DSBA 5122: Visual Analytics

Class 1: Intro to Visual Analytics

Ryan Wesslen

January 14, 2019

Welcome

Introduce yourself

- Name
- What's your academic and professional background?
- Programming (e.g., R, python, JavaScript, SQL) & visualization experience (e.g., Excel, ggplot2, Tableau, Spotfire)
- What are you interested to learn more about data visualizations?

Introduce yourself

- Name
- What's your academic and professional background?
- Programming (e.g., R, python, JavaScript, SQL) & visualization experience (e.g., Excel, ggplot2, Tableau, Spotfire)
- What are you interested to learn more about data visualizations?

To do by next class: Send me an introduction email rwesslen@uncc.edu answering these questions. It's up to you on length. This is my chance to learn about your background.

https://dsba5122-spring2019.netlify.com or
https://bitly.com/dsba5122

Let's get started!

Intro to Visual Analytics (and Information Visualization)

What is a visualization?

The use of computer-supported, interactive visual representations of data to **amplify cognition**.

- Card, Mackinley, and Shneiderman (1999)

Data visualization is separated into three disciplines:

- Scientific visualization (1987)
- Information visualization (1999)
- Visual analytics (2005)

Scientific visualization (SciVis)

Volume Visualization

Flow visualization

Information Visualization (InfoVis)

- Low-level, system 1 processing
- Intuitive, perceptual decision-making
- Single, isolated graph/chart
- Typically, non-expert or average user

Information Visualization (InfoVis)

- Low-level, system 1 processing
- Intuitive, perceptual decision-making
- Single, isolated graph/chart
- Typically, non-expert or average user

Goal: identify best external (visual) representation

InfoVis: Harrison et al., 2014

Which visualization best aids in a "just-noticeable" difference in correlation?

InfoVis: Harrison et al., 2014

Visual Analytics (VA)

- High-level, system 2 processing
- Effortful, deliberate decision-making
- Highly interactive, coordinated-multiple views
- Typically, expert or specialized knowledge

Visual Analytics (VA)

- High-level, system 2 processing
- Effortful, deliberate decision-making
- Highly interactive, coordinated-multiple views
- Typically, expert or specialized knowledge

Goal: "human-in-the-loop" decision-making (e.g., integrate with machine learning)

Visual Analytics (VA)

- High-level, system 2 processing
- Effortful, deliberate decision-making
- Highly interactive, coordinated-multiple views
- Typically, expert or specialized knowledge

Goal: "human-in-the-loop" decision-making (e.g., integrate with machine learning)

This course is called "Visual Analytics" but includes a lot on Information Visualization.

VA: VAiROMA by Cho et al., 2015

How to visualize Roman History from Wikipedia?

