

Data Visualization and Dashboards

P. BOILY

UNIVERSITY OF OTTAWA | FACULTY OF SCIENCE | DEPARTMENT OF MATHEMATICS AND STATISTICS
DATA ACTION LAB | IDLEWYLD ANALYTICS

Instructor – Patrick Boily

Employment

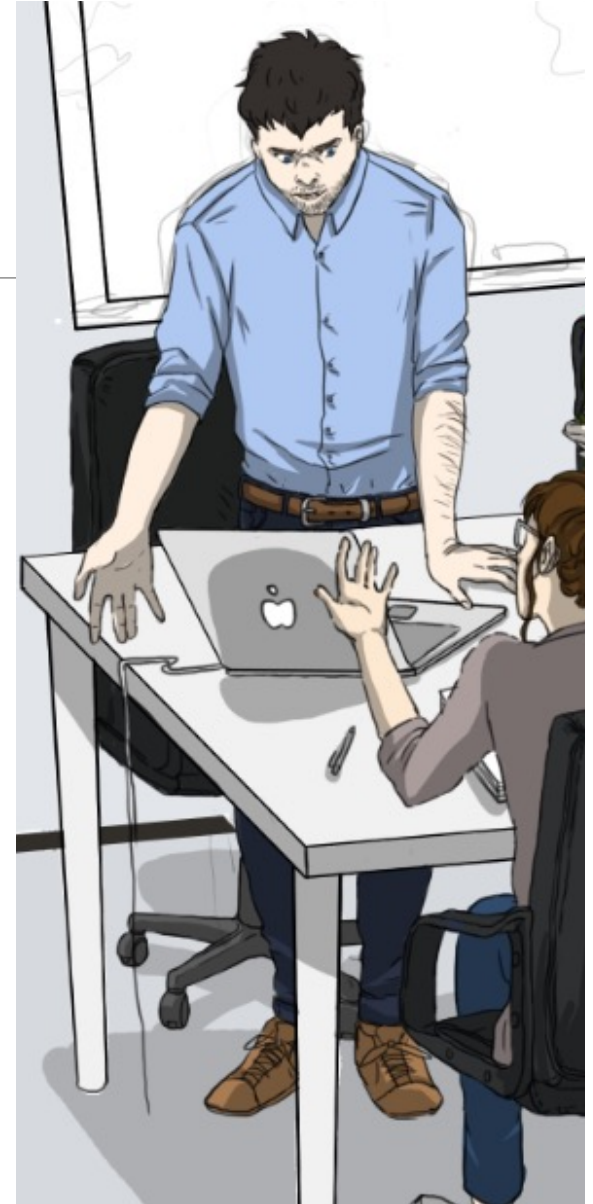
- Professor Math/Stat ['19 – now, uOttawa]
- President ['16 – now, Idlewyld Analytics]
- Manager and Senior Consultant ['12 – '19, CQADS, Carleton]
- Public Service ['08 – '12, ASFC | StatCan | TC | TPSGC]
- 60+ uni course; 250+ workshop days

Projects

- GAC; NWMO; CATSA; etc.
- 40+ projects

Specialization

- Data visualization; data cleaning (... unfortunately)
- Application of wide breadth of techniques to all kinds of data
- Mathematical/statistical modeling



Course Material

Course Webpage:

<https://data-action-lab.com/102-dvd>

Course Notes:

<https://www.practicedataviz.com>

Contact Info:

pboily@uottawa.ca

Slack Workspace:

<https://dspdi.slack.com>

Course Description

Prior to running analyses, data exploration is crucial to gain an understanding of the dataset and what it contains. Increasingly sophisticated visualization strategies can play an important role at this stage.

Visualization is also used at the end of the data analysis process to communicate data insights to an audience, often in the form of a dashboard.

In this course, participants will learn tool-agnostic fundamental strategies for visualizing and communicating data (storytelling) and be introduced to the grammar of graphics and the basics of dashboarding.

Following the course, the participants have the option of working on a guided project, getting feedback from the instructor.

Additional Information

Exposure to programming frameworks would be beneficial but not necessary. Participants must be comfortable (not necessarily experts) with the concepts introduced in *Data Science Essentials*.

Participants are required to bring a laptop/personal computer on which the current version of R/RStudio (Posit) are installed (for which they may require administrative authorisation to install packages).

Participants doing the guided project must be familiar with R, the tidyverse, and/or Python.

