

# Visualización de Información

Daniela Opitz

[dopitz@udd.cl](mailto:dopitz@udd.cl)

Instituto Data Science, Universidad del Desarrollo

Edición 2024

# Outline

- Introducción a la Visualización
  - ¿Qué es?
  - ¿Por qué la necesitamos?
- Herramientas
  - Anaconda
  - Entorno de trabajo
- Evaluaciones del Curso
- Bibliografía

# Introducción a la Visualización

# Motivación

Un video de Hans Rosling titulado The Joy of Stats

<https://www.youtube.com/watch?v=V8lbiiTF2P0>



Japón y otros que lo siguen directamente

**¿Opiniones?**

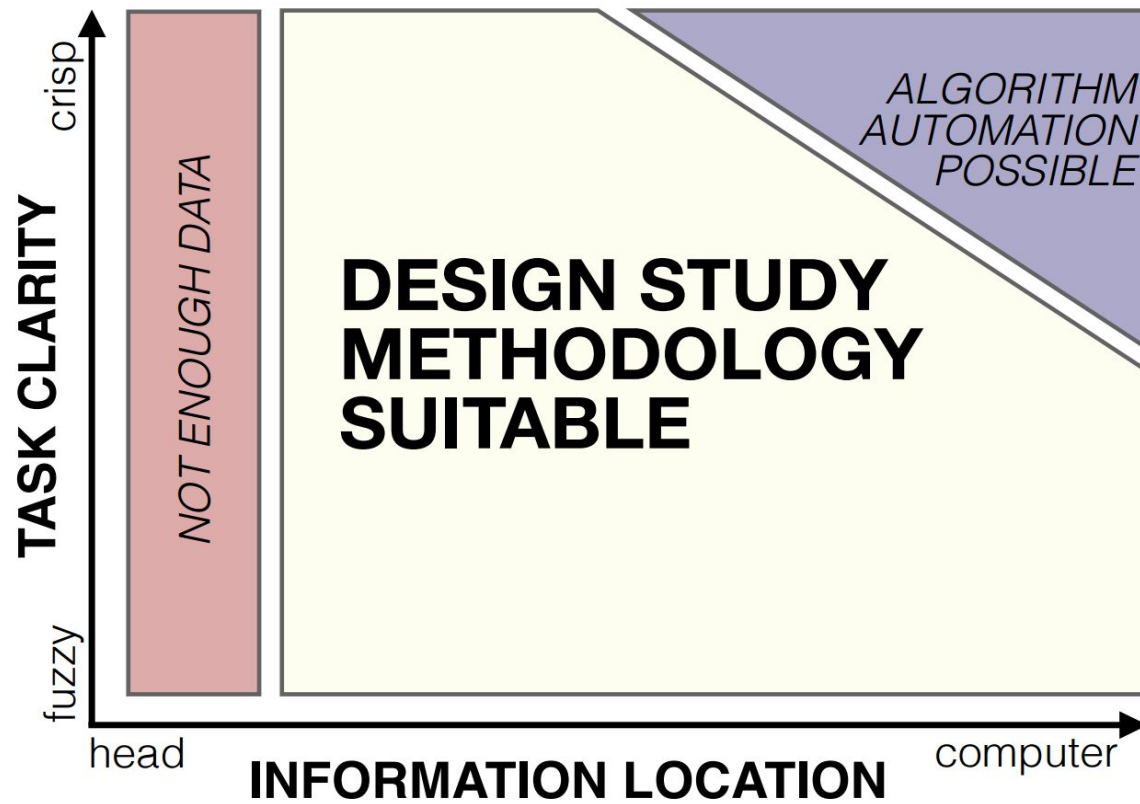
## Definición

Los **sistemas computacionales** de visualización proveen **representaciones visuales** de conjuntos de **datos** diseñadas para ayudar a las **personas** a efectuar tareas de manera más **efectiva**.

# ¿Cuándo necesitamos una visualización?

- Necesitamos aumentar la capacidad humana para obtener insights (no reemplazarla)
- Para muchos problemas de análisis las preguntas no están claras desde el inicio
- Algunas tareas no pueden ser automatizadas.





¿Cuándo es posible usar visualización?

