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 image visual channel graphical marks

Nominal, Ordinal & Quantitative

- N Nominal (labels or categories)
 - Operations: =, ≠
- O Ordered
 - Operations: =, \neq , <, >
- Q Interval (location of zero arbitrary)
 - Operations: =, ≠, <, >, -
 - 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?



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

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

NOMINAL 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

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 Erroneous Values Type Conversion Entity Resolution Data Integration

no measurements, redacted, ...? misspelling, outliers, ...? e.g., zip code to lat-lon diff. values for the same thing? effort/errors when combining data

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

Lesson: Iterative Exploration

Exploratory Process

Construct graphics to address questions
 Inspect "answer" and assess new questions
 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 size of Age groups in 1900 and 2000

5-year Age groups

Population size

U.S. Population change in different age groups from 1900 vs 2000





How does age distribution within the U.S. population compare between the years 1900 and 2000?



Age Group (years)



U.S Population by age a century apart

The plot of Population Number for Year broken down by Age. Color shows details about Year.



US Population over Age Distribution: 1900 vs. 2000

Age



Population Percentage



How has population distribution across age groups changed in the U.S. between years 2000 and 1900?

Year

How Has Age Distribution Changed in the US?





AGE DISTRIBUTION OVER TIME (1900 V 2000)

Age



How has the population distribution changed from 1900 to 2000?

Population Distribution by Age Group (1900 vs 2000)



1900 2000

Percentage of Population (%)



Age Binning



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

Year

Age, Sex, Year
Population Count





DEMOGRAPHIC MAKEUP OF THE UNITED STATES 1990 VS 2000



How Has The General Demographic Changed Over The Century?



What is the difference between males and females by year and age?





How 20th century wars affect the U.S. population trend

People



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)



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 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?



Male Population

(Source: US Census)

2000 Female Population 1900 Female Population

1900 Male Population 2000 Male Population



Comparision of Increase in population between Male and Female in from 1900 to 2000



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?





Comparing Age-Sex Distributions in the U.S. for 1900 and 2000

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?



CHANGE IN LONGEVITY DURING THE 20TH CENTURY IN THE U.S. BY SEX

■ 0-19 ■ 20-39 ■ 40-59 ■ 60-79 ■ 80+



Growth Rates

Which age group and sex combination had the most percentage increase from 1900 to 2000?

Source: U.S. Census Bureau via IPUMS



Age



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



Age

male & female population change from 1900 to 2000







Which Sex Group's Population Grew Faster in Year 1900-2000



How much did the population of the United States change between 1900 and 2000?

Age
Sex Ratios



How has the Population Gender Distribution Changed per Age Group in the U.S. over the Past Century?





Does Gender Affect Life Expectancy? (1900 vs 2000)

How Does the Difference Between Male and Female Populations in 2000 Vary by Age?



Simplification

How Has the Gender Makeup of the 90+ Year Old Population Changed as Life Expectancy Has Increased?



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



What was the gender distrubition 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?



Age

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 Census: A Century Apart

US Population Gender Gap





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



Population Increase from 1900 to 2000 for Each Age in Percent



Population Growth from 1900 to 2000



—— Sex 1 —— Sex 2





Dot / Scatter Plots





Female vs Male Population Changes in the U.S within a Century

🏙 Tableau - 442_a1

٥ Х

The Show Me

ATTR(Sex)

ATTR(Year)

O 1,900

2,000

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help G, G ₽₽ \rightarrow Į - Ø - T 🎝 Standard 🔻 •••• å 夺. \leftarrow + ⊖ + Pages Data Analytics iii Columns Copy of census2000 (Co.. I Rows III P -Dimensions Filters Age group proportion of total population given gender in 1900 vs 2000 # Year Measure Names Abc Measure Names Marks Ó Ao Shape Ŧ 8 ••• 6 Т Size Color Label 00 \Box 000 0 0 Tooltip Shape Detail Measure Values e0.0 **ATTR(Sex)** Ó ATTR(Year) Percentage of total population given gender 0.08 Measure Values Measures SUM(Age) # Age SUM(Number of Rec 0 # People SUM(People) 0.06 # Percentage of total popul... SUM(Percentage of . Ó # Sex =# Number of Records # Measure Values Ó 0.04 8 0 0 Ó 0 0.00 10 15 30 35 50 55 60 65 70 75 -5 5 20 25 40 45 Age Group Sheet 2 🖳 🕂 🗸 Data Source 304 marks 1 row by 1 column SUM of ATTR(Sex): 114 Highlighting on ATTR(Year) 🔲 🔀 🚸 🚖 💵 🗉 x 華 $\,\mathcal{P}\,$ Type here to search 0 9

IK K → → I 🔛 🔜 🔳

0

80

85

90

95

How does the gender distribution change in age groups within 1900 and 2000 census collections?





Circular Charts

Percentage of US Population by Age Group, 1900 vs 2000



Is the US population getting proportionately older?







*numerical data is in number of people



Other!



Population age distribution 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

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



% of Total People for each Census Year. Color shows details about Age (bin).
Simple Demographic Analysis Dashboard, 1900 vs. 2000

CSE 442 - WI 20 - Chenghao Zhu (Jessica) - chengz27



Sex Ratio

Distribution by Age Group

People



Mortality Rates Related to Alcoholism in the United States (1900 vs. 2000)

Death Rate (2000) 12.40 12.40 10.50 10.40 10.50 10.40 6.10 6.10 5.00 5.00 4.8 2.20 2.20 0.70 0.0.70 0.10 0.00 0.10 0.00 0.00 10 20 30 80 0 40 50 60 70 90

Legend



Population Growth of US in a Century

How Has the Distribution of the Population Between Age Groups Changed from 1900 to 2000 For Each Gender?



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









Age Demographic of the 2000 Census



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? Yin Yin Low

Age and Sex of the US Population from U.S. Census Bureau via IPUMS; Average life expectancy at birth (all races), United States (https://www.cdc.gov/nchs/data/hus/2010/022.pdf); India Age and Gender Census (http://censusindia.gov.in/Census_And_You/age_structure_and_marital_status.aspx); Switzerland Age and Gender Census (https://www.worldometers.info/demographics/switzerland-demographics/#age-structure)

Voung dependency ratio Total dependency ratio



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.

to improve (the data-ink ratio)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

Administrivia

Updates

A2: Exploratory data analysis (Tue 1/28)

- Choose dataset and identify questions of interest
 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



Source: Good Magazine



Source: *The Atlantic* 300 no. 2 (September 2007) Number of Classified U.S. Documents



EasyCheck-in is available at this airport. Easychectein

Washington Dulles Airport Map

Source: United Airlines *Hemispheres*

GEOGRAPHY



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



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



Source: *Wired Magazine*, September 2008 Edition <u>Music: Super Cuts (page 92)</u>