Learning Outcomes

At the end of this course, participants will be able to:

- develop strategies to explore a dataset using visual means
- discern which strategy and design to use to best represent the results from their data analysis
- create multivariate visualizations using ggplot2
- design simple conceptual dashboards
- learn how they can use storytelling notions to share their analysis insights

Course Outline

Data Visualization Concepts

1. Exploratory Data Analysis
2. Data Visualization for Communication
3. Visualization Catalogue
4. Chart Aesthetics
5. Dashboards

Storytelling with Data

6. Elements of Storytelling
7. Stories and Illustrations
8. Evolving a Storytelling Chart
9. Anatomy of Storytelling Dashboards

Session 1

Session 2

Session 3

Session 4

Course Outline

Visualizations with ggplot2

10. The Grammar of Graphics
11. Basics of ggplot2
12. Examples and Miscellanea

Session 1

Session 2

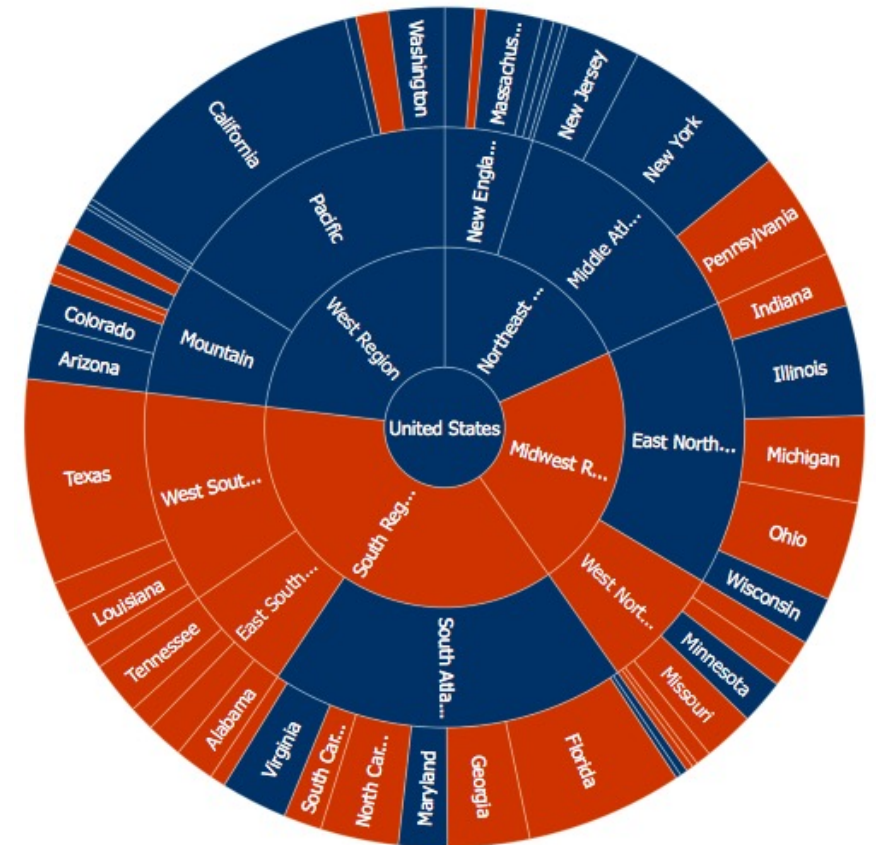
Session 3

Session 4

Data Visualization vs. Infographics

Data Visualization

- A **method**, as well as an item (**objective**)
- Typically focuses on the **quantifiable**
- Used to make sense of the data or to make it **accessible** (datasets can be massive and unwieldy)
- May be generated **automatically**
- The look and feel are less important than the **insights conveyed by the data**



Size Population Color Median Household Income

■ Low Income ■ High Income

Data Visualization vs. Infographics

Infographics

- Created for **story-telling** purposes (**subjective**)
- Intended for a **specific** audience
- **Self-contained** and discrete
- **Graphic design** aspect is key
- **Cannot** usually be re-used with other data
- Can incorporate **unquantifiable** information



