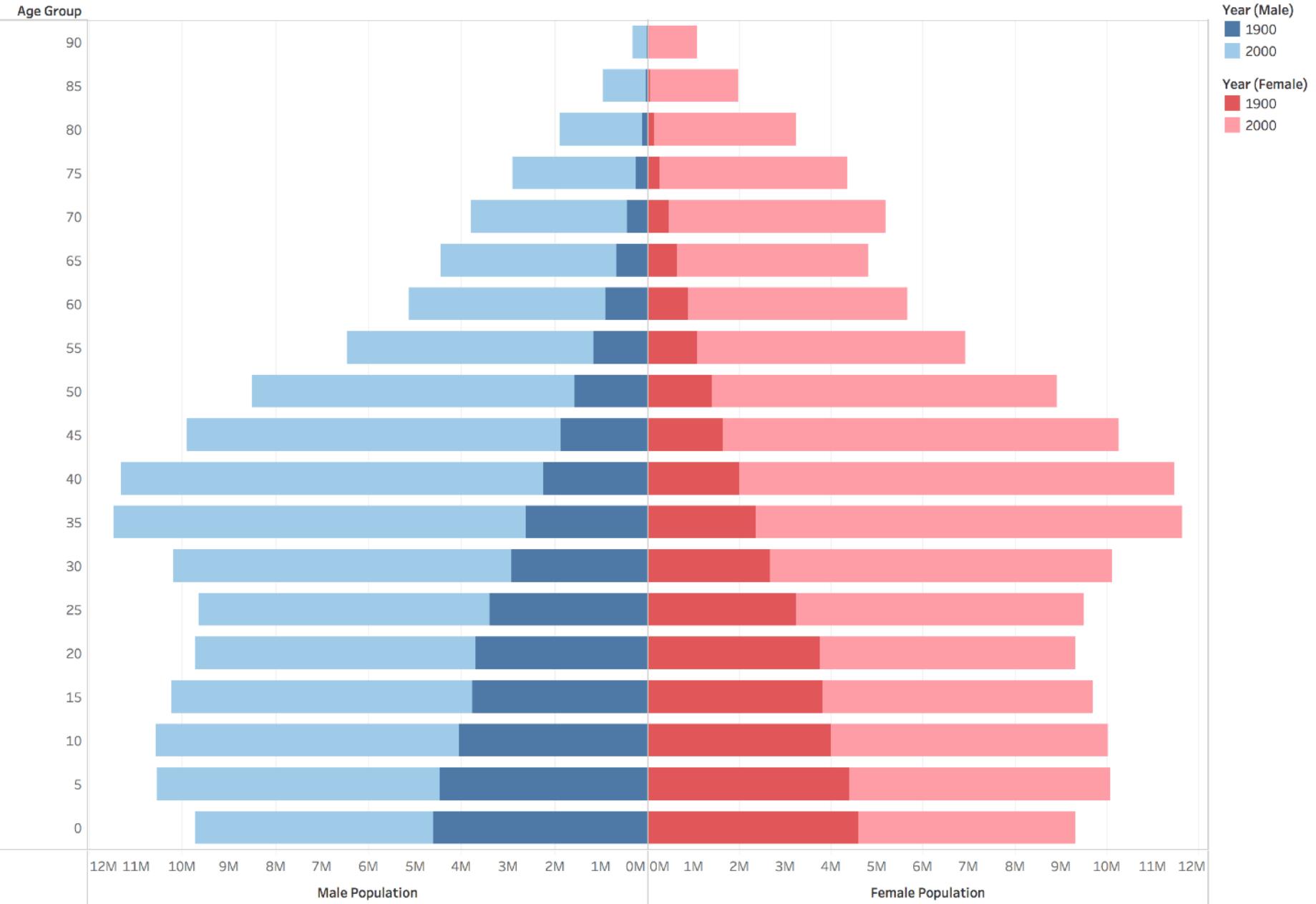


What is the gender and population structure across different age groups in 100 years?



Sum of Female Population and sum of Male Population for each Age (bin) broken down by Year (group). Color shows details about Gender. The marks are labeled by % of Total 1900 Population and % of Total 2000 Population. The view is filtered on Year (group), which keeps 1900 and 2000.

Population Structure Difference between 1900 and 2000



Sum of Male Population and sum of Female Population for each Age Group. For pane Sum of Male Population: Color shows details about Year (Male). For pane Sum of Female Population: Color shows details about Year (Female).

Growth of U.S. population of Female & Male

From 1900 TO 2000

- : Population of female in 1900
- : Population of male in 1900
- : growth of population from 1900 to 2000

Age

90+



80-89



70-79



60-69



50-59



40-49



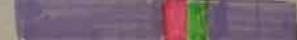
30-39



20-29



10-19



0-9



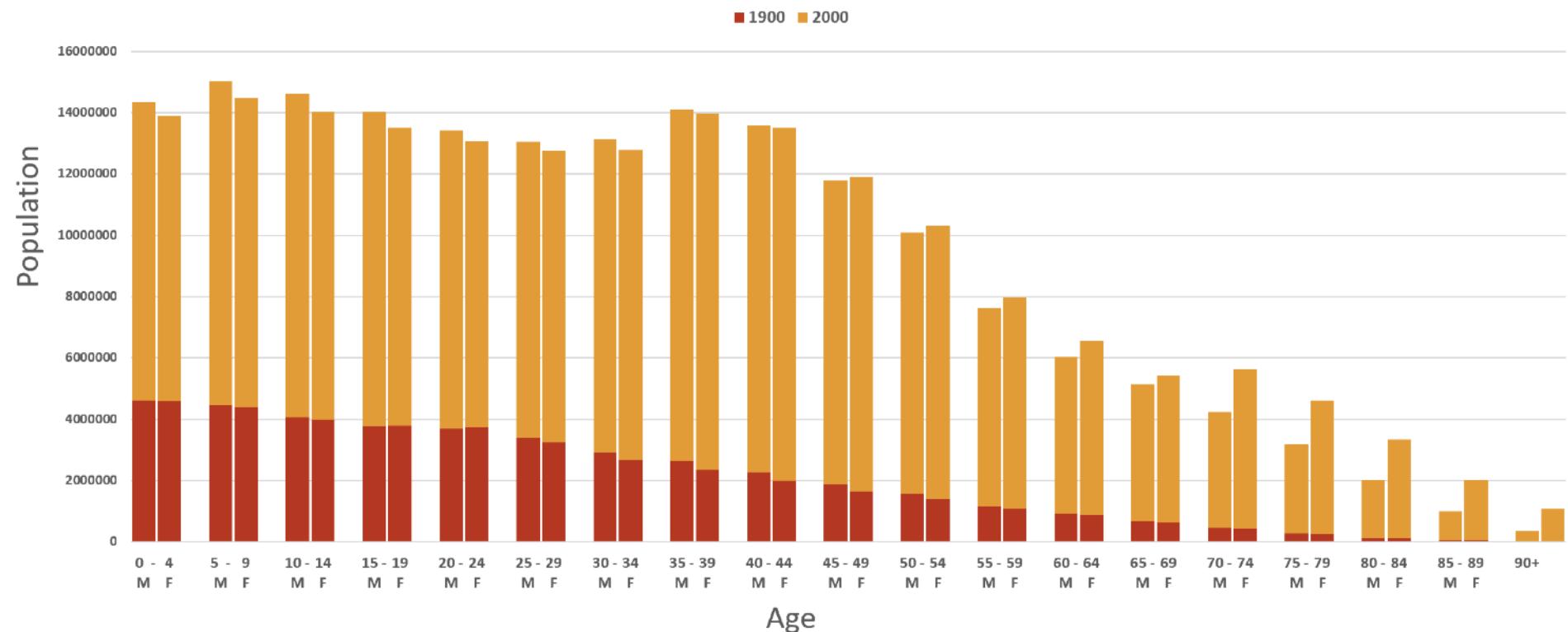
of people (million)

22M 20M 18M 16M 14M 12M 10M 8M 6M 4M 2M 0M

Female Population

Male Population

Comparing Male & Female Population by Age Group between 1900 and 2000



How has age distribution changed in the past century?

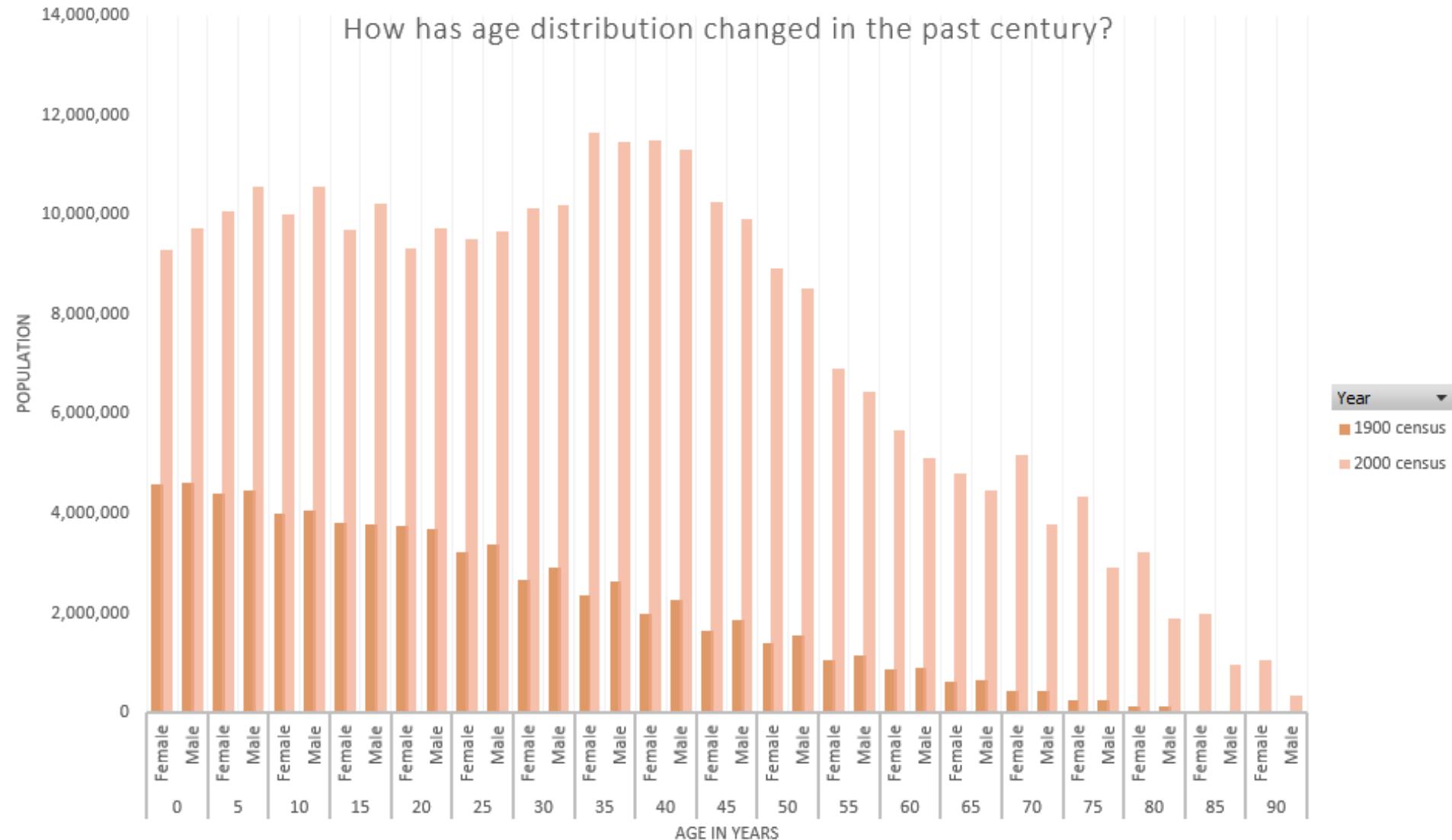


Figure 1

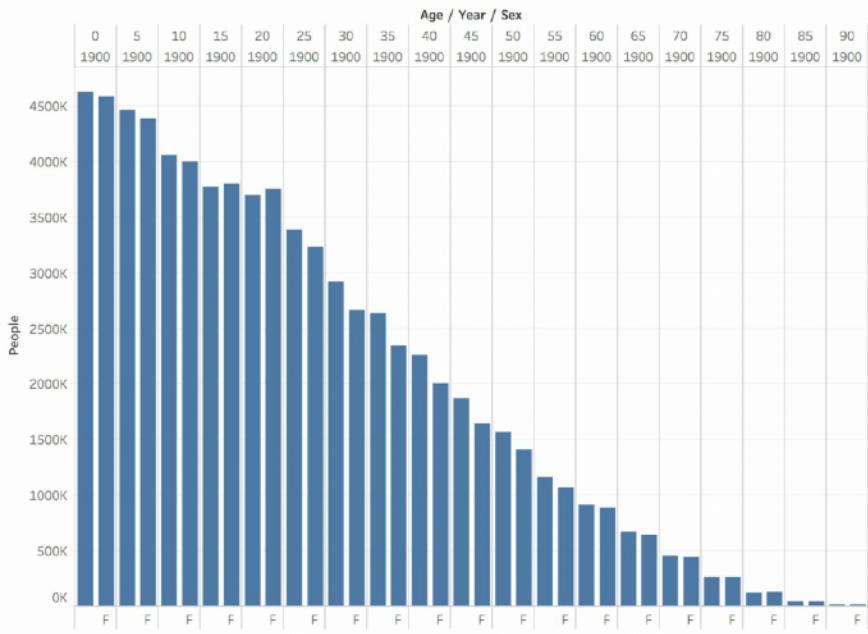


Figure 2

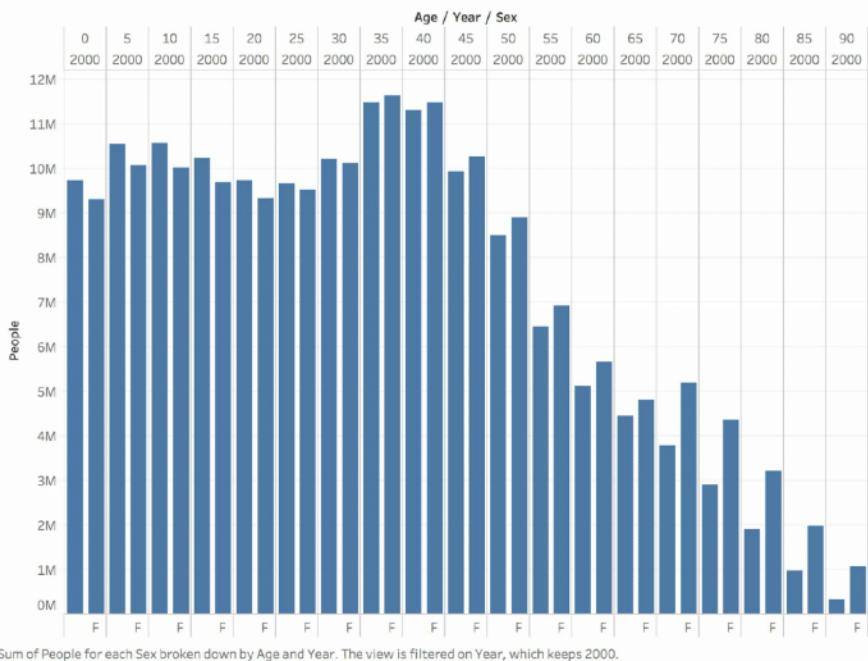
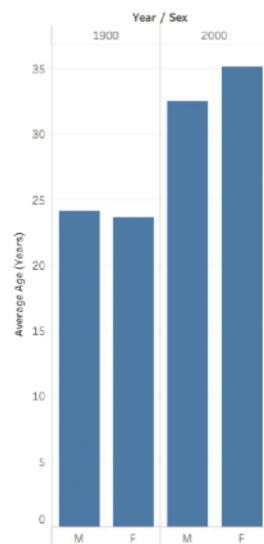
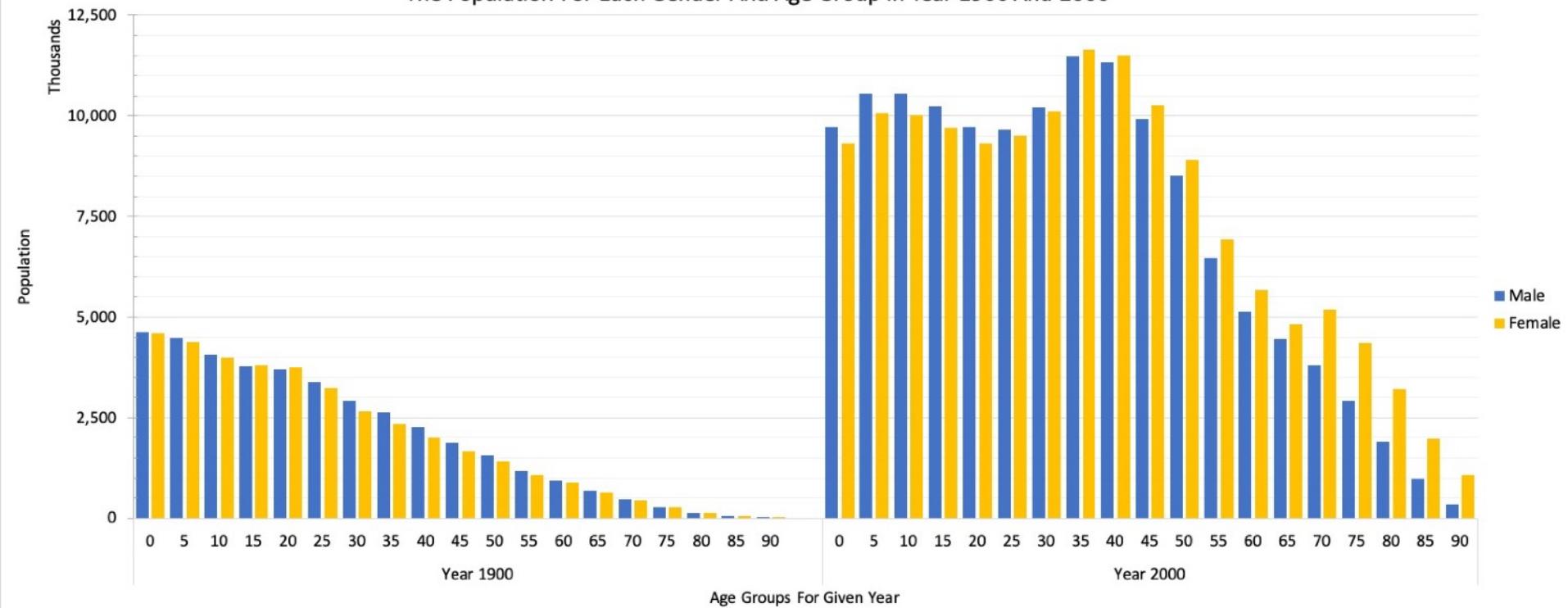


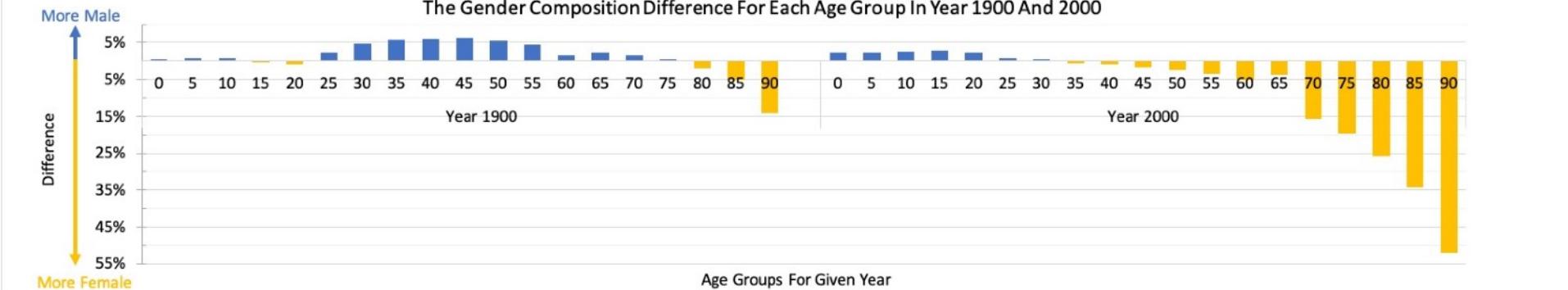
Figure 3



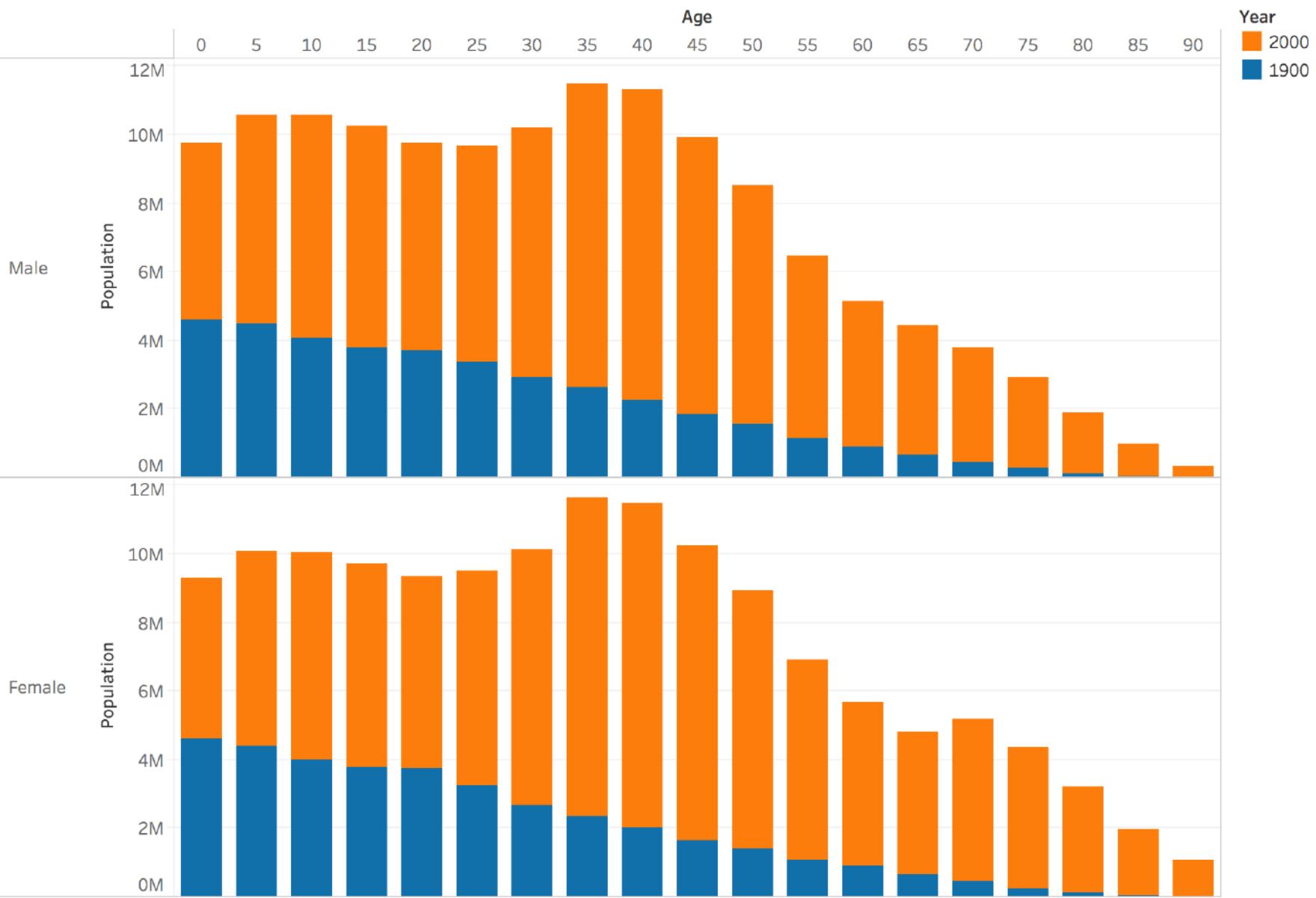
The Population For Each Gender And Age Group In Year 1900 And 2000



The Gender Composition Difference For Each Age Group In Year 1900 And 2000

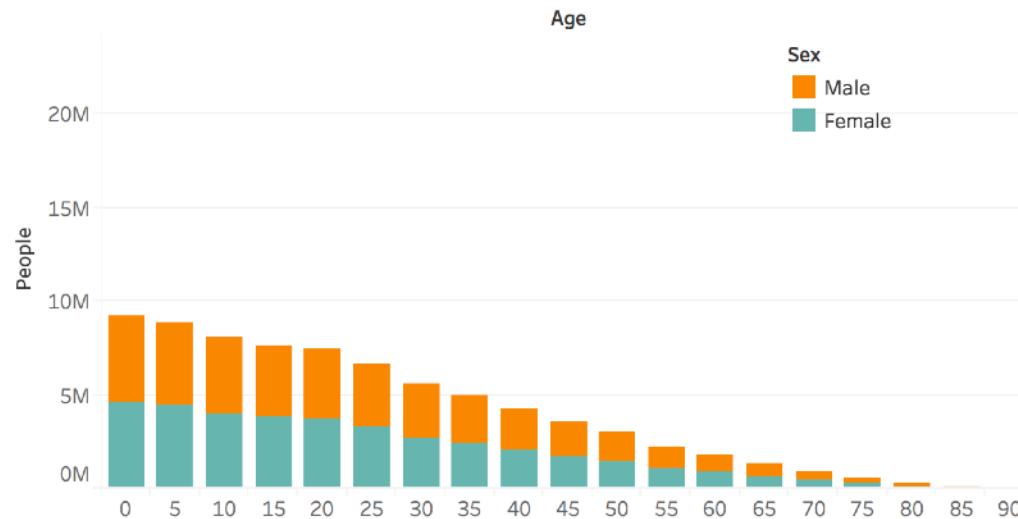


Male and Female Population Growth From 1900 to 2000

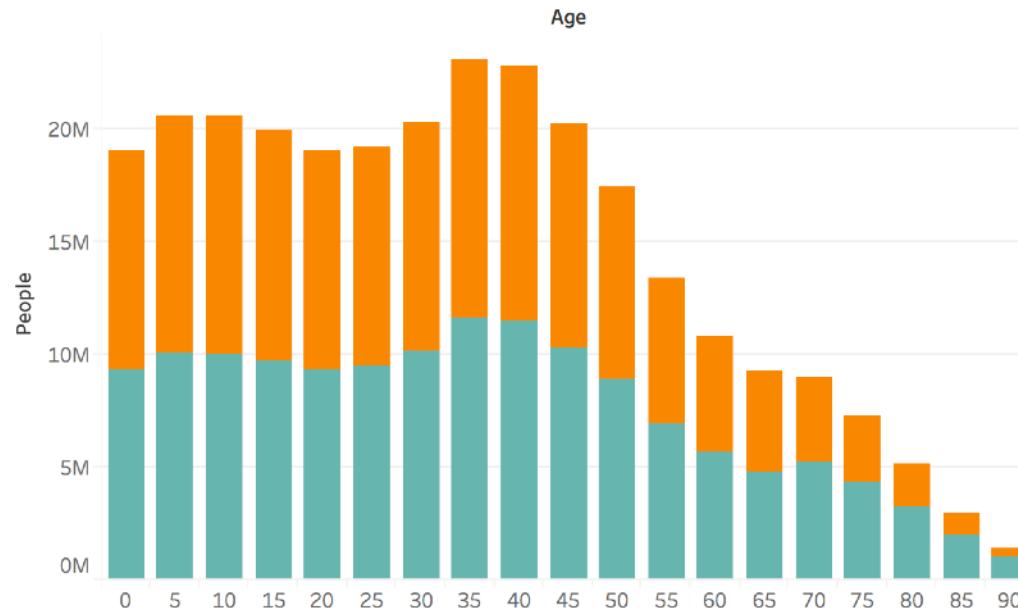


How have US population demographics changed over time?

1900



2000



Population Percentage

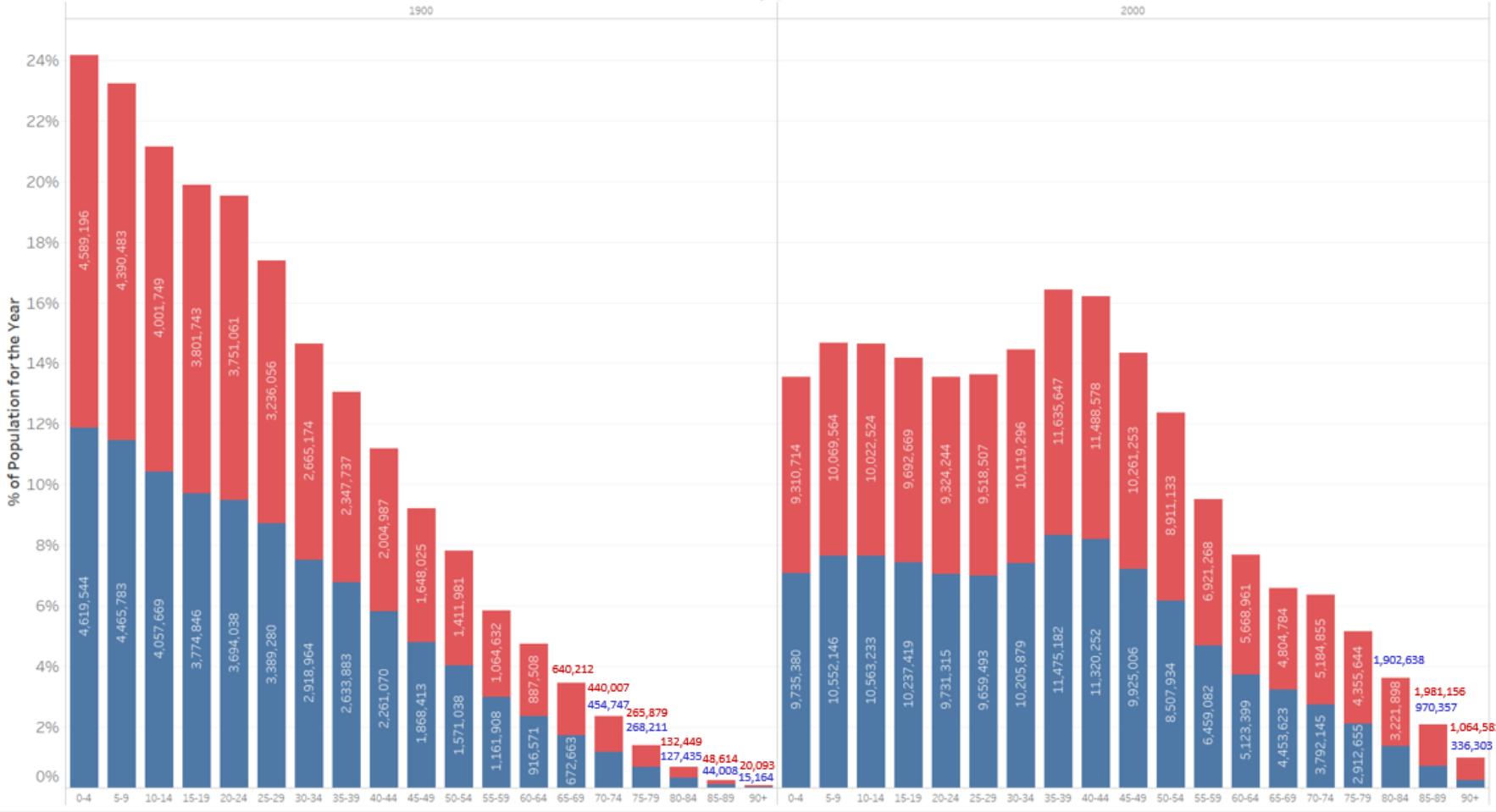
How Have Age and Sex Distributions Changed?

Year / Age Group

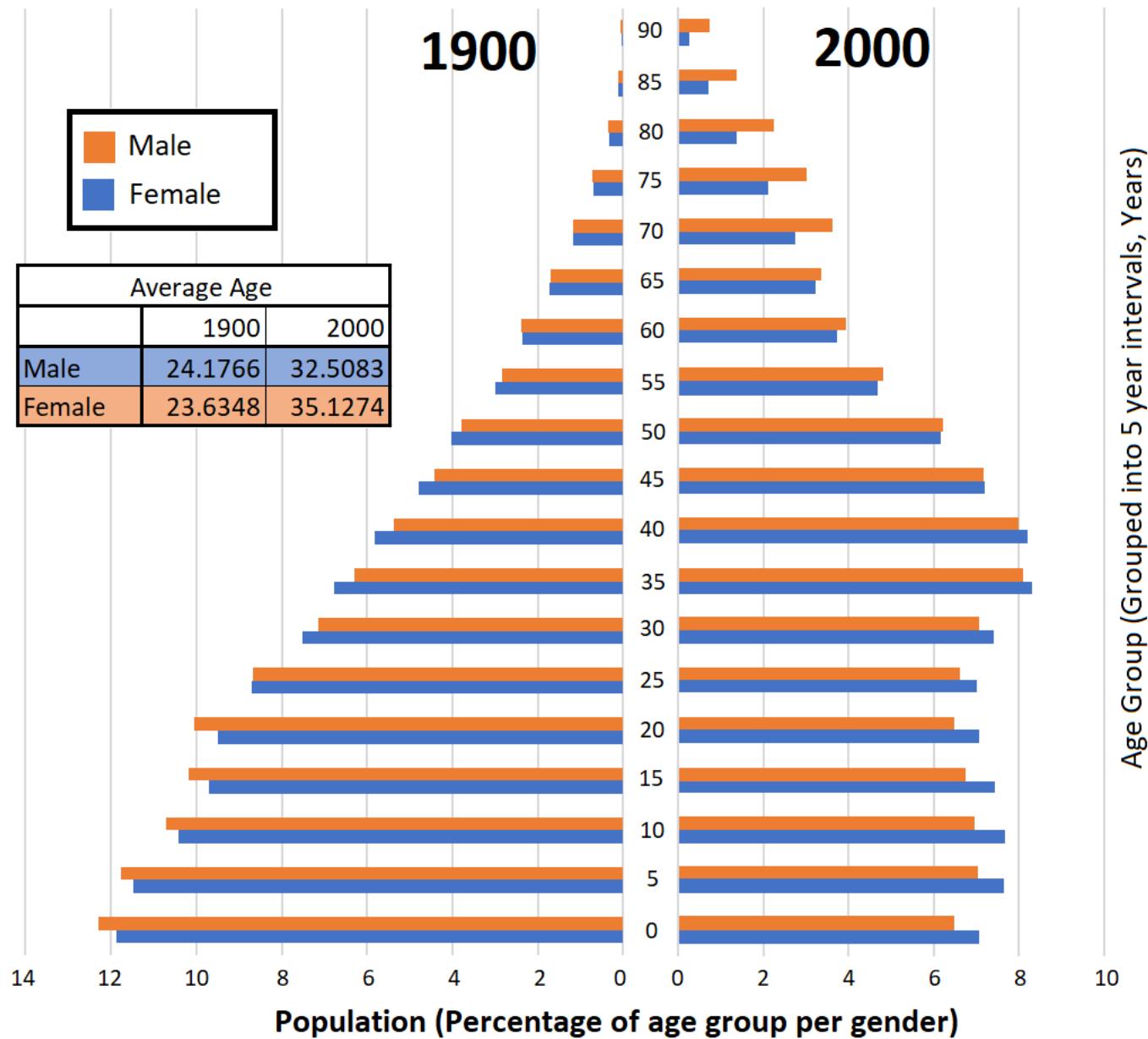
Sex
■ Female
■ Male

1900

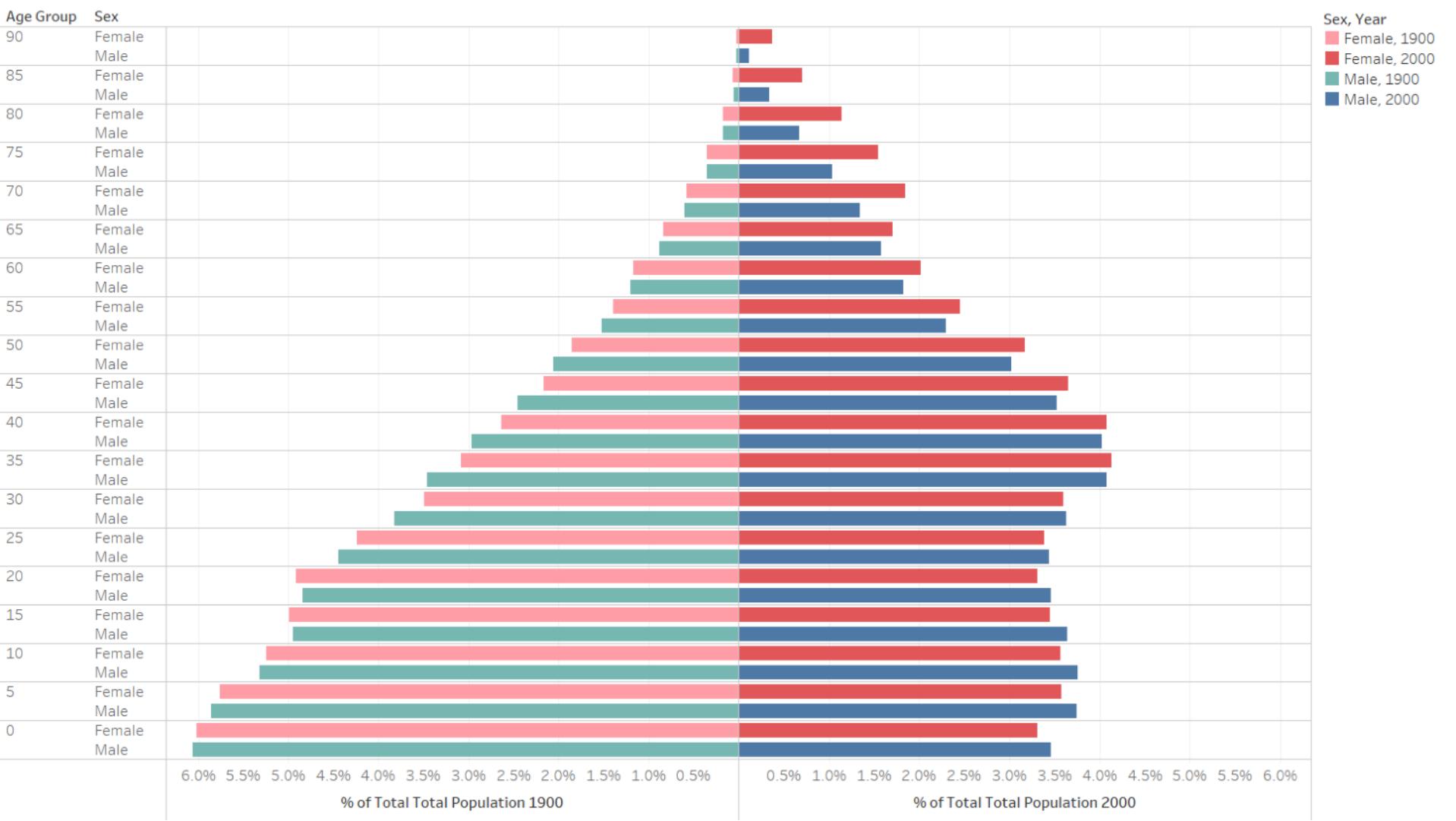
2000



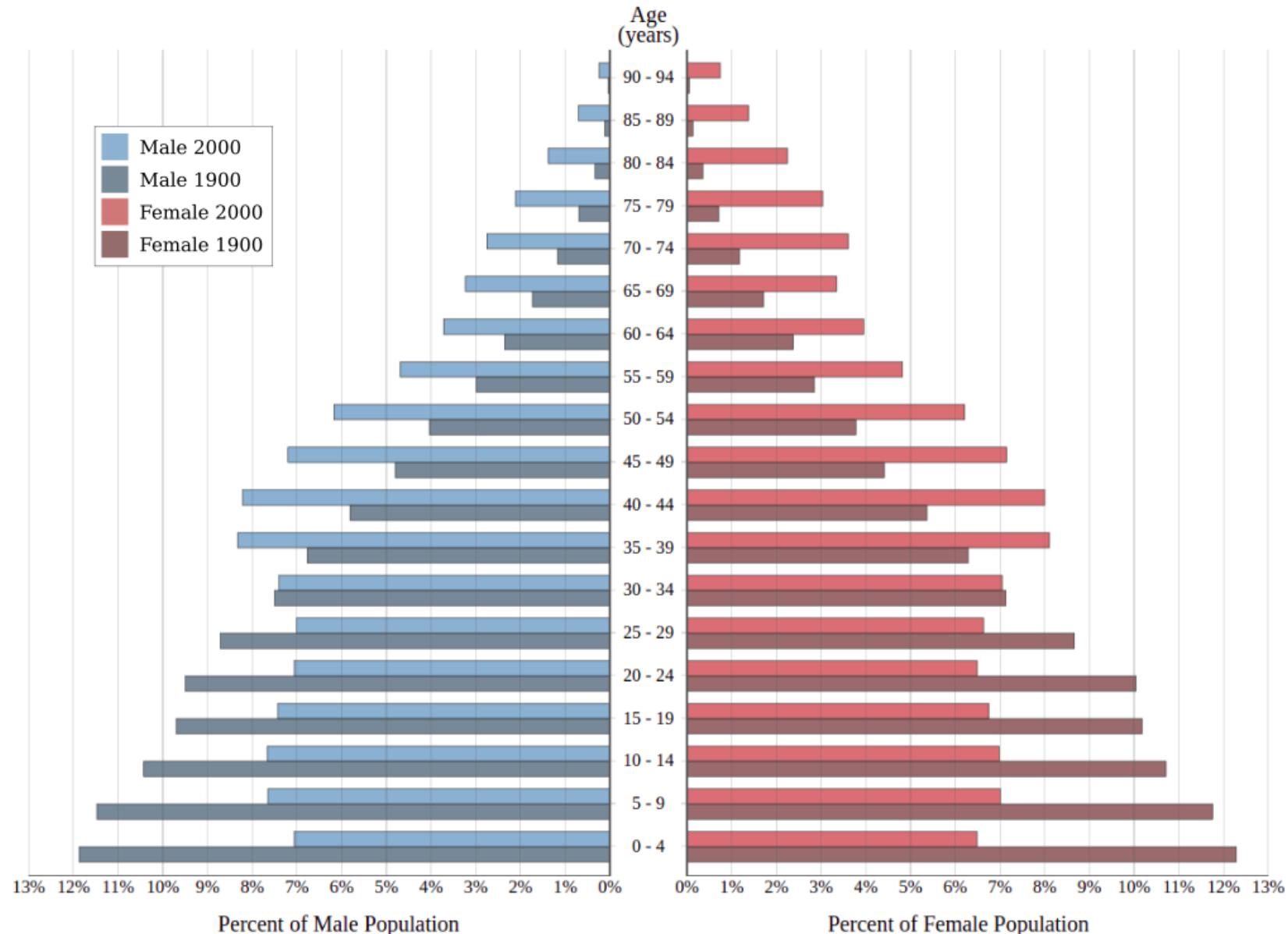
How Has the Distribution of Age shifted between 1900 and 2000?



How has the US Population Distribution across Age and Sex changed throughout the 20th Century?



