

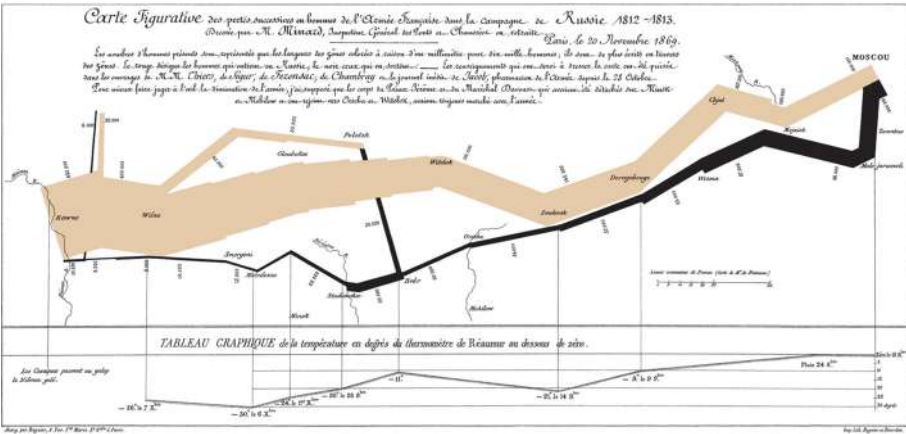
1. Pioneros de la visualización

Uno de los primeros trabajos en el ámbito de la visualización de datos es *The Visual Display of Quantitative Information*, escrito por Edward Tufte en 1983. Tufte es profesor de la Universidad de Yale y está considerado como uno de los pioneros de la visualización de datos. Tufte introdujo el uso de diagramas como metodología habitual para la descripción de datos y su análisis preliminar, como una herramienta más parte de la estadística. Es el inventor del concepto *chartjunk* (diagrama basura), para criticar el mal uso de las visualizaciones cuando no aportan nada a los datos representados.

En este trabajo, Tufte desgana una serie de visualizaciones de datos y las relaciona con el tipo de datos que se van a visualizar y el contexto en el cual fueron creadas, y también hace un repaso a algunas visualizaciones históricas, como el gráfico que describe las pérdidas del ejército francés durante la campaña de Napoleón en Rusia (1812-1813), creado por Charles Minard y mostrado en la figura 2. Tufte describe en detalle el uso de mapas, series temporales y su combinación, lo que él llama narrativas espacio-temporales, así como el uso de visualizaciones para mostrar relaciones entre elementos.

Pero lo más interesante del trabajo de Tufte es el concepto de *excelencia*, que define como la comunicación de ideas complejas con claridad, precisión y eficiencia. La excelencia es lo que proporciona al usuario de la visualización la mayor cantidad de ideas en el más corto espacio de tiempo mediante el mínimo uso de tinta y en el espacio más pequeño posible. La excelencia es casi siempre multivariada, no depende de una sola variable. Y, finalmente, la excelencia requiere contar la verdad sobre los datos, algo que desafortunadamente parece haberse perdido en muchas visualizaciones de carácter político, en las que se utilizan las visualizaciones como herramienta de manipulación. En este sentido, cualquier persona interesada en ser un «periodista de datos» debería adoptar las ideas de Tufte como principios básicos.

Figura 2. Mapa figurativo de las sucesivas pérdidas de hombres de la Armada Francesa en la campaña de Rusia 1812-1813, por Charles Minard (1869).



Fuente: Wikipedia

2. Antecedentes históricos

Aunque hay excelentes introducciones sobre qué es una visualización de datos y el significado del concepto a lo largo de la historia, la lectura recomendada para este apartado es el artículo «A brief history of data visualization», de Michael Friendly, publicado en el año 2006 como un capítulo de un manual de visualización de datos parte de una colección de libros de estadística, lo que da muestra de la importancia de la visualización como herramienta para el análisis de datos. Friendly es un autor muy prolífico en el ámbito de la visualización, impulsor del sitio web DataVis, en el cual se pueden encontrar muchos otros recursos relacionados, incluyendo artículos, libros y software.