Historical Charts

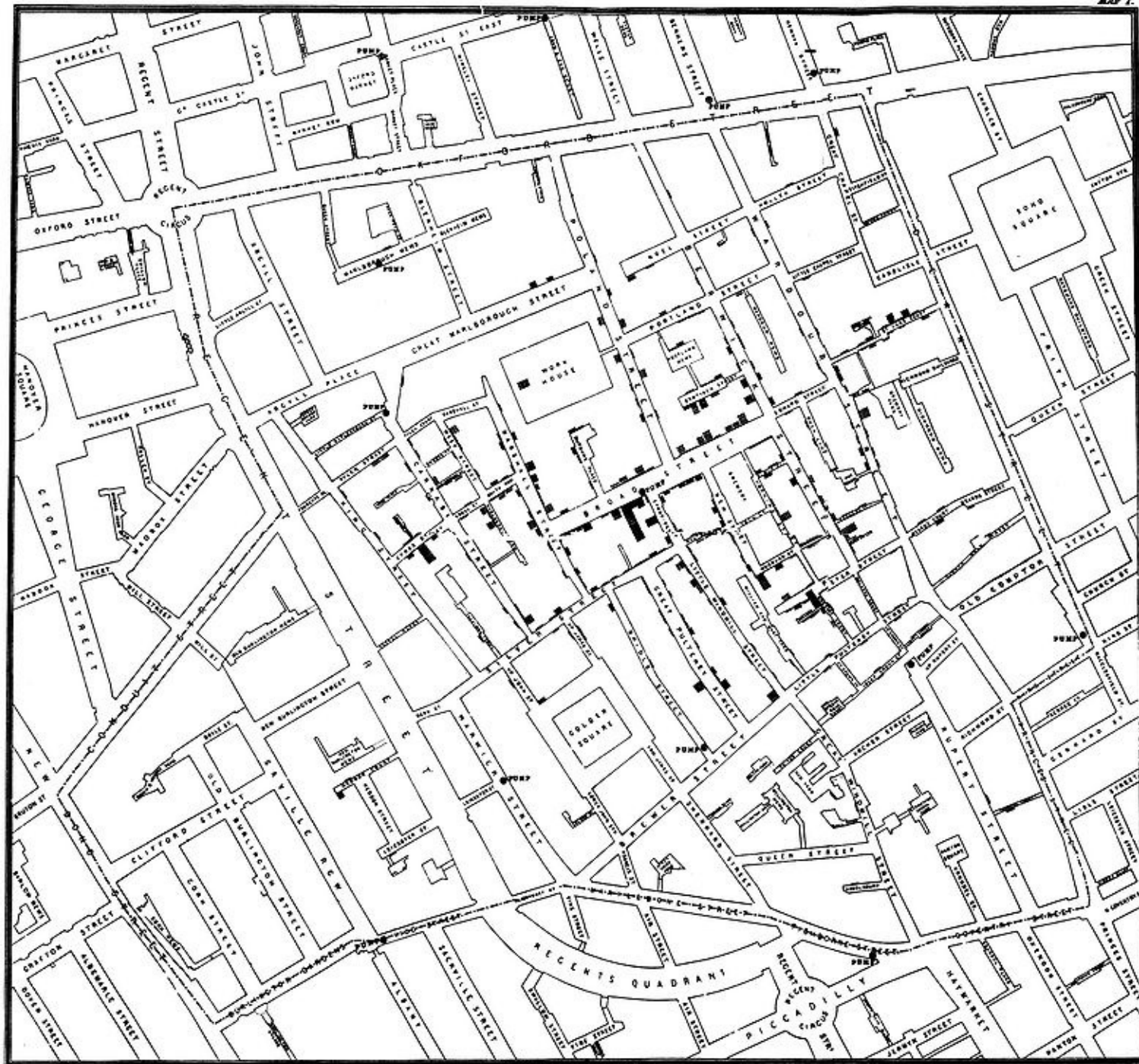
Data visualization is not confined to the recent past: charts have been used for many years to help **communicate information** and **tell stories**.

Due to the absence of technical tools, a lot of thought had to go into the design and creation of these visualizations.

Consequently, there is a lot we can (and **should**) learn to bring into the development of charts from a **design and storytelling perspective**.

London's Cholera Outbreak of 1854

Physician John Snow links the outbreak to a contaminated well by plotting number of cases on a map, jump-starting the science of epidemiology.



London's Cholera Outbreak of 1854

Physician John Snow links the outbreak to a contaminated well by plotting number of cases on a map, jump-starting the science of epidemiology.

John Snow's London Cholera Outbreak Map (1854)

