

# **DATA 230 Spring 2023**

## **Week 1**

# **Lecture 1:**

# **Data Visualization: What/Why/How?**

Dr. Seungjoon (Joon) Lee

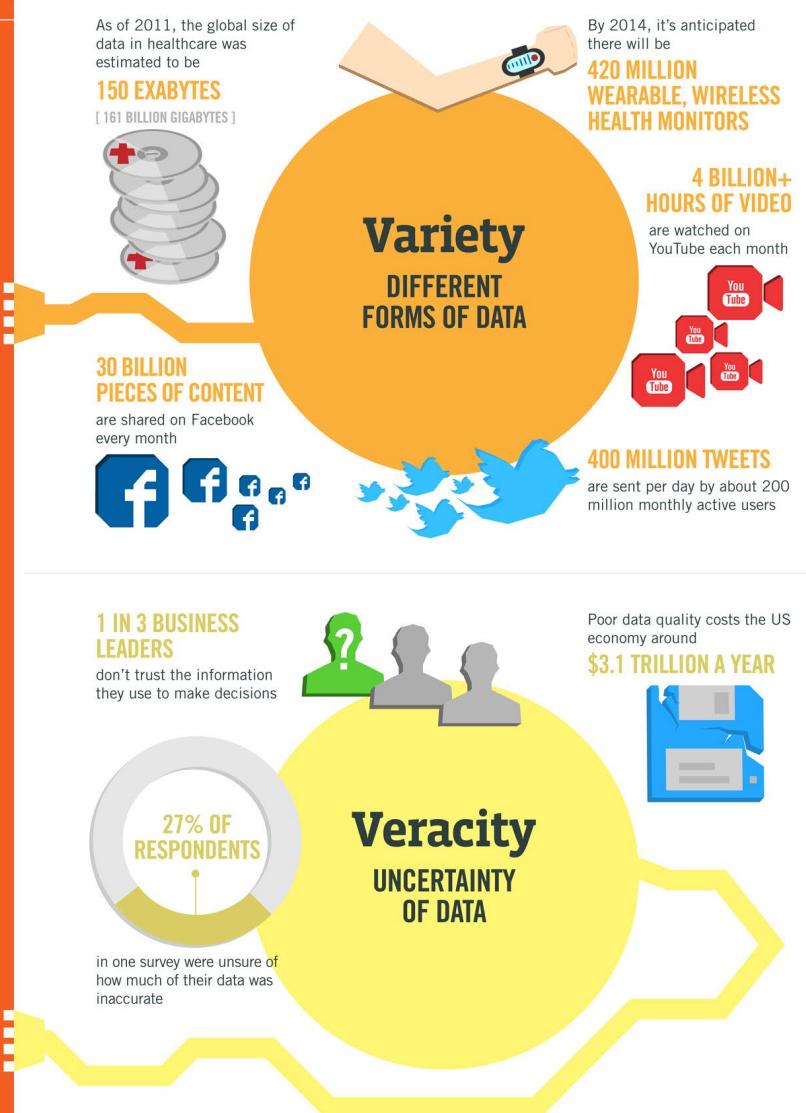
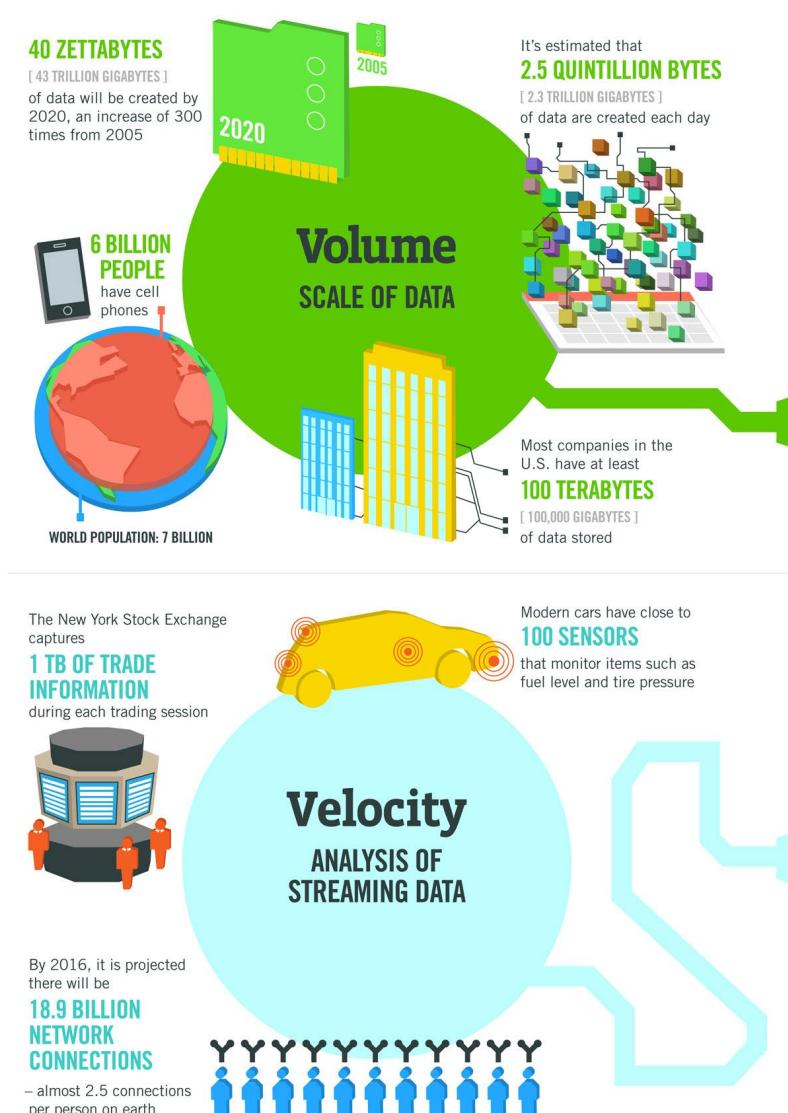
DATA 230 Data Visualization

# In DATA 220

- Mathematics and Statistics is **very important** to develop your career in data science/analytics.
- Domain Knowledge is also important!
- You have to have mathematical/statistical skillsets to analyze data effectively.

**Now, I am teaching DATA 230 ...**

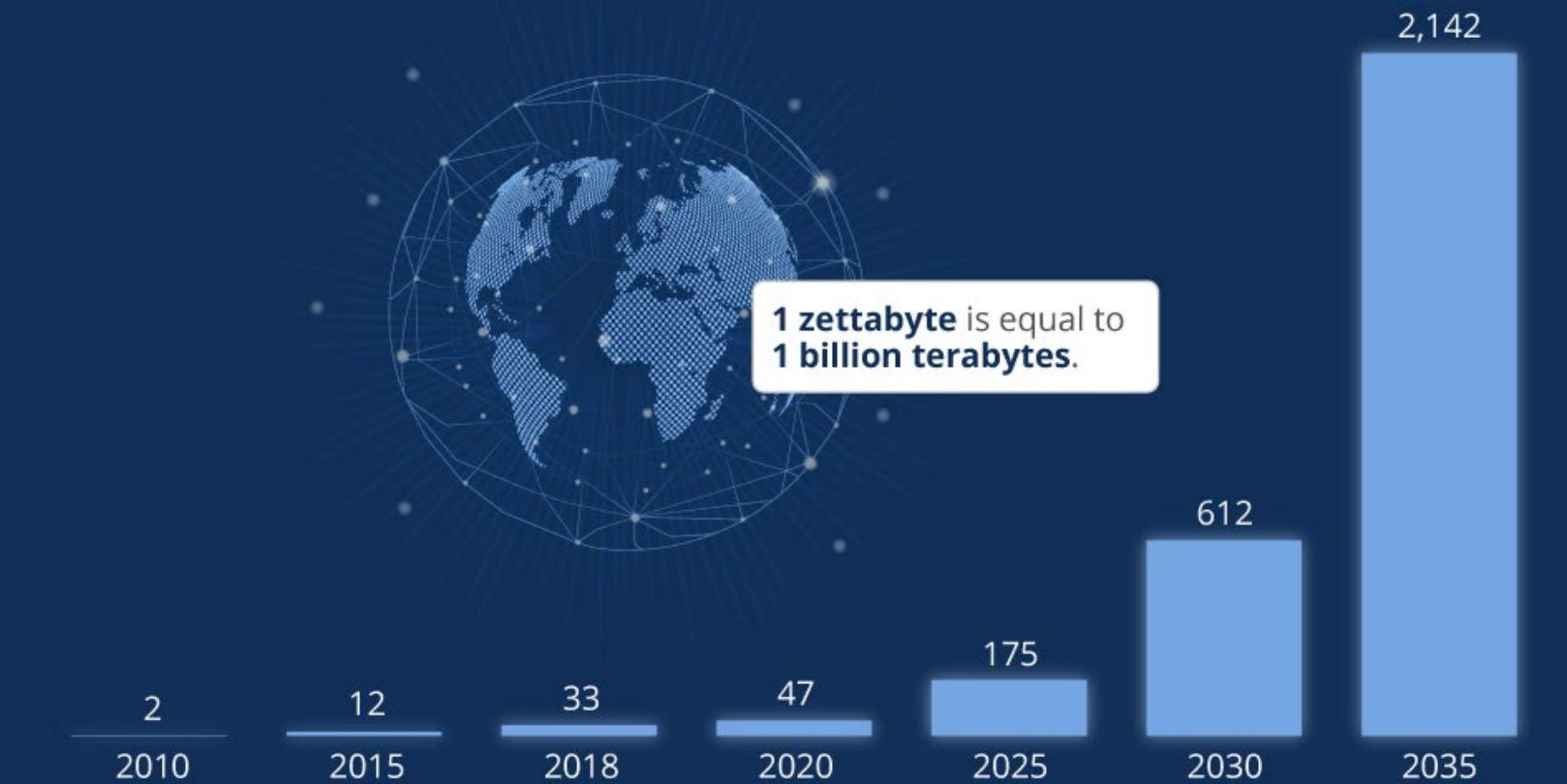
# Why Need Data Visualization?



