



UNIVERSIDAD
COMPLUTENSE
MADRID



Proyectos de Innovación y Mejora de la Calidad Docente Vicerrectorado de Evaluación de la Calidad

Convocatoria 2015, Proyecto núm. 164

«Videotutoriales de estadística aplicada a las Ciencias Sociales: un recurso formativo emergente en las actividades de enseñanza-aprendizaje»



Departamento de SOCIOLOGIA IV

Métodos y técnicas de investigación Aplicada y de Mercados y Teoría de la Comunicación



Proyectos de Innovación y Mejora de la Calidad Docente
Vicerrectorado de Evaluación de la Calidad

ESTADÍSTICA DESCRIPTIVA UNIVARIANTE MEDIANTE EL GRÁFICO DE CAJA Y BIGOTES

DESCRIPTIVE AND UNIVARIATE STATISTICS USING BOX-AND-WHISKER PLOT

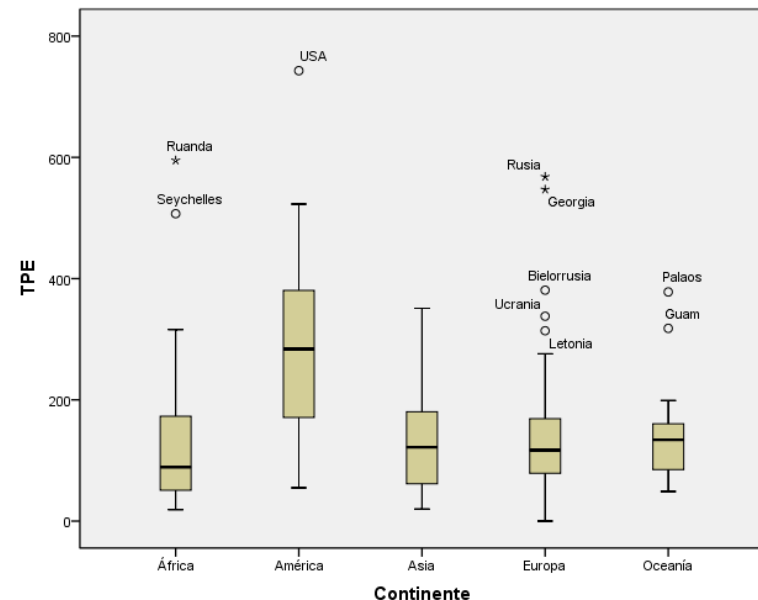
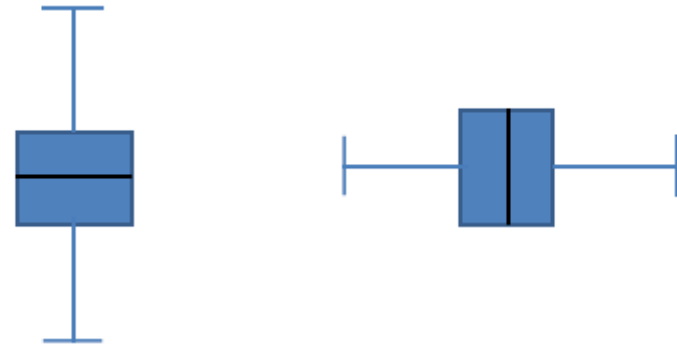
Esmeralda Ballesteros Doncel (Dto. Sociología IV – UCM)

En el gráfico de caja ¡TODO SON VENTAJAS!

El gráfico de caja y bigotes (en adelante gráfico de caja) se emplea como representación gráfica de variables cuantitativas.

Permite **resumir, describir y analizar** aspectos generales y particulares del indicador. En él quedan ilustrados los datos centrales, datos adyacentes y datos raros (atípicos y extremos, si los hubiera).

Su preferencia se debe a que es, simultáneamente, una herramienta sencilla y rigurosa de exploración-análisis de una distribución cuantitativa y, porque, además, permite establecer, en el mismo gráfico, **comparaciones entre subgrupos**.





Áreas de descripción estadística de una variable cuantitativa

| Tendencia central = RESUMIR | Posición no central = CATEGORIZAR |
|---|---|
| <ul style="list-style-type: none">○ Moda○ Mediana○ Media aritmética | <ul style="list-style-type: none">○ Cuartiles○ Deciles○ Percentiles |
| Dispersión = VARIABILIDAD | Forma = SIMETRÍA y CONCENTRACIÓN |
| <ul style="list-style-type: none">○ Varianza○ Desviación típica | <ul style="list-style-type: none">○ Asimetría○ Curtosis |



POSICIONES CUARTILES

Ejemplo, con $N = 25$

ORDENAR LA DISTRIBUCIÓN (menor a mayor o, de mayor a menor)

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

CREAR SEGMENTOS

MÍNIMO

MÁXIMO

50 % de unidades centrales

| | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------------------------------|--|--|--|--|--|----|--|--|--|--|--|----|--|--|--|--|----|--|--|--|--|--|--|--|
| | | | | | | Q1 | | | | | | Me | | | | | Q3 | | | | | | | |
| 25 % unidades con valores más bajos | | | | | | 7 | | | | | | 13 | | | | | 19 | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | |



CUARTIL 1

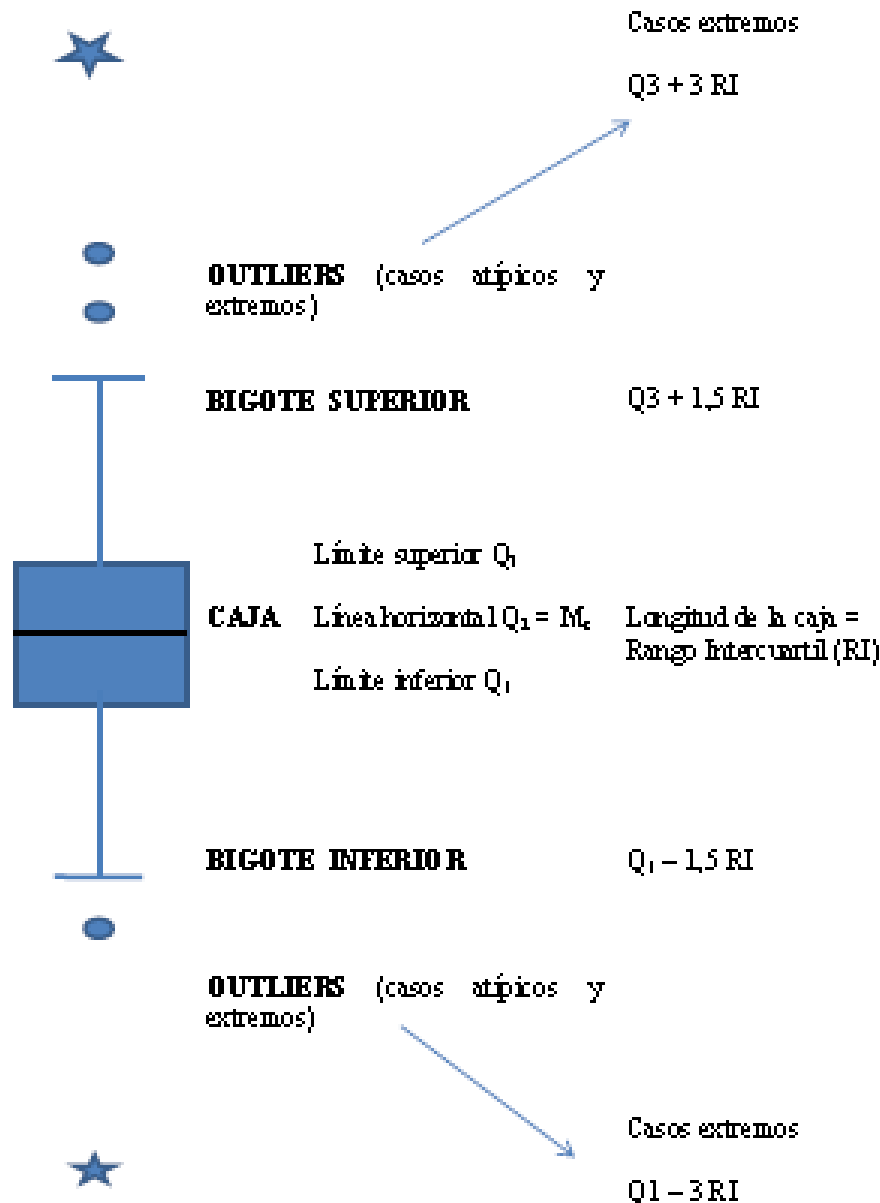
MEDIANA



CUARTIL 3



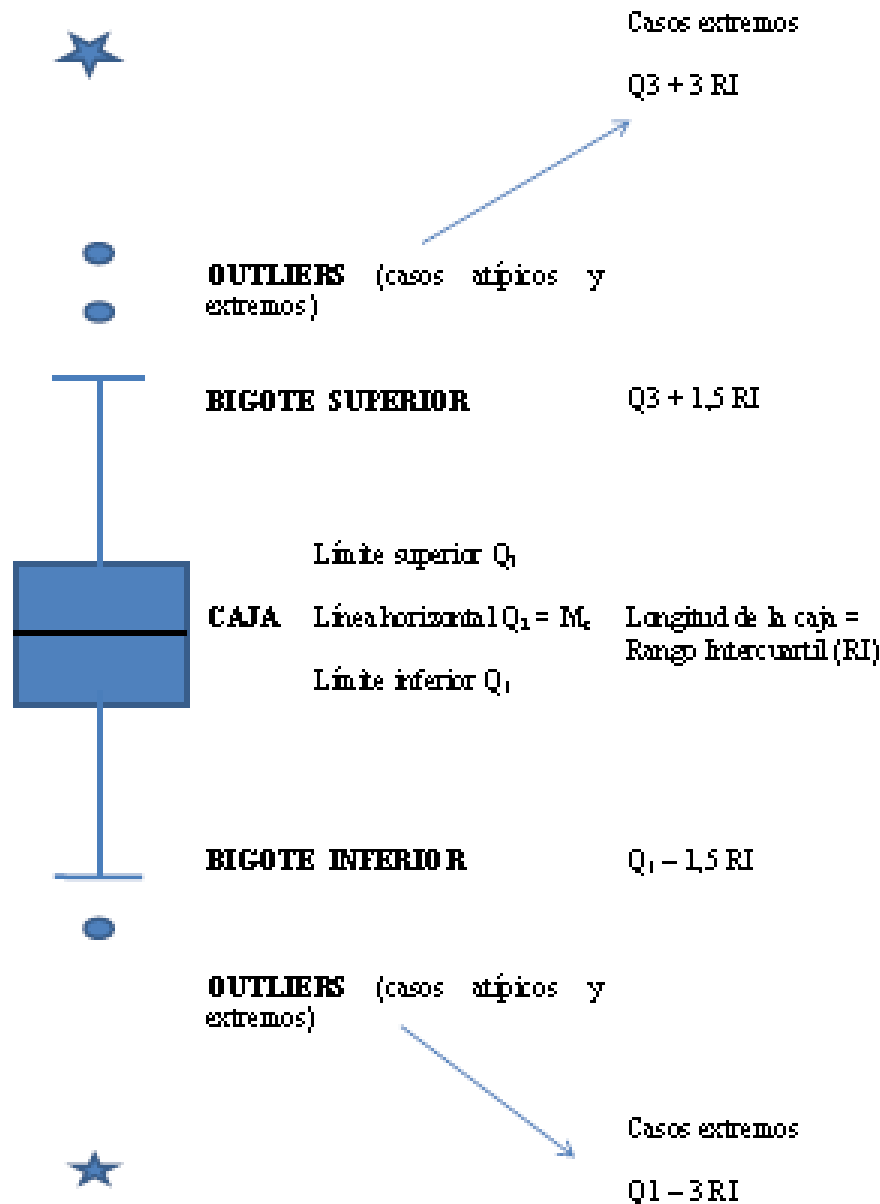
Elementos de un gráfico de caja



El gráfico de caja se construye en base a cinco medidas estadísticas: el valor mínimo, el valor máximo, la mediana, el primer cuartil y el tercer cuartil de los datos.