VA: VAiROMA by Cho et al., 2015

A *Very* Brief History of Data Visualizations

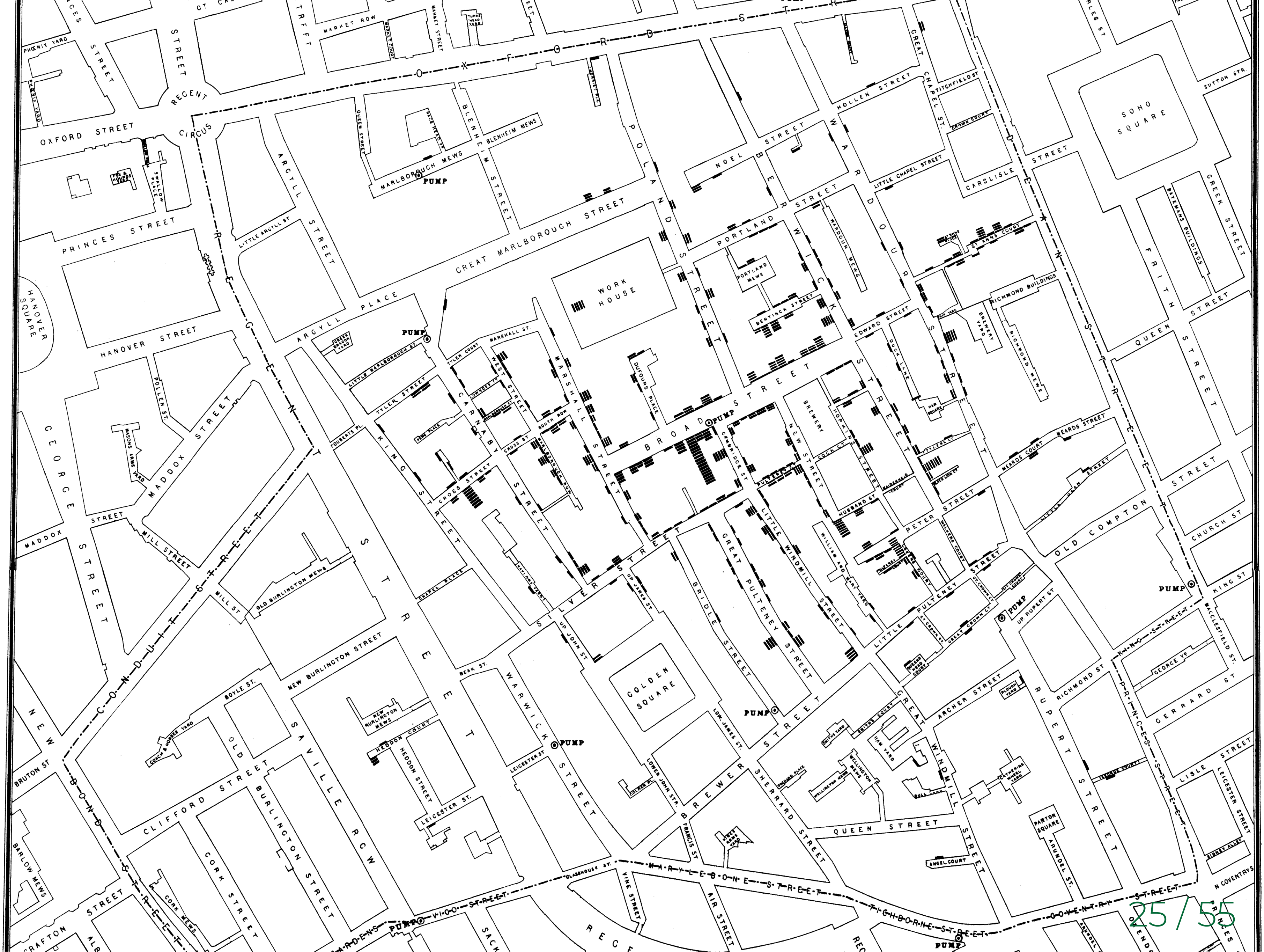
William Playfair, 1786

Cholera in London: 1854

John Snow

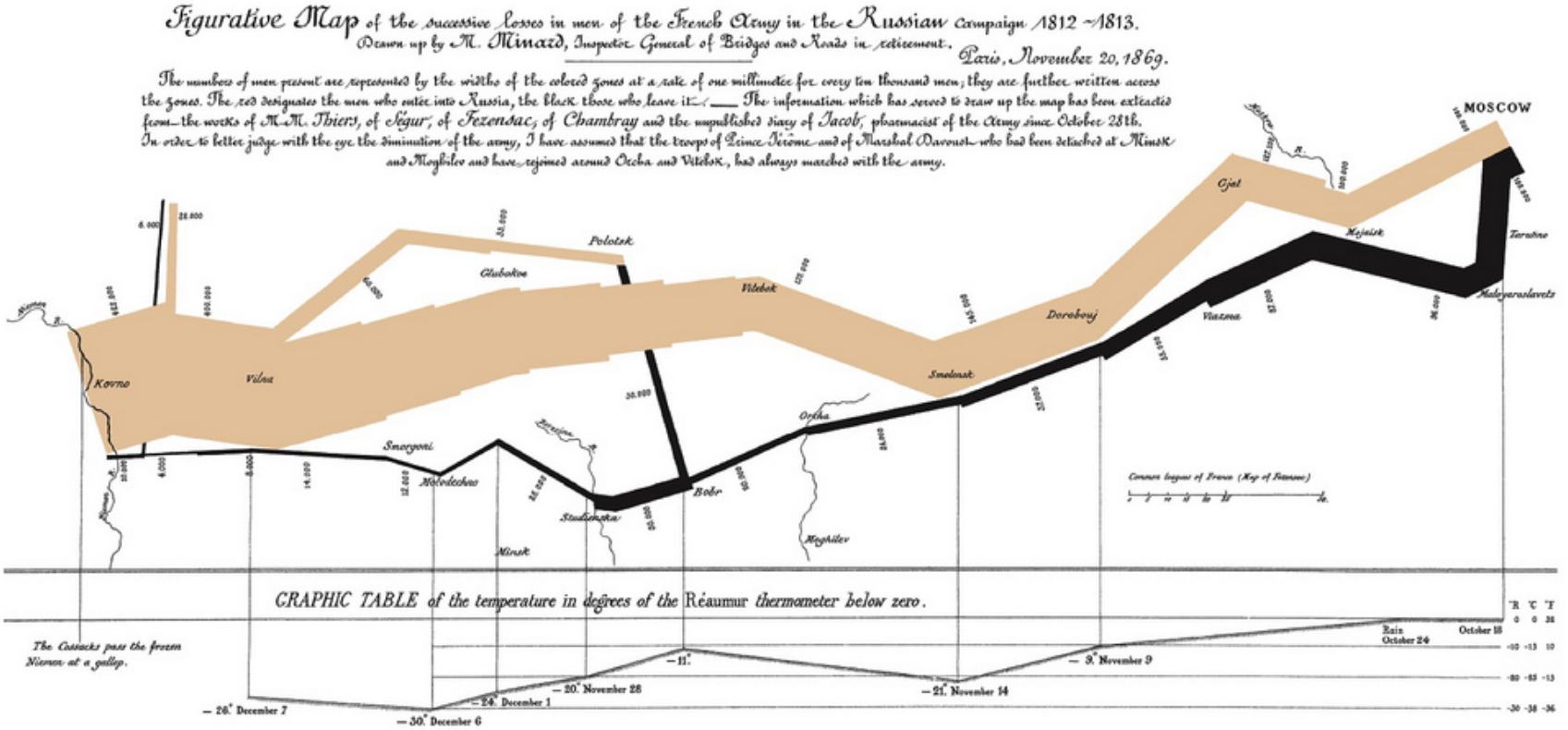
John Snow

The other John Snow


OXFORD STREET
REGENT CIRCUS
REGENT STREET
PRINCES STREET
HANOVER SQUARE
HANOVER STREET
GEORGE STREET
MADDOX STREET
MILL STREET
CONDUIT STREET
NEW BOND STREET
BRUTON ST
BARLOW MEWS
GRAFTON STREET
CLIFFORD STREET
OLD BURLINGTON STREET
OLD BURLINGTON MEWS
NEW BURLINGTON STREET
NEW BURLINGTON MEWS
BOYLE ST.
SAVILLE ROW
CORK STREET
CORK MEWS
VIGO STREET
ARGYLL STREET
LITTLE ARGYLL ST
ARGYLL PLACE
GREAT MARLBOROUGH STREET
MARLBOROUGH MEWS
BLENHEIM MEWS
BLENHEIM STREET
QUEEN STREET
POLAND STREET
NOEL STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WORK HOUSE
DUFOURS PLACE
BROAD STREET
MARSHALL STREET
MARSHALL ST.
WEST STREET
CARNABY STREET
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
KING STREET
CROSS STREET
CROSS ST
SOUTH ROW
SILVER STREET