Tamara Munzner. <https://www.cs.ubc.ca/labs/imager/tr/2012/dsm/>

# ¿Por qué hay computadores involucrados?

- Capacidad de procesamiento de datos.
  - Más rápido, más cantidad.
  - Análisis de variabilidad.
- Escalabilidad.
  - Las personas pueden realizar ciertas tareas para un dataset pequeño. ¿Son capaces de hacerlas para un dataset mucho más grande?
- Colaboración.
  - Tanto en el desarrollo como en el uso de un sistema de visualización.

# La visualización como representación externa

- Reemplaza cognición por percepción.
- Un diagrama puede organizar la información de manera que podamos realizar consultas de manera eficiente.

2009 Sales (thousands of U.S. \$)													
Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493	31,783
International	574	636	673	593	644	679	593	139	599	583	602	690	7,005
Total	2,557	2,979	3,266	2,876	3,218	3,517	2,975	2,773	3,537	3,322	3,585	4,183	38,788

Stephen Few

Why do we visualize Quantitative Data?

<https://www.perceptualedge.com/blog/?p=1897>

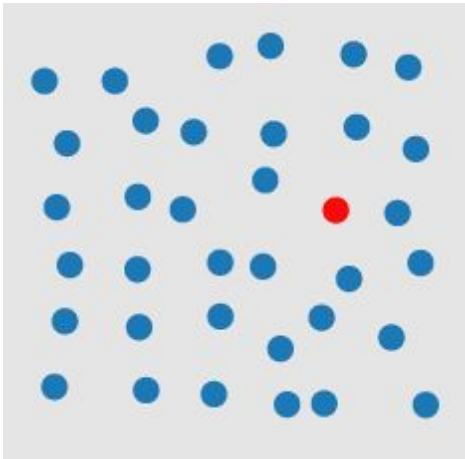


Stephen Few. Why do we visualize quantitative data?

<https://www.perceptualedge.com/blog/?p=1897>

# ¿Por qué depender de la visión?

- Es un sentido perceptualmente eficiente.
  - La comunicación con el cerebro es rápida.
  - Una gran proporción del procesamiento de información visual ocurre simultáneamente
  - En contraste, otros sentidos son secuenciales (por ej., sonido).
- Todavía no hemos desarrollado el entendimiento ni la tecnología para construir sistemas basados en el tacto o el olor tan efectivos como los visuales. Todavía no tenemos interfaces hápticas (tacto) mainstream.



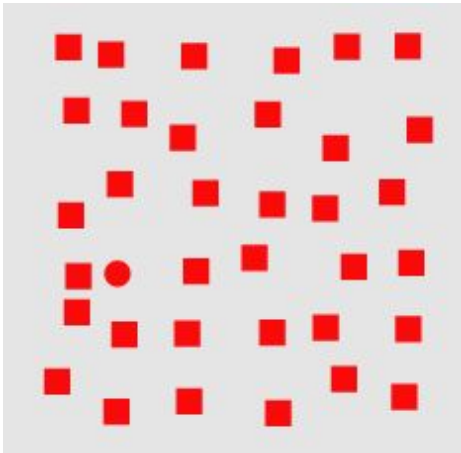
## ¿Dónde está el punto rojo?

Fuente:

Perception in Visualization

Christopher G. Healy

<https://www.csc2.ncsu.edu/faculty/healey/PP/>



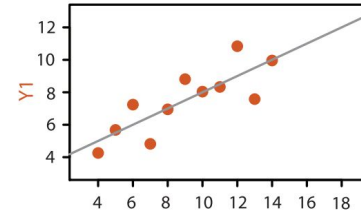
# ¿Por qué mostrar los datos en detalle?

- Las estadísticas descriptivas ocultan variabilidad que puede ser importante.
  - Puede inducir a errores de interpretación.
- Los datos agregados pueden ocultar patrones en los datos.
  - Limita la capacidad de encontrar algo que no esperábamos, porque la agregación usualmente se hace pensando en un fin específico.

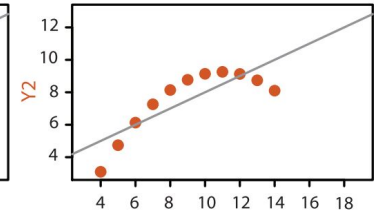
# ¿Por qué mostrar los datos en detalle?

