Start of Content

What The Tech?
A Summer Internship Program

Part 4
The Basics of Data Analysis and Visualization
&
Historic Examples of Data Visualization

Team Meeting

Team Meeting

- Icebreaker
  - Any volunteers?
- Group check-in
- Volunteer opportunity
- Next week
- Today’s agenda
- Check in with the surveys

Group Check-in

- How are we feeling about the projects?
- What is something that you feel good about?
- What is something you have questions about?
- What do you want to have accomplished by the end of the week?

Any Questions?

The Basics of Data Analysis and Visualization

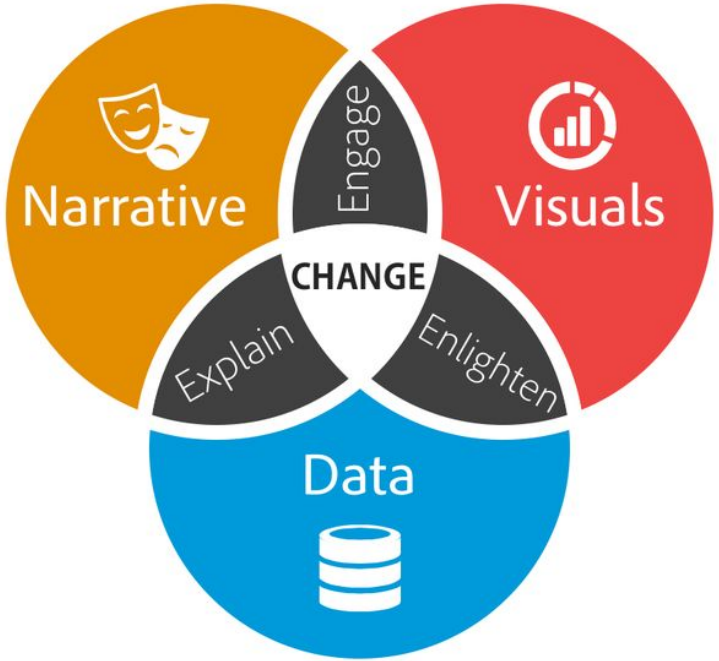
By Caleb Deitch and Confidence Oguebu, WTT? TA's!

Data Visualization and Storytelling

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps.

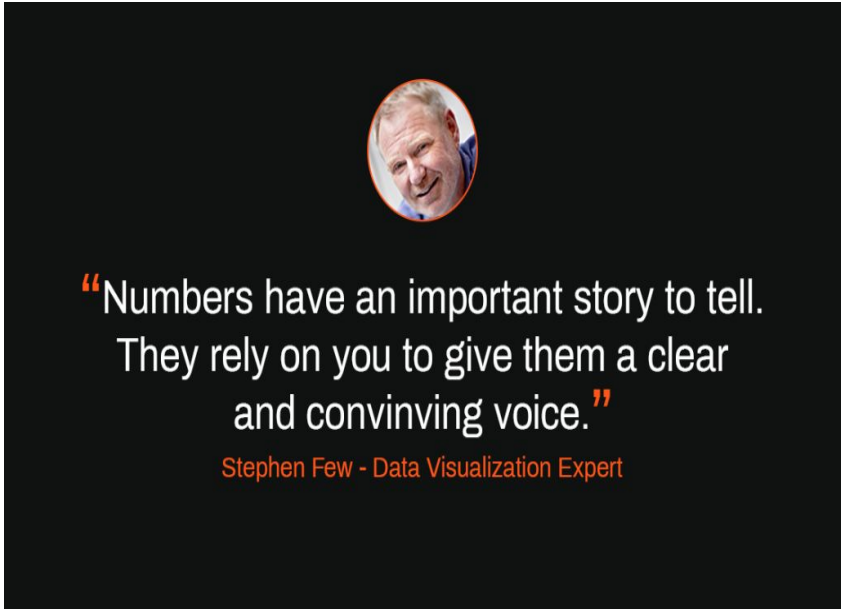
Storytelling through data implies understanding the data, doing exploratory data analysis in order to decide what story to tell, Identifying powerful and meaningful stories and Creating visualizations telling those stories.

The elements of good data storytelling are represented on the image.



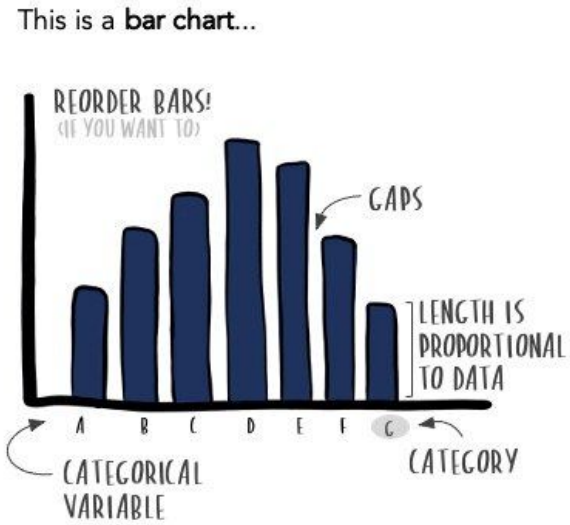
A Good Data Visual Should Have...

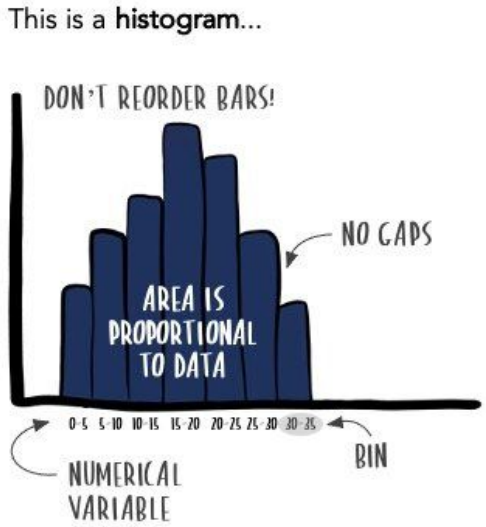
- **Clear Purpose:** Clearly define the purpose of the visualization. What question is it trying to answer? What story is it trying to tell?
- **Accurate Representation:** Ensure that the data is represented accurately. Avoid misleading scales, and other manipulations. Use proportions and scales that accurately reflect the data values.
- **Appropriate Chart Type:** Choose the right type of chart for the data.
- **Clear Labels**: Use Clear and Concise Labels, Titles and Subtitles.
- **Effective Use of Color:** Use contrasting colors to distinguish different data series or categories. Maintain color consistency.



Displaying Data

These types of graphs work for categorical data or numerical data that has been groups into bins

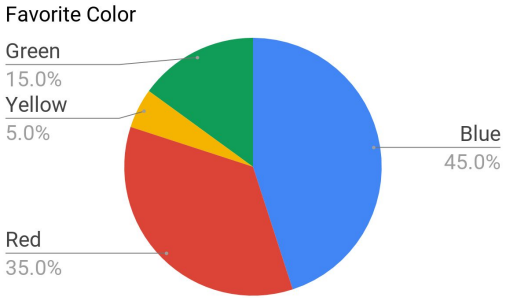




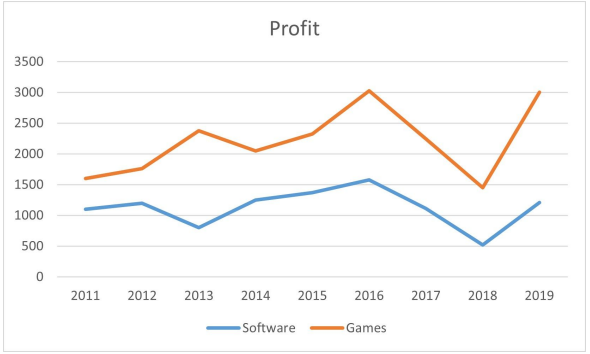
Bar Graph: displays categorical data with rectangular bars, where the length of each bar represents the value of the category.

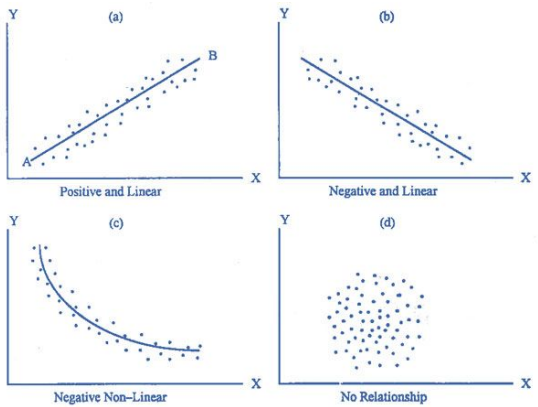
Histogram: shows the frequency distribution of a continuous dataset by grouping data into bins

Pie Chart: represents data as slices of a circular pie, where the size of each slice is proportional to the percentage of the whole.



Displaying Data



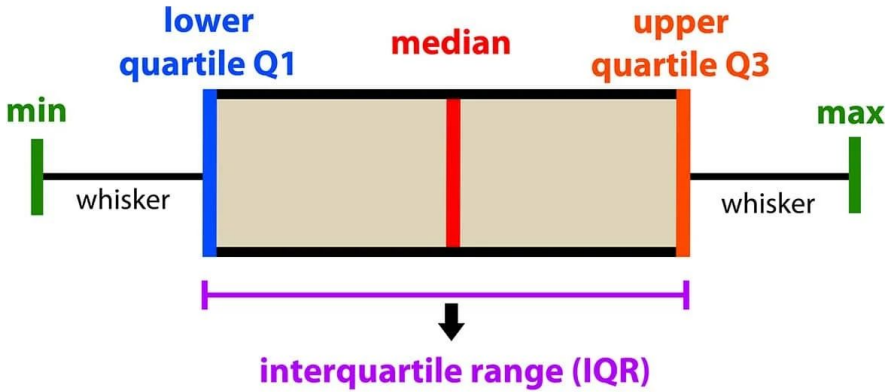


These types of graphs should only be used with numerical data

Scatter Plot: uses dots to represent the values of two different variables

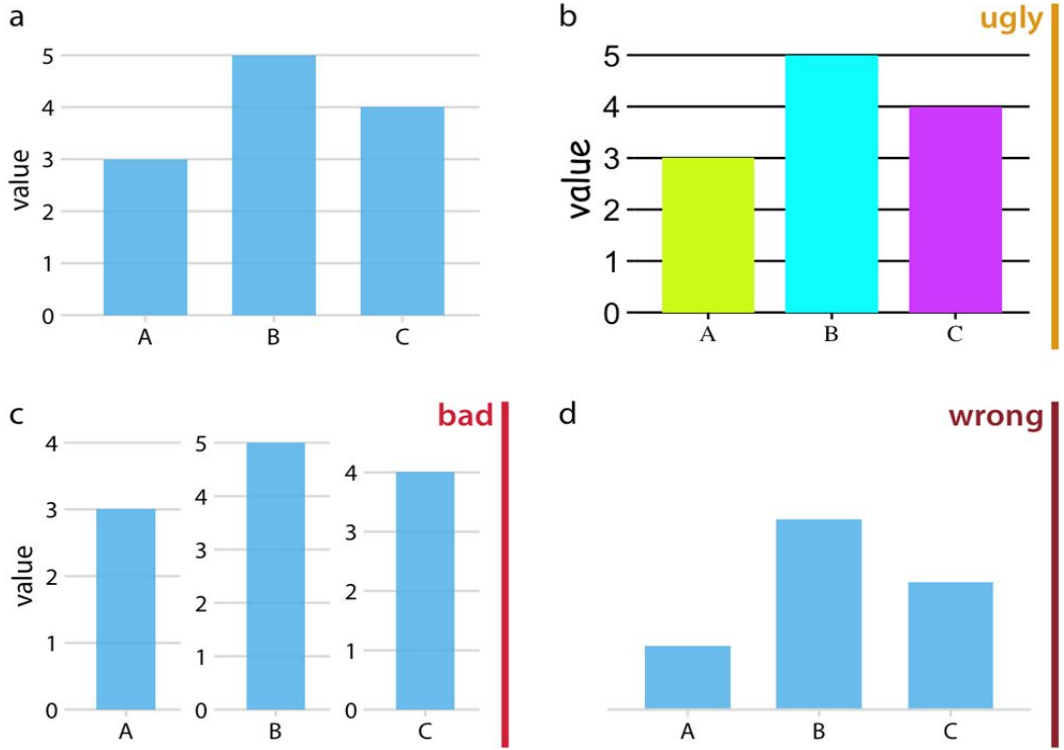
Box Plot: displays the distribution of a dataset, often highlighting outliers.

