CSE 442 - Data Visualization

A1 Review

Matthew Conlen University of Washington
Review: Image Models & EDA
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: =, ≠

O - Ordered
  - Operations: =, ≠, <, >

Q - Interval (location of zero arbitrary)
  - Operations: =, ≠, <, >, -
  - Can measure distances or spans

Q - Ratio (zero fixed)
  - Operations: =, ≠, <, >, -, %
  - Can measure ratios or proportions
Visual Encoding Variables

Position (x 2)
Size
Value
Texture
Color
Orientation
Shape
Others?

### Visual Encoding Variables

<table>
<thead>
<tr>
<th>XY 2 Dimensions du Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points</td>
</tr>
<tr>
<td>Lignes</td>
</tr>
<tr>
<td>Zones</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taille</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valeur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Les Variables de Séparation des Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain</td>
</tr>
<tr>
<td>Couleur</td>
</tr>
<tr>
<td>Orientation</td>
</tr>
<tr>
<td>Forme</td>
</tr>
</tbody>
</table>
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]

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area (Size)</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td>Density (Value)</td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td>Color Sat</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area (Size)</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>
## Effectiveness Rankings

**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

[MacKinlay 86]
### Effectiveness Rankings

[Mackinlay 86]

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td>Connection</td>
</tr>
<tr>
<td>Area (Size)</td>
<td></td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density (Value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color Sat</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Color Hue</strong></td>
<td>Texture</td>
<td></td>
</tr>
<tr>
<td>Texture</td>
<td>Connection</td>
<td>Containment</td>
</tr>
<tr>
<td>Connection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Containment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density (Value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color Sat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (Size)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis. Most of the time I’m lucky if I get to do any “analysis” at all.

Anonymous Data Scientist

[Kandel et al. ’12]
Data Quality Hurdles

Missing Data
- no measurements, redacted, …?

Erroneous Values
- misspelling, outliers, …?

Type Conversion
- e.g., zip code to lat-lon

Entity Resolution
- diff. values for the same thing?

Data Integration
- effort/errors when combining data

LESSON: Anticipate problems with your data.
Many research problems around these issues!
Lesson: Iterative Exploration

Exploratory Process
1. Construct graphics to address questions
2. Inspect “answer” and assess new questions
3. Repeat...

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

Show data variation, not design variation [Tufte]
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, Opposing Axes

**Extra Context:** Other countries, other years, mortality data
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, Year
Population Count
How has America's population distribution changed over the last century?

Population by Age Group: 1900

Population by Age Group: 2000
Population size of Age groups in 1900 and 2000

- **Year 1900**
- **Year 2000**

The graph shows the population size in 5-year age groups for the years 1900 and 2000. The bars for each age group are colored to differentiate between the two years.
How does age distribution within the U.S. population compare between the years 1900 and 2000?
U.S. Population by age a century apart

The plot of Population Number for Year broken down by Age. Color shows details about Year.
Is the Average Age of Americans Increasing?

- Number of People (Log Scale)
- Age Group

- 1900
- 2000
Population Percentage
How has population distribution across age groups changed in the U.S. between years 2000 and 1900?
How Has Age Distribution Changed in the US?

% of Total Population

Age (years)

Year
- 1900
- 2000

- 0.12
- 0.10
- 0.08
- 0.06
- 0.04
- 0.02
- 0.00
AGE DISTRIBUTION OVER TIME (1900 V 2000)

Percentage of Total Population

Age
How has the population distribution changed from 1900 to 2000?

Population Distribution by Age Group (1900 vs 2000)
How Have The Ages of Americans Changed From 1900 to 2000

Chart Area

Ages

Age Ratios

1900  2000
Age Binning
Adjustment of Age Group Proportions in the U.S. Population over Time

- Youth (0-20)
- Early adulthood (20-40)
- Middle age (40-65)
- Elderly (65+)

Number of Individuals

Year

1900

2000
Age, Sex, Year
Population Count
USA census in 1900 and 2000 comparison
How Has The General Demographic Changed Over The Century?