Anscombe's Quartet: Raw Data

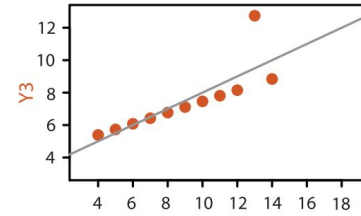
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



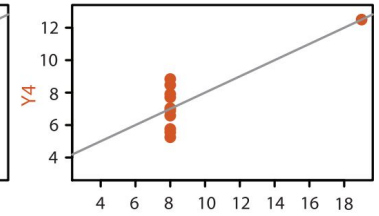
X1



X2



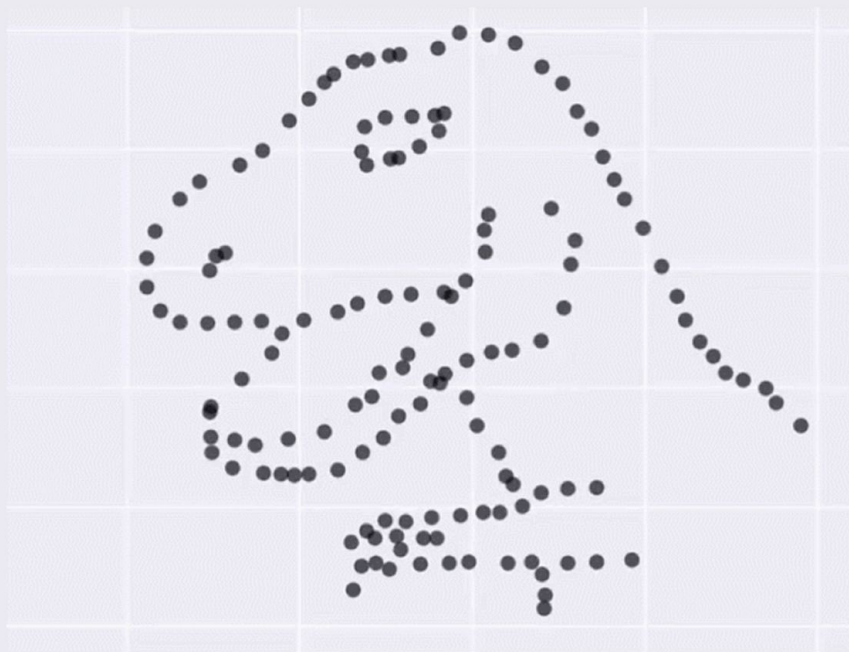
X3



X4

[https://en.wikipedia.org/wiki/Anscombe%27s\\_quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet)





Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing.

Justin Matejka, George Fitzmaurice. ACM SIGCHI Conference on Human Factors in Computing Systems, 2017.

# ¿Por qué usar interactividad?

- Una visualización estática solamente permite una única vista de los datos.
  - A medida que nos acercamos a los límites de personas y computadoras, la interactividad permite que lo que se muestra cambie: potencialmente infinitas vistas, cada una adaptada a los límites.
  - “InfoVis Mantra”: Overview first, zoom & filter, details on demand (Ben Shneiderman)
- “Una imagen vale mil palabras. Una interfaz de usuario vale mil imágenes” (B. Shneiderman también)