Morfología del gráfico de caja en caso de analizar una variable NORMALMENTE distribuida.

La caja



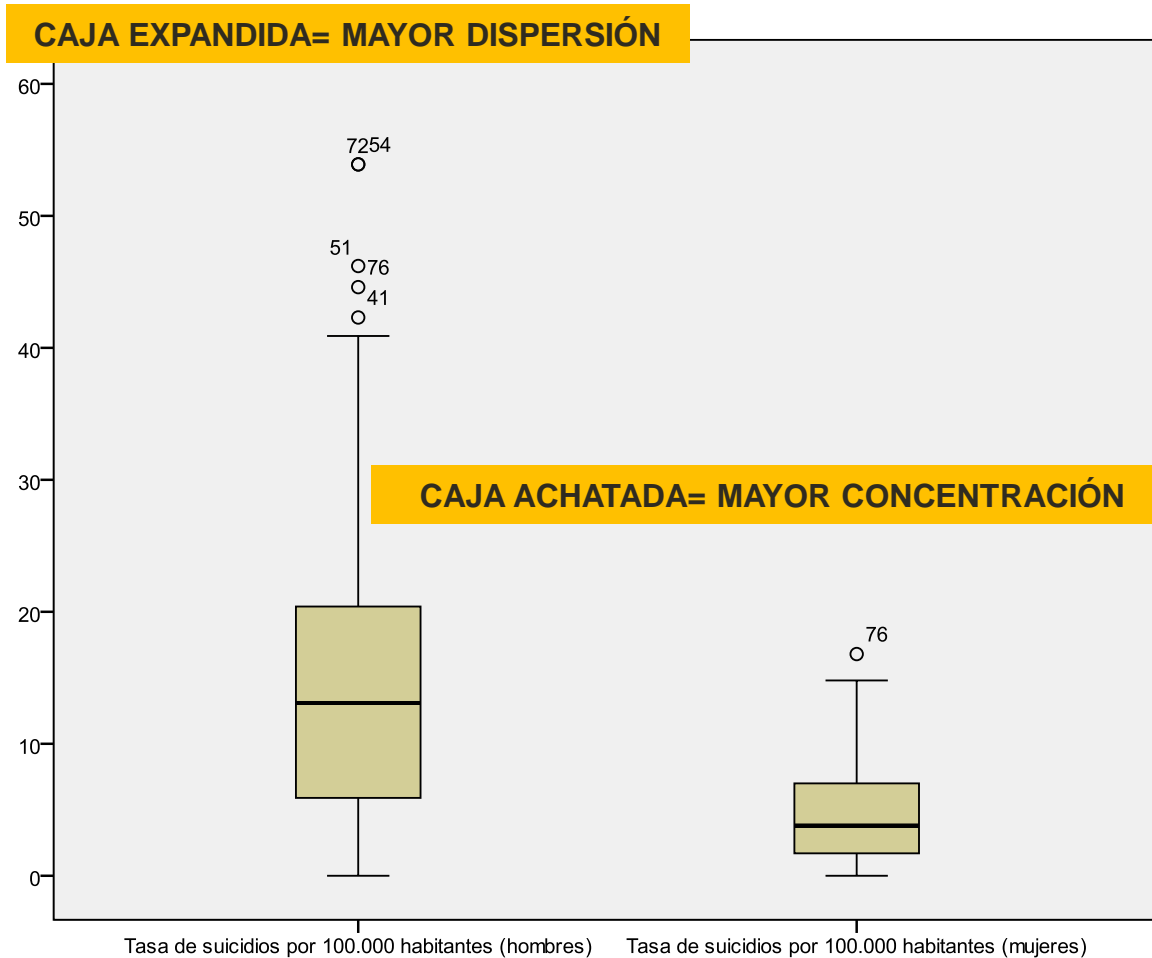
Lo más llamativo del gráfico es la «caja», esta se define a partir del **rango intercuartílico** ($Q_3 - Q_1$). De manera que el cuadrado/rectángulo contiene el 50 por ciento de los valores centrales.

La forma de la caja puede ser:

- a) Cuadrada
- b) Rectangular –achatada
- c) Rectangular -expandida

Las distintas morfologías describe si, el 50 por ciento de los valores centrales, se distribuyen de forma normal (a) o, por el contrario sus valores se concentran (b) o, se dispersan (c)

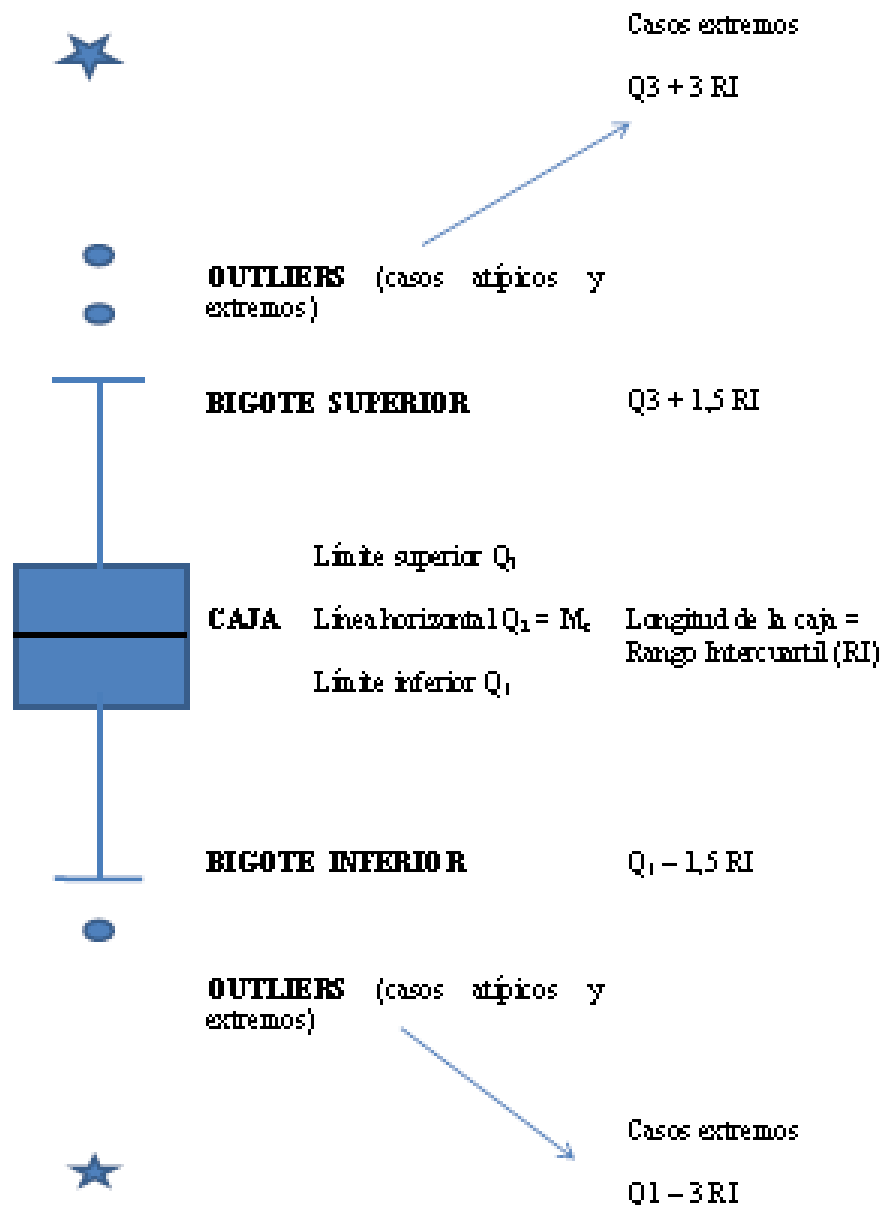
Morfología de la caja



En este gráfico de caja, en el que se analiza la tasa bruta de suicidio en el mundo, según el sexo de las personas, se aprecia que **“los hombres se suicidan mas que las mujeres”** (Ver valor diferencial de sus respectivas medianas, 13,10 frente a 3,8).

La diferente variabilidad de las distribuciones se expresa a través de la forma de las cajas. La mayor dispersión en la distribución de la tasa de suicidios en hombres, se ilustra con la expansión de la caja (distribución platicurtica) que contrasta con la mayor homogeneidad de las mujeres.

FUENTE: OMS, *Tasa bruta de suicidio por 100.000 habitantes* (2009), N= 86, tomado de (Ballesteros, 2013).

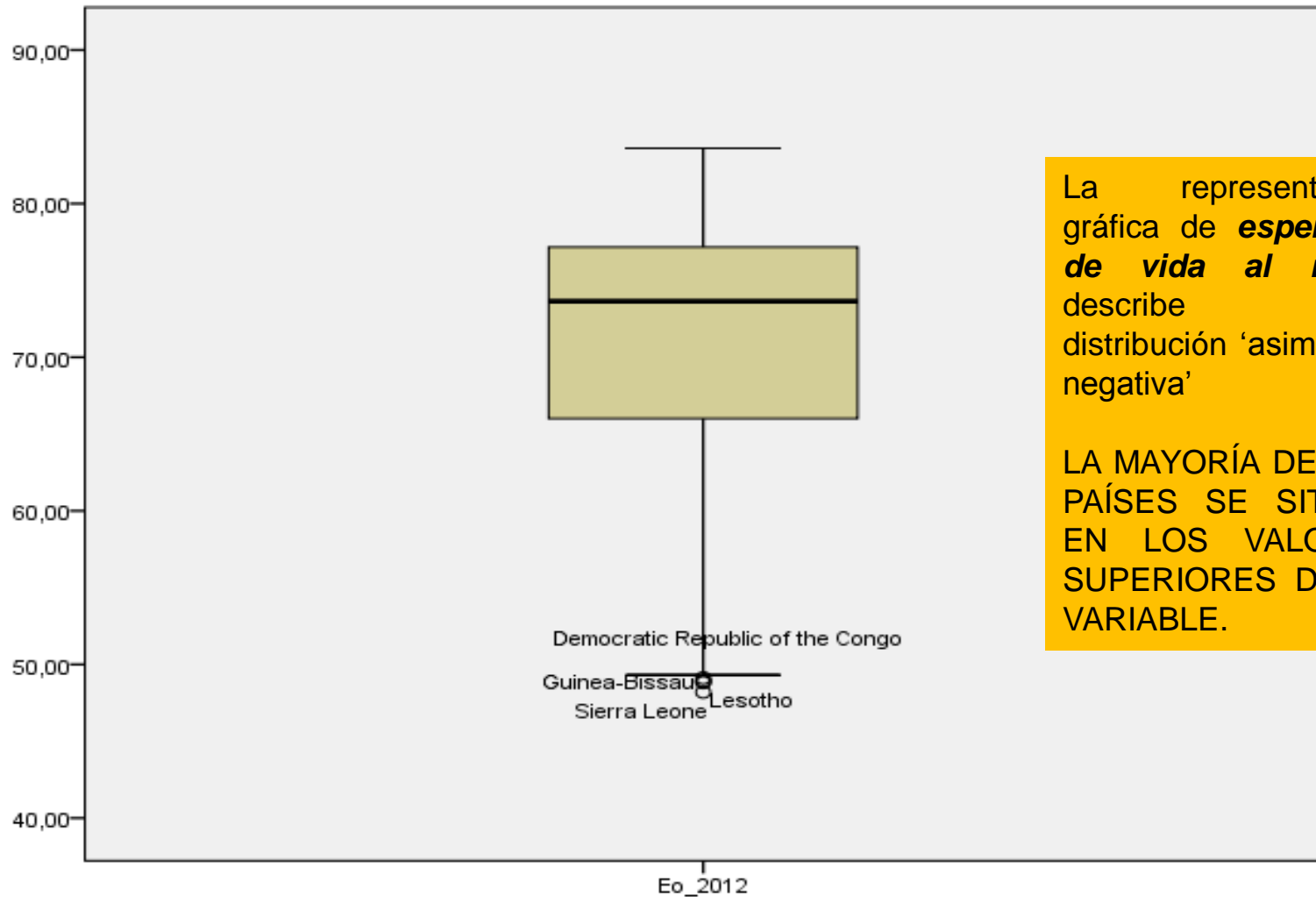


La posición de la mediana

La línea que divide la caja representa la **mediana** de la distribución y es un detector de simetría.