1900 Male vs Age

1900 Female vs Age

2000 Male vs Age

2000 Female vs Age
What is the difference between males and females by year and age?

**Male Difference Age and Year**

**Female Difference Age and Year**
Is there a trend change for the number of people in each age group between 1900 and 2000?
How 20th century wars affect the U.S. population trend

Age Group (binned into 5 year segments)
How has US gender population changed in different age groups from 1900 to 2000?
Why women had higher life expectancy than men during 2000?
The United States population and sex ratio by age (1990 vs. 2000)

Ratio = Male/Female population
How did the female and male population evolve respectively from 1900 to 2000?
How did the U.S. population growth rate change between 1900 and 2000?
How has the US Life Expectancy changed from 1900 to 2000?

- Population in 1900
- Population in 2000

- Age Group
- Male
- Female
How did U.S. Demographics (Population by Age/Sex) Change from 1900 to 2000?
What Are the Differences in Population Distribution in the U.S. in 1900 vs 2000?

Shown is the age distribution of the U.S.’s population—by gender—in 1900 and 2000.
How Have Population Distributions between Males and Females Changed from 1900 to 2000?

(Source: US Census)
Comparison of Increase in population between Male and Female in from 1900 to 2000

Males

Females

Age Group

Population

Males in 1900
Males in 2000

Females in 1900
Females in 2000

1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1e7
1e7
Is the U.S. Population Aging?

U.S. Demographics by Age

Economically Active Group
Population Percentage
How have the Age and Sex distribution of population changed between 1900 and 2000?
How has the percentage of young people in the United States changed from 1900 to 2000?
How Has the Male and Female US Age Distribution Changed From 1900 to 2000?

1900 US Male and Female Population by Age

2000 US Male and Female Population by Age
What is the Breakdown of the U.S. Population with Respect to Age & Sex According to the 1900 Census?

What is the Breakdown of the U.S. Population with Respect to Age & Sex According to the 2000 Census?
Comparing Age-Sex Distributions in the U.S. for 1900 and 2000

1900
(Total Population: 76,262,821)

2000
(Total Population: 281,420,717)
How did the population of each age group proportionally change from 1900 to 2000?
Age Binning
How Does the Number of Boys and Girls (under age 20) in 1900 Compare to that in 2000?

Number of People (in millions)

Boys

<table>
<thead>
<tr>
<th>Year</th>
<th>Ages 0-4</th>
<th>Ages 5-9</th>
<th>Ages 10-14</th>
<th>Ages 15-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Girls

<table>
<thead>
<tr>
<th>Year</th>
<th>Ages 0-4</th>
<th>Ages 5-9</th>
<th>Ages 10-14</th>
<th>Ages 15-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHANGE IN LONGEVITY DURING THE 20TH CENTURY IN THE U.S. BY SEX

DISTRIBUTION OF FIVE AGE GROUPS ACROSS EACH POPULATION

MALE POPULATION IN 1900: 38.9 MILLION
MALE POPULATION IN 2000: 137.9 MILLION
FEMALE POPULATION IN 1900: 37.3 MILLION
FEMALE POPULATION IN 2000: 143.6 MILLION
Growth Rates
Which age group and sex combination had the most percentage increase from 1900 to 2000?

Source: U.S. Census Bureau via IPUMS
How does the population percent growth compare between male and females by age group?
Population Growth Percentage in each age group from 1900 to 2000

- Male Grow Percentage
- Female Grow Percentage

<table>
<thead>
<tr>
<th>Age</th>
<th>Male Grow Percentage</th>
<th>Female Grow Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
male & female population change from 1900 to 2000

Sex

% Difference in Sum of People

female

male

Year / Age
1900-2000

% Difference in Sum of People

0K%

5K%

4K%

3K%

2K%

1K%

0K%

0K%

5K%

4K%

3K%

2K%

1K%

0K%

0K%

0K%
1900-2000 Population Growth Rate between male and female across ages

Age

% Increase in people from 1900 to 2000

Sex
Male
Female

0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90

1.11
1.03
1.36
1.39
1.60
1.74
1.63
1.49
1.94
2.50
2.80
3.36
3.96
4.01
4.73
4.31
5.23
4.42
5.31
4.56
5.50
4.59
5.39
5.62
6.50
7.34
10.78
9.86
15.38
13.93
23.33
21.05
39.75
21.18
51.98
Which Sex Group's Population Grew Faster in Year 1900-2000

Source:
U.S. Census Bureau via IPUMS ([ipums.org](http://ipums.org))
How much did the population of the United States change between 1900 and 2000?
Sex Ratios
How Does the Difference Between Male and Female Populations in 2000 Vary by Age?

