

La visualisation des données et les tableaux de bord

P. BOILY

UNIVERSITÉ D'OTTAWA | FACULTÉ DES SCIENCES | DÉPARTEMENT DES MATHÉMATIQUES ET DES STATISTIQUES
DATA ACTION LAB | IDLEWYLD ANALYTICS

Instructeur - Patrick Boily

Emploi

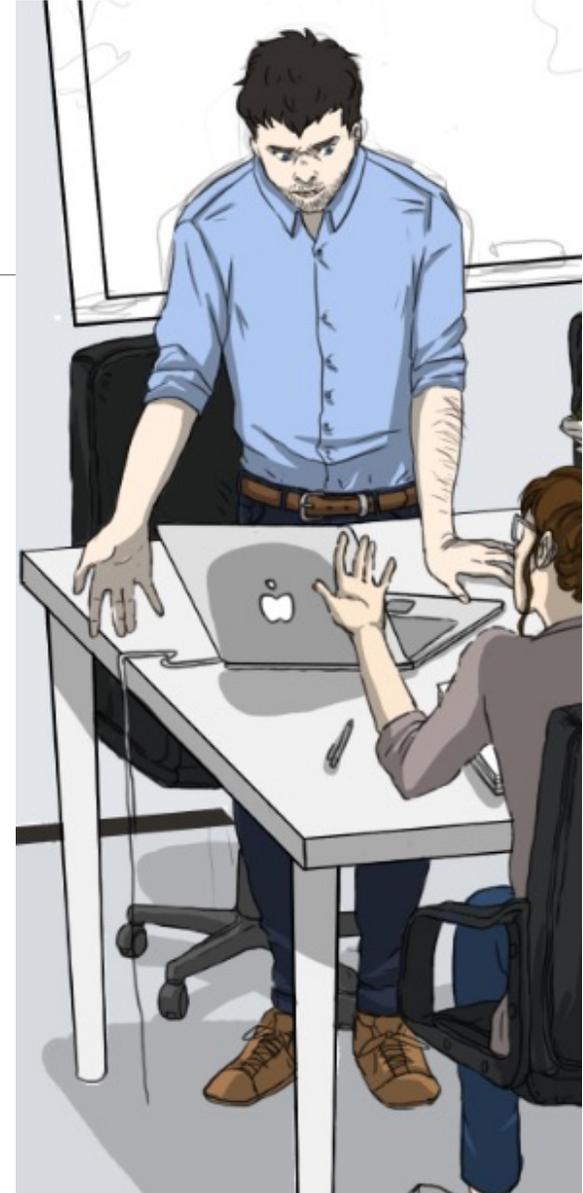
- Professeur Math/Stat [depuis '19, uOttawa]
- Président et conseiller principal [depuis '16, Idlewyld Analytics]
- Directeur et consultant principal ['12 - '19, CQADS, Carleton]
- Fonctionnaire GdC ['08 - '12, ASFC | StatCan | TC | TPSGC]
- 60+ cours universitaires ; 250+ jours d'atelier

Projets

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

Spécialisation

- Visualisation des données ; nettoyage des données (... malheureusement)
- Application d'un large éventail de techniques à tous les types de données
- Modélisation mathématique/statistique



Matériel de cours

Page web du cours :

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

Notes de cours :

<https://www.practicedataviz.com>

Contact :

pboily@uottawa.ca

Espace de travail Slack :

<https://dspdi.slack.com>

Description du cours

Avant d'exécuter des analyses, l'exploration des données est cruciale pour acquérir une compréhension de l'ensemble de données et de ce qu'il contient. Des stratégies de visualisation de plus en plus sophistiquées peuvent jouer un rôle important à cette étape.

La visualisation est également utilisée à la fin du processus d'analyse des données pour communiquer des informations sur les données à un public, sous la forme d'un tableau de bord.

Dans ce cours, les participants apprendront les stratégies fondamentales, sans outil, de visualisation et de communication des données (storytelling) et seront initiés à la grammaire des graphiques et aux concepts fondamentaux des tableaux de bord.

Après le cours, les participants ont la possibilité de travailler sur un projet guidé, en recevant un feedback de l'instructeur.

Informations supplémentaires

Une exposition aux cadres de programmation serait bénéfique mais pas nécessaire. Les participants doivent être à l'aise avec les concepts introduits dans *Les principes fondamentaux de la science des données*.

Les participants sont tenus d'apporter un ordinateur portable/personnel sur lequel la version actuelle de R/RStudio (Posit) est installée (pour lequel ils peuvent avoir besoin d'une autorisation administrative pour installer des paquets).

Les participants au projet guidé doivent être familiers avec R, tidyverse, Python.