CHAPEL PLACE
BEAK ST.
WARWICK STREET
UP. JOHN ST
UP. JAMES ST
GOLDEN SQUARE
LOW. JAMES ST
LOWER JOHN STR.
BRIDLE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
HUSBAND ST
PETER STREET
WILLIAM AND MARY YARD
LITTLE PULTENEY STREET
GREAT WINDMILL STREET
GREAT CROWN CT.
UP. RUPERT ST
RUPERT STREET
ARCHER STREET
BREWER STREET
SHERRARD STREET
QUEEN STREET
HEDDON COURT
HEDDON STREET
LEICESTER ST.
FRANCIS ST
AIR STREET
VINE STREET
GLASSHOUSE ST.
MARYLEBONE STREET
TICHBORNE STREET
COVENTRY STREET
N COVENTRY
ANGEL COURT
ARUNDEL ST.
PANTON SQUARE
WARDOUR STREET
WARDOUR MEWS
BERWICK STREET
EDWARD STREET
DUCK LANE
HOLLEN STREET
LITTLE CHAPEL STREET
GREAT CHAPEL ST.
TITCHFIELD ST
CARSLISLE STREET
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
MEARDS STREET
MEARDS COURT
DEAN STREET
SOHO SQUARE
SUTTON STR
BATEMANS BUILDINGS
GREEK STREET
FRITH STREET
QUEEN STREET
CHURCH ST
OLD COMPTON STREET
KING ST
MACCLESFIELD ST.
GEORGE YD
GERRARD STREET
PRINCES STREET
RICHMOND ST
LISLE STREET
LEICESTER STREET
NEWPORT ST
PUMP