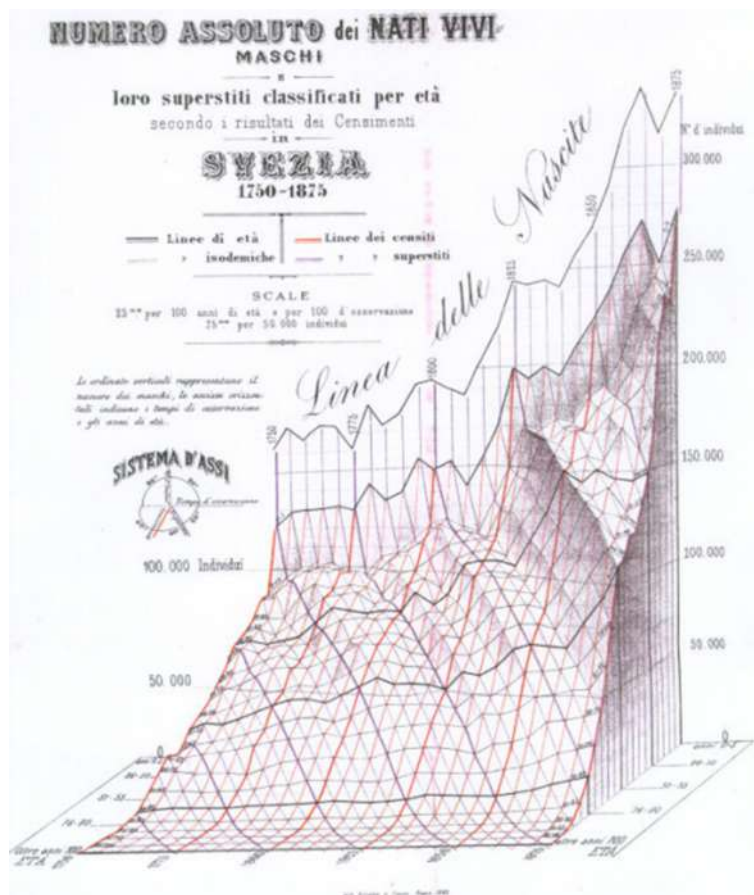
Michael Friendly

Sitio web: <http://www.datavis.ca/>

El artículo de Friendly está estructurado según una línea temporal, que incluye desde las primeras visualizaciones (de hecho, mapas y diagramas) previas al siglo XVII hasta la actualidad (a partir de 1975), donde la tecnología ha hecho posible la creación masiva de visualizaciones. De hecho, una revisión del artículo hoy día seguramente añadiría una nueva sección dedicada al *boom* que ha tenido la visualización recientemente, por la disponibilidad masiva de datos y la aparición de tecnologías que permiten la creación de visualizaciones de datos dinámicas, con la utilización de la web como medio.

Una de las etapas más interesantes destacadas por Friendly es la ocurrida en la segunda mitad del siglo XIX, cuando se desarrollaron muchas técnicas para el análisis estadístico, que eran aplicadas a todos los ámbitos de la planificación social, la industrialización, el comercio y el transporte. Esto provocó la aparición de muchas innovaciones en la visualización de datos, necesarias para poder explicar los datos y fenómenos tan complejos de la sociedad del momento. Un primer paso fue la utilización de elementos 3D proyectados como vía de escape del plano que hasta el momento limitaba las posibilidades, como el ejemplo de la figura 3. Otro fue la combinación de mapas con datos de cada región, de forma que en una misma representación se combinaban datos espaciales con otros temporales. Finalmente, el uso de gráficos para el análisis estadístico (la correlación era un concepto aún en desarrollo) permitió a Francis Galton y otros investigadores avanzar en la formalización de las observaciones realizadas y convertirlas en técnicas estadísticas, como muestra la figura 4. Un resultado de toda esta actividad fue la aparición de atlas estadísticos, es decir, informes de datos recopilados sobre casi todos los aspectos de la vida cotidiana acompañados de gráficos detallados. Friendly destaca la colección «Albums de Statistique Graphique», publicada anualmente por el Gobierno francés entre 1879 y 1897, y que fue discontinuada por su alto coste de producción, así como la realizada por el Gobierno de Estados Unidos entre 1872 y 1874.