Objectifs d'apprentissage

À la fin de ce cours, les participants seront en mesure de :

- développer des stratégies pour explorer un ensemble de données à l'aide de moyens visuels
- discerner la stratégie et le design à utiliser pour représenter au mieux les résultats de leur analyse de données
- créer des visualisations multivariées à l'aide de ggplot2
- concevoir des tableaux de bord conceptuels simples
- apprendre comment ils peuvent utiliser les notions de storytelling pour partager leurs idées d'analyse

Plan du cours

Les concepts de la visualisation des données

1. L'analyse exploratoire des données
2. La visualisation de données pour la communication
3. Un catalogue de visualisations
4. L'esthétique des graphiques
5. Les tableaux de bord

La mise en récit de données

6. Les éléments de la narration
7. Histoires et illustrations
8. L'évolution d'une mise en récit de données
9. L'anatomie des tableaux de bord narratifs

Session 1

Session 2

Session 3

Session 4

Plan du cours

La visualisation avec ggplot2

10. La grammaire des graphiques
11. Les bases de ggplot2
12. Exemples et miscellanea

Session 1

Session 2

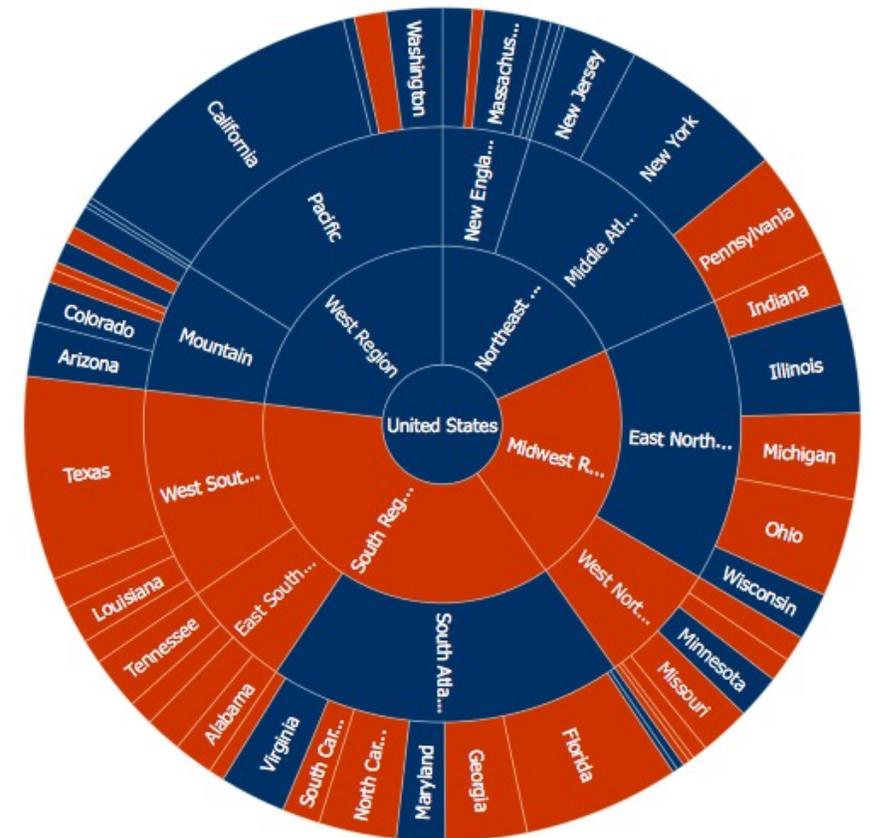
Session 3

Session 4

Visualisation de données vs. Infographies

Visualisation des données

- Une **méthode**, ainsi qu'un élément (**objectif**)
- Se concentre généralement sur le **quantifiable**
- Utilisé pour donner du sens aux données ou pour les rendre **accessibles** (les ensembles de données peuvent être massifs et difficiles à manier).
- Peut être généré **automatiquement**
- L'aspect et la convivialité sont moins importants que les **informations transmises par les données.**



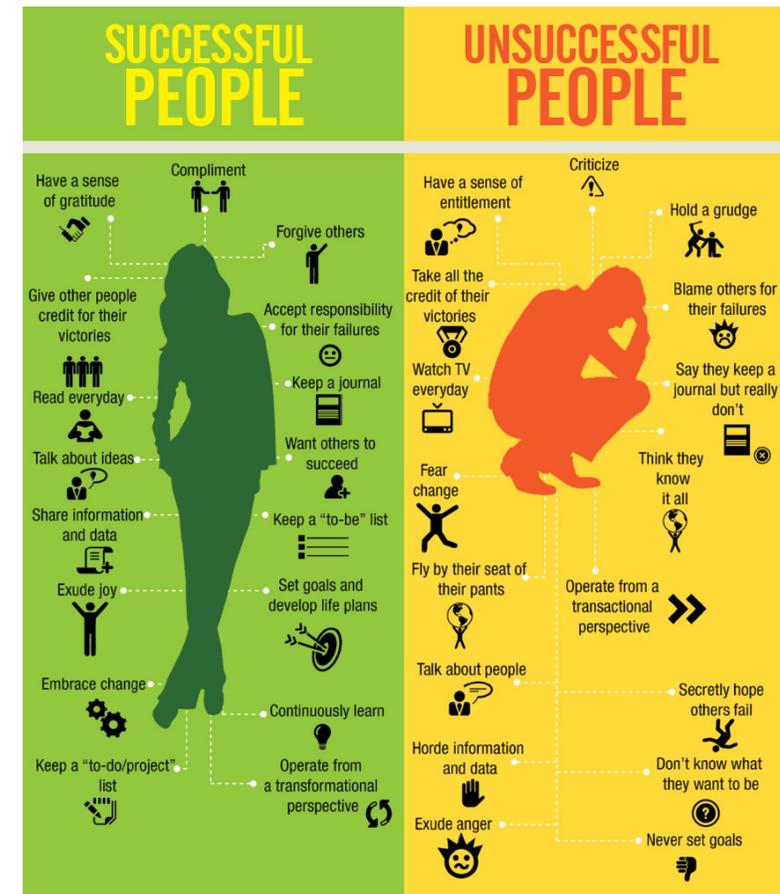
Size Population Color Median Household Income

