



Cenicafé

Centro Nacional de Investigaciones de Café

Ciencia y Pasión en cada grano



Cenicafé
Centro Nacional de Investigaciones de Café



Presentación de resultados en investigación agrícola:
entre preferencias, guías y criterios editoriales

Beatriz Eugenia Mira Rada - Biometría

La comunicación científica



Conferencia General
30ª reunión, París 1999

30 C/15
18 de agosto de 1999
Original: Inglés

Punto 4.6 del orden del día provisional

**DECLARACION SOBRE LA CIENCIA Y EL USO DEL SABER CIENTIFICO
Y PROGRAMA EN PRO DE LA CIENCIA:
MARCO GENERAL DE ACCION**



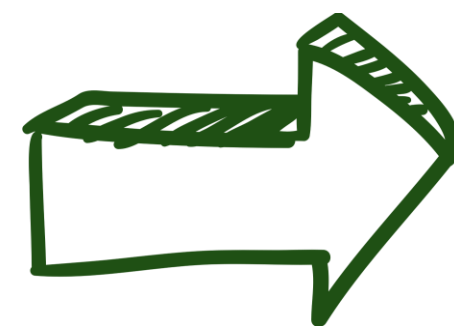
La comunicación de la ciencia es un proceso **cultural, diverso y colectivo**, cuya forma y contenido condicionan directamente la calidad de la democracia.

El método científico y la investigación en cenicafé



El método científico es una serie de pasos para la **investigación científica**

- ✓ Conocimiento verificado y viable
- ✓ Conocimiento replicable
- ✓ Base para generar mas conocimiento



Caficultor

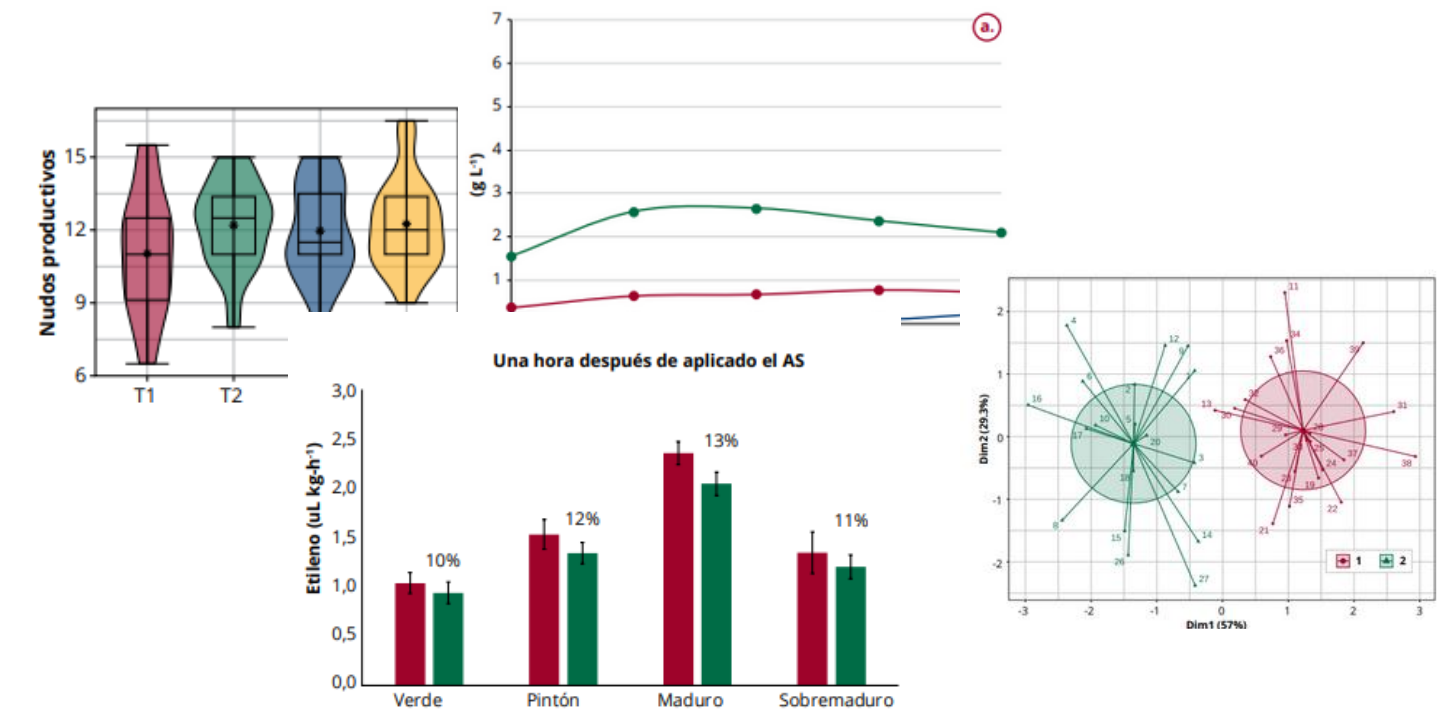


La presentación de información estadística tiene tres opciones de formato:



Formato gráfico

Consiste en la utilización de puntos, líneas y figuras que sirven para mostrar magnitudes, asociadas a una escala de medición, de manera que se facilita la comparación e interpretación de los datos estadísticos, sin que necesariamente se incluyan los valores numéricos.

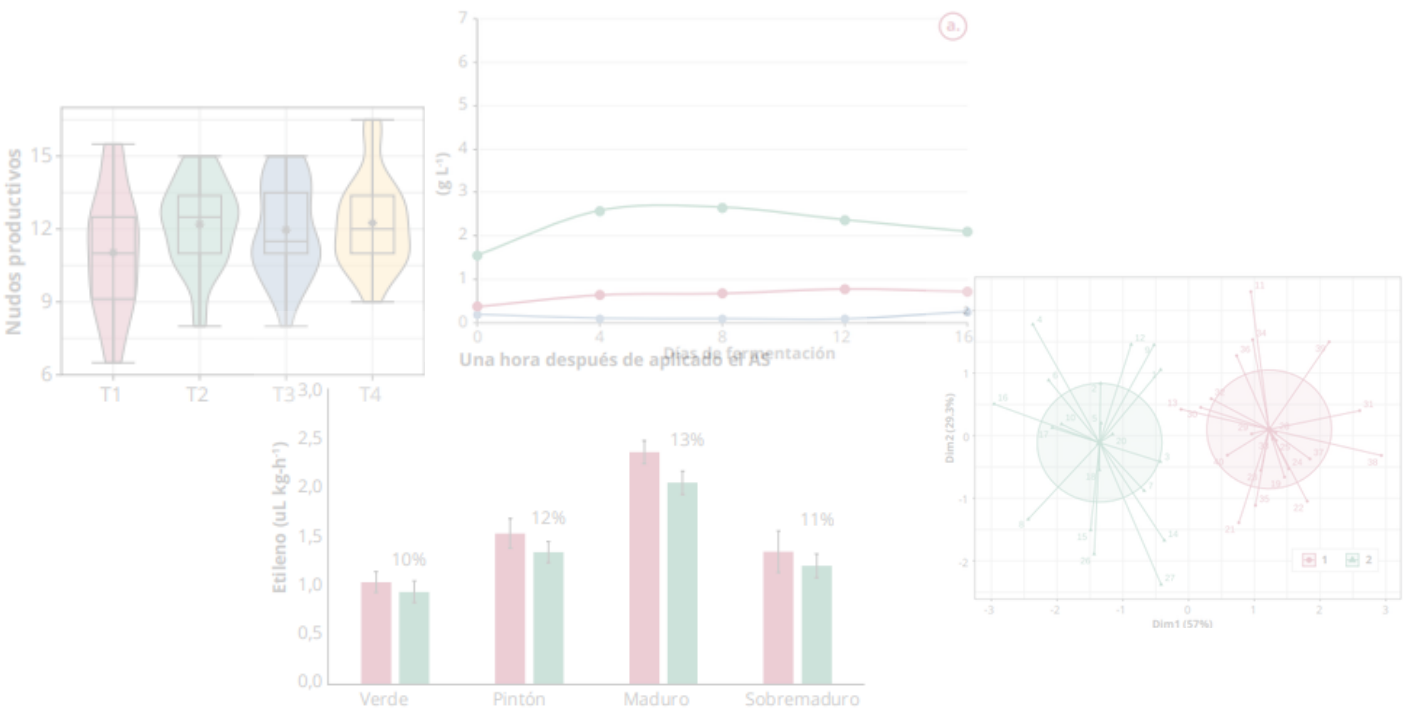


La presentación de información estadística tiene tres opciones de formato:



Formato gráfico

Consiste en la utilización de puntos, líneas y figuras que sirven para mostrar magnitudes, asociadas a una escala de medición, de manera que se facilita la comparación e interpretación de los datos estadísticos, sin que necesariamente se incluyan los valores numéricos.



Formato tabular

Consiste en el ordenamiento de los datos seleccionados, en filas y columnas, denominado cuadro estadístico.

Compuesto		Variable	Mínimo	Máximo	Media	Desv. típica	Valor-p	Error relativo de predicción
Lípidos	Lab		9,20	13,39	11,04	0,83	0,26	0,0100
	NIRS		9,14	12,96	10,93	0,70		
					5,32	0,28	0,51	0,0040
					5,30	0,22		
					7,85	0,77		
					7,87	0,45	0,85	0,0020
					1,15	0,14		
+ 0 horas					1,16	0,46	0,0105	
					0,86			
					0,88			
					37,66			
+ 4 horas					37,74			
					40,60			
+ 8 horas					40,62			
+ 12 horas								
Oleico	Lab		9,01	14,87	11,85			
	NIRS		10,01	13,58	11,89			
Esteárico	Lab		6,59	9,83	7,81			
	NIRS		6,46	9,11	7,77			
Araquídico	Lab		1,01	3,22	2,09			
	NIRS		0,90	2,97	2,07			

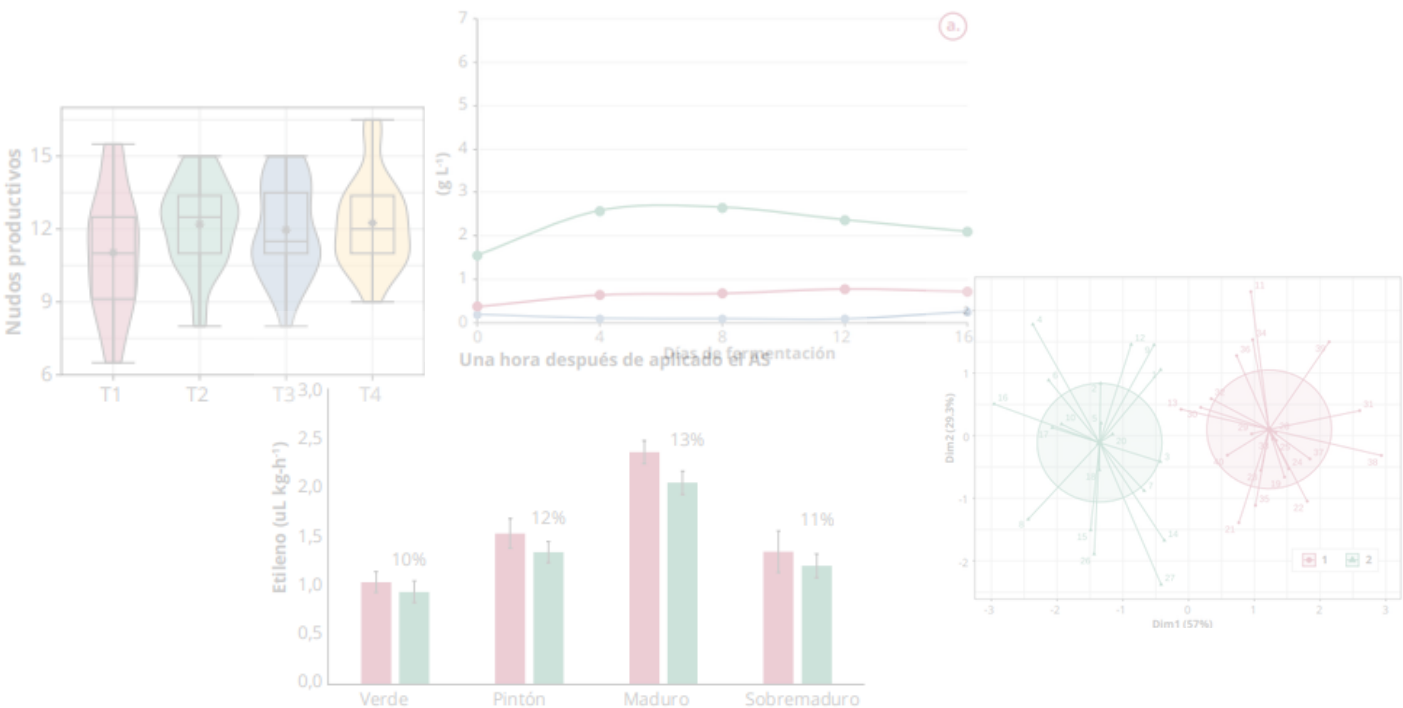
Estación Experimental El Rosario, Promedio Parcela 1, 2, 3. Etapa de levante							
Variable	Promedio	Máximo	Mínimo	Datos	DE	CV (%)	IC
Parámetros suelo a 15 y 45 cm de profundidad							
Humedad a 15 cm (m³ m⁻³) sin corregir	0,351	0,420	0,296	57	0,02	6,46	0,01
Humedad a 15 cm (m³ m⁻³) corregida	0,374	0,383	0,346	57	0,01	2,20	0,00
Temperatura a 15 cm (°C)	22,32	24,96	20,44	57	1,09	4,90	0,28
Conductividad a 15 cm (mS cm⁻¹)	0,096	0,213	0,050	57	0,05	48,56	0,01
Humedad a 45 cm (m³ m⁻³) sin corregir	0,412	0,461	0,352	57	0,03	6,53	0,01

La presentación de información estadística tiene tres opciones de formato:



Formato gráfico

Consiste en la utilización de puntos, líneas y figuras que sirven para mostrar magnitudes, asociadas a una escala de medición, de manera que se facilita la comparación e interpretación de los datos estadísticos, sin que necesariamente se incluyan los valores numéricos.