Si la línea se sitúa en el centro de la caja la distribución es **simétrica**. Si por el contrario se aproxima a los límites de la caja (Q_1 ó Q_3) tendremos un diagnóstico de **asimetría**, positiva o negativa respectivamente.

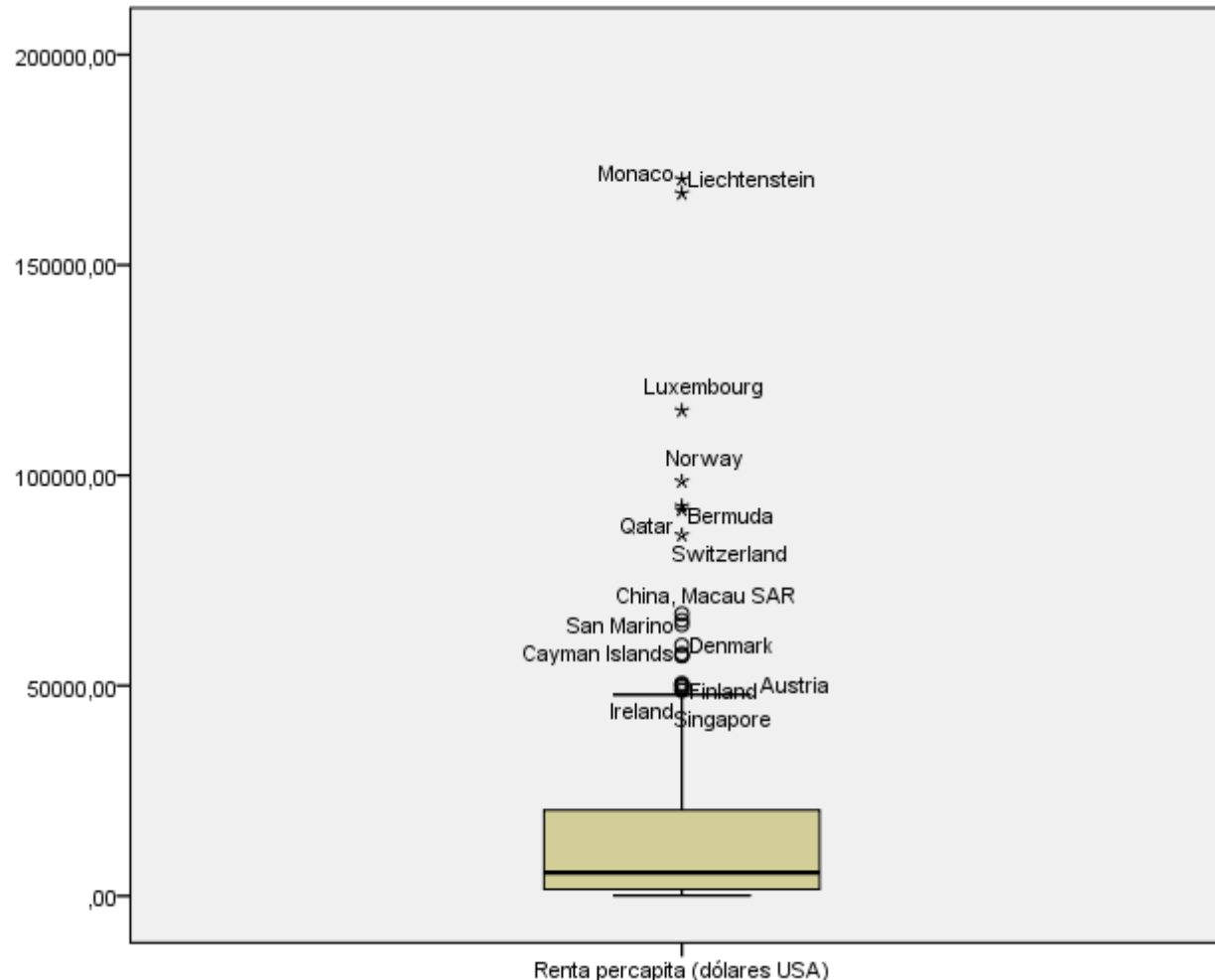
La posición de la mediana



La representación gráfica de **esperanza de vida al nacer** describe una distribución 'asimétrica negativa'

LA MAYORÍA DE LOS PAÍSES SE SITÚAN EN LOS VALORES SUPERIORES DE LA VARIABLE.

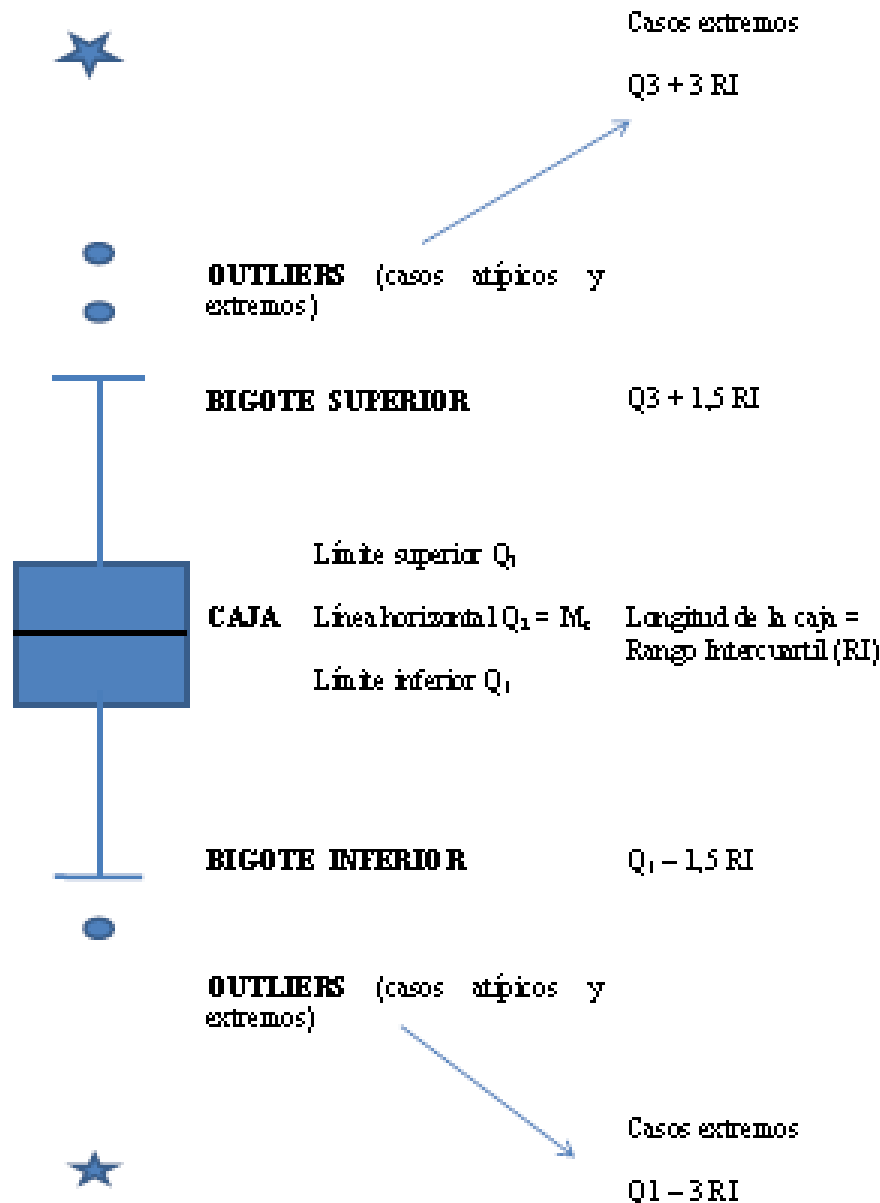
La posición de la mediana



La representación gráfica de **renta per capita** describe una distribución 'asimétrica positiva'.

LA MAYORÍA DE LOS PAÍSES SE SITUAN EN LOS VALORES INFERIORES DE LA VARIABLE.