- Males outnumber females in the 0-4 age range.
- Negative numbers indicate a decrease in male population compared to female population.
Simplification
How Has the Gender Makeup of the 90+ Year Old Population Changed as Life Expectancy Has Increased?

Year  

Percentage of Total U.S. Population 90 Years or Older

- Male
- Female

2000
1900
What was the gender distribution over different age groups in year 2000?
Lines / Area
What is the Trend of the U.S. population for Increasing Age Groups?

The trend of sum of People for Age.
How did the age distribution change from 1900 to 2000?

The graph shows the number of people by age and sex from 1900 to 2000. The x-axis represents age, and the y-axis represents the number of people. Two lines are depicted: orange for female and blue for male. The graph indicates a decrease in the number of people across all age groups for both sexes, with a significant decrease in the elderly population.
How has the gender distribution between different ages changed in a century from 1900 to 2000?

The population between males and females in various age groups from 0 to 90 in 1900 (left figure) and in 2000 (right figure).
Population Growth of Men and Women in 1900 vs. 2000

- Population Rate for Men in 1900
- Population Rate for Women in 1900
- Population Rate for Men in 2000
- Population Rate for Women in 2000
How did the ratio of females to males change from the beginning to the end of the 20th c.?
Age Group Distribution in U.S. Census Data
Change in population percentage by age group from 1900 to 2000
PERCENTAGE OF MEN OF TOTAL POPULATION IN 1900 AND 2000
Dot / Scatter Plots
Which gender has more population by age in 1900 and 2000 and how significant is the difference?
Female vs Male Population Changes in the U.S within a Century

Year 1900

Year 2000

Sex
- Female
- Male
Age group proportion of total population given gender in 1900 vs 2000
How does the gender distribution change in age groups within 1900 and 2000 census collections?
How does Male and Female Population Differ by Age and Year?

Population Number (in millions) vs Age

- 1900 Male
- 1900 Female
- 2000 Male
- 2000 Female
Circular Charts
Percentage of US Population by Age Group, 1900 vs 2000

1900:
- 0-19: 44.2%
- 20-64: 51.7%
- 65-up: 4.1%

2000:
- 0-19: 12.4%
- 20-64: 59.1%
- 65-up: 28.5%
Is the US population getting proportionately older?

US population in 1900

US population in 2000

- under 25
- age 25 - 44
- over 65
Change in Gender Distribution of US Population in 1900 Vs. 2000, by Age Group

<table>
<thead>
<tr>
<th>Year</th>
<th>Children (0 - 14 y.o.)</th>
<th>Early Working Age (15 - 24 y.o.)</th>
<th>Prime Working Age (25 - 54 y.o.)</th>
<th>Mature Working Age (55 - 64 y.o.)</th>
<th>Elderly (65+ y.o.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>49.7% Male 13,142,996 Female 12,981,428</td>
<td>50.3% Male 7,468,884 Female 7,552,804</td>
<td>47.6% Male 14,642,648 Female 13,313,960</td>
<td>48.4% Male 2,078,479 Female 1,952,140</td>
<td>49.4% Male 1,582,228 Female 1,547,254</td>
</tr>
<tr>
<td>2000</td>
<td>48.8% Male 30,850,759 Female 29,402,802</td>
<td>49.7% Male 13,142,996 Female 12,981,428</td>
<td>50.3% Male 61,093,746 Female 61,934,414</td>
<td>52.1% Male 11,582,481 Female 12,590,229</td>
<td>58.9% Male 14,367,721 Female 20,612,918</td>
</tr>
</tbody>
</table>

*numerical data is in number of people*
Other!
Population age distribution in 1900 and 2000

[Box plots comparing age distribution for males, females, and all genders in 1900 and 2000]
How has the age distribution of the U.S. population changed from 1900 to 2000?

