

# Visualización Interactiva de la Información

Luis Fernando Franco Jiménez

03 Estadios de la visualización

# Percepción Visual y Cognición

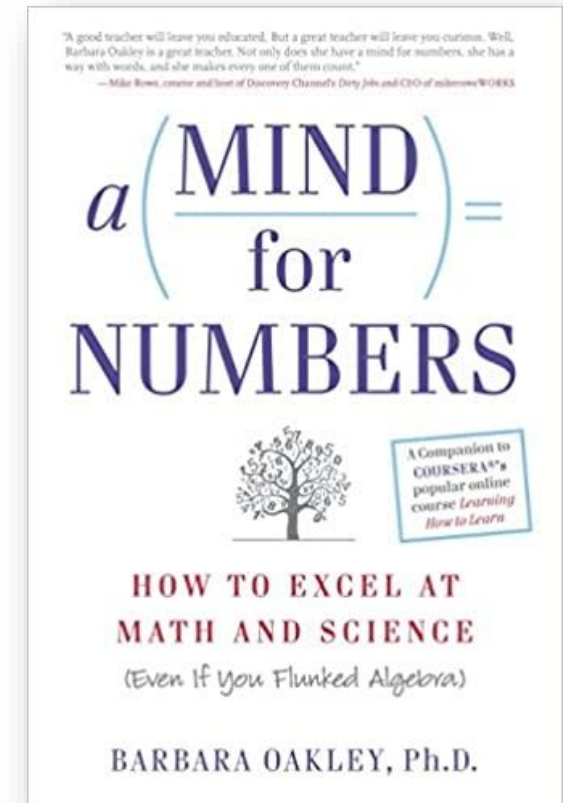
Tres tipos de memoria:

- Memoria Icónica
  - Fase “preatentiva” automática
  - Prácticamente inmediata
  - Detección de color y Ubicación
- Memoria de Corto Plazo
  - El cerebro procesa de forma “atentiva”
  - Selecciona partes específicas de la información
  - Tiene una capacidad limitada
  - Ordena la información en bloques (o *chunks*)
- Memoria de Largo Plazo
  - Principalmente se llega por procesos de repetición
  - Tiene capacidad prácticamente ilimitada

# Percepción Visual y Cognición

Tres tipos de memoria:

- Memoria Icónica
  - Fase “preatentiva” automática
  - Prácticamente inmediata
  - Detección de color y Ubicación
- Memoria de Corto Plazo
  - El cerebro procesa de forma “atentiva”
  - Selecciona partes específicas de la información
  - Tiene una capacidad limitada
  - Ordena la información en bloques (o *chunks*)
- Memoria de Largo Plazo
  - Principalmente se llega por procesos de repetición
  - Tiene capacidad prácticamente ilimitada



<https://amzn.to/2Y41yif>

# Percepción Visual y Cognición

Tres tipos de memoria:

- Memoria Icónica
  - Fase “preatentiva” automática
  - Prácticamente inmediata
  - Detección de color y Ubicación
- Memoria de Corto Plazo
  - El cerebro procesa de forma “atentiva”
  - Selecciona partes específicas de la información
  - Tiene una capacidad limitada
  - Ordena la información en bloques (o *chunks*)
- Memoria de Largo Plazo
  - Principalmente se llega por procesos de repetición
  - Tiene capacidad prácticamente ilimitada

