

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN
Y DOCUMENTACIÓN
GRADO EN
INFORMACIÓN Y DOCUMENTACIÓN
Trabajo de Fin de Grado

USO DE MÉTODOS GRÁFICOS EN PUBLICACIONES CIENTÍFICAS DE CIENCIAS DE LA DOCUMENTACIÓN

Nuria Benito García
Tutora: M^a Purificación Galindo Villardón

Salamanca, 2010

RESUMEN

Resumen

El uso de métodos gráficos estadísticos para la representación de los datos es una herramienta necesaria en cualquier investigación. Para la ampliación de los conocimientos gráficos, se realiza un breve recorrido por la literatura especializada y se enumeran una serie de pautas para su construcción. Con objeto de medir cuantitativamente la aplicación de las representaciones gráficas en las Ciencias de la Documentación, se ha seleccionado una muestra representativa de publicaciones científicas, tanto europeas como americanas, para extrapolar a la población los resultados del análisis estadístico. La muestra se compone de cinco publicaciones relevantes en la categoría temática *Information Science & Library Science* de JCR. A continuación se realiza una descriptiva de los datos en estudio, lo que permite efectuar una inferencia estadística que verifique la hipótesis planteada. Se concluye que es necesario potenciar el uso de métodos gráficos, en las Ciencias de la Documentación, puesto que una parte importante de los artículos revisados utilizan sólo gráficos descriptivos básicos.

Abstract

Using graphical and statistical methods for the representation of data is a necessary tool in any investigation. In order to increase the graphical knowledge, a review of the technical literature and the guidelines for its construction has been carried out along with this investigation. Seeking out the purpose of quantitatively measuring the application of graphical representation in Information Science, a representative selection of European and American scientific publications has been reviewed in order to extrapolate the results of the statistical analysis into the population. This selection is made up of five relevant publications from *Information Science & Library Science* that have been taken from JCR. A description of the data of the case study permits a statistical inference which verifies the hypothesis that has been set out. We have concluded that is necessary to promote the use of graphical methods in Information Science since the majority of the articles that have been analyzed only use basic descriptive graphics.

ÍNDICE

Capítulo 1: INTRODUCCIÓN	13
1.1. El papel que juega la Estadística en la Investigación.....	14
1.2. Estadística aplicada a las Ciencias de la Documentación.....	15
1.3. Objetivos del trabajo.....	17
Capítulo 2: EL LENGUAJE GRÁFICO	19
2.1. La importancia del lenguaje gráfico.....	19
2.1.1. El lenguaje gráfico.....	19
2.1.2. La fuerza e influencia de las representaciones gráficas.....	22
2.1.3. Comprensión Gráfica y Espacios matemáticos.....	25
2.2. Percepción gráfica.....	27
2.2.1. Ilusiones y distorsiones en la percepción gráfica	27
2.3. Construcción de un gráfico estadístico.....	30
2.3.1. Normas de representación	30
2.3.1.1. Normas básicas.....	31
2.3.1.1.1. Diagramas cartesianos.....	31
2.3.1.1.2. Gráficos Geométricos y Figurativos	32
2.3.1.1.3. Elementos de construcción	32
2.4. Representaciones gráficas.....	33
2.4.1. Gráficos estadísticos más usuales	34
2.4.1.1. Gráficos univariantes/bivariantes	34
2.4.1.2. Gráficos multivariantes.....	40
Capítulo 3: RECOGIDA DE DATOS: MATERIAL Y MÉTODO.....	49
3.1. Metodología.....	50
3.1.1. Metodología para la búsqueda de las fuentes de información.....	50
3.1.1.1. Criterios de búsqueda de las publicaciones periódicas...	51
3.1.1.2. Criterios de selección de las publicaciones periódicas....	51
3.1.1.3. Revistas seleccionadas.....	52
3.1.2. Metodología para el análisis estadístico de los datos obtenidos de las publicaciones.....	54
Capítulo 4: RESULTADOS.....	55
4.1. Análisis descriptivo.....	56
4.1.1. Análisis descriptivo de cada publicación periódica.....	56
4.1.1.1. Journal of the American Medical Informatics Association (E.E.UU).....	56
4.1.1.2. Journal of Informetrics (Países Bajos).....	58
4.1.1.3. Journal of Health Communication (E.E.UU)	60
4.1.1.4. Journal of the American Society for Information Science and Technology (E.E.UU)	62
4.1.1.5. El Profesional de la Información (España)	65
4.2. Análisis inferencial.....	68
4.3. Recomendaciones.....	71

Capítulo 5: CONCLUSIONES.....	72
Bibliografía	74

ÍNDICE DE FIGURAS

CAPÍTULO 2

Figura 2.1: Iconos de Internet.....	20
Figura 2.2: Primer gráfico estadístico.....	21
Figura 2.3: Paro registrado en el mes de Febrero del 2010.....	22
Figura 2.4: Tabla de distribución de frecuencias y diagrama de dispersión .	23
Figura 2.5: Diagrama de dispersión. Anscombe (1973).....	24
Figura 2.6: Polígono de frecuencias visto a diferentes escalas.....	25
Figura 2.7: Ejes cartesianos.....	26
Figura 2.8: Ilusión óptica.....	27
Figura 2.9: Ilusión de Poggendorff.....	28
Figura 2.10: Diagrama de barras.....	28
Figura 2.11: Ilusión geométrica de Titchener.....	29
Figura 2.12: Histograma en 3D.....	29
Figura 2.13: Diagrama de barras.....	35
Figura 2.14: Diagrama de escalera.....	35
Figura 2.15: Diagrama de sectores.....	36
Figura 2.16: Polígono de frecuencias.....	36
Figura 2.17: Histograma.....	37
Figura 2.18a: Box-plot.....	39
Figura 2.18b: Box-plot para las distintas categorías en las que la revista está considerada.....	39
Figura 2.19: Diagrama de dispersión según Talla y Peso.....	40
Figura 2.20: Curvas de Andrews.....	42
Figura 2.21: Diagramas de estrella que representan los datos de criminalidad de diversas ciudades de Estados Unidos.....	42
Figura 2.22a: Parámetros de los rasgos faciales de las Caras de Chernoff.....	43

Figura 2.22b: Caras de Chernoff	44
Figura 2.23: Pictograma	44
Figura 2.24: Cartograma	45
Figura 2.25: Diagrama de perfil radial	46
Figura 2.26: Redes de grafos	47
Figura 2.27a: Criterio de representación en coordenadas paralelas	47
Figura 2.27b: Perfil de un sujeto en coordenadas paralelas	48
Figura 2.27c: Visualización de la información multivariante de varios Individuos	48

CAPÍTULO 3

Figura 3.1: Box-plot. Posición que ocupa cada una de las publicaciones periódicas seleccionadas dentro de la categoría temática a la que pertenecen: "Information Science & Library Science"	53
--	----

CAPÍTULO 4

Figura 4.1: Representación gráfica de la distribución porcentual de los gráficos estadísticos utilizados en la revista JAMIA, en 2009.....	57
Figura 4.2: Representación gráfica de la distribución porcentual de los gráficos estadísticos utilizados en la revista JI, en 2009.....	59
Figura 4.3: Representación gráfica de la distribución porcentual de los gráficos estadísticos utilizados en la revista JHC, en 2009	61
Figura 4.4: Gráficos en la revista JHC, clasificados en el bloque "Otros".....	62
Figura 4.5: Representación gráfica de la distribución porcentual de los gráficos estadísticos utilizados en la revista JASIST, en 2009	64
Figura 4.6: Gráficos en la revista JASIST, clasificados en el bloque "Otros"	65
Figura 4.7: Representación gráfica de la distribución porcentual de los gráficos estadísticos utilizados en la revista EPI, en 2009.....	67
Figura 4.8: Representación gráfica conjunta de la distribución de uso de los diferentes tipos de gráficos en las revistas JHC, EPI, JAMIA, JI, JASIST	69
Figura 4.9: Vista frontal y vista girada de los perfiles de uso de los gráficos menos frecuentes.....	70

ÍNDICE DE TABLAS

CAPÍTULO 3

Tabla 3.1: Título, Factor de Impacto, Ranking y Cuartil de las publicaciones. Año 2008	52
---	----

CAPÍTULO 4

Tabla 4.1: Distribución de los tipos de gráficos en la revista JAMIA	57
Tabla 4.2: Distribución de los tipos de gráficos en la revista JI	59
Tabla 4.3: Distribución de los tipos de gráficos en la revista JHC	61
Tabla 4.4: Distribución de los tipos de gráficos en la revista JASIST	64
Tabla 4.5: Distribución de los tipos de gráficos en la revista EPI	67
Tabla 4.6: Distribución de frecuencias de las diferentes representaciones gráficas en las cinco revistas.....	68

CAPÍTULO 1

INTRODUCCIÓN

1. INTRODUCCIÓN

Elegir el Trabajo de Fin de Grado no ha sido una tarea trivial por varias razones: primero porque es mi primer trabajo de investigación, segundo por la dificultad que entraña elegir el tema y el tutor y, tercero porque es el puente entre mi actual vida de estudiante y mi futura vida como profesional.

Mi elección fue “Estadística”. A priori puede parecer una elección muy arriesgada, sin embargo ahora que ya he realizado el trabajo, creo que fue una decisión acertada.

Lo que me movió a hacer esta elección es el hecho de que cuando un profesional de las Ciencias de la Documentación tiene que abordar un trabajo de investigación, uno de los mayores hándicaps que se encuentra es realizar, o simplemente interpretar correctamente, el análisis estadístico de los datos. Pensé que trabajar en Estadística (Análisis de Datos) podría enriquecerme y así el paso hacia mi vida como profesional podría ser más sencillo.

Dentro del Análisis de Datos, la exploración descriptiva de los mismos es el primer paso para encontrar relaciones y sugerir hipótesis y, el análisis gráfico de la información, dentro de ese análisis descriptivo, es la forma más intuitiva y la que más abunda en todas las publicaciones científicas. Por esta razón el trabajo se centra en el tratamiento gráfico de los datos, en el campo de las Ciencias de la Documentación, evaluado a través de publicaciones científicas con visibilidad y calidad científica reconocida.

Aunque el primer gráfico estadístico registrado en la literatura data de 1644, el artículo publicado por Tukey en 1962, *"The Future of Data Analysis"*, fue el germen que proporcionó un inusitado auge de Métodos Gráficos en la Estadística, inaugurando una nueva era en este campo al otorgarles un papel central en análisis exploratorios.

Sin embargo, la importancia y protagonismo que entonces se preveía, no llegó a hacerse realidad hasta más tarde. Fue en la década de los 70 cuando aparecen publicaciones sobre el tema, tanto históricas (Royston, 1970), como de recapitulación (Fienberg, 1977), o de carácter novedoso (Chernoff, 1973; Tukey, 1977). Incluso tuvo lugar un simposio dedicado exclusivamente al tratamiento gráfico de los datos (Wang & Lake, 1978).

La aparición, desarrollo y generalizada utilización de los ordenadores fue y es, sin duda, la causa fundamental de este auge. Hoy día cualquier programa estadístico, por ejemplo SPSS, e incluso hojas de cálculo como Excel permiten realizar gráficos estadísticos con gran facilidad.

Este trabajo se ha estructurado de la siguiente manera: en el primer capítulo se comienza, tras una breve introducción, analizando el papel de la Estadística en la Investigación y particularmente en Ciencias de la Documentación; además se plantean los objetivos centrados en averiguar el grado de aplicación de las técnicas gráfico-estadísticas en las Ciencias de la Documentación. El capítulo 2 se dedica al Lenguaje Gráfico y su importancia

no sólo en los análisis estadísticos sino también en la Sociedad, como herramienta de comunicación. Se destacan las ventajas de su uso y se ponen de manifiesto también las limitaciones. Se presentan algunas ilusiones ópticas muy citadas en la literatura, y se enumeran las pautas para la correcta creación e interpretación de los gráficos estadísticos. Se describen además los gráficos univariantes y multivariantes más utilizados en las publicaciones.

Los capítulos 3 y 4 contienen la parte central de la investigación: en el 3º se describe la metodología, con especial énfasis en la selección de la muestra de revistas, se especifican los criterios de inclusión y exclusión y se analiza su representatividad; en el capítulo 4 se presentan los resultados, no solo desde un punto de vista cualitativo, sino realizando un análisis estadístico descriptivo e inferencial y presentándolos en forma gráfica. *“Gráficos para presentar los resultados del estudio de los gráficos utilizados en las revistas de Ciencias de la Documentación”*.