# Why Need Data Visualization?

## Global Data Creation is About to Explode

Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)



@StatistaCharts

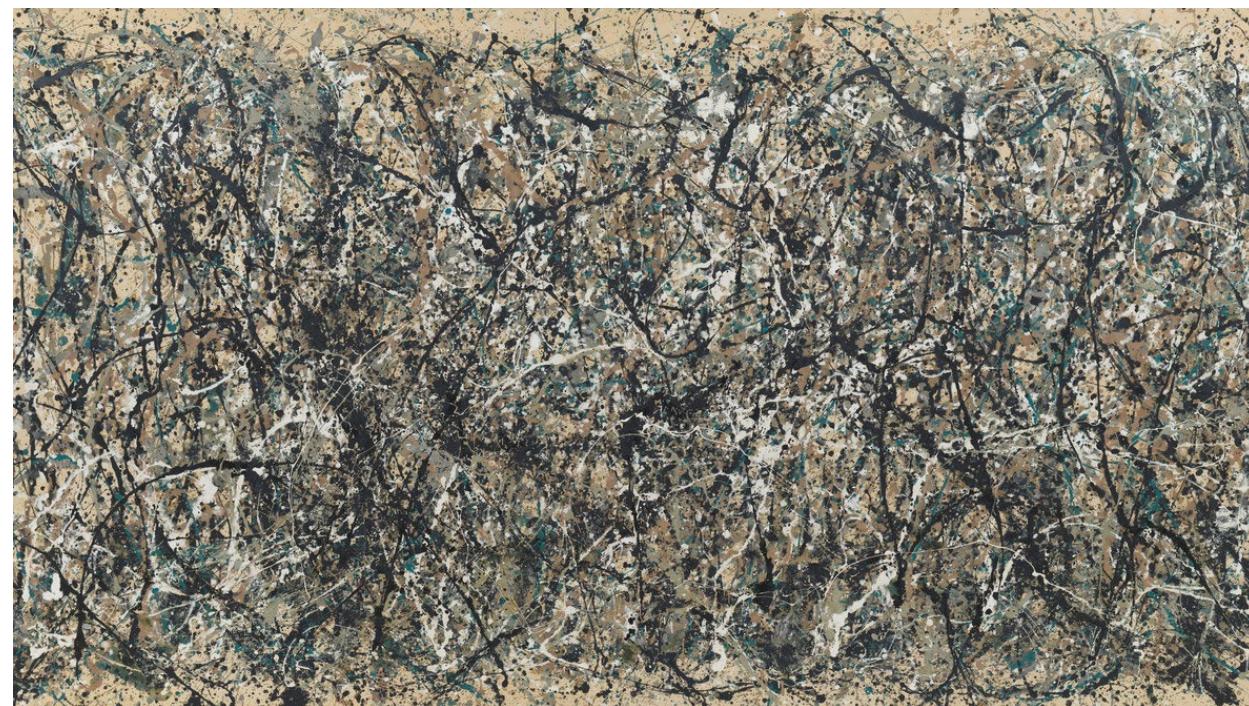
Source: Statista Digital Economy Compass 2019

statista

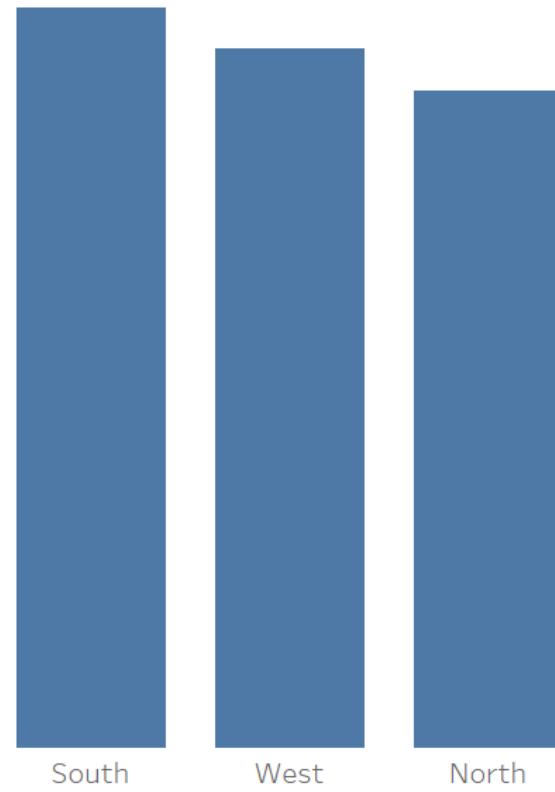
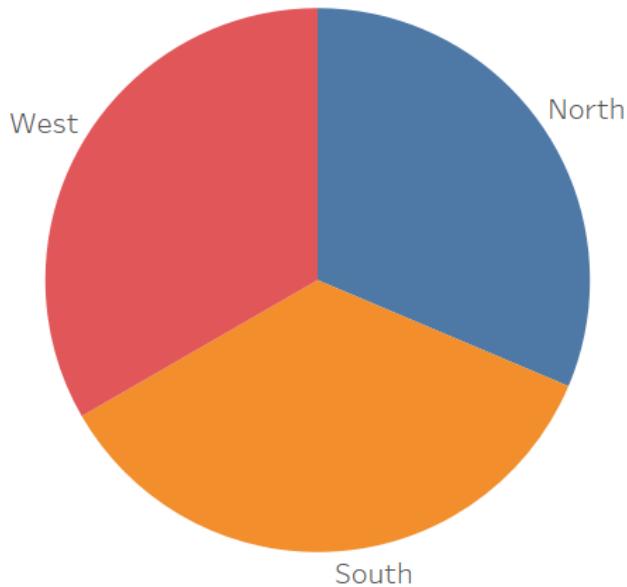
# Why Need Data Visualization?