Tufte's Perspective

1. Place data in an appropriate context for assessing **cause and effect**.
2. Making quantitatives comparisons (e.g., Workhouse & Brewery).
3. Considering alternative explanations and contrary cases.
4. Assessment for possible errors (e.g., compared with what?)



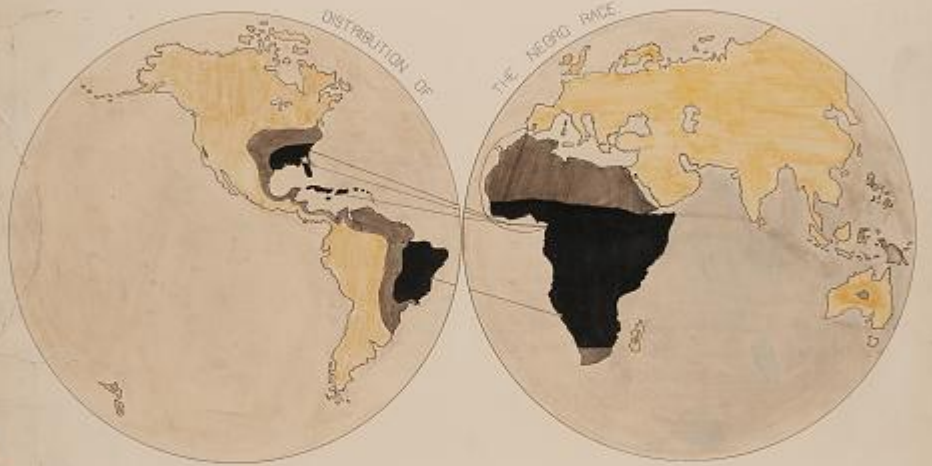
Napoleon's March to Moscow - Charles Joseph Minard - 1869

THE GEORGIA NEGRO.

A SOCIAL STUDY

BY

W.E.BURGHARDT DU BOIS.



≡ ROUTES OF THE AFRICAN SLAVE TRADE.

● THE STATE OF GEORGIA.