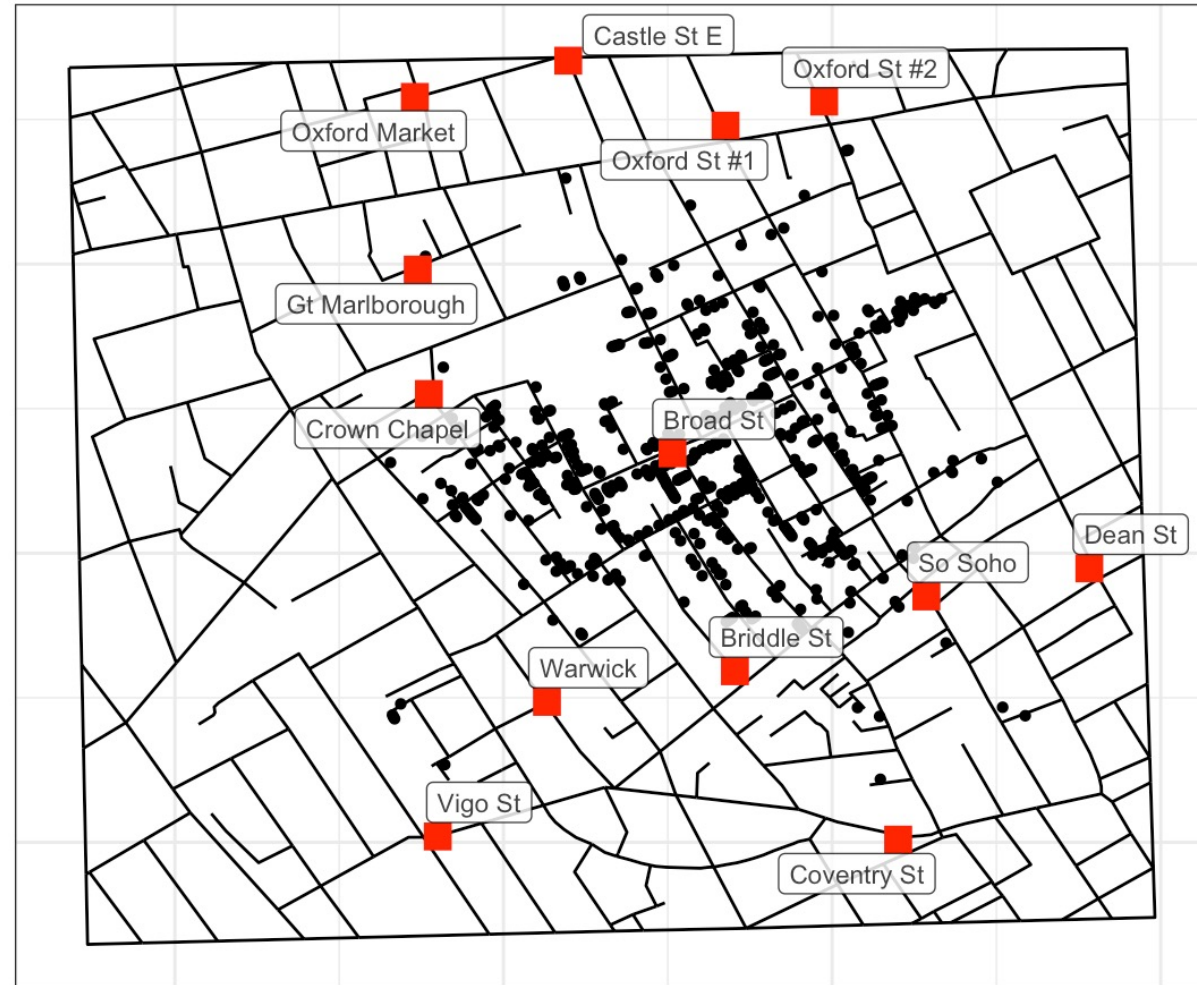
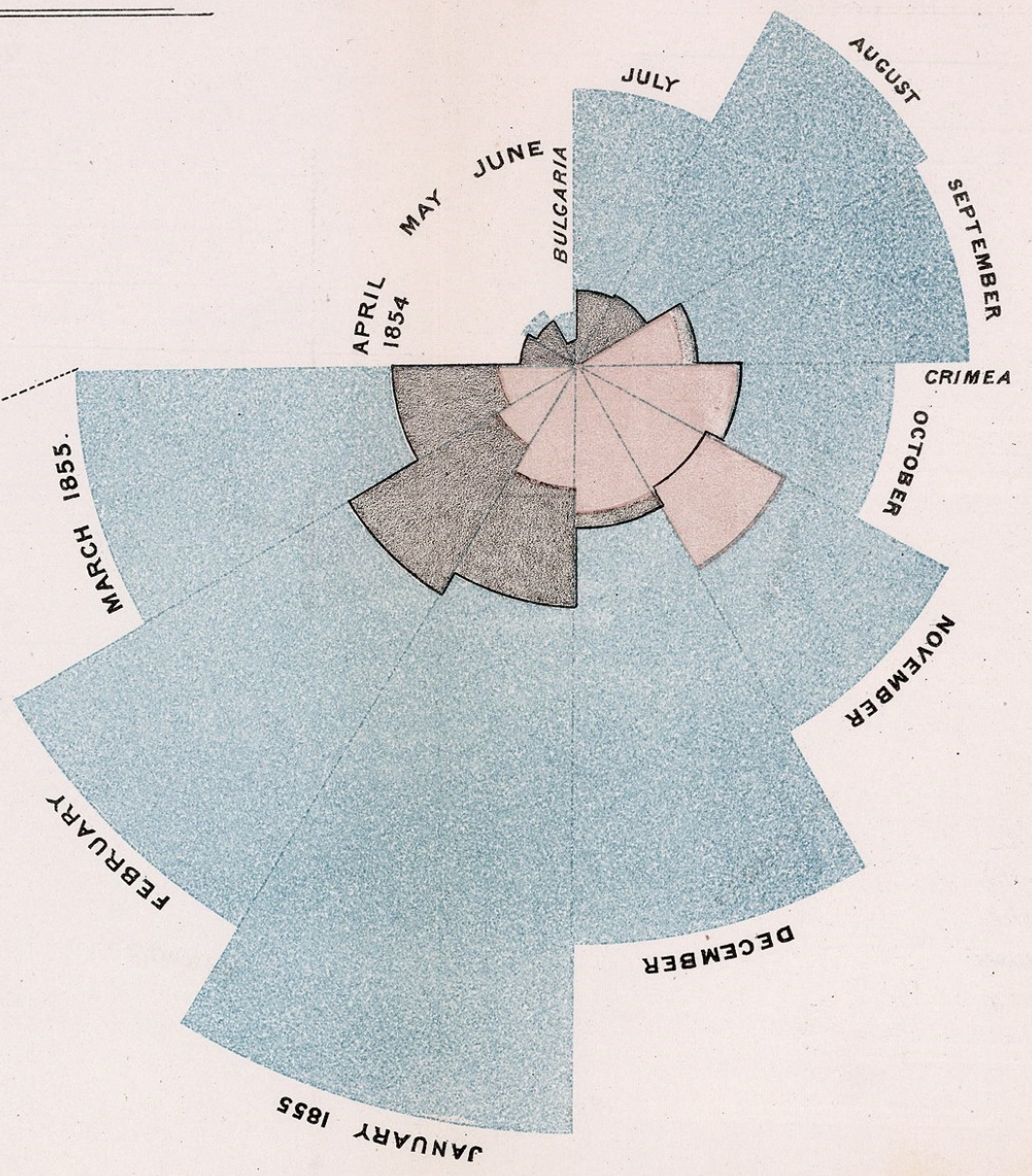
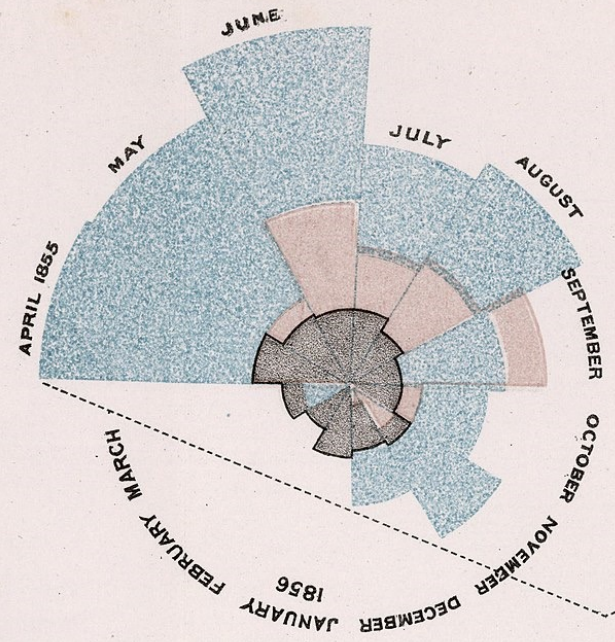