# Art



# Data Visualization



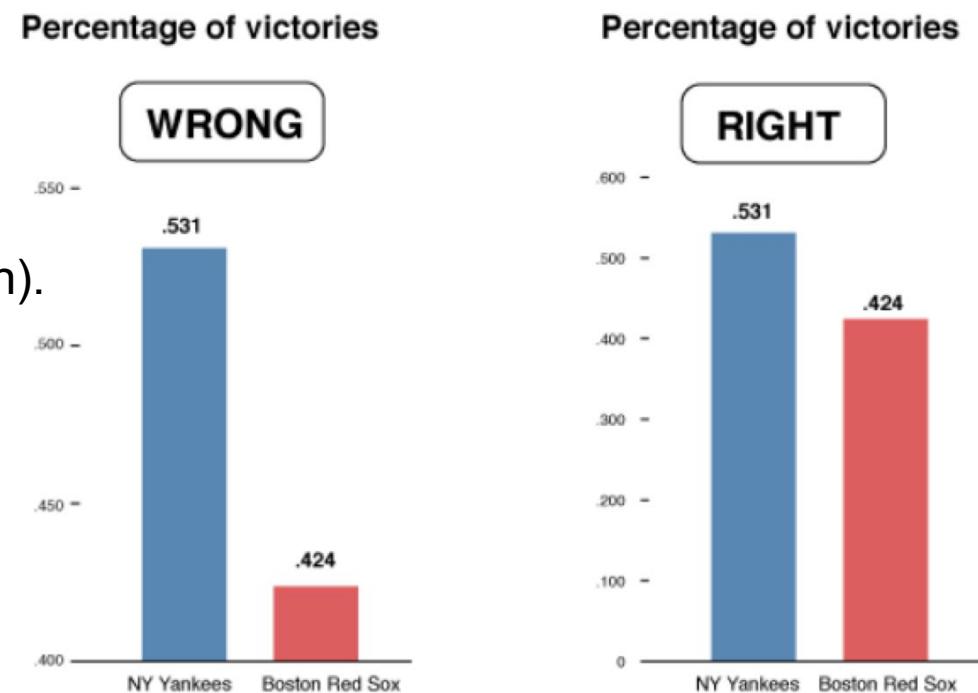
# Visualization and Mapping

1. Record information.
2. Analyze data to support reasoning (Exploratory visualization).

- Develop and assess hypothesis. - **Classical**
- Expand memory.
- Find the pattern. ... **Data-driven way**

3. Communicate information to others (Explanatory visualization).

- Persuade.
- Ethics.



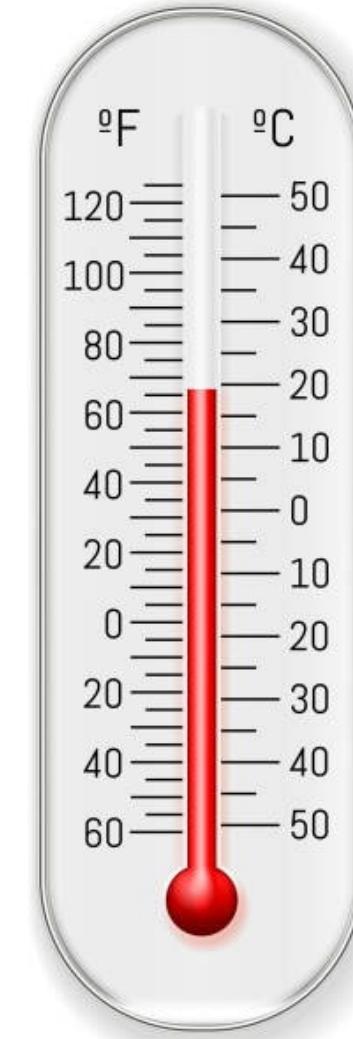
# Visualization and Mapping

- **Mapping** between the quantity of interest (or target or data) and its visual representation.
- **Arranging** between different visual representations in Big Data.
- Find **effective** mapping between data and its visual representation.  
→ Our final goal in this course!

# Visualization Example

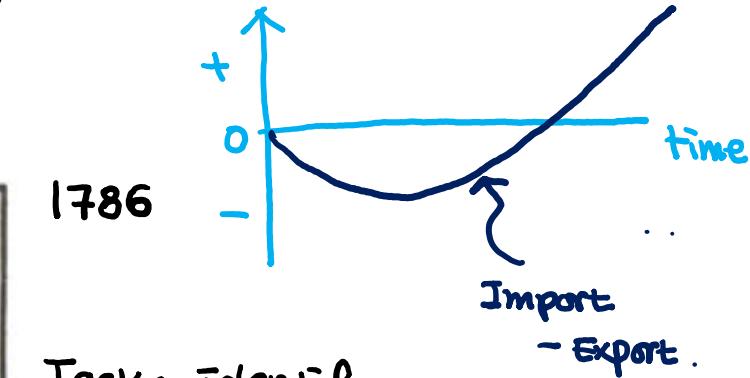
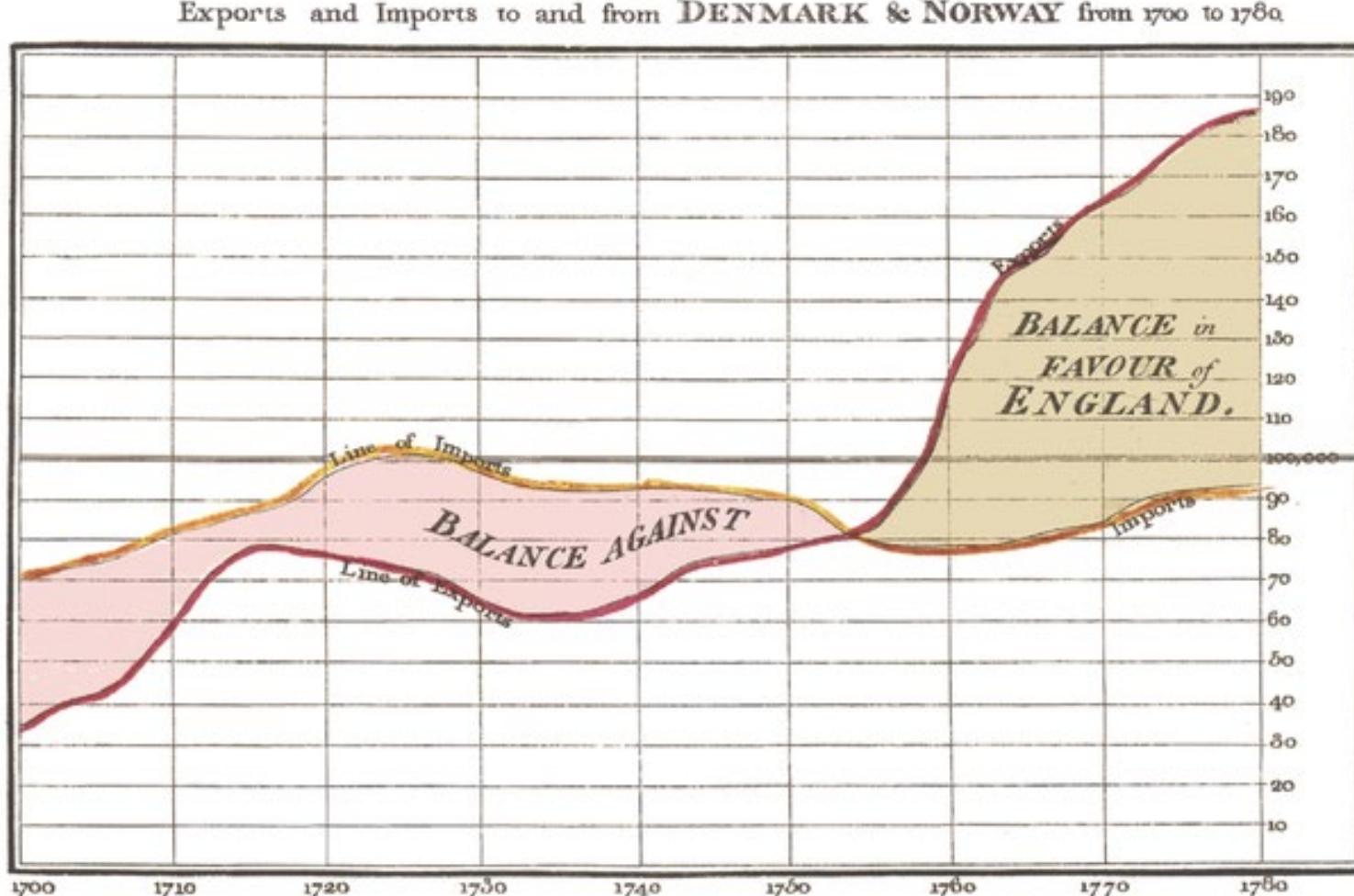


direction of wind  
→ direction to the  
rotation of the head



Temperature  
→ height of  
the volume of  
mercury .