El estudio termina con las conclusiones y la bibliografía.

1.1. El papel que juega la Estadística en la Investigación

¿Qué se entiende por Estadística? Estadística, derivado del latín estatus, significa estado, posición o situación. La Estadística, en un principio, era entendida como una herramienta para recoger y sintetizar datos que caracterizaban las condiciones predominantes en el Estado: por ejemplo, el número de nacimientos. Todo trabajo estadístico del siglo XIX y principios del XX tuvo como característica el estudio de las grandes masas.

La Estadística tiene ya muy poco en común con aquellos cementerios de números que hace años caracterizaban a esta ciencia. La idea clave en el pasado era la colección completa de datos. Toda la población era estudiada en censos cuidadosamente preparados y aparecían listas completas de todos los acontecimientos demográficos y económicos. También en nuestros días las estadísticas oficiales recogen los datos publicados por las agencias del gobierno; sin embargo, la Estadística es mucho más que eso.

A partir de los años 30, se produjo un nuevo giro en el desarrollo de la Estadística. Nació la Estadística Moderna, la Estadística Inductiva. La Estadística es el conjunto de los Métodos que nos permiten tomar decisiones razonables en caso de incertidumbre, es decir, da las bases para una teoría de la Decisión. La palabra clave de la Estadística Moderna es "Muestra".

La Estadística actualmente está considerada como la ciencia que estudia conjuntos de unidades taxonómicas (libros, artículos, pacientes, plantas, empresas, partidos políticos y un largo etc.), sobre las que se analizan variables cuantitativas y/o cualitativas, con el objetivo de describir y clasificar, en términos matemáticos, las unidades en estudio.

En el desarrollo de la ciencia en general y en especial en el campo de las Ciencias de la Documentación, la metodología estadística se ha convertido en una herramienta imprescindible para la obtención, análisis e interpretación de los datos que proceden de los estudios e investigaciones hechas al respecto. No existe investigación, proceso o trabajo encaminado a obtener información cuantitativa en general, en la que la estadística no juegue un papel central. Tradicionalmente se entendía que la Estadística requiere una alta y especializada formación en Matemáticas, pero afortunadamente eso no es cierto en nuestros días. El ordenador hace los cálculos matemáticos y el usuario basta con que conozca cómo se interpretan las salidas del ordenador, para realizar su trabajo.

1.2. Estadística aplicada a las Ciencias de la Documentación

El presente trabajo trata de ir más allá de la concepción típica que se tiene del profesional de Información y Documentación como mero ejecutor de tareas del proceso técnico y poner de manifiesto que con ayuda de las herramientas estadísticas puede, no solo recuperar la información, sino sacar conclusiones sobre las que basar la toma de decisiones futuras.

Varios autores han hablado de la importancia de este proceso entre los que cabe destacar a Moya-Anegón et al. (1996). Estos autores afirman “Si vamos abandonando el nivel intuitivo en el que hoy principalmente se mueve nuestra profesión, y pasamos a describir matemáticamente los procesos informativos, estaremos en condiciones de llegar a situaciones que hoy están muy lejos de alcanzarse. Nos estamos refiriendo a la capacidad de generar modelos”. En sentido similar se manifiesta Marín (1999): “a medida que el mundo se hace más complejo han surgido ciertos tipos de instituciones que actúan como generadoras, transformadoras y difusoras de la información. Como intermediarios de la información registrada y los usuarios, los científicos de la Biblioteconomía, la Información y la Documentación no sólo deben de ser capaces de localizar la información sino de interpretarla y evaluarla para los clientes”.

Y no solo la Estadística es importante con fines de investigación; en las tareas cotidianas, las técnicas cuantitativas resultan de gran ayuda: gestión de los centros, gestión de plazos para el cumplimiento de metas y objetivos, conocer qué resultados se producen con los recursos empleados, etc.

Marín (1999) afirma: “El documentalista es un consumidor de diversos datos, que estudia con objeto de mejorar su desempeño profesional o los servicios de la organización a la que pertenece. Además es de esperar que los documentalistas participen cada vez más en proyectos de investigación. Esos proyectos de investigación quizás consideren principalmente actividades de Bibliometría; pero los documentalistas también pueden participar en otros tipos de estudios estadísticos”.

Hay que tener en cuenta que en las Ciencias de la Documentación no sólo existen profesionales técnicos, sino que también existen profesionales dedicados a la investigación bibliológica. Una de las ramas en donde tener conocimientos estadísticos es importante, es en la Bibliometría.

Por Bibliometría se entiende lo siguiente: la Bibliometría es una parte de la Cienciometría que aplica métodos matemáticos y estadísticos a toda la literatura de carácter científico y a los autores que la producen, y tiene por objeto estudiar y analizar la producción científica.

La Bibliometría se fundamenta en las leyes bibliométricas, basadas en el comportamiento estadístico regular que a lo largo del tiempo han mostrado los diversos elementos que forman parte de la ciencia. Como se puede observar, la Estadística está presente en todas las áreas correspondientes a las Ciencias de la Documentación.

Martínez de Sousa (1989) la define como: “la técnica de la investigación bibliológica tiene por fin, por un lado, analizar el tamaño, crecimiento y distribución de la bibliografía en un campo determinado, y por otro, estudiar la estructura social de los grupos que la producen y la utilizan”. En función de estos dos aspectos la bibliometría tiene dos enfoques complementarios (López-Piñero, 1972):

- Análisis estadístico: análisis del tamaño, crecimiento y distribución de la bibliografía científica.
- Análisis sociométrico: análisis de las estructuras sociales de los grupos que producen, transmiten, consumen y utilizan la ciencia.

Es necesario dar un nuevo paso y admitir que las técnicas cuantitativas son esenciales en el ámbito de las Ciencias de la Documentación por múltiples razones: una buena gestión de los centros va a necesitar plantearse el cumplimiento de metas y objetivos, conocer el grado de cumplimiento de los mismos, elaborar nuevos objetivos, justificar el servicio que se presta, saber qué resultados producen con los recursos empleados, cómo se pueden dar más y mejores servicios con los mismos recursos, etc. (Moya-Anegón et al., 1996).

Como se ha puesto de manifiesto, es muy importante tener en cuenta que una apropiada gestión de la información no es sólo un proceso técnico. Es de gran utilidad, aunque no se esté realizando un trabajo de investigación, recurrir a métodos cuantitativos que ayuden a analizar dicha información.

1.3. Objetivos del trabajo

1.3.1. Objetivos generales

Los objetivos generales son los enunciados que expresan el propósito que fundamenta todo trabajo de investigación. Son los siguientes:

- Dar a conocer los diferentes tipos de gráficos que se pueden utilizar dependiendo del tipo de variables a estudiar.
- Averiguar el grado de aplicación de las técnicas gráfico-estadísticas en las Ciencias de la Documentación.

1.3.2. Objetivos específicos

- Estudiar el tratamiento gráfico de los datos, en el campo de las Ciencias de la Documentación, evaluado a través de publicaciones científicas con visibilidad y calidad científica reconocida.
- Analizar los tipos de gráficos empleados en las publicaciones periódicas seleccionadas en la muestra.

CAPÍTULO 2

EL LENGUAJE GRÁFICO

2. LENGUAJE GRÁFICO

El lenguaje gráfico posee la capacidad de expresar una idea de manera gráfica. Como cualquier lenguaje cuenta con un vocabulario, una gramática y una sintaxis propios y además, por el hecho de utilizar un lenguaje estructurado dota a los procedimientos de unas características. Algunas de ellas son:

- Reglas, lógica y estructura.
- Permite el intercambio de información.
- Permite la comunicación.
- Habilita la comparación.
- Actúa sobre la transmisión de conocimiento.

El uso del lenguaje gráfico implica un cambio de paradigma puesto que se cambia la concepción textual de un proceso o procedimiento por una concepción gráfica que conlleva la utilización del pensamiento gráfico. Esta abstracción puede parecer complicada pero si se percibe como un lenguaje que desde la antigüedad se usa puede que no parezca un gran cambio sino que además, se convendrá en que es una faceta de la comunicación que ya se tenía interiorizada.

2.1. La importancia del lenguaje gráfico

Las capacidades de percepción del ojo humano y las teorías e investigaciones acerca de los principios generales para las construcciones y representaciones gráficas son muy interesantes a la hora de comprender la adecuada representación de los datos mediante el lenguaje gráfico. Por esta razón se realiza un breve estudio preliminar de la evolución y percepción gráfica, y de la adecuada construcción de un gráfico, en búsqueda de una mejor y mayor comprensión y memorización de la información que dichas representaciones transmiten a los lectores.

2.1.1. El lenguaje gráfico

De aproximadamente 35.000 años a. C (Paleolítico superior) datan las primeras manifestaciones gráficas que aparecen en los techos y paredes de cuevas donde nuestros antepasados se refugiaban del calor, la lluvia y de los depredadores. Estas representaciones gráficas se basan en simples líneas serpentiformes de naturaleza y significado desconocido. Es por tanto lógico pensar, que una de las primeras formas de comunicación es la representación gráfica. Estas representaciones tienen sentido cuando transmiten información sobre situaciones, objetos e ideas.

Se piensa que las representaciones gráficas que se comenzaron a utilizar en la prehistoria fueron la base de posteriores lenguajes articulados más complejos. Este lenguaje prehistórico se sirve de numerosos símbolos y signos que han ido evolucionando con el tiempo, y que hoy en día se pueden encontrar en casi todas las manifestaciones de la actividad humana, en muchas ocasiones con significado universal. La creencia general en la actualidad es que la evolución de pictogramas e ideogramas desde lo concreto a lo abstracto fue lo que dio origen a la escritura (Schumandt-Besserat, 1978). No es exagerado por ello pensar que la expresión gráfica fue uno de los gérmenes de la Cultura.

El lenguaje gráfico es el conjunto de símbolos y convenios que permiten comunicar una información de la manera más eficiente posible (Gete-Alonso & del Barrio, 1990).

Detengámonos a pensar un momento en el mundo que nos rodea, y se podrá observar la importancia y uso del lenguaje gráfico en la vida cotidiana. Por ejemplo se pueden encontrar representaciones gráficas en las instrucciones del lavado de las prendas, en señales de tráfico, en los emblemas de las organizaciones, en el parte diario acerca del estado del tiempo, o incluso relajados en el sofá de casa mientras se navega por Internet o manejando aparatos de audio, DVD, etc. Muchas de estas representaciones son mundialmente conocidas, y es por tanto incoherente sustituirlas por oraciones o palabras, ya que aparte de tener mayor dificultad a la hora de expresar la idea conceptual asociada a la representación gráfica, también existe una pérdida de tiempo y espacio.

Un ejemplo de algunos de los lenguajes gráficos en la Sociedad en la que actualmente se vive puede observarse en la figura 2.1.



Figura 2.1: Iconos de Internet

La masiva utilización de iconos y representaciones gráficas en las acciones de la vida cotidiana, se debe en la mayoría de los casos a su utilización como herramienta de comunicación de datos y análisis estadístico. Este uso masivo ha hecho que las representaciones gráficas abandonen los libros especializados sobre la materia, encontrándose en la actualidad en los medios de comunicación social de todo tipo, e incluso es posible encontrarlas ya en obras dirigidas al gran público cuyo único fin es el de entretener.

El primer gráfico estadístico del que hasta ahora se tiene registro se remonta al S. XVII, y actualmente se discute si su aparición data del año 1628 o del año 1644. Su creador, Miguel Florencio vanLangren, quiso mostrar las estimaciones de las diferencias en longitud entre las ciudades de Toledo y Roma. Dichas estimaciones se calcularon utilizando distintos métodos cartográficos, astronómicos y geográficos en diferentes tiempos y por diferentes hombres entre los que se encontraba el famoso Ptolomeo.

El gráfico se conserva adjunto a un informe que fue presentado en 1624, por vanLangren a la corte de la infanta Isabel.