### U.S. Population Distribution by Age & Gender: 1900 vs. 2000

<table>
<thead>
<tr>
<th>Age Range</th>
<th>1900 (Male)</th>
<th>1900 (Female)</th>
<th>2000 (Male)</th>
<th>2000 (Female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90+</td>
<td>0.02M</td>
<td>0.02M</td>
<td>0.04M</td>
<td>0.05M</td>
</tr>
<tr>
<td>85-89</td>
<td>0.04M</td>
<td>0.05M</td>
<td>0.05M</td>
<td>0.06M</td>
</tr>
<tr>
<td>80-84</td>
<td>0.13M</td>
<td>0.13M</td>
<td>0.15M</td>
<td>0.16M</td>
</tr>
<tr>
<td>75-79</td>
<td>0.27M</td>
<td>0.27M</td>
<td>0.32M</td>
<td>0.33M</td>
</tr>
<tr>
<td>70-74</td>
<td>0.67M</td>
<td>0.64M</td>
<td>0.84M</td>
<td>0.81M</td>
</tr>
<tr>
<td>65-69</td>
<td>0.92M</td>
<td>0.89M</td>
<td>1.06M</td>
<td>1.04M</td>
</tr>
<tr>
<td>60-64</td>
<td>1.16M</td>
<td>1.06M</td>
<td>1.29M</td>
<td>1.27M</td>
</tr>
<tr>
<td>55-59</td>
<td>1.57M</td>
<td>1.41M</td>
<td>2.01M</td>
<td>1.93M</td>
</tr>
<tr>
<td>50-54</td>
<td>1.87M</td>
<td>1.65M</td>
<td>2.32M</td>
<td>2.24M</td>
</tr>
<tr>
<td>45-49</td>
<td>2.26M</td>
<td>2.00M</td>
<td>2.87M</td>
<td>2.74M</td>
</tr>
<tr>
<td>40-44</td>
<td>2.63M</td>
<td>2.35M</td>
<td>3.37M</td>
<td>3.21M</td>
</tr>
<tr>
<td>35-39</td>
<td>2.92M</td>
<td>2.67M</td>
<td>3.80M</td>
<td>3.65M</td>
</tr>
<tr>
<td>30-34</td>
<td>3.39M</td>
<td>3.24M</td>
<td>3.84M</td>
<td>3.72M</td>
</tr>
<tr>
<td>25-29</td>
<td>3.69M</td>
<td>3.75M</td>
<td>4.35M</td>
<td>4.23M</td>
</tr>
<tr>
<td>20-24</td>
<td>3.77M</td>
<td>3.80M</td>
<td>4.82M</td>
<td>4.70M</td>
</tr>
<tr>
<td>15-19</td>
<td>4.06M</td>
<td>4.00M</td>
<td>5.24M</td>
<td>5.12M</td>
</tr>
<tr>
<td>10-14</td>
<td>4.47M</td>
<td>4.39M</td>
<td>5.56M</td>
<td>5.44M</td>
</tr>
<tr>
<td>5-9</td>
<td>4.62M</td>
<td>4.59M</td>
<td>5.74M</td>
<td>5.61M</td>
</tr>
<tr>
<td>0-4</td>
<td>12M</td>
<td>10M</td>
<td>18M</td>
<td>16M</td>
</tr>
</tbody>
</table>

*Sum of Male Population and sum of Female Population for each Age Range broken down by Census Year.*

### Percent Distribution of the Total U.S. Population by Age: 1900 to 2000

- **1900**
  - 80+: 6.64%
  - 60-79: 17.04%
  - 50-59: 20.39%
  - 40-49: 26.22%
  - 30-39: 29.02%
  - 20-29: 29.48%
  - 0-19: 19.25%

- **2000**
  - 80+: 3.57%
  - 60-79: 12.90%
  - 50-59: 46.13%
  - 40-49: 26.49%
  - 30-39: 28.48%
  - 20-29: 25.09%
  - 0-19: 23.77%
How Has the Distribution of the Population Between Age Groups Changed from 1900 to 2000 For Each Gender?

Respective Segmentation Change for Each Gender from 1900 to 2000

The black line represents where the segmentation would not have changed over the 100-year period.

Female Population Segmentation by Age Group
- 1900:
  - Ages 0-24: 32%
  - Ages 25-49: 55%
  - Ages 50-74: 7%
  - Ages 75-95: 1%

- 2000:
  - Ages 0-24: 37%
  - Ages 25-49: 34%
  - Ages 50-74: 22%
  - Ages 75-95: 1%

Male Population Segmentation by Age Group
- 1900:
  - Ages 0-24: 38%
  - Ages 25-49: 34%
  - Ages 50-74: 1%
  - Ages 75-95: 21%

- 2000:
  - Ages 0-24: 37%
  - Ages 25-49: 53%
  - Ages 50-74: 1%
  - Ages 75-95: 4%
Questions: How have the US population demographics changed over a century? Is there any relationship in the trends of sex ratio and dependency ratio with the level of country economic development?