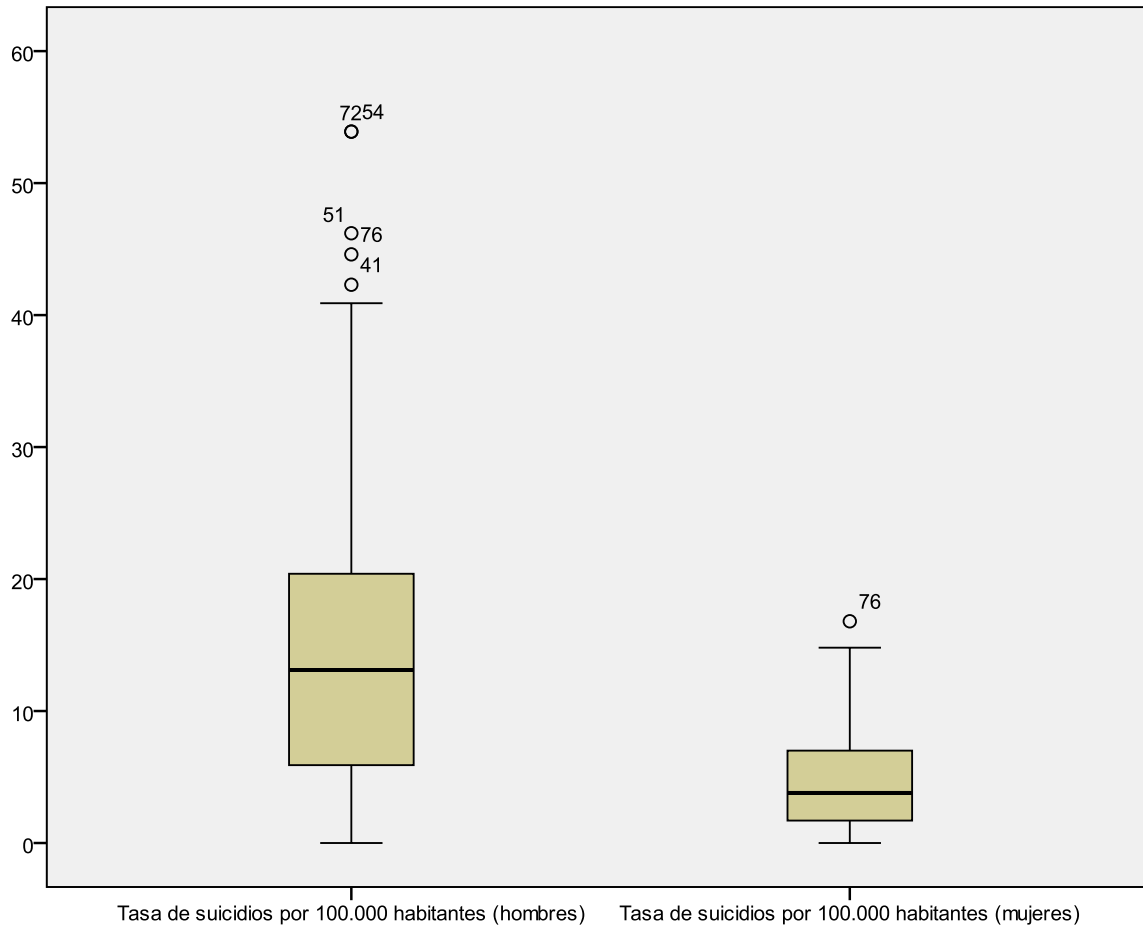
Los bigotes



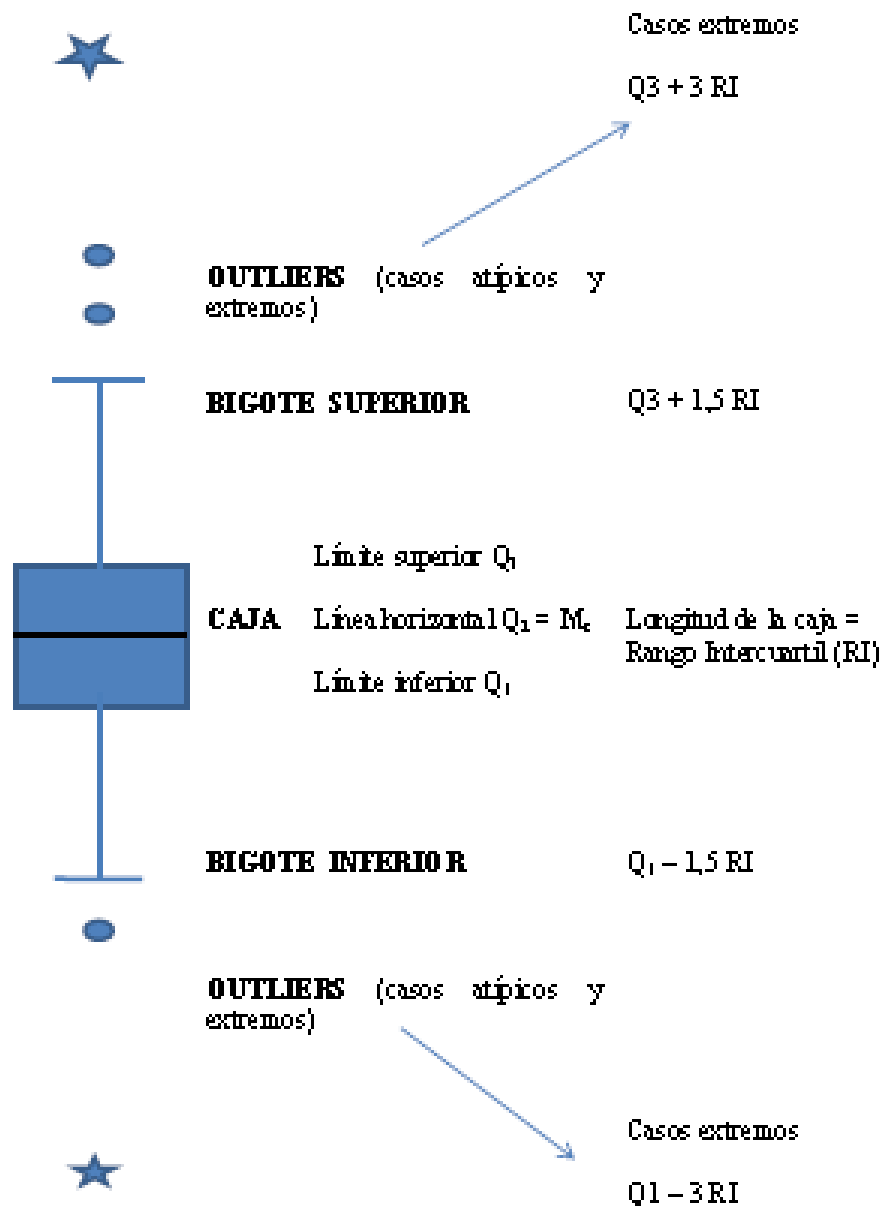
De la caja se prolongan dos segmentos (superior e inferior) denominados por Tukey **bigotes** –whisker- que señalan el límite para la detección de valores atípicos.

La longitud de los bigotes expresa la variabilidad de la distribución, en el 25 por ciento de los valores bajos (por debajo de Q_1) y altos (por encima de Q_3)

Longitud de los bigotes



La *tasa bruta de suicidio* entre hombres y mujeres muestra la diferencial variabilidad de las distribuciones, por encima y por debajo de los valores centrales..



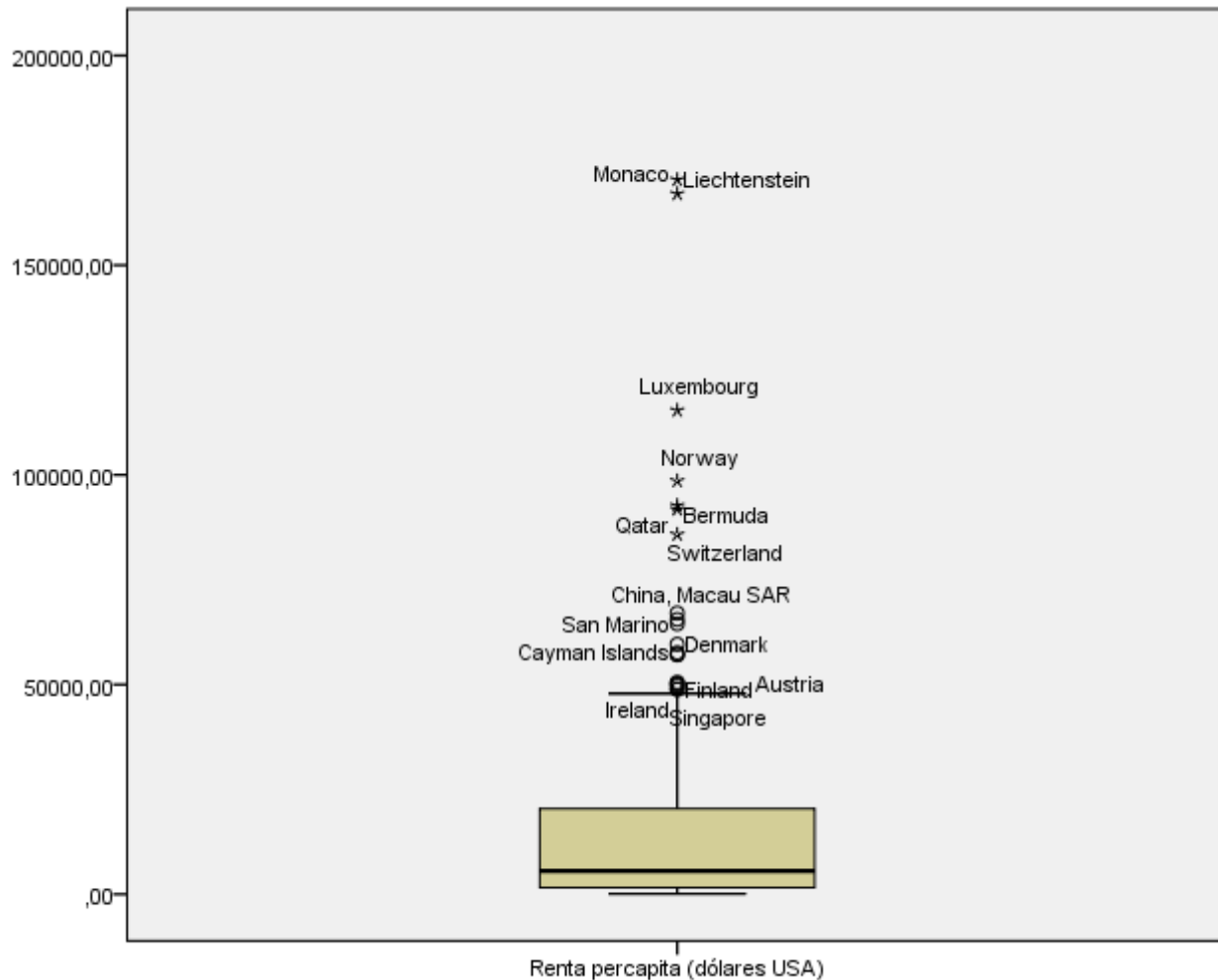
Más allá de los bigotes «los valores atípicos»

Los valores que se alejan más de lo esperado del centro de la distribución (RI) son clasificados como ATÍPICOS y EXTREMOS.

Se considera que un **valor** es **atípico** cuando excede 1,5 unidades de longitud de caja su distancia de los percentiles 25 (Q_1) y 75 (Q_3).

Se considera que un **valor** es **extremo** cuando excede 3 unidades de longitud de caja su distancia de los percentiles 25 (Q_1) y 75 (Q_3).