Formato tabular

Consiste en el ordenamiento de los datos seleccionados, en filas y columnas, denominado cuadro estadístico.

		Compuesto						Valor-p	Error relativo de predicción							
		Variable	Mínimo	Máximo	Media	Desv. típica										
Tratamiento	Mínimo	Lípidos	Lab	9,20	13,39	11,04	0,83	0,26	0,0100							
			NIRS	9,14	12,96	10,93	0,70									
		CQAT	Lab	4,52	5,86	5,32	0,28	0,51	0,0040							
			NIRS	4,77	5,81	5,30	0,22									
		Promedio	Lab	5,99	9,66	7,85	0,77	0,85	0,0020							
			NIRS	6,48	8,96	7,87	0,45									
		+ 0 horas	80,75	83,56	86,25	86,25	1,15	0,14	0,46	0,0105						
		Cafeína	Lab	0,66	1,55	1,15	0,14									
			Trigonelina	NIRS	0,91	1,44	1,16	0,11	Estación Experimental El Rosario. Promedio Parcela 1, 2, 3. Etapa de levante							
				Lab	0,32	1,14	0,86	0,13	0,10	0,0251	Promedio	Máximo	Mínimo	Datos	DE	CV (%)
NIRS	0,74		1,05	0,88	0,07	Variable										
+ 8 horas	81,25	83,67	86,25	86,25	40,62	37,66	1,12	Parámetros suelo a 15 y 45 cm de profundidad								
		Palmitico	NIRS	35,59	40,28	37,74	0,91	0,54	0,0006	0,351	0,420	0,296	57	0,02	6,46	0,01
			Lab	35,59	40,28	37,74	1,32									
		Linoleico	NIRS	38,28	42,96	40,62	0,96	0,59	0,0032	0,374	0,383	0,346	57	0,01	2,20	0,00
			Lab	9,01	14,87	11,85	1,18									
		Oleico	NIRS	10,01	13,58	11,89	0,85	0,59	0,0055	22,32	24,96	20,44	57	1,09	4,90	0,28
			Lab	6,59	9,83	7,81	0,67									
		Estearico	NIRS	6,46	9,11	7,77	0,57	0,59	0,0096	0,213	0,050	57	0,05	48,56	0,01	
			Lab	1,01	3,22	2,09	0,51									
		Araquidico	NIRS	0,90	2,97	2,07	0,41	0,82	0,0072	0,412	0,461	0,352	57	0,03	6,53	0,01
			Lab	0,90	2,97	2,07	0,41									



Texto

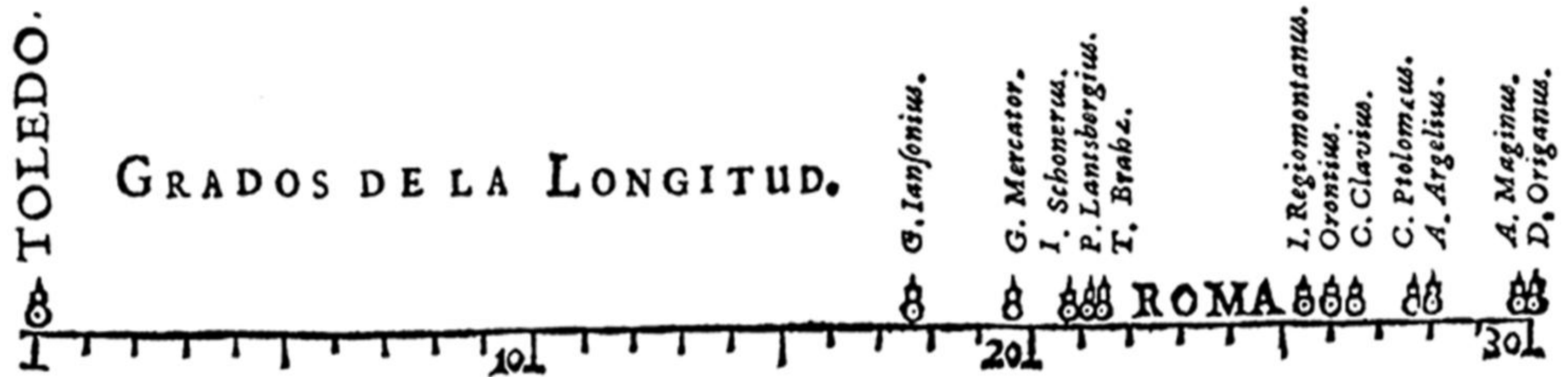
Enfatizar en algunos resultados, dar explicación adicional que contextualiza en casos particulares

Historia de la representación de datos estadísticos



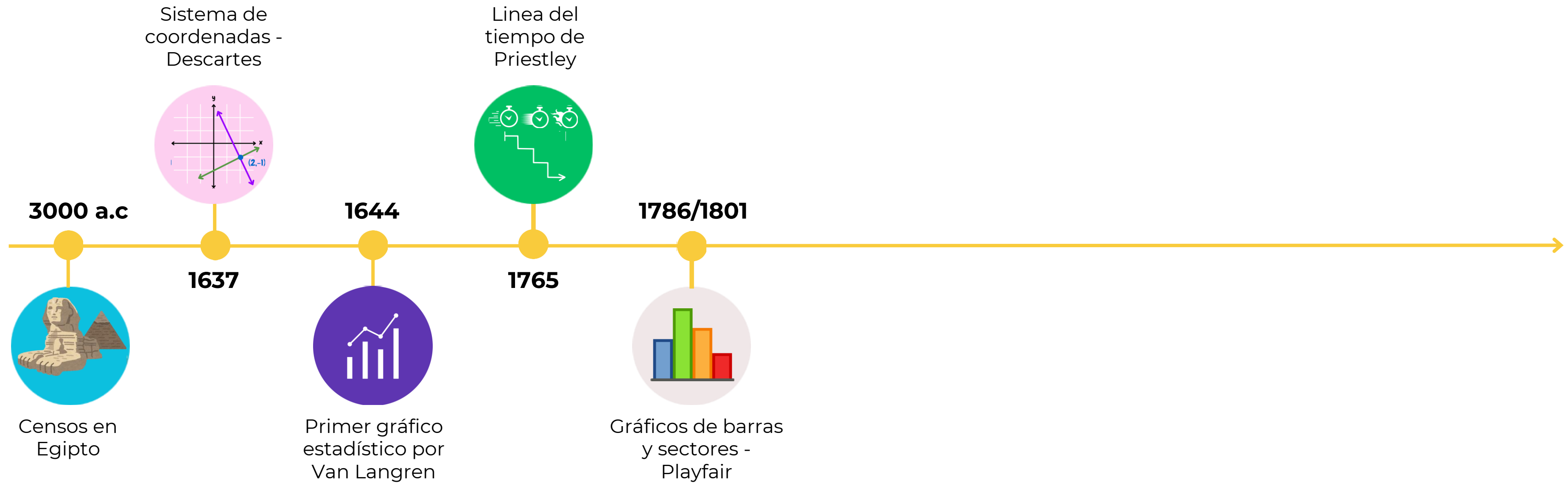


Variaciones en la determinación de la longitud entre la ciudad de Toledo (España) y Roma (Italia)