# Percepción Visual y Cognición

## Atributos Preatentivos

- Forma
- Orientación
- Tamaño
- Color
- Longitud
- Ancho



Para construir visualizaciones de **calidad** que logren que

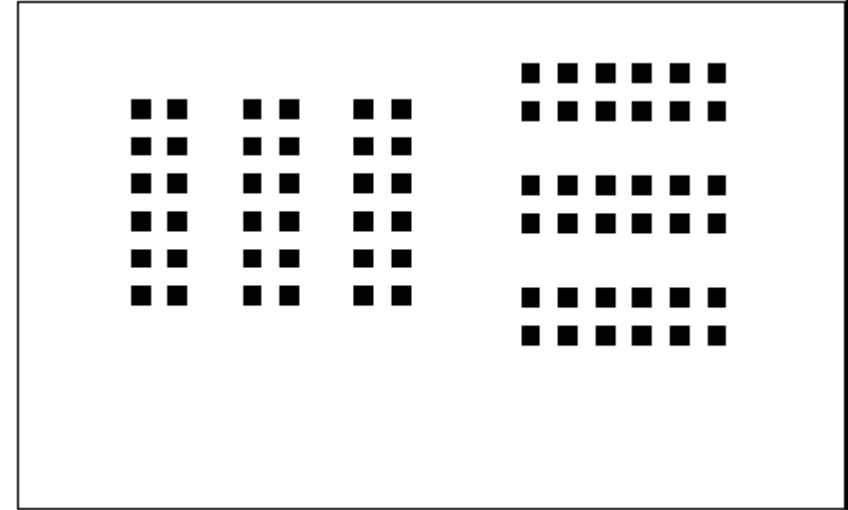
- Los datos sean fácilmente **interpretables**
- Sean **creíbles**
- Enfoquen a la audiencia en los puntos **más relevantes**

# Percepción Visual y Cognición

## Principios de Gestalt

- Principio de Proximidad

Percibimos los objetos que están **cerca entre si** como si fueran un **grupo**



# Percepción Visual y Cognición

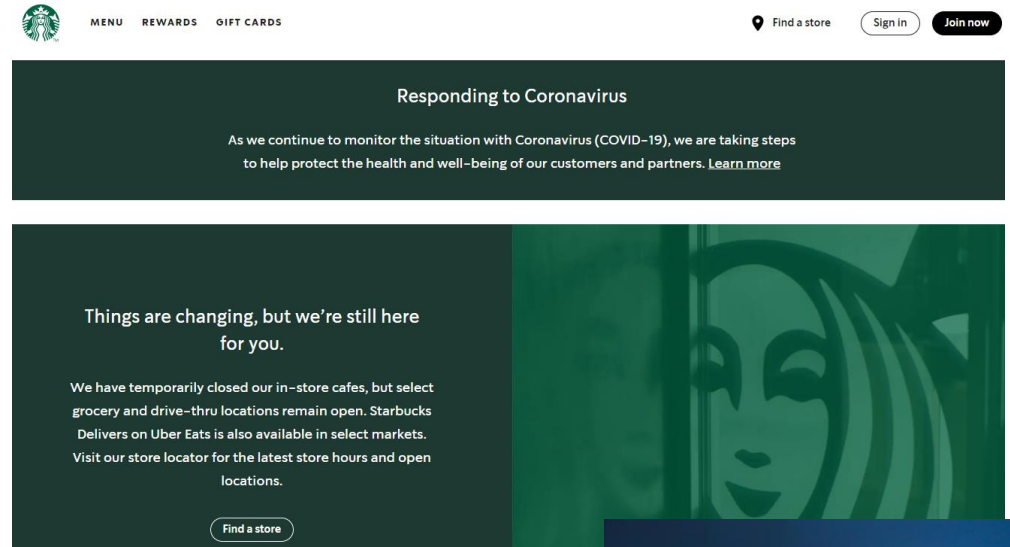
## Principios de Gestalt

- Principio de Proximidad
- Principio de Similitud

Tendemos a agrupar objetos con las mismas características

**“Los pájaros vuelan con los de su mismo tipo”**

Por esta razón, en mercadotecnia, se escogen ciertos colores, patrones, tipografía y demás elementos gráficos para crear una identidad de marca