Change of Age Distribution of Male and Female Sexes from 1900 to 2000 in the US



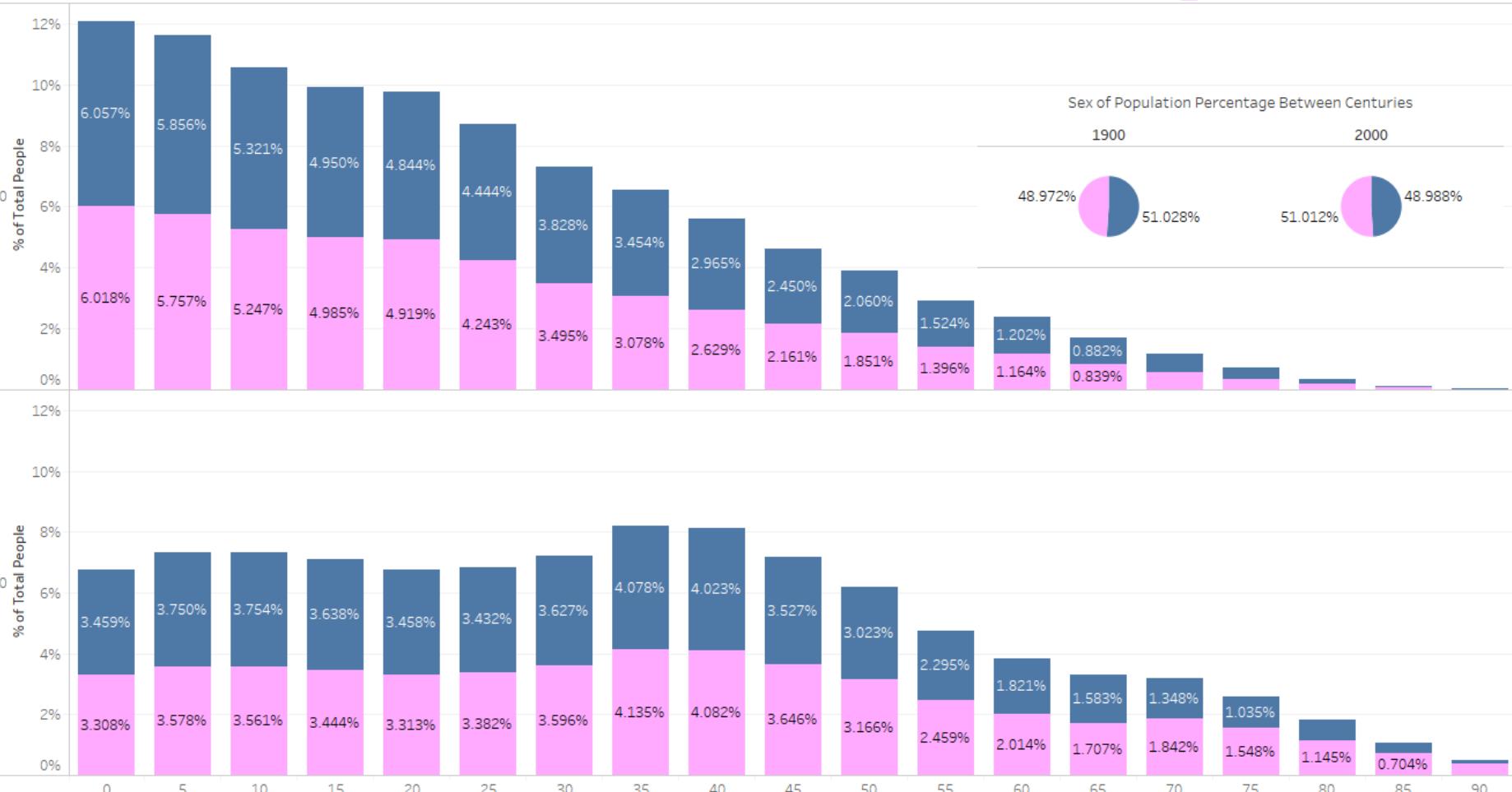
This graphic shows the distributions of age for each sex in each indicated year. Each data series (ex: "Male 2000", etc.) is self-contained, thus each bar in the graphic indicates its respective age group's population as a percentage of the total population of the series.

Age Groups Percentage of Population A Century Apart

Age

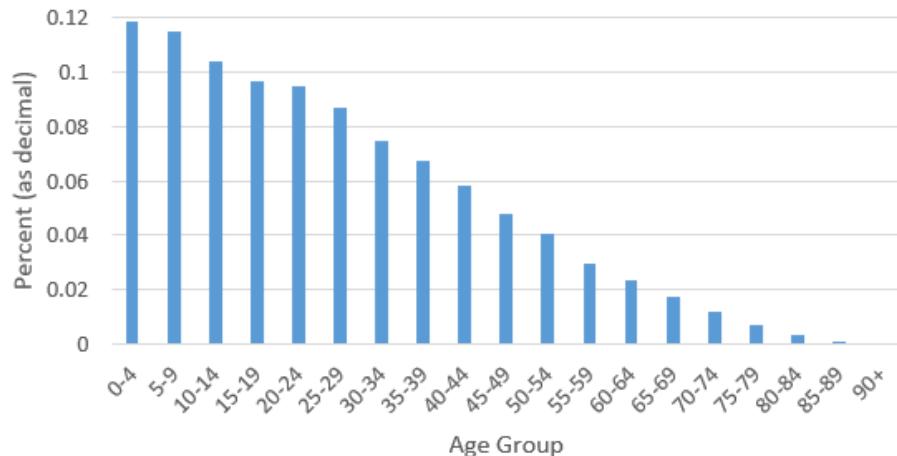
Sex

- Male
- Female

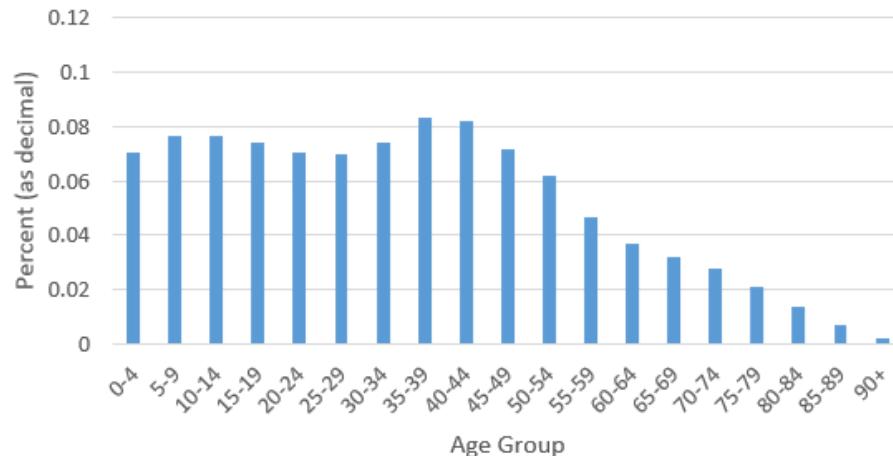


Distribution of US Population by Age Group Over Time

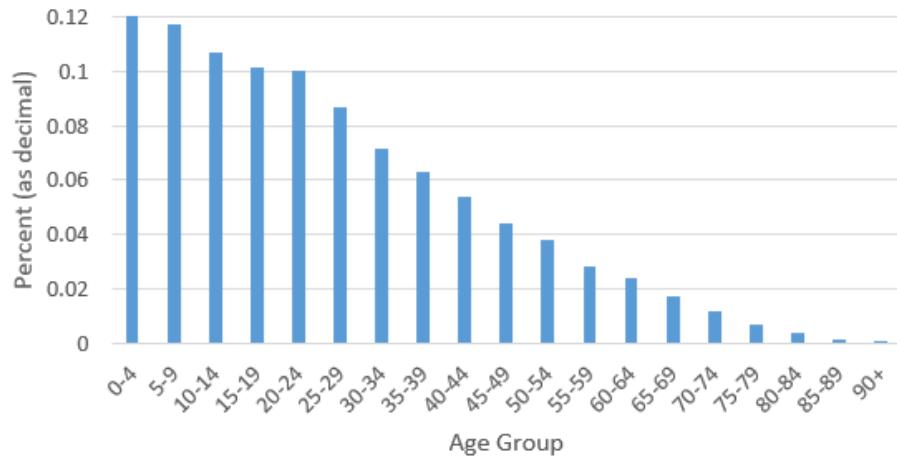
Distribution of Males by Age Group 1900



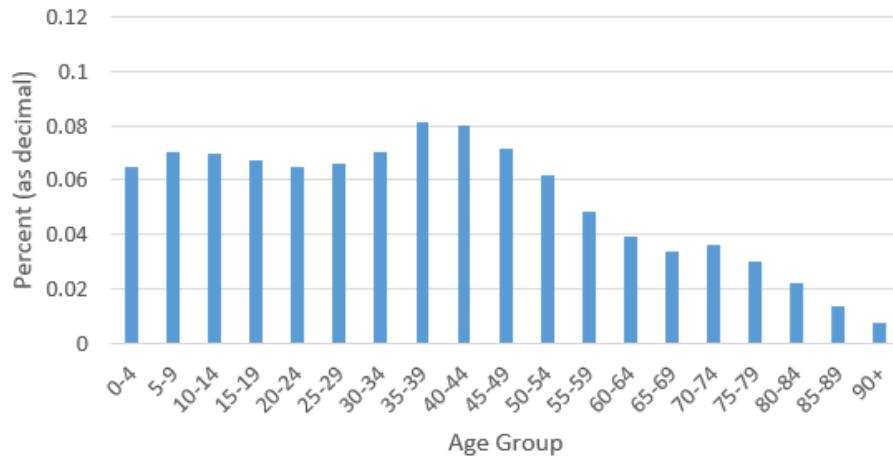
Distribution of Males by Age Group 2000



Distribution of Females by Age Group 1900

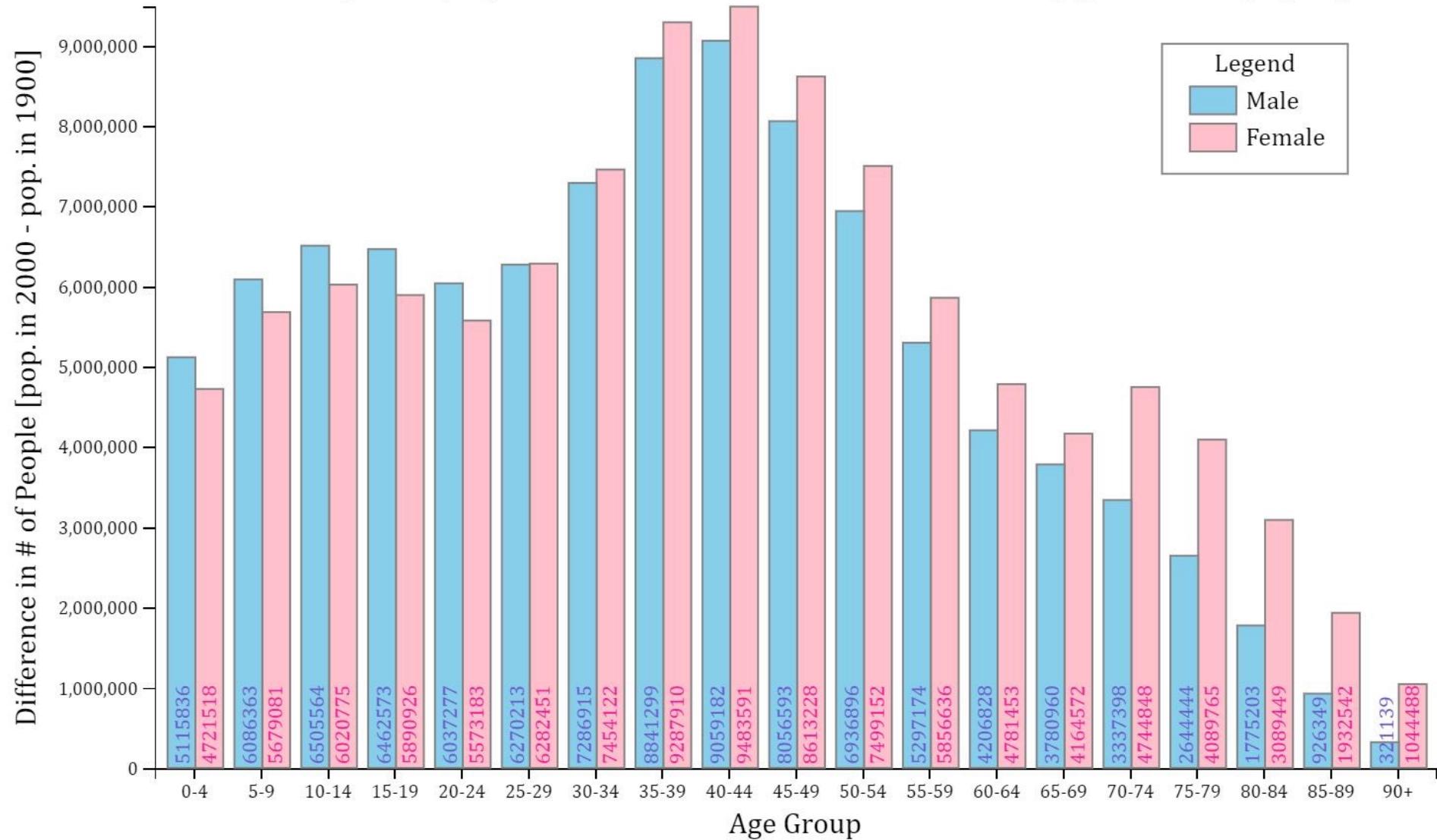


Distribution of Females by Age Group 2000

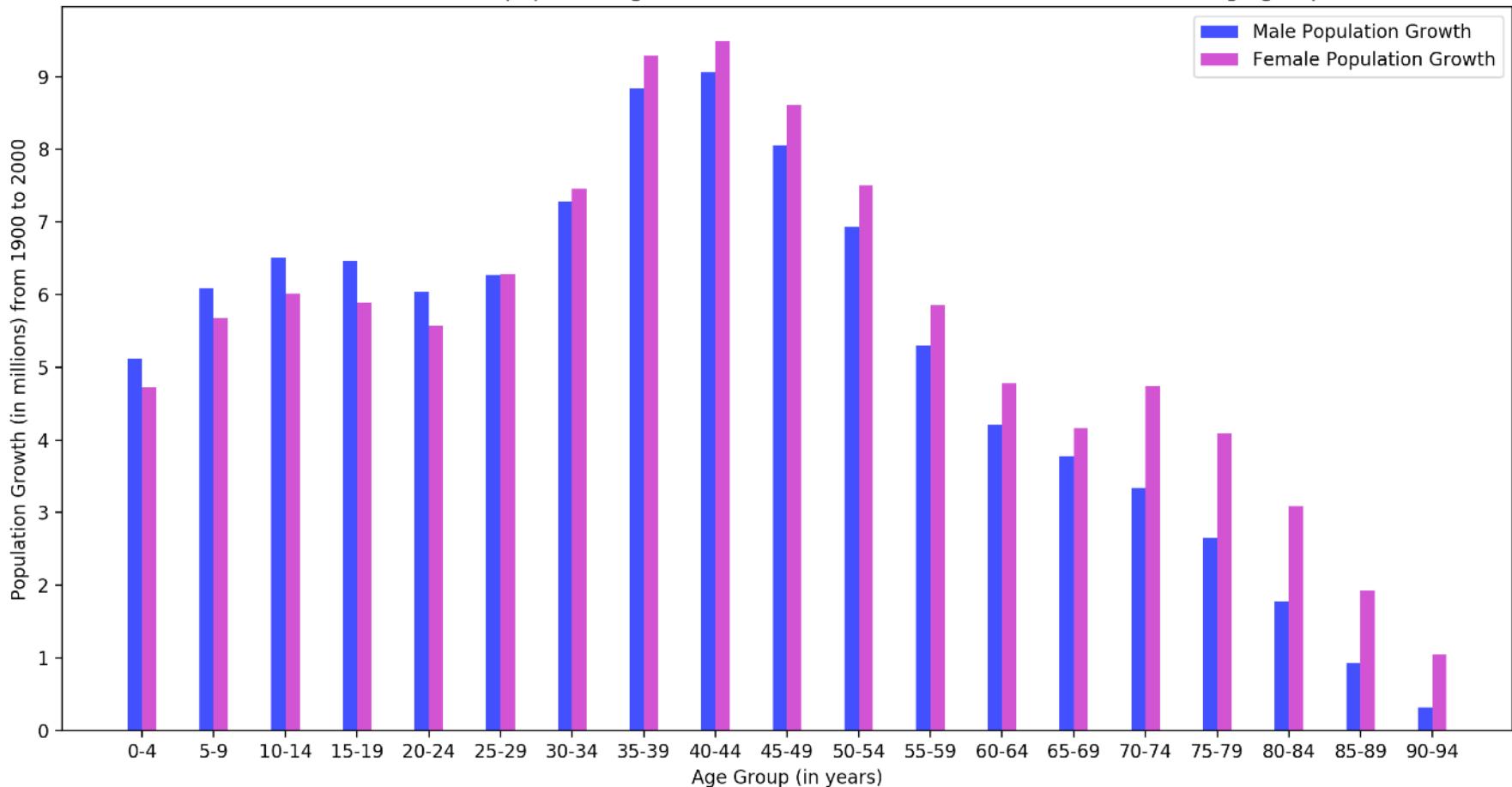


Growth Rates

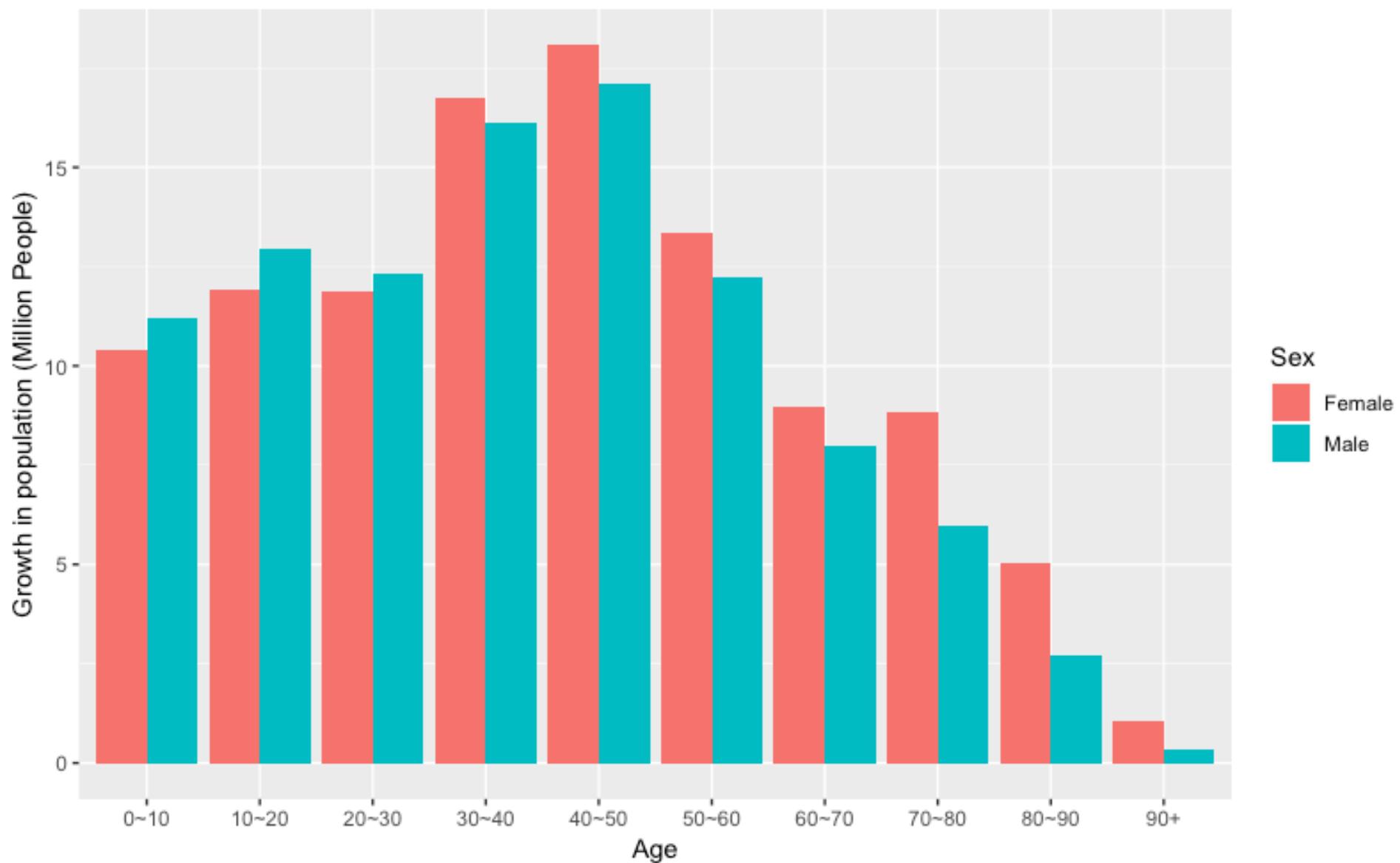
How many more people were alive in 2000 than 1900 in the U.S. by gender and age group?



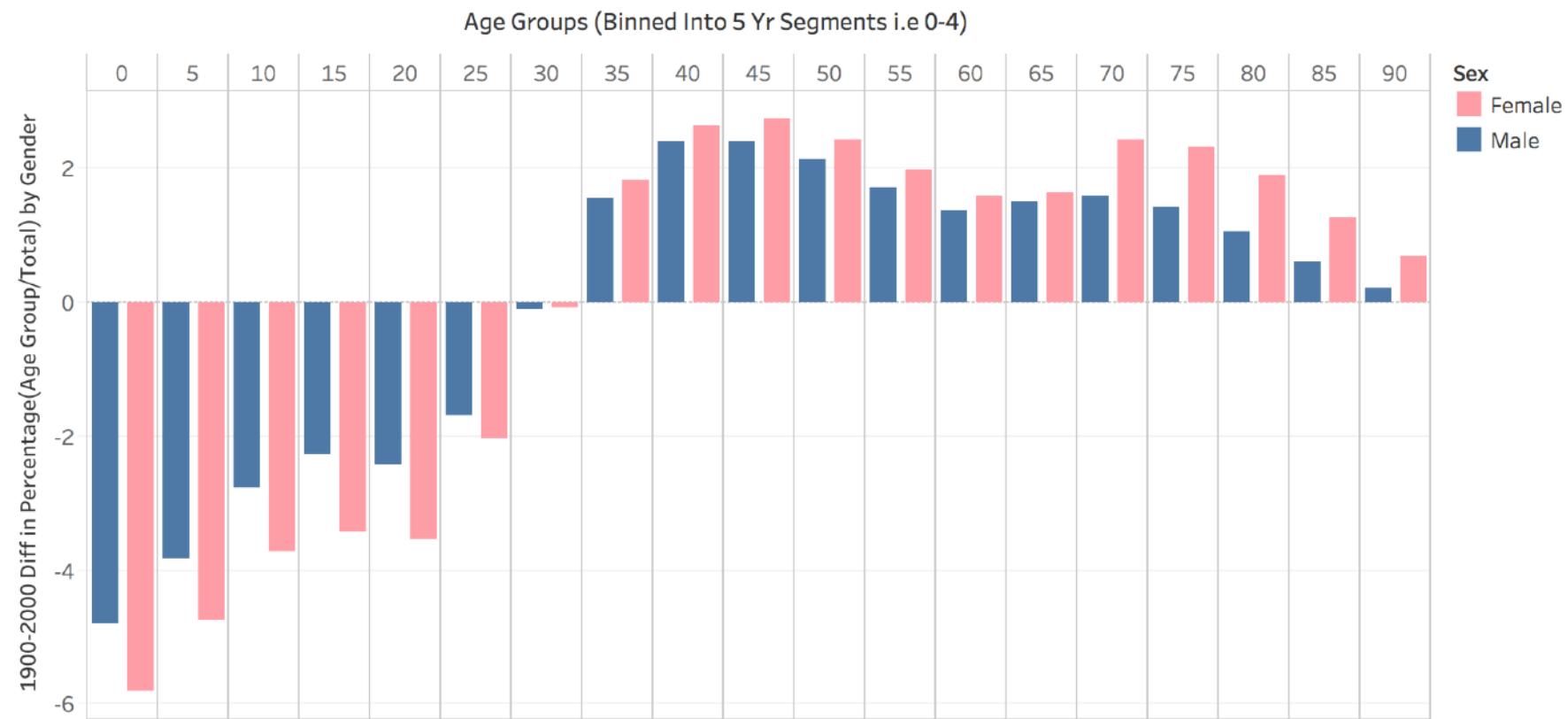
What is the difference in population growth between males and females across different age groups?



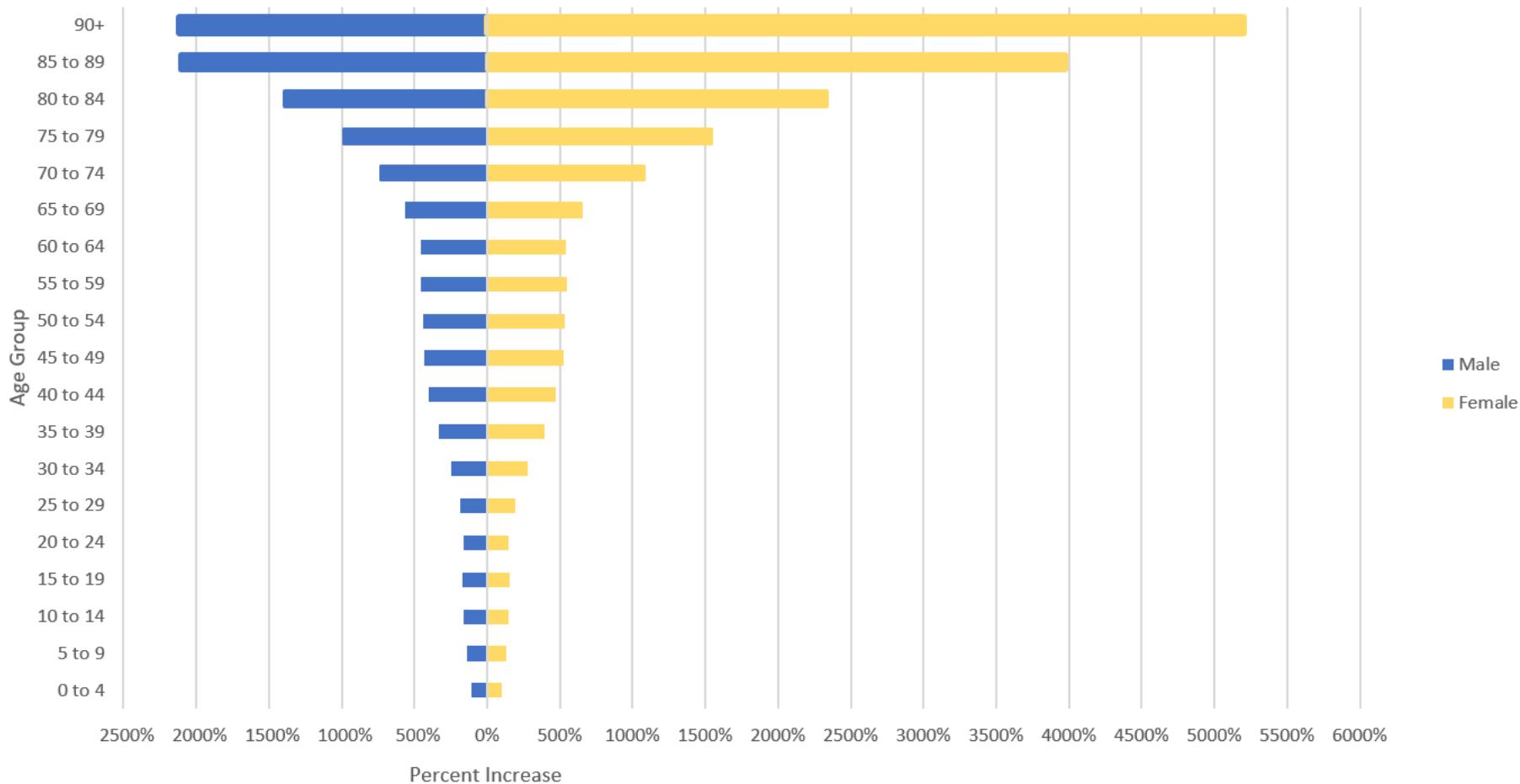
Did population of Male and Female of different age grow equally in 20th century?



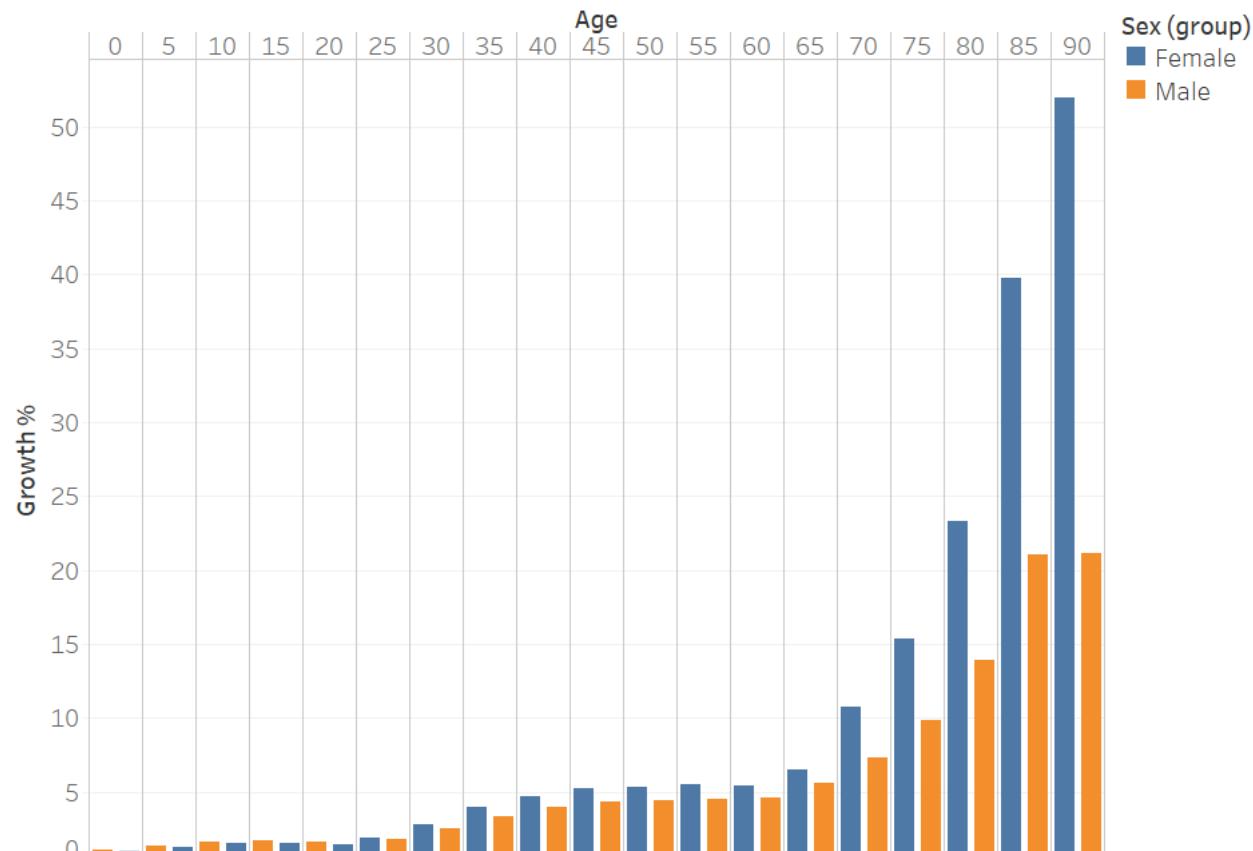
How has the age distribution in the U.S from 1900 to 2000 changed differently for males and females?



Percent Increase in Population by Age and Sex (1900 to 2000)

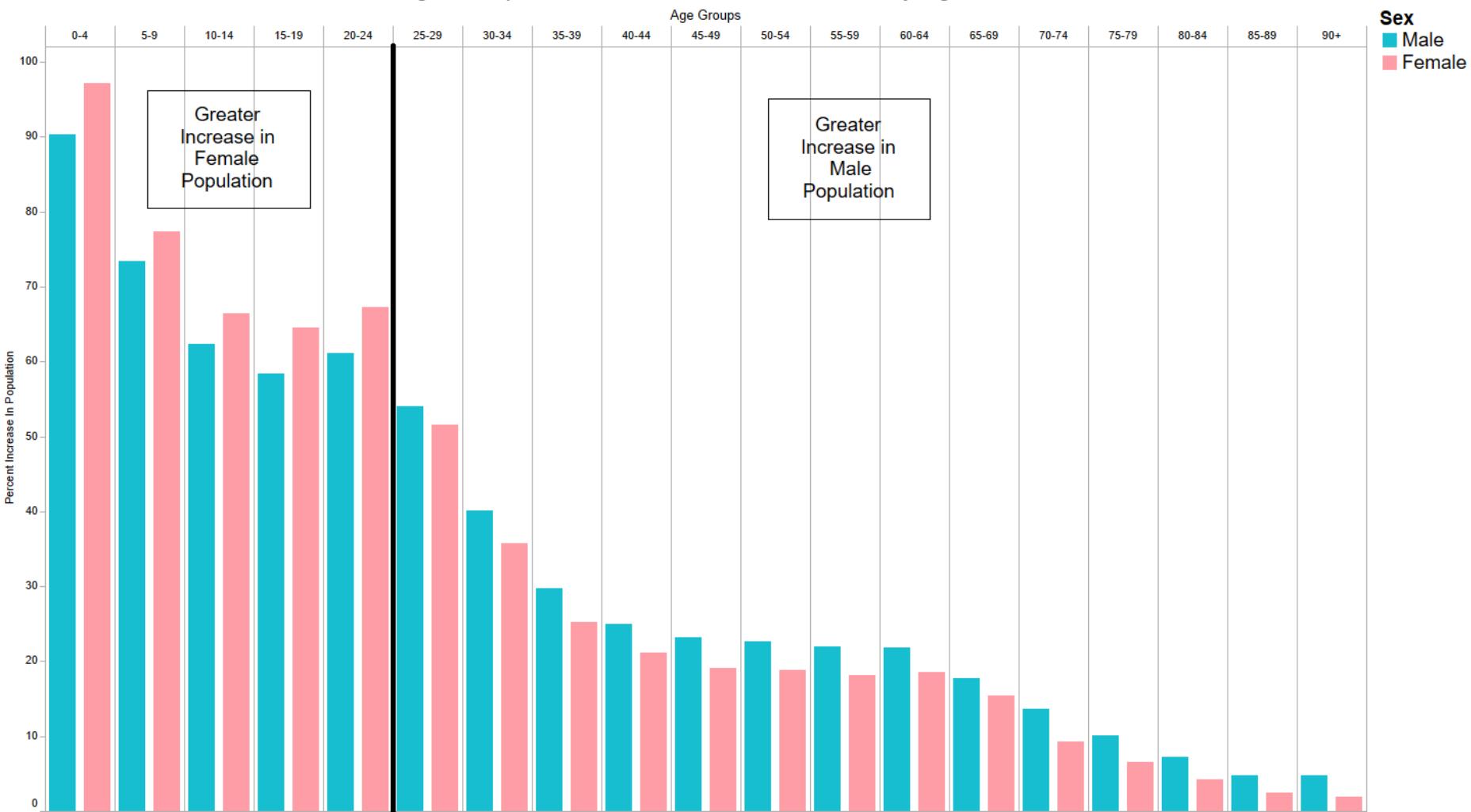


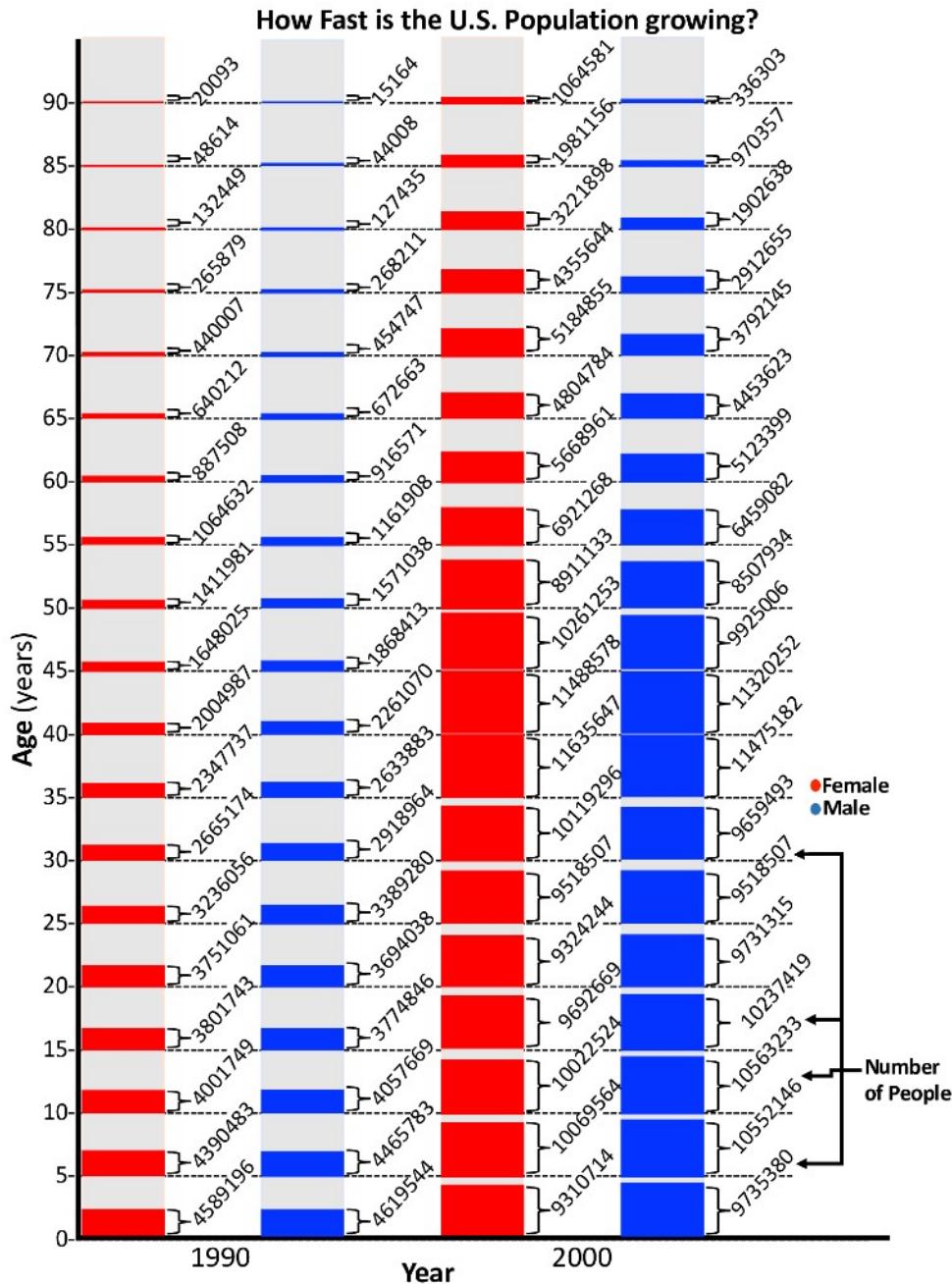
How Have Age and Sex Demographics Changed In the 20th Century?



Growth % for each Sex (group) broken down by Age. Color shows details about Sex (group).