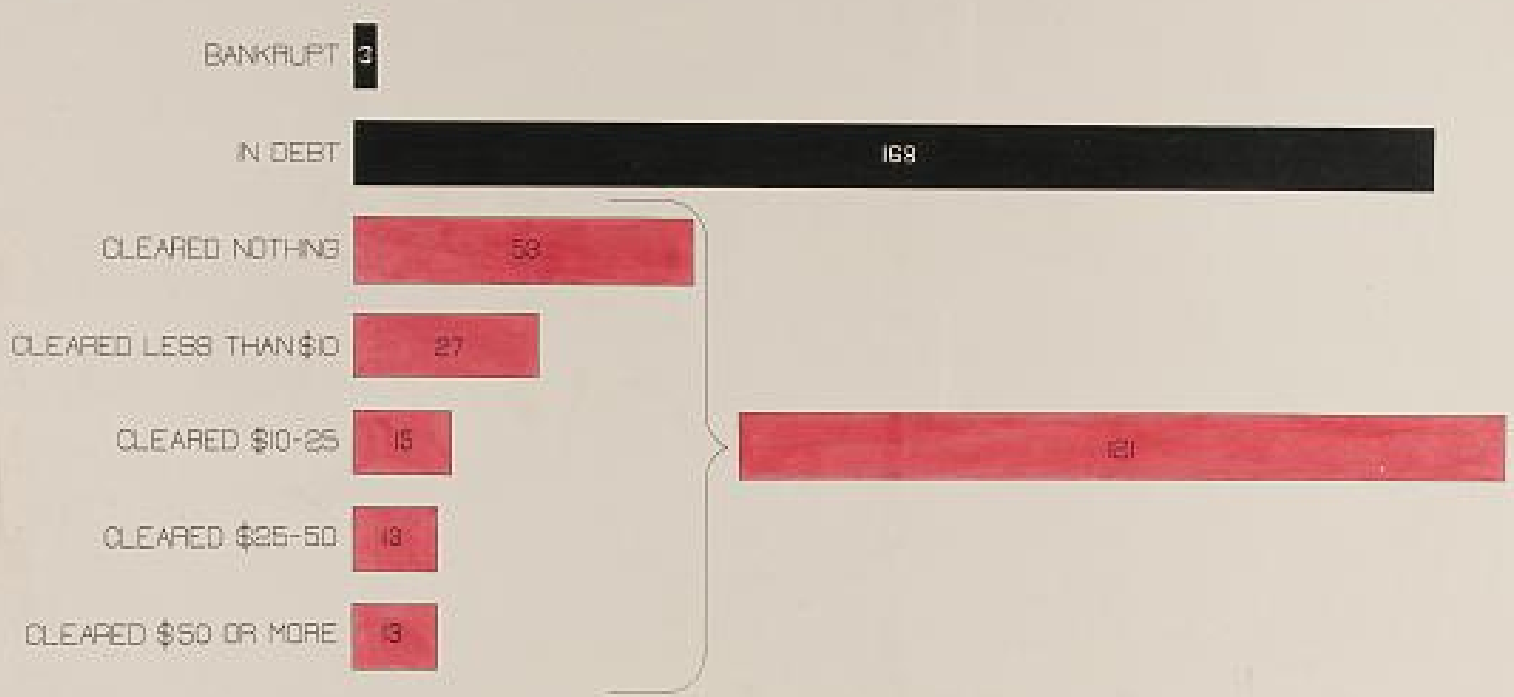
THIS CASE IS DEVOTED TO A SERIES OF CHARTS, MAPS AND OTHER DEVICES DESIGNED TO ILLUSTRATE THE DEVELOPMENT OF THE AMERICAN