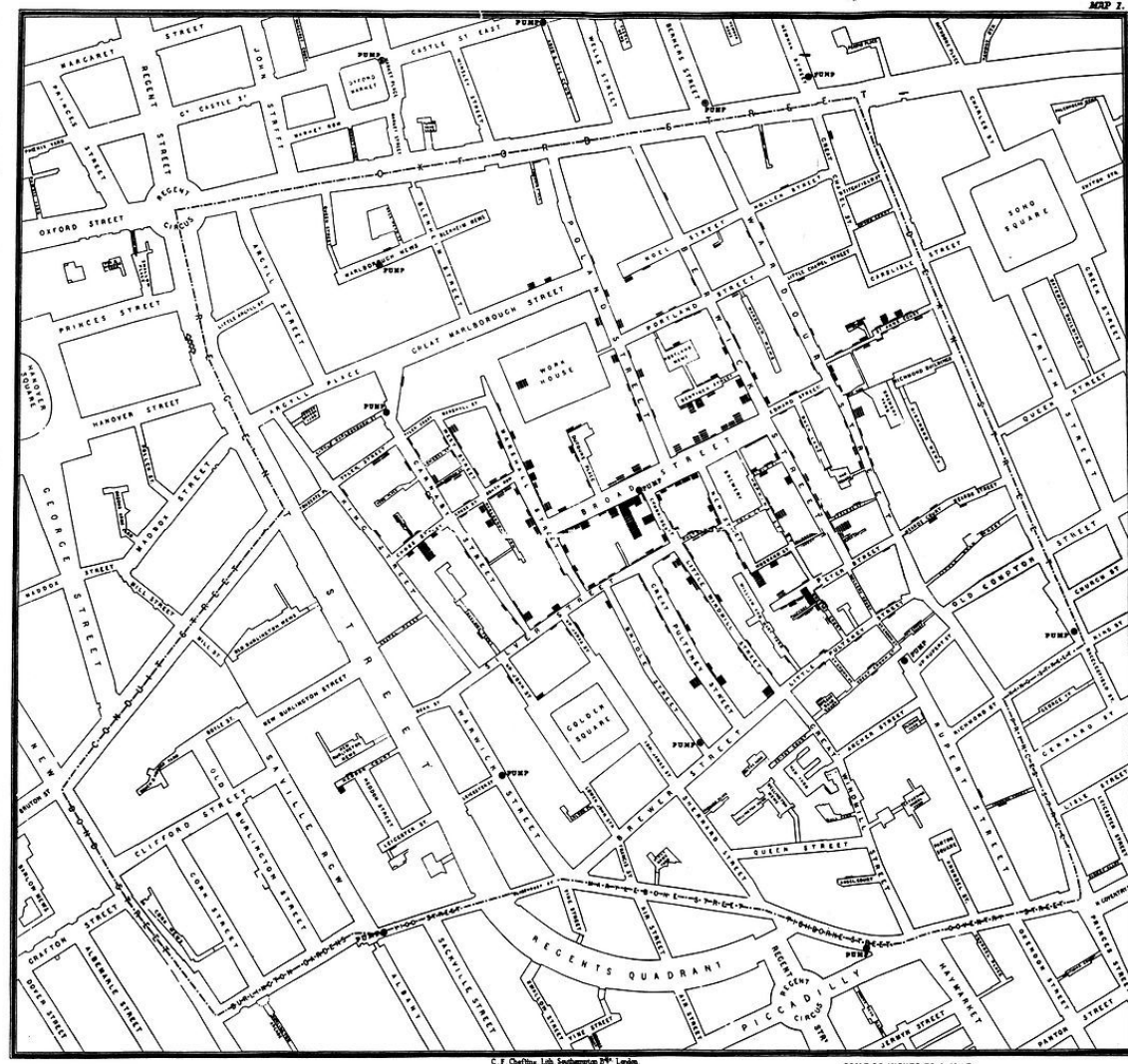
# ¿Por qué enfocarse en las tareas a realizar?

- La visualización **no es arte** (puede llegar a serlo). Nuestro propósito es crear **herramientas efectivas**.
  - La efectividad depende del contexto. Así, las tareas permiten restringir y enfocar el diseño.
  - Podemos descubrir algo que de otro modo no hubiésemos descubierto
- No todas las tareas están relacionadas a necesidades de negocio.
  - A veces la tarea es única y personal