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

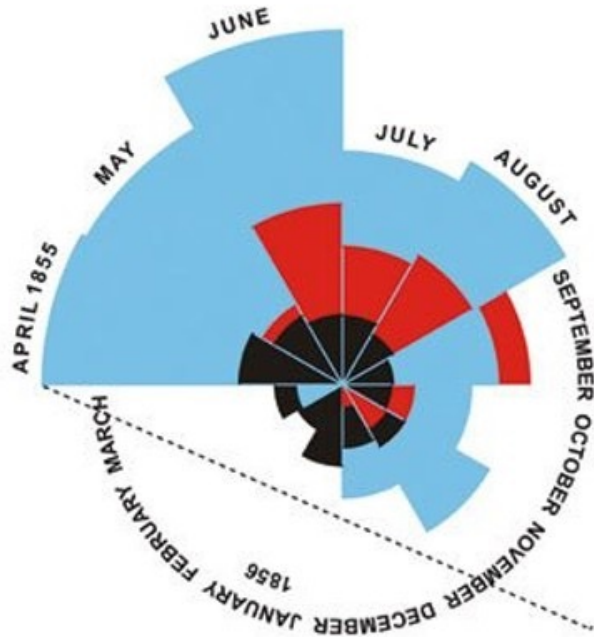
The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

