Lecture 4: Histograms

February 10, 2025

Course Administration

1. Policy brief proposal comments posted – alert me if you don't see them
2. Tutorials graded – more on next slide – grades up in BB
3. Reminder: Fully composed chart due class after holiday **Feb. 24**
   - if there is something you want to do, but can't figure out how
   - write it in words accompanying the graph
   - I want to be sure you know what to do
   - we can work on how to do it
   - look at the grading sheet for complete graph elements
4. Anything lingering?

Tutorial Feedback

- You need to do majority including questions at end to pass
- For future tutorials, put questions and answers at the top or make a clear heading
- Don't make me sort through your code!
- Make code and output for entire file
- See answer key about finding average state-level populations from Tutorial 1

General Policy Brief Proposal Feedback

Good work and interesting topics.

General Policy Brief Proposal Feedback

Good work and interesting topics.

Successful proposals

- have a point to make and a story to tell
- clearly set out the ≥ 2 data sources you're using
- explain how you're planning on aggregating data
- give a sense of having some thoughts about the graphics you'd like to do or the points you'd like to make
- for this class, aggregation does not mean merging together. it means going from one unit of observation to another

General Policy Brief Proposal Feedback

Good work and interesting topics.

Successful proposals

- have a point to make and a story to tell
- clearly set out the ≥ 2 data sources you're using
- explain how you're planning on aggregating data
- give a sense of having some thoughts about the graphics you'd like to do or the points you'd like to make
- for this class, aggregation does not mean merging together. it means going from one unit of observation to another

Do not

- report regression coefficients

General Policy Brief Proposal Feedback

Good work and interesting topics.

Successful proposals

- have a point to make and a story to tell
- clearly set out the ≥ 2 data sources you're using
- explain how you're planning on aggregating data
- give a sense of having some thoughts about the graphics you'd like to do or the points you'd like to make
- for this class, aggregation does not mean merging together. it means going from one unit of observation to another

Do not

- report regression coefficients

In the best final work, graphics drive narrative

Looking Forward to the Final Product

- Final product needs 5 to 8 graphics
- graphics with some basic descriptives often set the stage
- may be helpful to think about summary statistics before correlations
- with new data, good practice for you to match published summary stats
- as relevant, consider adding in decennial census/acs data to add demographics
- come see me about data sources

Looking Forward to the Final Product

- Final product needs 5 to 8 graphics
- graphics with some basic descriptives often set the stage
- may be helpful to think about summary statistics before correlations
- with new data, good practice for you to match published summary stats
- as relevant, consider adding in decennial census/acs data to add demographics
- come see me about data sources
- expect to have problems
- download any federal data **now**

Looking Forward to the Final Product

- Final product needs 5 to 8 graphics
- graphics with some basic descriptives often set the stage
- may be helpful to think about summary statistics before correlations
- with new data, good practice for you to match published summary stats
- as relevant, consider adding in decennial census/acs data to add demographics
- come see me about data sources
- expect to have problems
- download any federal data **now**
- next deadline: Lecture 5, one fully composed chart

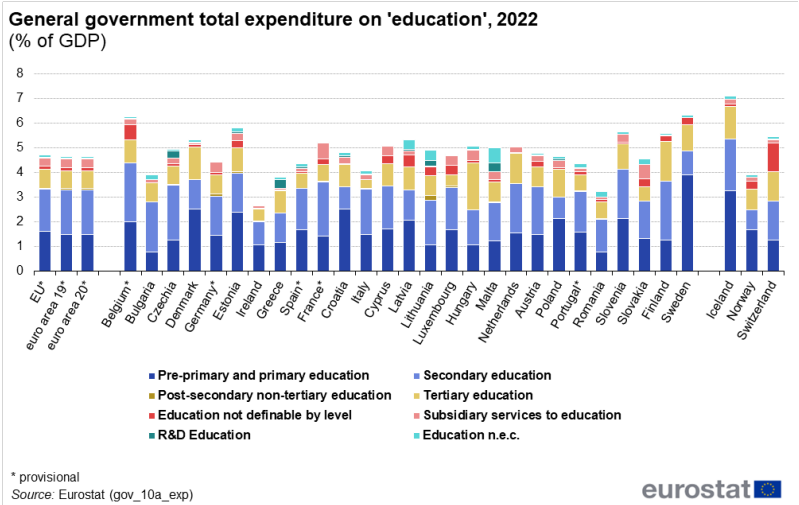
Next Week's Good Bad and Ugly

Find a histogram. Post by Wednesday noon. Post the link on the google sheet.

Finder	Commenter
Raquel L.	Katelyn H.

Next week is a holiday.