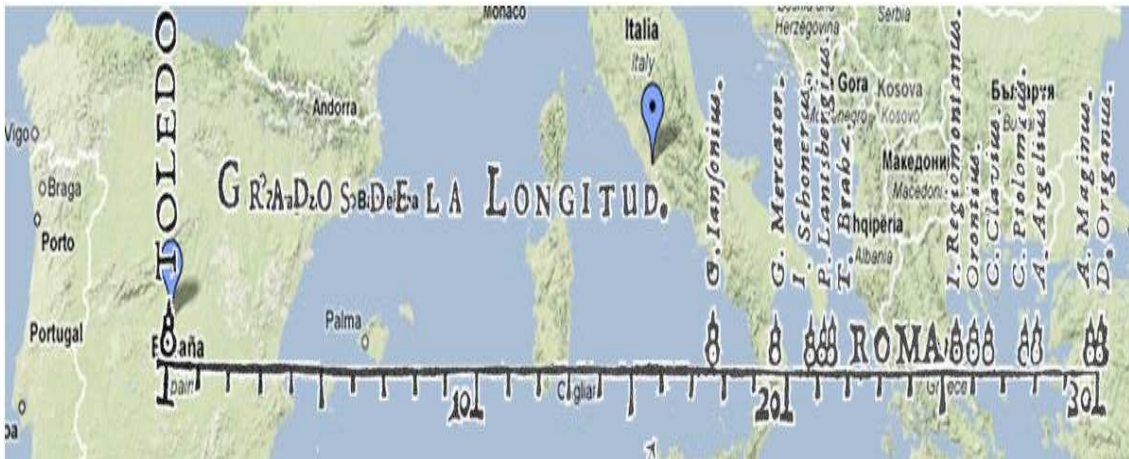


Figura 2.2: Primer gráfico estadístico datado

Este gráfico es el primer ejemplar conocido en mostrar el efecto del ordenamiento en la visualización de los datos, aunque el objetivo de vanLangren era mostrar las enormes diferencias entre las estimaciones de los más renombrados astrónomos y geógrafos de la historia.

Cierto es que en España la Estadística dista aún de ocupar un lugar como el que, por ejemplo, tiene en otros países europeos, americanos, etc., aunque la situación parece estar cambiando de manera evidente ya que cada vez más se recurre a datos y a análisis estadísticos para transmitir la información de manera más visual.

Hoy más que nunca se vive involucrado en una sociedad donde los gráficos, en base a datos y análisis estadísticos, están a la orden del día; analizar la evolución del paro (ver figura 2.3), la subida de los precios en el mercado, etc.

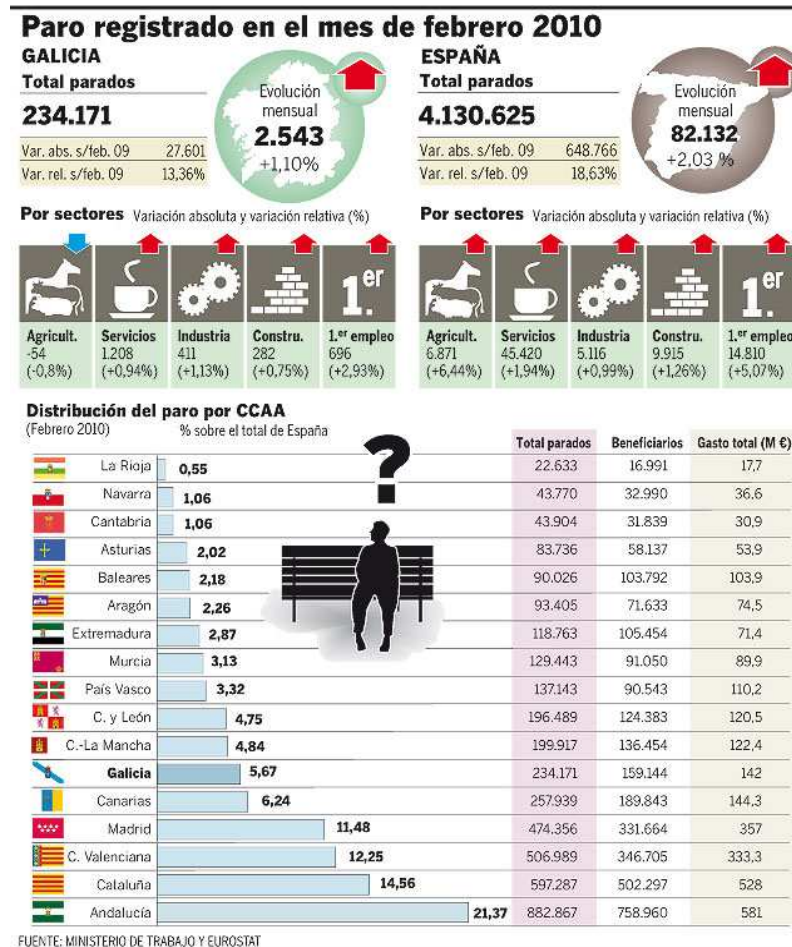


Figura 2.3: Paro registrado en el mes de Febrero del 2010

2.1.2. La fuerza e influencia de las representaciones gráficas

Es comúnmente aceptado por la comunidad científica que por lo general, una representación gráfica aporta mayor información acerca de las características y patrones de los datos, que un texto explicativo o una representación tabular de los mismos.

En el ejemplo siguiente, las características de los datos son mucho más evidentes cuando aparecen en un diagrama de dispersión (derecha), que cuando aparecen simplemente tabulados (izquierda), (figura 2.4). X e Y están relacionadas con una relación directa y el modelo lineal describiría adecuadamente esa relación.

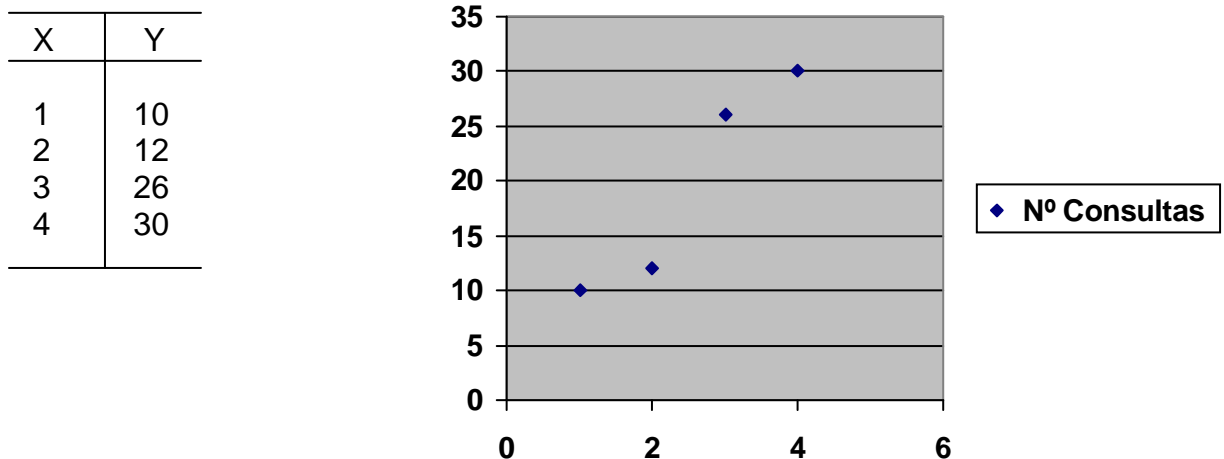


Figura 2.4: Tabla de distribución de frecuencias y diagrama de dispersión

Un estudio un poco más complejo, que aparece en los libros de Estadística, y está incluso recogido en la “Enciclopedia de Estadística”, es el que contiene los datos propuestos por Anscombe en 1973 (figura 2.5), el cual pone de manifiesto los riesgos de utilizar análisis estadísticos sin utilizar gráficos.

En este ejemplo, se parte de cuatro conjuntos de datos que pueden ser descritos por el mismo modelo matemático, una recta, que tiene exactamente la misma pendiente y el mismo término independiente. Por lo tanto, desde un punto de vista algebraico-analítico, los cuatro conjuntos serían identificables. Sin embargo, la representación gráfica de los residuos del modelo pone claramente de manifiesto que los cuatro conjuntos de datos, a pesar de ser descritos por el mismo modelo, recogen situaciones claramente diferenciadas.

El primer conjunto de datos está bien descrito por un modelo lineal, (gráfica superior izquierda), el segundo se ajusta realmente a un modelo curvilíneo, un polinomio, (gráfica superior derecha), el tercero (gráfica inferior izquierda), se ajusta a una línea recta pero presenta un outlier que sesga ligeramente la posición de la pendiente y el cuarto presenta un outlier con alta influencia, que no sólo sesga sino que claramente distorsiona la tendencia de la recta, haciendo que el modelo propuesto no sea en absoluto representativo.

De lo anterior se deduce, por tanto, que un estudio analítico puede ser confuso sino se acompaña de un estudio gráfico que evidencie los patrones internos en los datos. Desgraciadamente, las publicaciones científicas en todos los campos de la ciencia están llenas de ajustes de modelos en los que ni siquiera plantean el análisis gráfico de los residuales, confundiendo así el poder explicativo con el poder predictivo del modelo.

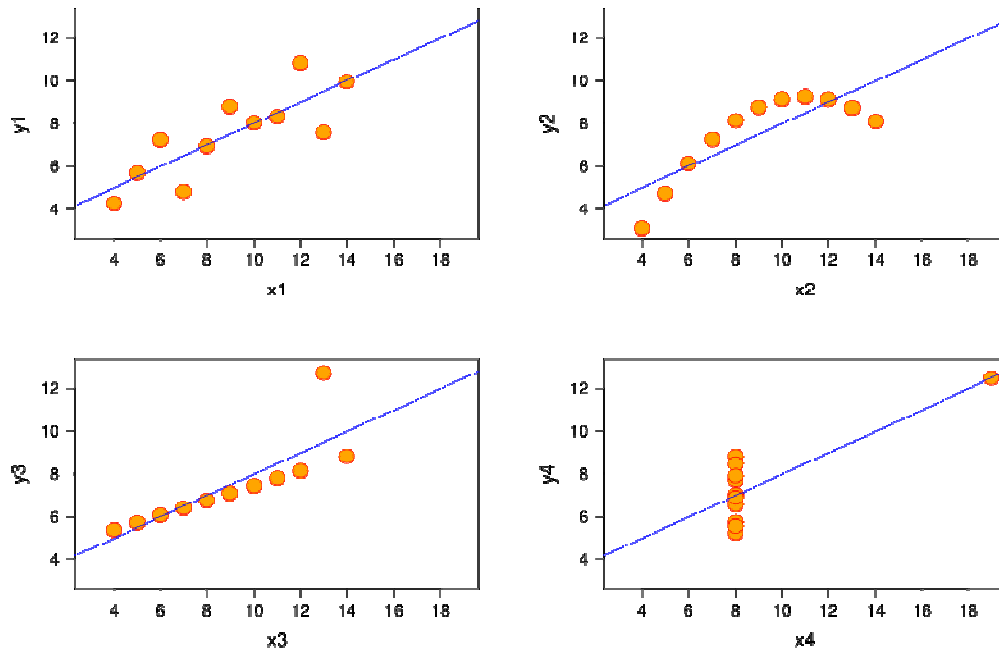


Figura 2.5: Diagrama de dispersión. Anscombe (1973)

Los Diagramas de Anscombe evidencian cómo cuatro grupos de datos que visualmente producen idénticas rectas de regresión e idénticos coeficientes de correlación y errores estándar, corresponden a casos muy diferentes. Así la información proporcionada puede hacer concluir resultados erróneos si los datos no se tienen en cuenta; algo muy fácil de evitar si se representan gráficamente.

Gran parte del poder e importancia de los Métodos de Representación Gráfica, es que nos fuerzan a percibir aquello que nunca esperábamos ver (Tukey, 1977).