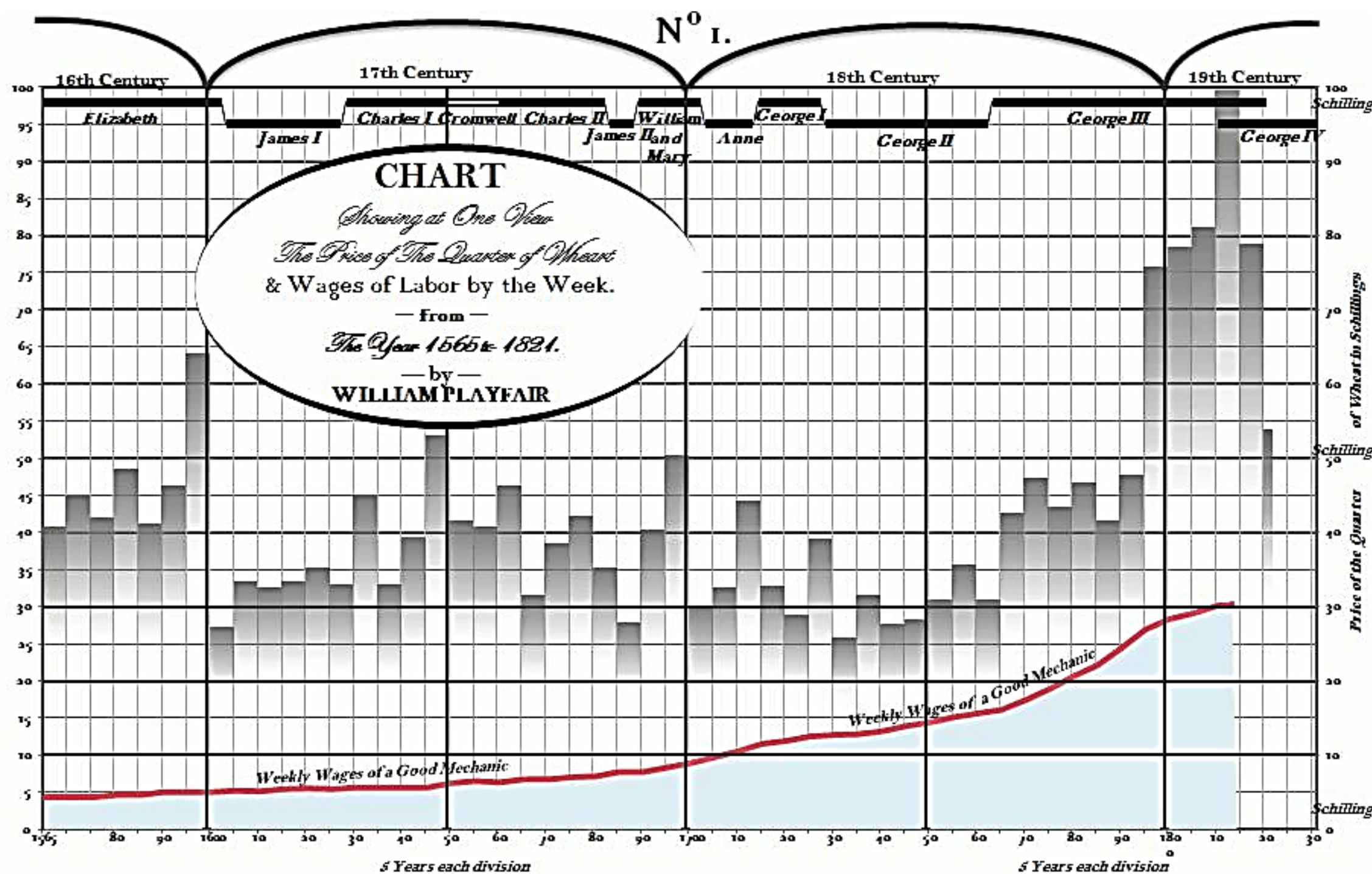


Michael F. Van Langren

Historia de la representación de datos estadísticos



William Playfair pionero de la estadística gráfica

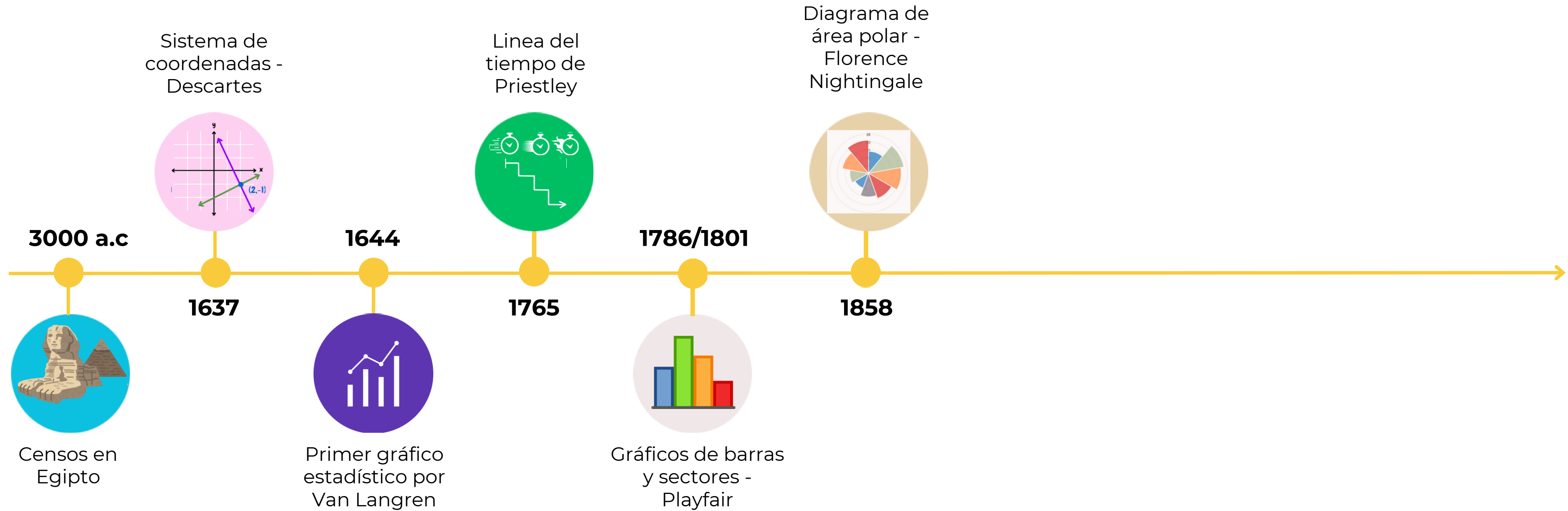


Principios de Playfair:

- 1 El método gráfico es una forma de simplificar lo tedioso y lo complejo.
- 2 Los hombres ocupados necesitan alguna clase de ayuda visual.
- 3 Un gráfico es más accesible que una tabla.
- 4 El método gráfico es concordante con los ojos.
- 5 El método gráfico ayuda al cerebro, ya que permite entender y memorizar mejor.



Historia de la representación de datos estadísticos



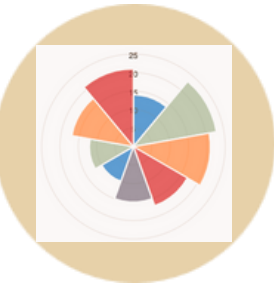
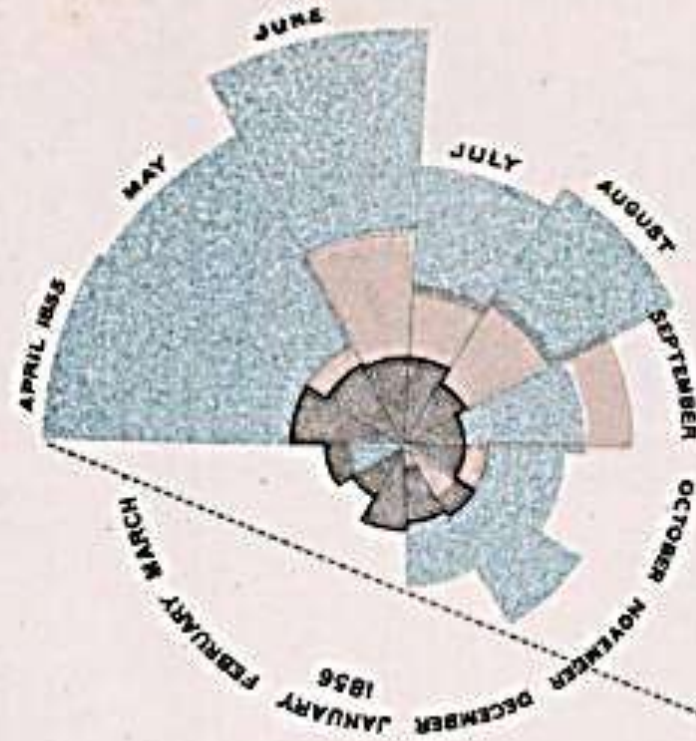
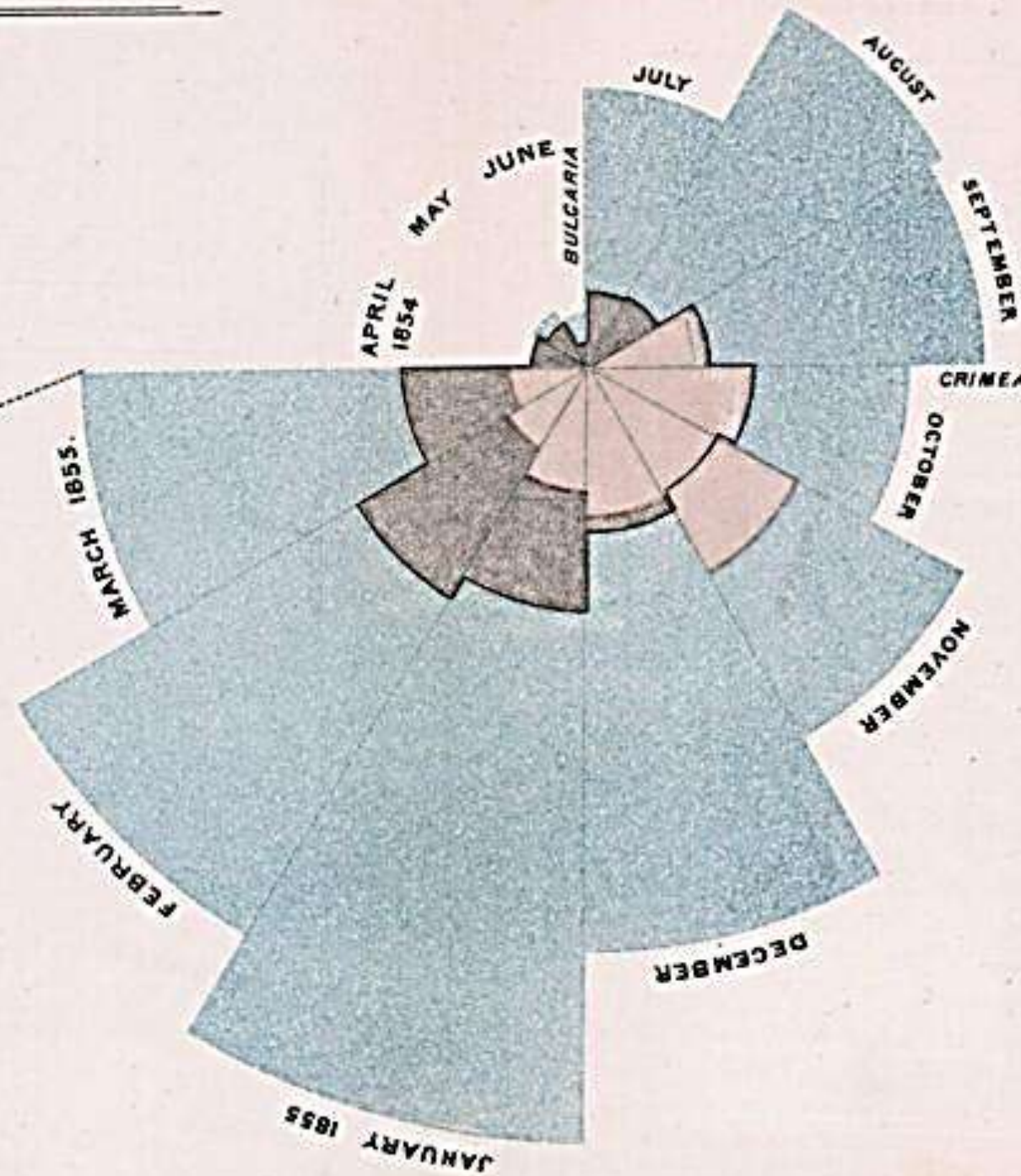


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



1.
APRIL 1854 to MARCH 1855.

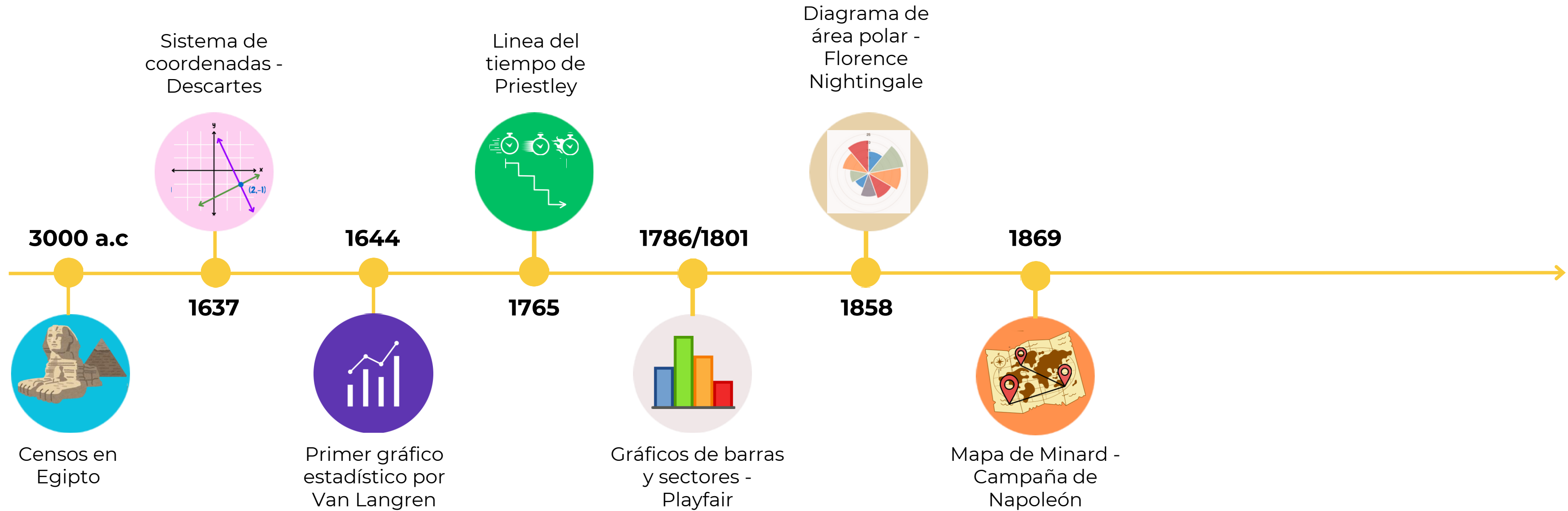


The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & 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.

Causas de muerte

- Enfermedades infecciosas
- Heridas
- Otras causas

Historia de la representación de datos estadísticos





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.