# Data Visualization Example



Task : Identify  
exports and imports

Mapping : amount of  
export/import  
→ line graph.

Arrangement :  
drew two lines in one  
graph.

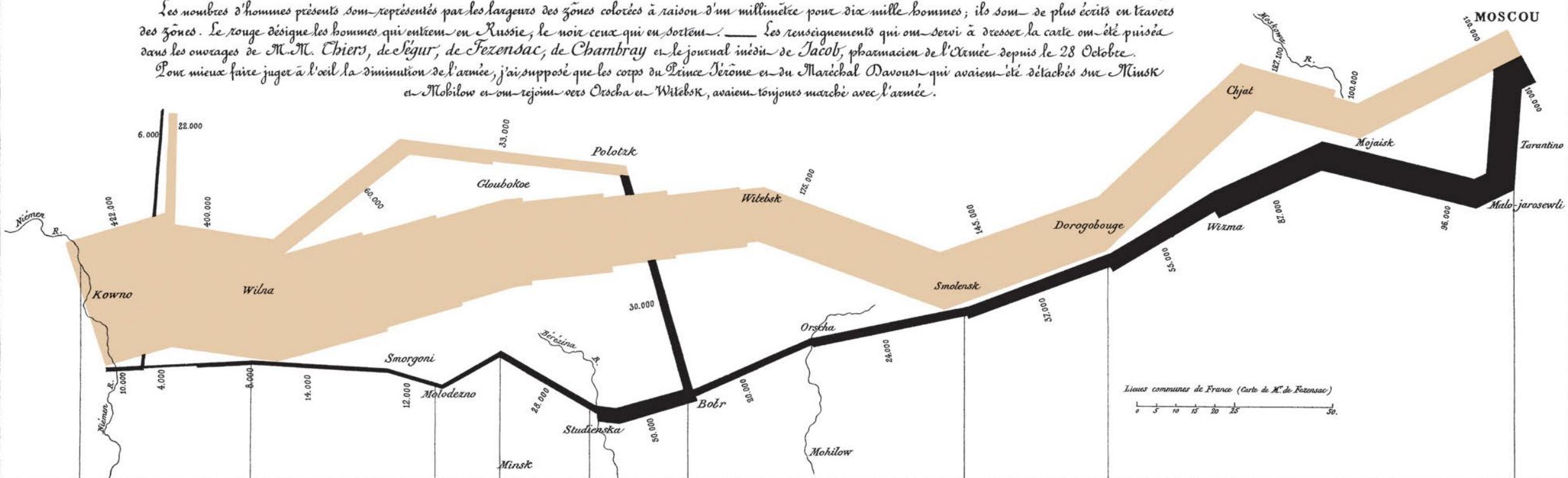
# Data Visualization Example

## Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite  
Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes ; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Segur, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Sébastien et du Maréchal Davout qui avaient été détachés sur Minsk et Mogilow et qui rejoignirent vers Orsha et Vitebsk, avaient toujours marché avec l'armée.



## TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop  
le Niemen gelé.

$-26^{\circ}$  le 7 X.<sup>bre</sup>  
 $-30^{\circ}$  le 6 X.<sup>bre</sup>

$-24^{\circ}$  le 1<sup>er</sup> X.<sup>bre</sup>  
 $-20^{\circ}$  le 28 9.<sup>bre</sup>