No obstante hay que señalar también que una utilización inadecuada (o una interpretación incorrecta) puede llevarnos a conclusiones que no se corresponden con la realidad tal como pone de manifiesto el siguiente ejemplo:

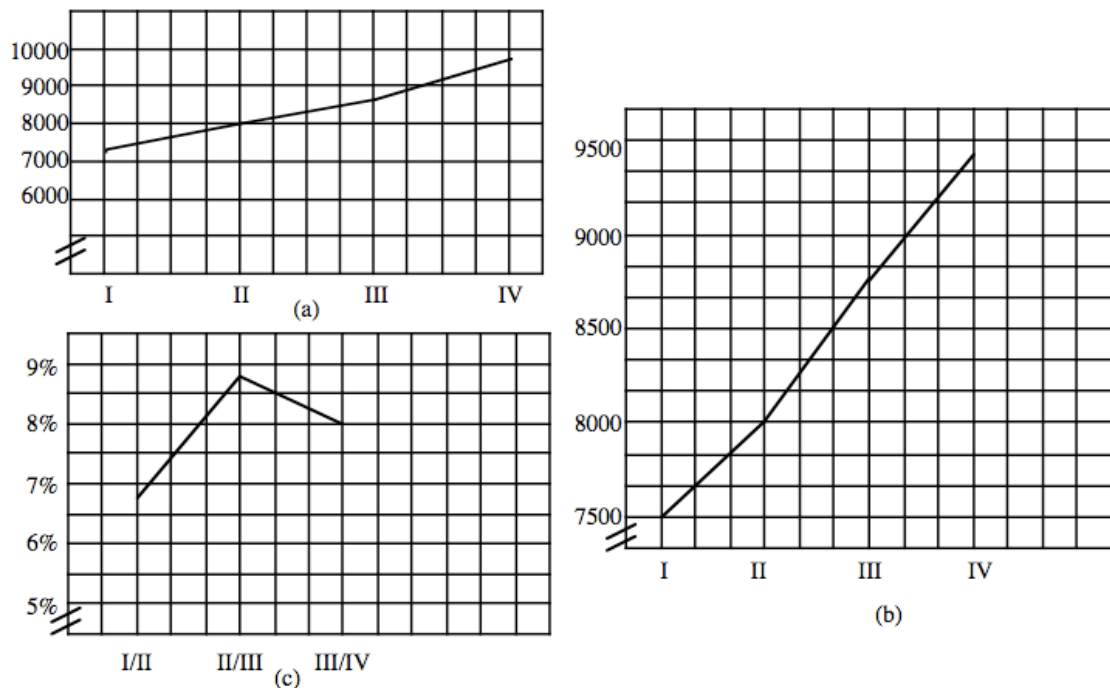


Figura 2.6: Polígono de frecuencias visto a diferentes escalas

La ascensión lenta de la curva (a) pone de manifiesto un crecimiento moderado. Los mismos datos pueden expresar un crecimiento explosivo y optimista (b). Se puede obtener una curva ascendente primero y descendente después, si se toman los índices de crecimiento de un período a otro en lugar de los números absolutos (c), (ver figura 2.6) (Adaptada de Swoboda, 1975).

Es pues evidente que tampoco conviene juzgar las imágenes solas, sino que deben considerarse también los números y las escalas. Lo más aconsejable es un análisis combinado: analítico y gráfico.

2.1.3. Comprensión Gráfica y Espacios Matemáticos

Un estudio de la utilización de los Métodos Gráficos resultaría incompleto si no se incluyesen aspectos tales como la comprensión y utilización de los mismos. Dado que son procedimientos gráficos descriptivos resulta interesante el tratamiento de todo lo relacionado con la percepción, comprensión gráfica y memorización de la información, así como de la construcción de los gráficos y la interpretación de los mismos.

2.1.3.1. Los Espacios Matemáticos y el Campo Visual

Dado que no existe un sólo espacio, sino varios, conviene distinguir los múltiples espacios matemáticos del espacio físico y el espacio perceptual visual para poder abordar el tema de la percepción visual.

Un espacio matemático se construye a partir de unos axiomas o postulados, que no es más que una fórmula bien formada de un lenguaje formal que se acepta sin demostración, como punto de partida para demostrar otras fórmulas, y se describe y define por su geometría. Algunas de estas geometrías pueden aplicarse, en mayor o menor medida, al espacio perceptual.

Existen muchos tipos de espacios matemáticos definidos por sus respectivas geometrías (proyectivo, euclídeo, topológico, etc.). De entre todos, el más conocido y utilizado es el Espacio Euclídeo Ordinario.

El Espacio Euclídeo se define con respecto a tres ejes cartesianos de referencia (z, y, x), correspondientes a tres dimensiones. Los dos primeros ejes se sitúan en un mismo plano y forman entre sí un ángulo de 90° , el tercer eje es perpendicular con respecto al plano correspondiente a los dos primeros ejes y pasa por su origen (figura 2.7).

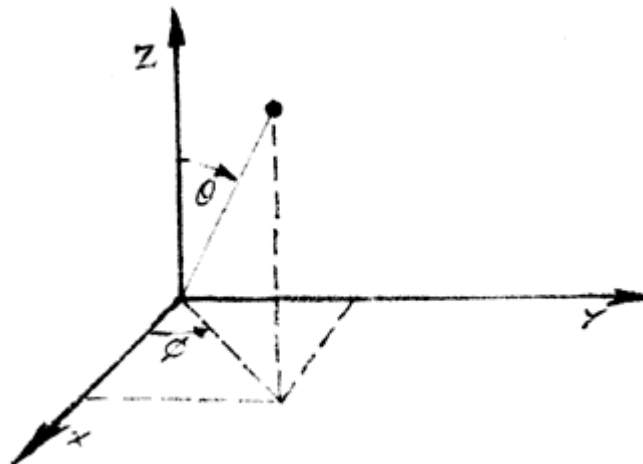


Figura 2.7: Ejes cartesianos

En el espacio Euclídeo, la situación de un punto queda determinada por sus tres coordenadas rectangulares relativas a los ejes de referencia; y los ángulos, paralelas y distancias se conservan durante los desplazamientos. Además gracias a la existencia de ejes de referencia es posible un procedimiento de medición. Admitir que el espacio físico es Euclídeo no equivale a que el espacio perceptual lo sea.

A pesar de esta controversia acerca de si el espacio perceptivo coincide con el espacio matemático Euclídeo es éste el que generalmente se utiliza para representar el mundo.

2.2. Percepción gráfica

La percepción gráfica abarca un extenso campo, con mucho terreno por explorar todavía, y son tan sugerentes estos fenómenos que desbordan el tratamiento especialista para formar parte de la más variada miscelánea de curiosidades y pasatiempos. La prodigalidad y mezcolanza de las imágenes pueden llegar a crear confusión.

2. 2.1. Ilusiones y distorsiones en la percepción gráfica

La percepción de los métodos gráficos por el ojo humano juega un papel muy importante a tener en cuenta en el Análisis de Datos. Cleveland (1985) afirma que “conocer el proceso de decodificación en el proceso perceptivo posibilitará una más adecuada construcción y utilización de los gráficos”. A pesar de que nuestra visión funcione de acuerdo con las mismas leyes ópticas que una máquina fotográfica percibimos las cosas de modo diferente. Por ejemplo, en la figura 2.8 se observa algo realmente curioso; mientras que en realidad la figura está estática, se puede llegar a percibir en movimiento.

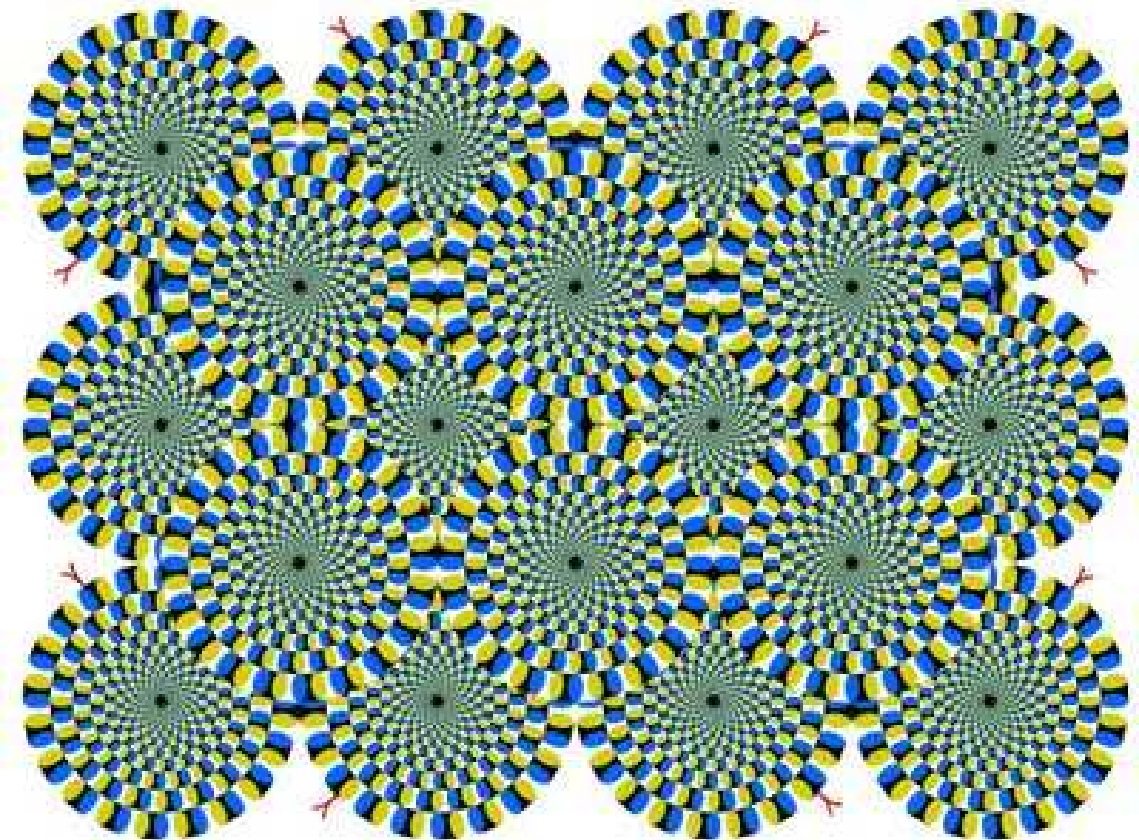


Figura 2.8: Ilusión óptica

No es tema del presente trabajo estudiar el proceso de la percepción, aunque resulta interesante y oportuno hablar de algunas de las famosas ilusiones geométricas, que no son más que figuras compuestas por líneas cuya longitud, orientación, dirección o curvatura se percibe erróneamente (Guillam, 1980).

Son escasos los estudios experimentales realizados que examinen la relación que puede existir entre las distorsiones perceptivas y el uso de los gráficos estadísticos.

Poulton (1985) ha investigado ilusiones similares a la clásica de Poggendorff (Figura 2.9), mediante experimentos que sugieren que las relaciones de líneas inclinadas sobre los ejes vertical y horizontal de los gráficos pueden producir errores de lectura, que se incrementan a medida que aumenta la distancia a los ejes.

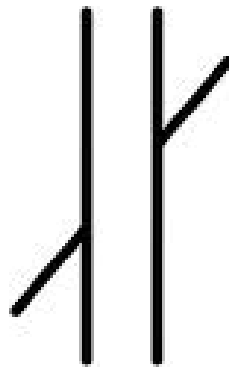


Figura 2.9: Ilusión de Poggendorff

Para reducir estas ilusiones geométricas propone que todos los gráficos muestren los cuatro ejes y que a su vez todos los ejes estén graduados.

Existen estudios que demuestran que la ilusión de la T invertida puede afectar sobre todo a los diagramas de barras realizados con segmentos o rectángulos de poca anchura. Una posible solución, como antes se ha comentado, es mostrar los cuatro ejes o bien realizar un entramado de fondo con la graduación de los mismos (figura 2.10).

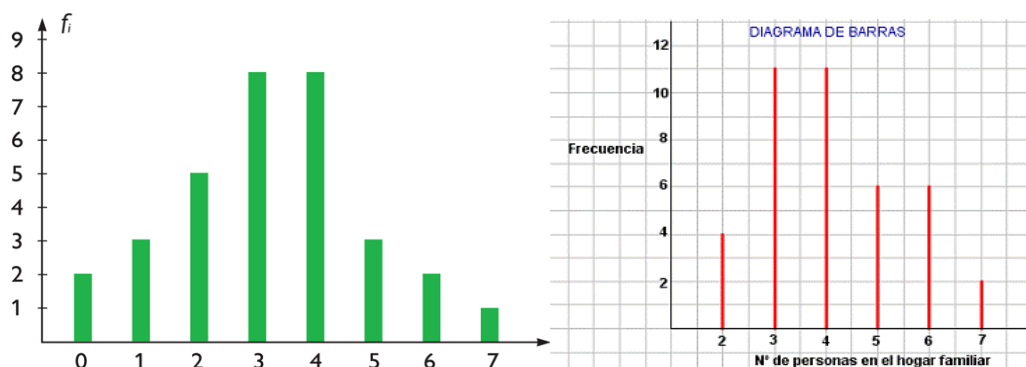


Figura 2.10: Diagramas de barras

Existen ilusiones geométricas en las que objetos iguales pueden parecer diferentes dependiendo de que los objetos circundantes sean grandes o pequeños. En la ilusión de Titchener (figura 2.11), los dos círculos interiores tienen el mismo tamaño pero parecen diferentes debido a los círculos que los rodean.