■ Low Income ■ High Income

Visualisation de données vs. Infographies

Infographies

- Créé à des fins de **narration (subjectif)**
- Destiné à un public **spécifique**
- **Autonome** et discret
- L'aspect **graphique** est essentiel
- **Ne peut** généralement **pas** être réutilisé avec d'autres données
- Peut intégrer des informations **non quantifiables**



Graphiques historiques

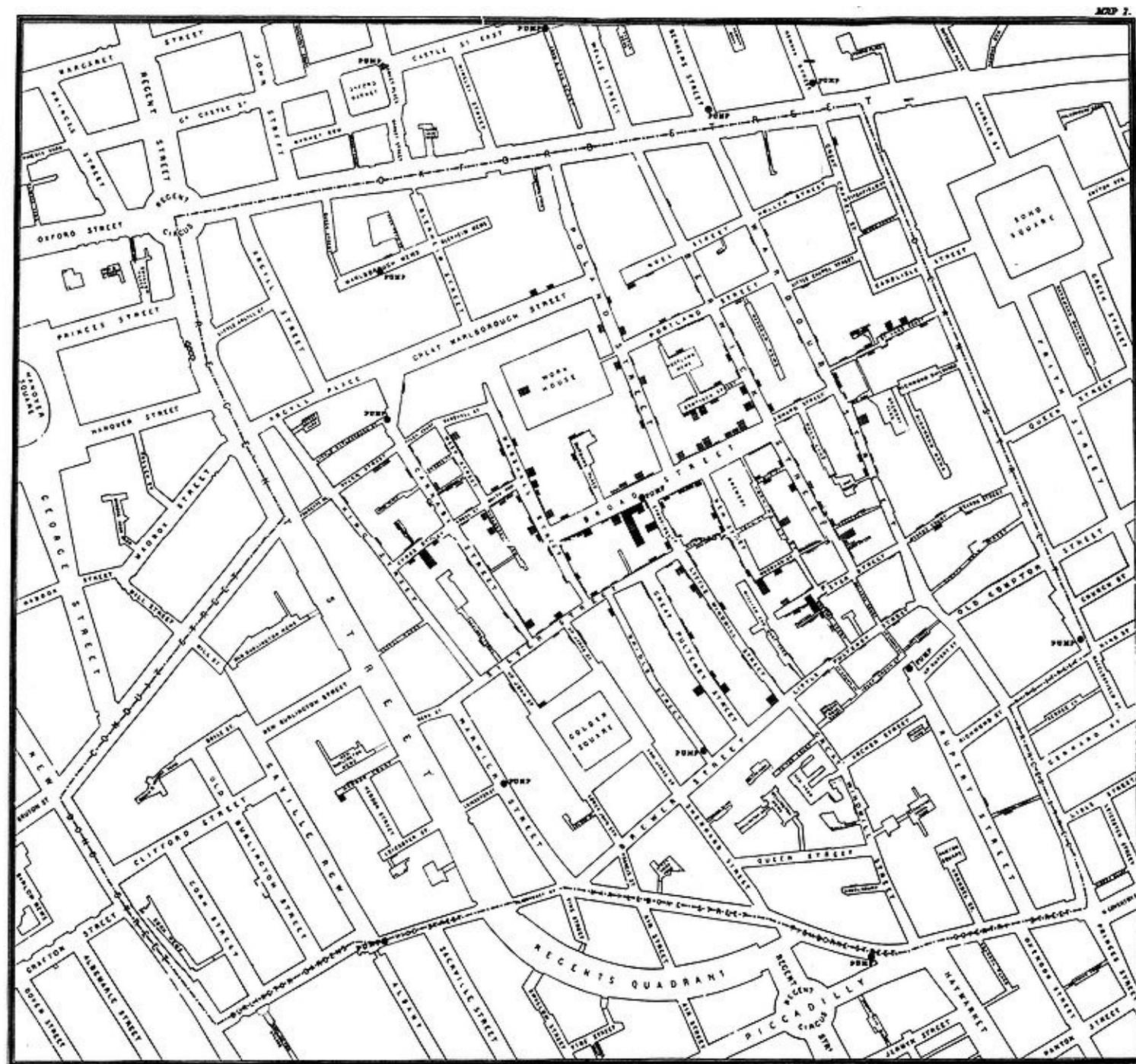
La visualisation des données ne se limite pas à un passé récent : les graphiques sont utilisés depuis de nombreuses années pour aider à **communiquer des informations** et à **raconter des histoires**.

En raison de l'absence d'outils techniques, il a fallu beaucoup de réflexion pour concevoir et créer ces visualisations.