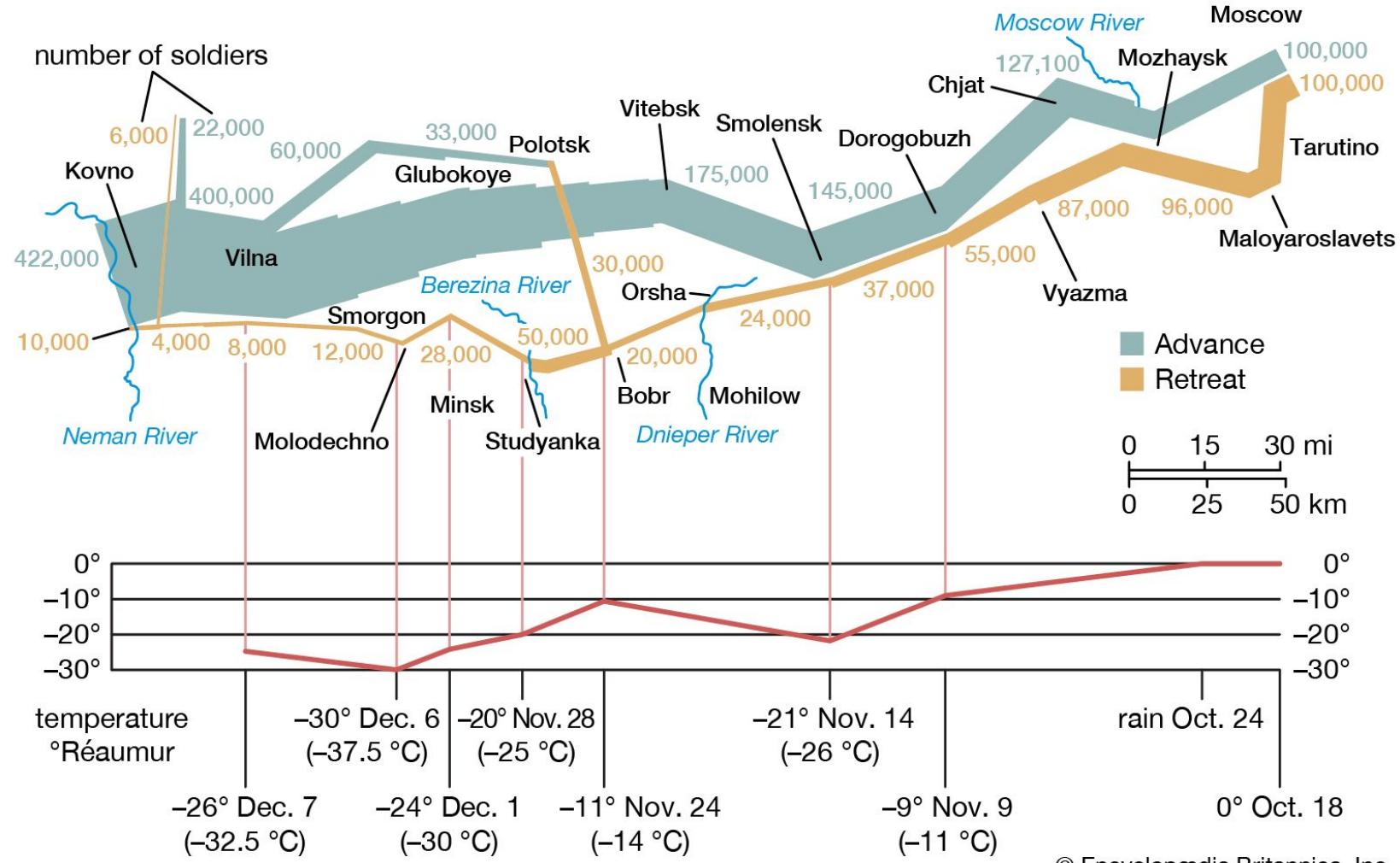
$-11^{\circ}$

$-21^{\circ}$  le 14 9.<sup>bre</sup>

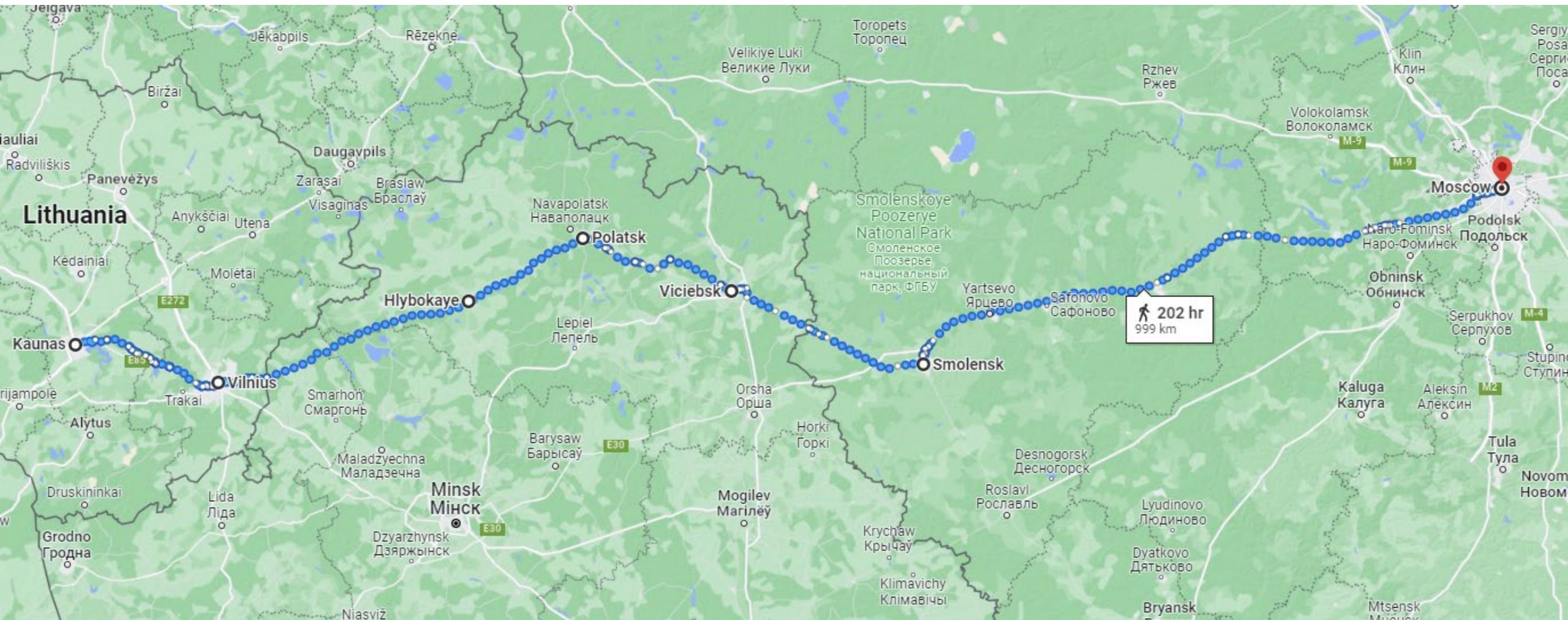
Zéro le 18 8.<sup>bre</sup>  
Pluie 24 8.<sup>bre</sup>  
5  
10  
15  
20  
25  
30 degrés

# Data Visualization Example

Based on Charles Minard's graph of Napoleon's Russian campaign of 1812.