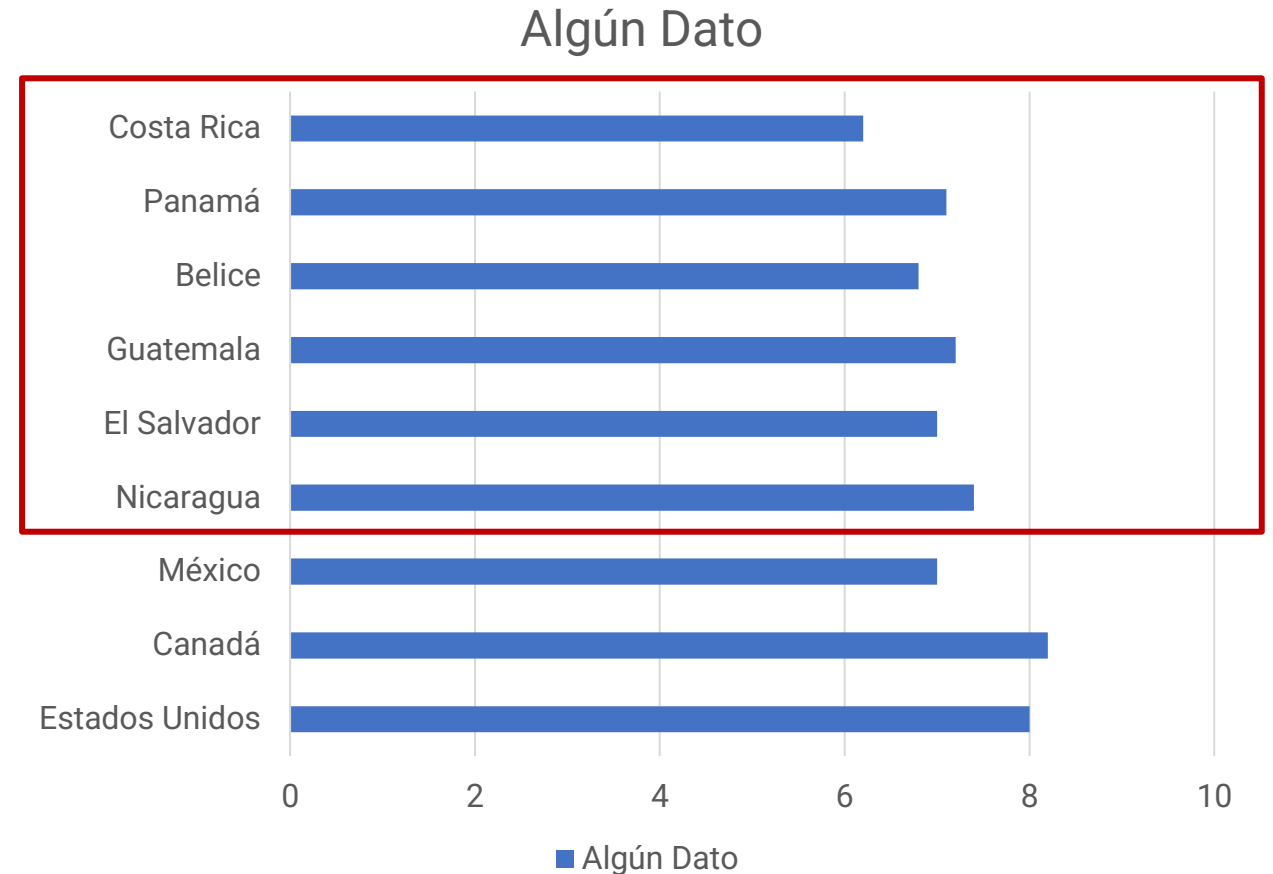


# Percepción Visual y Cognición

## Principios de Gestalt

- Principio de Proximidad
- Principio de Similitud
- Principio de Cercado

Cuando encontramos elementos que se encuentran encerrados en lo que parece ser un borde, percibimos que pertenecen a un grupo





# Percepción Visual y Cognición

## Principios de Gestalt

- Principio de Proximidad
- Principio de Similitud
- Principio de Cercado
- Principio de Continuidad

Se percibe a un objeto como parte de un todo

En mercadotecnia se usa en los espectaculares, que muestran solo una parte parcial de la imagen, dejando a la imaginación el que el logo es mucho más grande el espacio que lo contiene