Emma on Katelyn's Graph



From Eurostat, [link]

Which Graph for What Purpose?

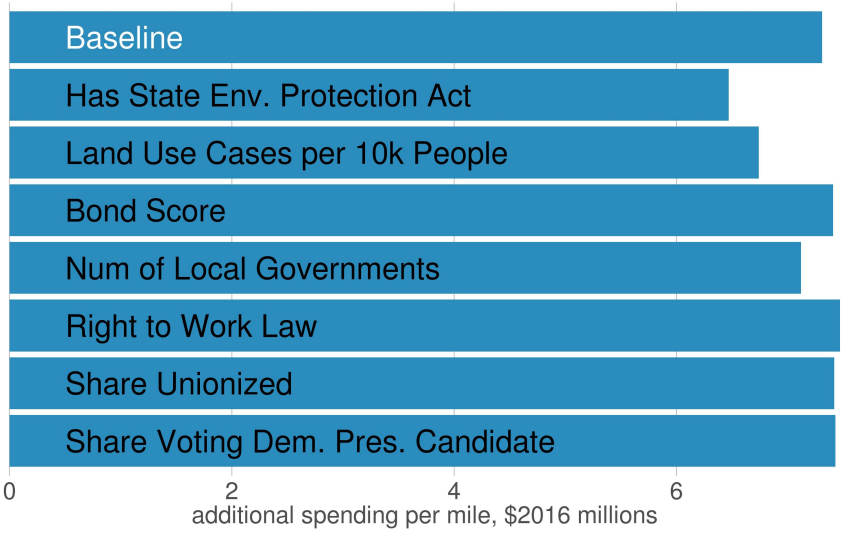
Few: Basic Ways to Convey Information Graphically

1. Bars
2. Lines
3. Boxes for distributions

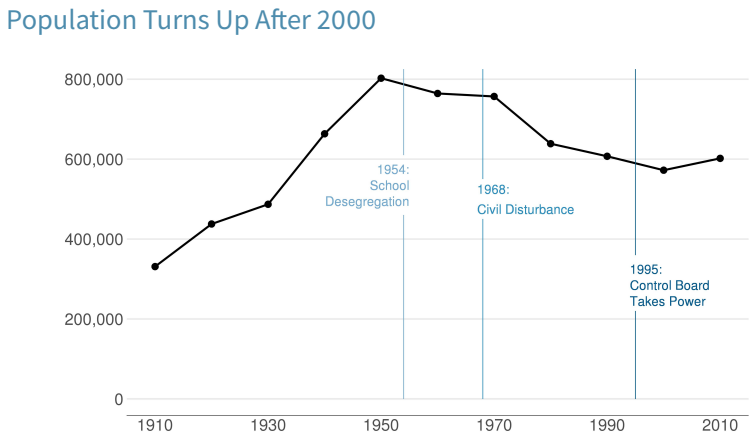
Few: Basic Ways to Convey Information Graphically

1. Bars
2. Lines
3. Boxes for distributions
4. Not for today
    - Points
    - Shapes with varying 2-D colors
    - Shapes with varying color intensity

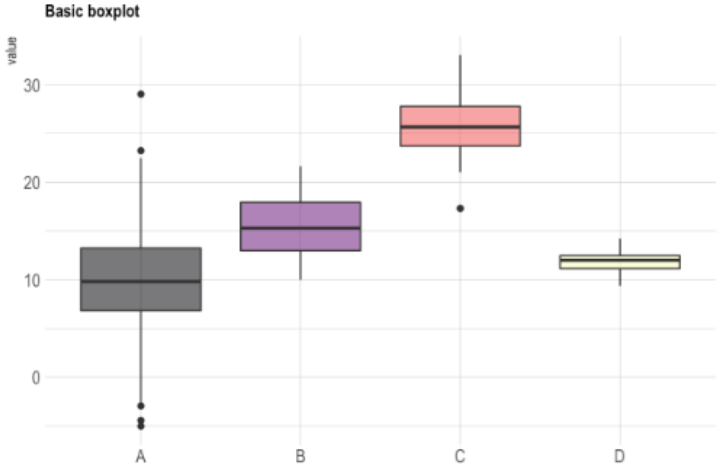
Bars


Baseline
Has State Env. Protection Act
Land Use Cases per 10k People
Bond Score
Num of Local Governments
Right to Work Law
Share Unionized
Share Voting Dem. Pres. Candidate
0
2
4
6
additional spending per mile, $2016 millions

Lines


Population Turns Up After 2000
800,000
600,000
400,000
200,000
0
1910
1930
1950
1970
1990
2010
1954:
School
Desegregation
1968:
Civil Disturbance
1995:
Control Board
Takes Power