NEGRO IN A SINGLE TYPICAL STATE OF THE UNITED STATES.

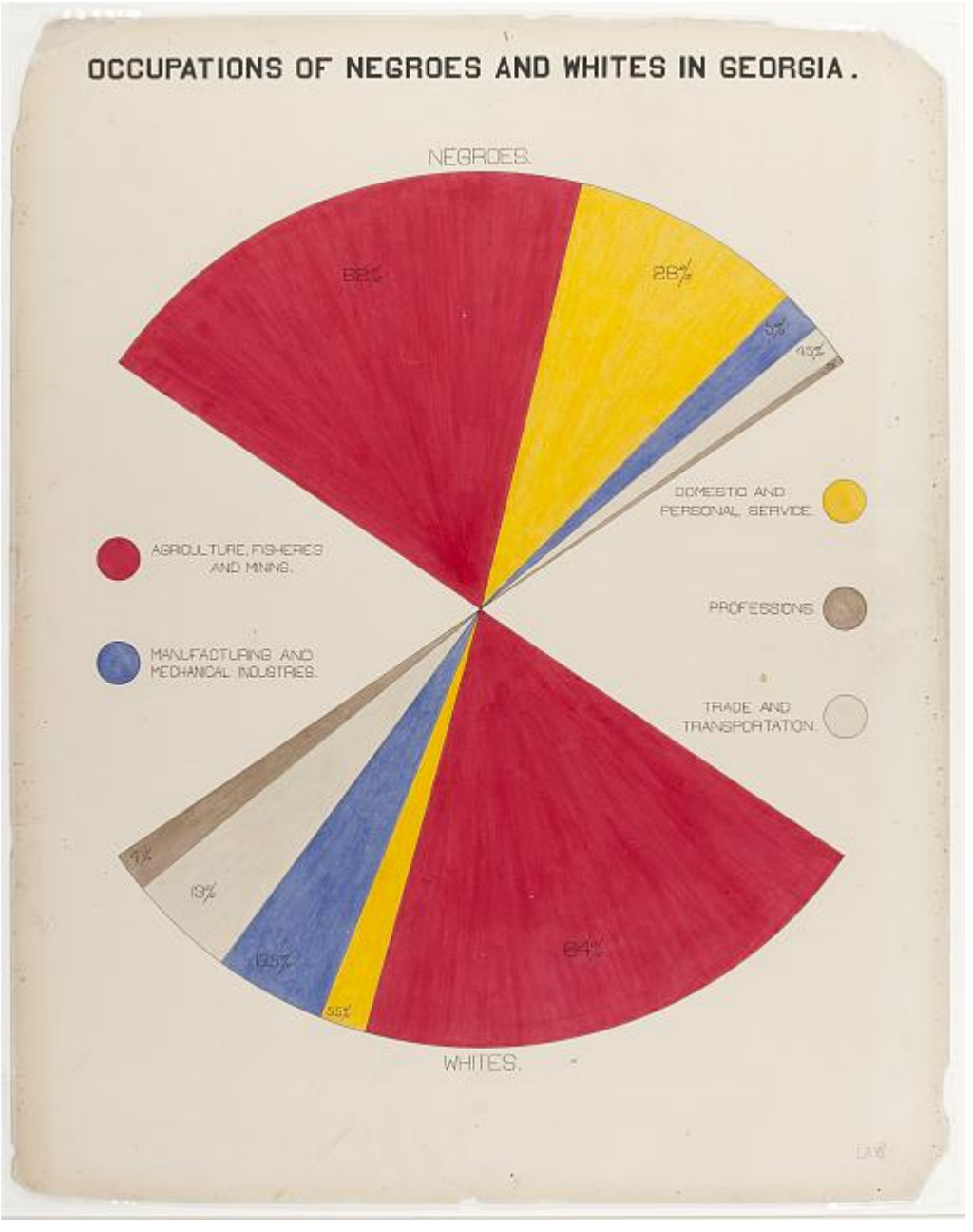
" THE PROBLEM OF THE 20TH CENTURY IS THE PROBLEM OF THE COLOR-LINE."