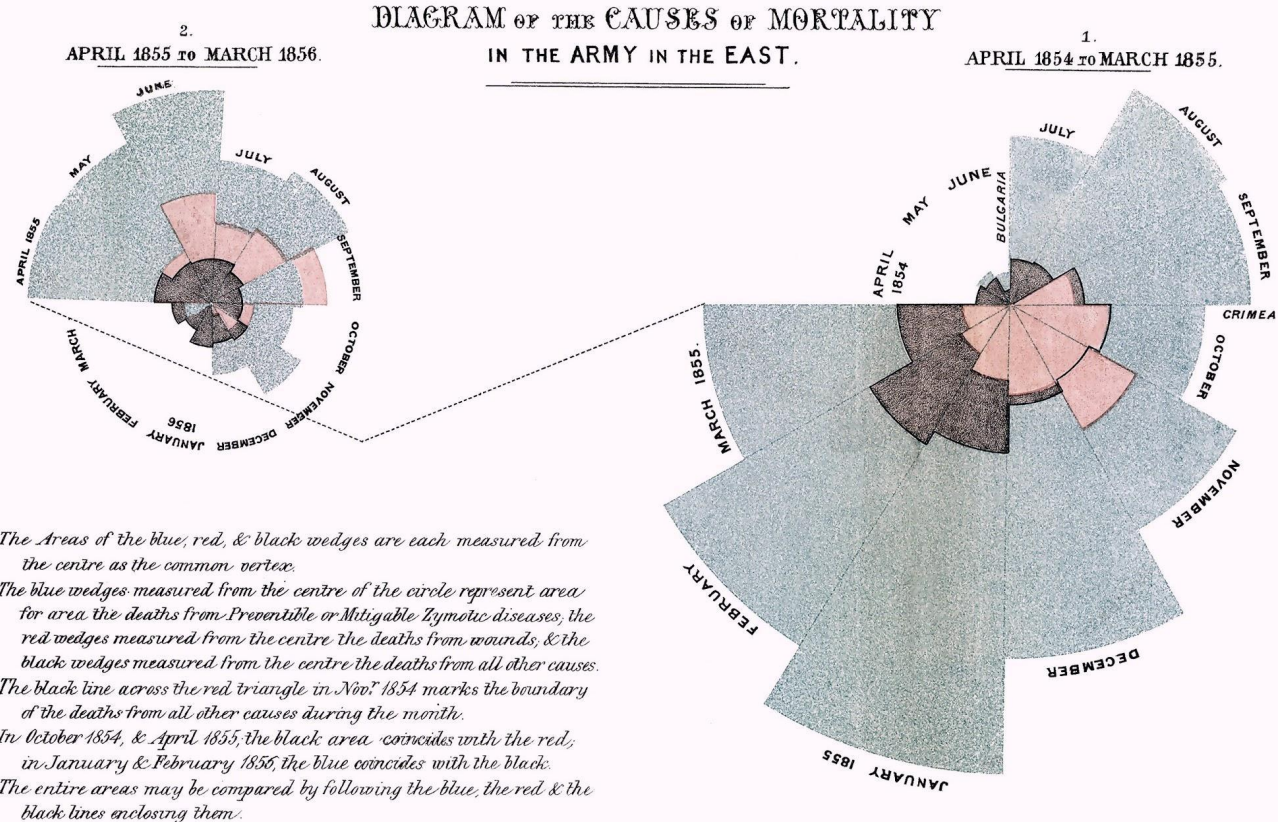
# **Ejemplos de Visualizaciones**

**Tarea: encontrar relación entre incidencia de cólera y fuentes de contagio.** Por John Snow en 1855.

[https://en.wikipedia.org/wiki/1854\\_Broad\\_Street\\_cholera\\_outbreak](https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak)



## Diagrama de la Rosa

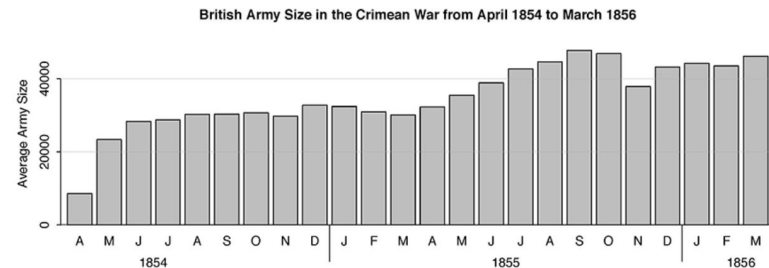
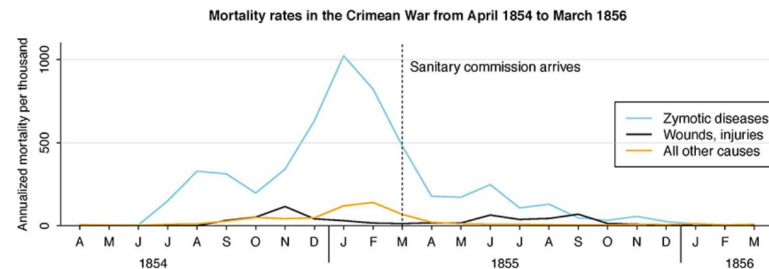
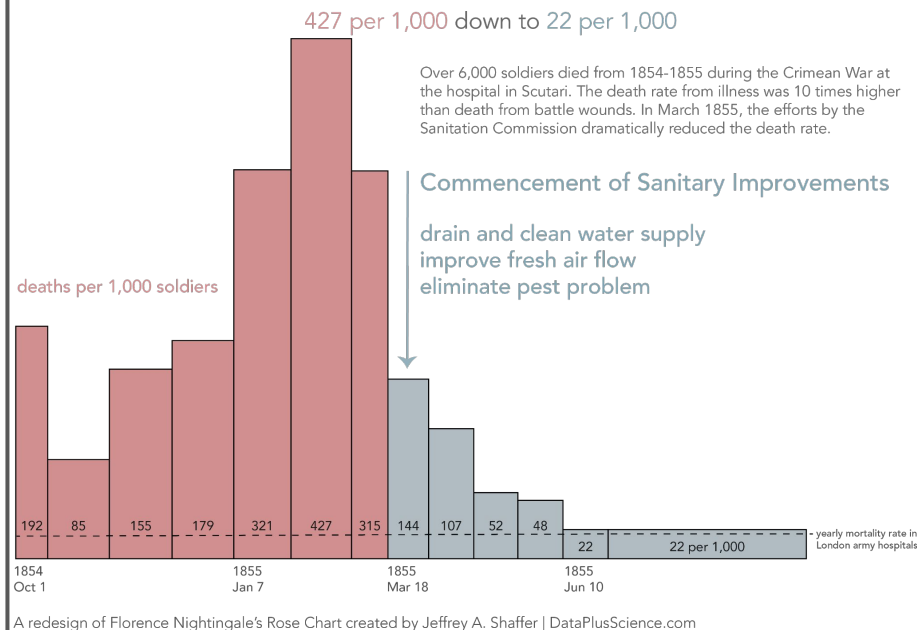