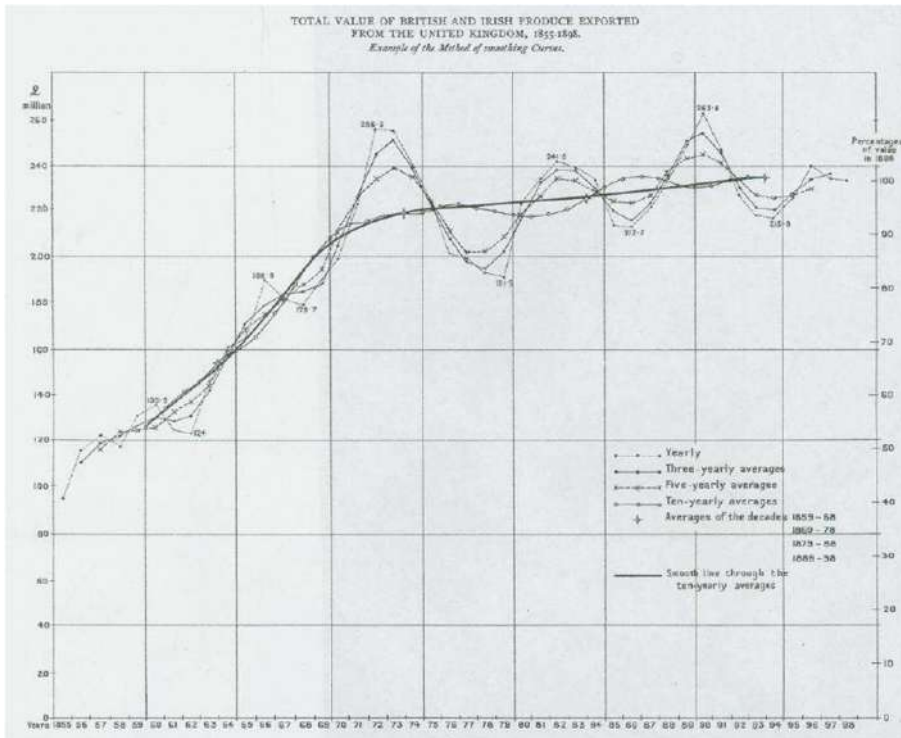
Figura 3. Población de Suecia entre 1750 y 1875, realizada por Luigi Perozzo en 1880.



Fuente: datavis.ca

Friendly destaca varios elementos que han sido clave en la evolución de lo que actualmente se conoce como visualización de datos. Por una parte, el desarrollo y formalización de herramientas estadísticas para el análisis de datos, y la consiguiente necesidad de representar los resultados obtenidos mediante estas. Por otra parte, la aparición de ordenadores y lenguajes de programación (como Fortran), que permitieron automatizar cálculos y crear las primeras representaciones gráficas para conjuntos de datos con cientos o miles de elementos.

Figura 4. Visualización del efecto del suavizado en series temporales, por Arthur Bowley, 1901.



Finalmente, Friendly resume la evolución de la visualización de datos sobre la base de la necesidad de resolver problemas concretos, relacionados con el deseo de visualizar fenómenos y relaciones entre elementos de forma diferente, y esto último es posible gracias al desarrollo de las metodologías (análisis estadístico) y las tecnologías (ordenadores).