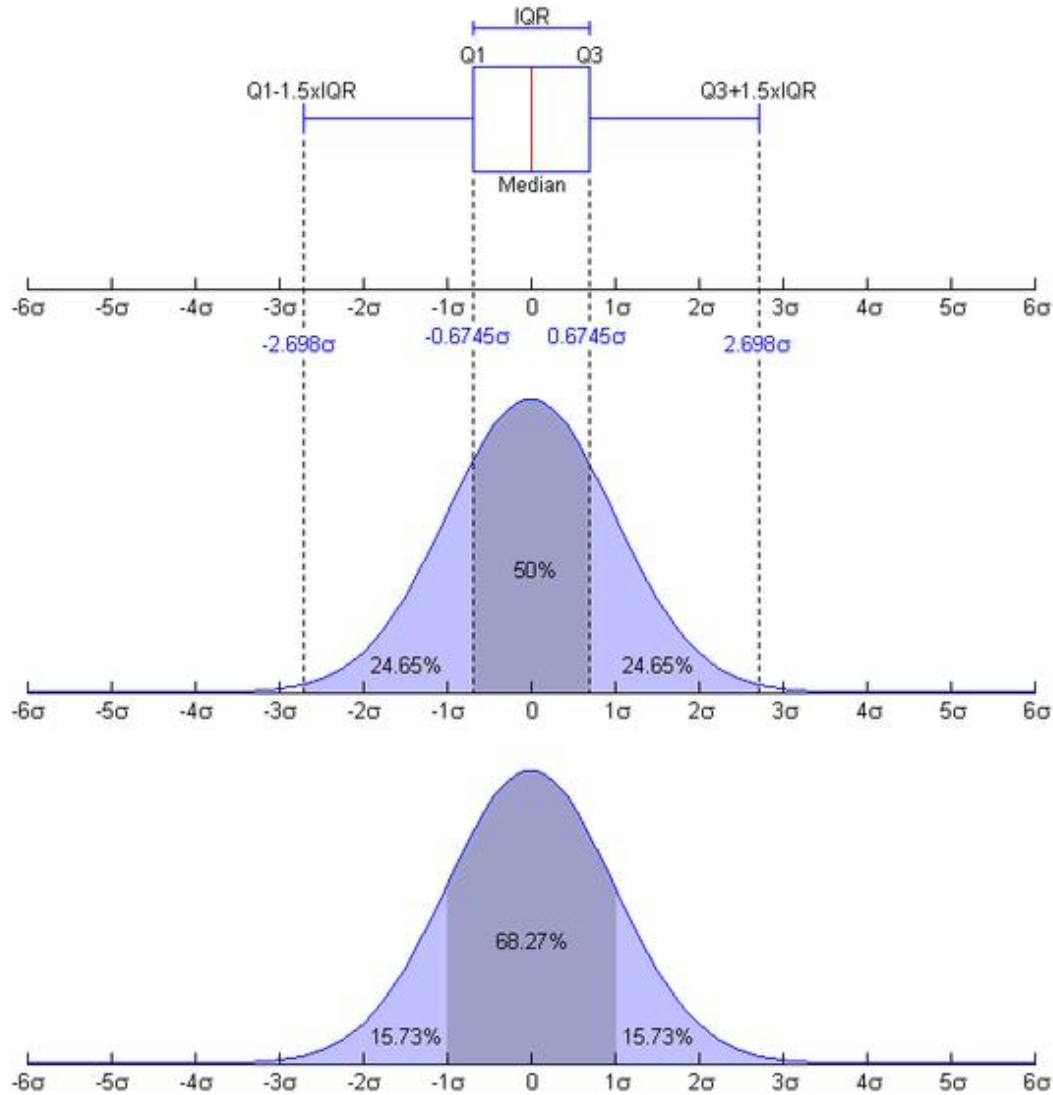
Valores atípicos y extremos



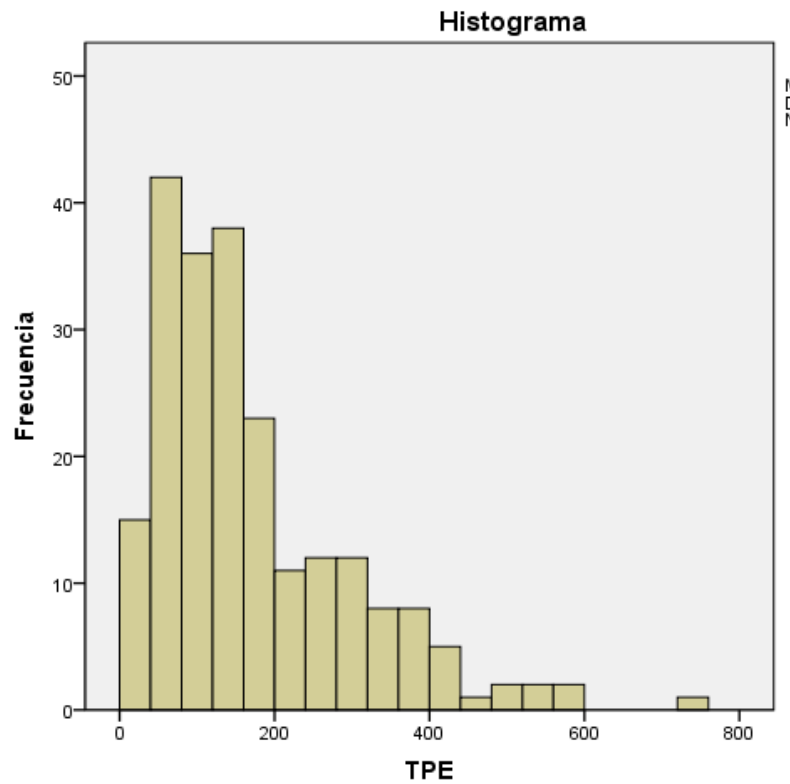
Los países identificados como **valores extremos** (rentas que se alejan en más de 3 unidades de longitud de caja del tercer cuartil) se representan con una estrella: Mónaco, Liechtenstein, Luxemburgo, Noruega, etc.

Los países identificados como **valores atípicos** (rentas que se alejan en más de 1,5 unidades de longitud de caja del tercer cuartil) se representan mediante un círculo: San Marino, Dinamarca, Finlandia, Austria, etc.

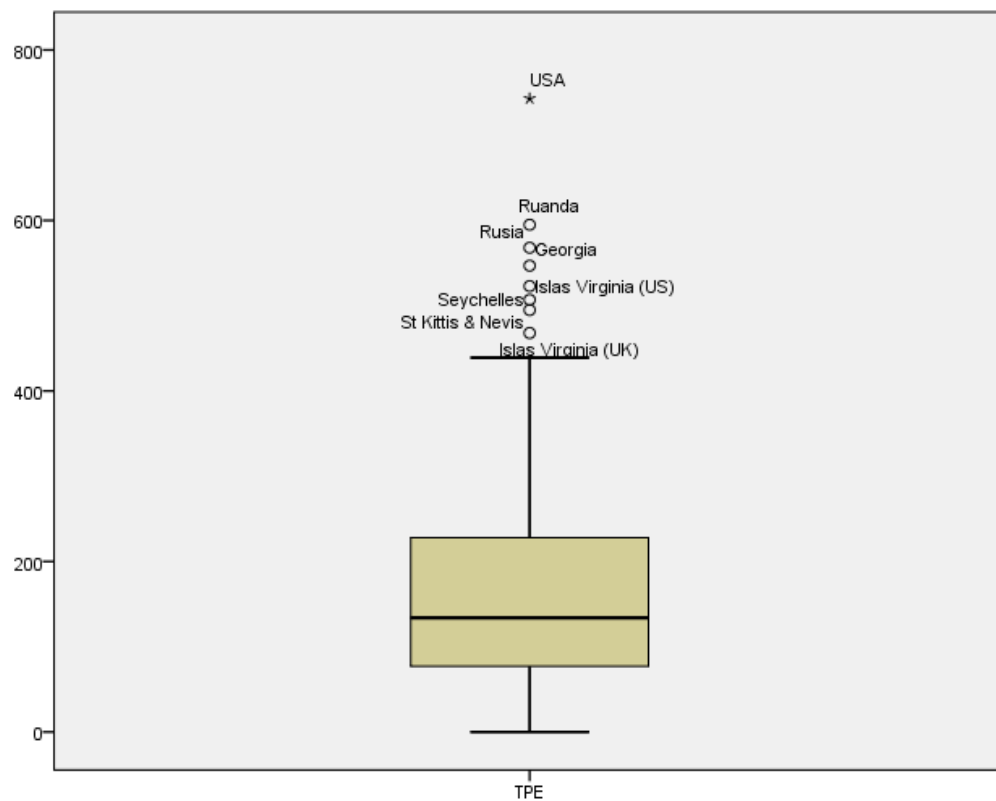
Correspondencia gráfico de caja y la curva normal



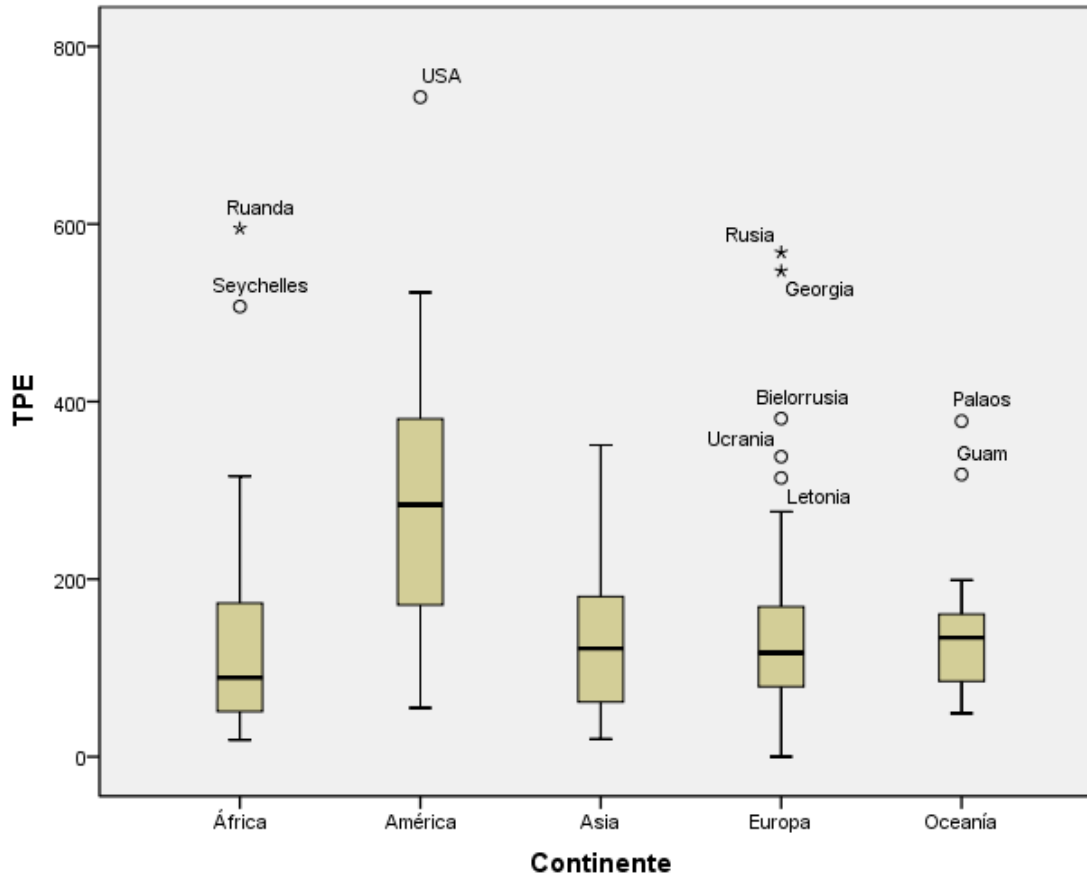
Fuente: “Box plot”, Wikipedia, Boxplot and a probability density function of a Normal $N(0,1\sigma^2)$ Population, disponible en línea https://en.wikipedia.org/wiki/Box_plot#/media/File:Boxplot_vs_PDF.svg



Histograma y gráfico de caja



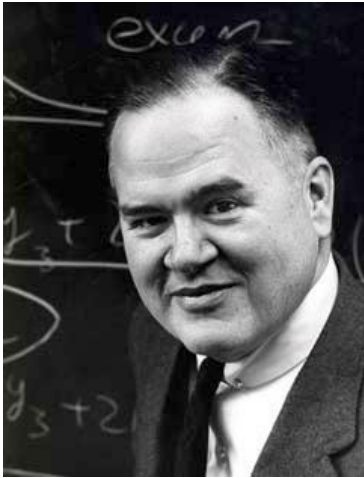
Perspectiva comparada



El gráfico de caja permite comparar, EN EL MISMO GRÁFICO, distintos subconjuntos referidos a la misma variable.

En este ejemplo, vemos las distribuciones de la tasa de población encarcelada, según continente.

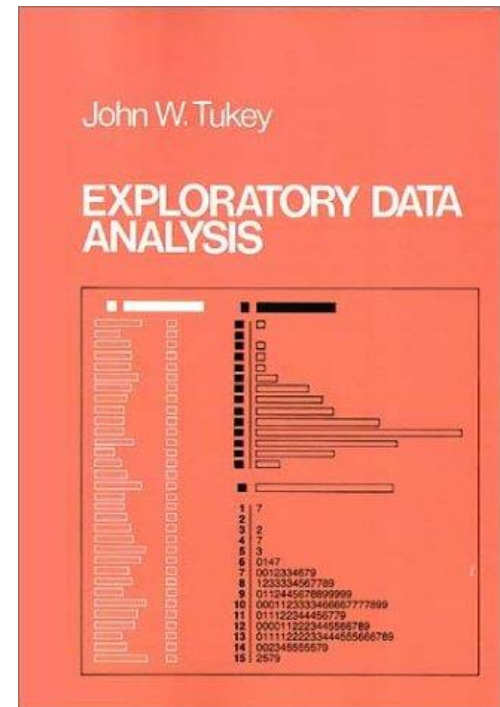
Cultura estadística



El gráfico de caja fue propuesto por John Wilder Tukey (1915-2000).

Tukey fue un polifacético científico: matemático, químico y estadístico. En 1977 publicó el libro «*Exploratory Data Analysis*» (popularizado por su acrónimo EDA) en el cual propone una nueva didáctica de la enseñanza de la estadística descriptiva mediante innovadoras herramientas. Para él, la representación gráfica y sintética de los datos era un inestimable potencial:

Exploratory data analysis is detective work- numerical detective work- or counting detective work- -or graphical detective work. A detective investigating a crime needs both tools and understanding (Tukey, 1977: 1)



Referencias y recursos de información para trabajar el ‘gráfico de caja’

BALLESTEROS DONCEL, Esmeralda (2013): “La estadística descriptiva como herramienta de análisis en la investigación social. Ejemplo de «actividad práctica» adaptada a la enseñanza del Espacio Europeo de Educación Superior”, *Números. Revista de Didáctica de las Matemáticas*, núm. 84, págs. 85-98, disponible en línea <http://eprints.ucm.es/23413/>

BRILLINGER, David R. (2011): “Data Analysis Exploratory”, *International Encyclopedia of Political Science*, págs. 530-537, SAGE.

LE GUEN, Monique (2002): “La boîte à moustaches de Tukey, un outil pour initier à l’Statistique”, *Bulletin de méthodologie sociologique*, núm. 73, p. 43-64. Recuperado el 7 de julio de 2012 <http://matisse.univ-paris1.fr/leguen/leguen2001b.pdf>

TUKEY, John W. (1977): *Exploratory Data Analysis*, London, Addison-Wesley (capítulos 1 y 2, págs. 1-56).

Actividad aplicada

«El mundo al revés: ¡Con voto pero sin voz!»



Inter-Parliamentary Union

For democracy. For everyone.