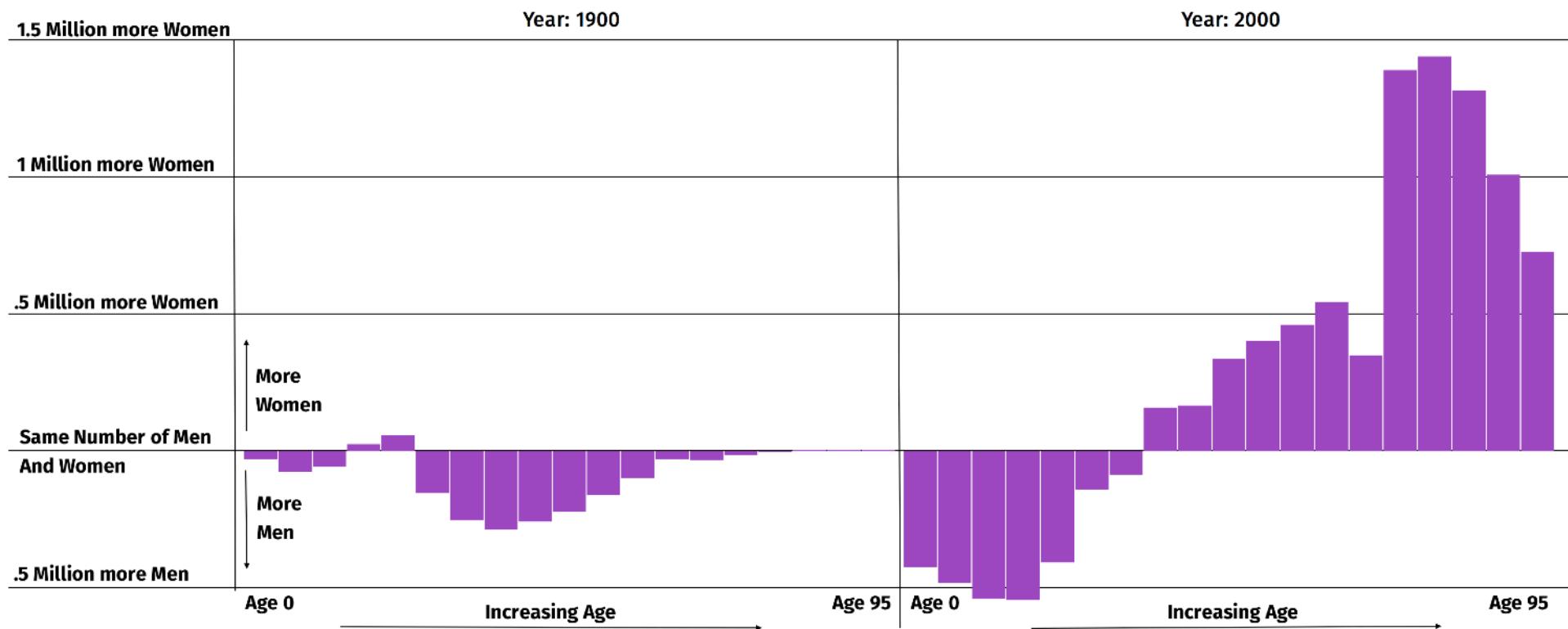
Percent Change in Population of USA from 1900 to 2000 by Age and Gender



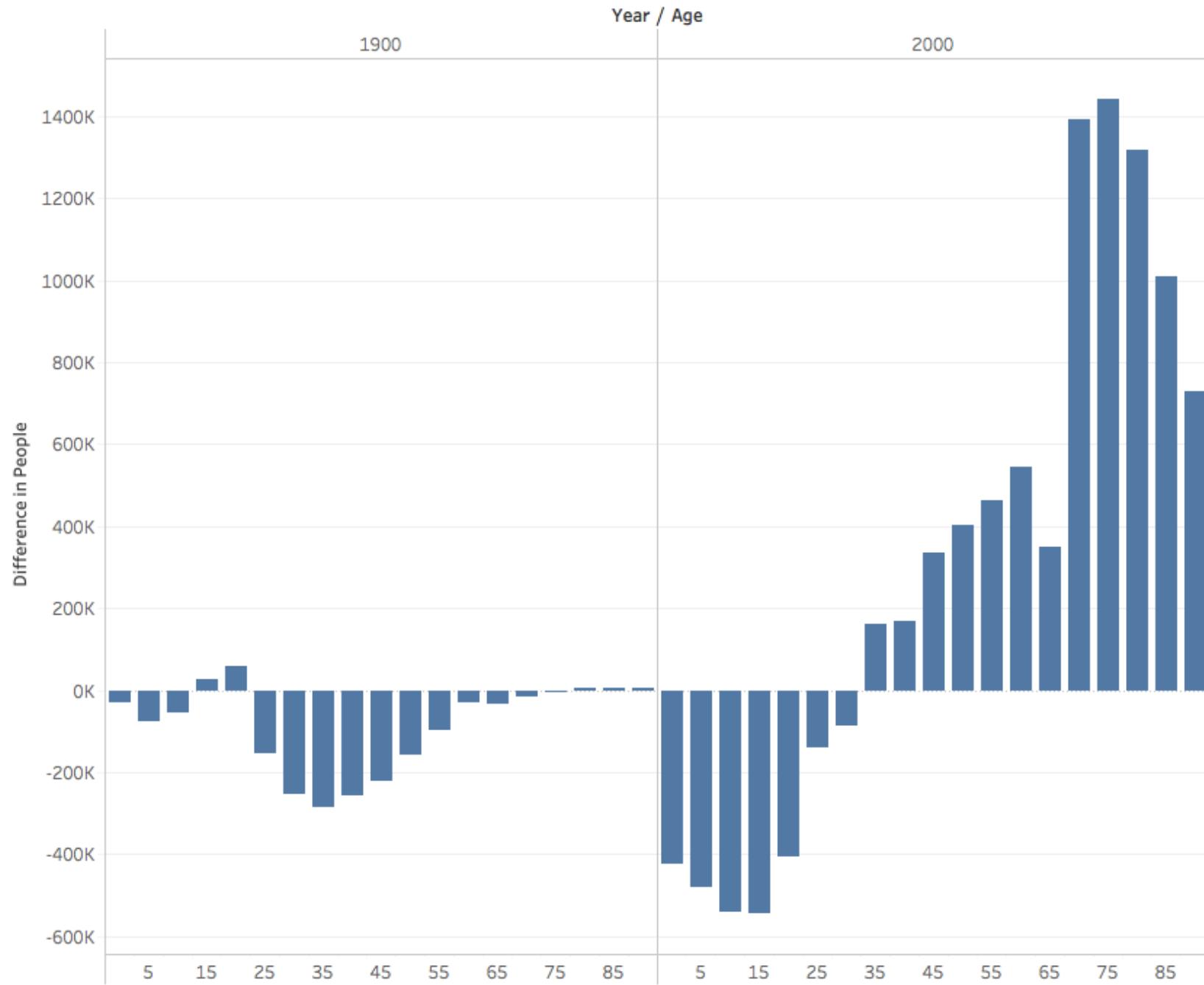


Sex Ratios

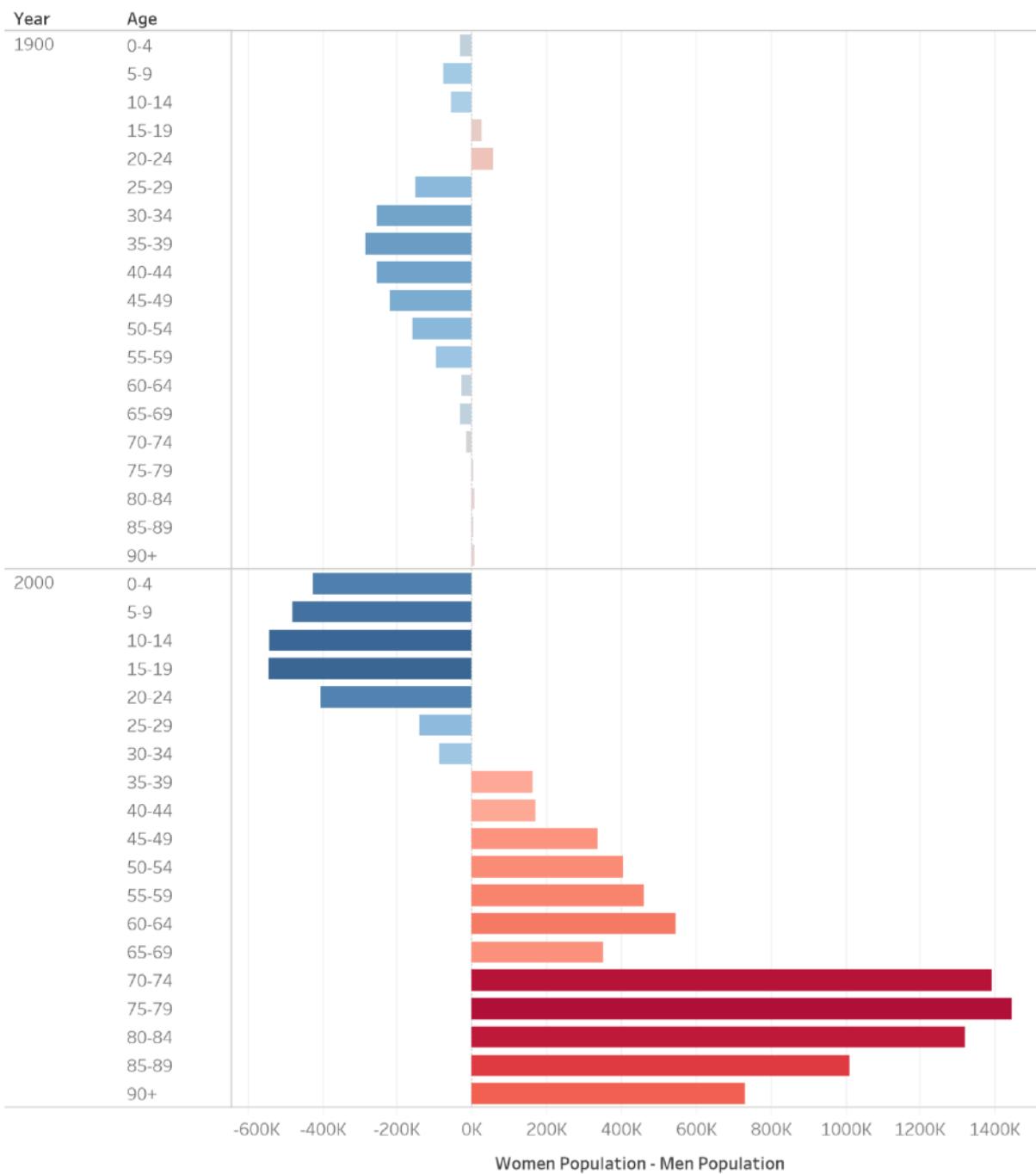
Was there a change in the population balance between men and women from 1900 to 2000?



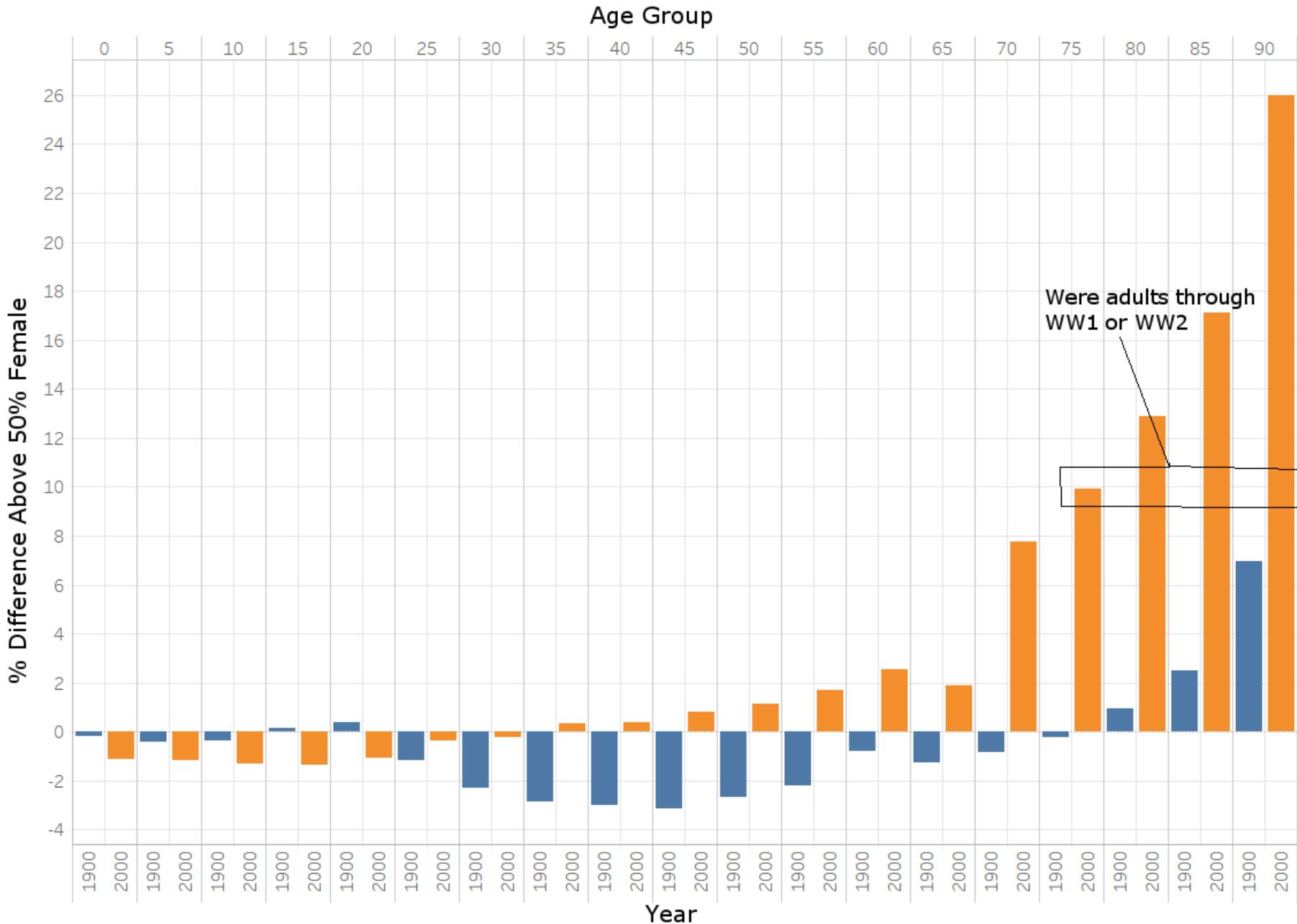
Female surplus in different age groups in year 1900 and 2000



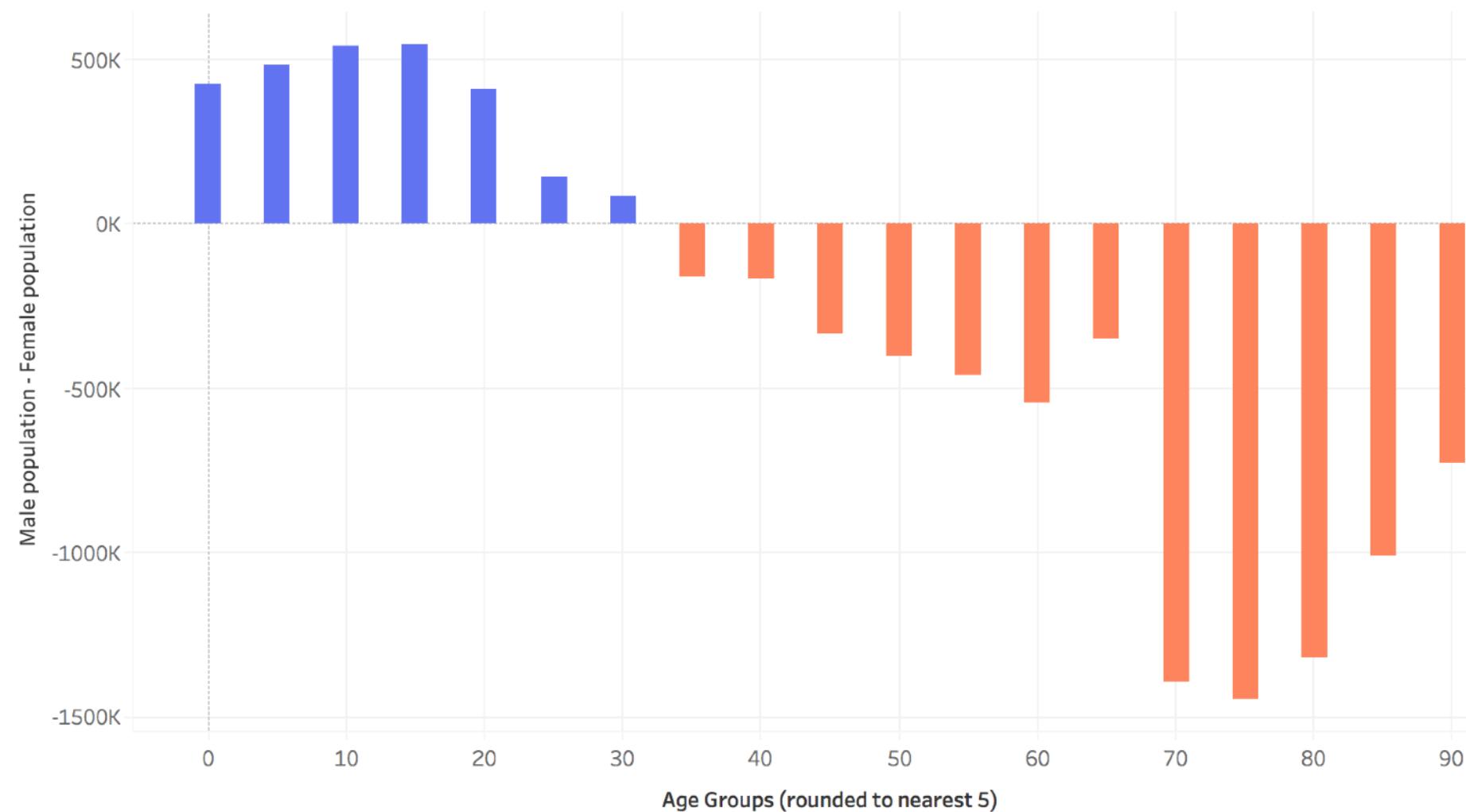
How do the differences in population of women and men change over time?



Sex Ratio per Age Group in 1900 and 2000

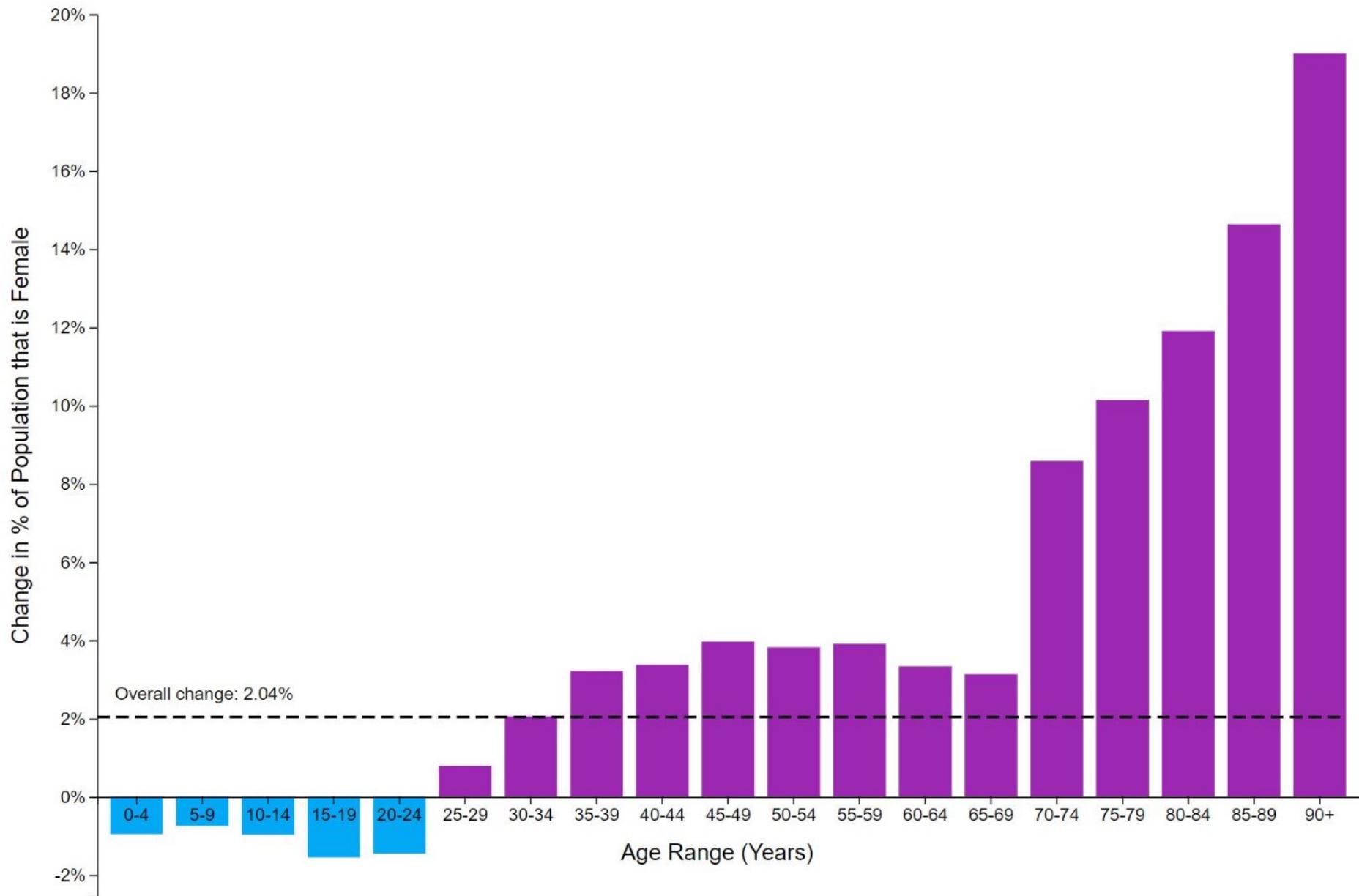


What is the population difference between genders trend across age groups in 2000?



The plot of sum of Difference for Age. Color shows details about Difference.

How has the balance between women and men in the US changed over the past 100 years?



Is there a relationship between age groups and the percentage of female population compared to male population?

Year

Age

% compared to Male Population

200%

150%

100%

50%

0%

1900

% compared to Male Population

200%

150%

100%

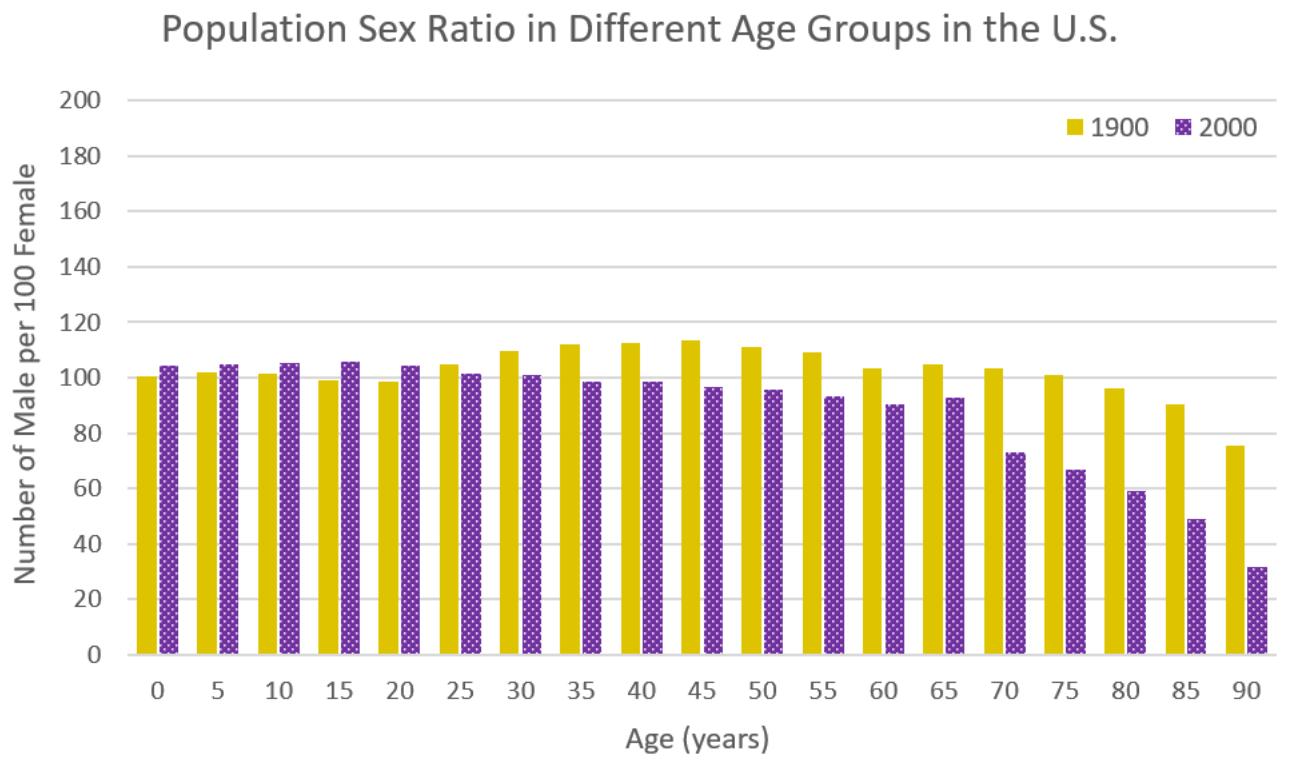
50%

0%

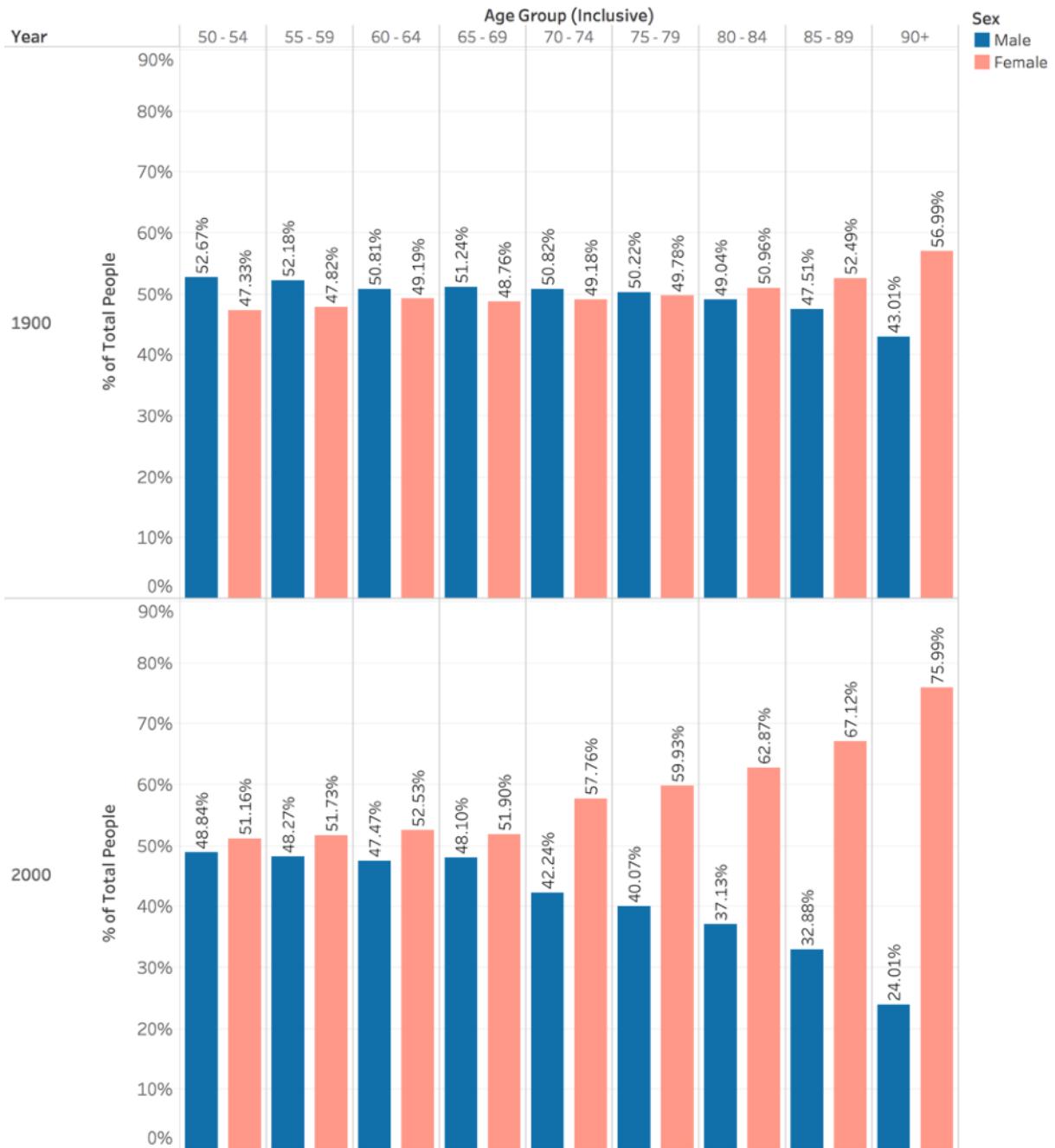
2000

0-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90+



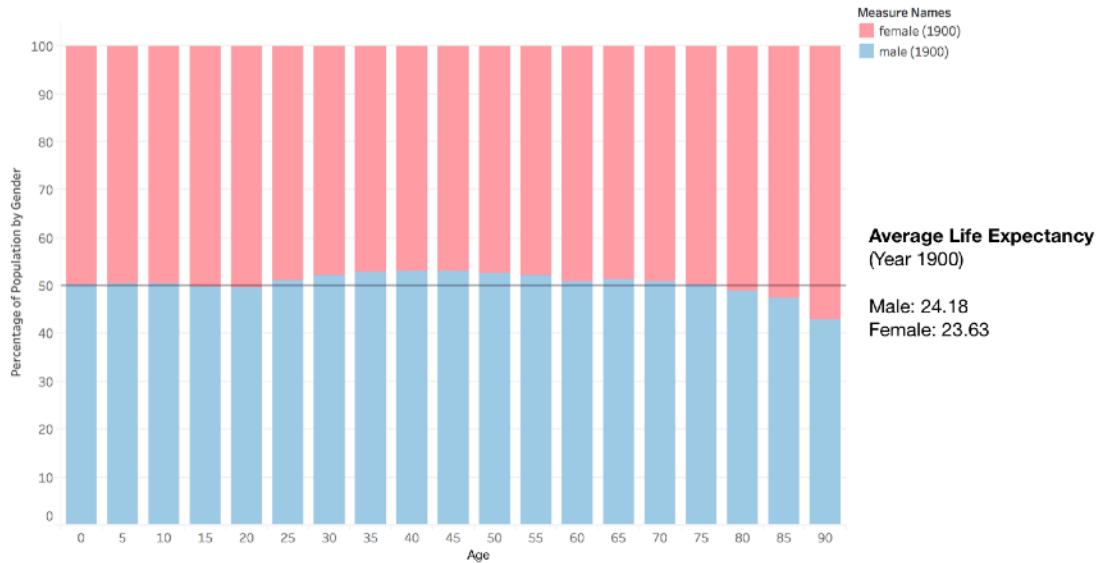


How did male-to-female ratios change between years 1900 and 2000 for USA persons over 50 years of age?

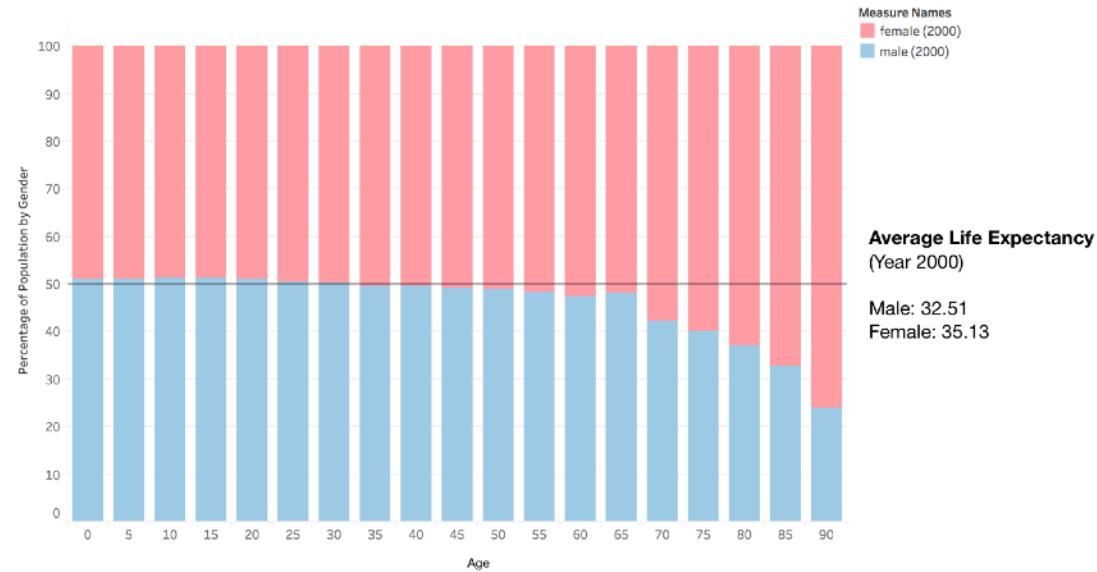


How has the life expectancy of men and women changed between 1900 and 2000?

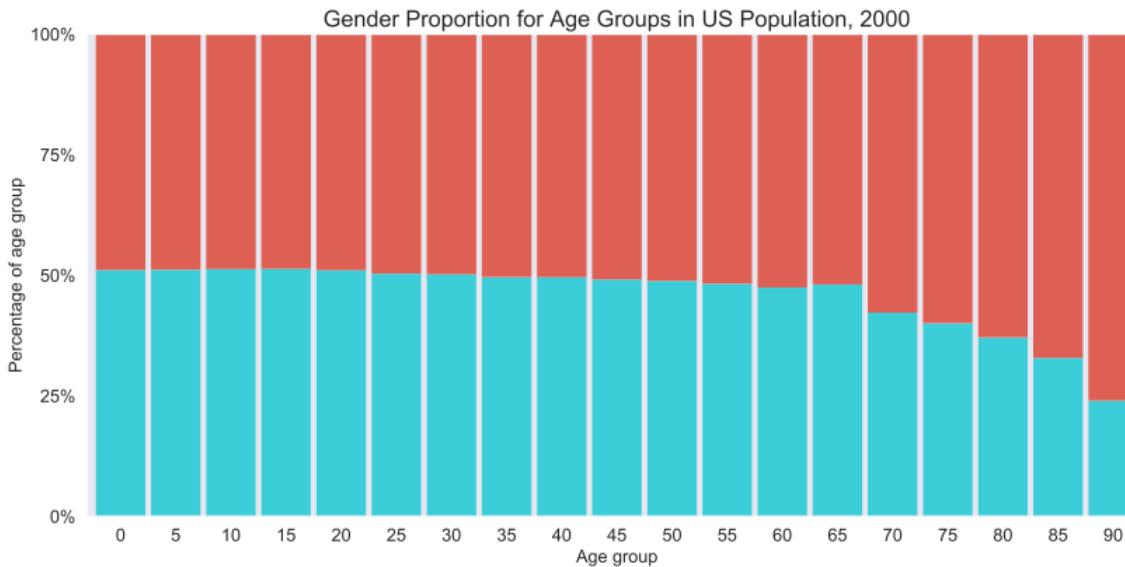
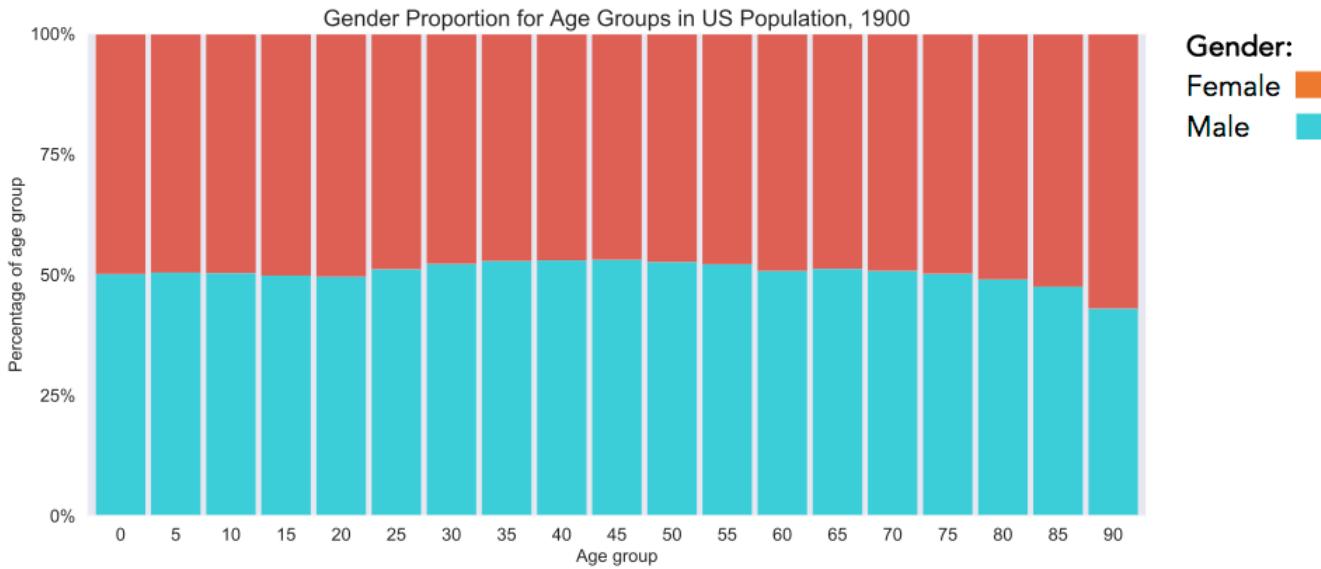
Population Ratio by Age (Year 1900)



Population Ratio by Age (Year 2000)

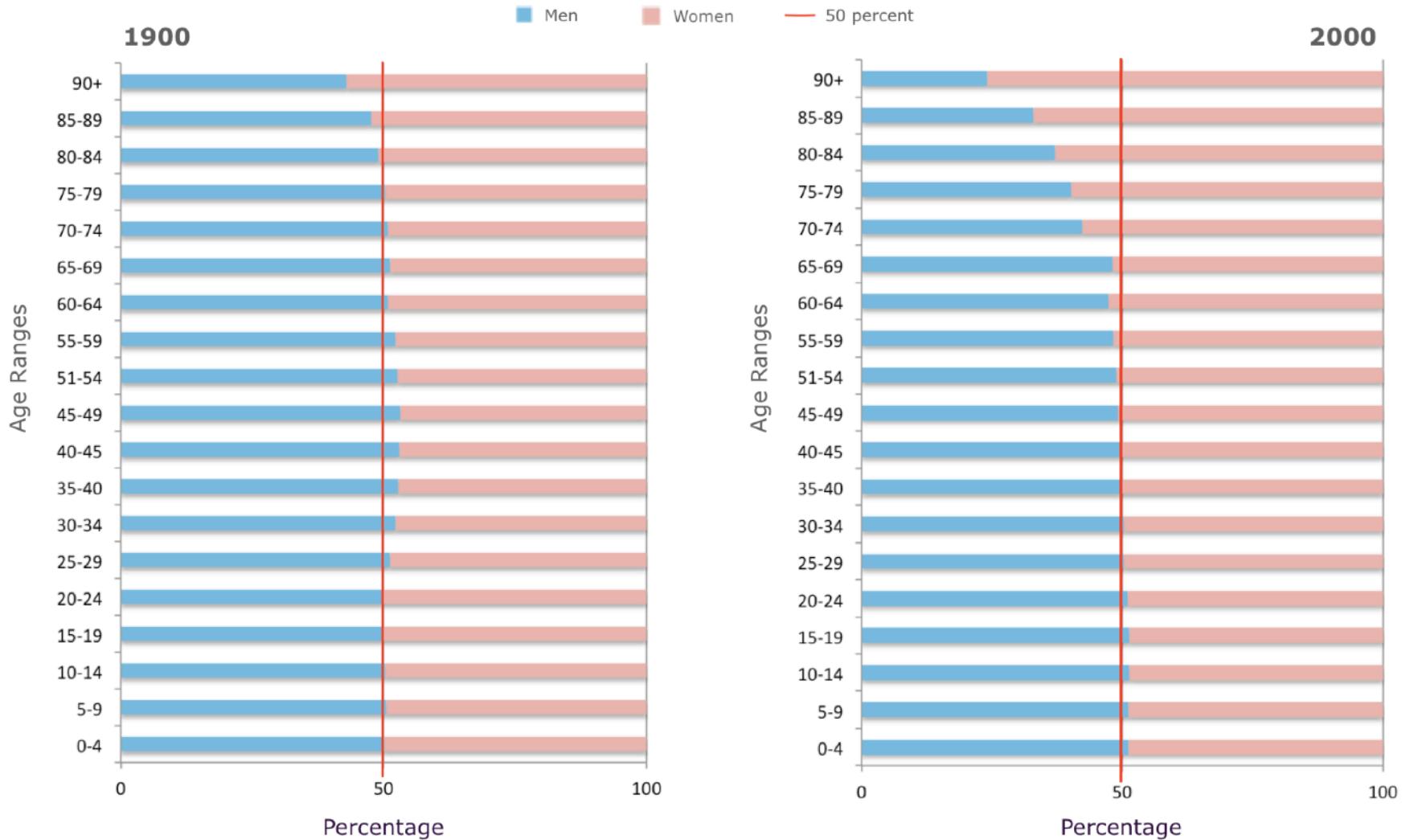


How Has The US Population Gender Breakdown Changed Between Two Centuries?

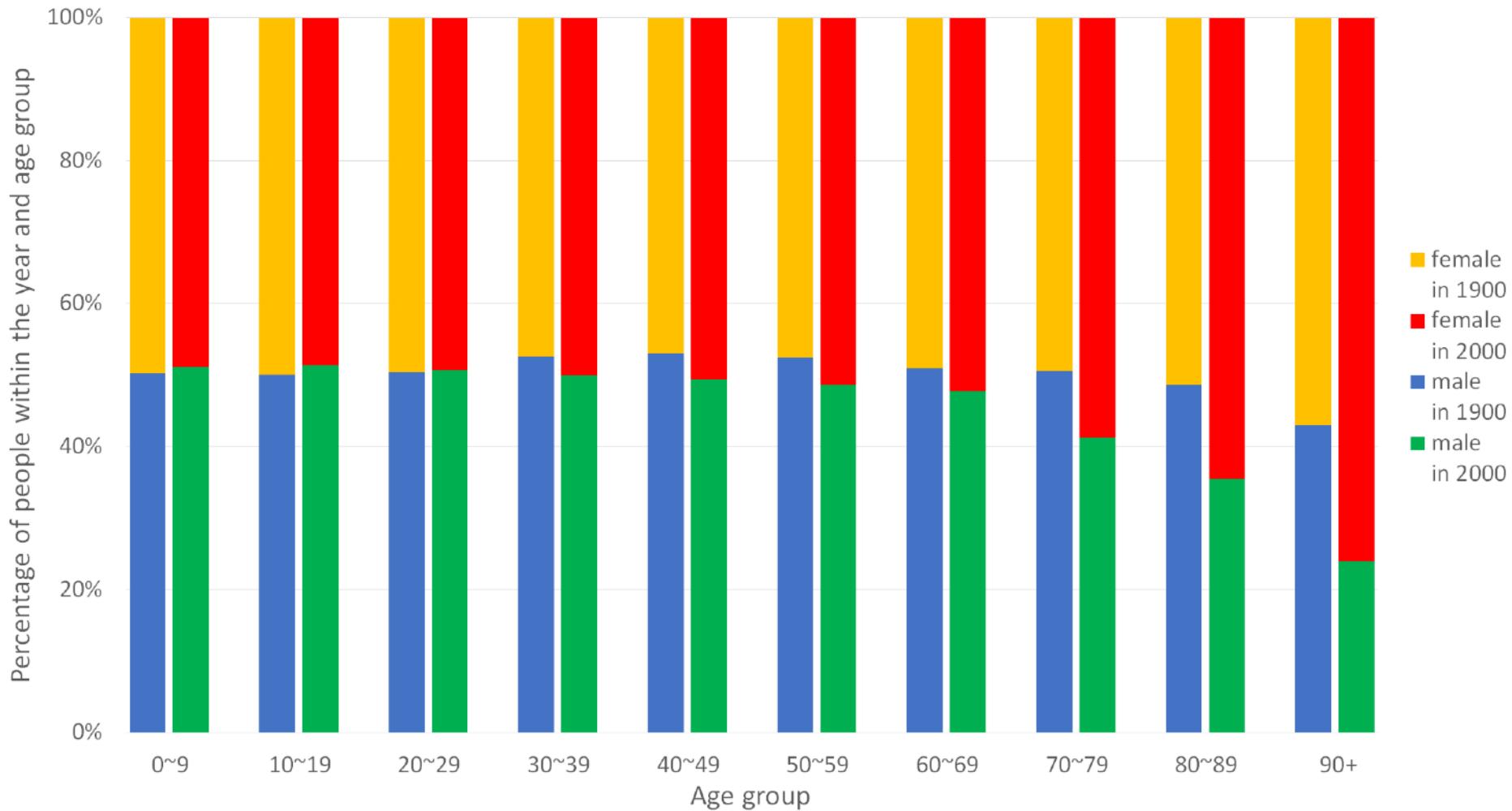


Is it really 50/50?

A look at the ratio of men to women in 1900 and 2000.



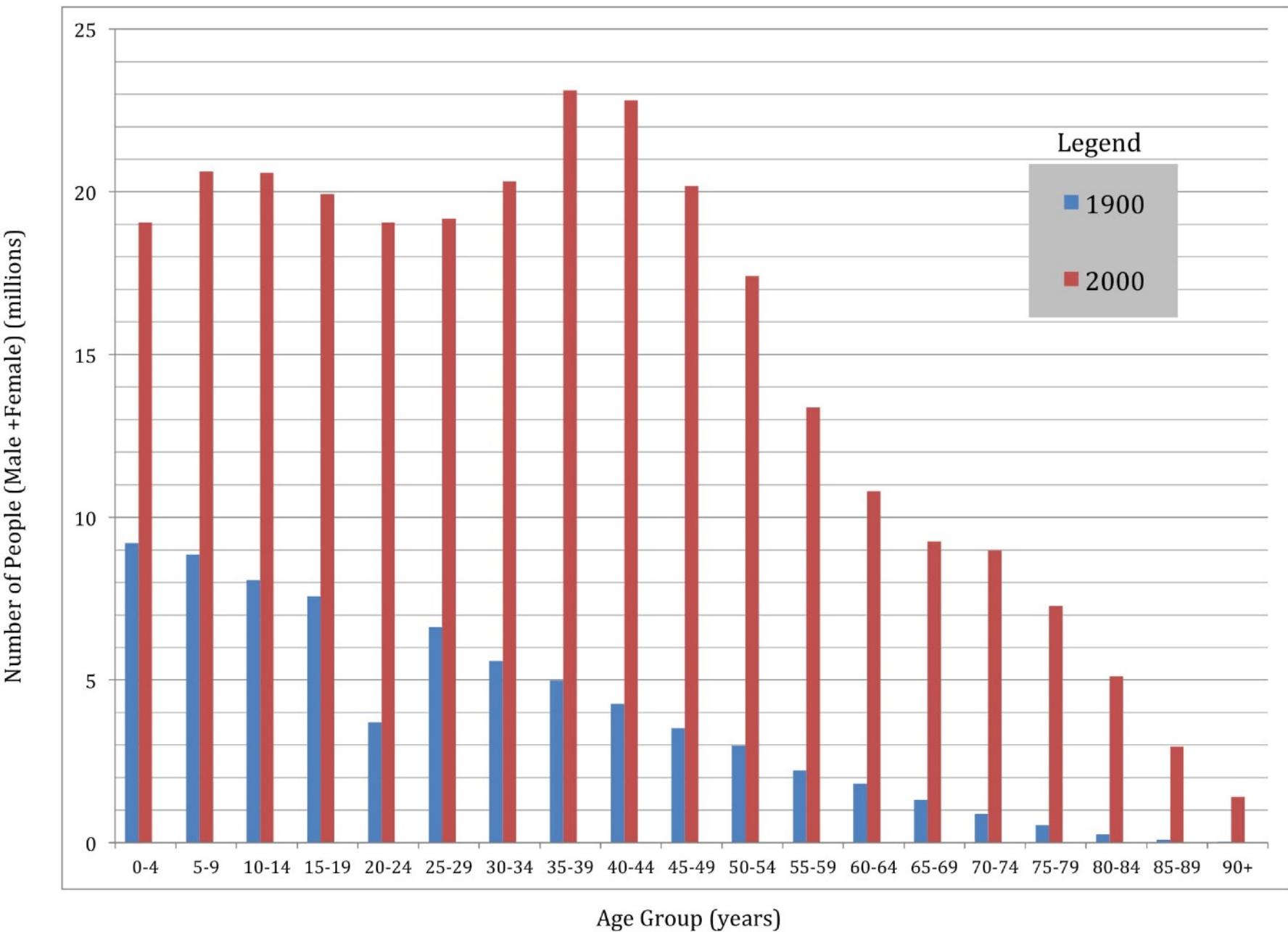
How has the gender composition changed for each age group in the last two centuries?