# Estadios de la Visualización

1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

**¿Será que el promedio de temperatura de un país *determine* su Producto Interno Bruto?**



# Estadios de la Visualización

1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

**¿Será que el promedio de temperatura de un país *determine* su Producto Interno Bruto?**

## Fuentes

- World Bank GDP ranking  
<https://www.kaggle.com/theworldbank/world-bank-gdp-ranking>
- Climate Change: Earth Surface Temperature Data  
<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>

# Investigación y Selección de Datos

- World Bank GDP ranking

<https://www.kaggle.com/theworldbank/world-bank-gdp-ranking>

	A	B	C	D	E
1	Gross domestic product 2017				
2					
3					(millions of
4		Ranking	Economy		US dollars)
5					
6	USA	1	United States		19,485,394
7	CHN	2	China		12,237,700
8	JPN	3	Japan		4,872,415
9	DEU	4	Germany		3,693,204
10	IND	5	India		2,650,725
11	GBR	6	United Kingdom		2,637,866
12	FRA	7	France		2,582,501
13	BRA	8	Brazil		2,053,595
14	ITA	9	Italy		1,943,835
15	CAN	10	Canada		1,647,120
16	RUS	11	Russian Federation		1,578,417 a
17	KOR	12	Korea, Rep.		1,530,751
18	AUS	13	Australia		1,323,421
19	ESP	14	Spain		1,314,314
20	MEX	15	Mexico		1,150,888
21	IDN	16	Indonesia		1,015,421
22	TUR	17	Turkey		851,549
23	NLD	18	Netherlands		830,573
24	SAU	19	Saudi Arabia		686,738
25	CHE	20	Switzerland		678,965
26	ARG	21	Argentina		637,430 b
27	SWE	22	Sweden		535,607
28	POL	23	Poland		526,466
29	BEL	24	Belgium		494,764

- Climate Change: Earth Surface Temperature Data

<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>

	A	B	C	D
1	dt	AverageTemperature	AverageTemperatureUncertainty	Country
2	1743-11-01	4.384	2.294	Å...land
3	1743-12-01			Å...land
4	1744-01-01			Å...land
5	1744-02-01			Å...land
6	1744-03-01			Å...land
7	1744-04-01	1.53	4.68	Å...land
8	1744-05-01	6.702	1.789	Å...land
9	1744-06-01	11.609	1.577	Å...land
10	1744-07-01	15.342	1.41	Å...land
11	1744-08-01			Å...land
12	1744-09-01	11.702	1.517	Å...land
13	1744-10-01	5.477	1.862	Å...land
14	1744-11-01	3.407	1.425	Å...land
15	1744-12-01	-2.181	1.641	Å...land
16	1745-01-01	-3.85	1.841	Å...land
17	1745-02-01	-6.575	1.36	Å...land
18	1745-03-01	-4.195	1.213	Å...land
19	1745-04-01	-0.966	1.172	Å...land
20	1745-05-01			Å...land
21	1745-06-01			Å...land
22	1745-07-01			Å...land
23	1745-08-01			Å...land
24	1745-09-01			Å...land
25	1745-10-01			Å...land
26	1745-11-01			Å...land
27	1745-12-01			Å...land
28	1746-01-01			Å...land
29	1746-02-01			Å...land
30	1746-03-01			Å...land

# Limpieza de Datos

Cuestionario No Oficial a Estudiantes (Responses) ☆ 📄  
File Edit View Insert Format Data Tools Form Add-ons Help Last edit was on March 31