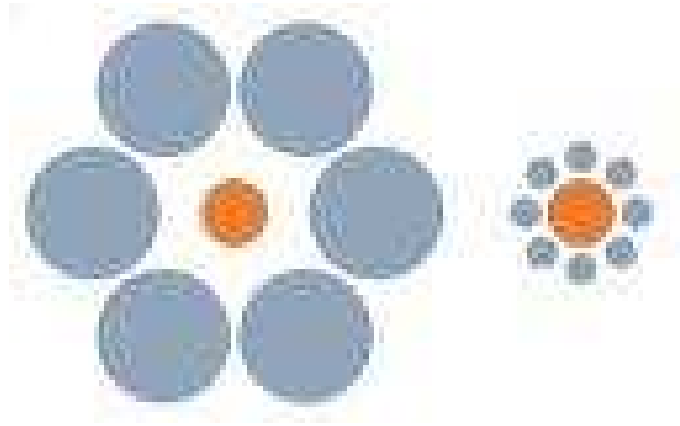


Figura 2.11: Ilusión geométrica de Titchener

Esta ilusión óptica puede afectar a aquellas representaciones gráficas que utilizan círculos de diferente tamaño como por ejemplo los diagramas de sectores o los mapas estadísticos.

En los gráficos estadísticos que se emplea la perspectiva, como por ejemplo en los histogramas en 3D, también se pueden crear ilusiones geométricas y ocasionar una percepción inadecuada de la información dando lugar a lecturas erróneas y no reales (figura 2.12). Como se puede observar, la información que intenta transmitir el siguiente gráfico no resulta clara para el lector, dando lugar a conclusiones erróneas.

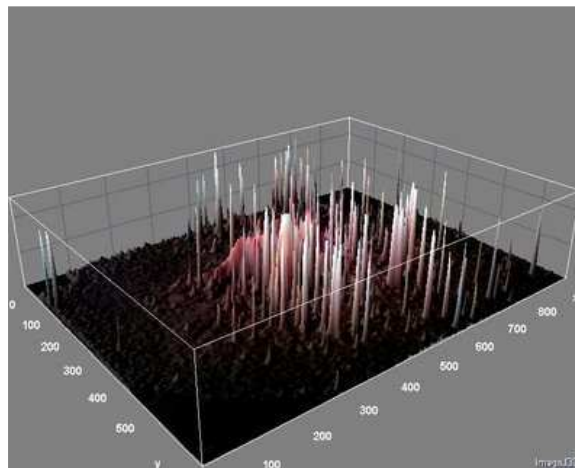


Figura 2.12: Histograma en 3D

2.3. Construcción de un gráfico estadístico

Toda manifestación gráfica depende del material empleado, pero tan importante como esto es la manera de utilizar y de presentar este mismo material. Cada persona utilizando idénticos instrumentos y material de trabajo imprime a su obra una serie de características que le son propias.

No existen muchos estudios realizados sobre cómo se deben realizar y construir los gráficos, y la mayoría carecen de fundamentación empírica. Entre ellos cabe citar los de Schmid (1983) y Bertin (1983) que presentan una gramática para la descripción de gráficos, según la cual, cada gráfico puede ser inequívocamente reducido desde la descripción basada en un pequeño número de elementos gramaticales y símbolos que recogen la descripción de cada gráfico.

McKinley (1987) incorpora en un programa de ordenador normas acerca de cómo se deben construir los gráficos, de forma que es capaz de generar gráficos expresivos y efectivos automáticamente.

A la hora de construir un gráfico, el investigador debe de reflexionar sobre el propósito del mismo y conocer los medios y tipos de métodos de representación de los que dispone. De este modo podrá alcanzar un resultado brillante y efectivo con los datos a estudio.

Una vez elegido el Método de representación adecuado, en función de las características de los datos y los medios con los que cuenta para alcanzar los objetivos planteados, es conveniente realizar pruebas (borrador) sobre la pantalla del ordenador, tanto al inicio del proceso como al finalizar el mismo; esto servirá para comprobar la correcta construcción del gráfico y la adecuada visualización de las características en el mismo (Ávila, 1993).

En el proceso de prueba deberán ser considerados todos los elementos que entran a formar parte de la construcción del gráfico y los datos con los que el gráfico ha sido construido. Una salida en papel de las pruebas realizadas consideradas óptimas, contribuirá a dar una idea de sí el objetivo buscado se ha conseguido. No es lo mismo una imagen sobre la pantalla que su impresión en papel.

Si el resultado es el buscado, y su construcción es correcta, es adecuado salvar el formato de dicho gráfico para posteriores representaciones gráficas.

2.3.1. Normas de representación

A menudo se confía en la intuición como guía cuando se ha de decidir si un gráfico es bueno o malo y no suele preocupar encontrarse ante un mal gráfico, ni de lo perjudicial que el mismo puede llegar a ser.

2.3.1.1. Normas básicas

Algunas de las reglas básicas para la representación gráfica son las siguientes:

- Búsqueda de la máxima fidelidad.
- Resultados claros y sencillos.
- Evitar carácter tendencioso.
- Emplear notaciones adecuadas.
- Resaltar los resultados más importantes.
- Realizar anotaciones aclaratorias.
- Identificar siempre el gráfico.

2.3.1.1.1. Diagramas cartesianos

Los diagramas cartesianos deberán ser construidos de acuerdo con las normas básicas de construcción gráfica ya señaladas, pero además, deberán cumplir criterios específicos:

- Los ejes coordenados: deberán ser debidamente identificados, indicando la variable que cada uno de ellos representa, así como las unidades o cualidades en que han sido medidas. Conviene en ocasiones dibujar los cuatro ejes (Poulton, 1985).
- El origen de coordenadas será especificado claramente con objeto de evitar un carácter tendencioso, sobre todo cuando el origen sea distinto del (0,0) (Avila, 1993).
- La escala apropiada de los ejes será escogida a partir del dominio de las variables y teniendo en cuenta el tamaño en que se construirá el gráfico.
- La división de los ejes será realizada de acuerdo con la escala escogida. Las marcas deberán ser claras y estar acompañadas de sus correspondientes valores; si éstos son numerosos, sólo deberán aparecer los más importantes.
- Uso de leyenda aclaratoria, dentro o al margen del gráfico, cuando en el mismo se representen por iconos o diseños distintos: varios sujetos, variables, etc.
- Pie explicativo e identificativo, en el margen inferior del gráfico y en la misma orientación que él, que identifique claramente el Método gráfico del que se trata y la correspondencia que representa, cuando ésta no esté especificada en el texto. En ocasiones, puede aparecer un título en la parte superior.

2.3.1.1.2. Gráficos Geométricos y Figurativos

Este tipo de gráficos han de ser construidos de acuerdo con las normas básicas de construcción gráfica ya señaladas. Pero además deben de cumplir los siguientes aspectos:

- Todas las figuras han de estar debidamente identificadas, bien con letras o números señalando el sujeto que cada una de ellas representa.
- La escala ha de ser la misma para todas las figuras.
- La asignación de variables, debidamente explicada en el texto, ha de realizarse siempre en el mismo orden y con las mismas variables para todos los individuos.
- Uso de leyenda aclaratoria, al margen del gráfico, cuando el mismo lo requiera.
- Pie explicativo e identificativo, en el margen inferior del gráfico, y en la misma orientación que él, que identifique claramente el Método gráfico del que se trata.

2.3.1.1.3. Elementos de construcción.

A la hora de construir un gráfico además de valorar las normas, se debe tener en cuenta los elementos de construcción que son: el entramado, el color, los tipos de diseño y el rotulado.

- Entramado: las tramas de un gráfico deben escogerse cuidadosamente, sobre todo con representaciones monocromáticas. Su función es la de distinguir o señalar aspectos del mismo y no distraer la atención de una parte en relación a otra.

No existen más normas que las del sentido común y los propios gustos a la hora de escoger un entramado. Debe prestarse gran atención a que los contrastes tengan fuerza suficiente; debe mantenerse un cierto equilibrio entre tonos claros y oscuros y evitar cualquier inicio de monotonía, de gris uniformizador. Es aconsejable realizar algunas pruebas con tramas diferentes, en general no se debe dar por buena la salida que, por defecto, ofrece el ordenador.

- Tipos de diseño: en ocasiones se utilizan diferentes tipos de símbolos para diferenciar los valores que toman los distintos individuos respecto de cada variable; éstos deben escogerse de tamaño similar, pero de formas diferentes.

Tampoco existen normas específicas para su elección. Quizás una buena norma sea elegir diseños discretos pero diferenciadores. No se deben utilizar diseños llamativos que centren toda la atención hacia ellos, al menos que sea su finalidad en busca de un objetivo concreto.

- **Color:** los colores resultan de gran utilidad en la construcción de representaciones gráficas, ya que no solo contribuyen a magnificarlos y embellecerlos sino que, al contrario que las tonalidades de grises o los entramados y diseños, no ahogan los detalles, los números, las etiquetas, ni el entramado. No obstante debe tratarse con cuidado, ya que los colores muy llamativos pueden ocasionar un efecto contraproducente. Los colores claros son más adecuados para resaltar sin recargar la imagen.
- **Rotulado:** a los nombres se debe una parte importante de la información presentada por el gráfico, y son de gran interés pues determinan hasta cierto punto la apariencia del mismo (Avila, 1993).

Una vez más no existen más normas que las citadas por los sentidos común y estético. En la medida de lo posible, el rotulado se colocará horizontalmente al gráfico para facilitar su lectura, siempre adjunto a él y no en otra página, aunque lo suficientemente separado como para apreciar directamente que se trata de un texto complementario.

El rotulado inserto en el gráfico debe colocarse de forma equilibrada en la consecución de una composición armónica en relación al trazado del gráfico en conjunto.

- Es imprescindible asegurar que, además de la composición equilibrada y armónica, el rotulado se encuentre claramente asignado a la característica y función para la que fue creado.
- Si se requiere, se alterará la posición del “etiquetado” para evitar superposiciones y aglomeraciones de letras o palabras y se optará por una identificación al margen del gráfico por medio de tramas, símbolos, letras etc. Éstas deberán ser claramente identificadas en la leyenda o en el texto explicativo situado bajo el gráfico. El tipo de letras, su estilo y tamaño son elección del creador del gráfico, es bueno elegir letras simples y no complicadas ni elaboradas.

2.4. Representaciones gráficas

La visión es la modalidad sensorial dominante del ser humano. Nuestro cerebro está altamente capacitado para el manejo de información visual, siendo capaz de reconocer y procesar imágenes gráficas con una simple inspección ocular.

Está comúnmente aceptado por la comunidad científica que, en general, una representación gráfica proporciona mayor información acerca de las características y patrones de los datos, que un texto o una presentación tabular de los mismos.

La información que proporcionan las tablas es muy completa, pero tiene como factor negativo que la lectura requiere un cierto tiempo y un cierto grado de entrenamiento. Está probado que los lectores al comenzar a leer un determinado artículo, da igual que sea científico o no, dirigen su vista en el siguiente orden: título, gráficos y, finalmente, a las tablas.

Los gráficos estadísticos constituyen una de las principales y más sencillas herramientas para exponer la información ya que tienen la suficiente capacidad de impactar en el lector con muy poco esfuerzo por su parte. Dan información global y rápida de los datos, son útiles incluso al investigador ya que permiten tener una idea general de los resultados, pudiendo sugerir nuevas hipótesis.

2.4.1. Gráficos estadísticos más usuales

Las representaciones gráficas que aparecen en las publicaciones científicas tienen generalmente un carácter descriptivo y se pueden clasificar en univariantes, bivariantes y multivariantes, según el número de variables que intervengan en el estudio.

En este trabajo, se prestará más atención a los gráficos uni y bivariantes porque son los que más aparecen en las publicaciones científicas de Ciencias de la Documentación. Sin embargo, se realizará una breve exposición de los gráficos multivariantes porque son muy utilizados en otras disciplinas y se estima que su uso crecerá en un futuro inmediato en el campo ya citado, tan pronto como se divulguen.

Esta forma de clasificarlo no es única pero se cree que es la más apropiada para el fin que persigue el presente estudio.