Line Graph: connects individual data points with straight lines to show trends or changes over time



What Do You Think?

The chart below represents a good visual and three visuals of poor quality. Any thoughts? Share them!



EXPORTING DATA FROM QUALTRICS

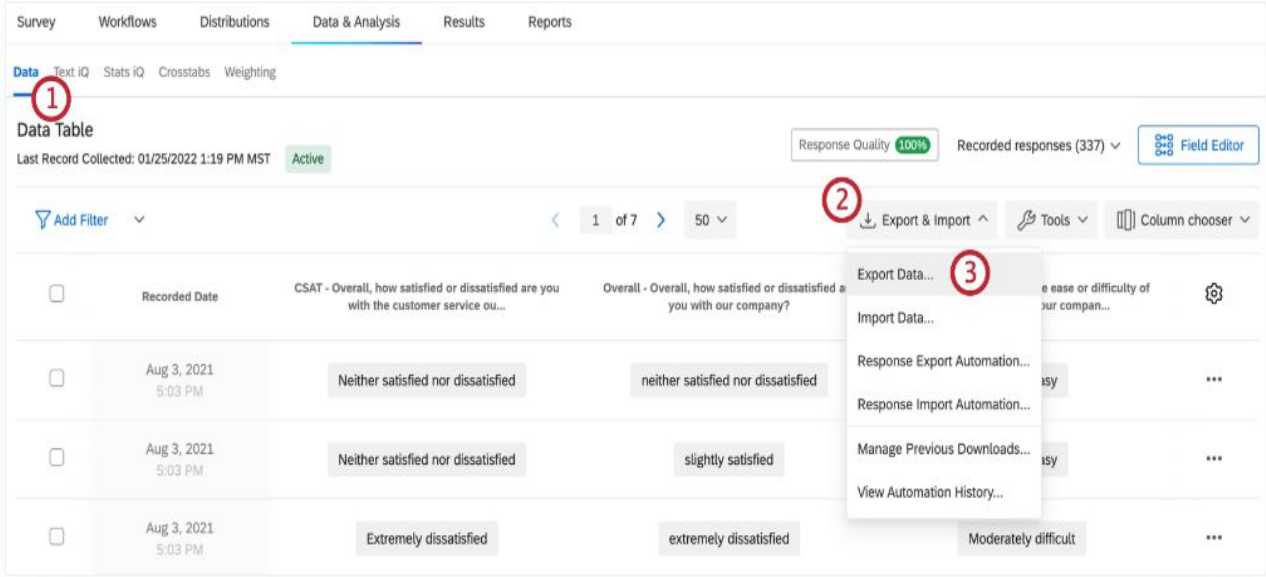
Exporting Data from a Qualtrics Survey

Tip: If you have a filter applied to data, you will export only the filtered data.

Step 1: Go to the Data section of the Data & Analysis tab.

Step 2: Click Export & Import.

Step 3: Select Export Data.



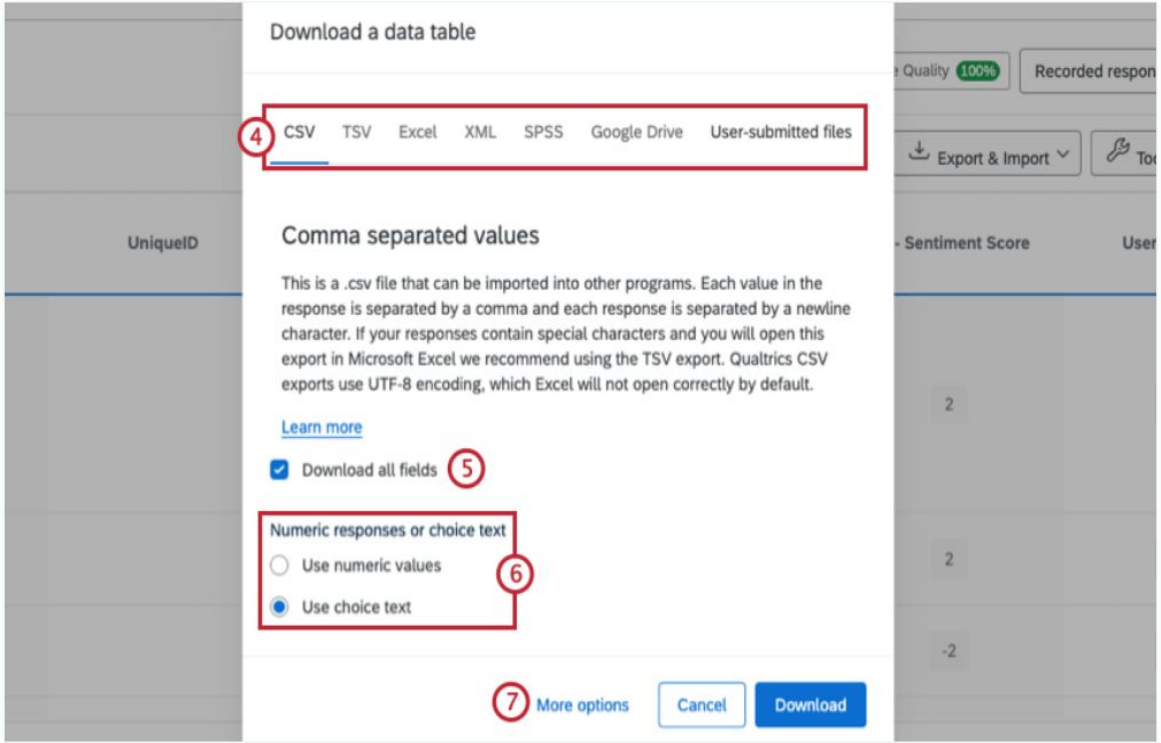
Exporting Data from a Qualtrics Survey

Step 4: Select your desired file format- CSV, TSV, Excel, XML, SPSS, Google Drive and User-submitted files.

Seep 5: Select Download all fields if you would like to export all of the data you collected.

Step 6: Decide whether you'd like to download data in a quantifiable, numeric format (Use numeric values) or with the answer choices as written (Use choice text).

Step 7: Click Download.



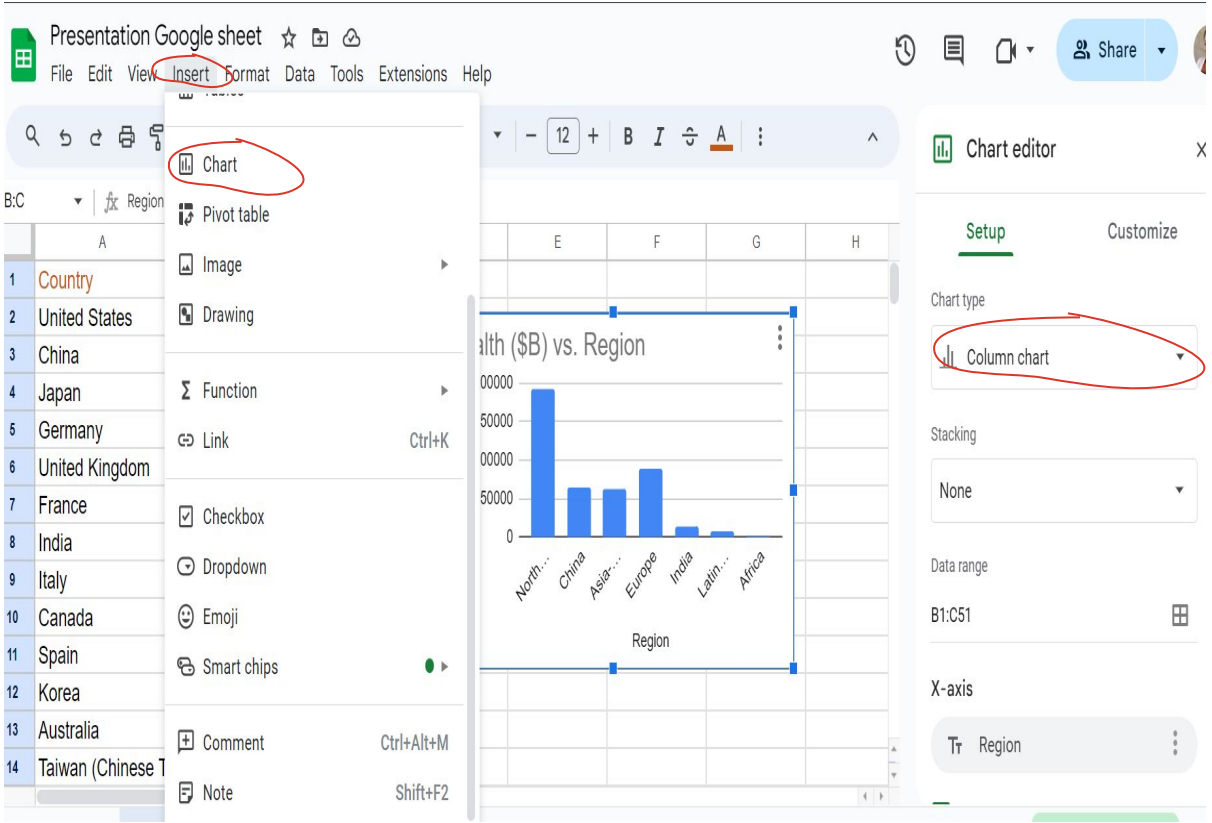
Data Visualization In Google Sheet

Step 1: Highlight the cells you want to visualize by clicking on the column letter(s).

Step 2: Click on Insert and select Chart.

Step 3: A default chart will pop up alongside the chart editor pane.

Step 4: Use the chart editor pane to select the desired/suitable chart for your visual, choose its aggregation, and fine tune the chart using labels and legends to effectively represent the data.



Multiple Choice/ Matrix

Split multi-value fields into columns changes how your data is exported (one column for each statement in matrix)

Selected:

each choice will include its own column in the dataset. On each participant's row, there will be a "1" in the columns of the choices they selected.