Boxes



Source: https://www.r-graph-gallery.com/89-box-and-scatter-plot-with-ggplot2.html

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison		
Time Series		
Ranking		
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	
Time Series		
Ranking		
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series		
Ranking		
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	
Ranking		
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	Bars falsely suggest independence
Ranking		
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	Bars falsely suggest independence
Ranking	Bars or Dots	
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	Bars falsely suggest independence
Ranking	Bars or Dots	Not lines!
Part-to-whole		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	Bars falsely suggest independence
Ranking	Bars or Dots	Not lines!
Part-to-whole	Bars or stacked bars	

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Nominal comparison	Bars, Points sparingly	Bars starting above 0
Time Series	Lines	Bars falsely suggest independence
Ranking	Bars or Dots	Not lines!
Part-to-whole	Bars or stacked bars	No pies!

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single		
Multiple		
Correlation		
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple		
Correlation		
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple	Bars or Dots	
Correlation		
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple	Bars or Dots	Two histograms together is hard!
Correlation		
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple	Bars or Dots	Two histograms together is hard!
Correlation	Points or paired bars	
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple	Bars or Dots	Two histograms together is hard!
Correlation	Points or paired bars	Rarely lines
Geospatial		

Types of Relationships You May Want to Show, 1 of 2

Relationship	Use	Avoid
Distribution		
Single	Histogram, dot plot, or density curve	
Multiple	Bars or Dots	Two histograms together is hard!
Correlation	Points or paired bars	Rarely lines
Geospatial	Wait for maps!	

Histograms

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

1. levels

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

1. levels
    - number of commuters by jurisdiction

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

1. levels
   - number of commuters by jurisdiction
2. shares

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

1. levels
   - number of commuters by jurisdiction
2. shares
   - share of commuters by jurisdiction

When Do You Want to Show a Histogram?

- For when you need to communicate beyond the mean
- Or show relative prevalence

When Do You Want to Show a Histogram?

- For when you need to communicate beyond the mean
- Or show relative prevalence
- Examples
  - How many people at each part of the income distribution?
  - How much of an outlier is Caitlin Clark?
  - How much variation is there in prices for airline tickets?
  - My work: distribution of condo prices vs other residential prices

When Do You Want to Show a Histogram?

- For when you need to communicate beyond the mean
- Or show relative prevalence
- Examples
  - How many people at each part of the income distribution?
  - How much of an outlier is Caitlin Clark?
  - How much variation is there in prices for airline tickets?
  - My work: distribution of condo prices vs other residential prices
- You give some examples!

When Do You Want to Show a Histogram?

- For when you need to communicate beyond the mean
- Or show relative prevalence
- Examples
  - How many people at each part of the income distribution?
  - How much of an outlier is Caitlin Clark?
  - How much variation is there in prices for airline tickets?
  - My work: distribution of condo prices vs other residential prices
- You give some examples!

But beware! Histograms are harder than bars.

Notionally, to Create a Histogram

- Take a variable
- Make bins by value
- Count the number of observations in each bin
- Plot bars with that number

Notionally, to Create a Histogram

Imaginary Income Data

Person	Income
A	4
B	11
C	12
D	3
E	0

Notionally, to Create a Histogram

Imaginary Income Data

Person	Income
A	4
B	11
C	12
D	3
E	0

Decide on bins

Notionally, to Create a Histogram

Imaginary Income Data

Person	Income
A	4
B	11
C	12
D	3
E	0

Decide on bins

Adding a Bin

Person	Income	Bin
A	4	1-5
B	11	11-15
C	12	11-15
D	3	1-5
E	0	0

Notionally, to Create a Histogram

Imaginary Income Data

Person	Income
A	4
B	11
C	12
D	3
E	0

Decide on bins

Adding a Bin

Person	Income	Bin
A	4	1-5
B	11	11-15
C	12	11-15
D	3	1-5
E	0	0

Binned Dataset

Bin	No.
0	1
1-5	2
6-10	0
11-15	2

Notionally, to Create a Histogram

Imaginary Income Data

Person	Income
A	4
B	11
C	12
D	3
E	0

Decide on bins

Adding a Bin

Person	Income	Bin
A	4	1-5
B	11	11-15
C	12	11-15
D	3	1-5
E	0	0

Binned Dataset

Bin	No.
0	1
1-5	2
6-10	0
11-15	2

Graph this one!