Par conséquent, il y a beaucoup de choses que nous pouvons (et **devrions**) apprendre à apporter dans le développement des graphiques d'un **point de vue de la conception et de la narration**.

L'épidémie de choléra de 1854 à Londres

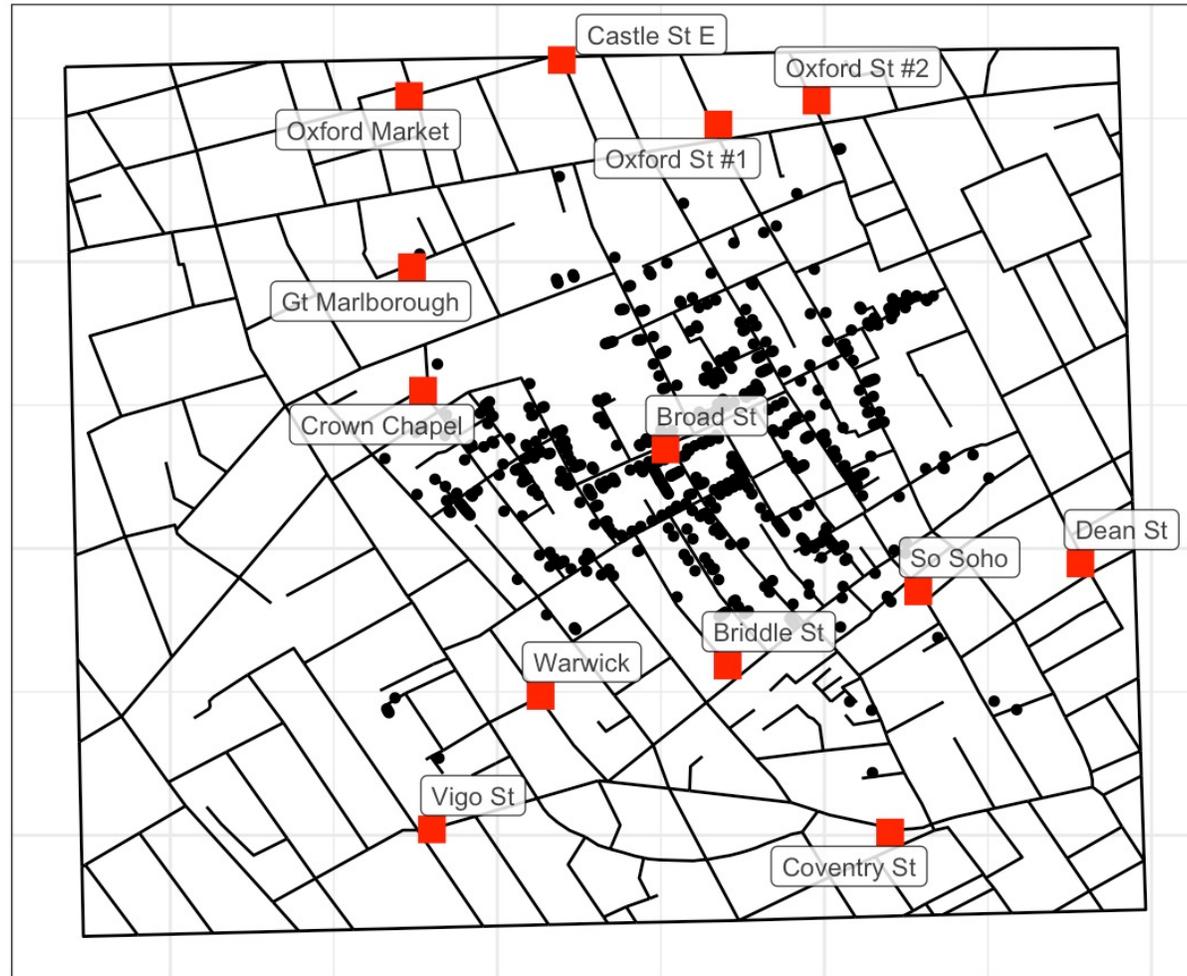
Le médecin John Snow établit un lien entre l'épidémie et un puits contaminé en reportant le nombre de cas sur une carte, ce qui marque le début de l'épidémiologie.



L'épidémie de choléra de 1854 à Londres

Le médecin John Snow établit un lien entre l'épidémie et un puits contaminé en reportant le nombre de cas sur une carte, ce qui marque le début de l'épidémiologie.