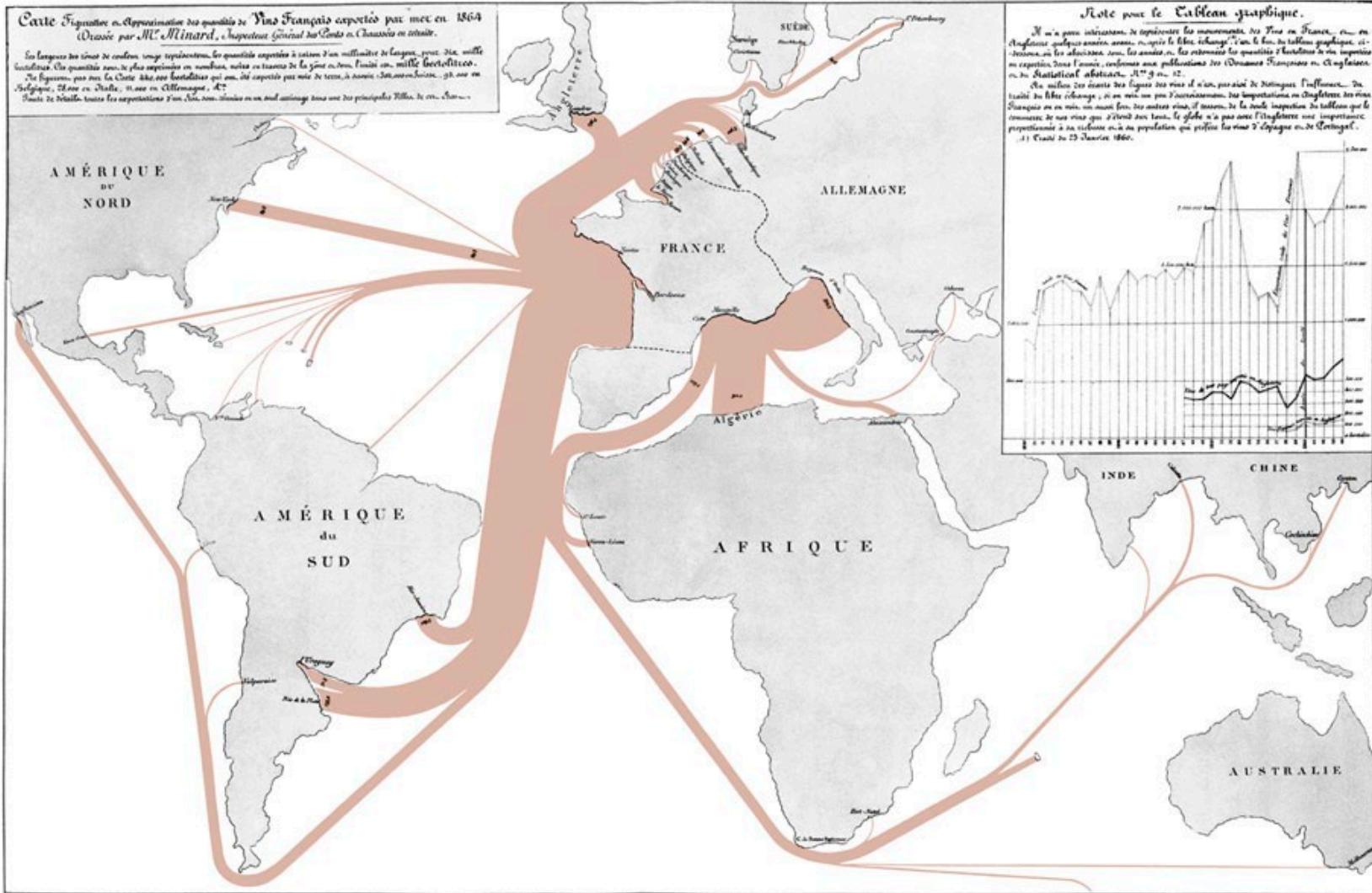


# Data Visualization Example



$$999/50=20\text{km}$$

# Data Visualization Example

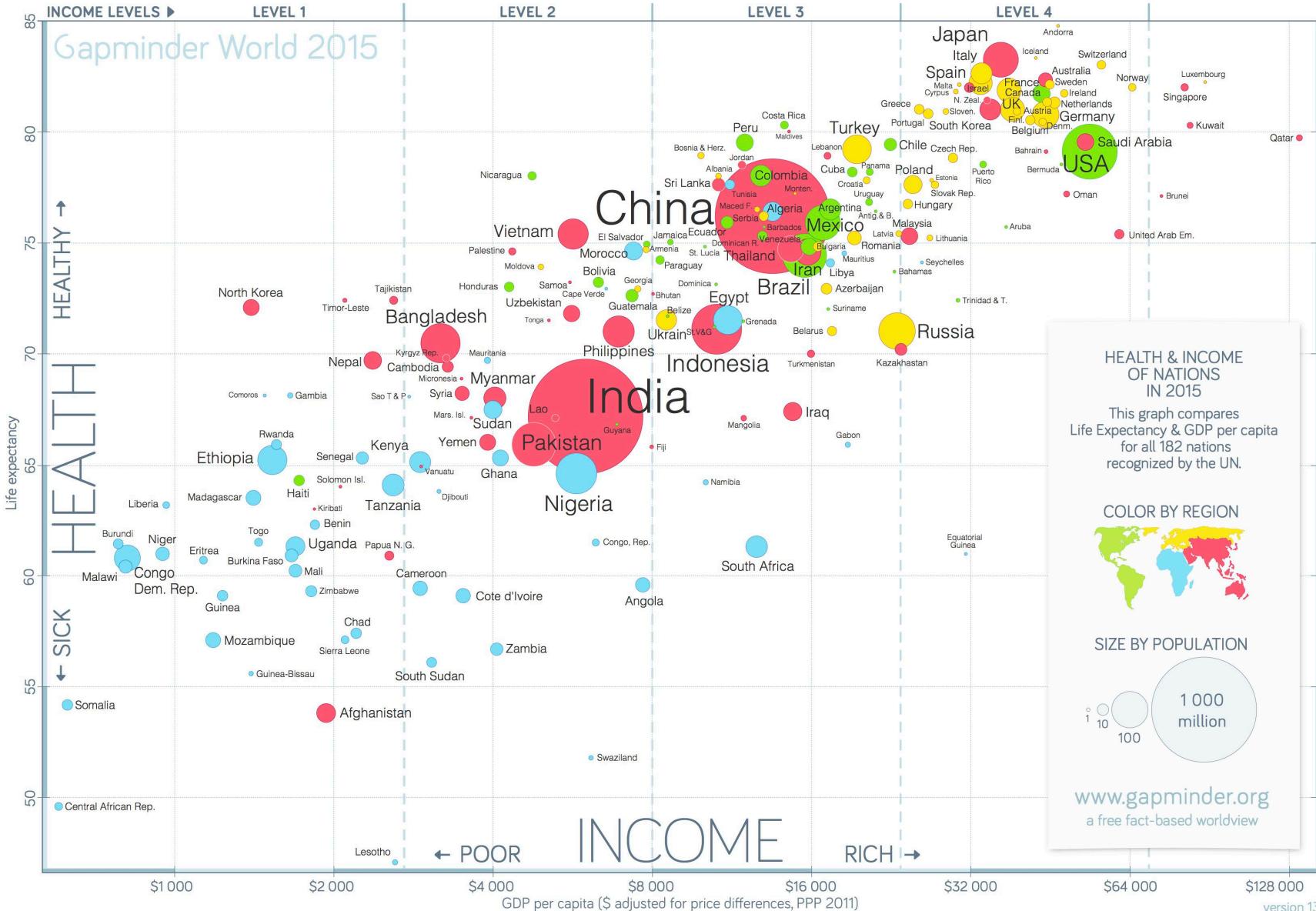


Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845-1869, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

Task: identify exports of  
French wine

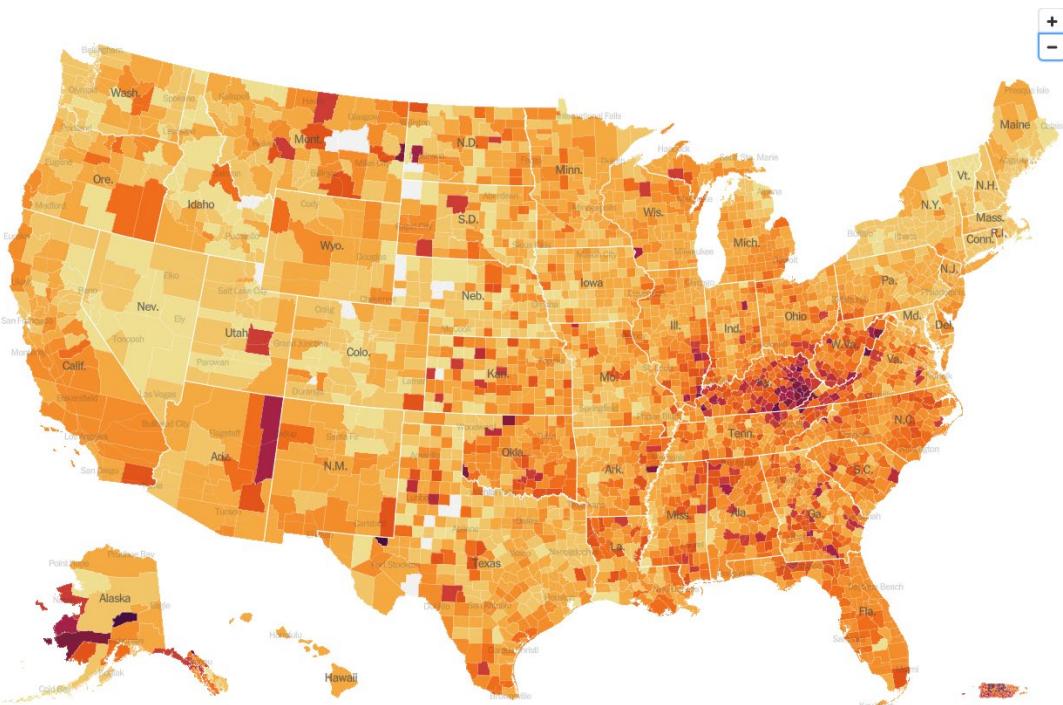
Mapping : amount of exports  
→ width.