Key Features of Histograms

- Looks like a bar chart
- But! unlike a bar chart, histogram bars touch, to indicate continuity
- Give me more examples of when a histogram would be useful

The Histogram Inventor

Karl Pearson (1857-1936) as a young man

The Histogram Inventor

Karl Pearson (1857-1936) as a young man

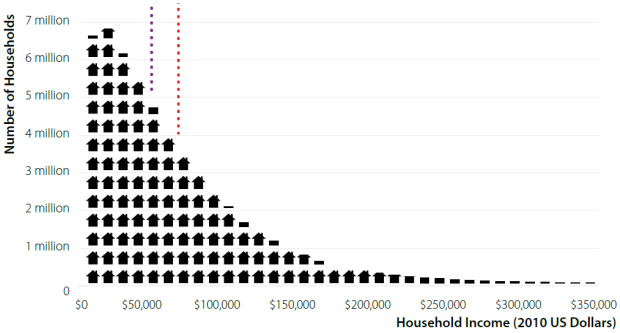
A big thinker

- father of mathematical statistics
- publishes first histogram, 1895
- fervent eugenicist
- early suffragist
- turned down knighthood due to socialist beliefs

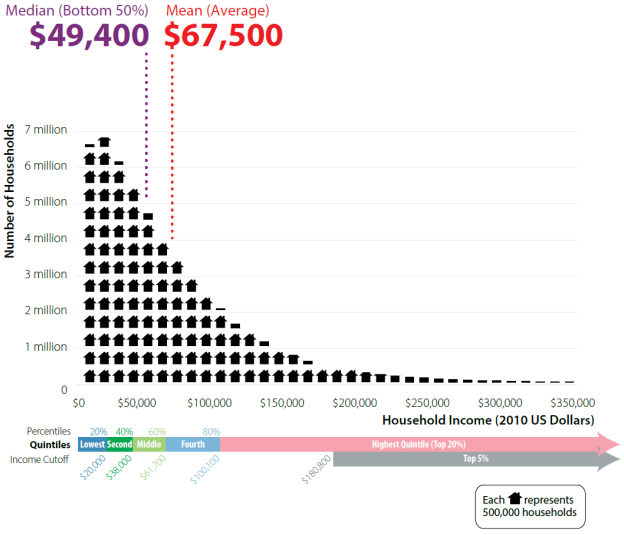
Histogram Examples

- Income distribution
- As a guide on a map
- Income distribution for DC MSA
- Condo price distribution

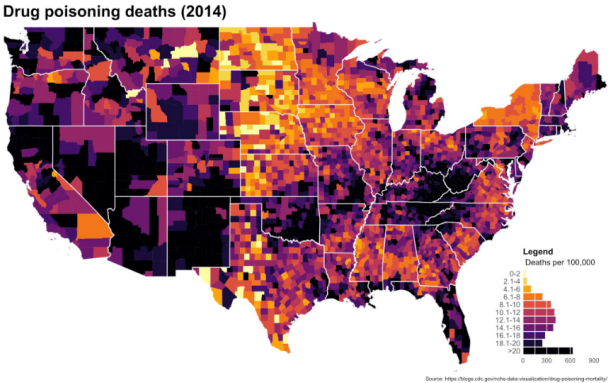
Mulbrandon's Income Histogram



Mulbrandon's Income Histogram

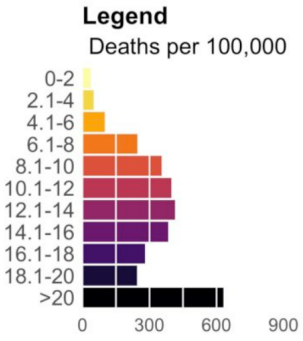


As a Map Legend



From https://mathewkiang.com/2017/01/16/using-histogram-legend-choropleths/

Legend, Now Visible

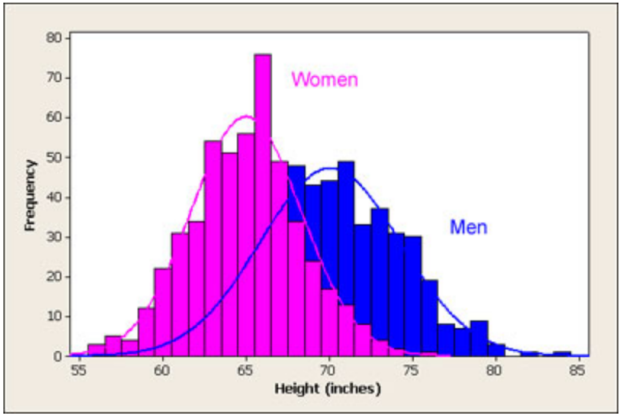


With code on how to do this! From Matthew Kiang's website

Density Curves: Smoothed Histograms

- Imagine many very thin bars
- This yields a curve
- Sometimes it is more helpful to draw the curve