100% \$ % .0 .00 123 Default (Arial) 10 B I U A

Timestamp	Nombre (sin apellidos, por favor)	¿En qué país vives?	¿En dónde te encuentras?	Experiencia en Visualización de Datos	Años de educación	¿Trabajas Actualmente?	¿Cuál es tu salario anual?	¿Cuál es tu interés principal?
11/17/2019 17:40:13	Frank	México	Torres de Satélite, Naulco	Experto		20 Empleo de Tiempo Completo	75000	
11/18/2019 18:42:19	Beatriz	México		Nula		6 Múltiples Empleos		Aprender a hacer
11/18/2019 18:47:54	Adriana	Canada	West Edmonton Mall, Edm	Regular		12 Trabajo por mi Cuenta (freelance)	68000	Aprender nue
11/18/2019 18:48:28	Salvador	México	Estadio Azteca, México	Regular		12 Trabajo por mi Cuenta (freelance)	77000	Curriculum
11/18/2019 18:49:11	Rocio	México	Desierto de los Leones, C	Buena		20 Empleo de Tiempo Completo	80000	Curiosidad
11/18/2019 18:49:56	Juan José	Guatemala		Básica		16 Empleo de Tiempo Completo	65000	Oportunidade
11/18/2019 18:50:37	Zeus	México	Angel de la Independenci	Básica		14 Empleo de Tiempo Completo	65000	Aprender nue
11/18/2019 18:51:45	Sergio	Canada	Parliament Hill, Ontario, C	Regular		18 Empleo de Tiempo Completo	102000	Preparar mejo
11/18/2019 18:52:48	Norma	Estados Unidos	Statue of Liberty, New York	Buena		20 Trabajo por mi Cuenta (freelance)	88000	Crear present
11/18/2019 18:54:03	César	USA	Disneyworld, Los Angeles	Básica		18 Medio Tiempo	25000	preparar infor
11/18/2019 18:54:43	Jesus	Israel	Los altos del Golán	Experto		30 Empleo de Tiempo Completo	1000000	conocer nuev
11/18/2019 18:55:34	Oscar	México	Torres de Satélite, Naulco	Básica		18 Medio Tiempo	38000	desarrollar ha
11/18/2019 18:56:07	Evelyn	Canada				10		conocer gente
11/18/2019 18:57:13	Carolina	EEUU	Cataratas del Niágara, Bu	Regular				Diagramar me
11/19/2019 11:02:15	Christian	México	Ciudad de México	Regular		20 Empleo de Tiempo Completo	60	aprender más
11/19/2019 11:32:52	Miguel Galvan	Mexico	Durango, Victoria de Dur	Básica		Empleo de Tiempo Completo	30000	Ampliar nuevo
11/19/2019 11:53:58	Unknown	Mexico	Tijuana	Regular		19 Empleo de Tiempo Completo	67200	Conocer softw
11/19/2019 12:04:26	Jesus	México	Guanajuato	Buena		3 Empleo de Tiempo Completo	40000	Conocer las r
11/19/2019 12:07:07	AMLO	México	CDMX	Nula		0 Empleo de Tiempo Completo	80000	Tener los dato
11/19/2019 12:57:58	Thom Yorke	México	CDMX	Básica		7 Empleo de Tiempo Completo	20000	Aprender a co
11/19/2019 19:11:03	mdrfr	Mexico	19.380647, -99.156002	Básica		25 Trabajo por mi Cuenta (freelance)	60	beneficio dire
11/19/2019 21:09:02	Mopez	México	Comalcalco, Tabasco	Regular		20 Múltiples Empleos	400000	Saber interpre
11/19/2019 21:48:10	Adrian	México	Durango	Regular		7 Empleo de Tiempo Completo	4000	Aprender a m

Vamos a revisar los datos que capturamos en la encuesta de Google (y los que generaron nuestros compañeros)

Para poder usar nuestros datos, es necesario limpiarlos, esto incluye, entre otras cosas:

- Eliminar datos duplicados
- Homologar formatos, unidades y nombres (unidades europeas vs. Americanas: tiempo, distancia, volumen, temperatura, etc.)
- Decidir qué hacer con los datos faltantes y erróneos ¿eliminarlos? ¿imputarlos?
- Adecuarlos para poder juntarlos con otros datos
- Tratar datos fuera de rango (outliers)
- Hacer ingeniería de características (*feature engineering*)