2.4.1.1. Gráficos univariantes/bivariantes

Los **Métodos Gráficos Univariantes** sólo exigen efectuar una transcripción geométrica de los datos correspondientes a una variable, en una representación gráfica.

2.4.1.1.1. Diagrama de barras

En un sistema de ejes de coordenadas cartesianas se representan en el eje de abscisas los valores de la variable. Posteriormente, sobre cada valor que ha tomado la variable se construye una barra vertical en la que la altura es proporcional a la frecuencia, ya sea ésta relativa o absoluta. Sobre el eje de

abscisas la escala de medida puede ser cualquiera y no coincidir con la escala del eje de ordenadas. Incluso el cero del eje de abscisas no tiene porque coincidir con el cero de la medida utilizada (figura 2.13).

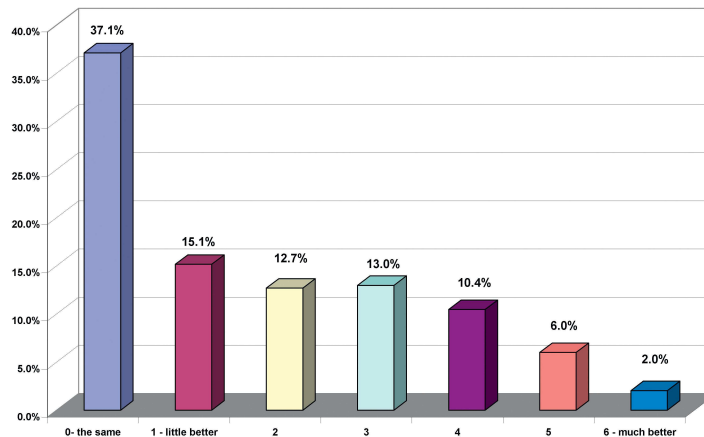


Figura 2.13: Diagrama de barras

Si lo que se quiere representar son las frecuencias acumuladas se hace exactamente igual, con los ejes cartesianos y levantando sobre cada valor de la variable una altura igual a la frecuencia acumulada, se unen mediante trazos horizontales el extremo de cada coordenada. Este diagrama recibe el nombre de diagrama de escalera (figura 2.14).

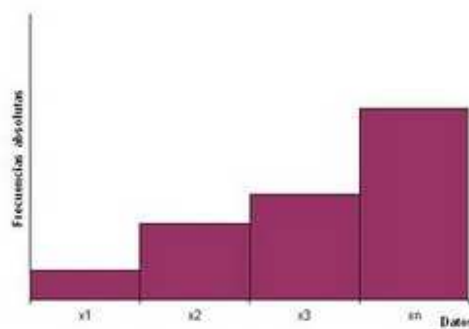


Figura 2.14: Diagrama de escalera

2.4.1.1.2. Diagrama de sectores

El diagrama de sectores sólo se puede utilizar en los casos que se quiera representar frecuencias absolutas o relativas, nunca para frecuencias acumuladas. En este gráfico se representan, en una circunferencia, sectores circulares cuyo ángulo central coincide con la frecuencia absoluta o relativa del elemento representando, mediante colores o incluyendo en cada sector el nombre del elemento a representar (figura 2.15). Puede utilizarse tanto para frecuencias agrupadas como no agrupadas. Previamente hay que calcular los

grados que corresponden a cada elemento multiplicando la frecuencia correspondiente a cada dato por el cociente entre 360° y el total de datos:

$$g_i = f_i \frac{360^\circ}{n}$$

Cuando en un diagrama de sectores ciertos datos presentan una frecuencia baja, en relación con los demás, su sector circular sería no detectable visualmente. Por ello, se une con otras frecuencias también bajas dándole el nombre de “Otros” o bien, si es posible, indicando todos los elementos que lo conforman.

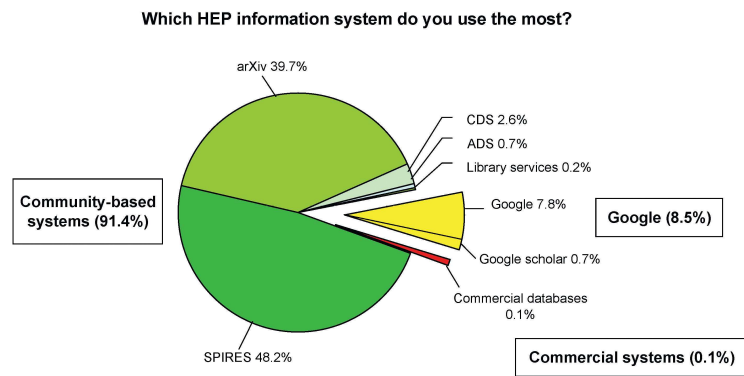


Figura 2.15: Diagrama de sectores

2.4.1.1.3. Polígonos de frecuencias

Sobre unos ejes cartesianos, análogos a los que conforman un diagrama de barras, se levanta una ordenada de altura idéntica a la frecuencia absoluta (o relativa) de dicho valor, uniendo posteriormente con una línea poligonal dichas ordenadas. La primera ordenada se une con el cero del eje de abscisas, teniendo en cuenta que si existe algún valor de la variable con frecuencia cero también ha de ser considerado y unido a dichos datos (figura 2.16). Igualmente se procedería con las frecuencias acumuladas.

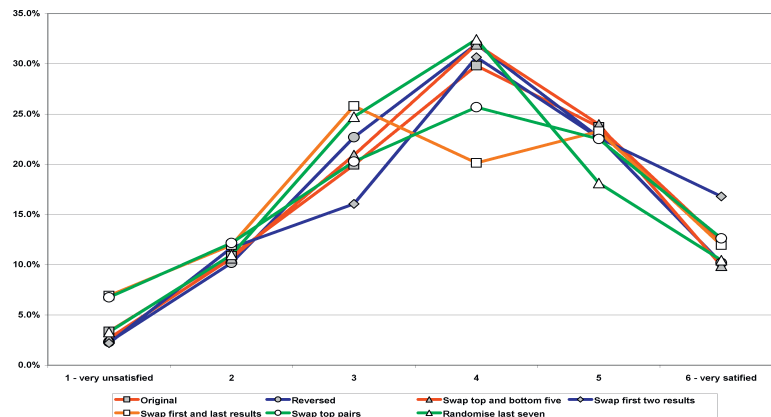


Figura 2.16: Polígono de frecuencias

2.4.1.1.4. Histograma para distribuciones de frecuencias agrupadas en intervalos

Este gráfico se construye levantando, sobre cada intervalo de la variable, un rectángulo de área proporcional a la frecuencia absoluta de dicho intervalo. En el caso de que los intervalos sean de amplitud constante, las alturas de los rectángulos serán iguales a las frecuencias absolutas respectivas, pues al ser las bases iguales las áreas son proporcionales a las alturas. En el caso de que las amplitudes de los intervalos sean diferentes, las alturas de los rectángulos deben calcularse dividiendo la frecuencia absoluta por la longitud del intervalo; ésta se puede representar por a_i :

$$a_i = \frac{f_i}{c_i}$$

y de esta forma, el área del rectángulo coincide con la frecuencia:

$$S_i = a_i c_i = \frac{f_i}{c_i} c_i = f_i$$

La altura a_i correspondería a la frecuencia correspondiente a cada unidad de medida de la variable en cada intervalo. Este gráfico se conoce también y con el nombre de densidad de frecuencia del intervalo (figura 2.17).

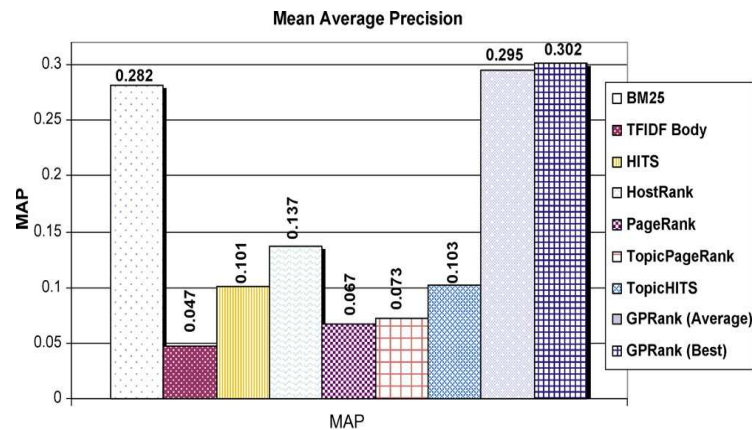


Figura 2.17: Histograma

2.4.1.1.5. Polígono de frecuencias (datos agrupados)

Si lo que se representan son las frecuencias no acumuladas, es decir datos agrupados, basta identificar el punto central del segmento superior de cada uno de los rectángulos que conforman el histograma y dibujar la línea poligonal abierta que pasa por los puntos.

2.4.1.1.6. Box-Plot

Esta representación gráfica, debida a Tukey (1977), puede ser traducida como “Caja con Bigotes” o “Representación Caja”, aunque se conoce usualmente con el nombre de **Box-plot** o bien **Box and Whiskers plot**.

Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos: mediana, primer cuartil, tercer cuartil, valor máximo y valor mínimo. Esta presentación visual, asocia las cinco medidas que suelen trabajarse de forma individual. Presenta información de la tendencia central, dispersión y simetría de los datos de estudio.

Por su facilidad de interpretación y construcción permite comparar a la vez varios grupos de datos sin perder información ni saturarse de ella.

El gráfico consiste en un rectángulo dispuesto verticalmente, que “encierra” el recorrido intercuartílico, de cuyo lado superior e inferior se derivan respectivamente, dos segmentos: uno hacia arriba y otro hacia abajo (ver figura 2.18a y 2.18b).

La caja y sus bigotes están ubicados paralelos a un eje rotulado. Sus partes son las siguientes:

- Límite superior o Noveno Decil: extremos superiores del bigote. Los datos por encima del límite superior se consideran atípicos o outliers.
- Tercer cuartil (Q.3): por debajo se encuentran el 75% de los datos.
- Mediana: coincide con el segundo cuartil (Q.2). Divide a la distribución en dos partes iguales. De este modo el 50% de los datos están por encima de la mediana y el otro 50% de los datos están por debajo de la mediana.
- Primer cuartil (Q.1): Por debajo de este cuartil se encuentran el 25% de los datos del estudio.
- Límite inferior ó Primer Decil: Es el extremo inferior del bigote. Los datos por debajo de este límite se consideran atípicos u outliers.
- Valores atípicos: Datos que están alejados del cuerpo principal. Pueden representar efectos de causas extrañas y pueden llegar a distorsionar la muestra.
- Media aritmética: es lo que tradicionalmente se conoce como promedio. Aunque la mayoría de los programas estadísticos no la representan en el Box Plot, puede representarse.

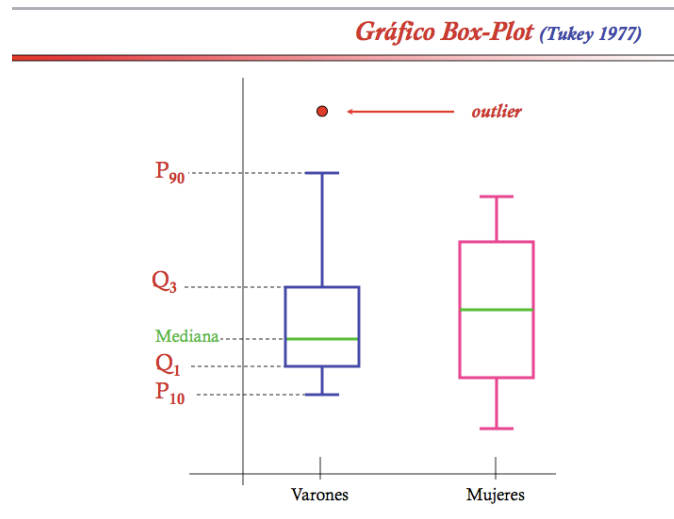


Figura 2.18a: Box-Plot

Este gráfico es muy conocido en Ciencias de la Documentación por ser un método simple de resumir la información. Proporciona una rápida impresión de las características más importantes de una distribución. Permite visualizar la dispersión de los factores de impacto en una categoría temática concreta.