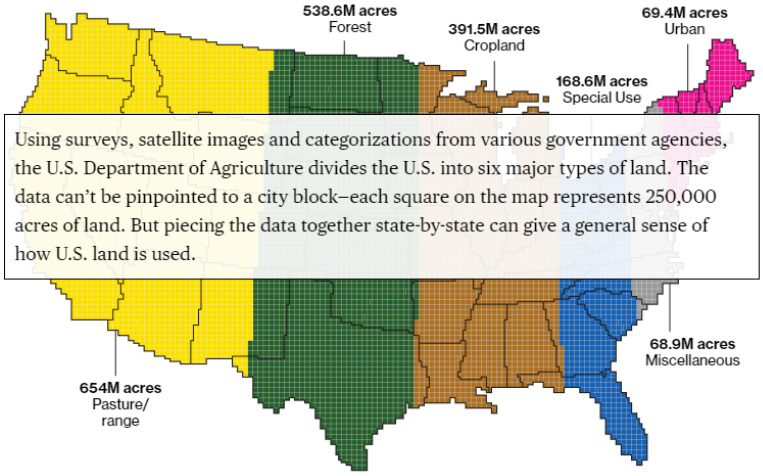
Height: Note the Curves



From http://www.usablestats.com/lessons/normal

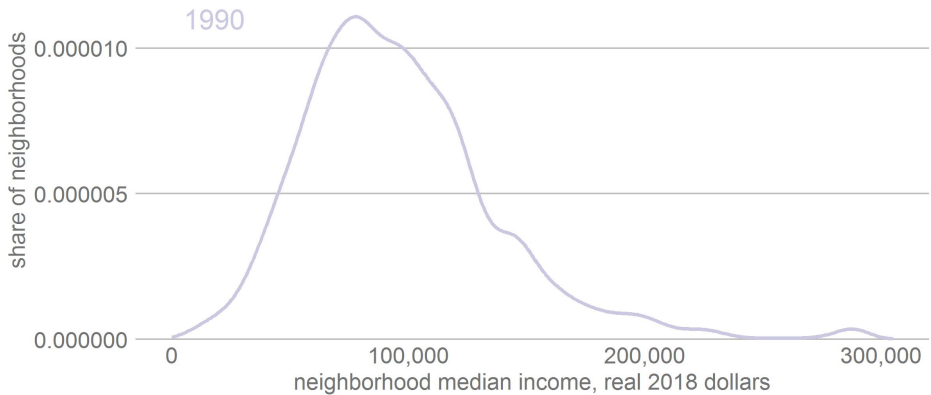
Land By Type: Between a Histogram and a Bar

Goal here is also histogram-like.



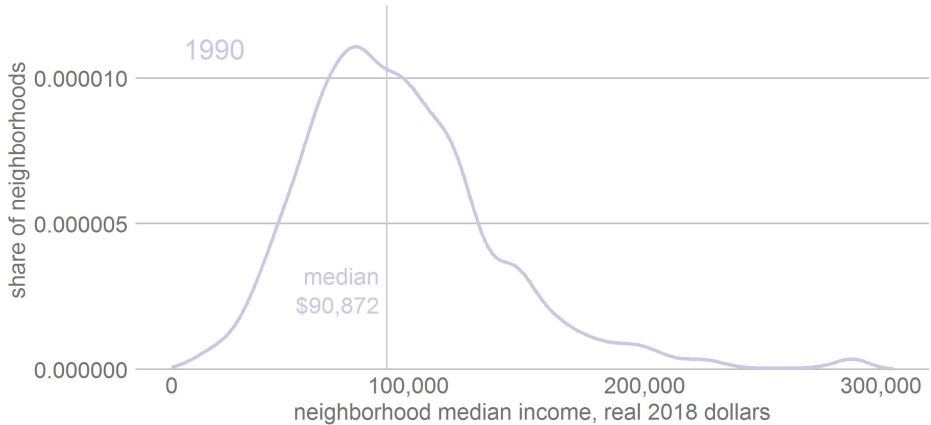
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



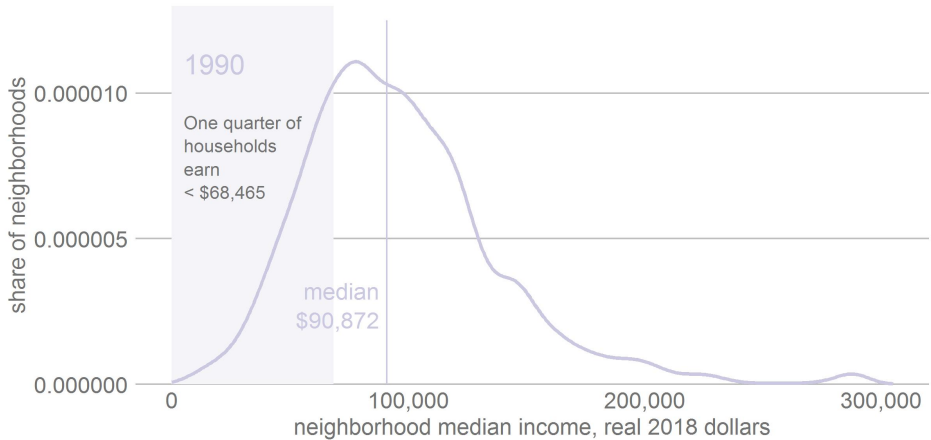
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