Eligibility - Home Chef	Eligibility - Dinnerly	Eligibility - HelloFresh
		HelloFresh
Home Chef		
Home Chef		
Home Chef	Dinnerly	

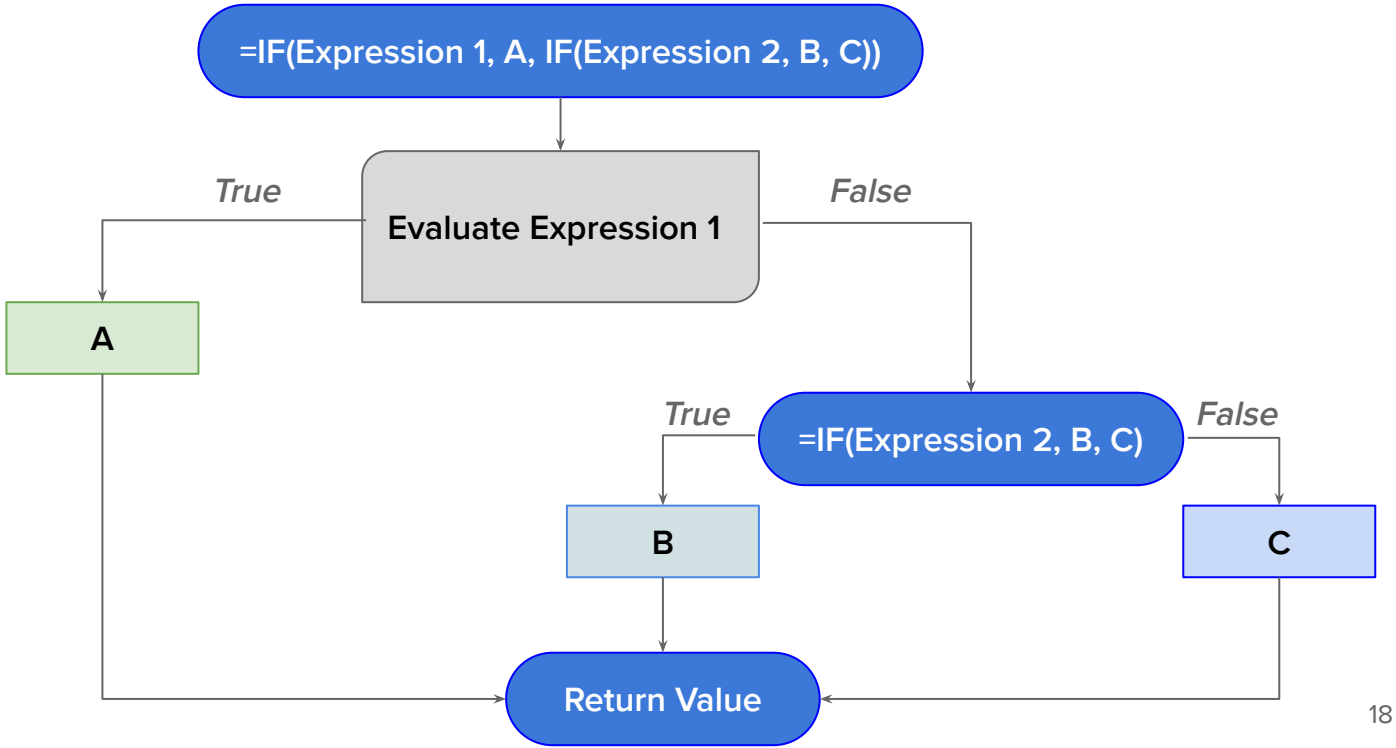
Deselected

question will have only one column in your dataset, and the values of the selected choices will be comma-separated beneath

Eligibility
YumYum
YumYum,Blue Apron,Some other brand
YumYum,Dinnerly
YumYum
YumYum,HelloFresh,Hungryroot
YumYum,HelloFresh

Nesting Functions in Google Sheets

- IF(logical _ expression, value _ if _ true, value _ if _ false)



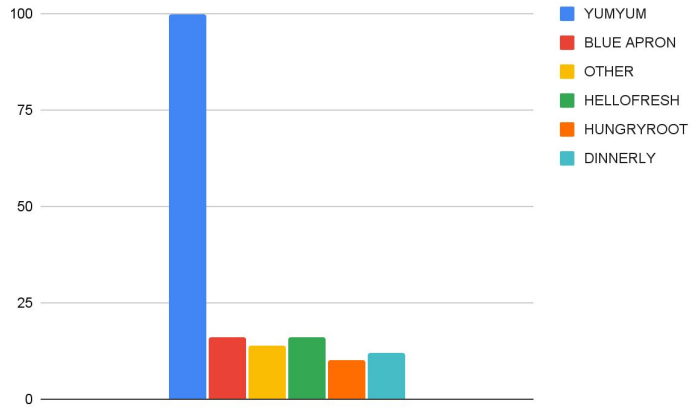
Bar Chart from choose multiple

C3 | =COUNTIF(B3:B103, "*YumYum*")

	A	B	C	D	E	F	G	H
1	ResponseId	Elig1						
2	Response ID	Eligibility	YUMYUM	BLUE APRON	OTHER	HELLOFRESH	HUNGRYROOT	DINNERLY
3	R_1TRFCbBhEzHmvjM	**YumYum**	100	16	14	16	10	12
4	R_3Qve7CmSkZCE8QK	YumYum,Blue Apron,Some other brand						
5	R_9oQz2iwW0uKRF5k	YumYum,Dinnerly						

=COUNTIF(range, condition) will find sum of records with condition true

Use “* *” to search for any case of the string (be careful)

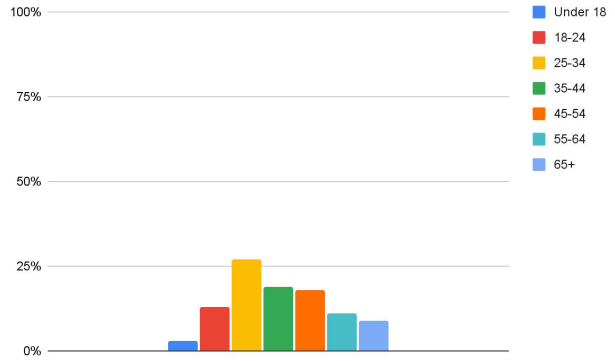


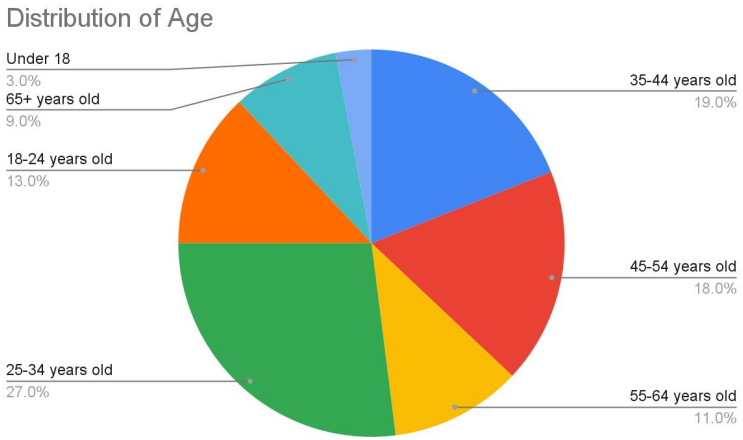
Histogram/ Pie Chart

Similar to Bar Chart

Use PercentIF() to get percentile

Make sure categories are distinct





Rank order

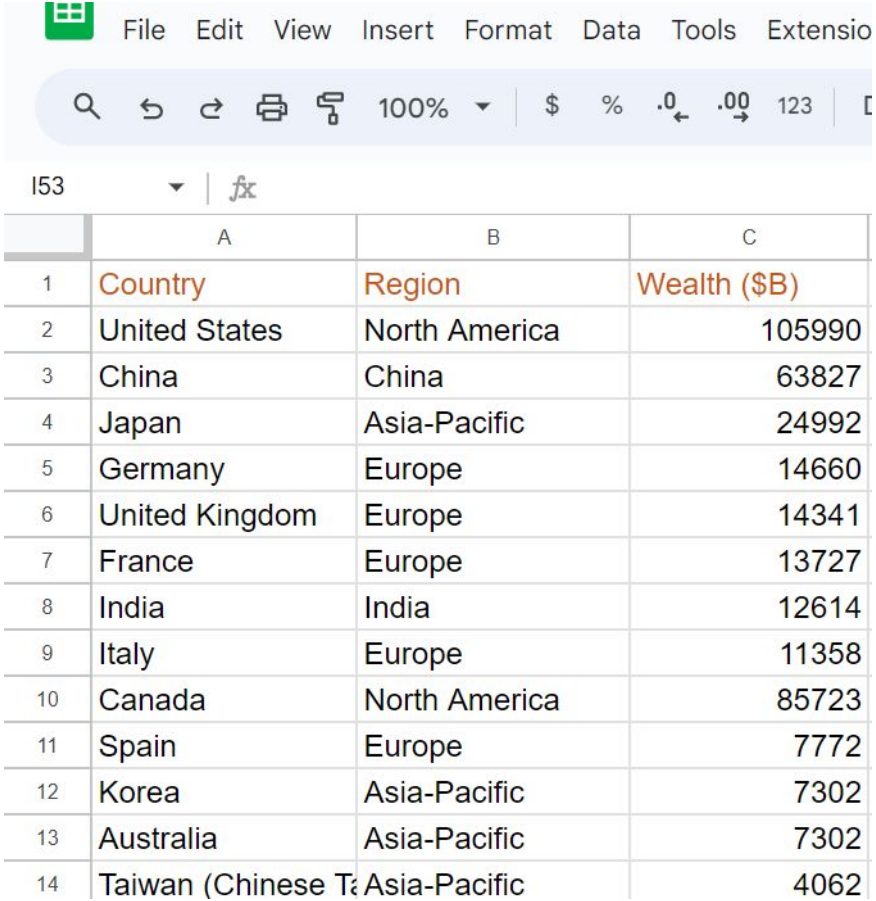
The downloaded dataset for a rank order question includes a column for every item being ranked. In each column, you'll see the ranking each participant awarded that particular item.

**** Mean, Median, Mode are important to look at for these******

Text Entry

Each text entry box in your survey will be shown as a column in your downloaded data, with a row for each response

Let's Dive In!



	A	B	C
1	Country	Region	Wealth ($B)
2	United States	North America	105990
3	China	China	63827
4	Japan	Asia-Pacific	24992
5	Germany	Europe	14660
6	United Kingdom	Europe	14341
7	France	Europe	13727
8	India	India	12614
9	Italy	Europe	11358
10	Canada	North America	85723
11	Spain	Europe	7772
12	Korea	Asia-Pacific	7302
13	Australia	Asia-Pacific	7302
14	Taiwan (Chinese Ta	Asia-Pacific	4062

- Represent the total wealth by region using a bar chart. Pick any colour of your choice and use appropriate titles/legend where necessary.
- Try doing same using a pie chart. Which of the two visuals do you think best represents the data and why?
- Now let's do something fun, using the countries and their Wealth ($B) represent this data using a line graph, a scatter plot, a bar graph, a pie chart and finally a geo map! Tell me what you think.