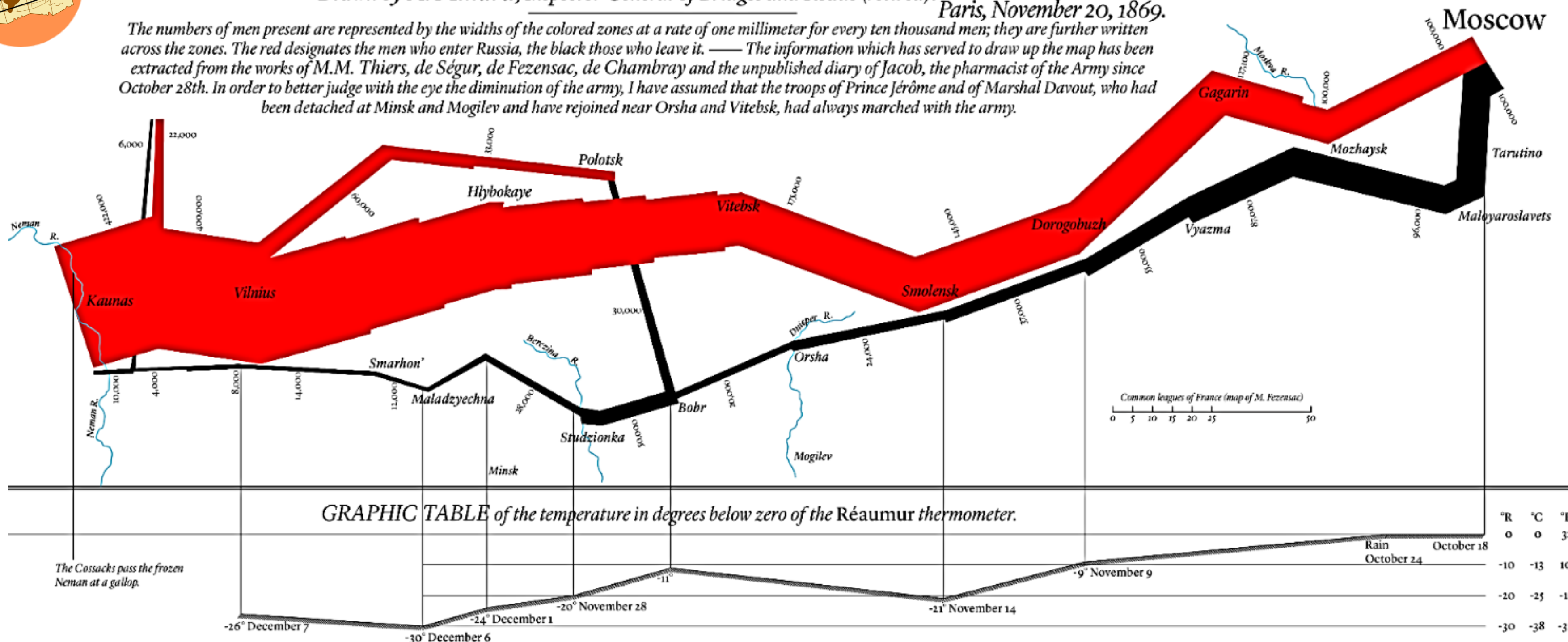
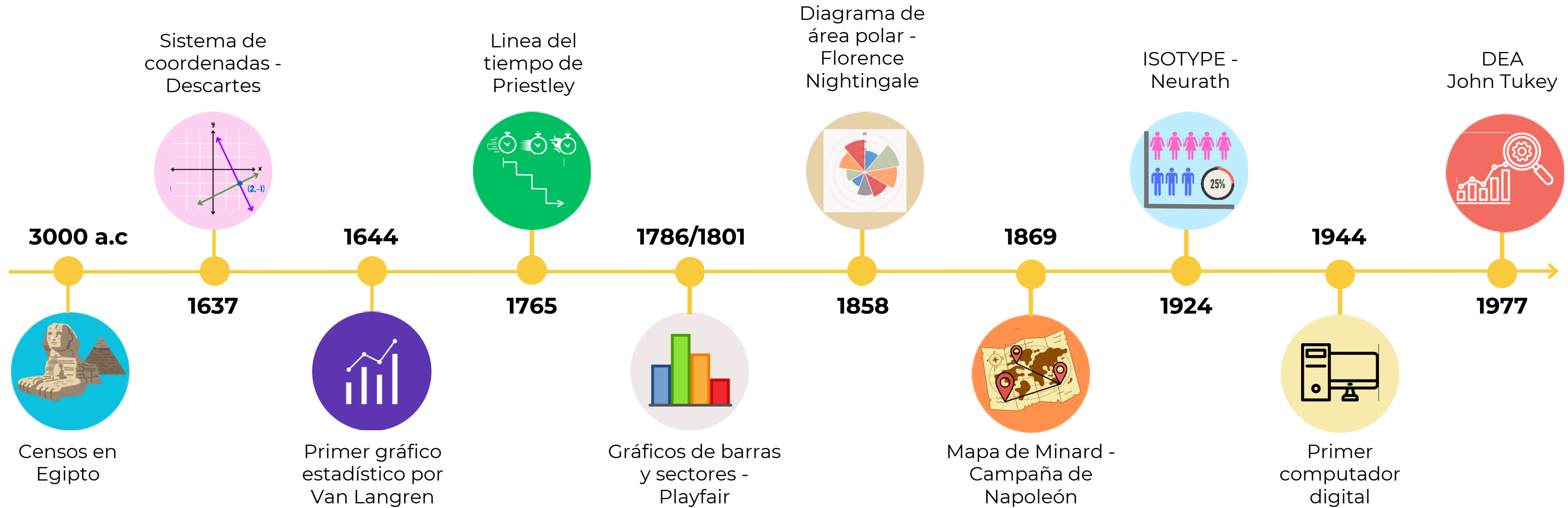


Gráfico de M. Minard sobre el Ejército de Napoleón en Rusia

Historia de la representación de datos estadísticos



Historia de la representación de datos estadísticos



The greatest value of a picture is
when it forces us to notice what we
never expected to see.

— John Tukey —

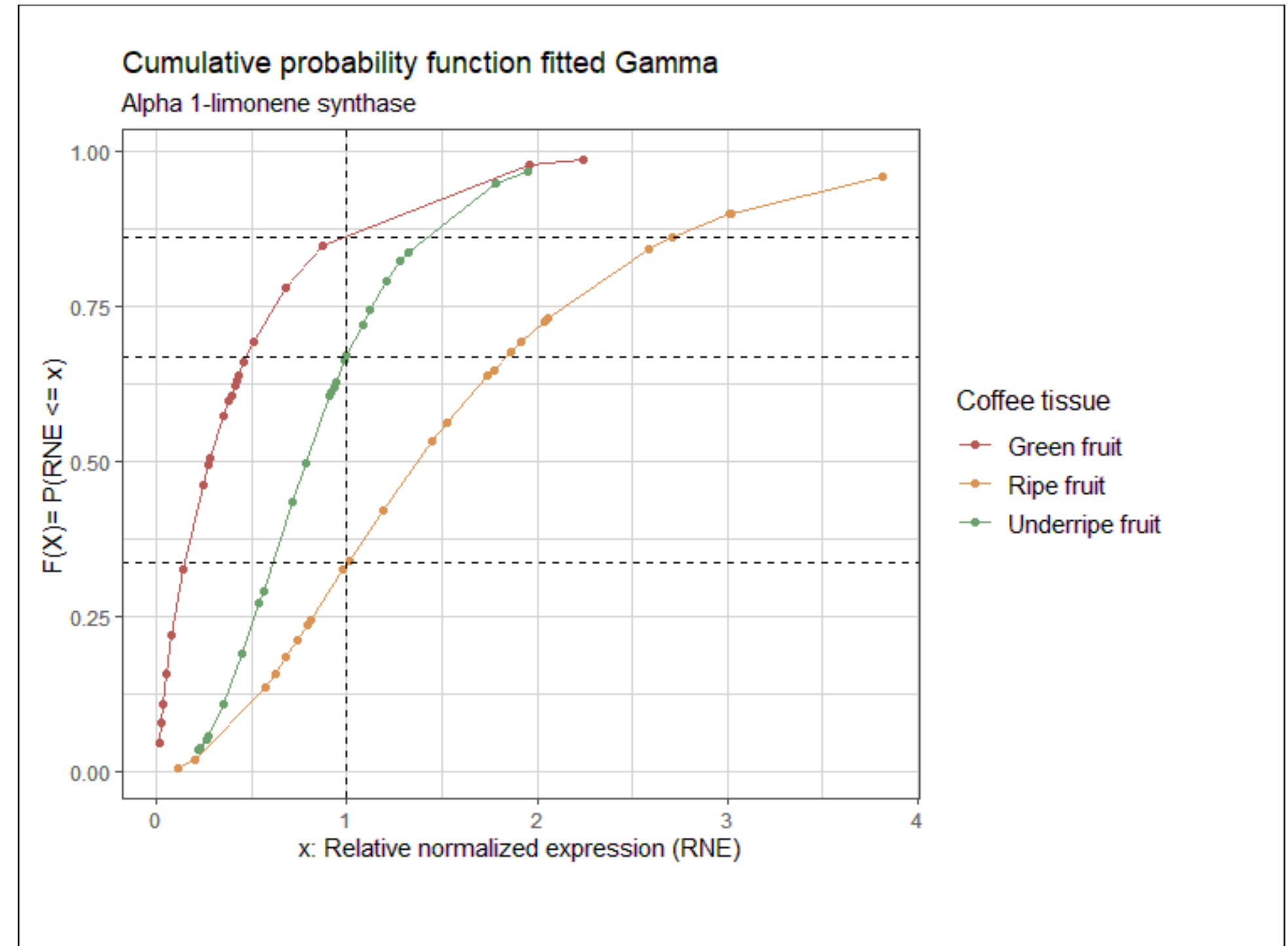
Graficar o no graficar, esa es la cuestión

La decisión de representar los datos de manera gráfica debería basarse en el propósito

Tipo de gráfico	Propósito
Propaganda	Mostrar lo aprendido por otras técnicas.
Analíticos	Qué puede estar ocurriendo.
Sustitutos de tablas	Lectura mas compleja, valores, escalas, etc.
Decorativos	Son bonitos, lucen en la presentación de la información.

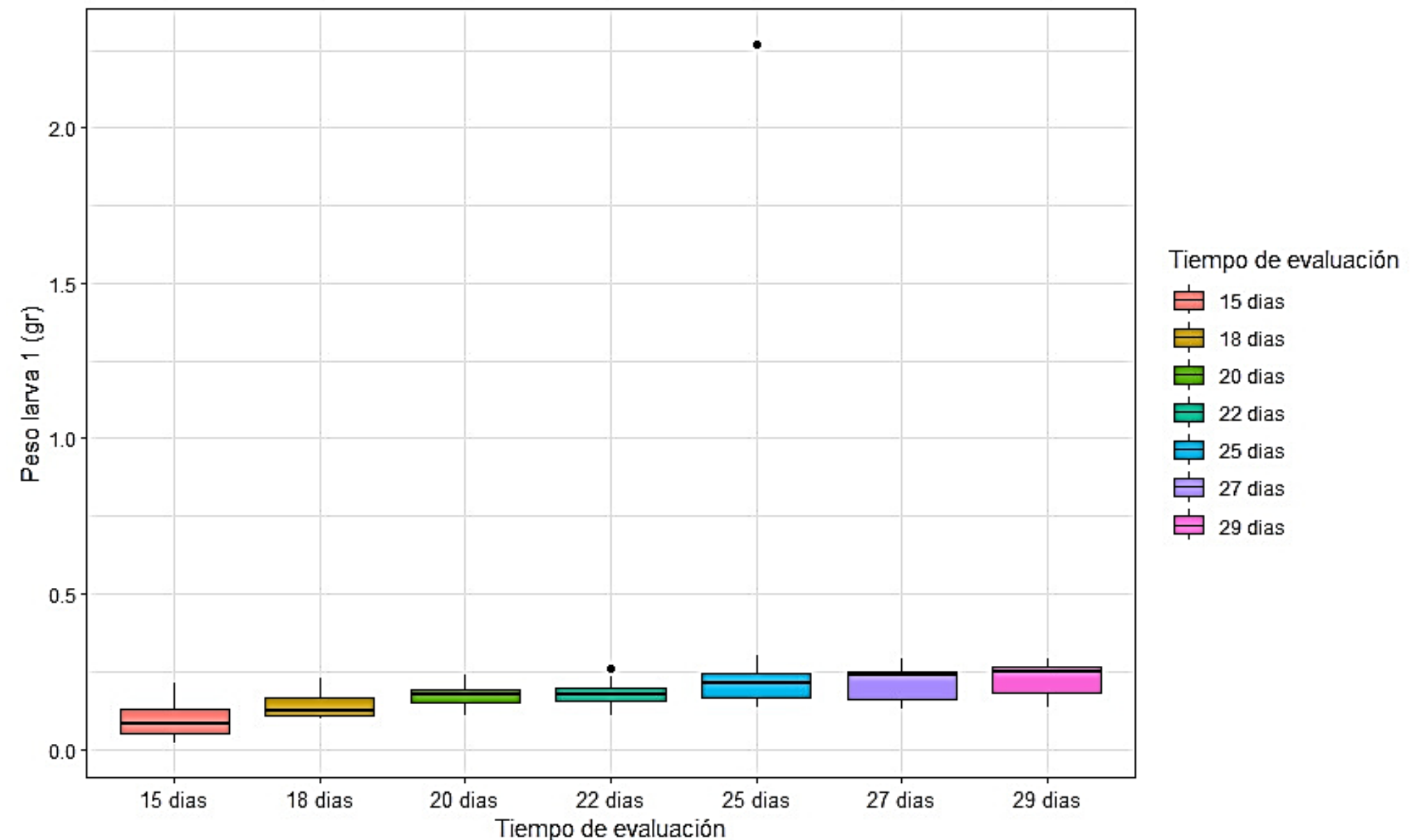
Lineamientos para la elaboración de un gráfico estadístico