**Tarea: visualizar el impacto en la reducción de muertes en los hospitales en las colonias inglesas tras una intervención de higiene.** Por Florence Nightingale en 1857.

[https://es.wikipedia.org/wiki/Florence\\_Nightingale](https://es.wikipedia.org/wiki/Florence_Nightingale)

# The Scutari Death Camp

simple sanitation improvements can save lives



The Crimean War mortality data as time series graphs. The use of different forms (line plot for death rates and bars for army size) is a visual cue that two sorts of data are being presented: monthly rates (multiplied by 12 to be annualized, following Nightingale's calculations) and absolute numbers. Courtesy: Andrew Gelman

**Dos rediseños de la visualización anterior.**



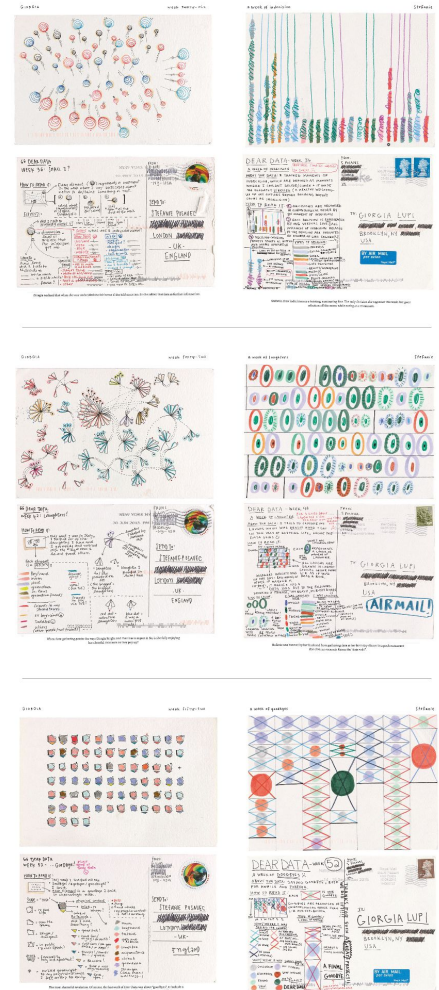
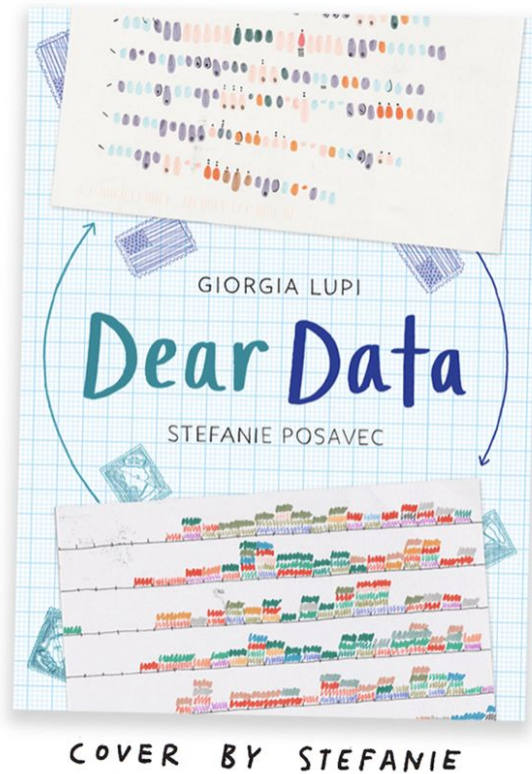
GIORGIA Stefanie