Content Halfway Point

One-on-One Meetings

Historic Examples of Data Visualization

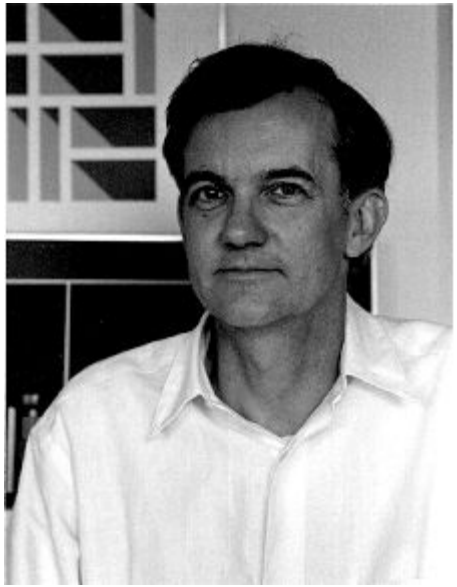
Historic Examples of Data Visualization

- So far you talked about…
  - Types of data
  - How we collect it
  - How we clean it
  - How we visualize it
  - Definitions of data visualization
  - Methodologies:
    - Charts
    - Graphs
- Now let’s see them in action, with two notable examples from the past and present.

Historic Examples of Data Visualization

- Edward Rolf Tufte
- American statistician and professor emeritus (fancy way of saying “retired) at Yale University
  - Political Science
  - Statistics
  - Computer science

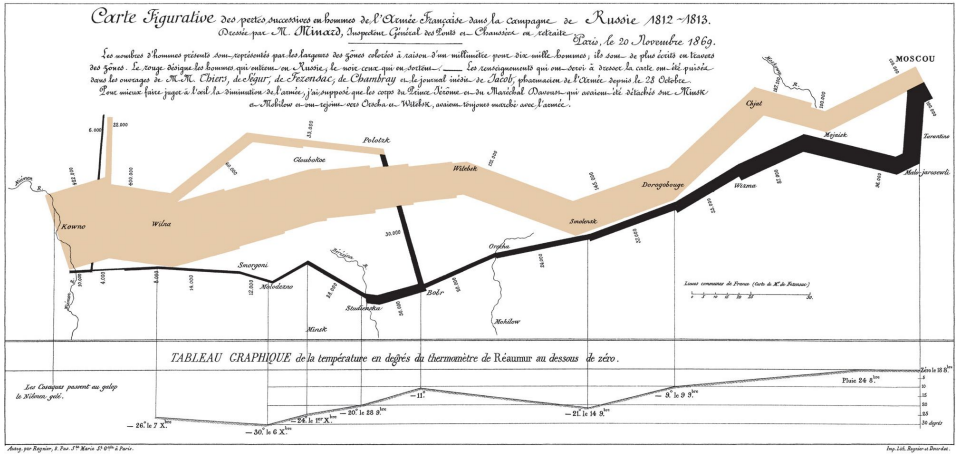
“He is noted for his writings on **information design** and as a pioneer in the field of **data visualization**.”



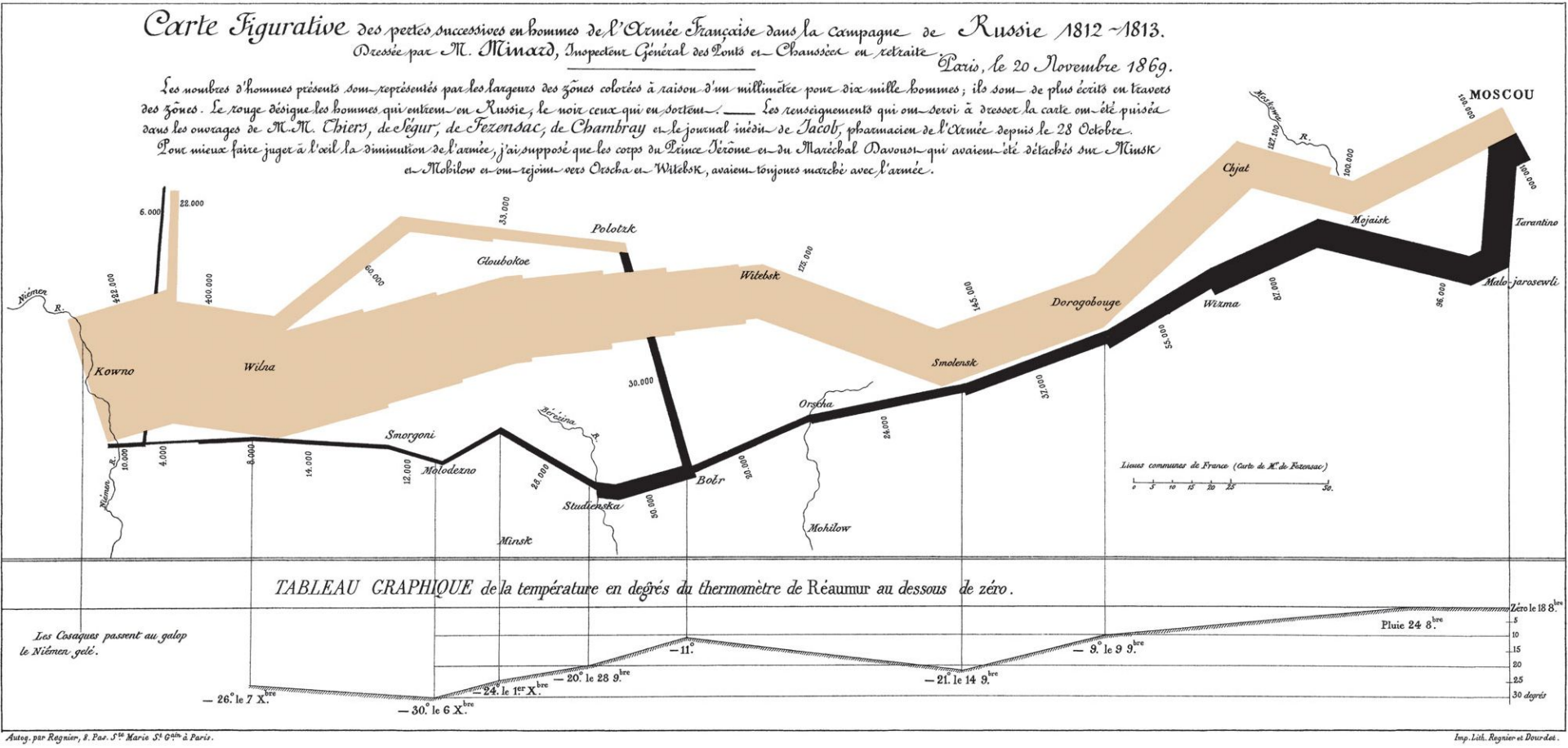
Historic Examples of Data Visualization

Data Visualization

- The practice of presenting data
- Usually it's raw data
- Many ways to organize it
  - Charts
  - Graphs
  - Matrices
  - Diagrams
  - Plots
  - Dashboards
- But it all has ONE GOAL... but what?
- **VISUALLY ENGAGING**
  - Why?



1861 illustration by Charles Joseph Minard. Depicts Napoleon's march to and retreat from Russia (1812-1813)



INFORMATION DESIGN

Historic Examples of Data Visualization

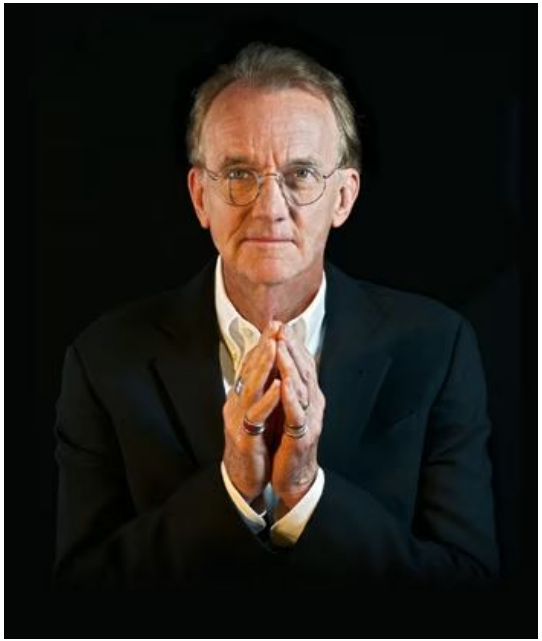
Data Visualization

- The practice of presenting data
- Usually it's raw data
- Many ways to organize it
  - Charts
  - Graphs
  - Matrices
  - Diagrams
  - Plots
  - Dashboards
- But it all has ONE GOAL... but what?
- **VISUALLY ENGAGING**
  - Why?

Information Design

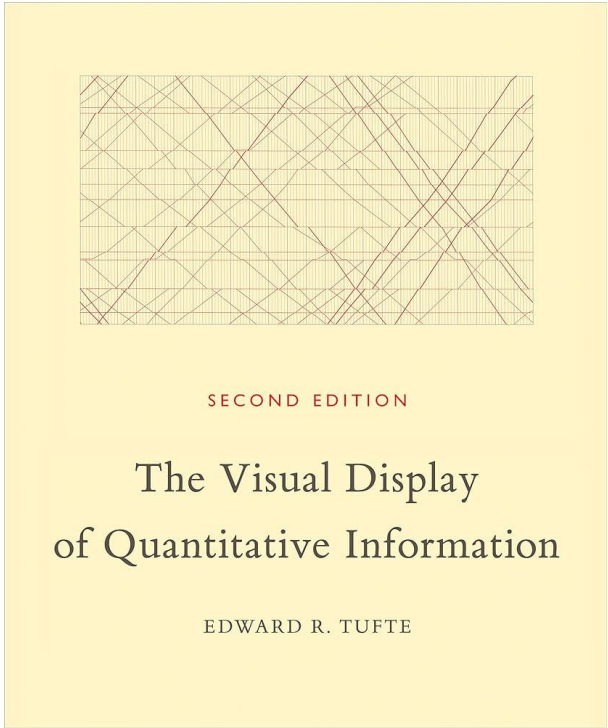
- Wikipedia: "The practice of presenting information in a way that fosters an efficient and effective understanding of the information."
- Many uses and applications across many careers
- All careers use data in some way
- Communicating that data is a key to success
- Information Design is often taught in graphic design programs