# Data Visualization Example



# Data Visualization Example

## Geography + color

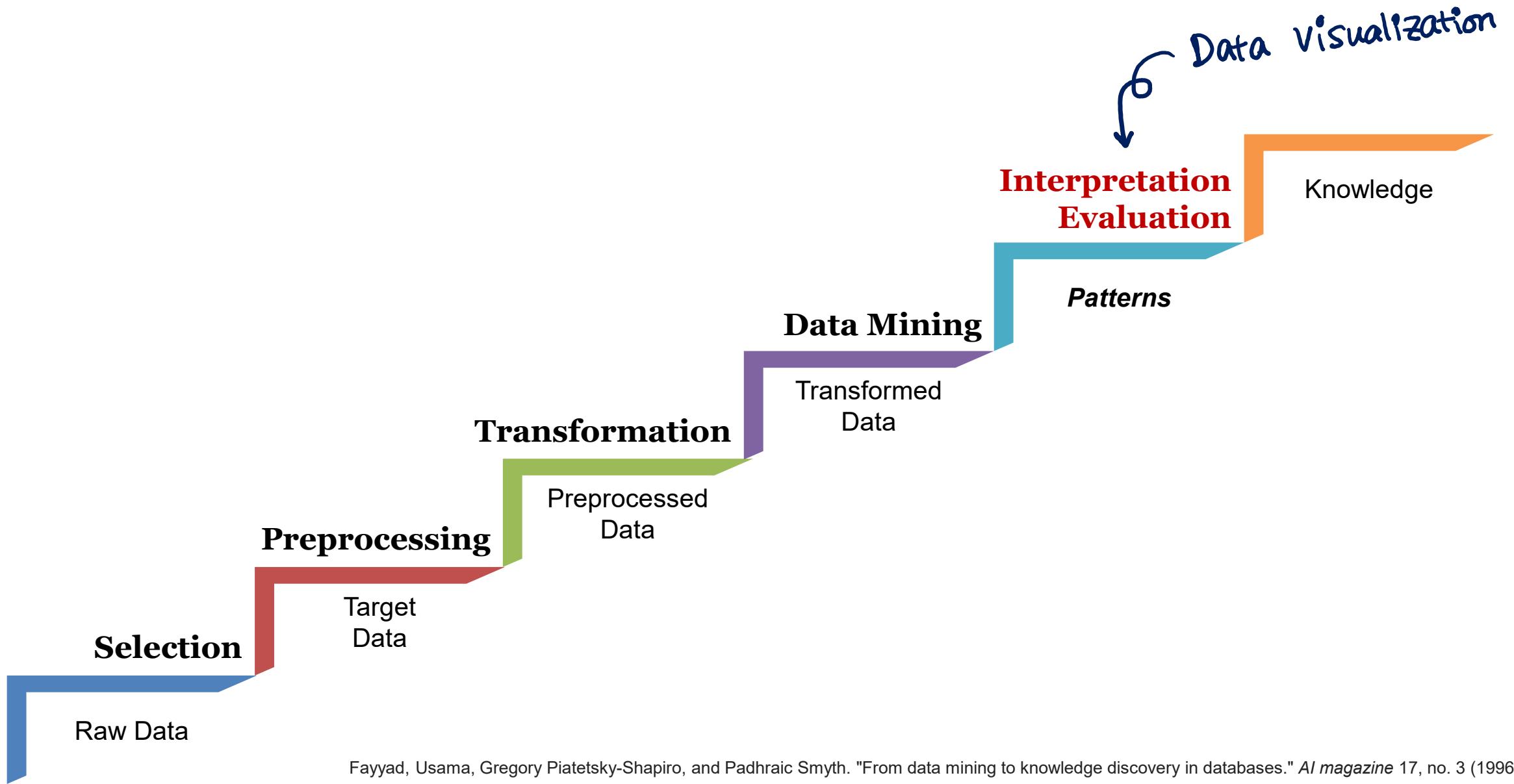


© Mapbox © OpenStreetMap

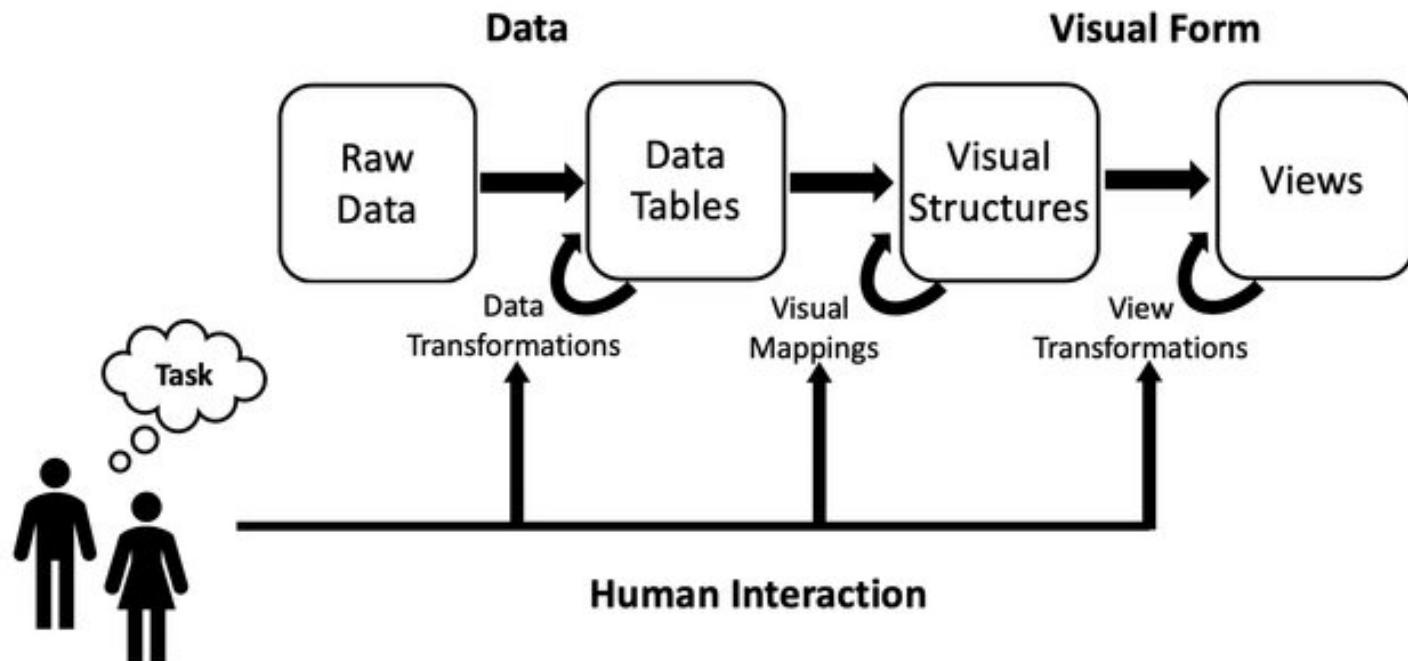
## Color and Size



# Knowledge Discovery in Databases (KDD)



# Reference Model



# Visual Popout

**Preattentive Processing:** low-level visual system can process information efficiently.

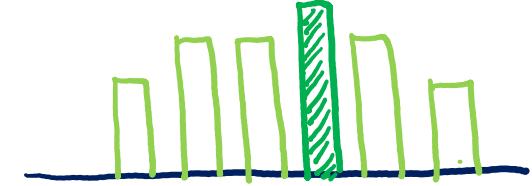
513514132103138486351568743538451561122546  
6546846548444844555546879846542354984846548  
1444845487896625484651875226533549844254864  
6654182489788552625489961154648466711533546

# Visual Popout

**Preattentive Processing:** low-level visual system can process information efficiently.

513514132103138486351568743538451561122546  
654684654844484455546879846542354984846548  
1444845487896625484651875226533549844254864  
6654182489788552625489961154648466711533546

# Visual Popout

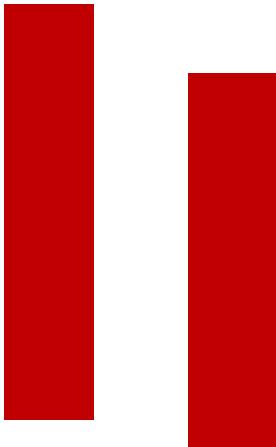


**Preattentive Processing:** low-level visual system can process information efficiently.

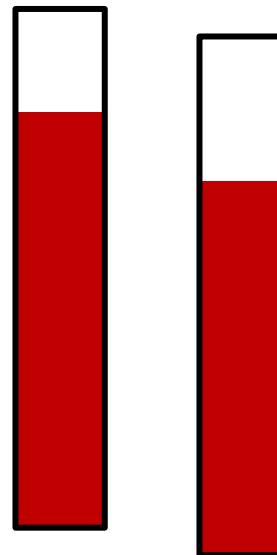
51**3**5141**3**210**3**1**3**8486**3**5156874**353**8451561122546  
6546846548444844555546879846542**3**54984846548  
14448454878966254846518752265**33**549844254864  
66541824897885526254899611546484667115**33**546

# Relative vs Absolute Judgement

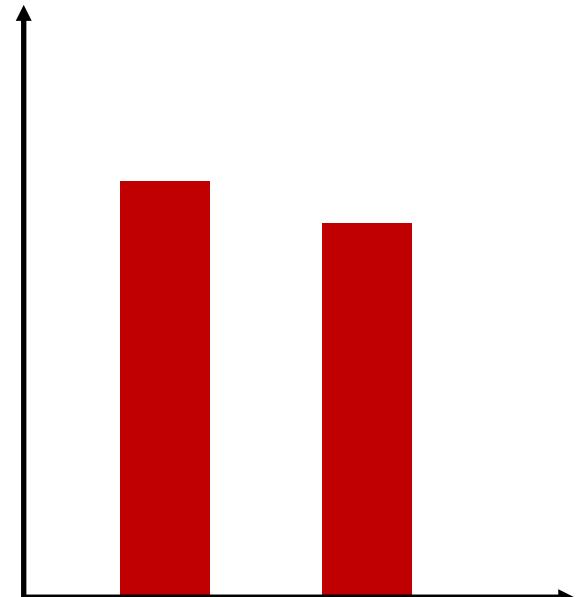
Frame and align.