John Snow's London Cholera Outbreak Map (1854)



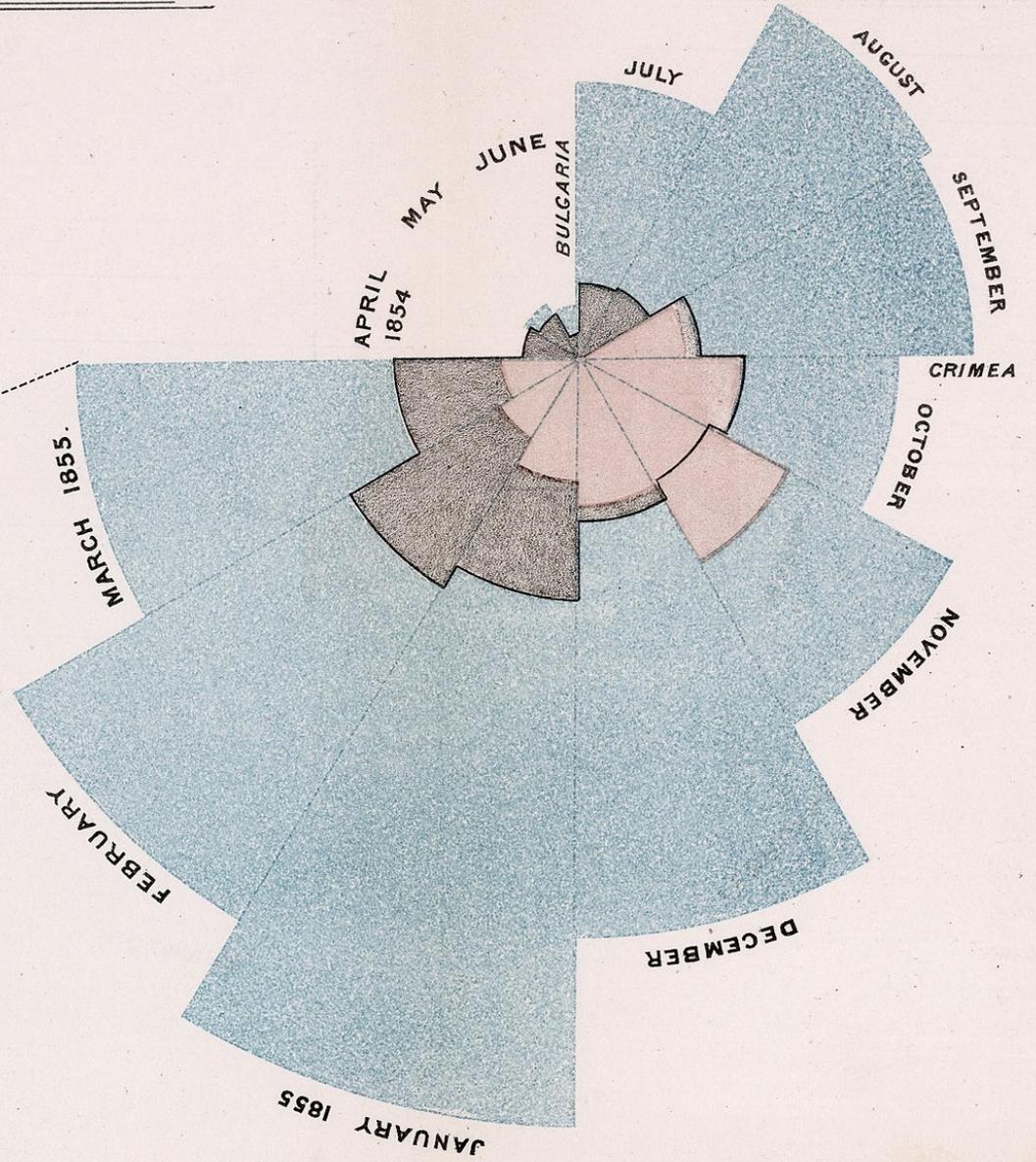
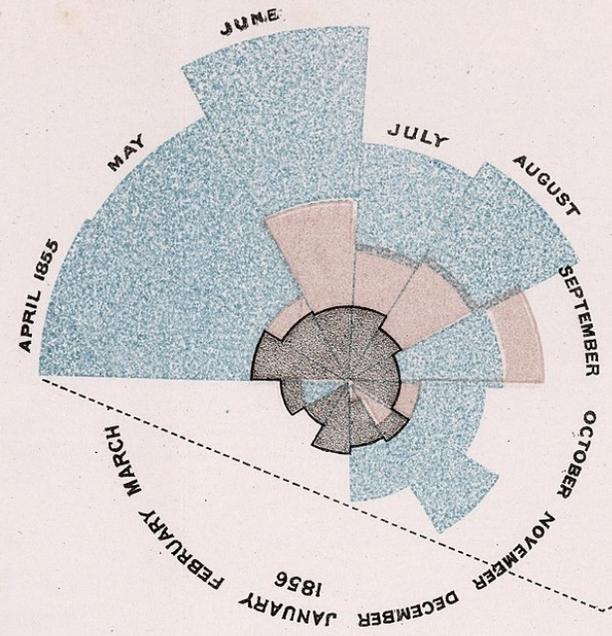
Locations

- deaths
- water pumps

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.