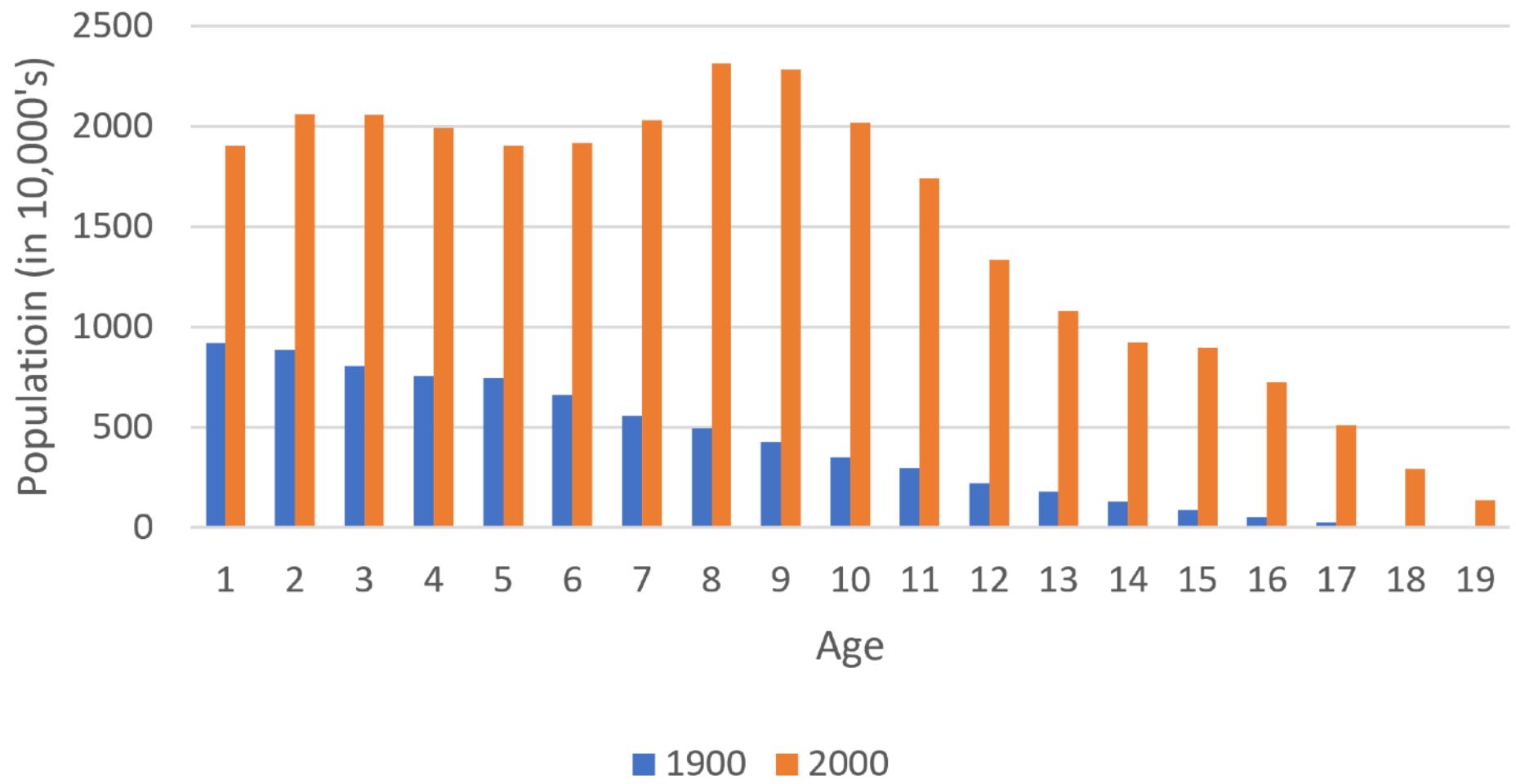
Age, Year

Population Count

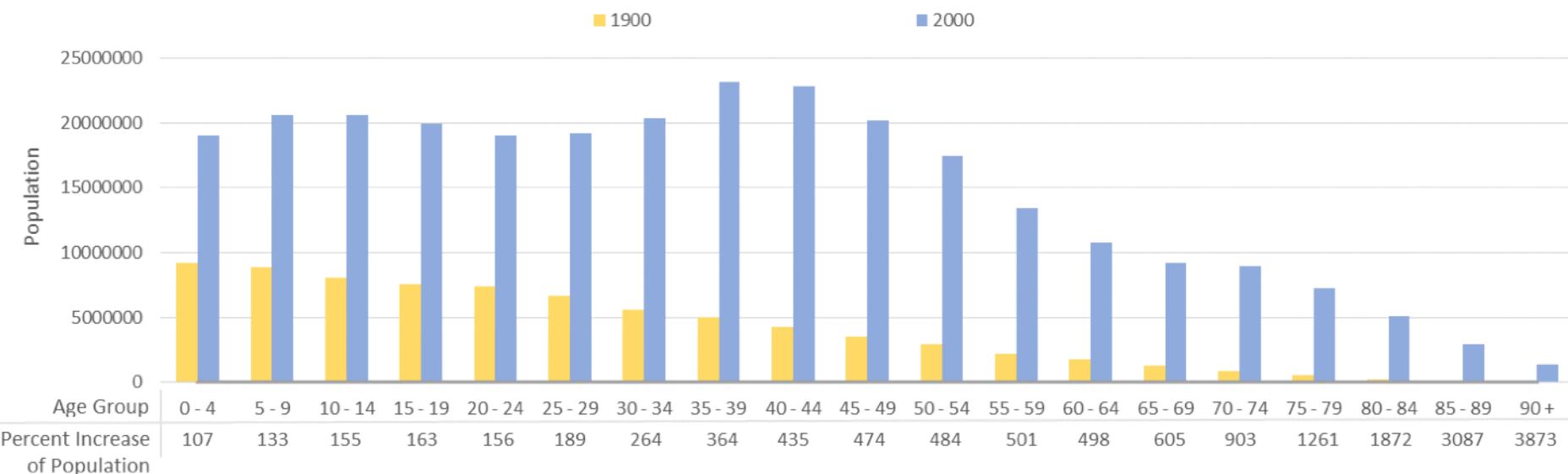
How does the number of people in each age group (binned into 5-year segments) differ from 1900 to 2000?



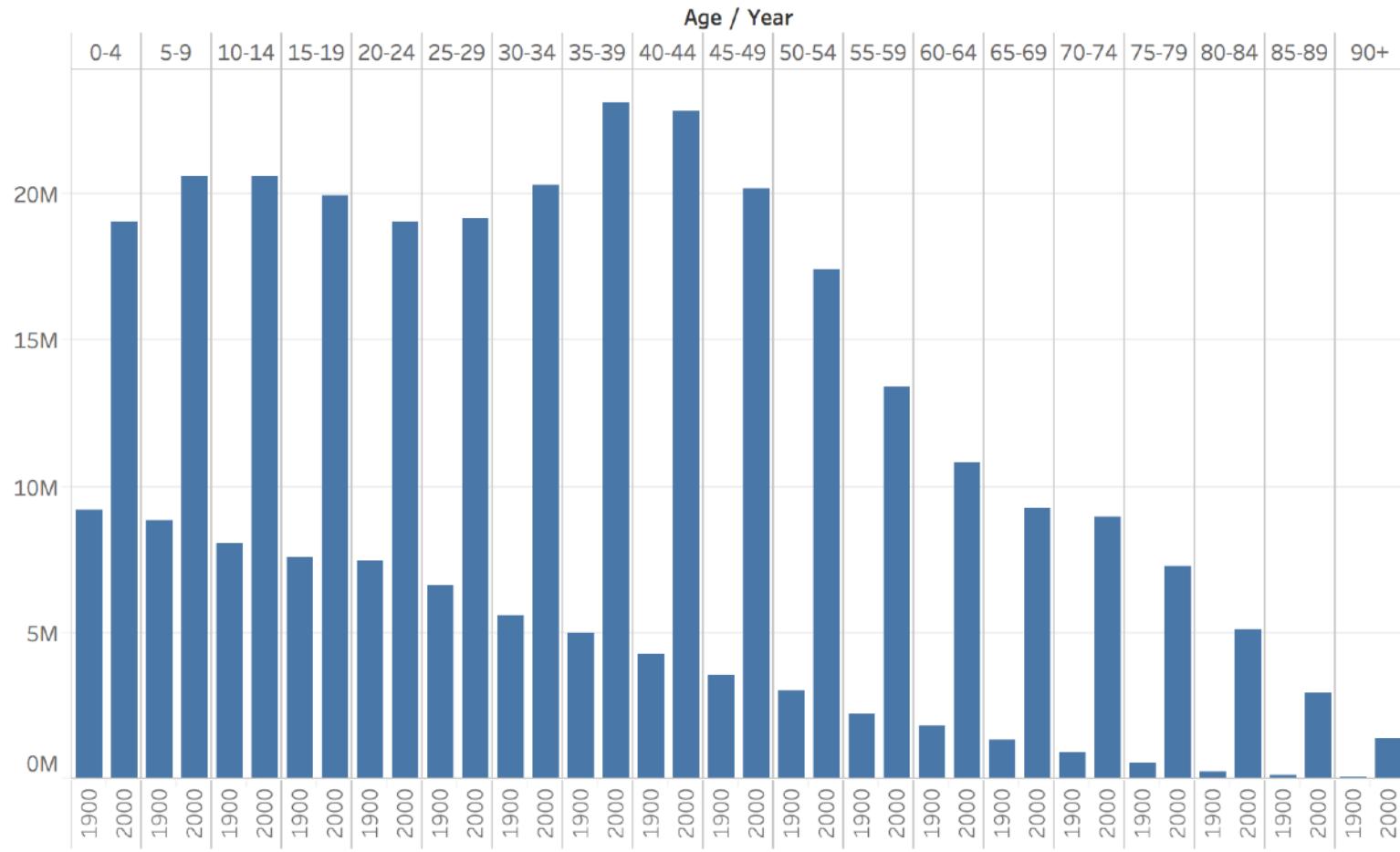
Are Americans Getting Older?



How Did The Population Across Age Groups Change From 1900 to 2000?



Population Change Over Time



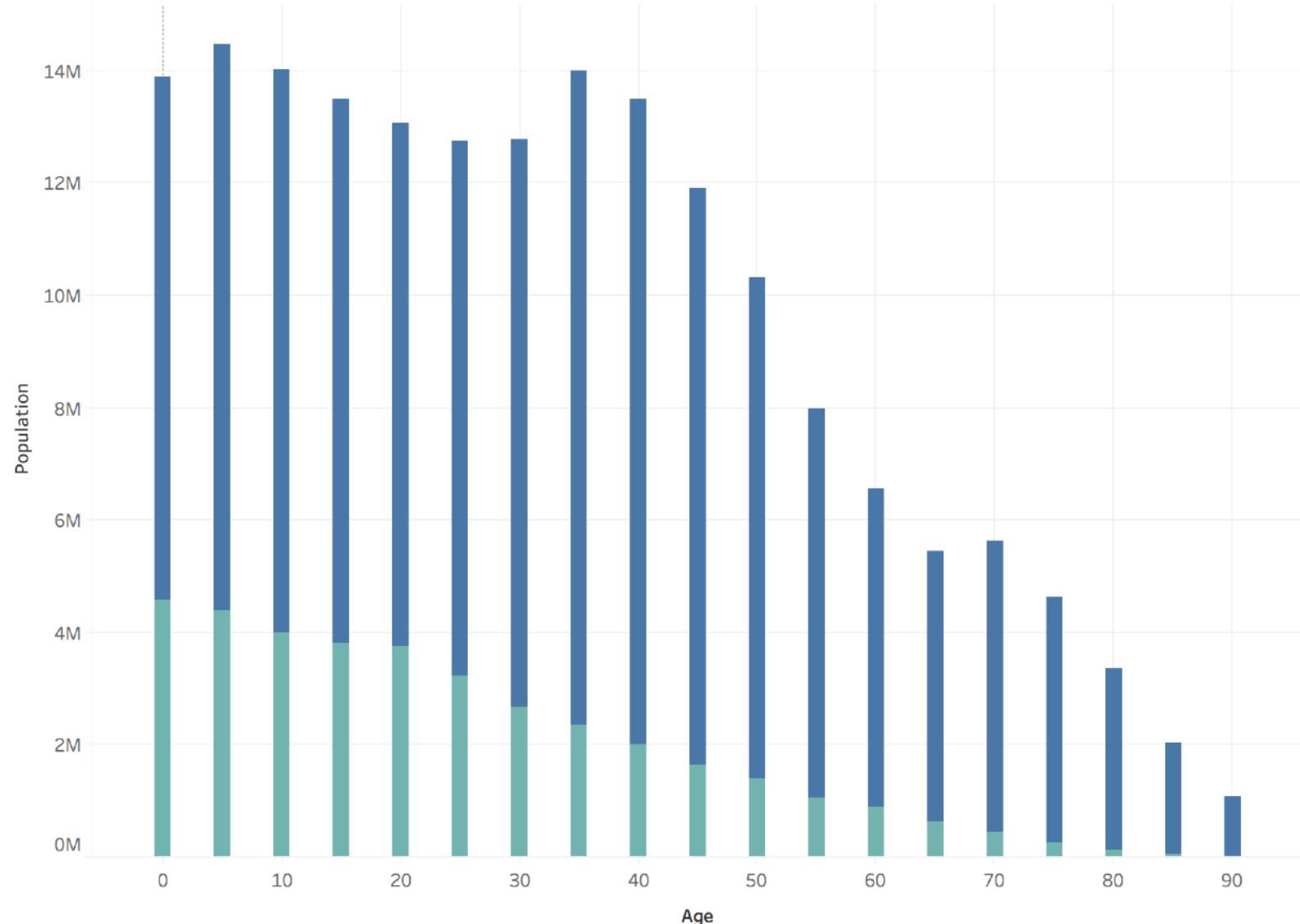
How does the age distribution in the U.S. change from 1900 to 2000?



How does the population, distributed by age, of men differ between 1900 and 2000?

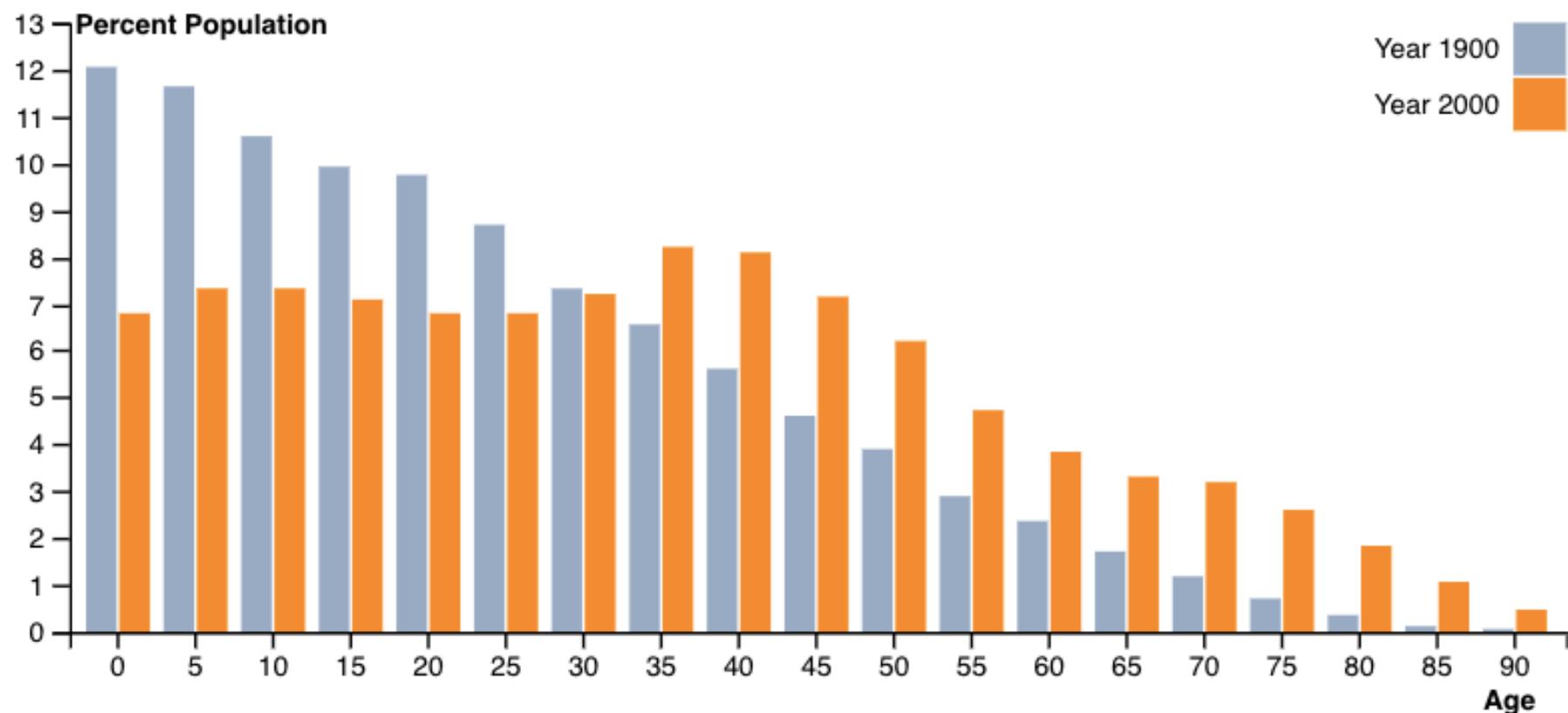
Measure Names

- Men in 2000
- Men in 1900

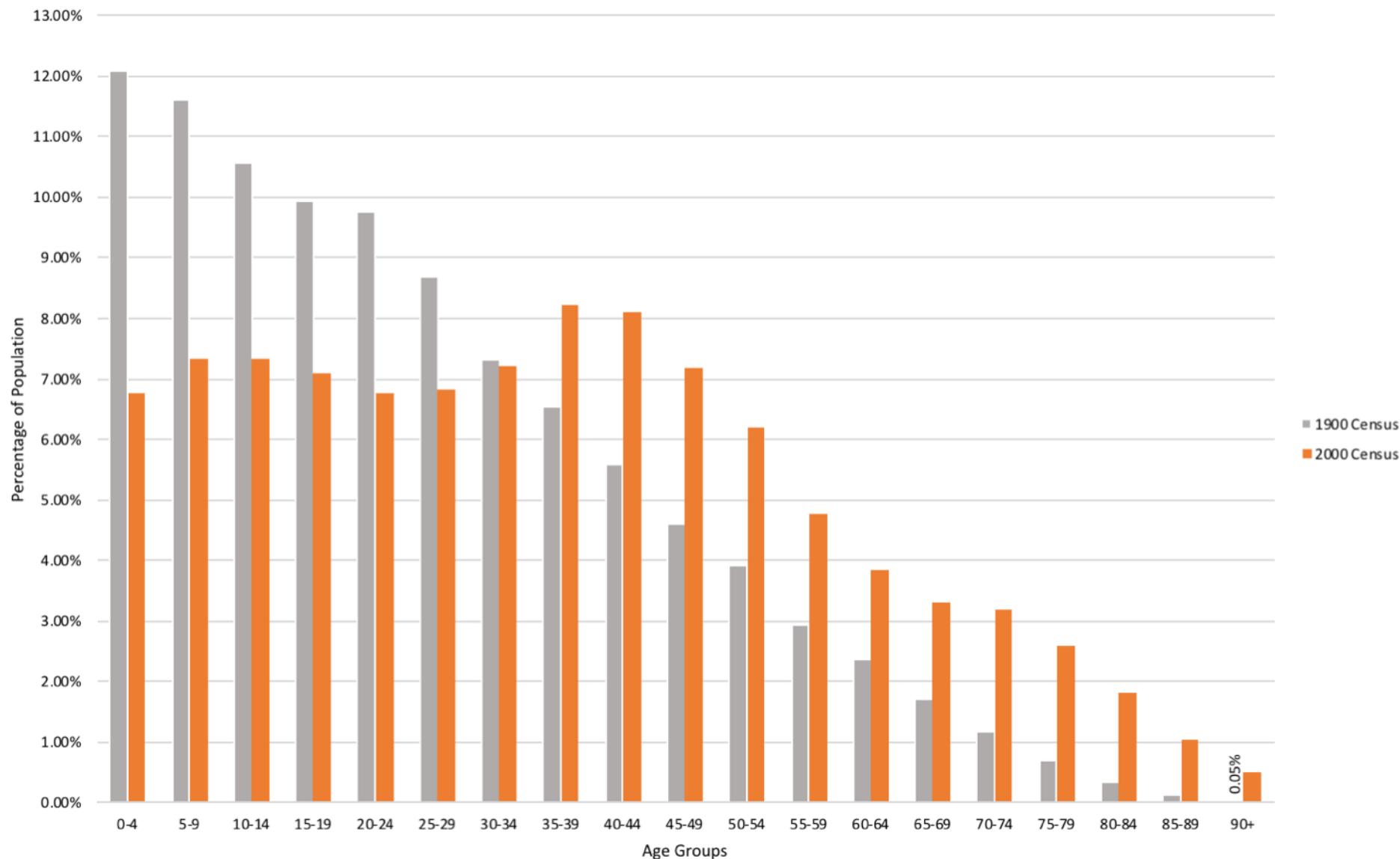


Population Percentage

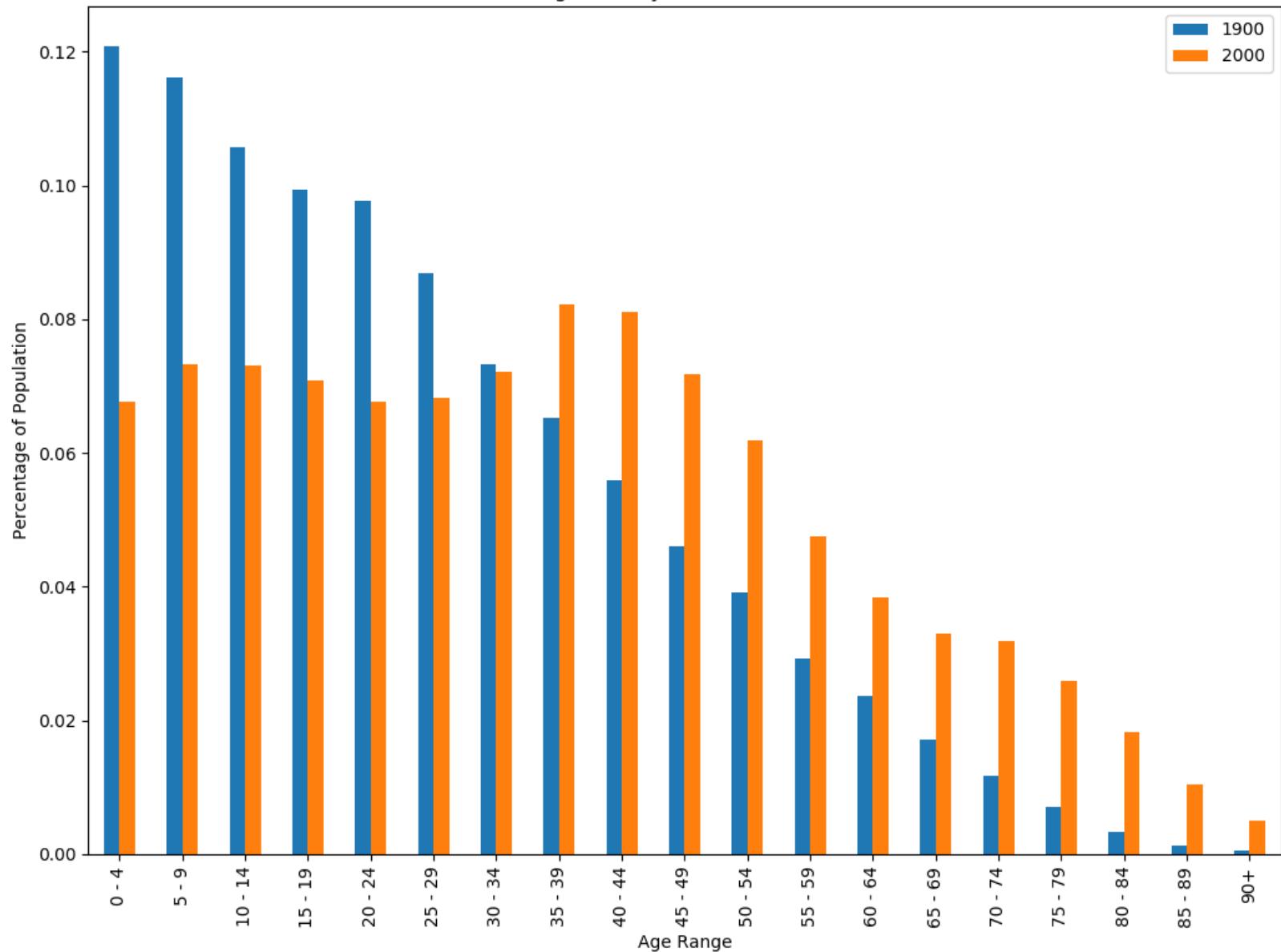
Age distribution comparison between 1900 and 2000.



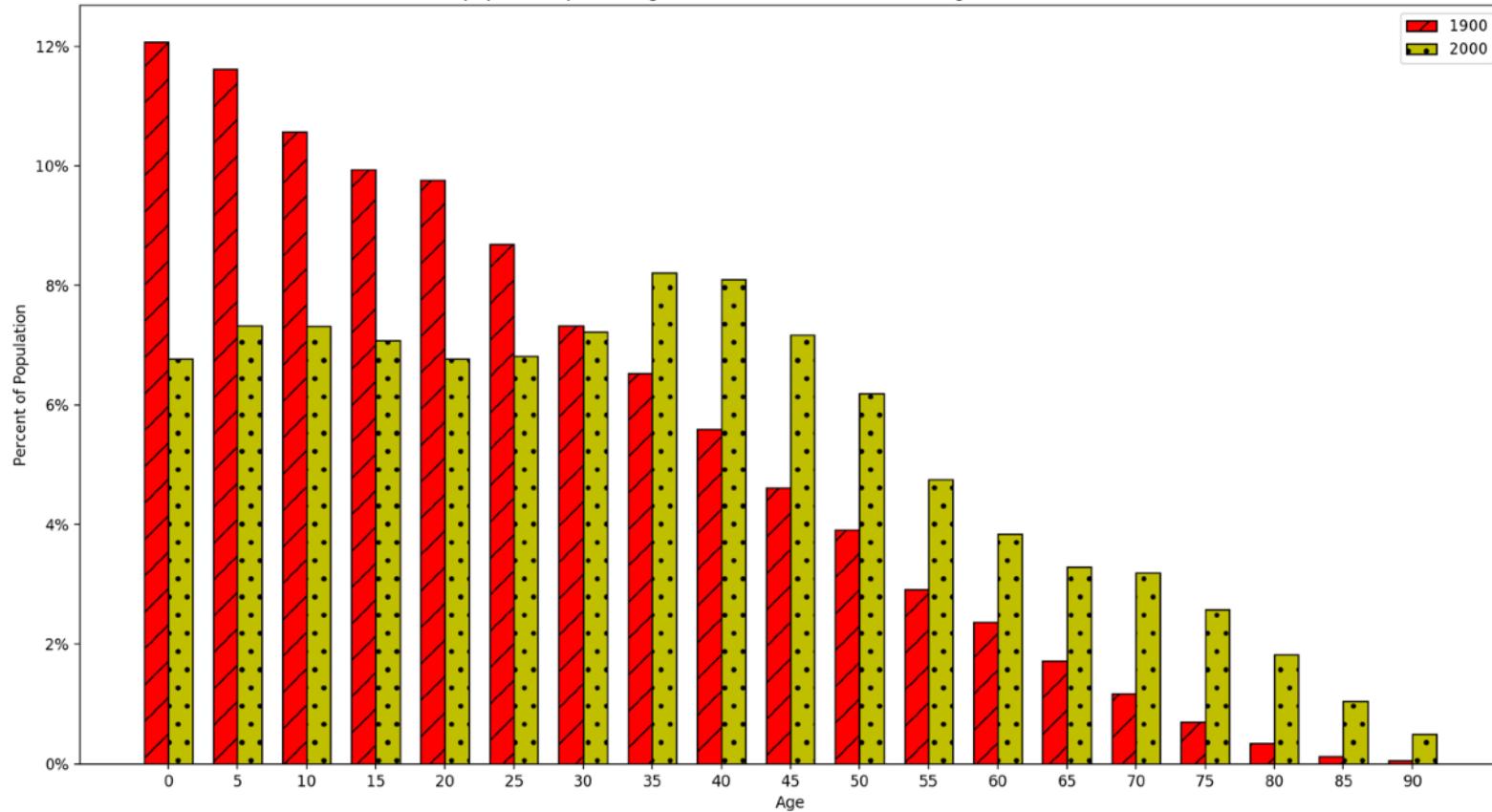
1900 vs 2000 : Are People in the United States Living Longer?



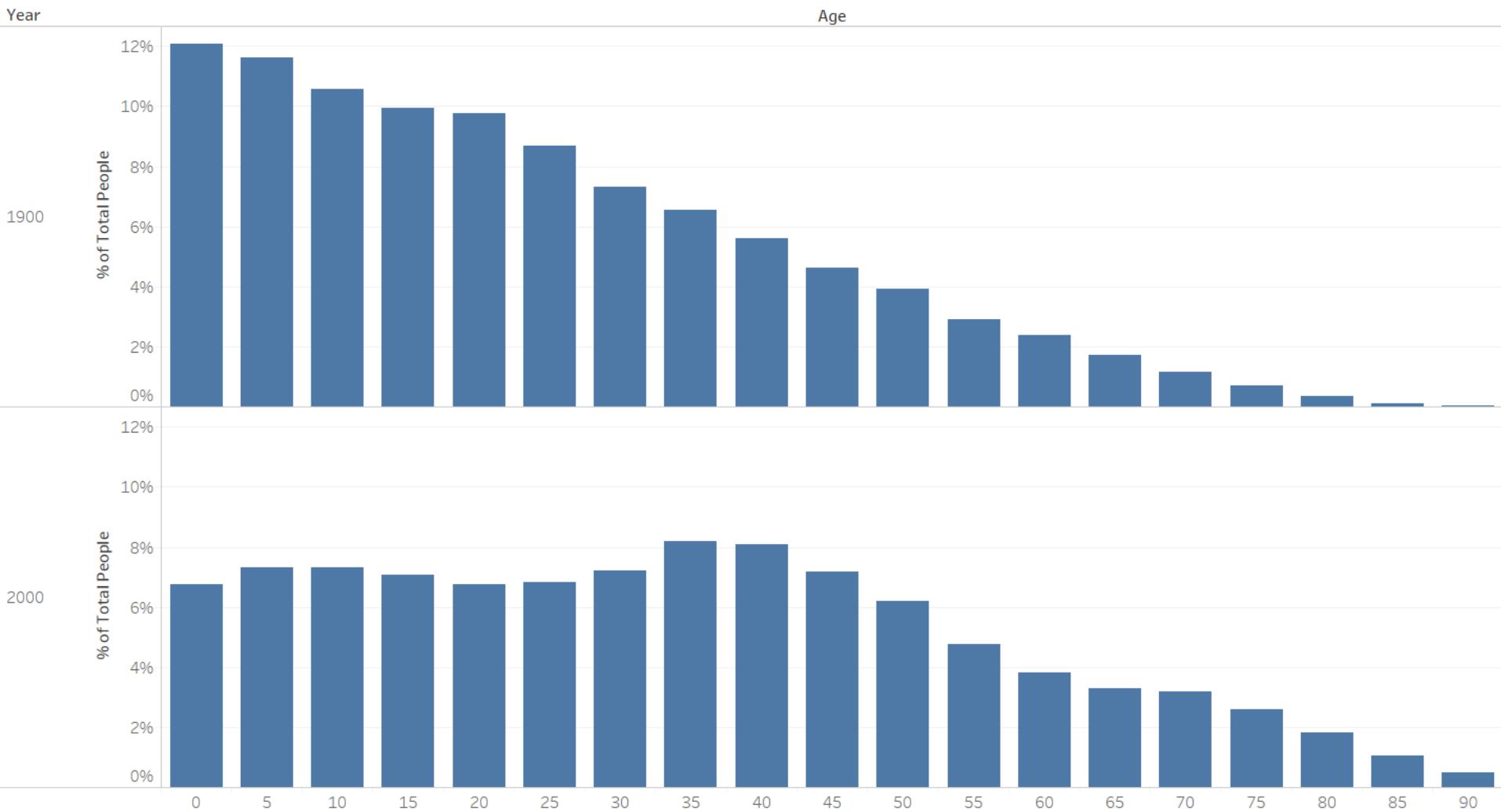
How does the age density differ between the centuries?



What is the population percentage difference between different age cohorts in 1900 and 2000?



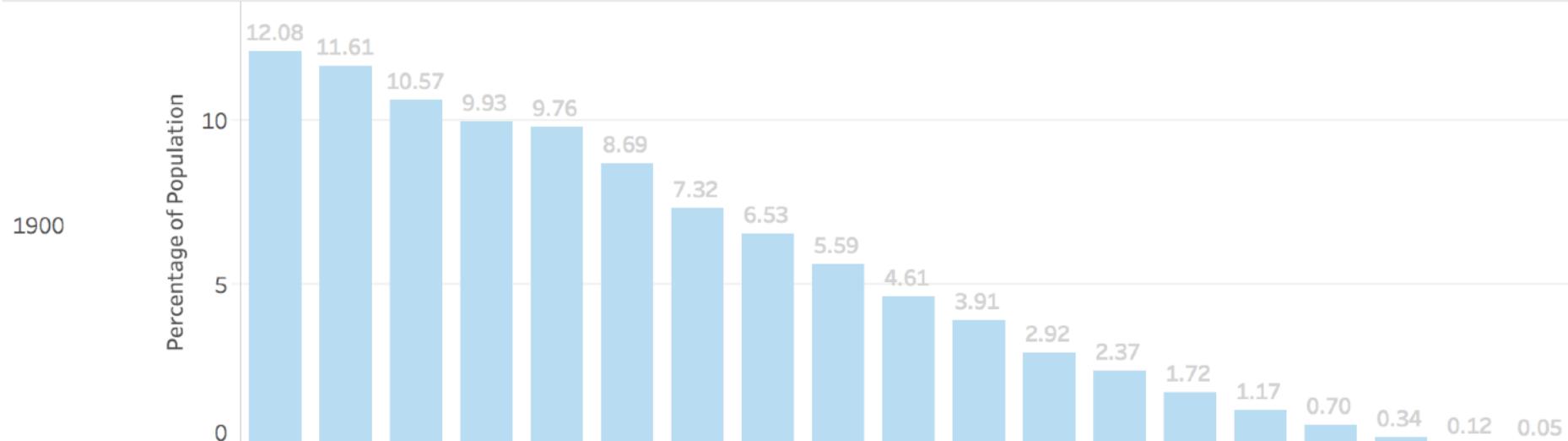
How Did Developments in the 20th Century Affect American Age Demographics?



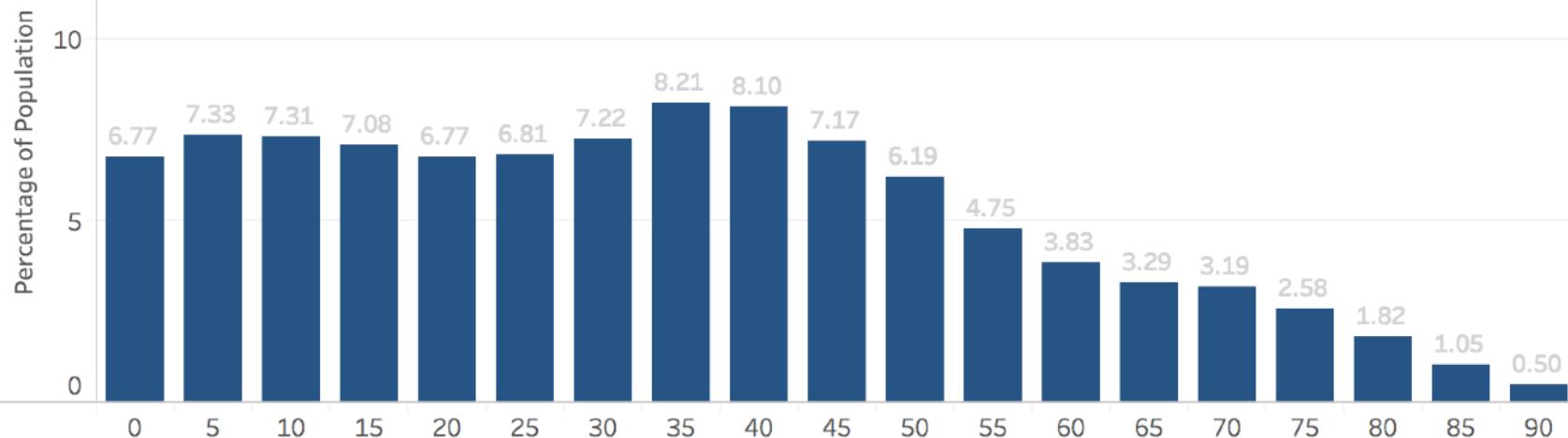
Rising Life Expectancy and Declining Birth Rate - Differences in Age Groups Between 1900 and 2000