RELATIVE NEGRO POPULATION OF THE STATES OF THE UNITED STATES.

750,000 NEGROES AND OVER
600,000—750,000
500,000—600,000
300,000—500,000
200,000—300,000
100,000—200,000
50,000—100,000
25,000—50,000
10,000—25,000
UNDER—10,000

CONDITION OF 300 NEGRO FARM TENANTS AFTER 1 YEAR'S TOIL.1898 .




OCCUPATIONS OF NEGROES AND WHITES IN GEORGIA.
NEGROES.
AGRICULTURE, FISHERIES AND MINING.
MANUFACTURING AND MECHANICAL INDUSTRIES.
DOMESTIC AND PERSONAL SERVICE.
PROFESSIONS.
TRADE AND TRANSPORTATION.
WHITES.

Watch Jeff Heer's 2009 "Brief History of Data Visualization".

Why look at Data?

"The greatest value of a picture is when it forces us to notice what we never expected to see." — John Tukey

Anscombe's Quartet (1973), Healy: Ch. 1

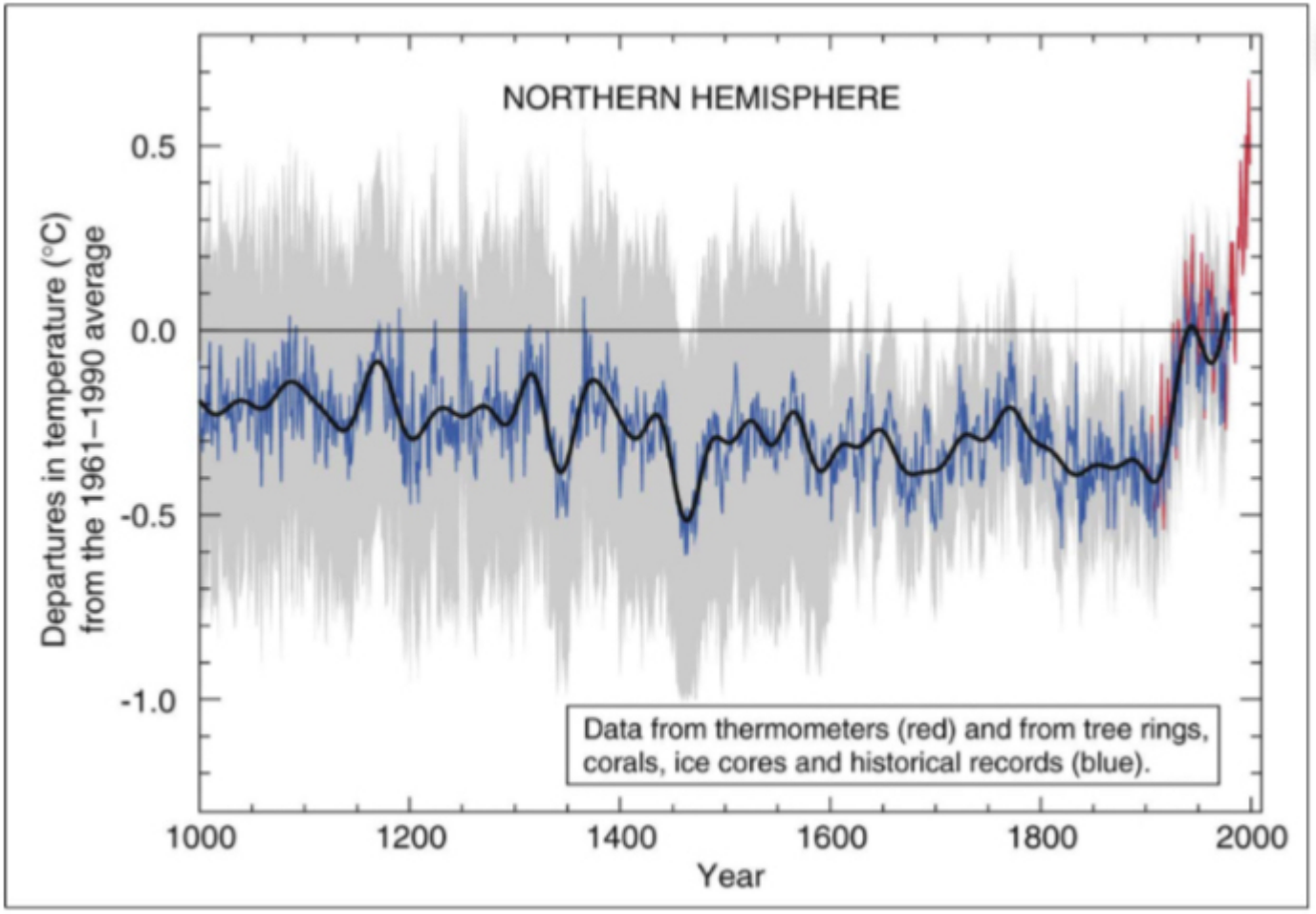
Datasaurus Dozen

Matejka and Fitzmaurice, 2017 / GitHub

Five Qualities of Great Visualizations

"But why do I never stumble on to an article or blog post showing ... how a visualization helped a group of doctors do something remarkable? Is it just because this stuff does not get reported or what?" -Enrico Bertini (NYU)

Ch. 2 of Cairo "The Truthful Art"


NORTHERN HEMISPHERE
Departures in temperature (°C)
from the 1961–1990 average
0.5
0.0
-0.5
-1.0
1000
1200
1400
1600
1800
2000
Year
Data from thermometers (red) and from tree rings,
corals, ice cores and historical records (blue).

Five Qualities of Great Visualizations

1. It is truthful.
2. It is functional.
3. It is beautiful.
4. It is insightful.
5. It is enlightening.

It is truthful

It is truthful

It is truthful

"If someone hides data from you, it’s probably because he has something to hide." (Cairo, Ch. 2)

"Truth and untruth aren’t absolutes. They are the extremes at either end of a spectrum" (Cairo, Ch. 2)

- Avoid self-deception
- Be honest with your audience

It is functional

It is beautiful

It is insightful

“The purpose of visualization is insight, not pictures" - Card, Mackinley, and Shneiderman (1999)

Two types of insight in data visualizations (Chang et al., 2009):

1. Eureka (or Aha!) moments

2. Knowledge-building

It is enlightening

"Some topics do matter more than others indeed because **they are more critical to the well-being of more people.**"

It is enlightening

"Some topics do matter more than others indeed because **they are more critical to the well-being of more people.**"

- What is the importance of your data visualization?

It is enlightening

"Some topics do matter more than others indeed because **they are more critical to the well-being of more people.**"

- What is the importance of your data visualization?
- Find topics that are important (and fun) to understand, especially in cases where you don't know the right question to ask (i.e., exploratory data analysis).