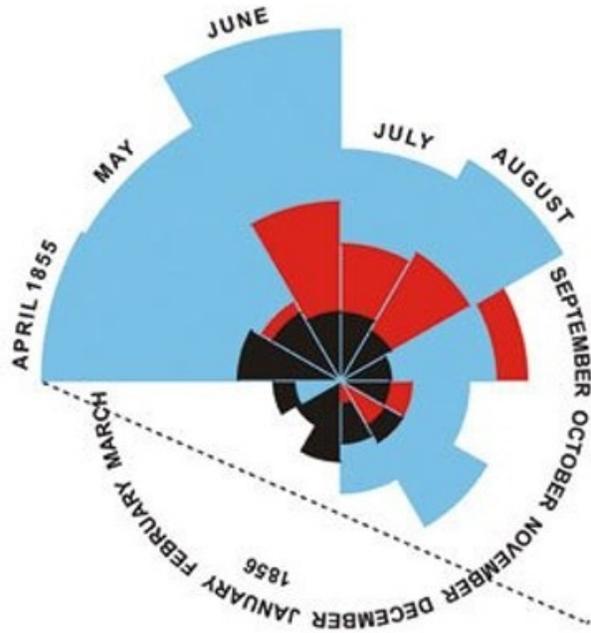
The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.

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

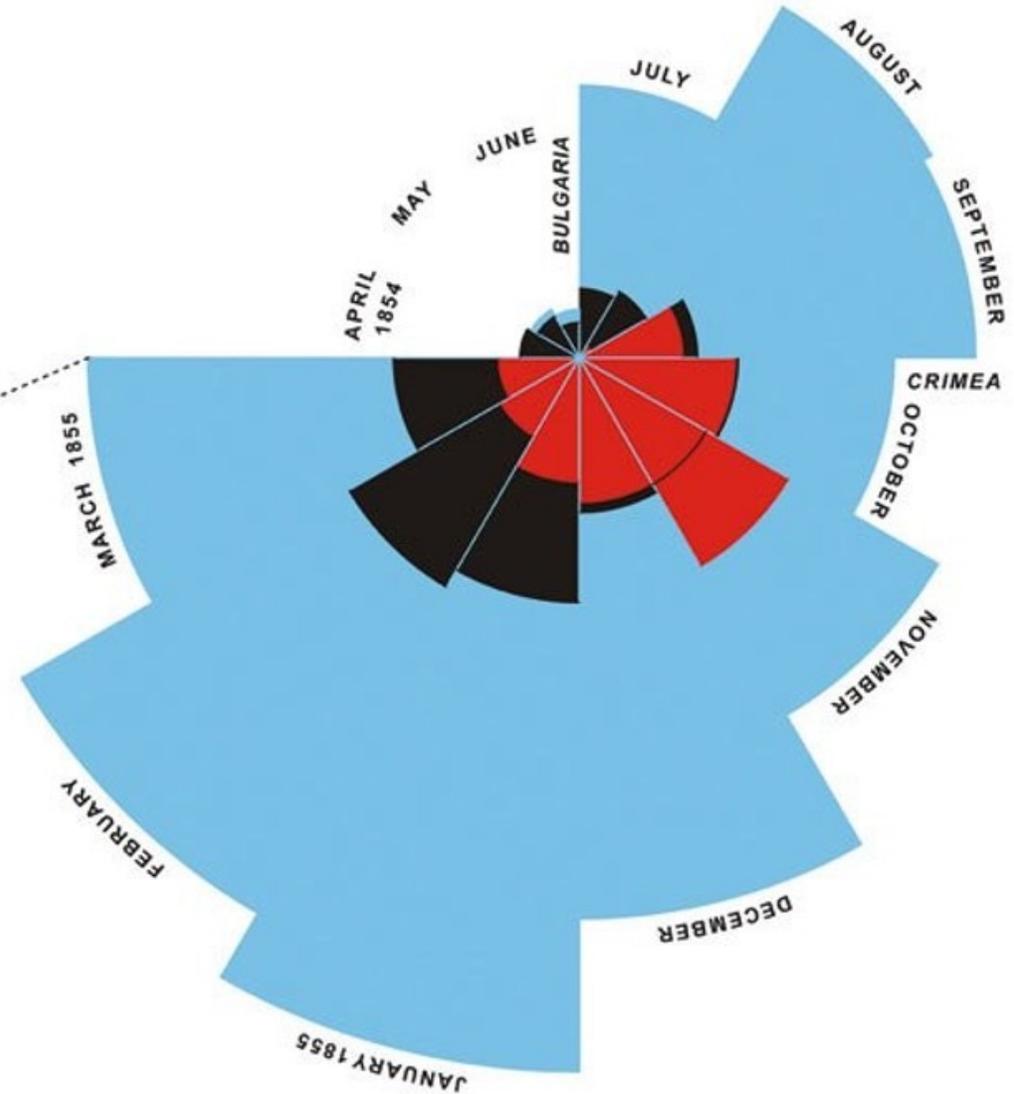
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