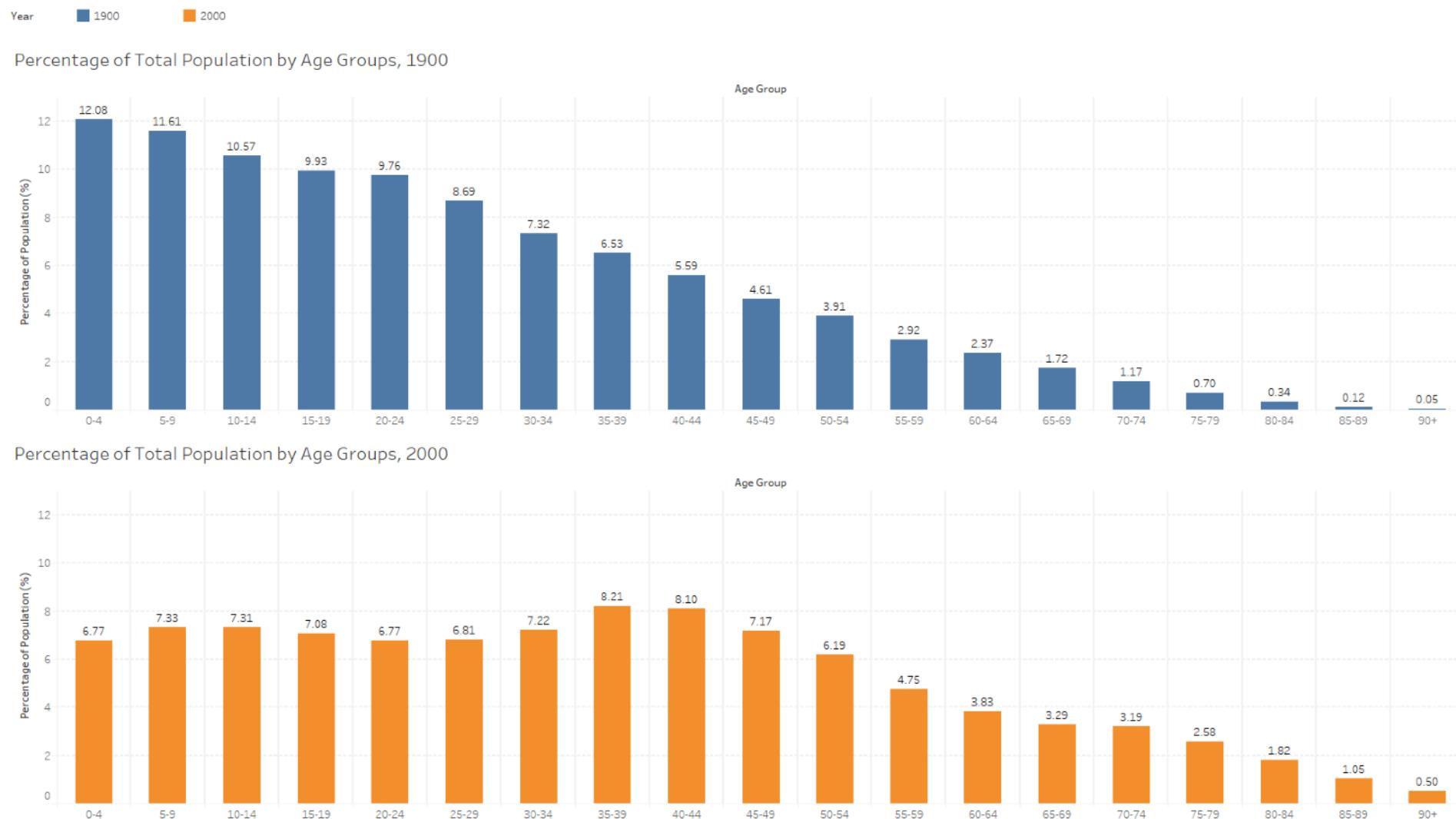
Year

Age



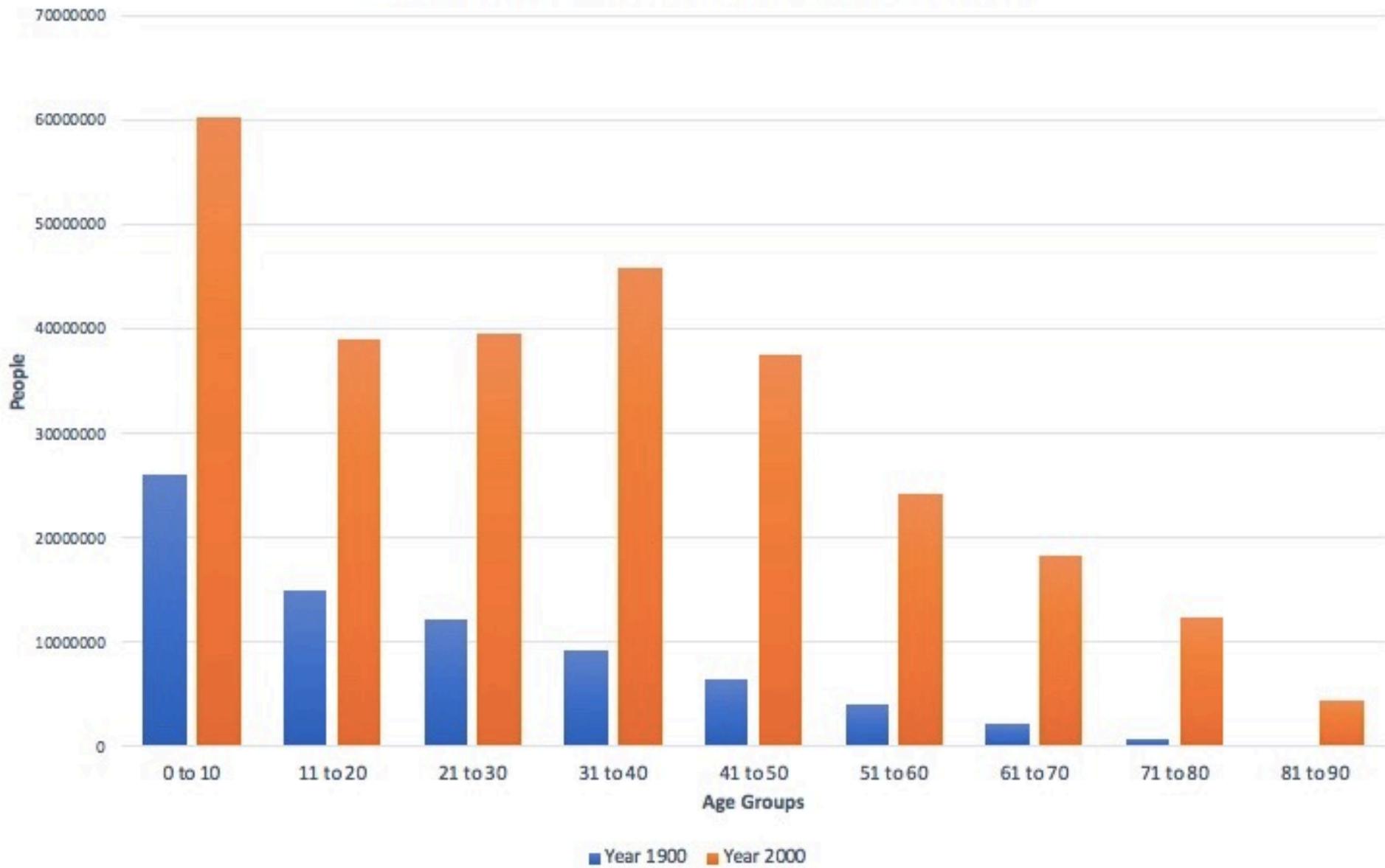
2000



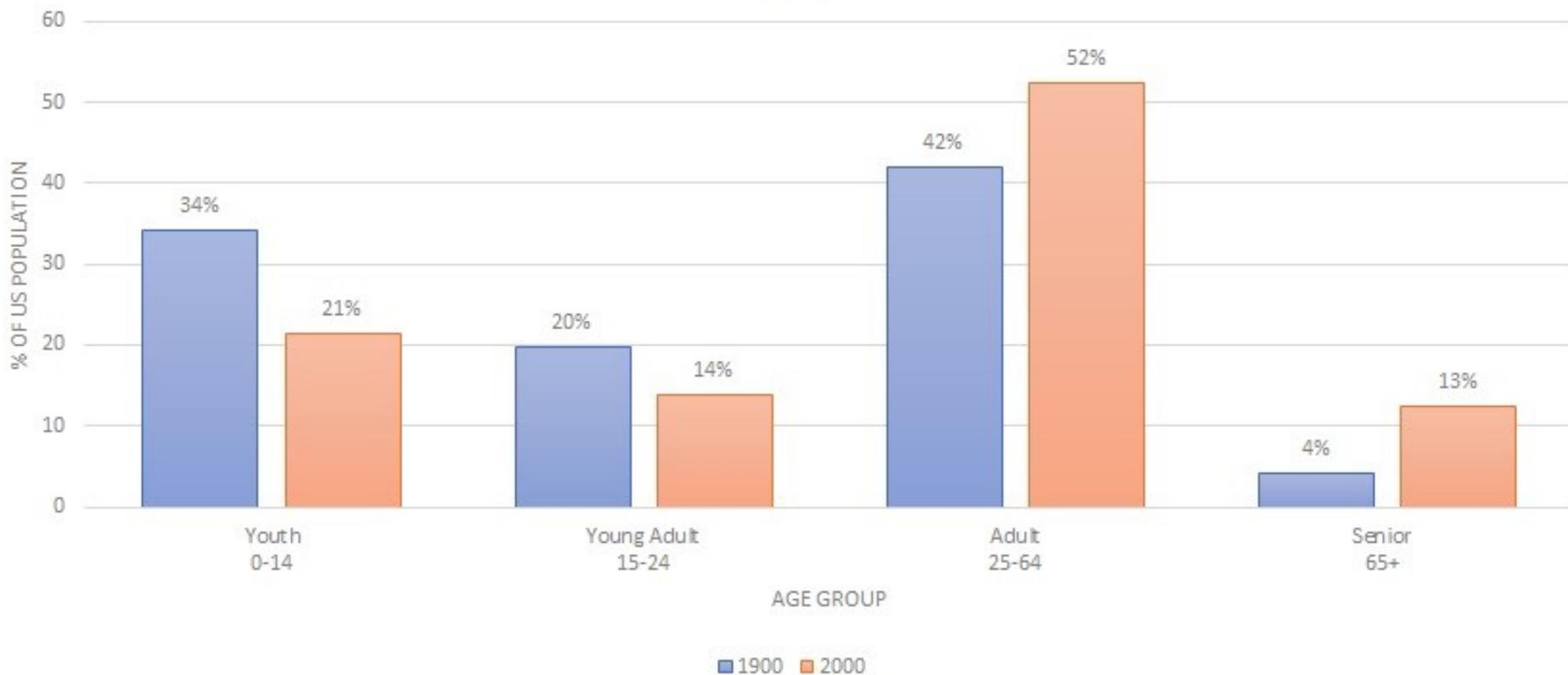


Age Binning

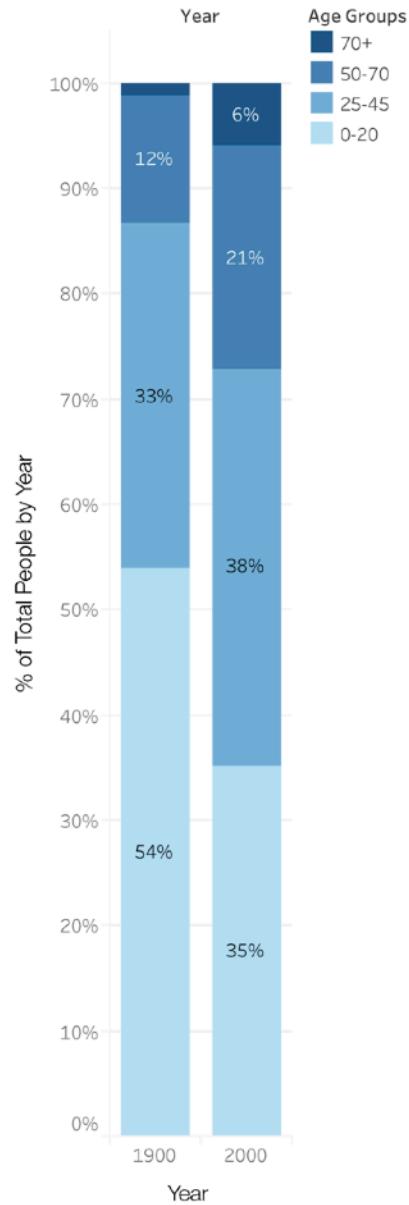
Difference of the U.S. Age Population in 1900 vs 2000



How do the percentages of the US population in various age groups differ in 1900 and 2000?

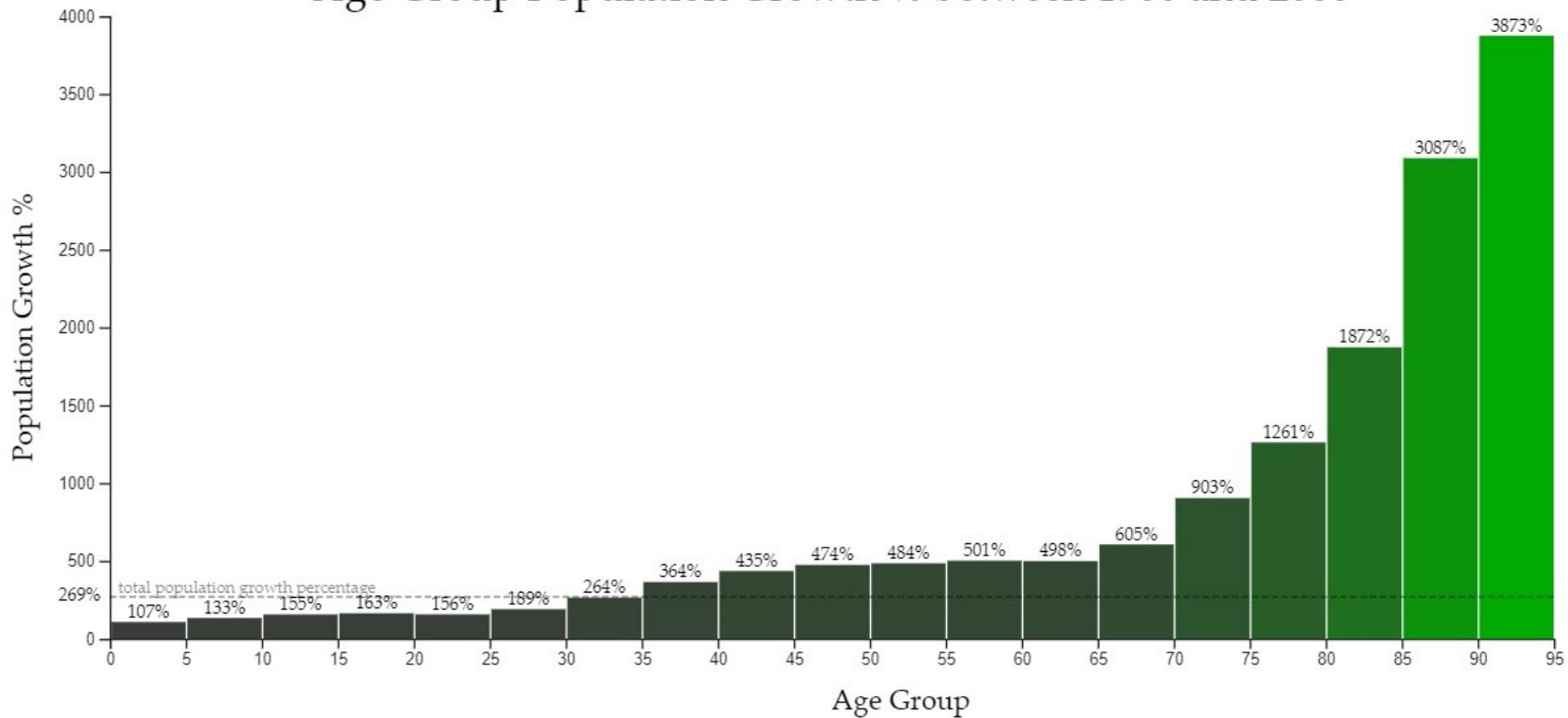


How has the age distribution in the U.S.
changed from 1900 to 2000?

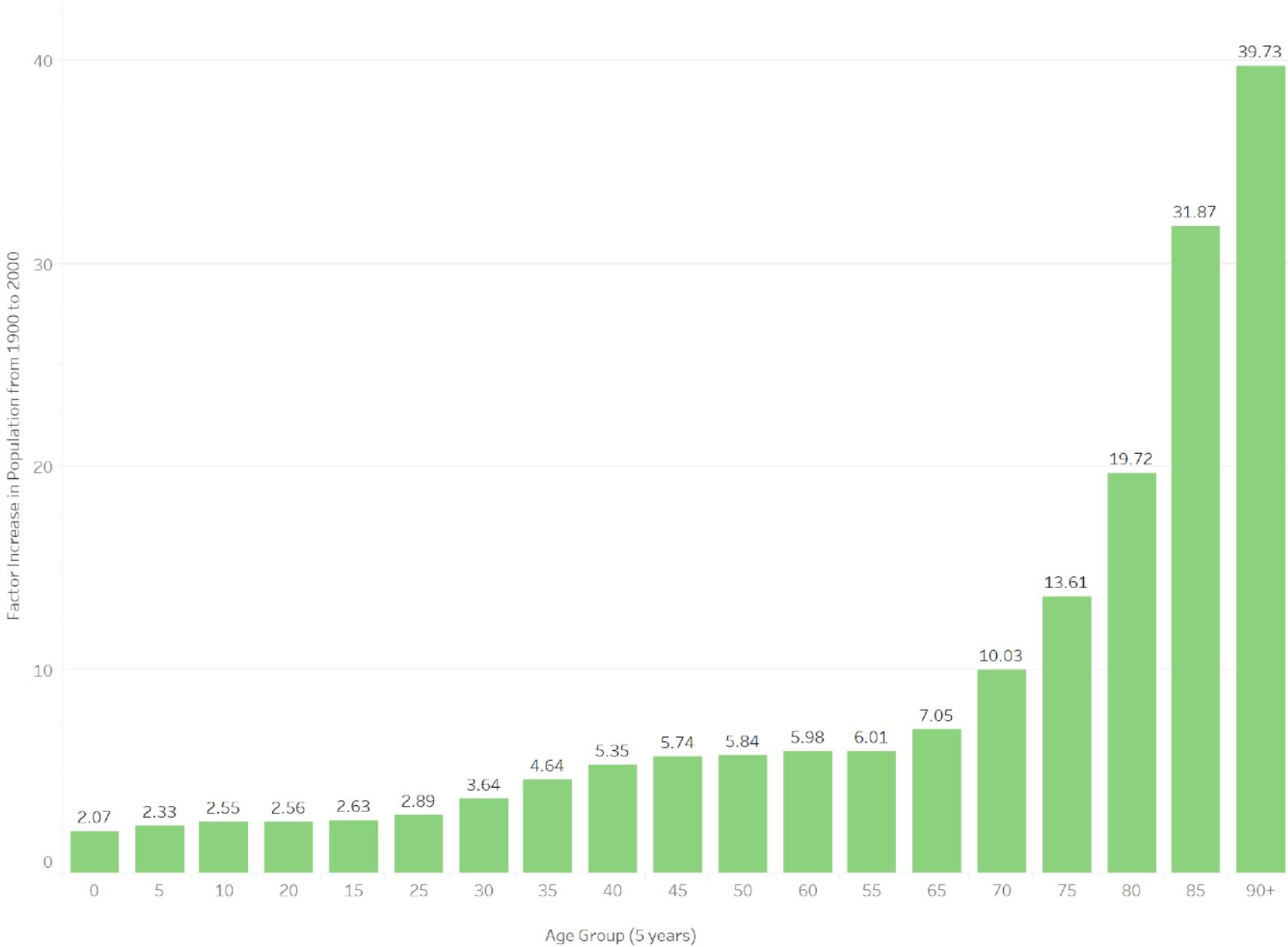


Growth Rates

Age Group Population Growth % between 1900 and 2000

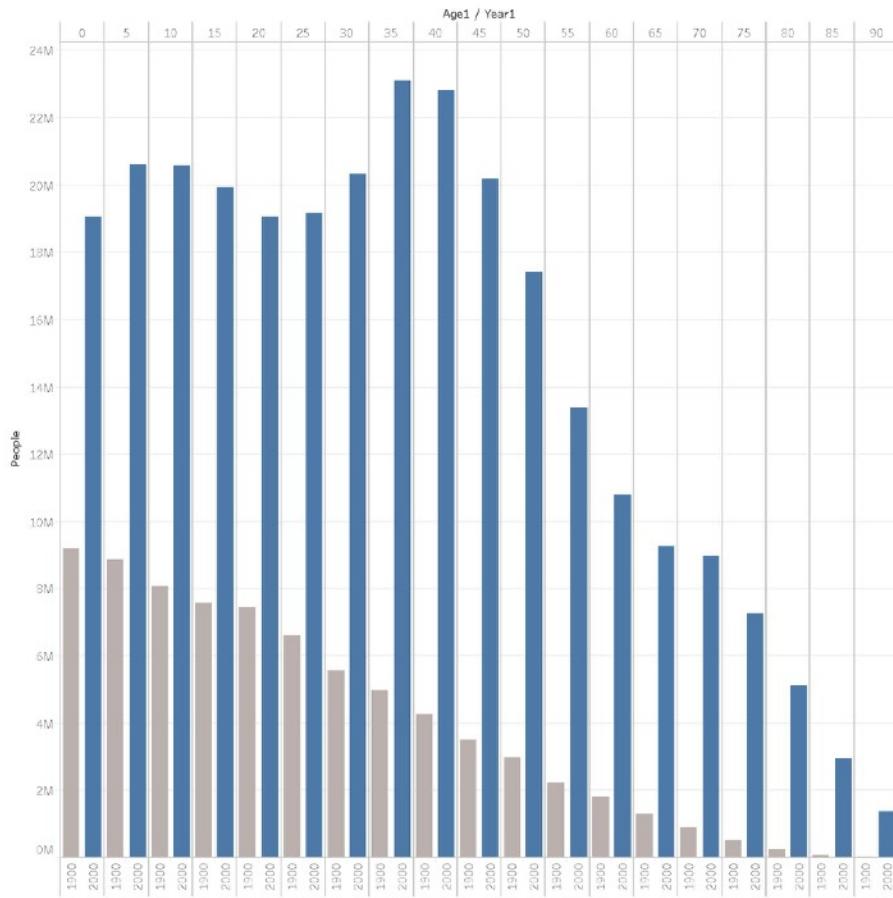


Which US Age Group is Growing the Fastest?

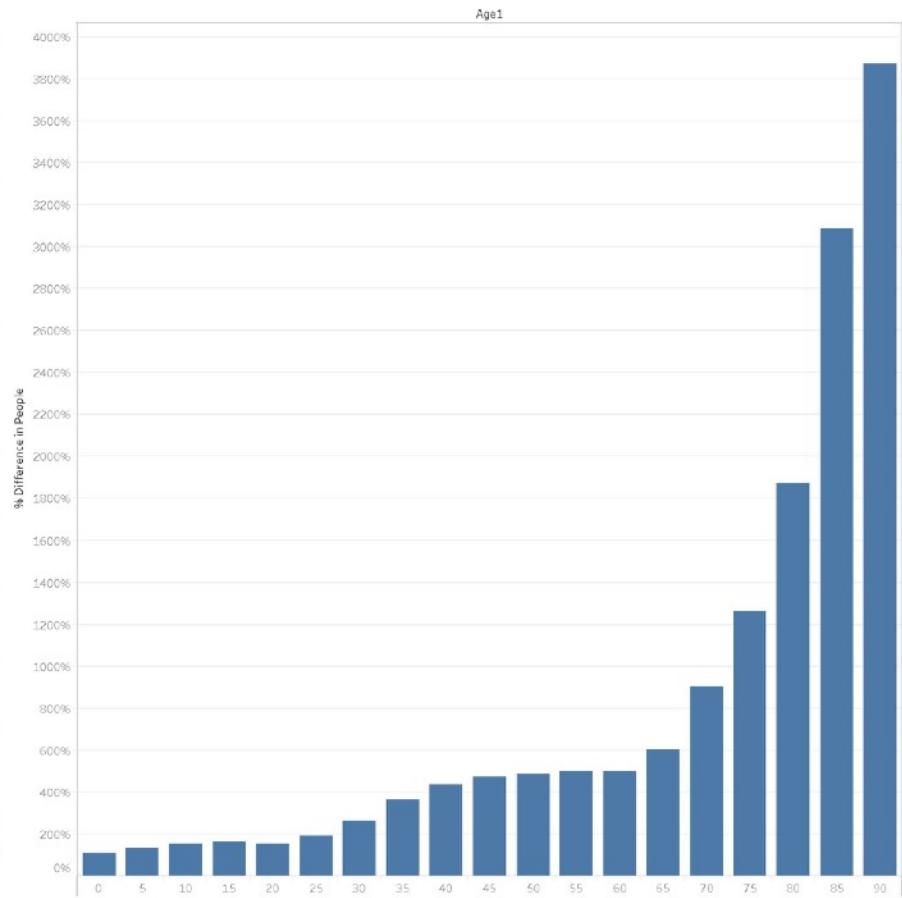


What Age Groups Saw the Largest Increase in Population from 1900 to 2000?

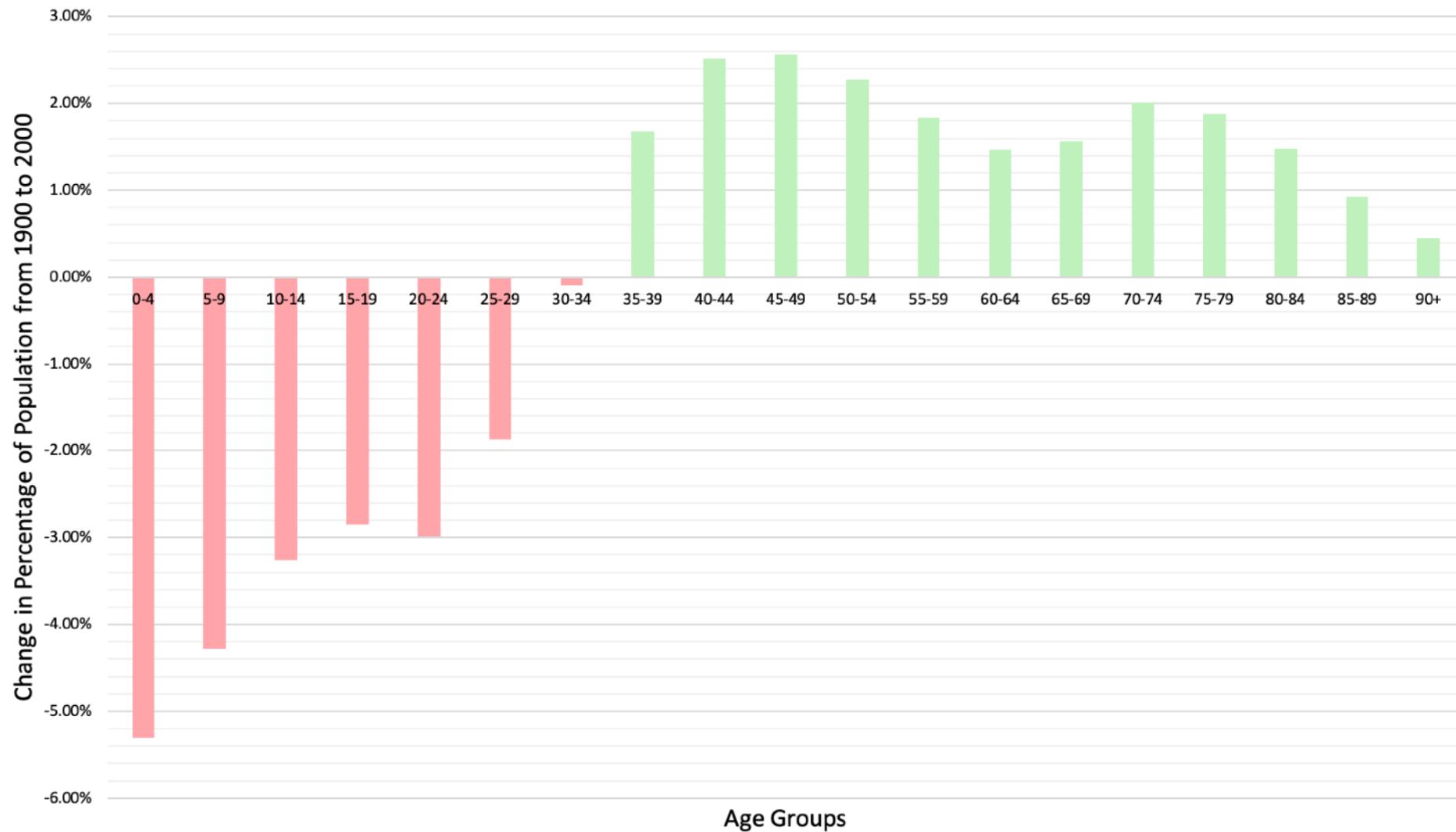
Count of People by Age and Year



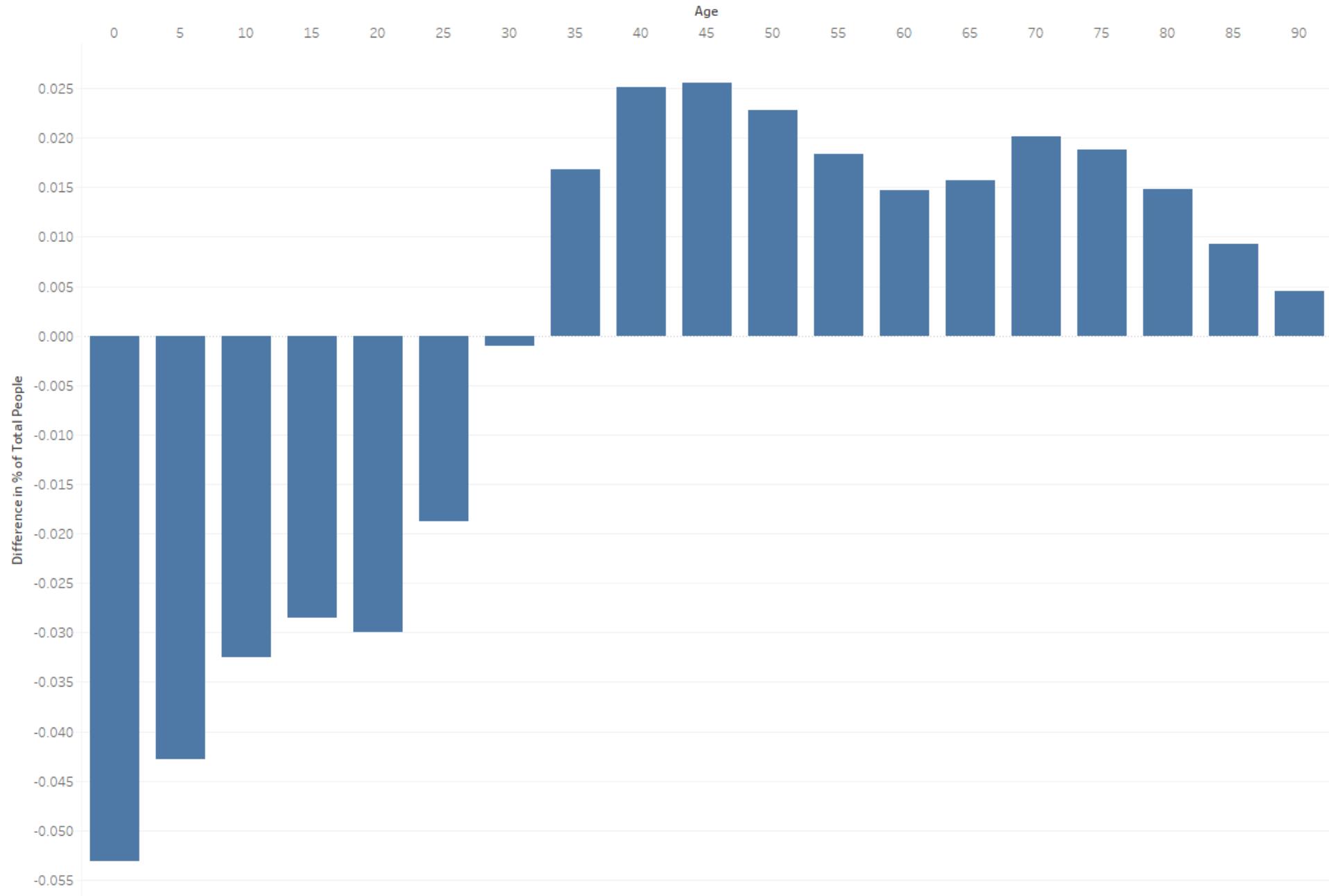
Percent Difference in People Count From 1900 to 2000 Organized by Age



Is the U.S. getting older?

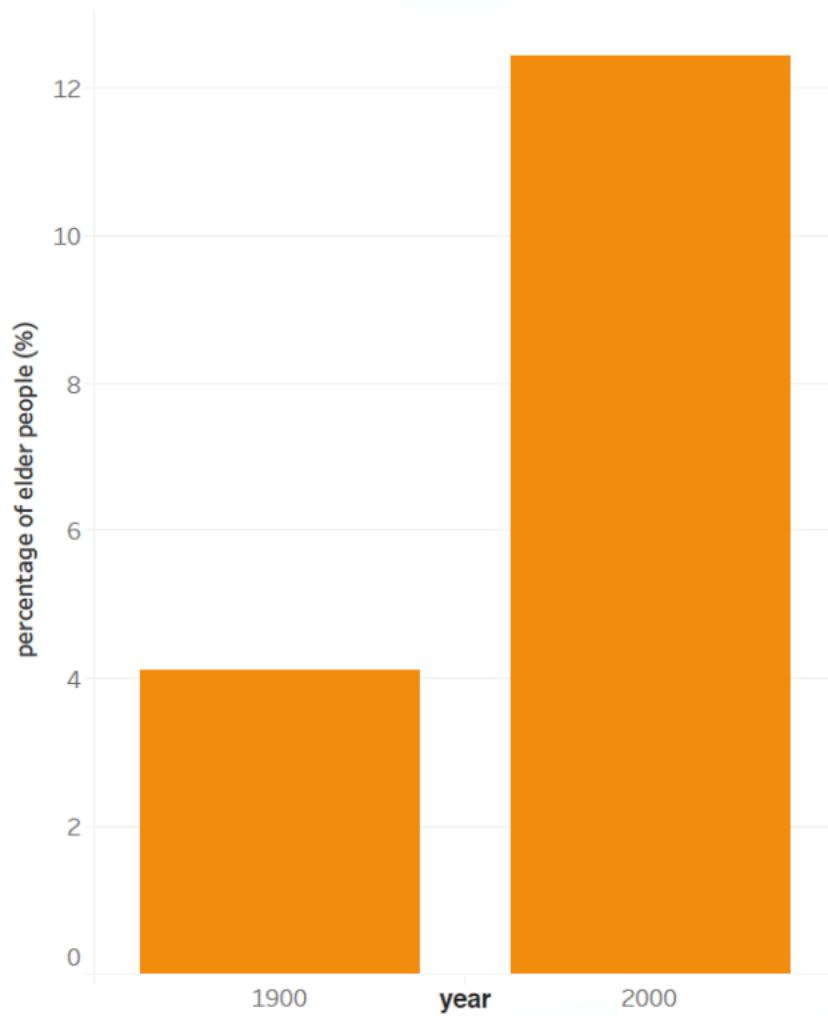


How does the population distribution by age in America differ in year 2000 compared to 1900?



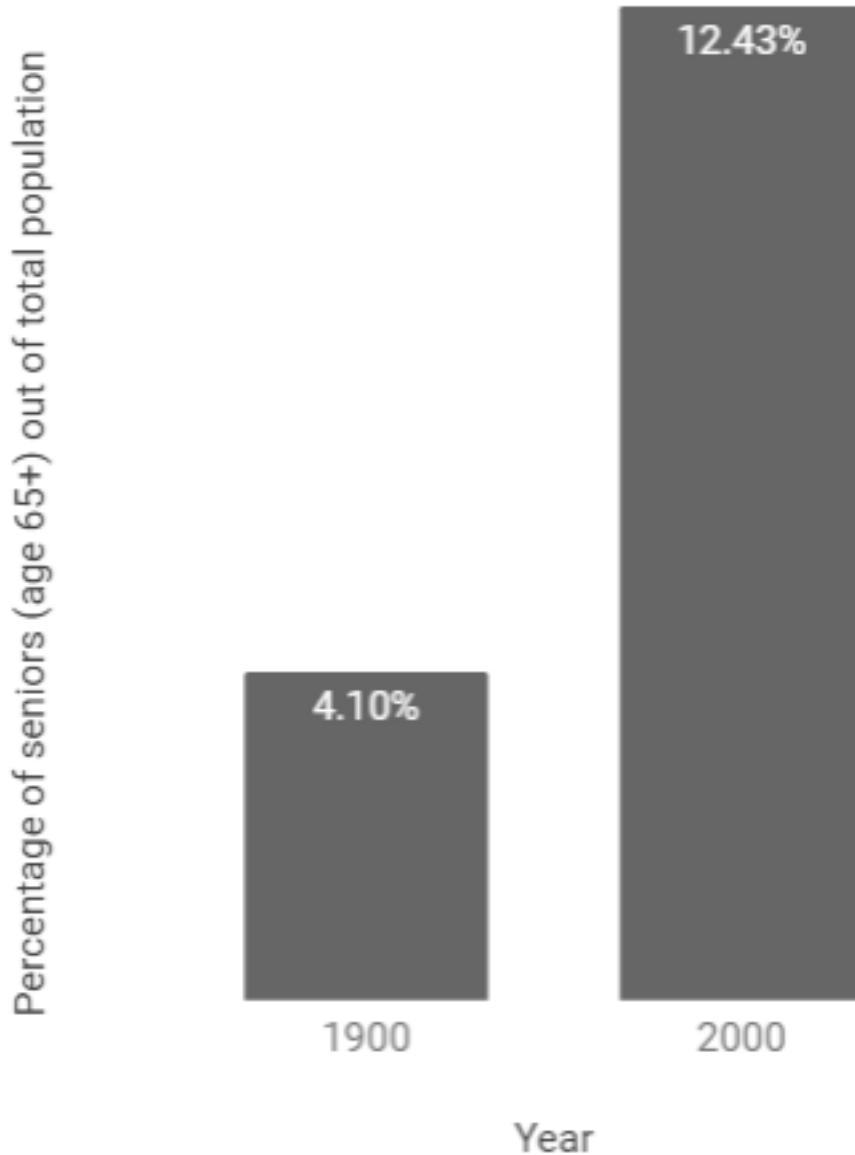
Simplification

How does the percentage of elder people in 1900 compare to that in 2000?

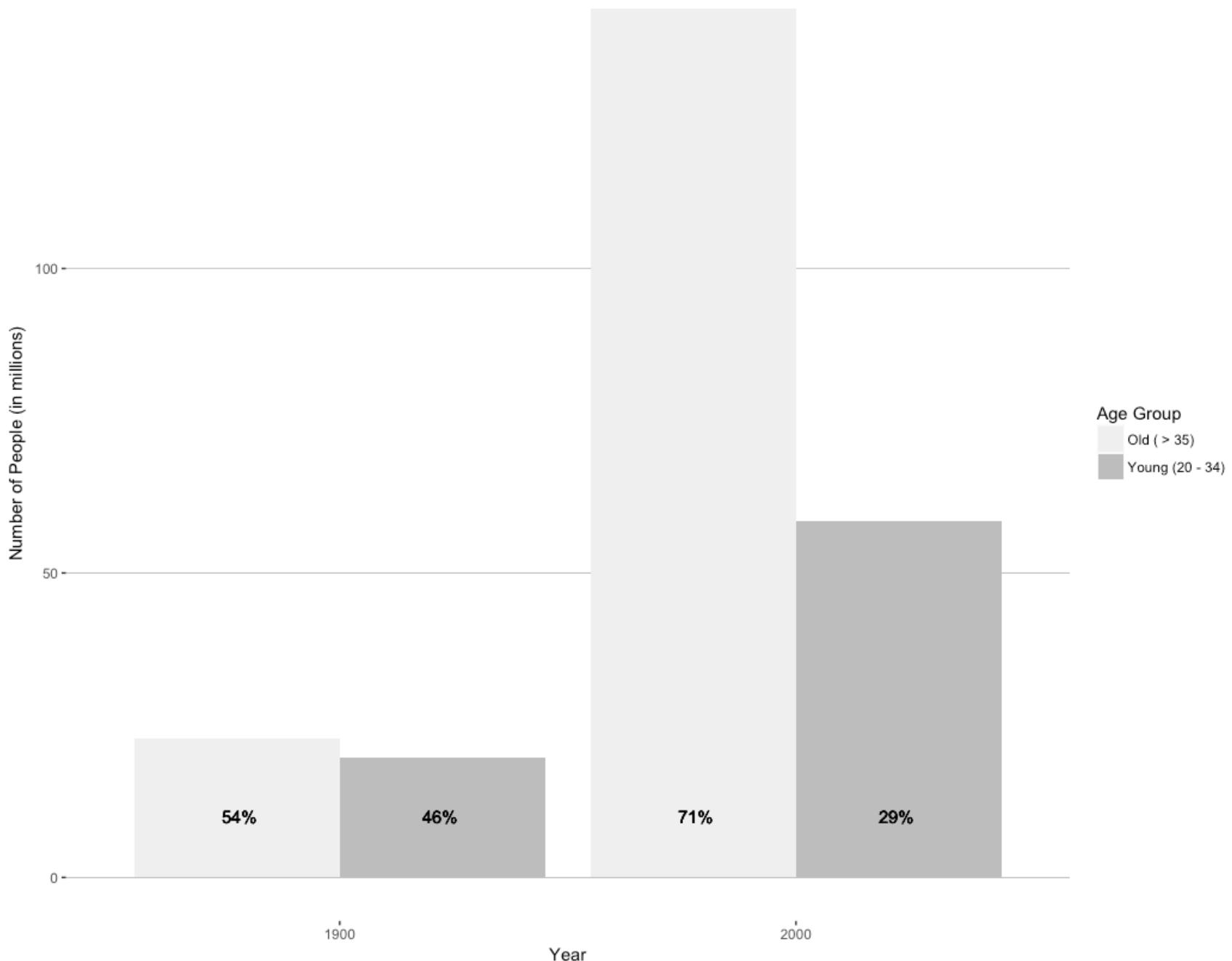


Note that according to WHO's definition, "elder people" here refer to those people who are over (and including) age 65.
(<http://www.who.int/healthinfo/survey/ageingdefolder/en/>)

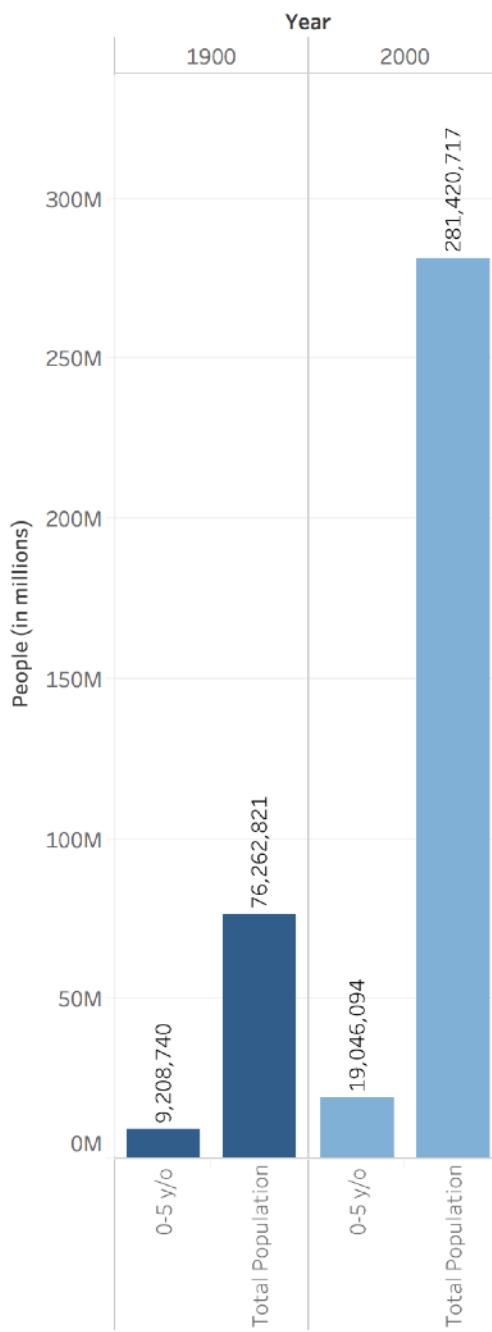
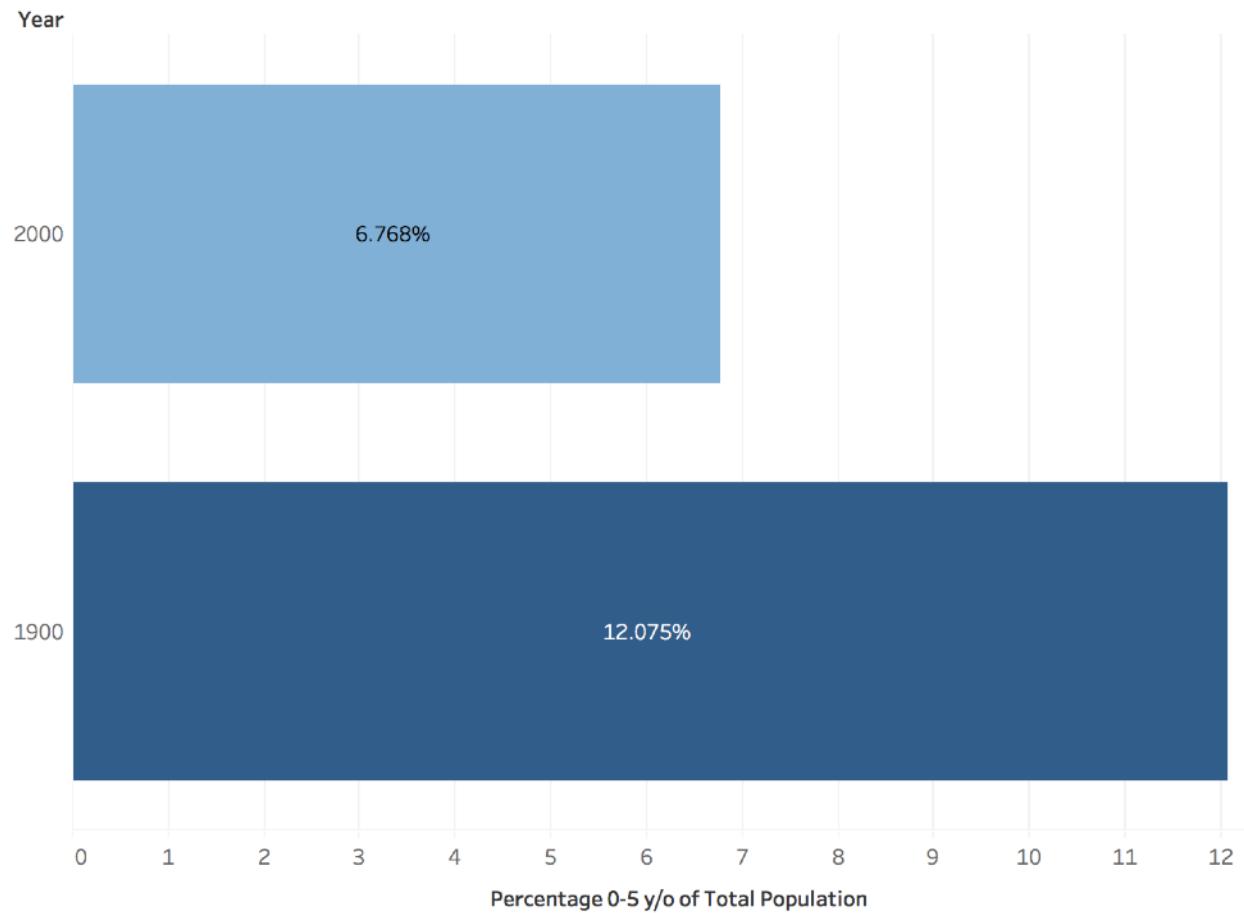
How has our senior population changed over a century?