# Limpieza de Datos

	A	B	C	D	E
1		Gross domestic product 2017			
2					
3					(millions of
4		Ranking	Economy		US dollars)
5					
6	USA	1	United States	19,485,394	
7	CHN	2	China	12,237,700	
8	JPN	3	Japan	4,872,415	
9	DEU	4	Germany	3,693,204	
10	IND	5	India	2,650,725	
11	GBR	6	United Kingdom	2,637,866	
12	FRA	7	France	2,582,501	
13	BRA	8	Brazil	2,053,595	
14	ITA	9	Italy	1,943,835	
15	CAN	10	Canada	1,647,120	
16	RUS	11	Russian Federation	1,578,417 a	
17	KOR	12	Korea, Rep.	1,530,751	
18	AUS	13	Australia	1,323,421	
19	ESP	14	Spain	1,314,314	
20	MEX	15	Mexico	1,150,888	
21	IDN	16	Indonesia	1,015,421	
22	TUR	17	Turkey	851,549	
23	NLD	18	Netherlands	830,573	
24	SAU	19	Saudi Arabia	686,738	
25	CHE	20	Switzerland	678,965	
26	ARG	21	Argentina	637,430 b	
27	SWE	22	Sweden	535,607	
28	POL	23	Poland	526,466	
29	BEL	24	Belgium	494,764	

World Bank GDP ranking

	A	B	C	D
1	dt	AverageTemperature	AverageTemperatureUncertainty	Country
2	1743-11-01	4.384	2.294	Å...land
3	1743-12-01			Å...land
4	1744-01-01			Å...land
5	1744-02-01			Å...land
6	1744-03-01			Å...land
7	1744-04-01	1.53	4.68	Å...land
8	1744-05-01	6.702	1.789	Å...land
9	1744-06-01	11.609	1.577	Å...land
10	1744-07-01	15.342	1.41	Å...land
11	1744-08-01			Å...land
12	1744-09-01	11.702	1.517	Å...land
13	1744-10-01	5.477	1.862	Å...land
14	1744-11-01	3.407	1.425	Å...land
15	1744-12-01	-2.181	1.641	Å...land
16	1745-01-01	-3.85	1.841	Å...land
17	1745-02-01	-6.575	1.36	Å...land
18	1745-03-01	-4.195	1.213	Å...land
19	1745-04-01	-0.966	1.172	Å...land
20	1745-05-01			Å...land
21	1745-06-01			Å...land
22	1745-07-01			Å...land
23	1745-08-01			Å...land
24	1745-09-01			Å...land
25	1745-10-01			Å...land
26	1745-11-01			Å...land
27	1745-12-01			Å...land
28	1746-01-01			Å...land
29	1746-02-01			Å...land
30	1746-03-01			Å...land

Climate Change: Earth Surface Temperature Data

Un problema que existe al utilizar herramientas como Google Drive o Microsoft Excel es que es muy difícil implementar **repetibilidad** en el proceso de limpieza y extracción de datos

Para poder usar nuestros datos, es necesario limpiarlos, esto incluye, entre otras cosas:

- Eliminar datos duplicados
- Homologar formatos, unidades y nombres (unidades europeas vs. Americanas: tiempo, distancia, volumen, temperatura, etc.)
- Decidir qué hacer con los datos faltantes y erróneos ¿eliminarlos? ¿imputarlos?
- Adecuarlos para poder juntarlos con otros datos
- Tratar datos fuera de rango (outliers)
- Hacer ingeniería de características (*feature engineering*)



# Recapitulando

## Hoy vimos

- Investigación y Selección de Datos
- Limpieza de Datos

## Actividades

- Descargar la tabla de Google Drive y limpiarla (nombres de países, eliminar o completar datos faltantes, etc.) La tabla se encuentra en: <http://bit.ly/vdd-data>
- Descargar e instalar Anaconda <https://www.anaconda.com> para tu sistema operativo
- Bajar el notebook de limpieza de datos y generar el archivo .csv