- ✓ • Título principal
- ✓ • Título secundario o subtítulo
- ✓ • Descripción del gráfico
- ✓✓ • Región de datos y símbolos
- ✓✓ • Eje horizontal y escala
- ✓✓ • Eje vertical y escala
- ✓✓ • Apuntadores
- ✓✓ • Descriptores de señales y marcadores



Lineamientos para la elaboración de un gráfico estadístico

- ✓ • Título principal
- ✓ • Título secundario o subtítulo
- ✓ • Descripción del gráfico
- ✓✓ • Región de datos y símbolos
- ✓✓ • Eje horizontal y escala
- ✓✓ • Eje vertical y escala
- ✓✓ • Apuntadores
- ✓✓ • Descriptores de señales y marcadores



Lineamientos para la elaboración de un gráfico estadístico

- ✓ • Título principal
- ✓ • Título secundario o subtítulo
- ✓ • Descripción del gráfico
- ✓✓ • Región de datos y símbolos
- ✓✓ • Eje horizontal y escala
- ✓✓ • Eje vertical y escala
- ✓✓ • Apuntadores
- ✓✓ • Descriptores de señales y marcadores

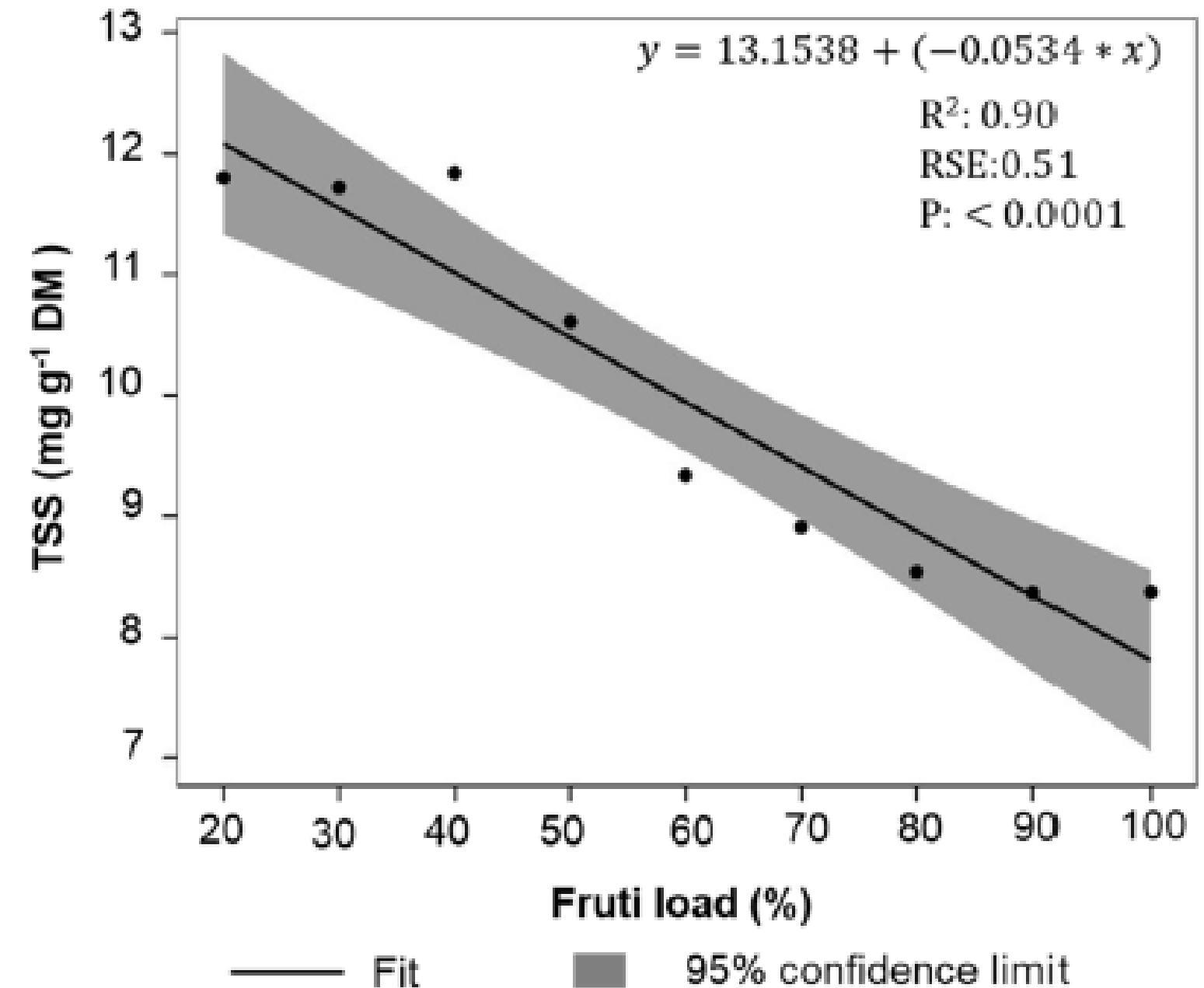


Fig. 5. Fit of linear model between the concentrations of total soluble sugars-TSS and fruit load measured in fully developed leaves. Values are reported as means ($n = 4$). RSE corresponds to residual standard error.

Lineamientos para la elaboración de un gráfico estadístico

- ✓ • Título principal
- ✓ • Título secundario o subtítulo
- ✓ • Descripción del gráfico
- ✓✓ • Región de datos y símbolos
- ✓✓ • Eje horizontal y escala
- ✓✓ • Eje vertical y escala
- ✓✓ • Apuntadores
- ✓✓ • Descriptores de señales y marcadores

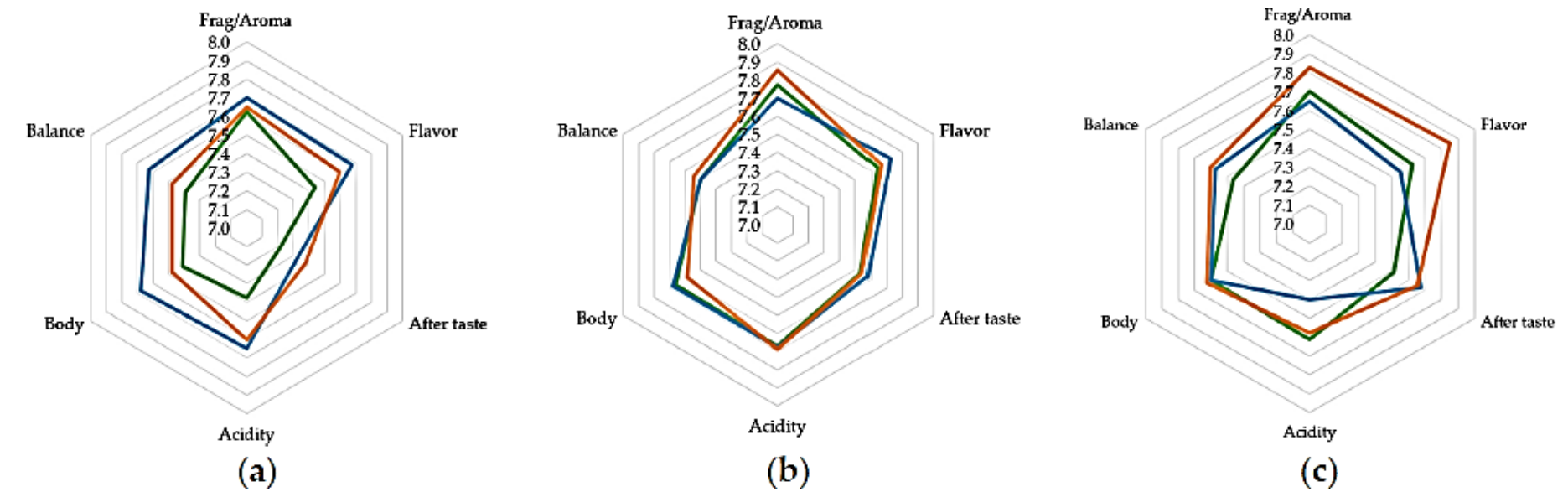


Figure 4. Average attribute scores for the Castillo (a), Cenicafé1 (b) and Tabi (c) varieties under spontaneous fermentation (green line) and temperature-controlled fermentation 15 °C (blue line) and 30° (orange line).

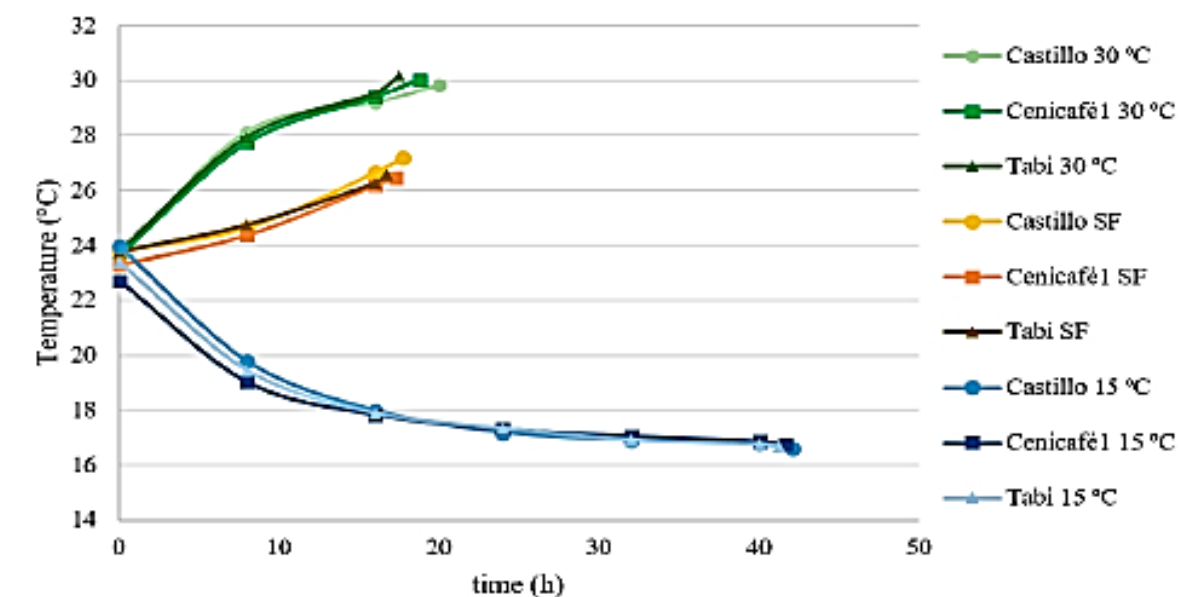


Figure 1. Behavior of the temperature of the coffee mass of three varieties of coffee under spontaneous fermentation and under temperature-controlled fermentation at 15 °C and 30 °C.