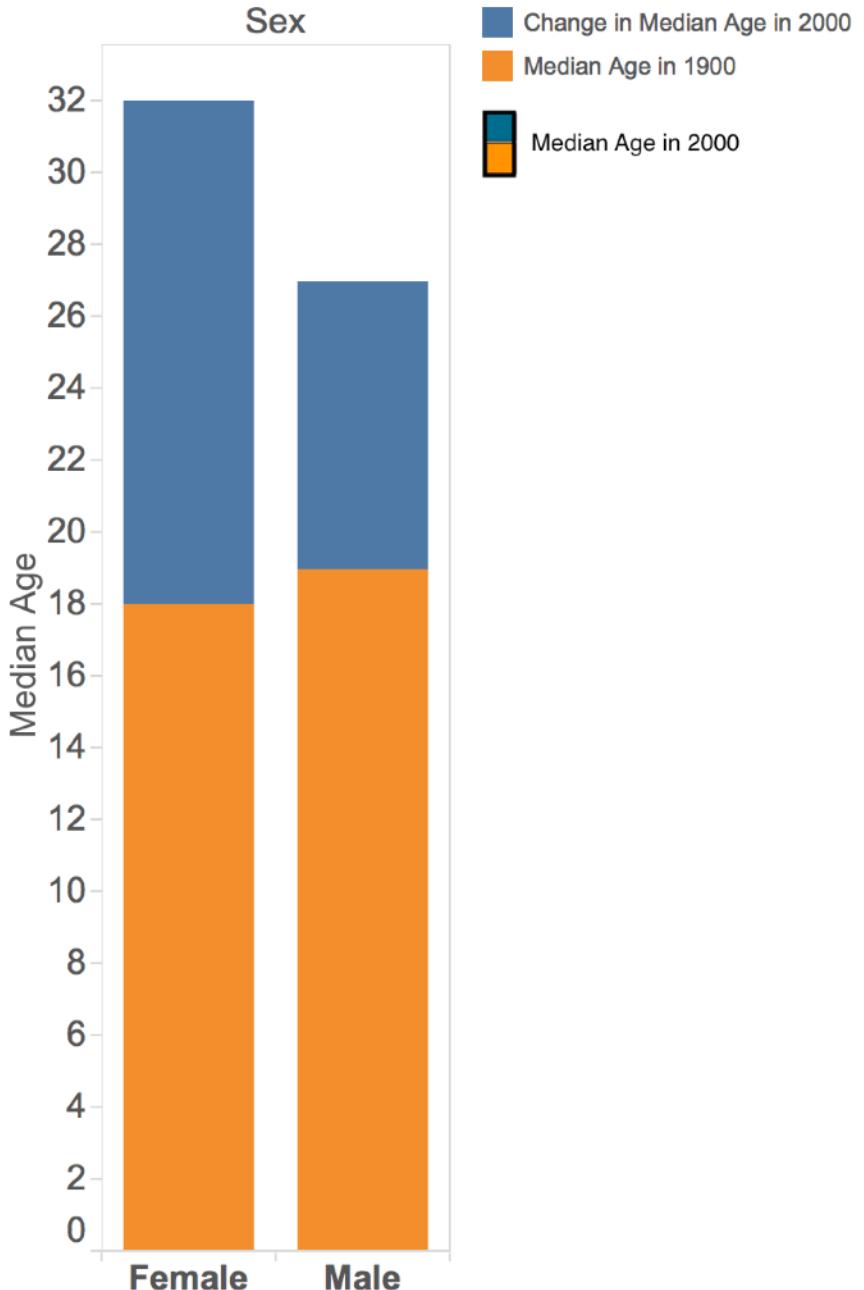
How has the proportion of young voting age people changed between 1900 and 2000?



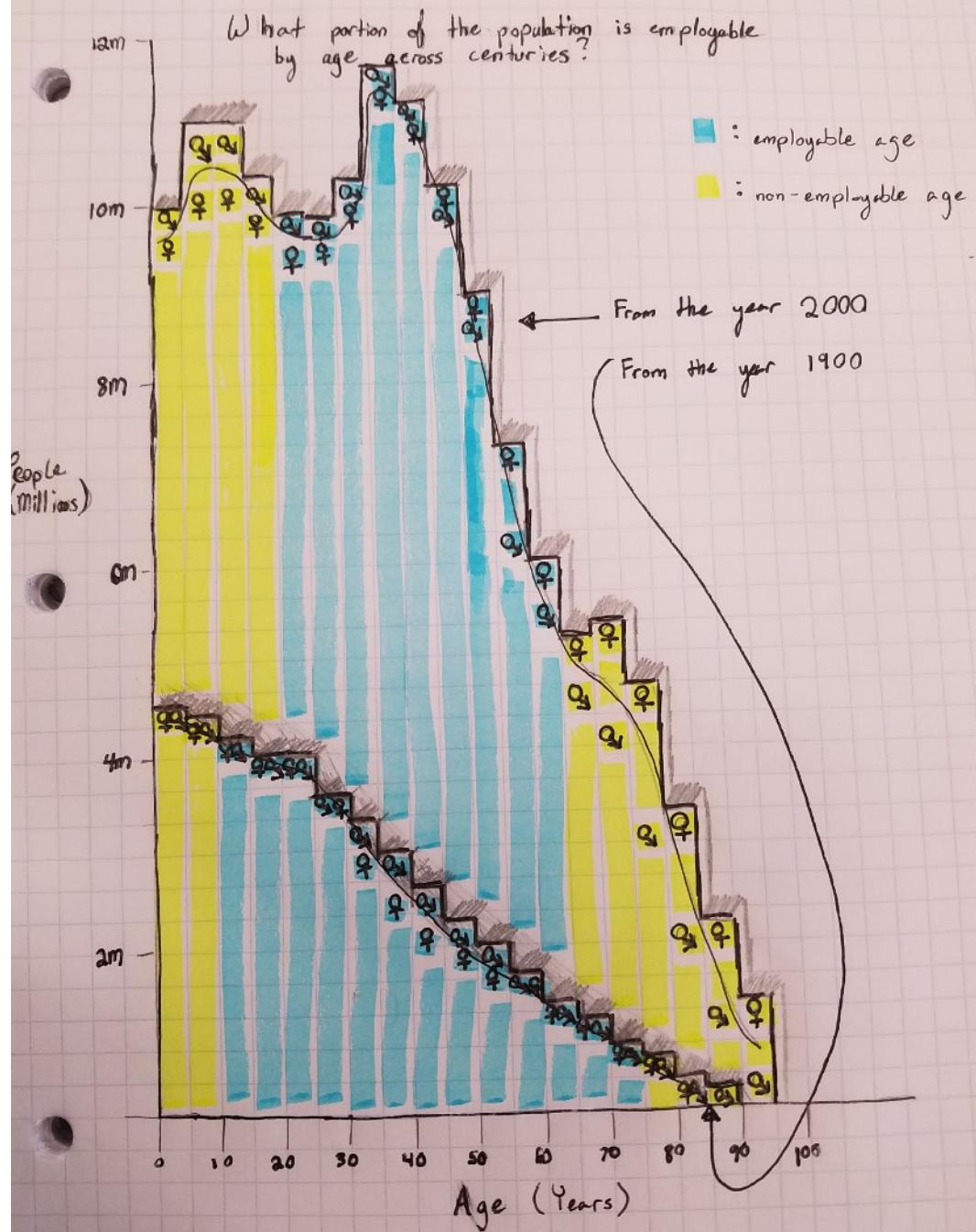
What percentage of the population is 0-5 years old in 1900 vs. 2000?



How did the Median Age change over the past century?



Employment



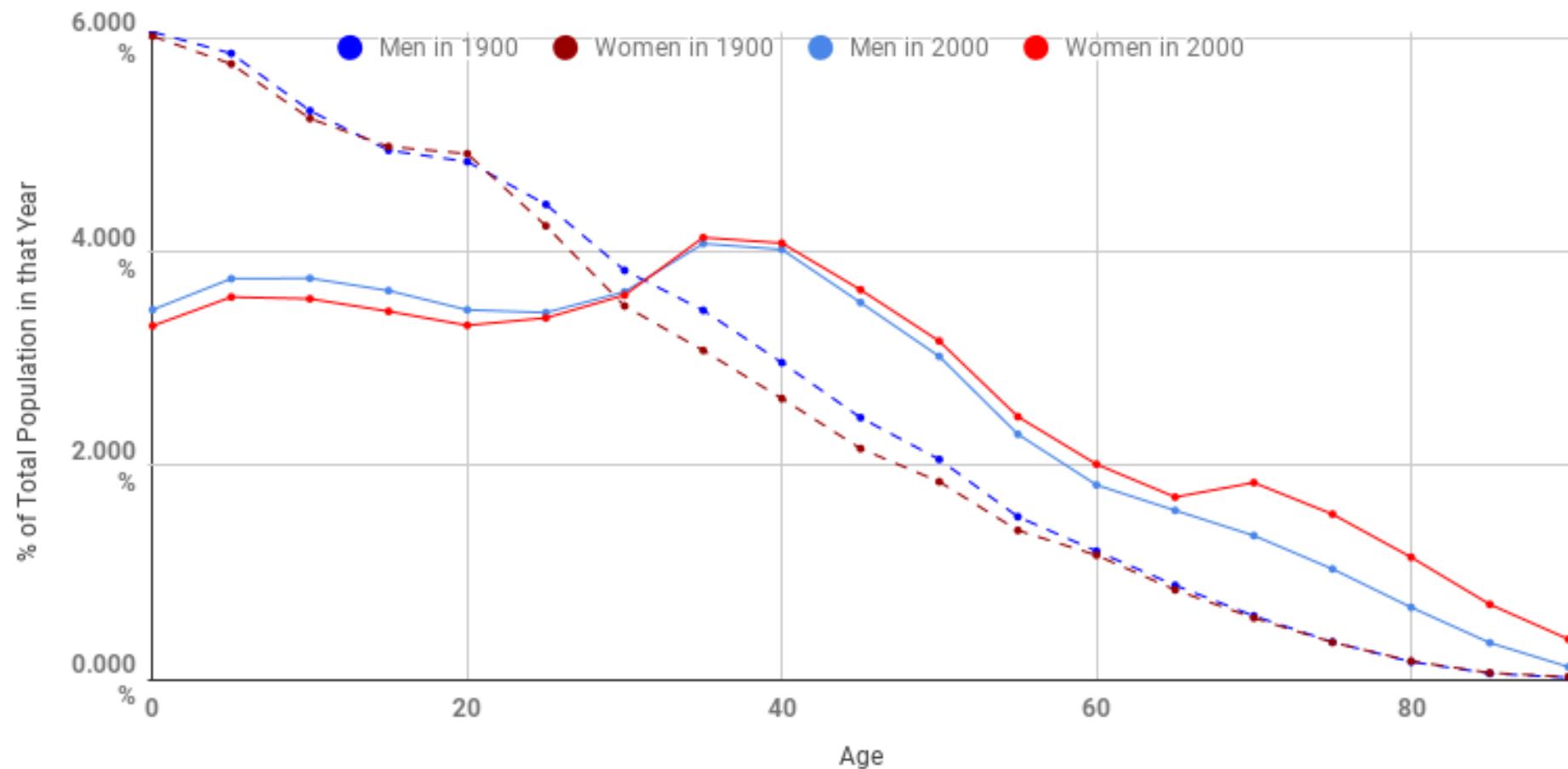
How does the work force in 1900 compare to the one in 2000?



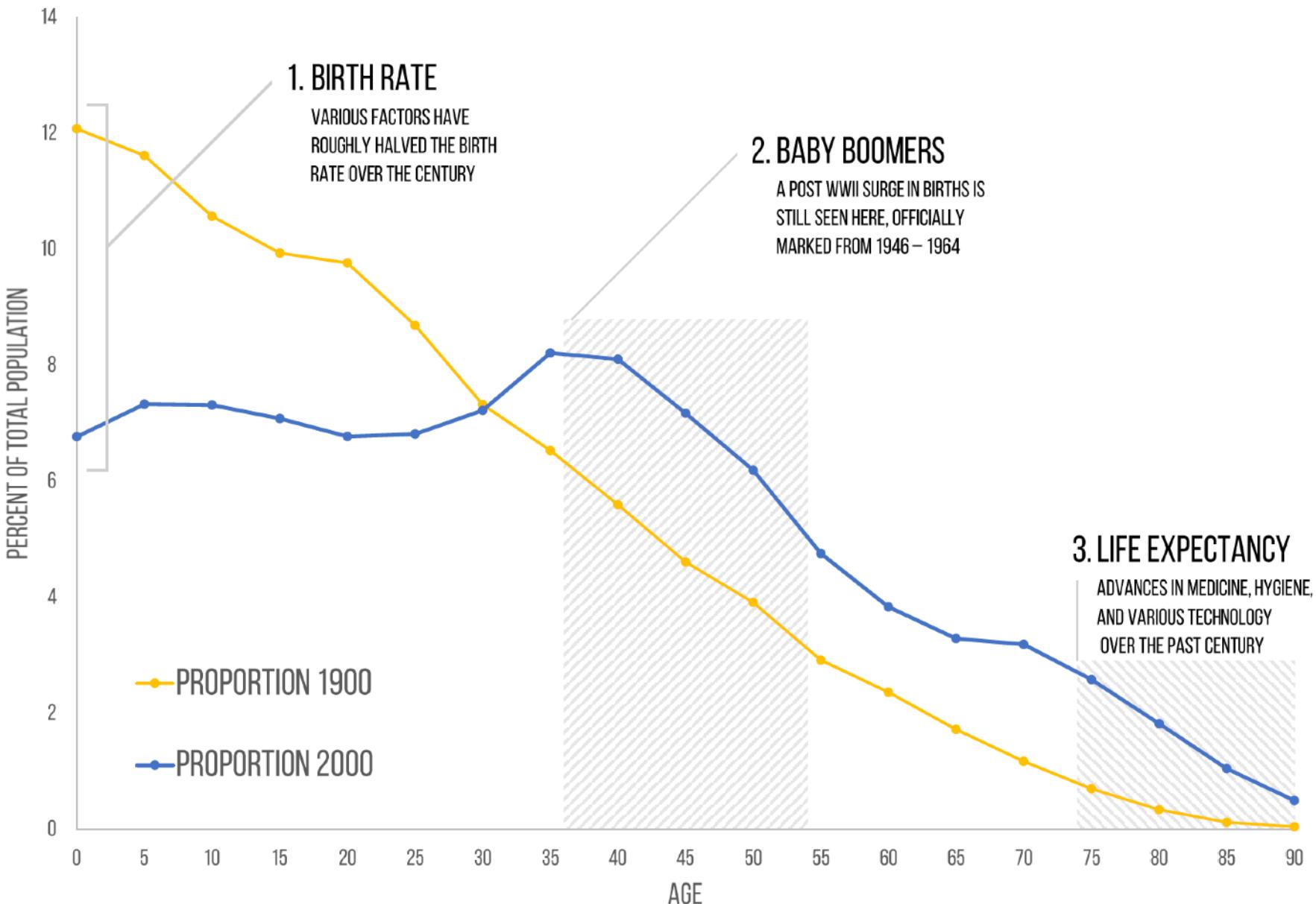
Lines / Area

Distribution of the U.S. Population

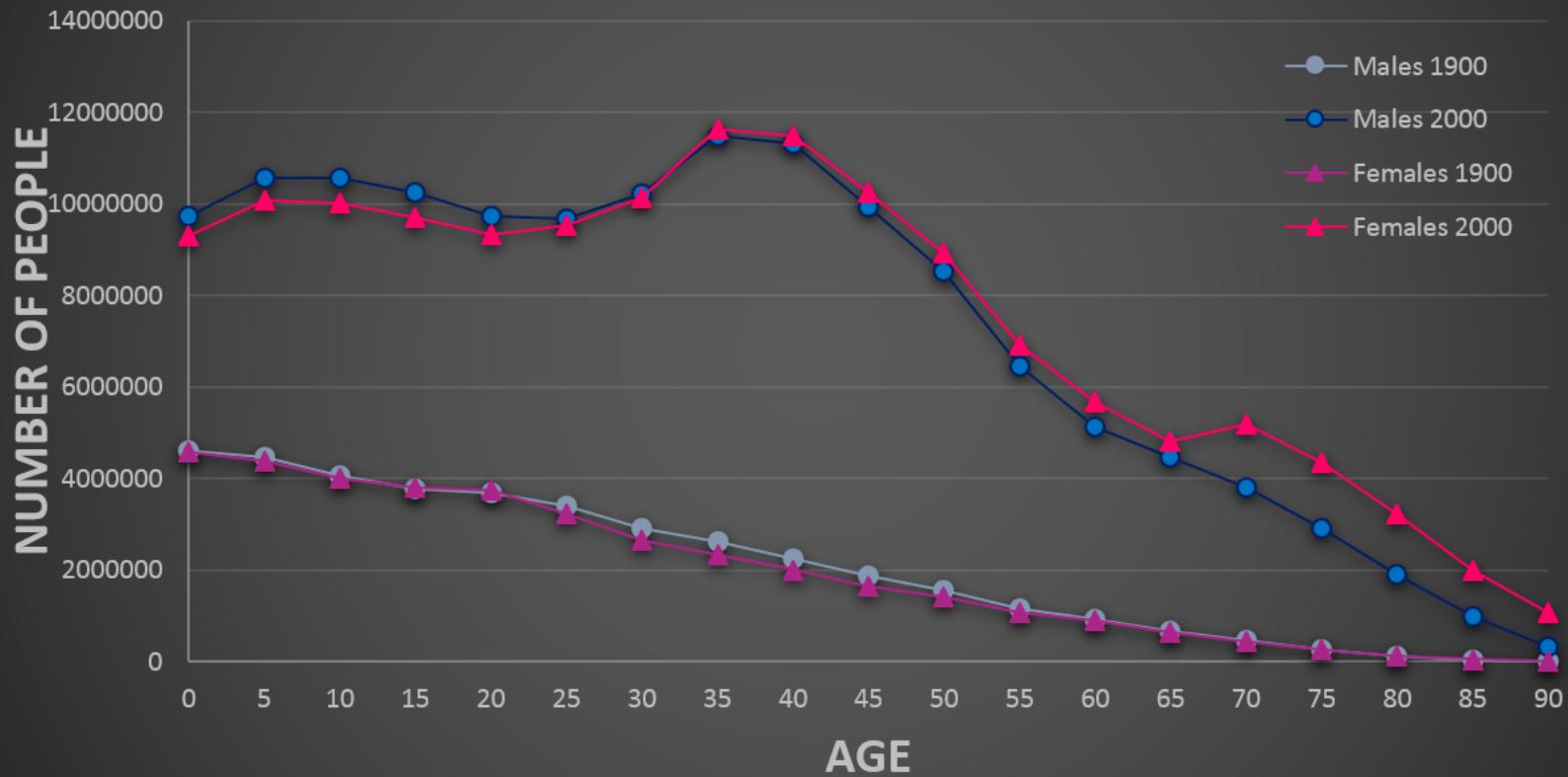
1900 vs 2000



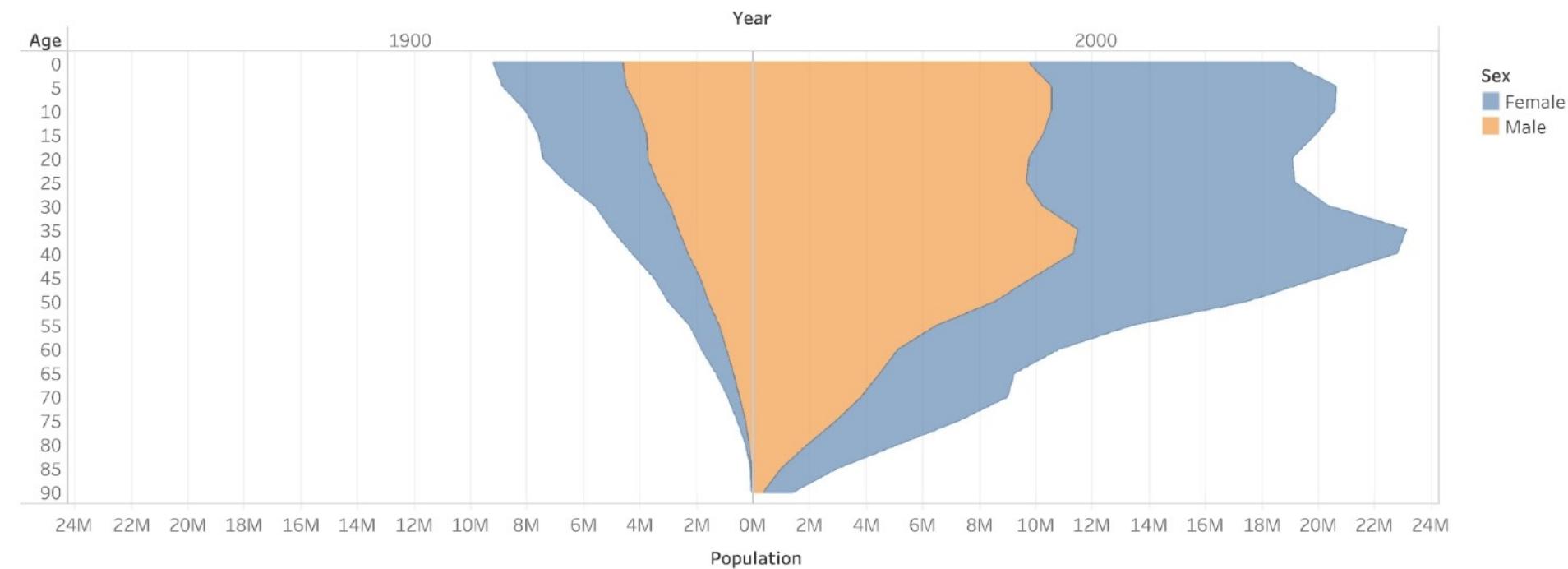
HOW HAS THE POPULATION DISTRIBUTION CHANGED OVER 100 YEARS?



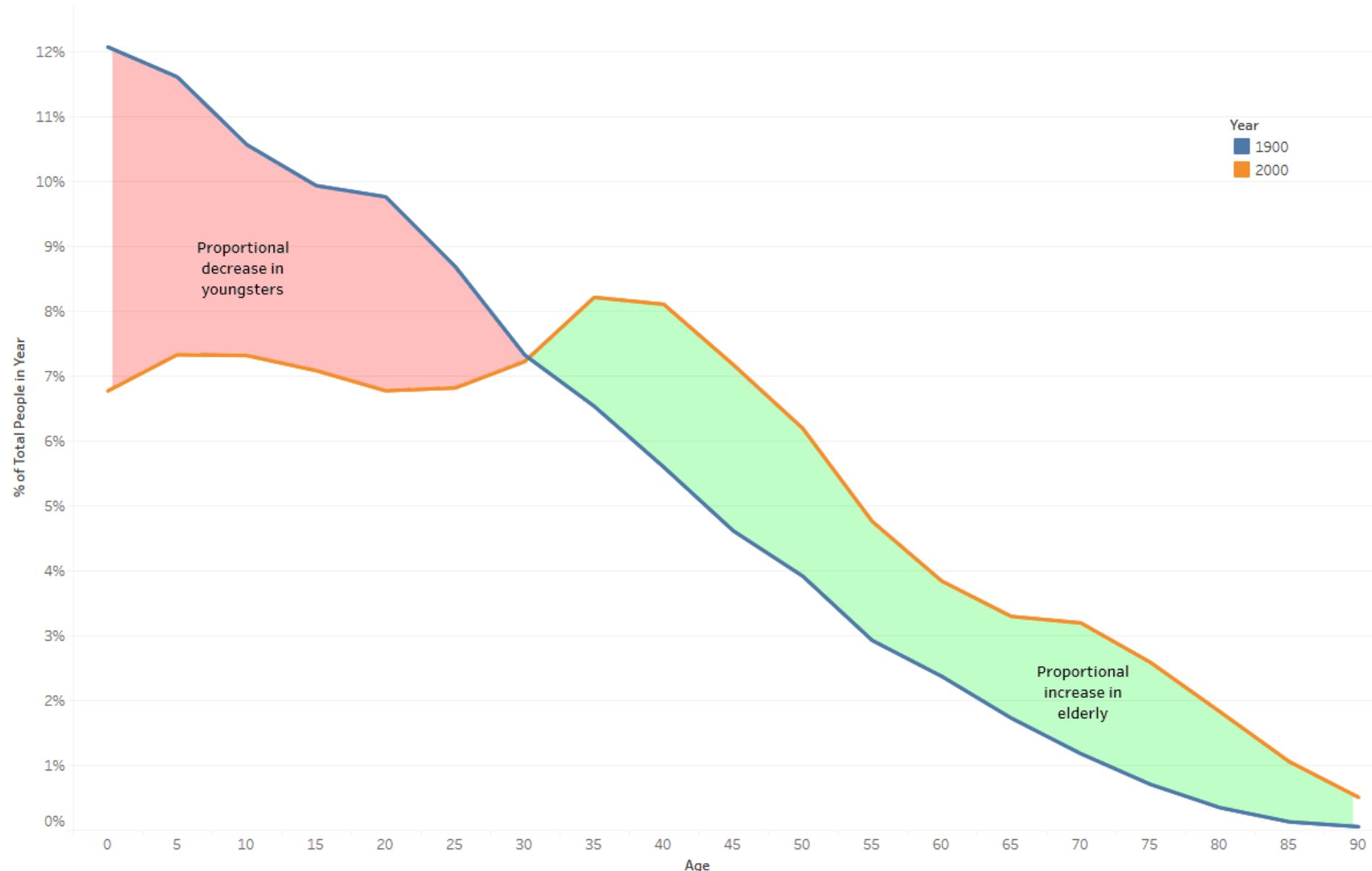
Do women outlive men in the 2000's vs 1900's?



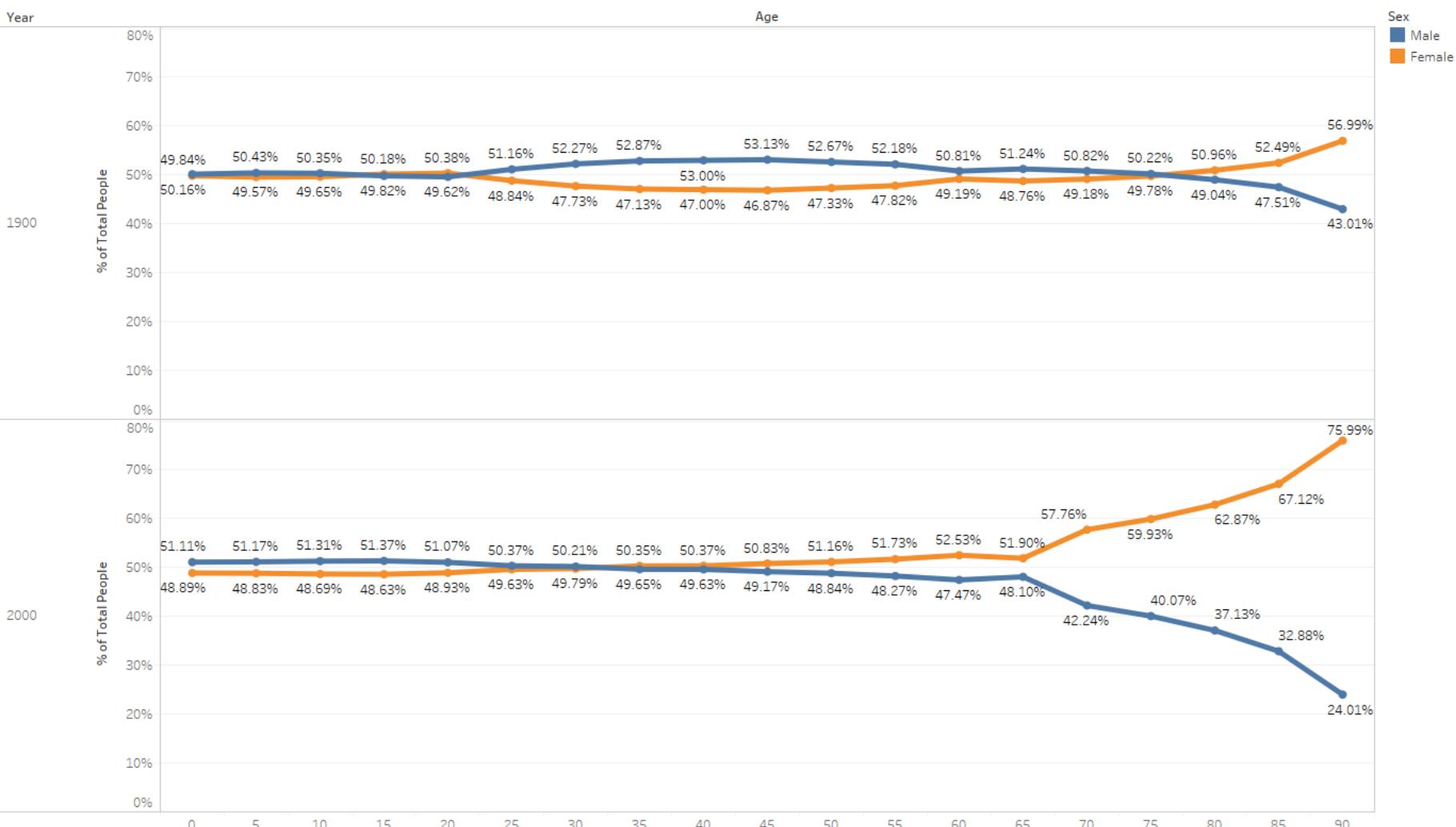
U.S. Population, 1900 vs. 2000



How has the proportion of youngsters (0-20) and elderly (60+) in the United States population changed from 1900 to 2000?

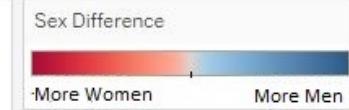


How did WWII affect US male population?

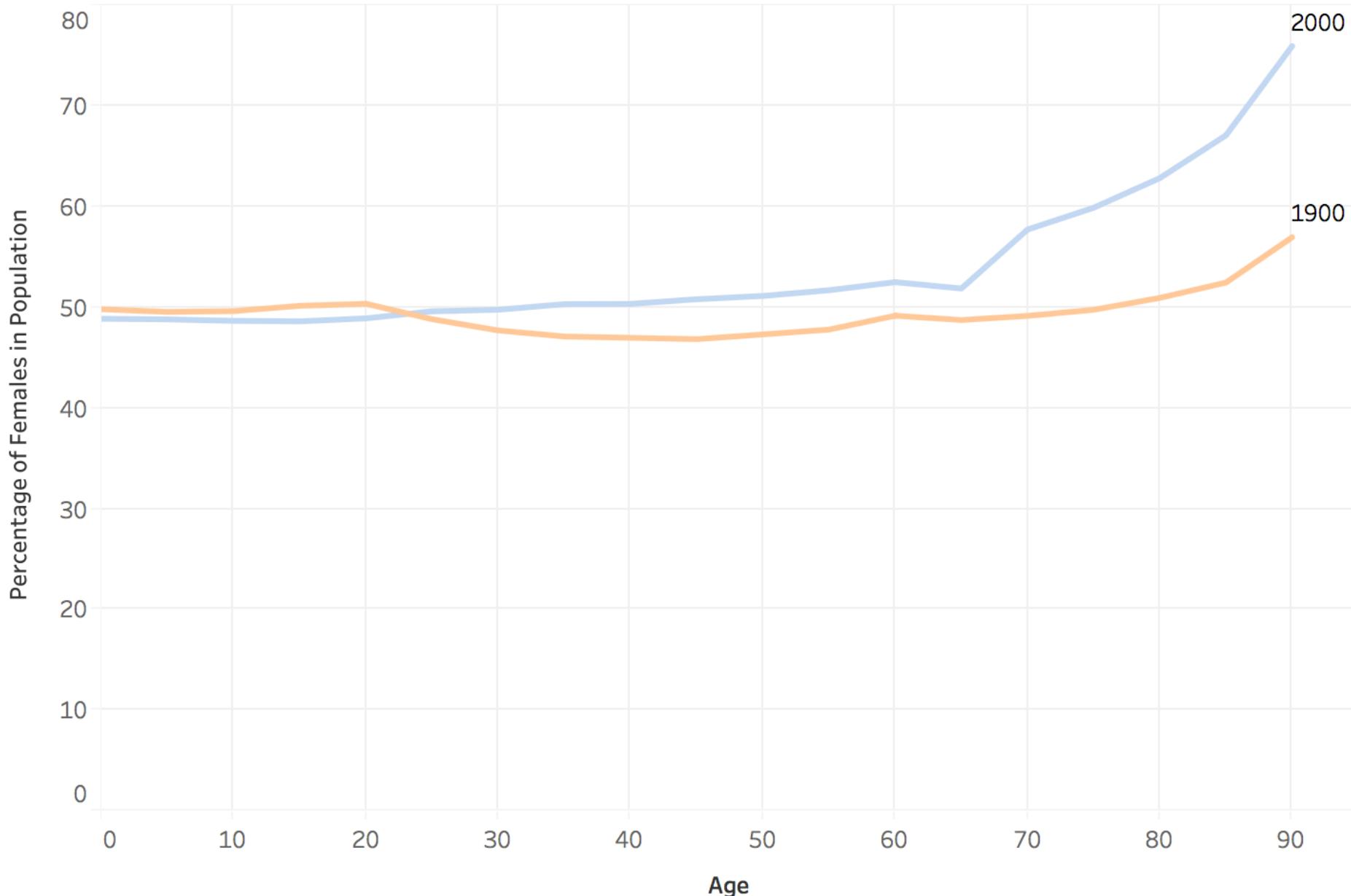


The trend of % of Total People for Age broken down by Year. Color shows details about Sex.

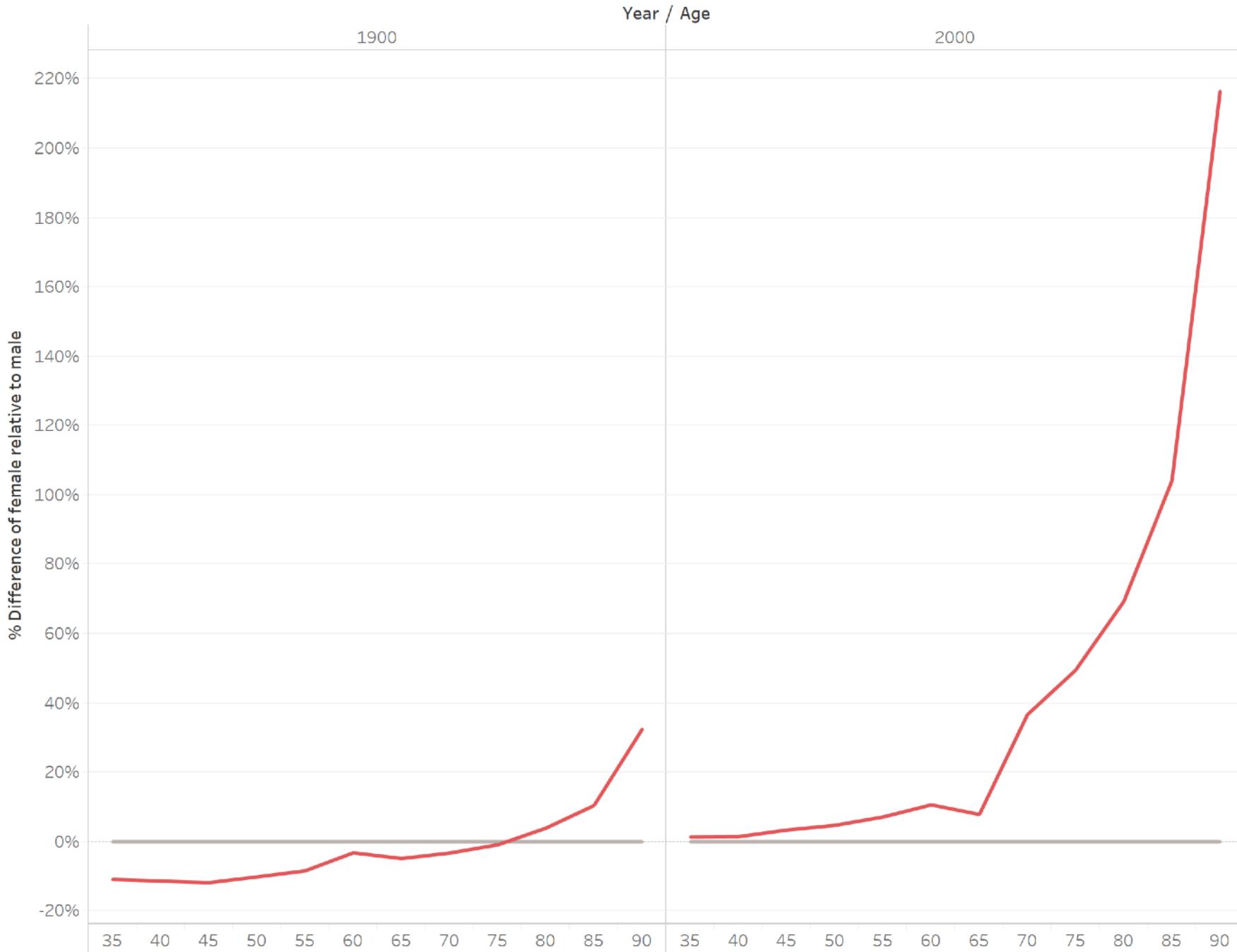
How did gender population differences by age group change over time?



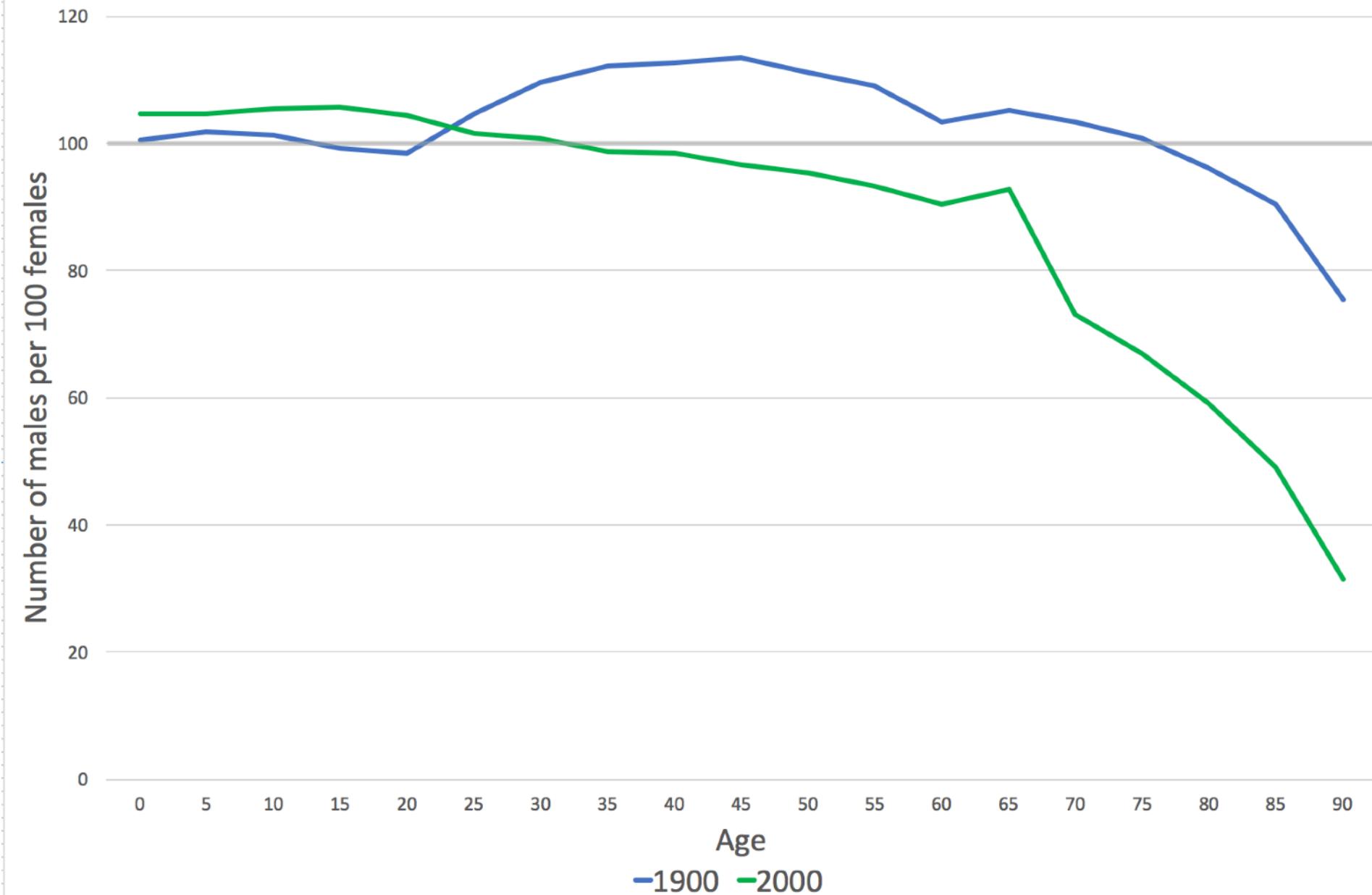
Percentage of Females in the Population Per Age Group and Census Year



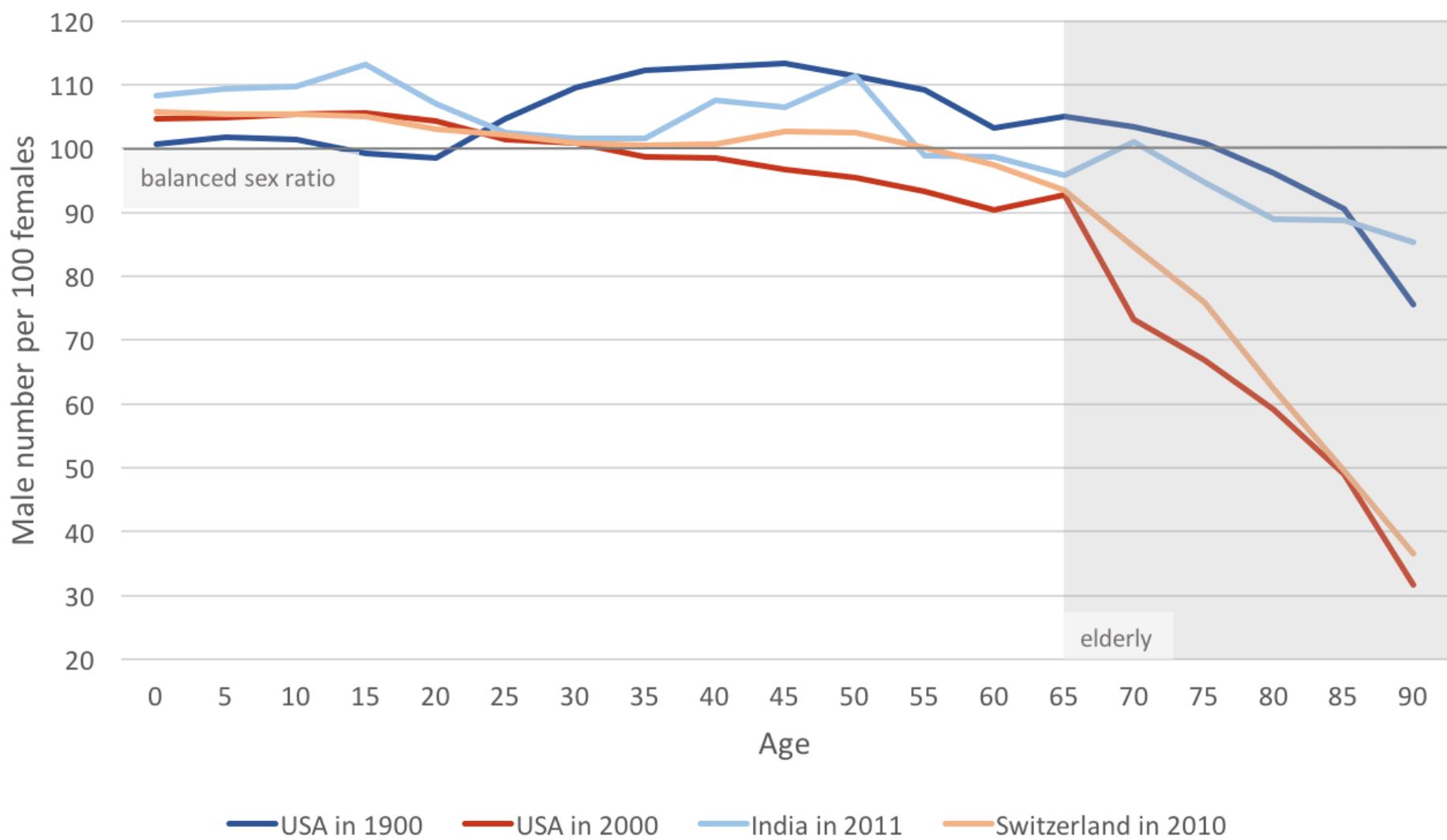
Does female out-live male even more now than 100 years ago



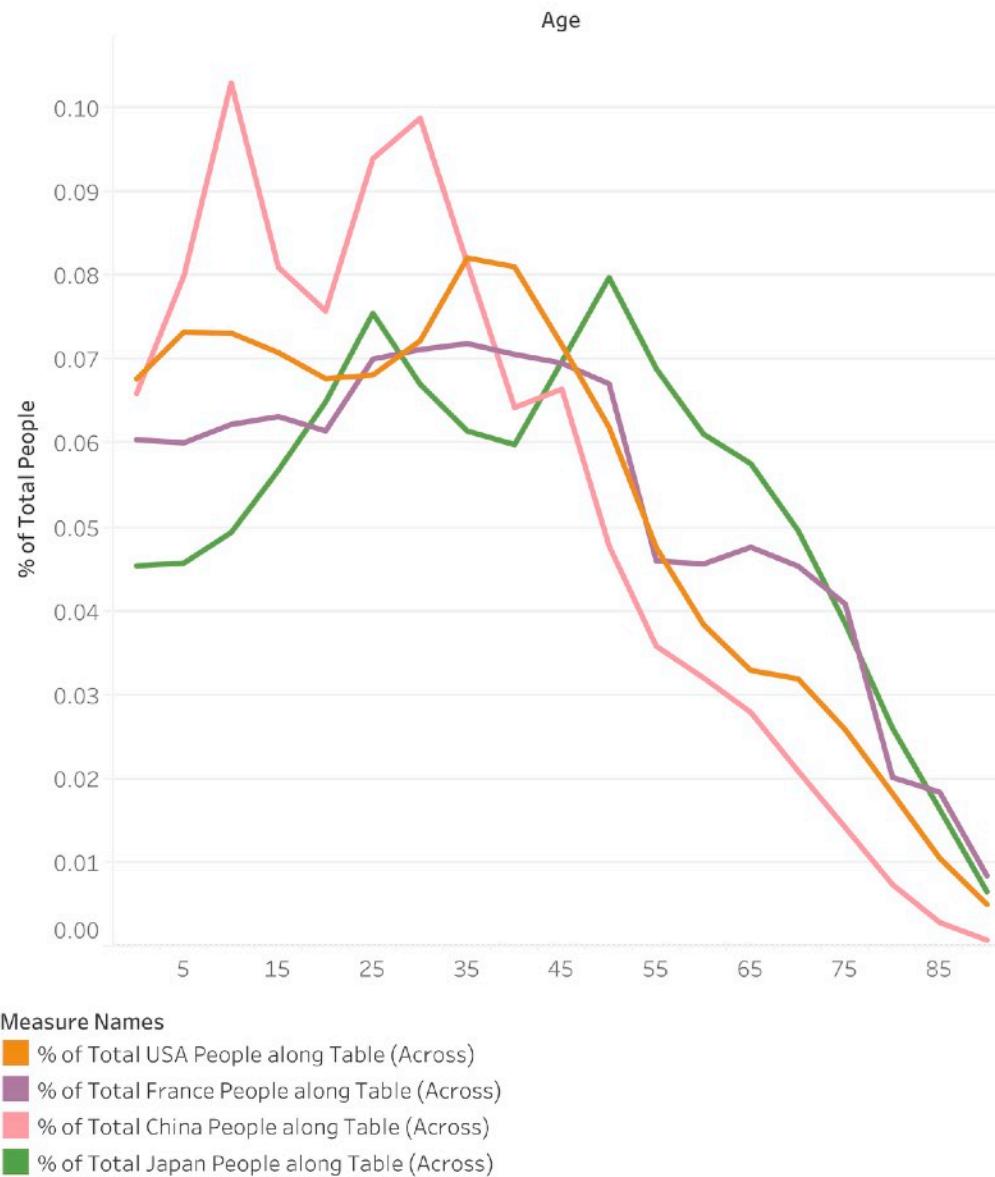
What is the Sex Ratio (number of males per 100 females) for 1900 & 2000 at Different Ages?