Como ejemplo se representa un box-plot para la revista "Statistics in Medicine", datos de 2008, la cual tiene un índice de impacto (resultado de dividir el número de citas recibidas en los dos años inmediatamente anteriores entre el número de artículos publicados en esos años) de 2.111.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
MATHEMATICAL & COMPUTATIONAL BIOLOGY	28	11	Q2
MEDICAL INFORMATICS	20	6	Q2
MEDICINE, RESEARCH & EXPERIMENTAL	82	43	Q3
PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH	105	42	Q2
STATISTICS & PROBABILITY	92	9	Q1

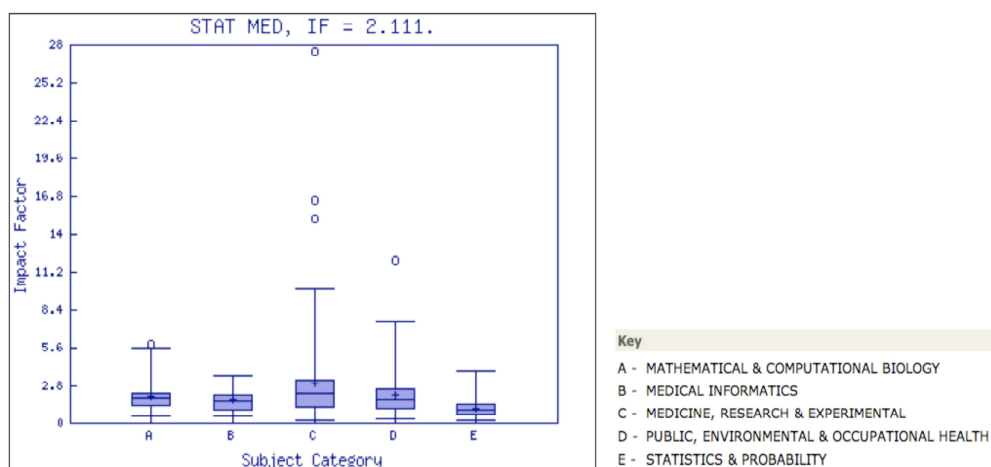


Figura 2.18b: Box Plot para las distintas categorías en las que la revista está considerada

representar la información del hiperespacio de partida en un subespacio de dimensiones reducidas (Componentes Principales, BIPLLOT, Multidimensional Scaling, Dendogramas, etc.).

La práctica totalidad de los Métodos Gráficos Multivariantes que se conocen no son otra cosa que “constructos pictóricos”, (también conocidos como gráficos pictoriales, figurativos o simplemente “Iconos”), formados por elementos geométricos (puntos, rectas, curvas, triángulos, círculos...) que varían en función de los valores que toman las variables en los individuos a los que representan.

Cuando los datos tienen más de dos o tres dimensiones, la representación gráfica se hace complicada: las dimensiones del plano no son capaces de acoger un mayor número de variables que los que acoge la representación cartesiana convencional o la tridimensional; por lo tanto se requiere otro tipo de representación (Vicente-Villardón et al., 1990).

Existen diferentes métodos para representar datos multivariantes, prácticamente tantos como autores se han ocupado del tema. No se realizará por ello una revisión exhaustiva ni una descripción detallada de los MGM; solamente se enumerarán algunas de las técnicas existentes en la actualidad. Únicamente se expondrán con cierto detalle las más importantes y actuales que permiten la representación gráfica de entidades definidas por varias variables. En este grupo se pueden incluir por ejemplo las Curvas de Andrews (Andrews, 1972), las Caras de Chernoff (Chernoff, 1973), etc.

2.4.1.2.1. Curvas de Andrews

Andrews (1972) propone una técnica muy simple conocida como “Curvas de Andrews” para obtener una representación visual de datos multivariantes, donde cada punto es representado en una suma de funciones trigonométricas.

Este método consiste en representar a cada individuo que presenta los valores muestrales (x_1, \dots, x_k) de las variables consideradas, mediante un grafo de la función periódica:

$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t)$$

para valores de t comprendidos entre $-\pi < t < \pi$.

Este tipo de funciones son fácilmente trazables mediante un dispositivo gráfico y un sencillo programa, obteniéndose, para los individuos considerados, series de curvas que permiten la interpretación posterior (ver figura 2.20).

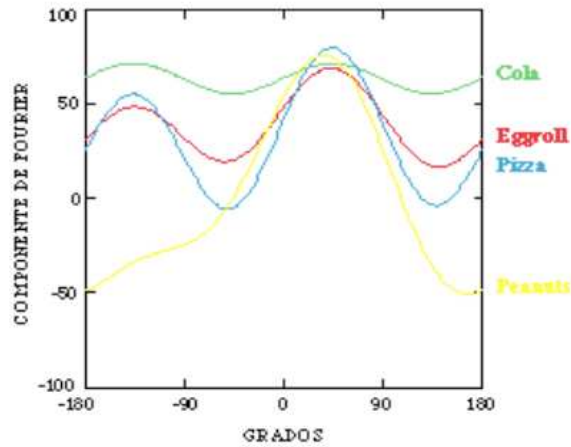


Figura 2.20: Curvas de Andrews

La conservación de las distancias euclídeas originales para las curvas permite afirmar que dos curvas similares a lo largo del recorrido corresponden a individuos con perfiles similares.

2.4.1.2.2. Diagrama de estrellas

Determinan perfiles configurados por segmentos que parten de un origen común y cuya longitud corresponde al valor que, para cada entidad, toma la variable a la cual dicho segmento representa (ver figura 2. 21).

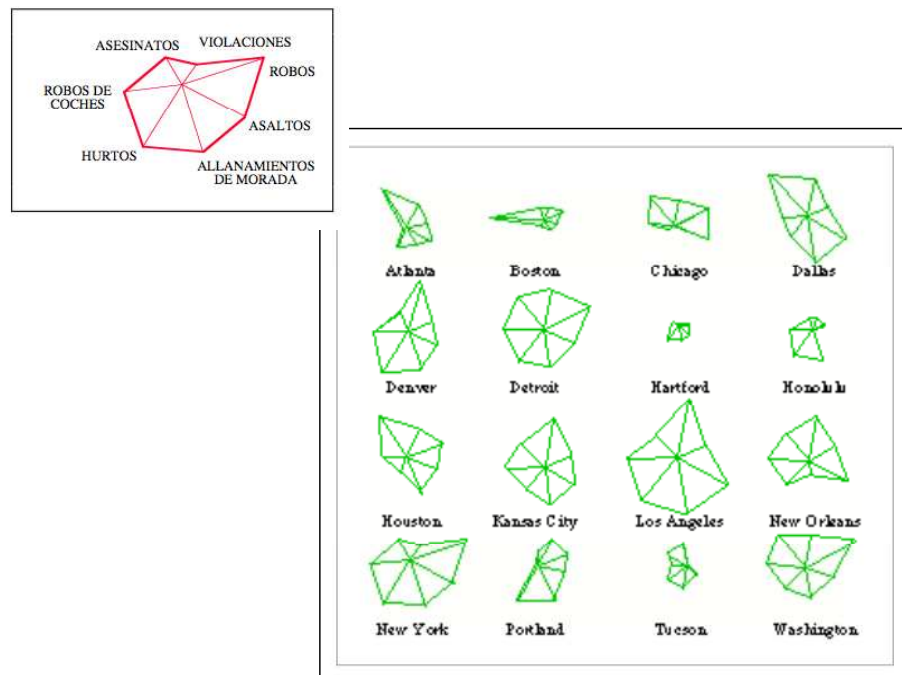


Figura 2.21: Diagramas de Estrella que representan los datos de criminalidad de diversas ciudades de Estados Unidos. (Datos originales de Everitt, 1993). Arriba izquierda, diagrama de estrella para la ciudad de Nueva York que muestra la asignación de las variables a cada segmento

2.4.1.2.3. Caras de Chernoff

Esta representación gráfica consiste en asignar un rasgo facial a cada una de las variables sobre una cara esquematizada. Este procedimiento gráfico, que requiere la utilización de un programa de ordenador para ser realizado, concentra la información de cada individuo en un rostro humano caricaturizado, cuyos rasgos reflejan cada una de las variables o "descriptores".

El método intenta aprovechar la capacidad humana para reconocer y diferenciar los rostros. Naturalmente la asignación de los rasgos a las variables puede ser arbitraria por lo que pueden obtenerse diferentes representaciones diferentes para el mismo conjunto de datos. La única limitación es el número de variables que se pueden utilizar (aproximadamente 20) aunque utilizando caras asimétricas es posible doblarlo. Figuras: 2.22a y 2.22b.

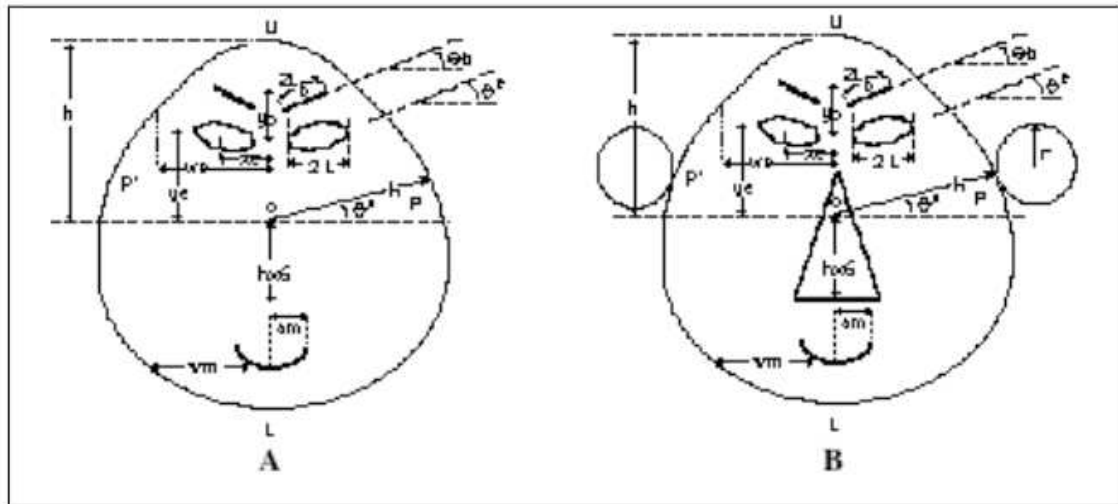


Figura 2.22a: Parámetros de los Rasgos faciales de las Caras de Chernoff, según Bruckner (1978). A.- Caras originales de CHERNOFF (1973). B.- Variante de Bruckner, (1978), que incluye nariz y oreja

A continuación se presenta la figura 2.22b que recoge un ejemplo de aplicación de Caras de Chernoff.

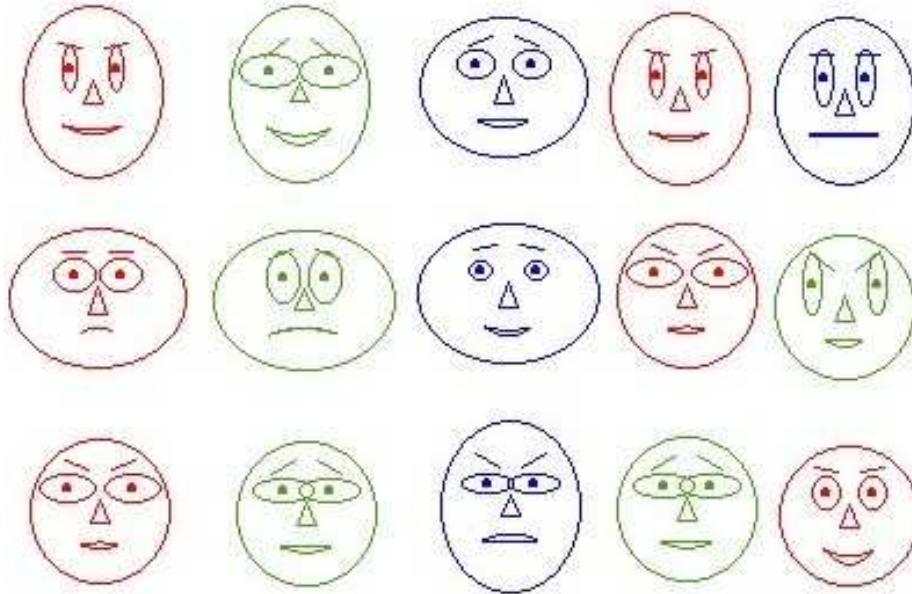


Figura 2.22b: Caras de Chernoff

Con este método resulta sencillo efectuar agrupaciones por similitudes de rasgos.