Historic Examples of Data Visualization

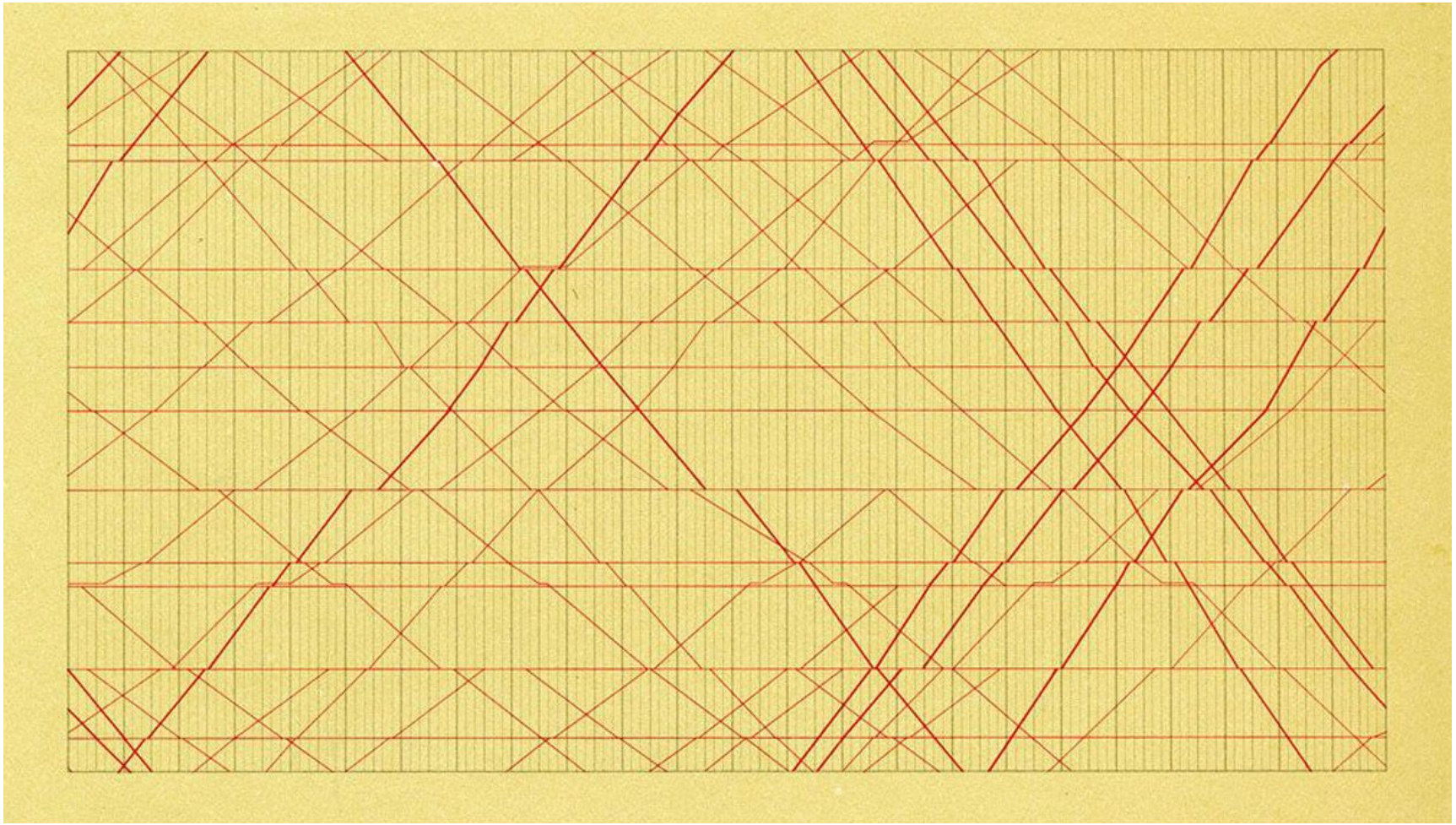


- Edward Tufte is perhaps THE big name
- His most famous book is *The Visual Display of Quantitative Information*
  - The catchiest book title of all time
- The book has become a bible of the trade

Historic Examples of Data Visualization

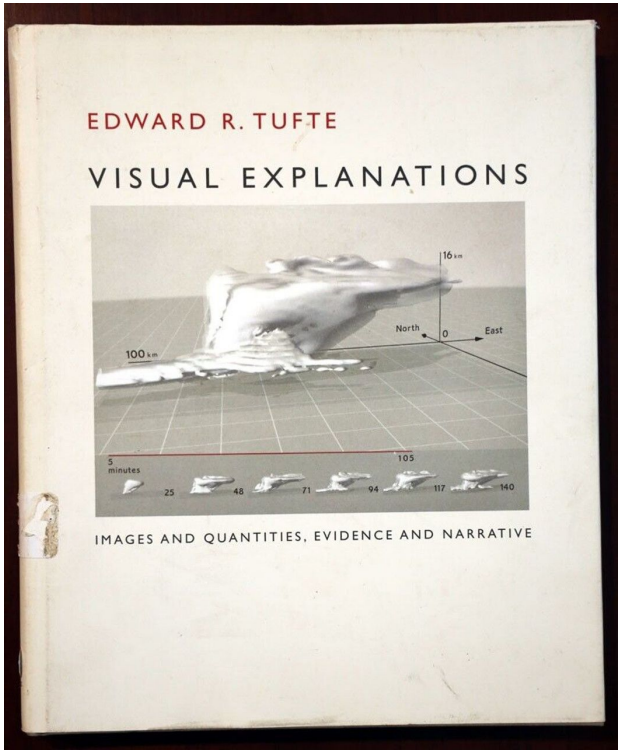


- Focuses on theory and practice of designing charts, tables, etc.
- Part history book/part manual for data display
- Topics include:
  - Data-ink ratio
  - Aesthetics
  - Do’s and don'ts of graphic design
  - Avoiding distracting clutter (*“Chartjunk,”* as Tufte calls it)



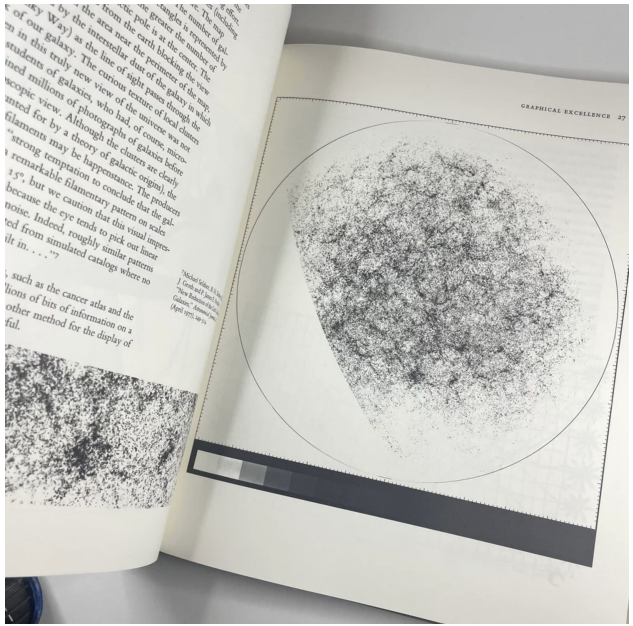
Paris-to-Lyon Train Schedule

Historic Examples of Data Visualization



- Tufte's next major book: *Visual Explanations* (1997)
  - Dan was kind enough to lend us his copy
- Focuses on (from the publisher)
  - pictures of verbs
  - the representation of mechanism and motion
  - process and dynamics
  - causes and effects
  - explanation and narrative
- Essentially a companion piece for *Quantitative Information*
- Pass the book around and ask yourselves…
  1. What survey data is most important?
  2. What story does it tell? About communities, trends, etc.
  3. How would you WANT to visualize it?
- Regroup in 15 minutes

Historic Examples of Data Visualization

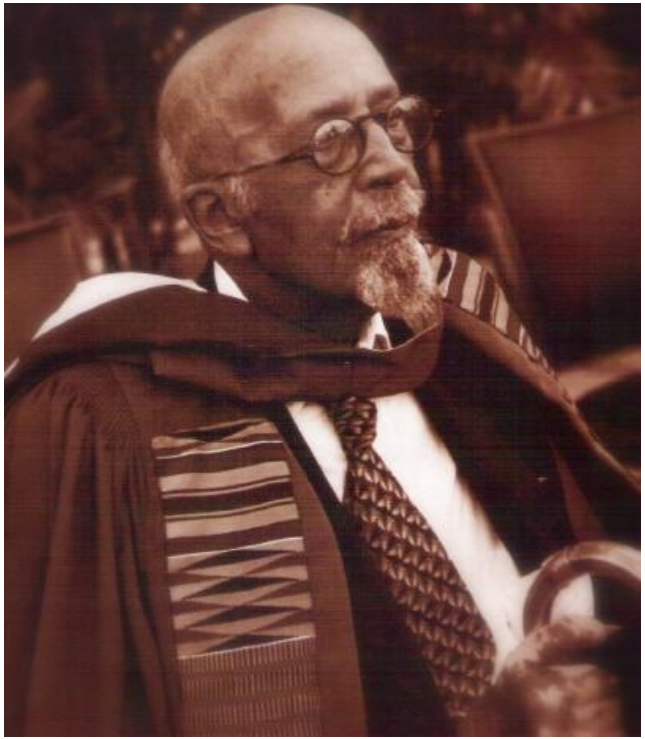


- *Visual Display* is a great book for many reasons
- One of the biggest: storytelling
- Applies to one of the most-used methods of data communication today: the slideshow
- Nancy Duarte (*slide:ology*) has 5 rules:
    1. Tell the truth,
    2. Get to the point,
    3. Pick the right tool for the job,
    4. Highlight what is important,
    5. Of course, keep it simple.
- Duarte and Tufte are right… but someone came first

Historic Examples of Data Visualization

- W.E.B. Du Bois (1868-1963) achieved many things as a civil rights activist:
  - First African American to earn a doctorate (Harvard, 1895)
  - *The Souls of Black Folk* (1903) was a landmark book that talked about Black society from a Black perspective

Historic Examples of Data Visualization



- Less known is his contribution to data science
- 1900: A cultural conference called “The Paris Exhibition, 1900” is held
  - The goal: show how people around the world live.
  - DuBois, a history professor at Atlanta University, helps develop an exhibit depicting the post-Civil War living conditions of African Americans

Historic Examples of Data Visualization

- Exhibit showcased progress and living conditions of Black people in 1900
- Exhibit included over sixty studies, graphics and photos taken by Du Bois and his team!
- These studies and graphics are in the Library of Congress
- Have uploaded to the LoC website.

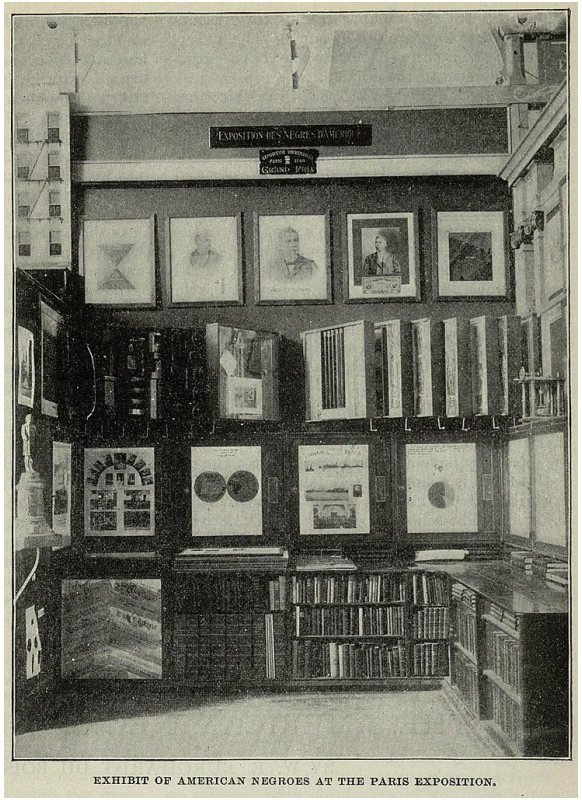
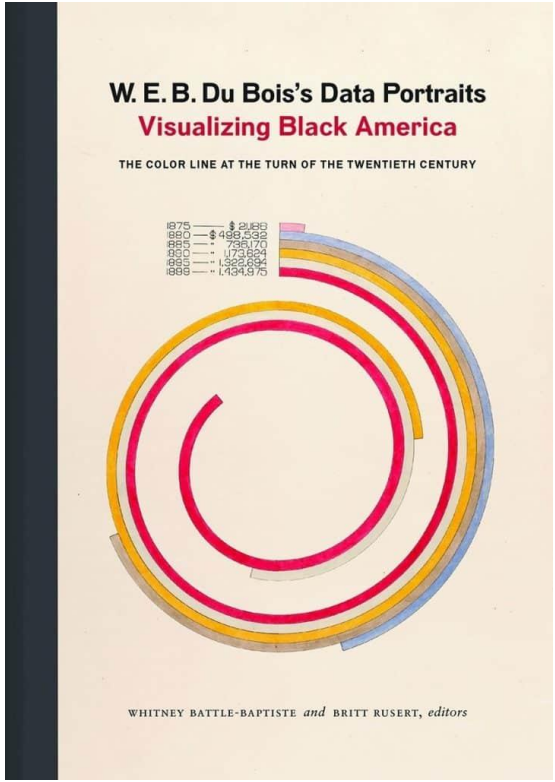


EXHIBIT OF AMERICAN NEGROES AT THE PARIS EXPOSITION.