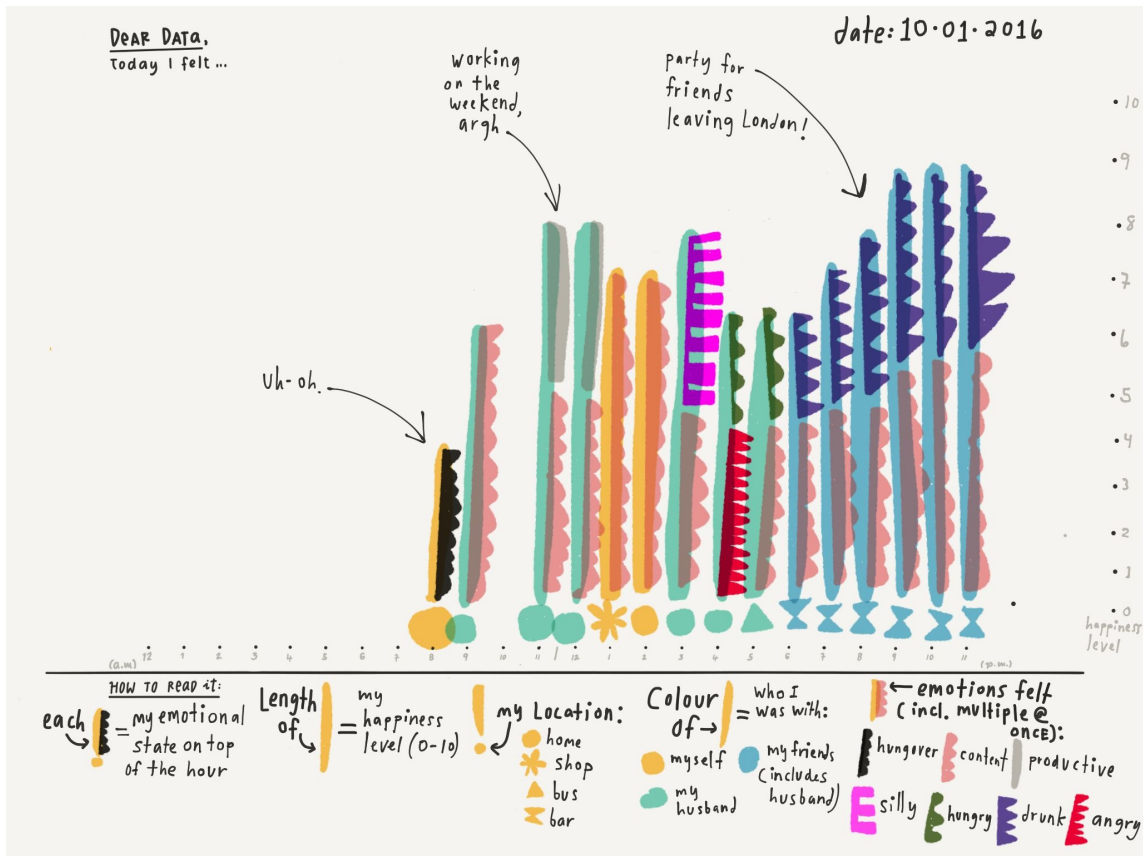


**Tarea:** Compartir información cotidiana a través de tarjetas postales para conocerse mejor.

**Dear Data.** Por Giorgia Lupi and Stefanie Posavec



# Dear Data





Pero la pregunta de verdad es

# ¿Por qué hay recursos limitados?

- Estamos sujetos a los límites humanos
  - Change Blindness: no vemos cambios grandes si estamos atentos a otra cosa.
  - Limitada memoria de corto plazo.
- Estamos sujetos a límites del medio como los límites del papel o del computador
- Y los límites de los datos
  - Suciedad, completitud, sesgos, etc.