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

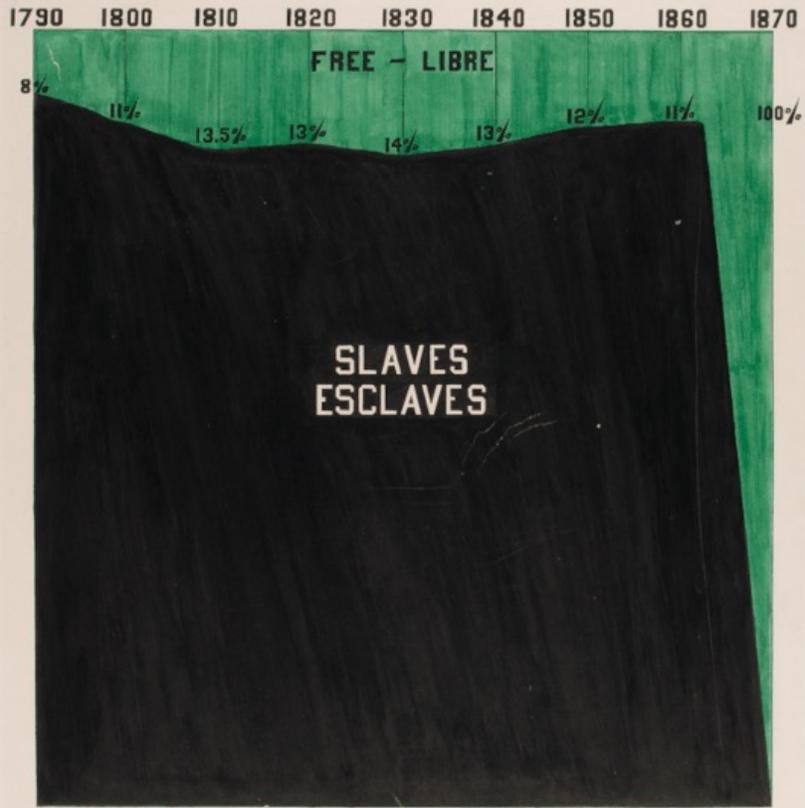
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month

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

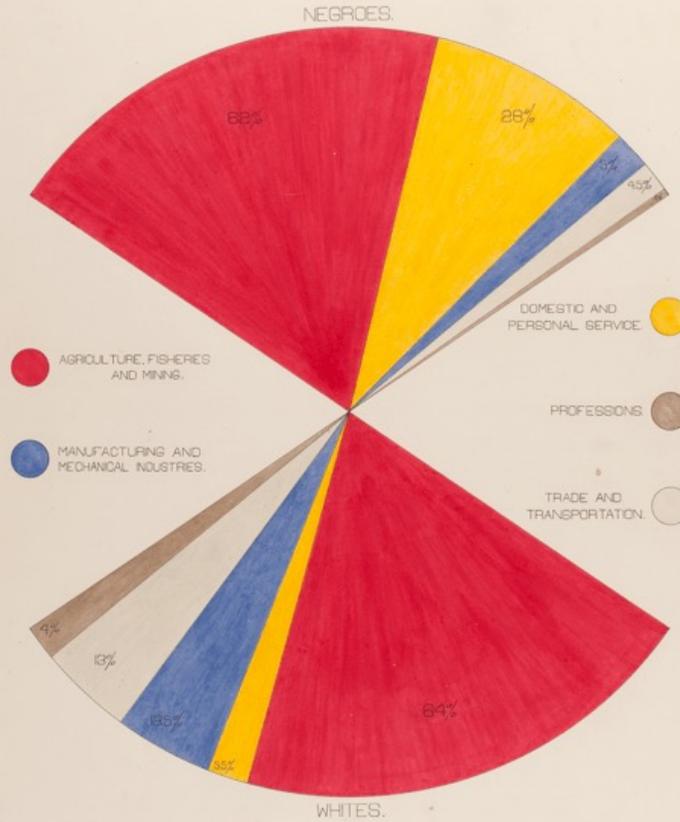
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

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.

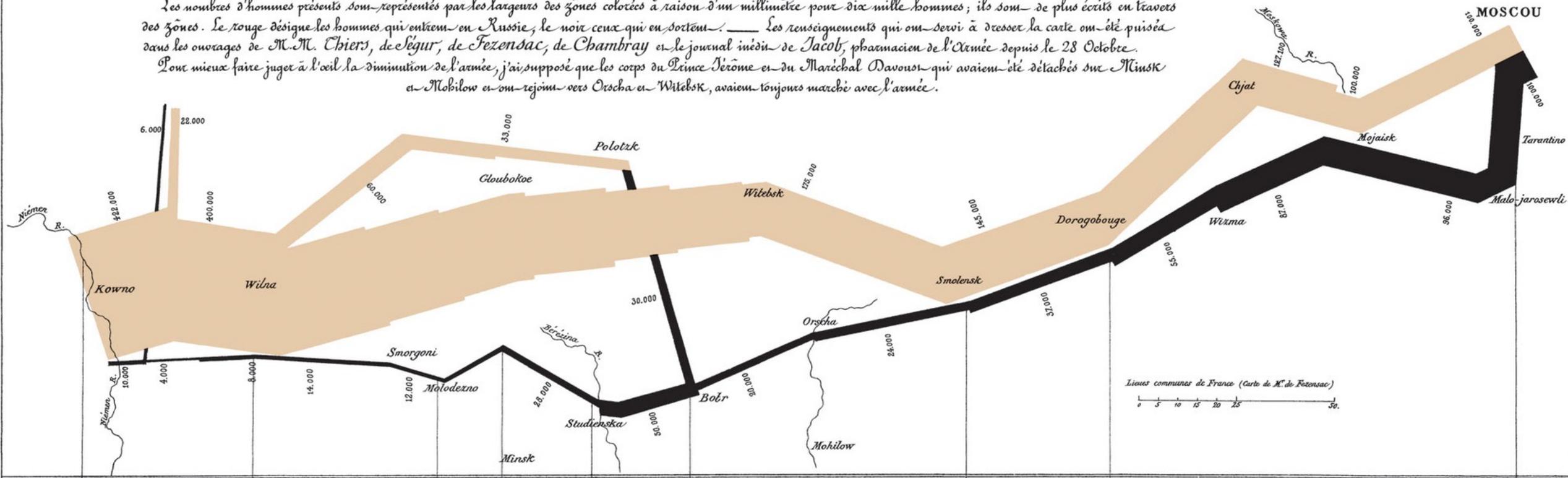
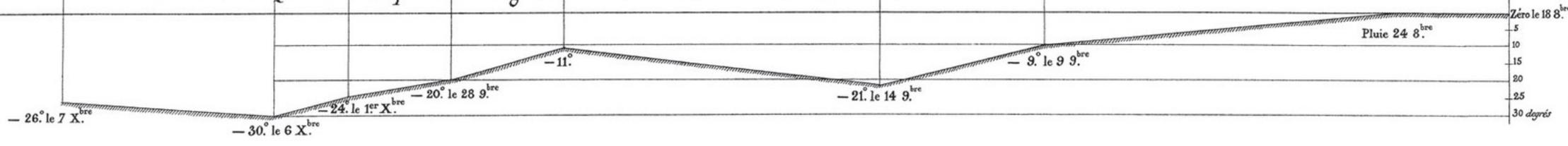