Data Sources:
- Age and Sex of the US Population from U.S. Census Bureau via IPUMS (https://www.census.gov/health/data/hse2019022.pdf)
- India Age and Gender Census (http://censusindia.gov.in/Census_Agd/Young_age_structure_and_marital_status.aspx)
- Switzerland Age and Gender Census (https://www.worldometers.info/demographics/switzerland-population-structure/)

Population Pyramids of the United States in 1900 and 2000

Relationship between the trends of sex ratio and dependency ratio with the level of country economic development.
How has the United States Age Distribution Changed From 1900 - 2000?

Age Groups:
- 0-14, Pre-employment
- 15-59, Working-Age Population
- 60+, Retired

United States Population vs. Year
- 1900: 6.4%
- 1950: 26.9%
- 2000: 21.4%

- 1900: 34.4%
- 1950: 60.9%
- 2000: 62.3%

- 1900: 59.2%
- 1950: 12.2%
- 2000: 16.3%
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.
Remove to improve (the data-ink ratio)
Administrivia
A2: Exploratory data analysis (Tue 1/28)

1. Choose dataset and identify questions of interest
2. Analysis
   - Start by cleaning data…
   - then get a broad overview. Sanity check the shape of things and look for any quality issues.
   - Dive deeper to answer specific questions and explore interesting features as they come up.

Deliverable: A sequence of annotated visualizations that clearly communicate your findings.
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

Average Salary: $47,974

State | Average Salary
---|---
CT  | 43.1%
DC  | 40.2%
CA  | 37.6%

State | Average Salary
---|---
NJ  | 38.7%
NY  | 37.7%
MI  | 46.7%
IL  | 35.9%
MA  | 30.6%
RI  | 59.8%
PA  | 65.2%

State | Average Salary
---|---
OR  | 53.7%
OH  | 40.8%
MN  | 34.9%
IN  | 47.3%
GA  | 23.9%
WA  | 37.9%
VA  | 29.4%

State | Average Salary
---|---
NH  | 34.5%
WI  | 52.1%
NV  | 35.7%
NC  | 25.0%
AZ  | 28.3%
SC  | 35.8%
ID  | 35.3%

State | Average Salary
---|---
KY  | 40.8%
ME  | 42.2%
AK  | 35.8%
WI  | 41.3%
IA  | 41.5%
UT  | 29.5%
NE  | 40.8%

State | Average Salary
---|---
AL  | 28.4%
MT  | 43.1%
WV  | 38.9%
OK  | 25.3%
MS  | 31.1%
ND  | 31.6%
SD  | 32.2%

State | Average Salary
---|---
NJ  | Avg. Salary: $48,400
NY  | $53,300
MI  | $55,900
IL  | $55,900
MA  | $55,900
RI  | $53,700
PA  | $55,900
OR  | $50,700
OH  | $54,000
MN  | $46,500
WA  | $49,700
VA  | $54,700
NH  | $43,400
WI  | $43,300
NV  | $42,900
NC  | $42,900
AZ  | $42,200
SC  | $42,100
ID  | $43,400
KY  | $41,800
ME  | $42,400
AK  | $41,300
WI  | $49,400
IA  | $49,400
UT  | $49,400
NE  | $49,400
AL  | $38,800
MT  | $38,000
WV  | $36,000
OK  | $37,100
MS  | $36,400
ND  | $36,400
SD  | $34,000

Source: Good Magazine

GoodMagazine.com

Transparency
Source: *The Atlantic* 300 no. 2 (September 2007)
Number of Classified U.S. Documents
Washington Dulles Airport Map
Source: United Airlines Hemispheres
IT WAS A VERY GOOD YEAR?

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

2005 90T
2004 91R
2003 92I
2002 95E
2001 96T
2000 78C
1999 88T
1998 85R
1997 94I
1996 90T
1995 94T
1994 95E
1993 90E
1992 93R
1991 94T
1990 94E

RATINGS

96-100 Extraordinary
90-95 Outstanding
80-89 Above average
70-79 Average
60-69 Below average
<60 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 long-lived.
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