Historic Examples of Data Visualization

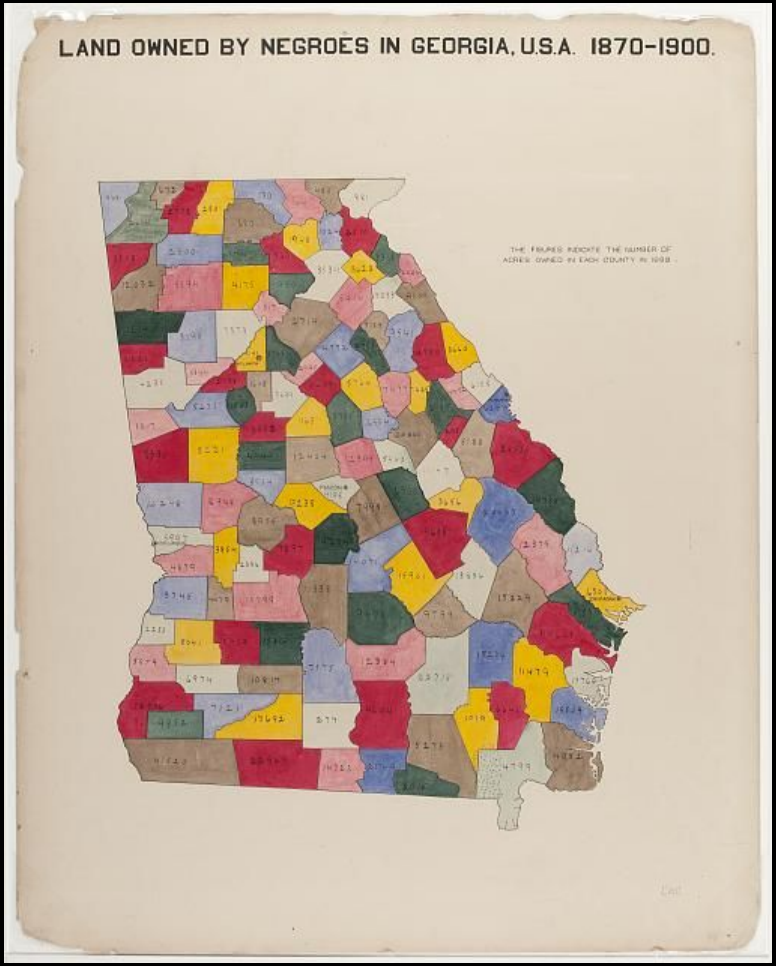


- They've also been published in the collection *W.E.B. Du Bois's Data Portraits: Visualizing Black America*
- I have a copy from Northeastern University's Snell Library

Historic Examples of Data Visualization

- These studies and graphics are in the Library of Congress
- Have also been uploaded to the LoC website.
- Feel free to open
  - **CW**: contains terms and words that are not acceptable today!
- Let's see some examples!

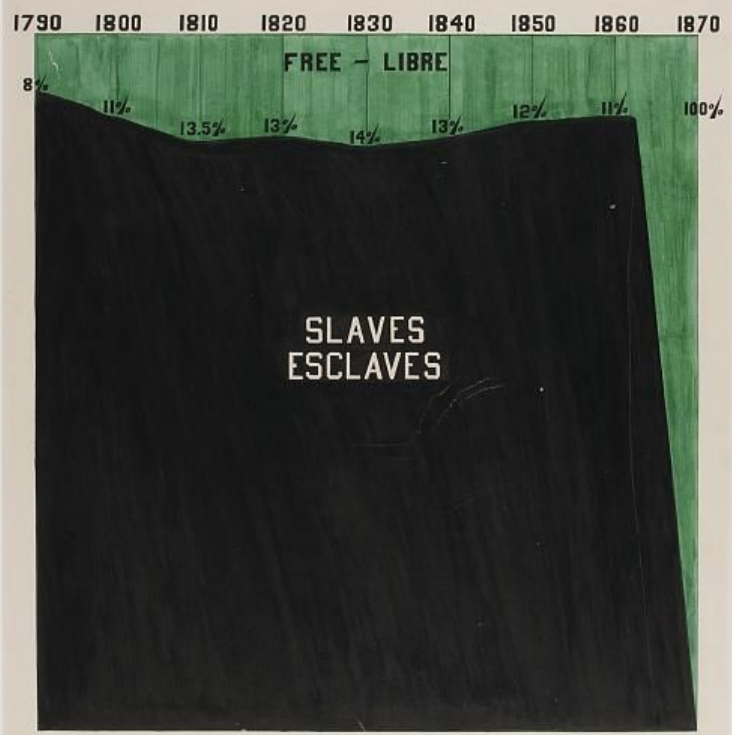
ACRES OF LAND OWNED BY NEGROES
IN GEORGIA.
1874
338,769
1875
1876
1877
1878
1879
1880
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1,062,223


LAND OWNED BY NEGROES IN GEORGIA, U.S.A. 1870-1900.
THE FIGURES INDICATE THE NUMBER OF ACRES OWNED IN EACH COUNTY IN 1900.

PROPORTION OF FREEMEN AND SLAVES AMONG AMERICAN NEGROES .

PROPORTION DES NÈGRES LIBRES ET DES ESCLAVES EN AMÉRIQUE .

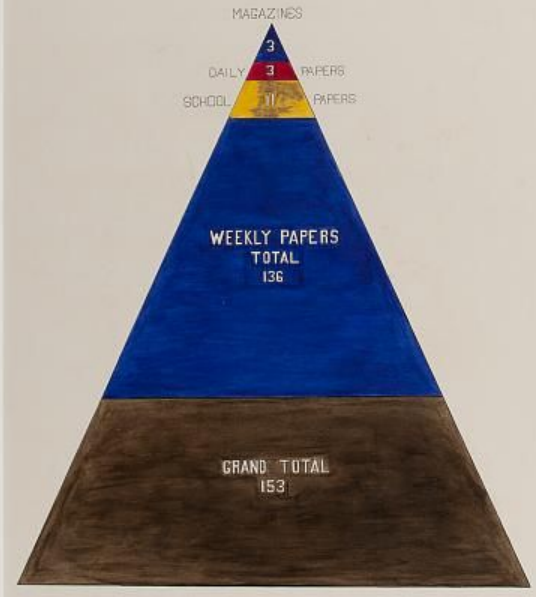
DONE BY ATLANTA UNIVERSITY .



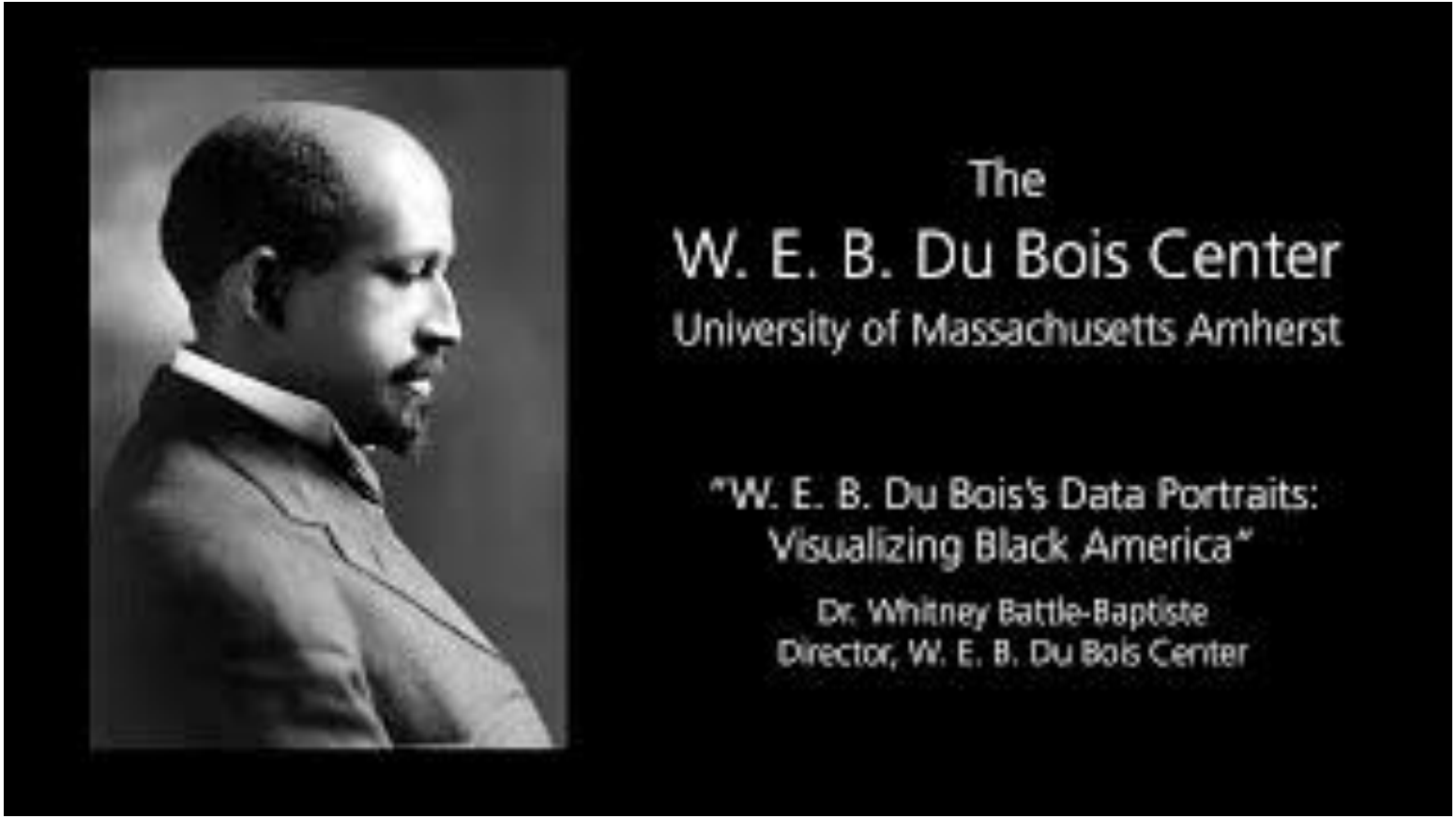
American Negro newspapers and periodicals.

Journaux et publications periodiques Nègres aux Etats Unis.