Unframed  
Unaligned



Framed  
Unaligned



Unframed  
Aligned

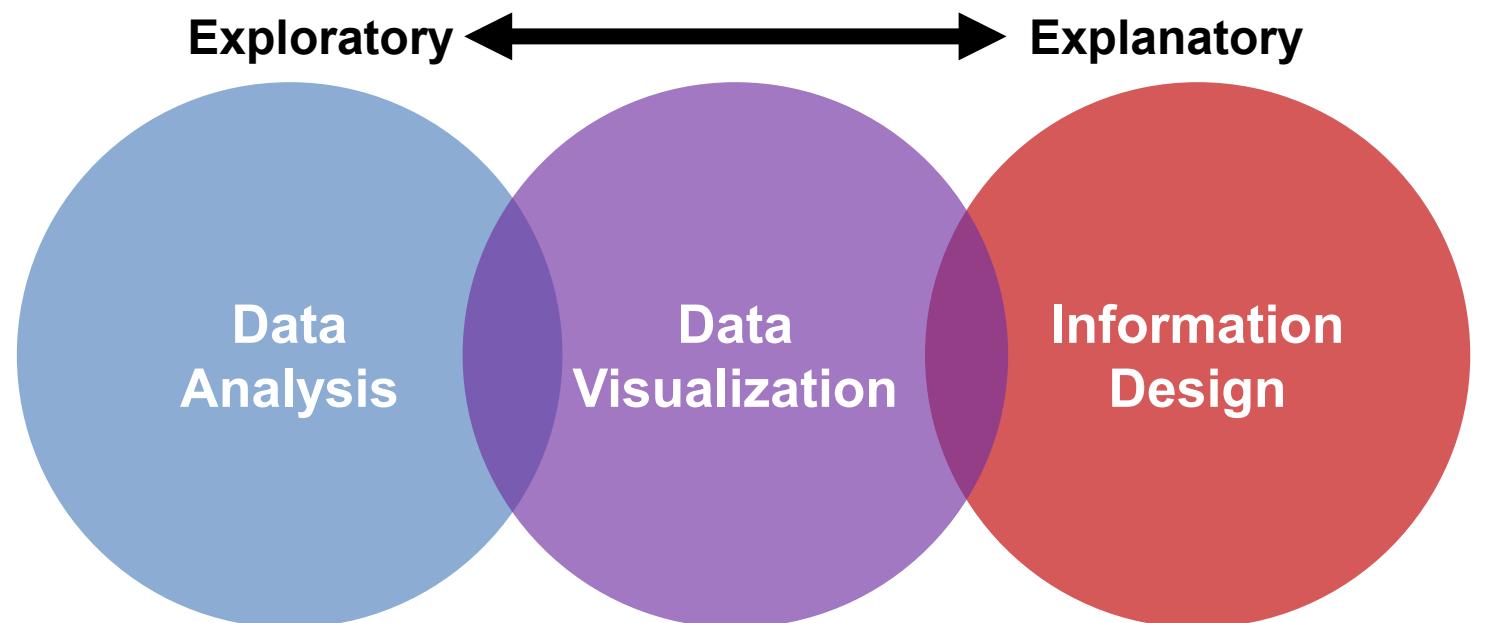
# Exploration and Explanation

## Exploratory visualization

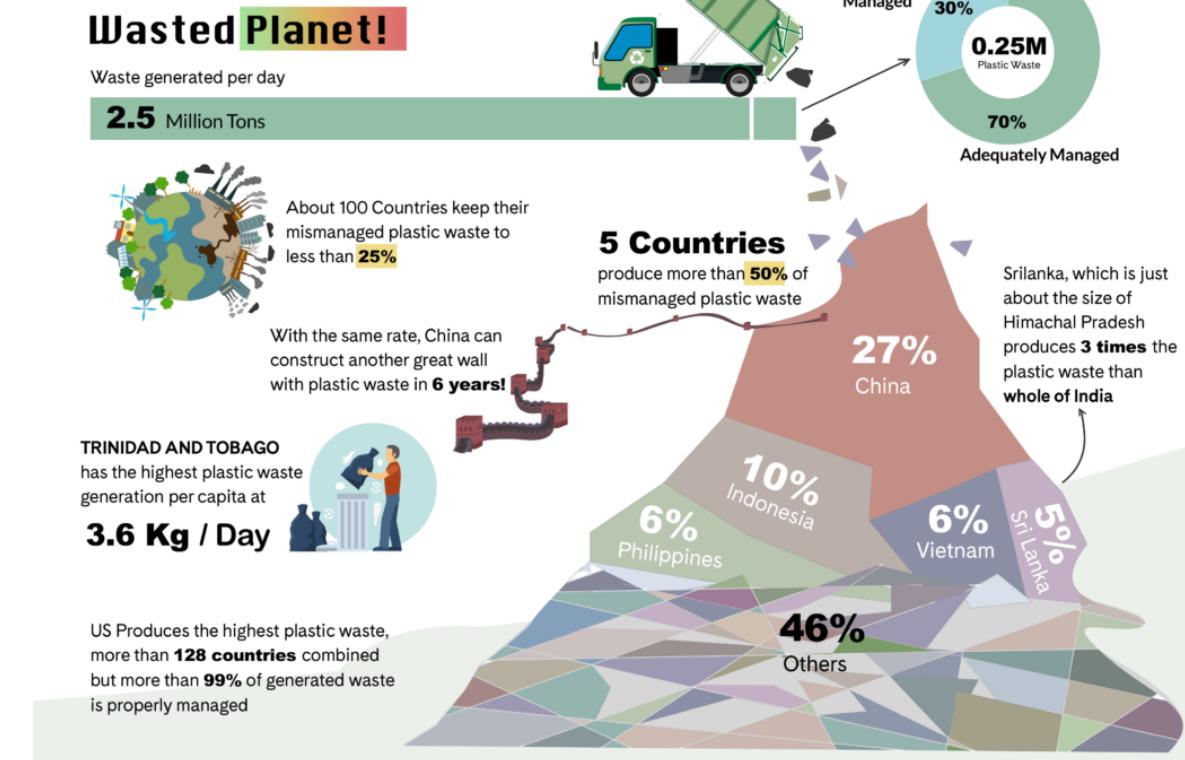
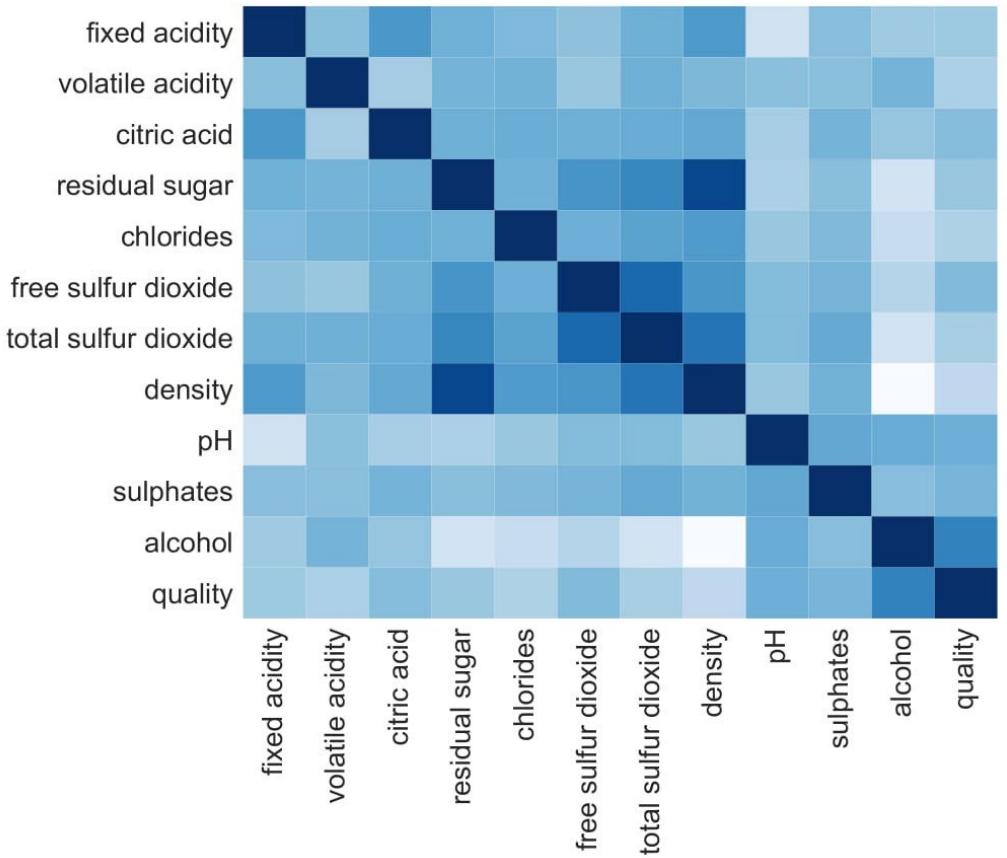
- Making sense of your data.

## Explanatory visualization

- Sharing information with an audience.



# Exploration and Explanation



# Interactive vs Static

- Data is too big and complicated to show in a single static view – check more information.
- Dashboard: show important information by data analyst – better understanding.



# **Summary**

## **We will learn various representation skills for Big Data**

- All the methods have pros and cons.
- There is no universal method for all types of data.

## **What will we learn, specifically?**

- One variable data.
- Two (or more) variable data.
- Spatiotemporal data.
- Result from models.
- Data ethics.
- Exploratory data analysis (EDA) and dashboard.
- Data reduction.