Participación de las mujeres en las Asambleas
Nacionales

Fuente de los datos estadísticos



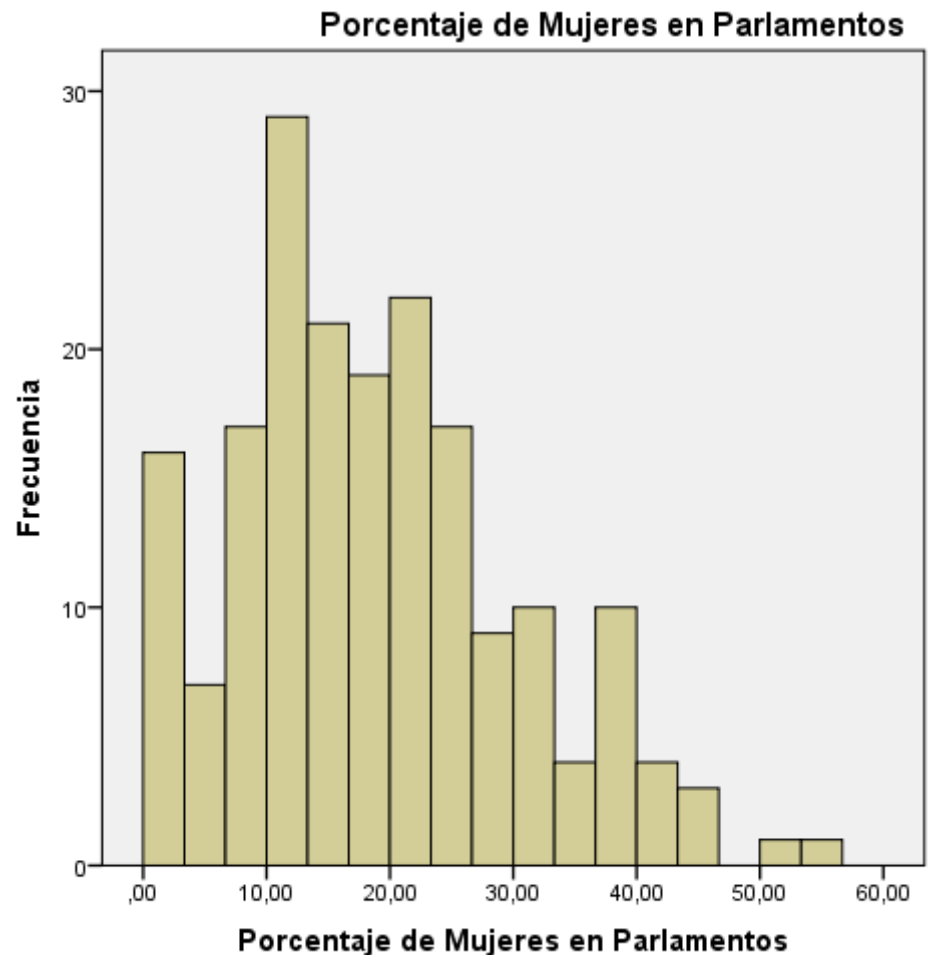
Desde 1997, la *Unión Interparlamentaria* divulga anualmente, a través de su website (ipu.org), información sobre la participación de las mujeres en los parlamentos nacionales de distintos países del mundo.

Los datos están accesibles y pueden descargarse libremente, para su análisis con software informático.

En esta actividad, la distribución está referida a información actualizada a 1/01/2013.

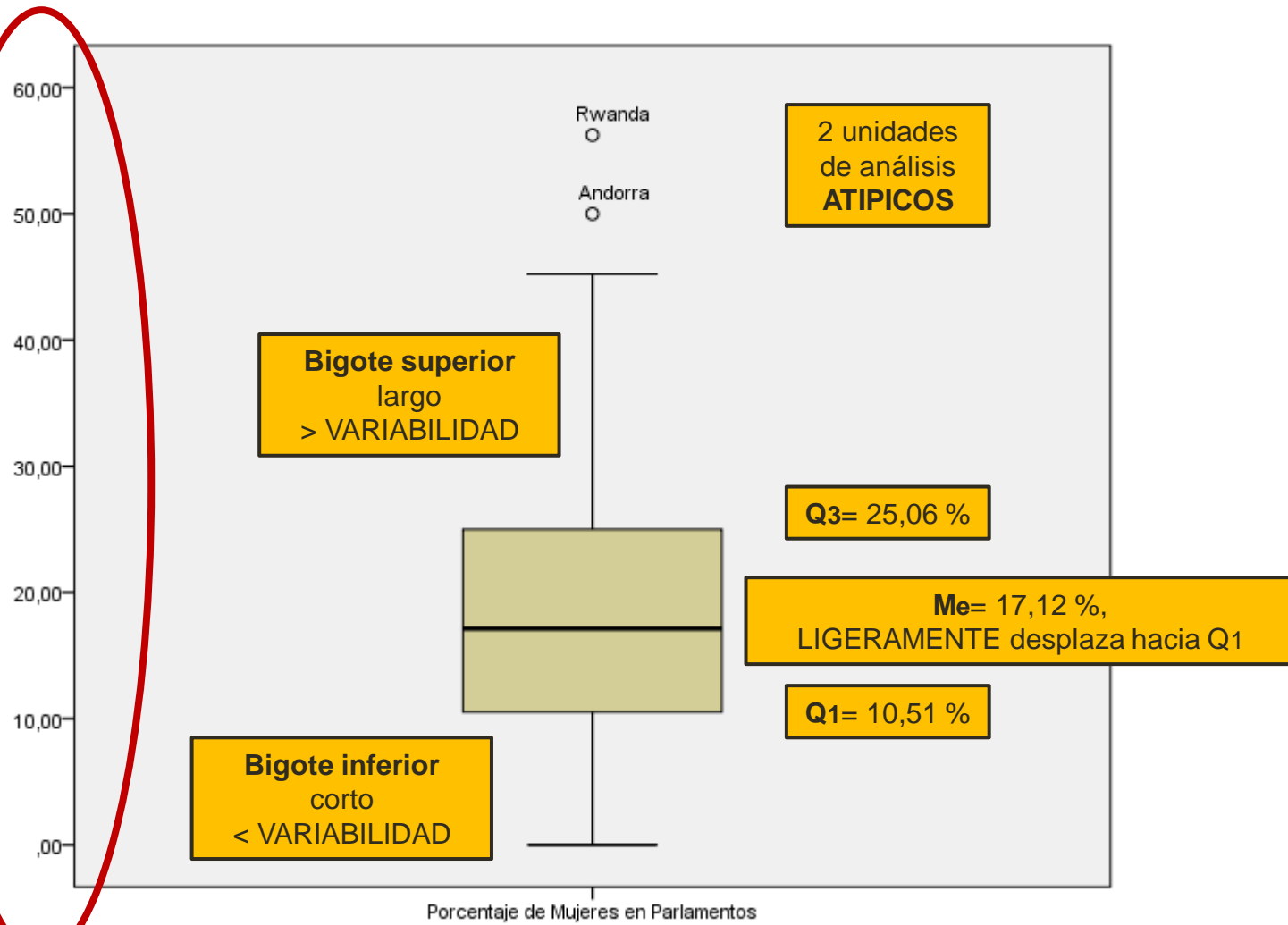
Recursos para el análisis de la distribución

| <i>Medidas estadísticas</i> | | % Mujeres en los parlamentos |
|-----------------------------|----|------------------------------|
| N | | 190 |
| Tendencia central | | |
| Media | | 18,92 |
| Mediana | | 17,12 |
| Dispersión | | |
| Desviación típica | | 11,48 |
| Varianza | | 131,71 |
| Coeficiente de variación | | 60,68 % |
| Rango | | 56,25 |
| Mínimo | | 0,00 |
| Máximo | | 56,25 |
| Forma | | |
| Asimetría | | 0,599 |
| Curtosis | | -0,014 |
| Posición no central | | |
| Percentiles | 25 | 10,51 |
| | 50 | 17,12 |
| | 75 | 25,06 |



El gráfico de caja y sus componentes

El **eje vertical** muestra la escala de valores de la variable que, en este caso se expresa como un porcentaje que oscila entre 0,0 (mínimo) y 56,3 (máximo).

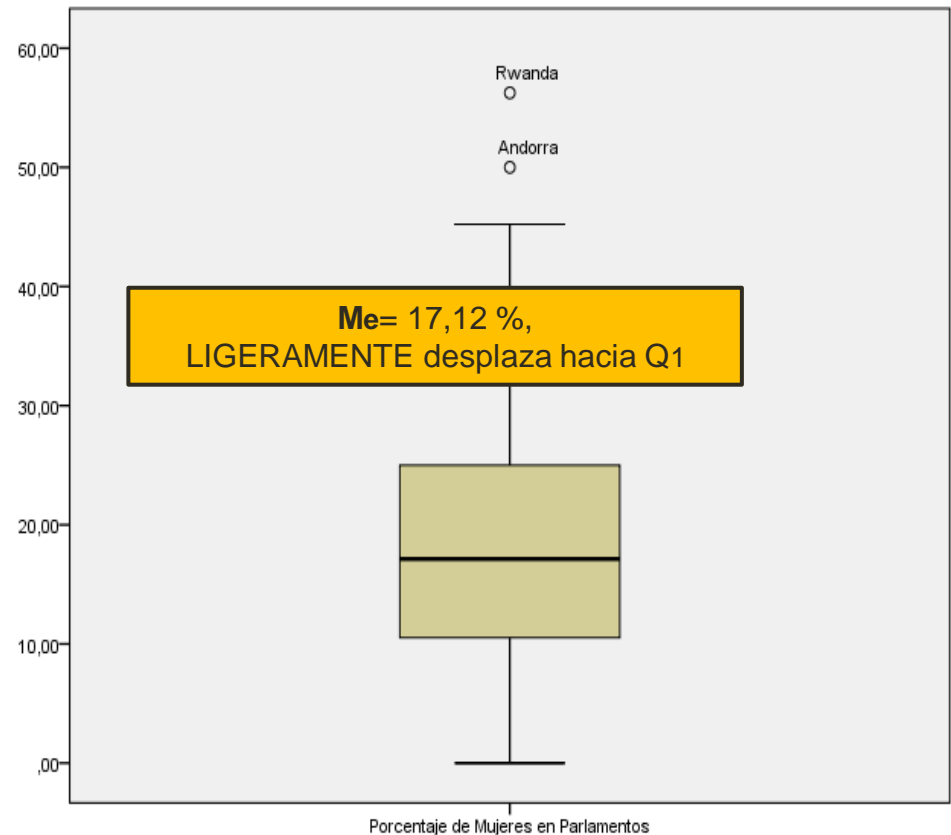


Estadísticos descriptivos

Tendencia central

El **promedio** de mujeres en los parlamentos es de tan sólo 18,9 por ciento.

La distribución está afectada por un diagnóstico de asimetría positiva, por lo que parece más adecuado referirse a la **mediana** que, rebaja el centro de gravedad en dos puntos porcentuales la presencia femenina en las asambleas nacionales (17,12 %).