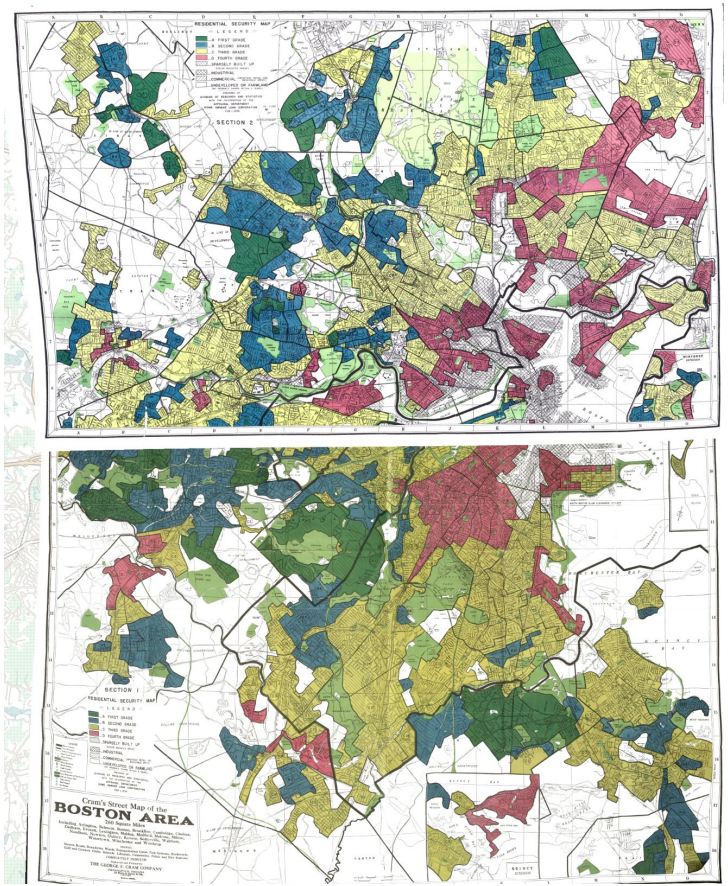
Done by Atlanta University.



Historic Examples of Data Visualization



Historic Examples of Data Visualization

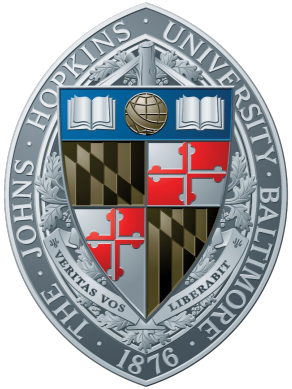


One more example: MAPS!

- Data Visualization: not just about graphs and charts
- Maps are also a method of visualizing data
- Can see data within data
  - Redlining:

"the process of denying services to residents of specific neighborhoods or communities (usually neighborhoods with non-white people) either by direct methods or by raising prices."

Historic Examples of Data Visualization



- The “Mapping Inequality” Project
- A collaboration between Virginia Tech, University of Maryland, University of Richmond and Johns Hopkins University
- Part of the **American Panorama** project that tracks american history through maps

Historic Examples of Data Visualization

https://dsl.richmond.edu/panorama/#maps

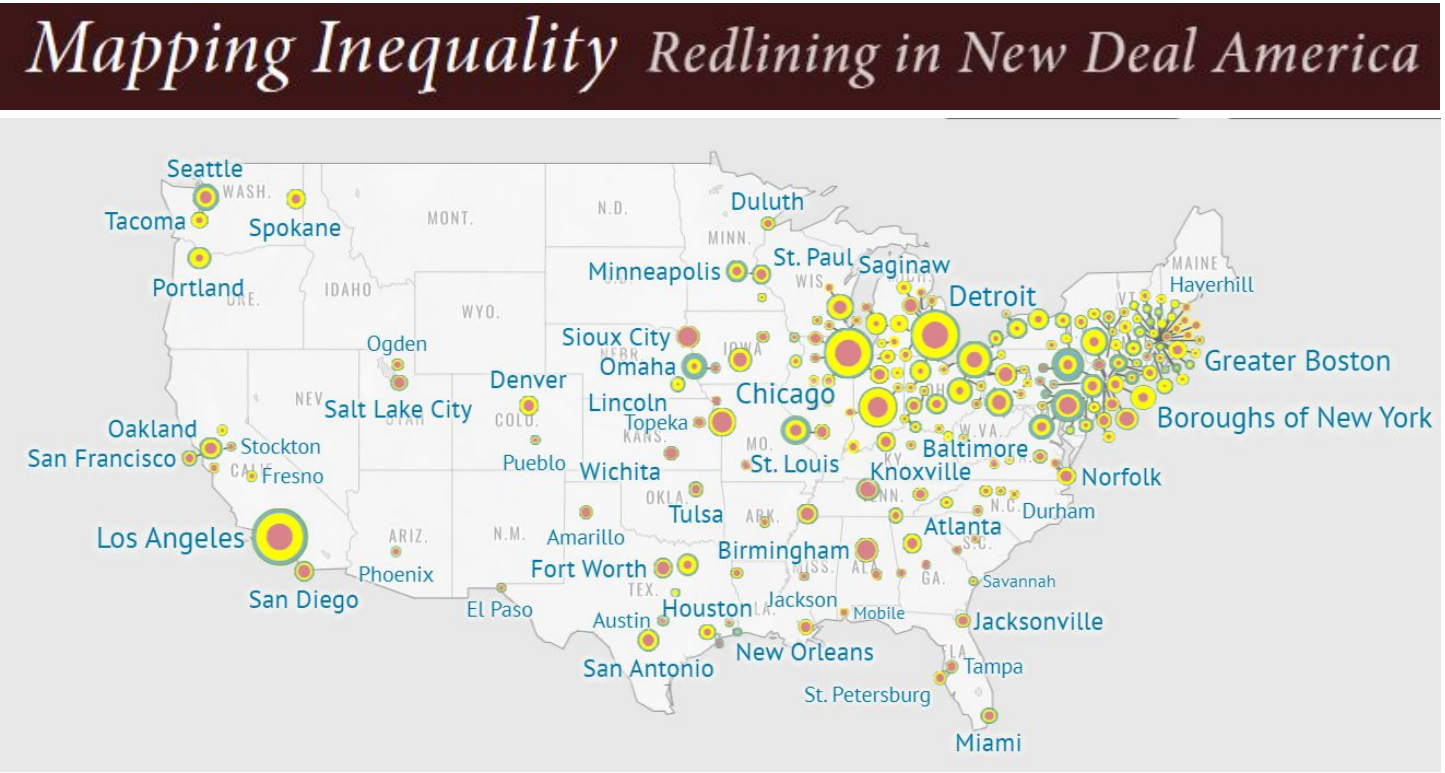
- The “Mapping Inequality” Project
- A collaboration between Virginia Tech, University of Maryland, University of Richmond and Johns Hopkins University
- Part of the **American Panorama** project that tracks american history through maps
- Feel free to open
  - **CW**: contains 1930s terms and words that are not acceptable today!

Historic Examples of Data Visualization

Background

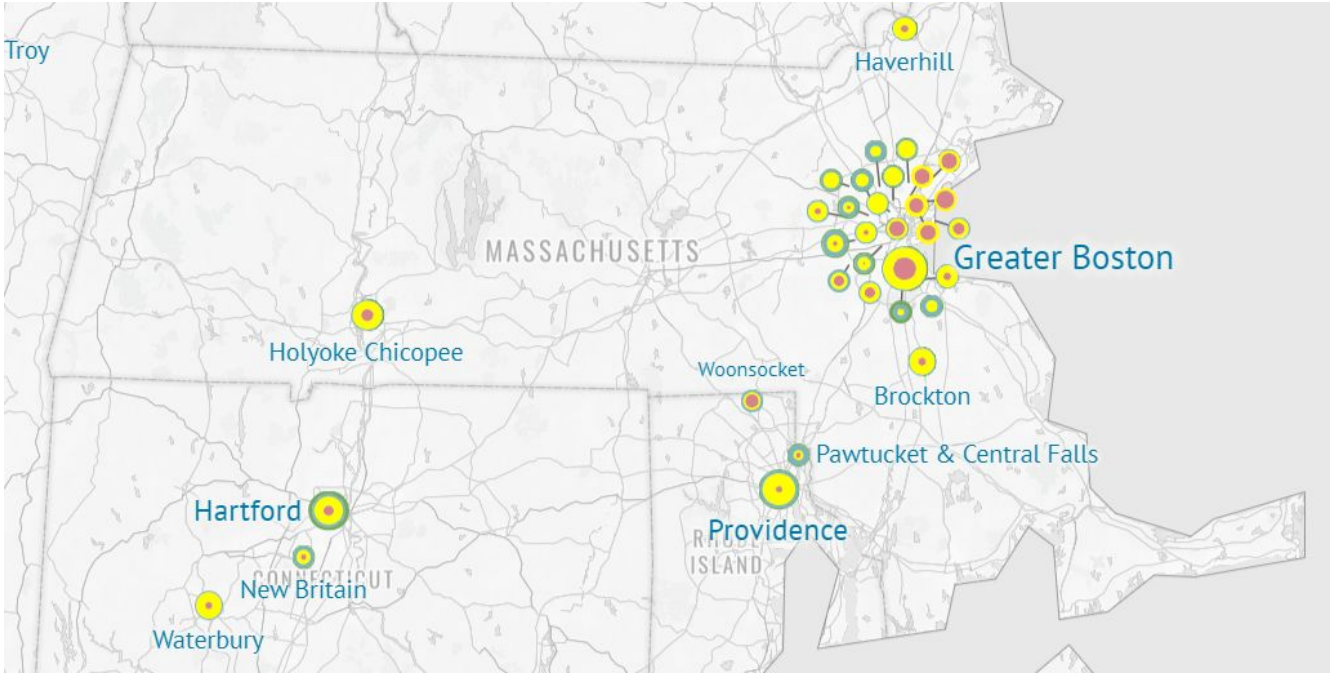
- In the 1930s, as part of the New Deal, the government creates the Home Owners Loan Corporation (HOLC).
- One of the causes of the Great Depression: collapse of the housing market
- HOLC sends agents to cities and towns to determine which areas would "safe" for banks to issue mortgages.
- The housing racism of 1937 lays the groundwork for the racism of 2023
    - Let's look at an example