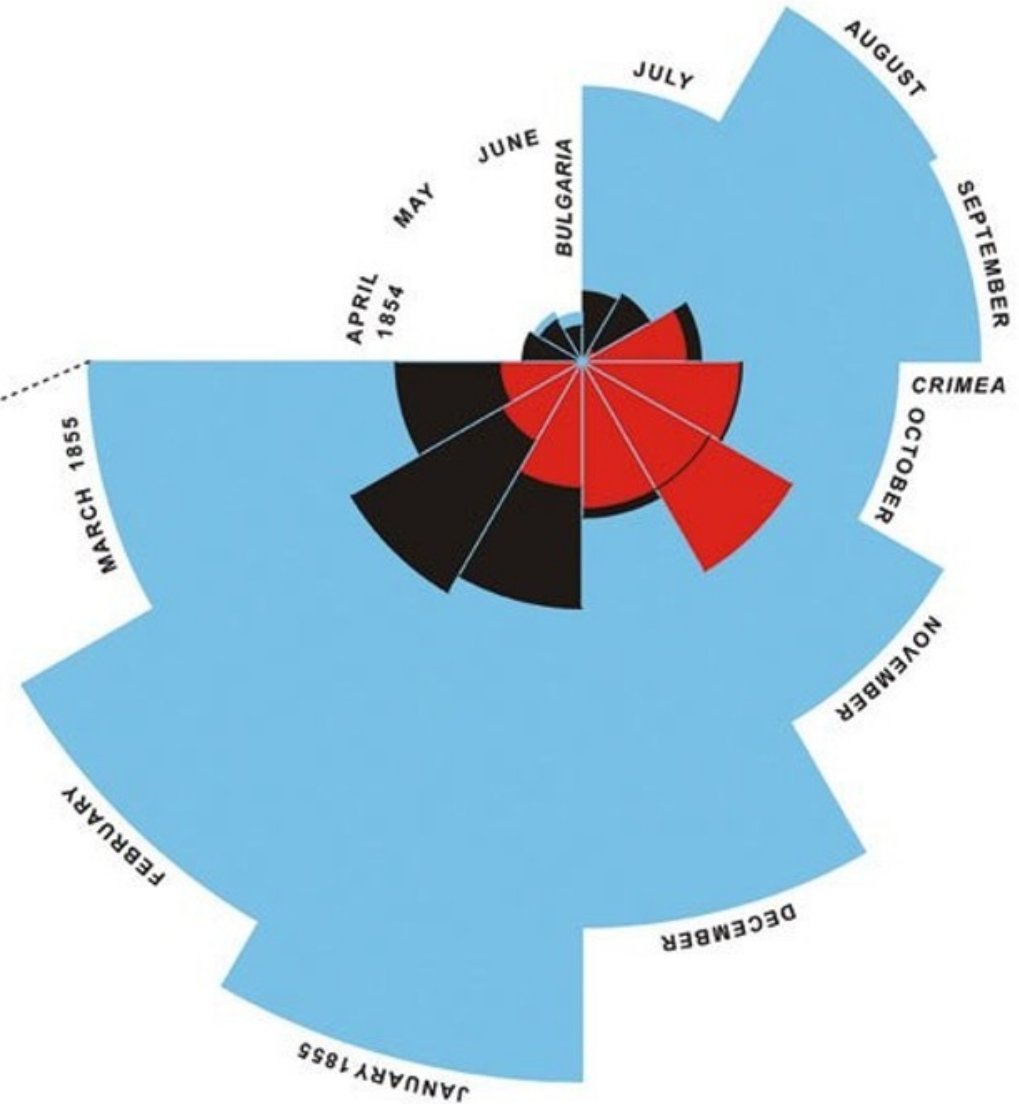
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST .

2.
APRIL 1855 TO MARCH 1856 .



1.
APRIL 1854 TO MARCH 1855 .



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes

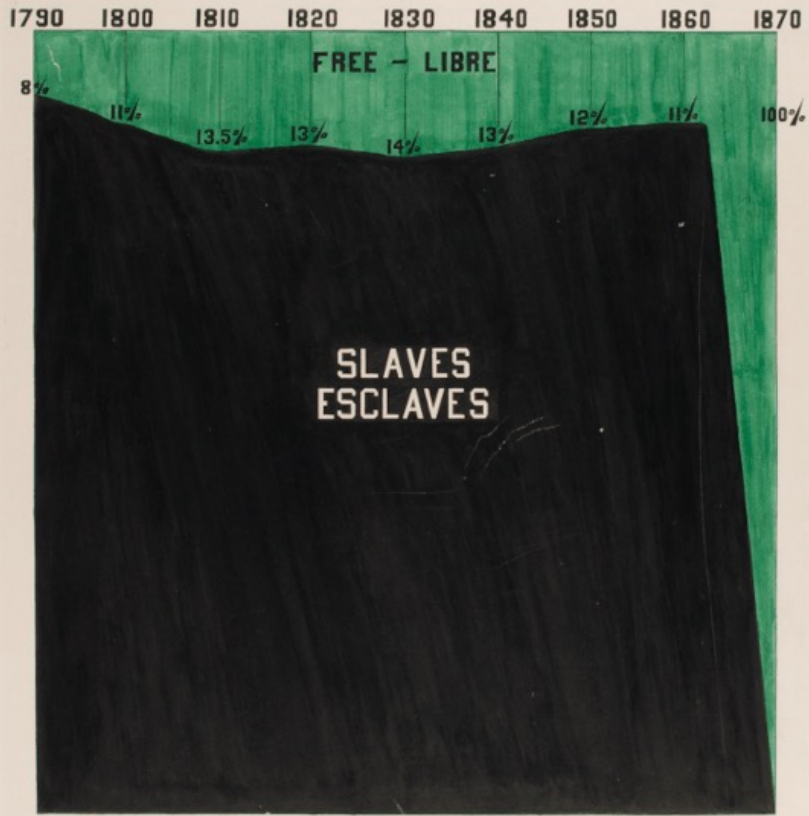
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black