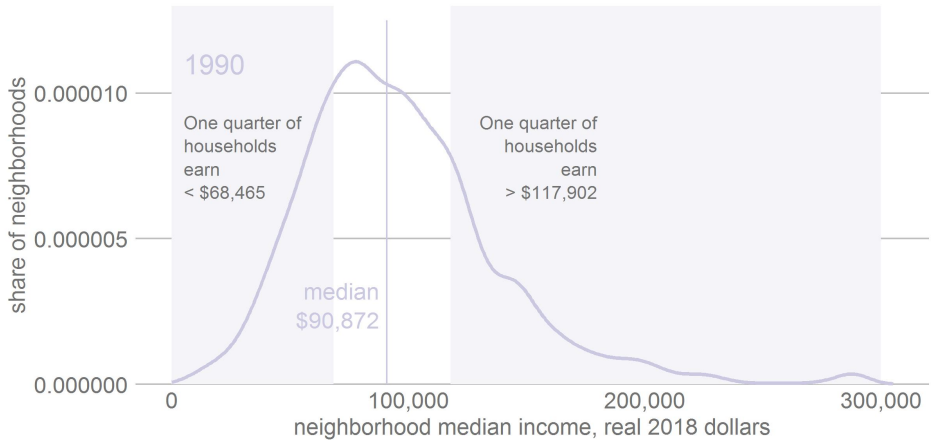
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



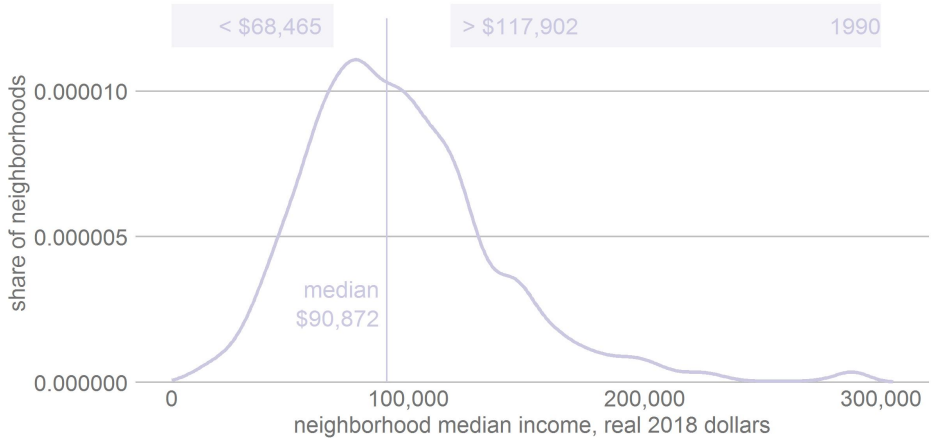
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



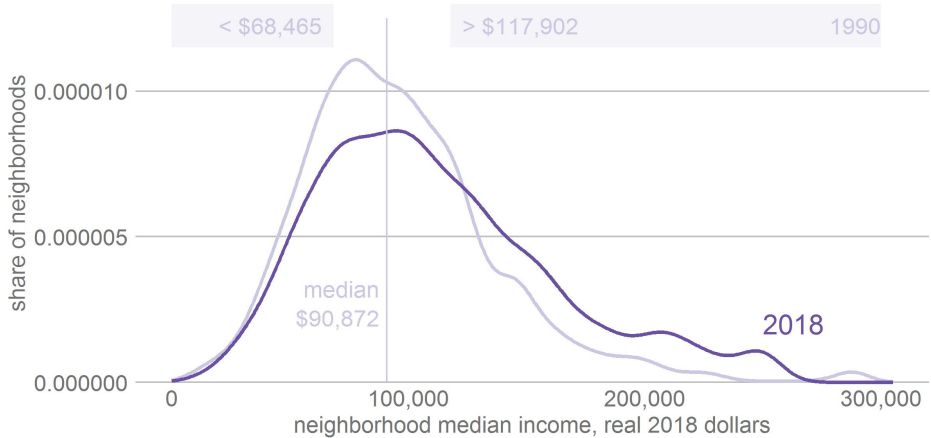
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