Lineamientos para la elaboración de un gráfico estadístico

- ✓ • Título principal
- ✓ • Título secundario o subtítulo
- ✓ • Descripción del gráfico
- ✓✓ • Región de datos y símbolos
- ✓✓ • Eje horizontal y escala
- ✓✓ • Eje vertical y escala
- ✓✓ • Apuntadores
- ✓✓ • Descriptores de señales y marcadores

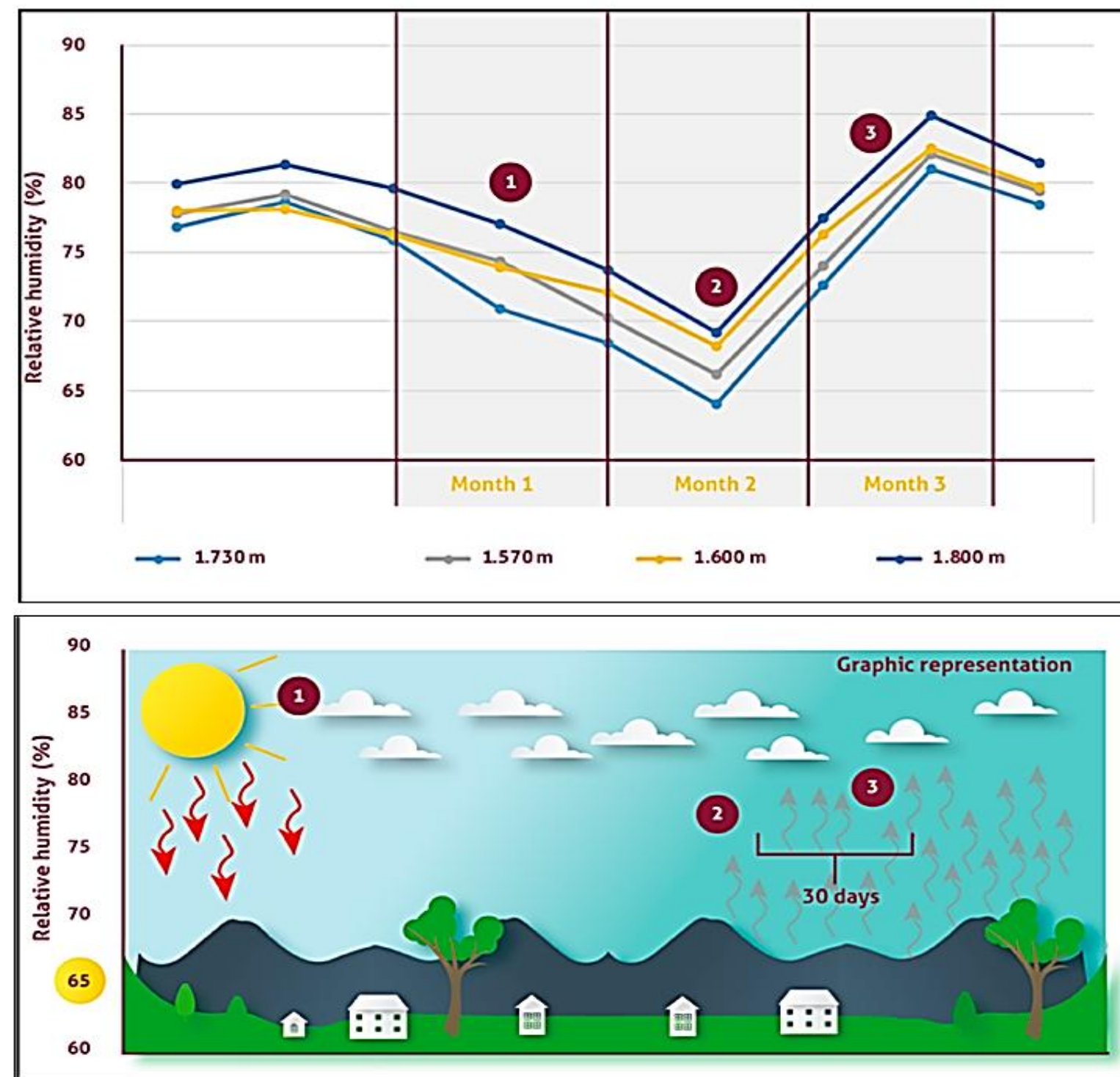


Figure 11. Explanatory diagram of the behavior of the climate in the regions where there is chamusquina infesting coffee. (1) Increase in the average temperature before the attack ≥ 19 and < 22 °C. (2). Decrease in relative humidity to less than 55% during the day. (3). Progressive increase in relative humidity up to 95% at night over a period of 30 days. (m = m.a.s.l)

Tipos de gráficos mas comunes

- **Gráficos para variables cuantitativas**

- Histograma
- Gráfico de barras
- Gráfico de dispersión
- Gráfico de líneas
- Gráfico de cajas y bigotes (boxplot)
- Gráfico de violín
- Gráfico de densidad

- **Gráficos para variables cualitativas**

- Gráfico de barras
- Gráfico de torta
- Nube de palabras

- **Gráficos para métodos estadísticos especializados**

- Gráfico de distribución de cuartiles o percentiles (qqplot)
- Dendograma
- Gráfico de codo
- Gráfico para análisis de residuales
- Mapas de calor

Tablas estadísticas

Forma de presentación de la información de manera compacta y que se supone autoexplicativa. La información se organiza de forma sistemática y lógica en filas y columnas de acuerdo con las características de los datos.

Al optar por presentar una tabla, es importante:



- ① Mostrar la información de una manera clara y resumida.
- ② Debe ser fácil de interpretar y no se debe saturar para facilitar su lectura.

Lineamientos para la elaboración tablas

- ✓• Título principal
- ✓• Título secundario o subtítulo
- ✓• Descripción de la tabla
- ✓✓• Ordenar las filas y las columnas por algún criterio o variable.
- ✓✓• Unidades de medida de variable o datos reportados
- ✓✓• Igual número de decimales en todas las cifras.
- ✓✓• Coma o punto como separador decimal (Unificar en toda la tabla y en el documento).
- ✓✓• No dejar celdas sin ningún contenido o vacías.

Table 4. Plant P concentration in response to the treatments

Soil	P ₂ O ₅ applied (g·plant ⁻¹)	P concentration (%)							
		Control	*SE	AMF	SE	PSF	SE	AMF+PSF	SE
Eutropept	0	0.22	0.00	0.18	0.03	0.27	0.02	0.25	0.00
	0.5	0.26	0.02	0.16	0.03	0.19	0.01	0.22	0.02
	1	0.23	0.02	0.24	0.01	0.20	0.00	0.25	0.01
	2	0.26	0.02	0.25	0.03	0.23	0.04	0.17	0.02
Dystropept	0	0.21	0.01	0.28	0.04	0.28	0.03	0.25	0.01
	0.5	0.28	0.02	0.34	0.03	0.25	0.00	0.31	0.02
	1	0.28	0.00	0.34	0.04	0.30	0.03	0.24	0.05
	2	0.23	0.01	0.23	0.02	0.28	0.03	0.24	0.05
Ultisol	0	0.22	0.00	0.22	0.00	0.17	0.02	nd	nd
	0.5	0.18	0.02	0.17	0.00	0.19	0.01	0.18	0.01
	1	0.19	0.02	0.20	0.01	0.17	0.02	0.19	0.01
	2	0.21	0.01	0.18	0.01	0.19	0.02	0.22	0.02
Fulvudand	0	0.18	0.00	0.16	0.02	0.21	0.00	0.18	0.00
	0.5	0.15	0.01	0.21	0.01	0.22	0.01	0.23	0.01
	1	0.22	0.03	0.22	0.01	0.19	0.00	0.20	0.02
	2	0.23	0.03	0.25	0.00	0.20	0.01	0.29	0.00

*Standard error. nd: null data

Lineamientos para la elaboración tablas

- ✓• Título principal
- ✓• Título secundario o subtítulo
- ✓• Descripción de la tabla
- ✓✓• Ordenar las filas y las columnas por algún criterio o variable.
- ✓✓• Unidades de medida de variable o datos reportados
- ✓✓• Igual número de decimales en todas las cifras.
- ✓✓• Coma o punto como separador decimal (Unificar en toda la tabla y en el documento).
- ✓✓• No dejar celdas sin ningún contenido o vacías.

CCC ID	Genotype	Percentage of Fruit Formation		
		Selfing (SF)	Fertile x Male-sterile (DC)	Male-sterile x Fertile (RC)
446	16.07/446ms565*	0.0	0.0	46.3
	16.07/446ms020*	0.0	0.0	21.0
	16.07/446ms149	1.0	0.0	61.7
	16.07/446ms158	0.0	0.5	73.2
	16.07/446ms248	0.5	0.5	54.1
376	16.07/376ms169	2.5	2.5	42.9
	16.07/376ms382	0.0	0.6	10.8
379	16.07/379ms762*	0.0	0.0	72.6
285	05.26/285ms424*	0.0	0.0	27.4
	05.26/285ms425	0.0	5.5	33.3
292	05.26/292ms734*	0.0	0.0	32.4
	05.26/292ms740*	0.0	0.0	43.6
343	05.26/343ms999	3.3	4.4	7.6
168	05.26/168ms463	0.0	1.0	26.6
194	05.26/194ms497*	0.0	0.0	27.4
480	05.26/480ms607	5.0	5.9	38.1
195	05.26/195ms623*	0.0	0.0	2.9
344	05.26/344ms2073*	0.0	0.0	27.8
318	05.26/318ms2148*	0.0	0.0	50.2
393	05.26/393ms197	0.0	1.9	23.9
386	16.08/386ms917	4.1	0.3	2.8
	16.08/386ms1333*	0.0	0.0	16.0
254	16.08/254ms1486	2.8	0.0	25.9

*Genotypes considered male-sterile, since the percentage of fruit formation in the SF and DC controlled pollination treatments was equal to zero, while in RC, it was greater than zero.

Lineamientos para la elaboración tablas

- ✓• Título principal
- ✓• Título secundario o subtítulo
- ✓• Descripción de la tabla
- ✓✓• Ordenar las filas y las columnas por algún criterio o variable.
- ✓✓• Unidades de medida de variable o datos reportados
- ✓✓• Igual número de decimales en todas las cifras.
- ✓✓• Coma o punto como separador decimal (Unificar en toda la tabla y en el documento).
- ✓✓• No dejar celdas sin ningún contenido o vacías.

Table 4: Descriptive statistic for P fixation capacity at a 0.2mg L⁻¹ concentration of P in solution, q (mg kg⁻¹ of soil), and maximum fixation capacity of P, b (mg kg⁻¹ of soil), per soil unit. Non-common letters indicate significant differences in the means of least squares according to the t-test at 5%.

Parameter	Value	Soil unit									
		CA		ER		LC		SB		SS	
q	Mean	59.32		174.41		72.19		301.32		340.91	
	Median	51.84	B	76.83	BA	38.03	BA	151.22	BA	190.86	A
	SE	11.52		70.51		36.28		151.49		101.59	
b	Mean	519.26		692.10		654.12		1352.52		1299.17	
	Median	501.24	B	545.19	BA	511.38	BA	721.05	A	825.23	A
	SE	68.10		69.42		168.64		544.29		320.77	

Group	No. of evaluated plants	Participation (%)	Preselected in the field (%)	Preselected in the laboratory (%)	Male-sterile genotypes (%)
Ethiopian origin	6,683	68.52	0.87	0.34	0.16
Interspecific hybrids	3,070	31.48	1.56	0.03	0.00
Total	9,753	100.00	2.43	0.24	0.11

Lineamientos para la elaboración tablas

- ✓• Título principal
- ✓• Título secundario o subtítulo
- ✓• Descripción de la tabla
- ✓✓• Ordenar las filas y las columnas por algún criterio o variable.
- ✓✓• Unidades de medida de variable o datos reportados
- ✓✓• Igual número de decimales en todas las cifras.
- ✓✓• Coma o punto como separador decimal (Unificar en toda la tabla y en el documento).
- ✓✓• No dejar celdas sin ningún contenido o vacías.

Table 1. Mean air temperature (Temp_{Mean}, °C), maximum air temperature (Temp_{Max}, °C), minimum air temperature (Temp_{Min}, °C), accumulated daily insolation (Insolation, W m⁻²), number of accumulated light hours (Daylength, W m⁻²), accumulated rainfall (Rainfall, mm) and number of days with rain (DaysRain, #) per flowering semester (Sem) for the four departments (DEP) evaluated in Colombia.

DEP	Sem	Temp _{Mean}	Temp _{Max}	Temp _{Min}	Insolation	Daylength	Rainfall	DaysRain
Cesar	1st	20.4	27.7	14.9	71,475	2142	680	58
	2nd	21.8	28.0	17.2	78,602	2285	1852	100
	3rd	22.9	27.5	17.2	71,463	2142	656	46
	4th	21.3	28.0	16.5	78,614	2285	1171	77
Mean		21.6	27.8	16.5	75,039	2214	1090	70
Caldas	1st	20.9	28.1	16.7	74,647	2169	1501	101
	2nd	21.2	28.6	17.0	77,194	2256	1793	118
	3rd	20.9	28.1	16.9	74,643	2169	1514	104
	4th	21.2	28.6	16.9	77,198	2256	1273	107
Mean		21.1	28.4	16.9	75,921	2213	1520	108
Quindío	1st	21.7	29.1	17.2	74,947	2172	1359	88
	2nd	22.2	29.8	17.6	77,008	2253	952	73
	3rd	21.7	29.5	17.4	74,944	2172	1707	92
	4th	22.0	29.7	17.5	77,011	2253	1040	75
Mean		21.9	29.5	17.4	75,978	2213	1265	82

Presentación de medidas y resultados de análisis (Texto)

La primera etapa del análisis estadístico se basa en la **exploración de los datos recolectados**. La descripción de las variables permite comprender cómo el investigador desarrolla el experimento, los hallazgos previos y cómo llega a la formalización de métodos mas complejos (inferencia) para dar las conclusiones del trabajo de investigación.