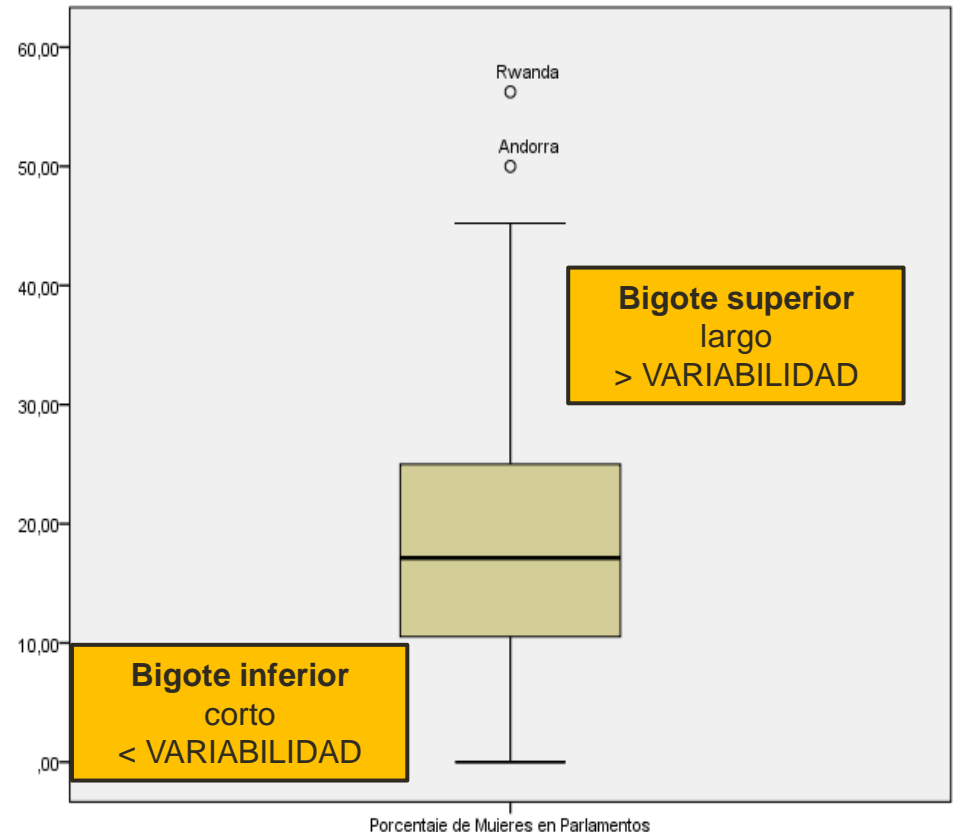


Estadísticos descriptivos

Dispersión

Desde el punto de vista de la variabilidad, el promedio de alejarse de la media aritmética es 11,48 puntos (desviación típica) y, por tanto, la distribución presenta una gran heterogeneidad (el coeficiente de variación -cociente entre S_x y promedio- alcanza un 60,62 %).

Ello implica que la participación de las mujeres en los parlamentos describe situaciones muy dispares que oscilan entre ninguna mujer en la Asamblea Nacional (0,0 % = mínimo, Arabia Saudí) a un máximo de 56,25 por ciento (Ruanda).

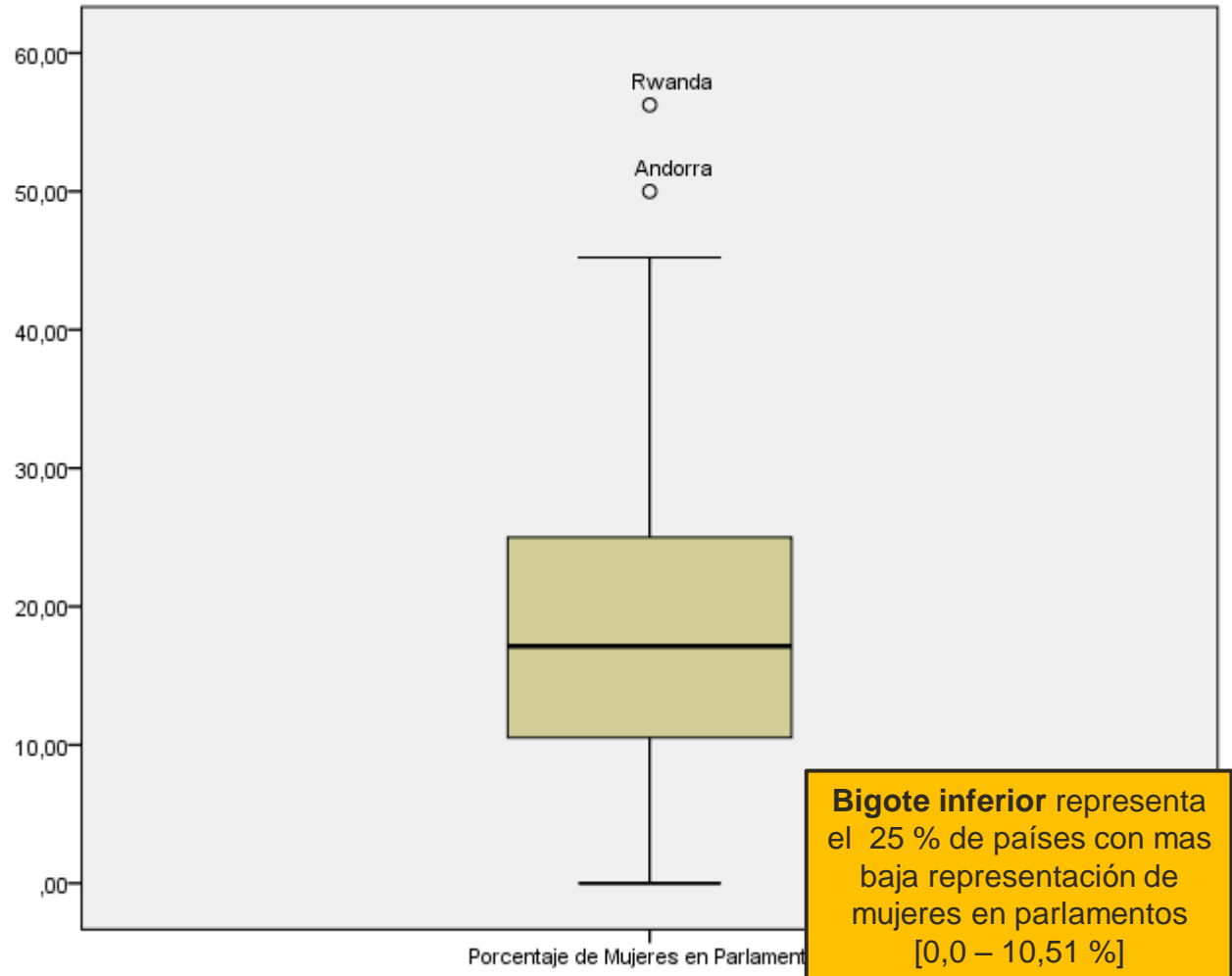


Estadísticos descriptivos

Posición no central

Este desolador diagnóstico encubre realidades muy diferentes entre países. El 25 por ciento de los países con más baja representación 0,0 - 10,5 %.

Asombra comprobar que Japón, uno de los países con mayor desarrollo económico y tecnológico del mundo, se sitúe entre los territorios más desiguales (7,92 %).