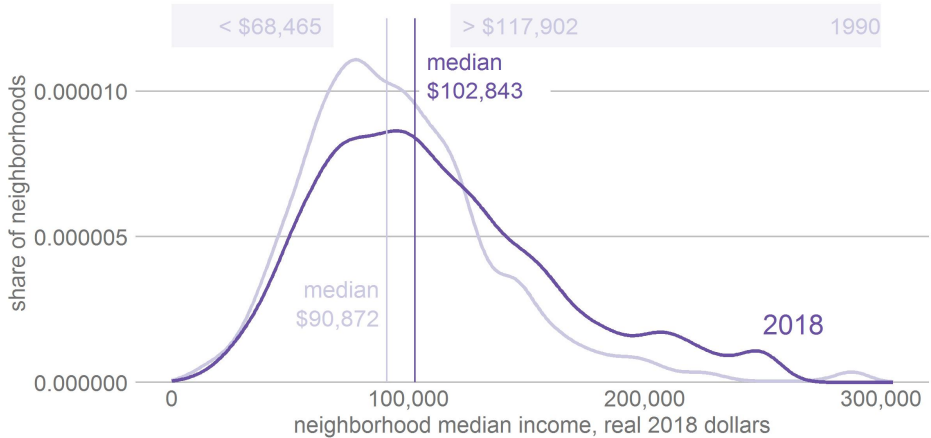
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



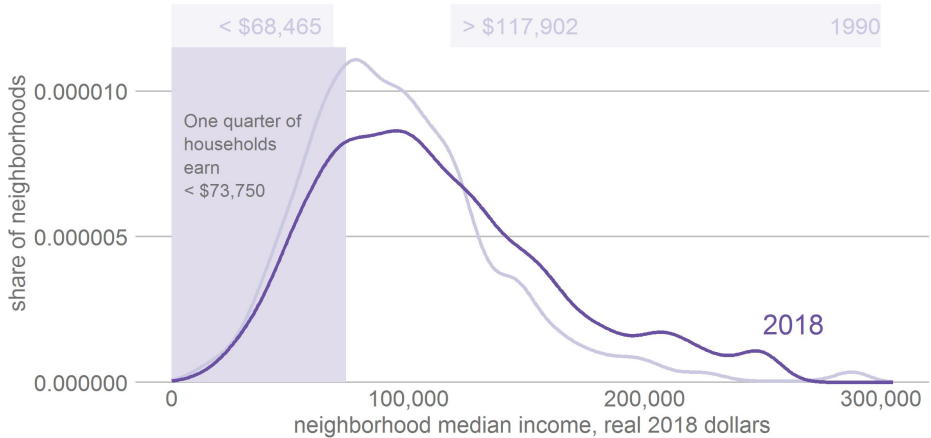
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



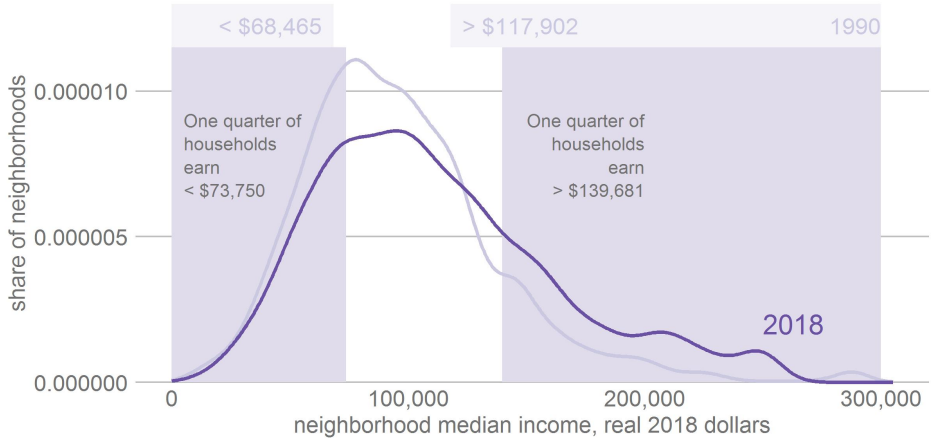
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