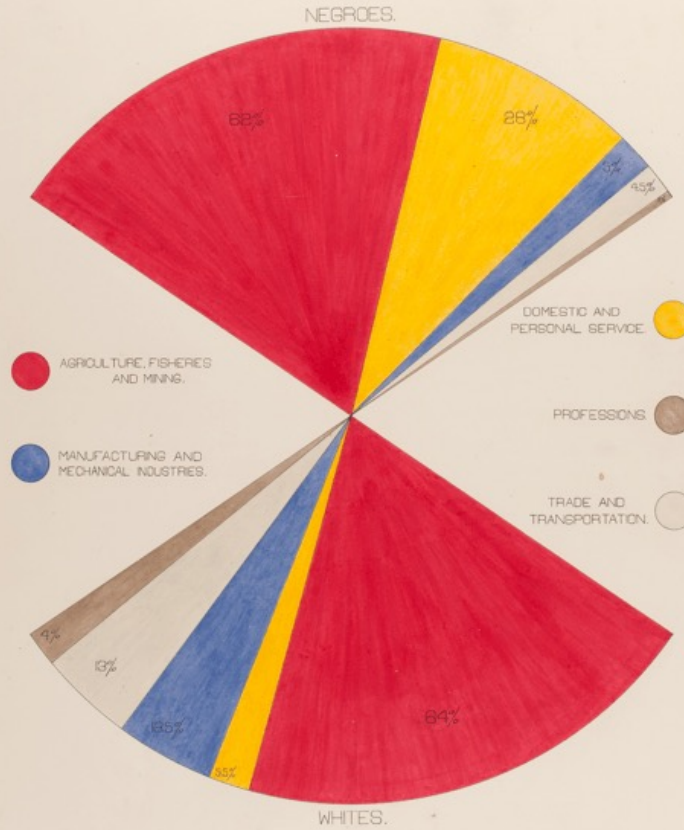
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

PROPORTION OF FREEMEN AND SLAVES AMONG AMERICAN NEGROES .
 PROPORTION DES NÈGRES LIBRES ET DES ESCLAVES EN AMÉRIQUE .

DONE BY ATLANTA UNIVERSITY .



OCCUPATIONS OF NEGROES AND WHITES IN GEORGIA .



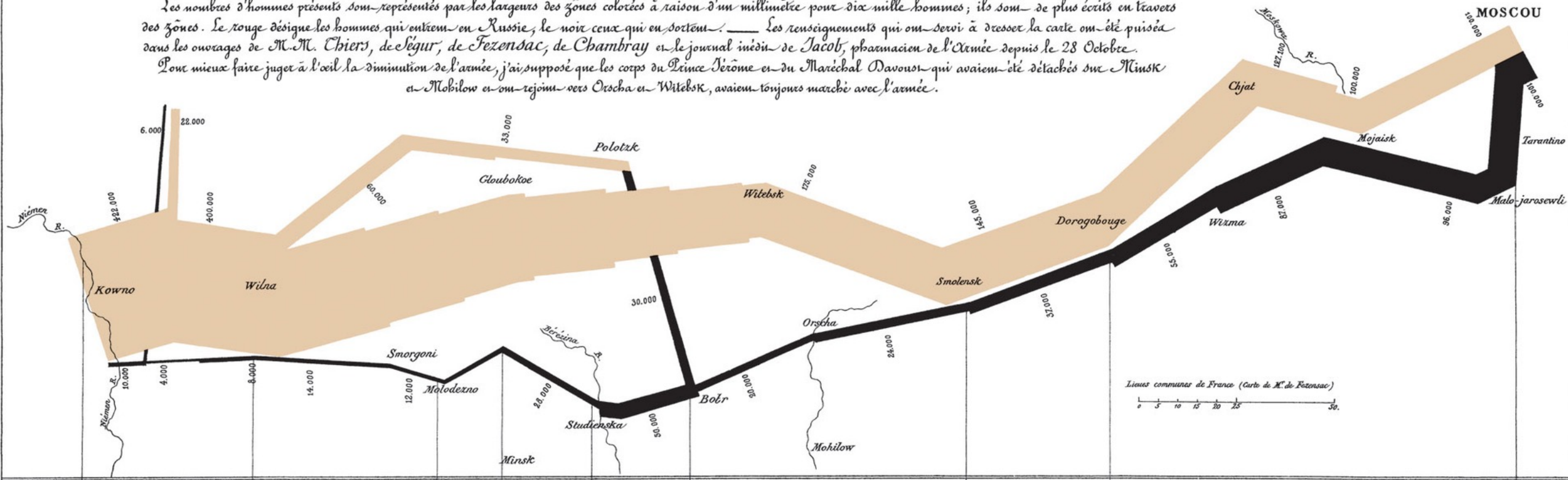
CITY AND RURAL POPULATION .
 1890 .