Historic Examples of Data Visualization

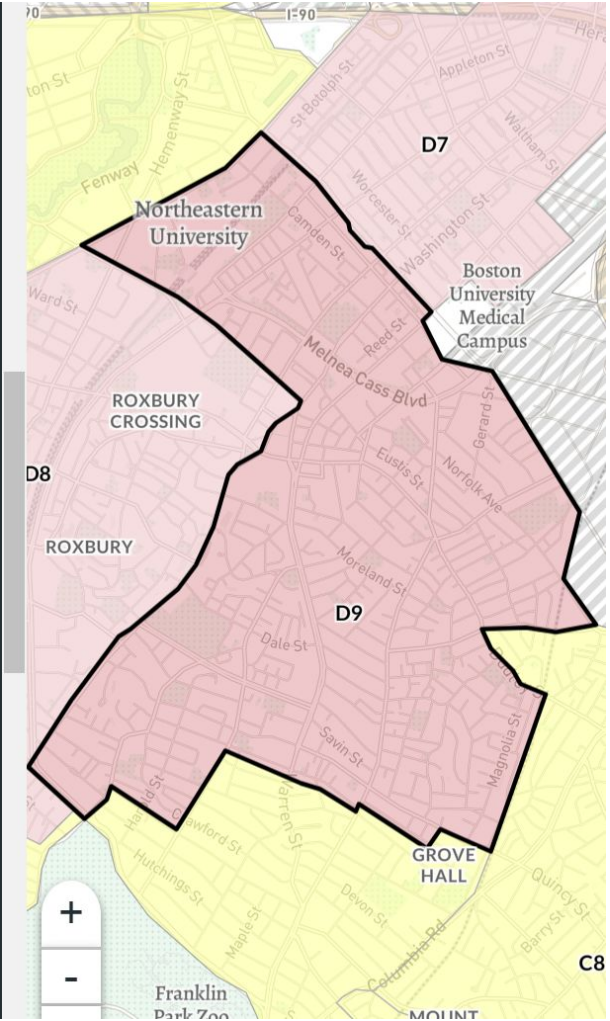


Historic Examples of Data Visualization

Roots of EJ communities can be traced to these decisions

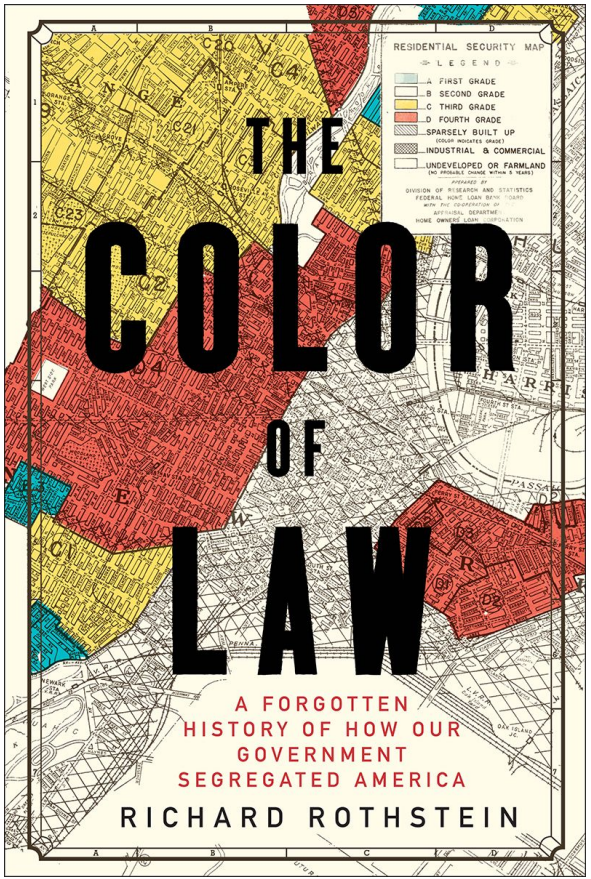


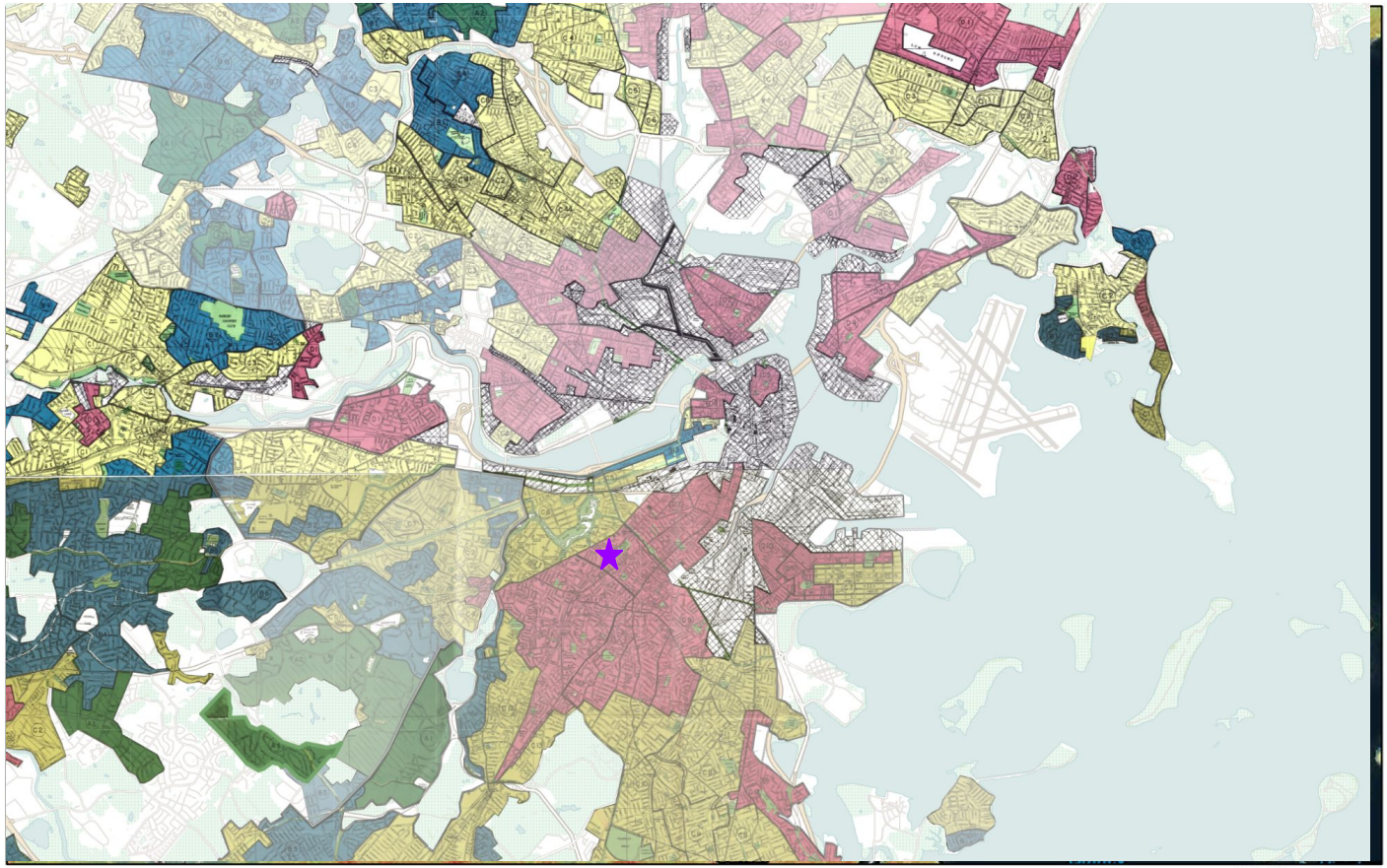
By contrast, the D9 area, in the heart of Roxbury, typified HOLC's use of racial characteristics to assess a neighborhood's investment trajectory. Here the surveyors found 50% foreign-born residents and 25% Black (the highest of any neighborhood in Boston proper), and marked the neighborhood's "cosmopolitan population" as one of its noteworthy "detrimental influences." This neighborhood displayed the hallmarks of a mixed area wherein racial attributes were taken by federal surveyors as indicators of uneven value. Black residents were "heavily concentrated" north of Ruggles Street and west of Washington Street, in what had historically been the meeting point between the Lower South End and Lower Roxbury, an area that had become the center of Boston's Black community in the first decades of the twentieth century. HOLC surveyors noted the neighborhood was "a large territory with some streets showing better experiences than the balance of the section." This tension between Roxbury's growing Black population and residents who sought the social and economic security afforded by whiteness would make Roxbury one of the key sites of white flight and racially-motivated disinvestment later in the twentieth century.



Historic Examples of Data Visualization

- This tool is important because it allows you to see…
  - How communities were divided on maps
  - How the people in charge thought (word choice, etc.)
  - What the government's priorities were
    - and think about whether or not they've changed
- Data visualization can sometimes say more than a person means to…
- Visualizations like Du Bois's are just one way we can measure and present findings.
- Programs like Excel, Tableau and others have made visualization even easier.





July 3, 2017

Study Area

High : 139.746

Low : 64.0012

Juridictions

1:135,743

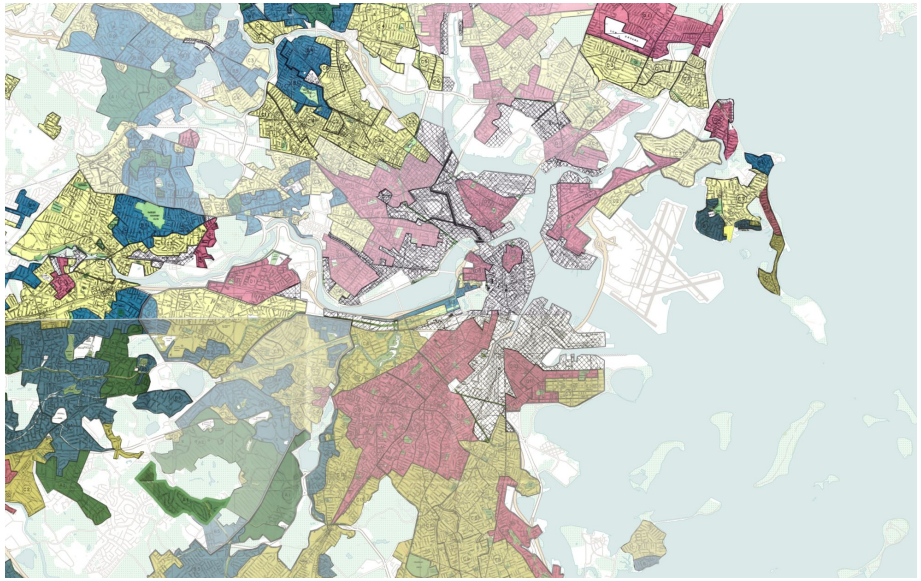
0 0.75 1.5 3 mi

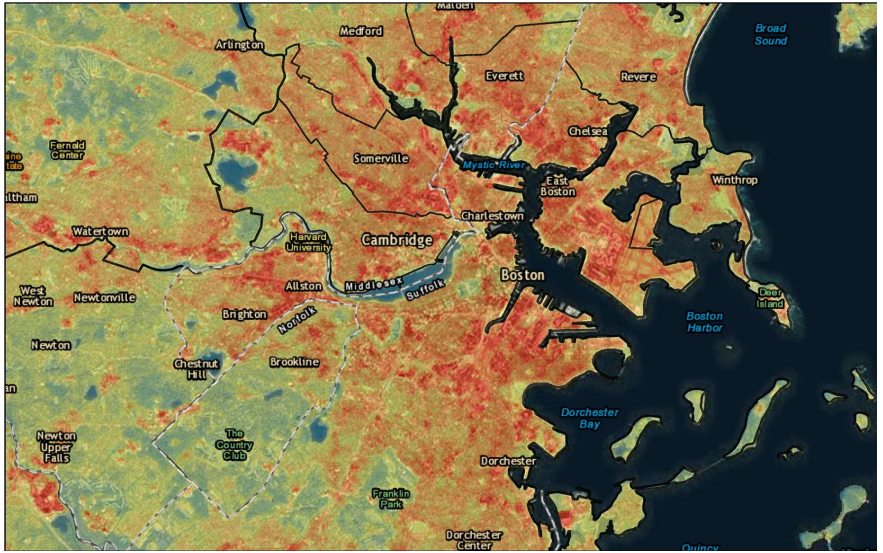
0 1.25 2.5 5 km

THE TRUST FOR PUBLIC LAND

The Trust for Public Land
Copyright 2017

Historic Examples of Data Visualization





- More to these visualizations than heat and property values
- Imagine if we could quantify the BIG ISSUES of this internship
- Where would access to them pop up?

Historic Examples of Data Visualization

Last thoughts...

- Maps, charts, bars, graphs... what do they all have in common?
- All ways to illustrate the same thing
- Application depends on circumstance

Wrap-up

Next Time

- Rough draft of report and slides
- We'll also do some practice sessions

End of Content