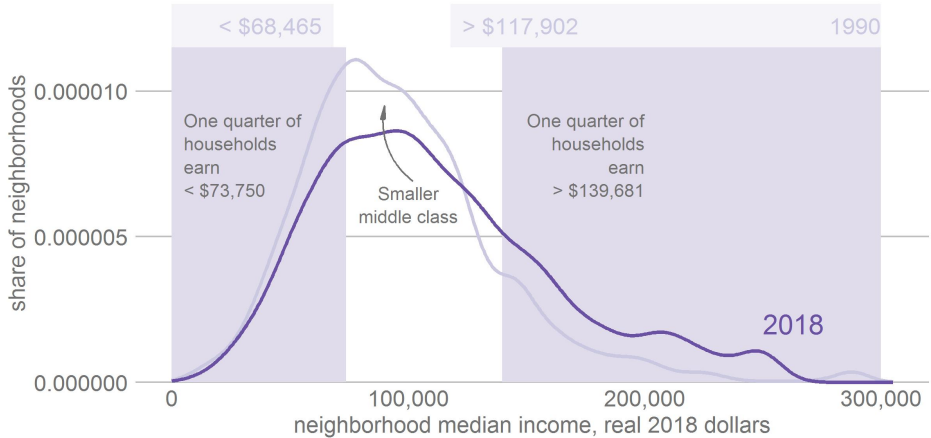
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out



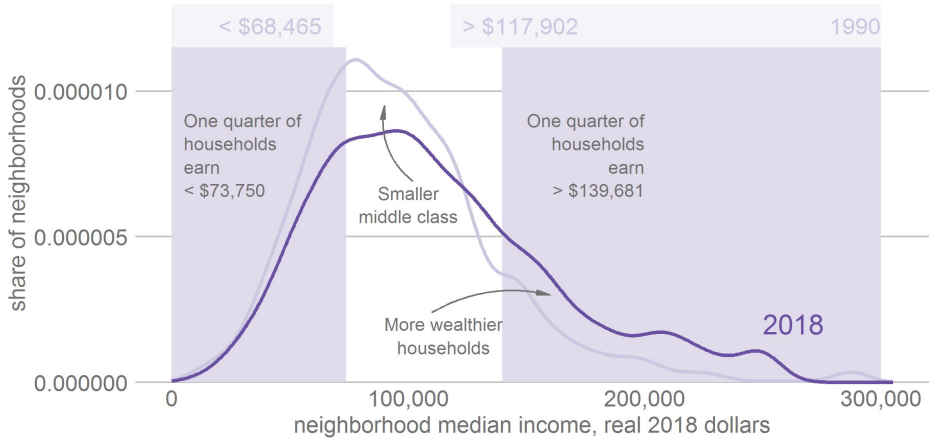
INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out

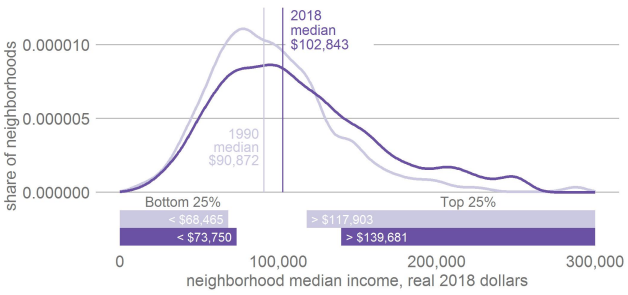


INCOME

Incomes Have Grown, But the Middle Class Has Hollowed Out

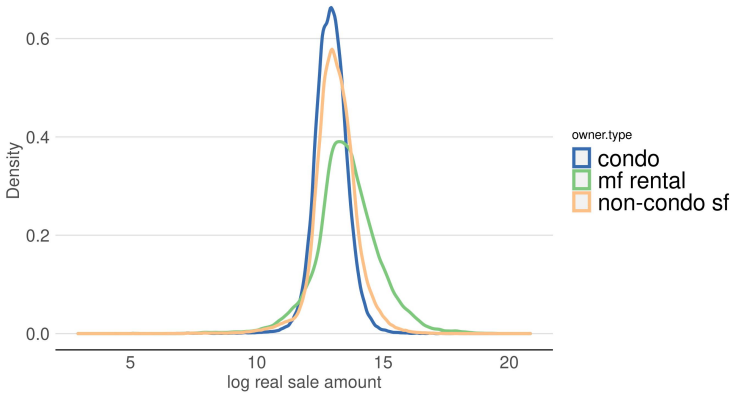


Income Distribution in the DC Metro Area Over Time



- print version
- was never satisfied with y axis
- light purple probably too light
- goal was to show 25th and 75th percentiles
- and change therein

Comparison of Value at the Tails of the Distribution



- want to see if low value SF homes exist
- and low value condos do not
- this graph is a first start
- but not sufficient

R Histograms

Next Class: Maps

- Not next week but the week after
- Turn in Tutorial 4
- Turn in fully composed chart assignment to piazza folder fully_composed_chart
- Manson, *Mapping and Society*, Chapter 3
- Look at linked dot density map from *Post*