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.

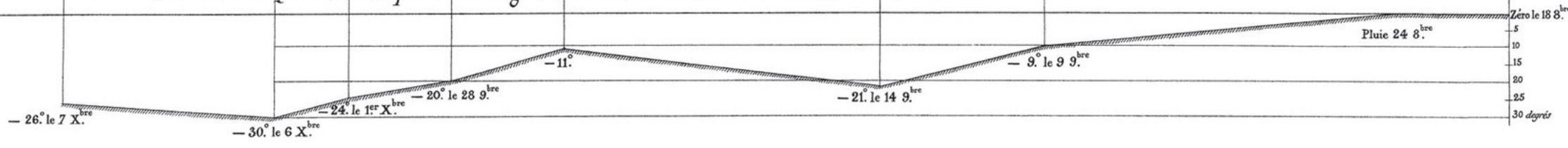
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit 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 Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Lignes communes de France (Carte de M. de Fezensac)
0 5 10 15 20 25 30

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

Les Cosaques passent au galop le Niémen gelé.



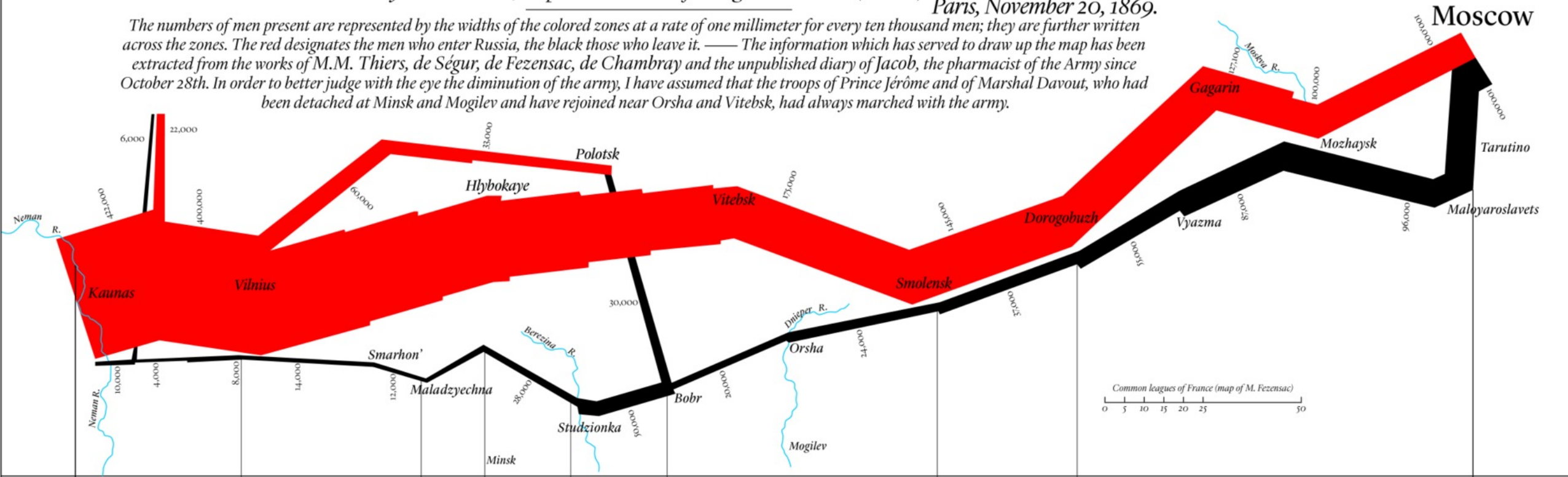
Minard's March to Moscow

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~ 1813

Drawn by M. Minard, Inspector General of Bridges and Roads (retired).

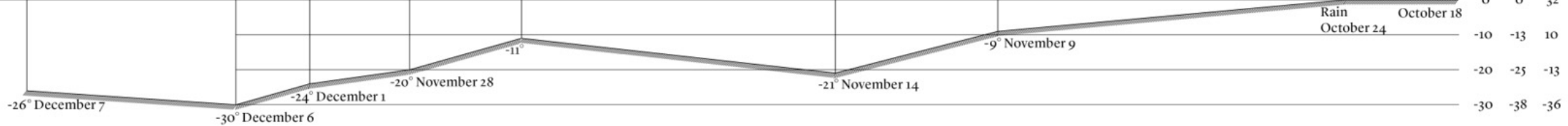
Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.



GRAPHIC TABLE of the temperature in degrees below zero of the Réaumur thermometer.

The Cossacks pass the frozen Neman at a gallop.



Minard's March to Moscow

Sister Courses

DATA SCIENCE ESSENTIALS

1. Non-Technical Aspects
2. Data Science Basics
3. Data Preparation
4. Data Engineering

INTRODUCTION TO MACHINE LEARNING

1. Statistical Learning
2. Classification
3. Clustering
4. Issues and Challenges