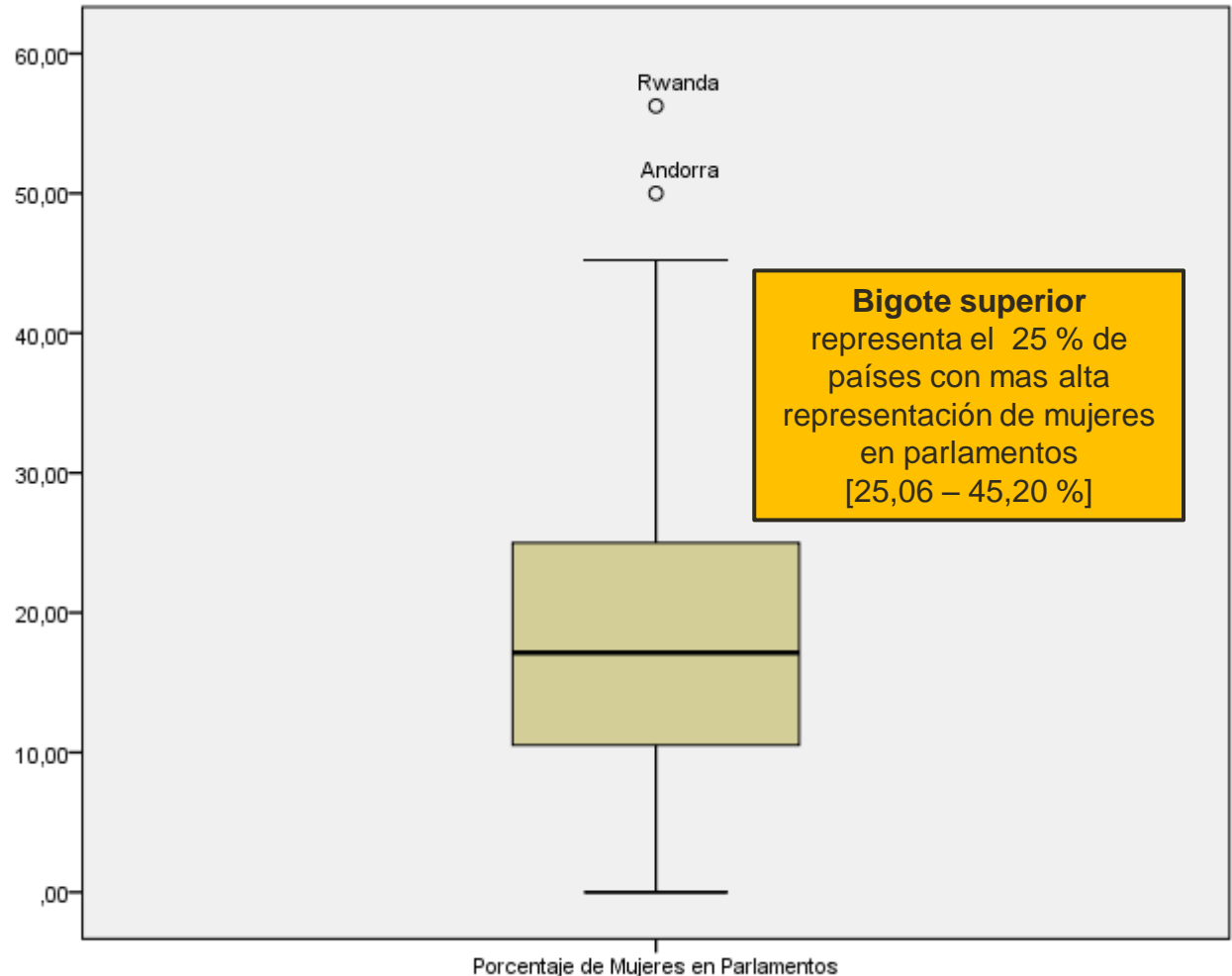


Estadísticos descriptivos

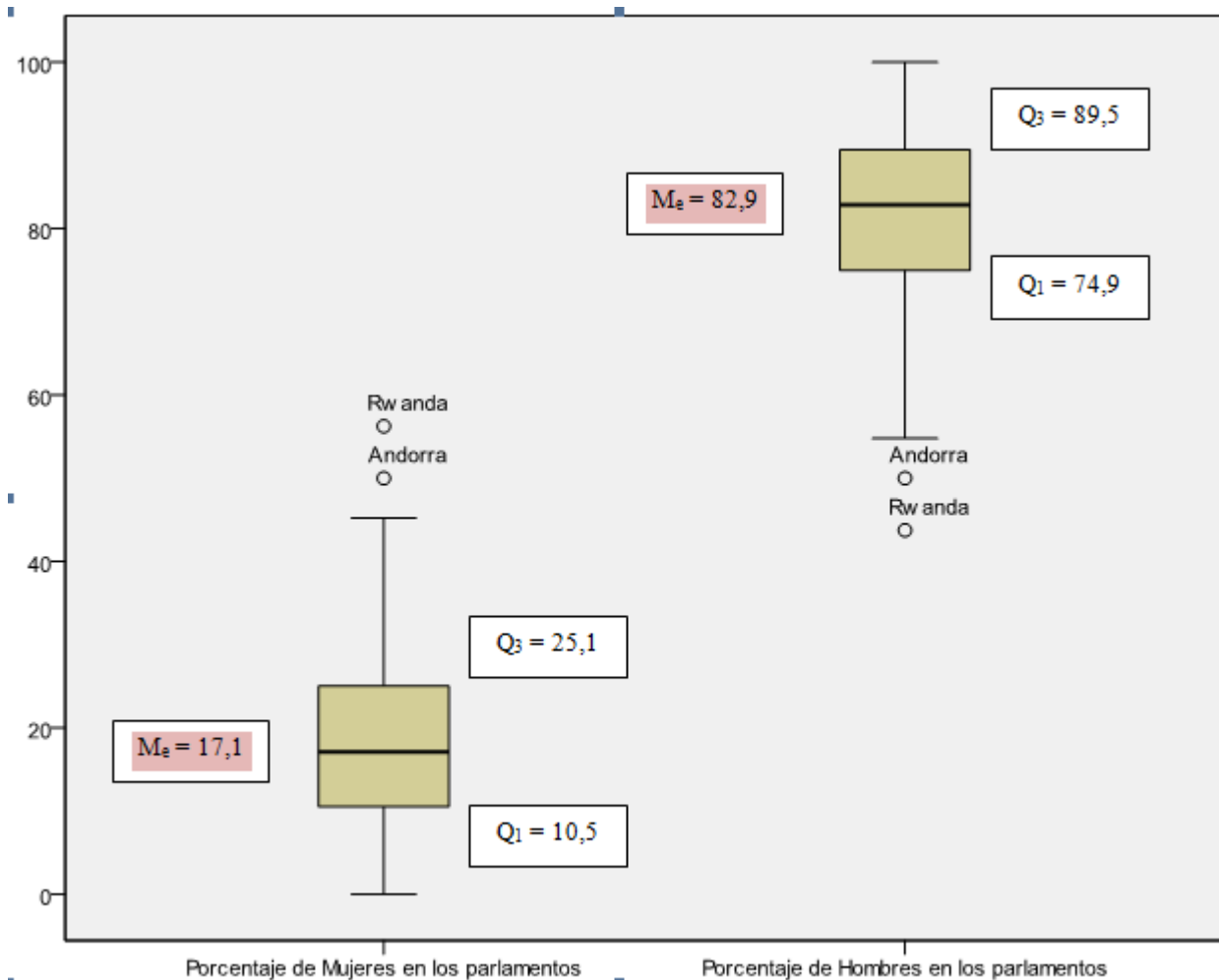
Posición no central

El 25 por ciento de los países con mayor representación 25,06 – 45,20 %.

Sorprende que sea Ruanda el territorio con mayor porcentaje de diputadas (56,25 %), por delante de los países nórdicos - Suecia, Finlandia, Islandia y Noruega-, vanguardia histórica de la participación parlamentaria de las mujeres.



Perspectiva comparada, según sexo



Si establecemos la comparación, según sexo, apreciamos un evidente diagnóstico de desigualdad.

Mientras la mediana de la representación parlamentaria de varones se sitúa en 82,9 %, la de las mujeres sólo alcanza el 17,1 %

La estadística como herramienta de análisis en la investigación social

«Reflexiones»



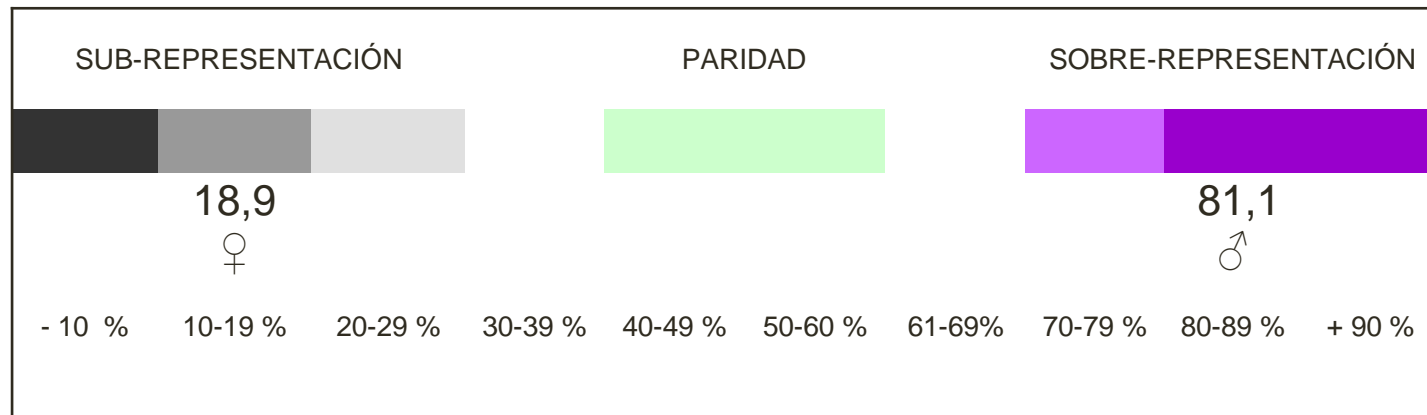
Inter-Parliamentary Union

For democracy. For everyone.

Participación de las mujeres en las Asambleas
Nacionales

Análisis socio-político ¿Qué implicaciones tiene?

El **promedio** de mujeres en las asambleas nacionales es de tan sólo 18,9 por ciento (frente al 81,1 % de los hombres), pese a que, al menos, el 50 por ciento del electorado es femenino.



Transcurridas dos décadas de la *Declaración de Atenas* (1992) las estadísticas confirman que, desde la perspectiva de género, se mantiene un déficit democrático en la práctica totalidad de los países en el mundo. Es decir, la mujeres están muy lejos de la paridad en la representación y en la administración de las naciones.

Análisis socio-político

El análisis de los datos estadísticos divulgados por la *Unión Interparlamentaria* permite identificar dos modelos en la incorporación de mujeres a los parlamentos:

- a) 'Vía lenta', resultado de una dilatada conquista histórica (modelo nórdico)
- b) 'Vía rápida', desarrollada por democracias jóvenes, que han constituido parlamentos tras un conflicto armado, estableciendo cuotas de presencia paritarias (Ruanda, Afganistán, etc.)

La cuestión es preguntar(se) o, preguntar(nos)

¿Cuánto tiempo tienen que esperar las mujeres para romper el dominio masculino en la política?

La especialista Drude Dahlerup (2012) no cree que el modelo nórdico deba ser la principal referencia, ni tan siquiera la estrategia a seguir. Aunque también admite que, la vía rápida no está exenta de controversia, pues forzar legalmente la paridad no garantiza la equidad entre los sexos.

Evolución 1945 -2013: un lento proceso

| Año | Núm. países | Porcentaje de mujeres en las Asambleas Nacionales |
|------|-------------|---|
| 1945 | 26 | 3,0 |
| 1955 | 61 | 7,5 |
| 1965 | 94 | 8,1 |
| 1975 | 115 | 10,9 |
| 1985 | 136 | 12,0 |
| 1995 | 176 | 11,9 |
| 2007 | 185 | 17,9 |
| 2013 | 190 | 18,9 |

Fuente: IPU, Women in politics

Resolución UN



La resolución A/RES/66/130 de Naciones Unidas (2011) insta a todos los Estados a que adopten medidas encaminadas a garantizar la participación paritaria de las mujeres en órganos de decisión política de carácter local, nacional e internacional, es decir, que puedan incorporarse con voz y voto a los núcleos de decisión económico, político y social. Puede que, quizá entonces, surjan nuevas sinergias y tendencias en el desarrollo de las sociedades contemporáneas.

Gracias por vuestra atención

«Esperamos que este video-tutorial os resulte de utilidad»

Anexos

LOS DATOS

Women in National Parliaments (01/1/2013)

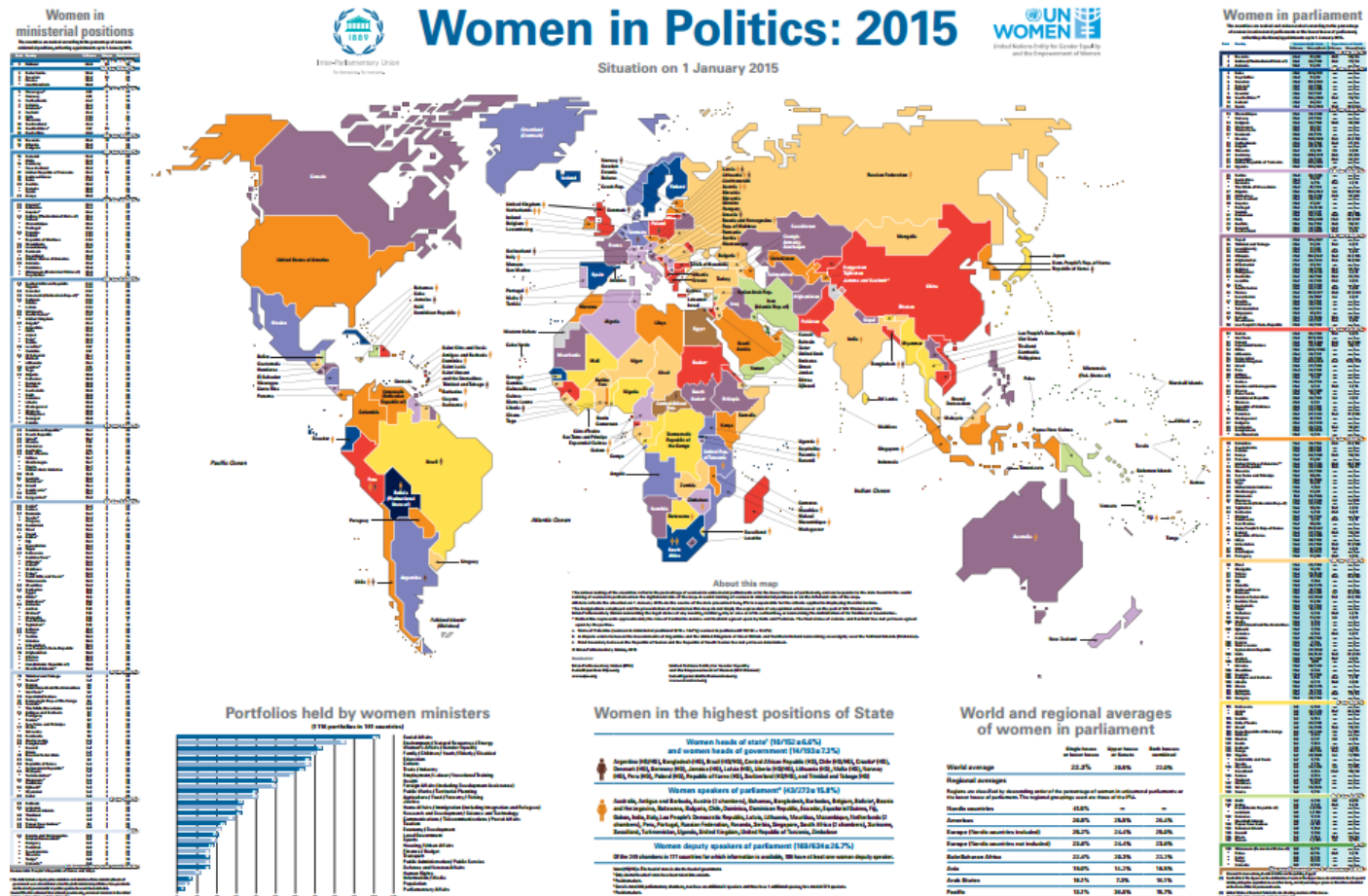
En el material didáctico complementario se facilita el fichero de datos empleado en la aplicación práctica en formato 'excel' (Women in Parliaments_DATOS)

RECORDAMOS que desde la página web de la IPU (<http://www.ipu.org>) se pueden descargar todas las series anuales, desde la sección 'PARLINE database on national parliaments'



Inter-Parliamentary Union

For democracy. For everyone.



Referencias y recursos de información

Declaración de Atenas/Athens Declaration (1992), disponible en línea:
<http://eurogender.eige.europa.eu/news/athens-declaration-1992>

DAHLEUP, Drude (2012): “On a fast track to gender balance in politics”, Women United, núm, 3, pp. 13-14, disponible en línea:
<http://www.fokuskvinner.no/Nyheter/2013/On-a-fast-track-to-gender-balance-in-politics/>

INTERNATIONAL ORGANIZATION OF PARLIAMENTS
(<http://www.ipu.org>)

UN, Resolución 66/130: La participación de la mujer en política/
Resolution adopted by the General Assembly (A/RES/66/130): Women
and political participation, disponible en línea:
<http://www.un.org/womenwatch/daw/documents/ga66.htm>