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.



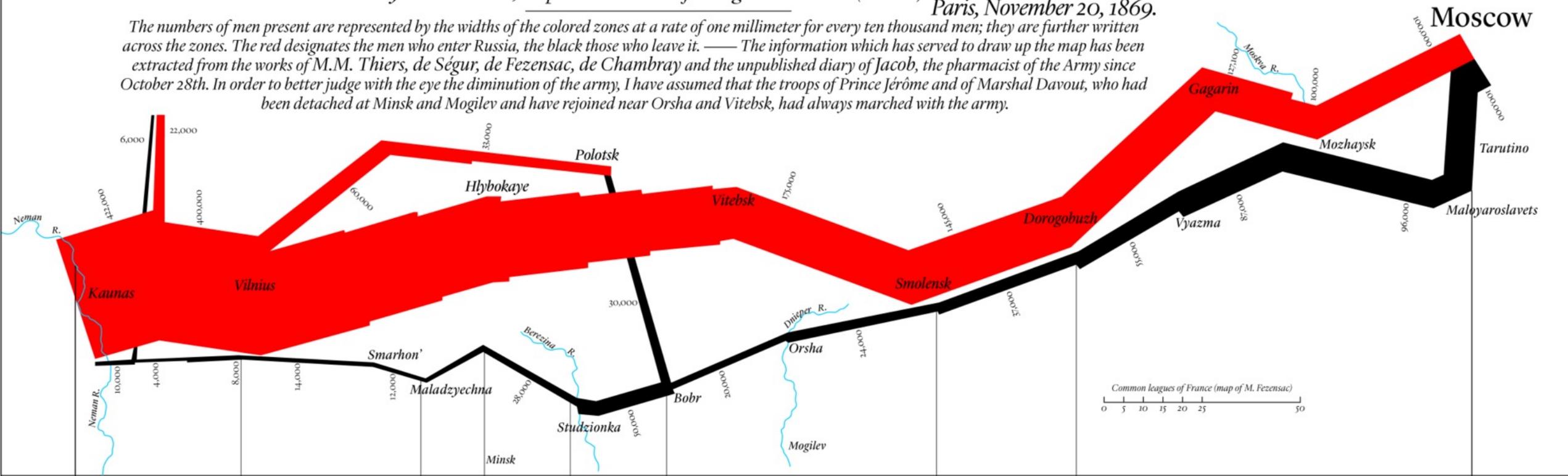
La marche vers Moscou

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).

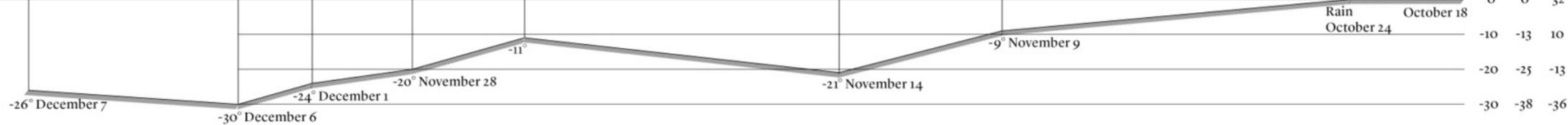
Paris, November 20, 1869.

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.



La marche vers Moscou

Cours sœurs

LES PRINCIPES FONDAMENTAUX DE LA SCIENCE DES DONNÉES

1. Les aspects non techniques
2. Les concepts fondamentaux
3. La préparation des données
4. L'ingénierie des données

INTRODUCTION À L'APPRENTISSAGE AUTOMATIQUE

1. L'apprentissage statistique
2. La classification
3. Le regroupement
4. Les problèmes et les défis