Medidas descriptivas

Medidas de tendencia central

- ✓ Media, mediana y moda

Medidas de variabilidad

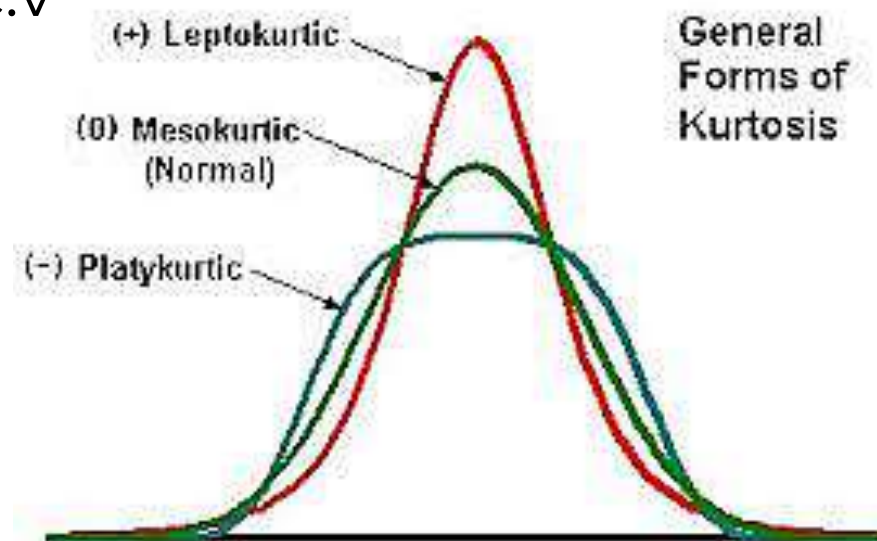
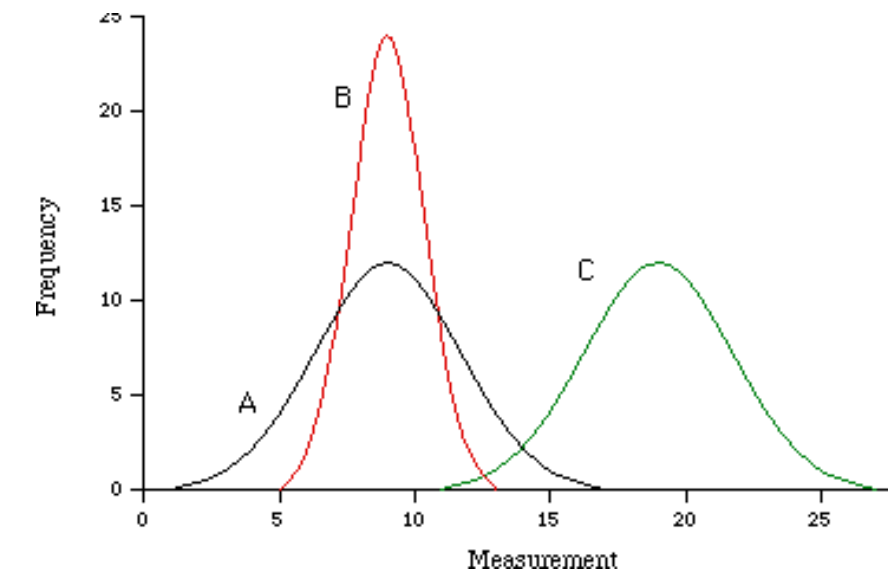
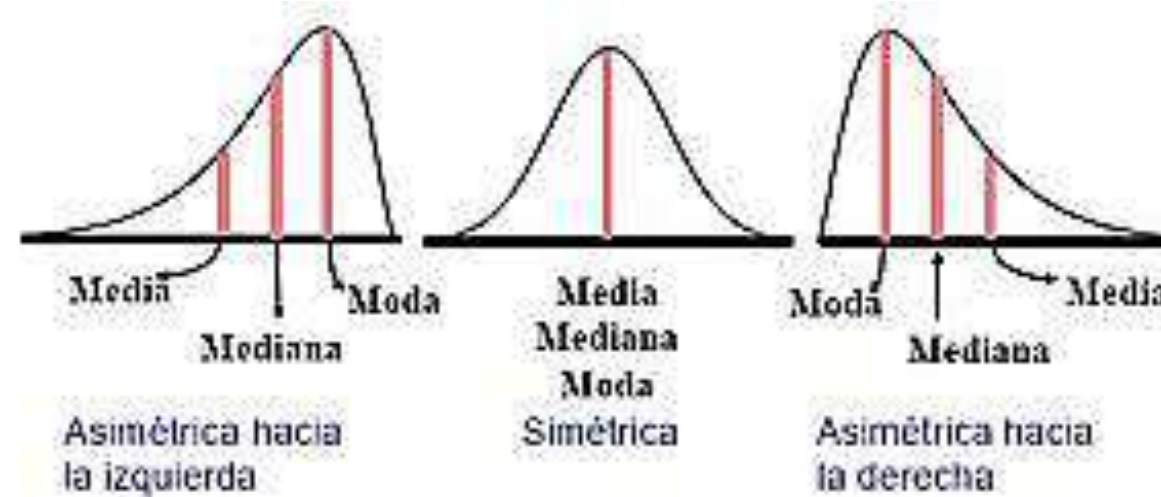
- ✓ Rango, varianza, desviación estándar, c.v

Medidas de posición

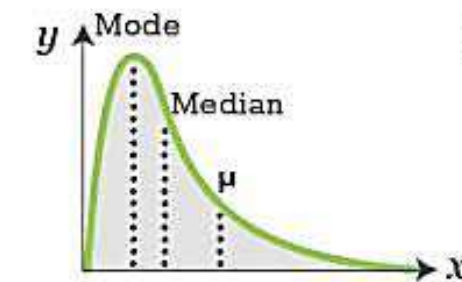
- ✓ Percentiles, deciles, cuartiles

Medidas de forma

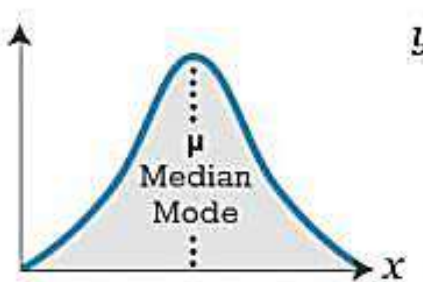
- ✓ Curtosis, simetría



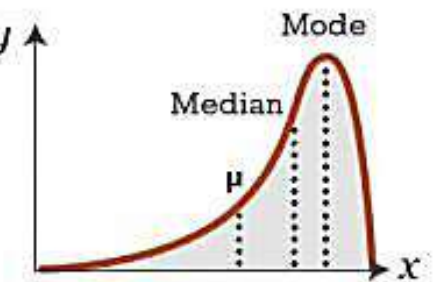
POSITIVE SKEW
(returns more likely to be negative)



NO SKEW
(returns equally likely to be positive and negative)



NEGATIVE SKEW
(returns more likely to be positive)



Presentación de medidas y resultados de análisis (Texto)

Medidas para hacer inferencia

En pruebas de hipótesis

- ✓ Nombre y estadístico de la prueba
- ✓ Valor de significancia (alfa)
- ✓ Valor de P (p-value)

En estimación puntual o por I.C

- ✓ Nivel de confianza del intervalo
- ✓ Error estándar del estimador

En ANOVA

- ✓ Diseño experimental y/o modelo
- ✓ Fuentes de variación (tto, blq, error, etc)
- ✓ Error cuadrático medio
- ✓ Valor de P (p-value)

En regresión

- ✓ Coeficientes estimados
- ✓ Coeficiente de determinación
- ✓ Pruebas de validación de supuestos
- ✓ Ecuación del modelo
- ✓ Valor de P (p-value)



Errores comunes en publicaciones científicas

Estadística descriptiva

- Medidas de centralidad sin medidas de dispersión.
- Medidas inadecuadas para la escala de la variable.
- Uso de medidas de forma para concluir distribución Normal de una variable.
- Histogramas para variables categóricas en vez de gráficos de barras.

Estimación

- Uso incorrecto de la aproximación normal para intervalos de confianza, especialmente con tamaños de muestra pequeños.

Pruebas de hipótesis

- Uso de valores tabulares inadecuados, ya sea por errores en el nivel de significancia, en los grados de libertad o en la distribución requerida.
- Pruebas de comparación de medias para factores cuyo efecto no resultó significativo, lo cual es considerado un error metodológico.

Total: **179 trabajos de grado**

172 trabajos de grado pregrado y 7 tesis de Maestría

En el **49,2%** de los trabajos se encontró al menos un error

Hernán, E. S., Guillermo, C. L., Fernando, P. D. J., José, A. J. J., & Alberto, R. R. J. (2006). *Evaluación de métodos estadísticos utilizados en trabajos de grado y tesis de los programas de la facultad de ciencias agropecuarias, en un período de tres años*. <https://repositorio.unal.edu.co/handle/unal/36619>

Errores comunes en publicaciones científicas

Diseño experimental

- Falta de aleatorización.
- Uso incompleto del análisis de varianza.
- Evaluación incorrecta de interacciones.
- Pruebas de comparación de medias cuando el análisis de varianza no es significativo.
- Realizar nuevos análisis de varianza sobre subconjuntos de tratamientos para buscar significancia.

Regresión y Correlación

- Incumplimiento de supuestos, extrapolaciones y uso de modelos inadecuados.
- Interpretar el coeficiente de correlación como prueba de causalidad.

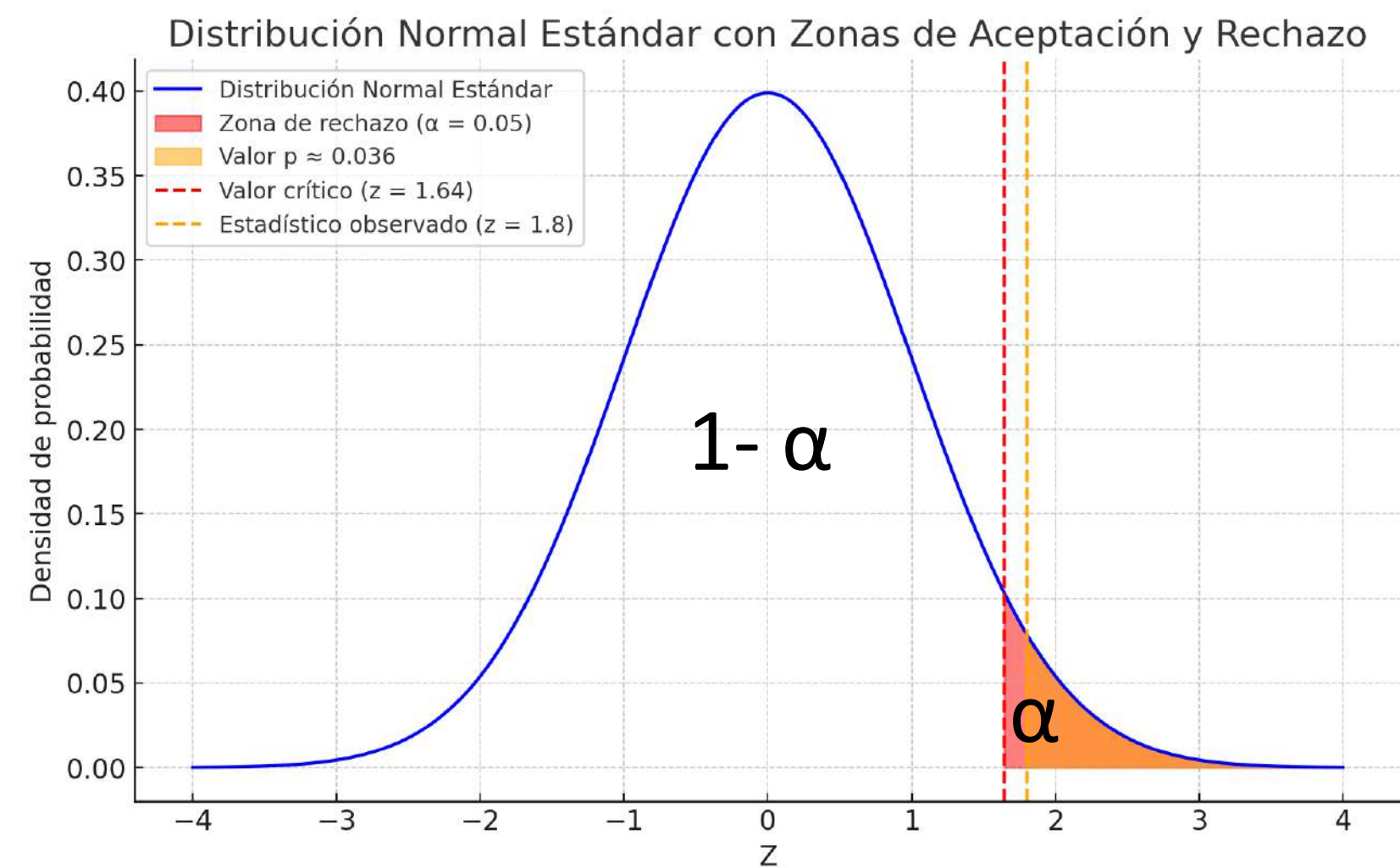
Análisis multivariado

- Elección de grupos en dendogramas, mezclar métodos de agrupación, no seguir los mismos criterios para la selección de grupos

Consideraciones

- El investigador tiene la libertad de presentar sus resultados de investigación, de acuerdo con la forma en la que él mismo se comunica y el público al que se comunica, sin embargo, debe procurar tener en cuenta los lineamientos técnicos básicos que respetan la presentación estadística de información.
- Los errores que se presentan en este contexto son más comunes de lo que pensamos, no siempre obedecen a fallas metodológicas o desconocimiento sobre los métodos estadísticos, la mayoría de las veces estos errores obedecen a fallas de comunicación, omisión o exceso de información, estrategias débiles de escritura y organización.
- En Cenicafé, elegir entre un gráfico, una tabla o describir los resultados, es una decisión que está determinada por el tipo de publicación. Los AVT y boletines muchas veces van a requerir, además de los lineamientos técnicos, de ayudas didácticas y atractivas para alcanzar la apropiación del conocimiento por parte de los caficultores.