Investigate The Correlation between Trends of Sex Ratios and Level of Social Development



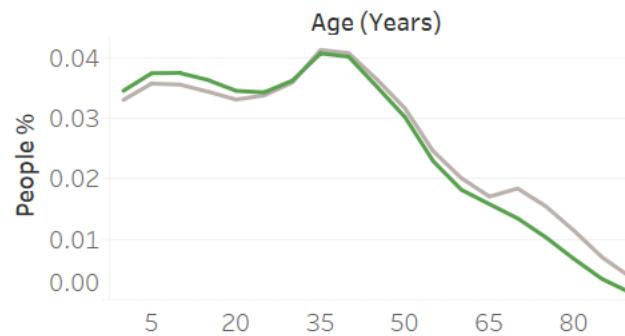
Did other countries experience the same "baby boom" that the US did?



People % vs. age comparing male & female: Year 1900



People % vs. age comparing male & female: Year 2000

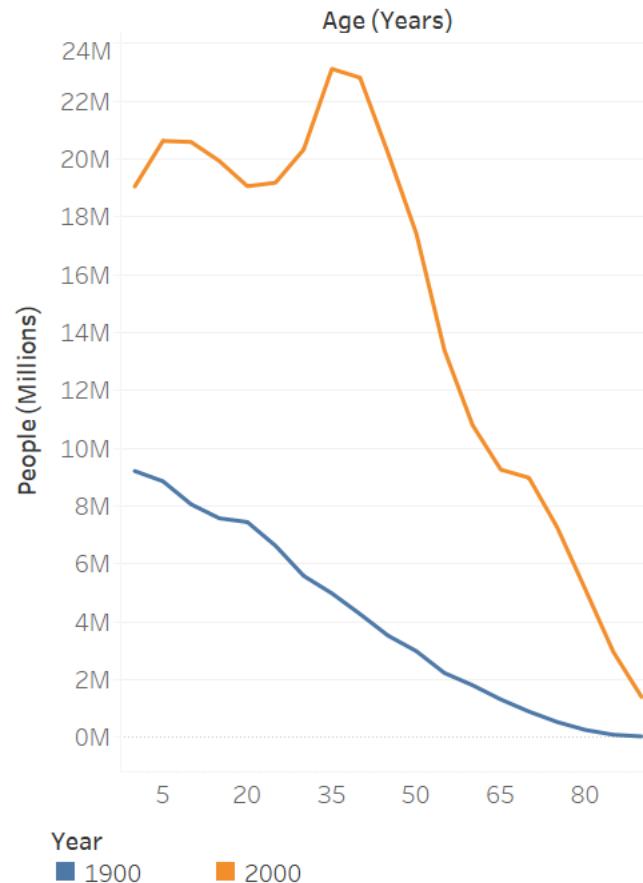


Sex

■ Male

■ Female

People vs age compared from 1900 and 2000



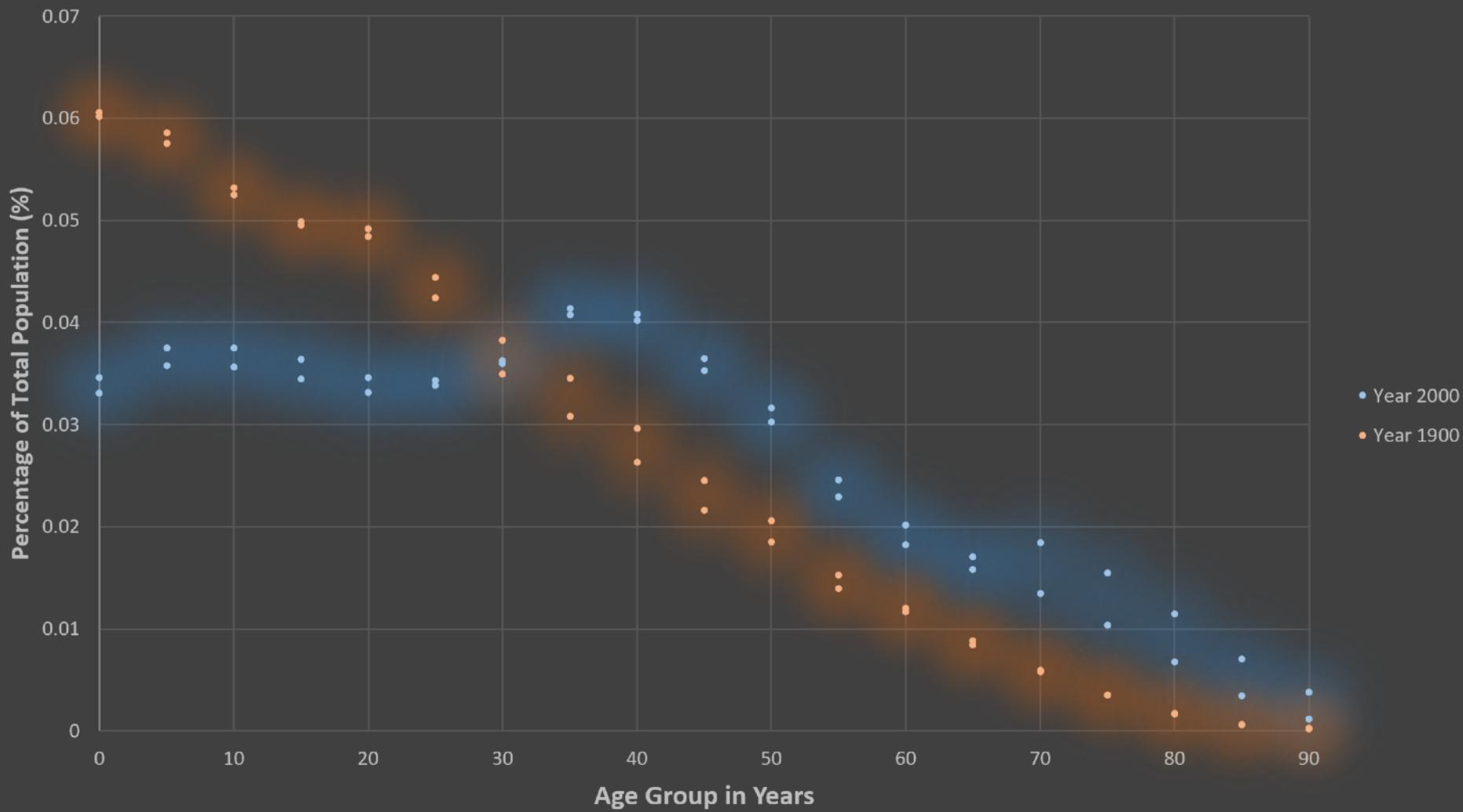
Year

■ 1900

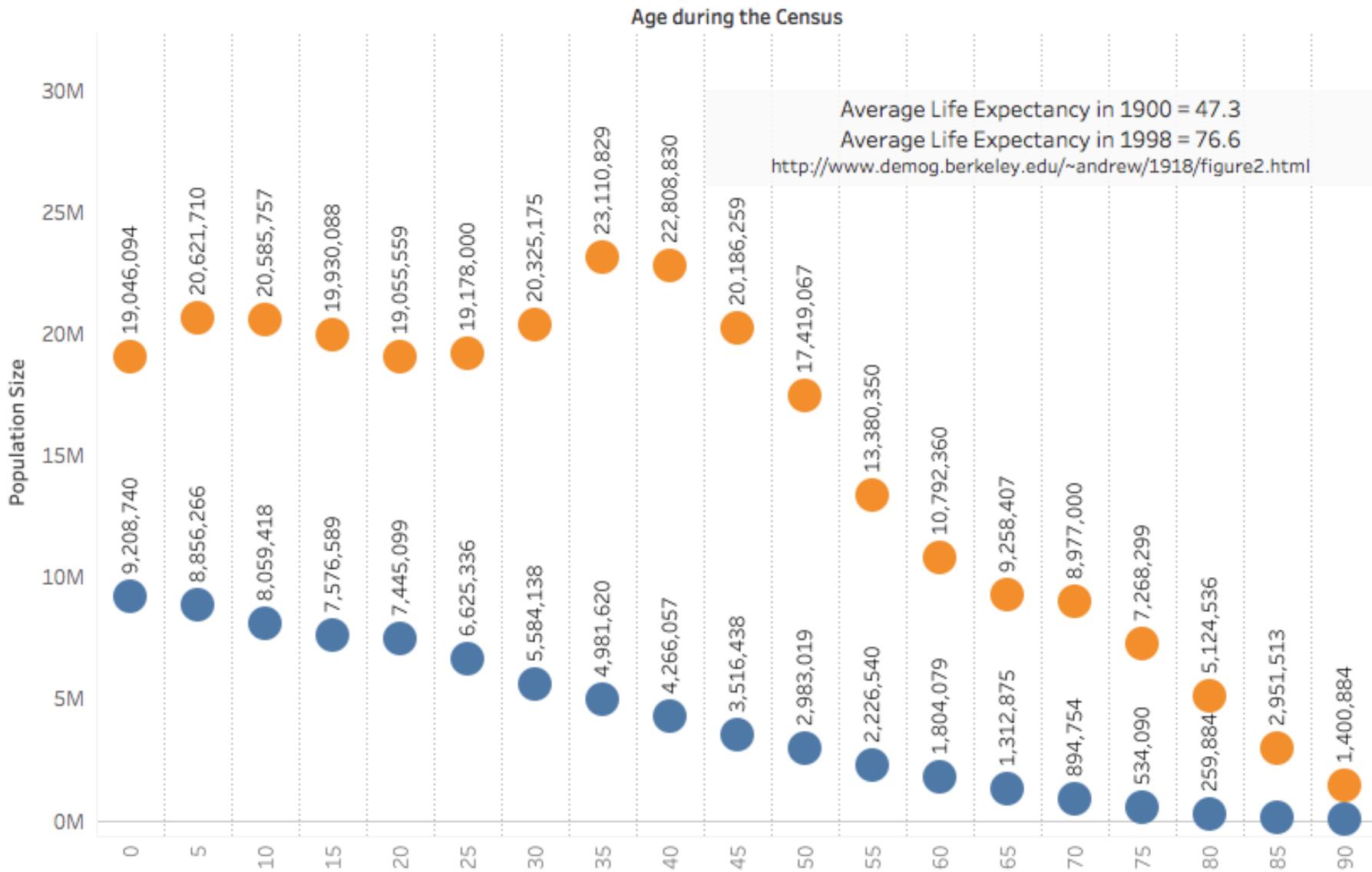
■ 2000

Dot / Scatter Plots

How Did America's Age Distribution Change Between 1900 and 2000?



What's the age of America's population?



Sum of Population Size for each Age during the Census. Color shows details about Year. The marks are labeled by sum of Population Size.

Year

1900

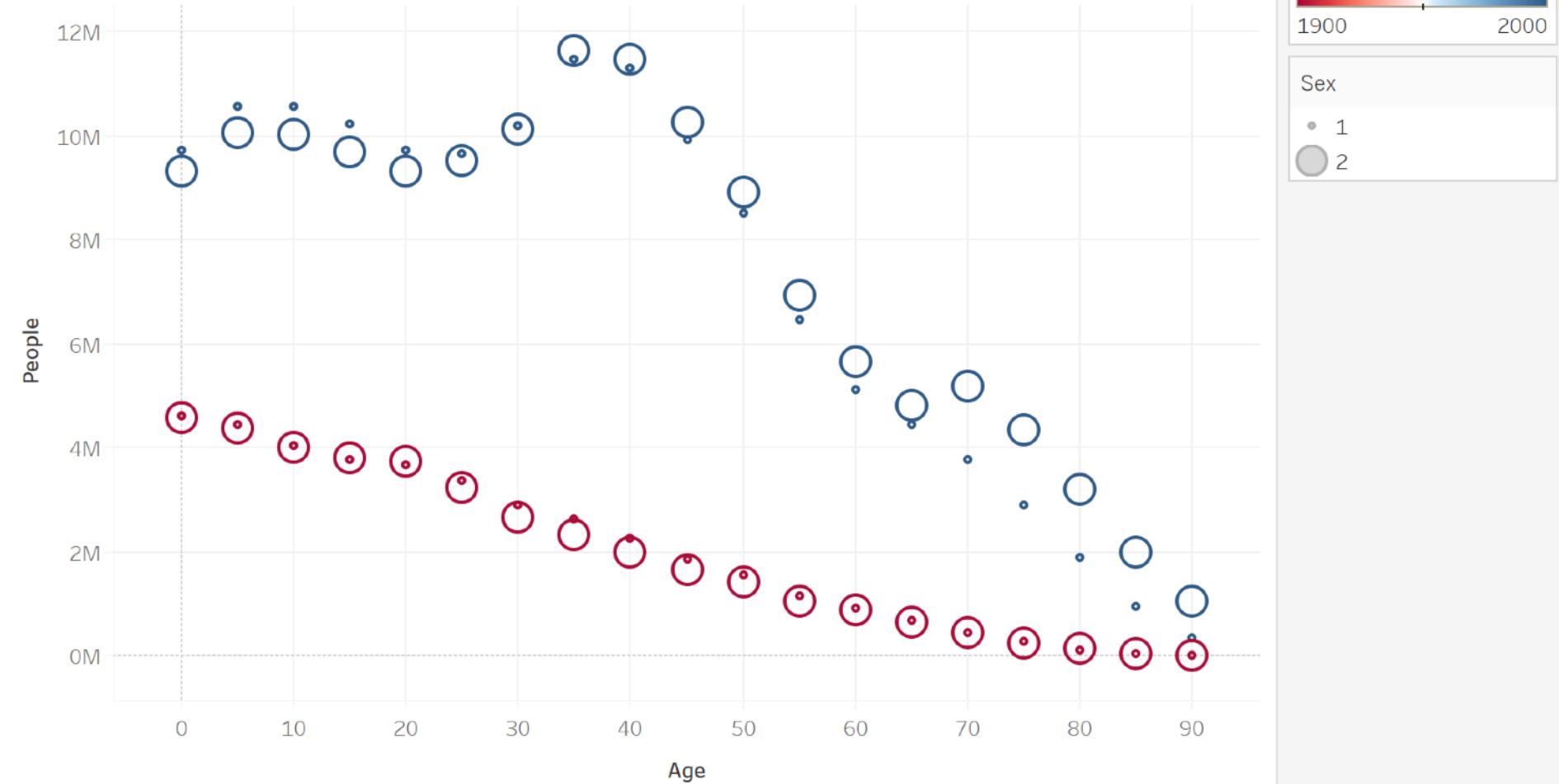
2000

Average Life Expectancy in 1900 = 47.3

Average Life Expectancy in 1998 = 76.6

<http://www демог.berkeley.edu/~andrew/1918/figure2.html>

Gender Disparity Between Age Groups in 1900 vs 2000



Is it possible that the support for same-sex marriage resulted from a gap between men and women?

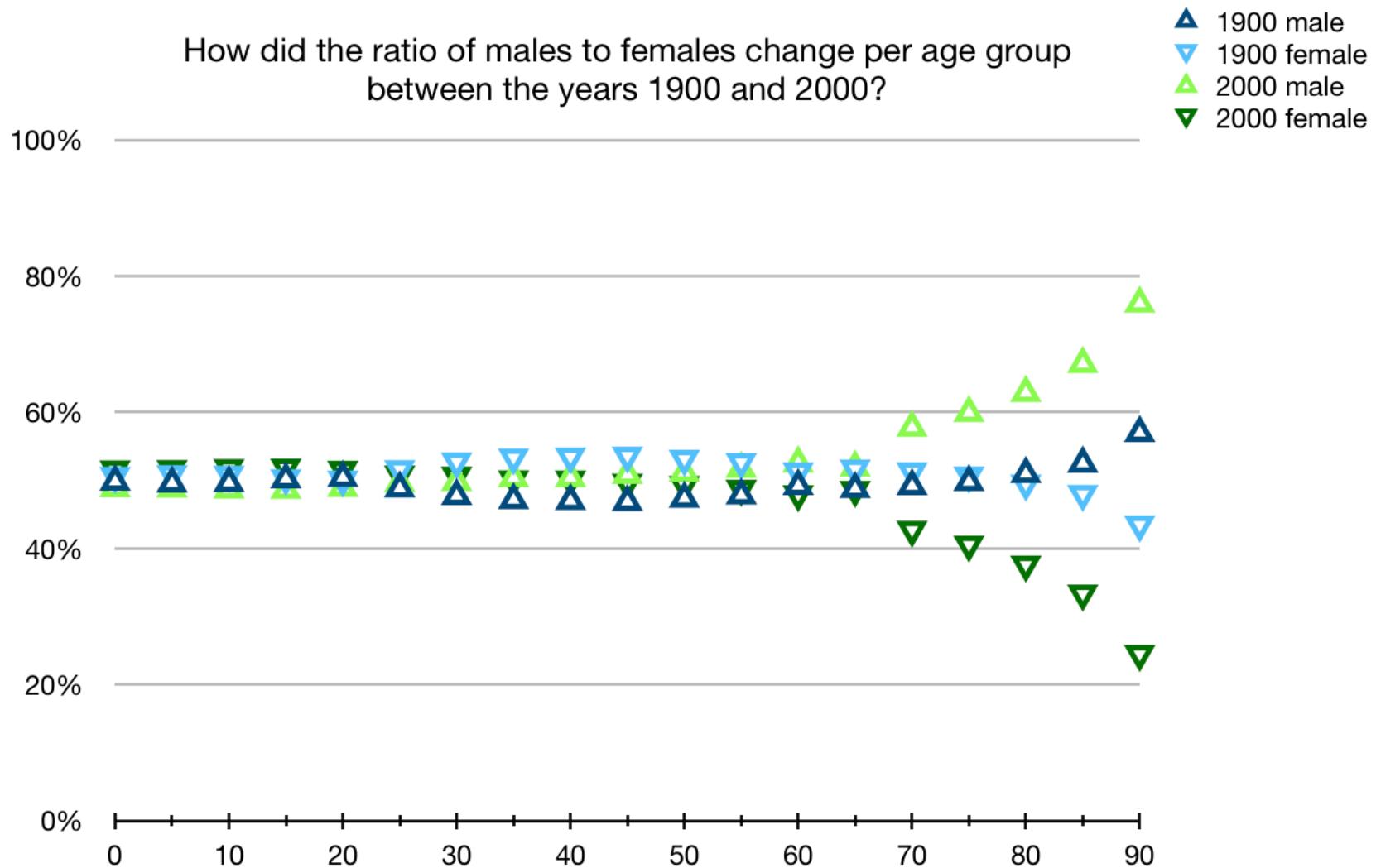


Sum of People for each Age broken down by Year. Color shows details about Sex. Size shows sum of ABS-Diff (sex difference).

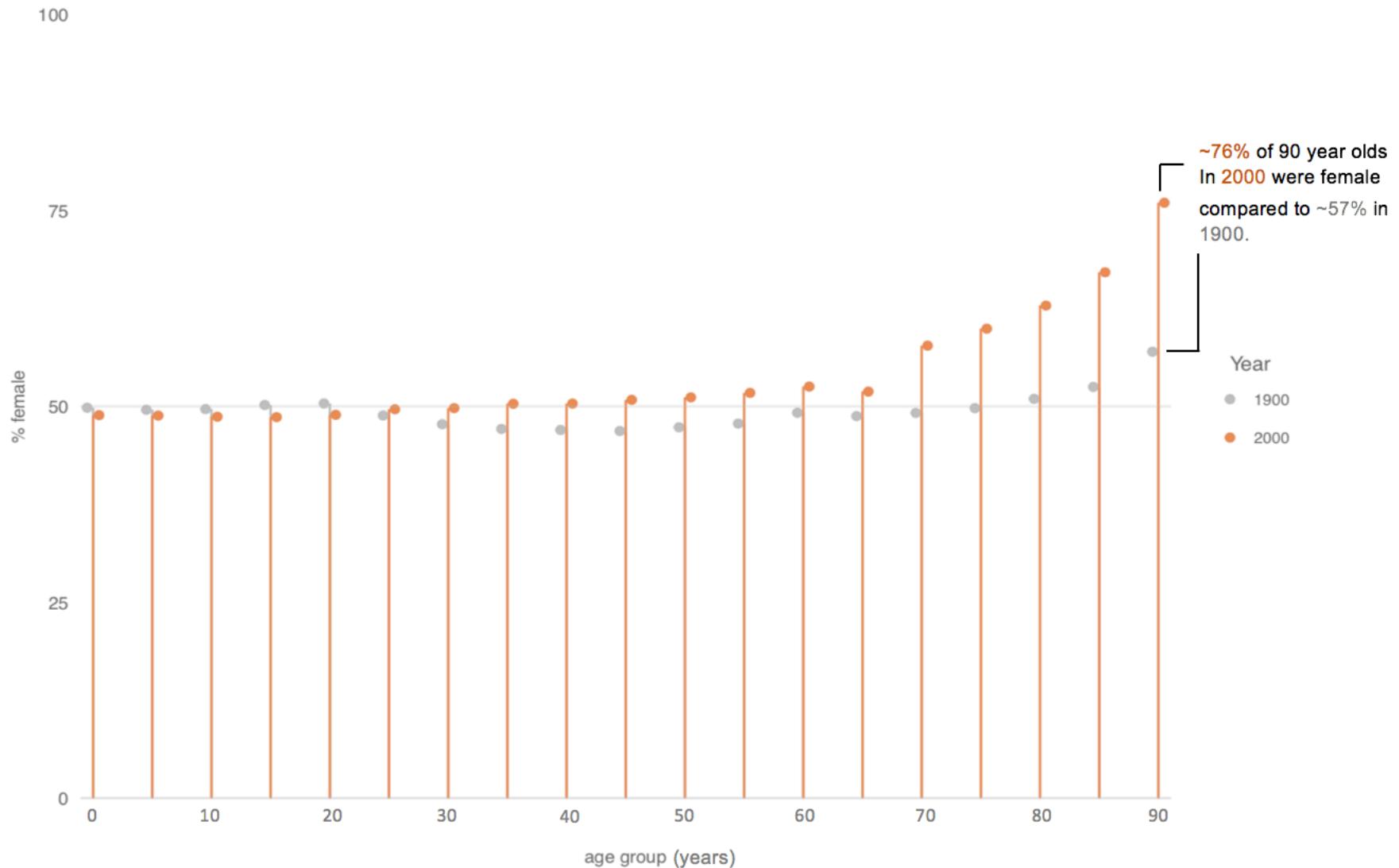
How has the male and female ratio changed over the century for each age group?



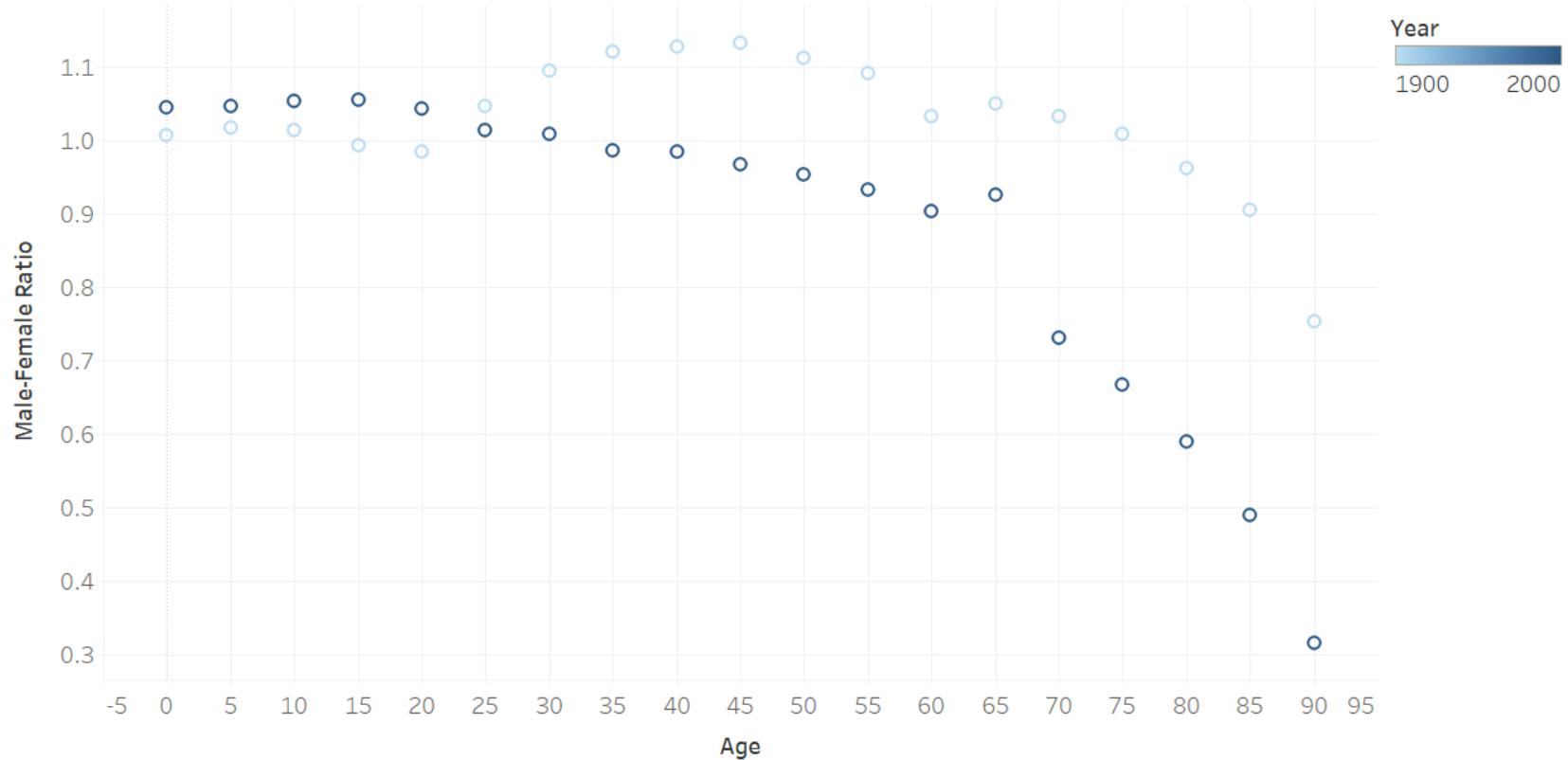
How did the ratio of males to females change per age group between the years 1900 and 2000?



How Does the Total US Population Represented by Females for Each Age Group in 2000 compare to 1900?



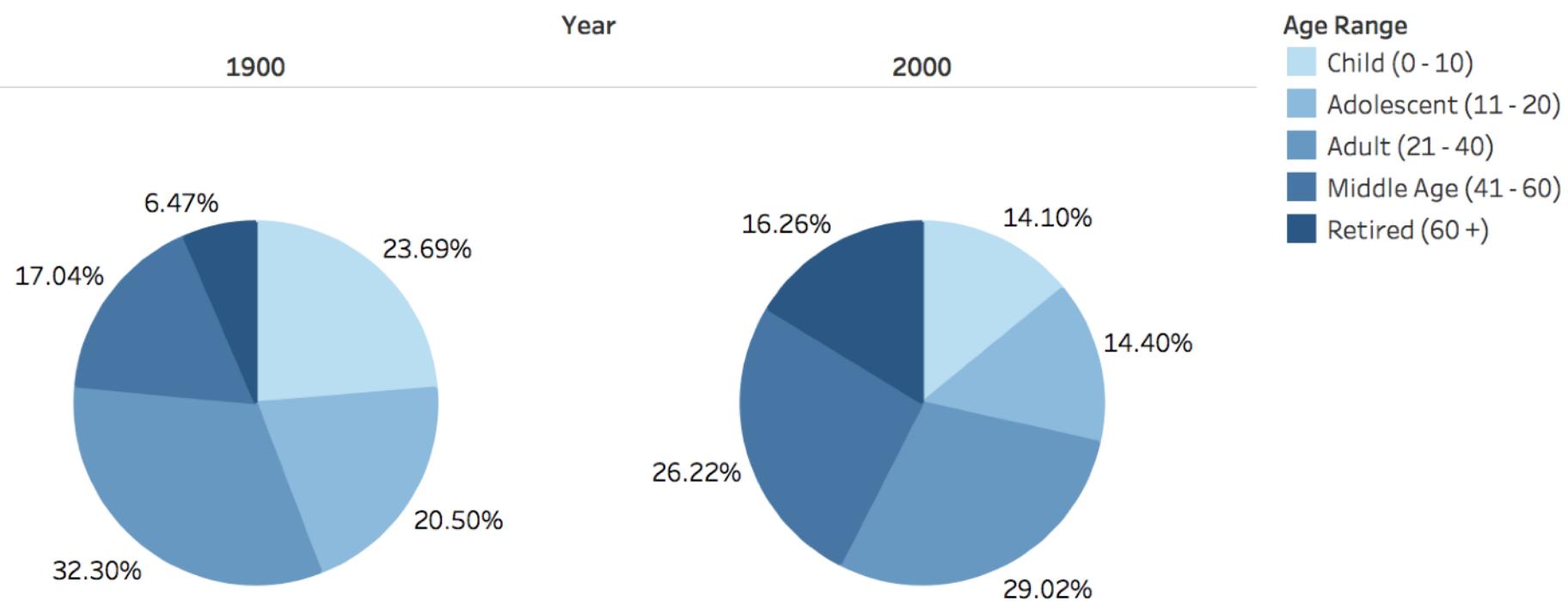
The Decreasing Male-Female Aging Ratio (based on 1900 and 2000 US Census Data)



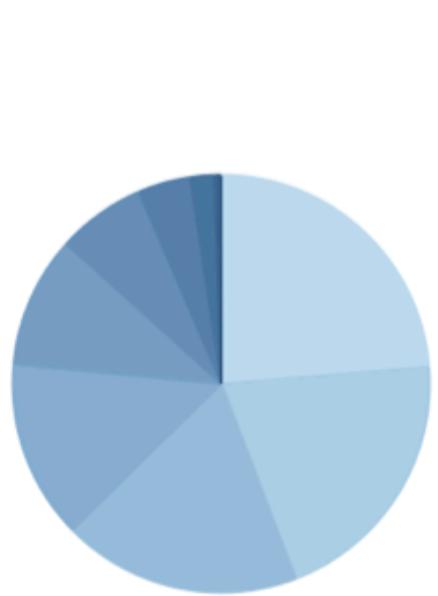
The plot of sum of Age for Male-Female Ratio. Color shows details about Year.

Circular Charts

Change in the Proportions of Different Age Groups in the Population from 1900 to 2000



How did the age composition of the U.S. population change over the twentieth century?



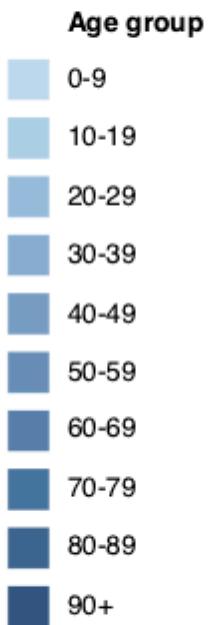
Year: 1900

Total population: 76,262,821

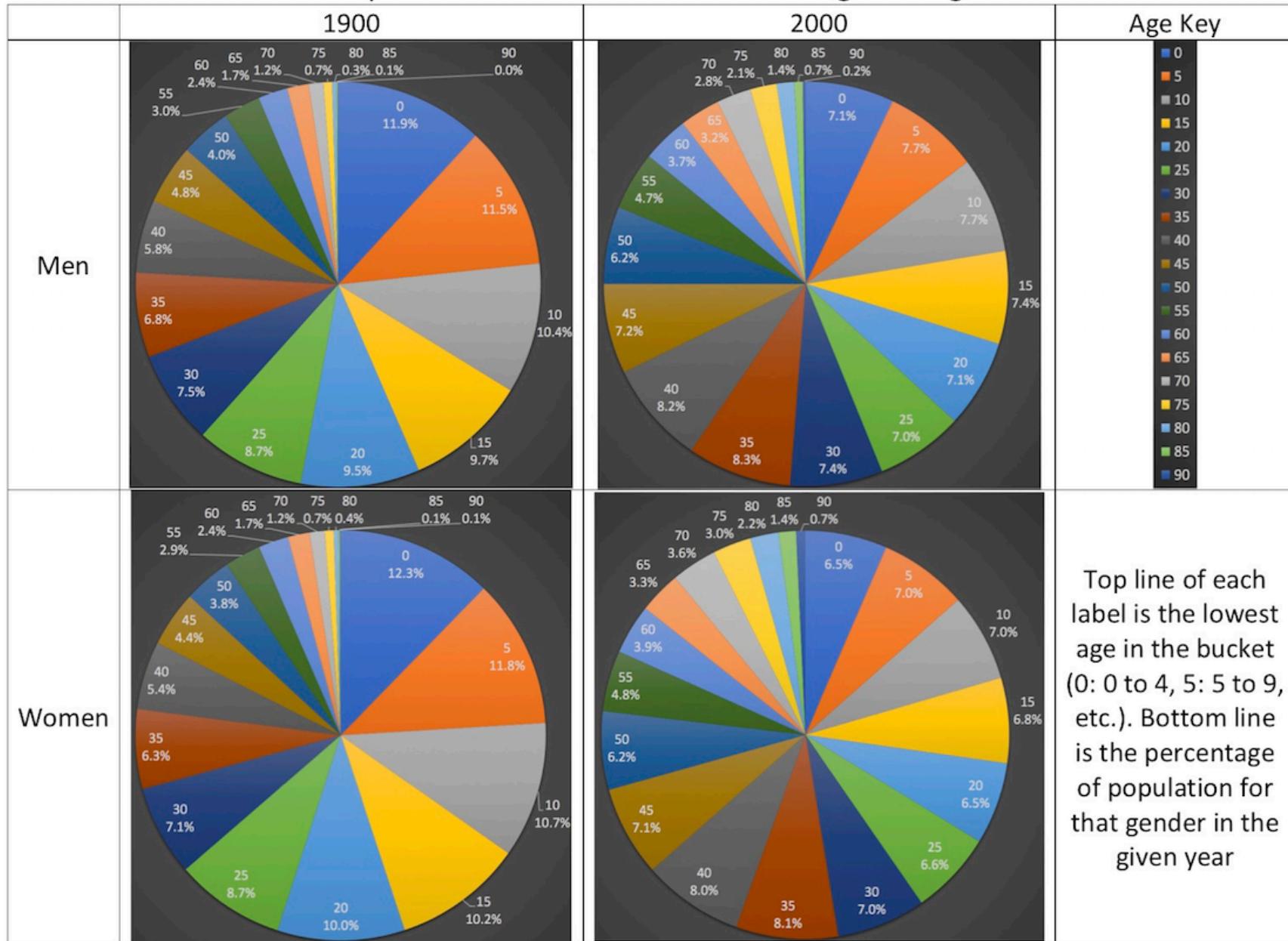


Year: 2000

Total population: 281,420,717

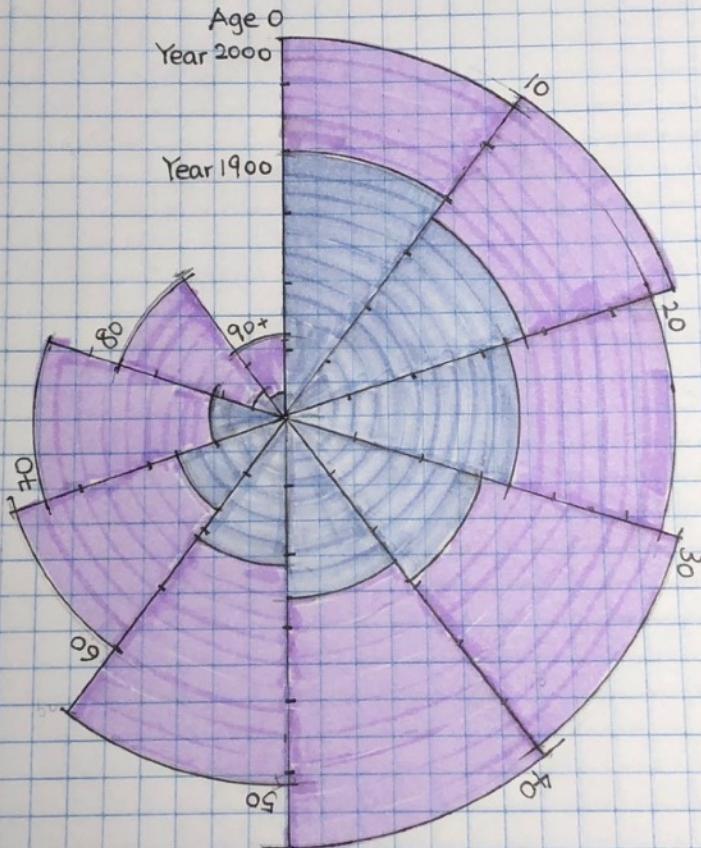


How Did Gendered Population Distribution Based on Age change from 1900 to 2000



A Century Apart

U.S. Population Age Composition and Growth



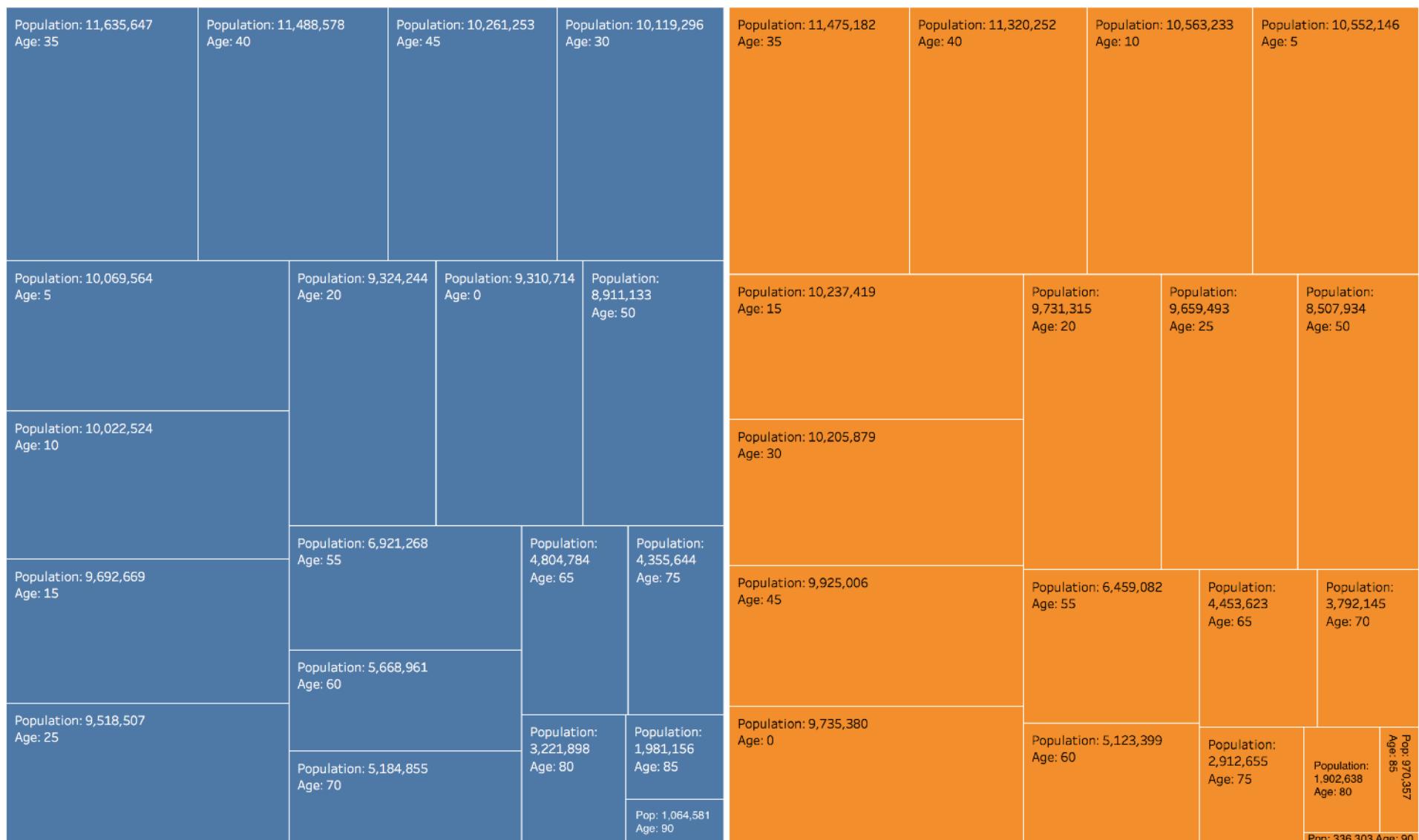
■ ~ 0.95 Million PPL

Year 2000 Total : 281,420,717 ppl

Year 1900 Total : 76,262,821 ppl

Other!