2.4.1.2.4. Pictogramas

Son dibujos que aluden a la distribución que se pretende estudiar y que mediante su tamaño, forma y configuración ofrecen una descripción, lo más característica posible. En cualquier caso, la superficie de la figura tiene que ser proporcional a la frecuencia, aunque existe también la posibilidad de que una figura represente un número determinado de frecuencias y entonces contenga este dato (ver figura 2.20). Este tipo de representaciones suelen utilizarse en las distribuciones cualitativas.

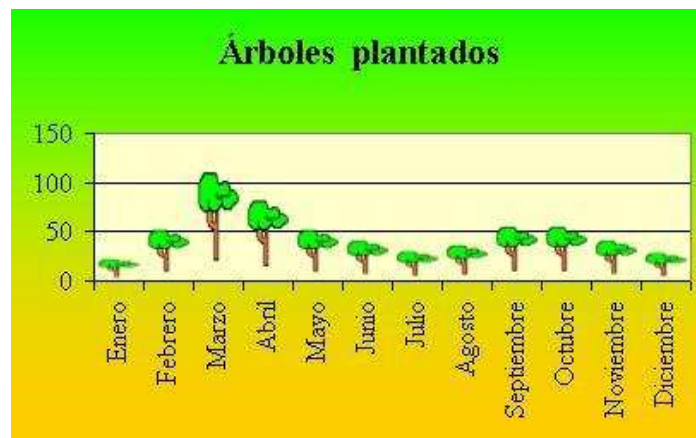


Figura 2.23: Pictograma

2.4.1.2.5. Cartogramas

Son las representaciones gráficas realizadas sobre mapas. Muestran el atributo estudiado en ciertas regiones, señalando las zonas con diferentes tramas o colores y poniendo de manifiesto las diferencias existentes entre las regiones del plano (ver figura 2.24). Estos gráficos se suelen utilizar para representaciones sobre las densidades demográficas de una nación, tipos de cultivos, renta per cápita, índices de lluvia, etc.

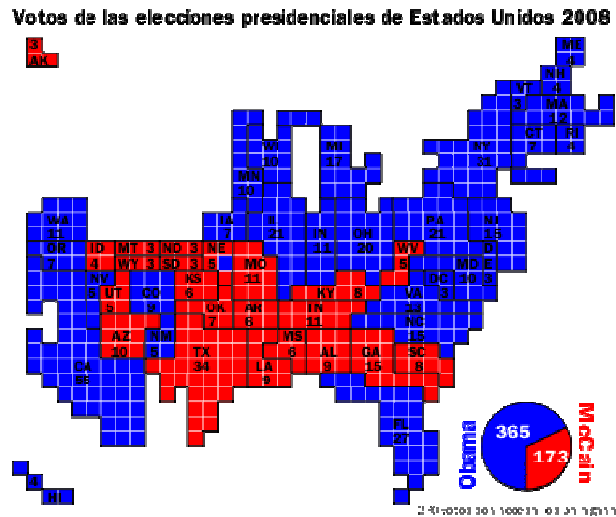


Figura 2.24: Cartograma

2.4.1.2.6. Diagrama de perfil radial

En este gráfico se trazan tantos radios como modalidades tenga la variable estudiada y después, sobre estos radios, se toma una distancia al centro proporcional a la frecuencia de cada modalidad. Posteriormente se unen los puntos extremos de cada radio y se obtiene un polígono cerrado, que es el perfil radial (figura 2.25).

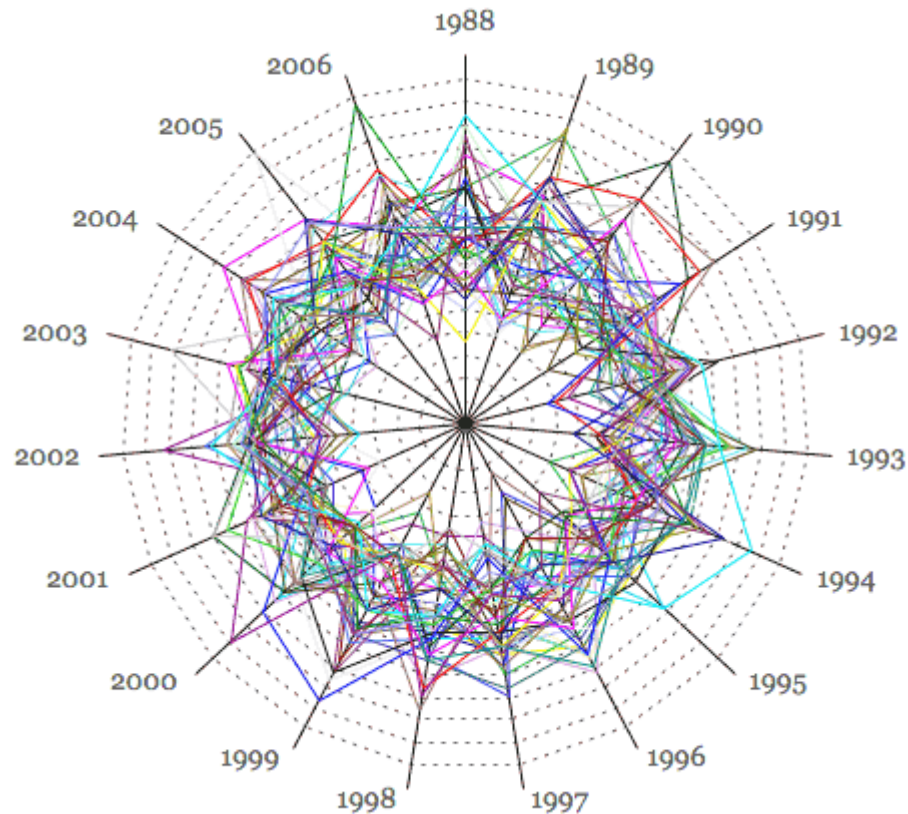


Figura 2.25: Diagrama de perfil radial

2.4.1.2.7. Grafos o Redes

Un tipo de gráficos muy utilizado en estudios de coautoría (entre otros) son las redes. Las redes tienen una base matemática que es la teoría de Grafos.

Un grafo, G , es un par ordenado *de vértices y aristas* $G = (V, A)$, donde V es el conjunto de vértices o nodos del grafo y A es un conjunto de aristas (las líneas que unen los nodos de un grafo y con la que se constituyen también caminos). Recorrer un grafo significa tratar de alcanzar todos los nodos que estén relacionados con uno que le asignara el nombre “nodo de salida” (ver figura 2.26).

Un estudio detallado puede encontrarse en Nooy, Mrvar y Batagelj (2005) y un ejemplo tomado del informe del “Proyecto EA 2006-0024. Programa de Estudios y Análisis Secretaría de Estado de Universidades e Investigación. Ministerio de Educación y Ciencia” (página 125), aparece en la gráfica siguiente.

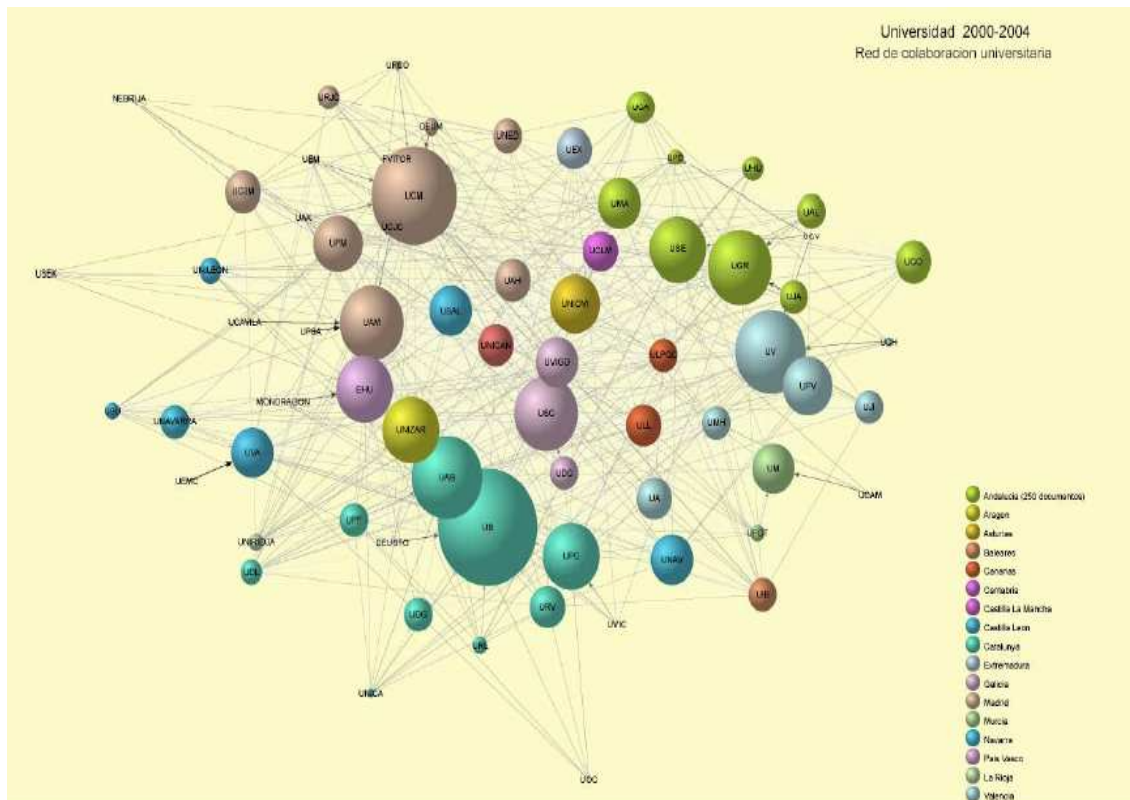


Figura 2.26: Redes de Grafos

2.4.1.2.8. Coordenadas paralelas

Uno de los métodos de representación gráfica de datos multivariantes más reciente es la representación en coordenadas paralelas. Se trata de un sistema de visualización que permite representar información multivariante en un plano bidimensional (ver figuras 2.27a, 2.27b, 2.27c). Cada eje representa una variable que puede ser continua o categórica.

La idea de la representación se recoge en la figura siguiente:

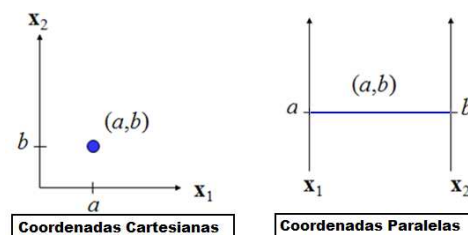


Figura 2.27a: Criterio de representación en coordenadas paralelas

Un individuo que tomara los valores (5, -5, 10, 15, 5, -10) quedaría representado con un perfil de la siguiente forma:

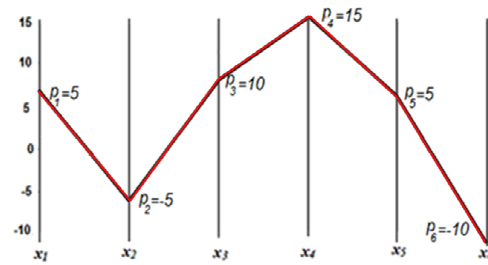


Figura 2.27b: Perfil de un sujeto en coordenadas paralelas

Si en lugar de un individuo se representaran muchos, el aspecto del gráfico en coordenadas paralelas sería el siguiente, en el cual lo que se buscan son patrones que nos permitan seleccionar sujetos con perfiles similares.

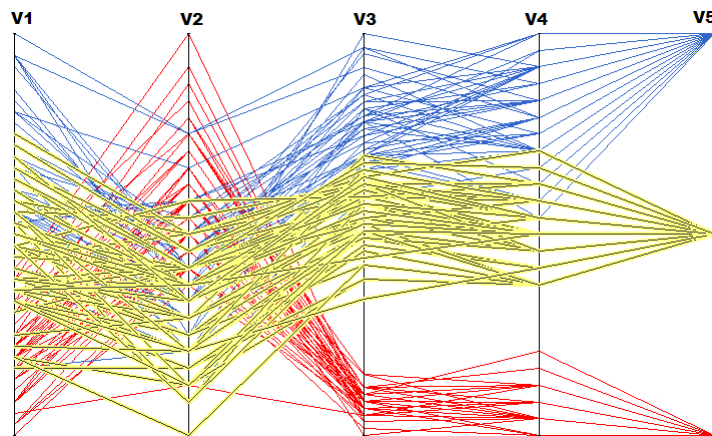


Figura 2.27c: Visualización de la información multivariante de varios individuos