## Resumen

Los **sistemas computacionales** de visualización proveen **representaciones visuales** de conjuntos de **datos** diseñadas para **ayudar a las personas** a **efectuar tareas** de manera más **efectiva**.

**¿Preguntas?**

# Herramientas

# Herramientas





# Herramientas (Anaconda)

- Mac:

[https://www.anaconda.com/products/distribution#Downloads~:text=64%2DBit%20Graphical%20Installer%20\(688%20MB\)](https://www.anaconda.com/products/distribution#Downloads~:text=64%2DBit%20Graphical%20Installer%20(688%20MB))

- Windows: [https://repo.anaconda.com/archive/Anaconda3-2022.10-Windows-x86\\_64.exe](https://repo.anaconda.com/archive/Anaconda3-2022.10-Windows-x86_64.exe)

- Linux:

[https://www.anaconda.com/products/distribution#Downloads~:text=64%2DBit%20\(x86\)%20Installer%20\(737%20MB\)](https://www.anaconda.com/products/distribution#Downloads~:text=64%2DBit%20(x86)%20Installer%20(737%20MB))

# **Evaluaciones**

# Evaluaciones

1- **Certamen 1 (teorico)**

2- **Certamen 2 (practico): proyecto 2 de visualización** grupal de un tema cotidiano o de relevancia personal, científica, industrial o pública usando un dataset público y **Python**

3- **Examen:** presentación presencial del proyecto 2 extendido y mejorado.

# Evaluaciones

certamen 1  
50%

certamen 2  
50%

nota de  
presentación a  
examen  
70%

examen  
(superior a 3.0)  
30%

# Certamen 2

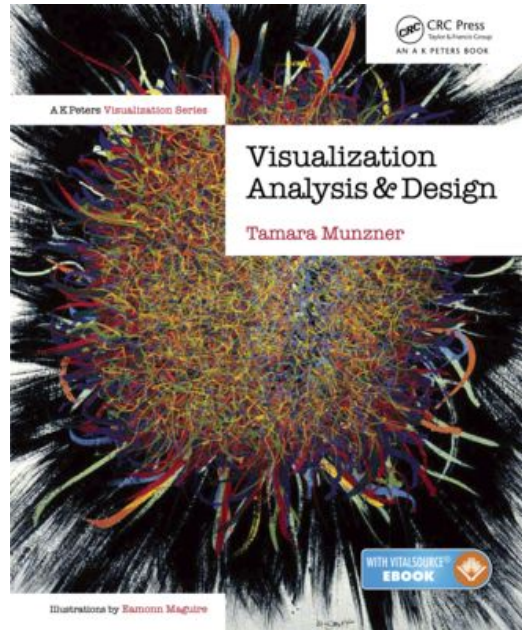
## **Compuesto de dos partes:**

- parte 1: avance semanal. Cada entrega tiene un 7 por defecto. Se promedian todas las entregas con la posibilidad de borrar sólo 1 nota. Algunas entregas serán presentaciones de lecturas.
- parte 2: evaluación del proyecto.

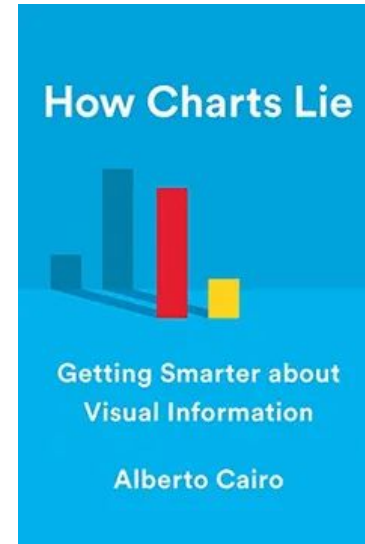
# Bibliografía

# Bibliografía

**Visualization & Analysis Design ([Tamara Munzner](#))**



**How Charts Lie ([Alberto Cairo](#))**



# Tareas Próxima Clase

- Pensar en alguna actividad cotidiana que les gustaría visualizar
- Pensar en un tema en el cual ustedes creen que pueda ser de utilidad tener una visualización. ¿Qué pregunta les gustaría responder?
- Instalar anaconda y abrir Jupyter
- Instalar numpy, pandas, matplotlib y seaborn