How does the 2000 U.S. population break down in terms of age and sex?

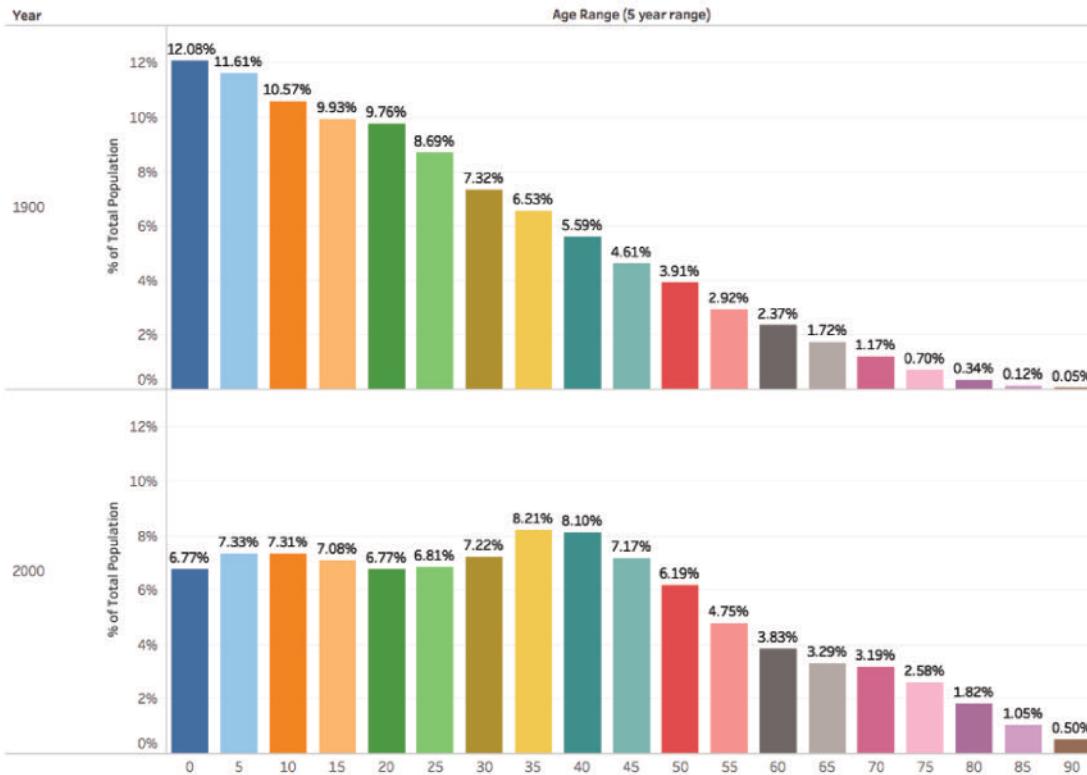


Sex

- Female
- Male

Was the United States younger in 1900, or in 2000?

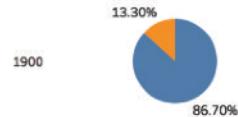
Percentage of U.S. Population distributed amongst 5-year Age Ranges



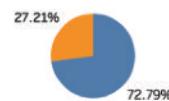
Population Percentage
of Youth (0-50)
compared with Seniors
(50+)

Age (group)
0-50
50+

Year



2000



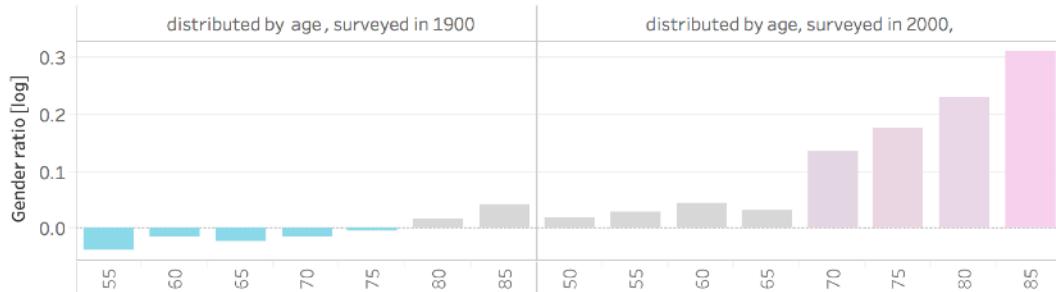
Did wars affect the ratio of women to men in the USA?

Gender ratio (# women / # men, not log)

0.882

2.298

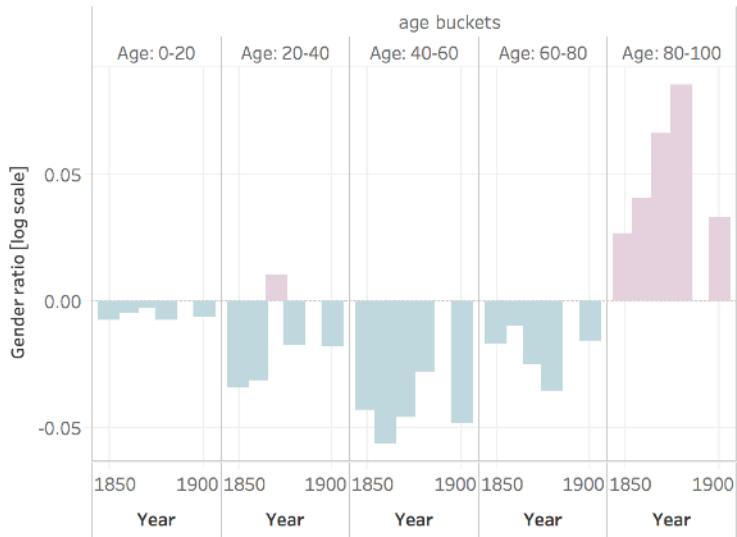
Female-male ratios across ages, as surveyed in 1900 and 2000



The Civil War took place in 1861-1865. Veterans of that war must've been at least 60 during the 1900 census. We see a higher female-male ratio for ages 60+ than for 55.

World War I took place in 1914-1918, the Korean war 1950-1953, and the Vietnam 1955-1975. Veterans of those wars must've been at least 70 by the 2000 census. We see a higher female-male ratio for ages 70+.

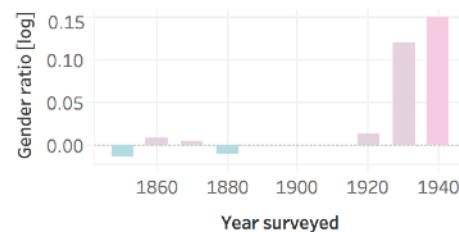
Gender ratio across the 19th century, bucketed by age groups



At high ages (80-100) do females become the majority. This implies biological lifespan accounts for a high female-male gender ratio.

The 1870 / Age: 20-40 is the only other point with more women than men. It is likely due to the civil war (estimated death toll: 600,000, armies: majority male)

People born in 1835-1844



The average soldiers of the civil war and WWII were ~26 years old. So people born in 1835-1844 capture the 'generation' of civil war soldiers, and people born in 1915-1924 capture the 'generation' of WWII soldiers.

People born in 1915-1924

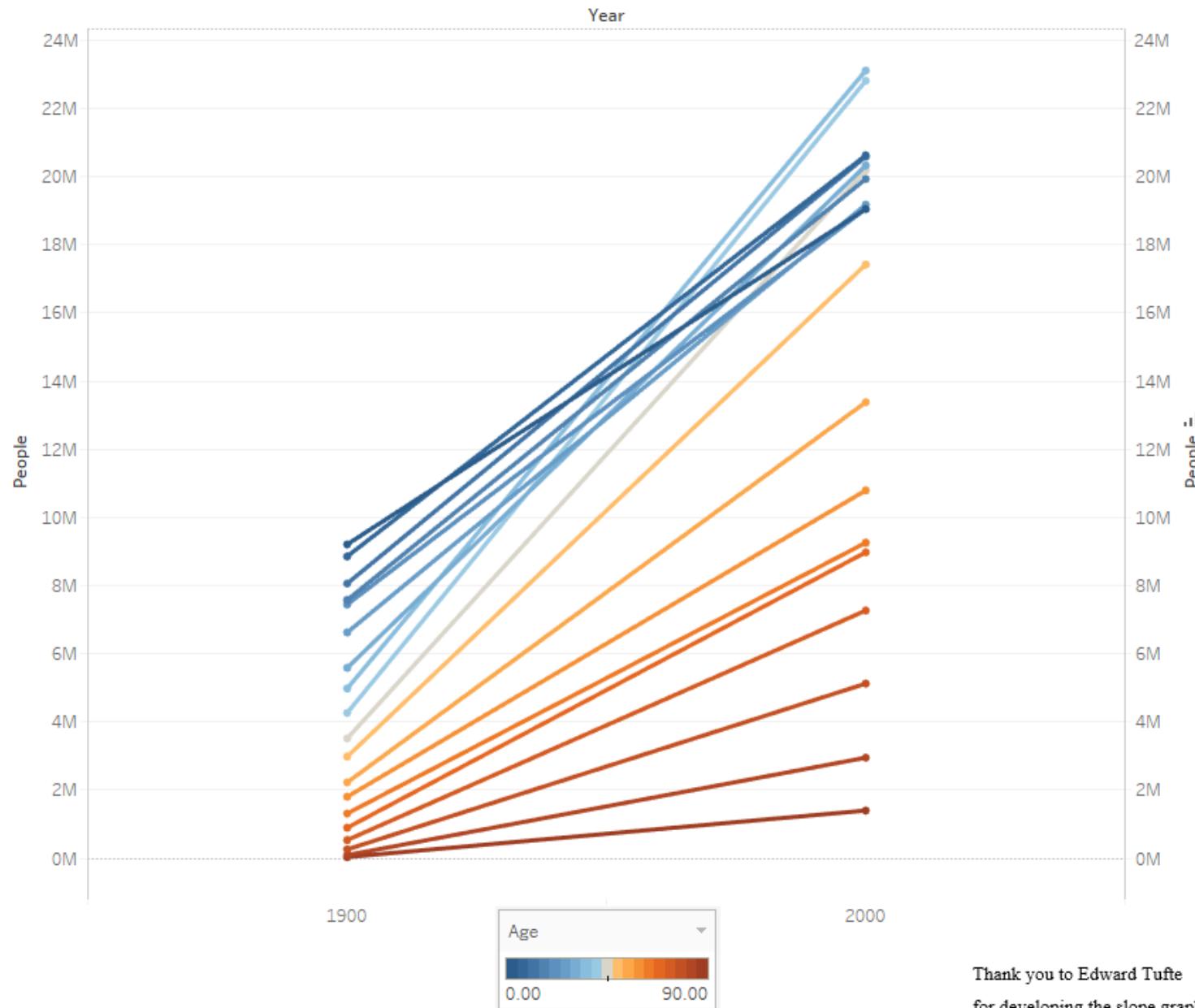


We see jumps in female-male ratios corresponding to the civil war and WWII timeframes, but the highest female-male gender ratios appear when the soldier generations become old.

In conclusion: The five wars with the highest American casualty counts took place in the 19th and 20th centuries. Wars seem to have an effect on gender ratios, although not as dramatically as biological lifespan.

How different is the age make-up of the U.S. comparing 1900 to 2000?

1900 vs. 2000: U.S. Population by Age from Two Snapshots

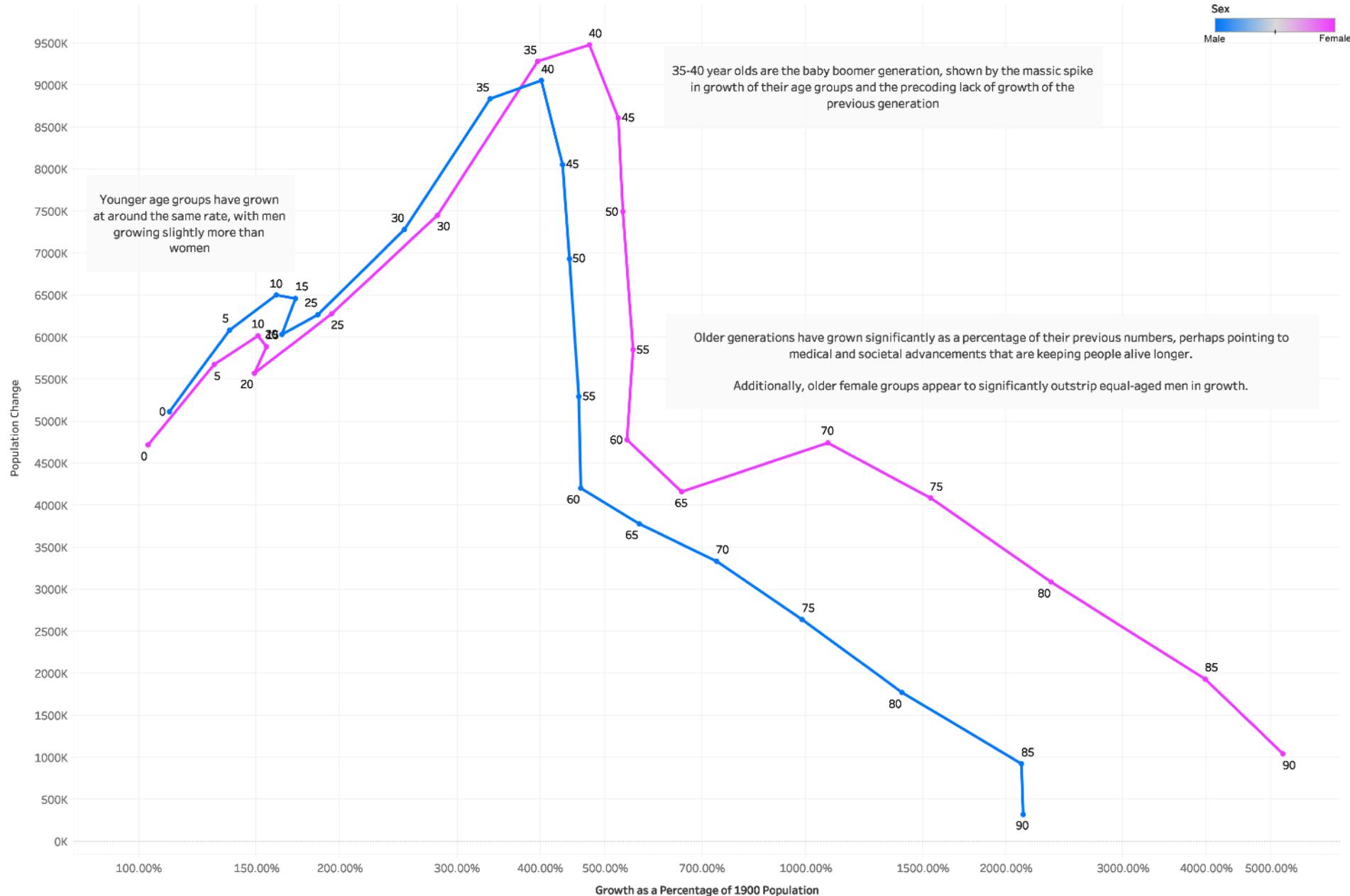


Thank you to Edward Tufte
for developing the slope graph.

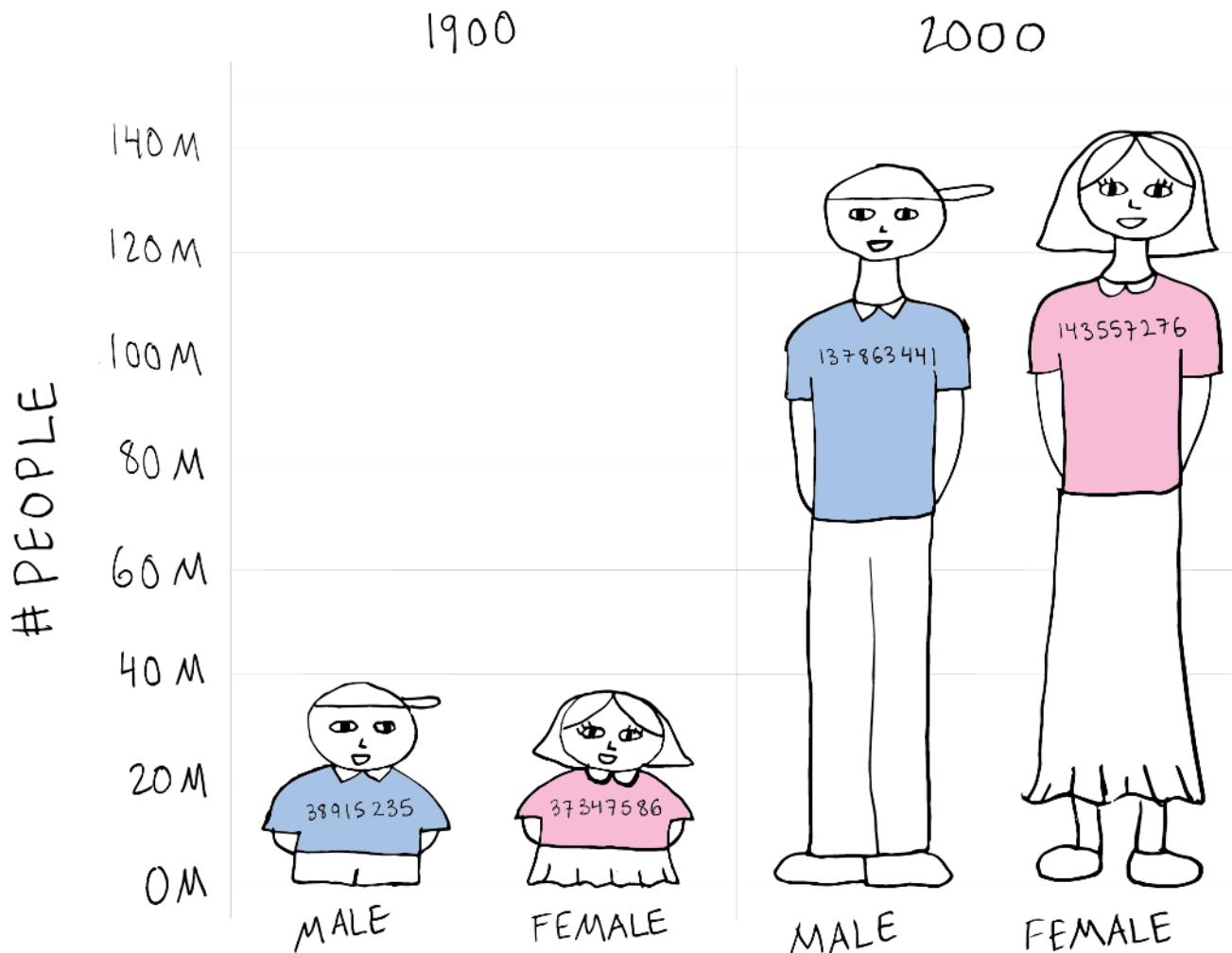
Each color represents a 5-year segment of ages from 0-4 to 90+ years old.

US Population Difference between 1900 and 2000 by Age and Gender

Sex
Male Female



HOW HAS THE POPULATION CHANGED BY SEX IN 100 YEARS?



What Is The Combined Weight Of All US Citizens



Re-Design Exercise

Re-Design Exercise

Task: Analyze and Re-design visualization

Identify data variables (N/O/Q) and encodings

Critique the design: what works, what doesn't

Sketch a re-design to improve communication

Be ready to share your thoughts with the class

Break into groups with those sitting near you
(~4 people per group)

Effectiveness Rankings

[Mackinlay 86]

QUANTITATIVE

Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL

Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL

Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume

Teacher Salaries: Is It Really That Bad?

National and State averages for K-12 Public-School Teachers



UNITED STATES

Avg. salary: \$47,674

Avg. vacation days: 63

HOURLY

Hours per week on-site: 35.5

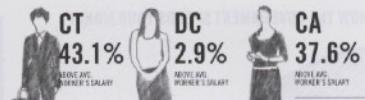
Public-School Teacher: \$34.06

Private-School Teacher: \$21.08

Average Worker: \$25.08

Police: \$22.64

Fire: \$17.91



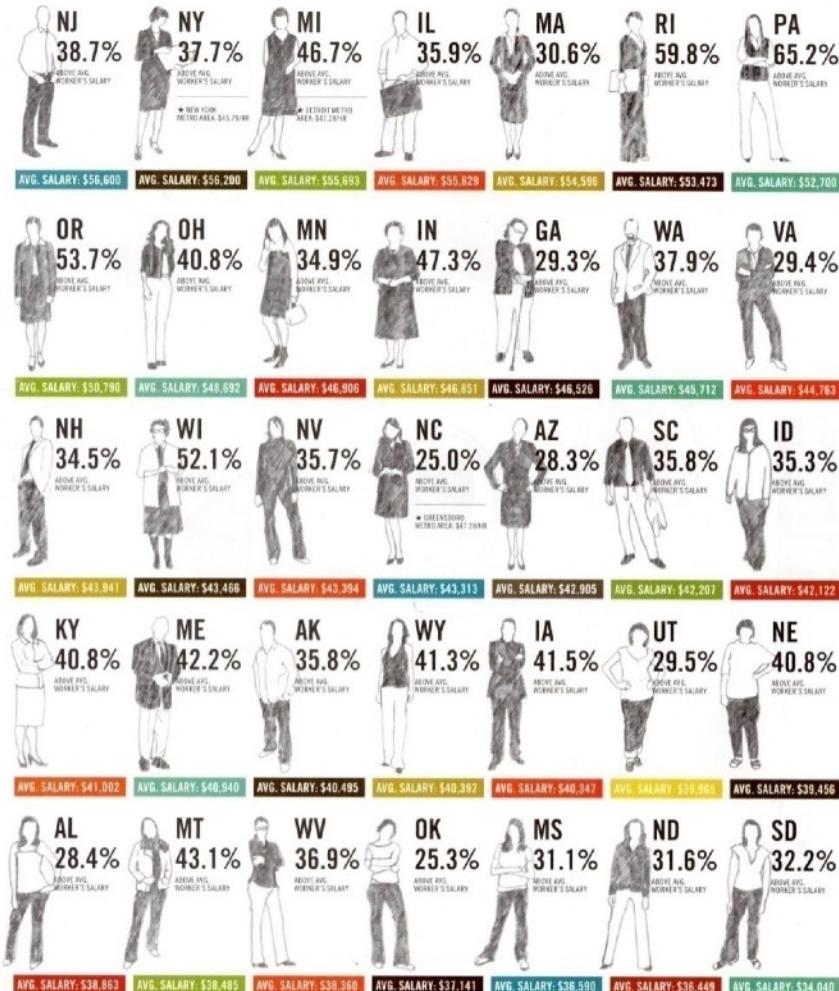
CANADA

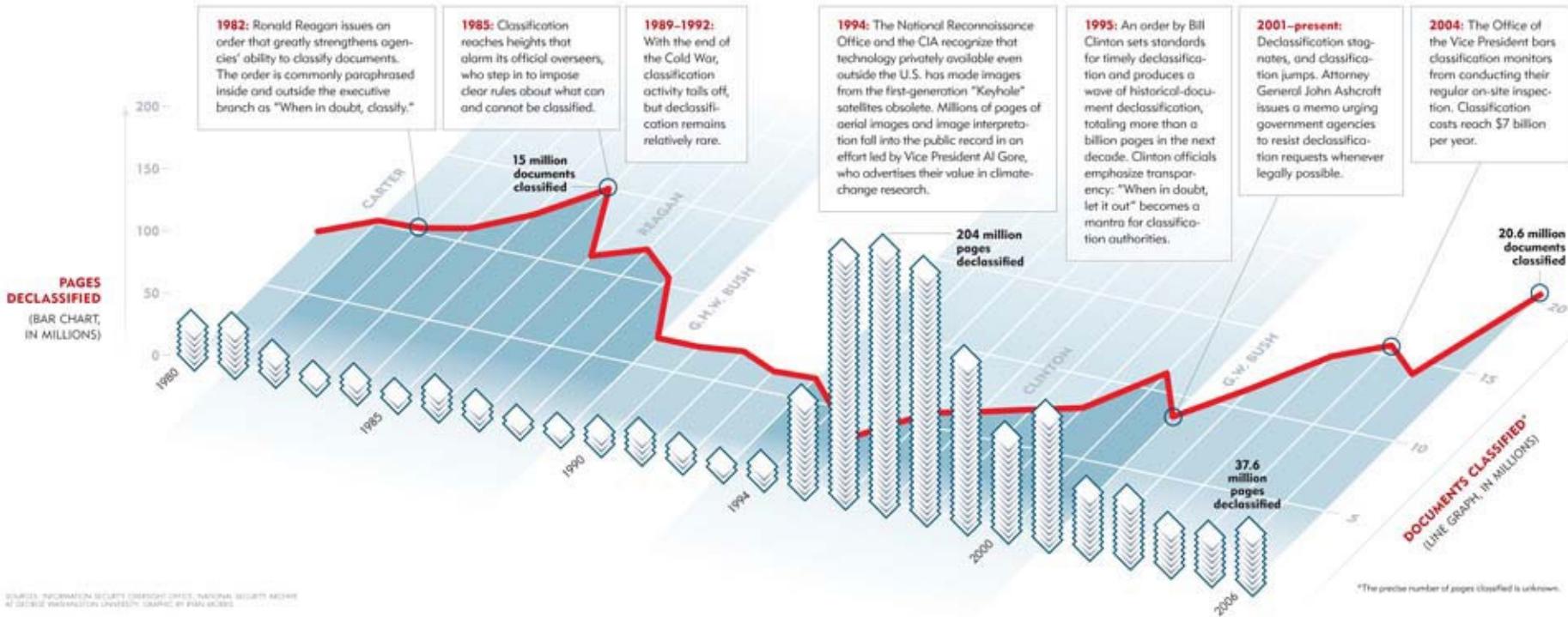
Avg. salary: \$43,000

Avg. vacation days: 50

Hourly: \$30.18

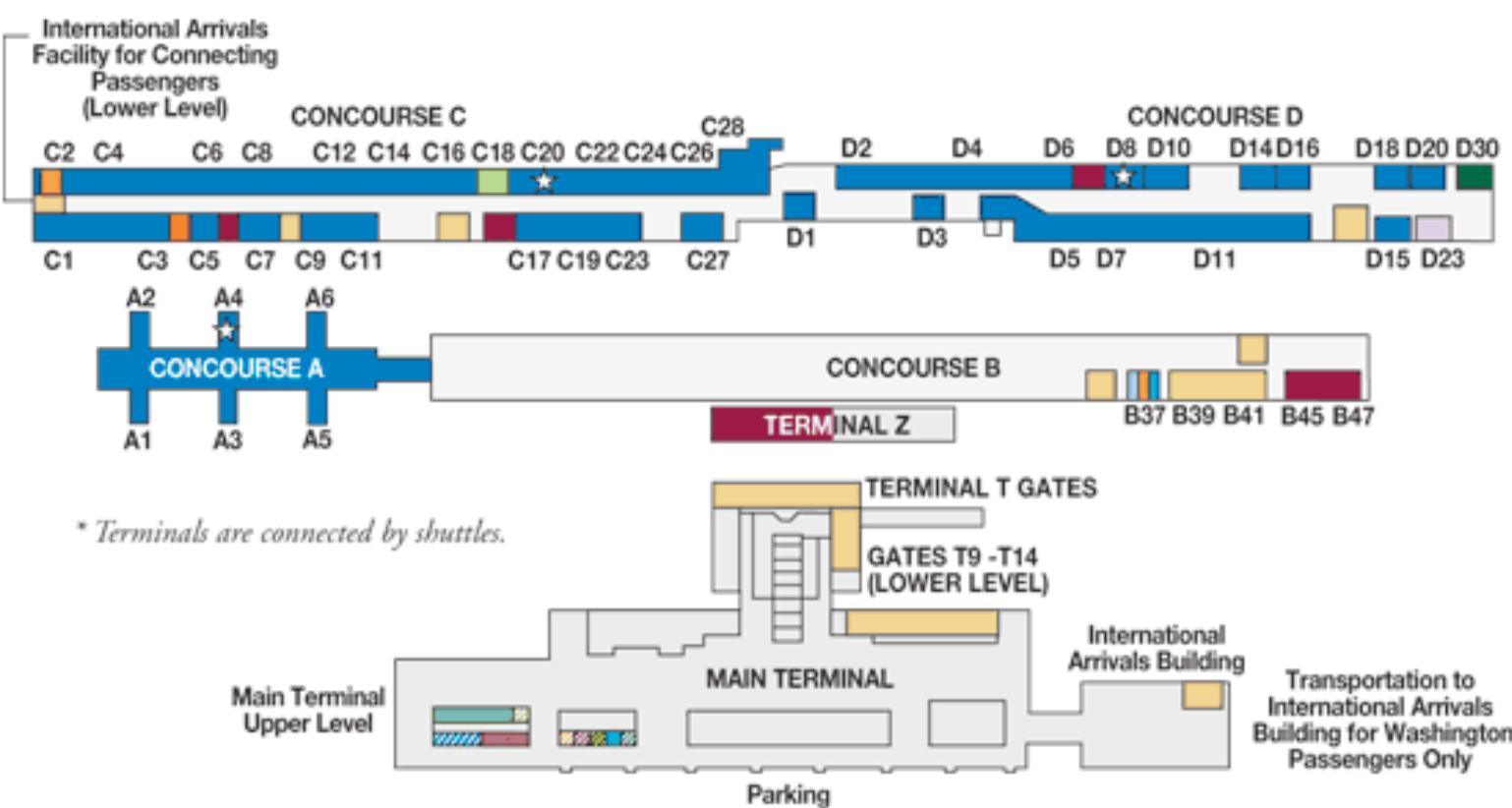
Hours per week on-site: 55.6





Source: *The Atlantic* 300 no. 2 (September 2007)

Number of Classified U.S. Documents



- | | | |
|-------------------------------------|------------------------------|-------------------------------|
| ■ United / TED Gate Area | ■ Lufthansa Check-in | ■ Austrian Airlines Gate Area |
| ■ United Premier Check-in | ■ Air Canada Gate Area | ■ SAS Gate Area |
| ■ United Check-in | ■ Air Canada Check-in | ■ BWIA Gate |
| ■ United International Check-in | ■ Mobile Lounge Dock | ■ South African Airways |
| ■ United Red Carpet Club | ■ ANA Check-in | ■ US Airways Gates |
| ■ United First International Lounge | ■ ANA Fuji Lounge/Gate Area | ★ United EasyCheck-in |
| ■ Lufthansa Gate Area | ■ Austrian Airlines Check-in | ■ US Airways Check-in |

EasyCheck-in is available at this airport.

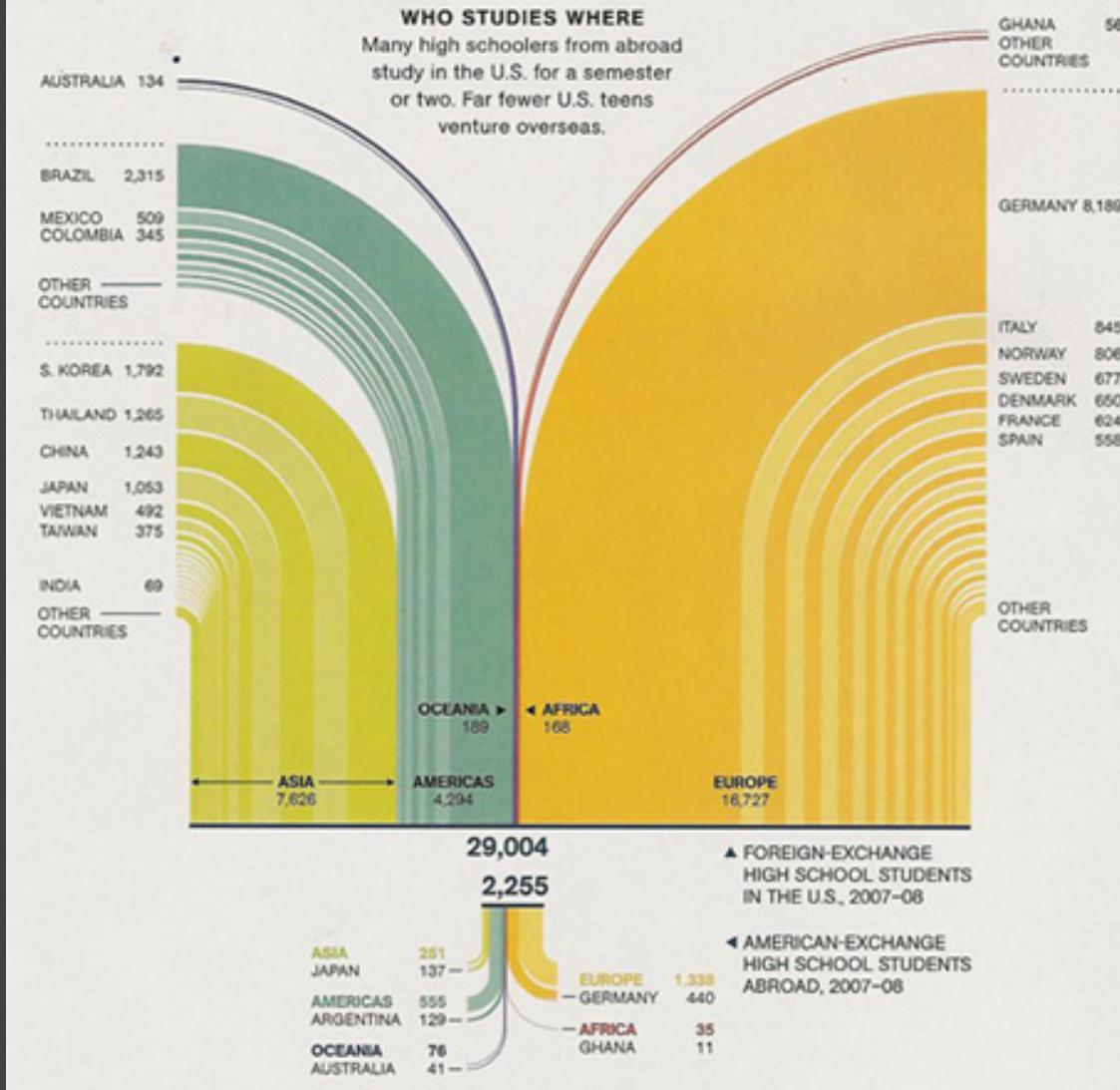


11/2006

Washington Dulles Airport Map

Source: United Airlines *Hemispheres*

GEOGRAPHY

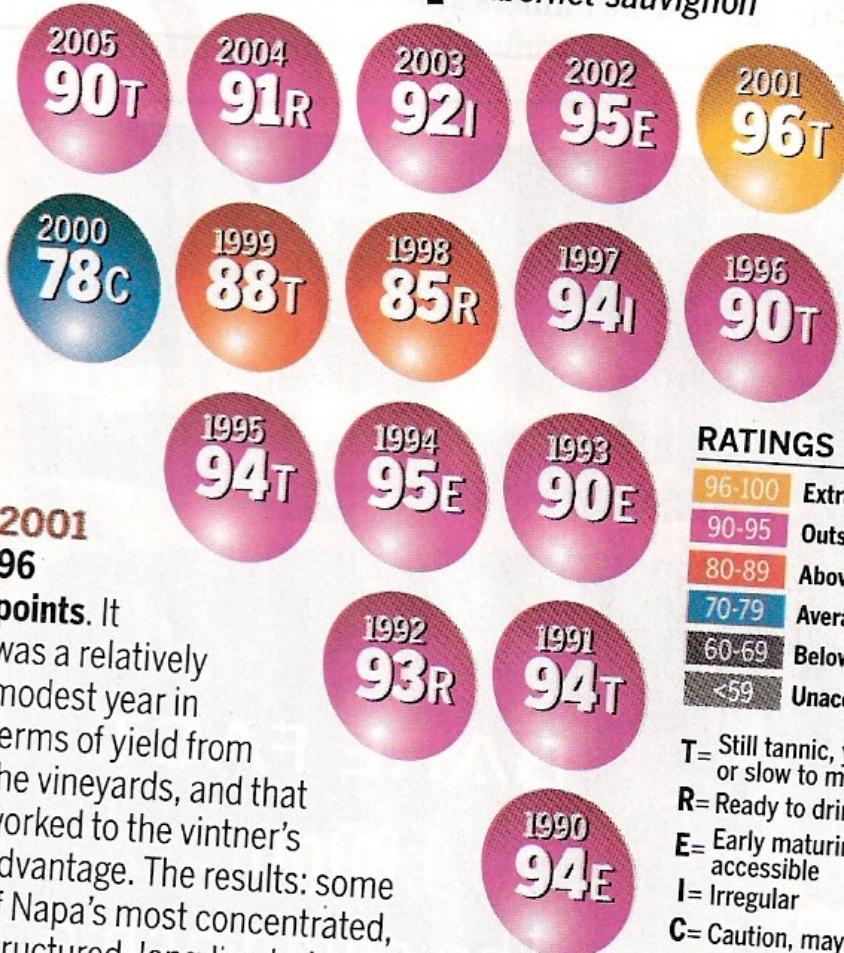


Source: *National Geographic*, September, 2008, p. 22.
 Silver, Mark. "High School Give-and-Take."

they are shaping up.

IT WAS A VERY GOOD YEAR?

Robert Parker's ratings for vintages of Napa Valley cabernet sauvignon



RATINGS

96-100	Extraordinary
90-95	Outstanding
80-89	Above average
70-79	Average
60-69	Below average
<59	Unacceptable

T= Still tannic, youthful, or slow to mature

R= Ready to drink

E= Early maturing and accessible

I= Irregular

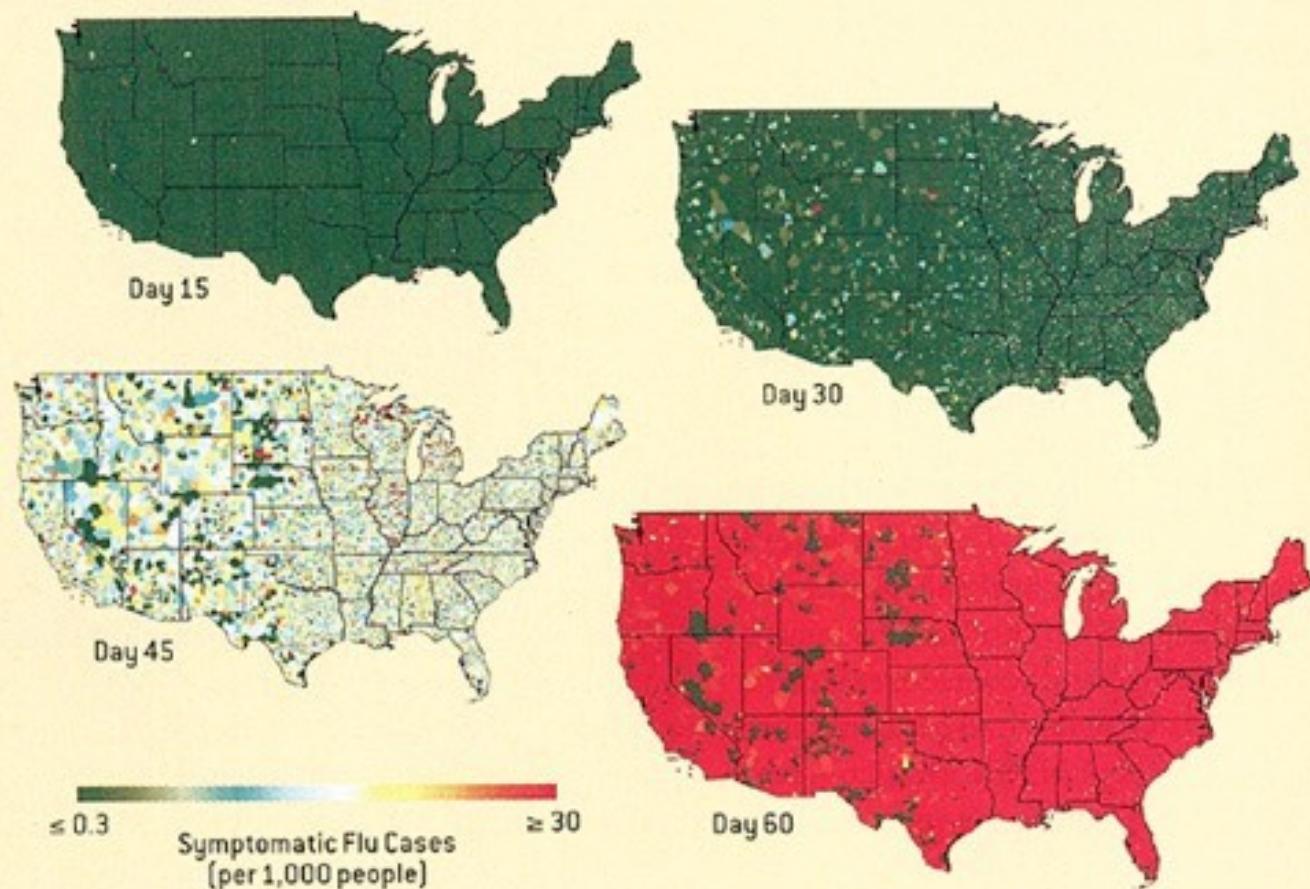
C= Caution, may be too old

2001
96
points. It was a relatively modest year in terms of yield from the vineyards, and that worked to the vintner's advantage. The results: some of Napa's most concentrated, structured, long-lived wines. Built for aging, they are rich, densely colored, fruity, and full-bodied.

Source: Business Week, June 18, 2007

Pandemic Flu Hits the U.S.

A simulation created by researchers from Los Alamos National Laboratory and Emory University shows the first wave of a pandemic spreading rapidly with no vaccine or antiviral drugs employed to slow it down. Colors represent the number of symptomatic flu cases per 1,000 people [see scale]. Starting with 40 infected people on the first day, nationwide cases peak around day 60, and the wave subsides after four months with 33 percent of the population having become sick. The scientists are also modeling potential interventions with drugs and vaccines to learn if travel restrictions, quarantines and other disruptive disease-control strategies could be avoided.



Preparing for a Pandemic

Source: *Scientific American*, 293(5). November, 2005, p. 50