Bonus track: Hablemos del P-value



$$H_0: \theta \leq \theta_0$$
$$H_a: \theta > \theta_0$$

Valor p (p-value):

- ✓ Mínimo valor de significancia con el cual se rechaza una H_0 .
- ✓ Probabilidad de encontrar un valor crítico mas grande que el valor calculado para rechazar la H_0 (en el caso de una prueba de cola derecha)

Decisión / Realidad	H_0 verdadera	H_0 falsa (H_1 verdadera)
No rechazar H_0	✓ Correcto (Confianza: $1 - \alpha$)	✗ Error Tipo II (β)
Rechazar H_0	✗ Error Tipo I (α)	✓ Correcto (Potencia: $1 - \beta$)

La película de terror

In February 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

Q: Why do so many colleges and grad schools teach $p = 0.05$?

A: Because that's still what the scientific community and journal editors use.

Q: Why do so many people still use $p = 0.05$?

A: Because that's what they were taught in college or grad school.

“We teach it because it's what we do; we do it because it's what we teach.” This concern was brought to the attention of the ASA Board.

“Lo enseñamos porque es lo que hacemos; lo hacemos porque es lo que enseñamos”

La película de terror

Por un breve momento en 2010, estuvo al borde de la gloria científica: había descubierto que los extremistas, literalmente, veían el mundo en blanco y negro.

Los resultados eran “claros como el día”, recuerda **Matt Motyl**, un estudiante de doctorado en psicología en la Universidad de Virginia, en Charlottesville.

Número de evaluaciones : 2000 participantes

Hipótesis : “Los políticos moderados perciben mejor los matices de gris que los extremistas”

Significancia: $\alpha = 0,05$

P-value = 0.01 **“muy significativo”**

La publicación en una revista de alto impacto parecía estar al alcance de Motyl.

La película de terror

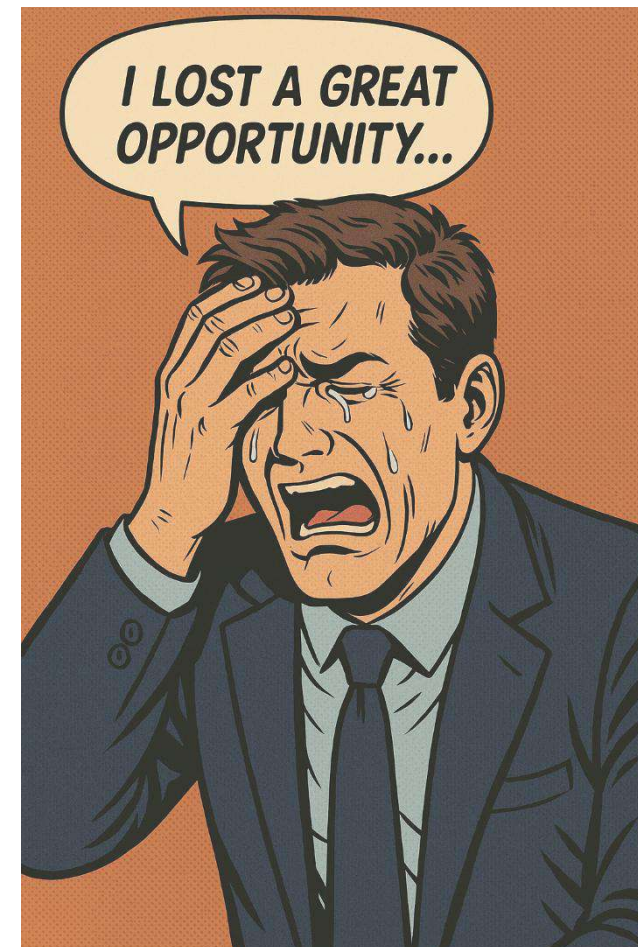
Pero entonces la realidad intervino . . .

Sensibles a las controversias sobre la reproducibilidad, Motyl y su asesor, Brian Nosek, decidieron replicar el estudio con más datos.

Resultado del nuevo experimento:

P-value = 0.59

“El efecto había desaparecido, y con él, los sueños de fama juvenil de Motyl”



La película de terror

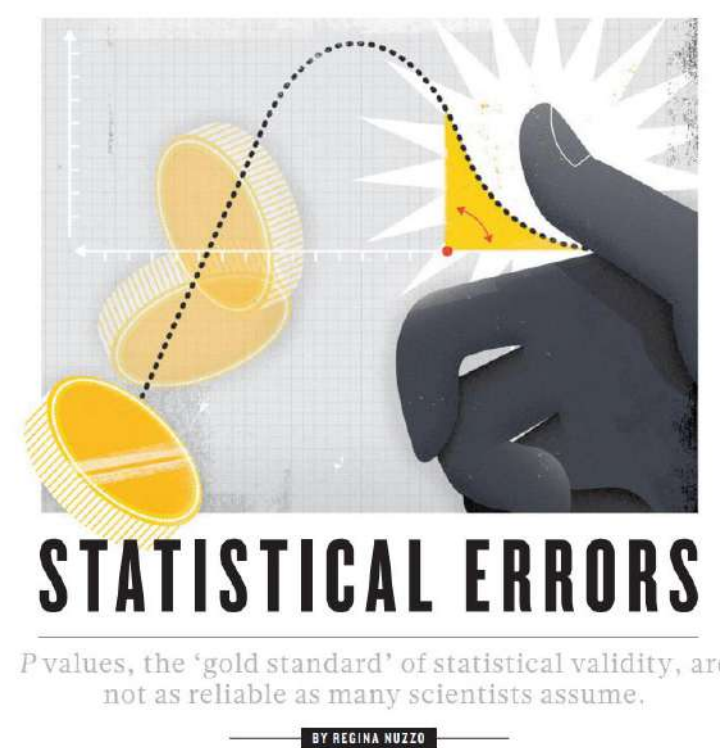
Los valores p siempre han tenido detractores.

- “ Son como mosquitos : molestos e imposibles de aplastar ”
- “ Son como el traje nuevo del emperador: llenos de problemas evidentes que todos ignoran”
- “ Son como la herramienta de un “galán intelectual estéril” que seduce a la ciencia pero la deja sin descendencia”

Un investigador incluso sugirió rebautizar la metodología como:

“**S**tatistical **H**ypothesis **I**nference **T**esting ”

El estadístico británico Ronald Fisher introdujo el valor p en la década de 1920, no lo concibió como una prueba definitiva.



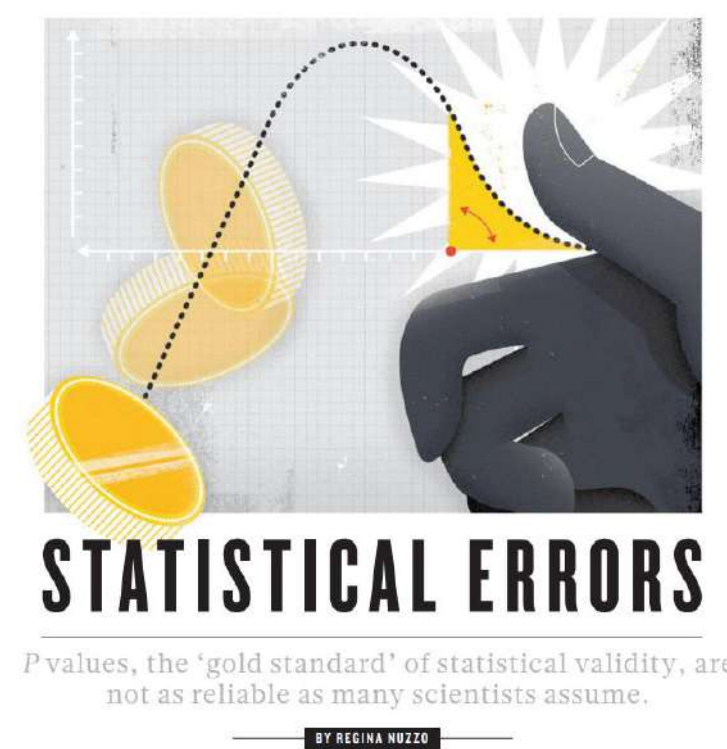
La película de terror

“A pesar de la aparente precisión del valor p , Fisher lo concibió como solo una parte de un proceso fluido y no numérico, que combinaba los datos con el conocimiento previo para llegar a conclusiones científicas.”

Intensión de Fisher:

Ofrecer una forma informal de juzgar si la evidencia era significativa en el sentido tradicional

1. Los investigadores primero formulaban una "hipótesis nula" que querían refutar.
2. Se asume que la hipótesis nula es cierta.
3. Calcular la probabilidad de obtener resultados al menos tan extremos como los observados. (p-value).



La película de terror

La pelea Neyman, Pearson Vs. Fisher

Jerzy Neyman y Egon Pearson introdujeron un marco alternativo para el análisis de datos:

- La potencia estadística.
- Los falsos positivos y los falsos negativos.

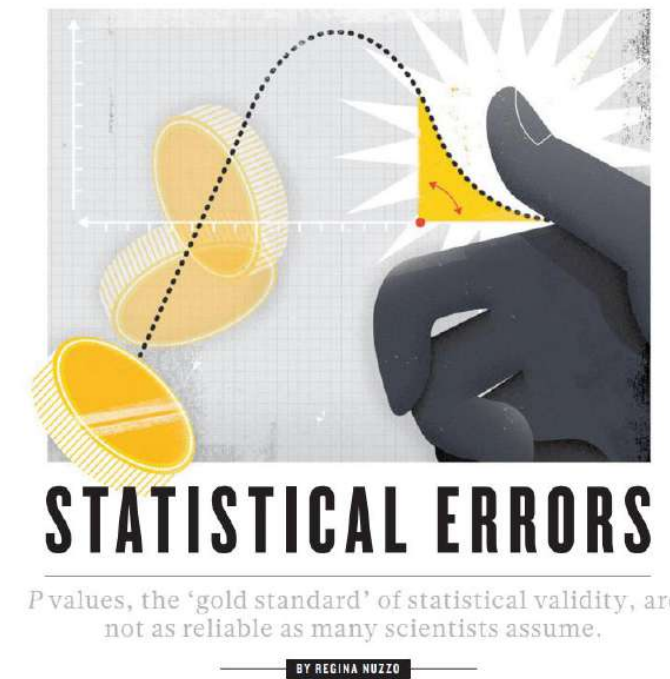
“De forma deliberada, dejaron fuera el valor p.”

Hay tres preguntas que un científico podría querer hacerse después de un estudio:

- “¿Cuál es la evidencia?”
- “¿Qué debería creer?”
- “¿Qué debería hacer?”

Richard Royall
Escuela de Salud Pública Bloomberg de Johns Hopkins
Baltimore, Maryland

“Los números son el punto de partida de la discusión científica, no su final.”



Consideraciones finales

- Si cuenta con un asesor estadístico cercano, consulte cuando tenga dudas sobre la forma de presentación de los resultados de análisis, este no tendrá la verdad revelada pero seguramente podrá fortalecer las ideas que usted ya ha considerado.
- Los comités editoriales de las diferentes revistas científicas tienen lineamientos específicos de acuerdo con los formatos y estilos que representan sus publicaciones, con toda seguridad, al someter un artículo a evaluación va a recibir algún tipo de comentario referente a mejorar o cambiar alguna forma de presentar sus resultados.

Cenicafé

Centro Nacional de Investigaciones de Café

Reserva Forestal Protectora Planalto

www.cenicafe.org



Cenicafé FNC



@cenicafe



cenicafé



CenicaféFNC



@cenicafefnc



MÁS FEDERACIÓN